

A BIODIVERSITY INFORMATICS APPROACH TO PREVENTING INVASIONS: USING A
WHOLE NON-NATIVE FLORA TO INVESTIGATE INTRODUCTION PATHWAYS AND
METHODS FOR INVASION TRACKING

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY
OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY IN BOTANY

DECEMBER 2021

BY

Kelsey Brock

Dissertation Committee:

Curtis Daehler, Chairperson

David Duffy

Lucas Fortini

Rakan Zahawi

Camilo Mora, University Representative

Keywords: biological invasions, plants, islands, biodiversity informatics, introduction pathways,
temporal trends, monitoring, frameworks

© Copyright 2021 by Kelsey Brock

All rights reserved

Acknowledgements

For funding my research, I am indebted to the Hawai‘i Invasive Species Council and the Pacific Islands Climate Adaptation Science Center. Additionally, I am grateful for the Torrey–Degener–Rock, Harold and Elizabeth St. John, Charles H. Lamoureaux botany scholarship programs for providing funding to publish my results. I also thank the Pacific Cooperative Studies Unit at the University of Hawai‘i at Mānoa for administrative support.

I thank my committee for lending their experience and insight, including Curt Daehler, David Duffy, Lucas Fortini, Zak Zahawi and Camilo Mora. Their attention greatly improved drafts of the chapters contained within this dissertation and helped me navigate my field of study. I am especially grateful to my advisor, Curt Daehler, for spending many hours discussing ideas and providing advice on projects; this guidance has helped make me a better researcher who is motivated to continue working on what I have started here.

My research is shaped by what I learned from many people I encountered during my time at the Kaua‘i Invasive Species Committee, who continue to guide me today. For stimulating my love of solving the messy problems one confronts when managing plant invasions, I thank the KISC staff and advisors, including Tiffani Keanini, Ray Kahaunaele, Adam Williams, Cleve Javier, Mugs Kaneolani, Jenny Kay and Joseph Kona for sharing their ideas and experience. I also thank Chuck Chimera for helping me understand weed risk assessments and for sharing data. For their never-ending knowledge and encouragement, I am really grateful to have met Tim Flynn and David Lorence at the National Tropical Botanical Garden. Thank you for helping me become a better botanist. Upon moving to O‘ahu to start my PhD, the Bishop Museum Botany Department became my go-to resource for understanding plant taxonomy and how biodiversity data are curated. I’m forever grateful to Barb Kennedy, Clyde and Pumehana Imada, Tim Gallaher, and Nick Walvoord for their wisdom and friendship.

Lastly, my family is my greatest asset, and I could not have accomplished any of this without them. Jared and Audrey, Mom and Dad, thank you for everything.

Abstract

Negative impacts from biological invasions continue to rise as non-native species spread around the globe. As the costs of controlling these species increases significantly after their establishment and spread, invasion biologists acknowledge that strategies to prevent invasions should be a key focus in addressing invasive species problems. However, numerous species may be introduced each year, making it difficult to prioritize would-be invaders among hundreds of other species. Thus, strategies that identify and regulate pathways for species introductions are needed to complement species-specific approaches. Furthermore, methods to track the rate at which new species establish and their fate after arrival are lacking, despite the need to assess risks from future invaders and evaluate the success of prevention strategies.

To address these issues, I gathered data on the date of naturalization for the entire naturalized flora of the Hawaiian Islands alongside data on their origins, native climate type, taxonomy and likely reason for introduction (introduction pathway). By comparing rates of naturalization for each introduction pathway with changes in socioeconomic factors, I reveal that Hawai‘i has received a diversity of plants from all over the world, and that the rate of ornamental plant naturalizations has risen dramatically since the mid-20th century, reflecting Hawai‘i’s shift to a tourism-based economy from an agricultural one. I also show that, although many naturalized plants may currently exist in Hawai‘i at lower elevations with warmer climates, a large proportion are native to climates similar to those found at higher elevations. This pattern is significant because Hawai‘i’s remnant native-dominated ecosystems exist primarily at higher elevations, indicating that preventing non-natives from establishing in native ecosystems is crucial to conservation of native species. Additionally, the rate of spread between islands has increased since the mid-20th century, highlighting the need to prevent inter-island spread.

To explore the pitfalls that arise when measuring rates of naturalization over time, I then focused on a relatively neglected factor that can bias analyses: time lags that accrue when processing data after their collection in the field. Using computer simulations and an analysis of a real-world case study of two independently collected datasets for the Hawaiian Islands, I illustrate how time lags interact with common data retrieval strategies to influence the interpretation of invasion trends. By doing so, I reveal that long lags due to insufficient field surveying can create the illusion of a sudden onset or exponential rise in naturalization rates, whereas lags in identification, reporting and compilation result in the deceptive appearance of an invasion slow down. I also show that harvesting data from already-compiled resources published by experts may introduce a temporal sampling bias because such works are sporadically produced, thereby introducing an additional lag between reporting and data compilation.

Lastly, I investigate a strategy for tracking species after their arrival by using Hawai‘i’s naturalized species checklist as a starting point and applying a well-recognized framework that categorizes the phase of a non-native species’ establishment along the introduction–invasion continuum. After finding that data deficiencies prevent hundreds of species from being categorized within the framework, I show that data from the Hawai‘i-Pacific Weed Risk Assessment can be reappropriated to predict whether a data-deficient species will progress along the continuum. Ultimately, I reveal that this predictive tool is a promising supplement to on-the-ground monitoring, especially when frequent field surveys are not feasible. Although globalization continues to facilitate numerous invasions, our ability to harness data and solve problems using biodiversity informatics is advancing rapidly. Here, I contribute to this progress by uncovering trends useful to policymakers and managers in invader-rich regions, while also offering guidance on how to improve methods used to measure those trends and track invasions.

Table of Contents

Acknowledgements	iii
Abstract	iv
List of Tables	viii
List of Figures	ix
Chapter 1 – Introduction	1
References	7
Chapter 2 – Plant naturalization trends reflect socio-economic history and show a high likelihood of inter-island spread in Hawai‘i	13
Abstract	13
<i>Box 1. Management Implications</i>	14
Introduction	15
Materials and Methods	18
Results and Discussion	23
<i>Composition and Origin of Naturalized Flora</i>	23
<i>Rates of Naturalization</i>	28
<i>Box 2. Post-1900 socio-economic events likely affecting plant naturalizations in Hawai‘i</i> .	32
<i>Inter-island Spread</i>	33
<i>Conclusion</i>	35
References	37
Figures	46
Chapter 3 –Temporal sampling biases in invasion trend analyses: data processing is just as important as field collection	51
Abstract	51
Introduction	52
Methods	56
<i>Varying Lag Lengths</i>	58
<i>Varying Harvest Points</i>	59
<i>Varying Record Retrieval Rate</i>	59
<i>Case Study</i>	60
Results	62
Discussion	66
<i>Conclusion</i>	72

References	73
Tables	80
Figures	84
Chapter 4 – Applying an invasion and risk framework to track non-native island floras and inform management: a case study of challenges and solutions in Hawai‘i.....	89
Abstract	89
Introduction	90
Methods.....	94
<i>Aligning the naturalized species checklist.....</i>	<i>94</i>
<i>Potential for Weed Risk Assessment scores to predict naturalization.....</i>	<i>96</i>
Results	97
<i>Alignment between the checklist and framework</i>	<i>97</i>
<i>Relationship between WRA score and naturalization</i>	<i>101</i>
<i>Predicting the fate of questionably naturalized species</i>	<i>102</i>
Discussion	102
<i>Ambiguous alignment of questionably naturalized species</i>	<i>103</i>
<i>WRA scores can assign a likely status for questionably naturalized species</i>	<i>105</i>
<i>Conclusions</i>	<i>113</i>
References	113
Tables	121
Figures.....	127
Chapter 5 – Conclusion	130
References	138
Appendix.....	142
Part 1—Guidance on status confidence rating	142
Part 2—Adapting a Regional Checklist to Track Invasion Statuses	143

List of Tables

Table 3.1 Quantification of trendlines using changepoint analysis with each each value representing an average of 10 runs	80
Table 4.1 Species in the “Questionably Naturalized” checklist category alongside their likely status category derived from WRA scores and time since last observation.	121
Table 4.2 Practical adaptation of the framework requiring minimal additional data collection beyond species checklists.	125

List of Figures

Figure 2.1 The main Hawaiian Islands, showing habitat status and climate types, which shows the overlap between temperate climate and native dominated ecosystems.	46
Figure 2.2 Number of naturalized plants in Hawai‘i as categorized by plant families containing > 15 species, life forms, and Köppen-Geiger climate types within their native range	47
Figure 2.3. Number of naturalized plants in Hawai‘i as categorized continent of origin, and the proportion of species from single or multiple continents per probable introduction pathway..	48
Figure 2.4 Naturalization of plants from 1900 to 2020, showing the accumulation of all species and species categorized by their likely introduction pathway alongside changes in real gross domestic product, human population, tourists, and acres of productive farmland	49
Figure 2.5 Inter-island spread of naturalized species over time in Hawai‘i, showing fluctuating rates of spread, and the proportion of species introduced each decade that have spread among islands.	50
Figure 3.1 Schematic representation of the pipeline for collecting and processing biodiversity data, including time lags that accumulate during the processing of naturalized species data as well as typical points at which the data are harvested for naturalization or invasion trend analyses..	84
Figure 3.2 A comparison of field surveying and data processing lags showing that field surveying lags affect the interpreted naturalization rate at invasion onset whereas data processing lags affect recent years..	85
Figure 3.3 The effect of shortening and lengthening both field surveying and data processing lags on the inferred trends of accumulating naturalized species in comparison to 1000 simulated species naturalizing at a constant rate between 1920 and 2020	86
Figure 3.4 The influence variable data collection has on inferred naturalization trendlines, including intentional methodology (how data are harvested) and incomplete sampling.....	87
Figure 3.5 Comparison of trends in the accumulation of naturalized plants in Hawai‘i calculated from two independently compiled first records datasets.	88
Figure 4.1 Alignment of Blackburn et al.’s (2011) scheme with categories in Hawai‘i’s non-native species checklist with colored polygons representing alignment or misalignment.	127
Figure 4.2 Distribution of non-naturalized and naturalized scores based on WRA and likelihood-only scoring	128
Figure 4.3 Relationship between the number of Hawaiian Islands naturalized with WRA and likelihood-only scores	129

Chapter 1 – Introduction

Of the numerous human impacts that have intensified with the expansion of modern civilization, the effects of redistributing life on our planet are some of the most damaging and difficult to measure (Sage 2020). Major impacts have been caused by anthropogenic introductions of invasive species - those whose establishment and spread outside of their native range cause harm to the environment, human health or the economy (Capinha et al. 2015, Reaser et al. 2019). Before humans established across the globe, the passage of species over large geographic barriers was rare, requiring events like the formation of land bridges after sea level decline or rare dispersal events over exceptionally long distances (Vermeij 1991, Gillespie et al. 2012). However, frequent transport of species across these barriers by humans, be it on purpose or accidentally, has rapidly accelerated the rate of global species exchange, with diverse accounts of catastrophic impacts. These impacts include species extinctions, changes to basic ecosystem functioning, loss of human life from events like wildfire or flooding, as well as lower food production (Gilbert and Levine 2013, Fusco et al. 2019, Diagne et al. 2021). Given the high cost of controlling invaders after they have already established, many biologists have emphasized the need to focus on prevention rather than the continuous control of established species to mitigate impacts (Hulme 2009). Thus, we are faced with an urgent need to prevent this global shuffling of species, but how do we achieve this?

Those who study the outcome of global species exchanges have found that while some species successfully invade, others fail to establish or maintain small, relatively innocuous populations (Blackburn et al. 2011). Thus, a prominent goal of invasion biology has been to predict which species will become invasive, with the hope that if we cannot prevent species

introductions entirely, then perhaps we can at least limit introductions to those unlikely to do harm (Roy et al. 2018). These efforts have led to the development of sophisticated species risk assessment methods that can be applied in a variety of regions (Pheloung et al. 1999, Daehler et al. 2004, Blackburn et al. 2014, Bacher et al. 2018), and it is likely that integrating risk assessments into biosecurity policies would mitigate many risks imposed by would-be invaders.

However, two major caveats impede our ability to solve problems caused by biological invasions using species-level risk assessments, with the first problem relating to implementation itself. Curating a list of blacklisted species to prevent introductions becomes a difficult task when faced with the extraordinary diversity of non-native species in transport, especially for policy-makers who must consider the costs of completing assessments and reinforcing policies (Hulme 2009, 2020). Furthermore, screening out long lists of species becomes impractical, especially when unintentional “hitchhiking” species and taxonomic issues such as misidentification further complicate detection (Jacobs et al. 2017).

A more recent solution to prevent invasions has been to identify major pathways for invasive species’ introductions, including information on the origin of introduced species as well as reasons for their introduction (i.e. pet trade, agriculture; Essl et al., 2015). Pathway-level risk assessments, which are non-specific in the species they target, stand to provide an extra layer of protection on top of species-level risk assessments. As regulating a handful of pathways is likely easier than regulating thousands of species, work is needed to identify which pathways could introduce more invaders in the future and how to stall the flow of species through them.

The second complication associated with employing species-level risk assessments is that, while tools are available to anticipate which species might establish and spread, the magnitude and type of impacts inflicted are much more difficult to predict (Novoa et al. 2020).

Impacts from invaders are frustratingly context dependent, and information about the receiving location (i.e. climate, species interactions) must be considered alongside the invading species' traits and introduction-related factors to estimate potential impacts (Pyšek et al. 2020). This knowledge gap is particularly problematic for regions with numerous non-native species, as managers tasked with eradicating potential invaders before they have a chance to spread must prioritize which species are deserving of their limited funds. Armed only with the knowledge that hundreds of already-present species could become invasive and no tools to predict impacts, managers often default to qualitative descriptions of an invader's reputation elsewhere. Given that the number of first time invaders continue to accumulate at an accelerating pace worldwide, overreliance on reputation not only fails to account for the context-dependent nature of impacts, but is likely to overlook invaders establishing outside of their native range for the first time (Seebens et al. 2017).

The slow implementation of policies to prevent invaders from crossing borders and the lack of tools to predict impacts emphasizes the importance of monitoring non-native species upon arrival (Latombe et al., 2017). Not only is regular on-the-ground surveillance necessary to detect new invasive species and document impacts, but tracking a species' distribution and stage of establishment is necessary to assess the feasibility of containing or eradicating recently established species. Furthermore, these data form the basis of standardized indicators that assess whether invasive species management strategies or policies are effective, as well as predictive tools needed to better estimate which species or pathways deserve the most attention.

Recent integration of the field of data science with the study of biodiversity, including the standardization and increased accessibility of data, has uncovered biodiversity informatics as a promising set of tools to study and monitor invasions (Hardisty et al. 2013, Latombe et al. 2017,

Seebens et al. 2020). A key advantage of biodiversity informatics is that numerous species (or even all known species) can often be included in analyses, and due to extraordinary efforts to digitize historic publications and specimens in natural history collections, it is also possible to discern temporal trends (Peterson et al. 2015). Analyses based on many non-native species are beneficial for understanding invasions, because although many non-native species may never gain an invasive reputation warranting specific studies of their impacts, the inclusion of relatively innocuous species can increase the robustness of analyses seeking to identify important patterns such as the geographic origins of invaders and the reasons they are introduced (Hardisty et al. 2013). Analyses of naturalization rates (i.e. species that form self-sustaining populations outside of human confinement) have been used to identify temporal trends and historic periods of mass invasion, which can be extrapolated to understand threats from future invaders (Blackburn et al. 2011, Seebens et al. 2018, Bonnamour et al. 2021). Furthermore, analyses of naturalization rates over time have been proposed as a key metric to inform invasive species policies to help meet the goals of the Convention on Biological Diversity, a global treaty which stipulates a 50% reduction in the rate of new introductions by 2030 (Latombe et al. 2017, McGeoch et al. 2021).

Nonetheless, most advances in biodiversity informatics have focused on native species (Groom et al. 2017). The majority of workflows for monitoring non-native biodiversity change have been developed in the last 5 years (Latombe et al. 2017, Groom et al. 2019, Seebens et al. 2020). More work is needed to improve data curation and develop methodologies to understand the rate at which species are naturalizing and how to track their populations over time. Given that poorer regions of the world likely have fewer resources to efficiently collect and report data on invasive species, tools that account for data deficiencies are critical to avoid pitfalls in analyses

and encourage wide adoption outside of wealthier countries (Magona et al. 2018, Latombe et al. 2019).

Here I explore aspects of invasion prevention by studying patterns of terrestrial non-native vascular plant diversity in the Hawaiian Islands. Challenges associated with invasion prevention are exemplified in Hawai‘i because, like many oceanic islands, Hawai‘i is an “invasion hotspot” in which numerous non-native species have already been introduced (Sax and Gaines 2008, Dawson et al. 2017). Furthermore, a statewide biosecurity plan to slow introductions is currently being implemented and could benefit from research identifying opportunities for prevention (HIBP 2016). Due to its remoteness, Hawai‘i also hosts many endemic and endangered ecosystems, which stand to lose much of their biotic uniqueness as non-native species are introduced (Castro et al. 2010, Baiser et al. 2012). This uniqueness makes it difficult to understand potential impacts because few analogous regions exist elsewhere in the world, which emphasizes the need to monitor species after their arrival.

Non-native plants provide excellent study subjects because they are estimated to be deliberately introduced more frequently than other taxonomic groups (Turbelin et al. 2017), offering an opportunity to examine the reasons invasive species are introduced and how these reasons change over time. However, it is often difficult to track the progression of plant species along the introduction-invasion continuum which can complicate efforts to monitor them after their arrival. This difficulty arises from difficulty in identifying and detecting new populations, their history in cultivation, and long life histories that equate to long lags periods between introduction and naturalization (Crooks 2005, Wilson et al. 2014).

This PhD dissertation investigates which non-native plants have naturalized in Hawai‘i and why they were introduced alongside their rate of naturalization. Additionally, I examine

methodologies for calculating rates of naturalization and tracking the status of species after they arrive, uncovering pitfalls that affect their usefulness and applying potential fixes. The second chapter starts by compiling data on the origin, date of first record and likely introduction pathway for the entire non-native flora of the main Hawaiian islands, analyzing how rates of naturalization have changed alongside historic social or economic changes in Hawai‘i. The third chapter seeks to further understand the temporal sampling biases lurking in first record data, using computer simulations and a real-world case study to examine how lags in data processing and differences in how researchers harvest data can affect our interpretation of invasion trends. Finally, the fourth chapter explores the possibility of converting a well-accepted framework of invasion stages into a practical system for tracking the status of non-native plants, using a species checklist as a starting point. This latter chapter also uses data amassed from weed risk assessments (Daehler et al. 2004) to determine if information on species traits and potential impacts can be used to predict a likely status for data-deficient species. This work therefore contributes to the global call for invasion prevention by using biodiversity informatics as a tool to uncover patterns and develop data analysis and curation methods that could help slow the rate of new arrivals and avert impacts from recently arrived species.

