ML 2015, Ch. 76. Art. 2, Sec. 6a Project Abstract

PROJECT TITLE: Sub-project #1: Garlic Mustard Biocontrol: Ecological Host Range of Biocontrol Agents
PROJECT MANAGER: Roger Becker
AFFILIATION: Department of Agronomy and Genetics, University of Minnesota
MAILING ADDRESS: 411 Borlaug Hall, 1991 Upper Buford Circle
CITY/STATE/ZIP: St Paul, MN 55108
PHONE: (612) 625-5753
E-MAIL: becke003@umn.edu
WEBSITE: http://appliedweeds.cfans.umn.edu/personnel

FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: ML 2015, Ch. 76. Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $570,173
AMOUNT SPENT: $570,173
AMOUNT REMAINING: $ 0

Sound bite of Project Outcomes and Results

We were integral in the release of *Ceutorhynchus scrobicollis* in Canada, the first biological control agent for garlic mustard in North America. We moved closer to federal regulatory approval to release *C. scrobicollis* and *C. constrictus* in the United States. When achieved, these will offer the first viable control of garlic mustard in Minnesota woodlands.

Overall Project Outcome and Results

Garlic mustard poses significant threats to our forest ecosystem. Research supported by this grant develops effective biological control of garlic mustard in Minnesota, the United States, and Canada, offering the first viable control option for this troublesome invasive plant. We gained a recommendation that *Ceutorhynchus scrobicollis* be considered for a release in the U.S. from the APHIS PPQ Technical Advisory Group. In follow-up consultation between USDA-APHIS-PPQ and USFWS, questions were generated that were intended to expedite writing the Biological Assessment for *C. scrobicollis*. Funding from this grant enabled us to address those questions with specific research on three federally listed species. COVID-19 altered our timeline, yet we will be submitting the third edition of the response in August 2021. This funding supported Entomology PhD candidate Mary Marek-Spartz analyze predictive tools used to determine the expected range of biological control insects introduced to a new region, define specific biological thresholds of *C. scrobicollis*, and develop a novel biennial stage-structured plant-herbivore population model. She improved the accuracy of this model through data generated in our monitoring efforts funded from this grant. Also supported on this grant, Project Scientist Dr. Katovich further defined the vernalization requirements for a garlic mustard which will greatly improve the accuracy of the projected range of garlic mustard in the US, a key factor in determining the risk of introducing specific biological control insects to North America. Additionally, she completed host specificity testing for *C. scrobicollis* and made significant progress towards completing the registration package for *C. constrictus*. We have a draft of the petition for the release of *C. constrictus* for biological control of garlic mustard. Due to technical difficulties in rearing threatened and endangered species out of their normal habitats, we will complete the few species needed at CABI, Delémont CH.

Project Results Use and Dissemination Knowledge gains have been distributed widely through professional and land manager meetings. Additionally, we presented our findings to our colleagues at the triennial International Symposium on the Biological Control of Weeds, hosted in 2018 by our cooperators from CABI, CH. Generations.py is a software program publicly available with a novel biennial component enabling modelers to improve predictions of the dynamics and biology of biennial organisms. We played a key role in the first release of a biological control insect for garlic mustard in North America. Additionally, four to six papers will be published in professional journals. A petition for the release of *C. constrictus* will be submitted to USDA APHIS PPQ TAG this fall or early next spring.

ML 2015, Ch 76 Art 2, Sec 6a Project Abstract
For the Period Ending December 31, 2021

PROJECT TITLE: **Subproject #2: Mountain pine beetle, Phase III: Protecting Minnesota**
PROJECT MANAGER: **Brian Aukema**
AFFILIATION: **University of Minnesota**
MAILING ADDRESS: **1980 Folwell Avenue**
CITY/STATE/ZIP: **St. Paul, MN 55118**
PHONE: **(612) 624-1847**
E-MAIL: **BrianAukema@umn.edu**
WEBSITE:
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER
ML 2015, Ch. 76, Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $445,347
AMOUNT SPENT: $445,347
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results

Repeated surveys did not find mountain pine beetle in Minnesota. Scant few individuals were captured dispersing far from active infestations in western states. We found that local bark beetles and predators do not optimally recognize the insect's chemical signals, however, suggesting that such components of invasion resistance might be low.

Overall Project Outcome and Results

Surveys over the course of this project did not detect any mountain pine beetle in Minnesota. Although absence data cannot rule out inappropriate lure choices, testing of a new lure within the Black Hills of South Dakota where mountain pine beetle is endemic found that the conventional lure worked well. No improvements were noted when testing a new formulation. Long distance dispersal transects revealed that mountain pine beetles can be captured up to 30 miles away from active tree-killing outbreaks, but these singletons represented a fraction of a fraction of the population. Dispersal pressure was much lower in the last year of the project when beetles returned to endemic levels, which is the norm in western forests for decades at a time. Thus, we expect that the risk of mountain pine beetle reaching Minnesota by blowing from infestations in the Black Hills of South Dakota, which is approximately 500 miles away from the nearest mature pine forests in Minnesota, is extremely low. If mountain pine beetle was to arrive in Minnesota, it would have to establish into an environment with new flora (species of pines) and fauna (other species of bark beetles as well as their predators) to which it had never been exposed. The only species of pine common to the Black Hills and Minnesota is Scots pine; exposures to the fungus that mountain pine beetle carries revealed strong localized responses of Scots pine to the inoculation sites with defensive chemicals known as monoterpenes. Surveys of Minnesota's community of bark beetles, competitors, and predators responding to lures of mountain pine beetle in comparison to similar in the Black Hills revealed nuanced, regional variations in responses, but overall strong fidelity to cures of predators associated with local prey. Thus, we expect that predators or competitors in Minnesota would not optimally recognize the aggregation pheromone of mountain pine beetle. In one case with direct comparative tests in the Black Hills, we noted that one of the most common bark beetles that would potentially compete with mountain pine beetle in Minnesota, *Ips grandicollis*, avoids the lure of mountain pine beetle. We did note a few mountain pine beetles in traps baited with the aggregation pheromone of *Ips grandicollis* when the traps were placed far from active infestations of mountain pine beetle. This finding suggests that mountain pine beetle could respond to such pheromones as a "last-ditch" effort to find habitat during endemic periods where there are insufficient numbers to mass-attack, colonize, and kill large trees. If true, mountain pine beetle could find an endemic niche in Minnesota's pine forests. Because we still lack knowledge about how mountain pine beetles persist in endemic states, and whether colonization densities might actually be lower in other species of Minnesota's

pines if they have lower defensive responses, continued vigilance against mountain pine beetle as a threat to Minnesota's pine forests is warranted.

Project Results Use and Dissemination

We have published one scientific paper from this work, with four more moving toward publication with peer-reviewed journal targets. We gave numerous regional, national, and even presentations as venues such as the Entomological Society of America, the IUFRO Conference on Biological Invasions in Forests, the North American Forest Insect Work Conference, North Central Forest Pest Workshop, Western Forest Insect Work Conference, Upper Midwest Invasive Species Conference, the Sustainable Forest Education Cooperative, State Forest Health Cooperators, Northern Advanced Silviculture Program, Minnesota Forest Industries, and MN Department of Natural Resources Forestry Team.

Forest Ecology and Management 497 (2021) 119455

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Colonization and reproduction of potential competitors with mountain pine beetle in baited logs of a new host for mountain pine beetle, jack pine

Zach M. Smith [a,*], Kevin D. Chase [a,b], Etsuro Takagi [c], Aubree M. Kees [a], Brian H. Aukema [a]

[a] Department of Entomology, University of Minnesota, 1980 Folwell Avenue, Saint Paul, MN 55108 USA
[b] Bartlett Tree Research Lab, Bartlett Tree Experts, 13768 Hamilton Road, Charlotte, NC 28278 USA
[c] Department of Tourism Science, Tokyo Metropolitan University, Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

ARTICLE INFO

Keywords:
Dendroctonus ponderosae
Ips grandicollis
Monochamus spp.
Pinus banksiana
Semiochemicals
Pheromones

ABSTRACT

The mountain pine beetle (*Dendroctonus ponderosae* Hopkins) is a bark beetle that is native to pine forests of western North America and the Black Hills of South Dakota. Recent eastward range expansion into stands of jack pine (*Pinus banksiana*) and associated hybrids with lodgepole pine (*Pinus contorta*) in western Canada has created concern that the insect will continue moving eastward. In the Great Lakes region, mountain pine beetle would encounter novel species of pines and associated insect fauna; interactions with which are largely unexplored. We baited logs of jack pine with lures for mountain pine beetle and *Ips grandicollis* (Eichhoff) alone and in combination in a 2 × 2 factorial design in the Black Hills of South Dakota. Both insects occur in this region, but not jack pine, a common species in the Great Lakes region of North America at risk of invasion by mountain pine beetle. We measured attraction and reproduction of insects that colonized the logs. *Ips grandicollis* were significantly more attracted to logs of jack pine baited with their aggregation pheromone, ipsenol, than unbaited logs or those baited with pheromones of mountain pine beetle and myrcene, a host volatile. Colonization by I. *grandicollis* was inhibited by the presence of lures for mountain pine beetle. We also found larvae of longhorn borers, likely - *Monochamus* spp., infesting logs. These borers, which act as competitors and facultative predators of bark beetles, were significantly attracted to logs baited with ipsenol over those baited with lures for mountain pine beetle. Our results suggest that if mountain pine beetle were to invade the Great Lakes Region, common bark and woodboring species such as *I. grandicollis* and longhorn borers would not compete with mountain pine beetles at tree-colonizing stages, and thus could pose little resistance to invasion.

1. Introduction

In forested ecosystems, bark beetles (Coleoptera: Curculionidae) provide critical ecosystem services and impact carbon dynamics by promoting wood deterioration, nutrient cycling, and biodiversity (Wood, 1982; Kurz et al., 2008; Mikkelson et al., 2013; Beudert, et al., 2014). Most species of bark beetles are termed "secondary" as they are innocuous and infest dying or stressed trees, where they consume the phloem tissues (Wood, 1982; Lindgren and Raffa 2013). Conversely, a minority of bark beetle species are capable of killing mature, live trees at landscape scales. As natural components of disturbance regimes, these "primary" species can alter successional trajectories of biomes (Taylor and Carroll, 2003; Raffa, et al., 2008).

Several environmental requirements must be met for tree-killing species of bark beetles to reach outbreak levels. The first condition is favorable climate. Depending on the species, conditions may include landscape-scale drought that stress host trees or warm minimum winter temperatures that foster brood survival (Safranyik, 1978; Carroll et al., 2004; Klutsch et al., 2017). The second requirement is an abundant supply of susceptible host trees (Safranyik, 1978). Forestry practices such as overstocking or under thinning can increase the likelihood of bark beetle outbreaks (Fettig et al. 2007). Finally, population growth of outbreaking species must exceed that of predators, parasitoids and other subcortical woodborers that compete for resources and exert regulatory effects. Each of these guilds may depress the reproductive capabilities of primary bark beetles at endemic levels (Rankin and Borden, 1991; Lindgren and Raffa, 2013; Aukema et al., 2016).

Competition may be especially pronounced among bark beetles as an endophytic feeding guild (Lindgren and Raffa, 2013), frequently mediating population dynamics between primary and secondary species

* Corresponding author.
E-mail address: smit7048@morris.umn.edu (Z.M. Smith).

https://doi.org/10.1016/j.foreco.2021.119455
Received 3 March 2021; Received in revised form 8 June 2021; Accepted 10 June 2021
Available online 30 June 2021
0378-1127/© 2021 Elsevier B.V. All rights reserved.

(Rankin and Borden, 1991; Aukema et al., 2016). Often, several sympatric species of bark beetles simultaneously infest the same tree (Rankin and Borden, 1991; Byers, 1989a). The overlap in the subcortical layer elicits the need for resource partitioning (Byers, 1989a). Interspecific competition can be avoided in two ways: temporally (e.g., exploiting differences in phenology) and spatially (e.g., infesting different parts of the tree). Host procurement and resource partitioning is frequently mediated by semiochemicals. Aggregation pheromones attract conspecifics, while inhibitory allomones push heterospecifics toward other resources (Paine et al., 1981; Byers, 1989b; Ayres et al., 2001). *Ips pini* (Say), *I. paraconfusus* (Lanier), and *Dendroctonus brevicomis* (LeConte), for example, may simultaneously infest a ponderosa pine (*Pinus ponderosae*) but spatially separate themselves throughout the bole (Birch and Wood 1975; Paine et al., 1981; Byers, 1989a). Despite temporal and spatial partitioning strategies among species, however, a broad range of overlap can still exist (Paine et al., 1981; Byers, 1989a; Ayres et al, 2001).

The mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae), is a bark beetle native to North America west of the Rocky Mountains and the Black Hills of South Dakota and Wyoming, USA (Safranyik and Carroll, 2006). While mountain pine beetle is a generalist, feeding on most species of sympatric pines in these regions, its principal hosts have historically been ponderosa pine (*Pinus ponderosae* Dougl. ex Laws) and lodgepole pine (*Pinus contorta* Dougl. ex Loud) (Safranyik and Carroll, 2006; Raffa et al, 2008). Outbreaks of mountain pine beetle can lead to landscape-level mortality of mature pines over many thousands of hectares (Safranyik and Carroll, 2006; Bentz et al., 2010; Hicke et al., 2012). While populations of mountain pine beetle were historically restricted in their northern range due to climatic and geological barriers, ameliorating winter temperatures have permitted northward range shifts (Cudmore et al., 2010, Cullingham et al., 2011). In recent years, British Columbia, Canada has experienced the largest outbreak of *D. ponderosae* in recorded history (Aukema et al., 2006), expanding the insect's range over the geoclimatic barrier of the Rocky Mountains into stands of jack pine (*Pinus banksiana* Lamb.) and its hybrids with lodgepole pine (*Pinus contorta* Douglas. ex Loud. Var. *latifolia* Engelm.) in northwestern Alberta (Kurz et al., 2008; Safranyik et al., 2010; Cullingham et al., 2011; De la Giroday et al., 2012; Lusebrink et al., 2013). Of emerging concern is potential range expansion of mountain pine beetle to the Great Lakes Region of North America with evolutionarily naïve host species that include jack pine, red pine (*Pinus resinosa* Ait.), eastern white pine (*Pinus strobus* L.), and naturalized Scots pine (*Pinus sylvestris* L.) (Cooke and Carroll, 2017; Rosenberger et al., 2017b). A recent study has shown that age classes and distributions of these species exist on the landscape in categories considered highly susceptible to this insect based on comparative studies with suitable and susceptible hosts in western North America (Windmuller-Campione, 2018). While little is known about susceptibility of live potential hosts in the threatened range, it is apparent that mountain pine beetle is able to colonize, tunnel, attract mates, and reproduce within cut logs of all four species of pine (Rosenberger et al., 2017b; Cale et al., 2017).

If mountain pine beetle were to arrive in the Great Lakes Region of North America, it may encounter species of secondary bark beetles and woodborers with which it shares no evolutionary history. This mixing would result in novel ecological interactions that may either facilitate or impede the invasion of mountain pine beetle. In regions where mountain pine beetle is endemic, for example, persistence of mountain pine beetle in the ecosystem is facilitated by species of secondary bark beetles that allow mountain pine beetle to co-colonize weakened trees in low numbers (Safranyik and Carroll, 2006; Smith et al., 2011; Burke and Carroll, 2016). When an environmental stress lowers stand resistance, increasing populations of mountain pine beetle begin partially or mass-attacking mature, live hosts in a population phase transition across the incipient-eruptive threshold (Safranyik and Carroll, 2006). At high levels, secondary species may then inhibit the population growth of mountain pine beetle. Rankin and Borden (1991), for example, showed that forcing a secondary bark beetle such as the pine engraver, *Ips pini* Say, to compete with mountain pine beetle in synchronously co-colonized logs resulted in fewer progeny for both species than if logs were infested independently.

The Great Lakes Region has several native species of secondary bark beetles such as the pine engraver, *I. pini*, the red turpentine beetle, *Dendroctonus valens* LeConte, and the eastern five-spined ips, *Ips grandicollis* Eichhoff. *Ips grandicollis* is an especially common, multivoltine secondary bark beetle native to the Great Lakes Region that typically infests dying or stressed pines (Erbilgin et al., 2002; Lombardero et al., 2006). It is unknown how mountain pine beetle and *I. grandicollis* might interact in the complex of naïve pine species if they were to become sympatric in the Great Lakes region. While pheromones of bark beetles of the genera *Dendroctonus* and *Ips* have been shown to inhibit each other's response in several cases (Rankin and Borden, 1991; Byers, 1989a; Symonds and Elgar, 2004), the responses of *I. grandicollis* and mountain pine beetle to each other's pheromones have never been studied. *Ips grandicollis* utilizes a species-specific aggregation pheromone, consisting mainly of ipsenol (2-methyl-6-methylideneoct-7-en-4-ol), to attract conspecifics and find mates (Witanachchi and Morgan, 1981). Mountain pine beetle uses both *trans*-verbenol and (±)-*exo*-brevicomin as its aggregation pheromones (Conn et al., 1983). In this study, we aim to elucidate how the colonization behavior of each species of bark beetle is affected by the presence of the other's aggregation pheromones in logs of jack pine. Further, we examine the patterns of reproduction of bark and woodboring beetles in logs baited with each type of pheromone. We hypothesize that there will be no cross attraction or inhibition between *I. grandicollis* and mountain pine beetle. Our goal is to understand whether cross attraction and thus competitive interactions might exist between *I. grandicollis* as a bark beetle native to the Great Lakes Region and the potential invader, mountain pine beetle.

2. Materials & methods

2.1. Host material

Jack pine was used as a representative novel species of pine native to the Great Lakes Region. Mountain pine beetle is known to colonize and reproduce in jack pine in western forests (Cullingham et al., 2011) and can similarly colonize and reproduce in cut logs of jack pine from Minnesota (Rosenberger et al., 2017a; Rosenberger et al., 2017b). Twelve trees of 25–27 cm in diameter at 1.4 m were harvested from a single-aged stand of jack pine approximately 40 years old at the University of Minnesota Cloquet Forestry Center (CFC) in Cloquet, Minnesota (46.704490° N, −92.525310° E) on 23 July 2018 and again on 22 July 2019. The site from which trees were harvested is classified as a fire dependent ecosystem in the Laurentian Mixed Forest province of the state, with a climate moderately modified by Lake Superior and lying within USDA Cold Hardiness Zone 4a (Reinikainen et al., 2015).

Harvested trees were visibly free of infestation of bark beetles or other insects. Each year, the twelve felled trees were cut into 48 sections approximately 56 cm in length. The ends of these logs were then immediately sealed with melted paraffin wax (Gulf Wax, Roswell, Georgia) to reduce the desiccation rate of the logs. Once the wax cooled, logs were enclosed in 13-gallon drawstring plastic bags and placed in the bed of a covered pickup truck to prevent unwanted infestation from pine-infesting insects. Each year, the logs were transported to the Black Hills National Forest near Rapid City, South Dakota (44.072086° N, −103.234799° E) where populations of mountain pine beetle and *I. grandicollis* currently exist sympatrically. The logs were deployed at 12 different sites in stands of mature ponderosa pine within 48 h of being felled. Sites were separated by at least 0.5 km. At each site, four logs were suspended at breast height using MIL-C-5040 Type III 550 lb. nylon paracord (Paracord Planet, Fargo, North Dakota) in a square formation at 25 m spacing between logs.

2.2. Experimental design

We utilized a 2x2 factorial design in each of the twelve sites each year as follows: each position within the square was randomly assigned a commercially available pheromone treatment (Alpha Scents, Inc., West Linn, Oregon). Two positions received *I. grandicollis* lure, while two did not. Two positions then also received lures for mountain pine beetle, while the other two did not, such that each site ended up with four treatments as follows: a mountain pine beetle lure (myrcene, 2-(2H-benzotriazol-2-yl)-4-methyl-phenol, butylated hydroxytoluene, *trans*-verbenol, and *exo*-brevicomin), an *I. grandicollis* lure containing ipsenol, both lure types, and a control with no lures in combination. Amounts and elution rates of lure compounds are shown in online supplementary material. Each lure packet was attached to the log using a nail through the outer bark.

Logs were exposed to field conditions from July 25-August 8 of the years 2018 and 2019 to correspond to the flight period of both mountain pine beetle (Safranyik and Carroll, 2006) and *I. grandicollis*, which is typically abundant throughout the summer (Erbilgin et al. 2002). Logs were then retrieved from the field and placed inside rearing tubes at the proximate Wheaton College Science Station (44.061693° N, −103.407476° E) under ambient temperature conditions where any colonizing insects were allowed to continue development undisturbed (Rosenberger et al, 2018).

For each annual replicate, we removed the logs from the rearing tubes and debarked them in two batches: one half after one month (i.e., mid-September), and the other half after one year (i.e., July of the following year) ($n = 24$ each time). Logs debarked in the first batch of each year were chosen by randomly selecting six of the twelve sites and then debarking all logs from that site to ensure equal sampling of treatments. We chose to debark logs at these two different time periods in order to elucidate differences in insect community composition and development after one month and after one year. Outer bark was removed with a wood chisel, exposing the internal larval galleries created by insects. All adults, larvae, and pupae of insects were counted from each gallery and placed in vials of 95% ethanol for identification. All bark beetle larvae found were assumed to be from the species that had constructed the parental gallery. After debarking, a clear plastic sheet was laid over the phloem layer of the peeled barked and markers of different colors were used to trace and color-code galleries of bark beetles, creating a map of subcortical utilization. A Scalex PlanWheel XL was used to measure the one-dimensional length of bark beetle parental galleries. In the second year of the study, most of the *I. grandicollis* brood in the initial set of peeled logs had matured to adults by the time logs were peeled in mid-September. Because we could not confidently distinguish initial colonizing beetles from new progeny that had reached adult life stages, we report total counts of bark beetles in each log for each year without assigning colonizer and brood status.

The remaining sets of logs ($n = 24$ each year) were peeled in July of the following year, approximately one year after field exposure to allow any univoltine species such as mountain pine beetle to complete development (Rosenberger et al., 2018). All insects were collected from within the tubes. Larval *Monochamus* spp. were found to have consumed most of the phloem resource, destroying most, if not all, of the bark beetle galleries within. Due to the amorphous nature of the cerambycid galleries we chose to use imageJ (Schneider et al., 2012) to measure the area of phloem consumed by cerambycids rather than gallery length. As such, data analyzed from logs left in rearing tubes for one year only includes the total area of phloem resource consumed by *Monochamus* spp. larvae. Larvae of different species of *Monochamus* cannot be identified to species, so were simply tallied as *Monochamus*.

2.3. Statistical analysis

For the logs peeled after six weeks of field exposure and incubation time each year, we constructed separate mixed effects models using response variables of the number of bark beetles, the number of galleries, the number of *Monochamus* spp. larvae and the total phloem surface area consumed. Each model used an ANOVA framework where the fixed effects incorporated the 2x2 factorial design (i.e., terms for mountain pine beetle pheromone lure, *I. grandicollis* pheromone lure, and their interaction) and a term for site was included as a random effect. A random effect term for year was also originally included in the models but low variation between years induced issues of model singularity, so the term for year was removed. Data from one site in the second year ($n = 4$ logs) was excluded from our analysis due to labeling error obscuring treatment. Square root transformations were used on the response variables to satisfy the assumptions of a normal distribution of residuals and homogenous variances. Assumptions were assessed by visual inspection of residual plots. Statistical significance was evaluated using $\alpha = 0.05$. All statistical analyses were conducted using R version 3.5.3 (R Core team, 2020).

3. Results

Ips grandicollis readily colonized the logs of jack pine in the field but numbers varied with lure type. The ipsenol treatment significantly enhanced colonization by *I. grandicollis* (Table 1). The two logs baited with ipsenol had a mean of 0.1 ± 0.3 SE ovipositional galleries of *I. grandicollis* per dm^2 per log compared to a mean of 0.001 ± 0.001 galleries/dm^2/log in the two logs without (Fig. 1A). Conversely, the presence of mountain pine beetle pheromone lures appeared to have a significant negative effect on the colonization behavior of *I. grandicollis* (Table 1). The two logs baited with myrcene and the aggregation pheromones of mountain pine beetle had a mean of 0.024 ± 0.006 ovipositional galleries of *I. grandicollis* per dm^2 log, compared to 0.078 ± 0.014 in the two logs without. There was a significant interaction effect between the two treatments (Table 1) as fewer ovipositional galleries of *I. grandicollis* were constructed in logs baited with both ipsenol and mountain pine beetle lures than would be expected if both main effects were additive (Fig. 1A). Thus, the presence of the mountain pine beetle lure inhibits the response of *I. grandicollis* to its own pheromone.

Unsurprisingly, the effect of pheromone treatments on the number of insects found under the bark was very similar to the effects noted on the number of ovipositional galleries found above (Table 1; Fig. 1B). We observed significantly more *I. grandicollis* inside logs baited with ipsenol with a mean of 3.0 ± 0.9/dm^2 compared to those logs without the pheromones. In contrast, very few *I. grandicollis* were found in logs baited with mountain pine beetle lures versus those without (Fig. 1B). Again, we observed a significant interaction effect between the two lure types: we found significantly fewer *I. grandicollis* in logs baited with both pheromones than would be expected with an additive effect of the two lures (Table 1; Fig. 1B). In these co-baited logs, we collected a mean of 0.70 ± 0.32 *I. grandicollis*/dm^2/log. These results confirm an aversive response from *I. grandicollis* to the pheromone lure of mountain pine beetle.

In logs left in rearing tubes for one month, significantly more

Table 1
Summary of 2 × 2 factorial statistical results from field experiments using logs of jack pine baited with *I. grandicollis* (IGR) × mountain pine beetle (MPB) lures, $n = 12$ sites of 4 treatments each in Black Hills of South Dakota, each of 2018 and 2019.

Response variable	IGR lure		MPB lure		IGR × MPB Interaction	
	$F_{1,31}$	P	$F_{1,31}$	P	$F_{1,31}$	P
I. grandicollis galleries	97.57	<0.001	12.23	<0.001	8.99	<0.005
I. grandicollis insects	42.73	<0.001	10.04	<0.001	5.62	0.024
Monochamus larvae	11.62	<0.001	0.77	0.370	8.41	0.007
Phloem area consumed	4.46	0.0420	1.18	0.290	2.76	0.110

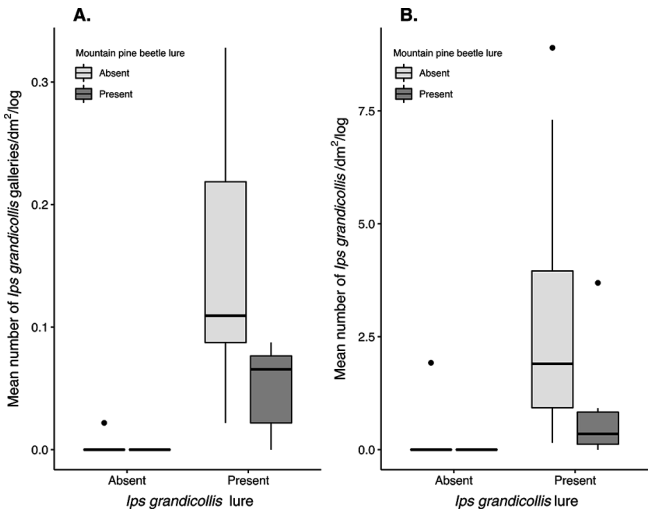


Fig. 1. Box and whisker plots of A) the number of galleries formed by *I. grandicollis* and B) the combined number of *I. grandicollis* adults, pupae, and larvae collected from logs baited with the different aggregation pheromones and peeled after thirty days in emergence tubes ($n = 44$ logs; mean surface area of a log was 45.74 dm^2). The upper whisker indicates the maximum value as the third quartile added to 1.5 times the interquartile range. The lower whisker represents the minimum value as the first quartile minus 1.5 times the interquartile range. The middle line of each box and whisker plot represents the median of the data set. Dots represent data points beyond plus or minus 1.5 times the interquartile range.

Monochamus beetle larvae were found in logs baited with ipsenol than those baited with mountain pine beetle aggregation pheromones (Table 1; Fig. 2). A mean of 0.42 ± 0.11 SE *Monochamus* larvae/dm^2 were found in logs baited with ipsenol compared to 0.31 ± 0.09 larvae/dm^2 found in logs without ipsenol. There was no overall effect of the mountain pine beetle lure on the numbers of *Monochamus* larvae (Table 1), with a mean of 0.28 ± 0.08 *Monochamus* larvae collected per dm^2. We did note a significant interaction effect between the two lure types (Table 1), with more longhorn borers in the logs with mountain pine beetle lure but no ipsenol than would be expected if the main effects were simply additive.

We did not find any mountain pine beetles in the logs peeled in the fall of each year or the sets that were allowed to develop for one year. Any potential colonizers may have been destroyed by *Monochamus*, however, as up to 90% of the surface area of the logs' phloem had been chewed apart by developing larvae. We noted small but statistically significant differences of phloem area consumed by *Monochamus* larvae between some lure treatments in logs left in rearing tubes for one year (Table 1; Fig. 3). In logs baited with ipsenol, *Monochamus* larvae consumed a mean of 2,015 ± 308 cm^2 which was significantly more than logs baited with no lures that had a mean area of 1,084 cm^2 ± 290 cm^2 of phloem consumed per log (IGR effect in Table 1). Logs baited with mountain pine beetle lures and those baited with both lures had 1733 cm^2 ± 262 cm^2 and 1,762 cm^2 ± 289 cm^2 of phloem consumed by *Monochamus* larvae respectively. Neither the effect of the MPB lure nor the interaction effect was significant (Table 1).

4. Discussion

Our results that the eastern five-spined pine engraver constructed fewer galleries and produced fewer offspring in logs that were baited with mountain pine beetle lures, even in the presence of their own aggregation pheromones, suggests that they will not compete directly mountain pine beetle if the latter were to arrive in the Great Lakes region. These two species have no historic sympatric association in the jack pine forests of Minnesota from where the logs originated. The

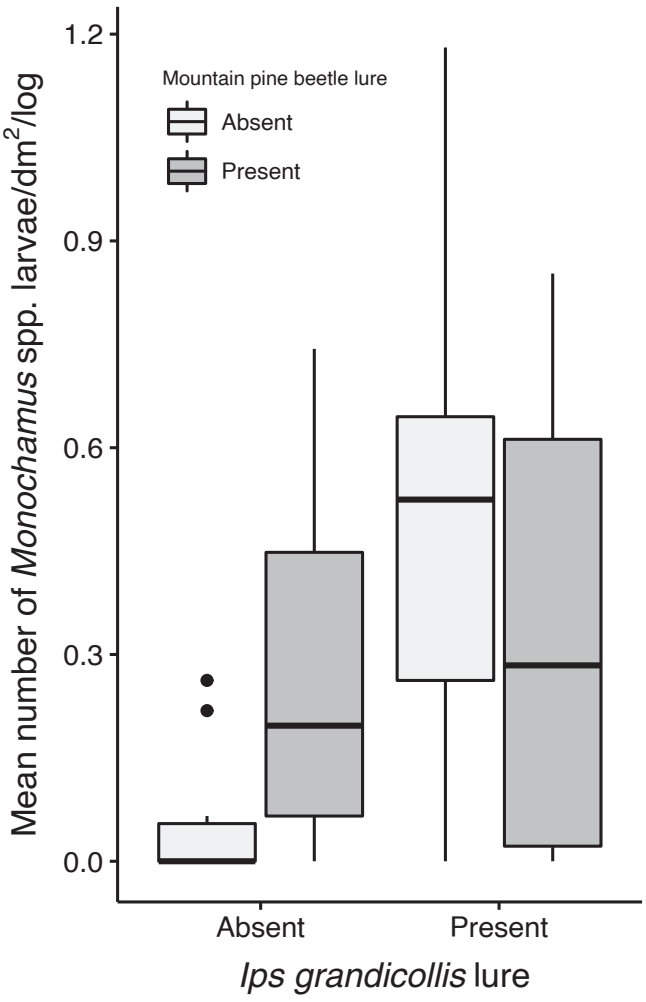


Fig. 2. Box and whisker plots of the number of cerambycid larvae collected from logs peeled after thirty days in emergence tubes of each treatment ($n = 44$ logs; mean surface area of a log was 45.74 dm^2). The upper whisker indicates the maximum value as the third quartile added to 1.5 times the interquartile range. The lower whisker represents the minimum value as the first quartile minus 1.5 times the interquartile range. The middle line of each box and whisker plot represents the median of the data set. Dots represent data points beyond plus or minus 1.5 times the interquartile range.

aversive behavior of eastern five-spined pine engraver to pheromones of mountain pine beetle reflects long conserved traits unique to both genera, consistent with interaction between *Ips* spp. and *Dendroctonus* spp. elsewhere. For example, Byers and Wood (1980) demonstrated that *Ips paraconfusus* and *Dendroctonus brevicomis* are both captured in traps in smaller quantities when in the presence of logs infested by both species rather than just conspecifics. Additionally, response of mountain pine beetle to its aggregation pheromones has been shown in both laboratory and field bioassays to be inhibited by the addition of ipsdienol, a component of the aggregation pheromone produced by *I. pini* (Hunt and Borden, 1988). Similarly, response of *I. pini* to its pheromone component ipsdienol is inhibited by the presence of mountain pine beetle aggregation pheromones involving myrcene, *trans*-verbenol, and *exo*-brevicomin or some combination of the three (Hunt and Borden, 1988).

The aversion of bark beetles to aggregation pheromones of other species facilitates pheromone-mediated niche partitioning, regulating interspecific competition within the tree (Paine et al., 1981; Byers, 1989a). Several species of *Ips* in the southern United States compete with the southern pine beetle, *Dendroctonus frontalis* (Zimmerman), for example (Stephen, 2011). Typically arriving after *D. frontalis*, *I. avulsis*

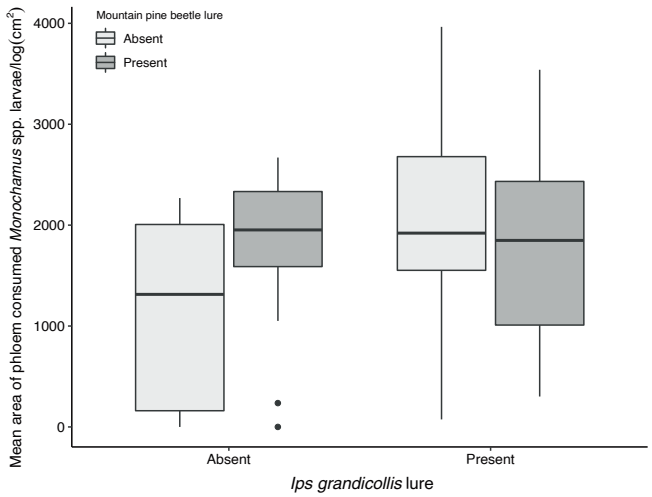


Fig. 3. Box and whisker plots of the area of phloem resource consumed by cerambycid larvae in logs of each treatment peeled after one year ($n = 48$ logs) in emergence tubes. Mean surface area of a log was 45.74 $dm^{2.}$ The upper whisker indicates the maximum value as the third quartile added to 1.5 times the interquartile range. The lower whisker represents the minimum value as the first quartile minus 1.5 times the interquartile range. The middle line of each box and whisker plot represents the median of the data set. Dots represent data points beyond plus or minus 1.5 times the interquartile range.

(Eichoff), *I. grandicollis,* and *I. calligraphus* (Germar) may all compete for resources within southern pine species, usually resulting in less phloem resource being consumed by each species when simultaneously infesting a tree than if they were individually infesting that tree (Stephen, 2011). While *D. frontalis* has been shown to be unresponsive to the pheromones produced by the southern *Ips* spp. complex, each of the *Ips* species in this system respond to pheromones of others *Ips* with only *I. grandicollis* responding to pheromones produced by *D. frontalis* (Svihra et al., 1980; Stephen, 2011). Interestingly, the response of *I. grandicollis* to pheromones of *D. frontalis* described by Svihra and others (1980) was one of attraction, which contrasts with our results. This pattern could indicate that sympatric coevolution of *I. grandicollis* with *D. frontalis* has resulted in the attraction of *I. grandicollis* to its pheromones while the lack thereof with mountain pine beetle has maintained its inhibitory response.

Most studies of pheromone responses in bark beetles have been conducted in areas of high populations, which may obscure responses to semiochemicals that vary with population density (Wallin and Raffa, 2004). Responses of both *I. pini* and *Ips latidens* (LeConte) to their respective aggregation pheromones, for example, are interrupted by verbenone, the anti-aggregation pheromone of mountain pine beetle, with increasing levels of interruption with increasing verbenone release rates (Borden et al., 1992; Miller et al., 1995). These examples of niche separation occur with bark beetles at high population levels. It is possible that facilitative interactions may emerge when at endemic population levels. Previous infestation of pines by *Pseudips mexicanus* Hopkins, for example, has been shown to create more suitable resources for mountain pine beetle when mountain pine beetle is in endemic population sizes. Mountain pine beetle was found to be attacking in higher densities while excavating similar sized galleries in trees previously infested by *P. mexicanus* rather than those not previously infested (Smith et al, 2011).

Species of *Monochamus* exist naturally in the Black Hills National Forest as well as the Great Lakes Region. *Monochamus clamator* is the most abundant sawyer beetle found in the Black Hills, SD and we suspect most larvae were of this species. Larvae of *Monochamus* beetles infest and feed on the phloem and xylem of pines. *Monochamus* spp. are thus potential subcortical competitors with mountain pine beetle and *I. grandicollis* as well as facultative predators (Dodds et al., 2001; Schenk and Benjamin, 1969; Schoeller et al., 2012). These findings are consistent with previous field and laboratory studies indicating that beetles in the *Monochamus* genus are kairomonally responsive to aggregation pheromones of *Ips* species (Rassati et al. 2012; Pajares et al. 2017; Chase et al. 2018) but not those of *Dendroctonus* (Dodds et al., 2001; Allison et al., 2003). These findings further indicate that species of *Monochamus* may also not act as significant agents of competition or predation on mountain pine beetle if it were to arrive in the Great Lakes Region but may instead compete more with native species of *Ips*.

While we noted aversion of *I. grandicollis* to mountain pine beetle lures, a lack of mountain pine beetles colonizing the logs precludes ability to make conclusions concerning the response of mountain pine beetle to pheromones of *I. grandicollis*. We also know little about potential responses of natural enemies of bark beetles to pheromones of mountain pine beetle in the Great Lakes Region (Pfammatter et al., 2015). It is possible that lack of coevolution with mountain pine beetle will correspond to a lack of response to the native complex of bark beetle predators and parasitoids, functionally releasing invasive populations of mountain pine beetle from the pressures of natural enemies. Additionally, our results suggest that insects such as the *I. grandicollis* may avoid trees being mass attacked by mountain pine beetle.

This research adds to existing bodies of literature of competition and interspecific pheromonal response of primary vs. secondary bark beetles as well as the response of Cerambycid beetles to the aggregation pheromones of bark beetles. Future work should focus on the kairomonal and/or allomonal responses of the diverse bark beetle predators as well as other potential competitors native to the Great Lakes Region to the aggregation pheromones produced by mountain pine beetle (Smith, 2021). Future work should also focus on testing and characterizing direct under-bark interactions where they exist, as neutral, facilitatory, or competitive interactions between bark beetle species mediate persistence at endemic levels and can influence population phase transitions (Rankin and Borden, 1991; Safranyik and Carroll, 2006; Smith et al, 2011). Understanding these interactions is especially important given the emerging threat to eastern North America by mountain pine beetle (Cudmore et al., 2010; Cullingham et al., 2011; Rosenberger et al., 2017b; Rosenberger et al., 2018).

CRediT authorship contribution statement

Zach M. Smith: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Supervision. **Kevin D. Chase:** Conceptualization, Methodology, Investigation, Writing - review & editing, Funding acquisition. **Etsuro Takagi:** Methodology, Investigation, Writing - review & editing. **Aubree M. Kees:** Conceptualization, Methodology, Investigation, Resources, Writing - review & editing, Project administration. **Brian H. Aukema:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Mara Short, Kristine Jecha, Cole Doolittle, Nicole Witt, and Stephanie Gunter for their assistance in field work and data collection as well as the Minnesota Department of Natural Resources for logistical support. Kyle Gill and Lane Johnson of the Cloquet Forestry Center assisted in acquiring and providing pine material. We thank Dr. Chris Keil, Bob Schryvers, and the rest of the staff at the Wheaton College Science Station for their warm hospitality and facilitating research space. This work was funded by the Minnesota Invasive Terrestrial Plant

and Pest Center. The comments of Drs. Robert Venette (US Forest Service), Robert Blanchette (UMN) and an anonymous reviewer greatly improved previous versions of this work.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foreco.2021.119455.

References

Allison, J.D., Morewood, W.D., Borden, J.H., Hein, K.E., Wilson, I.M., 2003. Differential bio-activity of *Ips* and *Dendroctonus* (Coleoptera: Scolytidae) pheromone components for *Monochamus clamator* and *M. scutellatus* (Coleoptera: Cerambycidae). Environ. Entomol. 32 (1), 23–30.

Aukema, B.H., Carroll, A.L., Zhu, J., Raffa, K.F., Sickley, T.A., Taylor, S.W., 2006. Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. Ecography 29 (3), 427–441.

Aukema, B.H., McKee, F.R., Wytrykush, D.L., Carroll, A.L., 2016. Population dynamics and epidemiology of four species of *Dendroctonus* (Coleoptera:Curculionidae): 100 years since J.M. Swaine. Can. Entomol. 148, 82–110.

Ayres, B.D., Ayres, M.P., Abrahamson, M.D., Teale, S.A., 2001. Resource partitioning and overlap in three sympatric species of *Ips* bark beetles (Coleoptera:Scolytidae). Oecologia 128, 443–453.

Bentz, Barbara, Regniere, Jacques, Fettig, Christopher, Hansen, E. Matthew, Hayes, Jane, Hicke, Jeffrey, Kelsey, Rick, Negron, Jose, Seybold, Steven, 2010. Climate change and bark beetles of the western united states and canada: direct and indirect effects. BioScience 60 (8), 602–613. https://doi.org/10.1525/bio.2010.60.8.6.

Beudert, B., Bassler, C., Thorn, S., Noss, R., Schroder, B., Dieffenbach-Fries, H., Foullois, N., Muller, J., 2014. Bark beetles increase biodiversity while maintaining drinking water quality. Conserv. Lett. 8 (4), 272–281.

Birch, M.C., Wood, D.L., 1975. Mutual inhibition of the attractant pheromone response by two species of *Ips* (Coleoptera: Scolytidae). J. Chem. Ecol. 1 (1), 101–113.

Borden, J.H., Devlin, D.R., Miller, D.R., 1992. Synomones of two sympatric species deter attack by the pine engraver, *Ips pini* (Coleoptera: Scolytidae). Can. J. For. Res. 22, 523–524.

Burke, J.L., Carroll, A.L., 2016. The influence of variation in host tree monoterpene composition on secondary attraction by an invasive bark beetle: Implications for range expansion and potential host shift by the mountain pine beetle. For. Ecol. Manag. 359, 59–64.

Byers, J.A., 1989a. Chemical ecology of bark beetles. Experientia. 45, 271–283.

Byers, J.A., 1989b. Behavioral mechanisms in reducing competition in bark beetles. Holarctic. Ecol. 12 (4), 466–476.

Byers, J.A., Wood, D.L., 1980. Interspecific inhibition of the response of the bark beetles, *Dendroctonus brevicomis* and *Ips paraconfusus*, to their pheromones in the field. J. Chem. Ecol. 6 (1), 149–164.

Cale, J.A., Muskens, M., Najar, A., Ishangulyyeva, G., Hussain, A., Kanekar, S.S., Klutsch, J.G., Taft, S., Erbilgin, N., 2017. Rapid monoterpene induction promotes the susceptibility of a novel host pine to mountain pine beetle colonization but not to beetle-vectored fungi. Tree Physiol. 37 (12), 1597–1610.

Carroll, A.L., Taylor, S.W., Régnière, J., Safranyik, L., 2004. Effects of climate and climate change on the mountain pine beetle. In: Proc. of the mountain pine beetle symposium: Challenges and solutions, Canadian For. Serv., Pacific Forestry Centre, Information Rep. BC-X-399, 223-232.

Chase, K.D., Stringer, L.D., Butler, R.C., Liebhold, A.M., Miller, D.R., Shearer, P.W., Brockerhoff, E.G., 2018. Multiple-lure surveillance trapping for *Ips* bark beetles, *Monochamus* longhorn beetles, and *Halymorpha halys* (Hemiptera: Pentatomidae). J. Econ. Entomol. 111 (5), 2255–2263.

Conn, J.E., Borden, J.H., Scott, B.E., Friskie, L.M., Pierce Jr., H.D., Oehlschlager, A.C., 1983. Semiochemicals for the mountain pine beetle, *Dendroctonus ponderosae* (Coleoptera: Scolytidae) in British Columbia: field trapping studies. Can. J. For. Res. 13 (2), 320–324.

Cudmore, T.J., Bjorklund, N., Carroll, A.L., Lindgren, B.S., 2010. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve host tree populations. J of Appl. Ecol. 47, 1036–1043.

Cooke, B.J., Carroll, A.L., 2017. Predicting the risk of mountain pine beetle spread to eastern pine forests: Considering uncertainty in uncertain times. For. Ecol. Manag. 396, 11–25.

Cullingham, C.I., Cooke, J.E.K., Dang, S., Davis, C.S., Cooke, B.J., Coltman, D.W., 2011. Mountain pine beetle host-range expansion threatens the boreal forest. Mol. Ecol. 20, 2157–2171.

De la Giroday, H.M.C., Carroll, A.L., Aukema, B.H., 2012. Breach of the northern Rocky Mountain geoclimatic barrier: initiation of range expansion by the mountain pine beetle. J. Biogeogr. 39 (6), 1112–1123.

Dodds, K.J., Graber, C., Stephen, F.M., 2001. Facultative intraguild predation by larval Cerambycidae (Coleoptera) on bark beetle larvae (Coleoptera: Scolytidae). Environ. Entomol. 30 (1), 17–22.

Erbilgin, N., Nordhein, E.V., Aukema, B.H., Raffa, K.F., 2002. Population dynamics of *Ips pini* and *Ips grandicollis* in red pine plantations in Wisconsin: within- and between-year associations with predators, competitors, and habitat quality. Environ. Entomol. 31 (6), 1043–1051.

Fettig, C.J., Klepzig, K.D., Billings, R.F., Munson, A.S., Nebeker, T.E., Negron, J.F., Nowak, J.T., 2007. The effectiveness of vegetation management practices for prevention and control of bark beetle infestations in coniferous forests of the western and southern United States. For. Ecol. Manag. 238, 24–53.

Hicke, J.A., Allen, C.D., Desai, A.R., Dietze, M.C., Hall, R.J., 2012. Effects of biotic disturbances on forest carbon cycling in the United States and Canada. Glob. Chang. Biol. 18, 7–34.

Hunt, D.W.A., Borden, J.H., 1988. Response of mountain pine beetle, *Dendroctonus ponderosae* Hopkins, and pine engraver, *Ips pini* (Say) to ipsdienol in southwestern British Columbia. J. Chem. Ecol. 14 (1), 277–293.

Klutsch, J.G., Shamoun, S.F., Erbilgin, N., 2017. Drought stress leads to systemic induced susceptibility to a nectrotrophic fungus associated with mountain pine beetle in *Pinus banksiana* seedlings. PLOSOne 12 (12), e0189203.

Kurz, W.A., Dymond, C.C., Stinson, G., Rampley, G.J., Neilson, E.T., Carroll, A.L., Ebata, T., Safranyik, L., 2008. Mountain pine beetle and forest carbon feedback to climate change. Nature 452, 987–990.

Lindgren, B.S., Raffa, K.F., 2013. Evolution of tree killing in bark beetles (Coleoptera: Curculionidae): trade-offs between the maddening crowds and a sticky situation. Can. Entomol. 145, 471–495.

Lombardero, M.J., Ayres, M.P., Ayres, B.D., 2006. Effects of fire and mechanical wounding on *Pinus resinosa* resin defenses, beetle attacks, and pathogens. For. Ecol. Manag. 225, 349–358.

Lusebrink, I., Erbilgin, N., Evenden, M.L., 2013. The lodgepole x jack pine hybrid zone in Alberta, Canada: a steppingstone for the mountain pine beetle on its journey east across the boreal forest? J. Chem. Ecol. 39, 1209–1220.

Mikkelson, K.M., Bearup, L.A., Maxwell, R.M., Stednick, J.D., McCray, J.E., Sharp, J.O., 2013. Bark beetle infestation impacts on nutrient cycling, water quality and interdependent hydrological effects. Bio. Geo. Chem 115, 1–21.

Miller, D.R., Borden, J.H., Lindgren, B.S., 1995. Verbenone: dose-dependent interruption of pheromone-based attraction of three sympatric species of pine bark beetles (Coleoptera: Scolytidae). Environ. Entomol. 24 (3), 692–696.

Paine, T.D., Birch, M.C., Svihra, P., 1981. Niche breadth and resource partitioning by four sympatric species of bark beetles (Coleoptera: Scolytidae). Oecologia. 48, 1–6.

Pajares, J.A., Alvarez, G., Hall, D.R., Ibarra, N., Hoch, G., Halbig, P., Cocos, D., Johansson, H., Schroeder, M., 2017. Attractants for management of the pine sawyer beetle *Monochamus sutor*, a potential vector of *Bursaphelenchus xylophilus*. J. Appl. Entomol. 141, 97–111.

Pfammatter, J.A., Krause, A., Raffa, K.F., 2015. Evaluating predators and competitors in Wisconsin red pine forests for attraction to mountain pine beetle pheromones for anticipators biological control. Environ. Entomol. 44 (4), 1161–1171.

Raffa, K.F., Aukema, B.H., Bentz, B.J., Carroll, A.L., Hicke, J.A., Turner, M.G., Romme, W.H., 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. BioSci. 58 (6), 501–517.

Rankin, L.J., Borden, J.H., 1991. Competitive interactions between the mountain pine beetle and the pine engraver in lodgepole pine. Can. J. For. Res. 21 (7), 1029–1036.

Rassati, D., Toffolo, E.P., Battisti, A., Faccoli, M., 2012. Monitoring of the pine sawyer beetle *Monochamus galloprovincialis* by pheromone traps in Italy. Phytoparasitica 40, 329–336.

Reinikainen, M.R., Nagel, L., Pike, C., Olesiak, R., Cuomo, G., 2015. Cloquet forestry center management plan. Staff Paper Series No. 244 Department of Forest Resources.

Rosenberger, D.W., Aukema, B.H., Venette, R.C., 2017a. Cold tolerance of mountain pine beetle among novel eastern pines: a potential for trade-offs in an invaded range? For. Ecol. Manag. 400, 28–37.

Rosenberger, D.W., Venette, R.C., Maddox, M.P., Aukema, B.H., 2017b. Colonization behaviors of mountain pine beetle on novel hosts: Implications for range expansion into northeastern North America. PLoSOne. 12 (5), e0176269. https://doi.org/10.1371/journal.pone.0176269.

Rosenberger, D.W., Venette, R.C., Aukema, B.H., 2018. Development of an aggressive bark beetle on novel hosts: Implications for outbreaks in an invaded range. J. Appl. Ecol. 55, 1526–1537.

R Core Team, 2020. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. URL https://www.R-project.org..

Safranyik, L., 1978. Effects of climate and weather of mountain pine beetle populations. In: Proceedings of symposium on theory and practice of mountain pine beetle management in lodgepole pine forests, 77-84.

Safranyik, L., Carroll, A.L., 2006. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests, pp. 3-66. In: L. Safranyik and B. Wilson (Eds), The mountain pine beetle: a synthesis of its biology, management and impacts on lodgepole pine. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada.

Safranyik, L., Carroll, A.L., Regniere, J., Langor, D.W., 2010. Potential range expansion of mountain pine beetle into the boreal forest of North America. Can. Entomol. 142 (5), 415–442.

Schenk, J.A., Benjamin, D.M., 1969. Notes on the biology of *Ips pini* in central Wisconsin jack pine forests. Ann. Etomol. Soc. Am. 62 (3), 480–485.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods. 9 (7), 671–675.

Schoeller, E.N., Husseneder, C., Allison, J.D., 2012. Molecular evidence of facultative intraguild predation by *Monochamus titillator* larvae (Coleoptera: Cerambycidae) on member of the southern pine beetle guild. Naturwissenschaften. 99, 913–924.

Smith, G.D., Carroll, A.L., Lindgren, B.S., 2011. Facilitation in bark beetles: endemic mountain pine beetle gets a helping hand. Agr. Forest Entomol. 13, 37–43.

Smith, Z.M., 2021. Responses by natural enemies to semiochemicals associated with mountain pine beetle (*Dendroctonus ponderosae*). Master's Thesis. University of Minnesota, p. 78.

Stephen, FM., 2011. Southern Pine Beetle Competitors. In: Coulson, R.N.; Klepzig, K.D. (Eds.) 2011. Southern Pine Beetle II. Gen. Tech. Rep. SRS-140. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station, 183-198.

Svihra, P., Paine, T.D., Birch, M.C., 1980. Interspecific olfactory communications in southern pine beetle. Naturwissenschaften 67, 518–520.

Symonds, M.R.E., Elgar, M.A., 2004. Species overlap, speciation and the evolution of aggregation pheromones in bark beetles. Ecol. Lett. 7, 202–212.

Taylor, S.W., Carroll, A.L., 2003. Disturbance, forest age, and mountain pine beetle outbreak dynamics in BC: A historical perspective. Natural Resources Canada, Canadian Forest Serve, Pacific Forestry Centre.

Wallin, K.F., Raffa, K.F., 2004. Feedback between individual host selection behavior and population dynamics in an eruptive herbivore. Ecol. Monogr. 74 (1), 101–116.

Windmuller-Campione, M.A., 2018. Assessing the future susceptibility of mountain pine beetle (*Dendroctonus ponderosae*) in the Great Lakes Region using forest composition and structural attributes. Can. J. For. Res. 48, 451–459.

Witanachchi, J.P., Morgan, F.D., 1981. Behaviour of the bark beetle, *Ips grandicollis* during host selection. Physiol. Entomol 6, 219–223.

Wood, S.L., 1982. The bark and ambrosia beetles of North and Central America (Coleoptera:Scolytidae), a taxonomic monograph. Great Basin Naturalist Memoirs 6, Provo, UT.

ML 2015, Ch 76 Art 2, Sec 6a Project Abstract
For the Period Ending December 31, 2021

PROJECT TITLE: Subproject #3: Biological control of the soybean aphid by *Aphelinus certus*
PROJECT MANAGER: George E. Heimpel
AFFILIATION: Dept. of Entomology, University of Minnesota
MAILING ADDRESS: 1980 Folwell Ave
CITY/STATE/ZIP: St. Paul, MN 55108
PHONE: 612-624-3480
E-MAIL: heimp001@umn.edu
WEBSITE: https://entomology.umn.edu/people/george-heimpel
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER
ML 2015, Ch. 76, Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $479,859
AMOUNT SPENT: $479,859
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results
Results of this study indicate that the parasitoid *Aphelinus certus* provides sufficient mortality of soybean aphids to substantially decrease the need to apply insecticides against this pest.

Overall Project Outcome and Result
Prior to the year 2000, the approximately seven million acres of soybeans in Minnesota suffered very little insect damage and were seldom subjected to insecticide applications. This changed with the arrival of the soybean aphid from Asia during that year. This aphid rapidly became the most important insect pest of soybeans due to its ability to substantially lower soybean yield when present at high densities on plants. This led to a 'new normal' that included widespread insecticide use in soybeans in Minnesota, with areas in excess of one million acres sprayed in bad aphid years. While predatory insects were capable of suppressing populations in some years, this level of control was not consistent. We noted the arrival of a new natural enemy of soybean aphid in Minnesota in 2011, however – the parasitoid *Aphelinus certus* – that appeared to have the potential to be a game changer. This insect lays its eggs into soybean aphids, and the developing larvae kill the aphids from within. Our main objective was to determine the extent to which this parasitoid could control populations of soybean aphids below the level that necessitates insecticide use. We also hoped to elucidate agronomic strategies that could lead to increased control by this parasitoid. Based upon a combination of laboratory, field and theoretical studies, we were able to show that *A. certus* is indeed capable for suppressing soybean aphid densities below the threshold levels that farmers use to initiate insecticide use. Our theoretical simulations suggested that such control occurs in approximately 10% of fields during a given year. These studies also pointed to overwintering success of the parasitoids as a critical factor determining the strength of aphid suppression. It therefore stands to reason that any agronomic factors that increase overwintering success improve the parasitoid's capability of suppressing soybean aphid.

Project Results Use and Dissemination
This research led to new analytical tools to analyze the ability of the parasitoid *Aphelinus certus* to control populations of the soybean aphid. It also provided novel information on the primary overwintering site of the parasitoid (within soybean fields) and aspects of its overwintering and diapausing strategy. This information can be used to predict when *A. certus* adults will emerge in a given field season. Lastly, the research quantified the extent of control provided by this parasitoid and generated novel hypotheses for how control can be improved.

We generated an analytical tool using a stage-based matrix modeling approach and published it in an open access Journal. This model can be modified based on environmental and life-history characteristics for this or similar host-parasitoid systems and the underlying R code is available upon request from the authors.

PLOS ONE

RESEARCH ARTICLE

A matrix model describing host–parasitoid population dynamics: The case of *Aphelinus certus* and soybean aphid

James Rudolph Miksanek*, George E. Heimpel

Department of Entomology, University of Minnesota, Saint Paul, Minnesota, United States of America

* miks0007@umn.edu

Check for updates

OPEN ACCESS

Citation: Miksanek JR, Heimpel GE (2019) A matrix model describing host–parasitoid population dynamics: The case of *Aphelinus certus* and soybean aphid. PLoS ONE 14(6): e0218217. https://doi.org/10.1371/journal.pone.0218217

Editor: Antonio Biondi, University of Catania, ITALY

Received: April 2, 2019

Accepted: May 28, 2019

Published: June 13, 2019

Copyright: © 2019 Miksanek, Heimpel. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available through the Data Repository for the University of Minnesota (DRUM) at hdl.handle.net/11299/203074.

Funding: This work was funded by grants to GEH from the Minnesota Invasive Terrestrial Plants and Pests Center (mitppc.umn.edu) through the Environment and Natural Resources Trust Fund, the North Central Soybean Research Program (ncsrp.com), and the Minnesota Soybean Research and Promotion Council (mnsoybean.org/MSRPC), and with support from the University of Minnesota Agricultural Experiment Station. The

Abstract

Integrating elements from life tables into population models within a matrix framework has been an underutilized method of describing host–parasitoid population dynamics. This type of modeling is useful in describing demographically-structured populations and in identifying points in the host developmental timeline susceptible to parasitic attack. We apply this approach to investigate the effect of parasitism by the Asian parasitoid *Aphelinus certus* on its host, the soybean aphid (*Aphis glycines*). We present a matrix population model with coupled equations that are analogous to a Nicholson–Bailey model. To parameterize the model, we conducted several bioassays outlining host and parasitoid life history and supplemented these studies with data obtained from the literature. Analysis of the model suggests that, at a parasitism rate of 0.21 d^{-1}, *A. certus* is capable of maintaining aphid densities below economically damaging levels in 31.0% of simulations. Several parameters—parasitoid lifespan, colonization timeline, host developmental stage, and mean daily temperature—were also shown to markedly influence the overall dynamics of the system. These results suggest that *A. certus* might provide a valuable service in agroecosystems by suppressing soybean aphid populations at relatively low levels of parasitism. Our results also support the use of *A. certus* within a dynamic action threshold framework in order to maximize the value of biological control in pest management programs.

Introduction

The ways in which demography, life history, interspecific interactions, and the biotic or abiotic characteristics of a habitat affect the dynamics of consumer–resource interactions may be investigated through simple experiments and ecological models [1]. Host–parasitoid systems are ideal for these studies not only because of their straightforward structure, but because of their application in the biological control of insect pests [2, 3]. Thus, population modeling has often been used to quantify the importance of parasitoids as natural enemies within a broad range of ecological and evolutionary processes [4]. Other approaches, such as life table analyses, also provide valuable insight into the effect of mortality imposed by parasitoids and other natural enemies on host populations [5, 6].

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Matrix population models are well adapted to studying structured populations [7–10], although they have seldom been used to describe aspects of host–parasitoid systems. Yet, these life table-based models have been successfully used in a variety of systems to address heterogeneity in populations and in identifying vulnerable aspects of the life history of a species, making them useful not only in theory, but in evaluating the impact of biological control agents in practice as well [3, 11–15]. For example, Lin and Ives [16] constructed a size-classified matrix model for soybean aphid (*Aphis glycines*) and the parasitoid wasp *Aphidius colemani*, showing that parasitoid preference for larger individuals tended to have the greatest impact on host population growth, and Mills [17] utilized a stage-structured matrix to identify the developmental stages of the codling moth (*Cydia pomonella*) most susceptible to increased parasitism pressure in a competitive environment.

We present a coupled stage-classified matrix model for a host–parasitoid system. Our method of coupling two species follows that of the more-familiar Nicholson–Bailey equations, $N_{t+1} = \lambda N_t e^{-aP_t}$ and $P_{t+1} = N_t(1 - e^{-aP_t})$, in which N and P are the host and parasitoid population densities, λ is the natural rate of increase for the host, and e^{-aP_t} is the escape function [18]. We parameterized the matrix model for the soybean aphid–*Aphelinus certus* system through a series of developmental and behavioral bioassays as well as with data from the literature. The purpose of this model—which may be adapted to describe other host–parasitoid systems—is to (1) evaluate the extent to which *A*. *certus* might suppress soybean aphid populations below damaging levels, (2) generate hypotheses related to the potential economic and environmental effects of *A*. *certus* in biological control of soybean aphid, and (3) investigate the dynamics of interacting stage-structured populations.

Materials and methods

Study system

Soybean aphid (*Aphis glycines* Matsumura; Hemiptera: Sternorrhyncha: Aphididae) is an important pest of soybean (*Glycine max* (L.) Merrill; Fabaceae) in North America, and often requires treatment with broad-spectrum insecticides that pose risk to an array of non-target organisms [19, 20]. The practice of biological control reduces risk to beneficial species (such as pollinators and natural enemies) as it complements or acts as an alternative to insecticide use [3, 21]. In North America, the biological control services provided by resident enemies offer some protection against soybean aphid (and reduce its overall environmental impact), but damaging outbreaks still occur [19, 22]. The importation and release of exotic specialized parasitoids from the native range of soybean aphid have been attempted unsuccessfully, and various hypotheses for this lack of establishment have been proposed, such as biotic interference [23], intraguild predation [24], and challenges related to overwintering [25, 26] and dispersal [27, 28].

The Asian parasitoid *Aphelinus certus* Yasnosh (Hymenoptera: Chalcidoidea: Aphelinidae) was evaluated for importation and release against soybean aphid. However, *A*. *certus* was determined to be an unsuitable classical biological control agent because it parasitized a broad range of aphid species during tests in quarantine [29, 30]. In or before 2005, *A*. *certus* was accidentally introduced into North America—possibly during a secondary invasion of soybean aphid—and has since spread throughout the north central United States and southeastern Canada [19, 31, 32]. Recent work on *A*. *certus* in Saint Paul, Minnesota, suggests that this parasitoid may be able to maintain soybean aphid populations below the economic threshold of 250 aphids per plant (the pest density at which management practices should be applied) [33], although a different study in the Montérégie area of Québec, Canada, found that *A*. *certus* only decreased peak aphid population densities (and cumulative aphid-days) by 1–7%, possibly due

to low early-season parasitism rates [34]. Thus, the overall impact of *A*. *certus* as a biological control agent of soybean aphid remains uncertain.

The matrix model

Following Caswell [10], a host population vector **n** (the abundance of each developmental stage in the host population) is projected to $t + 1$ (projection interval = 1 d) using a transition and fertility matrix **A** and the probability of escaping parasitism **H**, as well as with a temperature-scaling matrix $\mathbf{C}_{\mathrm{SBA}}$. Similarly, the parasitoid population vector **p** is projected with the transition and fertility matrix **W** and modified by a temperature-scaling matrix $\mathbf{C}_{\mathrm{Ac}}$. The model also includes a carrying capacity K for the total host population N following Allen [35] and Jensen [36–38], and this formulation includes the identity matrix **I** such that $\mathbf{HAC_{SBA}}-\mathbf{I}$ is analogous to the intrinsic rate of increase. Thus, the model takes the form

$$\mathbf{n}_{t+1} = \mathbf{n}_t + \frac{K - N_t}{K}(\mathbf{HAC_{SBA}} - \mathbf{I})\mathbf{n}_t$$

$$\mathbf{p}_{t+1} = \mathbf{WC_{Ac}p}_t \tag{1}$$

The matrices **A** and **W** represent the proportion of individuals in stage j (columns) surviving or transitioning to stage i (rows) from time t to $t + 1$. For the host, the survival probabilities (P_i), the transition probabilities (G_i), and the fertilities (F_i) are reduced by parasitism (g_i). Eq (2) details the host transition and fertility matrix **A** as well as the probabilities of escaping parasitism expressed in **H**

$$\mathbf{AH} = \begin{bmatrix} P_1 & 0 & 0 & 0 & F_5 \\ G_1 & P_2 & 0 & 0 & 0 \\ 0 & G_2 & P_3 & 0 & 0 \\ 0 & 0 & G_3 & P_4 & 0 \\ 0 & 0 & 0 & G_4 & P_5 \end{bmatrix} \begin{bmatrix} g_1 & 0 & 0 & 0 & g_5 \\ g_1 & g_2 & 0 & 0 & 0 \\ 0 & g & g_3 & 0 & 0 \\ 0 & 0 & g_3 & g_4 & 0 \\ 0 & 0 & 0 & g_4 & g_5 \end{bmatrix} \tag{2}$$

in which P_i, G_i, and F_i were calculated assuming a postbreeding census birth-pulse, so that the probability of observing an individual of a specific developmental stage is a function of the sampling period. $P_i = l(i)/l(i-1)$, $G_i = l(i)/l(i-1)$, and $F_5 = P_i g_i m_i$; l_i is the number or proportion of individuals surviving from $i - 1$ to i, m_i is per capita reproduction, and g_i is the proportion of hosts escaping parasitism. g_i was based on a type II functional response for parasitoids attacking hosts that was previously applied to *A*. *certus* by Frewin et al. [31] and takes the form

$$g_i = \exp\left(\frac{-a_i \alpha P_3 p_{3\varphi}}{1 + a_i \alpha T_h N}\right)\left(\frac{\sum_{i=1}^{5} n_i}{N}\right) \tag{3}$$

in which a_i is the fraction of all attacks on host stage i, α is the instantaneous search rate of the parasitoid, and T_h is the handling time. While $\sum_{i=1}^{5} n_i$ represents the total number of unparasitized hosts (the scalar sum of the host population vector **n**), N represents the entire host population, including both the unparasitized ($\sum_{i=1}^{5} n_i$) and parasitized, but still-living, hosts (the element p_1 in the parasitoid population vector **p**). Because only female parasitoids exert parasitism pressure on the host population, the element p_3 from the population vector **p** is multiplied by the proportion of adult parasitoids that are female, and is represented in Eq 3 as $p_{3\varphi}$. As only unparasitized individuals may be parasitized, the escape function is multiplied by the

relative number of available hosts, $\sum_{i=1}^{5} n_i/N$. Note that the stage-specific probability of escaping parasitism g_i was referred to as "p_i" by Lin and Ives [16]; the symbol for this variable was changed here for clarity as entries in our parasitoid population vector **p** are referred to as p_i in conventional matrix notation.

The transition and fertility matrix for the parasitoid (a combined egg and larval stage, mummy/pupal stage, and adult) is

$$\mathbf{W} = \begin{bmatrix} P_1 & 0 & F_3 \\ G_1 & P_2 & 0 \\ 0 & G_2 & P_3 \end{bmatrix} \tag{4}$$

in which P_i and G_i are calculated as before, with the exception of P_3, which incorporates host-density-dependent survival of adult parasitoids modeled using the Verhulst model of logistic growth (Miksanek JR & Heimpel GE, unpublished). Here, $P_3 = [l(3)/l(3-1)]\theta_1/[1+\theta_2 e^{\theta_3 N_{\text{total}}}]$, in which l is the proportion of parasitoids surviving as before, θ_1 is the maximum mean adult parasitoid lifespan, and θ_2 and θ_3 are shape and growth rate parameters. The fertility of adult parasitoids is $F_3 = (1-g_i)n_i/p_3$. Finally, the parasitoid survival and transition matrix **W** was additively decomposed to reflect the effects of the host carrying capacity on parasitoid eggs and larvae (as there is an equal probability of the carrying capacity affecting either parasitized or unparasitized hosts), such that the second line of Eq (1) becomes $\mathbf{p}_{t+1} = \begin{bmatrix} p_1 \\ 0 \\ 0 \end{bmatrix}_t + (\mathbf{WC_{Ac}} - I)\begin{bmatrix} p_1 \\ 0 \\ 0 \end{bmatrix}_t + \mathbf{WC_{Ac}}\begin{bmatrix} 0 \\ p_2 \\ p_3 \end{bmatrix}_t$

Offspring produced by parasitized hosts were added to the element n_1 (number of 1st stadium hosts) in the population vector $\mathbf{n}_{t+1}$ as $\sum_{i=i}^{5} p_1 J_i \frac{n_i}{N}$. Here, post-parasitism reproduction is accounted for by multiplying the stage-specific per capita reproduction of parasitized hosts (J_i) and the proportional stage structure n_i/N with the number of still-living parasitized hosts (p_1). This formulation approximates the stage structure of the parasitized host population by equating it to that of the unparasitized population.

A temperature-scaling matrix was implemented for the host ($\mathbf{C}_{\text{SBA}}$) and parasitoid ($\mathbf{C}_{\text{Ac}}$) in order to adjust population growth rates for temperatures outside of those used in laboratory assays. The temperature-scaling matrices take the form

$$\mathbf{C}_{\text{SBA}} = \begin{bmatrix} \mathrm{c}_{P1} & 0 & 0 & 0 & \mathrm{c}_5{}^4 \\ \mathrm{c}_1 & \mathrm{c}_{P2} & 0 & 0 & 0 \\ 0 & \mathrm{c}_2 & \mathrm{c}_{P3} & 0 & 0 \\ 0 & 0 & \mathrm{c}_3 & \mathrm{c}_{P4} & 0 \\ 0 & 0 & 0 & \mathrm{c}_4 & \mathrm{c}_5 \end{bmatrix} \qquad \mathbf{C}_{\text{Ac}} = \begin{bmatrix} \mathrm{c}_{P1} & 0 & \mathrm{c}_3{}^4 \\ \mathrm{c}_1 & \mathrm{c}_{P2} & 0 \\ 0 & \mathrm{c}_2 & \mathrm{c}_3 \end{bmatrix} \tag{5}$$

in which c_i represents a scaling function for the rate of increase and c_{P_i} is $[1-(1-P_i)\mathrm{c}_i]/P_i$. Thus, the term c_i adjusts sampling probabilities based on the temperature at which laboratory-conducted assays were performed as $\lambda(T)/\lambda(T_0)$, in which $T_0 = 25^{\circ}\mathrm{C}$. As $T \rightarrow T_{\max}$ (the upper temperature threshold for development), individuals have a decreasing probability of being resampled from t to $t+1$ ($\mathrm{c}_{P_i}P_i < P_i$) and an increasing probability of being sampled in the subsequent developmental stage ($c_iG_i > G_i$), with $\mathrm{c}_{P_i}P_i + c_iG_i = P_i + G_i$. Fertilities ($F_5$ for the host and F_3 for the parasitoid) exhibit the same trend exponentially, with c_i^4 providing the best fit for the matrix approximation of the native function. Our formulation of $\mathbf{C}_{\text{SBA}}$ and $\mathbf{C}_{\text{Ac}}$ was

necessary so that the population growth rate can follow temperature-dependent changes in juvenile development and adult survival and fertility. Direct application of the scaling function, e.g. $\mathbf{p}_{t+1} = \lambda(T)/\lambda(T_0)\mathbf{W}\mathbf{p}_t$, would yield the correct rate of population growth but only by adding or removing individuals from the population in a biologically unrealistic manner; although our formulation is an approximation, it holds from 5–30˚C, which spans the normal range of average historical daily temperatures during the modeling period.

Temperature-dependent development was added for the host following McCornack et al. [39] and for the parasitoid following Frewin et al. [31]. The McCornack et al. [39] model is a modified Logan [40] model that expresses the intrinsic rate of growth, r, as a function of temperature, and incorporates the upper development threshold ($T_{\max}$, the maximum lethal temperature), the range of thermal breakdown (Δ), and a constant ρ so that $r(T) = e^{\rho T} - e^{[\rho T_{\max} - (T_{\max} - T)/\Delta]}$. (For reference, the intrinsic rate of growth r was related to the natural rate of increase λ using the approximation $\lambda = e^r$.) The model used by Frewin et al. [31] was based on an earlier model by Briere et al. [41] (also based on Logan [40]) and estimates the intrinsic rate of growth r given an upper temperature threshold ($T_{\max}$), a lower temperature threshold (T_0), and a constant a so that $r(T) = aT(T - T_0)\sqrt{T_{\max} - T}$. These modified Logan [40] models build on improvements made by Lactin et al. [42] and are advantageous in that they decrease the number of necessary parameters while maximizing their biological relevance; Shi et al. [43] has since proposed a similar model based on physiological mechanisms (enzyme kinetics), but the McCornack et al. [39] and Frewin et al. [31] formulations were selected because they were parameterized for the host and parasitoid species used in our study.

Bioassays

Aphid development. Soybean aphids were observed to determine the amount of time required to reach maturity. Reproducing adult aphids from a mixed-aged colony raised at 25 ± 2˚C, 16:8 L:D, were transferred to the underside of an excised soybean leaflet. After 1.5 hr, 1st-stadium nymphs (n = 31) were transferred with a fine brush to the underside of a fresh excised leaflet from a V1–V2 soybean plant. Leaflets were positioned vertically with the stem placed in 3 cm^3 of moist, fine sand at the bottom of a 6 dram plastic vial that was ventilated by puncturing pinholes through the cap. Individual aphids were identified to developmental stage at 12 hr intervals until reaching reproductive maturity. Although nymphs and adults are visually and functionally similar, developmental stage can be distinguished by unique differences in antennal segmentation and caudal morphology; antennal segmentation increases from four (1st stadium) to five (2nd stadium) to six (3rd stadium and higher), and the caudum characteristically increases in size before tapering into an elongated teardrop shape at adulthood [44]. Additionally, 4th-stadium nymphs often exhibit the eyespots of well-developed embryos that may be seen through their integument. The presence of exuviae and analysis of exuvial antennal segmentation was also used to confirm stage transitions. The entire assay was conducted in a growth chamber at constant 25 ± 2˚C, 16:8 L:D.

Parasitoid development. Parasitoids were evaluated for their capacity to complete development on each of five apterous stages of soybean aphid (1st–4th stadia and adult). Mummies —the darkened exoskeletal remains of recently killed aphids that contain late larval parasitoids or pupae—were collected from laboratory colonies of *A. certus* maintained at 23 ± 2˚C, 16:8 L:D (first established in August 2011 with field-collected mummies from Saint Paul and Rosemount, Minnesota). Mummies were stored individually in 0.6 mL plastic microcentrifuge tubes supplied with a droplet of honey water (approx. 50 vol%). Each newly emerged female parasitoid (< 24 hr old, n = 59) was paired with a newly emerged male and observed for

copulatory behavior; after copulation, the male was discarded and the female left overnight. Each female was randomly assigned a treatment (one of the five host stages), and twenty soybean aphids of representative size and quality for that stage were transferred from a mixed-aged laboratory colony to the underside of a soybean leaflet placed in a plastic vial (as previously described). The aphids were allowed to settle for ten minutes, after which time the parasitoid was introduced into the tube and left to interact with the aphids for 24 hr. The parasitoid was then removed and the aphids were allowed to continue development. Aphids were checked daily for the formation of mummies, which were individually collected in 0.6 mL microcentrifuge tubes and observed at 3–12 hr intervals for the emergence of adult parasitoids. Hind tibia length was measured for a subsample (n = 194) of the emerged offspring as a proxy for size and fitness. The assay was conducted in a growth chamber at 25 ± 2˚C, 16:8 L:D. An ANOVA was used to compare the main effects of host stage, sex, and parental identity on parasitoid developmental time as well as hind tibia length. Tukey's *post hoc* was used to separate means for multiple comparisons. Differences in parasitoid sex ratio in response to different host stages were determined using a linear model with Tukey's *post hoc*, and the response (proportion male) was weighted based on brood size. These analyses were performed using the agricolae package and base R version 3.4.4 (The R Foundation for Statistical Computing, 2017).

Host-stage preference. To determine host-stage preference (defined as the deviation in the proportion of host stages attacked by a single female *A*. *certus* from random chance when all stages present are of equal abundance), *A*. *certus* mummies were collected and mated as before. At the start of the assay, a single female (n = 73) was allowed to exit the microcentrifuge tube onto the underside of a leaflet containing three each of the 1^{st}–4^{th} stadia and apterous adult soybean aphids in a plastic vial (as previously described). Only aphids of visually similar quality and of representative size for their stage were used in the assay. Each parasitoid was allowed to interact with aphids for two hours at 25 ± 2˚C, which provides sufficient time to locate and parasitize approximately one host (Miksanek JR, personal observation). Immediately after parasitoid exposure, aphids were stored in 70% ethanol and later dissected to recover parasitoid eggs. Host-stage preference was determined using the Friedman rank-sum test (the package stats in base R), which follows a χ^2 distribution. Parasitoids that did not oviposit during the assay were excluded from the analysis.

Post-parasitism reproduction. Aphids were assessed for their reproductive capacity following parasitism. A single 3^{rd}, 4^{th}, or adult stadium aphid was collected from the laboratory colony and transferred to the underside of a V1 soybean leaflet, which was situated in a plastic vial as previously described. 1^{st} and 2^{nd} stadium aphids were not included because preliminary testing revealed that these stages do not reproduce prior to mummification. A total of 105 vials were assembled, fifteen for each unparasitized (control) 3^{rd}, 4^{th}, and adult stadium soybean aphid and twenty for each parasitized 3^{rd}, 4^{th}, and adult stadium aphid. Adult *A*. *certus* (n = 60) were aspirated from a two-week old laboratory colony and placed individually into the appropriate vials. (In the colony, individual wasps had the opportunity to mate, acquire host handling experience, and feed on honeydew or host hemolymph, thus they were considered to be reproductively, behaviorally, and nutritionally prepared for the bioassay.) Each parasitoid was allowed to interact with its aphid for 24 hr, after which the parasitoid was removed. Aphids were observed daily for 8 d for the production of offspring, and nymphs were removed with each observation. The assay was performed in a growth chamber held at 25 ± 2˚C, 16:8 L:D.

Reproduction of parasitized and unparasitized adult hosts was analyzed with a cumulative link mixed effects model (CLMM). This approach consists of a multivariate analysis of variance with a logit link function that assesses ordinal response variables while accounting for random factors. The daily number of offspring was the response variable; treatment (parasitized or

control), initial host stage (3rd, 4th, or adult stadium), and number of days since exposure (1–7, a discrete variable) were fixed effects; and individual aphid was included as a random factor to account for repeated measures. An interaction term between treatment and day was included to account for any time-dependent effects of parasitism (i.e. delayed impact on host reproduction). Pairwise comparisons were determined using a *post hoc* test of least square means with a Bonferroni correction. An ANOVA was performed to compare the number of days between molts for parasitized and unparasitized aphids. Aphids that died within the parasitoid exposure period (e.g. due to host feeding or overstinging) were excluded from the analysis, and aphids in the parasitism treatment that did not mummify by the end of the seven-day study period were excluded as well. The CLMM was analyzed using the ordinal package in R, with ten quadrature points used for Gauss-Hermite likelihood approximation. The package emmeans was used as a *post hoc* test for pairwise comparisons of least square means.

Model analysis

Population dynamics. A 90 d period was simulated given a randomly selected initial number of individuals ranging from 0.3–1.82 hosts and 0.15–4.08 parasitoid mummies per plant. These values represent the range of early-season host and parasitoid densities sampled at four sites surveyed during 2017: Hitterdal, MN (47.0˚N, 96.2˚E), Starbuck, MN (45.6˚N, 95.7˚E), Appleton, MN (45.3˚N, 95.9˚E), and Pipestone, MN (44.0˚N, 95.9˚E) (United States). In order to reflect natural conditions, the initial stage structure for the host was juvenile-biased as colonizing soybean aphid alatae deposit a few offspring per plant without themselves settling [45]; the initial parasitoid population was similarly biased towards younger stages. Colonization timeline followed field observations: aphids were introduced on June 22nd and parasitoids were introduced 20 d later. Simulations were conducted in R and replicated 10000 times.

Parasitism and host suppression. Accurate comparisons of field data and ecological models requires clear differentiation of the various methods of measuring parasitism of a host population. *A. certus* and other aphid parasitoids are surveyed during their late larval and pupal stages because, at this point, their host has died, leaving behind a mummy (the darkened exoskeletal remains), which are easily sampled in field settings and identifiable to subfamily or genus [46]. However, the relative abundance of mummies—referred to as *mummy fraction*—is not synonymous with other measures of parasitism. To clarify this terminology, we use *parasitism rate* to denote a temporal unit of measurement expressing an absolute or proportional change in the individuals succumbing to parasitism over time [47]. In contrast, *percent* (or *proportion*) *parasitism* is a unitless measure that compares a subset of hosts (the parasitized) to the larger population at some point in time; percent parasitism is thus the *result* of a specific parasitism rate interacting with other competing rates (birth/death, immigration, dispersal, etc.), following van Driesche [48]. Operating under these definitions and following the format of the matrix model, we define *parasitism rate* as $\sum_{i=1}^{5}[(1-g_i)n_i/N]$, *percent parasitism* as $p_1/N \times 100\%$, and *mummy fraction* as $p_2/(N+p_2)$.

Sensitivity analyses. The influence of adult parasitoid lifespan, date of parasitoid colonization, host-stage preference, and mean daily temperature was evaluated in ecologically plausible parameter space. The effect of these parameters on host population densities was calculated as a percent difference in maximum host population density with and without the parasitoid present ("peak pest reduction"). The sensitivity analysis for adult parasitoid lifespan (uncoupled from host density) evaluated a mean adult parasitoid survival period of 2–26 d. Parasitoid colonization was evaluated from 2–32 d after host establishment. For host-stage preference, a total of 21 graded preferences were assessed, which ranged from a strong preference for early-stage juveniles (a_{1-5} = {0.50, 0.35, 0.10, 0.05, 0.00}), to no overall preference (a_i = 0.20), to a strong

preference for adults (a_{1-5} = {0.00, 0.05, 0.10, 0.35, 0.50}). The effects of mean daily temperature were assessed over a range of ± 3˚C compared to publicly available historical records from the station GHCND:USC00215204 located at (44.4706˚N, 95.7908˚E) in Marshall, MN. With the exception of the manipulated parameter, all parameters were the same as previously described and simulated using median starting densities for the host and parasitoid populations.

Results

Laboratory assays

Aphid development. All aphids successfully reached reproductive maturity within seven days. All adults began reproducing within 24 hours of their final molt, and most produced their first offspring within 12 hours; thus, a significant non-reproductive adult stage (referred to as S5 by Lin and Ives [16]) was not noted in our study.

Parasitoid development. Host stage affected the amount of time required for *A*. *certus* to complete development, both in terms of the time until host mummification ($F_{4,\ 455}$ = 17.23, $p < 0.001$) and time until adult parasitoid emergence ($F_{4,\ 455}$ = 18.87, $p < 0.001$) (Table 1: *Mean time to mummy* and *Mean time to emerge*). Both parasitoid sexes developed more slowly on 1st stadium hosts compared to adult hosts (Table 1: *Total development time*). Males developed more slowly than females ($F_{1,\ 455}$ = 17.38, $p < 0.001$) (Table 1: *Total development time*). The amount of time from mummification to emergence did not differ significantly between sexes ($F_{1,\ 455}$ = 0.06, p = 0.807) (Table 1: *Mean time to emerge*). There was an effect of experimental block on developmental rate (time to mummification: $F_{54,\ 455}$ = 6.52, $p < 0.001$; time to emergence: $F_{54,\ 455}$ = 4.89, $p < 0.001$) and size ($F_{52,\ 136}$ = 1.86, p = 0.002), indicating similarities among offspring of the same parental parasitoid. Offspring reared on adult hosts were smaller than those developing on other stages ($F_{4,136}$ = 8.37, $p < 0.001$), and males tended to be smaller than females ($F_{1,\ 136}$ = 12.66, $p < 0.001$) (Fig 1). A female-biased sex ratio was produced on most host stages; the proportion male was 0.37 ± 0.05, 0.48 ± 004, 0.52 ± 0.04, 0.27 ± 0.6, 0.38 ± 0.6 (mean ± SEM) on 1st, 2nd, 3rd, and 4th stadia and adult hosts, respectively. Although sex ratio varied with host stage ($F_{4,\ 44}$ = 6.49, $p < 0.001$), groups could not be separated *post hoc* by means of Tukey.

Table 1. Development time for *Aphelinus certus* on the various stages of soybean aphid separated by parasitoid sex with two-factor comparisons.

	Host-stage	Mean time to mummy (days ± SE) [a]		Mean time to emerge (days ± SE)[a]		Total development time (days ± SE) [a]		n
♀	1st stadium	6.18 ± .05	a	7.00 ± .07	a	13.18 ± .07	a	72
	2nd stadium	6.00 ± .05	abc	6.43 ± .08	d	12.43 ± .07	b	67
	3rd stadium	6.02 ± .04	abc	6.60 ± .07	bcd	12.61 ± .09	b	57
	4th stadium	6.08 ± .07	ab	6.47 ± .09	cd	12.55 ± .07	b	66
	adult	5.75 ± .07	de	6.76 ± .07	abc	12.51 ± .09	b	51
♂	1st stadium	6.07 ± .06	abc	6.93 ± .06	ab	13.00 ± .06	a	42
	2nd stadium	5.84 ± .07	cde	6.49 ± .09	cd	12.33 ± .07	b	61
	3rd stadium	5.93 ± .05	bcd	6.46 ± .08	cd	12.39 ± .09	b	56
	4th stadium	5.85 ± .11	bcde	6.65 ± .15	abcd	12.50 ± .07	b	20
	adult	5.57 ± .14	e	6.74 ± .13	abcd	12.30 ± .08	b	23
	Pooled:	5.97 ± .02		6.64 ± .03		12.61 ± .03	*Total*:	515

[a]Shared letters indicate no significant difference.

https://doi.org/10.1371/journal.pone.0218217.t001

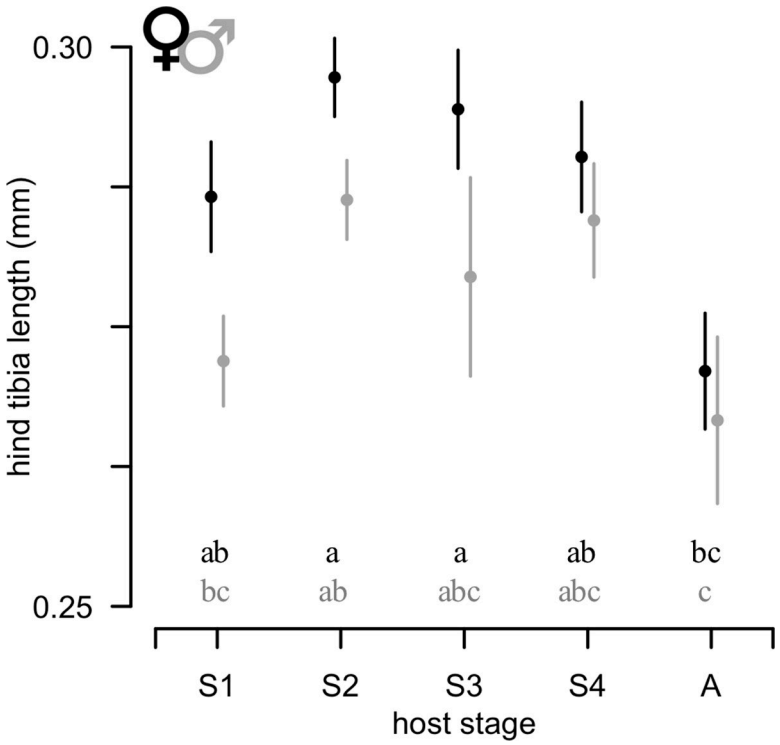


Fig 1. Hind tibia length as a function of host developmental stage. Black: female *A. certus*; gray: male. S1: 1st stadium aphid, S2: 2nd stadium, S3: 3rd stadium, S4: 4th stadium, A: adult. Mean ± SE; shared letters indicate no significant difference.

https://doi.org/10.1371/journal.pone.0218217.g001

Host-stage preference. Parasitoid eggs recovered from dissected aphids were typically located within the anterior abdomen or posterior thorax of the host. Host-stage preference for *A. certus* was 0.21, 0.23, 0.21, 0.17, and 0.19 for 1st–4th stadia and apterous adults respectively, but did not demonstrate a significant deviation in oviposition from random (Friedman test, $F_R = 0.640$, $p = 0.958$).

Post-parasitism reproduction. Parasitism by *A. certus* negatively affected soybean aphid reproduction (CLMM, likelihood ratio $\chi^2_{1,440} = 89.29$, $p < 0.001$) and varied by day ($\chi^2_{6,440} = 33.91$, $p < 0.001$) and with the host stage parasitized ($\chi^2_{2,440} = 45.84$, $p < 0.001$). Additionally, there was an interaction between treatment and day ($\chi^2_{6,440} = 195.44$, $p < 0.001$), indicating that the effect of parasitism on host reproduction changed over time (parasitism-induced changes in fertility did not begin until after the third day). The difference in reproduction between parasitized and control aphids was not statistically significant until four days after parasitism, at which time parasitized aphids were rendered infertile (Fig 2). Parasitism did not influence the amount of time between soybean aphid molts (3rd stadium to 4th stadium: $F_{1,20} = 1.34$, $p = 0.261$; 4th stadium to adult: $F_{1,38} = 0.229$, $p = 0.635$).

Matrix model parameterization

Results from the bioassays were supplemented with data from peer-reviewed literature to parameterize the matrix model. Survival and transition probabilities for soybean aphid nymphs were obtained from the aphid development data, and fertility of parasitized aphids was taken from the post-parasitism reproduction assay. Pooled data from the parasitoid development assay were used determine survival and transition probabilities for immature

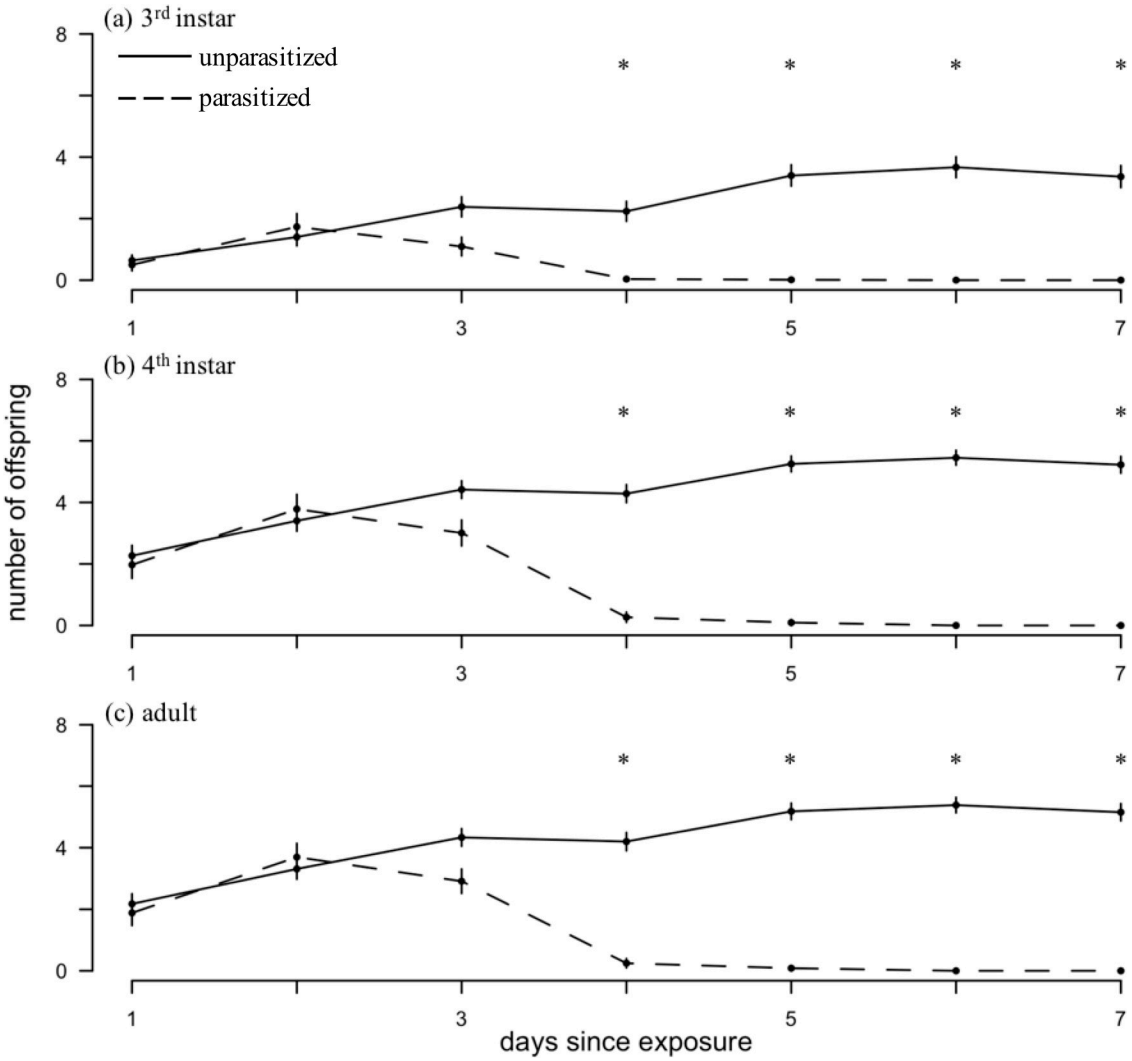


Fig 2. Daily reproduction of parasitized and unparasitized soybean aphids. (a) 3rd stadium, (b) 4th stadium, and (c) adult. Points plot least square means ± SE. Solid lines: unparasitized (control) aphids; dashed lines: parasitized aphids. Asterisks indicate significant differences between control and parasitized aphid reproduction on each day following parasitoid exposure (adjusted for multiple comparisons).

https://doi.org/10.1371/journal.pone.0218217.g002

parasitoids because, even though *A. certus* developed more slowly on 1st stadia hosts, this difference was relatively small. Parasitoid sex ratio data were obtained from the parasitoid development assay, and adult parasitoid survival was calculated using unpublished data (Miksanek JR & Heimpel GE, unpublished). Because *A. certus* did not exhibit a significant host-stage preference, the null hypothesis $a_i = 0.2$ was used in the model. The remaining parameters in the model were obtained from the literature, and all parameters used in the model are summarized in Table 2.

Model analysis

Population dynamics. Soybean aphid densities peaked just before day 45 of the simulation, which corresponds to the last week of July (Fig 3A). In 9.9% of simulations including *A. certus*, soybean aphid densities were below the economic threshold of 250 aphids per plant

Table 2. Values and sources of parameters used in the matrix model.

	Parameter	Symbol(s)	Value(s) (excluding units)	Source(s)
soybean aphid	juvenile survival probabilities	P_1, P_2, P_3, P_4	0.500, 0.143, 0.311, 0.205	bioassay
	adult survival probability	P_5	0.86	[16]
	transition probabilities	G_1, G_2, G_3, G_4	0.484, 0.857, 0.689, 0.795[a]	bioassay
	per capita reproduction	F_5	2.56	bioassay, [16]
	post-parasitism reproduction	J_1, J_2, J_3, J_4, J_5	0, 0, 0.563, 1.521, 1.471	bioassay
	temperature-curve	ρ, T_{max}, Δ	34.9, 7.1, 0.14	[39]
	carrying capacity	K	6000	[49] (lower estimate)
A. certus	egg+larval survival probability[b]	P_1	0.832	bioassay
	pupal survival probability	P_2	0.869	bioassay
	adult survival probability	P_3	0.932	Miksanek & Heimpel, unpub.
	transition probabilities	G_1, G_2	0.168, 0.131	bioassay
	sex ratio	–	0.412	bioassay
	host-stage preference	a_i	0.2 (H_0)	bioassay
	functional response	α, T_h	0.979, 0.045	[31]
	temperature: egg to mummy	a, T_0, T_{max}	1.19×10^{-4}, 7.8, 35.7	[31]
	temperature: mummy to adult	a, T_0, T_{max}	1.37×10^{-4}, 11.6, 36.9	[31]
	host-density-dependent survival	θ_1, θ_2, θ_3,	18.6, 13.5, –0.562	Miksanek & Heimpel, unpub.

[a]There is also a 0.016 probability of sampling a first stadium as a third stadium 24 hours later that was included in the model.

[b]Probability of being resampled as an egg or larva at time $t + 1$ assumes no mortality during this period because egg and larva survival were not measured during the assay.

https://doi.org/10.1371/journal.pone.0218217.t002

(the density at which pest management practices should be applied), and in 31.0% did not exceed the economic injury level of 674 aphids per plant (the density at which yield loss

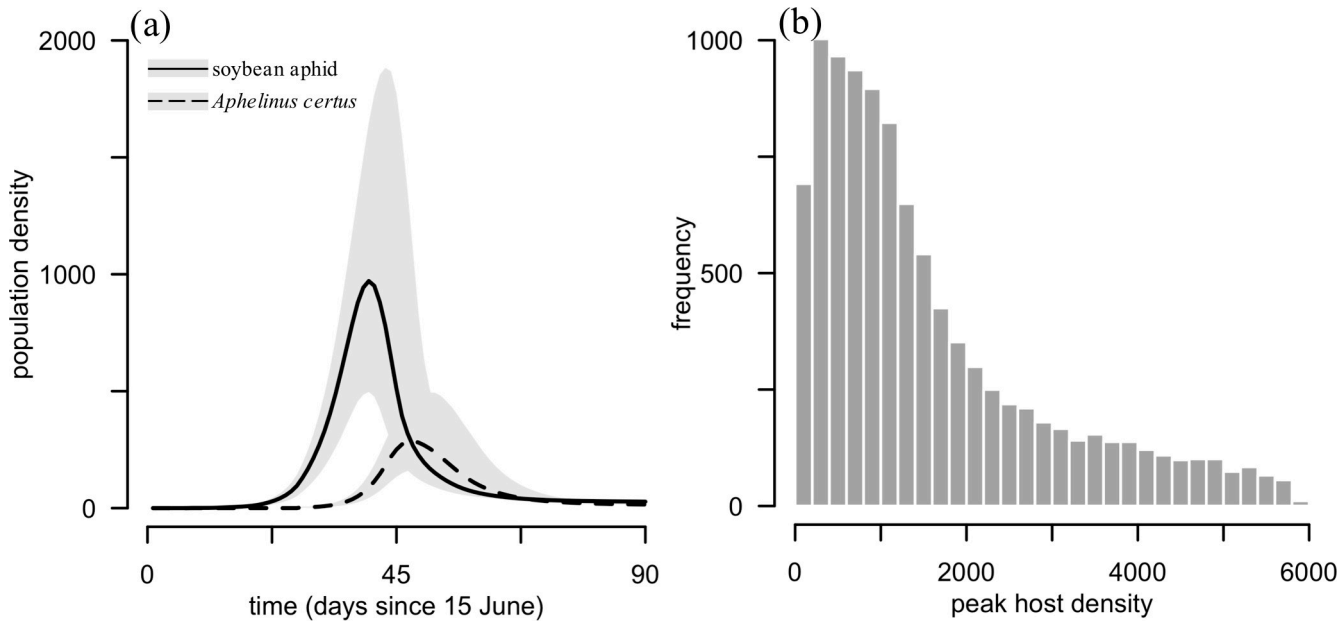


Fig 3. Population dynamics predicted by the matrix model. (a) Population dynamics of soybean aphid and *A. certus*. Black lines represent median densities with the interquartile (Q1–Q3) range shaded in gray. Solid line: soybean aphid (all living aphids); Dashed line: *Aphelinus certus* (all stages). Dotted horizontal line: economic threshold. (b) Histogram of peak aphid densities simulated from the model.

https://doi.org/10.1371/journal.pone.0218217.g003

exceeds management costs) (Fig 3b). Densities in simulations not including *A*. *certus* reached the carrying capacity of 6000 aphids per plant, and there was a 74.2 ± 0.2% decrease in peak aphid abundance in the presence of *A*. *certus*.

Parasitism and host suppression. Because the model was stage-structured for both the host and the parasitoid, parasitism rate at any time throughout the season can be equated with mummy fraction and percent parasitism (Fig 4). The parasitism rate associated with no host population growth ($\lambda = 1$, or the apex of peak aphid density for each of the 10000 simulations) was 0.208 ± 0.012 d^{-1} (mean ± SD), which equates to 11.3 ± 3.7% parasitism or 3.4 ± 1.4% mummies (Fig 4). Regardless of the method of measuring parasitism, parasitism increased with host density before decreasing as the host population declined; however, time-dependent measures of parasitism (percent parasitism and mummy fraction) exhibited a notable lag in comparison with parasitism rate (Fig 4).

Sensitivity analyses. Biological control efficacy of *A*. *certus* was greatest for long-lived parasitoids that colonized fields early and attacked hosts without a strong preference for either younger or older host stages. Longer-lived adult parasitoids had a higher impact on the aphid population, and the slope of this relationship was greatest when parasitoid longevity was less than 10 days (Fig 5a). Delaying the date of parasitoid introduction greatly reduced the effect of *A*. *certus* such that, for parasitoids colonizing fields more than a month after the arrival of soybean aphid, their effect was nearly zero (Fig 5B). An increase in parasitoid preference from younger to older hosts produced a concave response in peak pest reduction, indicating that parasitoids attacking all host stages indiscriminately have the greatest effect on aphid population dynamics (Fig 5C). Additionally, lower temperatures were more conducive to host suppression (Fig 5D). Finally, post-parasitism reproduction had a modest effect on peak

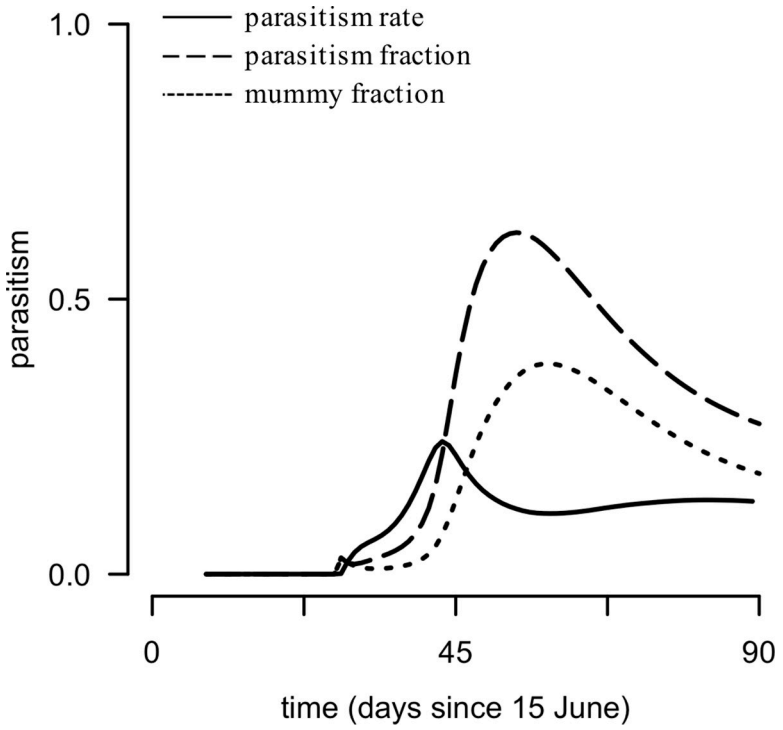


Fig 4. Comparing measures of parasitism in model simulations. Solid line: parasitism rate (d^{-1}); dashed line: parasitism fraction (×100 = percent parasitism [%]); dotted line: mummy fraction.

https://doi.org/10.1371/journal.pone.0218217.g004

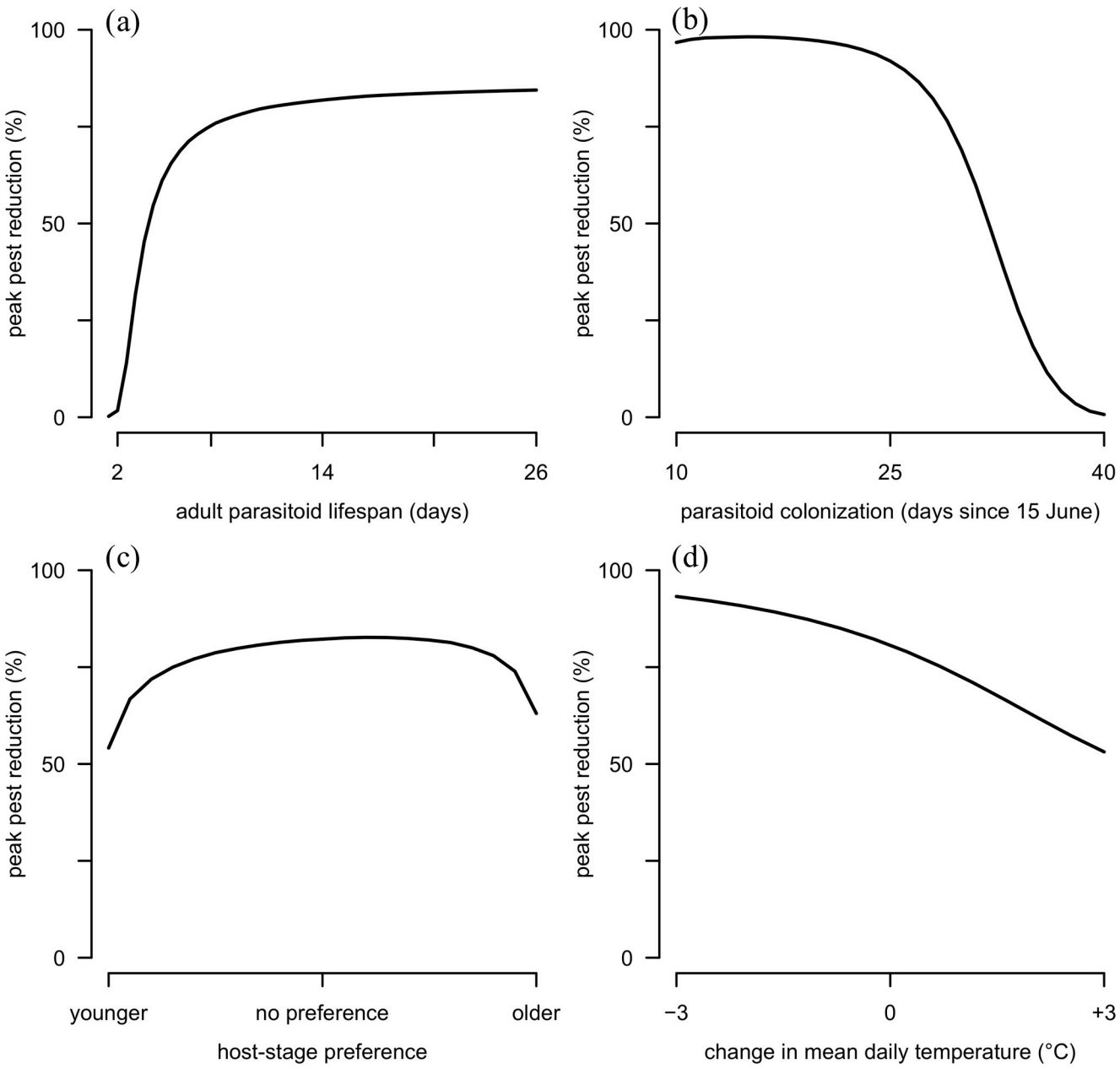


Fig 5. Sensitivity analyses. The effect of the parasitoid is shown as a percent reduction in peak host densities (solid line) as a function of (a) adult parasitoid lifespan, (b) the date of parasitoid colonization, (c) host-stage preference, and (d) mean daily temperature.

https://doi.org/10.1371/journal.pone.0218217.g005

population reduction; exclusion of this term from the model increased peak population reduction by 0.08% (no figure).

Discussion

The developmental bioassays revealed a host with a high capacity for growth and a parasitoid that exhibits optimal growth on intermediate host stages but without a significant preference for any individual stage. The fully parameterized matrix model predicted that *Aphelinus certus* reduces soybean aphid populations below the economic threshold in 9.9% of simulations and

below the economic injury level in 31.0% of simulations. Host suppression was predicted at a parasitism rate of 0.21 d^{-1}, which corresponds with 3.4% of the aphid population being visibly mummified; notably, because parasitism was dynamic, relatively low parasitism rates early in the season could still be associated with low peak host densities. Assuming that the 9.9% modeled reduction in fields exceeding threshold due to parasitism by *A*. *certus* is scalable, then *A*. *certus* might reduce insecticide applications by 1.8 million acres annually, saving $2.43/ha in management costs and contributing to a commensurate reduction in greenhouse gas emissions [22, 50–52]. Our assessment of *A*. *certus* supports the conclusions of Hallett et al. [53] in calculating the value of this parasitoid for implementation in a dynamic action threshold, which would adjust the traditional economic treatment threshold for a pest based on the relative abundance of its natural enemies [54].

Coupled host–parasitoid dynamics

The matrix model described increasing soybean aphid densities that peaked in late July. This pattern is characteristic of soybean aphid population dynamics in North America [45, 55]. In its native range in Asia, soybean aphid exhibits the same early/mid-season peak, although densities are considerably lower overall and midsummer migrations are of decreased importance [56, 57]. In our model, this unimodal pattern of soybean aphid abundance was driven by a single natural enemy, *A*. *certus*. Soybean aphid is limited by a suite of natural enemies in its native range [57, 58], and, in North America, *Harmonia axyridis*, *Coccinella septempunctata*, *Orius insidiosus* (debatably), and *A*. *certus* have been identified as important predators in certain landscapes [33, 53, 59–64].

Our model suggests that *A*. *certus* is capable of suppressing soybean aphid at a parasitism rate of 0.21 d^{-1} (i.e. parasitizing 21% of the total host population per day). This value is consistent with the 20–30% total daily parasitism range required for soybean aphid population suppression previously determined by Lin and Ives [16], but was relatively low in comparison to the field-estimated 42% parasitism rate proposed by Kaser and Heimpel [33]. This discrepancy may be due to different methods of analysis. The matrix model was analyzed using a non-equilibrium approach, and as a result, our model was able to show that parasitism rate fluctuates dynamically in response to aphid population densities, in which high mid-season parasitism rates followed low early-season parasitism, which contributed to an overall increase in percent parasitism over time. This time difference between increasing parasitism rates associated with host suppression and percent parasitism suggests that it may be difficult to identify the impact of *A*. *certus* in field settings until the pest population is already in decline.

Insights into host–parasitoid dynamics

The sensitivity analyses demonstrated that adult parasitoid lifespan, date of parasitoid colonization, host-stage preference, mean daily temperature, and post-parasitism reproduction all affect peak host densities to some degree, but the ways in which host-stage preference and post-parasitism reproduction influenced the system merit further discussion. Interestingly, we did not find host suppression to be at a maximum when parasitoids preferentially attacked the oldest host stages and we did not identify post-parasitism reproduction as a mechanism of notably increasing peak population densities, both of which are contrary to the results of Lin and Ives [16].

Host-stage preference. The parasitoid *A*. *certus* did not show a significant preference for any individual host developmental stage. Although many *Aphelinus* species readily accept all host stages, there is broad variability in host-stage preference [65–68]. While Lin and Ives [16] showed that preference for older host stages produces the lowest equilibrium host densities,

we found that the relationship between host-stage preference and peak host densities produced a different result in our nonequilibrium analysis. When preference for younger hosts became less pronounced, peak aphid densities began decreasing, which is consistent with parasitoids removing hosts of a higher reproductive value [16]. However, as preference for older individuals continued increasing, peak aphid densities began to rise again. In our model, the initial aphid population consists predominately of immature host stages, which mimics the conditions imposed by colonizing alate aphids at the beginning of the season [45]; thus, the adults are much scarcer, so parasitoid preference for late-stage hosts suggests that parasitism rate will be low until their relative abundance eventually increases and the host population approaches its stable stage structure.

Post-parasitism reproduction. Soybean aphid reproduction was decreased 72 hr after parasitism by *A*. *certus*, and soybean aphids were reproductively dead the following day. Compared to parasitism by the aphidiine *Aphidius colemani* [16], soybean aphids parasitized by *A*. *certus* reproduce a full day longer, but compared to parasitism by the aphidiine *Binodoxys communis* [69], post-parasitism reproduction was similar. Aphid parasitoids decrease host reproduction when they compete with developing host embryos for nutritional resources [70], which indirectly leads to embryonic degeneration via starvation [71]. Additionally, parasitoids influence the fertility of their hosts by venomous castration [72] or by directly feeding on embryos [71]. In response, parasitized aphids may allocate additional resources to any surviving embryos [69]. Lin and Ives [16] showed that continued reproduction by parasitized aphids during the early stages of parasitoid development produces a partially compensatory effect that leads to higher population growth rates compared with non-reproducing parasitized hosts, and this compensation may be high enough that parasitoids attacking adult aphids—especially older adults—do not affect the maximum growth rate (r_m) or doubling time of their host populations [73, 74]. However, our model did not indicate a strong effect of post-parasitism reproduction on peak soybean aphid densities. Instead, our analysis supports the hypothesis that total lifetime reproduction of aphids has little impact on population growth rates and that the reproductive output during early adulthood contributes disproportionately to population growth (e.g. van Steenis and El-Khawass [75] and references therein). As a result, preference for the oldest host stages allows for high survival for mid- to late-stage immature hosts, which then mature and begin reproducing before succumbing to parasitism themselves.

An alternative modeling approach

A different approach to modeling herbivorous pest species of annual crops—aphids in particular—involves a linear decline in the intrinsic rate of growth, *r*, due to bottom-up effects of decreasing plant quality as a result of plant phenology; this approach is termed the *decreasing* r *model* [76, 77]. The decreasing *r* model produces a distinctive bell-shaped population curve defined as $N_t = N_0 e^{r_{\max} t(1-0.5at)}$, in which *N* is aphid density, $r_{\max}$ is the maximum rate of population growth, *t* is time, and *a* is the rate of decline for *r*. Decreasing *r* was field-validated for soybean aphid by Costamagna et al. [77] and applied to a host–parasitoid system by Leblanc and Brodeur [34]. Both studies reported a high degree of success using this bottom-up model to describe population dynamics in the field even though soybean aphid dynamics have been previously linked to the strong top-down effect of predation [63].

Decreasing *r* may be incorporated into a matrix model as $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t\lambda^{-at}$, in which **n** is the aphid population vector, **A** is the transition and fertility matrix, and λ is the dominant eigenvalue of the matrix **A** representing the natural rate of population increase. Analysis of the matrix model (as described in *Materials and Methods*: *The matrix model*) with the addition of decreasing *r* (in which $a = 0.0247$ per Costamagna et al. [77]) predicts that peak soybean aphid

densities are reduced by 49.6 ± 0.2% in the presence of *A*. *certus* and, even in the absence of *A*. *certus*, do not exceed the economic injury level, suggesting that soybean aphid might be unlikely to be considered an economically damaging species in any scenario. Regardless, the biggest challenge to the decreasing *r* model in general is that it oversimplifies aphid population dynamics by imposing a season-long effect of plant phenology that confounds density-dependent effects of the aphid on host plant quality with density- and time-dependent changes in aphid behavior and physiology such as emigration, mid-summer migration, or a parthenogenic shift and migration to the primary host plant [78, 79]. Additionally, host population dynamics can be affected by hyperparasitoids and other higher-order natural enemies if, for example, they trigger avoidance behaviors in primary parasitoids or signal hosts of a reduced risk of parasitism, leading to increased reproduction (reviewed by Frago [80]). The overall course of soybean aphid colonization and growth throughout the season is also influenced by landscape-level resource availability, such as proximity to buckthorn or agricultural intensification [53, 81–84].

Final remarks

Our study highlighted the value of including host stage-specific parameters as well as parasitoid lifespan and colonization timeline in host–parasitoid population models. We also showed a negligible effect of post-parasitism reproduction on peak host densities, and that relatively low parasitism rates early in the season could maintain peak host densities below the economic injury level during the mid-season. Although there have been successes applying real-time monitoring protocols to assess the influence of natural enemies on pest population dynamics and adjust the economic thresholds accordingly (e.g. Hoffmann et al. [85]), such programs can face challenges in development and implementation and are not currently recommended for soybean aphid in the United States [86].

Acknowledgments

We would like to thank Dave Andow, Anthony Ives, Robert Koch, Mary Marek-Spartz, Robert Venette, and two anonymous reviews for volunteering their time to provide feedback during various stages of this work, and to Jonathan Dregni for feedback and maintenance of laboratory colonies of *Aphelinus certus* and soybean aphid. We are also grateful for Henry Davis and Komala Kanagala, who were essential in performing day-to-day laboratory operations.

Author Contributions

Conceptualization: James Rudolph Miksanek, George E. Heimpel.

Data curation: James Rudolph Miksanek.

Formal analysis: James Rudolph Miksanek.

Funding acquisition: George E. Heimpel.

Methodology: James Rudolph Miksanek.

Writing – original draft: James Rudolph Miksanek.

Writing – review & editing: James Rudolph Miksanek, George E. Heimpel.

References

1. Murdoch WW, Briggs CJ, Nisbet RM. Consumer–Resource Dynamics. Princeton: Princeton University Press; 2003.

2. Hassell MP. Host–parasitoid population dynamics. J Anim Ecol. 2000; 69:543–566.
3. Heimpel GE, Mills NJ. Biological Control: Ecology and Applications. Cambridge: Cambridge University Press; 2017.
4. Hochberg ME, Ives AR. Parasitoid Population Biology. Princeton: Princeton University Press; 2000.
5. Ryan RB. Evaluation of biological control: Introduced parasites of larch casebearer (Lepidoptera: Coleophoridae) in Oregon. Environ Entomol. 1990; 19:1873–1881.
6. Bellows Jr. TS, Van Driesche RG, Elkinton JS. Life-table construction and analysis in the evaluation of natural enemies. Annu Rev Entomol. 1992; 37:587–614.
7. Leslie PH. On the use of matrices in certain population mathematics. Biometrika. 1945; 33:183–212. https://doi.org/10.1093/biomet/33.3.183 PMID: 21006835
8. Birch LC. The intrinsic rate of natural increase of an insect population. J Anim Ecol. 1948; 17:15–26.
9. Lefkovitch L. The study of population growth in organisms grouped by stages. Biometrics. 1965; 21:1–18.
10. Caswell H. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd ed. Sunderland: Sinaur Associates, Inc. Publishers; 2001
11. Stark JD, Banks JE, and Acheampong S. Estimating susceptibility of biological control agents to pesticides: Influence of life history strategies and population structure. Biol Control. 2004; 29:392–398.
12. Evans JA, Davis AS, Raghu S, Ragavendran A, Landis DA, Schemske DW. The importance of space, time, and stochasticity to the demography and management of *Alliaria petiolata*. Ecol Appl. 2012; 22:1497–1511. PMID: 22908709
13. Meisner MH, Harmon JP, Ives AR. Temperature effects on long-term population dynamics in a parasitoid–host system. Ecol Monogr. 2014; 84:457–476.
14. Davis AS, Landis DA, Nuzzo V, Blossey B, Gerber E, Hinz HL. Demographic models inform selection of biocontrol agents for garlic mustard (*Alliaria petiolata*). Ecol Appl. 2006; 16:2399–2410. PMID: 17205913
15. Rand TA, Richmond CE, Dougherty ET. Using matrix population models to inform biological control management of the wheat stem sawfly, *Cephus cinctus*. Biol Control. 2017; 109:27–36.
16. Lin LA, Ives AR. The effect of parasitoid host-size preference on host population growth rates: An example of *Aphidius colemani* and *Aphis glycines*. Ecol Entomol. 2003; 28:542–550.
17. Mills N. Selecting effective parasitoids for biological control introductions: Codling moth as a case study. Biol Control. 2005; 34:274–282.
18. Nicholson AJ, Bailey VA. The Balance of Animal Populations.—Part I. Proc Zool Soc London. 1935; 105:551–598.
19. Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, Desneux N. Ecology and management of the soybean aphid in North America. Annu Rev Entomol. 2011; 56:375–399. https://doi.org/10.1146/annurev-ento-120709-144755 PMID: 20868277
20. Tilmon KJ, Hodgson EW, O'Neal ME, Ragsdale DW. Biology of the soybean aphid, *Aphis glycines* (Hemiptera: Aphididae) in the United States. J Integr Pest Manag. 2011; 2:A1–A7.
21. Heimpel GE, Ragsdale DW, Venette R, Hopper KR, O'Neil RJ, Rutledge CE, Wu Z. Prospects for importation biological control of the soybean aphid: Anticipating potential costs and benefits. Ann Entomol Soc Am. 2004; 97:249–258.
22. Heimpel GE, Yang Y, Hill JD, Ragsdale DW. Environmental consequences of invasive species: Greenhouse gas emissions of insecticide use and the role of biological control in reducing emissions. PloS One 2013; 8:e72293. https://doi.org/10.1371/journal.pone.0072293 PMID: 23977273
23. Chacón JM, Landis D, Heimpel GE. Potential for biotic interference of a classical biological control agent of the soybean aphid. Biol Control. 2008; 46:216–225.
24. Chacón JM, Heimpel GE. Density-dependent intraguild predation of an aphid parasitoid. Oecologia. 2010; 164:213–220. https://doi.org/10.1007/s00442-010-1611-7 PMID: 20369258
25. Asplen MK, Wyckhuys KAG, Heimpel GE. Parasitism of autumnal morphs of the soybean aphid (Hemiptera: Aphididae) by *Binodoxys communis* (Hymenoptera: Braconidae) on buckthorn. Ann Entomol Soc Am. 2011; 104:935–944.
26. Gariepy V, Boivin G, Brodeur J. Why two species of parasitoids showed promise in the laboratory but failed to control the soybean aphid under field conditions. Biol Control. 2015; 80:1–7.
27. Heimpel GE, Asplen MK. A "Goldilocks" hypothesis for dispersal of biological control agents. BioControl. 2011; 56:441–450.
28. Asplen MK, Chacón JM, Heimpel GE. Sex-specific dispersal by a parasitoid wasp in the field. Entomol Exp Appl. 2016; 159:252–259.

29. Kaser JM. Risk and efficacy in biological control: An evaluation of the aphid parasitoid *Aphelinus certus* in North America. PhD Dissertation, University of Minnesota. 2016.
30. Hopper KR, Lanier K, Rhoades JH, Coutinot D, Mercadier G, Ramualde N, et al. Host specificity of *Aphelinus* species considered for introduction to control *Diuraphis noxia*. Biol Control. 2017; 107:21–32.
31. Frewin AJ, Xue Y, Welsman JA, Broadbent AB, Schaafsma AW, Hallett RH. Development and parasitism by *Aphelinus certus* (Hymenoptera: Aphelinidae), a parasitoid of *Aphis glycines* (Hemiptera: Aphididae). Environ Entomol. 2010; 39:1570–1578. https://doi.org/10.1603/EN09312 PMID: 22546454
32. Heimpel GE, Frelich LE, Landis DA, Hopper KR, Hoelmer KA, Sezen Z, Asplen MK, Wu K. European buckthorn and Asian soybean aphid as components of an extensive invasional meltdown in North America. Biol Invasions. 2010; 12:2913–2931.
33. Kaser JM, Heimpel GE. Impact of the parasitoid *Aphelinus certus* on soybean aphid populations. Biol Control, 2018; 127:17–24.
34. Leblanc A, Brodeur J. Estimating parasitoid impact on aphid populations in the field. Biol Control. 2018; 119:33–42.
35. Allen LJS. A density-dependent Leslie matrix model. Math Biosci. 1989; 95:179–187. PMID: 2520184
36. Jensen AL. Simple density-dependent matrix model for population projection. Ecol Modell. 1995; 77:43–48.
37. Jensen AL. Matrix population model with density-dependent recruitment for assessment of age-structured wildlife populations. Bull Math Biol. 1997; 59:255–262.
38. Jensen AL. Sex and age structured matrix model applied to harvesting a white tailed deer population. Ecol Modell. 2000; 128:245–249.
39. McCornack BP, Ragsdale DW, Venette RC. Demography of soybean aphid (Homoptera: Aphididae) at summer temperatures. J Econ Entomol. 2004; 97:854–861. https://doi.org/10.1603/0022-0493(2004)097[0854:DOSAHA]2.0.CO;2 PMID: 15279264
40. Logan JA, Wollkind DJ, Hoyt SC, Tanigoshi LK. An analytic model for description of temperature dependent rate phenomena in arthropods. Environ Entomol. 1976; 5:1133–1140.
41. Briere JF, Pracros P, Le Roux AY, Pierre JS. A novel rate model of temperature-dependent development for arthropods. Environ Entomol. 1999; 28:22–29.
42. Lactin DJ, Holliday NJ, Johnson DL, Craigen R. Improved rate model of temperature-dependent development by arthropods. Environ Entomol. 1995; 24:68–75.
43. Shi P, Ge F, Sun Y, Chen C. A simple model describing the effect of temperature on insect developmental rate. J Asia-Pacific Entomol. 2011; 14:15–20.
44. Hodgson EW, Venette RC, Abrahamson M, Ragsdale DW. Alate production of soybean aphid (Homoptera: Aphididae) in Minnesota. Environ Entomol. 2005; 34:1456–1463.
45. Ragsdale DW, Voegtlin DJ, O'Neil RJ. Soybean aphid biology in North America. Ann Entomol Soc Am. 2004; 97:204–208.
46. Powell W. The identification of hymenopterous parasitoids attacking cereal aphids in Britain. Syst Entomol. 1982; 7:465–473.
47. Post van der Burg M, Powell LA, and Tyre AJ. Modeling parasitism rate and parasitism risk: An illustration using a colonially nesting songbird, the red-winged blackbird Agelaius phoeniceus. J Avian Biol. 2009; 40:263–270.
48. Van Driesche RG. Meaning of "percent parasitism" in studies of insect parasitoids. Environ Entomol. 1983; 12:1611–1622.
49. Catangui MA, Beckendorf EA, Riedell WE. Soybean aphid population dynamics, soybean yield loss, and development of stage-specific economic injury levels. Agron J. 2009; 101:1080–1092.
50. Ragsdale DW, McCornack BP, Venette RC, Potter BD, MacRae IV, Hodgson EW, O'Neal ME, Johnson KD, O'Neil RJ, DiFonzo CD, Hunt TE, Glogoza PA, Cullen EM. Economic threshold for soybean aphid (Hemiptera: Aphididae). J Econ Entomol. 2007; 100:1258–1267. https://doi.org/10.1603/0022-0493(2007)100[1258:ETFSAH]2.0.CO;2 PMID: 17849878
51. United States Department of Agriculture National Agricultural Statistics Service. 2015 Agricultural Chemical Use Survey. 2015.
52. Landis DA, Gardiner MM, van der Werf W, Swinton SM. Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes. Proc Natl Acad Sci. 2008; 105:20552–20557. https://doi.org/10.1073/pnas.0804951106 PMID: 19075234
53. Hallett RH, Bahlai CA, Xue Y, Schaafsma AW. Incorporating natural enemy units into a dynamic action threshold for the soybean aphid, *Aphis glycines* (Homoptera: Aphididae). Pest Manag Sci. 2014; 70:879–888. https://doi.org/10.1002/ps.3674 PMID: 24214819

54. Zhang W, Swinton SM. Incorporating natural enemies in an economic threshold for dynamically optimal pest management. Ecol Modell. 2009; 220:1315–1324.

55. Bahlai CA, Sikkema S, Hallett RH, Newman J, Schaafsma AW. Modeling distribution and abundance of soybean aphid in soybean fields using measurements from the surrounding landscape. Environ Entomol. 2010; 39:50–56. https://doi.org/10.1603/EN09127 PMID: 20146839

56. Liu J, Wu K, Hopper KR, Zhao K. Population dynamics of *Aphis glycines* (Homoptera: Aphididae) and its natural enemies in soybean in northern China. Ann Entomol Soc Am. 2004; 97:235–239.

57. Wu Z, Schenk-Hamlin D, Zhan W, Ragsdale DW, Heimpel GE. The soybean aphid in China: A historical review. Ann Entomol Soc Am. 2004; 97:209–128.

58. Wang C, Xiang N, Chang G, Zhu H. Studies on the soybean aphid, *Aphis glycines* Matsumura. Acta Entomol Sin. 1962; 11:31–43.

59. Rutledge CE, O'Neil RJ, Fox TB, Landis DA. Soybean aphid predators and their use in integrated pest management. Ann Entomol Soc Am. 2004; 97:240–248.

60. Rutledge CE, O'Neil RJ. *Orius insidiosus* (Say) as a predator of the soybean aphid, *Aphis glycines* Matsumura. Biol Control. 2005; 33:56–64.

61. Costamagna AC, Landis DA. Predators exert top-down control of soybean aphid across a gradient of agricultural management systems. Ecol Appl. 2006; 16:1619–1628. PMID: 16937822

62. Desneux N, O'Neil RJ, Yoo HJS. Suppression of population growth of the soybean aphid, *Aphis glycines* Matsumura, by predators: The identification of a key predator and the effects of prey dispersion, predator abundance, and temperature. Environ Entomol. 2006; 35:1342–1349.

63. Brosius TR, Higley LG, Hunt TE. Population dynamics of soybean aphid and biotic mortality at the edge of its range edge. J Econ Entomol. 2007; 100:1268–1275. https://doi.org/10.1603/0022-0493(2007)100[1268:PDOSAA]2.0.CO;2 PMID: 17849879

64. Donaldson JR, Myers SW, Gratton C. Density-dependent responses of soybean aphid (*Aphis glycines* Matsumura) populations to generalist predators in mid to late season soybean fields. Biol Control. 2007; 43:111–118.

65. Gerling D, Roitberg BD, Mackauer M. Instar-specific defense of the pea aphid, *Acyrthosiphon pisum*: Influence on oviposition success of the parasite *Aphelinus asychis* (Hymenoptera: Aphelinidae). J Insect Behav. 1990; 3:501–514.

66. Mueller TF, Blommers LHM, Mols PJM. Woolly apple aphid (*Eriosoma lanigerum* Hausm., Hom., Aphidae) parasitism by *Aphelinus* mali Hal. (Hym., Aphelinidae) in relation to host stage and host colony size, shape, and location. J Appl Entomol. 1992; 114:143–154.

67. Tang YQ, Yokomi RK. Biology of Aphelinus spiraecolae (Hymenoptera: Aphelinidae), a parasitoid of the spirea aphid (Homoptera: Aphididae). Environ Entomol. 1996; 25:519–523.

68. Bayram Y, Bayhan E. Life table of *Aphelinus paramali* Zehavi & Rosen, 1989 (Hymenoptera: Aphelinidae), a parasite of the melon aphid, *Aphis gossypii* Glover, 1877 (Hemiptera, Aphididae). Acta Biol Turc. 2016; 29:111–123.

69. Kaiser MC, Heimpel GE. Parasitoid-induced transgenerational fecundity compensation in an aphid. Entomol Exp Appl. 2016; 159:197–206.

70. Colinet H, Salin C, Boivin G, Hance TH. Host age and fitness-related traits in a koinobiont aphid parasitoid. Ecol Entomol. 2005; 30:473–479.

71. Polaszek A. The effects of two species of hymenopterous parasitoid on the reproductive system of the pea aphid, *Acyrthosiphon pisum*. Entomol Exp Appl. 1986; 40: 285–292.

72. Digilio MC, Isidoro N, Tremblay E, Pennacchio F. Host castration by *Aphidius ervi* venom proteins. J Insect Physiol. 2000; 46(6):1041–1050. PMID: 10802117

73. Campbell A, Mackauer M. The effect of parasitism by Aphidius smithi (Hymenoptera: Aphidiidae) on reproduction and population growth of the pea aphid (Homoptera: Aphididae). Can Entomol. 1975; 107:919–926.

74. Tang YQ, Yokomi RK. Effect of parasitism by Aphelinus spiraecolae (Hymenoptera: Aphelinidae) on development and reproduction of spirea aphid (Homoptera: Aphididae). Biol Control. 1996; 25:703–707.

75. van Steenis MJ, El-Khawass KAMH. Life history of *Aphis gossypii* on cucumber: Influence of temperature, host plant and parasitism. Entomol Exp Appl. 1995; 76:121–131.

76. Williams IS, van der Werf W, Dewar AM, Dixon AFG. Factors affecting the relative abundance of two coexisting aphid species on sugar beet. Agric For Entomol. 1999; 1:119–125.

77. Costamagna AC, Van der Werf W, Bianchi FJJA, Landis DA. An exponential growth model with decreasing r captures bottom-up effects on the population growth of *Aphis glycines* Matsumura (Hemiptera: Aphididae). Agric For Entomol. 2007; 9:297–305.

78. Lamb RJ, MacKay PA. Seasonal dynamics of a population of the aphid Uroleucon rudbeckiae (Hemiptera: Aphididae): Implications for population regulation. Can Entomol. 2017; 149:300–314.

79. Ríos Martínez AF, Costamagna AC. Effects of crowding and host plant quality on morph determination in the soybean aphid, *Aphis glycines*. Entomol Exp Appl. 2018; 166:53–62.

80. Frago E. Interactions between parasitoids and higher order natural enemies: Intraguild predation and hyperparasitoids. Curr Opin Insect Sci. 2016; 14:81–86. https://doi.org/10.1016/j.cois.2016.02.005 PMID: 27436651

81. Gardiner MM, Landis DA, Gratton C, DiFonzo CD, O'Neal M, Chacon JM, Wayo MT, Schmidt NP, Mueller EE, Heimpel GE. 2009. Landscape diversity enhances biological control of an introduced crop pest in the north-central USA. Ecol Appl. 2009; 19:143–154. PMID: 19323179

82. Noma T, Gratton C, Colunga-Garcia M, Brewer MJ, Mueller EE, Wyckhuys KAG, Heimpel GE, O'Neal ME. Relationship of soybean aphid (Hemiptera: Aphididae) to soybean plant nutrients, landscape structure, and natural enemies. Environ Entomol. 2010; 39:31–41. https://doi.org/10.1603/EN09073 PMID: 20146837

83. Liere H, Kim TN, Werling BP, Meehan TD, Landis DA, Gratton C. Trophic cascades in agricultural landscapes: indirect effects of landscape composition on crop yield. Ecol Appl. 2015; 25:652–661. PMID: 26214911

84. Whitney KS, Meehan TD, Kucharik CJ, Zhu J, Townsend PA, Hamilton K, Gratton C. Explicit modeling of abiotic and landscape factors reveals precipitation and forests associated with aphid abundance. Ecol Appl. 2016; 26:2600–2610.

85. Hoffmann MP, Wilson LT, Zalom FG, Hilton RJ. Parasitism of Heliothis zea (Lepidoptera: Noctuidae) eggs: Effect on pest management decision rules for processing tomatoes in the Sacramento Valley of California. Environ Entomol. 1990; 19:753–763.

86. Koch RL, Potter BD, Glogoza PA, Hodgson EW, Krupke CH, Tooker JF, et al. Biology and economics of recommendations for insecticide-based management of soybean aphid. Plant Health Prog. 2016; 17:265–269.

ML 2015, Ch 76, Art. 2, Sec. 6a Project Abstract
For the Period Ending December 31, 2021

PROJECT TITLE: MITPPC #4: Decreasing environmental impacts of soybean aphid management
PROJECT MANAGER: Robert Koch
AFFILIATION: University of Minnesota
MAILING ADDRESS: 1980 Folwell Ave.
CITY/STATE/ZIP: St. Paul, MN 55108
PHONE: 612-624-6771
E-MAIL: koch0125@umn.edu
WEBSITE: https://entomology.umn.edu/people/robert-koch
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER. ML 2015, Ch. 76, Art. 2, Sec. 6a.
APPROPRIATION AMOUNT: $570,000
AMOUNT SPENT: $570,000
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results
Management of soybean aphid relies on applications of broad-spectrum insecticides. This work aimed to decrease insecticide use and ameliorate associated environmental impacts through development of aphid-resistant soybean and advancement of remote scouting.

Overall Project Outcome and Results
The invasion of US soybean by the soybean aphid resulted in dramatic increases in insecticide use, which has increased production costs for farmers and environmental and human-health risks. This proposal takes a two-pronged approach (preventative and therapeutic) to improve management of the soybean aphid through decreased insecticide input, which will result in increased environmental and economic sustainability of soybean production. Integration of preventative and therapeutic pest management tactics is fundamental to integrated pest management (IPM). For preventative management, we advanced the development and availability of aphid-resistant soybean. This included advancement of numerous resistant soybean lines already in the soybean breeding pipeline, including commercial release of one line. Furthermore, numerous crosses were made to incorporate different combinations of aphid-resistance genes into soybean lines, and to test and advance them through the pipeline. Novel research was also performed to examine the variability in susceptibility of aphid populations to these aphid resistant lines. For therapeutic management, we advanced the ability to use remote sensing for soybean aphid through a series of field experiments and technological advancements. Through caged experiments and open-field experiments, we documented that aphid-induced stress to soybean can be detected from drone-based sensors. In addition, through additional caged experiments we found that typical levels of defoliation (<5%) from another insect, the Japanese beetle, is unlikely to affect the ability to scout for soybean aphid; however, higher levels of defoliation (>33%) could impact scouting for soybean aphid. In addition, we built hardware to host new algorithms for autopilots used to guide small drones for accurate and safe pest management missions. We have tested the algorithm in simulation and by post-processing data collected from flight tests. These advancements will help farmers prevent soybean aphid outbreaks through the use of aphid-resistant soybean and to more effectively respond to outbreaks through efficient drone-based scouting.

Project Results Use and Dissemination
An aphid-resistant variety stemming from the work has become commercially available. Results of this project have been actively disseminated to stakeholders and the scientific community. Project results were shared in extension presentations to farmers and agricultural professionals throughout the life of this project and a video was created for stakeholders. A publication for stakeholders listed available resistant soybean varieties. Updates on this work were also shared at several scientific conferences. This work has led to scientific publications on

remote sensing applications and technology (2019, 2020, 2021), and aphid-resistant soybean, and led to detection of a new soybean pest.

JOURNAL OF AIRCRAFT
Vol. 56, No. 4, July–August 2019

Observability and Performance Analysis of a Model-Free Synthetic Air Data Estimator

Kerry Sun,* Christopher D. Regan,† and Demoz Gebre-Egziabher‡
University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455

DOI: 10.2514/1.C035290

The performance, accuracy, and observability of a model-free angle of attack and angle of sideslip estimator are presented. The estimator does not require an aircraft dynamic model; rather, it only relies on measurements from a GPS receiver, an inertial measurement unit, and a pitot tube. The estimator is an inertial navigation system (INS)/GPS extended Kalman filter augmented with the states to account for wind and an additional measurement from a pitot tube. It is shown that the estimator is conditionally observable. Conditions (maneuvers) that enhance its observability are identified. A bound on the angle of attack and angle of sideslip estimate uncertainties is derived. The effect of INS/GPS, horizontal and vertical wind uncertainty on the accuracy of angle of attack and angle of sideslip estimate is assessed. Simulation and flight-test results of the method are presented. The results show that the $1-\sigma$ bound on a small, slow-flying unmanned aerial vehicle for angle of attack and sideslip angle estimates are about 5 and 3 deg, respectively.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

Nomenclature

$\boldsymbol{A}_{\alpha\beta}$	=	flow angle sensitivity matrix
a_x, a_y, a_z	=	accelerometer components
b_{ax}, b_{ay}, b_{az}	=	gyro biases
b_{gx}, b_{gy}, b_{gz}	=	accelerometer biases
C_b^n	=	direction cosine matrix from body frame to north–east–down frame
$\mathcal{G}_d$	=	discrete observability gramian
$\boldsymbol{H}$	=	linearized Jacobian measurement matrix
O	=	observability matrix
$\boldsymbol{P}$	=	state error covariance matrix
P_N, P_E, P_D	=	components of north–east–down position, m
$\boldsymbol{P}_{\alpha\beta}$	=	flow angle error covariance matrix
$\boldsymbol{R}$	=	measurement noise matrix
u, v, w	=	components of airspeed velocity in the body frame, m/s
u_g, v_g, w_g	=	components of airspeed velocity in the north–east–down frame, m/s
V_a	=	airspeed magnitude, m/s
$\boldsymbol{V}_a$	=	airspeed vector
V_N, V_E, V_D	=	components of inertial velocity, m/s
W_N, W_E, W_D	=	components of wind velocity, m/s
$\boldsymbol{w}$	=	process noise vector
$\boldsymbol{x}$	=	state vector
$\boldsymbol{y}$	=	measurement vector
α, β	=	angle of attack and angle of sideslip
$\boldsymbol{\eta}$	=	measurement noise vector
κ	=	condition number
$\sigma_{(\cdot)}$	=	standard deviation
$\tau_{(\cdot)}$	=	correlation time constant (a first-order Gauss–Markov process)
$\boldsymbol{\Phi}$	=	state-transition matrix
ϕ, θ, ψ	=	roll, pitch, and yaw angles

Subscripts

ad	=	accelerometer dynamic (time-varying) output error
aw	=	accelerometer white (uncorrelated) output error
gd	=	gyro dynamic (time-varying) output error
gw	=	gyro white (uncorrelated) output error
m	=	measured quantity

Superscripts

b	=	quantity in the body frame
L	=	lower bound
n	=	quantity in the north–east–down frame
U	=	upper bound

I. Introduction

AN AIRCRAFT'S performance and safety in flight depend, in part, on the magnitude of airspeed V_a and the orientation of the aircraft, as quantified by angle of attack α and angle of sideslip β. Accurate estimates of these parameters are essential for the efficient and safe operation of aircraft. Traditionally, V_a, α, and β are directly measured by sensors, but there is new interest in methods for indirectly estimating them. For example, because many small unmanned aerial vehicle (UAV) systems cannot accommodate sensors that directly measure these quantities due to size, weight, and power constraints (e.g., wind vanes), nontraditional approaches collectively known as synthetic air data systems are being considered [1]. Synthetic air data systems (SADS), in short, are algorithms for generating estimates of the air data triplets V_a, α, and β indirectly without using the traditional air data sensors such as pitot tubes and vanes. The current state-of-the-art synthetic air data systems use a high-fidelity vehicle dynamics model fused with inertial measurement unit (IMU) and GPS measurements to estimate V_a, α, and β [1–6]. However, a high-fidelity aircraft model is difficult to obtain, and the parameters of the aircraft model can change over time due to wear and tear of the airframe and propulsion system [1]. This makes model-free synthetic air data estimation methods an attractive alternative [7]. A model-free system would be particularly useful for system identification of small UAVs. In these applications, the aircraft dynamic model is the unknown that needs to be determined, and algorithms for doing this can greatly benefit from the availability of α and β estimates [8]. Another potential application for model-free approaches is fault detection and isolation of traditional air data system malfunctions [9–12].

Received 8 October 2018; revision received 13 January 2019; accepted for publication 29 January 2019; published online 14 March 2019. Copyright © 2019 by Kerry Sun, Christopher D. Regan, and Demoz Gebre-Egziabher. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. All requests for copying and permission to reprint should be submitted to CCC at www.copyright.com; employ the eISSN 1533-3868 to initiate your request. See also AIAA Rights and Permissions www.aiaa.org/randp.

*Graduate Research Assistant, Department of Aerospace Engineering and Mechanics; sunx0486@umn.edu. Student Member AIAA.

†Research Fellow, Department of Aerospace Engineering and Mechanics; rega0051@umn.edu.

‡Professor, Department of Aerospace Engineering and Mechanics; gebre@umn.edu. Associate Fellow AIAA.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

A. Prior Work

Some of the earlier SADS literature relevant to the work here is from the late 1980s. In these works, the idea to use non-air-data sensors such as IMU and GPS fused with vehicle dynamics models to estimate the air data triplet V_a, α, and β (either separately or combined) is explained. Most of the earlier work used vehicle dynamics models to estimate air data in both aircraft and spacecraft applications. We refer to this approach as the aerodynamic model-based SADS [1,2,6,13–22]. The aerodynamic model-based SADS are difficult to implement, however, because it is challenging to obtain accurate vehicle dynamics models possessing the fidelity needed to yield the required accuracy in α, β, and V_a estimates. To address this issue, model-free SADS have been proposed [4,5,7,23]. There are a couple of key challenges in the model-free SADS framework. First, an accurate three-dimensional (3-D) wind vector estimate is needed for the model-free SADS. In most of the previous work, either wind modeling was not considered [23], or the wind modeling is two-dimensional (horizontal wind only) [24]. Furthermore, most of the wind estimators require an aircraft dynamics model [25,26]. Another challenge that has not been addressed adequately in the literature is the observability issue. Showing under what conditions the estimator is observable has not been studied extensively. In the Kalman filtering setting, observability analysis is often used as a binary test to show whether a system is observable or not [27–30]. However, in the SADS problem, this is not sufficient because the fact that a system is observable does not guarantee that the state estimates will be accurate. This is the problem of quantifying observability, and there has been some recent work in this area [31–34]. Most of the work, however, analyzes observability as a function of the vehicle trajectory ex post facto instead of giving guidance as to what flight maneuver sequences to execute to enhance observability.

B. Contribution

There are two objectives of this paper. First, building on work first reported in [7], we will show that it is possible to design a model-free SADS estimator that can generate sufficiently accurate estimates of α, β, and 3-D wind vector without using an aerodynamic model of the aircraft. Second, we will present a covariance analysis that quantifies the accuracy of the model-free estimator as a function of inputs (e.g., sensor noise, vehicle maneuvers, etc.). To address the first objective, we show that the model-free SADS estimator is conditionally observable under the slow-varying wind assumption. A linear time-varying (LTV) observability analysis is also constructed to examine the flight maneuver sequence that yields a high degree of observability. The findings of this observability analysis are validated using simulation and experimental flight-test data. Related to the second objective, the covariance bounds for α and β are derived using the general law of propagation variances [35]. Monte Carlo analysis is performed on several parameters, including IMU and GPS sensor errors, to show how various error sources affect the accuracy of α and β.

C. Paper Organization

The remainder of this paper is organized as follows. Section II presents a brief description of the model-free SADS estimator developed in this paper. Sections III and IV present the observability analysis for the estimator. Section V presents simulation results assessing the estimator's performance, and Sec. VI addresses the limitations on the model-free estimator. In Sec. VII, the estimator is validated using flight-test data from a small UAV.

II. Estimator Description

In this section, a brief overview of the time update and measurement models for the model-free air data estimator is provided. The estimator is an extension of the 15-state, loosely coupled inertial navigation system (INS)/GPS extended Kalman filter (EKF) [36], which blends information from an IMU and GPS receiver. The INS/GPS filter's state vector is augmented by three additional states representing the components of the wind velocity vector. Therefore, the SADS filter state vector $\delta\boldsymbol{x} \in \mathbb{R}^{18\times 1}$ is given by

$$\delta\boldsymbol{x} = [\,\delta\boldsymbol{p}^n \quad \delta\boldsymbol{v}^n \quad \delta\boldsymbol{\varepsilon} \quad \delta\boldsymbol{b}_a \quad \delta\boldsymbol{b}_g \quad \delta\boldsymbol{W}^n\,]^T \tag{1}$$

where $\delta\boldsymbol{p}^n = [\,\delta P_N \quad \delta P_E \quad \delta P_D\,]^T$ and $\delta\boldsymbol{v}^n = [\,\delta V_N \quad \delta V_E \quad \delta V_D\,]^T$ are the position and velocity error vectors resolved in the north–east–down (NED) frame. The vector $\delta\boldsymbol{\varepsilon} = [\,\delta\varepsilon_N \quad \delta\varepsilon_E \quad \delta\varepsilon_D\,]^T$ represents the attitude errors, which are defined to be the small rotations between the actual NED frame and the estimated NED frame. The vectors $\delta\boldsymbol{b}_a = [\,\delta b_{ax} \quad \delta b_{ay} \quad \delta b_{az}\,]^T$ and $\delta\boldsymbol{b}_g = [\,\delta b_{gx} \quad \delta b_{gy} \quad \delta b_{gz}\,]^T$ are accelerometer and rate gyro triad output errors (bias). Finally, $\delta\boldsymbol{W}^n = [\,\delta W_N \quad \delta W_E \quad \delta W_D\,]^T$ is the error in the wind speed vector resolved about the NED frame.

The estimates of α and β are calculated using the following:

$$\alpha = \tan^{-1}\left(\frac{u}{v}\right), \qquad \beta = \sin^{-1}\left(\frac{v}{\sqrt{u^2+v^2+w^2}}\right) \tag{2}$$

where

$$[\,u \quad v \quad w\,]^T = \boldsymbol{C}_n^b[\boldsymbol{v}^n - \boldsymbol{W}^n] \tag{3}$$

and $\boldsymbol{C}_n^b = [\boldsymbol{C}_b^n]^T$. The measurement vector $\boldsymbol{y} \in \mathbb{R}^{7\times 1}$ consists of GPS position and velocity estimates along with the scalar airspeed V_a; it is given by

$$\boldsymbol{y} = [\,\boldsymbol{p}^n \quad \boldsymbol{v}^n \quad V_a\,]^T \tag{4}$$

where V_a is the true airspeed estimate determined using the pressure measurements from a pitot tube. Figure 1 shows the overall filter architecture. The time and covariance update equations for this filter are, for the most part, identical to those of the filter described in [36]. What is new is the dynamic model for the augmented states (wind) and the measurement model. Similar to the modeling of the accelerometer and gyroscope biases in the filter, the dynamics of the wind is modeled as a first-order Gauss–Markov model, motivated by [37]. The details of the Gauss–Markov model for the wind and sensors can be found in [37,38], respectively.

The linearized measurement error model used by the EKF is $\delta\boldsymbol{y}_i = \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i \approx \boldsymbol{H}_i\delta\boldsymbol{x}_i + \boldsymbol{\eta}_i$. Note that $V_a = \|\boldsymbol{v}^n - \boldsymbol{W}^n\|_2 + \eta_{V_a}$,

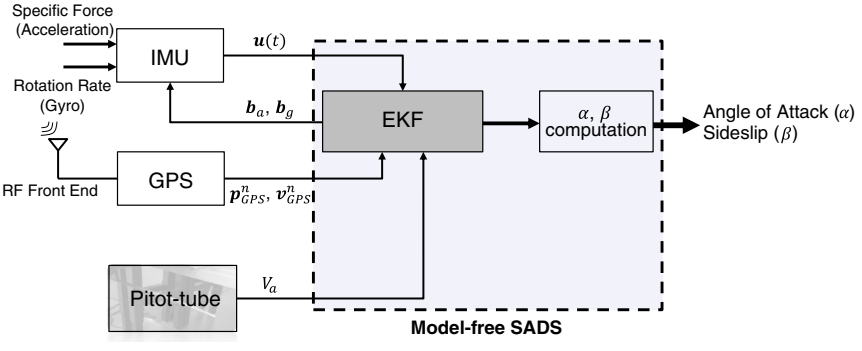


Fig. 1 Model-free SADS filter architecture.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

and $\hat{\boldsymbol{y}}_i$ is the predicted measurement. The measurement Jacobian $\boldsymbol{H}_i \in \mathbb{R}^{7\times 18}$ is given by the following:

$$\boldsymbol{H}_i = \begin{bmatrix} \boldsymbol{I}_3 & \boldsymbol{0}_3 & \boldsymbol{0}_{3\times 9} & \boldsymbol{0}_3 \\ \boldsymbol{0}_3 & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 9} & \boldsymbol{0}_3 \\ \boldsymbol{0}_1 & H_{v^n} & \boldsymbol{0}_{1\times 9} & H_{W^n} \end{bmatrix} \tag{5}$$

where H_{v^n} can be calculated as

$$H_{v^n} = \frac{1}{V_a}[\, V_N - W_N \quad V_E - W_E \quad V_D - W_D \,] \tag{6}$$

and $H_{W^n} = -H_{v^n}$. The measurement noise $\boldsymbol{\eta}$ is modeled as a zero mean, white Gaussian sequence with a diagonal covariance $\boldsymbol{R}$. That is,

$$\boldsymbol{R} = \mathrm{diag}([\, \sigma^2_{P_N} \quad \sigma^2_{P_E} \quad \sigma^2_{P_D} \quad \sigma^2_{V_N} \quad \sigma^2_{V_E} \quad \sigma^2_{V_D} \quad \sigma^2_{V_a} \,]) \tag{7}$$

The estimates and covariances of α and β are computed at each time step as illustrated by the α, β computation block in Fig. 1. Note that α and β are not part of the EKF state vector. Rather, they are functions of the EKF state vector, as given in Eqs. (2) and (3). Their covariance matrix $\boldsymbol{P}_{\alpha\beta}$ is computed using the general law of propagation variances [35] as follows:

$$\boldsymbol{P}_{\alpha\beta} = A_{\alpha\beta}\boldsymbol{P}A^T_{\alpha\beta} \tag{8}$$

where $\boldsymbol{P} \in \mathbb{R}^{18\times 18}$ is the covariance matrix of the EKF, and $A_{\alpha\beta} \in \mathbb{R}^{2\times 18}$ is the flow angle sensitivity matrix given by

$$A_{\alpha\beta} = \begin{bmatrix} \frac{\partial\alpha}{\partial P_N} & \frac{\partial\alpha}{\partial P_E} & \frac{\partial\alpha}{\partial P_D} & \frac{\partial\alpha}{\partial V_N} & \frac{\partial\alpha}{\partial V_E} & \frac{\partial\alpha}{\partial V_D} & \frac{\partial\alpha}{\partial\phi} & \frac{\partial\alpha}{\partial\theta} & \frac{\partial\alpha}{\partial\psi} & \frac{\partial\alpha}{\partial b_{ax}} & \frac{\partial\alpha}{\partial b_{ay}} & \frac{\partial\alpha}{\partial b_{az}} & \frac{\partial\alpha}{\partial b_{gx}} & \frac{\partial\alpha}{\partial b_{gy}} & \frac{\partial\alpha}{\partial b_{gz}} & \frac{\partial\alpha}{\partial W_N} & \frac{\partial\alpha}{\partial W_E} & \frac{\partial\alpha}{\partial W_D} \\ \frac{\partial\beta}{\partial P_N} & \frac{\partial\beta}{\partial P_E} & \frac{\partial\beta}{\partial P_D} & \frac{\partial\beta}{\partial V_N} & \frac{\partial\beta}{\partial V_E} & \frac{\partial\beta}{\partial V_D} & \frac{\partial\beta}{\partial\phi} & \frac{\partial\beta}{\partial\theta} & \frac{\partial\beta}{\partial\psi} & \frac{\partial\beta}{\partial b_{ax}} & \frac{\partial\beta}{\partial b_{ay}} & \frac{\partial\beta}{\partial b_{az}} & \frac{\partial\beta}{\partial b_{gx}} & \frac{\partial\beta}{\partial b_{gy}} & \frac{\partial\beta}{\partial b_{gz}} & \frac{\partial\beta}{\partial W_N} & \frac{\partial\beta}{\partial W_E} & \frac{\partial\beta}{\partial W_D} \end{bmatrix} \tag{9}$$

The partial derivatives with respect to the position $\boldsymbol{p}^n$, accelerometer $\boldsymbol{b}_a$, and gyro $\boldsymbol{b}_g$ biases are zero (see the Appendix). Therefore, the covariance of α and β depends on the covariance of the inertial velocity, attitude, and 3-D wind estimates. The matrix $A_{\alpha\beta}$ will be used in Sec. VI for error analysis.

III. Observability Analysis

In [7], it was shown that the SADS estimator is conditionally observable. For completeness, we will briefly discuss the results of the observability analysis from [7] because its results are a prerequisite for understanding the observability enhancement results discussed later in the paper. Recall that the mathematical models in the synthetic air data estimation problem [the INS/GPS equations in [36], Eqs. (2) and (3)] are nonlinear. The formulation of the EKF requires linearizing the mathematical models, and thus the resulting estimator falls in the domain of linear time-varying systems. Therefore, the observability analysis requires the combined use of nonlinear and linear tools. In this paper, we take a two-tiered approach to this analysis. The first tier involves assessing the observability of the state vector $\delta\boldsymbol{x} \in \mathbb{R}^{18\times 1}$ given by Eq. (1). This analysis was carried out in some detail in [7], and the details of that analysis will not be repeated here. The key results of this analysis were that $\delta\boldsymbol{x}$ is conditionally observable and its observability requires the following.

1) The airplane is accelerating so that the INS/GPS heading and gyro bias states become observable [36].

2) The wind speed vector $\boldsymbol{W}^n$ is quasi-static. The term quasi-static is used to mean that the variations in $\boldsymbol{W}^n$ are such that it is valid to assume that it remains constant over a small time window.

With respect to the second observability condition, the size of the time window in which $\boldsymbol{W}^n$ can be assumed constant depends on several factors. In the second-tier analysis (see Sec. IV), we provide mathematical relationships describing these and other conditions that enhance the observability of the SADS estimates.

IV. Observability Enhancement

Although the work in [7] provided a mathematically based observability analysis, the second-tiered analysis here will provide a more intuitive interpretation. The goal is to get insight into how observability can be improved or enhanced. This approach sheds some light on how to design flight maneuver sequences to optimize the degree of observability. To this end, the estimation problem is recast in a slightly different way. The objective is to estimate α, β, W_N, W_E, and W_D. Observability is an inherent property of the system that does not change by new definitions of the state vector. The recasting does not alter the observability, but it gives a better view into the problem by essentially eliminating the dynamic model and allowing us to recast the problem as a batch estimation process.

A. Change of Variables

Let $\boldsymbol{v}^b_g = [\, u_g \quad v_g \quad w_g \,]^T$ denote the velocity vector of the aircraft ground velocity in the body frame. Let $\boldsymbol{v}^b_a = [\, u \quad v \quad w \,]^T$ be the air velocity in the body frame. The kinematic relationship between ground velocity, air velocity, and wind velocity is described by the following:

$$\begin{bmatrix} u_g \\ v_g \\ w_g \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} + \boldsymbol{C}^b_n \begin{bmatrix} W_N \\ W_E \\ W_D \end{bmatrix} \tag{10}$$

The vector $\boldsymbol{v}^b_a$ can also be expressed in terms of airspeed V_a, angle of attack α, and side-slip angle β, as shown in Eq. (11):

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} V_a \cos\alpha\cos\beta \\ V_a \sin\beta \\ V_a \sin\alpha\cos\beta \end{bmatrix} \tag{11}$$

Note that $V_a \neq u$ because the stagnation pressure sensed by a pitot tube is relatively constant under small ($\leq \pm 10$ deg) variations of α and β [39].

Using a small-angle approximation for α and β and substituting Eq. (11) into Eq. (10) gives the following:

$$\begin{bmatrix} u_g - V_a \\ v_g \\ w_g \end{bmatrix} \approx \begin{bmatrix} 0 \\ V_a\beta \\ V_a\alpha \end{bmatrix} + C^b_n \begin{bmatrix} W_N \\ W_E \\ W_D \end{bmatrix} = [\, \boldsymbol{\Psi} \mid \boldsymbol{\Omega} \,] \left[\begin{array}{c} \boldsymbol{W}^n \\ \hline \alpha \\ \beta \end{array}\right] \tag{12}$$

where $\boldsymbol{\Psi} = C^b_n$, and $\boldsymbol{\Omega}$ is a function of V_a, shown next:

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(V_a) = V_a \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{13}$$

We will refer to Eq. (12) as the wind triangle kinematic relationship and use it as the measurement update equation of the following LTV system:

$$\mathrm{LTV}_K = \begin{cases} \boldsymbol{x}_{k+1} = \Phi_k\boldsymbol{x}_k + G_k\boldsymbol{u}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_{k+1} = H_k\boldsymbol{x}_k + \boldsymbol{\eta}_k \end{cases} \tag{14}$$

where $H_k = [\boldsymbol{\Psi}_k \mid \boldsymbol{\Omega}_k]$; $\boldsymbol{y}_k = [u_g(k) - V_a(k) \quad v_g(k) \quad w_g(k)]^T$; the lowercase k represents the time step (sensor sampling rate); and the uppercase K represents the Kth piecewise LTV system (see Sec. IV.E.1). The states of the LTV are $\boldsymbol{x}_k = [W_N \quad W_E \quad W_D \quad \alpha_k \quad \beta_k]^T$. We assume that the wind is constant for each LTV, but α and β are not. Because we are performing a nonstochastic observability analysis, the process noise $\boldsymbol{w}_k$ and measurement noise $\boldsymbol{\eta}_k$ are dropped. Also, there is no input to the system (i.e., $G_k = 0$). Our objective is to convert this LTV into a batch process. To this, we can assume that Φ_k is equal to a 5×5 identity matrix. Because we have already assumed that the winds are constant in a given LTV, the fact that $\Phi_k = I_{5\times5}$ implies that all the information about the changes in α and β are coming from the measurement update equation of the LTV, namely Eq. (12).

To convert any LTV to a batch process, we assume that wind is constant over a time window of width $l \cdot T_s$. The constant l is the number of time steps ($l \in [1, 2, \ldots, k, \ldots, n-1, n]$), and $T_s = t_k - t_{k-1}$ is the sampling interval. Now, the entire trajectory is a collection of piecewise LTV systems $\mathrm{LTV}_1, \mathrm{LTV}_2, \ldots \mathrm{LTV}_K, \ldots \mathrm{LTV}_N$, where $K \in [1, N]$, and the wind vector is assumed to be constant over the time span $l \cdot T_s$. If we now start stacking the measurement equation for several epochs, we obtain

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}}_{\mathcal{Y}} = \underbrace{\begin{bmatrix} \boldsymbol{\Psi}_1 & \boldsymbol{\Omega}_1 & \boldsymbol{0}_{3\times2} & \cdots & \cdots & \boldsymbol{0}_{3\times2} \\ \boldsymbol{\Psi}_2 & \boldsymbol{0}_{3\times2} & \boldsymbol{\Omega}_2 & \boldsymbol{0}_{3\times2} & \cdots & \boldsymbol{0}_{3\times2} \\ \boldsymbol{\Psi}_3 & \boldsymbol{0}_{3\times2} & \boldsymbol{0}_{3\times2} & \boldsymbol{\Omega}_3 & \cdots & \boldsymbol{0}_{3\times2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Psi}_n & \cdots & \cdots & \cdots & \cdots & \boldsymbol{\Omega}_n \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} \boldsymbol{W}^n \\ \hline \alpha_1 \\ \beta_1 \\ \hline \alpha_2 \\ \beta_2 \\ \hline \alpha_3 \\ \beta_3 \\ \hline \vdots \\ \alpha_n \\ \beta_n \end{bmatrix}}_{\mathcal{X}} \tag{15}$$

The vector $\mathcal{Y}$ has $3n$ elements. The vector $\mathcal{X}$ has $3 + 2n$ elements. For the preceding system to be solvable, $\mathcal{H}$ must have full rank, and $3n \geq 3 + 2n \Rightarrow n \geq 3$. Thus, as long as the wind remains constant over a time window of $3T_s$ (or, alternatively, each LTV is at least $3T_s$ wide in time) and $\mathcal{H}$ is full rank, then the SADS problem is observable. As long as attitude changes, meaning that $\boldsymbol{\Psi}_k$ are not constant, then $\mathcal{H}$ will have full rank. However, a simple "yes" or "no" on observability does not tell us anything about the quality (accuracy) of the estimates. What we want to know is how much do attitude and airspeed have to change to get an accurate estimate? To answer this question, we will look at the observability Gramian.

B. Maneuver Optimization Problem

To proceed with this analysis, we assume that sufficient time has elapsed, and the INS/GPS solution has converged such that the attitude errors are small. This means that C_n^b is known. The Gramian $\mathcal{G}_d$ over $l \cdot T_s$ is calculated to be using the discrete observability Gramian definition [Eq. (6) in [7]]:

$$\mathcal{G}_d = \mathcal{H}^T\mathcal{H} \tag{16}$$

As mentioned earlier, at least three different time steps are needed to achieve a rank of 5. However, if attitude and airspeed are not changed sufficiently over $3T_s$, then the matrix $\mathcal{H}$ of Eq. (15) may be ill-conditioned. If α and β are changing rapidly and significantly, then an even larger change in attitude and airspeed are required. We will start with a benign but unrealistic case of α and β being constant. This will provide a floor or the minimum attitude change required. We will relax this assumption later. Now, with the constant α and β assumption, Eq. (15) has dimensions of $3k \times 5$, and matrix $\mathcal{H}$ has the following form:

$$\mathcal{H}_c = \begin{bmatrix} \boldsymbol{\Psi}_1 & \boldsymbol{\Omega}_1 \\ \vdots & \vdots \\ \boldsymbol{\Psi}_k & \boldsymbol{\Omega}_k \\ \vdots & \vdots \\ \boldsymbol{\Psi}_n & \boldsymbol{\Omega}_n \end{bmatrix} \tag{17}$$

Hence, for a fixed time horizon T, the following Gramian is obtained:

$$\mathcal{G}_d = \mathcal{H}^T\mathcal{H} \text{ for nonconstant } \alpha \text{ and } \beta \tag{18}$$

$$= (\mathcal{H}_c)^T\mathcal{H}_c \text{ for constant } \alpha \text{ and } \beta \tag{19}$$

Krener and Ide [31] introduced measures of observability based on the condition number or the smallest singular value of the local observability Gramian. The condition number measures the degree of attenuation information experiences while traveling along the path from the measurements to the states. The larger the condition number, the more attenuated the information. The local observability Gramian can be calculated in the closed form in this case, and we will use this metric to quantify observability. Formally, measuring the degree of observability for the system can be formulated as the following optimization problem for the condition number of the observability Gramian $\kappa(\mathcal{G}_d)$, given next:

$$\begin{aligned} \underset{\boldsymbol{\psi}_{nb_1}, \ldots, \boldsymbol{\psi}_{nb_k}, \ldots, V_{a_1}, \ldots, V_{a_k}}{\operatorname{argmin}} \quad & \kappa(\mathcal{G}_d) \\ \text{s.t.} \quad & V_{a_{\min}} \leq V_a \leq V_{a_{\max}} \\ & \theta_{\min} \leq \theta \leq \theta_{\max} \\ & \phi_{\min} \leq \phi \leq \phi_{\max} \\ & \psi_{\min} \leq \psi \leq \psi_{\max} \end{aligned} \tag{20}$$

where $\boldsymbol{\psi}_{nb_k} = [\psi \quad \theta \quad \phi]^T$ and V_{a_k} are the set of Euler angles and airspeed at time step k. There are four linear constraints that represent, in part, the flight envelope of the aircraft. The goal is to minimize the condition number over a finite time horizon. Theoretically, the minimum condition number that can be obtained is 1, which is an indication that the given matrix is well conditioned and the states are well observed. The Gramian $\mathcal{G}_d$ for the constant α and β can become ill-conditioned due to the nature of the $\mathcal{H}_c$ matrix. The first three columns consist of the coordinate transformation matrix (with the norm equal to 1), but the last two columns are scaled by the airspeed (typically in the range of 15–25 m/s for small UAVs). This makes the entire matrix ill-conditioned. Therefore, the minimum acceptable condition number is a function of a particular aircraft's flight envelope. The minimum acceptable $\kappa_{\min}$ can be derived from the design requirements from users of the SADS. The definition of condition number of a matrix $\mathcal{G}_d$, denoted $\kappa(\mathcal{G}_d)$, is defined as

$$\kappa(\mathcal{G}_d) = \|\mathcal{G}_d\| \cdot \|\mathcal{G}_d{}^{-1}\| \tag{21}$$

where $\|\|$ is either the 2-norm or Frobenius norm of the Gramian [40]. Because the optimization problem in Eq. (20) is nonconvex (multiplication of sinusoidal functions), numerically solving it would not give us any meaningful insight (multiple local minimum solutions). Therefore, it is parsed into several suboptimization problems that are physically sound for practical aircraft maneuvers.