References

- Bacher S, Blackburn TM, Essl F, Genovesi P, Heikkilä J, Jeschke JM, Jones G, Keller R, Kenis M, Kueffer C, Martinou AF, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Roy HE, Saul W-C, Scalera R, Vilà M, Wilson JRU, Kumschick S (2018) Socio-economic impact classification of alien taxa (SEICAT). *Methods in Ecology and Evolution* 9: 159–168. <https://doi.org/10.1111/2041-210X.12844>
- Baiser B, Olden JD, Record S, Lockwood JL, McKinney ML (2012) Pattern and process of biotic homogenization in the New Pangaea. *Proceedings of the Royal Society B: Biological Sciences* 279: 4772–4777. <https://doi.org/10.1098/rspb.2012.1651>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Blackburn TM, Essl F, Evans T, Hulme PE, Jeschke JM, Kuhn I, Kumschick S, Markova Z, Mrugala A, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Ricciardi A, Richardson DM, Sendek A, Vila M, Wilson JRU, Winter M, Genovesi P, Bacher S (2014) A Unified Classification of Alien Species Based on the Magnitude of their Environmental Impacts. *Plos Biology* 12: 11. <https://doi.org/10.1371/journal.pbio.1001850>
- Bonnamour A, Gippet JMW, Bertelsmeier C (2021) Insect and plant invasions follow two waves of globalisation. *Ecology Letters* 24: 2418–2426. <https://doi.org/10.1111/ele.13863>
- Capinha C, Essl F, Seebens H, Moser D, Pereira HM (2015) The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348: 1248–1251. <https://doi.org/10.1126/science.aaa8913>
- Castro SA, Daehler CC, Silva L, Torres-Santana CW, Reyes-Betancort JA, Atkinson R, Jaramillo P, Guezou A, Jaksic FM (2010) Floristic homogenization as a teleconnected trend in oceanic islands. *Diversity and Distributions* 16: 902–910. <https://doi.org/10.1111/j.1472-4642.2010.00695.x>
- Crooks JA (2005) Lag times and exotic species: The ecology and management of biological invasions in slow-motion1. *Écoscience* 12: 316–329. <https://doi.org/10.2980/i1195-6860->

12-3-316.1

- Daehler CC, Denslow JS, Ansari S, Kuo H-C (2004) A Risk-Assessment System for Screening Out Invasive Pest Plants from Hawai'i and Other Pacific Islands. *Conservation Biology* 18: 360–368. <https://doi.org/10.1111/j.1523-1739.2004.00066.x>
- Dawson W, Moser D, van Kleunen M, Kreft H, Pergl J, Pysek P, Weigelt P, Winter M, Lenzner B, Blackburn TM, Dyer EE, Cassey P, Scrivens SL, Economo EP, Guenard B, Capinha C, Seebens H, Garcia-Diaz P, Nentwig W, Garcia-Berthou E, Casal C, Mandrak NE, Fuller P, Meyer C, Essl F (2017) Global hotspots and correlates of alien species richness across taxonomic groups. *Nature Ecology & Evolution* 1. <https://doi.org/10.1038/s41559-017-0186>
- Diagne C, Leroy B, Vaissière A-C, Gozlan RE, Roiz D, Jarić I, Salles J-M, Bradshaw CJA, Courchamp F (2021) High and rising economic costs of biological invasions worldwide. *Nature* 592: 571–576. <https://doi.org/10.1038/s41586-021-03405-6>
- Essl F, Bacher S, Blackburn TM, Booy O, Brundu G, Brunel S, Cardoso A-C, Eschen R, Gallardo B, Galil B, García-Berthou E, Genovesi P, Groom Q, Harrower C, Hulme PE, Katsanevakis S, Kenis M, Kühn I, Kumschick S, Martinou AF, Nentwig W, O'Flynn C, Pagad S, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Roques A, Roy HE, Scalera R, Schindler S, Seebens H, Vanderhoeven S, Vilà M, Wilson JRU, Zenetos A, Jeschke JM (2015) Crossing Frontiers in Tackling Pathways of Biological Invasions. *BioScience* 65: 769–782. <https://doi.org/10.1093/biosci/biv082>
- Fusco EJ, Finn JT, Balch JK, Nagy RC, Bradley BA (2019) Invasive grasses increase fire occurrence and frequency across US ecoregions. *Proceedings of the National Academy of Sciences* 116: 23594–23599. <https://doi.org/10.1073/pnas.1908253116>
- Gilbert B, Levine JM (2013) Plant invasions and extinction debts. *Proceedings of the National Academy of Sciences* 110: 1744–1749. <https://doi.org/10.1073/pnas.1212375110>
- Gillespie RG, Baldwin BG, Waters JM, Fraser CI, Nikula R, Roderick GK (2012) Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution* 27: 47–56. <https://doi.org/10.1016/j.tree.2011.08.009>
- Groom Q, Desmet P, Reyserhove L, Adriaens T, Oldoni D, Vanderhoeven S, Baskauf SJ,

- Chapman A, McGeoch M, Walls R, Wicczorek J, Wilson J, Zermoglio PF, Simpson A (2019) Improving Darwin Core for research and management of alien species. *Biodiversity Information Science and Standards* 3: e38084. <https://doi.org/10.3897/biss.3.38084>
- Groom QJ, Adriaens T, Desmet P, Simpson A, De Wever A, Bazos I, Cardoso AC, Charles L, Christopoulou A, Gazda A (2017) Seven recommendations to make your invasive alien species data more useful. *Frontiers in Applied Mathematics and Statistics* 3: 13.
- Hardisty A, Roberts D, The Biodiversity Informatics Community (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* 13: 16. <https://doi.org/10.1186/1472-6785-13-16>
- HIBP (2016) Hawai‘i Interagency Biosecurity Plan 2017-2027.
- Hulme PE (2009) Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology*: 835–847. [https://doi.org/10.1111/j.1365-2664.2006.01227.x@10.1111/\(ISSN\)1365-2664.KEYPERS](https://doi.org/10.1111/j.1365-2664.2006.01227.x@10.1111/(ISSN)1365-2664.KEYPERS)
- Hulme PE (2020) Plant invasions in New Zealand: global lessons in prevention, eradication and control. *Biological Invasions* 22: 1539–1562. <https://doi.org/10.1007/s10530-020-02224-6>
- Jacobs L, Richardson D, Lepschi B, Wilson J (2017) Quantifying errors and omissions in alien species lists: The introduction status of *Melaleuca* species in South Africa as a case study. *NeoBiota* 32: 89–105. <https://doi.org/10.3897/neobiota.32.9842>
- Latombe G, Canavan S, Hirsch H, Hui C, Kumschick S, Nsikani MM, Potgieter LJ, Robinson TB, Saul W-C, Turner SC, Wilson JRU, Yannelli FA, Richardson DM (2019) A four-component classification of uncertainties in biological invasions: implications for management. *Ecosphere* 10: e02669. <https://doi.org/10.1002/ecs2.2669>
- Latombe G, Pyšek P, Jeschke JM, Blackburn TM, Bacher S, Capinha C, Costello MJ, Fernández M, Gregory RD, Hobern D, Hui C, Jetz W, Kumschick S, McGrannachan C, Pergl J, Roy HE, Scalera R, Squires ZE, Wilson JRU, Winter M, Genovesi P, McGeoch MA (2017) A vision for global monitoring of biological invasions. *Biological Conservation* 213: 295–308. <https://doi.org/10.1016/j.biocon.2016.06.013>

- Magona N, Richardson DM, Roux JLL, Kritzing-Klopper S, Wilson JRU (2018) Even well-studied groups of alien species might be poorly inventoried: Australian *Acacia* species in South Africa as a case study. *NeoBiota* 39: 1–29.
<https://doi.org/10.3897/neobiota.39.23135>
- McGeoch M, Arlé E, Belmaker J, Buba Y, Clarke D, Essl F, García-Berthou E, Groom Q, Henriksen M, Jetz W, Kühn I, Lenzner B, Meyer C, Pagad S, Pili A, Roigé Valiente M, Seebens H, Tingley R, Vicente J, Winter M (2021) Policy-relevant indicators for invasive alien species assessment and reporting. <https://doi.org/10.1101/2021.08.26.457851>
- Novoa A, Richardson DM, Pyšek P, Meyerson LA, Bacher S, Canavan S, Catford JA, Čuda J, Essl F, Foxcroft LC, Genovesi P, Hirsch H, Hui C, Jackson MC, Kueffer C, Le Roux JJ, Measey J, Mohanty NP, Moodley D, Müller-Schärer H, Packer JG, Pergl J, Robinson TB, Saul W-C, Shackleton RT, Visser V, Weyl OLF, Yannelli FA, Wilson JRU (2020) Invasion syndromes: a systematic approach for predicting biological invasions and facilitating effective management. *Biological Invasions* 22: 1801–1820.
<https://doi.org/10.1007/s10530-020-02220-w>
- Peterson AT, Soberón J, Krishtalka L (2015) A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC Ecology* 15: 15.
<https://doi.org/10.1186/s12898-015-0046-8>
- Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management* 57: 239–251. <https://doi.org/10.1006/jema.1999.0297>
- Pyšek P, Bacher S, Kühn I, Novoa A, Catford JA, Hulme PE, Pergl J, Richardson DM, Wilson JRU, Blackburn TM (2020) MAcroecological Framework for Invasive Aliens (MAFIA): disentangling large-scale context dependence in biological invasions. *NeoBiota* 62: 407–461. <https://doi.org/10.3897/neobiota.62.52787>
- Reaser JK, Burgiel SW, Kirkey J, Brantley KA, Veatch SD, Burgos-Rodríguez J (2019) The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. *Biological Invasions*.
<https://doi.org/10.1007/s10530-019-02156-w>

Roy HE, Rabitsch W, Scalera R, Stewart A, Gallardo B, Genovesi P, Essl F, Adriaens T, Bacher S, Booy O, Branquart E, Brunel S, Copp GH, Dean H, D'hondt B, Josefsson M, Kenis M, Kettunen M, Linnamagi M, Lucy F, Martinou A, Moore N, Nentwig W, Nieto A, Pergl J, Peyton J, Roques A, Schindler S, Schönrogge K, Solarz W, Stebbing PD, Trichkova T, Vanderhoeven S, van Valkenburg J, Zenetos A (2018) Developing a framework of minimum standards for the risk assessment of alien species. *Journal of Applied Ecology* 55: 526–538. <https://doi.org/10.1111/1365-2664.13025>

Sage RF (2020) Global change biology: A primer. *Global Change Biology* 26: 3–30. <https://doi.org/10.1111/gcb.14893>

Sax DF, Gaines SD (2008) Species invasions and extinction: The future of native biodiversity on islands. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11490–11497. <https://doi.org/10.1073/pnas.0802290105>

Seebens H, Clarke DA, Groom Q, Wilson JRU, García-Berthou E, Kühn I, Roigé M, Pagad S, Essl F, Vicente J, Winter M, McGeoch M (2020) A workflow for standardising and integrating alien species distribution data. *NeoBiota* 59: 39–59. <https://doi.org/10.3897/neobiota.59.53578>

Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pysek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jager H, Kartesz J, Kenis M, Kreft H, Kuhn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajero K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* 8. <https://doi.org/10.1038/ncomms14435>

Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pysek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brouckhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guenard B, Jager H, Kartesz J, Kenis M, Kuhn I, Lenzner B, Liebhold AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R,

- Schindler S, Stajerova K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Turbelin AJ, Malamud BD, Francis RA (2017) Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography* 26: 78–92. <https://doi.org/10.1111/geb.12517>
- Vermeij GJ (1991) When Biotas Meet: Understanding Biotic Interchange. *Science* 253: 1099–1104. <https://doi.org/10.1126/science.253.5024.1099>
- Wilson JRU, Caplat P, Dickie IA, Hui C, Maxwell BD, Nuñez MA, Pauchard A, Rejmánek M, Richardson DM, Robertson MP, Spear D, Webber BL, van Wilgen BW, Zenni RD (2014) A standardized set of metrics to assess and monitor tree invasions. *Biological Invasions* 16: 535–551. <https://doi.org/10.1007/s10530-013-0605-x>

Chapter 2 – Plant naturalization trends reflect socio-economic history and show a high likelihood of inter-island spread in Hawai‘i

Citation: Brock KC, Daehler CC (2021) Plant naturalization trends reflect socioeconomic history and show a high likelihood of inter-island spread in Hawai‘i. *Invasive Plant Science and Management* 14: 135–146. <https://doi.org/10.1017/inp.2021.18>

Abstract

The composition of non-native floras is influenced by a region’s socio-economic history, yet rarely are these factors studied alongside plant naturalization rates over time. Such information is especially critical for archipelagos, which often host large numbers of non-native plants and would benefit from preventing inter-island spread. We compiled the first record of occurrence and first record of naturalization for all naturalized plants in Hawai‘i alongside data on their origin, native climate types, taxonomy, and likely introduction pathway, and compared rates of naturalization with socio-economic trends. We found that the rate of total plant naturalizations has increased at a roughly constant rate during the past century without any sign of plateauing. However, this relatively steady increase is underlain by notable fluctuations in naturalization rates for different introduction pathways, with ornamentals increasing recently while agriculture-related plants have decreased. Furthermore, this trend mirrors a shift from an agriculture-dominated economy to a tourism-based one associated with increases in both resident and tourist populations as well as general economic well-being. We further found that the average naturalized species spreads at a rate of 1.86 islands per decade, eventually occupying most major islands in the archipelago, and the rate of spread appears to be increasing since Hawai‘i’s economic shift. Our findings also emphasize the diversity of Hawai‘i’s non-native flora, which originates from a variety of climates, continents, and taxonomic groups. We demonstrated that many non-native species have native ranges that include temperate climates, which is important

because these climates typically co-occur with higher elevation, remnant patches of native-dominated ecosystems in Hawai‘i’. This study reveals trends that may help predict a species’ ability to naturalize and spread within and between islands, and we discuss management implications that may be extended to other regions.

Box 1. Management Implications

Hawai‘i’s non-native flora is cosmopolitan in origin, and we uncovered no evidence to suggest that regulating or inspecting imports from any one continent would prevent most naturalizations. Similarly, Hawai‘i’s naturalized plants are native to various climatic zones, and although managers may currently recognize most of these species as invaders of tropical lowlands, our study shows that almost half could be equally at home in cooler, mid-upper elevations that often host remnant native-dominated habitats. We recommend increasing efforts to reduce establishment opportunities for non-natives in native-dominated habitats by limiting dispersal and disturbance. We also found that a relatively constant rate of new plant naturalizations over time is underpinned by fluctuations in naturalization rates among different introduction pathways. In particular, the naturalization rate for ornamental plants is increasing exponentially while naturalization of agricultural species has declined, corresponding with a historic shift from agriculture to a tourism-based economy, which is associated with increased tourist and resident populations as well as general economic well-being. These trends may be especially relevant to other tropical islands that have undergone (or will soon undergo) similar socio-economic changes (Russell et al. 2017, Brown et al. 2021). Hawai‘i should expect more ornamental naturalizations, which emphasizes the need for on-the-ground monitoring because

future invaders that have recently become popular in the horticultural trade could have no history of introduction or invasiveness to inform risk assessments. Finally, we found that many plants spread quickly within the archipelago, and although most recent introductions appear to be present on just 1–2 islands, the average rate of spread appears to be increasing. Thus, managers may capitalize on the limited distribution of new arrivals and prevent inter-island introductions if they act quickly (e.g., noxious weed list updates). However, the large number of species in this 1–2 island pool heightens the need for prioritization schemes that identify monitoring and control targets posing the greatest risk of becoming invasive (Brock and Daehler 2020).

Introduction

Plants are deliberately introduced outside of their native ranges more frequently than other taxonomic groups (Turbelin et al. 2017), and although preventing introductions is ideal, little guidance exists for managing the plethora of species that have already arrived. One silver lining is that areas rich in non-native species offer a chance to analyze trends and identify factors that promote naturalizations. Regions that have experienced numerous past invasions may be in the best position to tailor their biosecurity efforts to avoid future threats as new invaders continue to increase worldwide (Seebens et al. 2017). Furthermore, their invasion histories may offer useful lessons to areas that have so far experienced fewer invasions but may face increasing threats in the future (Brown et al. 2021).

An increase in the number of new invaders over the past centuries is well documented in many regions (Seebens et al. 2017), but few studies have investigated the temporal dynamics of these trends and how they correlate with socio-economic activities (van Kleunen et al. 2020). This knowledge gap represents a significant problem for biosecurity efforts, as it is difficult to determine whether our perceptions about the types of species and pathways most in need of management match the current reality (Essl et al. 2011). More generally, questions have arisen about whether we should expect species to continue to naturalize at their historic rates at the regional scale (Meyerson and Mooney 2007). Increased regulation of imports and greater awareness of the negative impacts from non-native species might result in a decreased rate of naturalization over time, whereas more frequent economic trade worldwide may result in increasing naturalization rates. Additionally, Seebens et al. (2018) propose that the rate of accumulation for non-native species may begin to slow if species source pools are exhausted, and alternatively, a rise in emerging invaders may be explained by increased accessibility to new species source pools via the expansion of global trade networks. A shift or expansion in species source pools brought on by economic changes may maintain high rates of naturalization even in regions that already host large numbers of naturalized species.

The capacity for these species to spread once they arrive or remain locally naturalized is even less understood. This issue is particularly important for oceanic archipelagos, which offer discrete inter-island borders that should theoretically aid biosecurity efforts (Weigelt et al. 2013). In the absence of policies that limit inter-island spread, it is important to determine whether spread within an archipelago is truly limited, or if natural inter-island isolation is essentially erased by inter-island trade. Another challenge faced by high oceanic islands and mountain regions alike is that new plant species are often initially introduced to lowland regions, which are

often heavily disturbed and populated by humans (Pauchard et al. 2009, Rojas-Sandoval et al. 2020). Meanwhile, less disturbed native-dominated habitats are more likely to persist at inaccessible upper elevations that correspond to cooler climates, which may be inhabited by numerous endemic species of conservation concern. This pattern warrants investigation into whether non-native species can spread to cooler, higher-elevation areas dominated by natives. Some studies have attempted to address this question by developing species distribution models, but these analyses are often limited to a small number of species that are selected based on *a priori* decisions about their invasive behavior or impacts. The lack of whole-flora analyses precludes an answer to whether native species at higher elevations are at risk of being substituted by hundreds of non-natives introduced at lower elevations over time.