C. Two-Dimensional Versus Three-Dimensional Maneuver

As a concrete demonstration of how the insight gained from the observability analysis in the previous section can be used in affecting the performance of the SADS estimates, we exercise the estimator on a catalog of aircraft trajectories. A catalog of typical two-dimensional (2-D) and 3-D flight maneuvers is provided in Table 1. The double dashed lines in Table 1 mean that the variable is not being actively commanded; it only varies passively due to the coupled longitudinal and lateral motions. For example, if the heading angle is changing, then the roll angle will inevitably change. However, the roll angle can

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

Table 1 2-D and 3-D flight maneuver motion

Maneuver	V_a	ψ	θ	ϕ
		2-D case		
1	Constant	Constant	Varying	— —
2	Constant	Varying	Constant	— —
3	Constant	Constant	— —	Varying
4	Varying	Constant	Varying	— —
5	Varying	Varying	Constant	— —
6	Varying	Constant	— —	Varying
		3-D case		
1	Constant	Varying	Varying	— —
2	Varying	Varying	Varying	— —
3	Varying	Varying	Varying	Varying

return to zero while the heading remains at a new commanded value after the maneuver. We will consider the first 3-D case in Table 1 in detail to demonstrate the conditional observability analysis. All other cases can be analyzed in the same fashion. Some key results for this analysis are summarized at the end of this section.

D. Three-Dimensional Case Study Example

At least three different points in state space (i.e., $k = 1, 2, 3$) are needed for the Gramian $\mathcal{G}_d$ to achieve full rank. For the first case of 3-D motion, the $\mathcal{H}_c(1{:}3)$ matrix is used in the construction of $\mathcal{G}_d$ using Eq. (19). Because an aircraft can start from any orientation, the easiest orientation is picked as the reference point, which is to set all Euler angles to be 0 deg. The airspeed is set to be 15 m/s. For the second and third points in state space, ϕ is held to be the same angle, but ψ and θ are varied, giving the following $\mathcal{H}_c(1{:}3)$:

$$\mathcal{H}_c(1{:}3) = \begin{bmatrix} C_n^b(0,0,0 \text{ deg}) & \boldsymbol{\Omega}(V_a) \\ C_n^b\left(0 \text{ deg}, \frac{\theta_i}{2}, \frac{\psi_i}{2}\right) & \boldsymbol{\Omega}(V_a + \epsilon) \\ C_n^b(0 \text{ deg}, \theta_i, \psi_i) & \boldsymbol{\Omega}(V_a + \epsilon) \end{bmatrix} \tag{22}$$

where $\theta_i \in (-30, 30 \text{ deg})$, and $\psi_i \in (0, 180 \text{ deg})$. θ and ψ in the second row block of Eq. (22) are half of the values in the third row block. This is done to prevent a poor condition number. Also, small ϵ is necessary to prevent a poor condition number, and it is set to be 1 m/s. With $\mathcal{H}_c(1{:}3)$ covering all possible pitch and heading angles, Figs. 2 and 3a are obtained by calculating the 2-norm and F-norm condition number over a sweep of ψ and θ, respectively.

Figure 2 is the normalized 2-norm condition number for various ψ and θ. The normalized condition number is calculated by dividing every condition number by the minimum condition number. Figure 2 gives a guideline on how an airplane should maneuver to achieve a high degree of observability. For example, in Fig. 2, the red dot O represents the reference orientation $(\psi, \theta, \phi) = (0, 0, 0 \text{ deg})$, which is the initial of orientation of the aircraft. Now, suppose that the red dot A represents the final orientation [e.g., $(90, -20, 0 \text{ deg})$] of the aircraft in one flight maneuver sequence. A certain normalized condition number is achieved by this maneuver sequence from the initial orientation represented at O to the final orientation presented at A. The lower the condition number is, the higher degree of observability is achieved by the maneuver sequence.

Figure 3a is the normalized F-norm condition number for various ψ and θ. As can be seen in the figure, the dynamic range of the F-norm as a function of ψ and θ is larger than that for the 2-norm. For example, once the orientation deviates from point O, the condition number is immediately reduced to a flat region. The condition number is not further reduced if heading ψ and pitch angle θ are not changed together significantly. As such, the F-norm is a more sensitive indicator in the relationship between observability and aircraft maneuver aggressiveness.

Of course, this is a very optimistic analysis because α and β were assumed to be constant. When α and β are allowed to vary and $\mathcal{H}$ [Eq. (15)] is used to form the observability Gramian, Fig. 3b results. Figure 3b is effectively Fig. 3a with a "crown" of the surface removed. In other words, if α and β are changing, more aggressive attitude changes are needed to achieve a good degree of observability.

A similar analysis was carried out for the different cases listed in Table 1, and a summary of the key findings of this analysis follows.

1) Condition number does not improve significantly if heading angle stays at the reference value, regardless of how the pitch angle is changed. The same can be said if the pitch angle stays at the reference value while the heading angle is changed.

2) The minimum condition number is achieved by changing the aircraft's pitch and heading angle simultaneously (greatest gradient direction) as much as possible.

3) If the reference point is at (0, 0, 0 deg), then the figure is symmetric with respect to $\psi - \kappa_i$ or $\theta - \kappa_i$ (where $i = 2$ or F) planes. In other words, pitching down or up and changing heading

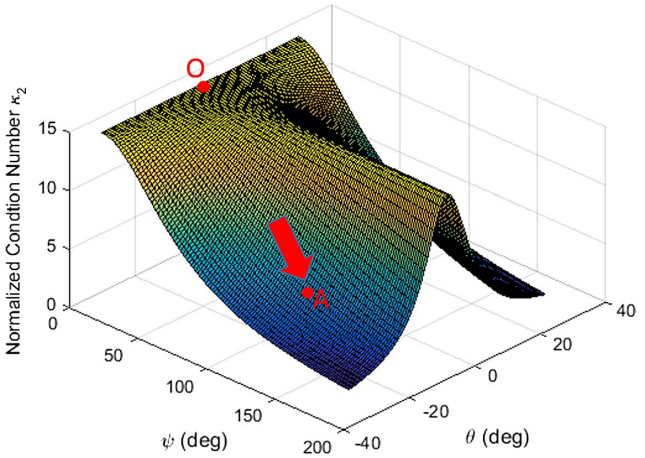
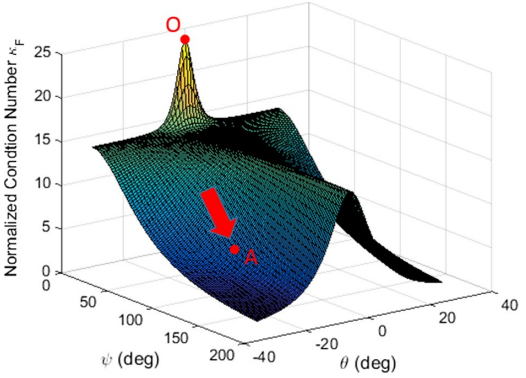
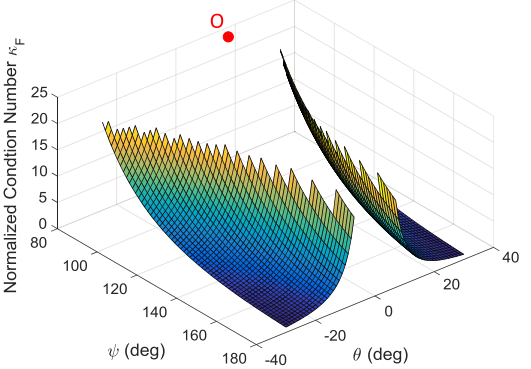


Fig. 2 Normalized 2-norm condition number over ψ and θ.



a) Normalized F-norm condition number over ψ and θ assuming α and β are constant over each LTV_K



b) Normalized F-norm condition number over ψ and θ assuming α and β changing over each LTV_K

Fig. 3 Normalized F-norm condition number over ψ and θ.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

clockwise and counterclockwise have the same effect on condition number.

4) Varying airspeed while changing the orientation of the aircraft is the best flight maneuver to improve the condition number.

The observability analysis presented here is independent of the aircraft. The analysis can be applied to any aircraft to study observability. Furthermore, the analysis is not derived based on the typical flight trajectories such as a circular or a climb/descent trajectory. Instead, the flight trajectory is decomposed into points of the state space that we will refer to as the maneuver frame in this paper. Each maneuver frame is defined by the aircraft's orientation and airspeed. A sequence of maneuver frames forms a flight trajectory. The condition number of an LTV observability Gramian is used as a measure of observability, and its variation as a function of flight trajectory is examined to show how an aircraft can maneuver to achieve a high degree of observability.

E. Time Aspect of Maneuverability

The time it takes to maneuver from point O to point A shown in Figs. 2 and 3a is aircraft-dependent. For example, it might take a slightly longer time for a larger aircraft to maneuver from point O to point A than it would for a smaller, more agile aircraft. As we will discuss next, the length of the time interval required to go from one maneuver frame to the next and how the wind varies during this time interval have an important bearing on the SADS estimator observability.

1. *Maneuver Frame Versus Estimator Rate*

Recall that the maneuver frame is defined as one snapshot of the aircraft maneuver in the inertial coordinates, characterized by aircraft's orientation and airspeed. The rate at which an aircraft moves from one maneuver frame to the next depends primarily on the aircraft's dynamics. The estimator rate, defined as the rate at which the SADS estimator computes the air data states, is related to the sensor sampling rate. It is the rate at which the underlying nonlinear equations are linearized to form the LTV system used by the EKF. In Fig. 4, the flow of maneuver frames and its relationship to estimator rate are shown graphically. If the wind velocity vector is constant, then all that is required for observability is that the aircraft go from maneuver frame O to maneuver frame A. The number of sampling frames that elapse during the movement from O to A would be immaterial. However, the fact that the wind velocity vector is not a constant puts an upper bound on how slowly the aircraft can move from maneuver frame O to A and still maintain observability of the SADS estimator.

2. *Wind Variation During Maneuvers*

Figure 5 is a sectional cut at $\psi = 90$ deg of Fig. 2. Let τ_M the be maneuvering time from point O to any other maneuver frame (i.e., A, B, C, or D). Let τ_W be the time constant of the wind, the average time of wind variation. Let τ_M^U be the upper bound of the maneuvering time and τ_W^L be the lower bound of the time constant of the wind. The following inequality has to be satisfied to have an observable system:

$$\tau_M \le \tau_M^U \le \tau_W^L \le \tau_W \tag{23}$$

Equation (23) states that maneuver time has to be less than the wind time constant. This supports the assumption of slow-varying wind. If an aircraft can maneuver fast enough and provide enough excitation

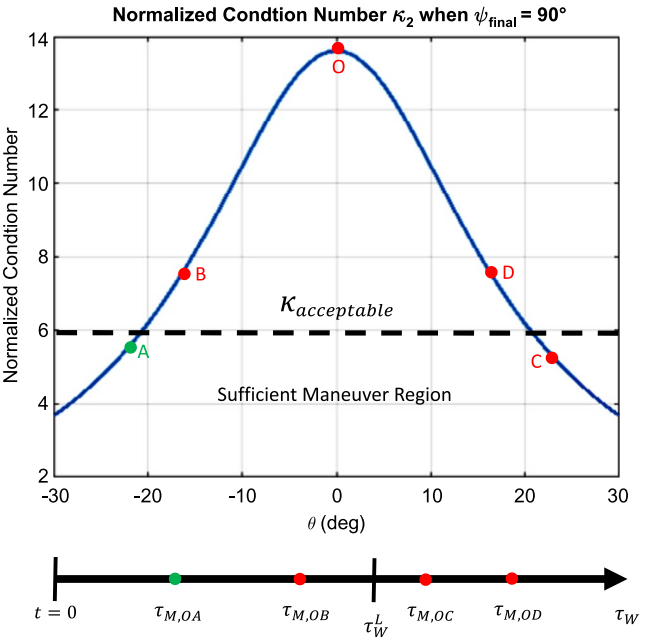


Fig. 5 Sectional cut of Fig. 2 when $\psi_{final} = 90$ deg.

before the wind changes, then the estimator is observable; the wind, angle of attack, and angle of sideslip can be estimated.

This concept is depicted graphically in Fig. 5, which shows four hypothetical trajectories in the maneuver space: OA, OB, OC, and OD. It also shows a timeline of maneuver times. For the estimator to be observable, the maneuver has to satisfy two criteria; it must have achieved an acceptable condition number, and the maneuver time τ_M has to be less than the lower bound of wind time constant τ_W^L. Assuming that the area below the dotted line, denoted as the sufficient maneuver region, is the region of acceptable condition numbers, then OA and OC have an acceptable condition number, whereas OB and OD do not. For the maneuver time τ_M, it can be seen that $\tau_{M,\mathrm{OA}}$ and $\tau_{M,\mathrm{OB}}$ are less than τ_W^L, whereas $\tau_{M,\mathrm{OC}}$ and $\tau_{M,\mathrm{OD}}$ are not. Thus, as summarized in Table 2, the trajectory from O to A is observable.

The wind time constant depends on the weather condition on a given day. There are models for the variation of τ_W. For example, the work in [37] provides a model with an upper bound for τ_W, which is approximately 400 s.

V. Validation and Simulation Study

A. Simulation Setup

A modified version of the nonlinear aircraft Simulink model described in [41,42] is used for the simulation analysis that is discussed next. The parameters of the aircraft in the Simulink model are for an Ultra Stick 120 UAV, whose equations of motion were developed in [43]. Different maneuver sequences are simulated to test the observability analysis. For the sake of brevity, only case 2 of the 3-D cases listed in Table 1 is discussed here. Note that case 2 is a more generalized version of case 1 shown in Sec. IV.D. The airspeed in case 2 is varied to help increase the degree of observability through more vertical excitations. Case 2 consists of two maneuver sequences

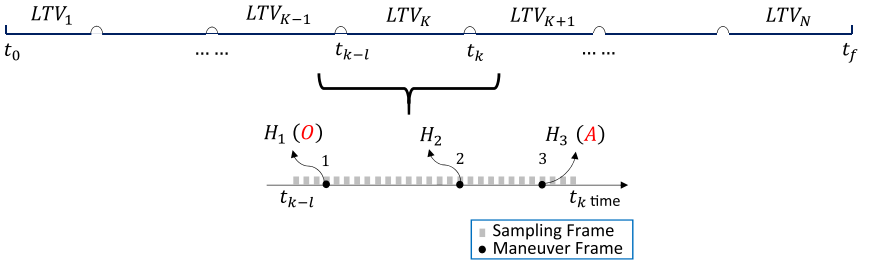


Fig. 4 Graphical depiction of the maneuver frame and sampling frame.

Table 2 Observability cases study based on normalized condition number $\kappa_{\text{acceptable}}$ and maneuver time τ_M

Case	$\kappa_{(\cdot)} < \kappa_{\text{acceptable}}$	$\tau_{M,(\cdot)} < \tau_W^I$	Observable maneuver?
OA	Yes	Yes	Yes
OB	No	Yes	No
OC	Yes	No	No
OD	No	No	No

(denoted maneuvers 1 and 2 in Table 1) executed sequentially. The wind direction is simulated to slowly vary horizontally from 30°W to 45°E in the first 75 s. The wind direction then is kept at 45°E. The downward wind is initially at 0.5 m/s and is then stepped to −0.5 m/s at 150 s. The actuator noise is set to be zero for simplicity. The details of the maneuver and wind conditions can be found in Table 3. The measurement noise and the wind parameters of the simulation are listed in Table 4. The GPS and IMU parameters are chosen based on a consumer-level GPS and IMU [36]. The measurement error for airspeed is set to be 10% of the airspeed. This is determined to be sufficient to cover observed errors in the pitot tube based on the real flight calibration results.

To estimate the changes in the wind accurately during flight, a high degree of observability needs to be achieved. Varying heading constantly and sinusoidally varying airspeed (which leads to varying pitch angle) simultaneously is sufficient to accurately estimate the changes in the horizontal wind direction (maneuver 1 in Table 3). The rate of heading change and airspeed have to be faster than the rate of the horizontal wind directional change.

Table 3 3-D case 2 simulation description

Name	Time, s	Description
Maneuver 1	0–75	Steady flight with varying heading ψ constantly (in a figure-of-eight pattern); sinusoidally varying airspeed V_a (±4.5 m/s from average)
Maneuver 2	75–300	Aggressive flight with varying heading ψ constantly (in a figure-of-eight pattern); sinusoidally varying airspeed V_a (±8 m/s from average)
Wind condition 1	0–75	Start blowing from 30° W, then direction is slowly changing to 45° E in 75 s; downward wind: 0.5 m/s
Wind condition 2	75–150	Stays at 45° E; downward wind: 0.5 m/s ($W_N = -3.5535$ m/s, $W_E = -3.5535$ m/s, $W_N = 0.5$ m/s)
Wind condition 3	150–300	Stays at 45° WE; downward wind: −0.5 m/s ($W_N = -3.5535$ m/s, $W_E = -3.5535$ m/s, $W_N = -0.5$ m/s)
Miscellaneous	0–300	Actuator noise off

Maneuver 2 contains several pitch and altitude changes. These are required to make the vertical component of wind W_D observable. In addition to constant heading change, maneuver 2 also features large airspeed changes. This is done by commanding the aircraft to climb and descend at the maximum rate supported by the air frame and power plant (changes around Δ8 m/s). Figure 6a shows the trajectory of the simulated flight. The color bar represents the magnitude of the altitude. Figure 6b shows the simulated time history of airspeed, altitude, and throttle. Observe the large altitude variation (45–120 m) and concomitant large vertical velocity excitation of maneuver 2.

Figure 7 sheds light on why maneuver 2 is necessary for all the wind components to be observable. To estimate the 3-D wind vector, the aircraft must fly with a sufficient amount of different inertial velocities, particularly with large directional variations. This is shown on the left side of Fig. 7, where the inertial velocity vectors (green color vectors) are used to estimate the 3-D wind vector (the blue color vector). The aircraft can achieve sufficient variation in inertial velocity, particularly in the vertical direction, by flying through p_1, p_2, and p_3.

The right side of Fig. 7 illustrates how the flight trajectory points p_1, p_2, and p_3 can be thought of geometrically as a problem of estimating the radius and center of the sphere. If we translate the wind vector and all the inertial velocity vectors to the origin of the NED frame, the difference between the wind vector and inertial velocity vector is the airspeed vector (red vector). The inertial velocity vectors point to the surface of the sphere. When there are sufficient inertial velocity vectors, the center and radius of the sphere can be estimated. The radius of the sphere represents the airspeed magnitude. The center of the sphere is used in the wind triangle equation to obtain the 3-D wind vector. The yellow ellipsoid only represents partial wind observability space if the downward direction is not fully swept by the inertial velocity vector. In other words, the yellow ellipsoid is deficient for the 3-D wind estimation. Furthermore, inaccurate spherical estimation can also result in ambiguity in estimating the direction of the downward wind. This indicates that maneuvers containing large variations in pitch, altitude, and heading are needed to estimate the 3-D wind vector.

Maneuver 1 is not sufficient enough to capture the change in the downward wind. The vertical inertial velocity component V_D has to be the same order of magnitude as the horizontal inertial velocity components V_N and V_E to detect the downward wind change. Therefore, maneuver 2 is employed to provide both horizontal and vertical velocity excitations in the second phase of the flight simulation.

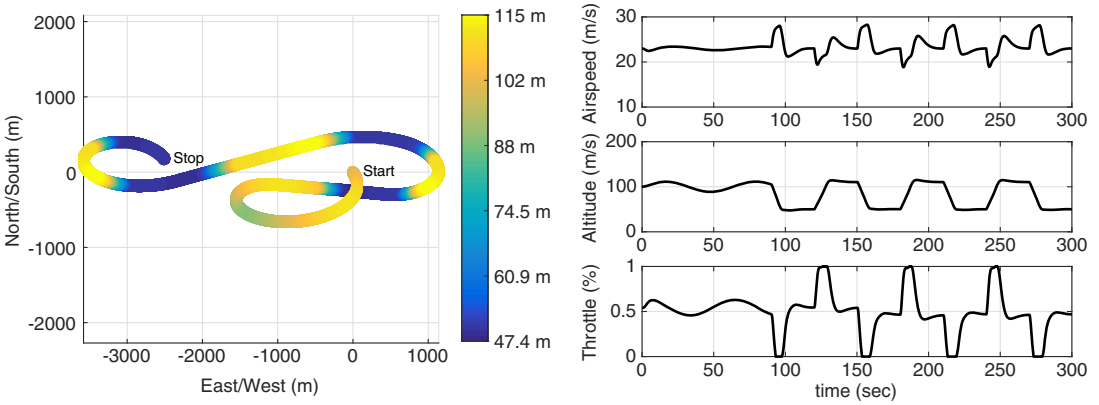
B. Simulation Results

Figure 8 shows the attitude estimates, attitude estimation error, and its $3-\sigma$ covariance bound for the entire flight. The uncertainty represented by $3-\sigma$ bound of the attitude decreases significantly at 75 s due to the longitudinal maneuver. The $3-\sigma$ bounds stop decreasing further and become oscillatory, especially in ψ. In real flight, the uncertainty bound for attitude should decrease further; the maneuvers

Table 4 Measurement and wind parameters setting for simulation

Parameter	Source	Variable	Simulation setting
Position	GPS position measurement	P_N, P_E, P_D	$\sigma_{P_D} = \sigma_{P_E} = \sigma_{P_D} = 3$ m
NED velocity	GPS velocity measurement	V_N, V_E, V_D	$\sigma_{V_N} = \sigma_{V_E} = \sigma_{V_D} = 0.2$ m/s
Airspeed	Pitot tube	V_a	$\sigma_{V_a} = 10\%\overline{V_a}$
Acceleration	IMU measurement	a_x, a_y, a_z	$\sigma_{\text{aw}} = 0.05$ m/s^2 $\sigma_{\text{ad}} = 5^{-3}$ g $\tau_a = 300$ s
Angular rate	IMU measurement	p, q, r	$\sigma_{\text{gw}} = 0.1$ deg/s $\sigma_{\text{gd}} = 360$ deg/h $\tau_g = 300$ s
Wind	Wind parameter	W_N, W_E, W_D	$\sigma_{W_N} = \sigma_{W_E} = 0.5$ m/s $\sigma_{W_D} = 2.0$ m/s $\tau_{W_N} = \tau_{W_E} = 1$ s $\tau_{W_D} = 10$ s

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290



a) Simulated trajectory

b) Airspeed, altitude, and throttle setting time history. The vertical bar on the right provides a color code for altitude changes

Fig. 6 Simulated trajectory, airspeed, altitude, and throttle setting.

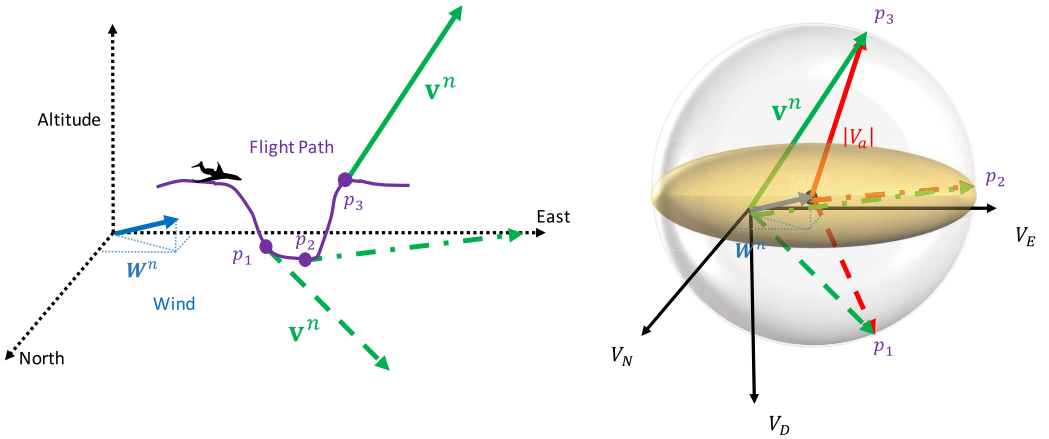
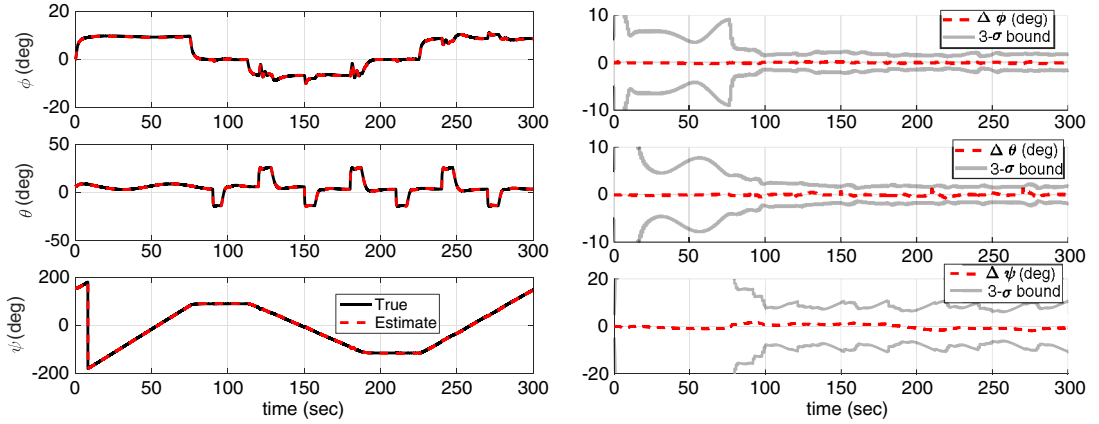


Fig. 7 Wind velocity estimation illustrated using sphere estimation.



a) Euler angle time history

b) Euler angle error and its 3-σ bound

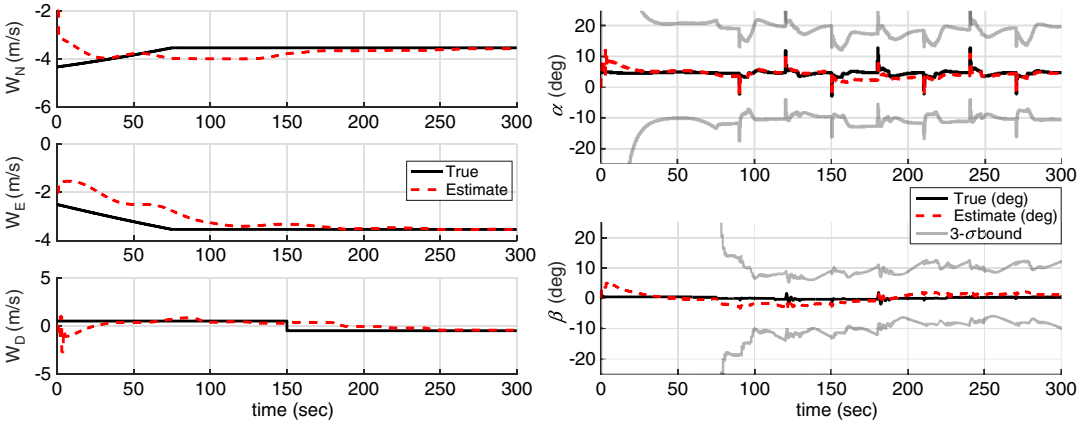
Fig. 8 Euler angle true and estimates and its $3-\sigma$ bounds.

will be more aggressive (e.g., include sharp turn) than the ones simulated.

Figure 9 shows the 3-D wind components, α, and β estimates. All the wind components converge to the true value after a sufficient amount of time and excitation. For the W_N and W_E components, the estimates slowly converge to the true value, whereas the magnitude of wind is changing over the first 75 s. The downward wind estimate W_D captures the vertical wind speed change at $t = 150$ s. The α estimate converged after about 20 s. When the downward wind changed direction at 150 s, the α estimate quickly converged to the true value. On the other hand, β estimate did not converge quickly to the true value. Stable convergence was achieved at $t = 210$ s. The accuracy of the attitude is critical to β estimation. For example, this can be observed by looking at the shape of the $3-\sigma$ bound for the β estimates shown in Fig. 9b; it mirrors the shape of the $3-\sigma$ bound of ψ.

The $1-\sigma$ of α and β estimate are determined to be 5.04 and 3.70 deg, respectively. The uncertainty of α and β is large and would

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290



a) Simulated vs. estimated wind vector **b) Simulated vs. estimated α and β**

Fig. 9 Wind vector (true and estimates), α, β (true and estimates), and its $3-\sigma$ bounds.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

generally be unacceptable. Note that the model-aided SADS estimator described in [1,3] reported a $3-\sigma$ covariance bound for the airspeed, angle of attack, and sideslip of 2 m/s, 3 deg, and 5 deg, respectively. The fact that the estimator proposed here is less accurate should not come as a surprise. After all, it is a model-free estimator and cannot leverage of information available from an aerodynamic model of the airplane. The following section explores the limitations of the estimates through sensitivity and error analysis. This analysis will give insight into how to improve the quality of the SADS estimates.

VI. Accuracy Limitation

The simulation of the previous section showed that the uncertainty of α and β can be rather large. In this section, we show how various factors affect the accuracy of α and β and why the minimum expected accuracy of α and β is mainly a function of the wind uncertainty and airspeed. The variance on the estimation errors of α and β [the diagonal entries of $\boldsymbol{P}_{\alpha\beta}$ in Eq. (8)] are given by

$$\boldsymbol{P}_{\alpha\beta}(1,1) = \sigma_\alpha^2 = \underbrace{\left(\frac{\partial\alpha}{\partial V_N}\sigma_{V_N}\right)^2 + \left(\frac{\partial\alpha}{\partial V_E}\sigma_{V_E}\right)^2 + \left(\frac{\partial\alpha}{\partial V_D}\sigma_{V_D}\right)^2 + \left(\frac{\partial\alpha}{\partial\phi}\sigma_\phi\right)^2 + \left(\frac{\partial\alpha}{\partial\theta}\sigma_\theta\right)^2 + \left(\frac{\partial\alpha}{\partial\psi}\sigma_\psi\right)^2}_{\text{INS/GPS}} + \underbrace{\left(\frac{\partial\alpha}{\partial W_N}\sigma_{W_N}\right)^2 + \left(\frac{\partial\alpha}{\partial W_E}\sigma_{W_E}\right)^2}_{\text{Horizontal Wind}} + \underbrace{\left(\frac{\partial\alpha}{\partial W_D}\sigma_{W_D}\right)^2}_{\text{Vertical Wind}} \tag{24}$$

$$\boldsymbol{P}_{\alpha\beta}(2,2) = \sigma_\beta^2 = \underbrace{\left(\frac{\partial\beta}{\partial V_N}\sigma_{V_N}\right)^2 + \left(\frac{\partial\beta}{\partial V_E}\sigma_{V_E}\right)^2 + \left(\frac{\partial\beta}{\partial V_D}\sigma_{V_D}\right)^2 + \left(\frac{\partial\beta}{\partial\phi}\sigma_\phi\right)^2 + \left(\frac{\partial\beta}{\partial\theta}\sigma_\theta\right)^2 + \left(\frac{\partial\beta}{\partial\psi}\sigma_\psi\right)^2}_{\text{INS/GPS}} + \underbrace{\left(\frac{\partial\beta}{\partial W_N}\sigma_{W_N}\right)^2 + \left(\frac{\partial\beta}{\partial W_E}\sigma_{W_E}\right)^2}_{\text{Horizontal Wind}} + \underbrace{\left(\frac{\partial\beta}{\partial W_D}\sigma_{W_D}\right)^2}_{\text{Vertical Wind}} \tag{25}$$

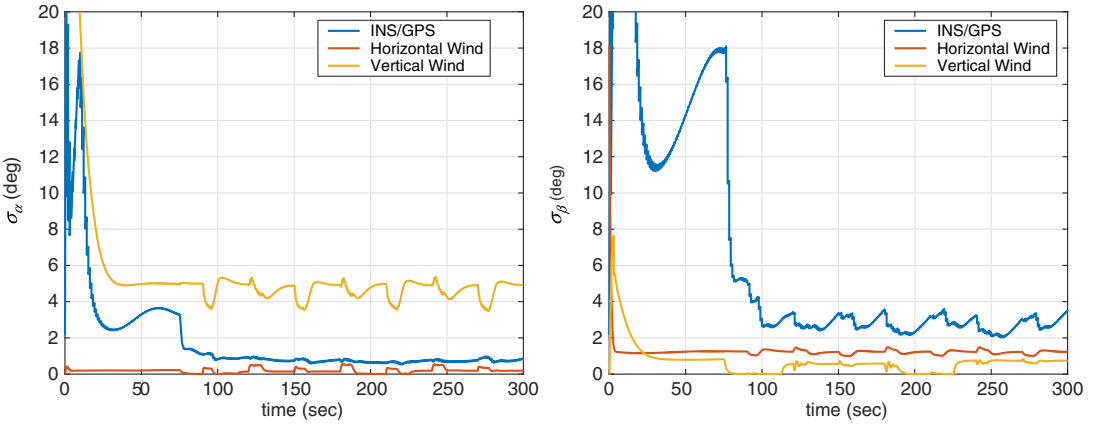
The off-diagonal terms of $\boldsymbol{P}_{\alpha\beta}$ representing correlations in the errors between the estimates of α and β are not discussed here. The key error sources are INS/GPS filter estimation errors and the uncertainty in the 3-D wind states. The $\sigma_{(\cdot)}$ in the first six terms of Eqs. (24) and (25) primarily represent the IMU errors and GPS measurement noise, and $\sigma_{W_{(\cdot)}}$ terms represent the wind modeling uncertainty.

Because all the derivatives and $\sigma_{(\cdot)}$ terms in both Eqs. (24) and (25) are time-varying, the time histories of the terms are examined. Figure 10 shows the error contributions of the three grouped terms in Eqs. (24) and (25) for σ_α and σ_β. The uncertainty of the downward wind estimate is seen to be the biggest uncertainty factor in σ_α and the smallest in σ_β. This points to one of the possible solutions for improving the accuracy of α and β: use higher-quality IMU and GPS sensors. In the following section, various IMU and GPS sensors are simulated to assess the impact on the accuracy of α and β.

A. Inertial Measurement Unit and Global Positioning System Contribution

The quality of the IMU and GPS measurement depend on the quality of the sensors. The output of each inertial sensor (gyro or accelerometer) of the IMU is modeled as an additive noise and bias. The additive noise is assumed to be wide-band noise with a given variance. The bias is assumed to be a first-order Gauss–Markov process, which is characterized by a variance and a time constant. For simplicity, we assume that all three accelerometers and all three gyros in the IMU are identical. Thus, a total of six parameters are required to model the IMU errors. For the GPS, an additive noise model is assumed in both position and velocity measurements. Thus, the impact on the accuracy of α and β can be examined by varying the error model parameters in a Monte Carlo analysis. Both IMU and GPS parameters and their possible ranges are listed in Table 5.

Figure 11 summarizes the key results from the Monte Carlo simulation conducted to assess how each parameter affects the accuracy of α and β. A starting point for these simulations is what we call the baseline filter. This is a filter whose sensor error



a) Various error contribution in σ_α b) Various error contribution in σ_β

Fig. 10 Various error contribution in σ_α and σ_β.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

characteristics are given in Table 4. For each Monte Carlo run, a particular error parameter in the baseline filter is perturbed to assess the impact on the accuracy of the estimates of α and β. The process and measurement noise matrices are changed according to the assigned parameter in each run. For each set of parameters, 20 Monte Carlo simulations are conducted. The averages of the final steady-state values of σ_α and σ_β for all 20 runs are plotted against various parameters.

Figure 11a and 11b show how the gyro and accelerometer parameters affect the accuracy of α and β. In Fig. 11a, σ_α is insensitive to both σ_{gw} and σ_{aw}, whereas σ_β is more sensitive to σ_{aw} than σ_{gw}. In Fig. 11b, both σ_α and σ_β are insensitive to σ_{gd} and σ_{ad}. What these results show is that care has to be taken when choosing the correct parameters. For example, a better IMU does not necessarily improve accuracy much in terms of the σ of α and β.

Figures 11c and 11d show how GPS position and velocity noise affect the accuracy of α and β. In Fig. 11c, σ_α stays around 5 deg and occasionally exceeds 5 deg. The phenomenon is more apparent in σ_β. σ_β often jumps to abnormal values when σ_{P_D} is greater than 1 m. Overall, both α and β are much more sensitive to σ_{P_D} than $\sigma_{P_{NE}}$.

Figure 11d shows that α and β are more sensitive to σ_{V_D} than $\sigma_{V_{NE}}$ in terms of the σ value. Only a small range of $\sigma_{V_{(\cdot)}}$ is shown because anything greater than those values causes the σ values to become much worse or even diverge. Comparing the range of the GPS position and velocity used in this analysis, it is found that α and β are more sensitive to GPS velocity measurement error than GPS position measurement error. The effect of GPS is greater than the effect of IMU on the accuracy of α and β, and having a better GPS can reduce uncertainty significantly. Therefore, to obtain measurements sufficient to achieve the accuracy shown in this paper using the proposed estimator, a high-price-range consumer-grade GPS and a middle-price-range IMU are needed.

Table 5 INS/GPS sensor error model parameters for tradeoff study

Sensor	Variable	Range (tactical to consumer grade [36])
GPS	$\sigma_{P_D}, \sigma_{P_E}, \sigma_{P_D}$	0.01–10 m
	$\sigma_{V_N}, \sigma_{V_E}, \sigma_{V_D}$	0.01–1 m/s
Accelerometer	σ_{aw}	0.05–0.1 m/s^2
	σ_{ad}	$5\times10^{-6} - 5\times10^{-3}$ g
	τ_a	50–300 s
Rate gyro	σ_{gw}	0.01–5 deg/s
	σ_{gd}	0.005–1080 deg/h
	τ_g	50–300 s

B. Wind Limitation

The previous section has narrowed down how IMU and GPS affect the accuracy of α and β. In this section, the effect of wind parameters on the accuracy of α and β is examined. It is stated earlier that the largest error source that affects the variance of α is the wind parameters, and it is also the secondary error source contributing to the variance of β. Furthermore, the uncertainty σ_{W_D} is the biggest error source entering the estimator.

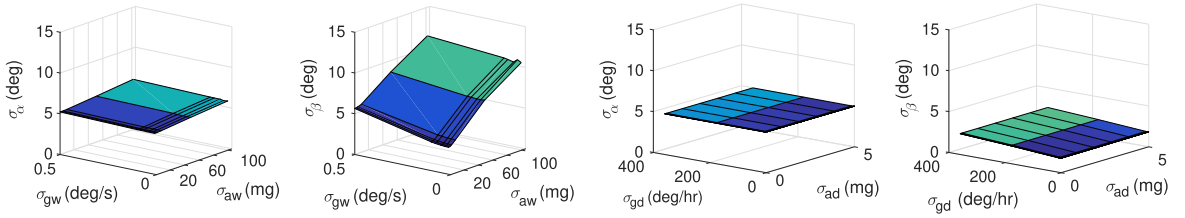
The wind is modeled by the first-order Gauss–Markov model, and the parameters are determined empirically. Ideally, the three components of the wind can be all equal to zero on a very calm day. However, there is usually a steady dominating horizontal wind. When this happens, the downward wind W_D (prevalent wind) is usually relatively small compared to the W_N and W_E components, and it is mainly dominated by wind gust and turbulence. To capture the downward wind in the presence of a dominating horizontal wind, the variance of the Markov process for the downward wind has to be larger than the variance of the Markov process for the horizontal wind. This relationship is shown in Eq. (26), where the variance of the downwind has to be greater than both of the variance of the north and east wind components, and it is empirically determined that the ratio of the variance between 1.5 and 10 works well:

$$\frac{2\sigma_{W_N}^2}{\tau_{W_N}} = \frac{2\sigma_{W_E}^2}{\tau_{W_E}} < \frac{2\sigma_{W_D}^2}{\tau_{W_D}} \tag{26a}$$

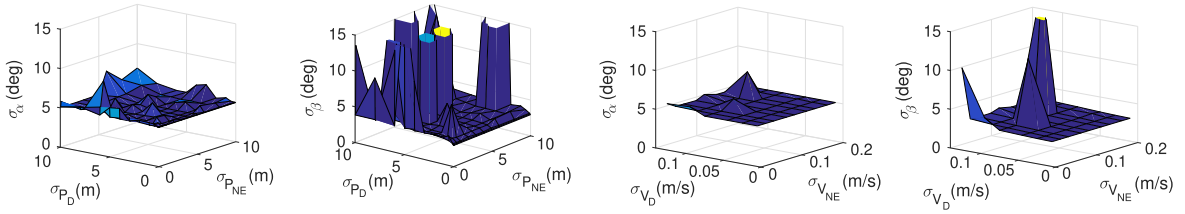
$$\frac{2\sigma_{W_D}^2/\tau_{W_D}}{2\sigma_{W_N}^2/\tau_{W_N}} \in (1.5, 10) \tag{26b}$$

The time constant τ_W reflects how fast the wind is changing over time. The maneuvering time τ_M has to be less than or equal to τ_W [Eq. (23)]. The UAV used in this paper has a time constant 0.14 s of the short-period mode, and the short-period mode reflects how fast the aircraft can quickly self-damp when the stick is briefly displaced. To satisfy Eq. (23), τ_{W_N} and τ_{W_E} are chosen to be 1 s. Similarly, the time constant of the phugoid mode is 9.53 s, and the phugoid is caused by a repeated exchange of airspeed and altitude. This relates to the vertical excitation in the downward wind direction; thus, τ_{W_D} is chosen to be 10 s. Although the analysis to determine parameters here is heuristic, more rigorous work can and should be done in the future.

With predetermined time constant τ_W, it is found that setting $\sigma_{W_N} = \sigma_{W_E} = 0.5$ m/s and $\sigma_{W_D} = 2$ m/s works well with both simulation and flight data when there is a dominating horizontal wind (3–6 m/s). If the flight is conducted in a very calm day, then the



a) Sensitivity analysis with respect to σ_{aw} and σ_{gw} b) Sensitivity analysis with respect to σ_{ad} and σ_{gd}



c) Sensitivity analysis with respect to $\sigma_{P_{NE}}$ and σ_{P_D} d) Sensitivity analysis with respect to $\sigma_{V_{NE}}$ and σ_{V_D}

Fig. 11 Sensitivity of α and β with various parameters in IMU and GPS.

performance of the estimator will be degraded because the wind condition does not match the presumed wind modeling parameter setting.

The estimator can be improved if the uncertainty in the wind is small. If there is an external sensor (e.g., accurate weather broadcast data) that can measure the 3-D wind with good accuracy, the accuracy of α and β can be improved. Improved wind models usually account for gust and turbulence, but it does not reduce the uncertainty in the wind. Therefore, for any synthetic air data estimation, the uncertainty in wind limits the accuracy of α and β.

C. Airspeed Limitation

For small UAVs, the airspeed is usually low and in the range of 12–30 m/s. The impact of airspeed on the SADS estimates can be determined by using the following error equations:

$$\delta\alpha = -\frac{w}{u^2+w^2}\delta u + \frac{u}{u^2+w^2}\delta w \tag{27}$$

$$\delta\beta = -\frac{uv}{V_a^2\sqrt{u^2+w^2}}\delta u + \frac{u^2+v^2}{V_a^2\sqrt{u^2+w^2}}\delta v - \frac{vw}{V_a^2\sqrt{u^2+w^2}}\delta w \tag{28}$$

If $u \gg v$ and $u \gg w$, Eqs. (27) and (28) can be further simplified as the following:

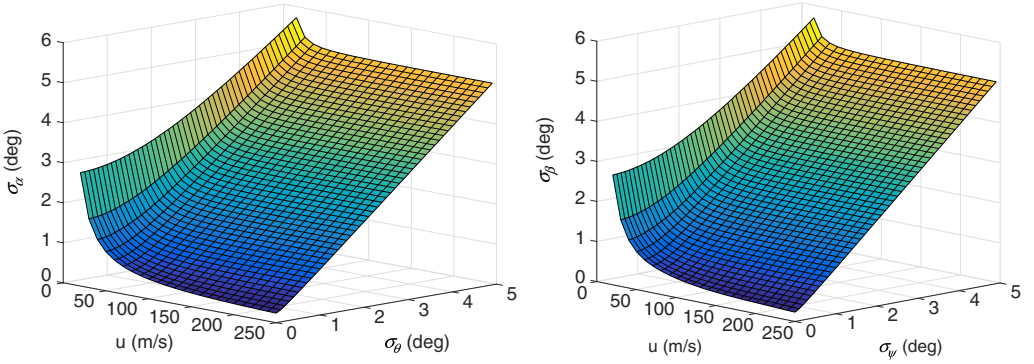
$$\delta\alpha \approx \frac{1}{u}\delta w = \delta\theta + \frac{1}{u}(\cos\theta\delta V_N - \cos\theta\delta W_N - \sin\theta\delta V_D + \sin\theta\delta W_D) \tag{29}$$

$$\delta\beta \approx \frac{1}{u}\delta v = -\delta\psi + \frac{1}{u}(-\sin\psi\delta V_N + \sin\psi\delta W_N + \cos\psi\delta V_E - \cos\psi\delta W_E) \tag{30}$$

From Eqs. (29) and (30), if the airspeed is large (represented by the u term), then the accuracy of α and β largely depends the accuracy of the attitude estimate. The uncertainty of inertial velocity and wind velocity are suppressed by the large airspeed. In other words, when the aircraft is operating at a high speed, the magnitude of the wind is much smaller than the airspeed, and so the wind uncertainty is relatively small.

By squaring each term in Eqs. (29) and (30), and assigning some typical values for $\delta V_{(\cdot)}$ and $\delta W_{(\cdot)}$ (the σ_W and σ_V values in Tables 4 and 5), and some reference point at $\theta = 5$ deg and $\psi = 0$ deg, the following trend shown in Fig. 12 is obtained for σ_α and σ_β. From Fig. 12, σ_α and σ_β are well below 1 deg when u is greater than 40 m/s, given an accurate σ_θ and σ_ψ (e.g., 0.5 deg) as an example.

Increasing airspeed would improve the accuracy of α and β if the attitude accuracy is reasonable. However, because small UAVs cannot usually operate over 25 m/s, σ_α and σ_β can only be reduced to some extent even if the attitude is relatively accurate.



a) Effects of airspeed and attitude on the accuracy of σ_α b) Effects of airspeed and attitude on the accuracy of σ_β

Fig. 12 Effects of airspeed and attitude on the accuracy of σ_α and σ_β.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

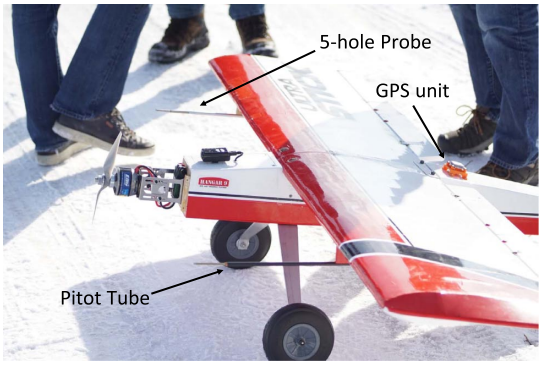


Fig. 13 Ultra Stick 120 setup.

VII. Flight Test

Experimental validation of the model-free synthetic air data estimator was accomplished by postprocessing the flight-test data collected using the UltraStick 120 testbed shown in Fig. 13. The goal of the flight test was to compare the performance of the estimator against the traditional air data system installed on the aircraft. The airframe is a low-cost fixed-wing radio-controlled aircraft with standard elevator, aileron, and rudder control surfaces. The physical and geometric properties can be found in [1]. The Ultra Stick 120 is equipped with a traditional pitot static system, a GPS receiver (u-blox-Neo-M8N), and an IMU (Invensense MPU-9250). These sensors are used to generate the α and β estimates. The ground truth for α and β is supplied by a five-hole probe [44,45]. A five-hole probe is a specially designed pitot tube that provides α, β, and V_a measurements. The five-hole probe is calibrated with wind-tunnel data initially and then calibrated in-situ to account for local flow conditions and alignment. In addition, an integrated INS/GPS system onboard provides position, inertial velocity, and attitude estimates.

Table 6 IMU and GPS sensor error model parameters for flight test

Sensor	Setting
GPS	$\sigma_{P_N} = \sigma_{P_E} = 3$ m, $\sigma_{P_D} = 6$ m
	$\sigma_{V_N} = \sigma_{V_E} = 0.5$ m/s, $\sigma_{V_D} = 1$ m/s
Accelerometer	$\sigma_{\text{aw}} = 0.05$ m/s^2
	$\sigma_{\text{ad}} = 0.01$ m/s^2
	$\tau_a = 100$ s
Rate gyro	$\sigma_{\text{gw}} = 0.00175$ rad/s
	$\sigma_{\text{gd}} = 0.00025$ rad/s
	$T_g = 50$ s

The flight data used in this paper were collected on 21 June 2018. Although the main objective of the test was not solely for the air data estimation, the flight maneuvers were designed to provide sufficient excitation for the estimation. Wind circles, push-over–pull-up (± 2 g), and rudder-induced steady forward slip maneuvers were conducted before other flight maneuver tests. The sensor error model parameters of IMU and GPS used for this flight test are given in Table 6. The wind parameters are the same as the one used in the simulation. The IMU runs at 50 Hz for the time update, whereas the GPS runs at 1 Hz for the measurement update.

Finally, to validate the wind estimation results, 3-D wind components are collected during the flight test by a remote anemometer located about 2.2 km away from the flight-test site. The anemometer is located at a height of 127.9 m, and its measurements are taken at 20 Hz. Although the measurements are not directly taken over the test field, it provides a reasonable check for the wind estimation, particularly in the wind direction.

Figures 14 and 15 show a portion of the flight-test results. Figure 14 shows the 3-D wind estimates and the mean of the wind measurements taken by the remote anemometer. The three standard deviations of the wind components are also calculated and plotted in Fig. 14. The magnitude of these estimates are very similar to the mean values. They are also within the three standard deviations. From Figs. 15a and 15b, it can be seen that the α estimate is reasonably good compared to the five-hole calibrated α measurement, and the estimate generally stays within the $3-\sigma$ bound. On the other hand, the β estimate momentarily goes out of the $3-\sigma$ bound right after a sharp turn at $t \approx 1275$ s. This sharp turn is manifested by the significant change in ϕ from a wing-level condition where β stays constant at about -5 deg. This is due to INS/GPS attitude estimation errors. To further analyze why the estimate is biased in this turn as well as to understand the behavior of the $3-\sigma$ bound of α and β, another

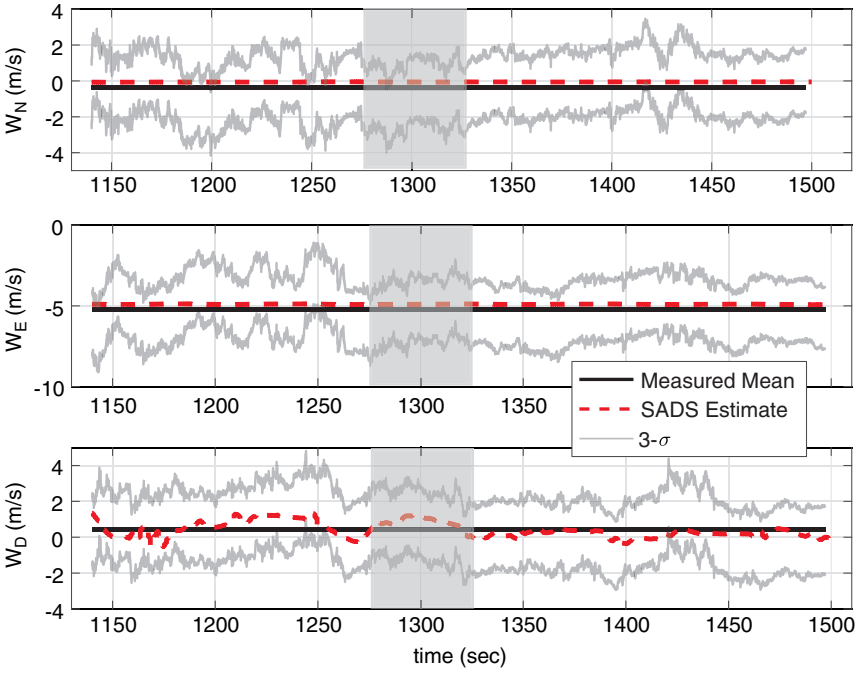
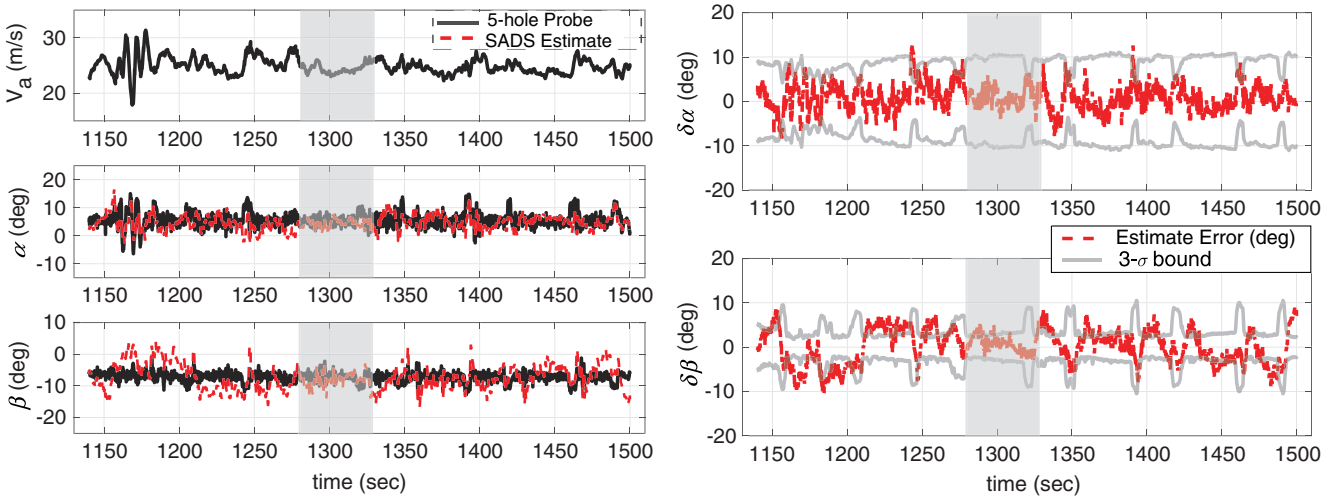


Fig. 14 Flight-test results: 3-D wind estimates.

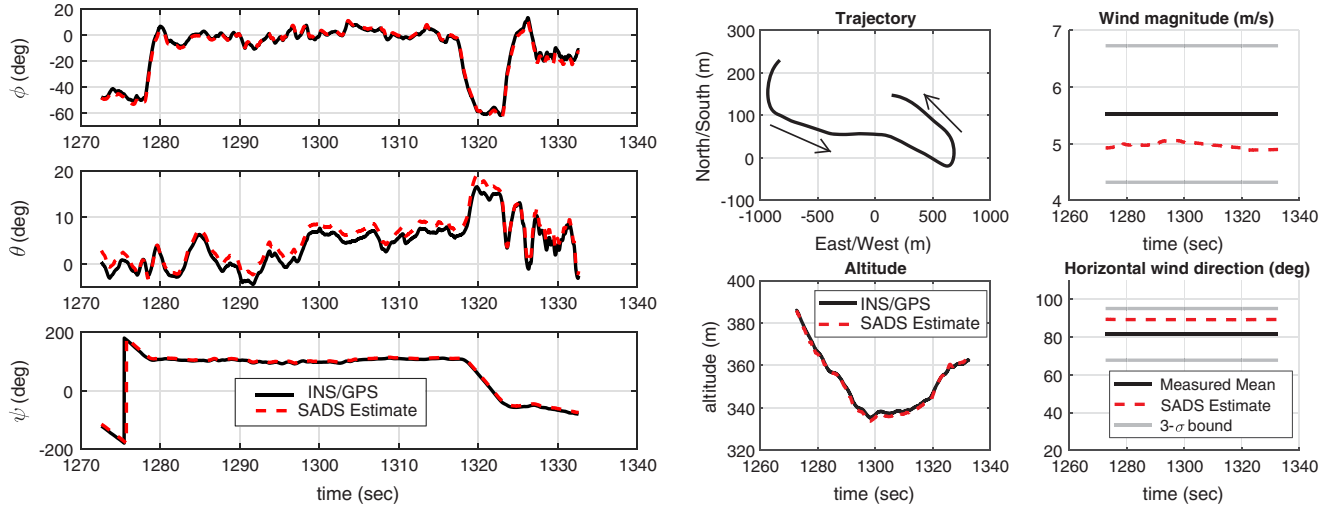
Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290



a) Flight test airspeed measurement and α and β estimates

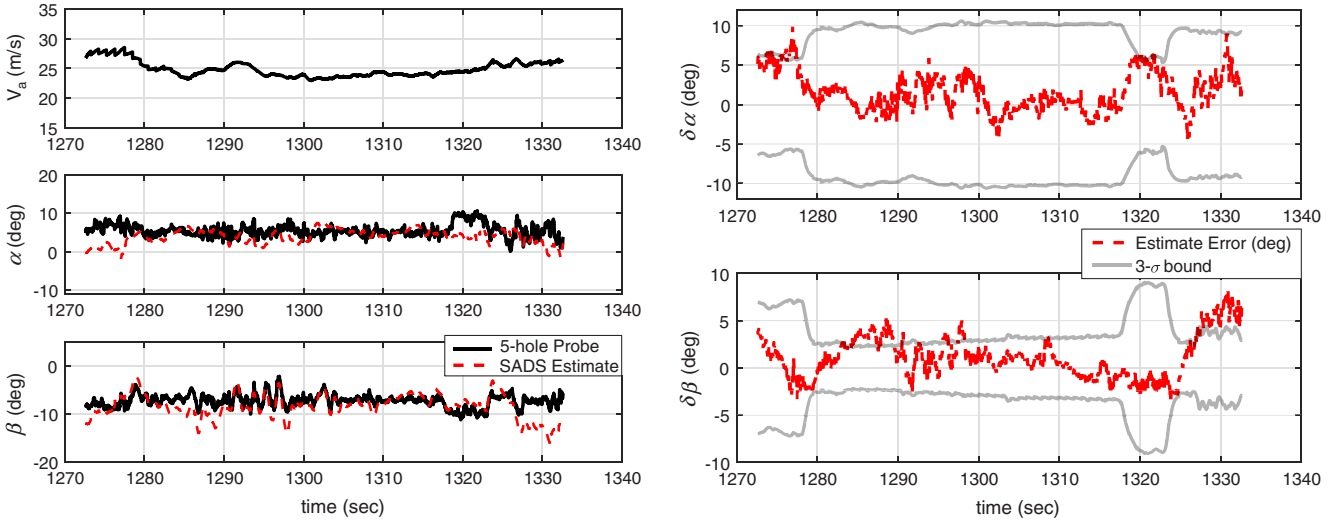
b) Flight test α and β estimate error and its $3-\sigma$ bound

Fig. 15 Flight-test results: α and β.



a) Highlighted flight test euler angle INS/GPS Nav solution vs. estimates

b) Highlighted flight test trajectory, altitude, wind estimates



c) Highlighted flight test airspeed measurement and α and β estimates

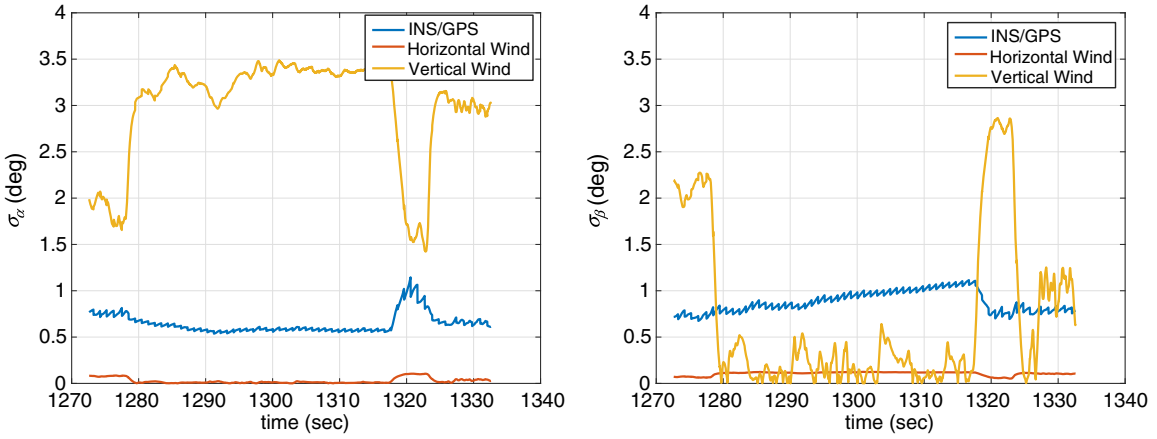
d) Highlighted flight test α and β estimate error

Fig. 16 Flight-test results of the highlighted gray region.

section of the flight-test results (around $t = 1300$ s and highlighted by the gray shading) is examined closely in Fig. 16.

The trajectory, altitude, wind magnitude, and direction estimates of the highlighted gray region are shown in Fig. 16b. The UAV just made a sharp turn before flying to the east direction. Then, the UAV was fighting against the wind before making another sharp turn back to the west direction. Figures 16c and 16d show the estimates, estimation errors, and $3-\sigma$ bound for α and β in the highlighted

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290



a) Error contribution to σ_α during the grey flight test region

b) Error contribution to σ_β during the grey flight test region

Fig. 17 Error contribution to σ_α and σ_β in the gray highlighted region of Figs. 15a and 15b.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

region. The $1-\sigma$ bounds for α and β shown in Fig. 16d are determined to be 3.07 and 1.28 deg, respectively, which is a significant improvement compared to the simulation results, due to the more aggressive maneuver performed in the real flight test. The α estimate is reasonably accurate before the sharp turn. Interestingly, the β estimate seems to be accurate during the turn but starts going out of the bound after the turn. Although this maneuver can provide good excitation for the estimator, the maneuver itself is aggressive, which can result in poor estimates in attitude, inertial velocities, and wind estimates. This can be seen in the deviation in the θ estimate from the INS/GPS solution as well as the changes in wind estimates. Also, the uncertainties in all inertial velocity estimates increase during the sharp turns (not shown here), which leads to errors in α and β.

The poor estimates can also be explained by the various error contributions shown in Fig. 17. When the attitude maneuvers are benign ($1280 \leq t \leq 1320$ s), the biggest error contribution to α comes from the uncertainty in the vertical wind estimate, and the second largest contribution comes from the INS/GPS filter's attitude estimation errors. The smallest error in σ_α comes from the uncertainty in the horizontal wind estimate. When the sharp turn occurs ($t \approx 1320$ s), the error from the vertical wind estimate drops drastically because the sharp turn provides a sudden huge excitation in the downward direction, whereas the uncertainty in both INS/GPS and horizontal wind increase. The INS/GPS filter normally does not provide good estimates when there is a sudden dynamic change and sensor biases are not accurately estimated. This explains the increase in the INS/GPS error. The increase of the uncertainty in the horizontal wind estimate is also caused by the sudden directional change of the aircraft. All the error contributors of σ_α return to the nominal level when the aircraft finishes the sharp turn.

The error contributions in σ_β behavior is slightly different from the simulation shown previously in Fig. 10. The error of INS/GPS ramps up during a relatively straight flight and decreases during the sharp turn. This is expected because the heading state ψ of INS/GPS becomes less observable when there is no vehicle acceleration (e.g., turning flight). Once the turn is made, the degree of observability increases again. The error contribution from the vertical wind increases significantly when the sharp turn occurs. This signals that the error from the vertical wind does play a huge factor in the β accuracy. That is, the β estimate should not be relied on when the error contribution from the vertical wind exceeds the INS/GPS error contribution.

VIII. Conclusions

This paper presented a method for estimating 3-D wind vector, angle of attack, and angle of sideslip without the aerodynamic model of the aircraft. The observability analysis shows that the system is conditionally observable, provided that wind speed and direction do not change faster than the aircraft dynamics. A sensitivity analysis was performed to show what factors affect the accuracy of the angle of attack and sideslip estimates. Specifically, the variance of the downward wind is the largest limiting factor to the angle of attack and angle of sideslip estimations. A low airspeed would also decrease the accuracy due to the relative increasing wind effect on the aircraft. Furthermore, the accuracy of the attitude estimation also plays a significant role in estimating angle of attack and angle of sideslip. The conclusion of this is that an inexpensive, model-free air data estimator can still provide reasonably accurate estimate of α and β ($1-\sigma$ bound for a slow-flying unmanned aerial vehicle of approximately 5 and 3 deg, respectively) in the absence of an air data system and a dynamics model of the aircraft. The method and results presented in this paper suggest that a synthetic air data systems estimator can potentially be used as part of a fault detection and isolation scheme for traditional air data systems.

Appendix: Derivation of Flow Angle Covariance

To calculate all the components of $A_{\alpha\beta}$, recall that α and β can be calculated as

$$\alpha = \tan^{-1}\frac{w}{u},\ \beta = \sin^{-1}\frac{v}{\sqrt{u^2+v^2+w^2}} \tag{A1}$$

Using the wind triangle kinematic relationship and Eq. (A1), all the components can be calculated as follows:

$$\begin{aligned}\frac{\partial\alpha}{\partial(\cdot)} &= \frac{\partial\alpha}{\partial u}\frac{\partial u}{\partial(\cdot)} + \frac{\partial\alpha}{\partial w}\frac{\partial w}{\partial(\cdot)}\\ \frac{\partial\beta}{\partial(\cdot)} &= \frac{\partial\beta}{\partial u}\frac{\partial u}{\partial(\cdot)} + \frac{\partial\beta}{\partial v}\frac{\partial v}{\partial(\cdot)} + \frac{\partial\beta}{\partial w}\frac{\partial w}{\partial(\cdot)}\end{aligned} \tag{A2}$$

where the partial derivative with respect to the position, accelerometer, and gyro biases is zero. For completion, all the partial derivatives involved are shown in Eqs. (A3) and (A4):

$$\begin{aligned}&\frac{\partial\alpha}{\partial u} = -\frac{w}{u^2+w^2} \quad \frac{\partial\alpha}{\partial v} = 0 \quad \frac{\partial\alpha}{\partial w} = \frac{u}{u^2+w^2}\\ &\frac{\partial\beta}{\partial u} = -\frac{uv}{V_a^2\sqrt{V_a^2-v^2}} \quad \frac{\partial\beta}{\partial v} = \frac{1-(v^2/V_a^2)}{\sqrt{V_a^2-v^2}} \quad \frac{\partial\beta}{\partial w} = -\frac{vw}{V_a^2\sqrt{V_a^2-v^2}}\end{aligned} \tag{A3}$$

$$\frac{\partial u}{\partial V_N} = \cos\theta\cos\psi, \qquad \frac{\partial u}{\partial V_E} = \cos\theta\sin\psi, \qquad \frac{\partial u}{\partial V_D} = -\sin\theta, \qquad \frac{\partial u}{\partial \phi} = 0$$

$$\frac{\partial u}{\partial \theta} = -\sin\theta\cos\psi(V_N - W_N) - \sin\theta\sin\psi(V_E - W_E) - \cos\theta(V_D - W_D)$$

$$\frac{\partial u}{\partial \psi} = -\cos\theta\sin\psi(V_N - W_N) + \cos\theta\cos\psi(V_E - W_E), \qquad \frac{\partial u}{\partial W_N} = -\cos\theta\cos\psi, \qquad \frac{\partial u}{\partial W_E} = -\cos\theta\sin\psi$$

$$\frac{\partial u}{\partial W_D} = \sin\theta, \quad \frac{\partial v}{\partial V_N} = \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi, \quad \frac{\partial v}{\partial V_E} = \sin\phi\sin\theta\sin\psi - \cos\phi\cos\psi, \quad \frac{\partial v}{\partial V_D} = \sin\phi\cos\theta$$

$$\frac{\partial v}{\partial \phi} = (\cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi)(V_N - W_N) + (\cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi)(V_E - W_E) + \cos\phi\cos\theta(V_D - W_D)$$

$$\frac{\partial v}{\partial \theta} = (\sin\phi\cos\theta\cos\psi)(V_N - W_N) + \sin\phi\cos\theta\sin\psi(V_E - W_E) - \sin\phi\sin\theta(V_D - W_D)$$

$$\frac{\partial v}{\partial \psi} = (-\sin\phi\sin\theta\sin\psi - \cos\phi\cos\psi)(V_N - W_N) + (\sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi)(V_E - W_E)$$

$$\frac{\partial v}{\partial W_N} = -(\sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi), \qquad \frac{\partial v}{\partial W_E} = -(\sin\phi\sin\theta\sin\psi - \cos\phi\cos\psi), \qquad \frac{\partial v}{\partial W_D} = -\sin\phi\cos\theta$$

$$\frac{\partial w}{\partial V_N} = \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi, \qquad \frac{\partial w}{\partial V_E} = \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi, \qquad \frac{\partial w}{\partial V_D} = \cos\phi\cos\theta$$

$$\frac{\partial w}{\partial \phi} = (-\sin\phi\sin\theta\cos\psi + \cos\phi\sin\psi)(V_N - W_N) + (-\sin\phi\sin\theta\sin\psi - \cos\phi\cos\psi)(V_E - W_E) - \sin\phi\cos\theta(V_D - W_D)$$

$$\frac{\partial w}{\partial \theta} = (\cos\phi\cos\theta\cos\psi)(V_N - W_N) + \cos\phi\cos\theta\sin\psi(V_E - W_E) - \cos\phi\sin\theta(V_D - W_D)$$

$$\frac{\partial w}{\partial \psi} = (-\cos\phi\sin\theta\sin\psi + \sin\phi\cos\psi)(V_N - W_N) + (\cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi)(V_E - W_E)$$

$$\frac{\partial w}{\partial W_N} = -(\cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi), \qquad \frac{\partial w}{\partial W_E} = -(\cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi), \qquad \frac{\partial w}{\partial W_D} = -\cos\phi\cos\theta \tag{A4}$$

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

Acknowledgments

The authors gratefully acknowledge UTC Aerospace Systems in Burnsville, Minnesota, for donating the five-hole pitot tubes that are used in this work. The authors also acknowledge the Minnesota Invasive Terrestrial Plants Center for financial support to research issues associated with increasing the reliability of small UAV technology used for surveying applications. The flight data described in this paper are collected and conducted by University of Minnesota under contract to Systems Technology Inc. as part of a Phase II Small Business Innovation Research (SBIR) program for NASA Langley Research Center that is investing Unmanned Aerial System (UAS) handling qualities. The authors would also like to thank NASA (via grant NNX15AV67G), University of Minnesota–The Office of the Vice President for Research–Minnesota's Discovery, Research, and InnoVation Economy (UMN OVPR MnDRIVE) Initiative, Sentera LLC, and LMCCR Legislature of the State of Minnesota for supporting this work. Finally, the authors also gratefully acknowledge the help of Curtis Olson, member of the UAV lab, for conducting flight tests, and the University of Minnesota Eolos Wind Energy Research Group for the wind measurements. However, any opinions, findings, conclusions, or recommendations in this paper are those of the authors and do not necessarily reflect views of any of the preceding organizations.

References

[1] Lie, F. A. P., and Gebre-Egziabher, D., "Synthetic Air Data System," *Journal of Aircraft*, Vol. 50, No. 4, 2013, pp. 1234–1249. doi:10.2514/1.C032177

[2] Nebula, F., Palumbo, R., Morani, G., and Corraro, F., "Virtual Air Data System Architecture for Space Reentry Applications," *Journal of Spacecraft and Rockets*, Vol. 46, No. 4, 2009, pp. 818–828. doi:10.2514/1.42485

[3] Lie, F. A. P., and Gebre-Egziabher, D., "Sensitivity Analysis of Model-Based Synthetic Air Data Estimators," *AIAA Guidance, Navigation, and Control Conference*, AIAA Paper 2015-0081, 2015. doi:10.2514/6.2015-0081

[4] Johansen, T. A., Cristofaro, A., SØrensen, K., Hansen, J. M., and Fossen, T. I., "On Estimation of Wind Velocity, Angle-of-Attack and Sideslip Angle of Small UAVs Using Standard Sensors," *2015 International Conference on Unmanned Aircraft Systems*, IEEE Publ., Piscataway, NJ, 2015, pp. 510–519. doi:10.1109/ICUAS.2015.7152330

[5] Wenz, A., Johansen, T. A., and Cristofaro, A., "Combining Model-Free and Model-Based Angle of Attack Estimation for Small Fixed-Wing UAVs Using a Standard Sensor Suite," *2016 International Conference on Unmanned Aircraft Systems*, IEEE Publ., Piscataway, NJ, 2016, pp. 624–632. doi:10.1109/ICUAS.2016.7502583

[6] Tian, P., and Chao, H., "Model Aided Estimation of Angle of Attack, Sideslip Angle, and 3D Wind Without Flow Angle Measurements," *2018 AIAA Guidance, Navigation, and Control Conference*, AIAA Paper 2018-1844, 2018. doi:10.2514/6.2018-1844

[7] Sun, K., Regan, C. D., and Egziabher, D. G., "GNSS/INS Based Estimation of Air Data and Wind Vector Using Flight Maneuvers," *2018 IEEE/ION Position, Location and Navigation Symposium*, IEEE Publ., Piscataway, NJ, 2018, pp. 838–849. doi:10.1109/PLANS.2018.8373461

[8] Klein, V., and Morelli, E. A., *Aircraft System Identification: Theory and Practice*, AIAA, Reston, VA, 2006, Chaps. 3, 6.

[9] Rossi, M., and Brunelli, D., "Autonomous Gas Detection and Mapping with Unmanned Aerial Vehicles," *IEEE Transactions on Instrumentation and Measurement*, Vol. 65, No. 4, 2016, pp. 765–775. doi:10.1109/TIM.2015.2506319

[10] Fravolini, M. L., del Core, G., Papa, U., Valigi, P., and Napolitano, M. R., "Data-Driven Schemes for Robust Fault Detection of Air Data System Sensors," *IEEE Transactions on Control Systems Technology*, Vol. 27, No. 1, Jan. 2019, pp. 1–15. doi:10.1109/TCST.2017.2758345

[11] Ossmann, D., "Enhanced Detection and Isolation of Angle of Attack Sensor Faults," *AIAA Guidance, Navigation, and Control Conference*, AIAA Paper 2016-1135, 2016. doi:10.2514/6.2016-1135

[12] Ossmann, D., Joos, H.-D., and Goupil, P., "Enhanced Sensor Monitoring to Maintain Optimal Aircraft Handling in Case of Faults,"

Journal of Guidance, Control, and Dynamics, Vol. 40, No. 12, 2017, pp. 3127–3137.
doi:10.2514/1.G002341

[13] Zeis, J. E. J., "Angle of Attack and Sideslip Estimation Using an Inertial Reference Platform," M.S. Thesis, U.S. Air Force Inst. of Technology, Wright-Patterson AFB, OH, June 1988.

[14] Colgren, R., "The Feasibility of Using an INS for Control System Feedbacks," *World Aviation Conference*, AIAA Paper 1998-5517, 1998.
doi:10.2514/6.1998-5517

[15] Colgren, R., Frye, M., and Olson, W., "A Proposed System Architecture for Estimation of Angle-of-Attack and Sideslip Angle," *Guidance, Navigation, and Control and Conference and Exhibit*, AIAA Paper 1999-4078, 1999.
doi:10.2514/6.1999-4078

[16] Colgren, R., and Martin, K., "Flight Test Validation of Sideslip Estimation Using Inertial Accelerations," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2000-4448, Aug. 2000.
doi:10.2514/6.2000-4448

[17] Colgren, R. D., "Method and System for Elimination and Correction of Angle of Attack and Sideslip Angle from Acceleration Measurements," Patent No. 6,273,370 B1 , Aug. 2001, http://www.freepatentsonline.com/6273370.html.

[18] Heller, S., Myschik, S., Holzapfel, F., and Sachs, G., "Low-Cost Approach Based on Navigation Data for Determining Angles of Attack and Sideslip for Small Aircraft," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2003-5777, 2003.
doi:10.2514/6.2003-5777

[19] Wise, K., "Flight Testing of the X-45A J-UCAS Computational Alpha-Beta System," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2006-6215, Aug. 2006.
doi:10.2514/6.2006-6215

[20] Myschik, S., Holzapfel, F., and Sachs, G., "Low-Cost Sensor Based Integrated Airdata and Navigation System for General Aviation Aircraft," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2008-7423, Aug. 2008.
doi:10.2514/6.2008-7423

[21] Murch, A., "A Flight Control System Architecture for the NASA AirSTAR Flight Test Infrastructure," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2008-6990, Aug. 2008.
doi:10.2514/6.2008-6990

[22] Karlgaard, C. D., and Schoenenberger, M., "Planetary Probe Entry Atmosphere Estimation Using Synthetic Air Data System," *Journal of Spacecraft and Rockets*, Vol. 55, No. 3, 2018, pp. 599–610.
doi:10.2514/1.A34115

[23] Morelli, E. A., "Real-Time Aerodynamic Parameter Estimation Without Air Flow Angle Measurements," *Journal of Aircraft*, Vol. 49, No. 4, 2012, pp. 1064–1074.
doi:10.2514/1.C031568

[24] Cho, A., Kim, J., Lee, S., and Kee, C., "Wind Estimation and Airspeed Calibration Using a UAV with a Single-Antenna GPS Receiver and Pitot Tube," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 47, No. 1, 2011, pp. 109–117.
doi:10.1109/TAES.2011.5705663

[25] Langelaan, J., Alley, N., and Neidhoefer, J., "Wind Field Estimation for Small Unmanned Aerial Vehicles," *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 4, 2011, pp. 1016–1030.
doi:10.2514/1.52532

[26] Lee, J. H., Sevil, H. E., Dogan, A., and Hullender, D., "Estimation of Maneuvering Aircraft States and Time-Varying Wind with Turbulence," *Aerospace Science and Technology*, Vol. 31, No. 1, 2013, pp. 87–98.
doi:10.1016/j.ast.2013.09.009

[27] Goshen-Meskin, D., and Bar-Itzhack, I. Y., "Observability Analysis of Piece-Wise Constant Systems. I. Theory," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 4, 1992, pp. 1056–1067.
doi:10.1109/7.165367

[28] Goshen-Meskin, D., and Bar-Itzhack, I. Y., "Observability Analysis of Piece-Wise Constant Systems. 2. Application to Inertial Navigation In-Flight Alignment [Military Applications]," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 4, 1992, pp. 1068–1075.
doi:10.1109/7.165368

[29] Rhee, I., Abdel-Hafez, M. F., and Speyer, J. L., "Observability of An Integrated GPS/INS During Maneuvers," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 40, No. 2, 2004, pp. 526–535.
doi:10.1109/TAES.2004.1310002

[30] Bageshwar, V. L., Gebre-Egziabher, D., Garrard, W. L., and Georgiou, T. T., "Stochastic Observability Test for Discrete-Time Kalman Filters," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 4, 2009, pp. 1356–1370.
doi:10.2514/1.38128

[31] Krener, A. J., and Ide, K., "Measures of Unobservability," *Proceedings of the 48th IEEE Conference on Decision and Control*, IEEE Publ., Piscataway, NJ, 2009, pp. 6401–6406.
doi:10.1109/CDC.2009.5400067

[32] Hinson, B. T., Binder, M. K., and Morgansen, K. A., "Path Planning to Optimize Observability in a Planar Uniform Flow Field," *2013 American Control Conference*, IEEE Publ., Piscataway, NJ, 2013, pp. 1392–1399.
doi:10.1109/ACC.2013.6580031

[33] Hausman, K., Preiss, J., Sukhatme, G. S., and Weiss, S., "Observability-Aware Trajectory Optimization for Self-Calibration with Application to UAVs," *IEEE Robotics and Automation Letters*, Vol. 2, No. 3, 2017, pp. 1770–1777.
doi:10.1109/LRA.2017.2647799

[34] Dianetti, A. D., Weisman, R., and Crassidis, J. L., "Observability Analysis for Improved Space Object Characterization," *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 1, 2018, pp. 137–148.
doi:10.2514/1.G002229

[35] Ghilani, C., *Adjustment Computations: Spatial Data Analysis*, 5th ed., Wiley, Hoboken, NJ, 2010, pp. 89–90, Chap. 6.

[36] Gleason, S., and Gebre-Egizabher, D., *GNSS Applications and Methods*, Artech House, Norwood, MA, 2009, pp. 149–172, Chap. 6.

[37] Berman, Z., and Powell, J. D., "The Role of Dead Reckoning and Inertial Sensors in Future General Aviation Navigation," *IEEE 1998 Position Location and Navigation Symposium*, IEEE Publ., Piscataway, NJ, 1998, pp. 510–517.
doi:10.1109/PLANS.1998.670206

[38] Xing, Z., "Over-Bounding Integrated INS/GNSS Output Errors," Ph.D. Thesis, Univ. of Minnesota, Twin Cities, Minnesota, MN, Sept. 2010.

[39] Gracey, W., "Measurement of Aircraft Speed and Altitude," NASA TR 1046, May 1980.

[40] Golub, G., and van Loan, C., *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, MD, 2013, pp. 71–73, Chap. 2.

[41] Taylor, W., "Aircraft Simulation Baseline 2014 v1," Univ. of Minnesota's Digital Conservancy Minnesota, 2014, http://hdl.handle.net/11299/163868 [retrieved Sept. 2018].

[42] Dorobantu, A., Johnson, W., Lie, F. A., Taylor, B., Murch, A., Paw, Y. C., Gebre-Egziabher, D., and Balas, G., "An Airborne Experimental Test Platform: From Theory to Flight," *2013 American Control Conference*, IEEE Publ., Piscataway, NJ, 2013, pp. 659–673.
doi:10.1109/ACC.2013.6579912

[43] Owens, B., Cox, D., and Morelli, E., "Development of a Low-Cost Sub-Scale Aircraft for Flight Research: The FASER Project," *25th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*, AIAA Paper 2006-3306, June 2006.
doi:10.2514/6.2006-3306

[44] "Rosemount Model 858 Flow Angle Sensors," Rosemount, Bulletin 1014, Burnsville, MN, 1988.

[45] Parameswaran, V., Jategaonkar, R., and Press, M., "Five-Hole Flow Probe Calibration from Dynamic and Tower Flyby Maneuvers," *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 80–86.
doi:10.2514/1.5708

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.C035290

JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 43, No. 5, May 2020

Check for updates

Two-Stage Batch Algorithm for Nonlinear Static Parameter Estimation

Kerry Sun* and Demoz Gebre-Egziabher†
University of Minnesota, Minneapolis, Minnesota 55455

https://doi.org/10.2514/1.G004713

A two-stage batch estimation algorithm for solving a class of nonlinear, static parameter estimation problems that appear in aerospace engineering applications is proposed. It is shown how these problems can be recast into a form suitable for the proposed two-stage estimation process. In the first stage, linear least squares is used to obtain a subset of the unknown parameters (set 1) and a residual sampling procedure is used for selecting initial values for the rest of the parameters (set 2). In the second stage, depending on the uniqueness of the local minimum, either only the parameters in the second set need to be re-estimated, or all the parameters will have to be re-estimated simultaneously, by a nonlinear constrained optimization. The estimates from the first stage are used as initial conditions for the second-stage optimizer. It is shown that this approach alleviates the sensitivity to initial conditions and minimizes the likelihood of converging to an incorrect local minimum of the nonlinear cost function. An error bound analysis is presented to show that the first stage can be solved in such a way that the total cost function will be driven to the optimal cost, and the difference has an upper bound. Two tutorial examples are used to show how to implement this estimator and compare its performance to other similar nonlinear estimators. Finally, the estimator is used on a 5-hole Pitot tube calibration problem using flight test data collected from a small unmanned aerial vehicle that cannot be easily solved with single-stage methods.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Nomenclature

a_x, a_y, a_z	=	body-axis translational acceleration
$b_{a_x}, b_{a_y}, b_{a_z}$	=	bias of body-axis translational acceleration
b_p, b_q, b_r	=	bias of body-axis rotational velocity
$\boldsymbol{f}$	=	nonlinear dynamic model
g	=	gravitational acceleration
$\boldsymbol{h}$	=	nonlinear measurement model
K_α, K_α	=	sensitivity coefficients of angle-of-attack and sideslip
$N(\mu, \sigma^2)$	=	normal (Gaussian) distribution with mean μ and standard deviation σ
P_s, P_t	=	static and dynamic pressures
$P_{\Delta\alpha}$	=	differential angle-of-attack pressure
$P_{\Delta\beta}$	=	differential sideslip angle pressure
p, q, r	=	body-axis rotational velocity
$\boldsymbol{R}$	=	noise covariance matrix
t	=	time
u, v, w	=	body-axis translational velocity
$\boldsymbol{u}$	=	input vector
V_a	=	airspeed
$\boldsymbol{x}$	=	state vector
$\boldsymbol{y}$	=	true output vector
α	=	angle of attack
β	=	sideslip angle
ρ	=	air density
ϕ, θ, ψ	=	Euler angles

Superscripts

$(\cdot)^T$	=	transpose
$(\cdot)^{-1}$	=	matrix inverse
$\hat{(\cdot)}$	=	estimate of $(\cdot)$
$(\cdot)^*$	=	optimal value of $(\cdot)$

Received 13 July 2019; revision received 26 December 2019; accepted for publication 30 December 2019; published online 17 February 2020. Copyright © 2019 by Kerry Sun and Demoz Gebre-Egziabher. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. All requests for copying and permission to reprint should be submitted to CCC at www.copyright.com; employ the eISSN 1533-3884 to initiate your request. See also AIAA Rights and Permissions www.aiaa.org/randp.