In this study we consider the entire non-native flora of the Hawaiian Islands, an archipelago system particularly apt for such an analysis due to numerous non-native species and a dynamic socio-economic history. Additionally, Hawai‘i’s high proportion of endemic species and cultural richness adds urgency to finding biosecurity strategies that can address invasive species problems. The trends and origins of rapidly accumulating non-native plants in Hawai‘i were examined between 1800 and 1980 by Wester (1992), who noted a correlation between economic activities and introduction rates over time. Since then, the number of known naturalized plants in Hawai‘i has more than doubled and significant effort has been made to compile data on the naturalization status of species on each island (Wagner et al. 1999, Imada 2019). These new data, combined with the global digitization of herbarium specimens, allows for a more detailed analysis of naturalization rates and further allows us to examine the spread of species throughout the archipelago. First, we investigate the composition of Hawai‘i’s flora, asking: do species tend to originate from certain plant families, continents, or climates, and how

many are capable of growing at higher elevation, cooler climate types where natives currently tend to dominate? Second, we analyze dates of first naturalization records alongside probable introduction pathway and corresponding socio-economic events, asking: have naturalization rates increased or decreased, and do these shifts correspond to socio-economic changes? Finally, we investigate dates of first known introduction to consider the number of islands on which non-native plants are present, asking: how does time or introduction pathway influence inter-island spread?

Materials and Methods

A list of naturalized species was compiled based on Imada (2019) for six islands in the Hawaiian archipelago: Kaua'i, O'ahu, Molokai, Lānai, Maui, Hawai'i. The island of Kaho'olawe and all of the Northwest Hawaiian Islands were excluded because they are uninhabited and geographically less complex relative to the other islands. The island of Ni'ihau was also excluded because, while it hosts a small population of people, plant surveys are infrequent due to its private ownership. We defined naturalized as non-native species that survive and reproduce naturally outside of cultivation over many generations (Richardson et al. 2000, Pyšek et al. 2004, Blackburn et al. 2011), and thus, 225 plants considered questionably naturalized according to Imada's list (2019) were excluded from analyses.

Year of first collection and year of naturalization were retrieved for all naturalized plant species in Hawai'i by examining digitized herbarium voucher labels and new naturalization reports (Evenhuis and Miller 2015), which often include biogeographic information and population status. Specimen data were retrieved by querying the Bernice Pauahi Bishop Museum (BISH), National Tropical Botanical Garden (PTBG) and Smithsonian U.S. National herbaria databases as well as the Global Biodiversity Information Facility (GBIF 2020), which included

specimens from an additional 81 institutions. This compiled dataset included 77,980 herbarium specimens of species on our naturalized checklist. Based on the recognized plant use(s) or non-use, each species was categorized according to their probable introduction pathway into Hawai'i as follows: accidental, food/medicine/textile, forestry, and ornamental/landscaping. These pathways were inferred mainly using the CABI Invasive Species Compendium (CABI www.cabi.org/isc/) or other early records of plant use, and descriptions on the voucher labels of early specimens. Continent of origin, plant family and taxonomic data were obtained from the Plants of the World database (POWO 2019), and life forms were classified as shrub, herb, tree, vine, succulent or aquatic based on descriptions from floras (Wagner et al. 1999, eFloras 2017) to analyze the relative proportion of each category within Hawai'i's non-native flora. We additionally explored other ways of categorizing continents of origin and compared the results for single versus multi-continent groupings. To determine if plants introduced through certain introduction pathways were more likely to originate from multiple continents, we compared the observed versus expected number of species ranging across 1, 2 or >2 continents using a chi-square test of independence. Post hoc analyses of this test and all subsequent chi-square tests were Bonferroni adjusted for multiple comparisons.

Naturalization rates were examined by plotting the accumulation of species over time based on year of naturalization using the `purrr` (Henry and Wickham 2020) and `ggplot2` (Wickham 2016) packages in R. We analyzed dates from 1910, which coincides with the beginning of plant voucher collection for local herbaria, until 2010, to exclude a potential lag time in depositing recently vouchered specimens. Rates of increase over time were calculated by dividing the increase in the number of naturalized plants by the time elapsed between dates. To determine whether naturalization rates have increased or decreased in recent years, we fitted the

data to linear, exponential, and saturating (logistic growth) curves using the `nls` function in the `stats` package in R v 4.0.3 (R Core Team 2020) and compared models using the Akaike Information Criterion (AIC). To determine whether sampling effort may have affected our results, we considered the number of native plant vouchers deposited at the Bishop Museum annually as a proxy for sampling effort over time. We analyzed collection trends between 1910 and 2010 and compared the number of vouchers collected with the number of new naturalization reports at 1-, 5- and 10-year intervals.

To understand what factors may have affected naturalization rates, we reviewed the literature describing Hawai‘i’s economic history and compared potentially significant events with naturalization trends. We assessed possible economic drivers of naturalization rates for each introduction pathway by conducting linear regression analyses in R (R Core Team 2020). Our independent variables included population, tourism, and agriculture during 1900–2020, measured as the number of residents, the number of tourists, and acres of farmland in production per decade. For the first two, we derived data from the Hawai‘i Department of Business, Economic Development & Tourism (DBEDT (2020a), DBEDT (2020b)), and the third was obtained data from the National Agricultural Statistics Service (NASS 2020). A fourth explanatory variable was real gross domestic product (GDP), which was adjusted for inflation to 2019 US dollars, with values prior to 1970 from Schmitt (1977) and thereafter from DBEDT (2020a). Where Shapiro-Wilks tests showed that data were not normally distributed, we log transformed these data prior to analysis. Additionally, we performed a Breusch-Godfrey test in conjunction with our linear regression analyses to ensure our inferences were not biased by serial autocorrelation of the residuals.

To understand the ability of non-native plants to invade Hawai‘i’s remaining native ecosystems, we compiled native range climate data for each non-native species based on the Köppen-Geiger climate classification (Beck et al. 2018). This system groups climates into five main classes: tropical (warm year-round; average monthly temperature $>18\text{ }^{\circ}\text{C}$ + significant precipitation), arid (very little annual precipitation, calibrated according to temperature), temperate (cool; at least one month between $10\text{--}18^{\circ}\text{C}$), continental (seasonally cold; with at least one month $<0^{\circ}\text{C}$) and polar (very cold; summers not $>10^{\circ}\text{C}$). Köppen-Geiger classifications are based on temperature and precipitation rather than geography, meaning that elevation and latitude may interact to produce a range of climate types not predictable by latitude alone (Beck et al. 2018). For instance, this system could classify a climate as temperate whether it is from temperate latitudes (between subtropical and polar geographical zones) or from temperate elevations within tropical or subtropical geographic zones. The potential for each species to grow in each of these conditions was inferred by retrieving the Köppen-Geiger climate classification for individual occurrence records in their native range such that species were allowed to be classified as having >1 climate type. Native range descriptions were harvested from numerous floristic studies compiled in the Plants of the World database (POWO 2019) using the taxize package in R (Chamberlain et al. 2020b) to construct native range polygons based on geopolitical boundaries. Occurrence records within these polygons were retrieved from the Global Biodiversity Information Facility (GBIF 2020) using the rgbif package in R (Chamberlain et al. 2020a) and filtered to exclude cultivated species and duplicate records. If 10 or more occurrences were retrieved with accurate location data (<1 km uncertainty), the kgc package (Bryant et al. 2017) was used to retrieve the Köppen-Geiger classification for each occurrence. Climate classifications based on fewer than 10 occurrences were double-checked against climate

and range descriptions in floras and the Plants of the World database (POWO 2019) and cross referenced with a 1-km resolution map of Köppen-Geiger climates (Beck et al. 2018). We then compared the resulting data set for the non-native flora to maps of native-dominated ecosystems in Hawai‘i (Jacobi et al. 2017) combined with Köppen-Geiger climate types (Beck et al. 2018) (Figure 2.1) using the GeoPandas package in python (Jordahl 2014).

To investigate whether the rate of inter-island spread varied among introduction pathways, and whether these rates have changed over time, we calculated the average number of inter-island introductions per decade. This rate was based on records of first occurrence on each island, which allowed us to determine the length of time each species took to be recorded on a new island. The occurrence dataset included the aforementioned herbarium specimens (described above), as well as research-grade citizen scientist observations from iNaturalist via GBIF (2020), which added 137,209 observations. We further determined whether species introduced earlier have spread to more islands than recently introduced species by tallying the number of islands where a species is present and where it has naturalized by analyzing records binned by 25-year time intervals, starting with the onset of consistent herbarium collection in 1910 and ending in 2010. To understand the extent to which species are simply present as opposed to fully naturalized, we relied on reports describing population status by field botanists on each island, which are compiled in Imada (2019). Species present in the occurrence data set but not reported as naturalized by field botanists were considered to be present, but not necessarily naturalized (e.g., only in cultivation). Thus, species were categorized species as either “present” or “not present” on an island, with present species further subcategorized as “naturalized” or “not naturalized” species. The observed versus expected number of species for each time period was then compared using chi-square tests of independence. If time has no effect on a species ability

to spread between islands, then we would expect the proportion of species on just 1–2 islands to be no larger for recently introduced species than for earlier introductions.

Results and Discussion

Composition and Origin of Naturalized Flora

Naturalized species represent at least 55% of the total plant diversity throughout the archipelago (excluding species found only in cultivation), and thus Hawai‘i joins the ranks of 82 other islands or island systems (26% of all islands worldwide) known to host more naturalized species than native species (Essl et al. 2019). The 1,488 species that have naturalized in the main Hawaiian Islands belong to 165 families and 789 genera, corresponding to a species density of 0.89 naturalized species per 10 km². Of these taxa, 70% are eudicots (including basal eudicots), 26% are monocots, 3% are pteridophytes and 1% are gymnosperms. Poaceae (206 species), Fabaceae (151 species) and Asteraceae (120 species) have the highest species richness, comprising 32% of all naturalized species. This composition is similar to both continental and insular floras in tropical or subtropical areas (Dodd et al. 2015, Rojas-Sandoval and Acevedo-Rodriguez 2015, Ansong et al. 2019), as well as patterns in the global naturalized flora. Although these families are among the most speciose families worldwide, others have calculated that Poaceae and Fabaceae are proportionally overrepresented relative to family size (Pyšek et al. 2017). An additional 22 families (13%) have 15 species or more (Figure 2.2A), while 50 families (31%) are represented by a single species. Twelve genera contain 10 or more species (1.5%), with the top five being *Eucalyptus* L'Hér. (Myrtaceae; 28 species), *Cyperus* L. (Cyperaceae; 22 species), *Paspalum* L. (Poaceae; 18 species), *Euphorbia* L. (Euphorbiaceae; 17 species), with *Crotalaria* L. (Fabaceae; 15 species) and *Eragrostis* Wolf (Poaceae; 15 species) tied for 5th place. However, the majority of genera (65%) contain a single naturalized species.

Naturalized species in Hawai‘i display a diversity of life forms, with 39% being herbs, 19% trees, 17% graminoids, 12% shrubs, 9% vines, 2% succulents, and 2% aquatic (Figure 2.2B). These patterns mirror those observed elsewhere; for instance, naturalized species in Puerto Rico and the Virgin Islands are predominantly herbs and trees, although they have proportionally more woody species (trees + shrubs) in comparison to Hawai‘i, which has somewhat more herbs and grasses (Rojas-Sandoval and Acevedo-Rodriguez 2015).

No continent emerges as a dominant donor of naturalized species, with 13% originating from Asia, 11% from North America, 10% from Oceania, 9% from South America, 8% from Africa, and 1% from Europe, and almost half of the species have native ranges that span multiple continents (47%; Figure 2.3A). This pattern is similar to that found by van Kleunen (2015) when quantifying plant exchanges globally, albeit with fewer European introductions (20.3% Asia, 17.1% North America, 15% Europe, 14.3% Africa, 12.8% South America), revealing the cosmopolitan nature of Hawai‘i’s non-native species source pool. The precise origin of naturalized species is difficult to determine because species may be sourced from anywhere within their ranges (many of which span multiple continents) and they may not have arrived directly from their home continent. However, the lack of a single geographical region identifiable as a dominant donor of species in Hawai‘i (even when broader, multi-continental ranges were explored) contrasts studies of other non-native floras. In New Zealand, for example, 49% of species are from Eurasia (Gatehouse 2008), and the Americas are the source of most of the non-native species in Ghana (Ansong et al. 2019), and the source of 50% of those in Puerto Rico and the Virgin Islands. Although the Americas are also the largest donor region for Hawai‘i (20%), they did not contribute much more than any other region. These comparatively even proportions indicate that a naturalized species in Hawai‘i has an almost equal likelihood of

originating from any continent, and that the non-native flora likely includes a diverse array of ecological adaptations and evolutionary histories.

Our investigation into the climatic origins of Hawai‘i’s naturalized species revealed that over 83% (1240 species) are native to temperate climates according to the Köppen-Geiger system, but most of these species show broad climate preferences, of which 66% also occur in tropical climates within their native range (Figure 2.2C). Only 16% of species appear to be obligate tropical species, although the majority of all species (64%) occur in this climate type in their native ranges. Fewer species are native to arid (31%), continental (12%) and polar (<1%) climates (Figure 2.2C). Thus, most non-natives are likely able to grow in tropical lowlands as well as in temperate areas in Hawai‘i. These findings are especially relevant alongside the patterns shown in Figure 2.1, where native-dominated ecosystems comprise only 37% of the land area in the main Hawaiian Islands, and 70% of these patches are classed as temperate, as they tend to persist at cool mid–upper elevations (Jacobi et al. 2017, Beck et al. 2018). Our data reveal that the current presence of intact, relatively uninvaded native ecosystems at these elevations is not explainable by a lack of non-native species that can tolerate these climates, and could be eventually be invaded by temperate species that have so far only established in Hawai‘i’s lowlands.

The majority (88%) of temperate patches in Hawai‘i are subclassified as having a temperate oceanic climate (*Cfb*), and 11% of the temperate patches are warm-summer Mediterranean (*Csb*), both of which are defined as having the warmest month averaging below 22 °C, although the former has no dry season while the latter has a dry summer season (Beck et al. 2018). In contrast, 95% of non-native dominated habitats exist in the tropical lowlands, and most species were likely introduced to these areas because they coincide with human settlements

[as has been demonstrated for other islands (Rojas-Sandoval et al. 2020)]. One limitation of climate classes is that each class spans a range of environmental gradients, and it cannot be assumed that species are climatically suited to the entire binned classes as mapped onto Hawai‘i (Figure 2.1). For example, the humid subtropical temperate subclass (*Cfa*) accounts for merely 0.5% of native-dominated habitats in Hawai‘i, and is defined as having no dry season and the warmest month averages over 22 °C (i.e. warmer than the *Cfb* and *Csb* classes; Beck et al. 2018). Thus, species restricted to the *Cfa* class (or warmer) in their native range are unlikely to be able to spread throughout Hawai‘i’s native-dominated temperate zones, although we determined that only 12% of naturalized species were so restricted. Nonetheless, 59% of species broadly classified as temperate in our study exactly match the dominant climate subclasses of Hawai‘i’s native-dominated temperate areas (*Cfb* and *Csb*), amounting to 49% of total naturalized species. Therefore, almost half of Hawai‘i’s naturalized species are probably capable of invading temperate native-dominated areas, and perhaps more if some species can spread in climate types not found within their native ranges. These findings are reflected by Daehler (2005), who found that hundreds of species have already naturalized above 1200 m, which roughly corresponds to the beginning of the temperate zone.

However, it is worth noting that our climate class analysis may be subject to the sampling biases associated with GBIF occurrence records, where sampling is often denser in wealthier nations. The variety of climate classes occupied by species native to affluent countries in our analyses are likely better represented than species native to poorer ones, and this may have underestimated the number of species that occupy tropical climate classes (as many poor nations exist in tropical areas). However, this bias is likely less important for temperate climate classes, which are well represented in developed countries like North America, Europe and Australia.

The majority of species were likely introduced deliberately, with 40% introduced for ornamental or landscaping purposes, 16% for food, medicine or textiles, 6% for forestry, and another 4% as cultivated species that do not fit within these categories. Likely accidental introductions account for 31% of all species, while the introduction pathway of another 4% of species is unknown. Probable introduction pathway is associated with the number of continents covered by a species' native range (χ^2 (df = 6, N = 1357) = 164.43, $p < 0.001$), and post hoc tests revealed that accidental introductions are significantly more likely to have a native range that spans multiple continents, with just 33% of accidentals being native to a single continent while 42% are native to 2 continents and another 24% are native to > 2 continents (Figure 2.3B). In contrast, plants selected for ornamental or landscaping purposes are significantly more likely to originate from single continents, with 64% from a single continent while 29% and 6% are native to 2 and > 2 continents, respectively. This trend is even more pronounced for forestry species, with 84% of species being native to a single continent with many of these introductions originating from Oceania.

This association between introduction pathway and native continental range suggests that introduction pathways may play different roles in how the homogenization of global biodiversity will unfold (McKinney and Lockwood 1999, Winter et al. 2009), and should be investigated further. The tendency for accidentally introduced species to have multi-continental ranges may be explained by these species having a larger geographic donor area, or by having life history characteristics that enable them to spread widely. Nonetheless, accidentally introduced species tend to already be represented on multiple continents by their native ranges, while the opposite is true for ornamental and forestry species. Horticulturalists may be more likely to seek out seemingly unique ornamentals, which may correspond to these species being native to a limited

geographical range. For example, we determined that 35 naturalized species in Hawai'i are endemic to other islands (data not shown). Thus, horticulture may play a disproportionate role in spreading species whose native ranges are small, which challenges the notion that the agents of global biotic homogenization are exclusively species with large native ranges (McKinney and Lockwood 1999, Newbold et al. 2018). Additionally, species that are rare worldwide may not always have a lower likelihood of being transported to another region, which has been assumed in models projecting global species accumulation rates (Seebens et al. 2020). Further investigation into this phenomenon would be particularly interesting for islands, which host globally unique native (i.e., endemic) species.

Rates of Naturalization

The increase in the number of naturalized species over time is roughly linear at an average rate of 12 species/year during 1910–2010 (Figure 2.4A), or possibly increasing at a modest rate. Comparisons between trendline shapes revealed that the data fit an exponential model ($AIC = 1059$) somewhat better than a linear line ($AIC = 1089$), although the growth constant of the fitted exponential line was small (1.27% per year; data not shown). Given this small rate of increase alongside the uncertainty associated with the true timing of naturalization, we chose to interpret the overall trend as approximately linear with the possibility that the naturalization rate has increased above 12 species/year in the last 30 years. Furthermore, we found no evidence to suggest that the increase in naturalized species reflects an increase in sampling effort over time, as the number of vouchers collected for herbaria was not correlated with the number of new naturalization reports ($R^2 = 0.03$, $p = 0.08$). Collection for herbaria has not increased over time, but instead shows variation with peaks in 1925–1935 and the 1970s and 1980s as well as notable slumps in 1938–1945 and 1990–2005. Increases in plant families and genera are also roughly linear over time, corresponding to an average increase of 6 genera/year

and 1.2 families/year (data not shown). The roughly linear accumulation of non-native species in Hawai‘i is comparable to findings for other islands such as New Zealand, which shows a slight increase in the rate of naturalization after 1940 (Hulme 2020), but contrasts with Puerto Rico and the Virgin Islands, which show a marked deceleration after the early 1900s (Rojas-Sandoval and Acevedo-Rodriguez 2015). Our data do not suggest that the overall rate of naturalized species accumulation is rapidly accelerating, which some have predicted due to expanded global trade networks in recent decades (Meyerson and Mooney 2007, Hulme et al. 2009). We also did not find evidence that the number of non-native species in Hawai‘i is beginning to approach saturation, which was recently projected for the Pacific Island region based on an analysis of potential species source pools and naturalization rates (Seebens et al. 2020).

While the overall naturalization rate has remained relatively constant over time, naturalization rates associated with different introduction pathways have varied and correlate with historic socio-economic activities (Figure 2.4B–F; Box 2). Notably, naturalization rates track a shift from an agriculture-based economy to one fueled by tourism. Plants introduced for food accumulated rapidly between 1910 and 1940, corresponding to 3.1 species/year, and then slowed to 1.24 naturalizations per year between 1940 and 2010, a 60% rate decrease. Subsequent model comparisons for food plant naturalization over time supported a saturating curve ($AIC = 803$) as a better fit over the linear model ($AIC = 891$). Rate of naturalization for food plants was positively correlated with acres of productive farmland ($R^2 = 0.47$, $p = 0.020$), which began to decrease in the early–mid 1900s. In contrast, the rate of naturalizations of ornamentals during the study period fits an exponential growth curve ($AIC = 842$) over a linear model ($AIC = 1131$), eventually overtaking food plants around 1960. Ornamental naturalizations then noticeably accelerated around 1980, increasing from 3 species/year before 1980, to 10.2 species/year after

1980 (a 240% increase), and their rate of naturalization is positively correlated with human population ($R^2 = 0.73$, $p < 0.001$), GDP ($R^2 = 0.68$, $p = 0.002$), and tourism ($R^2 = 0.68$, $p = 0.004$). Thus, the approximately linear accumulation of species overall is maintained by recent increases in the rate of naturalization for ornamental species, which compensates for the decreasing food plant naturalizations. Breusch-Godfrey tests confirmed that serial autocorrelation did not exist in our residuals for these, or any of the subsequent linear regression analyses examining the relationship between naturalization rate and socio-economic variables.

GDP, tourism and population positively covaried in our analyses (population vs. tourism: $R^2 = 0.80$, $p < 0.001$; GDP vs. tourism: $R^2 = 0.88$, $p < 0.001$; GDP vs. population: $R^2 = 0.82$, $p < 0.001$), and therefore, it is difficult to pinpoint whether the increased naturalization of ornamentals is a direct result of tourism or increased economic well-being that coincided with this economic shift. However, it seems likely that an increased demand for ornamentals may have partially arisen from a desire to increase the aesthetic appeal of the islands for tourists. Ayala (1991) describes a late 1900s paradigm shift in the international tourism industry to create large resort landscapes that are billed as alluring destinations unto themselves, citing specific examples of Hawaiian resorts that introduced thousands of plants and animals to generate a sense of exotism and luxury. Nevertheless, economic growth from tourism investments not only increased infrastructure to support visitors, but also improved the financial well-being of its growing population (Figure 2.4C–E). An increase in disposable incomes for both residents and local governments (via tax revenues) likely promoted spending on amenities and beautification not only for tourists. Furthermore, other studies confirm that naturalized plant richness is commonly correlated with measures of human wealth (Denslow et al. 2009). Regardless of the exact reasons for introduction, our analyses emphasize the role of the ornamental and

landscaping industry as an important source of naturalized species, and this finding is reflected by several studies of islands and continents alike where ornamentals make up the largest portion of naturalized floras (Crawley et al. 1996, Pyšek et al. 2002, Silva and Smith 2004, Dodd et al. 2015, Rojas-Sandoval and Acevedo-Rodriguez 2015). Our data from Hawai‘i, as well as similar findings in New Zealand (Hulme 2020), suggest that the dominance of ornamentals in naturalized floras may be a recent phenomenon, which could have implications for studies on how plants are exchanged globally if this trend is found to exist in other regions (van Kleunen et al. 2015).

Prior to being surpassed by ornamentals around 2000, accidentally introduced species were the most numerous group, maintaining a roughly linear trend between 1910 and 2010 that corresponds to an average rate of 3.3 accidental naturalizations/year (Figure 2.4B). These data fit a linear model ($AIC = 1086$) better than an exponential increase model ($AIC = 1234$), and slightly better than a saturating (logistic) model ($AIC = 1094$). This rate of naturalization is positively correlated with human population ($R^2 = 0.84$, $p < 0.001$), which similarly experienced linear growth. Although comparatively fewer in number, the rate of naturalization for forestry species best fits an exponential growth curve ($AIC = 564$) over a linear one ($AIC = 785$). Between 1910 and 1970, the mean rate of forestry species was 0.32 naturalized per year versus 1.49 species per year after 1970, representing a 365% increase. Nonetheless, it is surprising to find that forestry has contributed significantly fewer naturalized species overall than the other introduction pathways given the intense reforestation effort that took place in Hawai‘i in the early 1900s (Woodcock 2003; Box 1). The increased rate in recent years may reflect a substantial lag phase between introduction for major reforestation efforts in the early–mid 1900s and naturalization, which may be expected as woody species often take longer to mature (Daehler

2009). Alternatively, newer forestry projects in Hawai‘i, such as *Eucalyptus* plantings for biofuels (DeBell et al. 1989), may be the source of recent forestry naturalizations, and should be investigated further.

Box 2. Post-1900 socio-economic events likely affecting plant naturalizations in Hawai‘i

After Hawai‘i was annexed as a US territory in 1900, the US government recruited thousands of plantation workers from all over the world to tend sugar cane, which was Hawai‘i’s dominant industry in the early 1900s (La Croix 2019). Workers brought familiar food plants with them (La Croix 2019), including species consumed in Japan, Korea, the Philippines, Spain, Portugal, Puerto Rico, England, Germany and Russia. The sugar cane industry peaked in 1926 while pineapple production was expanding and during this time, accidental introductions likely took the form of agricultural weeds, which were able to naturalize among vast tracts of land recently converted to agricultural use (Perroy et al. 2016). Despite economic focus on crop production, plants cultivated commercially for direct human consumption were not very diversified (Perroy et al. 2016). However, livestock forage trials were conducted at agricultural research stations on multiple islands in the early 1900s and many grass species and other fodder plants were introduced that have since naturalized (Henke 1945, Work 1946).

As a result of intensive agricultural practices and spread by feral ungulates, the declining state of native forests was increasingly noticed in the early 1900s, raising concerns about the fresh water supply and soil erosion (Woodcock 2003). To solve this problem, tens of millions of trees and other plants were planted in forest reserves from 1910–1960 (especially

between 1919 and 1940), comprising over 1000 species (Skolmen 1980). However, an increased naturalization rate for forestry trees is not detectable until well after this period (~1960), and less than 10% of these forestry species are now considered naturalized (Imada 2019).

The Great Depression in the 1930s followed by World War II was associated with economic stagnation, and fear of encountering enemy ships reduced oceanic trade, which may have resulted in fewer plant introductions (La Croix 2019). However, a flood of armed service personnel may have accidentally introduced other species. Around the same period, agriculture slowed due to increased worker rights and the cost of labor, as well as competition with foreign markets (Linnekin 1997), which likely corresponded with fewer introductions of food plants. However, the attainment of US statehood in 1959 demarcates a shift to an economy dominated by the tourism industry and rapid commercial expansion, which is associated with a rapid increase in the naturalization of ornamental species (Linnekin 1997, La Croix 2019).

Hawai'i experienced its strongest economic growth between 1958 and 1973, and commercial flights to Hawai'i, which were previously only affordable by the upper class, suddenly became affordable to the middle class (La Croix 2019). This drastically increased the number of visitors and fueled the growth of infrastructure to accommodate them, likely increasing the demand for ornamentals to improve the aesthetic appeal of Hawai'i and its new tourism-based economy (Ayala 1991).

Inter-island Spread

We found that the average rate of inter-island spread for all species was 1.86 islands/decade (sd = 1.84) between 1910 and 2010, although rates of spread varied widely among

species. Additionally, the rate of inter-island spread appears to increase at roughly the same time as Hawai‘i’s socio-economic shift in the middle of the 20th century such that rates in recent years are > 3 islands/decade (Figure 2.5A). Rates of inter-island spread over time are somewhat variable among introduction pathways (analysis not shown), with the rate of inter-island spread for food plants remaining constant in comparison to other pathways. Additionally, the spread of forestry plants increased during the early 1900s, which may be attributable to mass reforestation efforts during this time, while inter-island spread for other pathways declined.

Two sources of sampling bias may have affected estimated rates of inter-island spread. First, our data show an elevated rate of inter-island spread at the very beginning of the 1900s, coinciding with the first consistent collection of plants for the Bishop Museum herbarium in Honolulu (Figure 2.4A); collectors likely found many long-naturalized species on multiple islands in quick succession, translating artificially into high rates of inter-island spread until those species were collected. Secondly, the apparent acceleration of inter-island spread could be the outcome of the increased focus on documenting naturalizing cultivated species due to increased public awareness of these plants as invasives (Tye 2006). However, our analysis of naturalization rates alongside socioeconomic trends lends credit to our conclusion that the ornamental naturalizations have legitimately increased. Both naturalization rate and inter-island spread rate for ornamentals coincide with the rise of GDP, tourism and human population growth in Hawai‘i, suggesting that socio-economic factors are at least partially responsible for these trends. Additionally, if inter-island spread trends for ornamentals was merely a sampling artefact, we might expect them to be the only group displaying an accelerated inter-island spread rate, but this is not the case. An increased rate also exists for accidentals, which have long been recorded

in floras (Hillebrand 1888), and likely also reflect increased inter-island trade due to economic activity (Denslow et al. 2009).

The distribution of species on multiple islands (within our 6-island data set) is significantly associated with time since first record, both when considering a species' naturalized distribution (χ^2 (df = 6, N = 1369) = 403.22, p < 0.001), or where it is present but not necessarily naturalized (χ^2 (df = 6, N = 1369) = 506.22, p < 0.001). For instance, of species first recorded in the archipelago prior to 1910, 68% are present and 62% are naturalized on 5–6 islands (Figure 2.5B–C). In contrast, most recent introductions have 1–2 island distributions. For example, 84% of species first recorded between 1985 and 2010 are naturalized on merely 1–2 islands, but when accounting for a species' presence alone (not necessarily naturalized), the proportion of species is smaller (75%). Although ornamental plants are more likely to be restricted to 1–2 islands, and accidentals and food plants are more likely to occur on 5–6 islands, these distributions appear to be the result of the timing of introduction because most ornamentals were introduced recently while the majority of food or accidental species were introduced earlier (analysis not shown). Thus, the number of islands where a plant is present does not vary according to introduction pathway after accounting for time. Our results support the findings of a recent analysis in Hawai'i showing that Weed Risk Assessment scores, which are mostly based on biological characteristics, are only weakly correlated with inter-island spread (Brock and Daehler 2020). This suggests that human activities are more important than biological traits when predicting inter-island spread for many species.

Conclusion

We show that Hawai'i's non-native plants are drawn from a diversity of continents, climates and taxonomic groups, and that introductions associated with past socio-economic

activities still form a major component of the current flora (Essl et al. 2011). Seebens et al. (2018) contend that the persistent increase in new naturalized species worldwide cannot be solely explained by the upkeep of current economic activities, but that expanding trade networks (e.g., via socio-economic shifts) can increase the likelihood that new species will be introduced and expand the number of geographic regions that species can be sourced from. Our analyses build on this idea by detailing the approximate timing of these socio-economic events and demonstrate the emergence of new dominant introduction pathways that allow new species to naturalize. The horticultural industry may be particularly adept at increasing access to new species source pools if plant enthusiasts continue to drive demand for new and unusual species (van Kleunen et al. 2018). After introduction to an archipelago, species may spread readily, both within an island because they tend to have broad climatic tolerances in their native ranges, and between islands because humans purposely or accidentally transport them. The Hawaiian archipelago needs stringent biosecurity plans to stem the accelerating inter-island spread of plants, which is bolstered by continuous new additions to the already-dominant non-native flora.

References

- Ansong M, Pergl J, Essl F, Hejda M, van Kleunen M, Randall R, Pyšek P (2019) Naturalized and invasive alien flora of Ghana. *Biological Invasions* 21: 669–683.
<https://doi.org/10.1007/s10530-018-1860-7>
- Ayala H (1991) Resort hotel landscape as an international megatrend. *Annals of Tourism Research* 18: 568–587. [https://doi.org/10.1016/0160-7383\(91\)90075-M](https://doi.org/10.1016/0160-7383(91)90075-M)
- Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* 5: 1–12. <https://doi.org/10.1038/sdata.2018.214>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JR, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Brock KC, Daehler CC (2020) Applying an invasion and risk framework to track non-native island floras: a case study of challenges and solutions in Hawai‘i. *NeoBiota* 62: 55–79. <https://doi.org/10.3897/neobiota.62.52764>
- Brown ME, Prieto RO, Corbin JD, Ness JH, Borroto-Páez R, McCay TS, Farnsworth MS (2021) Plant pirates of the Caribbean: is Cuba sheltered by its revolutionary economy? *Frontiers in Ecology and the Environment* 19:208–215. <https://doi.org/10.1002/fee.2311>
- Bryant C, Wheeler NR, Rubel F, French RH (2017) kgc: Koeppen-Geiger Climatic Zones. Available from: <https://CRAN.R-project.org/package=kgc> (October 29, 2020).
- CABI (www.cabi.org/isc/) Invasive Species Compendium.
- Chamberlain S, Barve V, McGlinn D, Oldoni D, Desmet P, Geffert L, Ram K (2020a) rgbif: Interface to the Global Biodiversity Information Facility API. R package version 3.3.0. Available from: <https://CRAN.R-project.org/package=rgbif>.

- Chamberlain S, Szoecs E, Foster Z, Arendsee Z, Boettiger C, Ram K, Bartomeus I, Baumgartner J, O'Donnell J, Oksanen J, Greshake Tzovaras B, Marchand P, Tran V, Salmon M, Li G, Grenié M (2020b) taxize: Taxonomic information from around the web. R package version 0.9.97. Available from: <https://github.com/ropensci/taxize>.
- Crawley MJ, Harvey PH, Purvis A (1996) Comparative Ecology of the Native and Alien Floras of the British Isles. *Philosophical Transactions: Biological Sciences* 351: 1251–1259. Available from: <http://www.jstor.org/stable/56200> (November 17, 2020).
- Daehler CC (2005) Upper-montane plant invasions in the Hawaiian Islands: Patterns and opportunities. *Perspectives in Plant Ecology Evolution and Systematics* 7: 203–216. <https://doi.org/10.1016/j.ppees.2005.08.002>
- Daehler CC (2009) Short Lag Times for Invasive Tropical Plants: Evidence from Experimental Plantings in Hawai'i. *PLOS ONE* 4: e4462. <https://doi.org/10.1371/journal.pone.0004462>
- DeBell DS, Whitesell CD, Schubert TH (1989) Using N₂-Fixing Albizia to Increase Growth of Eucalyptus Plantations in Hawai'i. *Forest Science* 35: 64–75. <https://doi.org/10.1093/forestscience/35.1.64>
- Denslow JS, Space JC, Thomas PA (2009) Invasive Exotic Plants in the Tropical Pacific Islands: Patterns of Diversity. *Biotropica* 41: 162–170. <https://doi.org/10.1111/j.1744-7429.2008.00469.x>
- Department of Business, Economic Development & Tourism (2020a) Hawai'i State Data Center Tables & Graphs (Data focused on Hawai'i). <http://census.Hawai'i.gov/folder-2005-10-13-2927/hsdc-tables-congdist/>.
- Department of Business, Economic Development & Tourism (2020b) State of Hawai'i Data Book Timeseries. https://dbedt.Hawai'i.gov/economic/databook/data_book_time_series/.
- Dodd AJ, Burgman MA, McCarthy MA, Ainsworth N (2015) The changing patterns of plant naturalization in Australia. *Diversity and Distributions* 21: 1038–1050. <https://doi.org/10.1111/ddi.12351>

- eFloras (2017) Efloras.org. Flora of China. Available from: <http://www.efloras.org> (December 26, 2017).
- Essl F, Dullinger S, Rabitsch W, Hulme PE, Hülber K, Jarošík V, Kleinbauer I, Krausmann F, Kühn I, Nentwig W, Vilà M, Genovesi P, Gherardi F, Desprez-Loustau M-L, Roques A, Pyšek P (2011) Socioeconomic legacy yields an invasion debt. *Proceedings of the National Academy of Sciences* 108: 203–207. <https://doi.org/10.1073/pnas.1011728108>
- Essl F, Dawson W, Kreft H, Pergl J, Pyšek P, Van Kleunen M, Weigelt P, Mang T, Dullinger S, Lenzner B, Moser D, Maurel N, Seebens H, Stein A, Weber E, Chatelain C, Inderjit, Genovesi P, Kartesz J, Morozova O, Nishino M, Nowak PM, Pagad S, Shu W-S, Winter M (2019) Drivers of the relative richness of naturalized and invasive plant species on Earth. *AoB PLANTS* 11. <https://doi.org/10.1093/aobpla/plz051>
- Evenhuis NL, Miller SE (2015) Twenty Years of the Records of the Hawai‘i Biological Survey. *Records of the Hawai‘i Biological Survey for 2014. Part II: Index. 1* Edited by Neal L. Evenhuis & Scott E. Miller. *Bishop Museum Occasional Papers* 117: 1–2.
- Gatehouse HA (2008) Ecology of the naturalisation and geographic distribution of the non-indigenous seed plant species of New Zealand.
- GBIF (2020) Global Biodiversity Information Facility. Available from: <https://www.gbif.org/> (October 29, 2020).
- Henke LA (1945) Protein sources and supplements for dairy cows in Hawai‘i. *Hawai‘i Agricultural Experiment Station Bulletin* 95: 26 pp.
- Henry L, Wickham H (2020) purrr: Functional Programming Tools. R package version 0.3.4. Available from: <https://CRAN.R-project.org/package=purrr>.
- Hillebrand W (1888) *Flora of the Hawaiian islands :a description of their phanerogams and vascular cryptogams*. Williams & Norgate ;, London :, 774 pp. <https://doi.org/10.5962/bhl.title.15785>

Hulme PE (2020) Plant invasions in New Zealand: global lessons in prevention, eradication and control. *Biological Invasions* 22: 1539–1562. <https://doi.org/10.1007/s10530-020-02224-6>

Hulme PE, Pyšek P, Nentwig W, Vilà M (2009) Will Threat of Biological Invasions Unite the European Union? *Science* 324: 40–41. <https://doi.org/10.1126/science.1171111>

Imada CT (2019) Hawaiian naturalized vascular plant checklist (February 2019 update). Bishop Museum Technical Report 69: 23 pp.