*Ph.D. Candidate, Department of Aerospace Engineering and Mechanics, Twin Cities; sunx0486@umn.edu. Student Member AIAA.

†Professor, Department of Aerospace Engineering and Mechanics, Twin Cities; gebre@umn.edu. Associate Fellow AIAA.

I. Introduction

THIS paper presents an algorithm for solving a class of nonlinear estimation problems that appear in aerospace guidance, navigation, and control. These nonlinear estimation problems appear in applications such as vehicle system identification; sensor calibration; and vehicle positioning, navigation, and timing (PNT). In the past, these problems have been solved either by standard estimators (e.g., the Kalman filter or its many variants [1,2]; maximum likelihood estimators [3]; or output-error minimization [4–6]) or, in many instances, by ad hoc approaches developed for the particular problem at hand. It is the claim of this paper that a large number of these nonlinear estimation problems have a similar mathematical structure that can be exploited in a two-stage estimator. This estimator can overcome the initial condition sensitivity problem, have good convergence, and, in many instances, have a guaranteed estimation bound on the total cost function. In this paper, we describe this nonlinear mathematical structure and discuss why it arises in many aerospace sensing and estimation problems. Subsequently, we develop an estimator designed to exploit this nonlinear structure and provide examples to demonstrate its performance.

The class of nonlinear estimation problems that are the subject of this paper have the following form:

$$\boldsymbol{z}_k = \boldsymbol{A}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 + \boldsymbol{b}(\boldsymbol{\xi}_2) + \boldsymbol{v}_k \tag{1}$$

where $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T \quad \boldsymbol{\xi}_2^T]^T$ is the vector of parameters to be estimated, $\boldsymbol{z}_k$ is a measurement vector at any discrete time t_k, and $\boldsymbol{v}_k$ is the noise vector corrupting the measurement at t_k. The matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ are functions of the unknown parameters $\boldsymbol{\xi}_2$ only. This mathematical form appears often in parameter estimation problems. As we show later in the paper, this form arises when embedded in the problem at hand is the standard sensor error model that relates *measured* quantities $\boldsymbol{z}_k$ to their *true* values $\boldsymbol{y}_k$ given by the following mathematical relationship from Refs. [5] [Eq. (10.13)] and [7] [Eqs. (4.15–4.17)]:

$$\boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{y}_k, \boldsymbol{\xi}) + \boldsymbol{v}_k = \boldsymbol{C}\boldsymbol{y}_k + \boldsymbol{n}_k + \boldsymbol{v}_k \tag{2}$$

In the standard sensor error model given above, the matrix $\boldsymbol{C}$ is a matrix whose entries are a function of unknown sensor parameters (scale factor errors, axis misalignment errors, etc.), and the vector $\boldsymbol{n}_k$ consists of unknown null-shifts (biases). Both $\boldsymbol{C}$ and $\boldsymbol{n}_k$ are functions

of the parameter $\boldsymbol{\xi}$. The vector $\boldsymbol{v}_k$ is independent Gaussian white measurement noise. In the Appendices of this paper, we provide a general canonical form and two examples that show how the form of Eq. (1) arises from Eq. (2).

The algorithm proposed in this paper exploits the structure in Eq. (1) by using a two-stage estimation scheme. In the first stage, we solve a linear least-squares problem for the parameter vector $\boldsymbol{\xi}_1$, where the remainder of the unknowns in the parameter vector $\boldsymbol{\xi}_2$ are held fixed at some predetermined values (i.e., using prior knowledge or systematically selected). In the second stage, depending on the uniqueness of the local minimum, we solve a constrained nonlinear optimization problem for either $\boldsymbol{\xi}_2$ only (and $\boldsymbol{\xi}_1$ can be determined consequently), or all of the unknowns ($\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$) simultaneously, by using the estimates from the first stage as the initial conditions for the optimization. As will be demonstrated later, this formulation overcomes initial condition sensitivity issues and leads to excellent convergence properties and, in many instances, guaranteed upper bounds on the cost function.

A. Prior Work

The idea of solving nonlinear estimation problems in two stages is not new, and some of the earliest work relevant to the discussion here dates from the early 1970s [8–10]. In particular, Golub and Pereyra [10] dealt with a nonlinear parameter estimation problem by solving only a subset of the total parameters in the first stage. They used the idea of removing "conditionally linear" parameters to separate linear and nonlinear parameters [11]. It was proved that all the critical points (local or global optima) of the first stage yield the same critical points as the nonlinear least-squares problem. When the nonlinear estimate is solved in the first stage, then the rest of the unknown can be solved for linearly. However, the numerical algorithm can be complex as it requires computing special derivatives of orthogonal projectors that have to be obtained for the efficient gradient descent optimization method to work.

Haupt et al. [12] proposed a two-step, recursive, and iterative estimation algorithm. The algorithm uses a change of variables to split the cost function into a linear problem in the first step and a nonlinear problem in the second step. The split is done in such a way that the first-step states become measurements for the second-step states. While this estimator is powerful and has been used successfully in many aerospace estimation problems, the underlying approach will not always lead to an optimal estimate, most notably when the second step cost function is nonconvex. Furthermore, as we show later, it is not always obvious (or even possible) how to split some problems into a linear and nonlinear step by a simple change of variables.

Another similar and highly effective two-step procedure was proposed by Alonso and Shuster [13] to solve the magnetometer calibration problem. Their approach "centers" the nonlinear measurement model into a linear model and solves a centered estimate in the first step. In the second step, it uses the centered estimate as an initial estimate to approximate the original estimated parameters. However, this algorithm is somewhat ad hoc in that it is very specific to the magnetometer calibration problem; the statistical properties of the estimation errors cannot be easily transferred to other general estimation problems. The Prony algorithm [14] is another example of an ad hoc estimation approach that has been used successfully in the problem of estimating frequency, amplitude, phase, and damping components of electrical power system response signals.

In the field of aerodynamic parameter estimation, the equation-error approach [5] is often used to obtain starting values for the model parameters before applying iterative methods such as output error [5], which is a maximum likelihood estimator for the problem where process noise is neglected. In other cases, measured states can be substituted in the first iteration of output error so that initial parameter estimates are not needed. Using the equation-error approach or substituting the measured states in the first iteration, followed by application of output error, can be also viewed as two-stage approaches.

B. Contribution

There are two main contributions of this paper. First, we show that there is a class of nonlinear estimation problems that arise in aerospace engineering applications that often have the mathematical structure of Eq. (1). Second, we exploit this nonlinear structure to develop an estimator that naturally leads to a procedure for selecting good initial conditions for a given problem and have comparable (and in some instances better) accuracy and convergence characteristics relative to other nonlinear estimators currently used in aerospace applications. We present two illustrative scalar examples to show how this estimator is implemented. Finally, we use this estimator to solve the problem of calibrating a 5-hole Pitot tube in flight. This problem is difficult to solve with a single-stage estimator due to the nonlinearity and nonzero wind condition.

C. Paper Organization

The remainder of this paper is organized as follows. Section II describes the proposed estimator. The description includes a detailed derivation of the estimator equations and error bounds. In Sec. III, the estimator is used to solve two simple examples. These examples are tutorial in nature and show how the estimator is implemented in practice and how its performance compares to other nonlinear estimators. Then, in Sec. IV, we use the estimator to solve the 5-hole Pitot tube calibration problem using flight test data collected from an unmanned aerial vehicle (UAV). Section V provides concluding remarks.

II. Estimator Formulation

In this section, we formulate the two-stage estimator, which is the subject of this paper. We start by noting that the general nonlinear measurement model with additive noise from estimation theory [1] can be written as follows:

$$\boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\xi}) + \boldsymbol{v}_k \tag{3}$$

Without loss of generality, we are posing this as a parameter estimation problem. As such, we have separated the parameters to be estimated, $\boldsymbol{\xi}$, from the states of the system $\boldsymbol{x}_k$. We assume that this measurement model can be recast (as shown by the canonical form and examples in the Appendices) into the form given by Eq. (1) or

$$\boldsymbol{z}_k = \boldsymbol{A}(\boldsymbol{x}_k,\boldsymbol{u}_k,\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 + \boldsymbol{b}(\boldsymbol{x}_k,\boldsymbol{u}_k,\boldsymbol{\xi}_2) + \boldsymbol{v}_k, \quad E\{\boldsymbol{v}_k\} = 0, \quad E\{\boldsymbol{v}_k\boldsymbol{v}_k^T\} = \boldsymbol{R} \tag{4}$$

where we assume that $\boldsymbol{u}_k$ and $\boldsymbol{x}_k$ for $k = 1, \ldots, N$ are known.

The measurement noise $\boldsymbol{v}_k$ is assumed to be independent, identically distributed Gaussian white noise. Thus the covariance matrix $\boldsymbol{R}$ is set to be diagonal and its entries are unknown. As noted earlier, the algorithm proposed in this paper exploits the structure of Eq. (4) as follows: First, we solve a linear least-squares problem for the parameter vector $\boldsymbol{\xi}_1$ where the remainder of the unknowns in the parameter vector $\boldsymbol{\xi}_2$ are held fixed at some appropriate and fixed values. The algorithm includes a method for assessing the appropriateness of candidate $\boldsymbol{\xi}_2$ values. This is called the *first stage*. In the following *second stage*, we solve a constrained nonlinear optimization problem for either $\boldsymbol{\xi}_2$ only ($\boldsymbol{\xi}_1$ can be subsequently determined) or else for all of the unknowns ($\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$) simultaneously, by using the estimates from the first-stage as the initial condition for the optimization. The choice of re-estimating either $\boldsymbol{\xi}_2$ only or else all the parameters in the second stage depends on the uniqueness of the local minimum. The determination is made empirically by a residual sampling procedure. This formulation leads to excellent convergence properties and, in many instances, guaranteed error bounds on the total cost function to be minimized. It should be noted that this is different from the two-step estimator proposed by Haupt et al. [12] in two fundamental ways. First, a change of variables is not required. Rather, the inherent structure of the problem is used in the two-stage process. Second, the Haupt/Kasdin estimator uses estimates from their first-step process (a linear problem) as measurements in the second-step process (nonlinear optimization). In the algorithm proposed here, the parameters are all estimated without having to formulate a pseudomeasurement by a change of variables.

To show why the proposed estimator works, we start by noting that the optimal estimate of the parameter vector $\boldsymbol{\xi}^*$ is the minimizer of the quadratic cost function $J(\boldsymbol{\xi})$ with a penalized term on the covariance noise matrix $\boldsymbol{R}$, which is nonlinear in $\boldsymbol{\xi}$ and given below:

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

$$\boldsymbol{\xi}^* = \arg\min_{\boldsymbol{\xi}\in\boldsymbol{\xi}_{\text{limit}}} J(\boldsymbol{\xi}) \tag{5}$$

$$J(\boldsymbol{\xi}) = J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \frac{1}{2}\sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_2)\|^2_{\boldsymbol{R}} + \frac{N}{2}\ell_n|\boldsymbol{R}| \tag{6}$$

where we drop $\boldsymbol{x}_k$ and $\boldsymbol{u}_k$ from $\boldsymbol{A}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\xi}_2)$ to simplify the notation. $\boldsymbol{\xi}_{\text{limit}}$ is the constraint that is imposed on $\boldsymbol{\xi}$. This cost function is essentially the maximum likelihood estimation without the constant term [3,5]. From Eq. (6), it is clear that for a given, fixed value of $\boldsymbol{\xi}_2$ (which implies $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ are known), solving for $\boldsymbol{\xi}_1$ is nothing more than the traditional, linear least-squares estimation problem if $\boldsymbol{R}$ is an identity matrix. Assuming that $\boldsymbol{R}$ is known for now (how the unknown $\boldsymbol{R}$ is handled is discussed in Sec. II.C), the accuracy of the estimate for $\boldsymbol{\xi}_1$, denoted as $\hat{\boldsymbol{\xi}}_1$, will depend on how accurate $\boldsymbol{A}(\boldsymbol{\xi}_2)$ is. This, in turn, depends on how close a particular $\boldsymbol{\xi}_2$ used to form $\boldsymbol{A}(\boldsymbol{\xi}_2)$, denoted as $\boldsymbol{\xi}_{2p}$, is to the optimal $\boldsymbol{\xi}_2^*$. If the initial guess $\boldsymbol{\xi}_{2p}$ is equal to $\boldsymbol{\xi}_2^*$, then the estimate of $\boldsymbol{\xi}_1$ resulting from the linear least-squares problem will be optimal. However, because $\boldsymbol{\xi}_2^*$ is not known, how can we decide whether a given value of $\boldsymbol{\xi}_{2p}$ is close to $\boldsymbol{\xi}_2^*$? We will answer this question by showing that the following are true:

1) The minimum of the cost function $J(\boldsymbol{\xi})$ is bounded from above and below by the error term E (E will be discussed in detail in Secs. II.A and II.B):

$$J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) - E \le J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_2^*) \le J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) \tag{7}$$

2) If $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ satisfy the Lipschitz condition and the domain of the state vector $\boldsymbol{\xi}$ is finite, then the cost function error E is bounded. Furthermore, the error term E is a function of $\boldsymbol{\xi}_{2p}$.

We will use these two points to develop a metric for assessing how close $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p})$ is to $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_2^*)$. This will be used to guide our selection of $\boldsymbol{\xi}_{2p}$, which will bring the cost function value in the first stage close to its optimal value. Once we are close enough to the minimum value of $J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, we carry out the second-stage optimization either on $\boldsymbol{\xi}_2$ only, or else on $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ simultaneously. The choice of determining whether to estimate one set or both sets can be empirically assessed by estimating trace of $\boldsymbol{R}$, denoted as $\text{Tr}[\boldsymbol{R}]$, in the first stage. If the estimated $\text{Tr}[\boldsymbol{R}]$ computed from a range of $\boldsymbol{\xi}_{2p}$ has a unique local minimum, then only $\boldsymbol{\xi}_2$ needs to be re-estimated. Otherwise, both $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ must be re-estimated simultaneously because the constraints for $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ in the sequential optimizing setting may not be valid. This will be explained further in Sec. II.C.

It is observed that, in some aerospace parameter estimation problems, $\boldsymbol{\xi}_2$ can be set to zero initially because it normally represents terms that are small biases or scale factor errors (c.f. Appendix A), and they are close to zero if the sensors are accurate. This information can also help determine $\boldsymbol{\xi}_{2p}$ qualitatively in addition to the quantitative procedure described in Sec. II.C. In the next section, we show why the two points noted above are true.

A. Bounding $J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$

To show that Eq. (7) is true, we expand the cost function in Eq. (6) as follows:

$$\begin{aligned} J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) &= \frac{1}{2}\sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_2)\|^2_{\boldsymbol{R}} + \frac{N}{2}\ell_n|\boldsymbol{R}| \\ &= \frac{1}{2}\sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_{2p}) - [\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1 - [\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})]\|^2_{\boldsymbol{R}} + \frac{N}{2}\ell_n|\boldsymbol{R}| \\ &\ge \underbrace{\frac{1}{2}\sum_{k=1}^{N} (\|z_k - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|^2_{\boldsymbol{R}} - \|[\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1\|^2_{\boldsymbol{R}} - \|\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|^2_{\boldsymbol{R}}) + \frac{N}{2}\ell_n|\boldsymbol{R}|}_{H(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)} \end{aligned} \tag{8}$$

The last inequality is obtained using the triangle inequality: $\|v + w\| \ge \|v\| - \|w\|$.

Thus, if we minimize both sides of Eq. (8), the following is obtained:

$$\begin{aligned} J^* &\ge \min_{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2} H(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \\ &= \underbrace{\min_{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2} \frac{1}{2}\sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|^2_{\boldsymbol{R}} + \frac{N}{2}\ell_n|\boldsymbol{R}|}_{J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p})} \\ &\quad \underbrace{- \max_{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2} \frac{1}{2}\sum_{k=1}^{N} (\|[\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1\|^2_{\boldsymbol{R}} + \|\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|^2_{\boldsymbol{R}})}_{E} \end{aligned} \tag{9}$$

From the equation above, we see that E is the error between the global optimal cost J^* and the minimum of the first-stage cost $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p})$ using a particular $\boldsymbol{\xi}_{2p}$.

By *definition* of the optimum cost, the following inequality is true:

$$J^* \triangleq J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_2^*) \le J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) \tag{10}$$

Equation (7) follows naturally from Eqs. (9) and (10). It should be noted that Eq. (7) *does not* imply that there is a value of $\boldsymbol{\xi}_2 = \boldsymbol{\xi}'_{2p}$ such that $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}'_{2p}) = J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) - E < J^*$. Recall that in this first stage we are selecting a value for $\boldsymbol{\xi}_2$ a priori and the free variable is $\boldsymbol{\xi}_1$. So for every value of $\boldsymbol{\xi}_2$ we select, the cost function for $\boldsymbol{\xi}_1$ changes. Instead, the point articulated by Eq. (9) is this: If the cost error E is small, then $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) \approx J^*$ and the result of the first-stage cost is very close to the true optimal cost. In other words, the second step is now just a fine-tuning of the first stage. In the next section, we derive bounds for the cost error E.

B. Bounding $E = E(\boldsymbol{\xi}_{2p})$

In general, it would be difficult to bound E unless we place some restrictions on the nature of the functions $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ as well as the state vector $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T \quad \boldsymbol{\xi}_2^T]^T$. Thus, we will assume that the following conditions hold true:

1) The norm of the unknown parameter $\boldsymbol{\xi}_1$ is bounded by ℓ_1: $\|\boldsymbol{\xi}_1\| \le \ell_1$.

2) The norm of the difference between $\boldsymbol{\xi}_{2p}$ and $\boldsymbol{\xi}_2^*$ is bounded by ℓ_2: $\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \le \ell_2$.

3) The nonlinear functions $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ are Lipschitz continuous functions and they satisfy the following:

$$\|\boldsymbol{A}(\boldsymbol{\xi}_2^*) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\| \le L_{\boldsymbol{A}}\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \text{ and}$$
$$\|\boldsymbol{b}(\boldsymbol{\xi}_2^*) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\| \le L_{\boldsymbol{b}}\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \text{ for } \boldsymbol{\xi}_2^* < \boldsymbol{\xi}_2 < \boldsymbol{\xi}_{2p}$$

where ℓ_1 and ℓ_2 are scalars, and $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{b}}$ are called Lipschitz constants (also scalars). The first two conditions are satisfied if the state vector $\boldsymbol{\xi}$ has a finite domain. This is a reasonable assumption in many engineering problems where the state vector represents some physical and measurable quantity. The upper bound ℓ_1 in the first

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

assumption represents the maximum value that $\boldsymbol{\xi}_1$ can achieve. The upper bound ℓ_2 in the second assumption represents the error between the initial guess and optimal value of $\boldsymbol{\xi}_2$. Thus, these two conditions are not very restrictive. The values of ℓ_1 and ℓ_2 can be usually estimated based on the prior knowledge. For example, the absolute value of a reasonable scale factor $\boldsymbol{\xi}_2$ should not be bigger than 0.5 (i.e., $-0.5 \le \boldsymbol{\xi}_2^* \le 0.5$ and this bound is very conservative). Then we can pick $\boldsymbol{\xi}_{2p}$ such that $|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}| \le 0.5 + |\boldsymbol{\xi}_{2p}| \le \ell_2$. $\boldsymbol{\xi}_{2p}$ should be chosen such that it is close to $\boldsymbol{\xi}_2^*$. If $\boldsymbol{\xi}_{2p}$ is set to be 1, then ℓ_2 can be set to 1.5 to upper bound $|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}|$. The third condition requiring the functions $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ to be Lipschitz continuous is not very restrictive either. Many mathematical functions used to model physical systems, such as the square root (real positive numbers under the square root), as well as sine and cosine functions, are Lipschitz continuous. Furthermore, $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{b}}$ can also be viewed as the derivative information of $\boldsymbol{\xi}_2$. If the selected $\boldsymbol{\xi}_{2p}$ approaches $\boldsymbol{\xi}_2^*$, then $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{b}}$ approach zero. With these three assumptions, we can upper bound the following two error terms:

$$\begin{aligned} E_1 &\triangleq \max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \|[\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1\| \le \max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \|\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\| \|\boldsymbol{\xi}_1\| \\ &\le \max_{\boldsymbol{\xi}_2} \|\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\| \ell_1 \le L_{\boldsymbol{A}} \|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \ell_1 \\ &\le L_{\boldsymbol{A}} \ell_2 \ell_1 \end{aligned} \tag{11}$$

$$E_2 \triangleq \max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \|\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\| \le L_{\boldsymbol{b}} \|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \le L_{\boldsymbol{b}} \ell_2 \tag{12}$$

where the first inequality in Eq. (11) comes from Cauchy–Schwarz inequality. Using Eqs. (11) and (12) we can derive an upper bound on the error E as follows:

$$\begin{aligned} E &= \max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \frac{1}{2} \sum_{k=1}^{N} (\|[\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1\|^2 + \|\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|^2) \\ &= \frac{1}{2} \sum_{k=1}^{N} (\max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \|[\boldsymbol{A}(\boldsymbol{\xi}_2) - \boldsymbol{A}(\boldsymbol{\xi}_{2p})]\boldsymbol{\xi}_1\|)^2 \\ &\quad + \frac{1}{2} \sum_{k=1}^{N} (\max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \|\boldsymbol{b}(\boldsymbol{\xi}_2) - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|)^2 \\ &= \frac{1}{2} \sum_{k=1}^{N} (E_1^2 + E_2^2) \le \frac{N}{2} (L_{\boldsymbol{A}}^2 \ell_2^2 \ell_1^2 + L_{\boldsymbol{b}}^2 \ell_2^2) = \frac{N}{2} \ell_2^2 (L_{\boldsymbol{A}}^2 \ell_1^2 + L_{\boldsymbol{b}}^2) \end{aligned} \tag{13}$$

where we dropped the subscript $\boldsymbol{R}$ without loss of generality. Equation (13) implies that for a fixed length of data set N, if the initial guess $\boldsymbol{\xi}_{2p}$ is close to the optimal $\boldsymbol{\xi}_2^*$ (i.e., $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{b}}$ approach zero) and the bounds are $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are small (i.e., ℓ_1 and ℓ_2 approach zero), then the first-stage optimization cost function is close to the original cost function (i.e., E consequently approaches zero). This means that the result of the first stage can bring the cost very close to the minimum global cost, which makes the second stage more likely to converge.

C. Selection of $\boldsymbol{\xi}_{2p}$

So how do we select $\boldsymbol{\xi}_{2p}$ so that E is small, thereby assuring that the second-stage optimization will lead to the correct solution? Although it is difficult to develop a prescriptive solution for selecting $\boldsymbol{\xi}_{2p}$, we can answer the following related question: How do we know if a given choice of $\boldsymbol{\xi}_{2p}$ is one that will increase the chances of convergence to the correct solution? To answer this question, we start by linearizing $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ with respect to $\boldsymbol{\xi}_2$ at $\boldsymbol{\xi}_{2p}$ as follows:

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{\xi}_2) &\approx \boldsymbol{A}(\boldsymbol{\xi}_{2p}) + \frac{\partial \boldsymbol{A}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \Delta\boldsymbol{\xi}_2 \\ \boldsymbol{b}(\boldsymbol{\xi}_2) &\approx \boldsymbol{b}(\boldsymbol{\xi}_{2p}) + \frac{\partial \boldsymbol{b}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \Delta\boldsymbol{\xi}_2 \end{aligned} \tag{14}$$

If $\boldsymbol{\xi}_{2p}$ is chosen such that the first-order terms are sufficient small and satisfy Eq. (15),

$$\begin{aligned} \left\| \frac{\partial \boldsymbol{A}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \Delta\boldsymbol{\xi}_2 \right\| &\le \left\| \frac{\partial \boldsymbol{A}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \right\| \ell_2 \ll \|\boldsymbol{A}(\boldsymbol{\xi}_{2p})\| \\ \left\| \frac{\partial \boldsymbol{b}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \Delta\boldsymbol{\xi}_2 \right\| &\le \left\| \frac{\partial \boldsymbol{b}(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2} \right\| \ell_2 \ll \|\boldsymbol{b}(\boldsymbol{\xi}_{2p})\| \end{aligned} \tag{15}$$

then the nonlinear parameter cost function in Eq. (6) can be approximated by

$$\begin{aligned} &\arg \min_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \frac{1}{2} \sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 - \boldsymbol{b}(\boldsymbol{\xi}_2)\|_{\boldsymbol{R}}^2 + \frac{N}{2} \ell_n |\boldsymbol{R}| \\ &\approx \arg \min_{\boldsymbol{\xi}_1} \frac{1}{2} \sum_{k=1}^{N} \|z_k - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\xi_1 - \boldsymbol{b}(\boldsymbol{\xi}_{2p})\|_{\boldsymbol{R}(\boldsymbol{\xi}_{2p})}^2 + \frac{N}{2} \ell_n |\boldsymbol{R}(\xi_{2p})| \end{aligned} \tag{16}$$

where $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$ is still unknown but it is a matrix that depends on $\boldsymbol{\xi}_{2p}$. Equation (16) implies that the linearized system cost is close to the original nonlinear system cost. The right-hand side of Eq. (16) can be solved using linear least squares by setting the unknown $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$ equal to the identity matrix. Note that $\boldsymbol{\xi}_{1p}$ is suboptimal (biased) in the first stage due to the unknown $\boldsymbol{R}$ and the bounding properties shown in Eqs. (7) and (13) do not change except $J(\boldsymbol{\xi}_1^*, \boldsymbol{\xi}_{2p}) = J(\boldsymbol{\xi}_{1p}, \boldsymbol{\xi}_{2p})$. The unknown $\boldsymbol{R}$ and $\boldsymbol{\xi}_2$ are solved optimally via the second-stage nonlinear optimization.

There may be one or more suboptimal pairs $(\boldsymbol{\xi}_{1p}, \boldsymbol{\xi}_{2p})$ obtained from solving the linearized system that has a cost value approximately equal to the optimal cost. Because estimating parameters using linear least square is not computationally expensive, we can sample a large pool of $\boldsymbol{\xi}_{2p}$ from the feasible set (constrained by $\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \le \ell_2$) to estimate the suboptimal $\boldsymbol{\xi}_{1p}$. Also, the parameter $\boldsymbol{\xi}_{2p}$ should satisfy Eqs. (17) and (18):

$$\frac{\|(\partial \boldsymbol{A}(\boldsymbol{\xi}_{2p})/\partial \boldsymbol{\xi}_2)\| \ell_2}{\|\boldsymbol{A}(\boldsymbol{\xi}_{2p})\|} \le T_1 \quad \text{and} \quad \frac{\|(\partial \boldsymbol{b}(\boldsymbol{\xi}_{2p})/\partial \boldsymbol{\xi}_2)\| \ell_2}{\|\boldsymbol{b}(\boldsymbol{\xi}_{2p})\|} \le T_2 \tag{17}$$

$$\sum_{i=1}^{m_A} \sum_{j=1}^{n_A} \frac{\partial \boldsymbol{A}(i,j)^2(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2 \partial \boldsymbol{\xi}_2^T} > \boldsymbol{0}_{n_{\xi_2} \times n_{\xi_2}} \quad \text{and} \quad \sum_{i=1}^{m_b} \frac{\partial \boldsymbol{b}(i)^2(\boldsymbol{\xi}_{2p})}{\partial \boldsymbol{\xi}_2 \partial \boldsymbol{\xi}_2^T} > \boldsymbol{0}_{n_{\xi_2} \times n_{\xi_2}} \tag{18}$$

where T_1 and T_2 are user defined and can be interpreted as percentage requirements, and n_{ξ_2} is the number of parameters in $\boldsymbol{\xi}_2$. The smaller the values of T_1 and T_2 (obtained through varying $\boldsymbol{\xi}_{2p}$), the tighter the error bound on E. Equation (17) ensures validity of linearization in the first stage and Eq. (18) enforces local convergence for iterative methods in the second stage.

Once we have chosen a set of $\boldsymbol{\xi}_{2p}$, we can estimate the residual vector history $\boldsymbol{v}_k$ for $k = 1, \ldots, N$ and use it to build a metric to find a suitable pair $(\boldsymbol{\xi}_{1p}, \boldsymbol{\xi}_{2p})$ for the second-stage nonlinear estimation. Namely, we find the suboptimal pair $(\hat{\boldsymbol{\xi}}_{1p}, \hat{\boldsymbol{\xi}}_{2p})$ by solving Eq. (19):

$$\begin{aligned} \arg \min_{\xi_{2p}} \mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})] &= \arg \min_{\boldsymbol{\xi}_{2p}} \sum_{k=1}^{N} \mathrm{Tr}(\boldsymbol{v}_k \boldsymbol{v}_k^T) \\ &\text{where } \boldsymbol{v}_k = z_k - \boldsymbol{A}(\boldsymbol{\xi}_{2p})\hat{\boldsymbol{\xi}}_{1p} - \boldsymbol{b}(\boldsymbol{\xi}_{2p}) \\ &\text{and } \hat{\boldsymbol{\xi}}_{1p} = (\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T(\boldsymbol{Z} - \mathcal{B}) \\ &\text{for } \boldsymbol{\xi}_{2p} \in S_{\boldsymbol{\xi}_2} \end{aligned} \tag{19}$$

where $\mathcal{A}$, $\mathcal{B}$, and $\boldsymbol{Z}$ are concatenations of $\boldsymbol{A}_k(\boldsymbol{\xi}_{2p})$, $\boldsymbol{b}_k(\boldsymbol{\xi}_{2p})$, and z_k, respectively, for $k = 1, \ldots, N$. $S_{\boldsymbol{\xi}_2}$ is a chosen set that satisfies the constraint $\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \le \ell_2$ and Eqs. (17) and (18). By minimizing the trace of $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$, we are essentially finding the suboptimal pair that

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

gives the smallest residual vector. We denote this method as the residual sampling procedure. Note that $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ is a similar measure of the error term E shown in Eq. (7), where E can be interpreted as a weighted residual least-squares error.

If the estimated $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ has a local minimum, then only $\boldsymbol{\xi}_2$ needs to be re-estimated in the second stage. Estimating only $\boldsymbol{\xi}_2$ also means that the search space in the nonlinear programming is significantly reduced. Once $\boldsymbol{\xi}_2^*$ and $\boldsymbol{R}$ are estimated alternately in the second stage, $\boldsymbol{\xi}_1^*$ is immediately calculated using weighted linear least squares. We also use $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$ to initialize $\boldsymbol{R}$ in the second stage, as shown in Eq. (16). If the estimated $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ does not have a unique local minimum (as shown in Sec. IV), both $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ should be re-estimated simultaneously in the second stage. This is because the sequential order of constraints may not be valid. Namely,

$$\min_{\xi_1 \in S_{\hat{\xi}_1},\, \xi_2 \in S_{\xi_2}} J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \neq \min_{\xi_2 \in S_{\xi_2}} \left[\min_{\xi_1 \in S_{\hat{\xi}_1}} J(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \right] \tag{20}$$

When $\hat{\boldsymbol{\xi}}_1$ from the inner minimization on the right-hand side of Eq. (20) cannot be uniquely determined, the outer minimization may not be able to arrest $\hat{\boldsymbol{\xi}}_1$ escaping from its own constraint. Though this inequality holds true in general, we observe that if the inner minimization has a unique solution (i.e., the error E is small) using a large sample of $\boldsymbol{\xi}_{2p}$ from the feasible set S_{ξ_2}, then both sides of Eq. (20) can be equal. In other words, because the search space of $\boldsymbol{\xi}_2$ in the inner minimization has been searched exhaustively via sampling, the chance of $\hat{\boldsymbol{\xi}}_1$ escaping from the outer minimization is small. Therefore, if we cannot clearly find a unique local minimum in the first stage represented by the inner minimization, we need to re-estimate $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ simultaneously by solving the left-hand side of Eq. (20). The estimates $\hat{\boldsymbol{\xi}}_{1p}$ and $\hat{\boldsymbol{\xi}}_{2p}$ from the first stage are still used as the initial condition, where $\hat{\boldsymbol{\xi}}_{2p}$ is any vector of the set that results in multiple local minima.

Though this residual sampling method is very crude, it does provide an excellent initial condition for the second stage, as will be demonstrated by examples in Secs. III and IV. One possible alternative of selecting $\boldsymbol{\xi}_{2p}$ would be to evaluate the Jacobian $\nabla J(\boldsymbol{\xi}_{2p})$ and iteratively update $\boldsymbol{\xi}_{2p}$ until $\nabla J(\boldsymbol{\xi}_{2p}) = 0$. However, this method is computationally expensive and prone to error when the nonlinear functions $\boldsymbol{A}(\boldsymbol{\xi}_{2p})$ and $\boldsymbol{b}(\boldsymbol{\xi}_{2p})$ are multidimensional and highly nonlinear. The effect of this selection of $\boldsymbol{\xi}_{2p}$ is depicted graphically in Fig. 1, where $\boldsymbol{\xi}_{2p}^{(j)}$ is a not good choice; it does not give the smallest $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ and it may cause the second stage to arrive the wrong minimum even though it is within the bound of ℓ_2. On the other hand, $\boldsymbol{\xi}_{2p}^{(i)}$ is a good choice because 1) it is the local minimum in the constraint set ℓ_2 and 2) the positive concavity (concave up) ensures local convergence.

Putting all of this together results in the following procedure for implementation of the proposed algorithm:

Step 1: Formulate the measurement equation to have the form given by Eq. (4).

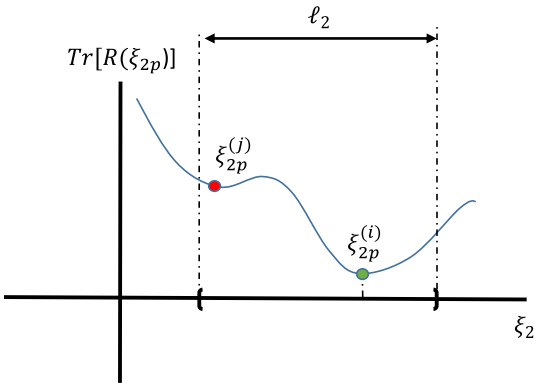


Fig. 1 Pictorial depiction of the effect of choices of $\boldsymbol{\xi}_{2p}$ on $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ value.

Step 2: Sample a large pool of $\boldsymbol{\xi}_{2p}$ from the feasible constraint set $\|\boldsymbol{\xi}_2^* - \boldsymbol{\xi}_{2p}\| \leq \ell_2$; those $\boldsymbol{\xi}_{2p}$ should also satisfy Eqs. (17) and (18).

Step 3: Estimate $\boldsymbol{\xi}_{1p}$ by minimizing the cost function (right-hand side of Eq. 16) using linear least squares with the unknown $\boldsymbol{R} = \boldsymbol{I}$. Then calculate the corresponding trace $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$.

Step 4: Find a suboptimal pair $(\hat{\boldsymbol{\xi}}_{1p}, \boldsymbol{\xi}_{2p})$ such that the corresponding $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ has a unique local minimum. If there exist multiple suboptimal pairs (similar numerical values), choose an arbitrary one from these suboptimal pairs. This completes the first stage.

Step 5: If there exists a unique local minimum from $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$, then solve for $\boldsymbol{\xi}_2$ only in the second stage with $\boldsymbol{\xi}_{2p}$ as the initial condition. Use $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$ to initialize $\boldsymbol{R}$ in the second stage. Once $\boldsymbol{\xi}_2^*$ and $\boldsymbol{R}$ are obtained, $\boldsymbol{\xi}_1^*$ immediately follows using weighted linear least squares. Otherwise, solve for both $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ simultaneously with the suboptimal estimate $(\hat{\boldsymbol{\xi}}_{1p}, \boldsymbol{\xi}_{2p})$ as the initial condition in the second stage. The nonlinear function can be minimized by any standard iterative method such as modified Newton–Raphson, Gauss–Newton, or Levenberg–Marquardt [15] method. We estimate $\boldsymbol{\xi}$ and $\boldsymbol{R}$ alternately until both $\boldsymbol{\xi}$ and the diagonal elements of $\boldsymbol{R}$ converge. This completes the second stage.

Though the measurement model in Eq. (4) resembles a Kalman filter (KF) or extended Kalman filter (EKF) measurement model equation, we find that it is not straightforward to make the proposed algorithm a stand-alone measurement equation in a recursive estimation. This is because of the nature of the first stage, where the optimality and separability of $\hat{\boldsymbol{\xi}}_1$ depends on $\hat{\boldsymbol{\xi}}_2$ generally in a nonlinear fashion. However, one can use the first stage of the proposed algorithm to estimate an initial condition with a small batch of data for an EKF or iterated-EKF (IEKF) filter. Then we can use $\dot{\hat{\boldsymbol{\xi}}} = 0$ as the parameter time update equation and linearize the measurement in Eq. (4) with respect to $\boldsymbol{\xi}$ to formulate the linearized measurement matrix needed for EKF or IEKF. This is demonstrated in Sec. IV of Ref. [16]. In the following section, we provide a demonstration on how to implement this estimator.

III. Two Tutorial Examples: Scalar Measurement Equations

To demonstrate the mechanics of using this estimator, gain some intuition into its operation, and compare its performance to other estimators, we solve the following static parameter estimation problem, which is a simplified version of the problem presented in Ref. [12] [Eq. (36)]:

$$z_k = f_k(\eta_k) + v_k = \underbrace{(1+a)\cos(\eta_k + b) + c}_{f_k(\eta_k)} + v_k \tag{21}$$

The variables a, b, and c (the coefficients of the nonlinear function f_k) are the parameters we want to estimate. In this particular case, we set the values of the parameters as follows: $a = 1$, $b = 0.1$, and $c = 1$. There are 100 scalar measurements z_k generated by varying η from 1 to 10 rad, incrementing by the same interval. The 100 measurement noise v_k is drawn from a normal distribution with mean of zero and a standard deviation of 0.3.

To use the estimator developed in this paper on Eq. (21), the scalar measurement model is recast into an affine problem by exploiting the structure of the nonlinear function $f_k(\eta_k)$ as shown below:

$$z_k = \underbrace{\begin{bmatrix}\cos(\eta_k + b) & 1\end{bmatrix}}_{\boldsymbol{A}(\boldsymbol{\xi}_2)} \underbrace{\begin{bmatrix}a\\ c\end{bmatrix}}_{\boldsymbol{\xi}_1} + \underbrace{\cos(\eta_k + b)}_{\boldsymbol{b}(\boldsymbol{\xi}_2)} + v_k \tag{22}$$

where $\boldsymbol{\xi}_1 = [a, c]^T$ and $\boldsymbol{\xi}_2 = b$.

Because there is only one parameter in $\boldsymbol{\xi}_2$, we can simply sweep a range of b to estimate $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$. Also, $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})] = \boldsymbol{R}(\boldsymbol{\xi}_{2p})$ for this problem because the measurement at each time step is a scalar. Figure 2 shows the estimated scalar value of $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$. It can be seen that $\boldsymbol{\xi}_{2p} = b = 0.08$ corresponds the minimum value of $\boldsymbol{R}(\boldsymbol{\xi}_{2p})$. We also observe that both $b = 0.08$ and its corresponding $\boldsymbol{R}(0.08)$

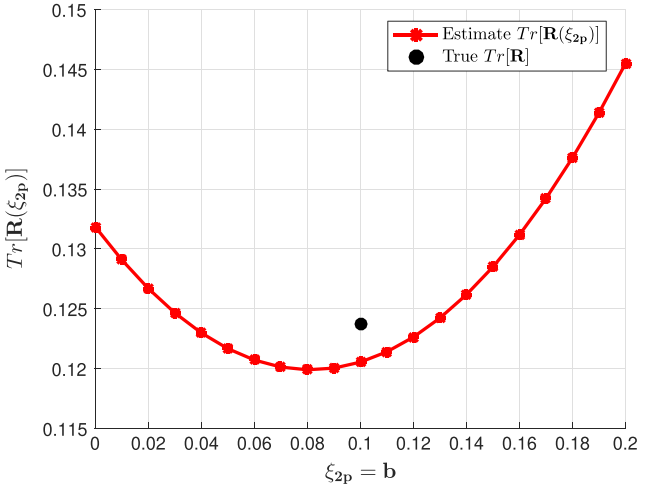


Fig. 2 Estimated Tr[$R(\xi_{2p})$] by sampling random of ξ_{2p} for scalar example 1.

are not same as the true values due to the measurement noise. Nonetheless, there exists a unique minimum, and so we will use $\xi_{2p} = 0.08$ to estimate ξ_{1p}.

With the unique measurement structure, valid linear approximation [Eq. (16)], and a unique local minimum (indicated by the positive definiteness of $\boldsymbol{A}(\xi_{2p})$ and $\boldsymbol{b}(\xi_{2p})$ [Eq. (18)], the parameters a and c are estimated in first stage by solving a linear least-squares problem, which minimizes the following cost function:

$$\hat{\xi}_{1p} = \arg\min_{\xi_1}\left(\frac{1}{2}\sum_{k=1}^{N}\|z_k - \boldsymbol{A}(0.08)\xi_1 - \boldsymbol{b}(0.08)\|_{\boldsymbol{I}}^2\right) \tag{23}$$

Because there exists a unique minimum as shown in Fig. 2, we use $\xi_{2p} = 0.08$ as the initial condition to estimate ξ_{12} in the second stage:

$$\hat{\xi}_2, \hat{\boldsymbol{R}} = \arg\min_{\xi_{12},\boldsymbol{R}} \frac{1}{2}([\boldsymbol{Z} - \mathcal{A}\xi_{11} - \mathcal{B}]^T \boldsymbol{W}[\boldsymbol{Z} - \mathcal{A}\xi_{11} - \mathcal{B}]) \tag{24}$$

where

$$\xi_2 = b_p = 0.08 \qquad \boldsymbol{Z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} \qquad \mathcal{A} = \begin{bmatrix} \boldsymbol{A}_1(\xi_2) \\ \vdots \\ \boldsymbol{A}_N(\xi_2) \end{bmatrix}$$
$$\mathcal{B} = \begin{bmatrix} \boldsymbol{b}_1(\xi_2) \\ \vdots \\ \boldsymbol{b}_N(\xi_2) \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{R}^{-1} & & \\ & \ddots & \\ & & \boldsymbol{R}^{-1} \end{bmatrix} \tag{25}$$

The sequential quadratic programming (SQP) algorithm is used to solve the optimization problem given by Eq. (24). Note that ξ_1 is calculated iteratively using weighted linear least squares inside the nonlinear cost solver, and so there is no need to initialize ξ_1 in the beginning of the second stage. The term ξ_{1p} in Eq. (23) is used to initialize the second stage if the local minimum is not unique (demonstrated later in Sec. IV). For the work reported in this paper, the SQP is implemented using the built-in MATLAB function `fmincon` [17]. An outer while-loop outside of `fmincon` is written to estimate $\boldsymbol{R}$ alternately with ξ_2 until the following is satisfied (Eq. (6.41e) in Ref. [5]):

$$\left|\frac{(\hat{r}_{jj})_k - (\hat{r}_{jj})_{k-1}}{(\hat{r}_{jj})_{k-1}}\right| < 0.05 \quad \forall\, j, j = 1, 2, \ldots, n_o \tag{26}$$

where $(\hat{r}_{jj})$ is the estimate of the jth diagonal element of the estimate $\hat{\boldsymbol{R}}$ and n_o is the number of the total diagonal terms. In this scalar example, $j = 1$ since $\boldsymbol{R}$ is a scalar. Once the optimal $\hat{\xi}_2^*$ and $\hat{\boldsymbol{R}}$ are obtained, $\hat{\xi}_1^*$ can be immediately solved using weighted linear least squares:

$$\hat{\xi}_1^* = (\mathcal{A}^T\boldsymbol{W}\mathcal{A})^{-1}\mathcal{A}^T\boldsymbol{W}(\boldsymbol{Z} - \mathcal{B}) \tag{27}$$

We will benchmark the performance of this estimator against the following pair of nonlinear estimators: 1) a classic, nonlinear program that solves for ξ_1 and ξ_2 simultaneously and 2) the Haupt/Kasdin two-step estimator described in Ref. [12].

A. Benchmark 1: Classic Nonlinear Programming

The first benchmark is nothing more than a solution to the optimization problem posed by the left-hand side of Eq. (16). The implementation of this benchmark differs from the algorithm proposed in this paper, because the initial conditions are selected randomly.

B. Benchmark 2: Haupt/Kasdin Two-Step Estimator

To implement the Haupt/Kasdin two-step estimator, we choose a new set of states by a change of variables such that Eq. (21) can be written as a linear measurement model shown below:

$$\begin{aligned} z_k &= \boldsymbol{H}_k\boldsymbol{f}(\xi) = \boldsymbol{H}_k\boldsymbol{y} + v_k \\ &= \underbrace{\begin{bmatrix}\cos\eta_k & -\sin\eta_k & \cos\eta_k & -\sin\eta_k & 1\end{bmatrix}}_{\boldsymbol{H}_k} \underbrace{\begin{bmatrix} a\cos b \\ a\sin b \\ \cos b \\ \sin b \\ c \end{bmatrix}}_{\boldsymbol{y}} + v_k \end{aligned} \tag{28}$$

The choice of change of variable is arbitrary and leads to the following cost function:

$$J_y = (\boldsymbol{Z} - \mathcal{H}\boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{Z} - \mathcal{H}\boldsymbol{y}) \tag{29}$$

where $\mathcal{H}$ is given by

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_N \end{bmatrix} \tag{30}$$

Note that, even though the choice of new variable $\boldsymbol{y}$ is arbitrary, it actually dictates the condition number of $\mathcal{H}$. If $\mathcal{H}$ is not well conditioned, the result of the first stage can be poor. For this particular problem, it can be problematic if the data length N is small. This is because columns 1 and 2 of $\boldsymbol{H}_k$ are same as columns 3 and 4, respectively, in Eq. (28). This is also a pitfall of benchmark 2. The first-step state $\boldsymbol{y}$ is estimated using the linear least-squares method. In the second step, the estimates of the first-step states $\hat{\boldsymbol{y}}$ are treated as the new measurements in the second stage. This leads to the following measurement equation:

$$\hat{\boldsymbol{y}} = \boldsymbol{f}(\xi) + \boldsymbol{e} \tag{31}$$

where the measurement noise $\boldsymbol{e}$ has covariance matrix $\boldsymbol{P}_y$. Once the estimate $\hat{\boldsymbol{y}}$ is obtained, the following cost function is minimized using an iterative nonlinear optimizer:

$$J(\xi) = [\hat{\boldsymbol{y}} - \boldsymbol{f}(\xi)]^T\boldsymbol{P}_y^{-1}[\hat{\boldsymbol{y}} - \boldsymbol{f}(\xi)] \tag{32}$$

This second-step cost function can be nonlinear and nonconvex. Thus, there is no guarantee that the solution is optimal. For a static problem, this essentially reduces to solving a set of simultaneous, nonlinear algebraic equations. In general, the solution for such problem is not unique.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

C. Performance Comparisons

A set of 1000 Monte Carlo (MC) simulation runs were used to assess the performance of the algorithm developed in this paper and compare it against the two benchmarks. For each MC run of the proposed algorithm, the initial value of the parameter b is determined to be 0.08 from the first stage, and used for the second step optimization. For the first benchmark (the classic nonlinear program), initial conditions for a and c were selected randomly from $N(0, 1^2)$ and b is drawn from $\sim N(0, 0.1^2)$. We also set the constraint for $b \in [0, 0.2]$ in the first benchmark for a fair comparison because we only sampled $\boldsymbol{\xi}_{2p}$ from a predetermined range (assumed to be due to prior knowledge). The second benchmark (Haupt/Kasdin two-step estimator) does not require an initialization for the first step states, but the initial values for a, b, and c are needed for the second stage. The same initial values from the 1000 runs in benchmark 1 were used for the initialization in the second stage in benchmark 2. Benchmark 2 does not require any constraint setting for b, according to Ref. [12].

Table 1 shows the MC results in terms of the percentage of times the algorithm converged to the correct solution. The correction solution is determined by taking the 2-norm between the estimated and true parameter vector, that is, less than 0.1. Both benchmark 1 and the proposed algorithm converged 100% of the time. Although this is not a theoretical proof that the correct solution is guaranteed by the algorithm developed in this paper, the comparison shows that it can yield equivalent or favorable results when compared with other nonlinear estimators. Table 2 shows the estimated parameter, standard deviation, and noise covariance versus the true values. The standard deviation in the proposed estimator is calculated by taking the square root of the diagonal of the inverse of the final Hessian matrix, which is one of the outputs from `fmincon`.

Note that it is not always obvious (particularly, in actual applications) whether the estimator has converged to the correct solution. This can be seen if we use the estimates for the parameters to construct a predicted measurement $\hat{z}$. That is, we apply $\hat{\boldsymbol{\xi}} = [\hat{\boldsymbol{\xi}}_1^T \quad \hat{\boldsymbol{\xi}}_2^T]^T$ to Eq. (21) to determine $\hat{z}$. Figure 3 plots 100 randomly selected estimated outputs out of the 1000 MC runs for the proposed algorithm and the two benchmarks. In the case of the Haupt/Kasdin two-step estimator (benchmark 2), we see that there are many instances where predicted measurement $\hat{z}$ is close to the observed measurement z, even though the estimates of a, b, and c used to generate $\hat{z}$ are *incorrect*. The fact that the solution has converged to the incorrect value is not visible in the output. This implies that the cost function used in the second step optimization of Haupt/Kasdin algorithm is nonconvex; it has multiple local minima that are sensitive to the values of the states used to initialize the optimization process.

The comparisons so far show that breaking the estimation process into two steps can improve the chance of converging to the correct solution. As the authors of Ref. [12] note, however, it may not always be possible to do this with the Haupt/Kasdin algorithm because of the mathematical structure of the problem at hand. To show this, we modify the estimator problem given by Eq. (21) slightly as follows:

$$z_k = f_k(\eta_k) + v_k = (1 + a)\cos(\eta_k(1 + b) + c) + d + v_k \quad (33)$$

Equation (33) can be recast into the suitable form for the proposed algorithm shown below:

$$z_k = \underbrace{[\cos((\eta_k + b) + c) \quad 1]}_{\boldsymbol{A}(\boldsymbol{\xi}_2)} \underbrace{\begin{bmatrix} a \\ d \end{bmatrix}}_{\boldsymbol{\xi}_1} + \underbrace{\cos((\eta_k + b) + c)}_{\boldsymbol{b}(\boldsymbol{\xi}_2)} + v_k \quad (34)$$

where $\boldsymbol{\xi}_1 = [a, d]^T$ and $\boldsymbol{\xi}_2 = [b, c]^T$. An additional unknown parameter d has been added to the measurement model. In this case, the Haupt/Kasdin estimator cannot be used, as the parameter b cannot be linearly separated by change of variables from η. For completeness, we ran another set of MC simulations to compare the performance of the proposed estimator and benchmark 1 on the modified in Eq. (33). We draw a and d from $N(0, 1^2)$ and b and c from $N(0, 0.1^2)$, respectively. We also re-draw from the noise term v_k from $N(0, 0.3^2)$.

Figure 4 shows estimated $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ from sampling b and c. It can be seen that there is clearly a local minimum $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ value. Therefore, we use the corresponding $\boldsymbol{\xi}_{2p} = [b_p, c_p]^T = [0.0556, 0.0808]^T$ as the initial condition for the second stage in the proposed estimator. We set the constraint for $b \in [0, 0.5]$ and $c \in [0, 1]$ in benchmark 1 for a fair comparison because we sampled those values to estimate $\mathrm{Tr}[\boldsymbol{R}(\boldsymbol{\xi}_{2p})]$ for the proposed algorithm (assumed to be due to prior knowledge). The results of this simulation are summarized in Tables 3 and 4 and Fig. 5. It can be seen that the correct percentage actually decreased due to the high nonlinearity for benchmark 1. There are still a number of incorrect solutions, whereas the proposed algorithm still converges to the correct value every time. We randomly plotted 100 corresponding time series of the predicted measurements out of the 1000 MC runs in Fig. 5. It can be seen that the predicted measurement (generated by estimated parameters from the nonlinear programming approach) can be incorrect.

These two tutorial examples show that the proposed estimator can work well if the starting initial guess $\boldsymbol{\xi}_{2p}$ is close to the true value. The estimates of the first stage essentially bring the total cost very close to the true cost, which makes the nonlinear, iterative optimization of the second state converge consistently. It does this by eliminating the randomness of the initial guesses for the parameters in either two benchmark methods.

Table 1 Monte Carlo simulation results for the measurement model in Eq. (21)

Estimation algorithm	Nonlinear programming	Haupt/Kasdin two-step estimator	Proposed algorithm
Correct solution, %	100	97.2	100

Note: The estimate parameter $\hat{\xi}$ is considered correct when $\|\xi_{\text{true}} - \hat{\xi}\|_2 \leq 0.1$.

Table 2 Estimation results from proposed algorithm for scalar example 1

Parameter	True	Estimate	Standard deviation
a	1	0.9512	0.0507
b	0.1	0.0812	0.0182
c	1	0.9415	0.0360
R	0.1237	0.1199	— —

Note: The true $\boldsymbol{R}$ is calculated using the 100 noise v samples.

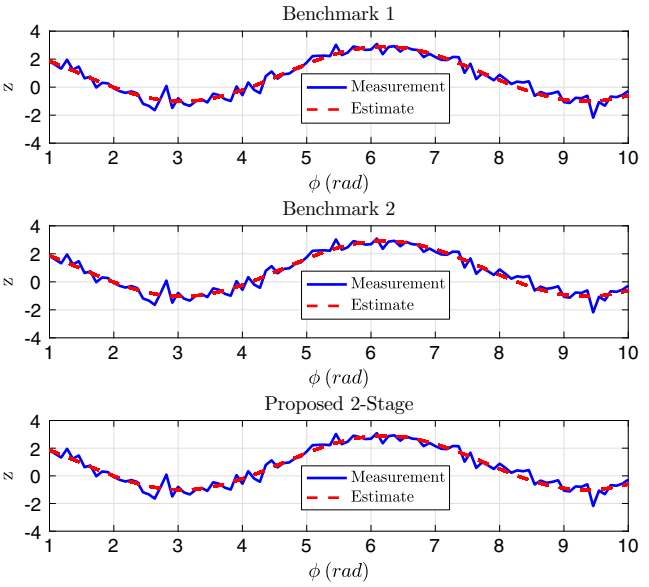


Fig. 3 Random 100 MC simulation results from 3 different methods using Eq. (21). In this case, all 3 estimates were effective compared with the measurement.

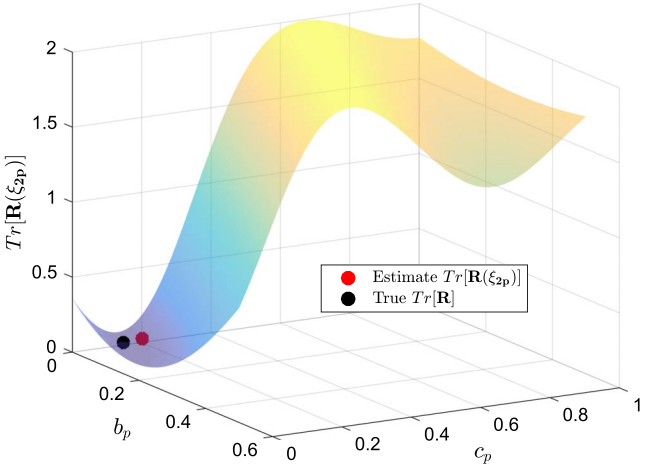


Fig. 4 Estimated Tr$[R(\xi_{2p})]$ by sampling random of ξ_{2p} for scalar example 2.

IV. Flight Test Example: 5-Hole Pitot Tube Calibration

Some parameter estimation problems, such as the magnetometer calibration and data compatibility problem, can be recast (shown in Appendix B) and solved with the proposed estimator. The magnetometer calibration and data compatibility problem can also be solved by well-known methods such as the Haupt/Kasdin two-step estimator and output error, respectively. However, there are other parameter estimation problems that cannot be easily solved with these known methods, because of the sensitivity to initial values. In this section, we demonstrate an aerospace application using the proposed estimator that overcomes the initial-value sensitivity issue.

In particular, we exercise the estimator on calibration of a 5-hole Pitot tube using flight test data for small UAV applications. The problem was previously investigated in Ref. [18] and is an excellent example that shows how conventional methods may suffer from an incorrect local minimum, due to a poor initial parameter guess. To briefly summarize, this is the problem of calibrating a 5-hole Pitot tube (i.e., finding error model parameters) using an existing navigation solution, such as inertial velocity and attitude. The calibration consists of estimating sensor scale factor, bias errors, installation misalignment error, and steady wind vector. One challenging part of this problem is that the wind vector cannot be assumed to be zero, due to the relatively slow airspeed (10–25 m/s) range relative to the wind speed (1–10 m/s). Single-stage estimators will not converge to the correct solution if the initial parameter guess is not close to the underlining true values. In particular, the typical zero-value initial guess for wind vector might not always result in consistent estimates

Table 3 Monte Carlo simulation results for the measurement model in Eq. (33)

Estimation algorithm	Nonlinear programming	Haupt/Kasdin two-step estimator	Proposed algorithm
Correct solution, %	97.8	— —	100

Note that there is no entry for the Haupt/Kasdin estimator, because the measurement model cannot be easily cast into a linear first step.
Note: The estimate parameter $\hat{\xi}$ is considered correct when $\|\xi_{\text{true}} - \hat{\xi}\|_2 \leq 0.1$.

Table 4 Estimation results from proposed algorithm for scalar example 2

Parameter	True	Estimate	Standard deviation
a	1	0.9980	0.0474
b	0.05	0.0546	0.0059
c	0.1	0.0895	0.0351
d	1	0.9897	0.0333
R	0.1016	0.1008	— —

Note: The true $\boldsymbol{R}$ is calculated using the 100 noise v samples.

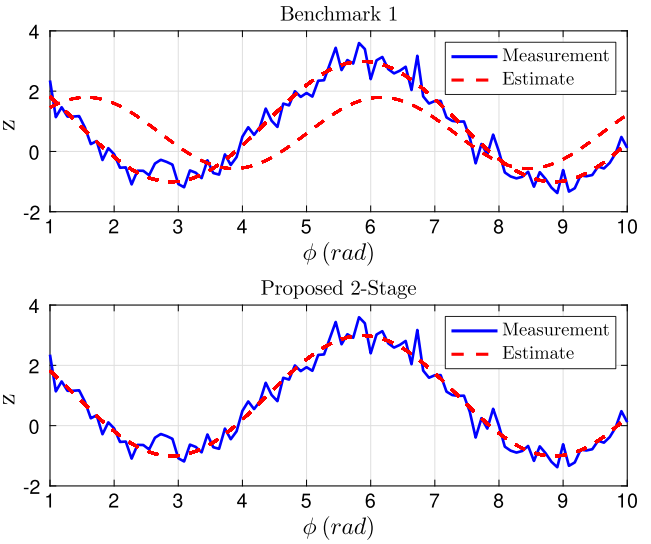


Fig. 5 Random 100 MC simulation results from 2 different methods using Eq. (33). Benchmark 1 occasionally fails to converge to the correct estimates, whereas the proposed algorithm works consistently.

(i.e., the same local minimum) due to the nonzero wind vector and high nonlinearity in the measurement model.

Because many of the details are discussed in detail in Ref. [18], we only present information required to facilitate understanding. The flight test was conducted on an Ultra Stick 120 UAV. The Ultra Stick 120 was initially used as a low-cost flight test platform at NASA Langley Research Center [19]. The Ultra Stick 120 is equipped with a traditional Pitot-static tube, a 5-hole probe [20], a GPS (u-blox-Neo-M8N), an Inertial Measurement Unit (IMU; Invensense MPU-9250), and a camera. The onboard software provides a Global Navigation Satellite System/Inertial Navigation System (GNSS/INS) navigation solution in real time through an open-source flight control system [21].

Equations (35a–35d) show the states, input, output, and parameters to be estimated. All the states in Eq. (35a) are assumed to be known or measured from the onboard navigation solution. The input $\boldsymbol{u}$ in Eq. (35b) is the direct pressure measurement from the 5-hole probe. The output $\boldsymbol{z}$ is the inertial velocity, resolved in the north–east–down (NED) frame, which is also from the navigation solution. The estimated parameter $\boldsymbol{\xi}$ includes airspeed scale factor λ_{V_a} and bias b_{V_a}, angle of attack and sideslip scale factors and biases $\lambda_\alpha, b_\alpha, \lambda_\beta, b_\beta$, installation misalignment angle ϵ_ϕ of the 5-hole probe rotated about the longitudinal axis of the fixed-wing aircraft, and the steady wind vector components W_N, W_E, W_D. Those parameters are known to be observable through various flight excitation (wind circle, pushover-pullup, pitch chirp, yaw chirp, rudder doublet, and multisines) as described in Ref. [18,22]. Table 5 summarizes the input design, time specifications, and where these data are used in the proposed algorithm [18]. Note that only those design inputs are used for the calibration—the estimated results are validated with the entire flight trajectory.

Table 5 Input design and time specification for calibration

Maneuver type	Time, s	Usage
Wind circle 1	[384, 408.2]	Stage 1
Wind circle 2	[411, 438.3]	Stage 1
Pushover-pullup (POPU)	[510.9, 530]	Stage 1
Multisine 1	[576, 596]	Stage 2
Multisine 2	[690, 711]	Stage 2
Multisine 3	[752, 772]	Stage 2
Multisine 4	[810, 830]	Stage 2
Pitch chirp 1	[867, 887]	Stage 2
Yaw chirp	[980.080, 1013.595]	Stage 2
Pitch chirp 2	[1041, 1061]	Stage 2
Rudder doublet	[1113, 1115]	Stage 2

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

$$\boldsymbol{x} = [\,p \quad q \quad r \quad b_{g_x} \quad b_{g_y} \quad b_{g_z} \quad \phi \quad \theta \quad \psi\,]^T \tag{35a}$$

$$\boldsymbol{u} = [\,P_{\Delta\alpha} \quad P_{\Delta\beta} \quad P_t \quad P_s\,]^T \tag{35b}$$

$$z = [\,V_N \quad V_E \quad V_D\,]^T \tag{35c}$$

$$\boldsymbol{\xi} = [\,\lambda_{V_a} \quad b_{V_a} \quad \lambda_\alpha \quad b_\alpha \quad \lambda_\beta \quad b_\beta \quad \epsilon_\phi \quad W_N \quad W_E \quad W_D\,]^T \tag{35d}$$

Equation (36) shows the air data error model. Though linear in the unknown parameters, it is determined to be sufficient for capturing the error dynamics in this 5-hole probe (Sec. II in [18]).

$$\begin{aligned} V_a &= (1+\lambda_{V_a})\sqrt{\frac{2(P_t-P_s)}{\rho}} + b_{V_a} \\ \alpha &= (1+\lambda_\alpha)\frac{P_{\Delta\alpha}}{K_\alpha(P_t-P_s)} + b_\alpha \\ \beta &= (1+\lambda_\beta)\frac{P_{\Delta\beta}}{K_\beta(P_t-P_s)} + b_\beta \end{aligned} \tag{36}$$

Equation (37) is the wind triangle equation resolved in the NED frame. The vector $\boldsymbol{V}$ and $\boldsymbol{W}$ are inertial vector and wind vector resolved in the NED frame. The vector $\boldsymbol{V}_{a,cg}$ is the airspeed vector (which consists of the body-axis translational components) at the center of IMU (in this case very close to the center of gravity, hence denoted with the subscript cg), and $\boldsymbol{V}_{a,s}$ is the airspeed vector at the 5-hole probe sensor location. The matrix C_b^n is the coordinate transformation from body frame to inertial frame, and $C(\epsilon_\phi)$ accounts for installation misalignment angle ϕ rotated about the longitudinal axis. Finally, $\boldsymbol{\omega}$ and $\boldsymbol{r}$ are the corrected rotational velocity and displacement vector from the center of the IMU in the UAV to the 5-hole probe sensor location. The exact formulation of C_b^n, $C(\epsilon_\phi)$, $\boldsymbol{\omega}$, and $\boldsymbol{r}$ are shown in Eqs. (38) and (39).

$$\boldsymbol{V} = C_b^n \boldsymbol{V}_{a,cg} + \boldsymbol{W} = C_b^n[C(\epsilon_\phi)\boldsymbol{V}_{a,s} - [\boldsymbol{\omega}]_\times \boldsymbol{r}] + \boldsymbol{W} \tag{37}$$

$$\boldsymbol{\omega} = \begin{bmatrix} p-b_p \\ q-b_q \\ r-b_r \end{bmatrix} \qquad \boldsymbol{r} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} \qquad C(\epsilon_\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\epsilon_\phi & \sin\epsilon_\phi \\ 0 & -\sin\epsilon_\phi & \cos\epsilon_\phi \end{bmatrix} \tag{38}$$

$$C_b^n = \begin{bmatrix} \cos\theta\cos\psi & \sin\phi\sin\theta\sin\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \tag{39}$$

We use the wind triangle equation in Eq. (37) as the measurement equation with assumed additive Gaussian white noise $[\,v_{V_N} \;\; v_{V_E} \;\; v_{V_D}\,]^T$ to represent measurement noise and to recast it into a suitable form for the two-stage estimator as follows:

$$z_k = \begin{bmatrix} V_N \\ V_E \\ V_D \end{bmatrix}_k = \underbrace{\begin{bmatrix} F\sqrt{\frac{2(P_t-P_s)}{\rho}} & F & I_3 \end{bmatrix}}_{\boldsymbol{A}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)} \underbrace{\begin{bmatrix} \lambda_{V_a} \\ b_{V_a} \\ W_N \\ W_E \\ W_D \end{bmatrix}}_{\boldsymbol{\xi}_1} + \underbrace{F\sqrt{\frac{2(P_t-P_s)}{\rho}} - C_b^n[\boldsymbol{\omega}]_\times \boldsymbol{r}}_{\boldsymbol{b}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)} + \underbrace{\begin{bmatrix} v_{V_N} \\ v_{V_E} \\ v_{V_D} \end{bmatrix}_k}_{\boldsymbol{v}_k} \tag{40}$$

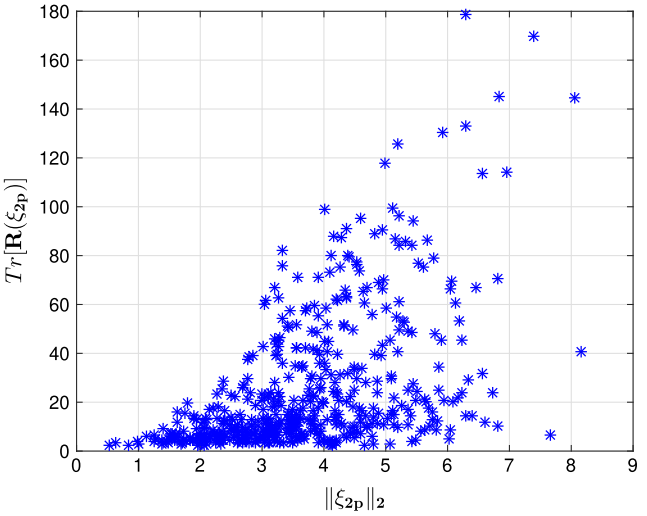


Fig. 6 Estimated Tr[$R(\xi_{2p})$] vs its corresponding 2-norm of ξ_{2p} using 500 random ξ_{2p}.

where the F is a 3-by-1 vector:

$$F = C_b^n C(\epsilon_\phi) \begin{bmatrix} \cos\alpha\cos\beta \\ \sin\beta \\ \sin\alpha\cos\beta \end{bmatrix} \tag{41}$$

The parameter vector $\boldsymbol{\xi}$ is now separated into $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ as shown in Eq. (42).

$$\boldsymbol{\xi}_1 = [\,\lambda_{V_a} \quad b_{V_a} \quad W_N \quad W_E \quad W_D\,]^T \tag{42a}$$

$$\boldsymbol{\xi}_2 = [\,\lambda_\alpha \quad b_\alpha \quad \lambda_\beta \quad b_\beta \quad \epsilon_\phi\,]^T \tag{42b}$$

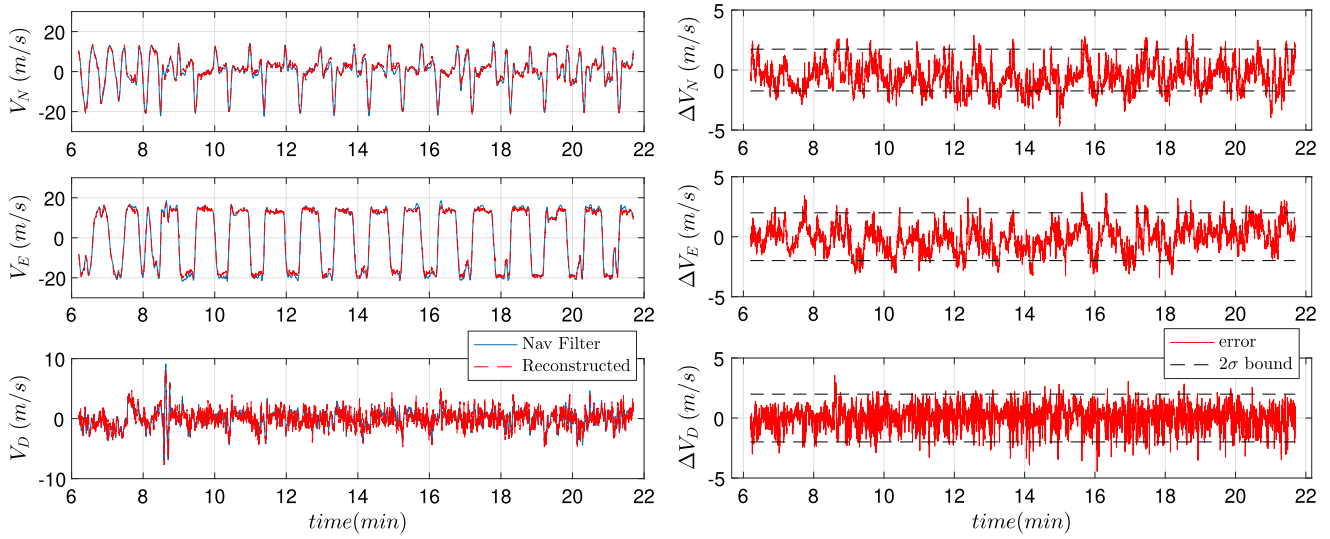
We estimate the parameter vector $\boldsymbol{\xi}$ using the proposed estimator. Intuitively, the proposed estimator works for this initial-condition-sensitive calibration problem because it isolates some of the nonzero parameters (e.g., wind vector) and minimizes the cost in the first stage until it is close to the optimal cost.

Because the parameter $\boldsymbol{\xi}_2$ is expected to be small, or at least bounded, we sampled 500 random $\lambda_\alpha, b_\alpha, \lambda_\beta, b_\beta$ from $N(0, 2^2)$ and ϵ_ϕ from $N(0, 0.3491^2)$ (standard deviation of 20 deg), respectively. Figure 6 shows Tr[$\boldsymbol{R}(\boldsymbol{\xi}_{2p})$] versus the 2-norm of $\boldsymbol{\xi}_{2p}$ using the 500 samples of $\boldsymbol{\xi}_2$. Notice that there is no unique local minimum (a flat region when $\|\boldsymbol{\xi}_{2p}\|_2 = 1$ to 4) using the sampled values, which means that the second-order condition is close to zero. This also means that the nonlinear estimator with respect to $\boldsymbol{\xi}_2$ only might not work well, because the inner optimization in Eq. (20) may not have brought the estimated cost close enough to the true cost, and so $\boldsymbol{\xi}_1$ cannot be uniquely determined in the minimal residual sense. Hence, we have to re-estimate $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ simultaneously with an initial guess $\boldsymbol{\xi}_{2p}$ and

Table 6 Parameter estimate, standard deviation, and constraint setting

Parameter	Two-stage (standard deviation)	Unit	Constraints used
λ_{V_a}	−0.1748 (0.0739)	— —	[−0.5, 0.5]
b_{V_a}	4.3553 (1.2672)	m/s	[−5, 5]
λ_α	0.2982 (0.3228)	— —	[−0.5, 0.5]
b_α	−2.4854 (0.5603)	deg	[−5, 5]
λ_β	−0.2673 (0.2379)	— —	[−0.5, 0.5]
b_β	−1.2980 (0.5164)	deg	[−5, 5]
ϵ_ϕ	−0.1380 (0.2699)	rad	[−0.2618, 0.2618]
W_N	−3.8038 (0.8704)	m/s	[−6, 6]
W_E	−2.4137 (0.9956)	m/s	[−6, 6]
W_D	−0.7168 (0.9952)	m/s	[−2, 2]

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713



a) Comparison between measured and corrected inertial velocity components

b) Error between measured and corrected inertial velocity components

Fig. 7 Measured vs corrected inertial velocity.

Table 7 Output root-mean-square error and measurement noise stand deviation ($\sqrt{\mathbf{diag}(\boldsymbol{R})}$)

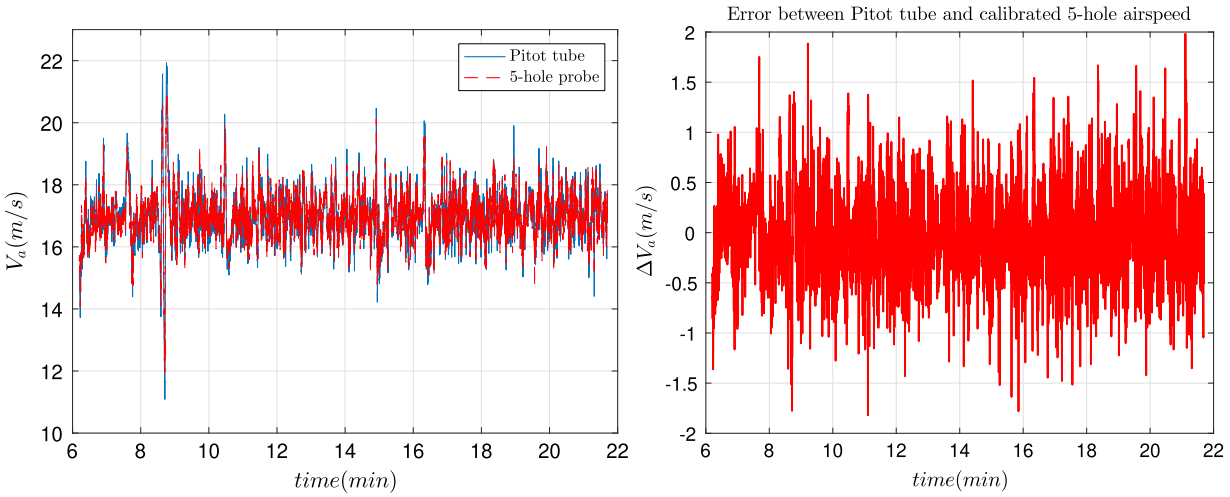
	Mean square error, m/s	Estimated standard deviation, m/s
V_N	1.2564	0.9875
V_E	1.1028	0.8645
V_D	0.8321	0.9207

the estimated $\hat{\boldsymbol{\xi}}_{1p}$ from the first stage. The initial guess of $\boldsymbol{\xi}_{2p} = [0.2, -1, -0.2, -1, -0.1]^T$ was determined to be a good initial condition from the 500 samples. Note that the 2-norm of $\boldsymbol{\xi}_{2p}$ should still be small based on Fig. 6.

Table 6 shows the final estimated parameters and the associated standard deviations in parentheses. It also lists the constraints used in the second stage, which was determined by the physical limitations of the system. The constraints of the wind vector were also refined based on the output of the first stage.

Figures 7a and 7b show the reconstructed (estimated) and measured (GNSS/INS solution) inertial velocity components and their error over the entire flight trajectory. Table 7 lists the root-mean-square error values of the estimated outputs and noise standard deviation from the estimated $\boldsymbol{R}$. The estimated output matches well with the measurement; the error plot is mostly bounded by the estimated 2 standard deviations. When using the single-stage method (benchmark 1—not shown), there is a large discrepancy between the reconstructed and measured inertial velocity, though the estimator was able to converge. This means that without good initial guess, the single-stage estimator may not always converge to the correct minimum. Estimating $\boldsymbol{\xi}_2$ only in the second stage also did not work well in terms of the error between measured and computed outputs.

Figure 8a shows the estimated airspeed from the 5-hole probe and the onboard airspeed measurement from an independently calibrated Pitot tube. The error between the estimated and measured airspeed is shown in Fig. 8b and the root-mean-square error was calculated to be 0.1241 m/s. The small error in airspeed when compared with another independent source also supports our claim that the proposed estimator worked well for this calibration problem.



a) Comparison between calibrated Pitot tube measurement and cablirated 5-hole probe estimate

b) Error calibrated Pitot tube measurement and cablirated 5-hole probe estimate

Fig. 8 Measured vs estimated 5-hole airspeed measurement.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

V. Conclusions

This paper presented a two-stage estimation algorithm for solving a class of nonlinear, parameter estimation problems that appear in aerospace engineering applications. This class of problems appears as a result of the mathematical form of the standard sensor error model used. Problems having this form can be recast into a problem that is linear with respect to a subset of the unknown parameters and nonlinear with respect to the remaining parameters. Implementation of the proposed estimator proceeds in two stages. In the first stage, linear least squares is used to obtain initial values for a subset of the unknown parameters and a residual sampling procedure is used for selecting initial values for the rest of the parameters. In the second stage, only a subset of the parameters needs to be re-estimated, and the rest of the parameters can be immediately calculated via weighted least squares. However, if a unique local minimum condition for the second stage cannot be determined, all the parameters have to be re-estimated simultaneously by a nonlinear constrained optimization. The examples provided in this paper show that this approach alleviates the initial condition sensitivity issue and minimizes the likelihood of converging to an incorrect local minimum of the nonlinear cost function. It also provides a technique for selecting initial conditions for a nonlinear measurement model that has the same canonical form. Furthermore, it was shown that, if the measurement model and unknown parameters satisfy certain conditions (i.e., Lipschitz continuity and finite domain), then the error in the final cost of the optimization has an upper bound.

Although the problems presented in this paper had static parameters, the algorithm can be used to find initial conditions with a mini-batch data set for dynamic problems as well. Therefore, this algorithm is yet one more tool available to the designer of estimators for nonlinear engineering problems.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Appendix A: Derivation of Canonical Form

The purpose of this Appendix is to show how the model structure given by Eq. (1) arises in aerospace estimation problems. The structure arises from what we refer to in this paper as the *standard sensor error models*. Although not referred to as such, its mathematical form is given in Ref. [5] [Eq. (10.13)] and Ref. [7] [Eqs. (4.16) and (4.17)] and it relates a vector measurement $\boldsymbol{z} \in \mathbb{R}_{3\times1}$ made by a sensor (e.g., an accelerometer triad) to the actual physical quantity being measured, denoted by $\boldsymbol{y} \in \mathbb{R}_{3\times1}$. Mathematically, it is the affine map from $\boldsymbol{y}$ to $\boldsymbol{z}$ given by

$$\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y} + \boldsymbol{n} + \boldsymbol{v} \tag{A1}$$

where the entries of $\boldsymbol{C} \in \mathbb{R}_{3\times3}$ represent systematic errors such as scale factor deviations and axes misalignments. The vector $\boldsymbol{n} \in \mathbb{R}_{3\times1}$ represents null-shifts (biases), and $\boldsymbol{v} \in \mathbb{R}^{3\times1}$ represents random, output noise normally modeled as a normal distribution with some given covariance. The entries of the matrix $\boldsymbol{C}$ and $\boldsymbol{n}$ are usually unknown parameters and need to be estimated. Discussion of the nature of the entries in $\boldsymbol{C}$, $\boldsymbol{n}$, and $\boldsymbol{v}$ is beyond the scope of this paper, but we refer the interested reader to the text by Ref. [7] (Chap. 4) for more details. In this Appendix, we are interested in the mathematical structure of $\boldsymbol{C}$, which is normally the product of multiple matrices, each representing a different type of error.

Let us consider a typical simple case where $\boldsymbol{C}$ is the product of two matrices: a misalignment error matrix, $\boldsymbol{C}_{\boldsymbol{\eta}}$, and scale factor error matrix, $\boldsymbol{C}_{\lambda}$. The subscript $\boldsymbol{\eta}$ represents the vector $\boldsymbol{\eta} = [\,\eta_1 \quad \eta_2 \quad \eta_3\,]^T$, whose entries are small misalignment errors between the triads $\boldsymbol{z}$ and $\boldsymbol{y}$. Because the entries in $\boldsymbol{\eta}$ are normally very small (i.e., $\eta_i \ll 1$, $i = 1, 2, 3$), the matrix $\boldsymbol{C}_{\eta}$ is approximated as a skew symmetric matrix of the vector $\boldsymbol{\eta}$. Similarly, the subscript λ represents the vector of scale factor errors $\boldsymbol{\lambda} = [\,\lambda_1 \quad \lambda_2 \quad \lambda_3\,]^T$. The scale factor errors $\lambda_i \ll 1$, $i = 1, 2, 3$ and appear on the diagonal of $\boldsymbol{C}_{\lambda}$. This leads to $\boldsymbol{C}$ having the following structure:

$$\begin{aligned}\boldsymbol{C} = \boldsymbol{C}_{\boldsymbol{\eta}}\boldsymbol{C}_{\lambda} &= \begin{bmatrix} 1 & -\eta_3 & \eta_2 \\ \eta_3 & 1 & -\eta_1 \\ -\eta_2 & \eta_3 & 1 \end{bmatrix}\begin{bmatrix} 1+\lambda_1 & 0 & 0 \\ 0 & 1+\lambda_2 & 0 \\ 0 & 0 & 1+\lambda_3 \end{bmatrix} \\ &= \begin{bmatrix} 1+\lambda_1 & -\eta_3 & \eta_2 \\ \eta_3 & 1+\lambda_2 & -\eta_1 \\ -\eta_2 & \eta_3 & 1+\lambda_3 \end{bmatrix}\end{aligned} \tag{A2}$$

where we have assumed $\eta_i\eta_j = \lambda_i\lambda_j = \lambda_i\eta_j = 0$ for $i = 1, 2, 3$. Note that if $\boldsymbol{\eta}$ is not small, it still can be recast into this structure.

This structure of the sensor output error affine map can be generalized if we replace $\boldsymbol{y}$ by $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\xi}')$ (so that it can include the known state $\boldsymbol{x}$ as well as the unknown parameters $\boldsymbol{\xi}$ and control inputs $\boldsymbol{u}$) and write it as:

$$\boldsymbol{z} = \left[\prod_{m=1}^{\infty} \boldsymbol{N}_m\boldsymbol{D}\right]\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\xi}') + \boldsymbol{n} + \boldsymbol{v} \tag{A3}$$

where $\boldsymbol{N}_m \in \mathbb{R}_{3\times3}$ are *nondiagonal* matrices and $\boldsymbol{D} \in \mathbb{R}_{3\times3}$ is a diagonal matrix. The product $\prod_{m=1}^{\infty}$ means that there can be infinitely many $\boldsymbol{N}$ matrices. In real applications, usually $m < 4$. The function $\boldsymbol{f}$ still can have unknown parameters associated with the input $\boldsymbol{u}$, but the number of unknowns in $\boldsymbol{f}$ is reduced due to factorization of the matrices $\boldsymbol{N}_m$. We denote the reduced parameter vector as $\boldsymbol{\xi}'$.