Jacobi JD, Price JP, Fortini LB, M GI Samuel, Berkowitz P (2017) Carbon Assessment of Hawai‘i Land Cover and Habitat Status Map. <https://doi.org/10.5066/F7DB80B9>

Jordahl K (2014) GeoPandas: a python tool for geographic data. URL: <https://github.com/geopandas/geopandas>.

van Kleunen M, Xu X, Yang Q, Maurel N, Zhang Z, Dawson W, Essl F, Kreft H, Pergl J, Pyšek P, Weigelt P, Moser D, Lenzner B, Fristoe TS (2020) Economic use of plants is key to their naturalization success. *Nature Communications* 11: 3201. <https://doi.org/10.1038/s41467-020-16982-3>

van Kleunen M, Essl F, Pergl J, Brundu G, Carboni M, Dullinger S, Early R, Gonzalez-Moreno P, Groom QJ, Hulme PE, Kueffer C, Kuhn I, Maguas C, Maurel N, Novoa A, Parepa M, Pyšek P, Seebens H, Tanner R, Touza J, Verbrugge L, Weber E, Dawson W, Kreft H, Weigelt P, Winter M, Klöner G, Talluto MV, Dehnen-Schmutz K (2018) The changing role of ornamental horticulture in alien plant invasions. *Biological Reviews* 93: 1421–1437. <https://doi.org/10.1111/brv.12402>

van Kleunen M, Dawson W, Essl F, Pergl J, Winter M, Weber E, Kreft H, Weigelt P, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cárdenas D, Cárdenas-Toro J, Castaño N, Chacón E, Chatelain C, Ebel AL, Figueiredo E, Fuentes N, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Meerman J, Morozova O, Moser D, Nickrent DL, Patzelt A, Pelsers PB, Baptiste MP, Poopath M, Schulze M, Seebens H, Shu

- W, Thomas J, Velayos M, Wieringa JJ, Pyšek P (2015) Global exchange and accumulation of non-native plants. *Nature* 525: 100–103.
<https://doi.org/10.1038/nature14910>
- La Croix SJ (2019) From First Canoe to Statehood: Eight Hundred Years of Economic and Political Change in Hawai‘i. *Australian Economic History Review* 59: 2–23.
<https://doi.org/10.1111/aehr.12171>
- Linnekin J (1997) *Consuming Cultures: Tourism and the Commoditization of Cultural Identity in the Island Pacific*. University of Hawai‘i Press, Honolulu. Available from:
<https://www.jstor.org/stable/j.ctt6wqr0d> (November 16, 2020).
- McKinney ML, Lockwood JL (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology & Evolution* 14: 450–453.
[https://doi.org/10.1016/s0169-5347\(99\)01679-1](https://doi.org/10.1016/s0169-5347(99)01679-1)
- Meyerson LA, Mooney HA (2007) Invasive alien species in an era of globalization. *Frontiers in Ecology and the Environment* 5: 199–208. [https://doi.org/10.1890/1540-9295\(2007\)5\[199:IASIAE\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[199:IASIAE]2.0.CO;2)
- National Agricultural Statistics Service (2020) USDA National Agricultural Statistics Service Hawai‘i Field Office. Available from:
https://www.nass.usda.gov/Statistics_by_State/Hawai'i/index.php/ (November 20, 2020).
- Newbold T, Hudson LN, Contu S, Hill SLL, Beck J, Liu Y, Meyer C, Phillips HRP, Scharlemann JPW, Purvis A (2018) Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide. *PLOS Biology* 16:e2006841. <https://doi.org/10.1371/journal.pbio.2006841>
- Pauchard A, Kueffer C, Dietz H, Daehler CC, Alexander J, Edwards PJ, Arévalo JR, Cavieres LA, Guisan A, Haider S, Jakobs G, McDougall K, Millar CI, Naylor BJ, Parks CG, Rew LJ, Seipel T (2009) Ain’t no mountain high enough: plant invasions reaching new elevations. *Frontiers in Ecology and the Environment* 7: 479–486.

<https://doi.org/10.1890/080072>

Perroy RL, Melrose J, Cares S (2016) The evolving agricultural landscape of post-plantation Hawai‘i. *Applied Geography* 76: 154–162. <https://doi.org/10.1016/j.apgeog.2016.09.018>

POWO (2019) Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Plants of the World Online. Available from: <http://www.plantsoftheworldonline.org> (October 29, 2020).

Pyšek P, Sádlo J, Mandák B (2002) Catalogue of alien plants of the Czech Republic. *Preslia* 74: 97–186. Available from: <http://www.cabdirect.org/cabdirect/abstract/20023095405> (March 23, 2020).

Pyšek P, Richardson DM, Rejmanek M, Webster GL, Williamson M, Kirschner J (2004) Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon* 53: 131–143. <https://doi.org/10.2307/4135498>

Pyšek P, Pergl J, Essl F, Lenzner B, Dawson W, Kreft H, Weigelt P, Winter M, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cardenas D, Cardenas-Toro J, Castano N, Chacon E, Chatelain C, Dullinger S, Ebel AL, Figueiredo E, Fuentes N, Genovesi P, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Maurel N, Meerman J, Morozova O, Moser D, Nickrent D, Nowak PM, Pagad S, Patzelt A, Pelsers PB, Seebens H, Shu WS, Thomas J, Velayos M, Weber E, Wieringa JJ, Baptiste MP, van Kleunen M (2017) Naturalized alien flora of the world: species diversity, taxonomic and phylogenetic patterns, geographic distribution and global hotspots of plant invasion. *Preslia* 89: 203–274. <https://doi.org/10.23855/preslia.2017.203>

R Core Team (2020) R: A language and environment for statistical computing, v 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and distributions* 6: 93–107.

- Rojas-Sandoval J, Acevedo-Rodriguez P (2015) Naturalization and invasion of alien plants in Puerto Rico and the Virgin Islands. *Biological Invasions* 17: 149–163.
<https://doi.org/10.1007/s10530-014-0712-3>
- Rojas-Sandoval J, Ackerman JD, Tremblay RL (2020) Island biogeography of native and alien plant species: Contrasting drivers of diversity across the Lesser Antilles. *Diversity and Distributions* n/a. <https://doi.org/10.1111/ddi.13139>
- Russell JC, Meyer JY, Holmes ND, Pagad S (2017) Invasive alien species on islands: impacts, distribution, interactions and management. *Environmental Conservation* 44: 359–370.
<https://doi.org/10.1017/s0376892917000297>
- Schmitt RC (1977) *Historical Statistics of Hawai‘i*. University Press of Hawai‘i, Honolulu.
- Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE, Kleunen M van, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J, Pyšek P, Winter M, Essl F (2020) Projecting the continental accumulation of alien species through to 2050. *Global Change Biology* online early.
<https://doi.org/10.1111/gcb.15333>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jager H, Kartesz J, Kenis M, Kreft H, Kuhn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajerovala K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* 8. <https://doi.org/10.1038/ncomms14435>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guenard B, Jager H, Kartesz J, Kenis M, Kuhn I, Lenzner B,

- Liebholt AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajerova K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Silva L, Smith CW (2004) A Characterization of the Non-indigenous Flora of the Azores Archipelago. *Biological Invasions* 6: 193–204. <https://doi.org/10.1023/B:BINV.0000022138.75673.8c>
- Skolmen RG (1980) *Plantings on the forest reserves of Hawai‘i 1910-1960*. Honolulu, USA
- Turbelin AJ, Malamud BD, Francis RA (2017) Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography* 26: 78–92. <https://doi.org/10.1111/geb.12517>
- Tye A (2006) Can We Infer Island Introduction and Naturalization Rates from Inventory Data? Evidence from Introduced Plants in Galapagos. *Biological Invasions* 8: 201–215. <https://doi.org/10.1007/s10530-004-3574-2>
- Wagner WL, Herbst DR, Sohmer SH (1999) *Manual of the Flowering Plants of Hawai‘i*, Vols. 1 and 2. University of Hawai‘i and Bishop Museum Press.
- Weigelt P, Jetz W, Kreft H (2013) Bioclimatic and physical characterization of the world’s islands. *Proceedings of the National Academy of Sciences of the United States of America* 110: 15307–15312. <https://doi.org/10.1073/pnas.1306309110>
- Wester L (1992) Origin and distribution of adventive alien flowering plants in Hawai‘i. In: Stone CP, Smith CW, Tunison JT (Eds), *Alien plant invasions in native ecosystems of Hawai‘i*. University of Hawai‘i Press, Honolulu. Coop. National Resources Studies Unit, University of Hawai‘i, Honolulu, 99–154.
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer International

Publishing. <https://doi.org/10.1007/978-3-319-24277-4>

Winter M, Schweiger O, Klotz S, Nentwig W, Andriopoulos P, Arianoutsou M, Basnou C, Delipetrou P, Didziulis V, Hejda M, Hulme PE, Lambdon PW, Pergl J, Pyšek P, Roy DB, Kuhn I (2009) Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proceedings of the National Academy of Sciences of the United States of America* 106: 21721–21725.
<https://doi.org/10.1073/pnas.0907088106>

Woodcock D (2003) To Restore the Watersheds: Early Twentieth-Century Tree Planting in Hawai'i. *Annals of the Association of American Geographers* 93: 624–635. Available from: <https://www.jstor.org/stable/1515500> (May 5, 2020).

Work SH (1946) Digestible nutrient content of some Hawaiian feeds and forages. University of Hawai'i Agricultural Experiment Station Technical Bulletin 4: 22 pp.

Figures

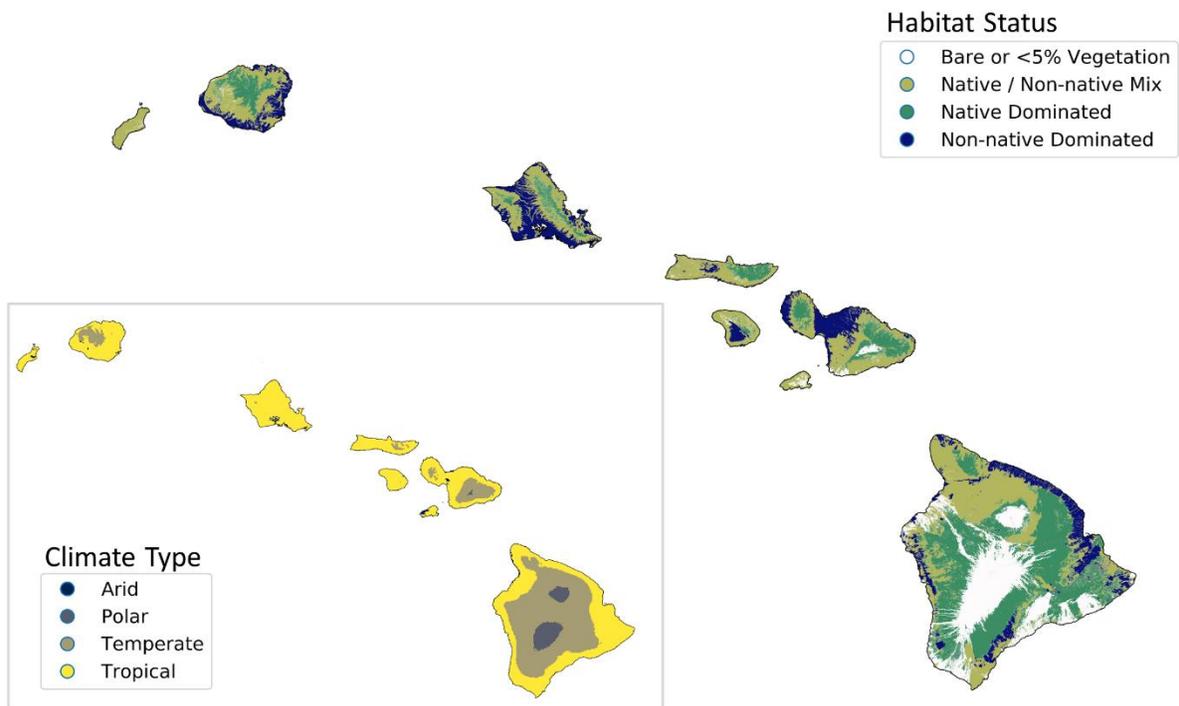


Figure 2.1 The main Hawaiian Islands, showing habitat status (data from Jacobi et al. 2017) and climate types (inset; data from Beck et al. 2018) which shows the overlap between temperate climate and native dominated ecosystems.

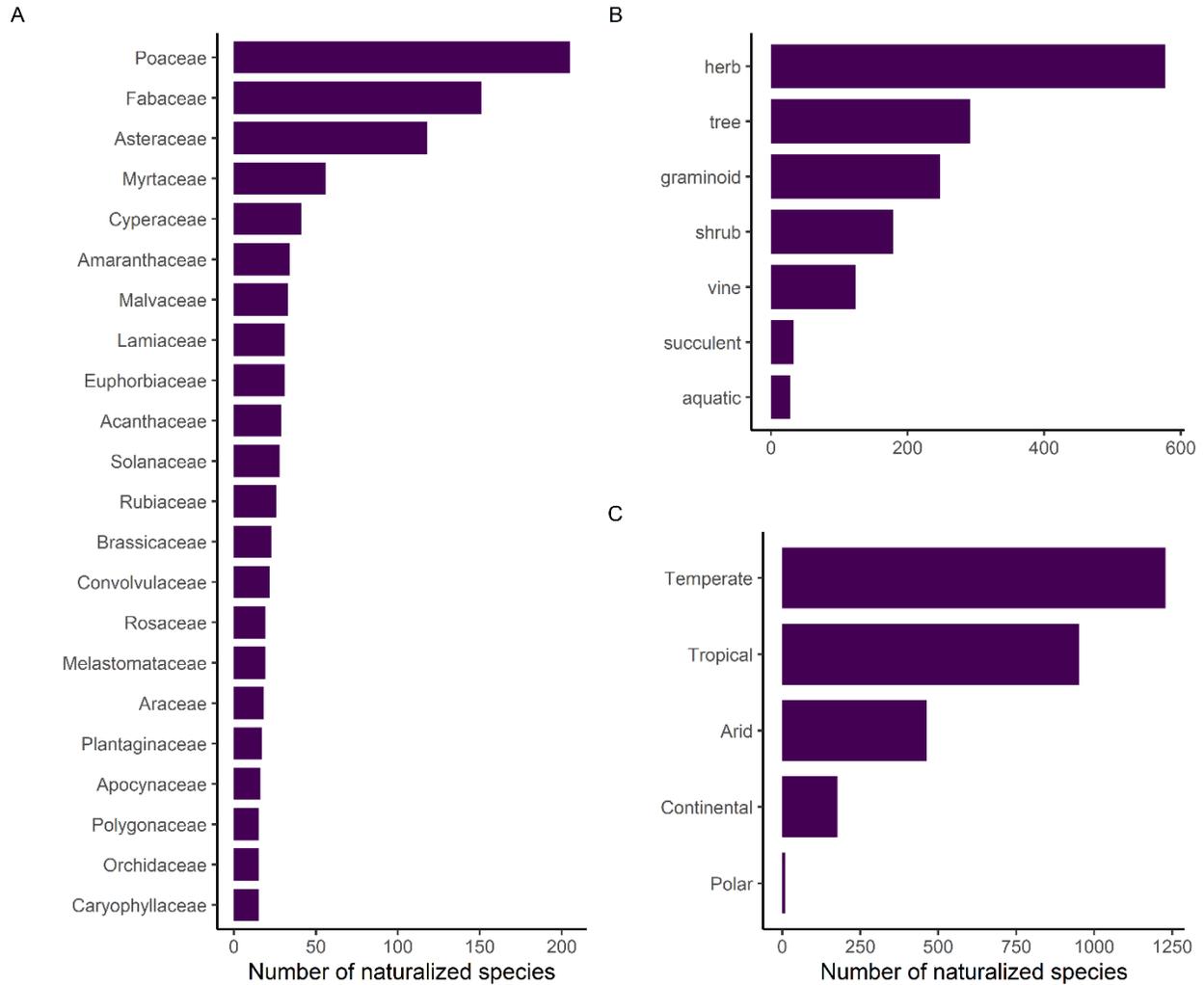


Figure 2.2 Number of naturalized plants in Hawai'i as categorized by: A) plant families containing > 15 species; B) life forms, and C) Köppen-Geiger climate types within their native range. Numbers add up to be greater than the total number of naturalized plants because species can occur in more than one climate type.

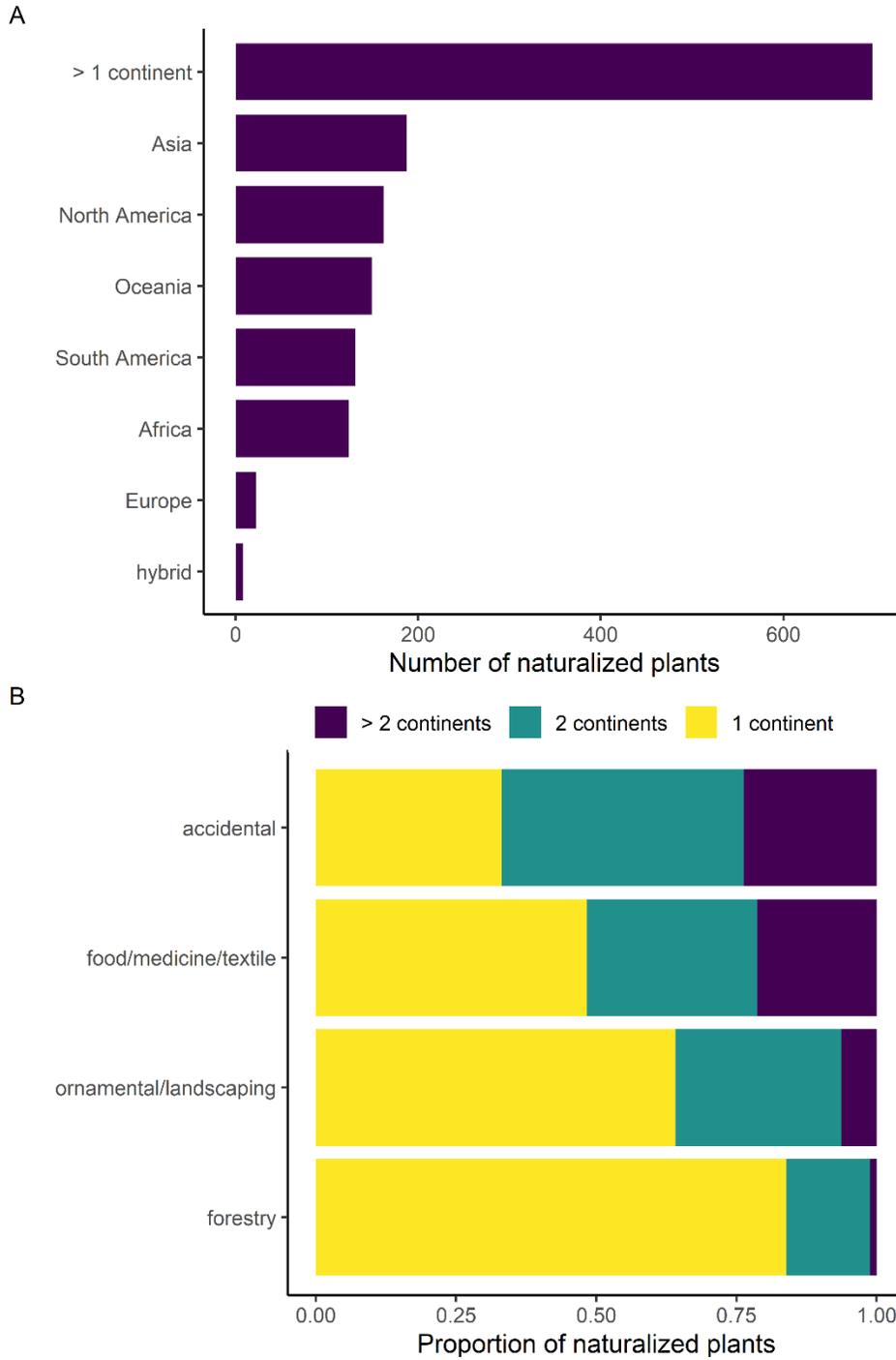


Figure 2.3. Number of naturalized plants in Hawai'i as categorized by: A) continent of origin, and B) the proportion of species from single or multiple continents per probable introduction pathway.

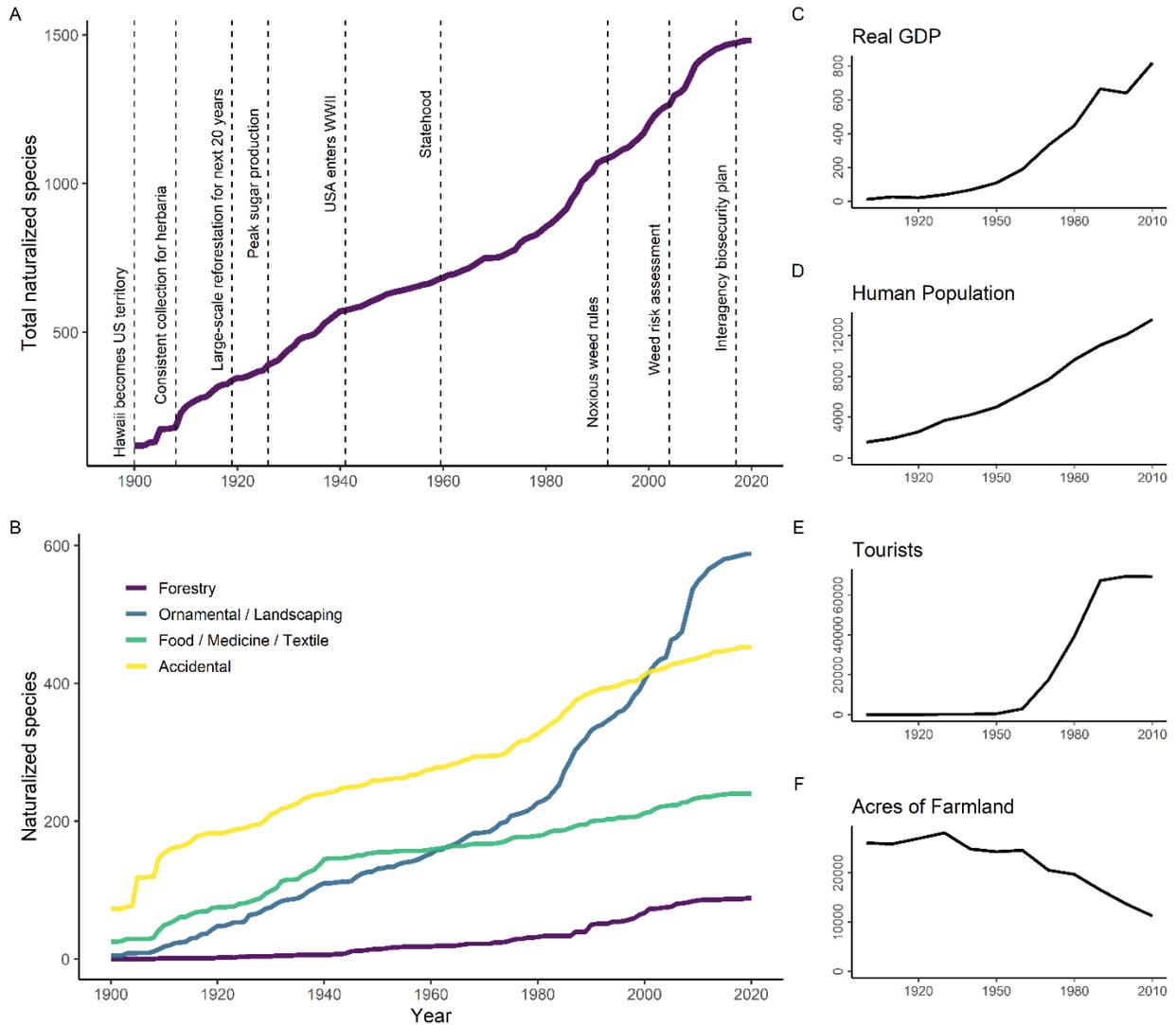


Figure 2.4 Naturalization of plants from 1900 to 2020, showing: A) accumulation of all species, alongside events in Hawai‘i’s history that may potentially affect naturalization rates, B) accumulation of all species, categorized by their likely introduction pathway. C-F show changes in real gross domestic product (in 2019-adjusted US\$1M), human population, tourists, and acres of productive farmland in Hawai‘i over time, each with y-axis units divided by 100.

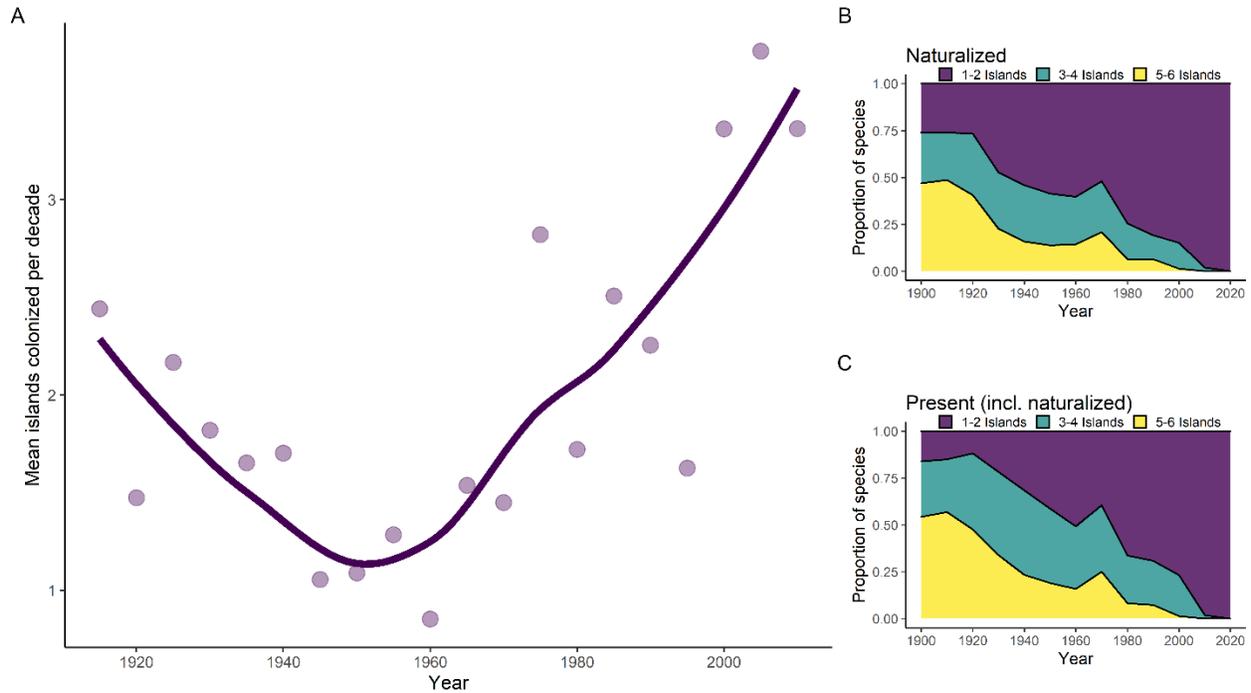


Figure 2.5 Inter-island spread of naturalized species over time in Hawai‘i, showing A) fluctuating rates of spread (averaged over 5-year intervals; line is LOESS smoothed to reduce noise), and the proportion of species introduced each decade that have spread among islands, with proportions calculated for B) species that have fully naturalized, and C) species that are present but not necessarily naturalized.

Chapter 3 –Temporal sampling biases in invasion trend analyses: data processing is just as important as field collection

Abstract

Invasion trend analyses based on the date of first record for new naturalized species are becoming common-place indicators of threats imposed by future invaders and the effectiveness of policy and management actions. Although the time lags between naturalization and detection are well-known, lags that accumulate while processing data after their collection in the field, including time spent identifying species and publishing a status report, may be an equally important source of bias. Here we construct a conceptual framework to understand data processing lags and how they are affected by the range of methods used by researchers to harvest first record data. Using computer simulations and a case study of two independently compiled datasets for naturalized plants of the Hawaiian Islands, we show that data processing lags are influenced by common data harvesting practices to produce the illusion of an invasion slowdown in recent years. This artefact is particularly concerning because it may lead to the belief that efforts to prevent invasions are succeeding when they are not. And, because this problem is more likely to plague geographic regions with fewer resources to track biodiversity change, analyses that compare trends between regions may distract attention from the areas most in need. Given that data processing lag lengths likely vary widely among regions and taxonomic groups, more work is needed to quantify them or find suitable proxies in order to improve analyses. However, we can make two recommendations to minimize the effect of these lags on analysis. First, harvesting dates from natural history collections is likely the most inconvenient yet least problematic source of data, as researchers need only worry about the amount of time that elapses between detection and identification. Second, we caution that the date of first record for

purposefully introduced species may be closer to the date of introduction than the date of naturalization in some cases, which can exacerbate the illusion of an invasion slowdown by incorporating some of the establishment lag into analyses. Our analyses emphasize the need to find ways to shorten data processing lags, such as investments in taxonomic expertise and improvements to biodiversity data infrastructure.

Introduction

The transportation of species outside of their native range is a major driver of global biodiversity change, and although invasions have clearly increased in the last century, we know little about how quickly non-native species have accumulated in different regions and whether invasions will continue at the same rate. As such, we struggle to answer the global call to assess threats imposed by future invaders and determine whether efforts to prevent them are succeeding (Essl et al. 2020). Recent studies have attempted to address this issue by analyzing dates of first record for species that have become naturalized—those forming permanent, self-sustaining populations outside of their native ranges (Blackburn et al. 2011). These naturalization rate studies have investigated a variety of taxonomic groups at both local and global scales to infer the shape of invasion trends over time (Rojas-Sandoval and Acevedo-Rodriguez 2015, Seebens et al. 2018, Muñoz-Mas and García-Berthou 2020), and at least one study has projected these trends to predict which geographical regions may face the highest increases (Seebens et al. 2021). Similar studies will likely be conducted in the future to provide policy-relevant information and take advantage of the increasing availability of non-native species data and their integration into large biodiversity data platforms (Dyer et al. 2017, Groom et al. 2019, Jetz et al. 2019, van Kleunen et al. 2019).

However, biodiversity data are well-known to have patchy taxonomic and geographic coverage because they are often collected, processed and published by many different people or institutions (Walpole et al. 2009, Hughes et al. 2021). Temporal sampling biases likely lurk in these datasets as well, yet crucial caveats for interpreting naturalization trends are rarely discussed beyond the well-recognized time lag that exists between naturalization and detection (i.e. field surveying lag; Belmaker et al., 2009; Boakes et al., 2010; Crooks et al., 1999; Solow & Costello, 2004). Less attention is paid to lags that are generated when processing data after their collection in the field, although their impact on invasion trend analyses may be equally important (Smith et al. 2018, Zenetos et al. 2019). Significant time may elapse between when a specimen is collected and identified, a report is published on its naturalized or invasive status, and information is compiled into a database or checklist (hereon collectively referred to as data processing lags; Figure 3.1).

Several studies recognize that data processing lags can create a perceived reduction in recent naturalization rates because researchers can only include species they are aware of, and therefore, recently detected species are excluded if a status report has not yet been published. To account for this artefact, many authors choose to truncate their study window by an arbitrary length of time, often excluding the most recent 15 years or less when estimating or projecting naturalization rates (Muñoz-Mas & García-Berthou, 2020; Seebens et al., 2017, but see Smith et al., 2018). However, these rule-of-thumb methods may be insufficient if data processing lags are longer than researchers realize (Zenetos et al. 2019), and prevent analysis of very recent trends that could be useful to policy makers. Limited funds to employ taxonomic experts and create infrastructure to house digital resources (i.e. data infrastructure) undoubtedly play a role in the length of data processing lags, or whether an effort is made to curate and communicate data on

new naturalized species at all. Without a better understanding of data processing lags, we are unable to understand whether plateaus that appear in invasion trend analyses might have arisen from slow data processing or an actual slowdown in the rate at which species are naturalizing (Crooks 2005, Byers et al. 2015). This distinction has enormous consequences for biosecurity planning.

In turn, we are unaware of any study that outlines how data processing lags are influenced by the range of ways first record data can be harvested, despite several authors reporting that they retrieved their data from a variety of sources (Seebens et al. 2017, Smith et al. 2018). For example, first records may be harvested at different points along the data processing pipeline, including identified specimens housed in natural history collections, species status reports, or compiled resources summarizing all known species alongside their naturalization dates (Figure 3.1). The convenience of harvesting first record data generally increases with each successive step in the data processing pipeline, but convenience may come at a cost if a compiled resource (such as a published flora) takes many years to produce, incurring lags that can lead to false conclusions about naturalization or invasion trends. The degree of processing undergone by each data source is also likely to influence how easily records can be retrieved and included in analyses, with some sources imparting a complete list of the known naturalized species in a region while others yielding only a portion. These differences may introduce a temporal sampling bias if researchers combine records harvested from convenient, highly processed sources with less processed, more inconvenient sources.

Another issue relating to lags concerns the definition of a “first record” itself and which time point is represented along the introduction–naturalization continuum (Blackburn et al. 2011). First record dates are often assumed to represent some time after a species’ actual

naturalization, although the accuracy of this assumption depends on whether the analyzed species remain contained in captivity for long periods or if the date of their introduction was well-recorded (Crooks 2005). Some vascular plants, for instance, are known to have persisted in cultivation for centuries before naturalization, and their introductions may be reasonably well documented and vouchered in some locations (Meares 1790, Crooks et al. 1999). Thus, when “first record” is defined as the date of a species’ first known occurrence rather than the date of first detection of a naturalized population, researchers run the risk of backdating the actual naturalization trend and introducing at least part of the establishment lag into their analyses (Figure 3.1).

Our aim is to provide a conceptual framework to understand data processing lags and examine how variability in harvesting and compiling first record data can impact the interpretation of naturalization trend analyses. First, we use simulated data to examine how estimates of naturalization rates are affected by 1) field surveying versus data processing lags and what happens when they shorten or lengthen over time, 2) harvesting data at different points along the data processing pipeline, and 3) varying rates of completeness in first record retrieval in analyses. Additionally, we construct trendlines based on dates that approximate a species’ introduction as well as its naturalization to illustrate their differences. Secondly, because any combination of these factors may be present in first record datasets, we compare two independently compiled datasets as a real-world example of how first record data can be harvested in different ways to yield different conclusions about invasion trends. We contrast first vascular plant records in Hawai‘i obtained from the Global Alien Species First Records Database (GASFR; Seebens et al. 2017) , which largely includes records from an expert-compiled all-species publication, with a dataset we compiled based on individual status reports and herbaria

vouchers. This latter comparison is timely because the GASFR dataset was recently used to project a declining invasion trend for vascular plants in the Pacific Island region (Seebens et al. 2021), offering an opportunity to explore how different data harvesting methods could affect conclusions about trends in the Pacific Islands.

Methods

To explore the simulated scenarios described below, we first created hypothetical naturalization dates for 1000 species accumulating at a constant rate (10 spp/year) between 1920 and 2020 (designated as the current year), yielding a dataset of naturalization dates to compare with dates depicted in Figure 3.1 (e.g. detection date). These naturalization dates were created by randomly assigning each species a year during this timeframe from a uniform distribution using the *stats* package in R (R Core Team 2020). An average lag time of 10 years was arbitrarily chosen for all scenarios (except where indicated) to represent each of the lags in Figure 3.1, with the lag for each species randomly selected (with replacement) from a normal distribution of 10,000 lag times ($\mu=10$ years, $\sigma=3.5$). We calculated the introduction year, detection year, identification year, report year, and the year in which species records are compiled by adding or subtracting lags from the hypothetical naturalization date for each of the 1000 species. For example, the identification date was calculated by adding naturalization–detection and detection–identification lags to the hypothetical naturalization year. We define “detection” as the date when a naturalized population has been both observed and documented, which may include the collection of voucher specimens for later identification (Reaser et al. 2019). As a search of the literature revealed that most researchers rely on naturalization or invasion status reports to determine which species should be included in analyses, we removed species having a report year >2020 for all simulations, unless stated otherwise. Trends in naturalization rates were

calculated by analyzing the number of species reported per year over time, and visualized by plotting the accumulation of species over time using the *purrr* (Henry and Wickham 2020) and *ggplot2* (Wickham 2016) packages in R. In most scenarios, analyses of these trends were conducted on both the introduction and detection dates alongside the “true” naturalization date.