Because the unknown parameters of $\boldsymbol{D}$ are on the diagonal and the unknown bias vector $\boldsymbol{n}$ is additive, this can be transformed into the following linear affine form:

$$\begin{aligned}\boldsymbol{z}_{3\times1} &= \underbrace{\left[\,\prod_{m=1}^{\infty}\boldsymbol{N}_m\mathcal{D}[\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\xi}')] \quad I_{3\times3}\,\right]}_{\boldsymbol{A}(\boldsymbol{\xi}_2)}\underbrace{\begin{bmatrix} \boldsymbol{D}(1,1) \\ \boldsymbol{D}(2,2) \\ \boldsymbol{D}(3,3) \\ \boldsymbol{n} \end{bmatrix}}_{\boldsymbol{\xi}_1} \\ &\quad + \underbrace{\prod_{m=1}^{\infty}\boldsymbol{N}_m\boldsymbol{f}_{3\times1}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\xi}')}_{\boldsymbol{b}(\boldsymbol{\xi}_2)} + \boldsymbol{v}\end{aligned} \tag{A4}$$

where $\boldsymbol{A}(\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{\xi}_2)$ contain all the parameters in $\boldsymbol{N}$ and $\boldsymbol{f}$, and $\boldsymbol{\xi}_1$ represents the rest of the unknown parameters. The operator $\mathcal{D}[\cdot]$ takes in a vector and returns a square matrix with elements of the vector on the diagonal. If there are more measurement vectors that have the same structure shown in Eq. (A4), they can be concatenated as follows:

$$\begin{aligned}\boldsymbol{Z}_{3n\times1} &= \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1(\boldsymbol{\xi}_{2,1}) & \boldsymbol{0}_{3\times6} & \cdots & \cdots & \boldsymbol{0}_{3\times6} & \boldsymbol{0}_{3\times6} \\ \boldsymbol{0}_{3\times6} & \boldsymbol{A}_2(\boldsymbol{\xi}_{2,2}) & \boldsymbol{0}_{3\times6} & \cdots & \boldsymbol{0}_{3\times6} & \boldsymbol{0}_{3\times6} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0}_{3\times6} & \boldsymbol{0}_{3\times6} & \cdots & \cdots & \cdots & \boldsymbol{A}_n(\boldsymbol{\xi}_{2,n}) \end{bmatrix} \\ &\quad \times \begin{bmatrix} \boldsymbol{\xi}_{1,1} \\ \boldsymbol{\xi}_{1,2} \\ \vdots \\ \boldsymbol{\xi}_{1,n} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_1(\boldsymbol{\xi}_{2,1}) \\ \boldsymbol{b}_2(\boldsymbol{\xi}_{2,2}) \\ \vdots \\ \boldsymbol{b}_n(\boldsymbol{\xi}_{2,n}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_n \end{bmatrix} \\ &= \boldsymbol{A}(\boldsymbol{\xi}_2)\boldsymbol{\xi}_1 + \boldsymbol{b}(\boldsymbol{\xi}_2) + \boldsymbol{V}\end{aligned} \tag{A5}$$

where $\boldsymbol{\xi}_1 = [\boldsymbol{\xi}_{1,1}, \boldsymbol{\xi}_{1,2}, \ldots, \boldsymbol{\xi}_{1,n}]^T$ and $\boldsymbol{\xi}_2 = [\boldsymbol{\xi}_{2,1}, \boldsymbol{\xi}_{2,2}, \ldots, \boldsymbol{\xi}_{2,n}]^T$. The combination of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ represents the total unknown parameters vector $\boldsymbol{\xi}$. Even though the total measurement vector $\boldsymbol{Z}$ in Eq. (5) has $3n$ number of elements, it does not have to be multiple of three, depending on the given measurement model. For example, quaternion-related measurements can have an even number of measurement equations.

There are many parameter estimation problems that can be recast into this canonical form in the field of aerospace engineering.

For example, magnetometers are used extensively in navigation, guidance, and control applications [16,23,24], and the measurement error model of magnetometer calibration can be reformulated into the form of Eq. (A4). Another application in aircraft system identification is data compatibility analysis [5]. Instrumentation errors from IMU and air data systems in both dynamic and measurement models can also be reformulated into this canonical form. Other applications, such as attitude estimation [25], air data calibration [18,26], and stereo vision systems [27,28], also have similar models that can be reformulated into this canonical form.

In Appendix B, we show how two classical estimation problems can be reformulated into the canonical form shown in Eq. (A4). The first example deals with the magnetometer calibration error model taken from Ref. [23]. The second example deals with dynamic model equations for aircraft data compatibility analysis from Ref. [5]. The first example deals only with a measurement error model, assuming that the time series is available. The second example considers unknown parameters from both dynamic and measurement error models.

It should be noted that though some problems can be recast into canonical form shown in Eq. (A4), it does not mean that the proposed method would necessarily be better than using conventional methods for parameter estimation. For example, even though the data compatibility problem can be solved by the proposed estimator, the proposed algorithm does not prove improved accuracy compared with the well-known output-error method. What is unique about the proposed algorithm is that it may resolve the initial-value sensitivity problem if the measurements can be recast in suitable form, as demonstrated by the 5-hole Pitot tube calibration example in Sec. IV.

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Appendix B: Application Examples

B.1. Magnetometer Calibration

Consider the following magnetometer error equation [23]:

$$\boldsymbol{h}^m = \boldsymbol{C}_\alpha \boldsymbol{C}_\eta \boldsymbol{C}_\lambda \boldsymbol{h}^b + \boldsymbol{n} + \boldsymbol{v} \tag{B1}$$

where $\boldsymbol{C}_\alpha$, $\boldsymbol{C}_\eta$, and $\boldsymbol{C}_\lambda$ are soft-iron, misalignment, and scale factor error matrices, respectively. $\boldsymbol{h}^b = [\,h_x^b \quad h_y^b \quad h_z^b\,]^T$ is the true field magnetic vector in the body axes of the vehicle, and $\boldsymbol{h}^m$ is the measured magnetic field vector. Null shifts or hard-iron biases are represented by the constant vector $\boldsymbol{n}$. The effect of wide-band, sampling, or sensor noise (uncorrelated noise) is represented by the vector $\boldsymbol{v}$. For details of this model, refer to Ref. [23]. Note that a more-complicated model can be found in Ref. [24], where time-varying parameters are included in the measurement model. The objective is to estimate the following model parameters:

$$\boldsymbol{\xi} = [\,\alpha_{ij} \quad \eta_i \quad \lambda_i \quad n_i\,]^T \tag{B2}$$

where i can be x, y, or z.

With simple algebraic manipulation, the following canonical form can be obtained:

$$\boldsymbol{z} = \boldsymbol{h}^m = \underbrace{\left[\begin{array}{c|c} \boldsymbol{C}_\alpha \boldsymbol{C}_\eta \begin{pmatrix} h_x^b & & \\ & h_y^b & \\ & & h_z^b \end{pmatrix} & \boldsymbol{I}_{3\times 3} \end{array}\right]}_{\boldsymbol{A}(\boldsymbol{\xi}_2)} \underbrace{\begin{bmatrix} \lambda_x \\ \lambda_y \\ \lambda_z \\ n_x \\ n_y \\ n_z \end{bmatrix}}_{\boldsymbol{\xi}_1} + \underbrace{\boldsymbol{C}_\alpha \boldsymbol{C}_\eta \begin{bmatrix} h_x^b \\ h_y^b \\ h_z^b \end{bmatrix}}_{\boldsymbol{b}(\boldsymbol{\xi}_2)} + \boldsymbol{v} \tag{B3}$$

where $\boldsymbol{\xi}$ are split into $\boldsymbol{\xi}_1 = [\,\lambda_i \quad n_i\,]^T$ and $\boldsymbol{\xi}_2 = [\,\alpha_{ij} \quad \eta_i\,]^T$. It can be clearly seen that Eq. (B3) has the same form as Eq. (A4).

B.2. Aircraft Data Compatibility

Another common application in aerospace engineering is data compatibility analysis. In particular, aircraft data compatibility analysis is a process of estimating and removing systematic instrumentation errors that create kinematic inconsistencies in the measured sensor data. The classic example from Ref. [5] is used to show how this application can also be transformed into the canonical form. The typical states $\boldsymbol{x}$, input $\boldsymbol{u}$, measurement $\boldsymbol{z}$, and set of typical parameters $\boldsymbol{\xi}$ for this problem are given by the following:

$$\boldsymbol{x} = [\,u \quad v \quad w \quad \phi \quad \theta \quad \psi\,]^T \tag{B4a}$$

$$\boldsymbol{u} = [\,a_x \quad a_y \quad a_z \quad p \quad q \quad r\,]^T \tag{B4b}$$

$$\boldsymbol{z} = [\,V_a \quad \beta \quad \alpha \quad \phi \quad \theta \quad \psi\,]^T \tag{B4c}$$

$$\boldsymbol{\xi} = [b_{a_x}\ b_{a_y}\ b_{a_z}\ b_p\ b_q\ b_r\ \lambda_{V_a}\ \lambda_\alpha\ \lambda_\beta\ b_{V_a}\ b_\alpha\ b_\beta\ \lambda_\phi\ \lambda_\theta\ \lambda_\psi\ b_\phi\ b_\theta\ b_\psi]^T \tag{B4d}$$

where $\lambda_\phi, \lambda_\theta, \lambda_\psi, b_\phi, b_\theta, b_\psi$ are scale factors and biases of Euler angles in addition to the parameters introduced in the earlier sections. The dynamic model for data compatibility analysis is

$$\begin{aligned} \begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix} &= \begin{bmatrix} 0 & r + b_r & -(q + b_q) \\ -(r + b_r) & 0 & p + b_p \\ q + b_q & -(p + b_p) & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\ &\quad + \begin{bmatrix} -g\sin\theta + a_x + b_{a_x} \\ g\sin\phi\cos\theta + a_y + b_{a_y} \\ g\cos\phi\cos\theta + a_z + b_{a_z} \end{bmatrix} \\ \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} &= \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\phi & -\sin\phi \\ 0 & \dfrac{\sin\phi}{\cos\theta} & \dfrac{\cos\phi}{\cos\theta} \end{bmatrix} \begin{bmatrix} p + b_p \\ q + b_q \\ r + b_r \end{bmatrix} \end{aligned} \tag{B5}$$

Finally, the measurement model outputs $\boldsymbol{z}$ are the airspeed, air flow angles, and Euler angles:

$$\begin{aligned} V_a &= (1 + \lambda_V)\sqrt{u^2 + v^2 + w^2} + b_{V_a} + v_{V_a} \\ \beta &= (1 + \lambda_\beta)\sin^{-1}\left(v/\sqrt{u^2 + v^2 + w^2}\right) + b_\beta + v_\beta \\ \alpha &= (1 + \lambda_\alpha)\tan^{-1}(w/u) + b_\alpha + v_\alpha \\ \phi &= (1 + \lambda_\phi)\phi + b_\phi + v_\phi \\ \theta &= (1 + \lambda_\theta)\theta + b_\theta + v_\theta \\ \psi &= (1 + \lambda_\psi)\psi + b_\psi + v_\psi \end{aligned} \tag{B6}$$

With some algebraic manipulation, the measurement output model can be recast into the canonical form as follows:

$$z_k \begin{bmatrix} V_a \\ \beta \\ \alpha \\ \phi \\ \theta \\ \psi \end{bmatrix}_k = \underbrace{\begin{bmatrix} V_{a_k} 1 & & & & & \\ & \beta_k 1 & & & & \\ & & \alpha_k 1 & & & \\ & & & \phi_k 1 & & \\ & & & & \theta_k 1 & \\ & & & & & \psi_k 1 \end{bmatrix}}_{\boldsymbol{A}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)} \underbrace{\begin{bmatrix} \lambda_{V_a} \\ b_{V_a} \\ \lambda_\beta \\ b_\beta \\ \lambda_\alpha \\ b_\alpha \\ \lambda_\phi \\ b_\phi \\ \lambda_\theta \\ b_\theta \\ \lambda_\psi \\ b_\psi \end{bmatrix}}_{\boldsymbol{\xi}_1} + \underbrace{\begin{bmatrix} V_k \\ \beta_k \\ \alpha_k \\ \phi_k \\ \theta_k \\ \psi_k \end{bmatrix}}_{\boldsymbol{b}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)} + \underbrace{\begin{bmatrix} v_{V_a} \\ v_\beta \\ v_\alpha \\ v_\phi \\ v_\theta \\ v_\psi \end{bmatrix}_k}_{\boldsymbol{v}_k} \tag{B7}$$

where $\boldsymbol{\xi}$ are split into $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ as follows:

$$\boldsymbol{\xi}_1 = [\lambda_V \quad \lambda_\alpha \quad \lambda_\beta \quad b_V \quad b_\alpha \quad b_\beta \quad \lambda_\phi \quad \lambda_\theta \quad \lambda_\psi \quad b_\phi \quad b_\theta \quad b_\psi]^T \tag{B8a}$$

$$\boldsymbol{\xi}_2 = [b_{a_x} \quad b_{a_y} \quad b_{a_z} \quad b_p \quad b_q \quad b_r]^T \tag{B8b}$$

The airspeed V_k, angle of sideslip β_k, and angle of attack α_k in $\boldsymbol{A}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)$ and $\boldsymbol{b}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\xi}_2)$ are calculated by the state $\boldsymbol{x}_k$ shown in Eq. (B9).

$$\begin{aligned} V_k &= \sqrt{u_k^2 + v_k^2 + w_k^2} \\ \beta_k &= \sin^{-1}\left(v_k / \sqrt{u_k^2 + v_k^2 + w_k^2}\right) \\ \alpha_k &= \tan^{-1}(w_k/u_k) \end{aligned} \tag{B9}$$

To use the proposed estimator, all the states $\boldsymbol{x}_k$ for $k = 1, \ldots, N$ have to be known, which is a downside of this algorithm. Also, data compatibility problems are not particularly sensitive to initial conditions. It was well known that the zero initial condition is sufficient to solve such problems via output error. Nevertheless, the proposed estimator is a viable and convenient alternative.

Both magnetometer calibration and aircraft data compatibility analysis examples reveal that a common nonlinear parameter estimation problem can be transformed into an affine linear model as shown in Eq. (A5). The unknown parameters are separated into two sets with simple algebraic manipulation. With this canonical form, the proposed estimator can be used to solve the parameter estimation problem accurately and consistently.

Acknowledgments

The authors gratefully acknowledge the Minnesota Invasive Terrestrial Plants and Pests Center (MITPPC) through the Minnesota Environment and Natural Resources Trust Fund for surveying applications. The authors also gratefully acknowledge Peter Seiler of the Department of Aerospace Engineering & Mechanics at the University of Minnesota, Twin Cities, for his insightful critique, technical discussions, and comments in support of developing and validating the work presented in this paper.

References

[1] Stengel, R. F., *Optimal Control and Estimation*, Dover, Mineola, NY, 1994, Chaps. 2, 4.

[2] Simon, D., *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley–Interscience, Hoboken, NJ, 2006, Chaps. 5–14.

[3] Kay, S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, 1993, Chap. 7.

[4] Jategaonkar, R., *Flight Vehicle System Identification: A Time Domain Methodology*, Vol. 216, AIAA, Reston, VA, 2006, Chap. 4.

[5] Klein, V., and Morelli, E., *Aircraft System Identification: Theory and Practice*, AIAA Education Series, AIAA, Reston, VA, 2006, Chap. 6.

[6] Grauer, J. A., "Real-Time Data-Compatibility Analysis Using Output-Error Parameter Estimation," *Journal of Aircraft*, Vol. 52, No. 3, 2015, pp. 940–947. https://doi.org/10.2514/1.C033182

[7] Groves, P. D., *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, 2nd ed., Artech House, Boston, 2013, Chap. 4.

[8] Lawton, W. H., and Sylvestre, E. A., "Elimination of Linear Parameters in Nonlinear Regression," *Technometrics*, Vol. 13, No. 3, 1971, pp. 461–467. https://doi.org/10.1080/00401706.1971.10488810

[9] Guttman, I., Pereyra, V., and Scolnik, H. D., "Least Squares Estimation for a Class of Non-Linear Models," *Technometrics*, Vol. 15, No. 2, 1973, pp. 209–218. https://doi.org/10.1080/00401706.1973.10489035

[10] Golub, G., and Pereyra, V., "The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate," *SIAM Journal on Numerical Analysis*, Vol. 10, No. 2, 1973, pp. 413–432. https://doi.org/10.1137/0710036

[11] Bates, D., and Watts, D., *Nonlinear Regression Analysis and Its Applications*, Wiley Series in Probability and Statistics—Applied Probability and Statistics Section Series, Wiley, New York, 1988. pp. 78–79.

[12] Haupt, G. T., Kasdin, N. J., Keiser, G. M., and Parkinson, B. W., "Optimal Recursive Iterative Algorithm for Discrete Nonlinear Least-Squares Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 19, No. 3, 1996, pp. 643–649. https://doi.org/10.2514/3.21669

[13] Alonso, R., and Shuster, M. D., "TWOSTEP: A Fast Robust Algorithm for Attitude-Independent Magnetometer-Bias Determination," *The Journal of the Astronautical Sciences*, Vol. 50, No. 4, 2002, pp. 433–451.

[14] Hauer, J. F., Demeure, C. J., and Scharf, L. L., "Initial Results in Prony Analysis of Power System Response Signals," *IEEE Transactions on Power Systems*, Vol. 5, No. 1, 1990, pp. 80–89. https://doi.org/10.1109/59.49090

[15] Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, 1963, pp. 431–441. https://doi.org/10.1137/0111030

[16] Crassidis, J. L., Lai, K.-L., and Harman, R. R., "Real-Time Attitude-Independent Three-Axis Magnetometer Calibration," *Journal of Guidance, Control, and Dynamics*, Vol. 28, No. 1, 2005, pp. 115–120. https://doi.org/10.2514/1.6278

[17] "fmincon," MathWorks, 2019, https://www.mathworks.com/help/optim/ug/fmincon.html [retrieved 5 Oct. 2019].

[18] Sun, K., Regan, C. D., and Gebre-Egziabher, D., "A GNSS/IMU-Based 5-Hole Pitot Tube Calibration Algorithm," *AIAA Scitech 2019 Forum*, AIAA Paper 2019-0360, 2019. https://doi.org/10.2514/6.2019-0360

[19] Owens, B., Cox, D., and Morelli, E., "Development of a Low-Cost Sub-Scale Aircraft for Flight Research: The FASER Project," AIAA Paper 2006-3306, 2006. https://doi.org/10.2514/6.2006-3306

[20] *Rosemount Model 858 Flow Angle Sensors, Bulletin 1014*, Rosemount, Inc., Burnsville, MN, 1988.

[21] "Unversity of Minnesota UAV Laboratories," 2019, https://www.uav.aem.umn.edu/resources/goldy-iii [retrieved 1 Oct. 2019].

[22] Sun, K., Regan, C. D., and Gebre-Egziabher, D., "Observability and Performance Analysis of a Model-Free Synthetic Air Data Estimator," *Journal of Aircraft*, Vol. 56, No. 4, 2019, pp. 1471–1486. https://doi.org/10.2514/1.C035290

[23] Gebre-Egziabher, D., "Magnetometer Autocalibration Leveraging Measurement Locus Constraints," *Journal of Aircraft*, Vol. 44, No. 4, 2007, pp. 1361–1368. https://doi.org/10.2514/1.27118

[24] Springmann, J. C., and Cutler, J. W., "Attitude-Independent Magnetometer Calibration with Time-Varying Bias," *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 4, 2012, pp. 1080–1088. https://doi.org/10.2514/1.56726

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

[25] Crassidis, J. L., Markley, F. L., and Cheng, Y., "Survey of Nonlinear Attitude Estimation Methods," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 1, 2007, pp. 12–28. https://doi.org/10.2514/1.22452

[26] Jurado, J. D., and McGehee, C. C., "Complete Online Algorithm for Air Data System Calibration," *Journal of Aircraft*, Vol. 56, No. 2, 2019, pp. 517–528. https://doi.org/10.2514/1.C034964

[27] Chu, C.-C., Lie, F. A. P., and Gebre-Egziabher, D., "Dual Hypothesis Filter for Robust INS/Camera Fusion," *Proceedings of the 2013 International Technical Meeting of the Institute of Navigation*, Inst. of Navigation, Manassas, VA, 2013, pp. 792–802.

[28] Johnson, D. T., Nykl, S. L., and Raquet, J. F., "Combining Stereo Vision and Inertial Navigation for Automated Aerial Refueling," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 9, 2017, pp. 2250–2259. https://doi.org/10.2514/1.G002648

Downloaded by UNIVERSITY OF MINNESOTA on April 8, 2022 | http://arc.aiaa.org | DOI: 10.2514/1.G004713

Journal of Economic Entomology, 114(3), 2021, 1336–1344
doi: 10.1093/jee/toab058
Advance Access Publication Date: 3 May 2021
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Plant Resistance

Variation in Soybean Aphid (Hemiptera: Aphididae) Biotypes Within Fields

S. J. Bhusal,[1] R. L. Koch,[2] and A. J. Lorenz[1,3]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108, USA, [2]Department of Entomology, University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, Saint Paul, MN 55108, USA, and [3]Corresponding author, e-mail: lore0149@umn.edu

Subject Editor: Kristopher Giles

Received 29 August 2020; Editorial decision 3 March 2021

Abstract

Soybean aphid (*Aphis glycines* Matsumura (Hemiptera: Aphididae)) has been a major pest of soybean in North America since its detection in this continent in 2000 and subsequent spread. Although several aphid resistance genes have been identified, at least four soybean aphid biotypes have been discovered, with three of them being virulent on soybean cultivars with certain soybean aphid resistance genes. These biotypes are known to vary across years and locations, but information on their variation within single fields is limited. An investigation was conducted to study the variation of soybean aphid biotypes within single townships and fields in Minnesota. Screening of 28 soybean aphid isolates collected from seven soybean fields (six soybean fields in Cairo and Wellington Townships of Renville County, MN and one field in Wilmar Township of Kandiyohi County, MN) revealed the existence of multiple known biotypes of soybean aphid within single fields of soybean. We found up to three biotypes of soybean aphid in a single field. Two biotypes were found in five fields while only one field had only a single biotype. Three isolates presented reactions on a panel of resistant and susceptible indicator lines that were different from known biotypes. These results highlight the importance of characterizing soybean aphid biotypes in small geographical areas and utilizing generated knowledge to develop soybean cultivars pyramided with multiple resistance genes. The outcome will be decreased use of insecticides, thereby improving economic and environmental sustainability of soybean production.

Keywords: soybean aphid biotype, resistance gene, virulence

Soybean aphid (*Aphis glycines* Matsumura), native to eastern Asia, was first detected in the United States in 2000. Soybean aphid has spread across at least 30 states of the United States and three provinces of Canada (Ragsdale et al. 2011). Soybean aphid is a major pest of soybean in Minnesota, causing significant loss of yield (up to 40%) and reduced seed quality (Ragsdale et al. 2007). Soybean aphids damage plants by sucking sap from plant tissues, resulting in stunted plant growth and decreased photosynthetic and transpiration rates (Wang et al. 1962, Macedo et al. 2003, Ragsdale et al. 2004). Growth of sooty mold on aphids' excretion (i.e., honeydew) further worsen plant photosynthesis and transpiration (Wu et al. 2004). Plant pathogenic viruses such as soybean mosaic virus and alfalfa mosaic virus can be transmitted by soybean aphids, contributing to yield reduction (Hill et al. 2001).

Management of soybean aphid has been heavily dependent on use of chemical insecticides, resulting in a dramatic increase of insecticide use in soybean fields (Ragsdale et al. 2011, Hodgson et al. 2012). Sole reliance on insecticides is not an ideal option as it can cause unintended environmental impacts, kill beneficial insects, trigger frequent pest outbreak, and lead to the development of insecticide resistance (Pedigo and Rice 2009). Hanson et al. (2017) and Menger et al. (2020) reported soybean aphids that had developed resistance to widely used insecticides.

Deployment of aphid-resistant soybean varieties, on the other hand, can be an effective, economical, and eco-friendly option by providing a preventive measure against aphid outbreaks. The development of biotypes capable of overcoming aphid-resistant varieties complicates the deployment of varietal resistance as an effective management tool. The evolution of biotypes is generally common within aphid species because of several characteristics related to their life cycle and feeding habits (Michel et al. 2011). For example, soybean aphid completes its sexual reproduction on buckthorn (*Rhamnus*

© The Author(s) 2021. Published by Oxford University Press on behalf of Entomological Society of America.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

spp., a primary host) during fall and overwinters there (Ragsdale et al. 2004), which may provide a source of inoculum of different biotypes within a geographical area. Since soybean aphids were first discovered in North America, four biotypes have already been identified and characterized (Fig. 1). Biotype 1 is avirulent to soybean genotypes that carry any genes of soybean aphid resistance known as *Rag* (Resistance to *Aphis glycines*) genes. Biotype 2 is virulent on soybean genotypes with the *Rag1* (Kim et al. 2008) and biotype 3 is virulent to soybean genotypes with *Rag2*, *Rag3*, *rag4*, or *rag1c* (Hill et al. 2010). Biotype 4 is virulent to *Rag1*, *Rag2*, and *Rag1* + *Rag2* genotypes (Alt and Ryan-Mahmutagic 2013). Widescale surveying of soybean aphid biotypes in North America reported that specific biotypes are not limited to specific geographical locations, but are widely distributed (Cooper et al. 2015, Crossley and Hogg 2015, Alt et al. 2019).

Knowledge on variation in biotypes within a single field or a township is limited. Using microsatelite molecular markers, Michel et al. (2009) found genetic variation among clones within single aphid populations collected at single sampling sites, but biotyping was not performed. Providing information on the frequency and distribution of different biotypes within single townships and even single fields would be informative for soybean breeders assessing the usefulness of new cultivars with *Rag* genes and farmers growing such cultivars as part of an integrated pest management system. Such information could also contribute to the development of models forecasting soybean aphid population shifts in soybean fields, and guide selection of aphid-resistant varieties for future cultivation. Therefore, we aimed to investigate the variation in known soybean aphid biotypes at the level of single fields spread across single townships in Minnesota.

Methods

Preliminary (Pilot) Test

A pilot study was performed in 2017 by collecting soybean aphid isolates from two different fields approximately 3 km apart in northwestern Minnesota (Roseau); two isolates from central Minnesota (one isolate each from Saint Paul and Rosemount, about 30 km apart); and two isolates from two different fields approximately 3 km apart in southwestern Minnesota (Lamberton). The six collected isolates were quarantined for 72 h in a growth chamber to

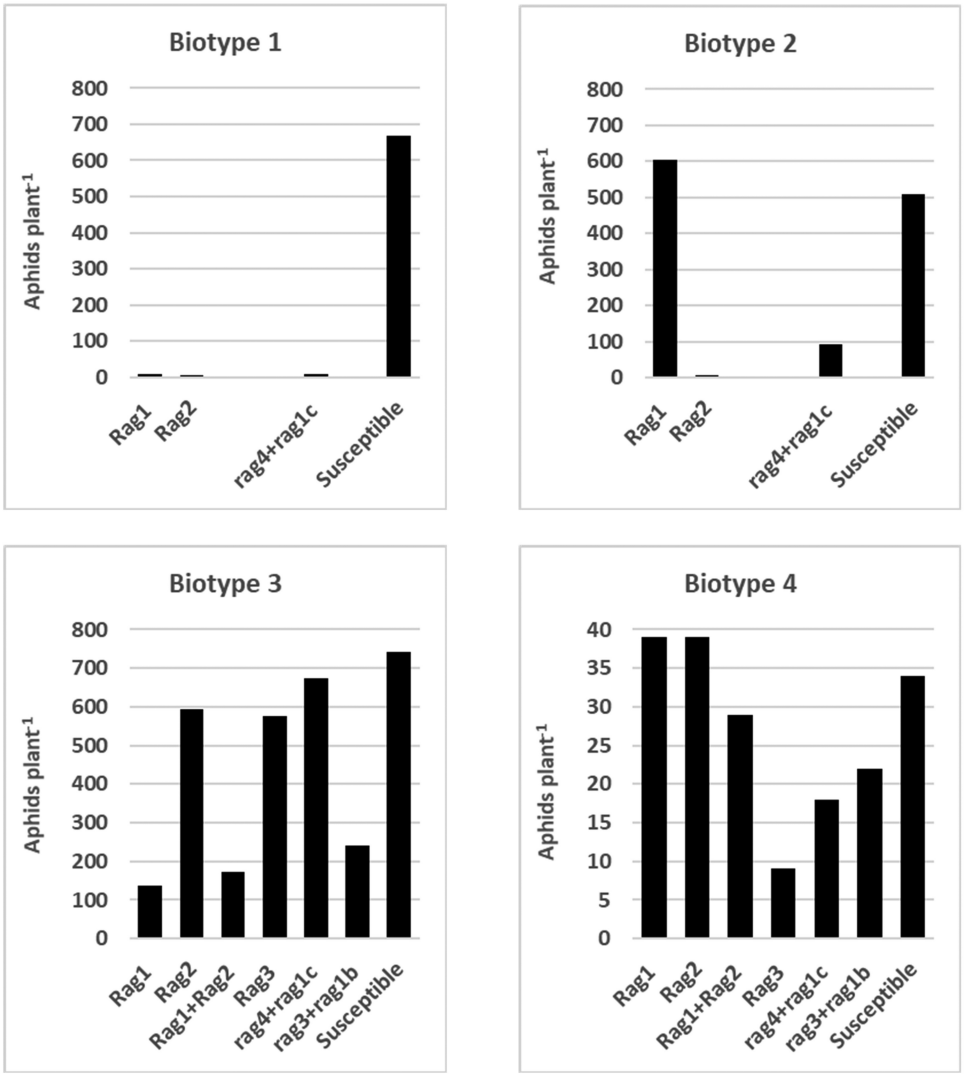


Fig. 1. Soybean aphids per plant on sets of soybean indicator lines carrying known resistance to *Aphis glycines* (*Rag/rag*) genes. Data were obtained from the original studies reporting the discovery of the biotypes. For biotypes 1 and 2 data were obtained from Kim et al. (2008). For biotype 3 data were obtained from Hill et al. (2010), and data were obtained from Alt and Ryan-Mahmutagic (2013) for biotype 4.

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

develop parasitoid-free colonies. One percent agar media was prepared in advance and transferred to small clear plastic cups (1 oz capacity, Item #: 999P100C, Webstaurant Store, Lancaster, PA) with a fresh leaf disc (1.5 inches diameter) of aphid-susceptible soybean, Sheyenne, placed on top of the media. Approximately 20 adult soybean aphids were transferred from a single plant per site to a leaf disc in agar media and covered with a perforated lid. Aphids were allowed to reproduce over a period of 24 h in growth chamber at 25°C, a photoperiod of 16:8 (L:D) h, and approximately 70% humidity. After 24 h, adult aphids were removed and newly born nymphs were quarantined for another 48 h. Any aphids showing symptoms of parasitism were removed and the healthy nymphs were transferred to soybean plants of an aphid-susceptible cultivar, Sheyenne, for reproduction in collapsible cages (13.5 × 13.5 × 24 inches, BioQuip Products, Inc., Rancho Dominguez, CA).

The soybean aphid isolates were screened against a panel of indicator soybean lines including a susceptible check and six lines carrying different aphid resistance genes (Table 1). Many of these indicator lines were used to characterize aphid biotypes by previous studies (Hill et al. 2010, Alt and Ryan-Mahmutagic 2013, Cooper et al. 2015). The screening tests were conducted with no-choice experiments by using clear-cup cages (Bhusal et al. 2013) in the growth chamber (maintained at 25°C, 16:8 (L:D) h period, and approximately 70% humidity) with three replicates in a randomized complete block design. Such no-choice experiments specifically assess effects of antibiosis on the aphids. Other experimental details of plant maintenance, aphid inoculation and maintenance, aphid counts, and statistical analysis are described under, *Biotype Testing*, section below.

Sampling Design

Based on preliminary results finding different biotypes in the same region, we designed our sampling strategy to investigate variation of soybean aphid populations within and between fields of single townships in Minnesota. Four separate isolates were collected from each field sampled. In 2018, three fields were sampled in Cairo Township of Renville County and one field in Wellington Township of Renville County (Table 2). In 2019, we sampled two fields in Cairo Township of Renville County and one field in Wilmar Township of Kandiyohi County (Table 2). The fields in Cairo and Wellington Townships of Renville County in 2018 and 2019 are displayed in Fig. 2. In total, 28 soybean aphid isolates were collected from seven fields. For each isolate, we collected approximately 20 adult aphids from a single plant (preferably from a single leaf). The collected isolates were quarantined to rear parasite-free aphid populations as described above.

Biotype Testing

Collected isolates were biotyped using the panel of indicator lines (Table 1). We performed a whole-plant bioassay in growth chambers, because bioassays with whole plants differentiate the biotypes more accurately than using detached-leaf assays (Lagos-Kutz et al. 2020). Three replicates of the biotyping were performed, with each replicate forming a complete block of the indicator lines listed in Table 1. In each replication, the indicator lines were randomized such that the susceptible check was always placed in the middle and the indicator lines carrying aphid resistance gene(s) were randomized around the susceptible check inside a rearing cage. This allowed uniform movement of aphids between the susceptible check and surrounding indicator lines. Such bioassays performed over the duration described below enable assessment of the combined effects of antixenosis and antibiosis on the aphids.

Three seeds of each line were planted in a 10.2 cm × 10.2 cm × 10.2 cm plastic pot filled with Berger BM2 germination mix (Berger Horticultural Products Ltd., Sulphur Springs, TX). After germination, plants were thinned so that each pot contained two soybean seedlings. At the unifoliate stage (Fehr and Caviness 1977), 10 mixed-aged apterous aphids were transferred to the first unfolding trifoliate leaf of each of the two plants in each pot. The experimental arrangements with aphid-infested plants were maintained in fine mesh insect rearing cages in growth chambers at 25°C, a photoperiod of 16:8 (L:D) h, and approximately 70% humidity. Plants were bottom-irrigated in the holding trays to avoid interference of irrigation water to soybean aphid infestations. Total number of aphids on each plant was counted at 2 wk after inoculation. Mean aphid counts per plant per replication were calculated and mean number of aphids across the replications of each treatment were plotted in the bar graphs. Analysis of variance was performed separately for each isolates using a model including soybean line and replication as fixed effects. Bonferroni correction was used to separate means at $P < 0.05$. Aphid isolates in resistant indicator lines were determined virulent when their infestation was similar to susceptible checks or different from other resistant indicator lines.

Results

Pilot Test

The results of the pilot test are presented in Fig. 3 and Supp Table 1 (online only). 'Roseau Field 1' isolate was similar to soybean aphid biotype 1 as it was virulent to the susceptible check (Sheyenne) but avirulent to all soybean indicator lines (Kim et al. 2008). However, 'Roseau Field 2' isolate was virulent to the soybean indicator lines that carried aphid resistance genes *Rag2* (PI 200538), *Rag3* (PI 567543C), or *rag4* + *rag1c* (PI 567541B) demonstrating responses similar to biotype 3 (Hill et al. 2010). 'Saint Paul' isolate was similar to biotype 1 except it was partially virulent to *Rag3*. 'Rosemount' isolate and 'Lamberton Field 2' isolate were virulent to *Rag1*, *Rag2*, and the *Rag1* + *Rag2* combination similar to biotype 4 (Alt and

Table 1. A panel of indicator soybean lines used for biotype screening

Line	Maturity group	Gene(s) of aphid resistance[a]	Reference
PI 548663 (Dowling)	VIII	*Rag1*	Hill et al. (2006)
PI 200538	VIII	*Rag2*	Hill et al. (2009)
Viking 2188	II	*Rag1* and *Rag2*	Albert Lea Seed House, Albert Lea, MN
PI 567543C	III	*Rag3*	Zhang et al. (2010)
PI 567541B	III	*rag4* and *rag1c*	Zhang et al. (2009)
PI 567598B	III	*rag3* and *rag1b*	Bales et al. (2013)
Sheyenne	0	None	

[a]The presence of upper case R in the gene name indicates the allele conferring resistance is dominant, whereas a lower case r indicates the allele conferring resistance is recessive.

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

Table 2. Details of fields sampled for soybean aphid isolates (four isolates collected per field) from Minnesota in 2018 and 2019

Sampling location				GPS coordinates			
Field	Township	County	Sampling date	Isolate 1	Isolate 2	Isolate 3	Isolate 4
Field 1	Cairo	Renville	4 August 2018	44.470N, 94.720W	44.469N, 94.717W	44.473N, 94.717W	44.473N, 94.718W
Field 2	Cairo	Renville	4 August 2018	44.481N, 94.719W	44.481N, 94.716W	44.480N, 94.717W	44.479N, 94.720W
Field 3	Cairo	Renville	4 August 2018	44.500N, 94.701W	44.500N, 94.698W	44.502N, 94.700W	44.503N, 94.701W
Field 4	Wellington	Renville	4 August 2018	44.555N, 94.737W	44.556N, 94.732W	44.552N, 94.733W	44.550N, 94.735W
Field 5	Cairo	Renville	14 August 2019	44.509N, 94.693W	44.513N, 94.692W	44.514N, 94.686W	44.511N, 94.689W
Field 6	Cairo	Renville	14 August 2019	44.531N, 94.735W	44.530N, 94.732W	44.529N, 94.727W	44.532N, 94.726W
Field 7	Wilmar	Kandiyohi	5 August 2019	45.130N, 95.096W	45.129N, 95.098W	45.129N, 95.095W	45.130N, 95.094W

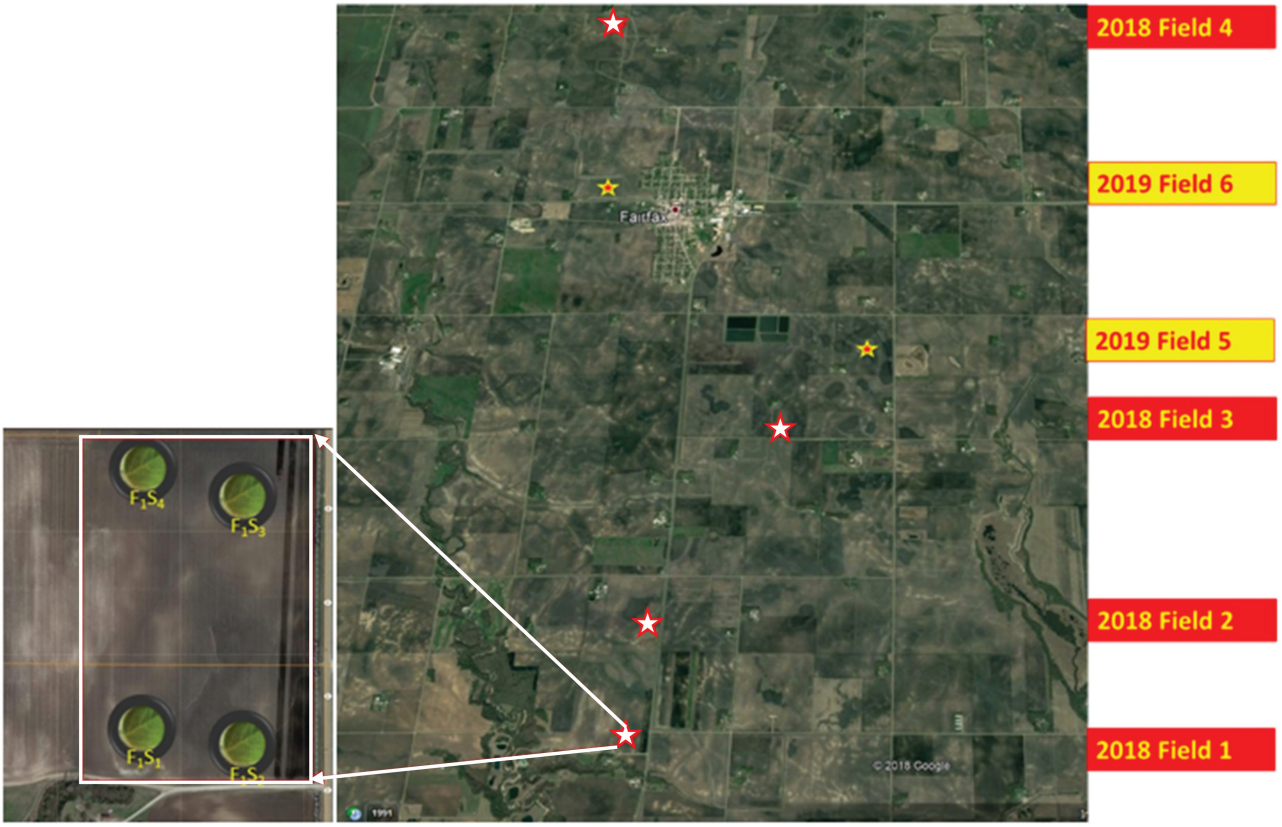


Fig. 2. Sampling sites (fields) of aphid isolates collected during 2018 and 2019 field season (Field 7 in Wilmar Township is not shown in the map because of its geographical distance from these field sites). Four isolates in all the fields were sampled to represent the corresponding field in different fashion such as from four corners in one field and three corners and deep in the field in other. The picture in the inset shows the location of four isolates sampled in field 1 in Cairo Township in 2018.

Ryan-Mahmutagic 2013). The 'Lamberton Field 2' isolate was collected from a *Rag1* + *Rag2* soybean cultivar growing at the University of Minnesota's Southwest Research and Outreach Center. Whereas 'Lamberton Field 1' isolate was found to be virulent to *Rag2*.

Biotype Screening Within Fields and Townships

Screening of aphid isolates against biotype indicator lines revealed different soybean aphid biotypes within fields and townships (Tables 3 and 4; Fig. 4). In 2018, three isolates collected from 'Field 1' in Cairo Township were avirulent to all of the resistant indicator lines and virulent to Sheyenne, the susceptible check. The remaining isolate from that field was virulent to susceptible check and *Rag2*, and partially virulent to all of the other indicator lines. Out of four isolates collected from 'Field 2', one isolate was avirulent to all of the resistant indicator lines, two isolates appeared to be virulent to *Rag1*, and the remaining isolate was virulent to *Rag2*, *Rag3*, or *rag4* + *rag1c*, indicating three different types of aphid populations in 'Field 2'. Two isolates from 'Field 3' and three isolates from 'Field 4' were avirulent to all of the resistant lines. The remaining isolates in these two fields were virulent to the indicator line carrying *Rag2*. Based on these results, out of 16 aphid isolates collected in 2018, nine isolates were similar to soybean aphid biotype 1 (Kim et al. 2008), two isolates were similar to biotype 2 (Kim et al. 2008), and one isolate was similar to biotype 3 (Hill et al. 2010). Three isolates were virulent to *Rag2*, but were unlike biotype 3 because they were avirulent to *Rag3* and the *rag4* + *rag1c* combination. One isolate ('Field 1–Isolate 2') was inconclusive (Table 4; Fig. 4).

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

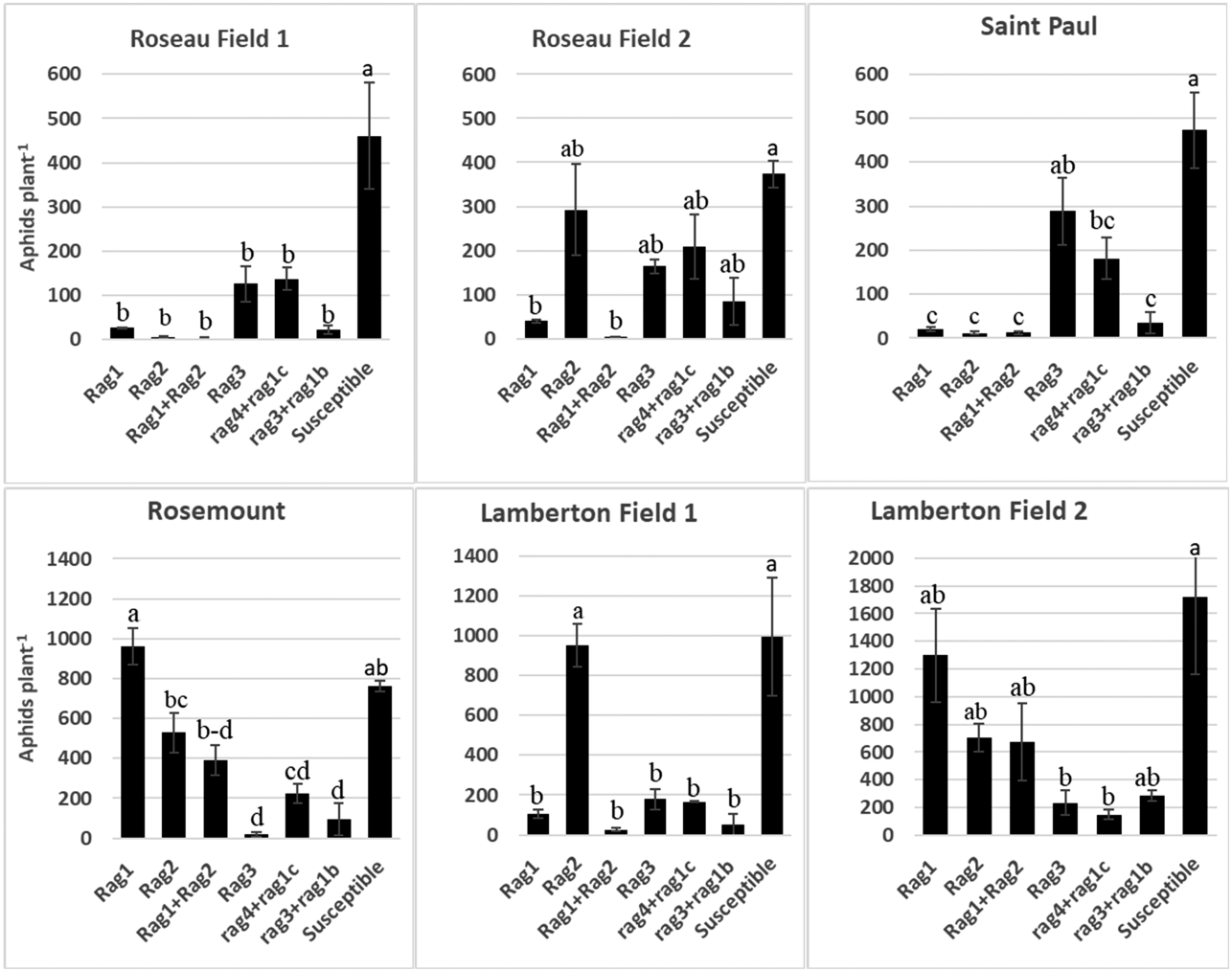


Fig. 3. Mean number of aphids per plant on different indicator lines observed in bioassay of the aphid isolates collected as part of the 2017 pilot test. Error bars represent the standard error of the means. Same letters above each bar within individual isolates are not different by LSD ($P > 0.05$).

In 2019, three isolates from 'Field 5', two isolates from 'Field 6', and two isolates from 'Field 7' were avirulent to all of the resistant indicator lines similar to biotype 1. One isolate from 'Field 5', two isolates from 'Field 6', and one isolate from 'Field 7' were virulent to *Rag1* demonstrating similar responses to biotype 2. Whereas 'Field 7–Isolate 1' was virulent to *Rag1* and *Rag2* individually, but not virulent to any resistant lines including the *Rag1* + *Rag2* combination, which suggests the presence of a mixture of biotype 1 and biotype 2 populations in that isolate. In summary, biotype 1 and biotype 2 were prevalent in all fields sampled in 2019 (Table 4; Fig. 4). The detailed results of screening of 2018 and 2019 collections are presented in Supp Table 2 (online only).

Discussion

A biotype is a population of an insect species that can survive on, reproduce on, and/or cause injury to a plant which is resistant to other populations of that insect species (Dogimont et al. 2010). The biotypic differentiation of soybean aphid is based on their reaction to soybean containing different genes of aphid resistance. Specific mechanisms of biotypic virulence are not well known (Alt et al. 2019) but the basis of virulence is hypothesized to involve the secreted effector proteins (Coates et al. 2020). Using a small number of single-nucleotide polymorphisms genotyped on different isolates, Wenger and Michel (2013) did not find significant genetic differentiation between isolates of soybean aphid biotype 1 and biotype 2 collected across northern Ohio. Based on this result, the authors speculated that variation in virulence does not stem from the development of a single, genetically distinct lineage of soybean aphid, but rather could be developed from ubiquitous genetic sources of virulence whereby virulence genes are broadly distributed in aphid populations at a low frequency. Other possible explanations include nongenetic causes of virulence and epigenetic variations. Similarly, Crossley and Hogg (2015) analyzed the clonal lineages of aphid isolates collected in Wisconsin during 2012 and 2013 and found that 41% and 8% of the isolates collected in 2012 and 2013, respectively, have matching multilocus genotypes. The dynamics of late-season dispersal and migration of soybean aphids to their primary winter host for sexual reproduction may cause admixture of a diverse sexual gene pool, and their clonal amplification in the following spring and summer may increase the heterozygosity in the aphid population (Orantes et al. 2012, Wenger and Michel 2013). A recent study that used whole-genome resequencing combined with population genomic analyses on different soybean aphid biotypes found that only a very small number of genomic regions were divergent between biotype 1 and biotypes 2, 3, and 4 (Coates et al. 2020). This finding suggests that a small number of loci control variation in virulence among soybean aphid biotypes.

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

Table 3. Virulence of collected soybean aphid isolates against different indicator lines

Aphid isolates	Indicator lines						
	PI 548663 (*Rag1*)	PI 200538 (*Rag2*)	Viking 2188 (*Rag1* + *Rag2*)	PI 567543C (*Rag3*)	PI 567541B (*rag4* + *rag1c*)	PI 567598B (*rag3* + *rag1b*)	Sheyenne (NA)
2018							
Field 1–Isolate 1	A	A	A	A	A	A	V
Field 1–Isolate 2	PV	V	PV	A	A	PV	V
Field 1–Isolate 3	A	A	A	A	A	A	V
Field 1–Isolate 4	A	A	A	A	A	A	V
Field 2–Isolate 1	A	V	A	V	PV	A	V
Field 2–Isolate 2	A	A	A	A	A	A	V
Field 2–Isolate 3	V	A	A	A	A	A	V
Field 2–Isolate 4	V	A	A	A	A	A	V
Field 3–Isolate 1	A	A	A	A	A	A	V
Field 3–Isolate 2	A	V	A	A	A	A	V
Field 3–Isolate 3	A	V	A	A	A	A	V
Field 3–Isolate 4	A	PV	A	A	A	A	V
Field 4–Isolate 1	A	A	A	A	PV	A	V
Field 4–Isolate 2	A	PV	A	A	A	A	V
Field 4–Isolate 3	A	V	A	A	PV	A	V
Field 4–Isolate 4	A	A	A	A	A	A	V
2019							
Field 5–Isolate 1	PV	A	A	A	A	A	V
Field 5–Isolate 2	V	A	A	A	A	A	V
Field 5–Isolate 3	A	A	A	A	A	A	V
Field 5–Isolate 4	A	A	A	A	A	A	V
Field 6–Isolate 1	A	A	A	A	A	A	V
Field 6–Isolate 2	A	A	A	A	A	A	V
Field 6–Isolate 3	V	A	A	A	A	A	V
Field 6–Isolate 4	V	A	A	A	A	A	V
Field 7–Isolate 1	V	V	A	A	A	A	V
Field 7–Isolate 2	A	A	A	A	A	A	V
Field 7–Isolate 3	V	A	A	A	A	A	V
Field 7–Isolate 4	A	A	A	A	A	A	V

A = avirulent; V = virulent; PV = partially virulent.

Table 4. Number of aphid isolates from each field that displayed virulence reactions similar to known soybean aphid biotypes 1–4

Sampling fields	Biotypes				
	1	2	3	4	Other[a]
2018					
Field 1	3				1
Field 2	1	2	1		
Field 3	2				2
Field 4	3				1
2019					
Field 5	3	1			
Field 6	2	2			
Field 7	2	1			1

[a]Aphid isolates displayed different responses than already known biotypes of soybean aphid for such as 'Field 3–Isolate 3', 'Field 3–Isolate 3', and 'Field 4–Isolate 3' (Fig. 4). 'Field 1–Isolate 2' was inconclusive. 'Field 7–Isolate 1' potentially mixture of soybean aphid biotype 1 and 2. Detailed results are presented in Supp Table 2 (online only).

Current knowledge on the variation of soybean aphid biotypes at a small spatial scale is very limited. Several studies have found variation of soybean aphid biotypes between different geographical regions. Cooper et al. (2015) found variation in soybean aphid biotypes from year to year and across U.S. locations. The widest spectrum of variability in soybean aphid virulence was found in Wisconsin, which was the first state to report the soybean aphid in North America (Ragsdale et al. 2011, Cooper et al. 2015). Crossley and Hogg (2015) found all four soybean aphid biotypes (i.e., biotypes 1, 2, 3, and 4) in 42, 21, 17, and 4%, respectively, of 24 aphid collection sites in Wisconsin in 2013. Our study also found higher variation of aphid biotypes in 2018 which had overall higher levels of aphid infestation statewide compared to 2019.

This investigation of soybean aphid biotype variation within a small spatial scale was rooted from the results of a pilot test of screening aphid isolates collected from northwestern, central, and southeastern Minnesota in 2017. The pilot test revealed that multiple biotypes of soybean aphid were prevalent in Minnesota and even in individual regions (Fig. 3; Supp Table 1 [online only]). Two aphid isolates, which were collected only about 3 km apart, presented reactions similar to two different biotypes of soybean aphid both in northwestern and in southeastern Minnesota. There were also two different biotypes in central Minnesota. However, the aphid isolates in central Minnesota were collected from two different cities approximately 30 km apart. These results suggest we found three known biotypes as well as one isolate different than known biotypes of soybean aphid in a small collection of aphid isolates from a single field season. Alt et al. (2019) reported all four known biotypes of soybean aphid in Minnesota as a part of the large collection of aphid isolates in the United States and Canada during 2011–2013. Such variation at a small spatial scale observed in the pilot study triggered

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

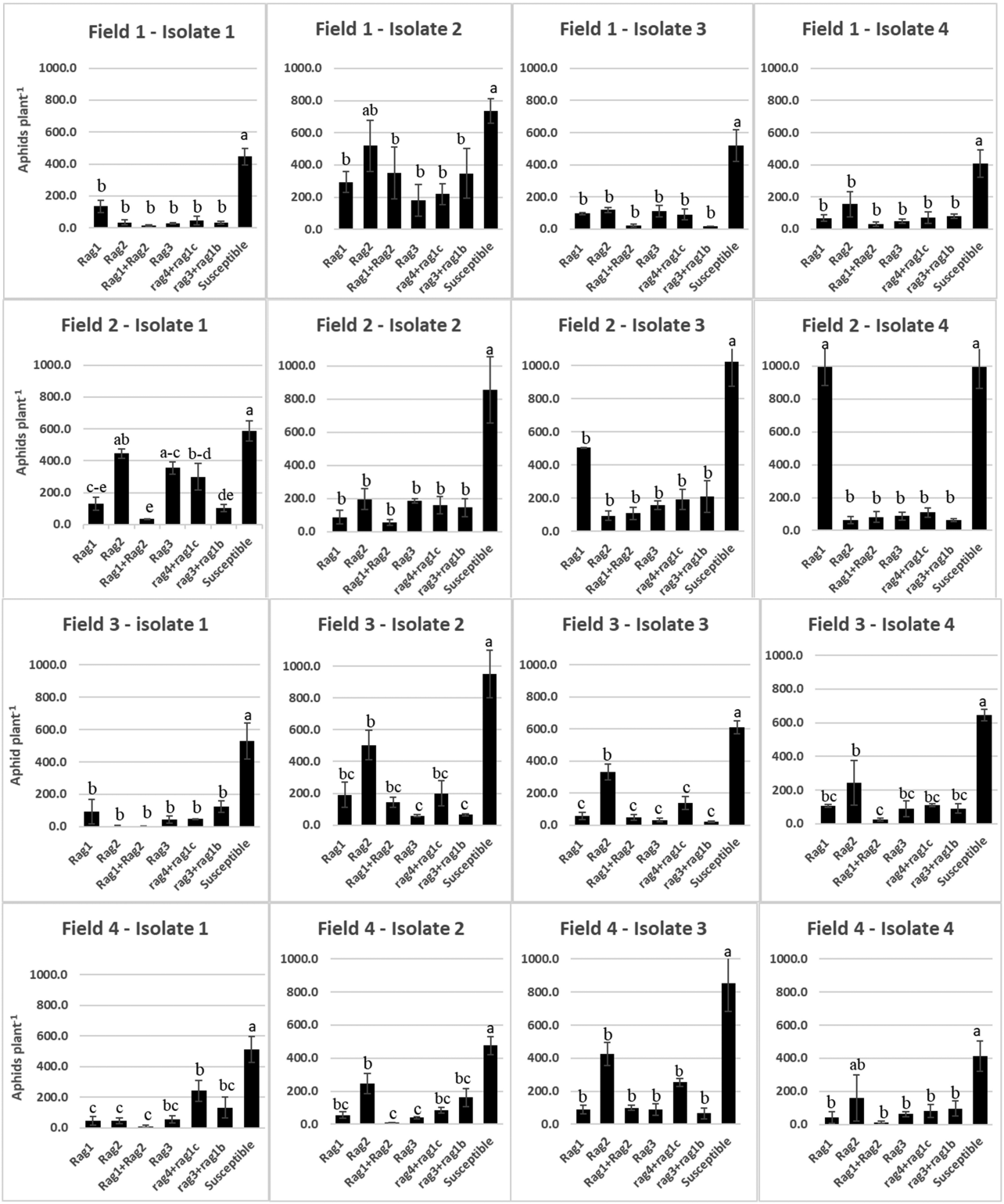


Fig. 4. Mean number of aphids per plant on different indicator lines observed in bioassay of the aphid isolates collected in 2018 and 2019 field season. Error bars represent the standard error of the means. Same letters above each bar within individual isolates are not different by LSD ($P > 0.05$).

an investigation into the variation of soybean aphid biotypes within township and within single fields.

Screening of aphid isolates collected in 2018 and 2019 revealed the presence of multiple biotypes of soybean aphid not only within a single township, but also within a single field. In 2018, four types of soybean aphid populations were prevalent including aphid populations similar to biotype 1, 2, and 3. The fourth type of aphid populations were different from previously known biotypes. The isolates different from previously known biotypes were virulent to *Rag2*, but unlike biotype 3 (Hill et al. 2010), they were avirulent to

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

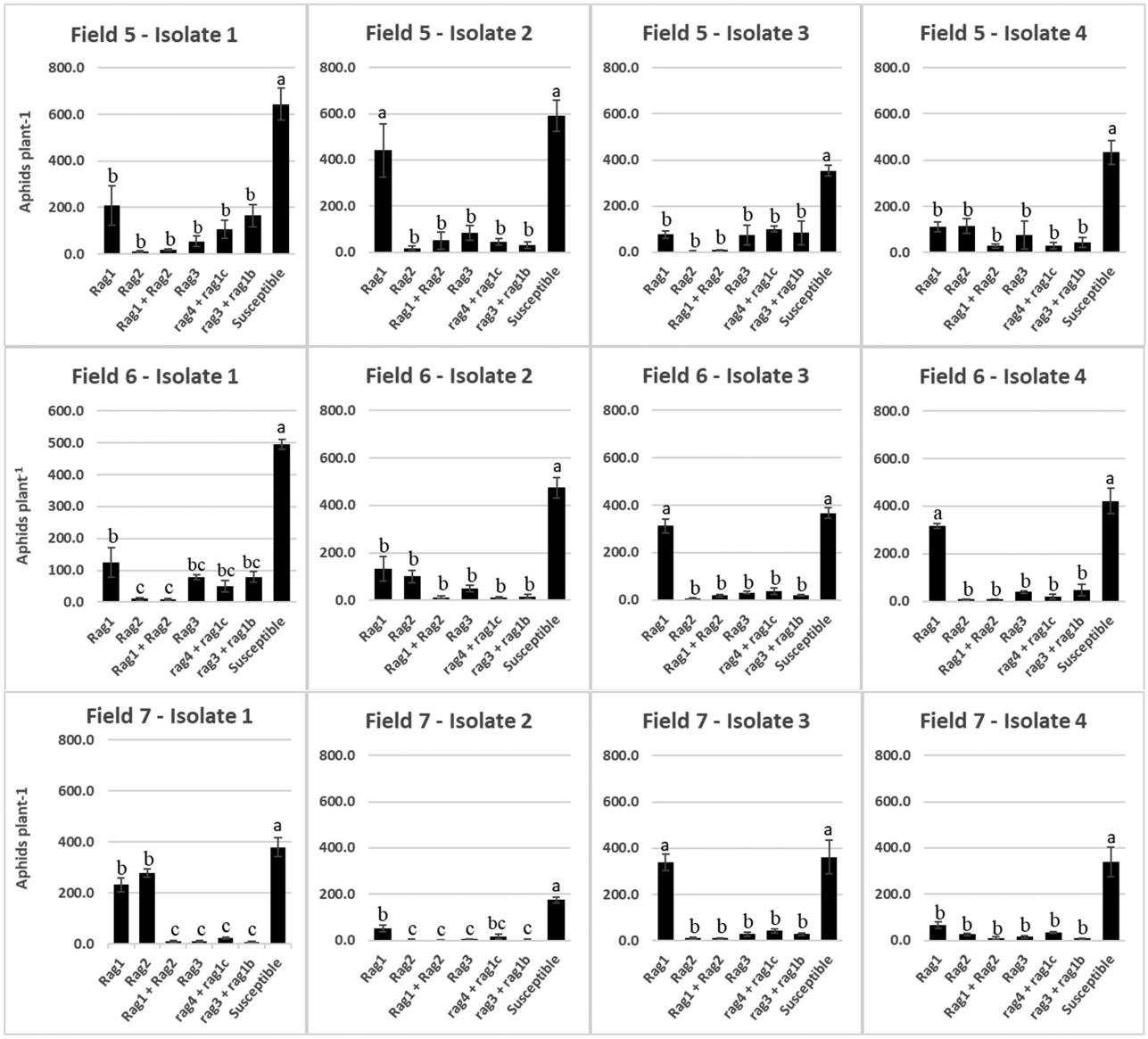


Fig. 4. Continued.

Rag3 and the *rag4* + *rag1c* combination (Table 4; Fig. 4). Lagos-Kutz (2020) also found a soybean aphid clone from Wooster, OH which readily colonized *Rag2* but did not colonize *Rag1* or *Rag3*. However, the authors did not screen this aphid clone against other resistant genes. The inconclusive results of an isolate from 'Field 1–Isolate 1' may have been attributed by the presence of intrapopulation or intrabiotypic variability as observed by Alt et al. (2019).

Soybean aphid infestation in Minnesota was not as widespread in 2019 compared to 2018, and therefore we were able to collect soybean aphid isolates from only two fields in Cairo Township of Renville County, MN, in 2019. We also collected aphid isolates from one field of its neighboring county, Kandiyohi. The aphid populations collected in 2019 were similar to either biotype 1 or biotype 2. Both types of aphid populations were present in all three fields sampled, but the populations similar to biotype 1 were more prevalent in each case (Table 4). The 'Field 7–Isolate 1', which was virulent to both *Rag1* and *Rag2* but avirulent to the *Rag1* + *Rag2*, makes it distinct from biotype 4. Because we collected several adult aphids from a single plant, the isolate may have been a mixture of two different biotypes. It is likely to have admixture of different populations/biotypes in natural aphid isolates due to their dispersal throughout the season (Wenger and Michel 2013). Alt et al. (2019) also reported intrabiotypic variability. It is important to further study genetic differences between potentially new soybean aphid biotypes to determine virulence attributes.

Detailed knowledge of how soybean aphid biotypes vary at different spatial scales will inform integrated pest management strategies for the control of this agriculturally damaging insect species. Previous studies of soybean aphid biotype variation were conducted at large geographical scales (e.g., state or regional [multistate] levels). Our current investigation examined variation of biotypes at the scale of townships and single fields. We found up to three different biotypes of soybean aphid in a single field. Six out of seven fields sampled were found to have more than one biotype. We found all four known biotypes of soybean aphid populations in Minnesota as well as populations that were different from known biotypes. These results highlight the importance of continually discovering new sources of aphid resistance

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

and developing soybean cultivars pyramided with multiple resistance genes. Varieties with pyramided *Rag* genes are crucial in integrated pest management of soybean aphid because they confer stronger and more durable resistance, and significant yield protection (McCarville et al. 2014, Zhang et al. 2018). Such pyramided resistant varieties and new genetic sources conferring resistance to multiple biotypes will help in decreasing insecticide inputs thereby improving economic and environmental sustainability of soybean production.

Supplementary Data

Supplementary data are available at *Journal of Economic Entomology* online.

Acknowledgments

We would like to acknowledge James Menger for helping in preparing and arranging logistic of the investigation. We would also like to thank Donn Vellekson for sampling soybean aphids from Roseau, MN. Bruce Potter and Steve Commerford are also acknowledged who helped in finding locations of aphid infestation. We are much thankful to Tom Palmer, Tom Runcke, and Jason Fussy who allowed us to collect aphids in their field. Funding for this investigation was provided by the Minnesota Invasive Terrestrial Plants and Pests Center through the Minnesota Environment and Natural Resources Trust Fund.

References

Alt, J., and M. Ryan-Mahmutagic. 2013. Soybean aphid biotype 4 identified. Crop Sci. 53: 1491–1495.

Alt, J., M. Ryan, and D. W. Onstad. 2019. Geographic distribution and intrabiotypic variability of four soybean aphid biotypes. Crop Sci. 59: 84–91.

Bales, C., G. Zhang, M. Liu, C. Mensah, C. Gu, Q. Song, D. Hyten, P. Cregan, and D. Wang. 2013. Mapping soybean aphid resistance genes in PI 567598B. Theor. Appl. Genet. 126: 2081–2091.

Bhusal, S. J., G. L. Jiang, K. J. Tilmon, and L. S. Hesler. 2013. Identification of soybean aphid resistance in early maturing genotypes of soybean. Crop Sci 53: 491–499.

Coates, B. S., J. D. Hohenstgein, R. Giordano, R. K. Donthu, A. P. Michel, E. W. Hodgson, and M. E. O'Neal. 2020. Genome scan detection of selective sweeps among biotypes of the soybean aphid, *Aphis glycines*, with differing virulence to resistance to *A. glycines* (*Rag*) traits in soybean, *Glycine max*. Insect Biochem. Molec. 124: 103364.

Cooper, S. G., V. Concibido, R. Estes, D. Hunt, G. L. Jiang, C. Krupke, B. McCornack, R. Mian, M. O'Neal, V. Poysa, D. Prischmann-Voldeseth, *et al.* 2015. Geographic distribution of soybean aphid biotypes in the United States and Canada during 2008–2010. Crop Sci. 55: 2598–2608.

Crossley, M. S., and D. B. Hogg. 2015. *Rag* virulence among soybean aphids (Hemiptera: Aphididae) in Wisconsin. J. Econ. Entomol. 108: 326–338.

Dogimont, C., A. Bendahmane, V. Chovelon, and N. Boissot. 2010. Host plant resistance to aphids in cultivated crops: genetic and molecular bases, and interactions with aphid populations. C. R. Biol. 333: 566–573.

Fehr, W. R., and C. E. Caviness. 1977. Stages of soybean development. Iowa Agric. Home Econ. Exp. Stn. Spec. Rep. 80: 1–12.

Hanson, A. A., J. Menger-Anderson, C. Silverstein, B. D. Potter, I. V. MacRae, E. W. Hodgson, and R. L. Koch. 2017. Evidence for soybean aphid (Hemiptera: Aphididae) resistance to pyrethroid insecticides in the upper midwestern United States. J. Econ. Entomol. 110: 2235–2246.

Hill, J. H., R. Alleman, D. B. Hogg, and C. R. Grau. 2001. First report of transmission of soybean mosaic virus and alfalfa mosaic virus by *Aphis glycines* in the New World. Plant Dis. 85: 561.

Hill, C. B., Y. Li, and G. L. Hartman. 2006. A single dominant gene for resistance to the soybean aphid in the soybean cultivar Dowling. Crop Sci. 46: 1601–1605.

Hill, C. B., K. S. Kim, L. Crull, B. W. Diers, and G. L. Hartman. 2009. Inheritance of resistance to the soybean aphid in soybean PI 200538. Crop Sci. 49: 1193–1200.

Hill, C. B., L. Crull, T. K. Herman, D. J. Voegtlin, and G. L. Hartman. 2010. A new soybean aphid (Hemiptera: Aphididae) biotype identified. J. Econ. Entomol. 103: 509–515.

Hodgson, E. W., B. P. McCornack, K. Tilmon, and J. J. Knodel. 2012. Management recommendations for soybean aphid (Hemiptera: Aphididae) in the United States. J. Integ. Pest Manag. 3: E1–E10.

Kim, K. S., C. B. Hill, G. L. Hartman, M. A. R. Mian, and B. W. Diers. 2008. Discovery of soybean aphid biotypes. Crop Sci. 48: 923–928.

Lagos-Kutz, D., M. L. Pawlowski, B. W. Diers, S. R. Purandare, K. J. Tilmon, and G. L. Hartman. 2020. Virulence of soybean aphid, *Aphis glycines* (Hemiptera: Aphididae) clones on detached leaves and whole plants. J. Kansas Entomol. Soc. 92: 497–511.

Macedo, T. B., C. S. Bastos, L. G. Higley, K. R. Ostlie, and S. Madhavan. 2003. Photosynthetic responses of soybean to soybean aphid (Homoptera: Aphididae) injury. J. Econ. Entomol. 96: 188–193.

McCarville, M. T., M. E. O'Neal, B. D. Potter, K. J. Tilmon, E. M. Cullen, B. P. McCornack, J. F. Tooker, and D. A. Prischmann-Voldseth. 2014. One gene versus two: a regional study on the efficacy of single gene versus pyramided resistance for soybean aphid management. J. Econ. Entomol. 107: 1680–1687.

Menger, J., P. Beauzay, A. Chirumamilla, C. Dierks, J. Govloski, P. Glogoza, K. Hamilton, E. W. Hodgson, J. J. Knodel, I. V. MacRae, and D. T. Pezzini. 2020. Implementation of a diagnostic-concentration bioassay for detection of susceptibility to pyrethroids in soybean aphid (Hemiptera: Aphididae). J. Econ. Entomol. 113: 932–939.

Michel, A. P., W. Zhang, J. Kyo Jung, S. T. Kang, and M. A. Mian. 2009. Population genetic structure of *Aphis glycines*. Environ. Entomol. 38: 1301–1311.

Michel, A. P., O. Mittapalli, and M. A. R. Mian. 2011. Evolution of soybean aphid biotypes: understanding and managing virulence to host-plant resistance, pp. 355–372. *In* A. Sudaric (ed.), Soybean – molecular aspects of breeding. InTech – Open Access Publisher, Rijeka, Croatia.

Orantes, L. C., W. Zhang, M. A. Mian, and A. P. Michel. 2012. Maintaining genetic diversity and population panmixia through dispersal and not gene flow in a holocyclic heteroecious aphid species. Heredity (Edinb). 109: 127–134.

Pedigo, L. P., and M. E. Rice. 2009. Entomology and pest management, 6th ed. Pearson Prentice Hall, Inc., Upper Saddle River, NJ.

Ragsdale, D. W., D. J. Voegtlin, and R. J. O'Neil. 2004. Soybean aphid biology in North America. Ann. Entomol. Soc. Am. 97: 204–208.

Ragsdale, D. W., B. P. McCornack, R. C. Venette, B. D. Potter, I. V. Macrae, E. W. Hodgson, M. E. O'Neal, K. D. Johnson, R. J. O'Neal, C. D. Difonzo, *et al.* 2007. Economic threshold for soybean aphid (Hemiptera: Aphididae). J. Econ. Entomol. 100: 1258–1267.

Ragsdale, D. W., D. A. Landis, J. Brodeur, G. E. Heimpel, and N. Desneux. 2011. Ecology and management of the soybean aphid in North America. Annu. Rev. Entomol. 56: 375–399.

Wang, C., N. Siang, and G. Chang. 1962. Studies on the soy-bean aphid, *Aphis glycines* Matsumura. Acta Entomol. Sinica. 11: 31–44.

Wenger, J. A., and A. P. Michel. 2013. Implementing an evolutionary framework for understanding genetic relationships of phenotypically defined insect biotypes in the invasive soybean aphid (*Aphis glycines*). Evol. Appl. 6: 1041–1053.

Wu, Z., D. Schenk-Hamlin, W. Zhan, D. W. Ragsdale, and G. E. Heimpel. 2004. The soybean aphid in China: a historical review. Ann. Entomol. Soc. Am. 97: 209–218.

Zhang, G., C. Gu, and D. Wang. 2009. Molecular mapping of soybean aphid resistance genes in PI 567541B. Theor. Appl. Genet. 118: 473–482.

Zhang, G., C. Gu, and D. Wang. 2010. A novel locus for soybean aphid resistance. Theor. Appl. Genet. 120: 1183–1191.

Zhang, S., Z. Wen, C. DiFonzo, Q. Song, and D. Wang. 2018. Pyramiding different aphid-resistance genes in elite soybean germplasm to combat dynamic aphid populations. Mol. Breed. 38: 29.

Downloaded from https://academic.oup.com/jee/article/114/3/1336/6262345 by guest on 05 April 2022

Journal of Integrated Pest Management, (2021) 12(1): 44; 1–4
https://doi.org/10.1093/jipm/pmab038
Brief Communication

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

First Reports of *Macrosaccus morrisella* (Lepidoptera: Gracillariidae) Feeding on Soybean, *Glycine max* (Fabales: Fabaceae)

Robert L. Koch,[1,3,] Joseph Moisan-De Serres,[2] and Arthur V. Ribeiro[1]

[1]Department of Entomology, University of Minnesota, Saint Paul, MN 55108, USA, [2]Laboratoire d'Expertise et de Diagnostic en Phytoprotection, Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec, QC, G1P 3W8, Canada, and [3]Corresponding author, e-mail: koch0125@umn.edu

Subject Editor: Jeffrey Davis

Received 20 September 2021; Editorial decision 1 November 2021

Abstract

Macrosaccus morrisella (Fitch) is a tiny leaf-mining moth native to North America. In this publication, we provide first reports of this insect feeding on soybean, *Glycine max* (L.) Merr., in Canada and the United States, describe its injury to soybean, and briefly review leaf miners associated with soybean. Further efforts related to *M. morrisella* should focus on its geographic extent of infestation of soybean, impacts to soybean, and ecology in the agricultural landscape.

Key words: leaf miner, host range, pest, soybean

Soybean, *Glycine max* (L.) Merr. (Fabales: Fabaceae), is an important crop to North American agriculture. Because soybean is a non-native crop, the herbivorous insects feeding on it in North America represent native species that have adapted to soybean and exotic species that have invaded (Kogan 1981, Kogan and Turnipseed 1987). Here we provide first reports of infestations of soybean in Canada and the United States by a native leaf-mining moth *Macrosaccus morrisella* (Fitch) (Lepidoptera: Gracillariidae).

Macrosaccus morrisella is widely distributed across eastern North America (Braun 1908, Davis and De Prins 2011). The identification of its life stages is detailed by Braun (1908) and Davis and De Prins (2011). The adults are tiny moths measuring 6–7 mm (Braun 1908). The front wings are patterned with orange, white and gray-black markings (Davis and De Prins 2011). Larvae reach about 4.7 mm in length and are pale green to white in color (Davis and De Prins 2011). The pupae are about 3.6 mm in length (Davis and De Prins 2011). *Macrosaccus* spp. feed on plants in the Fabaceae family (Davis and De Prins 2011). In particular, *M. morrisella* is known to feed on American hogpeanut, *Amphicarpaea bracteata* (L.) Fernald (Fabales: Fabaceae), and sickleseed fuzzybean, *Strophostyles leiosperma* (Torr. & A. Gray) Piper (Fabales: Fabaceae), which are both vining plants with trifoliate leaves (Davis and De Prins 2011). The observations presented below suggest that *M. morrisella* has adapted to feed on soybean.

Sample Collection and Identification of Insects

In Québec, Canada, white-colored blotch-type leaf mines were observed over several years. On 9 August 2016, heavily mined soybean leaves (e.g., Figs. 1 and 2) were collected from a field in Saint-Paul and sent to the Laboratoire d'Expertise et de Diagnostic en Phytoprotection (LEDP) for identification. During the following week, infested leaves were collected in Saint-Ambroise-de-Kildare (15 August 2016), Berthierville (18 August 2016), and Nicolet (18 August 2016), and sent to the LEDP. More than 10 mines could be seen on some leaflets. Additional soybean leaves with blotch-type leaf mines were collected from a field in Princeville on 8 July 2019 and from fields in Saint-Gervais and Yamachiche in 2021. Adult insects were reared from infested soybean leaves from each of these locations in Québec.

In Minnesota, United States, leaf mines similar to those in Québec were observed in soybean fields near St. Paul and Rosemount on 13 and 17 August 2021, respectively (e.g., Fig. 3). At both locations, infestations were light with only scattered plants showing symptoms of infestation. From the St. Paul and Rosemount fields, soybean leaves with blotch-type mines were collected. For each leaf, the number of mines, and length and width of each mine was recorded. Then, each mine was inspected under a dissecting scope to determine the presence and life stages of insects. The tip of an insect pin was used to carefully open intact mines. Pupae were reared to adults. Among the 45 infested leaves collected from these two fields in Minnesota, 37 leaves had one mine, five leaves had two mines, and three leaves had three mines. These mines average 12.5 ± 0.6 (±SEM) mm long and 7.3 ± 0.3 mm wide. Of the 56 mines observed across these leaves, 31 mines were empty, 19 mines had larvae (18 mines with one larva each, and one mine with three larvae), and six mines had pupae (five mines with one pupa each, and one mine with two pupae). Five adults later emerged from the seven pupae.

© The Author(s) 2021. Published by Oxford University Press on behalf of Entomological Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jipm/article/12/1/44/6438101 by guest on 05 April 2022

Morphological identification of adult specimens was performed using descriptions from Davis and De Prins (2011). Male genitalia were dissected and placed in 85% lactic acid and heated for 25 min at about 90°C, and observed using a Zeiss Discover V20 at 150×magnification. In addition, specimens from each location were subjected to DNA barcoding. DNA extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Following DNA extraction, amplification of the COI gene was conducted with primers LCO1490 and HCO2198 in reactions containing: 25.00 µl Invitrogen Platinum Hot Start PCR 2× MasterMix, 19.00 µl H_2O, 0.5 µl 10 µM forward primer, 0.5 µl 10 µM reverse primer, and 5 µl template DNA. Thermal cycler conditions were: 95°C for 3 min, 45 cycles of 95°C for 45 s, 53°C for 45 s, and 72°C for 45 s, and a final extension of 72°C for 3 min. Amplification success was verified using QIAxcel Advanced System (Qiagen). Purification and Sanger Sequencing were completed at the SANGER Sequencing Platform of the CHU de Québec-Université Laval Research Center on a ABI 3730xl DNA Analyzer. The sequences were trimmed, aligned, and assembled using Geneious v2021.2.1.

Observations and Significance

The leaf-mining insects obtained from soybean leaves collected in Québec and Minnesota were determined to be *M. morrisella* based on morphological identification and DNA barcoding. The blotch-type leaf mines of *M. morrisella* observed here on soybean (Figs. 1–3) were unlike the mines of *Odontota horni* Smith (Coleoptera: Chrysomelidae) (Kogan and Kogan 1979, Buntin 1994, Hodgson et al. 2012). Davis and De Prins (2011) described the mines of *M. morrisella* as: 'The mine begins as an elongate serpentine track on the abaxial side of the leaflet. This enlarges to an elongate-oval, whitish blotch which eventually becomes

Fig. 1. Blotch-type leaf mines caused by *Macrosaccus morrisella* in a soybean field. Image credit: J. Moisan-De Serres.

Fig. 3. Intact blotch-type leaf mine caused by *Macrosaccus morrisella* on abaxial (lower) side of a soybean leaflet. Image credit: R. L. Koch.

Fig. 2. Soybean leaves heavily infested with *Macrosaccus morrisella*. Image credit: J. Moisan-De Serres.

Downloaded from https://academic.oup.com/jipm/article/12/1/44/6438101 by guest on 05 April 2022

strongly tentiform'. The mines from *M. morrisella* in soybean did not cross the midribs or main lateral veins of the leaves, and one or more margins of individual mines were often defined by the midrib or lateral veins (Figs. 1–4). The lower (i.e., abaxial) surface of the mines was generally white in color and very thin (Figs. 1–3, 4B, and C). Sometimes the upper (i.e., adaxial) surface of the mine was slightly raised (i.e., tentiform) (Figs. 2 and 4A). In addition, the outer adaxial surface of the mines often had numerous small light-colored spots (Figs. 2 and 4A), which appeared to be caused by small feeding pits from larvae feeding on the inner side of that leaf surface (Fig. 5A). This spotted appearance was less apparent or absent for mines with small larvae. The inside of intact mines contained dark-colored frass from the larvae (Fig. 4C). Silken enclosures were found in mines with large larvae preparing to pupate and mines containing pupae (Fig. 5B). Most of the empty mines contained remnants of these silken enclosures and sometimes pupal exuviae, suggesting the adults had already emerged.

Leaf-mining insects are known to have physical and physiological impacts on their host plants (Liu et al. 2015). Three native leaf miners

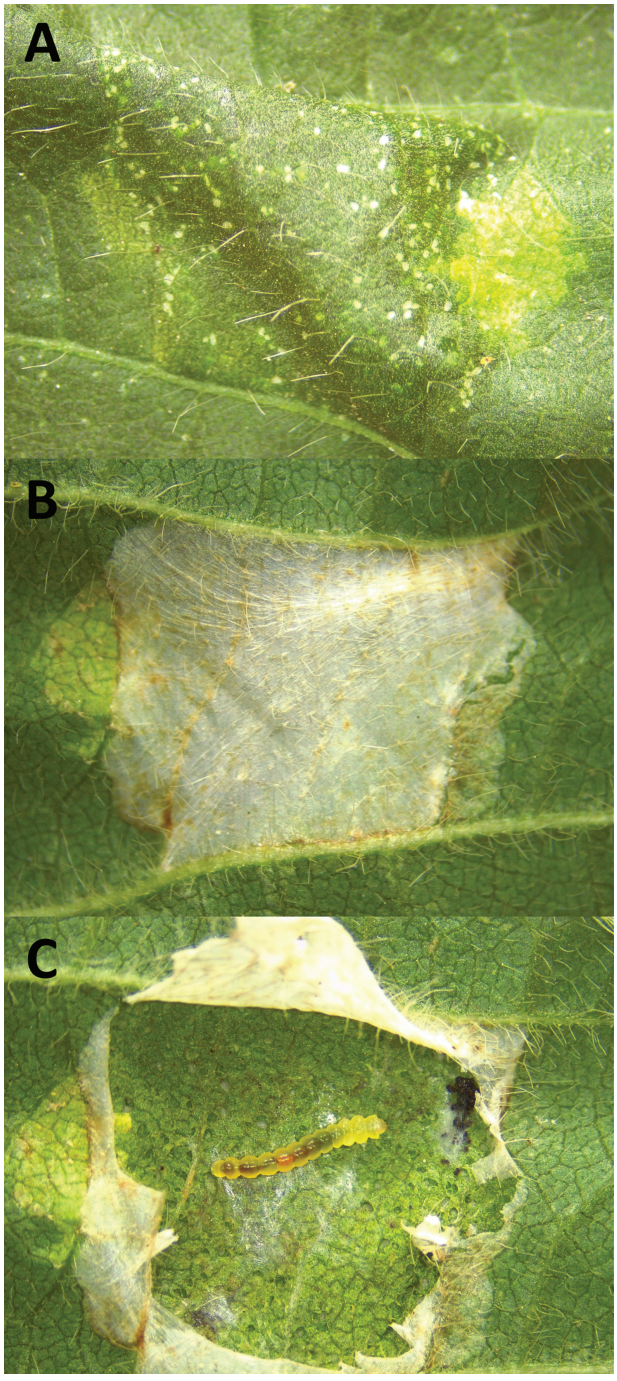


Fig. 4. Injury on the adaxial (upper) side (A), and intact (B) and open leaf mine showing a *Macrosaccus morrisella* larva (C) on the abaxial (lower) side of a soybean leaflet. Image credit: R. L. Koch.