To quantify how trendlines based on detection or introduction date differed from the “true” trend of species naturalizing at a constant rate over time, we conducted an optimal changepoint analysis using the Pruned Exact Linear Time (PELT) algorithm developed by Killick et al. (2012) available in the *changepoint* package in R (Killick and Eckley 2014). The PELT algorithm searches sequential data for multiple changepoints using a penalized likelihood-based approach that calculates pre- and post-change ratio tests on the mean and variance, maximizing the likelihood over the number and position of the changepoints. We also employed the Changepoints for a Range of Penalties (CROPS) method (Haynes et al. 2017) in the *changepoint* package to help select changepoints that minimize over- or underfitting models. Ultimately, these methods divided the naturalization rate trendline into segments so that we could compare the mean and variance for each time period. This allowed us to quantify the length of artificially low naturalization rate periods at the beginning and end of the data collection window—artifacts created by the time lags—as well as the rate of naturalization for a truncated segment after these low periods were removed. As these quantities are somewhat affected by the randomness of which lag lengths are selected in each simulation, we repeated the process of generating lags and searching for changepoints 10 times, comparing average values for: length of the artificially low start and end segments (# of years) as well as their naturalization rate (species/year), number of species lost from non-elapsed time lags, and the naturalization rate for the truncated segment (species/year).

Varying Lag Lengths

To understand how field surveying lags might affect the interpretation of naturalization trends in comparison with data processing lags, we compared four scenarios. In the first scenario, we added no lags, reflecting a situation where naturalization events are detected and reported immediately for all species. In the second scenario, we added only a field surveying lag, mimicking a scenario where some amount of time passes before a naturalized species is detected and evidence of its naturalization is collected, but where a status report is produced immediately. The third scenario represents the inverse of the second, where we added only a data processing lag, creating a situation where all species are detected immediately, but where some time elapses until species are identified and reported. In the fourth scenario, we added both field surveying as well as data processing lags.

To explore the effect of increasing or decreasing data processing lags over time, we first simulated a scenario where both field surveying and data processing lags do not change over time. The trendlines produced from this constant lag scenario were compared to six additional scenarios, including both the increase and decrease of field survey lags alone, data processing lags alone, and both field surveying and data processing lags together. These scenarios were simulated in the same way as the constant lag scenario mentioned above, but instead of randomly drawing lag times from a normal distribution with a mean of 10 years, we simulated a lag shortening to emulate a scenario where the efficiency of field surveying and/or data processing has improved over time, and similarly, a lag lengthening to reflect declining efficiency. The shortening was done by randomly selecting a lag length from a normal distribution with a mean of 10 years for 1920 ($\sigma=3.5$, $N=10,000$), and decreasing the mean linearly over the 100 year

period such that data processing lags were sampled from a distribution with a mean of 1 month ($\sigma=0.30$, $N=10,000$), with the ranges inverted for the process of lag lengthening (i.e. 1 month in 1920 and 10 years in 2020).

Varying Harvest Points

To understand the relationship between harvesting data at different points during data and time lags, we manipulated the number of species that were included based on whether their time lags had elapsed in time to be harvested. To simulate a scenario where species are included in analyses based on examining identified specimens deposited in natural history collections, we removed all species with an identification date >2020 , reasoning that researchers could not include species that had not been identified yet. This was compared with scenarios where we removed species having a report date >2020 , emulating a situation where researchers could not access natural history collections easily, and thus could only know about species that had been described in naturalization or invasion status publications. We additionally created a scenario where researchers harvested data from only already-compiled (convenient), checklist-style resources published periodically by regional or taxonomic experts. In this situation, we simulated that the expert published their compilation in 2000 (year selected arbitrarily) with access to both identified specimens as well as published reports, thus including all species identified by 2000.

Varying Record Retrieval Rate

To reflect that different data sources vary in how findable or harvestable they are, we explored four scenarios to understand the effects of retrieving different percentages of first records. In this context, we define record retrieval as the proportion of the total number of species that have been reported as naturalized (as opposed to the proportion of species that have actually naturalized). First, we simulated two scenarios where 85% and 50% of records were retrieved, with each species having an equal likelihood of being selected, emulating a scenario

where records are sampled evenly over time. We then simulated a scenario where record retrieval was lower in the earlier part of the century, reflecting a scenario where early records are difficult to find, perhaps because vouchered specimens have been lost or destroyed or because status report publications have not been digitized. We also simulated a scenario where record retrieval is lower in the later part of the century, which may be the case if publications are more readily produced for well-known invaders, which tend to have earlier introduction dates, or if periodic expert-compiled resources have not been published recently. These scenarios were simulated by randomly sampling 50% of records between 1920 and 1970, and 85% between 1970 and 2020 (and vice versa), with 1970 chosen because it coincides with increased biodiversity monitoring and data infrastructure on a global scale (Butchart et al., 2010).

Case Study

To understand how different data collection methods might produce different trendlines, we compared data for vascular naturalized plants compiled in version 1.2 of the Global Alien Species First Records (GASFR) Database, assembled by Seebens et al. (2017, 2021), with our own dataset ((Brock and Daehler 2021)). To account for taxonomic synonyms between the two datasets, each species name was updated to the accepted name recognized by Kew's Plants of the World database (POWO 2019) using the taxize package in R (Chamberlain and Szöcs 2013). We then plotted an accumulated species trendline for our data as well as the GASFR data.

Records for Hawai'i in the GASFR database were compiled from publications as well as online databases, including the Centre of Agriculture and Biosciences International Invasive Species Compendium (CABI) and Global Invasive Species Database (GISD). These dates appear to include the very first year a species was recorded in Hawai'i, even if that record was from cultivation and not yet known to be naturalized. In contrast, our dataset was compiled by first

consulting a species checklist for Hawai‘i (Imada 2019), and subsequently harvesting the year of naturalization from digitized voucher labels from 83 herbaria (Bernice Pauahi Bishop Museum, National Tropical Botanical Garden, the Smithsonian US National Herbarium, and other specimens contained in the Global Biodiversity Information Facility; (Brock and Daehler 2021)). We considered specimens to represent the first date of naturalization if they were referred to in naturalization reports or if details about their population or naturalized status were written on the specimen label. To determine whether differences in trendlines can be explained by using the very first date of record (including cultivation) versus the first date of naturalization, we created an “adjusted” version of the GASFR data where we ensured that the first record date was consistent with the date of the first known naturalized specimen.

To compare naturalization trends produced by each dataset, we assessed the fit of naturalization rate data (species/ 5 years) to different naturalization rate models and selected the best fit using the Akaike information criterion (AIC) using the `nls` function in the `stats` package in R (R Core Team 2020). Five models were compared, including: a constant rate of naturalization [linear model; $y = a + bt$], an accelerating rate of naturalization [exponential model; $y = ae^{bt}$], a declining rate of naturalization [logistic model; $y = L/(1 + e^{-k(x-x_0)})$], or a hump-shaped trend, where rates increase and then decrease [Weibull model; $a(c/b)^*(t/b)c e^{-(t/b)c}$]. To understand temporal sampling biases in each dataset, we considered all of the records reported in the GASFR data as well as our dataset to represent the total knowable first records available. We then classified each record as being present in our dataset only, the GASFR dataset only, or present in both publications. We then calculated the proportion of each class for every five year period between 1900 and 2020, and visualized the results using `ggplot` in R (Wickham 2016).

Results

By simulating the field surveying and data processing lags separately, we showed that field surveying lags alone create an artificially low naturalization rate in the earliest years of analysis (Figure 3.2, Table 3.1). The field lags resulted in the illusion of a decade of low naturalizations (~2 naturalizations per year) before reaching the rate of ~10 per year reflecting the “true” rate for the entire century. A complementary illusion occurred when we only included data processing lags, which portrayed a deceleration of naturalizations from ~10 to ~2 per year in the final decade. Our simulation including both field surveying and data processing lags demonstrates an additive effect in which they impose artificially flat naturalization rates at both ends of the inferred trendline, resulting in a shorter middle segment reflecting the true naturalization rate. We also noted a marked decrease in the variance of the data as they approach periods of artificially low naturalization rates. This arises because, in the case of field surveying lags, only species with small lags are recorded at first, such that the number of species per year is low. Similarly, in the case of data processing lags, species with long lags have not been reported yet, and increasingly become excluded from analyses (Figure 3.2; Table 3.1).

Lags that change over time also impact the inferred naturalization rate. Our simulations showed that when field surveying lags decrease over time, the trend based on the year of detection is steeper than our true naturalization rate, indicating a rate of 11 species/year as opposed to 10 for the truncated middle segment (Figure 3.3B, Table 3.1). Inversely, the detection trend is shallower (9 species/year) when field surveying lags increase (Figure 3.3E). In the latter case, the detection trend included virtually no initial flat part of the curve, as expected, because little time passed between naturalization and detection around 1920. When we set data processing lags to decrease (Figure 3.3C), we nearly eliminated the artificial flat tail end of the

curve because the species that contributed to the artificial plateau have been included in the analysis. For instance, under the scenario with constant lags over time, 286 species are excluded because their lags had not elapsed by the time of analysis, whereas only 154 species are excluded when data processing lags are shortened, translating part of the ending plateau into a linearly increasing segment that more resembles the true naturalization trend. However, unlike changes in the length of field survey lags over time, changing data processing lags does not alter the inferred naturalization rate of the truncated segment. An artificially steep trendline is inferred in our simulation where both lag types shorten over time (Figure 3.3D), and an artificially shallow trendline when all lags increase (Figure 3.3G). In the former case, the shortening lags combined to make nearly 93% of naturalized species apparent in the curve, as lags for most species have been allowed to elapse.

When examining the influence of harvesting data at different points on the interpretation of naturalization trends, we show that harvesting data earlier in the pipeline can coincide with a greater retrieval of records. Gathering data from identified vouchers in a natural history collection represents data that is available between the specimen identification and reporting phases in the pipeline represented in Figure 3.1, which need not be published to be included in analyses. By harvesting data at this point, over 82% of naturalized species are available for analysis (Figure 3.4A, Table 3.1). But when harvesting data from multiple published reports, we recover only 71% of species records (Figure 3.4B). Our simulation of harvesting data from an expert-compiled resource published in 2000 fared the worst, recovering merely 63% of naturalizations (Figure 3.4C). These numbers reflect species lost from non-elapsed lags that manifest as artificial plateaus in recent decades, with harvest points later in the pipeline accruing a longer plateau. In our study simulating all lags to average 10 years, the plateau is about 30

years long when awaiting a compiled resource, and <10 years long if harvesting identified specimen data.

When simulating the effect of different percentages of missing records, we found that not only did retrieving fewer records have the obvious problem of recovering a smaller number of species, but it also depressed the trendline so that the inferred naturalization rate was lower (Figures 4D-E, Table 3.1). This was less of a problem in our scenario where records were randomly removed evenly over time, as the overall trendline for 85% and 50% retrieval resembled the trendline for 100% retrieval, albeit showing a lower rate. However, varying the retrieval rate between 1920–1970 and 1970–2020 fundamentally changed the trendline shape. Retrieving fewer samples between 1920–1970 created a sigmoidal curve (which would appear exponential if sufficiently truncated to remove the artificial plateau at the end of the period), while retrieving fewer samples between 1970–2020 created the illusion of a saturating trend.

In all simulations that varied lag lengths, harvest points or record retrieval rate, we noted that trendlines calculated from the introduction date have a longer plateau at the end of the analysis window in comparison to trendlines calculated from the detection date (Figures 3-4, Table 1). This pattern reflects the problem of lag accumulation, where all lags from introduction to status reporting must elapse before researchers know to include these taxa in their analyses. In contrast, trendlines calculated using the detection date have shorter end-of-window plateaus because only the lags between detection and reporting must elapse. Instead, trendlines based on detection accrue a longer artificial level period at the beginning of the analysis window—which is not observed in introduction data trendlines. The short accelerating period at the beginning of the introduction date trendline is a result of introduction–naturalization lags being randomly selected from a normal distribution, with some lags longer than average.

When comparing trendlines calculated from first records within the GASFR database versus our own dataset, we find that the GASFR data depicted a saturating trend in the accumulation of species between 1900 and 2020 (hump-shaped curve of the naturalization rate; $AIC = 219.83$) while our dataset best fit a linear model, showing a continuous rate of naturalizations ($AIC = 247.85$). Adjusting the GASFR data such that it represented dates when naturalization evidence was collected rather than the very first date of record still revealed a saturating line (hump-shaped curve; $AIC = 221.23$), although it adjusted the overall slope of the line such that the rate of naturalization was higher. Thus, the change in the rate of naturalization preceding the leveling-off period is more abrupt in comparison to the original GASFR data (Figure 3.5A). This somewhat reflects our simulation shown in Figure 3.4A, where using an earlier date (i.e., introduction date versus detection date) back-dates the true naturalization trend, such that the trendline appears to approach saturation earlier.

As expected, the record retrieval in the GASFR dataset for each five-year period was lower than in our dataset, reflecting the inclusion of 816 first records in the GASFR dataset versus 1494 in ours. However, we noted that a substantial temporal sampling bias towards records earlier in the century was responsible for the saturating trendline inferred by Seebens et al. (2021) using the GASFR data, and that this trend persisted even after the records were adjusted to use the date evidence was collected on their naturalization, similar to the simulation depicted in Figure 3.4G. This temporal bias arises because 80% of their data were retrieved from an expert-compiled resource published in 1992 (Wester 1992), which itself only includes data up until 1985, similar to the simulation in Figure 3.4C. This artefact manifests as a plateau at the end of the analysis window after 1985, when many more species exist in the pipeline awaiting compilation and publication by experts before each record can be easily harvested and added to

the GASFR database. The combination of uneven sampling over time alongside the apparent preference for periodic expert-compiled resources is compounding and produces an even more convincing illusion that the naturalization rate in Hawai'i is declining. In contrast, our dataset did not include a few species that were noted in the GASFR dataset because these species have since been considered questionable in their naturalization status, and are treated differently in more recent floras and checklists (Wagner et al. 2005, Imada 2019), so the trend from the adjusted GASFR dataset is slightly higher in the earliest part of the 20th century.

Discussion

We demonstrate that data processing lags can have a significant influence on naturalization trendlines inferred from first record datasets, and that these lags are influenced by common data harvesting methods to produce conclusions that vary among researchers who collect data in different ways. Much effort is currently being invested to develop indicators of global biodiversity trends, with calls to concentrate on first record data in recent decades to inform invasive species policy (McGeoch et al. 2021). However, we show that a disregard for data processing lags may be especially problematic when trying to understand recent invasion trends. Unlike field surveying lags, long data processing lags increase the likelihood of falsely interpreting a recent invasion slowdown, which may lead to the conclusion that policies to prevent invasion and eradicate incipient species are working when they are not. Furthermore, comparing trends between regions (e.g. Seebens et al., 2021) may sway policy makers to allocate funds where invasions appear to be on the rise, even though investing in data infrastructure in regions with apparently saturating trends might reveal that they also face rising invasive threats (Moura et al. 2018). Thus, analyses of invasion trends in recent decades may be a poor indicator

of invasion policy effectiveness and a faulty tool to predict future invasion rates if data processing lags are not accounted for.

In contrast, field surveying lags have more potential to affect our understanding of invasion onset in a given region, making them more important when examining correlations between peaks of invasions and historical events (Mangiante et al. 2018, Bonnamour et al. 2021). Ultimately, both field surveying lags and data processing lags likely exist in real-world datasets, which suggests that the beginning and end of invasion trendlines should be interpreted with a considerable degree of caution. Methods to account for temporal sampling biases imposed by field surveying lags have been developed by incorporating models of the introduction and discovery processes (Solow and Costello 2004, Belmaker et al. 2009). However, similar fixes to account for data processing lags will be difficult to create because, in the case of correcting for field surveying lags, models seek to back-date known dates of detection to obtain a more representative naturalization date, whereas records lost to slow data processing are simply not known and therefore cannot be subject to correction.

Our study also confirms the findings of other authors that the progressive shortening or lengthening of field surveying lags over time can have a large impact on the perceived rate of naturalization (Costello and Solow 2003, Mangiante et al. 2018). However, unlike field surveying lags, data processing lags by themselves have little effect on the inferred rate of naturalization, as long as the inferred trendline is truncated such that the artificially low naturalization rates at the beginning and end of trendlines are trimmed off. In fact, abruptly shortening data processing lags (which would hopefully occur after major investments in biodiversity infrastructure) can “rescue” the trendline by removing the ending plateau with no impact on the naturalization rate. However, increasing or decreasing both lag types together can

exacerbate issues initiated by field surveying lags. For instance, in scenarios where both lag types decrease over time, the steeper trendline is inferred over a longer period of time as the artificial ending plateau decreases, which allows the artificially elevated trends to persist until the present. These long persisting and recent trends may mislead researchers into falsely interpreting an elevated rate of naturalization.

Both our simulated and real-world analyses show that when researchers are faced with the option to harvest first record data at different points during data processing (Figure 3.1), each option introduces a different set of data processing lags that should be considered. For instance, researchers harvesting dates directly from the labels of identified specimens in natural history collections need only worry about the data processing lag between a specimen's collection and identification, while those harvesting dates from published reports need to additionally consider the data processing lag between identification and report publication. Moreover, these data harvest points differ in how easy it is for researchers to retrieve first records, introducing variability in how closely the analyzed species list resembles reality for a given time period. Our comparison of the GASFR dataset for Hawai'i showed a temporal sampling bias against recent decades that arose from preferentially harvesting already compiled data (thereby accumulating a reporting–compilation lag), which may have been used because harvesting first records from an already-compiled resource was much simpler than searching through numerous specimens or publications. Furthermore, regional or taxonomic experts compiling these resources likely have better access to specimens and historic publications, and are able to compile a species list that is much more complete than a researcher unfamiliar with the local biota and data infrastructure, especially if most of the information is housed locally or in obscure locations. However, this

discrepancy can be minimized by building free online platforms that increase the ease of retrieving a complete, up-to-date lists of species and their first records.

The degree of bias that results from harvesting data at different points in the data processing pipeline depends on the length of data processing lags, and if our simulations had used 10-day data processing lags instead of 10 years, harvesting data at different points in the data processing pipeline would have made little difference. However, the fact that many invasion trend researchers truncate their analyses to account for data processing suggests that these lags are already recognized as non-trivial (Seebens et al. 2017, Mangiante et al. 2018, Gladstone et al. 2020). Even in a small and comparatively wealthy region like Hawai‘i, where recent political and scientific interest has bolstered its ability to monitor new naturalizations compared to most places elsewhere (Evenhuis and Miller 2015), the average lag between detection and status publication is 8 years, and ranges between <1 to 87 years (calculated by subtracting first naturalized specimen date from publication date since 1990; data not shown). More research is needed to quantify data processing lags so that we can further understand how they influence invasion trend analyses and identify the minimum number of taxonomic experts and data infrastructure needed to make valid inferences about biodiversity change. Because few studies have attempted to quantify data processing lags, we have little understanding of how these lags vary between regions, taxonomic groups or among species, which are likely to influence how first record trends are interpreted (Zenetos et al. 2019). For instance, simulating an increase in the variance of lag lengths or skewing the distribution of lags (i.e. mimicking a scenario where most species are reported quickly while others take decades) makes the change from the “true” naturalization rate into the artificial plateau in recent years appear more gradual (data not

shown), which may further increase the tendency for researchers to interpret these artefacts as naturally occurring.