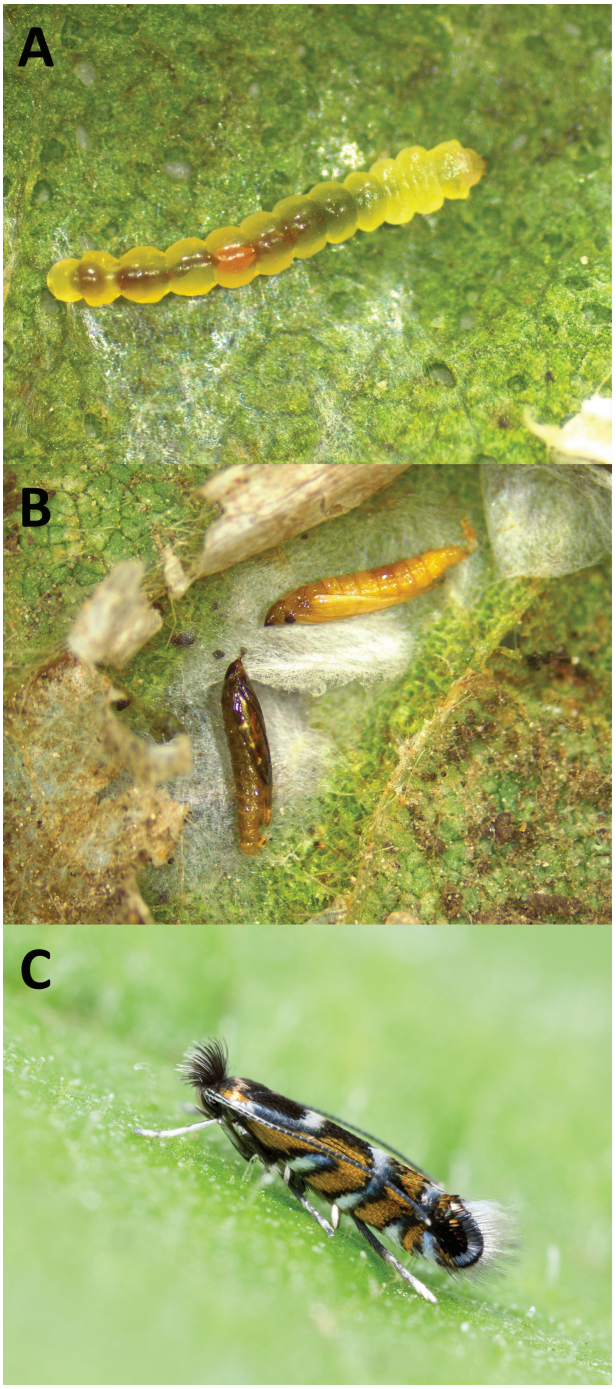


Fig. 5. Larva (A), pupae (B), and adult (C) of *Macrosaccus morrisella* from soybean. Image credits: R. L. Koch for (A) and (B); J. Moisan-De Serres for (C).

Downloaded from https://academic.oup.com/jipm/article/12/1/44/6438101 by guest on 05 April 2022

(i.e., soybean leaf miner, *Odontota horni* Smith, locust leaf miner, *O. dorsalis* (Thunberg), and *Sumitrosis rosea* (Weber) (Coleoptera: Chrysomelidae)) are known to feed on soybean in North America (Buntin and Pedigo 1982, McPherson and Ravlin 1983); however, these leaf miners are rarely of economic significance to soybean production (Buntin 1994). In contrast, the leaf-mining insect *Aproaerema modicella* (Deventer) (Lepidoptera: Gelechiidae) is a significant pest of soybean production in Africa, Asia, and Australia (Gaur and Mogalapu 2018, Buthelezi et al. 2021). Furthermore, within the family Gracillariidae (Lepidoptera), *Caloptilia soyella* (Deventer) in China and Japan, *Phodoryctis caerulea* (Meyrick) in Indonesia, *Porphyrosela homotropha* Vári in Ethiopia (De Prins and De Prins 2006–2021), and *Porphyrosela* (*Lithocolletis*) *aglaozona* (Meyrick) (Bailey 2007) have been reported to feed on soybean. Because the leaf-mining niche in North American soybean appears underutilized (Buntin 1994), the potential adaptation to soybean by a leaf-mining lepidopteran like *M. morrisella* requires further examination (e.g., Kogan 1981).

Relatively little is known about the ecology and management of *M. morrisella*. However, this species is known to be attacked by two parasitoids, *Pediobius albipes* (Provancher) and *Sympiesis marylandensis* Girault (Eulophidae: Hymenoptera) (Peck 1985, Maier 1988, De Prins and De Prins 2006–2021). Knowledge of the impacts and management of the chrysomelid leaf miners of soybean in North America (Buntin 1994) and lepidopteran leaf miners of soybean in Africa, Asia, and Australia (Gaur and Mogalapu 2018, Buthelezi et al. 2021) could provide a foundation for development of management plans for *M. morrisella*, if infestations persist and intensify. Future research should assess its geographic extent of infestation of soybean, impacts to soybean, and ecology in the agricultural landscape.

Acknowledgments

We thank Gunnar Morris and Siwen Ding (University of Minnesota) for help collecting specimens in Minnesota, and Jean-François Landry (Agriculture and Agri-Food Canada) for providing help in the initial identification of *M. morrisella*. The data collected from Minnesota stemmed from observations made while sampling fields for a project funded by the Minnesota Soybean Research and Promotion Council and the Minnesota Invasive Terrestrial Plants and Pests Center through the Minnesota Environment and Natural Resources Trust Fund.

References Cited

Bailey, P. T. 2007. Pests of field crops and pastures: identification and control. CSIRO Publishing, Collingwood, VIC, Australia.

Braun, A. F. 1908. Revision of the North American genus *Lithocolletis* Hübner. Trans. Am. Entomol. Soc. 34: 269–357.

Buntin, G. D. 1994. Soybean leafminer, pp. 79–80. *In* L. G. Higley and D. J. Boethel (eds.), The handbook of soybean insect pests. The Entomological Society of America, Lanham, MD.

Buntin, G. D., and L. P. Pedigo. 1982. Foliage consumption and damage potential of *Odontota horni* and *Baliosus nervosus* (Coleoptera: Chrysomelidae) on soybean. J. Econ. Entomol. 75: 1034–1037.

Buthelezi, M. N., D. E. Conlong, and G. E. Zharare. 2021. Review of the biosystematics and bio-ecology of the groundnut/soya bean leaf miner species (Lepidoptera: Gelechiidae). Austral Entomol. 60: 276–284.

Davis, D. R., and J. De Prins. 2011. Systematics and biology of the new genus *Macrosaccus* with descriptions of two new species (Lepidoptera, Gracillariidae). ZooKeys 98: 29–82.

De Prins, J., and W. De Prins. 2006-2021. Global taxonomic database of Gracillariidae (Lepidoptera). World Wide Web electronic publication. http://www.gracillariidae.net [accessed 15 August 2021].

Gaur, N., and S. Mogalapu. 2018. Pests of soybean, pp. 137–162. *In* Omkar (ed.), Pests and their management. Springer Nature Singapore Pte Ltd., Singapore.

Hodgson, E., A. Sisson, D. Mueller, L. Jesse, E. Saalau-Rojas, and A. Duster. 2012. Field crops insects. Iowa State University Extension and Outreach, Ames.

Kogan, M. 1981. Dynamics of insect adaptations to soybean: impact of integrated pest management. Environ. Entomol. 10: 363–371.

Kogan, M., and D. D. Kogan. 1979. *Odontota horni*, a hispine leaf miner adapted to soybean feeding in Illinois. Ann. Entomol. Soc. Am. 72: 456–461.

Kogan, M., and S. G. Turnipseed. 1987. Ecology and management of soybean arthropods. Annu. Rev. Entomol. 32: 507–538.

Liu, W. H., X. H. Dai, and J. S. Xu. 2015. Influences of leaf-mining insects on their host plants: a review. Collect. Bot. 34: e005.

Maier, C. T. 1988. Gracillariid hosts of *Sympiesis marylandensis* (Hymenoptera: Eulophidae) in New England. Ann. Entomol. Soc. Am. 81: 728–732.

McPherson, R. M., and R. W. Ravlin. 1983. Locust leaf miner development on soybean in Virginia. J. Ga. Entomol. Soc. 18: 58–60.

Peck, O. 1985. The taxonomy of the Nearctic species of *Pediobius* (Hymenoptera: Eulophidae), especially Canadian and Alaskan forms. Can. Entomol. 117: 647–704.

Downloaded from https://academic.oup.com/jipm/article/12/1/44/6438101 by guest on 05 April 2022

Journal of Economic Entomology, 113(2), 2020, 779–786
doi: 10.1093/jee/toz306
Advance Access Publication Date: 29 November 2019
Research

Detection of Stress Induced by Soybean Aphid (Hemiptera: Aphididae) Using Multispectral Imagery from Unmanned Aerial Vehicles

Zachary P. D. Marston,[1,5] Theresa M. Cira,[1] Erin W. Hodgson,[2] Joseph F. Knight,[3] Ian V. MacRae,[4] and Robert L. Koch[1]

[1]Department of Entomology, University of Minnesota, 1980 Folwell Ave., Saint Paul, MN 55108, [2]Department of Entomology, Iowa State University, 2213 Pammel Dr., Ames, IA 50011, [3]Department of Forest Resources, University of Minnesota, 1530 Cleveland Ave N, Saint Paul, MN 55108, [4]Department of Entomology, University of Minnesota, Northwest Research and Outreach Center, 2900 University Ave., Crookston, MN 56716, and [5]Corresponding author, e-mail: mars0368@umn.edu

Subject Editor: Silvia Rondon

Received 30 May 2019; Editorial decision 23 October 2019

Abstract

Soybean aphid, *Aphis glycines* Matsumura (Hemiptera: Aphididae), is a common pest of soybean, *Glycine max* (L.) Merrill (Fabales: Fabaceae), in North America requiring frequent scouting as part of an integrated pest management plan. Current scouting methods are time consuming and provide incomplete coverage of soybean. Unmanned aerial vehicles (UAVs) are capable of collecting high-resolution imagery that offer more detailed coverage in agricultural fields than traditional scouting methods. Recently, it was documented that changes to the spectral reflectance of soybean canopies caused by aphid-induced stress could be detected from ground-based sensors; however, it remained unknown whether these changes could also be detected from UAV-based sensors. Small-plot trials were conducted in 2017 and 2018 where cages were used to manipulate aphid populations. Additional open-field trials were conducted in 2018 where insecticides were used to create a gradient of aphid pressure. Whole-plant soybean aphid densities were recorded along with UAV-based multispectral imagery. Simple linear regressions were used to determine whether UAV-based multispectral reflectance was associated with aphid populations. Our findings indicate that near-infrared reflectance decreased with increasing soybean aphid populations in caged trials when cumulative aphid days surpassed the economic injury level, and in open-field trials when soybean aphid populations were above the economic threshold. These findings provide the first documentation of soybean aphid-induced stress being detected from UAV-based multispectral imagery and advance the use of UAVs for remote scouting of soybean aphid and other field crop pests.

Key words: remote sensing, reflectance, unmanned aerial vehicle, multispectral, crop scouting

Soybean, *Glycine max* (L.) Merrill (Fabales: Fabaceae), is currently the most widely grown field crop in the United States, with 36.5 million hectares producing over 120 million metric tons of grain in 2017 (USDA-NASS 2018). The United States continues to lead the world in soybean production and the north-central United States accounts for over 75% of the nation's production (USDA-NASS 2018). Over the past two decades, however, there has been a dramatic change in soybean production in the north-central United States, due to an invasive species, the soybean aphid, *Aphis glycines* Matsumura (Hemiptera: Aphididae). Prior to the invasion by soybean aphid in 2000, there were few insects reaching levels causing economic injury to soybean and fewer than 0.1% of soybean fields in the north-central United States were sprayed with insecticide (Ragsdale et al. 2011, Hodgson et al. 2012). However, by 2006, there was more than a 130-fold increase of insecticide applications to soybean in the region (Ragsdale et al. 2011, Hodgson et al. 2012). This increase was largely due to the soybean aphid's ability to rapidly reproduce and reduce soybean yields (Beckendorf et al. 2008, Ragsdale et al. 2011).

Soybean aphid damages soybean by inserting piercing-sucking mouthparts into the phloem of the plants to extract photosynthate (Tilmon et al. 2011). This feeding can decrease yield through plant stunting, decreased leaf area, reduced pod and seed number, decreased seed weight and oil concentrations, and even plant death (Hill et al. 2004, Mensah et al. 2005, Ragsdale et al. 2007, Beckendorf et al. 2008). Furthermore, soybean aphids excrete honeydew on leaf surfaces while feeding, which can promote the growth of sooty mold (Tilmon et al. 2011) and further reduce yields by inhibiting photosynthesis (Hill et al. 2010). Soybean aphid infestations left untreated

© The Author(s) 2019. Published by Oxford University Press on behalf of Entomological Society of America. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

have been documented to reduce soybean yields by more than 40% (Ragsdale et al. 2007). Soybean aphid is considered the most economically important insect pest of soybean in the north-central United States (Ragsdale et al. 2007, Hurley and Mitchell 2017) and considerable research has been performed to develop cost-effective management strategies for soybean aphid (Hodgson et al. 2004, 2007, 2012; Ragsdale et al. 2007; Ragsdale et al. 2011).

Current management recommendations involve routine scouting of soybean fields to monitor soybean aphid populations (Hodgson et al. 2004). Routine scouting is needed because widespread outbreaks of soybean aphid are erratic, and the timing of colonization can fluctuate (Hodgson et al. 2012). When aphid populations reach an economic threshold of 250 aphids per plant, chemical control is recommended to prevent aphids from reaching the economic injury level of ~674 aphids per plant (Ragsdale et al. 2007, Koch et al. 2016). While many farmers follow these recommendations, some farmers are reluctant to adopt these practices because the scouting process can be arduous and time consuming (Bueno et al. 2011). A more efficient binomial sampling plan was established in 2004, but further testing of this method found that correct management decisions were attained only 79% of the time, the other 21% of the time decisions to apply insecticide were made before aphid populations reached the economic threshold (Hodgson et al. 2004, 2007; Ragsdale et al. 2011). Furthermore, current scouting practices do not provide complete coverage of a field creating the potential to miss areas heavily infested with soybean aphid. The difficulty associated with counting aphids within a large field of densely planted soybean and the lack of coverage provided by current scouting methods has led some farmers to use prophylactic applications of insecticides rather than base chemical treatment on estimates of aphid populations in the field (Olson et al. 2008). This prophylactic method of control can increase production costs and risk for development of insecticide resistance, and be detrimental to nontarget organisms and water quality (Song and Swinton 2009, Bueno et al. 2011, Koch et al. 2016). Incorporating remote sensing offers the potential to improve management of soybean aphid by decreasing the effort and cost of scouting while increasing field coverage, which may increase adoption of management practices based on estimates of in-field pest abundance and thereby decrease unnecessary pesticide applications.

Remote sensing has been used to provide valuable insight into crop management for over 50 yr (Hatfield et al. 2008). Remote sensing for agriculture passively obtains information about within field variability in the health of a crop by relating electromagnetic, or spectral, reflectance to plant biological components and physiology, such as foliar pigment content, cellular structure, water content, as well as canopy coverage and architecture (Pinter et al. 2003). One of the most commonly used spectral indices for remote sensing in agriculture is the normalized difference vegetation index (NDVI). The NDVI is particularly helpful because it combines red reflectance with near-infrared reflectance (NIR). Red reflectance is an indicator of chlorophyll content of the plant canopy and active photosynthesis; and NIR provides information about the cellular structure and intracellular air spaces within leaves, overall canopy coverage, and above ground biomass (Hatfield et al. 2008). When these wavelengths are combined in an index, like NDVI, it provides a measure of overall plant health and has frequently been correlated with crop yield (Ma et al. 2001).

In soybean, different forms of crop stress, such as nutrient deficiencies (Milton et al. 1991, Bai et al. 2018), soybean cyst nematode, *Heterodera glycines* Ichinohe (Tylenchida: Heteroderidae) (Nutter et al. 2002, Bajwa et al. 2017), soybean sudden death syndrome, *Fusarium virguliforme* O'Donnell & Aoki (Hypocreales: Nectriaceae) (Bajwa et al. 2017, Hatton et al. 2017, Hatton 2018, Herrmann et al. 2018, Menke 2018), weed pressure (Koger et al. 2003, Chang et al. 2004, Henry et al. 2004, Gray et al. 2009), and drought stress (Pinter et al. 2003, Jackson et al. 2004, O'Shaughnessy et al. 2011) affect the spectral reflectance of the plants, which can be detected through remote sensing. The biophysical principles behind remote sensing have remained generally consistent over the past 50 yr; however, the technology used to record this information has not. Previously, remote sensing in agriculture used either ground-based systems, which are often restricted by small mapping swaths and limited transportability, or satellites and piloted aircraft which have been expensive, low-resolution, and limited by atmospheric conditions and orbital periods. (Lelong et al. 2008, Zhang and Kovacs 2012). More recently, however, unmanned aerial vehicles (UAVs) equipped with ultra-high spatial resolution multispectral sensors have become increasingly available to consumers and promise low-cost near real-time image acquisition for use in agricultural applications (Nebiker et al. 2008).

Recently, it was reported ground-based remote sensing of soybean was capable of detecting stress to plants caused by soybean aphid (Alves et al. 2015, 2019). However, it remains unknown if the stress caused by soybean aphid to soybean can be detected from UAV-based sensors. Therefore, the goal of this research was to determine whether soybean aphid-induced stress can be detected from UAV-based multispectral sensors. The results of this research will help to identify how UAV-based remote sensing can be incorporated into current crop scouting practices to improve scouting efficiency and adoption of IPM strategies.

Materials and Methods

Caged-Plot Experiments

Research trials were conducted in 2017 and 2018 at both the University of Minnesota Outreach, Research, and Education Park in Rosemount, MN (44° 44′ 1.2804″ N, 93° 5′ 4.2288″ W) and at the Iowa State University Northern Research Farm in Kanawha, IA (42° 55′ 51.3408″ N, 93° 47′ 32.4168″ W) (Table 1). Twenty-four plots were established at each location in both years. Each plot was created by planting soybean in two 2.5-m-long rows with 76.2-cm row spacing at a seeding rate of 345,000 seeds per hectare. Of these 24 plots, a subset of 12 plots were selected at each location that were not inoculated with pathogens at the time of planting. In Rosemount, soybean variety MN1410R2F5-121 was planted on 8 May 2017 and 10 May 2018. In Kanawha, soybean variety Syngenta S24-K2 was planted on 24 April 2017 and 18 May 2018. No fertilizer was applied at either location and weeds were managed by applying pre-emergent herbicide followed by hand weeding after growth stage VE (Fehr and Caviness 1977). At growth stage V3, plants were carefully inspected for soybean aphids, any aphids found were recorded and removed from plants either by hand or with an insecticide treatment, then PVC frames (1.5 × 2.5 m) were placed over each plot and covered with NO-SEE-UM mesh cages (Quest Outfitters, Sarasota, FL) to prevent aphid colonization. In Rosemount in both 2017 and 2018, soybean aphids infested the plots before cages were placed in the field. To remove aphids in 2017, all plots were sprayed with λ-cyhalothrin (116 ml product per ha, Warrior II with Zeon Technology, Syngenta, Greensboro, NC) on 15 June and again with a formulated mixture of λ-cyhalothrin and thiamethoxam (328 ml product per ha, Endigo ZC, Syngenta, Greensboro, NC) on 27 June. To remove aphids in 2018, all plots were sprayed with a formulated mixture of λ-cyhalothrin and thiamethoxam (328 ml product per ha, Endigo ZC, Syngenta) on

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

Table 1. Description of experimental details used to determine whether soybean aphid-induced stress could be detected with UAV-based remote sensing

Experiments	Location	Treatments	Sample Size	Year	Sample Dates
Caged-Plot Trials	Rosemount, MN	Aphid infested vs. uninfested	$n = 12$	2017	8 and 22 Aug.
	Rosemount, MN	Aphid infested vs. uninfested	$n = 11$	2018	8 and 13 Aug.
	Kanawha, IA	Aphid infested vs. uninfested	$n = 12$	2018	10 and 22 Aug.
Open-Field Trials	Rosemount, MN Plot 1	Treated with insecticide vs. Untreated	$n = 16$	2018	15 and 22 Aug.
	Rosemount, MN Plot 2	Treated with insecticide vs. Untreated	$n = 16$	2018	15 and 22 Aug.

1 June. Cages were inspected weekly to ensure plants stayed free of aphids until the planned infestation.

The 12 plots at each location were divided into two treatments, aphid-free and aphid-infested, in a completely randomized design. In both 2017 and 2018, the aphid-infested treatment was infested with soybean aphids at growth stage R3. In 2017, the infestation procedure consisted of infesting each plot with 200 mixed-stage aphids (i.e., nymphs + wingless adults) on 17 July in both locations. Due to poor aphid establishment in infested cages in Kanawha in 2017, each infested plot in 2018 received 400 mixed-stage aphids on 16 July in Rosemount and 18 July in Kanawha.

These infestations were accomplished by pinning leaf cuttings, each with 25 mixed-stage soybean aphids, to the abaxial side of the uppermost fully expanded trifoliate of plants within the caged plots. These infested leaf cuttings were evenly spaced throughout the plot. In 2017, each infested plot received 8 leaf cuttings and in 2018 each infested plot received 16 leaf cuttings. Soybean aphids were taken from a laboratory colony to infest cages at both locations in 2017 and in Kanawha in 2018. Several caged plots in Rosemount in 2018 were blown open during a storm on 17 June, and consequently three plots were naturally infested with soybean aphid before the intentional infestation. While two of these plots had fewer than 60 aphids per plant on 9 July, one plot exceeded the economic threshold (i.e., 250 aphids per plant) and was removed from the experiment, resulting in 6 aphid-infested plots and 5 aphid-free plots in Rosemount in 2018. Aphids from the plot that was removed from the experiment were used to artificially infest the remaining aphid-infested plots at that location in 2018.

Before and after infestation, aphid densities for each plot were estimated weekly from 6 June to 23 August 2017 in Rosemount and 13 June to 23 August 2017 in Kanawha. In 2018, aphid densities for each plot were estimated weekly from 25 June to 13 August 2018 in Rosemount and 5 July to 22 August 2018 in Kanawha. In order to assess the aphid populations in each plot, the fine-mesh cages were temporarily removed to allow counting and were replaced after the counts were recorded. The two rows of each plot were visually divided into four evenly spaced sections and one plant was randomly selected from each section of row, for a total of eight plants per plot. The number of aphids on each of these plants was assessed by nondestructive, whole-plant counts. Caution was taken to minimize the risk of inadvertent transfer of aphids to aphid-free plots.

Open-Field Experiments

Open-field experiments were conducted in 2018 at the University of Minnesota Outreach, Research, and Education Park in Rosemount, MN (Table 1). A commercial field was planted on 17 May 2018 with soybean variety Asgrow-AG1435 at a seeding rate of 368,000 per hectare and 76.2-cm row spacing. No fertilizer was applied, and weeds were managed by applying labeled rates of pre-emergent herbicide on 21 May 2018, and postemergent herbicide on 22 June 2018. Any weeds found within the sample area after postemergent herbicide application were removed by hand to ensure there were no weeds present during image acquisition. Two 0.4-hectare plots were established by tilling a 1.5-m alley around uniform areas within the field on 13 August. When soybean aphid densities reached an average of 250 aphids per plant, a strip measuring 15.9-m by 63.6-m in each of these 0.4-hectare plots was sprayed with a formulated mixture of λ-cyhalothrin and thiamethoxam (328 ml product per ha, Endigo, Syngenta) to create different levels of aphids within each plot.

A 16-cell grid was created within each 0.4-hectare plot with each cell measuring 15.9 × 15.9 m. Within each cell, aphid densities were estimated from 8 to 10 plants selected in a stratified random method on 15 and 22 August 2018. The number of aphids on each plant were assessed by destructive, whole-plant counts. The global positioning system (GPS) coordinates of each plant counted for aphids was recorded with a handheld GPS unit (GPSmap 62s, Garmin Ltd, Olathe, KS) with a GPS signal accuracy of less than 3 m.

Spectral Reflectance Measurements

Canopy spectral reflectance measurements were recorded weekly from caged plots at both locations between 8 and 22 August in 2017 and 2018. Canopy spectral measurements were recorded from 0.4-hectare open plots on 15 and 22 August 2018. Imagery was recorded with a nadir-facing multispectral camera (Quad Multispectral Sensor, Sentera Inc., Minneapolis, MN) mounted on an UAV (Solo, 3DR, Berkeley, CA). The multispectral camera was attached to the UAV via a vibration plate to minimize distortion in the imagery caused by UAV movement. The multispectral camera was equipped with a standard red, green, and blue color light sensor (1.2MP CMOS RGB), and customized to include a narrowband red sensor (1.2MP CMOS Mono 675 ± 12.5 nm), a narrowband near-infrared sensor (1.2MP CMOS Mono 775 ± 12.5 nm), and a broadband near-infrared sensor (1.2MP CMOS Mono 825 ± 100 nm). These customized bands were selected based on previous ground-based remote sensing work for soybean aphid, and preliminary analysis of band simulation and optimization for soybean aphid (Alves

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

et al. 2015, 2019). All sensors had global shutter, a lens focal length of 6.05 mm, and were set to autoexposure during image capture. Flights were automated using the open-source software, Mission Planner (available and maintained ardupilot.org), and performed in a cross-grid pattern with 80% forward overlap and 70% sidelap, at an altitude of 50 m in 2017 (3.2-cm Pixel GSD), and at 40 m in 2018 (2.5-cm Pixel GSD) on both the caged and open plots following recommendations included in the Sentera documentation packet (Quad Multispectral Sensor Documentation Packet, Sentera Inc.).

Imagery was recorded within 2.5 h of solar noon to minimize the effect of solar angle and shadowing on the crop canopy. Furthermore, all imagery was recorded after canopy closure to minimize the effect of bare ground and shadowing within the spectral measurements. All caged plots and open fields were also scouted prior to image acquisition to ensure there were no other confounding factors within the sampled areas such as disease, drought stress, nutrient deficiency, or other common stressors.

In an attempt to minimize atmospheric effects on the recorded imagery, flights were only flown when light conditions were uniform, such as cloudless days or at times when no visible clouds were moving between the sun and the crop canopy. To avoid potential damage to plants or handling effects, imagery was recorded immediately after the mesh cages were removed from the caged-plots and before performing aphid counts and all aphid counts in the open fields were recorded either the day before or the day after UAV-based spectral measurements were taken. Reference panels with known reflectance properties were placed in the field prior to each flight to aid in converting camera Digital Number (DN) values to relative reflectance through the empirical line method (Smith and Milton 1999).

Image Processing

Before images were processed, each image was visually inspected for quality ensuring no hotspots or banding was observed in the imagery. Imagery was then normalized for autoexposure, stitched into orthomosaics, and converted to relative reflectance. Prior to correction for autoexposure, images were cast from 8-bit depth to 16-bit depth in MATLAB Image Processing Toolbox (MathWorks Inc., Massachusetts) to avoid saturation during autoexposure correction. Images were then corrected for exposure time, digital gain, and analog gain using;

$$\text{Normalized DN} = \frac{\text{DN}}{(\text{Exposure Time} * \text{Digital Gain} * 2^{\text{Analog Gain}})}$$

(Quad Multispectral Sensor Documentation Packet, Sentera Inc.).

The images were then stitched using Pix4Dmapper (Pix4D SA, Lausanne, Switzerland) using the camera model parameters recommended within the Sentera documentation packet, and the Sentera template for processing options (Quad Multispectral Sensor Documentation Packet, Sentera Inc.). Coefficients from a linear fit between observed DN and known reflectance of the reference panels were used with the empirical line method for reflectance calibration of the Pix4D generated orthomosaics (Smith and Milton 1999).

Analyses

For the caged-plot experiment, plot reflectance values were extracted from the stitched orthomosaics using ImageJ (Version 1.52k, National Institutes of Health, Bethesda, MD). The area of reflectance extracted from each plot was centered over the middle of the two rows and was the same size for each plot, ensuring not to include pixels of the cage frame around each plot. The aphid counts were converted to cumulative aphid days (CAD), which is a measure of the cumulative aphid stress caused to the plants over time. CAD was calculated following the methods proposed by Ruppel (1983) and adapted for aphids by Hanafi et al. (1989).

We selected narrowband NIR (775 ± 12.5 nm), narrowband red (675 ± 12.5 nm), and the vegetation index NDVI for analyses as these have previously been identified as affected by soybean aphid-induced stress in ground samples (Alves et al. 2015). NDVI was calculated as NDVI=[(NIR–red)/(NIR+red)] (Rouse Jr. et al. 1973). Simple linear regression followed by analysis of variance (ANOVA) (R Development Core Team 2013) was used to determine whether CAD had an effect ($\alpha = 0.05$) on aerially measured spectral reflectance (Alves et al. 2015, 2019). Visual assessment of residual plots indicated log10 transformation of CAD was required to meet statistical assumptions for linear regression analysis.

For the open-field experiments, the cells in each 0.4-hectare plot consisted of 4 cells over the insecticide-treated area, and 12 cells over the untreated area. Only the narrowband NIR (775 ± 12.5 nm) sensor was used for analysis of the open-field experiments, because it was found to be the optimal band for detecting aphid-induced stress in previous research and was unlikely to be affected by insecticide treatments (Alves et al. 2017, 2019). Reflectance data within each cell were equally cropped on all sides to avoid any edge effects around the borders of the 0.4-hectare plot and between the insecticide-treated area and untreated area. Reflectance values were extracted in R (raster package, Hijmans 2017, R Development Core Team 2013). Only aphid counts within the cells after cropping were averaged to obtain a single average value of aphids per plant for each cell. Simple linear regression followed by ANOVA (R Development Core Team 2013) was used to determine whether average number of aphids per plant had an effect ($\alpha = 0.05$) on spectral reflectance values. CAD was not used for this experiment because aphid counts were taken on too few sample dates. Residual plots were visually inspected to ensure assumptions were met for linear regression analysis.

Results

Caged Plots

On 8 and 22 August 2017 in Rosemount, 8 and 13 August 2018 in Rosemount and 22 August 2018 in Kanawha, aphids surpassed the economic injury level (>6,500 CAD) within the infested cages (Ragsdale et al. 2007). However, on 10 August 2018 at the time spectral measurements were taken in Kanawha, aphids had not reached the economic injury level. Across sites and years, red reflectance was not associated with CAD (Table 2), except for 13 August 2018 in Rosemount where red reflectance increased with increasing CAD (Table 2). On all dates where soybean aphid populations reached the economic injury level, NIR reflectance decreased with increasing CAD (Table 2). On 10 August 2018 in Kanawha, which was before soybean aphid reached the economic injury level, NIR reflectance was not associated with CAD (Table 2). When CAD reached the economic injury level there were also decreases in NDVI values with increasing CAD (Table 2), except in Kanawha on 22 August 2018 where there was a marginal decrease in NDVI values with increasing CAD (Table 2). On 10 August 2018 in Kanawha, there was no association between NDVI and CAD.

Open-Field Experiments

Plot 1 had 347 ± 69 (mean ± SEM) aphids per plant in the treated areas and 861 ± 42 aphids per plant in the untreated areas on 15 August 2018. On this date, NIR reflectance decreased by 0.5%

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

Table 2. Model estimates from simple linear regressions for the effect of log10-transformed CAD on soybean canopy red reflectance, NIR, and NDVI recorded from an UAV from caged-plots in Rosemount, MN, 2017, and Rosemount, MN and Kanawha, IA, 2018

Wavelength/Index	Site	Date	Intercept	Slope	F-value$_{df}$	P-value	R^2
Red (675 ± 12.5 nm)	Rosemount, MN	8 Aug. 2017	2.927	0.013	$0.008_{(1,10)}$	0.784	0.008
	Rosemount, MN	22 Aug. 2017	1.755	0.053	$2.755_{(1,10)}$	0.128	0.216
	Rosemount, MN	8 Aug. 2018	2.068	0.089	$2.010_{(1,9)}$	0.190	0.183
	Rosemount, MN	13 Aug. 2018	0.996	0.175	$6.812_{(1,9)}$	0.028*	0.431
	Kanawha, IA	10 Aug. 2018	1.767	−0.010	$0.217_{(1,10)}$	0.652	0.021
	Kanawha, IA	22 Aug. 2018	1.663	7.83×10^{-5}	$0.000_{(1,10)}$	0.997	0.000
NIR (775 ± 12.5 nm)	Rosemount, MN	8 Aug. 2017	73.588	−3.144	$10.900_{(1,10)}$	0.008*	0.522
	Rosemount, MN	22 Aug. 2017	51.955	−4.331	$5.791_{(1,10)}$	0.037*	0.367
	Rosemount, MN	8 Aug. 2018	71.727	−7.680	$21.800_{(1,9)}$	0.001*	0.708
	Rosemount, MN	13 Aug. 2018	89.487	−11.504	$45.250_{(1,9)}$	<0.001*	0.834
	Kanawha, IA	10 Aug. 2018	49.253	−0.234	$0.186_{(1,10)}$	0.676	0.018
	Kanawha, IA	22 Aug. 2018	39.918	−0.980	$6.496_{(1,10)}$	0.029*	0.394
NDVI $(\frac{NIR - Red}{NIR + Red})$	Rosemount, MN	8 Aug. 2017	0.925	-5.30×10^{-3}	$5.400_{(1,10)}$	0.042*	0.351
	Rosemount, MN	22 Aug. 2017	0.968	−0.023	$7.373_{(1,10)}$	0.022*	0.424
	Rosemount, MN	8 Aug. 2018	0.982	−0.026	$17.550_{(1,9)}$	0.002*	0.661
	Rosemount, MN	13 Aug. 2018	1.045	−0.033	$26.770_{(1,9)}$	<0.001*	0.748
	Kanawha, IA	10 Aug. 2018	0.930	1.58×10^{-5}	$0.000_{(1,10)}$	0.92	0.000
	Kanawha, IA	22 Aug. 2018	0.920	−0.002	$4.389_{(1,10)}$	0.063	0.305

*Indicates a significant effect of CAD on canopy reflectance (α = 0.05). Otherwise, CAD had no significant effect on reflectance.

Table 3. Model estimates from simple linear regressions for the effect of soybean aphid density (aphids per plant) on soybean canopy NIR recorded by UAV from open fields in Rosemount, MN, 2018

Wavelength	Site	Date	Intercept	Slope	F-value$_{df}$	P-value	R^2
NIR	Plot 1	15 Aug. 2018	58.286	−0.005	$11.929_{(1,14)}$	0.004*	0.460
775 ± 12.5 nm	Plot 1	22 Aug. 2018	55.562	−0.011	$78.900_{(1,14)}$	<0.001*	0.849
	Plot 2	15 Aug. 2018	54.570	−0.003	$3.723_{(1,14)}$	0.074	0.210
	Plot 2	22 Aug. 2018	60.945	−0.010	$10.668_{(1,14)}$	0.006*	0.432

*Indicates a significant effect of soybean aphid density on soybean canopy reflectance (α = 0.05). Otherwise, soybean aphid density had no significant effect on reflectance.

per 100 aphids per plant (Table 3). On 22 August 2018 in plot 1, the treated area had 81 ± 27 aphids per plant on average, while the untreated area had 798 ± 31 aphids per plant. On this date, NIR reflectance decreased by 1.0% per 100 aphids per plant (Table 3).

In plot 2, the treated portion of the plot had 226 ± 53 aphids per plant and the untreated portion had 665 ± 82 aphids per plant on average on 15 August 2018. On this date, NIR reflectance decreased marginally by 0.3% per 100 aphids per plant (Table 3). On 22 August 2018 in plot 2, the treated portion of the plot had 119 ± 24 aphids per plant, and the untreated portion of the plot had 400 ± 41 aphids per plant. On this date, NIR reflectance decreased by 1.1% per 100 aphids per plant (Table 3).

Discussion

The use of UAVs for agriculture has immense potential to improve decision making for crop management by providing high temporal and spatial resolution information on soils, crop nutrients, pests, moisture, and yield (Canis 2015). Previous research showed the potential for use of remote sensing for soybean aphid through ground-based research (Alves et al. 2015, 2019). The findings from the present experiments provide the first documentation of UAV-based remote detection of soybean aphid-induced stress in soybean. Further work is needed to determine whether ground-based remote sensing for other pests and cropping systems may also extend to UAV-based approaches.

These findings advance the use of UAVs and remote sensing as actionable tools for scouting soybean aphid. On caged soybean plants, soybean aphid caused significant decreases in NIR reflectance and NDVI, but there were generally no changes to red reflectance, except for in the caged-plot experiment on 13 August 2018. These results, showing a decrease in NIR reflectance and NDVI caused by soybean aphid-induced stress, agree with previous ground-based remote sensing work on soybean aphid (Alves et al. 2015, 2019) and are consistent with remote sensing findings of other species of aphids and other hemipterans in grain crops (Mirik et al. 2007; Elliott et al. 2007, 2009, 2015; Prabhakar et al. 2011, 2013). When significant relationships were detected between CAD and reflectance, aphid populations were relatively high (i.e., above EIL). Because significant associations in the caged experiment were only observed when aphid populations exceeded the EIL, using linear regressions of spectral data for detecting soybean aphid-induced stress may not identify stress early enough to make actionable decisions and prevent economic injury. Further work is required to determine whether actionable decisions can be made from this research, specifically, it needs to be determined if these measured changes in reflectance can be used to classify aphid pressure as above or below treatment thresholds and differentiate aphid-induced stress from other stressors. Currently, these findings suggest that remote sensing may aid in

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

conventional scouting by directing the field scout to areas with decreased NIR reflectance.

The use of insecticides or cages to manipulate pest populations in remote sensing experiments may incorporate potential confounding factors by affecting the relevance of the results to production conditions. In a recent experiment, Alves et al. (2017) found that certain insecticides can have an effect on leaf-level reflectance in the visible portion of the spectrum, but they did not find an effect of insecticides on NIR reflectance (Alves et al. 2017); therefore, insecticide impacts on the NIR spectral measures in the present studies were assumed to be minimal. Furthermore, the pattern of aphid-induced change in NIR reflectance found in open plots also held for caged plots, which suggests the measured effects on NIR reflectance were the result of soybean aphid-induced stress and are robust to these experimental manipulations.

The increase in red reflectance seen in caged plots on 13 August 2018 contrasts previous findings for ground-based remote sensing of soybean aphid, which showed either no change or a decrease in red reflectance for leaf-level measurements (Alves et al. 2015). In soybean and other plants, an increase in red reflectance is commonly associated with reductions in chlorophyll (Chappelle et al. 1992, Gitelson et al. 2003). Soybean aphid feeding may cause a reduction in chlorophyll content of soybean (Diaz-Montano et al. 2007), which could explain the increase in red reflectance exhibited on 13 August 2018. However, chlorophyll measurements were not recorded in the present study.

Sooty mold is a common sign associated with soybean aphid infestations in the field (Koch et al. 2016). There were a number of cages and spots within field plots in which sooty mold started to develop on the honey-dew-coated leaves. Previously it has been documented in greenhouse experiments that citrus leaves covered in sooty mold showed an increase in red reflectance (Summy and Little 2008). The presence of sooty mold is another potential explanation for the significant increase in red reflectance seen on 13 August 2018, when many of the aphid-infested plots had sooty mold from aphid populations above economic injury level.

Our experiments showed that the decrease in NDVI values caused by soybean aphid-induced stress was largely driven by decreased NIR reflectance. This suggests NIR alone may be suitable for mapping soybean aphid-induced stress in soybean fields. Many other causes of soybean stress have been documented to affect NIR reflectance and the visible spectrum (Vigier et al. 2004, Gazala et al. 2013, Bajwa et al. 2017). Soybean aphid generally did not affect the red portion of the spectrum, so there may be potential for the differentiation of soybean aphid-induced stress from other forms of stress in soybean by using combinations of wavelengths as has been attempted in several cropping systems (Yuan et al. 2014, Bajwa et al. 2017). Diseases such as soybean cyst nematode and sudden death syndrome have been documented to affect NIR reflectance and NDVI values in similar ways to soybean aphid-induced stress, but these diseases also sometimes affect the visible spectrum in ways not observed for soybean aphid in this study (Bajwa et al. 2017, Hatton 2018, Herrmann et al. 2018, Menke 2018). The development of tools to differentiate between the spectral response caused by these diseases and soybean aphid-induced stress is ongoing.

More research is necessary in order to develop an actionable management system including remote sensing for soybean aphid. Early uses of remote sensing for scouting soybean aphid will likely rely on spectral data to identify areas with stressed plants followed by ground-truthing, because remote sensing data is often more meaningful when combined with ground data (Casady and Palm 2002; Liaghat and Balasundram 2010). However, there is potential for remote sensing to improve detection of soybean aphid-induced stress and differentiate it from other types of stress encountered within a field.

Acknowledgments

We thank all those who helped collect data in the field including: James Menger, Madelaine Bartz, Kendra Moran, Courtney Garrison Hickey, Alissa Geske, Pheylan Anderson, Claire Lotzer, Traci Eicholz, Narayan Bhagroo, Julia Stuartman, Daniela Pezinni, Obiratanea da Silva Queiroz, Rafael Carlesso Aita, Nadia Bueno, Arthur Vieira and Gregory VanNostrand, and Kimon Karelis for his invaluable help at the research site in Rosemount, MN. This project was supported by grants from the Agriculture and Food Research Initiative (competitive grant no. 2016-70006-25828) of the USDA National Institute of Food and Agriculture, the University of Minnesota MnDRIVE initiative, and the Minnesota Invasive Terrestrial Plants and Pests Center through the Minnesota Environment and Natural Resources Trust Fund.

References Cited

Alves, T. M., I. V. Macrae, and R. L. Koch. 2015. Soybean aphid (Hemiptera: Aphididae) affects soybean spectral reflectance. J. Econ. Entomol. 108: 2655–2664.

Alves, T. M., Z. P. Marston, I. V. MacRae, and R. L. Koch. 2017. Effects of foliar insecticides on leaf-level spectral reflectance of soybean. J. Econ. Entomol. 110: 2436–2442.

Alves, T. M., R. D. Moon, I. V. MacRae, and R. L. Koch. 2019. Optimizing band selection for spectral detection of *Aphis glycines* Matsumura in soybean. Pest Manag. Sci. 75: 942–949.

Bai, G., S. Jenkins, W. Yuan, G. L. Graef, and Y. Ge. 2018. Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. Front. Plant Sci. 9: 1002.

Bajwa, S., J. Rupe, J. Mason, S. G. Bajwa, J. C. Rupe, and J. Mason. 2017. Soybean disease monitoring with leaf reflectance. Remote Sens. 9: 127.

Beckendorf, E. A., M. A. Catangui, and W. E. Riedell. 2008. Soybean aphid feeding injury and soybean yield, yield components, and seed composition. Agron. J. 100: 237–246.

Bueno, A. de F., M. J. Batistela, R. C. O. de F. Bueno, J. de B. França Neto, M. A. N. Nishikawa, and A. Libério Filho. 2011. Effects of integrated pest management, biological control and prophylactic use of insecticides on the management and sustainability of soybean. Crop Prot. 30: 937–945.

Canis, B. 2015. Unmanned aircraft systems (UAS): commercial outlook for a new industry. Congressional Research Service (report no. R44192).

Casady, W. W., and H. L. Palm. 2002. Precision agriculture: remote sensing and ground truthing. University of Missouri Extension, Columbia, MO. https://extension.missouri.edu/p/EQ453 (accessed 30 April 2019).

Chang, J., S. A. Clay, D. E. Clay, and K. Dalsted. 2004. Detecting weed-free and weed-infested areas of a soybean field using near-infrared spectral data. Weed Sci. 52: 642–648.

Chappelle, E. W., M. S. Kim, and J. E. McMurtrey. 1992. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. Remote Sens. Environ. 39: 239–247.

Diaz-Montano, J., J. C. Reese, W. T. Schapaugh, and L. R. Campbell. 2007. Chlorophyll loss caused by soybean aphid (Hemiptera: Aphididae) feeding on soybean. J. Econ. Entomol. 100: 1657–1662.

Elliott, N., M. Mirik, Z. Yang, T. Dvorak, M. Rao, J. Michels, T. Walker, V. Catana, M. Phoofolo, K. Giles, and T. Royer. 2007. Airborne multispectral remote sensing of Russian wheat aphid injury to wheat. Southwest. Entomol. 32: 213–219.

Elliott, N., M. Mirik, Z. Yang, D. Jones, M. Phoofolo, V. Catana, K. Giles, and G. J. Michels. 2009. Airborne remote sensing to detect greenbug stress to wheat. Southwest. Entomol. 34: 205–211.

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

Elliott, N. C., G. F. Backoulou, M. J. Brewer, and K. L. Giles. 2015. NDVI to detect sugarcane aphid injury to grain sorghum. J. Econ. Entomol. 108: 1452–1455.

Fehr, W. R., and C. E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Cooperative Ext. Serv. Iowa State Univ., Ames, IA. pp. 11.

Gazala, I. F., R. N. Sahoo, R. Pandey, B. Mandal, V. K. Gupta, R. Singh, and P. Sinha. 2013. Spectral reflectance pattern in soybean for assessing yellow mosaic disease. Indian *J*. Virol. 24: 242–249.

Gitelson, A. A., Y. Gritz, and M. N. Merzlyak. 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160: 271–282.

Gray, C. J., D. R. Shaw, and L. M. Bruce. 2009. Utility of hyperspectral reflectance for differentiating soybean (*Glycine max*) and six weed species. Weed Technol. 23: 108–119.

Hanafi, A., E. B. Radcliffe, and D. W. Ragsdale. 1989. Spread and control of potato leafroll virus in Minnesota. J. Econ. Entomol. 82: 1201–1206.

Hatfield, J. L., A. A. Gitelson, J. S. Schepers, and C. L. Walthall. 2008. Application of spectral remote sensing for agronomic decisions. Agron. J. 100: S117-S131.

Hatton, N. M. 2018. Use of small unmanned aerial system for validation of sudden death syndrome in soybean through multispectral and thermal remote sensing. M.S. thesis, Kansas State University, Manhattan, KS.

Hatton, N. M., E. Menke, A. Sharda, D. Van der Merwe, and W. Schapaugh Jr. 2017. Comparison of aerial and ground remote sensing to quantify sudden death syndrome in soybeans. In 2017 ASABE Annual Invitational Meeting (p.1), 16–19 July 2017, Spokane, WA. American Society of Agricultural and Biological Engineers, St. Joseph, MI.

Henry, W. B., D. R. Shaw, K. R. Reddy, L. M. Bruce, and H. D. Tamhankar. 2004. Spectral reflectance curves to distinguish soybean from common cocklebur (*Xanthium strumarium*) and sicklepod (*Cassia obtusifolia*) grown with varying soil moisture. Weed Sci. 52: 788–796.

Herrmann, I., S. Vosberg, P. Ravindran, A. Singh, H.-X. Chang, M. Chilvers, S. Conley, P. Townsend, I. Herrmann, S. K. Vosberg, *et al*. 2018. Leaf and canopy level detection of *Fusarium virguliforme* (sudden death syndrome) in soybean. Remote Sens. 10: 426.

Hijmans, R. J. 2017. Raster: geographic data analysis and modeling. R package version 2.6–7. https://CRAN.R-project.org/package=raster/index.html

Hill, C. B., Y. Li, and G. L. Hartman. 2004. Resistance to the soybean aphid in soybean germplasm. Crop Sci. 44: 98–106.

Hill, C. B., L. Crull, T. K. Herman, D. J. Voegtlin, and G. L. Hartman. 2010. A new soybean aphid (Hemiptera: Aphididae) biotype identified. J. Econ. Entomol. 103: 509–515.

Hodgson, E. W., E. C. Burkness, W. D. Hutchison, and D. W. Ragsdale. 2004. Enumerative and binomial sequential sampling plans for soybean aphid (Homoptera: Aphididae) in soybean. J. Econ. Entomol. 97: 2127–2136.

Hodgson, E., B. McCornack, K. Koch, D. Ragsdale, K. Johnson, M. O'Neal, E. Cullen, H. Kraiss, C. DiFonzo, and L. Behnken. 2007. Field validation of speed scouting for soybean aphid. Crop Manag. 6: 1–8.

Hodgson, E. W., B. P. McCornack, K. Tilmon, and J. J. Knodel. 2012. Management recommendations for soybean aphid (Hemiptera: Aphididae) in the United States. J. Integr. Pest Manag. 3: E1–E10.

Hurley, T., and P. Mitchell. 2017. Value of neonicotinoid seed treatments to US soybean farmers. Pest Manag. Sci. 73: 102–112.

Jackson, T. J., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy, and E. R. Hunt. 2004. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sens. Environ. 92: 475–482.

Koch, R. L., B. D. Potter, P. A. Glogoza, E. W. Hodgson, C. H. Krupke, J. F. Tooker, C. D. DiFonzo, A. P. Michel, K. J. Tilmon, T. J. Prochaska, *et al*. 2016. Biology and economics of recommendations for insecticide-based management of soybean aphid. Plant Heal. Prog. 17: 265–269.

Koger, C. H., L. M. Bruce, D. R. Shaw, and K. N. Reddy. 2003. Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (*Ipomoea lacunosa)* in soybean (*Glycine max*). Remote Sens. Environ. 86: 108–119.

Lelong, C. C., P. Burger, G. Jubelin, B. Roux, S. Labbé, and F. Baret. 2008. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. Sensors (Basel). 8: 3557–3585.

Liaghat, S., and S. K. Balasundram. 2010. A review: the role of remote sensing in precision agriculture. Am. J. Agric. Biol. Sci. 5: 50–55.

Ma, B. L., L. M. Dwyer, C. Costa, E. R. Cober, and M. J. Morrison. 2001. Early prediction of soybean yield from canopy reflectance measurements. Agron. J. 93: 1227–1234.

Menke, E. J. 2018. Using spectral reflectance in soybean breeding: evaluating genotypes for soybean sudden death disease resistance and grain yield. M.S. thesis, Kansas State University, Manhattan, KS.

Mensah, C., C. DiFonzo, R. L. Nelson, and D. Wang. 2005. Resistance to soybean aphid in early maturing soybean germplasm. Crop Sci. 45: 2228–2233.

Milton, N. M., B. A. Eiswerth, and C. M. Ager. 1991. Effect of phosphorus deficiency on spectral reflectance and morphology of soybean plants. Remote Sens. Environ. 36: 121–127.

Mirik, M., G. J. Michels, S. Kassymzhanova-Mirik, and N. C. Elliott. 2007. Reflectance characteristics of Russian wheat aphid (Hemiptera: Aphididae) stress and abundance in winter wheat. Comput. Electron. Agric. 57: 123–134.

Nebiker, S., A. Annen, M. Scherrer, and D. A. Oesch. 2008. Light-weight multispectral sensor for micro UAV - opportunities for very high resolution airborne remote sensing. Int Arch Photogram Rem Sens Spat. Inf. Sci. 37(B1): 1193–1200.

Nutter, F. W., G. L. Tylka, J. Guan, A. J. Moreira, C. C. Marett, T. R. Rosburg, J. P. Basart, and C. S. Chong. 2002. Use of remote sensing to detect soybean cyst nematode-induced plant stress. J. Nematol. 34: 222–231.

Olson, K. D., T. M. Badibanga, and C. DiFonzo. 2008. Farmers' awareness and use of IPM for soybean aphid control: Report of survey results for the 2004, 2005, 2006, and 2007 crop years. Staff Pap. Report. Dep. Appl. Econ., Univ. Minn., St. Paul, MN.

O'Shaughnessy, S. A., S. R. Evett, P. D. Colaizzi, and T. A. Howell. 2011. Using radiation thermography and thermometry to evaluate crop water stress in soybean and cotton. Agric. Water Manag. 98: 1523–1535.

Pinter, P. J., Jr., J. L. Hatfield, J. S. Schepers, E. M. Barnes, M. S. Moran, C. S. T. Daughtry, and D. R. Upchurch. 2003. Remote sensing for crop management. Photogramm. Eng. Remote Sens. 69: 647–664.

Prabhakar, M., Y. G. Prasad, M. Thirupathi, G. Sreedevi, B. Dharajothi, and B. Venkateswarlu. 2011. Use of ground based hyperspectral remote sensing for detection of stress in cotton caused by leafhopper (Hemiptera: Cicadellidae). Comput. Electron. Agric. 79: 189–198.

Prabhakar, M., Y. G. Prasad, S. Vennila, M. Thirupathi, G. Sreedevi, G. Ramachandra Rao, and B. Venkateswarlu. 2013. Hyperspectral indices for assessing damage by the solenopsis mealybug (Hemiptera: Pseudococcidae) in cotton. Comput. Electron. Agric. 97: 61–70.

Ragsdale, D. W., B. P. McCornack, R. C. Venette, B. D. Potter, I. V. MacRae, E. W. Hodgson, M. E. O'Neal, K. D. Johnson, R. J. O'Neil, C. D. DiFonzo, *et al*. 2007. Economic threshold for soybean aphid (Hemiptera: Aphididae). J. Econ. Entomol. 100: 1258–1267.

Ragsdale, D. W., D. A. Landis, J. Brodeur, G. E. Heimpel, and N. Desneux. 2011. Ecology and management of the soybean aphid in North America. Annu. Rev. Entomol. 56: 375–399.

R Development Core Team, R. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rouse, J. W., Jr., R. H. Haas, D. W. Deering, and J. A. Schell. 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Prog. Report no. 7, Texas A&M University, College Station, TX.

Ruppel, R. F. 1983. Cumulative insect-days as an index of crop protection. J. Econ. Entomol. 76: 375–377.

Smith, G. M., and E. J. Milton. 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. Int. J. Remote Sens. 20: 2653–2662.

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

Song, F., and S. M. Swinton. 2009. Returns to integrated pest management research and outreach for soybean aphid. J. Econ. Entomol. 102: 2116–2125.

Summy, K. R., and C. R. Little. 2008. Using color infrared imagery to detect sooty mold and fungal pathogens of glasshouse-propagated plants. HortScience. 43: 1485–1491.

Tilmon, K. J., E. W. Hodgson, M. E. O'Neal, and D. W. Ragsdale. 2011. Biology of the soybean aphid, *Aphis Glycines* (Hemiptera: Aphididae) in the United States. J. Integr. Pest Manag. 2: 1–7.

United States Department of Agriculture – National Agricultural Statistics Service (USDA-NASS). 2018. Agricultural statistics 2018. https://www.nass.usda.gov/Publications/Ag_Statistics/2018/Complete%20Publication.pdf (accessed 30 April 2019).

Vigier, B. J., E. Pattey, and I. B. Strachan. 2004. Narrowband vegetation indexes and detection of disease damage in soybeans. IEEE Geosci. Remote Sens. Lett. 1: 255–259.

Yuan, L., Y. Huang, R. W. Loraamm, C. Nie, J. Wang, and J. Zhang. 2014. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. F. Crop. Res. 156: 199–207.

Zhang, C., and J. M. Kovacs. 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. 13: 693–712.

Downloaded from https://academic.oup.com/jee/article/113/2/779/5637354 by guest on 05 April 2022

Received: 2 October 2020 | Revised: 10 May 2021 | Accepted: 6 July 2021

DOI: 10.1002/navi.440

ORIGINAL ARTICLE

WILEY ION

Air data fault detection and isolation for small UAS using integrity monitoring framework

Kerry Sun | **Demoz Gebre-Egziabher**

Department of Aerospace Engineering and Mechanics, University of Minnesota - Twin Cities, Minneapolis, MN 55455, United States

Correspondence
Kerry Sun, Department of Aerospace Engineering and Mechanics, University of Minnesota - Twin Cities, Minneapolis, MN.
Email: sunx0486@umn.edu

Funding information
Minnesota Invasive Terrestrial Plants and Pests Center (MITPPC)

Abstract

A Fault Detection and Isolation (FDI) algorithm is developed to protect against Water-Blockage (WB) pitot tube failure in the safety-critical Air Data System (ADS) used on small Unmanned Aircraft Systems (UAS). The algorithm utilizes two identical Synthetic Air Data Systems (SADS) as the basis for state estimation. Each SADS works independently with a pitot tube while sharing an IMU and GNSS receiver. The fault detection is designed using the integrity monitoring framework, and the isolation is obtained via independent fault detection channels. The ADS requirements are established, and the WB failure mode is analyzed based on real faulty air data. A new residual-based test statistic is introduced, and the link among the test statistic, observability matrix, and Minimal Detectable Error (MDE) are examined. Finally, a flight data set with a known water-blockage fault signature is used to assess the algorithm's performance in terms of the air data protection levels and alert limits.

KEYWORDS
fault detection and isolation, integrity monitoring, synthetic air data system, UAS

1 | INTRODUCTION

A reliable Air Data System (ADS) plays a vital role in an aircraft's safety and performance. While ADS provides the measurement of various parameters, airspeed V_a, angle-of-attack α, and angle-of-sideslip β are the main parameters that define the flight envelope. As shown in Figure 1, airspeed is the speed of an aircraft relative to the air, and angle-of-attack and angle-of-sideslip are the flow angles relative to the aircraft. Air data is usually measured onboard by accurate air data sensors such as the pitot-static tube and the angle vane. Also, reliability analyses such as Fault Tree Analysis (FTA) and Failure Mode and Effect Analysis (FMEA) are often used to help identify fault modes and certify the redundant hardware systems.

The analytical redundancy approach is particularly useful for small Unmanned Aircraft Systems (UAS) due to the Size, Weight, And Power (SWAP) requirements. However, there is a lack of analytical methods to certify analytical redundancy. As emerging technologies such as Urban Air Mobility (UAM) (Vascik et al., 2018), or UAS operations either in the Line-of-Sight (LOS) or Beyond Visual Line-Of-Sight (BVLOS) (Cour-Harbo, 2017; Fang et al., 2018;

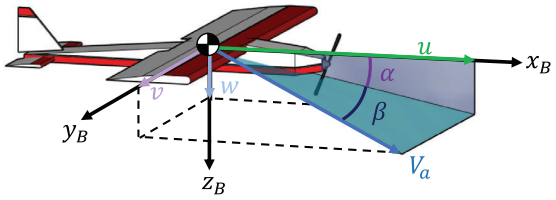


FIGURE 1 Illustration of air data triplet: airspeed V_a, angle-of-attack α, and angle-of-sideslip β [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

 © 2021 Institute of Navigation

Johnson et al., 2017; McCrink & Gregory, 2018; Yapp et al., 2018) mature, the need for rigorous and certifiable analytical redundancy methods will increase. In these applications, ADS is one of the safety-critical subsystems which needs to be certifiable and meet safety requirements.

Most of small UAS usually have one or two pitot tubes as the only sensors in its ADS. For example, the recent Part-135 certification process requires any unmanned air carriers (i.e., package delivery) to have at least one heated pitot tube (Federal Aviation Administration, 2019). However, many small UAS cannot afford to have a heated pitot tube onboard due to its cost and power requirements. Without the heating system, the low-cost pitot tubes on many small Unmanned Aerial Vehicles (UAVs) are prone to Water-Blockage (WB) faults. This is why many small UAV operations, such as agricultural surveying and construction inspection, cannot be carried out reliably during the rainy days. In Figure 2, a typical inexpensive pitot tube [10 US dollars to 20 US dollars (JDrones, 2020; Eagle Tree Systems, 2020)] is shown. It can be seen that the pitot tube is connected to a transducer via plastic tubes. The setup is simple and used by many UAVs but prone to the WB faults since there is no built-in drainage or heating system in the pitot tube. Water can enter the pitot tube on flights during foggy or rainy days, which fully or partially block the stagnation ports and affect the transducer's pressure readings.

To improve the safety and reliability of the ADS and minimize the number of redundant and multiple sensors, one approach being considered is called a Synthetic Air Data System (SADS) (Lie & Gebre-Egziabher, 2013; Sun et al., 2019b). A SADS is an estimator that calculates air data quantities using non-air data sensors such as the GNSS, IMU, magnetometer, and mathematical model of the aircraft.

SADS is a form of analytical redundancy that can help detect and deal with faults in the traditional ADS of small UAS. SADS can also potentially be coupled with one or two air data sensors to resolve the low-reliability issue. In fact, SADS has already been implemented in some commercial aircraft such as the Boeing 787 (Austrailian Transport Safety Bureau, 2015). The use of SADS is also being considered by many other aircraft designs at this time (Federal Democratic Republic of Ethiopia, Ministry of Transport, Aircraft Accident Investigation Bureau, 2020; Komite Nasional Keselamatan Transportasi, Republic of Indonesia, 2018; SeekingAlpha, 2019).

In what follows, we give a brief overview of the prior work on air data Fault Detection and Isolation (FDI) in the literature. We also explain why the existing methods are not adequate to certify ADS on small UAS and why the Integrity Monitoring (IM) framework can potentially solve this problem.

1.1 | Prior work

Air data FDI using advanced control and estimation algorithms has renewed interest over the last decade due to the recent advancements in the safety-critical UAV applications. These air data FDI techniques can be roughly separated into three categories: model-based, model-free, and data-driven algorithms.

The model-based algorithms typically leverage the dynamic model of the aircraft (Freeman et al., 2013; Hansen & Blanke, 2014; Ossmann et al., 2017). For example, Freeman et al. (2013) designed an airspeed fault detection algorithm using the aerodynamic model of the aircraft as well as linear robust H_∞ filters to detect faults, reject disturbance, and provide robustness to the modeling errors.

The model-free algorithms mainly rely on the sensor information and the kinematic models of the vehicles (Eubank et al., 2010; Guo et al., 2018; Lu et al., 2016; Van Eykeren & Chu, 2014). The model-free algorithms also often use Kalman Filter (KF)-based estimation techniques and the innovation χ^2 test to determine faults. An illustrative example is shown in Lu et al. (2016). They use an adaptive three-step unscented KF to detect and isolate air data faults.

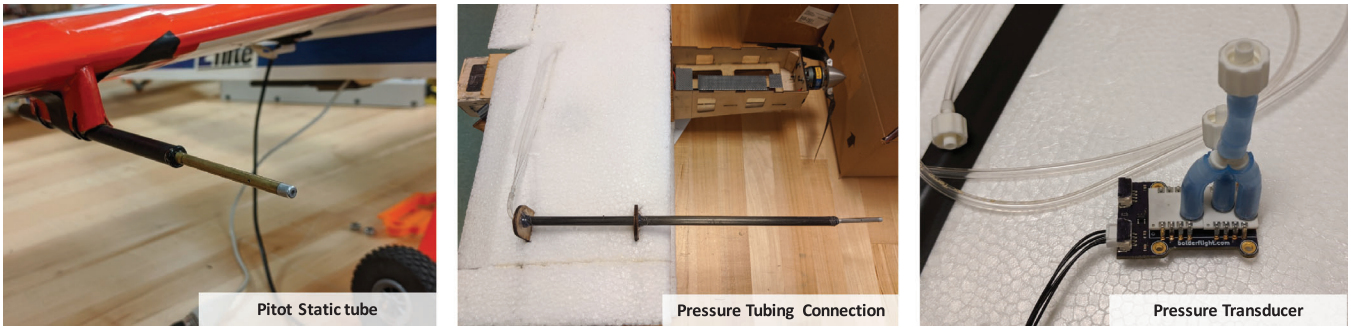


FIGURE 2 Entire ADS using a pitot-static tube, pressure tubes, and a pressure transducer for a small UAS [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

The data-driven algorithms primarily rely on large amounts of data to develop reliable input-output methods (e.g., autoregressive models, neural networks). These models are then used to detect faults through inconsistency check (Borup, 2018; Fravolini et al., 2017, 2018; Rohloff et al., 1999).

These air data FDI algorithms are primarily concerned with how to detect faults accurately, however, they do not provide a framework for ensuring the reliability of the air data fault detection algorithms from a requirement point of view. That is, they do not address the following question: *Can we design an air data FDI algorithm to satisfy a given set of system requirements such as integrity and continuity, and provide statistical protection levels for the air data estimates?*

Recent work (Freeman, 2014; Hu & Seiler, 2015; Kotikalpudi et al., 2020) has made progress towards certification of analytically redundant systems via reliability analysis. In the work here, we borrow tools often used in the field of integrated GNSS navigation to help design air data fault detection algorithms to ensure reliability.

One of the standard techniques for fault detection of the safety-critical aerospace navigation systems is called Receiver Autonomous Integrity Monitoring (RAIM) (Brown & Chin, 1998; Lee, 1986; Parkinson & Axelrad, 1988; Sturza, 1988). RAIM methods are used for safety-critical applications such as GNSS-based precision landing systems for aircraft (Khanafseh et al., 2014; Tanıl et al., 2017a, 2017b; Walter et al., 2008) and have been the subject of a significant amount of work for the last two decades. RAIM methods have also been used recently to provide Integrity Monitoring (IM) for other navigation systems such as the Simultaneous Localization and Mapping (SLAM) problem (Arana et al., 2019a, 2019b; Bhamidipati & Gao, 2019).

The basic idea of RAIM is to leverage redundant measurements at every time step [see snapshot detection scheme (Brown & Chin, 1998; Parkinson & Axelrad, 1988; Sturza, 1988)] or sequentially (Joerger & Pervan, 2013) to come up with probabilistic measures to detect faults and provide statistical bound to protect the state estimate. The advantage of this approach is that it provides the means for rigorous integrity risk computation. It uses redundant measurements to achieve fault detection capability and quantify the impact of undetected faults on state estimation errors. Another advantage is that the calculation of the threshold is based on probability, not selective tuning. Unlike RAIM, in many of the aforementioned works, it is often seen that a particular threshold is handpicked for a given application without rigorous probabilistic calculation.

However, the rigorous IM framework has not always been implemented on emerging non-PNT applications. This is partly due to the sensor measurements' inhomogeneity or the non-linearity of dynamics in many applications such as UAS. Many systems, such as ADS, have limited redundant and heterogeneous measurements at every time step. This limitation sometimes makes the snapshot of residual-based detection function infeasible to determine faults. And while linearization errors are usually small in the GNSS applications, measurement models in other systems are generally highly nonlinear, and the linearization errors' can be large. Therefore, the non-linearity might have a significant effect on the existing IM techniques. Lastly, many systems have observability issues [unobservable states (Kassas & Humphreys, 2014) or conditionally observable states (Sun et al., 2019b)], and this is usually overlooked when dealing with IM in GNSS applications.

Another aspect of fault detection in IM is the selection of an appropriate fault detector or test statistic. The goal of a fault detector is to detect fault quickly without raising too many false alarms. For real-time applications, online fault detectors are preferred. There are many online fault detectors such as the simple residual thresholding, KF innovation χ^2 test, least-squares residual-based χ^2 test [or commonly referred as the snapshot RAIM (Brown & Chin, 1998; Parkinson & Axelrad, 1988; Sturza, 1988) in navigation literature], Sequential Probability Ratio (SPRT), Cumulative Sum (CUSUM) test, and Generalized Likelihood Ratio (GLR) test.

Residual thresholding is the most straightforward test as it only requires a threshold to determine whether the data has exceeded the nominal level. The KF innovation χ^2 test is suitable for any KF-based state estimation, and the least-squares residual-based χ^2 test is a good choice when redundant measurements are available. The three methods mentioned above deal with linear or quadratic functions of the residual.

On the other hand, both SPRT and CUSUM tests are well-known for their nonlinear stopping rules (Gustafsson, 2000). For example, the standard one-sided SPRT test requires three tuning parameters: drift ν, threshold h, and reset level a. The basic idea of a one-sided SPRT test is to test whether the test statistics have drifted away significantly from the threshold. The drift parameter ν is used to subtract from the test statistic to control the drift's level, and the parameter a is used to prevent negative drift.

Similarly, the one-sided CUSUM test is the same as the one-sided SPRT test with the reset level $a = 0$. Several variations of both SPRT and CUSUM tests can be found in the literature. However, both tests require hand-tuning for the desired outcome. The GLR is also a powerful nonlinear test for fault detection, but it usually requires the knowledge of Probability Density Functions (PDF) under different hypotheses. In this paper, the KF-based detectors will be utilized with some additional novel improvement.

1.2 | Contribution

This paper provides four main contributions to the air data FDI literature: First, we design a dual pitot tube air data fault detection and isolation system that can be easily implemented on most UAVs. Second, we expand on sequential IM techniques in the Kalman filter setting to evaluate the integrity risk for the designed fault detection algorithm.

Specifically, we show how to deal with the limited redundant measurement problem and establish an analytical relationship among the residual-based test statistic, the Linear Time-Varying (LTV) observability matrix, and the Minimum Detectable Error (MDE). We also show how monitoring the observability of the system can potentially help rule out false alarms. Furthermore, we generalize the IM performance trade-off design procedure so that we can use it to evaluate other pitot tube failure modes.

Third, we also show how to establish alert limits and protection level bounds for the angle-of-attack and sideslip states. Lastly, we demonstrate our algorithm's capability using a recorded flight data in which a known WB pitot tube failure occurred.

1.3 | Paper organization

The remainder of this paper is organized as follows. Section 2 presents a brief description of the model-free SADS estimator used for the dual pitot tube air data system design developed in this paper. Section 3 presents the air data system requirements needed for the fault algorithm design. Section 4 describes the WB failure mode used in this work. Section 5 presents the fault detection design and analysis, which includes the derivation of the residual-based test statistic and its relation to the observability matrix, the MDE design and analysis, and the IM performance trade-off design procedure. Section 6 derives the alert limits and protection level calculations of angle-of-attack α and sideslip β. Section 7 presents the flight results and its associated detection performance. Concluding remarks and future outlook are given in Section 8.

2 | DUAL PITOT TUBE AIR DATA SYSTEM DESIGN

2.1 | System architecture

For the development that follows, we propose a dual pitot tube air data system for small UAS. The architecture consists of two identical SADS estimates of α and β by fusing airspeed measurements from the pitot tube with information from an IMU and a GNSS receiver (Sun et al., 2018, 2019a, 2019b).

Each SADS utilizes its own pitot tube, but both SADS share a GNSS receiver and IMU. Sharing the GNSS receiver and IMU reduces cost and software complexity. The design can be easily expanded to architecture with dual GNSS receivers and IMU units.

Each SADS can detect faults independently (i.e., identify and isolate the fault source) and is designed to satisfy the given system performance requirements (i.e., integrity and continuity requirements). The two SADS together provide recovery capability for a single faulty pitot tube failure via simple decision logic. This ADS system also provides accurate estimates and protection levels for the synthetic angle-of-attack α and sideslip β. The entire dual ADS design is illustrated graphically in Figure 3.

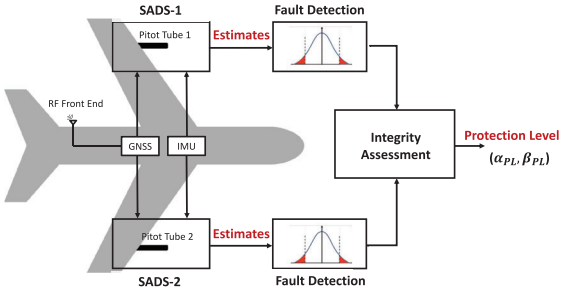


FIGURE 3 Dual air data fault detection design [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

In comparison to the commonly used triple-redundancy ADS design (Yeh, 1996), the dual pitot tube fault-tolerant ADS has a unique advantage: it only includes two small and inexpensive pitot tubes by leveraging a so-called dynamic redundancy approach (Isermann, 2005), which can be easily installed on the UAVs. In the case of a single pitot tube fault, the dual ADS can shut off the faulty pitot tube and continue its nominal operation using the secondary pitot tube (sometimes referred to as a *hot standby*).

Additionally, the two independent SADS filters can be implemented asynchronously on the hardware. It is a more fault-tolerant design choice in comparison to other filter methods [e.g., solution separation method (Joerger et al., 2014)], which in this case would use both pitot tube measurements simultaneously. Note that failure of a single pitot tube would lead to potential loss of control even if multiple filters (e.g., a bank of KF filters for various pitot tube failure modes) were used. Though simultaneous failure of both pitot tubes could occur under the same rainy condition, we considered that simultaneous-failure case beyond the scope of this paper and can be considered in future work. Table 1 summarizes the decision logic for the dual pitot tube ADS fault detection design.

TABLE 1 Dual Pitot tube ADS decision logic for each scenario

	SAD-1	SAD-2	Decision
Scenario 1	Nominal	Nominal	Nominal operation
Scenario 2	Nominal	Faulty	Use SAD-1 and raise alarm
Scenario 3	Faulty	Nominal	Use SAD-2 and raise alarm
Scenario 4	Faulty	Faulty	Terminate mission and land

Since SADS is the one of key components in the proposed algorithm, we will briefly go over some details to facilitate the understanding in the next subsection. For more details, please refer to Sun et al. (2019b).

2.2 | Synthetic Air Data System (SADS)

The synthetic SADS estimator is an extension of the 15-state, loosely-coupled INS/GNSS EKF (Gleason & Gebre-Egizabher, 2009), which blends information from an IMU and GNSS receiver. The INS/GNSS filter's state vector is augmented by three additional states representing the components of the wind velocity vector. Therefore, the SADS filter states, expressed in the error state vector $\delta\mathbf{x} \in \mathbb{R}^{18\times1}$, is given by:

$$\delta\mathbf{x} = \left[\delta\mathbf{p}^T \ \delta\mathbf{v}^{nT} \ \delta\boldsymbol{\psi}_{nb}^{nT} \ \delta\mathbf{b}_a^T \ \delta\mathbf{b}_g^T \ \delta\mathbf{W}^{nT}\right]^T \tag{1}$$

where $\delta\mathbf{p} = [\delta L \ \delta\lambda \ \delta h]^T$ is the position error vector in latitude, longitude, and altitude, $\delta\mathbf{v}^n = [\delta V_N \ \delta V_E \ \delta V_D]^T$ is velocity error vector resolved in the North-East-Down (NED) frame, denoted by the superscript n. The vector $\delta\boldsymbol{\psi}_{nb}^n = [\delta\phi \ \delta\theta \ \delta\psi]^T$ represents the attitude errors which are defined to be the small rotation angles between the actual NED frame and the estimated NED frame. The subscript nb indicates the positive direction is defined as being from the NED frame (n-frame) to the body frame (b-frame). The vectors $\delta\mathbf{b}_a = [\delta b_{ax} \ \delta b_{ay} \ \delta b_{az}]^T$ and $\delta\mathbf{b}_g = [\delta b_{gx} \ \delta b_{gy} \ \delta b_{gz}]^T$ are the accelerometer and rate gyro triad output bias error vectors, respectively. Finally, $\delta\mathbf{W}^n = [\delta W_N \ \delta W_E \ \delta W_D]^T$ is the error in the wind velocity vector resolved in the NED frame.

The synthetic SADS estimator synthesizes an estimate of α and β without using the α and β sensor measurements. The synthetic estimates of α and β are calculated using the EKF state estimates as follows:

$$\alpha = \tan^{-1}\left(\frac{u}{v}\right), \quad \beta = \sin^{-1}\left(\frac{v}{\sqrt{u^2+v^2+w^2}}\right) \tag{2}$$

where:

$$\begin{bmatrix} u & v & w \end{bmatrix}^T = \mathbf{C}_n^b(\boldsymbol{\psi}_{nb}^n)[\mathbf{v}^n - \mathbf{W}^n] \tag{3}$$

The $\mathbf{C}_n^b(\boldsymbol{\psi}_{nb}^n)$ is the coordinate transformation from NED to the body frame. The measurement vector $\mathbf{z} \in \mathbb{R}^{7\times1}$, shown in Equation (4), consists of position $\mathbf{p}$ and velocity $\mathbf{v}^n$ estimates from the GNSS receiver, along with the true airspeed V_a estimate determined using the pressure measurements from the pitot tube.

$$\mathbf{z}_k = \left[\mathbf{p}^T \ \mathbf{v}^{nT} \ V_a\right]_k^T \tag{4}$$

The time and covariance update equations for this filter are, for the most part, identical to those of the filter described in Gleason and Gebre-Egizabher (2009). What is new is the dynamic model for the augmented states (wind) and the measurement model.

Similar to the modeling of the accelerometer and gyroscope biases in the filter, the dynamics of the wind are modeled as a first-order Gauss-Markov model, motivated by Berman and Powell (1998). The details of the Gauss-Markov model for the wind and sensors can be found in (Berman & Powell, 1998) and (Xing, 2010), respectively. However, for the sake of completeness we re-state the process noise matrix $\mathbf{R}_w$ that accounts for accelerometer, gyroscope, and wind, respectively, for a quick reference:

$$\mathbf{R}_w = diag\left(\left[\sigma^2_{\mathbf{w}_a} \ \sigma^2_{\mathbf{w}_g} \ 2\sigma^2_{\mathbf{w}_{ad}}/\tau_{ad} \ 2\sigma^2_{\mathbf{w}_{gd}}/\tau_{gd} \ 2\sigma^2_{\mathbf{w}_{Wd}}/\tau_{Wd}\right]\right) \in \mathbb{R}^{15\times15} \tag{5}$$

where $\sigma^2_{\mathbf{w}_a}$ and $\sigma^2_{\mathbf{w}_g}$ are the accelerometer and gyroscope *white noise* variances, respectively. The parameter $\sigma^2_{\mathbf{w}_{ad}}$, $\sigma^2_{\mathbf{w}_{gd}}$, and $\sigma^2_{\mathbf{w}_{Wd}}$ are the accelerometer, gyroscope, and wind *random walk* variances, respectively. The τ_{ad}, τ_{gd}, and τ_{Wd} are the associated time constants defined in the first-order Markov process.

The linearized measurement model used by the EKF is $\delta\mathbf{z}_k = \mathbf{H}_k\delta\mathbf{x}_k + \boldsymbol{\upsilon}_k$, where $\boldsymbol{\upsilon}_k$, the measurement noise vector, is assumed to follow a normal distribution with zero mean and covariance $\mathbf{R}$, denoted as $N(\mathbf{0}, \mathbf{R})$. The measurement Jacobian $\mathbf{H}_k \in \mathbb{R}^{7\times18}$ is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times9} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3\times9} & \mathbf{0}_3 \\ \mathbf{0}_1 & H_{\mathbf{v}^n} & \mathbf{0}_{1\times9} & H_{\mathbf{W}^n} \end{bmatrix} \tag{6}$$

where the first two block rows map the EKF states into the GNSS position and velocity measurement errors, and the last block row maps the EKF states into the airspeed measurement error. The matrix $H_{\mathbf{v}^n}$ is derived from linearizing the nonlinear airspeed measurement model

$V_a = ||\mathbf{v}^n - \mathbf{W}^n||_2 + \upsilon_{V_a}$. The matrix $H_{\mathbf{W}^n}$ is similarly derived and happens to be equal to $-H_{\mathbf{v}^n}$. The matrix $H_{\mathbf{v}^n}$ is shown in the following:

$$H_{\mathbf{v}^n} = \frac{1}{V_a}\begin{bmatrix} V_N - W_N & V_E - W_E & V_D - W_D \end{bmatrix} \tag{7}$$

The associated measurement noise covariance $\mathbf{R}$ is shown as follows:

$$\mathbf{R} = diag\left(\begin{bmatrix} \sigma^2_{P_N} & \sigma^2_{P_E} & \sigma^2_{P_D} & \sigma^2_{V_N} & \sigma^2_{V_E} & \sigma^2_{V_D} & \sigma^2_{V_a} \end{bmatrix}\right) \in \mathbb{R}^{7\times 7} \tag{8}$$

where the diagonal of $\mathbf{R}$ contains the position, velocity, and airspeed noise variances.

2.3 | Observability consideration

An advantage of this SADS estimator is that it does not use the aircraft dynamic model, and it provides synthetic α and β estimates as well as their covariances. Specifically, unlike the model-based SADS, which uses the aerodynamic model of the aircraft and six degree-of-freedom dynamic equations, this mode-free SADS estimator only relies on the kinematic equation and sensor measurements. However, this estimator is conditionally observable as analyzed in detail in Sun et al.'s earlier work (2019b). Briefly, ensuring observability of this estimator requires the following two conditions (i.e., conditionally observable):

1. The airplane must be accelerating so that the INS/GNSS heading and gyro bias states become observable (Gleason & Gebre-Egizabher, 2009)
2. The wind vector $\mathbf{W}^n$ must be *quasi-static*. The term *quasi-static* means that the variations in $\mathbf{W}^n$ are assumed to be negligibly small over a small time window whose size is defined in (Sun et al., 2019b).

The second condition is required to ensure that changing airspeed and the wind states $\mathbf{W}^n$ separately observable (i.e., the wind triangle relationship). Furthermore, the quality of estimates, in part, depends on the degree of observability.

The degree of observability is determined quantitatively by analyzing the condition number of observability Gramian in Sun et al. (2019b). Since the synthetic estimate is conditionally observable, the ability to detect air data system faults is also conditional. One of this paper's key contributions is to show how observability is related to the fault test statistic, which is explained and demonstrated in Sections 5 and 7, respectively.

3 | AIR DATA SYSTEM REQUIREMENT

To quantify the air data fault detection performance, we start with given system requirements, such as integrity risk I_{req} and continuity risk C_{req}. The integrity risk is the probability that a hazardous fault goes undetected, and continuity risk is the probability that an alarm is issued about the presence of a fault when in fact there is no fault. Mathematically, they are defined as (Pervan, 1996):

$$I_{req} \triangleq P_{MD} + \sum_{DF} P(MI|DF)P_{DF} \tag{9}$$

$$C_{req} \triangleq P_{FA} + \sum_{DF} P(NI|DF)P_{DF} \tag{10}$$

where P is shorthand for *probability of*, and MD, FA, DF, MI, and NI stand for missed detection, false alarm, detected failure, missed-identified failure, and non-isolable failure, respectively.

Since the primary focus here is to deal with fault detection against pitot tube faults only, the second terms on the right-hand side of Equations (9) and (10) are ignored. These terms are associated with the isolation problem, which deals with all possible fault modes associated with the pitot tube.

In our case, we assume the pitot tube only experiences one particular failure mode, so the fault detection and isolation are essentially achieved at the same time. We also rely on the simple decision logic (Table 1) to recover from the faulty pitot tube experiencing the WB failure. Therefore, we limit our scope to the following performance requirement:

$$\begin{aligned} I_{req} &\approx P_{MD} \\ &= P(\text{Pitot tube fault not sensed} \\ &\quad |\text{Pitot tube has failed due to WB}) \end{aligned} \tag{11}$$

$$\begin{aligned} C_{req} &\approx P_{FA} \\ &= P(\text{issuing an alarm}|\text{no pitot tube failure}) \end{aligned} \tag{12}$$

As stated in Equations (11) and (12), the probability of missed detection P_{MD} represents the probability of not detecting a pitot tube failure given the pitot tube has indeed failed. Similarly, the probability of false alarm P_{FA} is the probability of issuing an alarm when there is no pitot tube failure. Also, if the pitot tube is working properly, this hypothesis is denoted as H_0. If the pitot tube is not working correctly, this hypothesis is denoted as H_1. The details of the particular WB pitot tube fault mode is

analyzed in Section 4. Also, since currently there is no universally accepted the numerical values of I_{req} and C_{req} for the WB pitot tube failure to our knowledge, we treat them as the trade-off variables in the Section 5.

4 | WATER BLOCKAGE FAILURE MODE

As mentioned in Section 1, low-cost pitot tubes such as the one shown in Figure 2 are susceptible to a failure model called Water Blockage (WB) during foggy or rainy days. This is a failure mode where the water particles in the air can enter through the front of the pitot tube and accumulate, leading to a reduction in total pressure either slowly or abruptly, as illustrated in Figure 4.

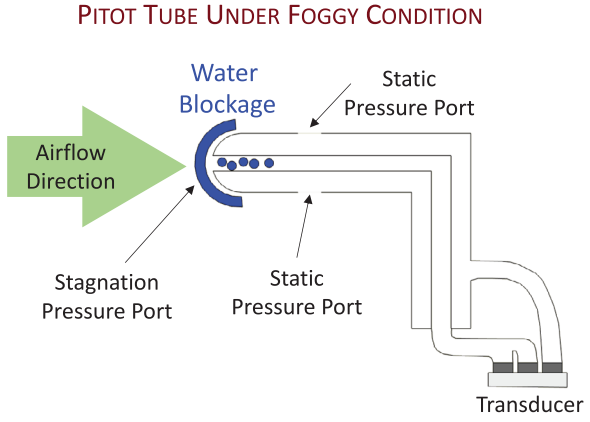


FIGURE 4 Illustration of the pitot static tube experiencing WB fault scenario [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

The airspeed V_a is typically calculated based on Bernoulli's principle as follows[1]:

$$V_a = \sqrt{\frac{2(P_t - P_s)}{\rho}} = \sqrt{\frac{2\Delta P}{\rho}} \tag{13}$$

where P_t is the stagnation or total pressure, P_s is the static pressure, and ρ is the air density. A partially blocked pitot tube would affect ΔP, which often leads to an airspeed drop. The size of the airspeed drop can vary significantly based on how much water is clogging the pitot tube.

Figure 5 shows a time history of two different faulty airspeed data sets from two different UAVs. The first faulty airspeed data (the top figure in Figure 5) comes from an agricultural inspection experiment. The flight data was collected by Sentera LLC.

The UAV took off around 570 secs, but the airspeed quickly decreased at 614 secs due to the WB faulty pitot

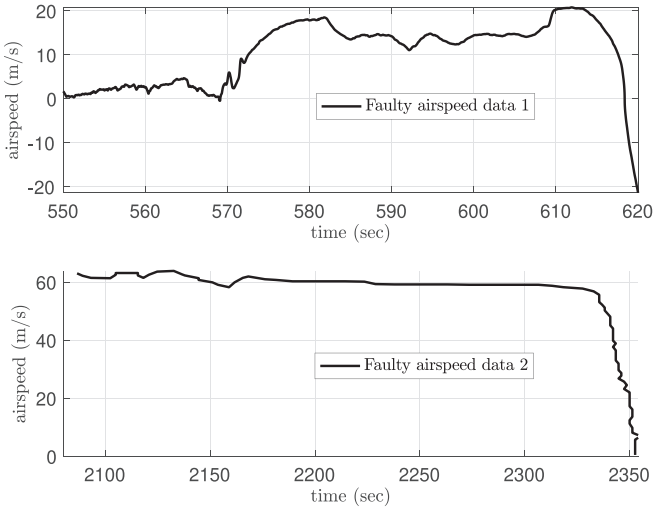


FIGURE 5 Two faulty airspeed data sets due to WB

tube. The other faulty airspeed data shown here[2] is reproduced from Hansen & Blanke (2014). The airspeed experiences a sharp drop at 2335 secs due to the water-clogged pitot tube.

Even though two different UAVs operate at different flight conditions (i.e., the nominal airspeed from the second set is almost three times higher than the first one), both pitot tubes experience a similar pressure drop rate. Those two airspeed drops' slope is estimated to be 2.5 m/s^2 and 3 m/s^2, respectively. Although this profile could be different from different pitot tubes, we will use them as the fault profile from which we need protection for the illustration in this paper.

Ideally, a larger faulty airspeed sample size would be required to represent the WB faulty pitot tube fault characteristics. However, it is difficult to obtain faulty airspeed data due to the WB failure in flight since: 1) the precise occurrence (i.e., timestamp) of the WB fault is usually unknown, and 2) faulty airspeed data is sensitive information and generally not shared in the public domain. In fact, to the best of our knowledge, this is the first paper that utilizes more than one set of faulty airspeed data due to the WB failure mode.

The airspeed measurement model under the faulty condition shown in Equation (14) is used for the SADS estimator. The WB fault mode is modeled as an unknown linear ramp fault, denoted as f_{V_a}, and the nominal airspeed is calculated by taking the euclidean norm (2-norm) of the difference between the inertial velocity $\mathbf{v}^n$ and wind vector $\mathbf{W}^n$ in the navigation frame. The airspeed measurement noise υ_{V_a} is modeled as the white Gaussian noise. The noise variance is typically unknown after the

[1] The temperature and altitude effect are not considered in this study

[2] The second airspeed data shown here is digitally extracted from the paper for illustration

fault occurs. Here we assume the noise variance stays the same before and after the fault:

$$V_a = ||\mathbf{v}^n - \mathbf{W}^n||_2 + f_{V_a} + \upsilon_{V_a} \tag{14}$$

Also, we assume the fault component f_{V_a} affects the nominal airspeed measurement continuously after water clogs in the tube. In other words, the water is assumed to stay in and continue clogging the pitot tube. A timeline for the fault scenario is described in Figure 6.

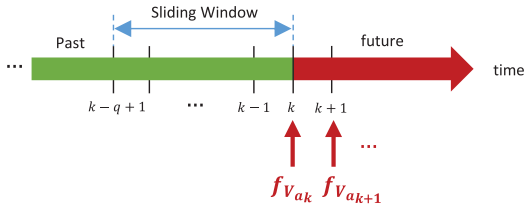


FIGURE 6 Fault scenario: the green portion of the timeline is fault-free section and the red is the faulted case [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

The fault vector $f_{V_{a_k}}$ starts entering at the time k and persists for future times. The test statistics developed in Section 5 uses a sliding window of size q to detect faults. This sliding window moves forward into the red region as time continues. Note that the sliding window would start right away when the measurement update of the KF filter reaches enough measurement for its detector. Hence the fault will most likely fall into the sliding window since the water blockage fault usually happens after the takeoff.

The methodology determining the minimum detectable faulty component f_{V_a} for the fault detection design is presented in Section 5. The minimum detectable airspeed fault depends on various factors, such as the measurement sampling rate T_s, the sliding window size q of the fault test statistic, the integrity risk P_{MD}, and continuity risk requirements P_{FA}. An IM trade-off design procedure is also presented to show how various factors can affect the design choice based on a given set of requirements.

5 | FAULT DETECTION DESIGN AND ANALYSIS

In this section, we first discuss the choice of Kalman filter based test statistics. Specifically, in addition to the conventional innovation-based KF test statistic, we introduce a sequential residual-based test statistic. We show how this test statistic is related to the observability matrix. Second, we discuss how the test statistics are generally designed to meet the integrity requirements.

The Minimum Detector Error (MDE) metric is used to link the integrity requirements and the specific air data system requirements. The MDE metric is determined through the Detector Operating Characteristic (DOC) curves. We also provide a general design procedure to determine the acceptable MDE and detection time τ for different failure modes. Lastly, we examine the quality of the proposed KF residual-based test statistic through the MDE analysis.

5.1 | Kalman filter based test statistics

Since the fault detection design in this paper relies on the Kalman filter based estimation, both innovation and least-squares residual-based χ^2 tests are considered for the fault detector design. Kalman filter based test statistics are widely used in the field of GNSS applications.

5.1.1 | Innovation-based test statistic

The most popular method is called the normalized innovation squared χ^2 test (Bar-Shalom et al., 2002). It uses the innovation vector and its covariance to form a test statistic, which follows a central χ^2 distribution with Df $= mq$ under the fault-free hypothesis H_0 as shown below:

$$D_{\gamma,k}|H_0 = \sum_{j=k-q+1}^{k} \boldsymbol{\gamma}_j^T \mathbf{S}_j^{-1} \boldsymbol{\gamma}_j \sim \chi^2(mq) \tag{15}$$

where $D_{\gamma,k}$ is test statistic at the time step k, $\boldsymbol{\gamma}_k$ is the innovation vector calculated using $\boldsymbol{\gamma}_k = \boldsymbol{z}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k^-)$, the matrix $\boldsymbol{S}_k$ is the innovation covariance calculated from $\boldsymbol{H}_k \boldsymbol{P}_k^- \boldsymbol{H}_k^T + \boldsymbol{R}_k$, m the number of measurements at the time step k, and q is the sliding window size. The vector $\hat{\boldsymbol{x}}_k^-$ is the predicted estimate, and $\boldsymbol{P}_k^-$ is the covariance of the KF prediction.

A suitable threshold for this test can be computed by using the inverse chi-square cdf $F_{\chi^2}^{-1}$ given the desired probability of false alarm and the appropriate Df as follows:

$$T_\gamma = F_{\chi^2}^{-1}(1 - P_{FA}, \text{Df}) = F_{\chi^2}^{-1}(1 - P_{FA}, mq) \tag{16}$$

The innovation-based test statistic's effectiveness depends on the quality of the KF-predicted estimates (linear prediction), KF measurements (sensor quality), and length of q. However, one weakness of the innovation-based test statistic is that it cannot be analyzed easily if there is a fault. This is because all the information embedded in the innovation vector and its covariance, and the contribution from a fault cannot be parsed out analytically. In other words, it would be more beneficial

if we could find a test statistic that is both analyzable and informative.

5.1.2 | Residual-based test statistic

The other common fault detector for KF-based estimation is the least-squares residual-based χ^2 test. However, the residual-based χ^2 test is not applicable to the SADS considered here because there is no redundant airspeed measurement at every time step; past measurements will be required for the fault detection design instead of using the well-established snapshot RAIM method. To overcome this issue, we formulate a KF residual-based test in the following:

$$D_{\mathbf{r},k}|H_0 = \mathbf{r}_{k-q:k}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}_{k-q:k} \sim \chi^2(mq-n) \tag{17}$$

where n is the number of the states in the KF and $\mathbf{r}_{k-q:k}$ is stacked residual vector from the past time step $k-q$ to the current time step k, denoted as $\mathbf{r}_{k-q:k} = \begin{bmatrix} \mathbf{r}_{\mathbf{k-q}}^{\mathbf{T}} & \mathbf{r}_{k-q+1}^T & \cdots & \mathbf{r}_k^T \end{bmatrix}^T$. Each residual $\mathbf{r}_k$ is computed from the difference between the measurements and a posteriori estimate $\boldsymbol{x}_k^+$ using $\mathbf{r}_k = \boldsymbol{z}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k^+)$. The matrix $\boldsymbol{\Sigma}$ is the covariance matrix of the weighted least residual vector $\mathbf{r}_{k-q:k}$ is shown in Equation (18):

$$\boldsymbol{\Sigma} = \mathbf{I}_{q\times q} \otimes \mathbf{R} + \mathbf{Q}_{w,k-q:k} (\mathbf{I}_{q\times q} \otimes \mathbf{R}_w) \mathbf{Q}_{w,k-q:k}^T \tag{18}$$

where $\otimes$ is the Kronecker tensor product and $\mathbf{R}_w$ is the process noise matrix. The matrix $\mathbf{Q}_{w,k-q:k}^T$ is realized through the batch linear system realization shown in Appendix A. The sliding window residual-based test statistic is used here instead of a one-time step residual test because the number of measurements m is less than the number of states n for the SADS.

In other words, the popular snapshot RAIM method from GNSS does not work here since no redundant measurements are available at each time step. The χ^2 test requires $\mathrm{Df} = mq - n > 0$, therefore the threshold for the residual-based test statistic is calculated as follows:

$$T_{\mathbf{r}} = F_{\chi^2}^{-1}(1-P_{FA}, \mathrm{Df}) = F_{\chi^2}^{-1}(1-P_{FA}, mq-n) \tag{19}$$

The sequential residual-based χ^2 fault detection test statistic has similar properties to the snapshot RAIM method. Also, we make a connection between this residual-based test statistic and the LTV observability matrix. This is done by connecting a window of measurement to a past state vector $\boldsymbol{x}_{k-q}$ using weighted least squares. The state vector $\boldsymbol{x}_{k-q}$ can be estimated by applying weighted linear least squares to the batch linear system shown in Equation (A2):

$$\hat{\boldsymbol{x}}_{k-q} = \boldsymbol{\mathcal{O}}_{k-q:k}^* \mathbf{Z}_{k-q:k} = \hat{\boldsymbol{x}}_{k-q}^+ \tag{20}$$

where $\hat{\boldsymbol{x}}_{k-q}^+$ signifies the posteriori estimate from the past measurement from time step $k-q$ to the current measurement k. The matrix $\boldsymbol{\mathcal{O}}_{k-q:k}^*$ is calculated as:

$$\boldsymbol{\mathcal{O}}_{k-q:k}^* = (\boldsymbol{\mathcal{O}}_{k-q:k}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{O}}_{k-q:k})^{-1} \boldsymbol{\mathcal{O}}_{k-q:k}^T \boldsymbol{\Sigma}^{-1} \tag{21}$$

where $\mathcal{O}_{k-q:k}$ is the LTV observability matrix shown in Equation (A7). The residual vector $\mathbf{r}_{k-q:k}$ can be subsequently expressed in the following:

$$\mathbf{r}_{k-q:k} = \begin{bmatrix} \mathbf{r}_{\mathbf{k-q}} \\ \mathbf{r}_{k-q+1} \\ \vdots \\ \mathbf{r}_k \end{bmatrix} = (\mathbf{I}_{mq\times mq} - \boldsymbol{\mathcal{O}}_{k-q:k} \boldsymbol{\mathcal{O}}_{k-q:k}^*) \mathbf{Z}_{k-q:k} \tag{22}$$

Under the fault-free H_0 hypothesis, we can also write $\mathbf{r}_{k-q:k}$ as follows:

$$\mathbf{r}_{k-q:k} = (\mathbf{I}_{mq\times mq} - \boldsymbol{\mathcal{O}}_{k-q:k} \boldsymbol{\mathcal{O}}_{k-q:k}^*) \left[\mathbf{Q}_{w,k-q:k} \mathbf{W}_{k-q:k} + \mathbf{V}_{k-q:k} \right] \tag{23}$$

where $\mathbf{r}_{k-q:k}$ follows a normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\mathbf{W}_{k-q:k}$ and $\mathbf{V}_{k-q:k}$ are defined in Equation (A4) and (A5) respectively.

Using Equation (22), we can also write $D_{\mathbf{r},k}$ as follows:

$$\begin{aligned} D_{\mathbf{r},k} &= \mathbf{r}_{k-q:k}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}_{k-q:k} \\ &= \mathbf{Z}^T (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T \boldsymbol{\Sigma}^{-1} (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) \mathbf{Z} \\ &= \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) \mathbf{Z} \end{aligned} \tag{24}$$

where the subscripts $mq \times mq$ and $k-q:k$ are dropped to shorten the notation. The last equality of Equation (24) is obtained because both $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)$ are symmetric, and the matrix $\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*$ is idempotent. The formal proof of this matrix equality is given in Appendix B.

The mathematical revelation in Equation (24) shows that the KF residual-based test statistic is a function of the observability matrix. Furthermore, the matrix $\boldsymbol{\mathcal{O}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{O}}$ inside of $\boldsymbol{\mathcal{O}}^*$ is the discrete weighted observability Gramian or the Fisher information matrix (Bar-Shalom et al., 2002).

This test statistic has a distinct advantage: it gives users a tool to analyze how the system's observability affects the test statistic $D_{\mathbf{r},k}$ given a sliding window of the

measurement from time step k-q to k. By analyzing the observability Gramian, we can tell how well the current system is observable. We can make useful statements between the effect of the test statistic and the motion of the vehicle (indirectly represented by the observability matrix). Furthermore, it can also be used to eliminate false alarms, as demonstrated in Section 7. Hence, we also call this test statistic the observability-based test statistic.

5.1.3 | Limitation of snapshot RAIM test statistic

It is worth noting that the test statistic in Equation (24) is different from the well-known RAIM-like $\sum_{j=k-q+1}^{k} \mathbf{r}_j^T \mathbf{R}_j^{-1} \mathbf{r}_j = \sum_{j=k-q+1}^{k} \boldsymbol{z}_j^T \mathbf{R}_j^{-1} (I - \mathbf{H}_j \mathbf{H}_j^*) \boldsymbol{z}_j$ test statistic, which also follows a central χ^2 with degrees-of-freedom $mq - n$ under H_0 hypothesis.

This test statistic does not account for the process noise from the time update step in the KF prediction step. It loses all the dynamic information between the KF measurement update. Again, though the snapshot $\mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k$ test statistic is often used for GNSS integrity monitoring, it is inapplicable when dealing with the system considered here which does not have redundancy measurements at each time step.

5.2 | Minimum Detectable Error (MDE) design

Before we proceed with determining the Minimum Detector Error (MDE), we will discuss some concepts and define some terms that will be used later. Any fault detection algorithm's goal is to detect credible faults before they lead a hazardous situation (e.g., loss of the aircraft, collision with terrain).

For a given UAV, the stall angle of attack α_{stall} and the minimum airspeed at which the airplane can fly $V_{a,stall}$ (stall speed) are synonymous. We will assume we are dealing with an electrically powered UAV, so its mass does not change during flight. Thus, the flight detection algorithm we design will have to detect faults before the estimated airspeed falls below $V_{a,stall}$.

Since the UAV's operating speed V_a is generally not constant during a given flight, the allowed drop in the estimate of airspeed (before the airplane is outside of the safe-flight envelope) is not constant either. To simplify the design, we use the average operating speed $\bar{V}_a$ as an approximation. We will call the difference between average operating airspeed $\bar{V}_a$ and $V_{a,stall}$ the airspeed Allowable Error, or AE for short.

Given a WB-fault profile, we can determine the time required for the airspeed estimate to drop below the stall speed. We call this τ_{max}, and the fault detection algorithm must detect a WB fault in a time shorter than τ_{max}.

Finally, we call the smallest airspeed estimation error that can be detected consistently (quantified by the missed detection and false alarm rate probabilities) the MDE. The fault detection algorithm ensures that MDE< AE and raises the alarm when the detection time less than τ_{max} after the onset of a pitot tube failure.

5.2.1 | MDE and τ_{max} determination

By examining the slope of the faulty airspeed data in Figure 5, it is determined that the minimum faulty airspeed drop rate is about 2.5 m/s^2. For the particular UAV used in the flight experiment, the average operating airspeed is about 17.5 m/s, and the stall speed is about 10 m/s. The difference 7.5 m/s between $\bar{V}_a$ and $V_{a,stall}$ translates to the maximum detection time $\tau_{max} = 3s$ as follows:

$$\tau_{max} = \frac{\text{AE}}{2.5} = \frac{\bar{V}_a - V_{a,stall}}{2.5} = \frac{17.5 - 10}{2.5} = 3s \tag{25}$$

If the fault detection algorithm fails to detect the fault before the fault exceeds AE, then the algorithm is ineffective against the allowable fault. Mathematically, the AE should satisfy the following:

$$\text{MDE}(P_{FA}, P_{MD}) \leq \text{AE}(\bar{V}_a, V_{a,stall}, \tau_{max}) \tag{26}$$

where the lower bound MDE is a function of P_{FA} and P_{MD}. The AE = 7.5 m/s is reasonable, but a tight bound since the UAV should be able to recover even if the airspeed drops below the stall speed as long as there is a sufficient altitude for recovery. Hence, a larger AE can be found based on the average operating altitude. Nevertheless, we will use 7.5 m/s as the upper bound for AE in the following analysis.

5.2.2 | MDE and sampling rate

We model the airspeed fault as a linear ramp fault using the minimum drop rate of 2.5 m/s^2. The fault detector should still catch any rate that is higher 2.5 m/s^2 since a larger fault would result in a quicker detection.

Since χ^2 based tests are used for the fault detection, one of the requirements for χ^2 test is that the degree-of-freedom (Df) has to be greater than zero. For example, the least-squares residual-based fault detection method (Brown & Chin, 1998) requires the number of the measurement m to be greater than the state n. Because we

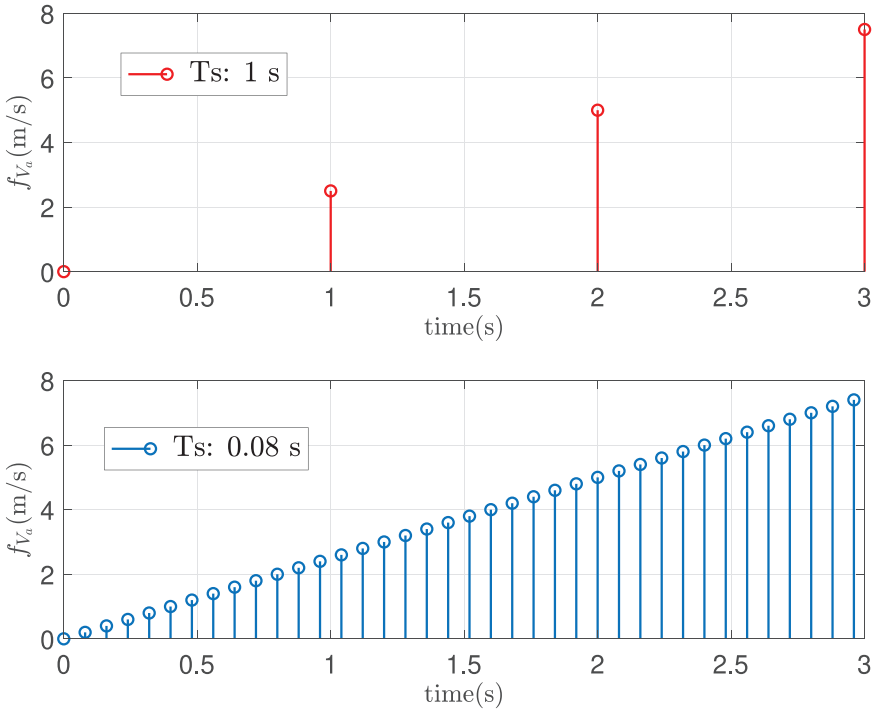


FIGURE 7 Simulated airspeed linear ramp fault profile at two different sampling rates [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

are doing sequential measurement update for the airspeed (i.e., $m = 1$ at each time step k) and we only assume the fault comes from the pitot tube, the minimum number of airspeed measurements needed for the residual-based χ^2 test is 19 since the number of states is 18.

Therefore, we need to collect at least 19 airspeed measurements (i.e., sliding window size $q = 19$) to detect fault within the maximum allowable detection time. Sampling too fast would also degrade the KF's performance because measurements closely spaced in time would cause the innovation vector to be highly correlated with itself, which violates the uncorrelatedness assumption.

Figure 7 shows the simulated airspeed linear ramp fault profiles at 1 s and 0.08 s sampling time over 3 s using the 2.5 m/s airspeed drop rate. The sampling time $T_s = 1$ s might be too slow for detection as the number of faulted measurement is too small for detection test statistic to respond before the time exceeds τ_{max}.

On the other hand, the sampling time $T_s = 0.08$ s is sufficient even if we only use the measurements after the fault occurs to form the test statistic[3]. Therefore, an appropriate sampling time should be chosen for the airspeed measurement based on the AE and the measurement uncorrelatedness assumption.

[3] Since a sliding window of measurement is used, the test statistics always has enough nominal measurements for χ^2 test before the fault occurs. If the faulty measurement enters the sliding window too slowly, the test statistics can be ineffective

For the small UAV we are dealing with here, the sampling time $T_s = 0.08$ s is found to be effective for good estimation and detection performance.

5.2.3 | MDE and detector operating characteristic curves

A Detector Operating Characteristic (DOC) curve (Sturza, 1988) is a graphical plot that illustrates the power of a discrimination threshold given various fault modes. It is used to understand the trade-off between the false alarm and missed detection rate, and the effectiveness of the designed threshold.

The DOC curves are obtained by plotting various P_{MD} and P_{FA} at different fault sizes. The DOC is the similar to the Receiver Operating Characteristic (ROC) curve, except the y-axis of ROC is the probability of detection P_D ($P_D + P_{MD} = 1$). Mathematically, the DOC curve is calculated using Equation (27):

$$P_{MD} = F_{\chi^2(\lambda)}(T, \mathrm{Df}) = F_{\chi^2(\lambda)}\left(F^{-1}_{\chi^2}(1 - P_{FA}, \mathrm{Df}), \mathrm{Df}\right) \tag{27}$$

where T is the designed threshold and determined by P_{FA}. The functions F_{χ^2} and $F_{\chi^2(\lambda)}$ are central and non-central χ^2 Cumulative Distribution Functions (CDF), respectively, with Df degrees-of-freedom. The non-centrality parameter λ represents the sum of the non-zero means. In this case,

it is sum of the biases vectors f_{V_a} over the sliding window q, as shown below:

$$\lambda = \sum_{j=k-q+1}^{k} \frac{f_{V_{a_j}}^2}{\sigma_{V_{a_j}}^2} \tag{28}$$

The expression for λ here is greatly simplified due to the sequential measurement update procedure. In the next subsection, we will generalize λ expression for dealing with various inhomogeneous measurements under the standard KF measurement update setting.

We use measurements over a short period for λ instead of a single measurement. Not only does this satisfy the Df requirement as mentioned earlier, but also a single faulty measurement may be ineffective. For example, if we wait until the fault grows to 7.5 m/s at 3 seconds, and use this measurement to form the test statistic (e.g., the innovation-based χ^2 test statistic only requires one measurement at minimum), then it would be too late in issuing the alarm. Hence we accumulate measurements over a short period and use them to form the detection test statistic.

For a fixed size sliding window q, λ can be calculated using a sliding window of the normalized $f_{V_{a_j}}^2/\sigma_{V_{a_j}}^2$ from the past time step $k-q+1$ to the current time step k. We define two parameters $\overline{\text{MDE}}$ and MDE in Equations (29) and (30):

$$\overline{\text{MDE}} \triangleq \sqrt{\lambda} = \sqrt{\sum_{j=k-q+1}^{k} \frac{f_{V_{a_j}}^2}{\sigma_{V_{a_j}}^2}} \tag{29}$$

$$\begin{aligned} \text{MDE} &\triangleq max\left(f_{V_{a_{k-q+1}}}, f_{V_{a_{k-q+2}}}, \ldots f_{V_{a_{k-1}}}, f_{V_{a_k}}\right) = f_{V_{a_k}} \\ &= f_{V_{a_{max}}} \end{aligned} \tag{30}$$

where $\overline{\text{MDE}}$ is the square root of the sum of normalized fault vectors and MDE the represents the magnitude of the largest fault in the sliding window. The definition of MDE is compatible with the one mentioned in Equation (26) because it represents the maximum tolerable fault size in the sliding window.

Figure 8 shows $\overline{\text{MDE}}$ using a window size $q = 19$ and a constant $\sigma_{V_a} = 1.75$ m/s. Each stem represents the sum of the past 19 normalized faulty measurements at sampled time k, where each faulty measurement is simulated based on the linear ramp fault profile shown in the bottom subfigure of Figure 7. The $\overline{\text{MDE}}$ represents the sum of the past normalized faulty measurements.

Figure 9 shows the DOC contour plot for a single degree-of-freedom: $\text{Df} = mq - n = 1 \times 19 - 18 = 1$. The plot is created by sweeping different λ over a wide range of P_{FA} and P_{MD}. The contour color is represented by the largest

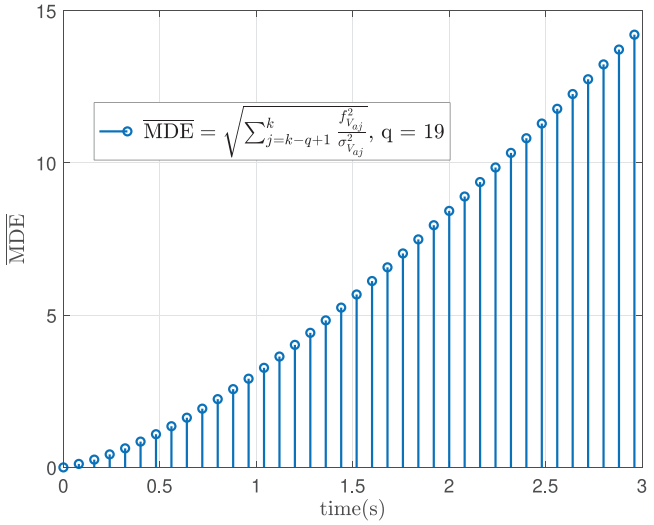


FIGURE 8 $\overline{\text{MDE}}$ over 3 secs period using the linear ramp fault profile in Figure 7 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

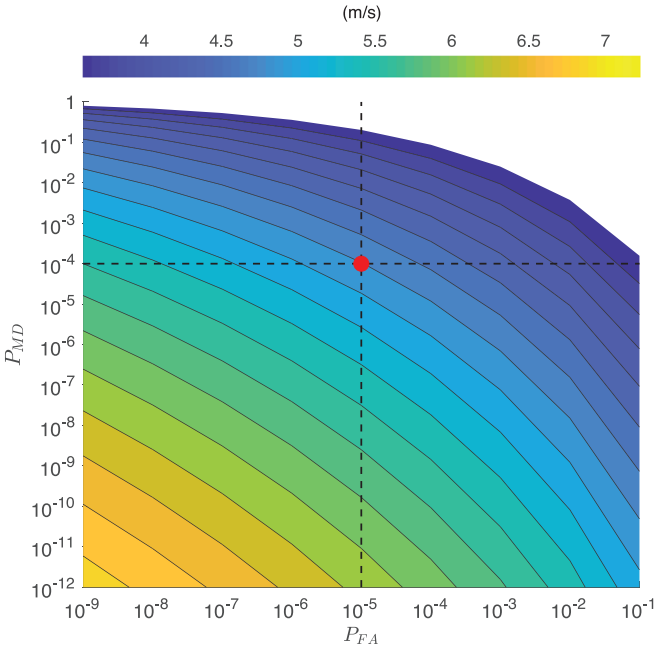


FIGURE 9 DOC curves where the color represents the associated MDE [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

minimum detectable airspeed error $f_{V_{a_{max}}}$ in the sliding window.

If the red dot represents a design choice $(P_{FA}, P_{MD}) = (10^{-5}, 10^{-4})$, the minimum detectable error requirement MDE is 5.0 m/s, which satisfies the inequality in Equation (26). In other words, the fault detection algorithm can detect the airspeed fault f_{V_a} after it reaches 5.0 m/s, which corresponds to a detection time $\tau = 2$ s starting from the beginning of the fault profile shown in the

TABLE 2 The sliding window $\overline{\text{MDE}}$ (m/s) and its associated MDE (m/s) in the square bracket for Df = 1

P_{FA}/P_{MD}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
10^{-1}	2.93 [3.6]	3.97 [3.6]	4.74 [3.6]	5.36 [3.8]	5.91 [4.0]	6.40 [4.2]	6.84 [4.4]	7.26 [4.6]	7.64 [4.8]
10^{-2}	3.86 [3.6]	4.90 [3.6]	5.67 [3.8]	6.29 [4.2]	6.84 [4.4]	7.33 [4.6]	7.78 [4.8]	8.19 [5.0]	8.57 [5.2]
10^{-3}	4.57 [3.6]	5.61 [3.8]	6.38 [4.2]	7.01 [4.4]	7.56 [4.8]	8.04 [5.0]	8.49 [5.2]	8.90 [5.4]	9.29 [5.4]
10^{-4}	5.17 [3.6]	6.22 [4.2]	6.98 [4.4]	7.61 [4.8]	8.16 [5.0]	8.64 [5.2]	9.10 [5.4]	9.51 [5.6]	9.89 [5.8]
10^{-5}	5.70 [4.0]	6.74 [4.4]	7.51 [4.8]	8.14 [5.0]	8.68 [5.2]	9.17 [5.4]	9.62 [5.6]	10.03 [5.8]	10.41 [6.0]
10^{-6}	6.17 [4.2]	7.22 [4.6]	7.98 [5.0]	8.61 [5.2]	9.16 [5.4]	9.65 [5.6]	10.10 [5.8]	10.51 [6.0]	10.89 [6.2]
10^{-7}	6.61 [4.4]	7.75 [4.8]	8.42 [5.0]	9.04 [5.4]	9.59 [5.6]	10.08 [5.8]	10.53 [6.0]	10.94 [6.2]	11.32 [6.4]
10^{-8}	7.01 [4.4]	8.06 [5.0]	8.82 [5.2]	9.45 [5.6]	9.99 [5.8]	10.48 [6.0]	10.93 [6.2]	11.34 [6.4]	11.73 [6.4]
10^{-9}	7.39 [4.6]	8.44 [5.2]	9.20 [5.4]	9.83 [5.6]	10.37 [6.0]	10.86 [6.2]	11.31 [6.4]	11.72 [6.4]	12.11 [6.6]

bottom sub-figure in Figure 7. Both of those numbers satisfy the given constraint as follows:

$$\begin{aligned} \text{MDE} = 5m/s \leq 7.5m/s = \text{AE} \\ \tau = 2s \leq 3s = \tau_{max} \end{aligned} \tag{31}$$

Of course, this is an ideal situation given the χ^2 test statistic is assumed to come from a perfect zero mean, unit variance white sequences. Nevertheless, the analysis provides a systematic way of assessing the fault detection capability for a realistic ramp airspeed fault profile.

Table 2 shows both $\overline{\text{MDE}}$ and MDE values over range of P_{FA} and P_{MD}. The numbers in the square bracket are the corresponding MDE values. It is seen that the MDE decreases as P_{FA} and P_{MD} increase and vice versa. This trend is correct and intuitive because a more stringent integrity and continuity requirement would enforce the detection function to catch a fault reliably at a larger MDE since a larger airspeed fault would trigger the detection function to cross the threshold more easily.

The final sliding window size was chosen to be 19 for both detectors. The sensitivity analysis is done for different sliding window sizes, and it is observed that increasing q from the minimum window size (i.e., 19) does not significantly change the DOC curves. Also, too big of a window size would make the test statistic function sluggish because it tends not to respond slower than the latest change in the measurement.

5.2.4 | Trade-off analysis procedure

The above subsections complete the determination of the MDE design based on the integrity requirements (I_{req}, C_{req}) and physical air data system requirement (AE, τ_{max}). Our MDE design can also be used for different pitot tube failure modes (e.g,. stuck or oscillatory fault). Note that analyzing a complete set of pitot tube failure modes would make the MDE calculation statistically more reliable, which is the subject of future studies. In what follows, we summarize the necessary IM performance trade-off design procedures:

1. Determine a suitable sampling rate for the measurement
2. Determine a reasonable AE based on the realistic fault mode profile (e.g., constant, ramp, or oscillatory)
3. Determine a feasible set of requirements I_{req} and C_{req} that satisfies the constraint in Equation (26) using DOC curves
4. If I_{req} and C_{req} are satisfied, then record the MDE and τ for assessing the fault detection performance
5. If I_{req} and C_{req} are not satisfied, return step 3. If the pair (I_{req}, C_{req}) is given as a hard requirement, then AE needs to be relaxed (return step 1)

5.2.5 | Important trade-off factors

The design procedure is also presented in a flowchart shown in Figure 10. It can be seen that when MDE and τ requirements are not satisfied, either performance requirements I_{req} and C_{req} are needed to be redefined, or AE and τ_{max} are needed to be re-adjusted. We also list the essential trade-off factors and their relationships that need to be considered for the IM performance trade-off design:

1. Sampling time versus auto-correlation
2. Effectiveness of test statistics versus sliding window size
3. System performance requirements (I_{req}, C_{req}) versus maximum allowable operating conditions (AE, τ_{max})
4. Sliding window size versus maximum allowable operating conditions (AE, τ_{max})

These variables are closely interconnected, and ultimately we need to find a set of (AE, τ_{max}) that is suitable for the given requirements (I_{req}, C_{req}). The choice of

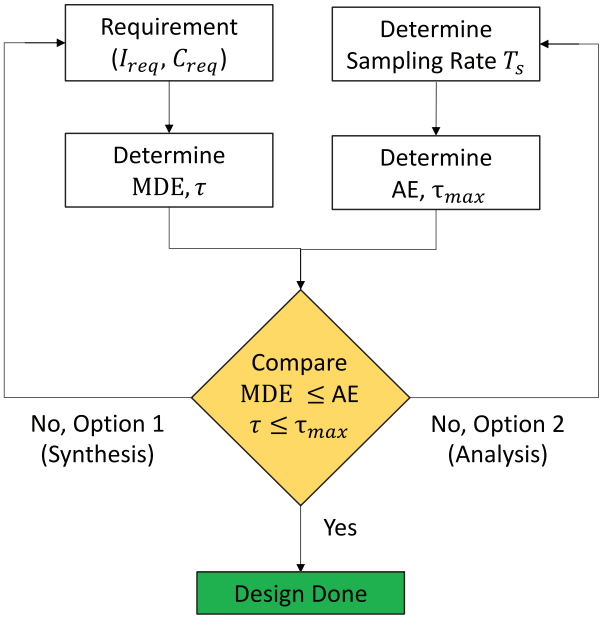


FIGURE 10 Trade-off design procedure [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

(AE, τ_{max}) would largely depend on the size of the sliding window, effectiveness of the chosen detectors (or test statistics), and sampling rate.

The fundamental limitation of change detection is that the design is a compromise between detecting true changes and avoiding false alarms. A poor design gives either a slow filter (no alarm from the test statistic) or a fast filter (many false alarms), and that is the worst can happen.

5.3 | MDE analysis for KF residual-based test statistic

The quality of a test statistic is often determined by its MDE (Grosch et al., 2017; Sturza, 1988). As demonstrated in Section 5.2.3, given a set of (P_{FA}, P_{MD}), the MDE (e.g., additive bias) can be found. The MDE informs us of the limits of the fault detection capability. In this section, we show the general expression of MDE for the residual-based test statistic. First, we can express $\mathbf{r}_{k-q:k}$ under the faulty H_1 hypothesis:

$$\mathbf{r}_{k-q:k} = (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)[\mathbf{Q}_w\mathbf{W} + \mathbf{V} + \mathbf{F}] \tag{32}$$

where $\mathbf{F}$ contains at most q steps since the start of the fault as defined in Equation (A6). $\mathbf{F}$ is the fault vector used to represent faults for different types of measurements. If we use both GNSS and airspeed measurement to update KF at the same time, then $\mathbf{F}$ can contain at most mq non-zero values. In this case, since the only faulty measurement is assumed to come from the pitot tube, we can further extract the faulty vector and maintain the correct matrix size by applying a column vector $\mathbf{E}_7$:

$$\mathbf{r}_{k-q:k} = (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\left[\mathbf{Q}_w\mathbf{W} + \mathbf{V} + (\mathbf{E}_7^T\mathbf{F})\mathbf{E}_7\right] \tag{33}$$

where $\mathbf{E}_7 = [\boldsymbol{e}_7^T \quad \boldsymbol{e}_7^T \quad \dots \quad \boldsymbol{e}_7^T]^T \in \mathbb{R}^{7q\times 1}$. The unit vector $\mathbf{e}_7$ is used because the airspeed measurement is on the seventh row shown in Equation (4). The size of $\mathbf{I}$ is now $7q \times 7q$. Then, the expectation of the residual $\mathbf{r}_{k-q:k}$ is calculated as follows since $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{V}]$ are zero respectively:

$$\mathbb{E}\left[\mathbf{r}_{k-q:k}\right] = (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)(\mathbf{E}_7^T\mathbf{F})\mathbf{E}_7 \tag{34}$$

Under the faulted hypothesis H_1, the residual-based test statistic follows a non-central χ^2 distribution with Df equal to $mq - n$:

$$D_{\mathbf{r},k}|H_1 \sim \chi^2_{(\lambda)}(mq - n) \tag{35}$$

The non-centrality parameter λ can be computed as:

$$\begin{aligned}\lambda &= \mathbb{E}\left[\mathbf{r}_{k-q:k}^T\boldsymbol{\Sigma}^{-1}\mathbf{r}_{k-q:k}\right] \\ &= \mathbf{E}_7\mathbf{F}^T\mathbf{E}_7^T(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{E}_7^T\mathbf{F}\mathbf{E}_7 \\ &= \mathbf{E}_7\mathbf{F}^T\mathbf{E}_7^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{E}_7^T\mathbf{F}\mathbf{E}_7\end{aligned} \tag{36}$$

In general, fault can come from any measurement or a combination of different measurements, and the fault characteristic depends on the particular type of measurement used. For an inhomogeneous fault vector $\mathbf{F}$, λ can be expressed as $\mathbf{F}^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{F}$ without any loss of generality. This might create a challenge if isolation is required. Hence, we continue with the expression shown in Equation (36). The parameter λ can be determined by solving Equation (27) for given a set of P_{FA}, P_{MD}, and Df.

Assuming we have an exact q-step pitot tube faults in the vector $\mathbf{F}$, we can project λ to a q-step detectable error for the pitot tube by reformulating Equation (36):

$$\sum_{j=k-q+1}^{k} f_{V_{a_j}}^2 = \frac{\lambda}{\mathbf{E}_7^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{E}_7} \tag{37}$$

This formulation assumes the fault happens in every time step in the sliding window. In reality, the fault can occur at any time, but it can be only accounted up to q

steps. Therefore, the following definition of MDE of f_{V_a}, denoted as $\text{MDE}_{f_{V_a}}$, is conservative if q is large, but less conservative when compared to the definition of MDE in Equation (30). The relationship between $\text{MDE}_{f_{V_a}}$ and MDE is shown below:

$$\begin{aligned}\text{MDE}_{f_{V_a}} &\triangleq \frac{1}{q}\sqrt{\sum_{j=k-q+1}^{k} f_{V_{a_j}}^2} \\ &= \frac{1}{q}\frac{\sqrt{\lambda}}{\sqrt{\mathbf{E}_7^T\boldsymbol{\Sigma}^{-1}(\mathbf{I}-\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{E}_7}} \\ &\leq \frac{\sqrt{\sum_{j=k-q+1}^{k} f_{V_{a_{max}}}^2}}{q} = f_{V_{a_{max}}} = \text{MDE}\end{aligned} \tag{38}$$

The $\text{MDE}_{f_{V_a}}$ can be interpreted as an average fault over the sliding window and is smaller than the MDE defined in Section 5.2.3. The difference between $\text{MDE}_{f_{V_a}}$ and MDE depends on the sliding window size and the fault profile. If a small sliding window size and a slow ramp fault profile are used, the MDE in the previous subsection is not a bad choice for the fault detection design.

6 | PROTECTION LEVEL CALCULATION

In the previous section, we introduced the test statistics and MDE design and analysis for the air data fault detection system. In this section, we derive a new protection level for α and β. We first introduce the definition of the alert limit and protection level in the context of synthetic air data, then we give the formal definition of the protection levels.

6.1 | Protection level and alert limit

We define the alert limit of angle-of-attack and sideslip angles needed for the protection level calculation. Alert Limit, denoted as AL, is usually defined as the maximum error in a state estimate that can be tolerated before a system is considered hazardous. Protection Level, denoted as PL, is defined as the guaranteed upper bound of the estimation error uncertainty σ. In theory, we want the probability of the state estimate error ϵ being greater than AL to be extremely low to assure integrity. Practically, we can also formulate the integrity requirement by using PL as follows[4]:

$$P(|\epsilon| > PL|H_i)P(H_i) \leq P_{MD} \text{ for } i = \{0,1\} \tag{39}$$

As a consequence, if we have $PL < AL$, the integrity requirement will be met. The error uncertainty σ is usually inflated by a factor K. This inflated error uncertainty $K\sigma$ is the protection level PL.

For a given integrity risk requirement P_{MD}, the PL are calculated. Protection Level PL is a function of the sensing system, and AL is a function of the operation. In other words, PL and AL are independent from each other.

In the case of pitot tube failures being considered here, we are interested in detecting WB faults before they result in the air data estimate (i.e., α and β) being outside of the safe-flight envelope, thereby leading to a control system (or a pilot in the case of a manned aircraft) to execute unnecessary but potentially hazardous maneuvers.

For the purpose of simplicity, we will assume the safe-flight envelope is a rectangle where the upper edge of the safe-flight envelope is defined by the UAVs maximum angle of attack α_{max} (which is the stall angle of attack α_{stall}). The lower edge is defined by a minimum angle of attack α_{min}, which is due to some aircraft structural considerations. The left and right edges are defined by β_{min} and β_{max}, which are derived either from aerodynamic control or structural strength limits. Therefore, we define the alert limit α_{AL}^{abs} and β_{AL}^{abs} as the *absolute* nominal safe operating region here since the true reference α and β are typically not available on UAVs. For the UAV considered here, the lower and upper bound of α_{AL}^{abs} and β_{AL}^{abs} is given in the following:

$$\begin{aligned}\alpha_{AL}^{\text{abs}} &= [\alpha_{min}, \alpha_{max}] = [-20°, 15°] \\ \beta_{AL}^{\text{abs}} &= [\beta_{min}, \beta_{max}] = [-30°, 30°]\end{aligned} \tag{40}$$

These boundaries of the safe-flight envelope form the absolute alert limits for the fault detection algorithm and are generally not the same for different UAVs.

Note that the protection level bound α_{PL} represents a deviation from the true state and the alert limit α_{AL} represents the *relative* error tolerance in the GNSS applications. However, the *absolute* alert limit α_{AL}^{abs} in this case is still valid. To see why this is the case, consider the following

[4] This expression is simplified from the following: $P(|\epsilon| > PL, D < T|H_i)P(H_i) = P(|\epsilon| > PL|H_i)P(D < T|H_i)P(H_i) \leq P(|\epsilon| > PL|H_i)$ $P(H_i) \leq P_{MD}$, where D is a detector function or test statistic, and T is a designed threshold

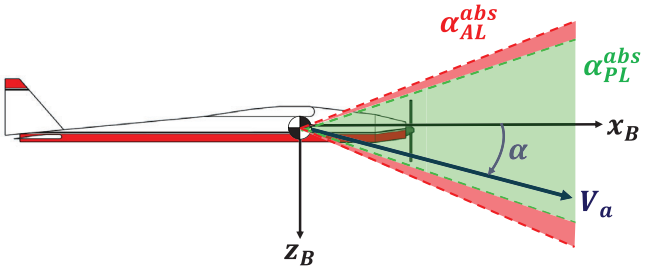


FIGURE 11 Depiction of protection level and alert limit of α [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

mathematical equivalence:

$$\alpha_{PL} < \alpha_{AL} \Rightarrow \alpha_{PL} + \hat{\alpha} < \alpha_{AL} + \hat{\alpha} \Rightarrow \alpha_{PL}^{\text{abs}} < \alpha_{AL}^{\text{abs}} \tag{41}$$

where α_{AL} represents the alert limit in the relative sense, and the absolute alert limit is defined as $\alpha_{AL}^{\text{abs}} \triangleq \alpha_{AL} + \hat{\alpha}$. Similarly, the absolute protection level is defined as $\alpha_{PL}^{\text{abs}} \triangleq \alpha_{PL} + \hat{\alpha}$. Hence, the new definition of α_{AL}^{abs} does not conflict with the typical definition of α_{AL}. In general, I_{req} is a set of discrete probability values representing various fault modes, and the probability of missed detection P_{MD_α} for α, as an example, would be part of the integrity budget I_{req}. Nevertheless, we will use the same value for the protection level calculation since we are only dealing with one fault mode, that is, $P_{MD_\alpha} = P_{MD_\beta} = P_{MD} = I_{req}$.

Figure 11 graphically depicts both protection level α_{PL}^{abs} and alert limit α_{AL}^{abs} in relation to α. Ideally, α_{PL}^{abs} should rarely go over α_{AL}^{abs} due to the small integrity risk requirement. When α_{PL}^{abs} does exceed α_{AL}^{abs}, we can safely conclude it is highly likely to have been the result of a faulty pitot tube and the integrity of α is lost.

6.2 | Protection level for synthetic air data

The fault detection algorithm needs to guarantee (at a certain level of confidence) that estimation error in α and β do not exceed their alert limits. Various slope-based PL calculations have been used to protect horizontal and vertical state errors against single (Walter & Enge, 1995; Brown & Chin, 1998; Milner & Ochieng, 2011) or multiple (Pervan et al., 1998; Angus, 2006; Blanch et al., 2009; Jiang & Wang, 2014) GNSS faults. However, these methods are developed for solving the redundant measurement problem.

Furthermore, the PL is usually calculated to protect states in an EKF. We develop a PL method that can protect the states derived from the EKF states. In particular, we calculate the PL for the synthetic angle-of-attack α and sideslip β estimates.

Under the fault-free hypothesis H_0, the PL is calculated by inflating state errors by a factor of K. In particular, the PL of α and β are calculated as follows:

$$\begin{aligned}\alpha_{PL,H_0} &= K_{\alpha,0}\sqrt{\mathbf{e}_1^T\mathbf{A}_{\alpha\beta}\mathbb{E}[\delta\boldsymbol{x}_k\delta\boldsymbol{x}_k^T]\mathbf{A}_{\alpha\beta}^T\mathbf{e}_1} \\ &= K_{\alpha,0}\sqrt{\mathbf{e}_1^T\mathbf{A}_{\alpha\beta}\boldsymbol{P}_k\mathbf{A}_{\alpha\beta}^T\mathbf{e}_1} = K_{\alpha,0}\sigma_\alpha\end{aligned} \tag{42}$$

$$\begin{aligned}\beta_{PL,H_0} &= K_{\beta,0}\sqrt{\mathbf{e}_2^T\mathbf{A}_{\alpha\beta}\mathbb{E}[\delta\boldsymbol{x}_k\delta\boldsymbol{x}_k^T]\mathbf{A}_{\alpha\beta}^T\mathbf{e}_2} \\ &= K_{\beta,0}\sqrt{\mathbf{e}_2^T\mathbf{A}_{\alpha\beta}\boldsymbol{P}_k\mathbf{A}_{\alpha\beta}^T\mathbf{e}_2} = K_{\beta,0}\sigma_\beta\end{aligned} \tag{43}$$

where $K_{\alpha,0}$ and $K_{\beta,0}$ under H_0 are calculated as follows:

$$K_{(\cdot),0} = Q^{-1}(P_{MD}/2) \tag{44}$$

where Q is the tail distribution function of the standard normal cdf. The matrix $\boldsymbol{P}_k$ is the state covariance from the EKF and $\mathbf{A}_{\alpha\beta}$ is the flow angle propagation transformation matrix specified in Sun et al. (2019b). The unit vectors $\mathbf{e}_1 = [1, 0]^T$ and $\mathbf{e}_2 = [0, 1]^T$ are used to extract the diagonal terms of $\mathbf{A}_{\alpha\beta}\boldsymbol{P}_k\mathbf{A}_{\alpha\beta}^T$.

In the presence of fault under the hypothesis H_1, PL needs to be increased to account for the faulty airspeed component f_{V_a}. The formulation is done as follows according to (Brown & Chin, 1998; Angus, 2006):

$$\alpha_{PL,H_1} = \text{slope}_\alpha\sqrt{\lambda_U} + K_{\alpha,1}\sigma_\alpha = \frac{\sigma_\alpha}{\sqrt{D_{\mathbf{r},k}}}\sqrt{\lambda_U} + K_{\alpha,1}\sigma_\alpha \tag{45}$$

$$\beta_{PL,H_1} = \text{slope}_\beta\sqrt{\lambda_U} + K_{\beta,1}\sigma_\beta = \frac{\sigma_\beta}{\sqrt{D_{\mathbf{r},k}}}\sqrt{\lambda_U} + K_{\beta,1}\sigma_\beta \tag{46}$$

where slope_α and slope_β represent the ratio between the α or β state error and the standard deviation of the test statistic $D_{\mathbf{r},k}$. The notion of the maximum slope (Brown & Chin, 1998) is not applicable here since there is no redundant airspeed at each time step to calculate a set of slopes. Specifically, since the accumulated sequential residual or innovation vectors come from the single pitot tube, so the source of the fault is always known and singular. In other words, the multiple hypothesis tests of determining which GNSS measurement is faulty by calculating the maximum slope does not apply here. The slope value in this study depends on the test statistic window size and severity of the fault profile.

The inflation factors $K_{\alpha,1}$ and $K_{\beta,1}$ under H_1 can be calculated as follows:

$$K_{(\cdot),1} = Q^{-1}\left(\frac{P_{MD}}{2P_{H_1}}\right) \tag{47}$$

where P_{H_1} represents the failure rate of the pitot tube due to water blockage. Generally, the failure rate of the pitot tube due to a particular fault is calculated based on rigorous sensor characterization testing. Since the failure rate testing is not done for this study, we assume $K_{(\cdot),1} = K_{(\cdot),0} = Q^{-1}(P_{MD}/2)$ for both α and β. This assumption is valid but conservative as $K_{(\cdot),1}$ is generally greater than $K_{(\cdot),0}$ in general. The parameter λ_U is the upper confidence bound for the maximum q-step airspeed faults, and it is computed as follows:

$$\begin{aligned}\lambda &= \mathbf{E}_7\mathbf{F}^T\mathbf{E}_7^T\boldsymbol{\Sigma}^{-1}(\mathbf{I}-\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\mathbf{E}_7^T\mathbf{F}\mathbf{E}_7 \\ &\leq \mathbf{E}_7\mathbf{F}^T\mathbf{E}_7^T\boldsymbol{\Sigma}^{-1}\mathbf{E}_7^T\mathbf{F}\mathbf{E}_7 \\ &\leq \mathbf{E}_7\mathbf{F}^T\mathbf{E}_7^T\boldsymbol{R}^{-1}\mathbf{E}_7^T\mathbf{F}\mathbf{E}_7 \\ &\leq \sum_{j=k-q+1}^{k}\frac{f_{V_{a_j}}^2}{\sigma_{V_{a_j}}^2} = \sqrt{\overline{\mathrm{MDE}}} = \lambda_U\end{aligned} \tag{48}$$

This upper bound λ_U can make the protection levels overly conservative if the size of the sliding window q is large. Hence, the tightness of the protection level is another factor for choosing an appropriate q.

FIGURE 12 PHX by Sentera LLC [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

It is worth noting that many small UAVs do not have angle vane sensors to provide a set of α and β. The protection levels of α and β are particularly useful when the angle vane sensor is not available. We can monitor α and β based on the synthetic air data estimates and protect the vehicle from exceeding α and β flight envelopes due to the pitot tube failure.

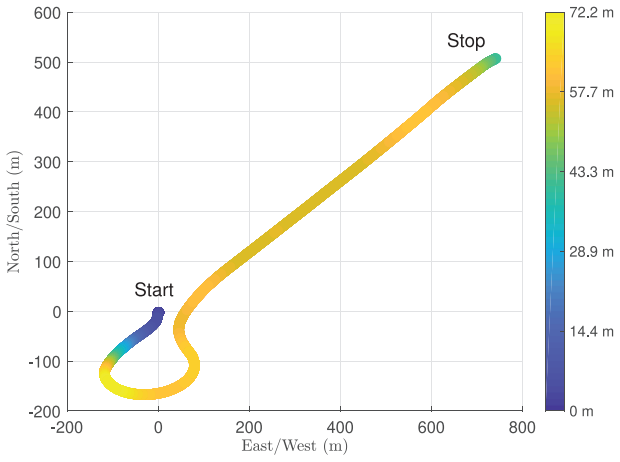
7 | FLIGHT DATA TESTING

The fault detection algorithm is tested using a flight data set recorded by a UAV. The UAV is the Sentera Phoenix (PHX) shown in Figure 12. It has an Eagle Tree pitot tube (Eagle Tree Systems, 2020) attached to the right wing to measure airspeed. PHX utilizes a similar version of Pixracer autopilot (Pixhawk FMU-V4) for control and navigation.

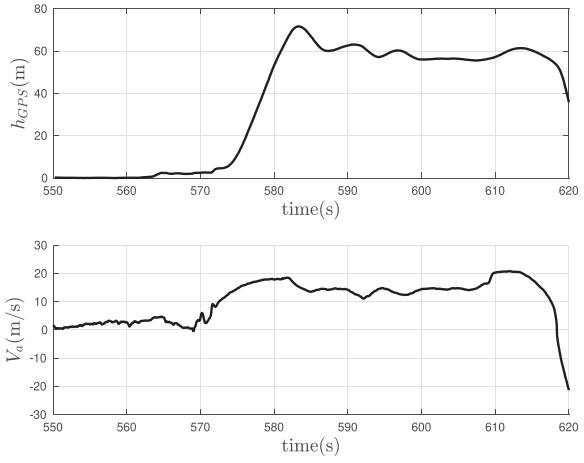
For this particular flight operation, the UAV crashed 45 seconds into the flight due to a water-plugged faulty pitot tube. This flight data is suitable for the fault detection algorithm analysis since it contains a known WB pitot tube fault signature.

Also, since the UAV has only one pitot tube, only one SADS is employed to show the flight results. Ideally if this UAV carried two pitot tubes and two independent SADS, then it would follow the decision logic in Table 1 to raise alarm and switch to SADS-2 when SADS-1 detected and isolated the fault airspeed measurement.

Figure 13(a) and 13(b) show the trajectory, altitude, and airspeed over the short flight period before the crash. The UAV takes off around 570 s and circles up to about 70 m, then flies straight northwest direction to the edge of the crop field. It is seen in Figure 13b that the airspeed of UAV experiences a sharp drop right after 618 s. The airspeed



(a) Recorded Trajectory using GPS data.



(b) Recorded altitude and airspeed over time.

FIGURE 13 Trajectory and airspeed information over the time of flight [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

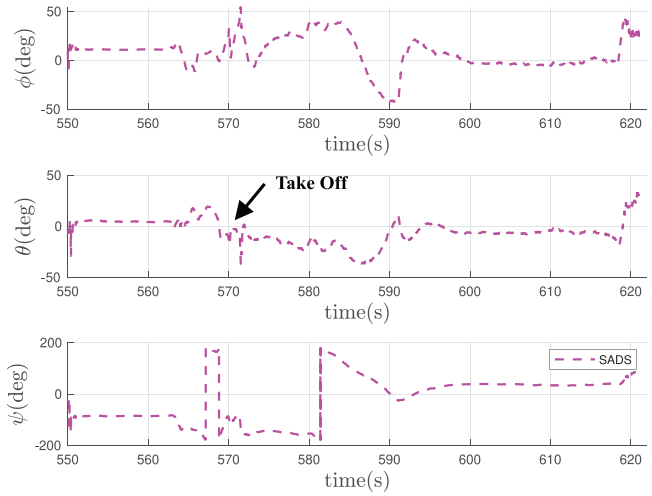


FIGURE 14 SADS attitude estimates over time [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

measurement eventually drops to a negative value. The last recorded altitude is around 35 m by the onboard autopilot.

Figure 14 shows the attitude estimates from the SADS filter. Notice that the Euler angle estimates start to change abruptly at the end of the flight due to the faulty pitot tube. For example, the pitch angle θ and roll angle ϕ increased dramatically around 618 s. The high value θ indicates the UAV might have been pitching up and stalling.

Figure 15 shows three different test statistics and their corresponding thresholds from the actual takeoff timestamp of 570 s to the end of the flight.

The top sub-figure shows the sliding window innovation-based test statistic D_γ over time where the window size q is 19. The bottom sub-figure shows a residual-based test statistic $D_{\mathbf{r}}$ and a Geometric Moving Average (GMA) residual-based test statistic $D_{\mathbf{r}}^{\text{GMA}}$, where the window size q is also 19 for both. The $D_{\mathbf{r}}^{\text{GMA}}$ is defined as the same as $D_{\mathbf{r}}$ except the residual vector $\mathbf{r}_{k-q:k}^{\text{GMA}}$ is written as follows:

$$\mathbf{r}_{k-q:k}^{\text{GMA}} = \left[\lambda^q \mathbf{r}_{k-q}^T \ \lambda^{q-1}\mathbf{r}_{k-q+1}^T \ \ldots \ \lambda \mathbf{r}_{k-1}^T \ \mathbf{r}_k^T\right]^T \tag{49}$$

In Equation (49), it is seen that the $D_{\mathbf{r}}^{\text{GMA}}$ uses an exponential forgetting factor μ^i to weigh on the past measurements less than the recent. The innovation-based threshold T_γ and residual-based $T_{\mathbf{r}}$ are different from each other due to the difference in Df, even though the same size of the sliding window is used.

It is seen that the innovation-based test statistic exceeds its threshold three times. It appears that the first occurrence at 593 s is most likely a false alarm given the fact we know the pitot tube WB fault occurs at the end of the flight. Furthermore, by examining the attitude in Figure 14, the large ϕ and sudden change in θ suggests a momentary accelerated stall.

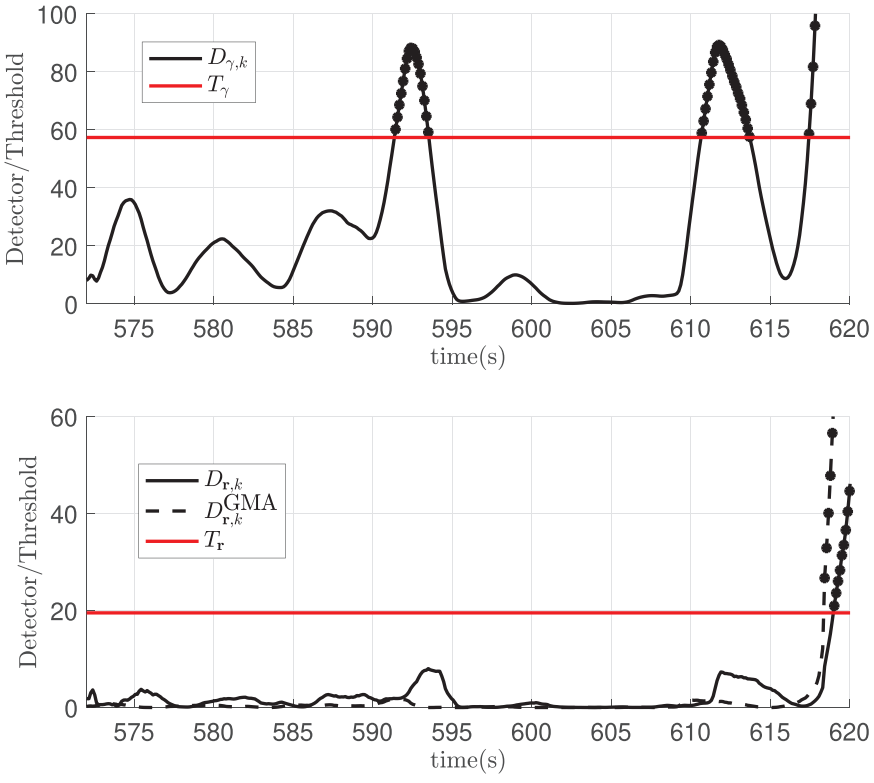


FIGURE 15 Detection variables over time [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

It is well known that the innovation-based test statistic is sensitive to the highly dynamic motion. The UAV experiences a large roll change at 593 s, which may cause D_γ to think there might be a fault in the system. The second occurrence might come from a sudden increase in airspeed at 610 s. Though we do not know if the fault has occurred yet by visually examining the airspeed plot, D_γ indicates there might be a fault.

Notice D_γ goes below the threshold at 614 s before rising to cross the threshold again at 618 s for the third time. This unstable behavior is not a good characteristic for a detector because it will be issuing many false alarms and failing to provide a crisp decision of a fault's occurrence.

It is also well known that the innovation-based test statistic is sensitive to sensor noise scaling (Gustafsson, 2000), and the innovation covariance $\mathbf{S}$ inside of D_γ is sensitive to noise scaling. The relationship is illustrated by Equation (50): $\mathbf{S}$ is affected by noise covariance matrices $\mathbf{R}_w$ and $\mathbf{R}$, and the initial state covariance $\mathbf{P}_0$. Small η can make D_γ sensitive to noise even though the state estimates are not affected.

$$\left.\begin{aligned} \bar{\mathbf{R}} &= \eta\mathbf{R} \\ \bar{\mathbf{R}}_w &= \eta\mathbf{R}_w \\ \bar{\mathbf{P}}_0 &= \eta\mathbf{P}_0 \end{aligned}\right\} \Rightarrow \begin{aligned} \bar{\mathbf{P}} &= \eta\mathbf{P} \\ \bar{\mathbf{S}} &= \eta\mathbf{S} \end{aligned} \tag{50}$$

Though not shown here, by scaling the process noise $\sigma_{\mathbf{w}_{Wd}}$ of the wind vector without adversely changing the performance of the estimator, the first two crossing occurrences of D_γ can be suppressed. However, hand-tuning is not recommended because it might adversely affect other estimates.

On the other hand, both residual-based test statistics $D_\mathbf{r}$ and $D_\mathbf{r}^{\text{GMA}}$ exceed their corresponding thresholds only once at the end of the flight. The threshold value for $D_\mathbf{r}$ and $D_\mathbf{r}^{\text{GMA}}$ are the same, hence only one $T_\mathbf{r}$ is plotted. Though small increases at 593 s and 612 s can be seen in $D_\mathbf{r}$, $D_\mathbf{r}$ did not cross its threshold.

The residual-based test statistic is less sensitive to both highly dynamic motion and noise scaling due to the larger weighting factor Σ. Furthermore, the GMA test statistic $D_\mathbf{r}^{\text{GMA}}$ appears to be even less sensitive. $D_\mathbf{r}^{\text{GMA}}$ is discounting many early measurements in the sliding window. Also, it rises faster than $D_\mathbf{r}$ when crossing the threshold, indicating a shorter fault detection time. When comparing detection time alone, the innovation does appear to have the fastest detection at the third crossing.

The GMA technique used here illustrates that sometimes different techniques can be used to improve the baseline detector function. For example, we may combine the CUSUM and residual-based test statistics to improve detectability. Ultimately, it is a trade-off between the detection time and false alarm when choosing a detection test statistic. In this case, it appears the GMA residual-based test statistic $D_\mathbf{r}^{\text{GMA}}$ is the best for the faulty pitot tube detection since it crosses the threshold at the correct incident and provides a good detection time.

In order to see how observability can actually play an important role in fault detection, we also monitor the observability Gramian of the system over time. Different metrics can measure the degree of discrete observability Gramian, such as the determinant, trace, or the condition number (Summers et al., 2016; Avant & Morgansen, 2019). In work here, we utilize the condition number κ to quantify the degree of observability. That is:

$$\kappa\left[\mathcal{G}_{d,k-q:k}\right] \triangleq \kappa\left[\boldsymbol{\mathcal{O}}_{k-q:k}^T\boldsymbol{\mathcal{O}}_{k-q:k}\right] \tag{51}$$

where $\mathcal{G}_{d,k-q:k}$ the discrete observability Gramian using the information from the time step k-q to the current k. In general, if the condition number of the observability Gramian is large, it means the states are not well observed. The observability Gramian here only requires a finite horizon instead of the infinite horizon. This calculation is done to monitor the recent motion of the dynamics of the vehicle. It also reduces the computational burden for the computer processor.

Figure 16 shows the normalized condition number of observability Gramian over time. The poor condition number is expected before takeoff since the observability

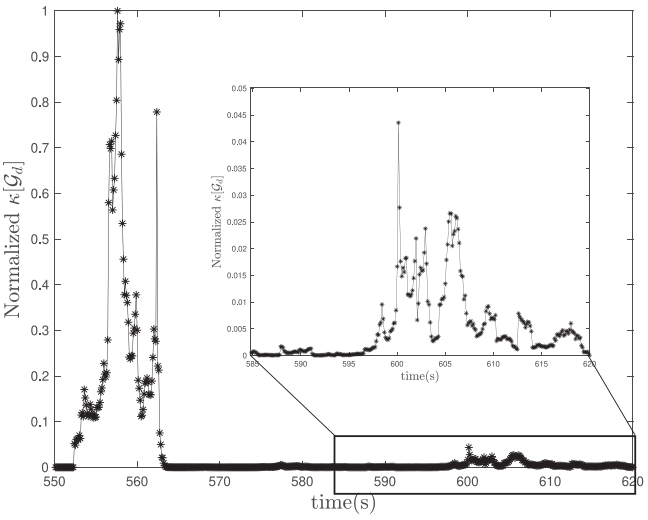


FIGURE 16 Normalized condition number of observability Gramian

Gramian rank is deficient. The condition number quickly reaches near zero after 570 s (takeoff). However, it is seen that the condition number got a bit worse from 600 s to 607 s. This change is perhaps due to the straight level flight (no heading change), or some water started to enter the pitot tube before it became evident.

The second claim is deduced based on examining all the EKF state estimates at 600 s. Since there is no abnormal phenomenon from the position, velocity, and attitude estimates around 600 s, the IMU and GNSS measurements are assumed to be nominal. Hence, the culprit is either the poor observability due to the trajectory or the pitot tube.

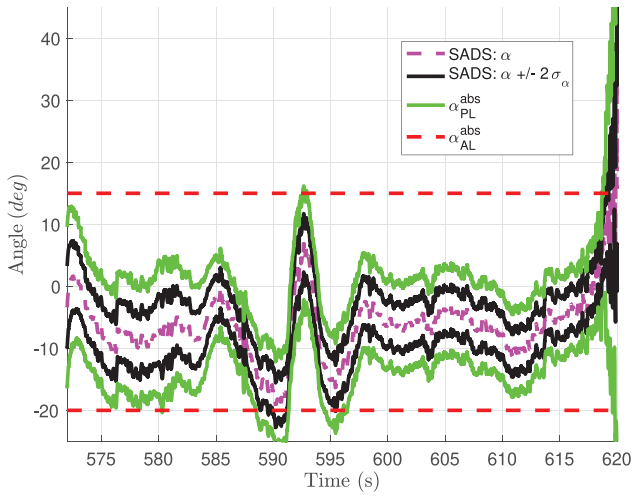
A poor condition number can potentially trigger the fault detection algorithm to raise the alarm even though there is no real fault. If there is a fault that is about to happen, then monitoring the observability Gramian can potentially warn the system before the fault occurs. If a poor condition number is a result of the straight-level flight, then weak observability can mask the fault detectability if a fault occurs during this time. Therefore, a close examination should be carried out when using observability to detect the actual fault.

Figures 17(a) and 17(b) show the α and β estimates, 2σ uncertainty bounds, protection levels α_{PL} and β_{PL}, and alert limits α_{AL} and β_{AL}. The protection level bound is for the single SAD. The protection level α_{PL} goes outside the alert limit momentarily at 590 s, 592 s, and 595 s, and exceeds both lower and upper alert limit at the end of the flight. The s-turn causes the first three crossings before heading to northwest direction. The last one is caused by the faulty pitot tube. It can be seen that the α increases drastically at the end of the flight, which leads to the stall and eventual crashing. Sideslip β also experiences a similar change during the s-turn. The sideslip estimate β changes from positive to negative, then positive again around 590 s, which is intuitively correct based on the s-turn and the northeast tail wind direction.

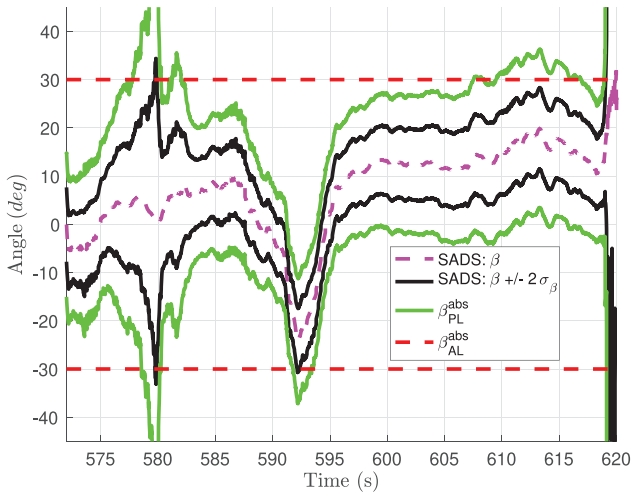
The protection level β_{PL} exceeds the alert limit a couple of times and eventually goes outside the alert limit at the end of the flight as expected. The protection level of α and β can be used to check the validity of ADS integrity, which is much more useful than just looking at the confidence uncertainty bound represented by 2σ. For example, we can see that the α and its $2\sigma_\alpha$ seem to be acceptable from 593 s to 596 s, but the integrity of the α estimate is actually lost during this time based on the protection level. This loss of integrity can be used as another flag for the validity of the α estimate.

8 | CONCLUSION AND FUTURE OUTLOOK

A dual pitot tube ADS is designed with fault detection and isolation capability for small UAS. The purpose of this algorithm is to provide a reliable ADS and recovery strategy for safe drone operations in case of the pitot tube water-blockage fault under rainy and foggy conditions. The fault detection algorithm is designed to detect faults based on the given integrity requirements and known water-blockage fault profile.



(a) Angle-of-attack α estimate, 2 σ uncertainy bound, protection level and alert limit.



(b) Sideslip β estimate, 2 σ uncertainy bound, protection level and alert limit.

FIGURE 17 Flow angle α and sideslip β estimate, uncertainty bound, protection level, and alert limit [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

Systematic procedures and various factors are laid out to show how to design the fault detection algorithm. Though the stringent performance requirement, such as 10^{-7}, may not be realizable to certify ADS in small UAS due to the low-cost sensors onboard, this IM approach allows engineers to assess the performance of the FDI algorithm from the requirement point of view. The high-performance requirements can potentially be achieved for the ADS in small UAS when highly accurate sensors (e.g., good GPS and IMU), good control design (reject external disturbances), and sensible path planning (observability-aware trajectory design) are employed.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Minnesota Invasive Terrestrial Plants and Pests Center (MITPPC) through the Minnesota Environment and Natural Resources Trust Fund for financial support to conduct research associated with increasing the reliability of small UAV technology used for surveying applications. The authors also gratefully acknowledge Todd Colten and Sentera LLC for donating the flight data for the air data fault detection research.

ORCID

Kerry Sun https://orcid.org/0000-0002-5310-2815
Demoz Gebre-Egziabher https://orcid.org/0000-0001-6623-9602

REFERENCES

Angus, J. E. (2006). Raim with multiple faults. *NAVIGATION*, *53*(4), 249–257. https://doi.org/10.1002/j.2161-4296.2006.tb00387.x

Arana, G. D., Hafez, O. A., Joerger, M. & Spenko, M. (2019a). Recursive integrity monitoring for mobile robot localization safety. *International Conference on Robotics and Automation (ICRA)*, (pp. 305–311). https://doi.org/10.1109/ICRA.2019.8794115

Arana, G. D., Joerger, M. & Spenko, M. (2019b). Efficient integrity monitoring for kf-based localization. *International Conference on Robotics and Automation (ICRA)*, (pp. 6374–6380). https://doi.org/10.1109/ICRA.2019.8794362

Austrailian Transport Safety Bureau. (2015). *Erratic airspeed indications Boeing 787-8, VH-VKE (Technical Report)*. Austrailian Transport Safety Bureau. https://www.atsb.gov.au/media/5773029/ao-2015-149-final.pdf

Avant, T., & Morgansen, K. A. (2019). *Observability properties of object pose estimation*. American Control Conference (ACC), (pp. 5134–5140). https://doi.org/10.23919/ACC.2019.8814791

Bar-Shalom, Y., Li, X., & Kirubarajan, T. (2002). *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0471221279.ch5

Berman, Z. & Powell, J. D. (1998, April). The role of dead reckoning and inertial sensors in future general aviation navigation. In *IEEE Position Location and Navigation Symposium*, (pp. 510–517). https://doi.org/10.1109/PLANS.1998.670206

Bhamidipati, S. & Gao, G. X. (2019, September). SLAM-based integrity monitoring using GPS and fish-eye camera. *Proc. of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, (pp. 4116–4129). https://doi.org/10.33012/2019.17117

Blanch, J., Walter, T., & Enge, P. (2009). RAIM with optimal integrity and continuity allocations under multiple failures. *IEEE Transactions on Aerospace and Electronic Systems*, *46*(3), 1235–1247. https://doi.org/10.1109/TAES.2010.5545186

Borup, K. T. (2018). Air Data Estimation for Small Unmanned Aircraft. (Doctoral dissertation, Norwegian University of Science and Technology). http://hdl.handle.net/11250/2565573

Brown, R. G., & Chin, G. (1998). *GPS RAIM: Calculation of thresholds and protection radius using chi-square methods ; a geometric approach*. The Institute of Navigation Monograph Series.

Harbo, A. L. Cour- (2017). Quantifying risk of ground impact fatalities of power line inspection bvlos flight with small unmanned aircraft. *International Conference on Unmanned Aircraft Systems (ICUAS)*, (pp. 1352–1360). https://link.springer.com/article/10.1007/s10846-018-0853-1

Eagle Tree Systems (2020). Prandtl Pitot Tube Kit.

Eubank, R., Atkins, E., & Ogura, S. (2010). *Fault detection and fail-safe operation with a multiple-redundancy air-data system*. AIAA Guidance, Navigation, and Control Conference. https://doi.org/10.2514/6.2010-7855

Fang, S., O'Young, S., & Rolland, L. (2018). Development of small UAS Beyond-Visual-Line-Of-Sight (BVLOS) flight operations: System requirements and procedures. *Drones*, *2*, 13. https://doi.org/10.3390/drones2020013

Federal Aviation Administration. (2019). 14 AC Part 135 Air Carrier and Operator Certification. https://www.ecfr.gov/cgi-bin/text-idx?rgn=div5&node=14%3A3.0.1.1.11&se14.3.135_1144#se14.3.135_1149

Federal Democratic Republic of Ethiopia, Ministry of Transport, Aircraft Accident Investigation Bureau. (2020, March). Aircraft Accident Investigation Bureau Interim Report. http://www.aib.gov.et/wp-content/uploads/2020/documents/accident/ET-302InterimInvestigationReportMarch92020.pdf

Fravolini, M. L., del Core, G., Papa, U., Valigi, P., & Napolitano, M. R. (2017). Data-driven schemes for the robust fault detection of aircraft air data system sensors. *IEEE Transactions on Control Systems Technology*, *27*(1). https://doi.org/10.1109/TCST.2017.2758345

Fravolini, M. L., Napolitano, M. R., Core, G. D., & Papa, U. (2018, September). Experimental interval models for the robust fault detection of aircraft air data sensors. *Control Engineering Practice*, *78*, 196–212. http://www.sciencedirect.com/science/article/pii/S0967066118302326

Freeman, P. (2014). Reliability assessment for low-cost unmanned aerial vehicles. (Ph.D. thesis, University of Minnesota). https://hdl.handle.net/11299/170136

Freeman, P., Seiler, P., & Balas, G. J. (2013). Air data system fault modeling and detection. *Control Engineering Practice*, *21*(10), 1290–1301. https://doi.org/10.1016/j.conengprac.2013.05.007

Gleason, S., & Gebre-Egizabher, D. (2009). *GNSS Applications and Methods, Chapter 6,7*. Norwood, MA: Artech Houser.

Grosch, A., Crespillo, O. G., Martini, I. & Günther, C. (2017). Snapshot residual and Kalman filter based fault detection and exclusion schemes for robust railway navigation. *European Navigation Conference (ENC)*, (pp. 36–47). https://doi.org/10.1109/EURONAV.2017.7954171

Guo, D., Zhong, M., & Zhou, D. (2018). Multisensor data-fusion-based approach to airspeed measurement fault detection for unmanned aerial vehicles. *IEEE Transactions on Instrumentation and Measurement*, *67*(2), 317–327. https://doi.org/10.1109/TIM.2017.2735663

Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*. John Wiley & Sons, Inc.

Hansen, S., & Blanke, M. (2014). Diagnosis of airspeed measurement faults for unmanned aerial vehicles. *IEEE Transactions on Aerospace and Electronic Systems*, *50*(1), 224–239. https://doi.org/10.1109/TAES.2013.120420

Hu, B., & Seiler, P. (2015). A probabilistic method for certification of analytically redundant systems. *International Journal of Applied Mathematics and Computer Science*, *25*(1), 103–116. http://eudml.org/doc/270194

Isermann, R. (2005). *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer Berlin Heidelberg. https://books.google.com/books?id=6yUfoZhGMY0C

JDrones. (2020). Pitot Tube and Silicon Tube. Last accessed: May 29, 2020. http://store.jdrones.com/AirSpeed_Pitot_tube_for_UAV_p/senpito01.htm

Jiang, Y., & Wang, J. (2014). A new approach to calculate the vertical protection level in A-RAIM. *Journal of Navigation*, *67*(4), 711–725. https://doi.org/10.1017/S0373463314000204

Joerger, M., Chan, F., & Pervan, B. (2014). Solution separation versus residual-based raim. *NAVIGATION*, *61*(4), 273–291. https://doi.org/10.1002/navi.71

Joerger, M., & Pervan, B. (2013). Kalman filter-based integrity monitoring against sensor faults. *Journal of Guidance, Control, and Dynamics*, *36*(2), 349–361. https://doi.org/10.2514/1.59480

Johnson, M., Jung, J., Rios, J., Mercer, J., Homola, J., Prevot, T., Mulfinger, D., & Kopardekar, P. (2017). *Flight test evaluation of an unmanned aircraft system traffic management (utm) concept for multiple beyond-visual-line-of-sight operations*. Twelfth USA/Europe Air Traffic Management Research and Development Seminar.

Kassas, Z. M., & Humphreys, T. E. (2014). Observability analysis of collaborative opportunistic navigation with pseudorange measurements. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 260–273. https://doi.org/10.1109/TITS.2013.2278293

Khanafseh, S., Roshan, N., Langel, S., Chan, F., Joerger, M., & Pervan, B. (2014). *GPS spoofing detection using RAIM with INS coupling, (pp. 1232–1239)*. IEEE/ION Position, Location and Navigation Symposium. https://doi.org/10.1109/PLANS.2014.6851498

Komite Nasional Keselamatan Transportasi, Republic of Indonesia. (2018). *Aircraft accident investigation report - pt. lion mentari airlines boeing 737-8 (max) pk-lqp tanjung karawang, west java republic of indonesia (Technical Report)*. Komite Nasional Keselamatan Transportasi. https://reports.aviation-safety.net/2018/20181029-0_B38M_PK-LQP_PRELIMINARY.pdf

Kotikalpudi, A., Danowsky, B. P., & Seiler, P. J. (2020). *Reliability analysis for small unmanned air vehicle with algorithmic redundancy*. AIAA Scitech Forum. https://doi.org/10.2514/6.2020-0739

Lee, Y. C. (1986). Analysis of range and position comparison methods as a means to provide GPS integrity in the user receiver. *Proc. of the 42nd Annual Meeting of The Institute of Navigation*. https://www.ion.org/publications/abstract.cfm?articleID=12197

Lie, F. A. P., & Gebre-Egziabher, D. (2013). Synthetic air data system. *Journal of Aircraft*, *50*(4), 1234–1249. https://doi.org/10.2514/1.C032177

Lu, P., Van Eykeren, L., Van Kampen, E., De Visser, C. C., & Chu, Q. P. (2016). Adaptive three-step kalman filter for air data sensor fault detection and diagnosis. *Journal of Guidance, Control, and Dynamics*, *39*(3), 590–604. https://doi.org/10.2514/1.G001313

McCrink, M., & Gregory, J. W. (2018). Design and development of a high-speed uas for beyond line-of-sight operation. *AIAA Information Systems-AIAA Infotech at Aerospace*. https://doi.org/10.2514/6.2018-0750

Milner, C. D., & Ochieng, W. Y. (2011). Weighted raim for apv: The ideal protection level. *Journal of Navigation*, *64*(1), 61–73. https://doi.org/10.1017/S0373463310000342

Ossmann, D., Joos, H., & Goupil, P. (2017). Enhanced sensor monitoring to maintain optimal aircraft handling in case of faults. *Journal of Guidance, Control, and Dynamics*, *40*(12), 3127–3137. https://doi.org/10.2514/1.G002341

Parkinson, B., & Axelrad, P. (1988, August). Autonomous GPS integrity monitoring using the pseudorange residual. *NAVIGATION*, *35*(2), 255–274. https://doi.org/10.1002/j.2161-4296.1988.tb00955.x

Pervan, B. (1996). *Navigation integrity for aircraft precision landing using the global positioning system*. (Ph.D. thesis, Stanford University). https://web.stanford.edu/group/scpnt/gpslab/pubs/theses/BorisPervanThesis96.pdf

Pervan, B., Pullen, S., & Christie, J. (1998). A multiple hypothesis approach to satellite navigation integrity. *NAVIGATION*, *45*(1), 61–84. https://doi.org/10.1002/j.2161-4296.1998.tb02372.x

Rohloff, T. J., Whitmore, S. A., & Catton, I. (1999). Fault-tolerant neural network algorithm for flush air data sensing. *Journal of Aircraft*, *36*(3), 541–549. https://doi.org/10.2514/2.2489

SeekingAlpha. (2019). Here Are The Disturbing Internal Emails Boeing Just Released About The 737 MAX. https://seekingalpha.com/article/4316117-are-disturbing-internal-emails-boeing-just-released-737-max. Last accessed: June 06, 2020. https://seekingalpha.com/article/4316117-are-disturbing-internal-emails-boeing-just-released-737-max

Sturza, M. A. (1988). Navigation system integrity monitoring using redundant measurements. *NAVIGATION*, *35*(4), 483–501. https://doi.org/10.1002/j.2161-4296.1988.tb00975.x

Summers, T. H., Cortesi, F. L., & Lygeros, J. (2016). On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, *3*(1), 91–101. https://doi.org/10.1109/TCNS.2015.2453711

Sun, K., Regan, C. D. & Egziabher, D. G. (2018). GNSS/INS based estimation of air data and wind vector using flight maneuvers. In *2018 IEEE/ION Position, Location, and Navigation Symposium (PLANS)*, (pp. 838–849). https://doi.org/10.1109/PLANS.2018.8373461

Sun, K., Regan, C. D., & Gebre-Egziabher, D. (2019a). *A GNSS/IMU-based 5-hole pitot tube calibration algorithm*. AIAA Scitech Forum. https://doi.org/10.2514/6.2019-0360

Sun, K., Regan, C. D., & Gebre-Egziabher, D. (2019b). Observability and performance analysis of a model-free synthetic air data estimator. *Journal of Aircraft*, *56*(4), 1471–1486. https://doi.org/10.2514/1.C035290

Tanıl, C., Khanafseh, S., Joerger, M., & Pervan, B. (2017a). An INS monitor to detect GNSS spoofers capable of tracking vehicle position. *IEEE Transactions on Aerospace and Electronic Systems*, *54*(1), 131–143. https://doi.org/10.1109/TAES.2017.2739924

Tanıl, C., Khanafseh, S., & Pervan, B. (2017b). Detecting global navigation satellite system spoofing using inertial sensing of aircraft disturbance. *Journal of Guidance, Control, and Dynamics*, *40*(8), 2006–2016. https://doi.org/10.2514/1.G002547

Van Eykeren, L., & Chu, Q. P. (2014). Sensor fault detection and isolation for aircraft control systems by kinematic relations. *Control Engineering Practice*, *31*, 200–210. http://www.sciencedirect.com/science/article/pii/S0967066114000884

Vascik, P. D., Hansman, R. J., & Dunn, N. S. (2018). Analysis of urban air mobility operational constraints. *Journal of Air Transportation*, *26*(4), 133–146. https://doi.org/10.2514/1.D0120

Walter, T. & Enge, P. (1995). Weighted RAIM for precision approach. *Proc. of the 8th International Technical Meeting of the Satellite Division of The Institute of Navigation*. pp. 1995-2004. https://www.ion.org/publications/abstract.cfm?articleID=2524

Walter, T., Enge, P., Blanch, J., & Pervan, B. (2008). Worldwide vertical guidance of aircraft based on modernized GPS and new integrity augmentations. *Proceedings of the IEEE*, *96*(12), 1918–1935. https://doi.org/10.1109/JPROC.2008.2006099

Xing, Z. (2010). *Over-Bounding Integrated INS/GNSS Output Errors*. (Ph.D. thesis, University of Minnesota Twin Cities). https://hdl.handle.net/11299/101401

Yapp, J., Seker, R. & Babiceanu, R. (2018). Providing accountability and liability protection for UAV operations beyond visual line of sight. IEEE Aerospace Conference, (pp. 1–8). https://doi.org/10.1109/AERO.2018.8396532

Yeh, Y. C. (1996). Triple-triple redundant 777 primary flight computer. *IEEE Aerospace Applications Conference. Proceedings*, *1*, 293–307. https://doi.org/10.1109/AERO.1996.495891

How to cite this article: Sun, K., & Gebre-Egziabher, D. (2021). Air data fault detection and isolation for small UAS using integrity monitoring framework. *NAVIGATION*, *68,* 577–600. https://doi.org/10.1002/navi.440

APPENDIX A: BATCH LINEAR SYSTEM REALIZATION

In this appendix, we present one form of the batch linear system realization to faciliate the connection between the LTV observability matrix and the residual-based test statistic shown in Section 5. We start with a model of the discrete linear dynamic system with an unknown additive fault:

$$\begin{aligned}\mathbf{z}_k &= \mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k + \mathbf{f}_k\\ \mathbf{x}_{k+1} &= \boldsymbol{\Phi}_k\boldsymbol{x}_k + \boldsymbol{\Gamma}_k\mathbf{w}_k\end{aligned} \tag{A1}$$

where $\mathbf{f}_k$ is the deterministic additive fault vector, and $\mathbf{x}_k$, $\mathbf{z}_k$, $\mathbf{w}_k$ and $\boldsymbol{v}_k$ are defined in Section 2.2. The vector $\mathbf{w}_k$ is the process noise vector and is assumed to follow $N(\mathbf{0}, \mathbf{R}_w)$. The matrices $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Gamma}_k$ are the state transition matrix and discrete noise coefficient matrix, respectively.

Only additive faults are chosen for this study to analyze the feasibility of air data fault detection capability. Nonlinear fault with more sophisticated models should be considered in the future.

Since there are not enough redundant airspeed measurements at each time step k, we obtain the following batch realization by stacking all the measurement vectors from the past time $k-q$ to the current time step k in Equation (A2):

$$\begin{aligned}\mathbf{Z}_{k-q:k} &= \boldsymbol{\mathcal{O}}_{k-q:k}\boldsymbol{x}_{k-q} + \mathbf{Q}_{w,k-q:k}\mathbf{W}_{k-q:k}\\ &\quad + \mathbf{V}_{k-q:k} + \mathbf{F}_{k-q:k}\end{aligned} \tag{A2}$$

The matrices $\mathbf{Z}_{k-q:k}$, $\mathbf{W}_{k-q:k}$, $\mathbf{V}_{k-q:k}$, $\mathbf{F}_{k-q:k}$, $\boldsymbol{\mathcal{O}}_{k-q:k}$ and $\mathbf{Q}_{w,k-q:k}$ are defined in the following:

$$\mathbf{Z}_{k-q:k} = \begin{bmatrix}\mathbf{z}_{\mathbf{k-q}}^{\mathbf{T}} & \mathbf{z}_{k-q+1}^{T} & \dots & \mathbf{z}_{k}^{T}\end{bmatrix}^{T} \tag{A3}$$

$$\mathbf{W}_{k-q:k} = \begin{bmatrix}\mathbf{w}_{\mathbf{k-q}}^{\mathbf{T}} & \mathbf{w}_{k-q+1}^{T} & \dots & \mathbf{w}_{k}^{T}\end{bmatrix}^{T} \tag{A4}$$

$$\mathbf{V}_{k-q:k} = \begin{bmatrix}\mathbf{v}_{\mathbf{k-q}}^{\mathbf{T}} & \mathbf{v}_{k-q+1}^{T} & \dots & \mathbf{v}_{k}^{T}\end{bmatrix}^{T} \tag{A5}$$

$$\mathbf{F}_{k-q:k} = \begin{bmatrix}\mathbf{f}_{\mathbf{k-q}}^{\mathbf{T}} & \mathbf{f}_{k-q+1}^{T} & \dots & \mathbf{f}_{k}^{T}\end{bmatrix}^{T} \tag{A6}$$

$$\boldsymbol{\mathcal{O}}_{k-q:k} = \begin{bmatrix}\mathbf{H}_{\mathbf{k-q}}\\ \mathbf{H}_{k-q+1}\boldsymbol{\Phi}_{k-q}\\ \vdots\\ \mathbf{H}_{k}\boldsymbol{\Phi}_{k-1}\dots\boldsymbol{\Phi}_{k-q}\end{bmatrix} \tag{A7}$$

$$\mathbf{Q}_{w,k-q:k} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}_{k-q+1}\boldsymbol{\Gamma}_{k-q} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}_{k-q+2}\boldsymbol{\Phi}_{k-q+1}\boldsymbol{\Gamma}_{k-q} & \mathbf{H}_{k-q+2}\boldsymbol{\Gamma}_{k-q+1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{H}_k\boldsymbol{\Phi}_{k-1}\dots\boldsymbol{\Phi}_{k-q+1}\boldsymbol{\Gamma}_{k-q} & \mathbf{H}_k\boldsymbol{\Phi}_{k-2}\dots\boldsymbol{\Phi}_{k-q+1}\boldsymbol{\Gamma}_{k-q+1} & \dots & \mathbf{H}_k\boldsymbol{\Gamma}_{k-1} & \mathbf{0} \end{bmatrix} \tag{A8}$$

The matrix $\boldsymbol{\mathcal{O}}_{k-q:k}$ is the discrete LTV observability matrix over a sliding window. This batch realization [extension of the linear time-invariant system in (Isermann, 2005)] formulation connects the stacked measurements $\mathbf{Z}_{k-q:k}$ to the past state vector $\boldsymbol{x}_{k-q}$ and its associated observability matrix $\boldsymbol{\mathcal{O}}_{k-q:k}$ nicely.

Batch realization is not unique; Joerger and Pervan (2013) define a different formulation where the past measurements are a function of all the past states. However, the measurement model in Joerger and Pervan (2013) is not explicitly expressed as a function of the observability matrix.

APPENDIX B: PROOF OF MATRIX EQUALITY

In this appendix, we prove the following equation is true:

$$(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) = \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) \tag{B1}$$

where $\boldsymbol{\mathcal{O}}$ is the observability matrix shown in Equation (A7), $\boldsymbol{\mathcal{O}}^*$ is left pseudoinverse of $\boldsymbol{\mathcal{O}}$ shown in Equation (21), and $\boldsymbol{\Sigma}$ is the weighting matrix shown in Equation (18). Note that $\boldsymbol{\Sigma}$ is symmetric by construction, that is, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$

Proof. First, we show $\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)$ is symmetric:

$$\begin{aligned}
\left(\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)\right)^T &= (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T\boldsymbol{\Sigma}^{-1} \\
&= (\mathbf{I} - (\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T)\boldsymbol{\Sigma}^{-1} \\
&= (\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} \\
&= \left(\mathbf{I} - \left(\left(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}\right)^{-1}\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\right)^T\boldsymbol{\mathcal{O}}^T\right)\boldsymbol{\Sigma}^{-1} \\
&= \left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-T}\boldsymbol{\mathcal{O}}^T\right)\boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-T}\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-1}\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^* \\
&= \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)
\end{aligned} \tag{B2}$$

where the 7th equality is achieved because $\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}}$ is also symmetric.

Now we have the following fact:

$$(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) \tag{B3}$$

With the equality above, we can express Equation (B1) as follows:

$$\begin{aligned}
(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) &= (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} \\
&= (\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1}
\end{aligned} \tag{B4}$$

Now we proceed to prove the following: $(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) = \mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T$.

This can be shown to be true by moving the left-hand side of equation to the right as follows:

$$\begin{aligned}
(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) - (\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) &= 0 \\
(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \mathbf{I} + \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) &= 0 \\
(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) &= 0
\end{aligned} \tag{B5}$$

We also realize the following equation is true:

$$\begin{aligned}
\boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T &= \left((\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-1}\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\right)^T\boldsymbol{\mathcal{O}}^T \\
&= \boldsymbol{\Sigma}^{-T}\boldsymbol{\mathcal{O}}(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-T}\boldsymbol{\mathcal{O}}^T \\
&= \boldsymbol{\Sigma}^{-T}\boldsymbol{\mathcal{O}}(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-1}\boldsymbol{\mathcal{O}}^T \\
&= \boldsymbol{\Sigma}^{-T}\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\mathcal{O}}^{-T}\boldsymbol{\mathcal{O}}^T \\
&= \mathbf{I}
\end{aligned} \tag{B6}$$

Note that $(\boldsymbol{\mathcal{O}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{O}})^{-1} = \boldsymbol{\mathcal{O}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\mathcal{O}}^{-T}$ is justified because $\boldsymbol{\mathcal{O}}^T$ is right-invertible. However, the right inverse (not Moore-Penrose) of $\boldsymbol{\mathcal{O}}^T$ is not unique.

By substituting Equation (B6) into Equation (B5), we show that $(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T) = \mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T$ is indeed true. Finally, we arrive the matrix property that we want to prove:

$$\begin{aligned}
(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*) &= (\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)^T(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} \\
&= (\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)(\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} \\
&= (\mathbf{I} - \boldsymbol{\mathcal{O}}^{*T}\boldsymbol{\mathcal{O}}^T)\boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{O}}^*)
\end{aligned} \tag{B7}$$

ML 2015, Ch 76 Art 2, Sec 6a Project Abstract
For the Period Ending December 31, 2021

PROJECT TITLE: **Subproject #5: Optimizing tree injections against emerald ash borer**
PROJECT MANAGER: **Brian Aukema**
AFFILIATION: **University of Minnesota**
MAILING ADDRESS: **1980 Folwell Avenue**
CITY/STATE/ZIP: **St. Paul, MN 55118**
PHONE: **(612) 624-1847**
E-MAIL: **BrianAukema@umn.edu**
WEBSITE:
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER
ML 2015, Ch. 76, Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $320,000
AMOUNT SPENT: $320,000
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results

Emerald ash borer continues to spread and devastate Minnesota's urban forests, but deploying the right types of insecticides to ash trees in the right ways can offer tree conservation and protection with minimal risk to non-target organisms such as bees that visit flowers and worms that decompose leaves.

Overall Project Outcome and Results

Emerald ash borer is an invasive insect that kills mature ash trees and has been spreading within Minnesota since its detection in 2009. Ash is a major component of many of Minnesota's urban forests. Injections of insecticides into ash trees can preserve trees indefinitely, but raises concerns for non-target organisms in the environment such as bees and earthworms. For this study, we injected subsets of 1200 trees located in eight different cities in Minnesota with two different insecticides. We specifically tested products that were not neonicotinoids that have presented past risks to pollinators. Insecticides were injected into the trunks in summer of 2017, with periodic reapplications until 2020 while we measured crown health of each tree each summer until 2021. The original site selections were in cities with low pressure from emerald ash borer. We found over the four years of the study that injecting only half of the trees in a given site gave good protection to all trees. We were unable to determine, however, whether this associational protection (i.e., preservation of canopy in an untreated tree when proximate to a treated tree), winter mortality to EAB, or some combination of both was responsible for the site-wide excellent conditions that persisted five years after EAB was present in these communities. Measurements of tree phenology such as leaf out and leaf drop showed that insecticides did not alter the timing of such events. One of the insecticides, emamectin benzoate, showed excellent protection of ash seeds against seed weevils by the third year of the study, without affecting seed viability. We also canvassed the insect communities that visited the trees and harvested leaves for feeding trials with nontarget organisms, and measured chemical concentrations in the leaves. We found that insects communities were similar between treated versus untreated trees across seasons, that bees preferred visiting synchronously flowering plants such as flowering crab apples and rhododendrons versus ash trees, that trunk-injected chemicals were not reliably detected in all plant parts after injection, and that invertebrates such as worms showed no reduction in reproduction or feeding on treated leaves. As such, we concluded that detrimental effects of the insecticides tested on non-target organisms are not likely to be ubiquitous or widespread. In summary, when homeowners or communities are selecting a product to preserve urban ash trees, we recommend emamectin benzoate as a suitable and effective alternative to neonicotinoid-based products.

Project Results Use and Dissemination

This work has been submitted for publication at two peer-reviewed journals, with two more submissions planned. The work has been presented at regional, national, and international venues including workshops and conferences such as the Shade Tree Short Course, the Entomological Societies of Canada and America, the IUFRO Conference on Biological Invasions of Forests, the North American Forest Insect Work Conference, the Upper Midwest Invasive Species Conference, the USDA Interagency Annual Forum, and the North Central Forest Pest Workshop. A number of presentations were also given to local community forestry and resource manager groups throughout the project, and we enjoyed a high number of interactions with members of the public while working in their communities.

Forest Ecology and Management 511 (2022) 120144

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Effects of systemic insecticides against emerald ash borer on ash seed resources

Dorah M. Mwangola [a,*], Aubree M. Kees [a], Donald M. Grosman [b], Brian H. Aukema [a]

[a] Department of Entomology, University of Minnesota, St. Paul, MN, United States
[b] Arborjet Inc., Woburn, MA, United States

ARTICLE INFO

Keywords:
Emerald ash borer
Agrilus planipennis
Ash tree
Fraxinus
Ash seed weevil
Lignyodes spp.
Systemic insecticides
Conservation of genetic resources

ABSTRACT

Emerald ash borer (EAB), *Agrilus planipennsis*, is an invasive insect that was accidentally introduced to North America from Asia. It continues to spread rapidly across North America and is responsible for the death of tens of millions of ash trees (*Fraxinus* spp.). All North American species of ash are susceptible to EAB infestation threatening the ash resource and diversity. Measures such as systemic insecticide treatments in urban forests and collection of ash seeds provide a means of conserving genetic resources. Understanding the effect these insecticides could have on ash seed viability is therefore important to informing conservation efforts. Another potential concern for effective conservation of ash seeds is the ash seed weevil (*Lignyodes* spp.) whose larvae develop in and feed on ash seeds. However, the effect of EAB insecticides on weevil infestation levels in ash seeds has not been investigated to date. Our study investigated the effect of two systemic insecticide treatments, azadirachtin and emamectin benzoate, on levels of ash seed weevil infestation, seed germination ability, and seed germination time of seeds collected from boulevard trees of green ash (*F. pennsylvanica* Marsh.) in cities in Minnesota from 2017 to 2019. Weevil infestation levels were similar between untreated and treated trees in 2017 and 2018. In 2019, the weevil prevalence in untreated trees was on average 17% and 30% higher than in azadirachtin and emamectin benzoate-treated trees respectively. Weevil infestation data suggests that repeated insecticide treatments at labelled rates can reduce seed weevils that target germplasm. Additionally, insecticide treatments did not affect ash seed germination rates between treatments. These results suggest that systemic insecticides may be effective at conserving the seed resource in addition to known benefits such as canopy preservation.

1. Introduction

The emerald ash borer, *Agrilus planipennsis* Fairmaire (Coleoptera: Buprestidae) is a wood boring beetle accidentally introduced to North America from Asia on wood packing material (Cappaert et al., 2005). Hundreds of millions of ash trees (*Fraxinus* spp.) have died from emerald ash borer infestation (USDA Forest Service and Michigan State University, 2022). Ash trees are of cultural importance to indigenous peoples, provide homes for wildlife, act as nutrient sources for other arthropods, and are used for wood products such as baseball bats and furniture. In urban centers, *Fraxinus* species are popular street trees that foster increased property value and human health benefits (Gandhi and Herms, 2010a, 2010b; Donovan et al., 2013; Herms and McCullough, 2014). In natural forests, emerald ash borer causes reproductive isolation of surviving ash trees, reduces seed production in infested trees, and slows regeneration rates of new saplings (Kashian and Witter, 2011; Klooster et al., 2014; O'Brien et al., 2017).

A combination of both *in situ* and *ex situ* measures have been employed in conserving ash genetic resource in the wake of the invasion of emerald ash borer to North America. The *in situ* approach is the use of systemic insecticide treatments that kill larvae feeding underneath the bark and adults that feed on leaves after emergence. This approach

* Corresponding author at: Department of Entomology, University of Minnesota, St. Paul, MN, 219 Hodson Hall, 1980 Folwell Ave, Falcon Heights, MN 55108, United States.
E-mail address: mwang022@umn.edu (D.M. Mwangola).

https://doi.org/10.1016/j.foreco.2022.120144
Received 24 December 2021; Received in revised form 25 February 2022; Accepted 3 March 2022
Available online 16 March 2022
0378-1127/© 2022 Elsevier B.V. All rights reserved.

maintains mature trees on the landscape, reducing the impacts of the loss of ash trees (O'Brien et al., 2017; Flower et al., 2018). The *ex situ* approach entails the collection of ash seeds for seed banks and breeding material (Widrlechner, 2010; Koch et al., 2011).

The conservation of genetic resources is an important component of sustainable management of forests especially in the face of environmental, pest and disease challenges facilitated by climate change and globalization (Rajora and Mosseler, 2001; St.Clair and Howe, 2011; Ramsfield et al., 2016). A primary goal in the genetic conservation of forest trees is the preservation of genetic diversity among populations. Genetic diversity enables the continued survival and success of a population by providing a population the ability to adapt and evolve in response to changes in the environment or biotic threats such as pests and diseases (Rajora and Mosseler, 2001; Šijačić-Nikolić et al., 2014). Furthermore, genetic variation in forest tree populations is a cumulation of years of different evolutionary processes and can be difficult to recreate if lost (Rajora and Mosseler, 2001). Genetic resources are also important in future breeding efforts for desirable/valuable traits such as resistance to diseases and pests (St.Clair and Howe, 2011; Woodcock et al., 2019).

The types of conservation measures for genetic resources can be grouped into two types: *in situ* and *ex situ* conservation. *In situ* (in site) involves conserving a species in its native habitat, e.g., conserving mature trees in natural populations or in protected parks or setting up artificial populations without controlled selection. *Ex situ* (out of site) conservation involves conserving the species away from its native habitat; e.g. seed bank collections, pollen bank collections and seedlings maintained in plantations (Ledig et al., 1998; Šijačić-Nikolić et al., 2014). The benefits of using *in situ* measures include maintaining a species' role in its environment, allowing for natural evolutionary forces to continue to act on the species' population. Such measures can be more affordable than maintaining *ex situ* populations. *Ex situ* measures provide protection for populations at risk in their natural habitat until resistance or tolerance is obtained from a pest or disease, but may limit the amount of diversity that can be conserved. Given both the pros and cons of each strategy, a cautious conservation strategy would implement both measures (Ledig, 1986; Rajora and Mosseler, 2001; St.Clair and Howe, 2011; Pritchard et al., 2014).

While much focus continues to be on emerald ash borer as a primary tree killer, ash seed weevils, *Lignyodes* spp. (Coleoptera: Curculionidae), feed on seeds of trees belonging to the Oleaceae such as *Fraxinus* spp. and *Syringa* (lilac) (Barger and Davidson, 1967; Clark, 1980; Dix, 1986). These insects are found in North America and Eurasia (Clark et al., 1977; Clark, 1980, 1982; Haran et al., 2019) and could affect the effective conservation of ash seed resources (Solomon et al., 1993; Knight et al., 2010). *Lignyodes bischoffi* (Blatchley), for example, is widespread in North America, including the Great Lakes region, and is known to infest seeds of green ash (*F. pennsylvanica* Marsh.), black ash (*F. nigra* Marsh.) and white ash (*F. americana* Linnaeus) (Clark, 1980). Additionally, *L. bischoffi* weevils are invasive in Europe and have been observed infesting introduced green ash, European ash (*F. excelsior* Linnaeus), manna ash (*F. Ornus* Linnaeus) and the narrow-leaved ash (*F. angustifolia* subsp. *oxycarpa* Wild.) (Poiras, 1998; Wanat and Mocarski, 2008; Arzanov, 2013). Female ash seed weevils oviposit a single egg into a single ash seed encased in its fruiting body (samara) in the summer or fall, whereupon larvae hatch and feed on the seed contents. Larvae emerge from the seed in the summer, fall or spring and burrow into the soil. There, they pupate and emerge as adults in the subsequent summer (Barger and Davidson, 1967; Solomon et al., 1993).

The effect of merging the two *in situ* and *ex situ* measures, i.e. insecticide treatment of ash trees and seed collection, has not been investigated. The primary aim of this study was to analyze the effect of systemic insecticides on ash seed resource collected in cities in Minnesota. The specific objectives were to analyze the effect of two systemic insecticides against emerald ash borer, i.e. emamectin benzoate and azadirachtin, on (i) ash seed weevil prevalence and (ii) ash seedling germination time and ability.

2. Materials and methods

2.1. Sites

In July 2017, twelve urban sites with trees with visible signs and symptoms of early-stage infestation of emerald ash borer such as crown thinning and epicormic shoot growth were selected in Minnesota, USA (Table 1). At each site, we selected mature boulevard green ash trees of *F. pennsylvanica* with more than 70% crown present and administered one of two different systemic insecticides; emamectin benzoate (n=395 trees; Tree-äge G4®, Arborjet, Woburn, MA) or azadirachtin (n=200 trees; AzaSol®, Arborjet, Woburn, MA). A total of 678 other trees within these sites were left as untreated controls. Treated trees were assigned according to a gradient, but trees assigned to this study were randomly selected. The ratio of treated: untreated trees in each site was 50:50, and each site contained approximately 100 trees (Table 1). The insecticides were administered using a pressurized injection system (the QUIK-jet AIR® tree injection system, Arborjet, Woburn, MA). Each treated tree received a dose of 0.2 g AI per 2.54 cm diameter at breast height (DBH). The mean (±SE) DBH of all trees in the study sites from which seeds were collected from was 44 ±1.0 cm and was similar across the three treatments ($ANOVAF_{2,178}$=1.27, P=0.28).

Insecticide treatments were carried out from 26 to 30 June 2017 and 31 July to 8 August 2017 for emamectin benzoate and azadirachtin respectively. In subsequent years treatments were reapplied according to the manufacturer's instructions. The trees receiving emamectin benzoate were re-treated 26 to 28 June 2019 on an every other year schedule and the trees receiving azadirachtin were re-treated 4 to 8 June 2018, 10 to 17 June 2019 and 15 to 25 June 2020 on an every-year schedule.

Table 1
Cities in central and southeastern Minnesota, USA with experimental sites, 2017–2020, with the number of emamectin benzoate-treated, azadirachtin-treated and untreated ash trees.

Insecticide	City	Number of trees	
		Treated	Untreated
Emamectin benzoate	St. Paul	50	65
	Roseville	24	33
	Rochester	48	60
	Eagan	63	50
	Mendota Heights	40	44
	Maple Grove	65	58
	Lake City	60	65
	Coon Rapids	45	41
Azadirachtin	St. Paul	50	91
	Eagan	50	55
	Maple Grove	50	54
	Lake City	50	53

Table 2
The number of ash trees from which seeds were sampled for each treatment in October 2017, 2018 and 2019 from across eight cities in Minnesota, USA.

Year	Azadirachtin	Emamectin benzoate	Untreated	Total
2017	15	7	20	42
2018	31	45	52	128
2019	29	48	58	135

2.2. Ash seed collection

Ash seeds are contained in fruiting bodies called samaras that develop from fertilized female ash flowers. Ash seeds were collected in fall from ash trees when the samaras had changed from green to yellow to brown from 28 September to 26 October 2017(n=42 trees), 12 to 28 October 2018 (n=128 trees) and 10 to 28 October 2019 (n=135 trees). All the green ash trees in the 12 study sites were surveyed for seeds and seeds were sampled from all trees that had seeds in that year. The number of trees sampled for each treatment are shown in Table 2. A pole pruner was used to cut branches with seeds and then seeds were placed in brown paper bags and stored at room temperature.

2.3. Ash seed weevil infestation level in ash seeds

In December 2017, 2018, and 2019, a random collection of n=10 seeds for each of six trees treated with azadirachtin, emamectin benzoate, or untreated were peeled to calculate the frequency of seeds with ash seed weevil larvae. We did not try to determine if the larvae were alive (and potentially in diapause) or dead, as it was difficult to remove the seed casing without injuring or killing the insects. Generalized linear mixed effect models were used to analyze the effect of insecticide treatment, a categorical variable, weevil prevalence (i.e., the proportion of seeds that contained ash weevil larvae) as the binomial response variable on a per seed basis. A term for individual trees was incorporated as a random effect with a logit link function and binomial family specified in the model. The effect of treatment on weevil prevalence was analyzed with ANOVAs and, where treatment effects were observed, pairwise differences between treatments were explored using Tukey's means comparison test using the multcomp package in R (Hothorn et al., 2008). All data analysis was conducted in R (R Core Team, 2018).

2.4. Ash seed germination assay

Seeds were randomly selected from trees from each treatment (n=15 trees) and peeled in December 2017, 2018 & 2019 and stored in brown envelopes (2.54 cm L by 2.54 cm W) at room temperature until spring the subsequent year. On 27 March 2018, 2 April 2019 and 2 April 2020 the peeled seeds were placed in small (5.08 cm W by 7.62 cm H) clear Ziploc bags containing moist soil (Sun Gro® Horticulture, growing mix OMRI, Agawam, MA), which contains Canadian sphagnum peat moss, coarse perlite and dolomite lime. The Ziploc bags were left in refrigerator at 4 degrees Celsius for 90 days to allow the seeds to undergo cold stratification. In July, a subset of the seeds (n=6/tree in 2018 and n=5/tree in 2019 and 2020) were planted in 225-cell seedling starter trays filled with soil (Sun Gro® Horticulture, growing mix OMRI, Agawam, MA) to ¾ the height of each cell. Cell dimensions were 3.81 cm W by 5.72 cm H. The seedlings were watered once a week and covered with clear plastic bags to maintain a relative humidity of 50–60%. The trays were placed on a laboratory bench at room temperature and florescent light was used as a source of light. The first day of the germination assay was the day the seeds were planted, and the end date of the germination assay was the last day germination occurred before no new germination was observed for five consecutive days. The germination assays were concluded over a period of 30 to 34 days in the three years they were carried out; in 2018 from 10th July to 10th August (33 days), in 2019 from 5th July to 7th August (34 days) and in 2020 from 17th July to 15th August (30 days). Two events were recorded throughout the assay; germination ability, defined as whether a seedling germinated, and germination time defined as the day the cotyledon emerged above the surface of the soil.

Separate generalized linear mixed effect ANOVA models were used to analyze the effects of insecticide treatment on the response variables of seedling germination ability and time to germination. For seedling germination ability, we used a logit link function for this binomial response variable (i.e. did or did not germinate) and also incorporated a term for individual trees as a random effect. For seedling germination time, we again incorporated individual trees as a random effect and used a normal data distribution family. Analytical assumptions of linear models were examined through visual examination of residual plots. All data analysis was conducted in R (R Core Team, 2018).Where treatment effects were observed, pairwise differences between treatments were explored using Tukey's means comparison test using the multcomp package in R (Hothorn et al., 2008).

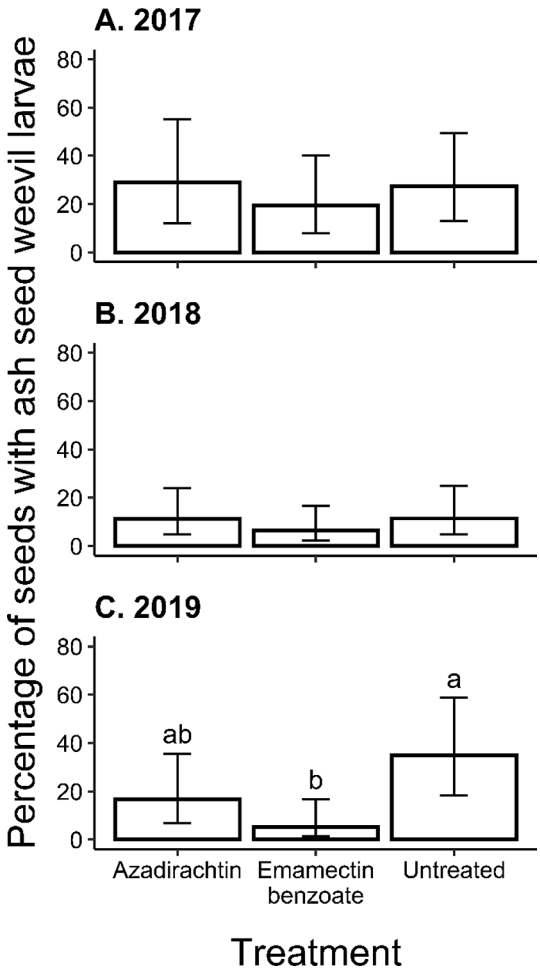


Fig. 1. Bar graphs of percentage of ash seeds with ash seed weevil larvae, i.e. weevil prevalence observed across seeds collected from ash trees treated with one of two insecticides, i.e. azadirachtin or emamectin benzoate and untreated trees in **A.** 2017 **B.** 2018 and **C.** 2019 in 8 cities in central and south east Minnesota (n=10 seeds per tree, from n=6 trees for each treatment). Error bars show 95% confidence intervals about means. Different letters indicate significant treatment differences (Tukey's means comparison test); ANOVA summaries provided in Table 3.

Table 3
Effects of insecticide treatment (azadirachtin, emamectin benzoate, or untreated ash trees) on weevil prevalence, seedling germination time, and germination ability. Seeds were examined or collected from trees in October 2017, 2018 and 2019 from across eight cities in Minnesota, USA.

Response	Figure	Year	Treatment Effect	
			χ^2_2	*P*
Weevil prevalence	1	2017	0.70	0.71
		2018	1.03	0.60
		2019	8.44	0.02
Seedling germination time	2	2018	4.41	0.11
		2019	0.82	0.66
		2020	0.99	0.61
Seedling germination ability	3	2018	1.18	0.55
		2019	0.61	0.74
		2020	5.32	0.07

3. Results

3.1. Ash seed weevil larvae prevalence in ash seeds

The ash seed weevil larval prevalence across all trees ranged from 0% to 70% in 2017, 0% to 30% in 2018 and 0% to 70% in 2019. The average weevil prevalence across all treatments in 2017 was 28% (±3.0 SE). The mean prevalence across all treatments dropped to10% (±2.0) in 2018, but then rose to 21% (±9.0) in 2019. Weevil prevalence was similar across all treatments; i.e. azadirachtin-treated trees, emamectin benzoate-treated trees and untreated trees in 2017 (Fig. 1A; Table 3) and 2018 (Fig. 1B). In 2019, the average weevil prevalence was highest in seeds collected from untreated trees (37%), followed by the trees treated with azadirachtin (20%) and then those treated with emamectin benzoate, where only 7% of the seeds contained a weevil larva. The untreated trees had a comparable weevil prevalence to azadirachtin-treated trees but were significantly greater than trees injected with emamectin benzoate (Fig. 1C).

We reared a weevil larva to an adult to aid in confirming species identity. We covered a peeled seed from an untreated tree containing a live larva with a thin layer of soil in a petri dish (50 by 9 mm, Falcon® Corning®, Corning, NY). The petri dish was placed in a 4 degrees Celsius fridge for 3 months. The petri dish was then removed from the fridge and placed at room temperature for 21 days until an adult weevil emerged. The weevil was confirmed as *Lignyodes bischoffi* (Blatchley).

3.2. Ash seed germination assay

Although we treated the seeds similarly each year, germination time was similar but percentage of seeds that germinated was variable. The germination time across all treatments, i.e. azadirachtin-treated trees, emamectin benzoate-treated trees and untreated trees was similar in all three years (Fig. 2; Table 3). Across all treatments seeds germinated in a mean time of 32 ±1 days, after planting if they were going to germinate at all.

The percentage of ash seeds that germinated (germination ability) across all treatments collected from an individual tree ranged from 0% to 100% in all three years (2018 to 2020). The average percentage of seeds that germinated across all treatments was 53% (±3.0) in 2018, it was 39% (±1.0) in 2019 and it was 81% (±4.0) in 2020. Germination ability was similar across all treatments i.e. azadirachtin-treated trees, emamectin-benzoate treated trees and untreated trees in 2018, 2019, and 2020 (Fig. 3; Table 3).

4. Discussion

Our data suggests that repeated years of the application of insecticide treatments against EAB can reduce the prevalence of seed-feeding weevils in ash trees, thus assisting in additional preservation of ash seed germplasm (Fig. 1). After three years of insecticide treatments the average weevil prevalence in 2019 was 30% and 17% higher in untreated trees than in trees treated with emamectin benzoate and azadirachtin respectively, although the reduction in weevil prevalence among trees treated with azadirachtin was not statistically significant. We postulate that this sink effect emerges over several years due to localized mortality of adult weevils that feed on treated ash leaves and limited dispersal capacity of the species. While feeding behaviour of ash weevil adults is not well elucidated in literature, the ligustrum weevil *(Ochyromera ligustri* Warner), a close relative that belongs to the same tribe (Tychiini), feeds on the seeds of the Chinese or hedge privet (*Ligustrum sinense* Lour., Oleaceae) during its larval stage and the leaves during its adult stage (Cuda and Zeller, 1998; Cuda et al., 2005). If ash seed weevils generally re-infest the same trees from year to year, a lower weevil prevalence would emerge where adult mortality accrues on trees

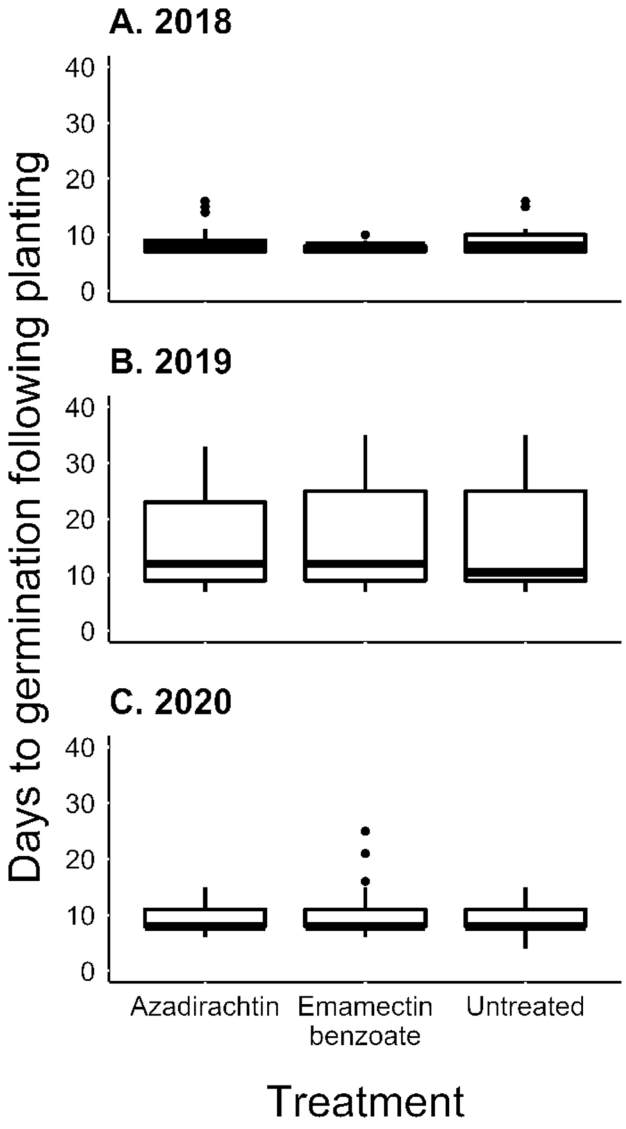


Fig. 2. Box and whisker plots of number of days to germination (i.e, the day the cotyledon emerged above surface of the soil) after cold treatment and planting. Ash seeds were collected in 8 cities in central and southeast Minnesota and planted in **A**. 2018 **B**. 2019 and **C.** 2020. The upper whisker indicates the maximum value as the third quartile added to 1.5 times the interquartile range. The lower whisker represents the minimum value as the first quartile minus 1.5 times the interquartile range. The middle line of each box and whisker plot represents the median of the data set. Dots represent data points beyond plus or minus 1.5 times the interquartile range. In 2018, n=6 seeds were planted per tree and in 2019 and 2020 n=5 seeds were planted per tree.

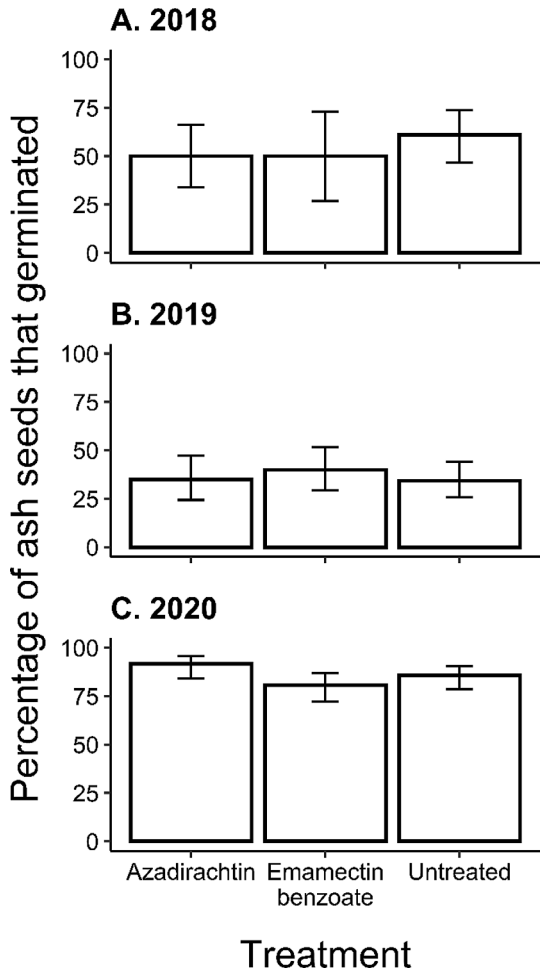


Fig. 3. Bar graphs of percentage of ash seeds that germinated from ash seeds collected from trees in 8 cities in central and southeast Minnesota and planted in **A**. 2018 **B**. 2019 and **C**. 2020. In 2018, n=6 seeds were planted per tree and in 2019 and 2020 n=5 seeds were planted per tree. Error bars show 95% confidence intervals about mean.

treated with insecticide *versus* weevil populations in untreated trees where they continue to proliferate. Localized feeding and restricted dispersal is not uncommon among seed-feeding weevils. Adults of the mango seed weevil (*Sternochetus mangiferae* Fabricius), which is a pest of mangoes in Africa, Asia, North America, Oceania, Europe and South America (EPPO, 2021), for example, usually remain on the same tree from which they had emerged such that dispersal and movement relies on transportation of seed and fruit (Subramanyam, 1926; Jarvis, 1946; Balock and Kozuma, 1964). The higher mortality noted in the emamectin benzoate vs. azadirachtin treatments could due to lower toxicity of azadirachtin, or perhaps differences in the mode of action of the two insecticides. Emamectin benzoate is a neurotoxin while azadirachtin can act as an antifeedant, growth disruptor, and/or reduce reproduction (Jansson et al., 1997; Mordue and Nisbet, 2000; Shawir et al., 2014).

Our study shows that the two systemic insecticides used against emerald ash borer, emamectin benzoate and azadirachtin, did not affect germination time and ability of ash seeds (Fig. 2; Fig. 3). Most research on phytotoxic effects of insecticides on seed germplasm have focused on direct contact of insecticides with seeds (Parween et al., 2016). In a laboratory germination study conducted by Olofinboba and Kozlowski (1982), for example, germination (i.e. cotyledon emergence) was either inhibited or completely prevented by direct application of three trunk systemic insecticides to seeds of Aleppo pine (*Pinus halepensis* Mill.). Potential mechanism for the limitation or inhibition of seed germination by insecticides include prevention of the germination of pollen that, aside from preventing seed development, could result in development of seeds void of endosperms that are thus inviable (Meyer et al., 1973; Sutherland et al., 1984; Annila and Heliövaara, 1991; Kimura et al., 1996). Furthermore, insecticides could affect the production of and activity of hormones and enzymes involved in seed germination (Singh et al., 1982; Bashir et al., 2014).

Previous work investigating the translocation of trunk injected systemic insecticides has focused on detecting insecticide presence in the bark, vascular system, roots and leaves of trees (Mota-Sanchez et al., 2009; Cevenini and Minelli, 2010; Tanis et al., 2012). In the absence of published data that systemic insecticides are translocated to the seed germplasm, we conclude that the insecticide either does not reach the seed germplasm or is present only at trace levels. Insecticide concentration can be a proponent of phytotoxic effects observed in seed germination (Werner, 1974) but we did not observe phytotoxic effects on germination time and ability between seeds from treated and untreated trees after three years of insecticide application.

Our work suggests a number of avenues for future research. First, feeding bioassays in the laboratory with adult weevils on treated foliage could ascertain mechanisms behind the field patterns of decreasing weevil prevalence observed over three years (e.g., toxicity vs. aversion). Second, little is known about dispersal capacity of the adults, and/or whether they could feed on other members of the Olaceae prior to oviposition on *Fraxinus* spp. Both of these factors could affect long-term population reduction strategies. Finally, there is little known about mortality factors of these weevils that affect their population dynamics. We noted a marked decrease in weevil abundance across the three treatments in the second year of the study (Fig. 1), but are unsure whether overwintering mortality or other causes were responsible for this pattern.

As emerald ash borer continues to kill ash trees in North America, there is a need to conserve the ash genetic resource until a time when the insect is no longer a threat or more permanent measures have been put in place to re-introduce ash into the landscape. Seed bank collections allow for the ash genetic variation to be conserved, while systemic insecticides treatments allow ash trees to remain on the landscape where they can continue to provide key ecological services. Our results show that there is potential in combining both genetic conservation measures. The insecticide treatments against emerald ash borer reduced potential loss of seed germplasm by curtailing the presence of seed feeding weevils after multiple years of insecticide application. Moreover, it did not appear to have an adverse effect on ash seed germination time and ability.

CRediT authorship contribution statement

Dorah M. Mwangola: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization. **Aubree M. Kees:** Methodology, Investigation, Writing – review & editing, Visualization. **Donald M. Grosman:** Methodology, Writing – review & editing. **Brian H. Aukema:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Bade Turgut, Joe Ure, Elgin Lee, Mara Short, Kristine Jecha, Tenzin Dothar, Nicole Witt, Piper Haynes, Stephanie Gunter and Cole Doolittle for technical and field help (UMN). Jennifer Burington with the Minnesota Department of Agriculture (MDA) and city foresters of the city of Saint Paul, Rochester, Maple Grove, Coon Rapids, Maple Grove, Roseville,

Lake City and Eagan for continued support and correspondence throughout the study. Funding for this project was provided by the Minnesota Invasive Terrestrial Plants and Pests Center through the Minnesota Environment and Natural Resources Trust Fund appropriation ML 2015, Ch. 76, Art. 2, Sec. 6a. We appreciate the comments of two anonymous reviewers who improved earlier drafts of this work.

References

Annila, E., Heliövaara, K., 1991. Chemical control of cone pests in a Norway spruce seed orchard. Silva Fenn. 25, 56–67.

Arzanov, Y.G., 2013. *Lignyodes bischoffi* Blatchley, 1916 (Curculionidae)-A new species of invasive weevils for Russia. Russ. J. Biol. Invasions 4 (4), 209–211. https://doi.org/10.1134/S2075111713040024.

Balock, J.W., Kozuma, T.T., 1964. Notes on the Biology and Economic Importance of the Mango Weevil, Sternochetus mangiferae (Fabricius), in Hawaii (Coleoptera: Curculionidae).

Barger, J., Davidson, R., 1967. A life history study of the ash seed weevils, *Thysanocnemis bischoffi* Blatchley and *T. helvola* LeConte (Coleoptera: Curculionidae). Ohio J. Sci. 67, 123–127.

Bashir, F., Zahid, F., Iqbal, M., 2014. Growth performance, photosynthetic efficiency and pigment concentration of Glycine max (L.) Merr., as affected by alphamethrin, a synthetic pyrethroid insecticide. Trends Biotechnol. Biol. Sci 1, 29–35.

Cappaert, D., McCullough, D.G., Poland, T.M., Siegert, N.W., 2005. Emerald ash borer in North America: A research and regulatory challenge. Am. Entomol. 51, 152–165. https://doi.org/10.1093/ae/51.3.152.

Cevenini, L., Minelli, A., 2010. Translocation of active ingredient using three trunk injection methods. Acta Hortic. (881), 409–412. https://doi.org/10.17660/ActaHortic.2010.881.64.

Clark, W.E., 1982. Classification of the weevil tribe Lignyodini (Coleoptera, Curculionidae, Tychiinae), with revision of the genus Plocetes. Trans. Am. Entomol. Soc. 108, 11–151.

Clark, W.E., 1980. Revision of Nearctic weevils of the genus *Lignyodes* DEJEAN (Coleoptera: Curculionidae). Trans. Am. Entomol. Soc. 103, 273–326.

Clark, W.E., Whitehead, D.R., Warner, R.E., 1977. Classification of the weevil subfamily Tychiinae, with a new genus and species, new combinations, and new synonymy in Lignyodini (Coleoptera: Curculionidae). Coleopt. Bull. 31, 1–18.

Cuda, J.P., Zellar, M.C., Thomas, M.C., 2005. Ligustrum Weevil (suggested common name), *Ochyromera ligustri* Warner (Insecta: Coleoptera: Curculionidae: Tychiinae: Tychiini: Endaeina). EDIS. https://doi.org/10.32473/edis-in620-2005.

Cuda, J.P., Zeller, M.C., 1998. First record of *Ochyromera ligustri* (Coleoptera: Curculionidae) from Chinese privet in Florida. Florida Entomol. 81, 582–584. https://doi.org/10.2307/3495965.

Dix, M.E., 1986. Lilac seed, an alternate host for the ash seed weevil, *Lignyodes bischoffi* (Blatchley) (Coleoptera: Curculionidae). J. Kansas Entomol. Soc. 59, 389–390.

Donovan, G.H., Butry, D.T., Michael, Y.L., Prestemon, J.P., Liebhold, A.M., Gatziolis, D., Mao, M.Y., 2013. The relationship between trees and human health: Evidence from the spread of the emerald ash borer. Am. J. Prev. Med. 44 (2), 139–145. https://doi.org/10.1016/j.amepre.2012.09.066.

EPPO, 2021. *Sternochetus mangiferae* (World distribution) EPPO Global Database [WWW Document]. URL https://gd.eppo.int/taxon/CRYPMA/distribution (Accessed June 17, 2021).

Flower, C.E., Fant, J.B., Hoban, S., Knight, K.S., Steger, L., Aubihl, E., Gonzalez-Meler, M. A., Forry, S., Hille, A., Royo, A.A., 2018. Optimizing conservation strategies for a threatened tree species: In situ conservation of white ash (*Fraxinus americana* L.) genetic diversity through insecticide treatment. Forests 9, 202. https://doi.org/10.3390/f9040202.

Gandhi, K.J.K., Herms, D.A., 2010a. North American arthropods at risk due to widespread Fraxinus mortality caused by the alien Emerald ash borer. Biol. Invasions 12 (6), 1839–1846. https://doi.org/10.1007/s10530-009-9594-1.

Gandhi, K.J.K., Herms, D.A., 2010b. Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. Biol. Invasions 12 (2), 389–405. https://doi.org/10.1007/s10530-009-9627-9.

Haran, J., Cocquempot, C., Anderson, R., 2019. *Lignyodes helvolus* (LeConte, 1876), une espèce associée aux Frênes nouvellement établie en France continentale (Coleoptera Curculionidae). L'Entomologiste 75, 309–311.

Herms, D.A., McCullough, D.G., 2014. Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. Annu. Rev. Entomol. 59 (1), 13–30. https://doi.org/10.1146/annurev-ento-011613-162051.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biometrical J 50 (3), 346–363. https://doi.org/10.1002/bimj.200810425.

Jansson, R.K., Brown, R., Cartwright, B., Cox, D., Dunbar, D.M., Dybas, R.A., Eckel, C., Lasota, J.A., Mookerjee, P.K., Norton, J.A., Peterson, R.F., Starner, V.R., White, S., 1997. Emamectin benzoate: a novel avermectin derivative for control of lepidopterous pests. Chem. Control 171–177.

Jarvis, H., 1946. Pests of the mango. Queensl. Agric Jour 62, 10–14.

Kashian, D.M., Witter, J.A., 2011. Assessing the potential for ash canopy tree replacement via current regeneration following emerald ash borer-caused mortality on southeastern Michigan landscapes. For. Ecol. Manage. 261 (3), 480–488. https://doi.org/10.1016/j.foreco.2010.10.033.

Kimura, P.H., Okamoto, G., Hirano, K., 1996. Effects of gibberellic acid and streptomycin on pollen germination and ovule and seed development in muscat bailey A. Am. J. Enol. Vitic. 47, 152–156.

Klooster, W.S., Herms, D.A., Knight, K.S., Herms, C.P., McCullough, D.G., Smith, A., Gandhi, K.J.K., Cardina, J., 2014. Ash (*Fraxinus* spp.) mortality, regeneration, and seed bank dynamics in mixed hardwood forests following invasion by emerald ash borer (Agrilus planipennis). Biol. Invasions 16 (4), 859–873. https://doi.org/10.1007/s10530-013-0543-7.

Knight, K.S., Karrfalt, R.P., Mason, M.E., 2010. United States Department of Agriculture Methods for Collecting Ash (*Fraxinus* spp.) Seeds.

Koch, J.L., Carey, D.W., Knight, K.S., Poland, T., Herms, D.A., Mason, M.E., 2011. Breeding strategies for the development of emerald ash borer-resistant North American ash, in: Proceedings of the 4th International Workshop on Genetics of Host-Parasite Interactions in Forestry. pp. 235–239.

Thomas Ledig, F., 1986. Conservation strategies for forest gene resources. For. Ecol. Manage. 14 (2), 77–90. https://doi.org/10.1016/0378-1127(86)90093-9.

Ledig, F.T., Vargas-Hernández, J.J., Johnsen, K.H., 1998. The conservation of forest genetic resources: Case histories from Canada, Mexico, and the United States. J. For. 96, 32–41. https://doi.org/10.1093/jof/96.1.32.

Meyer, B.S., Anderson, D.B., Bohning, R.H., Fratianne, D.G., 1973. Introduction to Plant Physiology, Second. ed. D. Van Nostrand Company.

Mordue(Luntz), A.J., Nisbet, A.J., 2000. Azadirachtin from the neem tree Azadirachta indica: its action against insects. An. da Soc. Entomológica do Bras. 29 (4), 615–632.

Mota-Sanchez, D., Cregg, B.M., McCullough, D.G., Poland, T.M., Hollingworth, R.M., 2009. Distribution of trunk-injected 14C-imidacloprid in ash trees and effects on emerald ash borer (Coleoptera: Buprestidae) adults. Crop Prot. 28 (8), 655–661. https://doi.org/10.1016/j.cropro.2009.03.012.

O'Brien, E.M., Herms, D.A., Gardiner, M.M., Goebel, P.C., Michel, A., 2017. Conserving Ash (*Fraxinus*) Populations and Genetic Variation in Forests Invaded. The Ohio State University.

Olofinboba, M.O., Kozlowski, T.T., 1982. Effects of three systemic insecticides on seed germination and growth of *Pinus halepensis* seedlings. Plant Soil 64 (2), 255–258. https://doi.org/10.1007/BF02184257.

Parween, T., Jan, S., Mahmooduzzafar, S., Fatma, T., Zahid, &, Siddiqui, H., Siddiqui, Z. H., 2016. Selective Effect of Pesticides on Plant-A Review. Crit. Rev. Food Sci. Nutr. 56, 160–179. https://doi.org/10.1080/10408398.2013.787969.

Poiras, A.A., 1998. Catalogue of the Weevils (Coleoptera, Curculionoidea) and their host plants in the Republic of Moldova. Pensoft Pub. https://doi.org/10.1649/0010-065x(2000)054[0548:cotwat]2.0.co;2.

Pritchard, H.W., Moat, J.F., Ferraz, J.B.S., Marks, T.R., Camargo, J.L.C., Nadarajan, J., Ferraz, I.D.K., 2014. Innovative approaches to the preservation of forest trees. For. Ecol. Manage. 333, 88–98. https://doi.org/10.1016/j.foreco.2014.08.012.

R Core Team, 2018. R: A Language and Environment for Statistical Computing.

Rajora, O.P., Mosseler, A., 2001. Challenges and opportunities for conservation of forest genetic resources. Euphytica 118, 197–212. https://doi.org/10.1023/A:1004150525384.

Ramsfield, T.D., Bentz, B.J., Faccoli, M., Jactel, H., Brockerhoff, E.G., 2016. Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts. Forestry 89 (3), 245–252. https://doi.org/10.1093/forestry/cpw018.

Shawir, M.S., Abdel-Latif Abbassy, M., Salem, Y.M., 2014. Laboratory Evaluation of some Insecticides against Larval and Adult Stages of Red Palm Weevil's Rhynchophorus ferrugineus (Olivier). Alexandria Sci. Exch. J. An Int Q. J. Sci. Agric. Environ. 35, 75–79. https://doi.org/10.21608/asejaiqjsae.2014.2579.

Šijačić-Nikolić, M., Milovanović, J., Nonić, M., 2014. Conservation of Forest Genetic Resources. In: Ahuja, M.R., Ramawat, K.G. (Eds.), Biotechnology and Biodiversity. Springer International Publishing, Cham, pp. 103–128. https://doi.org/10.1007/978-3-319-09381-9_7.

Singh, V.K., Mathur, M., Mathur, S.N., 1982. Phyto-toxicity of the insecticide, phorate on germination of vigna mungo. Agric. Biol. Chem. 46 (6), 1681–1682. https://doi.org/10.1080/00021369.1982.10865307.

Solomon, J.D., Leininger, T.D., Wilson, A.D., Andernon, R.L., Thompson, L.C., McCracken, F.I., 1993. Ash pests: Guide to Major Insects, Diseases, Air Pollution Injury, and Chemical Injury.

St.Clair, J.B., Howe, G.T., 2011. Strategies for conserving forest genetic resources in the face of climate change. Turk. J. Botany 35, 403–409. https://doi.org/10.3906/bot-1012-98.

Subramanyam, C.K., 1926. A note on the life history of *Cryptorhynchus mangiferae* Fabricius. Madras Agric. Dep. Yearb. 1925, 29–36.

Sutherland, J.R., Woods, T.A.D., Miller, G.E., 1984. Effect of selected insecticides and fungicides on germination of Douglas-fir and white spruce pollen. Tree Plant. Notes 35, 22–24.

Tanis, S.R., Cregg, B.M., Mota-Sanchez, D., McCullough, D.G., Poland, T.M., 2012. Spatial and temporal distribution of trunk-injected 14C-imidacloprid in Fraxinus trees. Pest Manag. Sci. 68 (4), 529–536. https://doi.org/10.1002/ps.2281.

USDA Forest Service and Michigan State University, 2022. emeraldashborer.info [WWW Document]. URL http://www.emeraldashborer.info (Accessed February 22, 2022).

Wanat, M., Mocarski, Z., 2008. Current range of the ash seed weevil *Lignyodes bischoffi* Blatchley, 1916 (Coleoptera: Curiculionidae) in Poland. Polish J. Entomol. 77, 177–182.

Werner, R.A., 1974. Penetration and Persistence of Systemic Insecticides in Seeds and Seedlings of Southern Pines. J. Econ. Entomol. 67 (1), 81–84. https://doi.org/10.1093/jee/67.1.81.

Widrlechner, M.P., 2010. Building a comprehensive collection of Ash germplasm, in: Proceedings of the 4th Global Botanic Gardens Congress. pp. 1–10.

Woodcock, P., Marzano, M., Quine, C.P., 2019. Key lessons from resistant tree breeding programmes in the Northern Hemisphere. Ann. For. Sci. 76, 51. https://doi.org/10.1007/s13595-019-0826-y.

ML 2015, Ch. 76. Art. 2, Sec. 6a Project Abstract

PROJECT TITLE: Sub-project #6: Distribution and traits of the fungal pathogen *Fusarium virguliforme* that influence current and future risks to soybean and other legumes in Minnesota

PROJECT MANAGERs: Dean Malvick and Kathryn Bushley
AFFILIATION: University of Minnesota
MAILING ADDRESS: Department of Plant Pathology, 495 Borlaug Hall; and Department of Plant and Microbial Biology, 140 Gortner Laboratory, University of Minnesota
CITY/STATE/ZIP: St Paul, MN 55108
PHONE: 612-625-5282 (D. Malvick)
E-MAIL: dmalvick@umn.edu
WEBSITE: https://plpa.cfans.umn.edu/people/dean-malvick

FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: M.L. 2015, Chp.76, Sec. 2, Subd. 6a

APPROPRIATION AMOUNT: $ 383,581
AMOUNT SPENT: $383,581
AMOUNT REMAINING: $ 0

Sound bite of Project Outcomes and Results

This project has discovered factors that influence the ability of the fungus *Fusarium virguliforme* to become established as a destructive pathogen on crops in new areas of Minnesota. The results are foundational to understanding this pathogen and contribute to managing the diseases it causes on soybean and other crops.

Overall Project Outcome and Results

The fungal pathogen *Fusarium virguliforme*, which causes sudden death syndrome (SDS) on soybean and root rot of other legumes, is an expanding problem for crop producers in Minnesota. Our research team has made discoveries regarding the pathogen's ability to spread and cause disease. First, a survey has confirmed the spread of the pathogen for the first time into seven counties in central and western MN. Second, studies of nutrient use suggest that *F. virguliforme* grows on a larger number of carbon and nitrogen sources than many other fungi in crop fields, likely giving it a competitive advantage. Analysis of competition between *F. virguliforme* and other fungi from crop fields revealed that while several fungi can inhibit its growth, multiple others are overcome by the pathogen, indicating it is a good competitor in soil and roots. Third, we determined it can survive to -40°C and thus its spread is not likely limited by cold temperatures. Fourth, in field and greenhouse experiments investigating host range, multiple crop species (black bean, pinto bean, kidney bean, and pea) showed symptoms of disease, and multiple other plant species were infected asymptomatically. Fifth, we completed a study and a publication on genetic and pathogenic variation among *F. virguliforme* populations in Minnesota and the Midwest. While genetic groups did not correspond to aggressiveness, three genetic clusters were identified, with two clusters likely contributing most to spread of this fungus. Sixth, we completed initial analysis of genomes from 35 isolates to investigate genes involved in pathogenicity and abilities to invade new environments. The projected trained one M.S. level and one postdoctoral level scientist, expanding expertise for addressing invasive plant pathogens. This project significantly advances fundamental and applied knowledge of *F. virguliforme* that can be harnessed for disease management and risk analysis by scientists, agricultural professionals, and crop producers.

Project Results Use and Dissemination

This project has discovered multiple factors that influence the ability of *F. virguliforme* to spread and become established as a destructive pathogen on crops in new areas. Results have been presented via University of Minnesota Extension programs to key agricultural professionals and crop producers across Minnesota that contribute to managing this pathogen and the crop diseases it causes. Results have also been presented at scientific conferences and are being published in scientific journals.

ML 2015, CH. 76, Art. 2, Sec. 6a, MITPPC Sub-Project #17 Project Abstract
For the Period Ending June 30, 2020

PROJECT TITLE: MITPPC Sub-project #7: Tools to Distinguish Native from Exotic Reed Canary Grass
PROJECT MANAGER: Neil O. Anderson
AFFILIATION: Dept. of Horticultural Science, University of Minnesota
MAILING ADDRESS: 286 Alderman Hall, 1970 Folwell Avenue
CITY/STATE/ZIP: Saint Paul, MN 55108
PHONE: 612.624.6701
E-MAIL: ander044@umn.edu
WEBSITE: n/a
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: M.L. 2015, Chp. 76, Sec. 2, Subd. 6a

SUB-PROJECT AMOUNT: $248,848
AMOUNT SPENT: $248,848
AMOUNT REMAINING: $0

SOUND BITE: This project used genetic techniques to find that most reed canarygrass in Minnesota is native to the state and not from Europe. The implications of this surprising result for land management are still being discussed.

Overall Project Outcome and Results

The goal of this project was to use molecular markers to determine native vs. exotic reed canary grass status in various locations across Minnesota growing along rivers (Des Moines, Minnesota, Mississippi, Red, Roseau, St. Croix), in fields, as commercially-grown cultivars (forage, ornamental), and preserved historic specimens in herbaria (<1940, presumed native) and a corollary set of samples from rivers in the Czech Republic as exotic comparisons (Activity 1); along Minnesota transportation corridors (highways) existing during the 1920s-1930s (Dust Bowl era) and Minnesota lakes (Bush, Cedar, Como, Phalen, Mille Lacs, Minnetonka, Square, White Bear) and Central Park (Activity 2). Due to Covid-19 travel restrictions, we were unable to get permission to collect along additional lakes. The number of plants analyzed totaled 3,430 (Activities 1,2). Plant DNA was extracted from each sample to determine genomic markers of short DNA sequences (2,889 highly differentiated single nucleotide polymorphisms, SNPs, out of 16,902 total markers) to distinguish native vs. exotic status. Genetic analysis of reed canarygrass showed that river populations are native Minnesota or North American types. Herbarium samples as well those from a native, unplowed field (Roseau, MN) were genetically similar to wild collections from five Minnesota rivers; forage cultivars in commercial fields (Roseau, MN) and along the Roseau River formed a separate group. The exotic central European populations were distinctly different from all native MN groups. Most variation is within (98.8%), rather than among (1.2%), populations, suggesting little divergence and a high level of shared genetic markers. Across the state, Minnesota rivers had 2-32 genetic variants present, some of which were shared among rivers. Thus, the majority of Minnesota reed canarygrass, while invasive, is native in origin and not exotic (European). Thus, based on this study, all of MN reed canarygrass is native; Tribal and State managers may choose to preserve this species.

Project Results Use and Dissemination

Dissemination of native vs. exotic status of all *Phalaris* results has been reported on the Department of Horticultural Science website, that of the PI's, as well as in all PIs/co-PIs Experts at umn.edu links. As many as 11 abstracts were published in national and international meetings, along with corollary public posters sessions or seminar talks to varied audiences of academics, land managers, students, and/or the public-at-large. We have kept State and Tribal Land Managers informed on the native status of MN reed canarygrass and have initiated discussions on approaches to managing this native species yet invasive. The investment by the state on control

measures for this invasive grass warrant careful consideration of best management approaches to maintaining the native genetic diversity yet not encouraging the invasive spread of this grass into managed areas. Results were also communicated to the scientific community in peer-reviewed journal articles.

ML 2015, CH. 76, Art. 2, Sec. 6a, MITPPC Sub-Project #8 Project Abstract
For Period Ending January 31, 2021

PROJECT TITLE: Sub-Project #8. Accurate detection and integrated treatment of oak wilt *(Bretziella fagacearum)* in Minnesota

PROJECT MANAGER: Jeannine Cavender-Bares
AFFILIATION: University of Minnesota
MAILING ADDRESS: Dept. of Ecology, Evolution and Behavior, 1479 Gortner Ave
CITY/STATE/ZIP: Saint Paul MN 55108
PHONE: 612-624-6337
E-MAIL: cavender@umn.edu
WEBSITE: https://cbs.umn.edu/cavender-bares-lab/home
FUNDING SOURCE: Environment and Natural Resources Trust Fund

APPROPRIATION AMOUNT: $ 357,420
AMOUNT SPENT: $ 357,420
AMOUNT REMAINING: $ 0

Sound bite of Project Outcomes and Results

This project developed methods and approaches for better detection of oak wilt using spectroscopic technology and documented best practices to prevent spread of the disease.

Overall Project Outcome and Results

Our team has made substantial progress on the development of methods and approaches for accurate detection of oak wilt in Minnesota forest using spectroscopic technology. We have also documented best practices for management efforts to prevent spread of the disease. Specifically, we have completed physiological experiments demonstrating the disease can be differentiated from other stress factors under controlled conditions (Activity 1). A manuscript on the greenhouse seedling experiment using leaf and whole plant spectroscopic data to differentiate oak wilt from bur oak blight and drought has been published in *Tree Physiology*. We have advanced analyses and ground-truthing of AVIRIS NG airborne imagery including model development and spectral index development for stress physiology in response to the oak wilt disease (Activity 2). In an outdoor field experiment using naturally growing oak saplings at the Cedar Creek Ecosystem Science Reserve, oak saplings were inoculated with oak wilt and compared to healthy saplings using leaf and canopy spectroscopy. Results indicate that physiological disease symptoms can be readily detected using spectral sensors at both leaf and canopy scales using statistical models and simple indices from spectral features linked to physiological stress. Lastly, treatments were completed at 20 oak wilt sites with a new "double plow line" to prevent spread of the disease through root grafts. Initial assessments indicate the approach is highly effective, but a final determination will be made 5 years after treatment, beyond the life of this project (funding secured from USDA Forest Service). Two postdoctoral scientists, a technical scientist, a first-year graduate student and two undergraduate research assistants received training and mentoring during the project.

Project Results Use and Dissemination

Our team has disseminated new knowledge from this project to local, regional, national and international audiences. A significant peer-reviewed publication has already come this project (Beth Fallon, Anna Yang, Cathleen Lapadat, Isabella Armour, Jennifer Juzwik, Rebecca A Montgomery, Jeannine Cavender-Bares. 2020. Spectral differentiation of oak wilt from foliar fungal disease and drought is correlated with physiological changes. *Tree Physiology* 40(3): 377–390, https://doi.org/10.1093/treephys/tpaa005). Others are in development. The team delivered 11 talks, three posters, and one field tour to professional audiences. In addition, the project was featured in *The Minnesota Daily* and Market Science (scientific engagement at farmers' markets).

ML 2015 Project Abstract

PROJECT TITLE: MITPPC Sub-project #9Characterizing dispersal of larval gypsy moth to improve quarantine regulations
PROJECT MANAGER: Brian Aukema
AFFILIATION: University of Minnesota, MITPPC
MAILING ADDRESS: 1992 Folwell
CITY/STATE/ZIP: St. Paul, MN 55108
PHONE: 612-626-1914
E-MAIL: hkoop@umn.edu
WEBSITE: mitppc.umn.edu]
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: ML 2015, Ch. 76m Sec. 2, Subd. 6a

APPROPRIATION AMOUNT: $35,000

Overall Project Outcomes and Results

The European gypsy moth is an invasive insect that feeds on over 300 species of trees and shrubs. Management guidelines within a national quarantine recommend that wood products, such as timber being harvested and moved from the forest, are staged within a buffer zone area devoid of any host vegetation during harvesting operations. This buffer zone reduces the likelihood that late instar gypsy moth larvae will pupate nearby, emerge as adults, mate, and lay eggs on the wood. In practice, this buffer zone is 100 feet in radius, but this distance was established with limited understanding of the movement ecology of gypsy moth larvae. We conducted laboratory experiments at the University of Minnesota to determine how host type and food deprivation affected movement of gypsy moth caterpillars. During outbreaks, food can become scarce as larvae strip trees of foliage. Larvae were raised on one of five foods: oak, tamarack, Norway maple, sugar maple, or artificial diet. Subsets of larvae were also deprived of food for zero, 24, or 48 hours. After the food deprivation period, late instar larvae were placed on the servosphere. Larvae raised on oak, a preferred host, were unlikely to move unless starved. They moved farther and faster the longer they were starved. In contrast, when larvae were raised on less preferred hosts, they were more likely to move without prior starvation. These results suggest that feeding on optimal hosts provides gypsy moth larvae with the energy and nutritional requirements to move more quickly to more food when there is none immediately available. Thus, risks of larvae crossing a regulatory buffer zone may increase where an outbreak results in complete defoliation of preferred hosts. Results from this laboratory study were integrated with a federally-funded field study to inform best management practices of this invasive species in Minnesota.

Project Results Use and Dissemination

This work has been shared with national regulatory officials at USDA APHIS who are revising the national gypsy moth management handbook at a time when the insect continues to invade Minnesota. This work was also been presented at two conferences with resource managers and other research staff:

Upper Midwest Invasive Species Conference Oct 15-18, 2018. Rochester, MN. Wittman, J.T., Kees, A.M., and B.H. Aukema. Characterizing the movement behavior of gypsy moth (*Lymantria dispar*) caterpillars using a servosphere.

ESA/ESC/ESBC Joint Annual Meeting Nov 11-14, 2018. Vancouver, BC. Wittman, J.T. and B.H. Aukema. Effects of host foliage on the movement behavior of larvae of gypsy moth *Lymantria dispar.*

One scientific journal article has been submitted:

Wittman, J.T. and B.H. Aukema. (201x) Foliage type and availability alters the movement behavior of late instar European gypsy moth *Lymantria dispar* (Lepidopera: Erebidae).*Journal of Insect Behavior* Submitted Nov 22, 2018.

ML 2015 Ch 76, Art 2, Sec 6a Project Abstract
For the Period Ending June 30, 2021

PROJECT TITLE: Subproject #10: Overwintering, Migration and Development of Cost-Effective Practical Management Strategies
for the Invasive Spotted Wing Drosophila in MN
PROJECT MANAGER: Mary Rogers
AFFILIATION: UMN Dept of Horticultural Science
MAILING ADDRESS: 1970 Folwell Ave
CITY/STATE/ZIP: St. Paul, MN, 55108
PHONE: 612-624-8871
E-MAIL: roge0168@umn.edu
WEBSITE: https://rogerslab.cfans.umn.edu
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER
ML 2015, Ch. 76, Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $478,876
AMOUNT SPENT: $478,876
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results
Our project developed new cost-effective methods to help growers manage damage and reduce yield loss caused by the invasive Spotted-wing drosophila in small fruit while reducing pesticide use. Additionally, we have gained basic knowledge on the behavior and flight capabilities of this pest that will contribute to future management strategies.

Overall Project Outcome and Results
Spotted wing drosophila (*Drosophila suzukii*, SWD) is an invasive fly that lays eggs in intact, ripening fruit such as blueberries, strawberries, and raspberries. This pest has caused considerable economic losses for small fruit growers. First detected in MN in 2012, SWD threatens 750 acres of raspberries, strawberries, grapes, and blueberries, in addition to its 5,000 high tunnel operations statewide. At the start of our project, current control tactics were limited to repeat applications of broad-spectrum insecticides that failed to adequately protect fruit from infestation, in addition to posing risks to the environment. Additionally, we faced gaps in understanding the basic biology ad behavior or SWD, such as migration and overwintering in Minnesota, which hindered our ability to recommend appropriate management strategies. To address this, we proposed three goals: 1) develop SWD forecasting tool using local migration and overwintering data; 2) investigate efficacy of alternative management techniques; and 3) research economic impact and develop decision making tools. As a result of our work, we have indirect evidence showing that SWD may be overwintering and little evidence that the SWD has the flight capabilities for long-distance movement. We learned that physical exclusion can effectively reduce SWD damage and is cost-effective for small farms and reduces the need for insecticide sprays. Our work on biopesticides and novel repellants shows promising results in the lab but is less consistent in the field, warranting new methods to increase field efficacy. Economically, we found that SWD is responsible for at least $2 million in losses annually to raspberry growers alone, establishing the need for management for the statewide fruit industry, and growers can benefit from adopting physical exclusion and biological based pesticides. Our science-based management recommendations for this best improves overall sustainability of small fruit production in Minnesota.

Project Results Use and Dissemination
Our project has resulted in six peer-reviewed publications in scientific journals, eight academic presentations, over thirty talks to grower audiences and dozens of online newsletters, articles, and blog submissions, and a grower decision making tool. Grower recommendations are available on the FruitEdge website and archives on

the UMN Extension Fruit and Vegetable News. Through this work, we have leveraged an additional $750,000 in federal funds to further develop sustainable production and pest management techniques for small fruit in Minnesota.

ML 2015, Ch 76, Art. 2, Sec. 6a Project Abstract
For the Period Ending December 31, 2021

PROJECT TITLE: Subproject #11: Will Future Weather Favor Minnesota's Woody Invaders?
PROJECT MANAGER: Peter Reich
AFFILIATION: University of Minnesota
MAILING ADDRESS: 1530 Cleveland Ave N
CITY/STATE/ZIP: Saint Paul, MN, 55108
PHONE:
E-MAIL: preich@umn.edu
WEBSITE:
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION:
MINNESOTA INVASIVE TERRESTRIAL PLANTS AND PESTS CENTER
ML 2015, Ch. 76, Art. 2, Sec. 6a

APPROPRIATION AMOUNT: $514,325
AMOUNT SPENT: $514,325
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results

Our findings tell the story of how exotic honeysuckle and buckthorn have invaded Minnesota forests, how and why new areas are likely to be invaded in the future, and how we may be able to mitigate invasion using native tree species.

Overall Project Outcome and Results

Glossy buckthorn, common buckthorn, tatarian honeysuckle, and morrow's honeysuckle are woody species that have been introduced to Minnesota forests from other continents. All four species frequently dominate forests and exclude native plant species. Warming temperatures and continued dispersal of these species are likely to significantly increase their abundance throughout Minnesota, especially in northern Minnesota. However, most effort by researchers and managers alike has been given to reactive measures against invasion instead of increased understanding of invasion processes and/or preventative measures. This project evaluated the climate sensitivity of these four invasive species in a way that provides for more accurate threat assessment of each throughout the state and provides tools for Minnesotans to potentially slow invasion into new areas and protect Minnesota's forests. We analyzed growth rings of 274 trees to determine how quickly invasive species spread and characterize how native and invasive species have responded to past growing conditions. We found that growth rates of invasive buckthorn and honeysuckle are most similar to native cherries and ashes in southern Minnesota, but that the invasive species already are growing much faster than those native species in northern Minnesota. Within a forest, we found that buckthorn tended first to invade hilltops and subsequently spread to low-lying areas at a rate of 3-4m yr^{-1} (slower than honeysuckle, which spread at 6 m yr^{-1}). We experimentally assessed 10 native species in addition to the four invaders to determine which are favored by changing temperature and rainfall patterns (i.e. their responses to future climate). We found invasive and more southern native species to be favored by warming conditions in terms of their growth and survival, whereas more northern native species were often strongly disfavored. We established programs to detect current invasion at fine-scale spatial resolution and predict future invasion based on the findings above, and set up long-term experiments to test the ability of tree plantings to slow invasion into new areas.

Project Results Use and Dissemination

Results from this project were disseminated through multiple avenues, including conference presentations, journal articles, and popular media. Principally, dissemination efforts focused on academic journals. We have submitted one manuscript detailing results from Activity 2 for peer review. Three other manuscripts related to the project are in preparation and will be submitted during the spring of 2022. We are also collaborating with National Geographic for a feature on work supported by this grant, primarily results associated with Activity 2.

ML 2015, CH. 76, Art. 2, Sec. 6a, MITPPC Sub-Project #12 Project Abstract
For the Period Ending January 30, 2020

PROJECT TITLE: MITPPC Sub-project #12: Developing robust identification assays for *Amaranthus palmeri* in seed mixture
PROJECT MANAGER: Don Wyse
AFFILIATION: Department of Agronomy and Plant Genetics
MAILING ADDRESS: 411 Borlaug Hall, 1991 Upper Buford Crescent
CITY/STATE/ZIP: St. Paul, MN 55108
PHONE: 612-625-7064
E-MAIL: wysex001@umn.edu
WEBSITE: https://agronomy.cfans.umn.edu/department-directory/donald-l-wyse
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: ML 2015, Ch. 76, Art. 2, Sec. 6a

SUB-PROJECT AMOUNT: $208,230
AMOUNT SPENT: $ 208,230
AMOUNT REMAINING: $ 0

SOUND BITE: This project created a highly reliable test for detecting Palmer Amaranth, in individual plants and pools of seed. The test is expected to be commercially available and will be an important tool for Minnesota farmers, crop consultants, and agronomic specialists. Palmer amaranth can reduce corn and soybean yields by 80-90%.

Overall Project Outcome and Results

Palmer Amaranth is an aggressive and prolific weed species that poses a major ecological and economic risk to growers in the state of Minnesota. Closely related to other pigweed species, Palmer has a far more severe impact on agricultural row cropping systems. Early identification of Palmer Amaranth is critical, as it has developed resistance to some of the most widely used herbicides; ALS-inhibitors, PPO-inhibitors, and glyphosate. Visual identification of Palmer Amaranth against other pigweed species is difficult, which has led to the use of genetic testing becoming the standard for identifying Palmer.

To address this emerging challenge we collected at team of weed science experts from the University of Minnesota, Colorado State University and the Minnesota Department of Agriculture. This team developed an improved genetic test to maximize the robustness and reliability of Palmer Amaranth identification for both individual plants and bulk seed screenings. To achieve this, our team collected Pigweed samples across the United States as well as Mexico, South America and Africa. We extracted DNA samples from a total of 24 populations of Palmer amaranth and 42 non-Palmer pigweeds, resulting in DNA from over 2,000 individual plants. We sequenced more than 800 of these samples through the University of Minnesota Genomic center to search for genetic differences between Palmer and the other species. These differences served as a target for developing a set of genetic markers that can be used for species identification. Once developed the genetic markers were tested against 1,250 pigweed samples to assess their performance.

The final result is a highly reliable test for (>99.7% accuracy) for detecting Palmer Amaranth, both for individual plants and pools of seed. This test will be an important tool for Palmer control for Minnesota growers, crop consultants, and other agronomic specialists. The test is expected to be commercially available in 2020.

Project Results Use and Dissemination

This project resulted in the development of a robust, highly accurate and easy to use assay for the identification of Palmer Amaranth against other pigweed species. This assay is commercially viable, and a patent was filed for the two markers developed solely at UMN on March 27th, 2020 (Patent #63,000,946). Collaborators at Colorado State University has stated their interest in licensing the tests and offering them as part of a comprehensive Pigweed seed testing service.

This project has also resulted in the creation of a large body of genetic sequence data for Pigweeds assembled from across a wide geographic range. This data will be a valuable resource for future work on Palmer Amaranth and related pigweed species, and will be made publically available through NCBI.

Results of this project were shared at a seminar at Colorado State University, a session at a joint meeting of the Upper Midwest Invasive Species Conference and North American Invasive Species Management Association, and an Applied Plant Science seminar at the University of Minnesota. A YouTube video about the project was created for general audiences as well.

Active presence during regional Palmer Amaranth conference calls allowed us to keep neighboring states appraised of our progress and will be one avenue of announcement for when the Palmer identification test when it becomes commercially available.

A peer reviewed journal article is currently in writing to be submitted to Frontiers in Plant Science, and presentations are being prepared for sessions at two upcoming conferences: the Ecological Society of America in August 2020 and the Upper Midwest Invasive Species Conference in October 2020..

ML 2015, CH. 76, Art. 2, Sec. 6a, MITPPC Sub-Project #13 Project Abstract
For Period Ending July 31, 2021

PROJECT TITLE: MITPPC #13: Terrestrial invasive species prioritization
PROJECT MANAGER: Amy Morey
AFFILIATION: Minnesota Invasive Terrestrial Plants and Pests Center & UMN Dept. of Entomology
MAILING ADDRESS: 1980 Folwell Ave, Hodson Hall Rm 219
CITY/STATE/ZIP: St. Paul, MN 55108
PHONE: 406-698-7684
E-MAIL: morey041@umn.edu
WEBSITE: n/a
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: M.L. 2015, Ch. 76, Sec. 2, Subd. 6(a)

APPROPRIATION AMOUNT: $74,461
AMOUNT SPENT: $74,461
AMOUNT REMAINING: $0

Sound bite of Project Outcomes and Results

This project produced written risk evaluations of 77 terrestrial invasive species requested for review by MITPPC stakeholders, and assisted with the 2020 update of the MITPPC prioritization analysis. Thorough review of species allows MITPPC to be dynamic and transparent in how it responds to emerging TIS threats and stakeholder concerns.

Overall Project Outcome and Results

In 2017, the Minnesota Invasive Terrestrial Plants and Pests Center undertook an expansive research prioritization to systematically evaluate threats posed by a wide array of terrestrial invasive invertebrates, plants, and plant pathogens and created the document, "*Minnesota's Top 124 Terrestrial Invasive Plants and Pests: Priorities for Research*," which has provided guidance on funding MITPPC research projects in subsequent years. Since its publication, many terrestrial invasive species (TIS) have been suggested for further review by stakeholders. The movement of TIS into Minnesota and their potential harms calls for a thorough review of suggested species.

Following methodology developed in the 2017 document, this MITPPC project evaluated 77 plant, invertebrate, and plant pathogen species submitted for review by stakeholders. The results of each evaluation are incorporated into the Analytic Hierarchy Process (AHP) model used by MITPPC to rank and prioritize the TIS that threaten Minnesota's terrestrial ecosystems. AHP is a form of multi-criteria decision analysis that makes the process of selecting the highest priority threats consistent and transparent. AHP has been used by many agencies and organizations to facilitate complex decision making. This project provided an update to the research priorities for the MITPPC 2020 call for proposals, which included ~45 new or re-reviewed species. Evaluations completed after the 2020 update will contribute to an anticipated 2022 prioritization update.

The project also enhanced and updated the evaluation of species such that multi-page documents are produced for each species to clearly outline the information used for characterizing the potential threat posed to Minnesota. Evaluation documents will be made available for stakeholder feedback. These changes allow MITPPC to continue to be dynamic and transparent in how it responds to emerging TIS threats and stakeholder concerns.

Project Results Use and Dissemination

During this project, evaluations of 77 species were conducted. Results from approximately 45 of these evaluations were analyzed and included in the 2020 updated prioritization, a summary of which can be found in the document "*Minnesota's Top Terrestrial Invasive Plants & Pests for Research: An Expanded Prioritization*" (https://mitppc.umn.edu/invasive-species-prioritization). The remaining evaluations will contribute to an anticipated update to the research priorities in 2022. Evaluation documents are in the process of becoming

publically available for feedback from MITPPC stakeholders on the MITPPC website (https://mitppc.umn.edu/invasive-species-prioritization).