We also caution that dates analyzed in various first record datasets may correspond to different points along the introduction–naturalization continuum, and our analyses show that this may result in a longer ending plateau in comparison to trendlines based on the date of the first naturalized specimen. However, the significance of this discrepancy likely depends on the taxonomic group being investigated. For example, first record data are often assumed to have been collected after their actual naturalization (Seebens et al. 2017), even though the date representing the first specimen collected for purposefully cultivated plants may be closer to the introduction date for certain regions, depending on whether cultivated plants are regularly surveyed and collected. Hawai‘i, for instance, has a long and well-documented history of collecting cultivated plant specimens (Neal, 1965; Staples & Herbst, 2005) and some plants are likely to have remained contained to their cultivated settings for much longer than the 10 year average in our simulations (Crooks et al., 1999; Daehler, 2009). Moreover, several studies show that purposefully cultivated species often comprise a large portion of naturalized floras, and thus, including cultivated occurrences in first record datasets could significantly influence the shape of inferred trendlines by pre-dating the real naturalization trendline and contributing to the illusion of an earlier invasion slowdown (Crawley et al. 1996, Dodd et al. 2015, Rojas-Sandoval and Acevedo-Rodriguez 2015, Brock and Daehler 2021).

Our real-world comparison between the GASFR dataset and our own emphasizes that data processing lags are influenced by data harvesting techniques in complex ways, and that their effects are difficult to identify when interpreting trends. By fitting trendlines to naturalization data and considering potential source pools of plants in the Pacific Island region, Seebens et al.

(2021) recently projected a slowdown in Pacific plant naturalizations. However, we suspect that this trend says more about inadequate investment in surveying, biodiversity data infrastructure and active research in the region rather than actual invasion trends. Although we cannot definitively say that Hawai‘i’s non-saturating trend, as demonstrated by our analysis presented here, is representative of the Pacific Island region as a whole, we suggest that Seeben’s et al’s (2021) conclusions would have changed if they had harvested Hawai‘i’s data beyond those appearing in published sources, considering that records from Hawai‘i comprised 73% of their Pacific Island data and therefore weighed heavily on their trendline. Although some records from databases such as CABI and GISD were incorporated into the GASFR database for recent decades, the record retrieval rate was much lower than for the expert-compiled, all-species resource published by Wester in 1992, and could not be fixed by employing a rule-of-thumb truncation of 15 years. We recommend that researchers use caution when combining data sources that cover different time periods, especially if one of the sources attempts to compile a complete species list, as this will almost definitely violate the assumption that records are being sampled evenly across time. After recognizing temporal disparities in sampling effort, researchers are then faced with two options: either truncate the study window to coincide with the sampling period of a particular source, or replicate their exact methodology to retrieve first records (i.e. sifting through natural history collections like in Wester (1992)) until the end of the desired time window to incorporate more recent records.

Nonetheless, we commend the GASFR database as a step in the right direction and suggest that in addition to developing global biodiversity health indicators such as invasion trendlines, we should also develop metrics that indicate the health of biodiversity data infrastructure. These developments could be facilitated by a centralized system where institutions

housing naturalized species data could access best practices and report average data processing times, as well as fluctuations in funding, taxonomic expertise, and investments in digital infrastructure. To shorten data processing lags, we propose that digitized specimen data from natural history collections be annotated and linked to continuously updated online databases designed to advance the research and management of biological invasions (Lendemer et al. 2020).

Conclusion

Although data processing lags remain largely overlooked relative to field surveying lags, we show that they should be recognized as a considerable source of bias when inferring invasion trends. We argue for the importance of seeking ways to shorten them, such as investing in taxonomic expertise and improved data infrastructure. Of particular concern is that long data processing lags increase the likelihood that researchers will conclude that an invasion slowdown has occurred in recent years, potentially influencing invasive species policy and management. While these artefacts can be eliminated by truncating the analysis window, we can offer no rule of thumb for how much truncation is necessary, as the average length of lags will vary widely among regions, and different lags will come into play depending on which data sources are harvested. However, we suggest that using the date of the first naturalized specimen (rather than the first date observed), and compiling lists of species to analyze based on identified specimens in natural history collections will reduce the risk of interpreting an invasion slowdown due to data processing artefacts. Future research identifying proxies for data processing lag lengths would be especially useful for research conducted on a large scale, and could additionally help quantify the minimum infrastructure needed to monitor invasive species.

References

- Belmaker J, Brokovich E, China V, Golani D, Kiflawi M (2009) Estimating the rate of biological introductions: Lessepsian fishes in the Mediterranean. *Ecology* 90: 1134–1141. <https://doi.org/10.1890/07-1904.1>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLOS Biology* 8: e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Bonnamour A, Gippet JMW, Bertelsmeier C (2021) Insect and plant invasions follow two waves of globalisation. *Ecology Letters* 24: 2418–2426. <https://doi.org/10.1111/ele.13863>
- Brock KC, Daehler CC (2021) Plant naturalization trends reflect socioeconomic history and show a high likelihood of inter-island spread in Hawai'i. *Invasive Plant Science and Management* 14: 135–146. <https://doi.org/10.1017/inp.2021.18>
- Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA, Baillie JEM, Bomhard B, Brown C, Bruno J, Carpenter KE, Carr GM, Chanson J, Chenery AM, Csirke J, Davidson NC, Dentener F, Foster M, Galli A, Galloway JN, Genovesi P, Gregory RD, Hockings M, Kapos V, Lamarque J-F, Leverington F, Loh J, McGeoch MA, McRae L, Minasyan A, Morcillo MH, Oldfield TEE, Pauly D, Quader S, Revenga C, Sauer JR, Skolnik B, Spear D, Stanwell-Smith D, Stuart SN, Symes A, Tierney M, Tyrrell TD, Vié J-C, Watson R (2010) Global Biodiversity: Indicators of Recent Declines. *Science* 328: 1164–1168. <https://doi.org/10.1126/science.1187512>
- Byers JE, Smith RS, Pringle JM, Clark GF, Gribben PE, Hewitt CL, Inglis GJ, Johnston EL, Ruiz GM, Stachowicz JJ, Bishop MJ (2015) Invasion Expansion: Time since introduction

- best predicts global ranges of marine invaders. *Scientific Reports* 5: 12436.
<https://doi.org/10.1038/srep12436>
- CABI (www.cabi.org/isc/) Invasive Species Compendium.
- Chamberlain SA, Szöcs E (2013) taxize: taxonomic search and retrieval in R. *F1000Research* 2.
<https://doi.org/10.12688/f1000research.2-191.v2>
- Costello CJ, Solow AR (2003) On the pattern of discovery of introduced species. *Proceedings of the National Academy of Sciences* 100: 3321–3323.
<https://doi.org/10.1073/pnas.0636536100>
- Crawley MJ, Harvey PH, Purvis A (1996) Comparative Ecology of the Native and Alien Floras of the British Isles. *Philosophical Transactions: Biological Sciences* 351: 1251–1259.
- Crooks JA (2005) Lag times and exotic species: The ecology and management of biological invasions in slow-motion1. *Écoscience* 12: 316–329. <https://doi.org/10.2980/i1195-6860-12-3-316.1>
- Crooks JA, Soulé ME, Sandlund O (1999) Lag times in population explosions of invasive species: causes and implications. *Invasive species and biodiversity management* 24: 103–125.
- Daehler CC (2009) Short Lag Times for Invasive Tropical Plants: Evidence from Experimental Plantings in Hawai'i. *PLOS ONE* 4: e4462. <https://doi.org/10.1371/journal.pone.0004462>
- Dodd AJ, Burgman MA, McCarthy MA, Ainsworth N (2015) The changing patterns of plant naturalization in Australia. *Diversity and Distributions* 21: 1038–1050.
<https://doi.org/10.1111/ddi.12351>
- Dyer EE, Redding DW, Blackburn TM (2017) The global avian invasions atlas, a database of alien bird distributions worldwide. *Scientific Data* 4: 170041.
<https://doi.org/10.1038/sdata.2017.41>

- Essl F, Latombe G, Lenzner B, Pagad S, Seebens H, Smith K, Wilson JRU, Genovesi P (2020) The Convention on Biological Diversity (CBD)'s Post-2020 target on invasive alien species – what should it include and how should it be monitored? *NeoBiota* 62: 99–121. <https://doi.org/10.3897/neobiota.62.53972>
- Evenhuis NL, Miller SE (2015) Twenty Years of the Records of the Hawai'i Biological Survey. Records of the Hawai'i Biological Survey for 2014. Part II: Index. 1 Edited by Neal L. Evenhuis & Scott E. Miller. Bishop Museum Occasional Papers 117: 1–2.
- GISD (2021) Global Invasive Species Database. Available from: www.issg.org/database.
- Gladstone NS, Bordeau TA, Leppanen C, McKinney ML (2020) Spatiotemporal patterns of non-native terrestrial gastropods in the contiguous United States. *NeoBiota* 57: 133.
- Groom Q, Desmet P, Reyserhove L, Adriaens T, Oldoni D, Vanderhoeven S, Baskauf SJ, Chapman A, McGeoch M, Walls R, Wiczorek J, Wilson J, Zermoglio PF, Simpson A (2019) Improving Darwin Core for research and management of alien species. *Biodiversity Information Science and Standards* 3: e38084. <https://doi.org/10.3897/biss.3.38084>
- Haynes K, Eckley IA, Fearnhead P (2017) Computationally Efficient Change-point Detection for a Range of Penalties. *Journal of Computational and Graphical Statistics* 26: 134–143. <https://doi.org/10.1080/10618600.2015.1116445>
- Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, Yang Q, Zhu C, Qiao H (2021) Sampling biases shape our view of the natural world. *Ecography* 44: 1259–1269. <https://doi.org/10.1111/ecog.05926>
- Imada CT (2019) Hawaiian naturalized vascular plant checklist (February 2019 update). Bishop Museum Technical Report 69: 23 pp.
- Jetz W, McGeoch MA, Guralnick R, Ferrier S, Beck J, Costello MJ, Fernandez M, Geller GN, Keil P, Merow C, Meyer C, Muller-Karger FE, Pereira HM, Regan EC, Schmeller DS, Turak E (2019) Essential biodiversity variables for mapping and monitoring species

populations. *Nature Ecology & Evolution* 3: 539–551. <https://doi.org/10.1038/s41559-019-0826-1>

Killick R, Eckley I (2014) *changept*: an R package for changepoint analysis. *Journal of Statistical Software* 58: 1–19.

Killick R, Fearnhead P, Eckley IA (2012) Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association* 107: 1590–1598.

van Kleunen M, Pyšek P, Dawson W, Essl F, Kreft H, Pergl J, Weigelt P, Stein A, Dullinger S, König C, Lenzner B, Maurel N, Moser D, Seebens H, Kartesz J, Nishino M, Aleksanyan A, Ansong M, Antonova LA, Barcelona JF, Breckle SW, Brundu G, Cabezas FJ, Cárdenas D, Cárdenas-Toro J, Castaño N, Chacón E, Chatelain C, Conn B, Dechoum M de S, Dufour-Dror J-M, Ebel AL, Figueiredo E, Fragman-Sapir O, Fuentes N, Groom QJ, Henderson L, Inderjit, Jogan N, Krestov P, Kupriyanov A, Masciadri S, Meerman J, Morozova O, Nickrent D, Nowak A, Patzelt A, Pelter PB, Shu W, Thomas J, Uludag A, Velayos M, Verkhosina A, Villaseñor JL, Weber E, Wieringa JJ, Yazlık A, Zeddani A, Zykova E, Winter M (2019) The Global Naturalized Alien Flora (GloNAF) database. *Ecology* 100: e02542. <https://doi.org/10.1002/ecy.2542>

Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2020) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* 70: 23–30. <https://doi.org/10.1093/biosci/biz140>

Mangiante MJ, Davis AJS, Panlasigui S, Neilson ME, Pflingsten I, Fuller PL, Darling JA (2018) Trends in nonindigenous aquatic species richness in the United States reveal shifting spatial and temporal patterns of species introductions. *Aquatic Invasions* 13: 323–338. <https://doi.org/10.3391/ai.2018.13.3.02>

McGeoch M, Arlé E, Belmaker J, Buba Y, Clarke D, Essl F, García-Berthou E, Groom Q, Henriksen M, Jetz W, Kühn I, Lenzner B, Meyer C, Pagad S, Pili A, Roigé Valiente M,

- Seebens H, Tingley R, Vicente J, Winter M (2021) Policy-relevant indicators for invasive alien species assessment and reporting. <https://doi.org/10.1101/2021.08.26.457851>
- Meares J (1790) Voyages made in the years 1788 and 1789, from China to the north west coast of America. London : Logographic Press, 372 pp. Available from: <https://open.library.ubc.ca/collections/bcbooks/items/1.0315360> (September 14, 2020).
- Moura MR, Costa HC, Peixoto MA, Carvalho ALG, Santana DJ, Vasconcelos HL (2018) Geographical and socioeconomic determinants of species discovery trends in a biodiversity hotspot. *Biological Conservation* 220: 237–244. <https://doi.org/10.1016/j.biocon.2018.01.024>
- Muñoz-Mas R, García-Berthou E (2020) Alien animal introductions in Iberian inland waters: An update and analysis. *Science of The Total Environment* 703: 134505. <https://doi.org/10.1016/j.scitotenv.2019.134505>
- Neal MC (1965) *In Gardens of Hawai‘i*. Bishop Museum Press, Honolulu.
- POWO (2019) *Plants of the World Online*. Facilitated by the Royal Botanic Gardens, Kew. *Plants of the World Online*. Available from: <http://www.plantsoftheworldonline.org> (October 29, 2020).
- R Core Team (2020) *R: A language and environment for statistical computing*, v 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reaser JK, Burgiel SW, Kirkey J, Brantley KA, Veatch SD, Burgos-Rodríguez J (2019) The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. *Biological Invasions*. <https://doi.org/10.1007/s10530-019-02156-w>
- Rojas-Sandoval J, Acevedo-Rodriguez P (2015) Naturalization and invasion of alien plants in Puerto Rico and the Virgin Islands. *Biological Invasions* 17: 149–163. <https://doi.org/10.1007/s10530-014-0712-3>

- Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE, Kleunen M van, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J, Pyšek P, Winter M, Essl F (2021) Projecting the continental accumulation of alien species through to 2050. *Global Change Biology* 27: 970–982. <https://doi.org/10.1111/gcb.15333>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pysek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jager H, Kartesz J, Kenis M, Kreft H, Kuhn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajero K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* 8. <https://doi.org/10.1038/ncomms14435>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pysek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guenard B, Jager H, Kartesz J, Kenis M, Kuhn I, Lenzner B, Liebhold AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajero K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Smith RM, Baker RHA, Collins DW, Korycinska A, Malumphy CP, Ostojá-Starzewski JC, Prior T, Pye D, Reid S (2018) Recent trends in non-native, invertebrate, plant pest establishments in Great Britain, accounting for time lags in reporting. *Agricultural and Forest Entomology* 20: 496–504. <https://doi.org/10.1111/afe.12282>
- Solow AR, Costello CJ (2004) Estimating the Rate of Species Introductions from the Discovery

- Record. *Ecology* 85: 1822–1825. <https://doi.org/10.1890/03-3102>
- Staples G, Herbst D (2005) *A tropical garden flora: plants cultivated in the Hawaiian Islands and other tropical places*. Bishop Museum Press., Honolulu, HI.
- Wagner WL, Herbst DR, Lorence DH (2005) *Flora of the Hawaiian Islands*
Website. <http://botany.si.edu/pacificislandbiodiversity/Hawaiianflora/index.htm>.
- Walpole M, Almond REA, Besançon C, Butchart SHM, Campbell-Lendrum D, Carr GM, Collen B, Collette L, Davidson NC, Dulloo E, Fazel AM, Galloway JN, Gill M, Govers T, Hockings M, Leaman DJ, Morgan DHW, Revenga C, Rickwood CJ, Schutyser F, Simons S, Stattersfield AJ, Tyrrell TD, Vié J-C, Zimsky M (2009) Tracking Progress Toward the 2010 Biodiversity Target and Beyond. *Science* 325: 1503–1504.
<https://doi.org/10.1126/science.1175466>
- Wester L (1992) Origin and distribution of adventive alien flowering plants in Hawai‘i. In: Stone CP, Smith CW, Tunison JT (Eds), *Alien plant invasions in native ecosystems of Hawai‘i*. University of Hawai‘i Press, Honolulu. Coop. National Resources Studies Unit, University of Hawai‘i, Honolulu, 99–154.
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer International Publishing. <https://doi.org/10.1007/978-3-319-24277-4>
- Zenetos A, Gratsia E, Cardoso A-C, Tsiamis K (2019) Time lags in reporting of biological invasions: the case of Mediterranean Sea. *Mediterranean Marine Science* 20: 469–475.
<https://doi.org/10.12681/mms.20716>

Tables

Table 3.1 Quantification of trendlines using changepoint analysis with each each value representing an average of 10 runs

Manipulated Variables						Responding Variables							
Lags	Data Harvest Point	Species Retrieval Rate	Species records not retrieved	Figure	Date used for analysis	Species lost from non-elapsd lags	Length of starting segments (years)	Rate for starting segments (mean spp/year \pm SD)	Length of truncated segment (years)	Rate for truncated segment (mean spp/year \pm SD)	Length of ending segments (years)	Rate for ending segments (mean spp/year \pm SD)	Rate for Entire Line
No Lags Exist	NA	NA	0	All Figures ("True" Naturalization Rate)	Hypothetical Naturalization Date	NA	0	NA	100	10.00 \pm 3.50	0	NA	10
Only Field Surveying Lags Exist	Species Reported as Naturalized by 2020	100%	0	Figure 3.2B	Detection Date	92	12	1.75 \pm 1.22	88	9.96 \pm 3.22	0	NA	9.08
Only Data Processing Lags Exist	Species Reported as Naturalized by 2020	100%	0	Figure 3.2C	Detection Date	95	0	NA	88	9.96 \pm 3.35	12	2.09 \pm 1.25	9.05
Both Field Surveying and Data Processing Lags Exist	Species Reported as Naturalized by 2020	100%	0	Figure 3.2D	Detection Date	189	13	2.28 \pm 1.50	74	9.99 \pm 3.29	13	2.32 \pm 1.34	8.11
All Lags Constant Over Time	Species Reported as	100%	0	Figure 3.3A, 4B	Approximated Introduction Date	286	4	3.47 \pm 1.93	66	9.94 \pm 3.27	42	0.73 \pm 0.71	6.32

	Naturalized by 2020				Detection Date		12	1.64 ± 1.30	70	9.95 ± 3.20	21	1.08 ± 0.93	7.25
Field Surveying Lags Shorten Over Time	Species Reported as Naturalized by 2020	100%	0	Figure 3.3B	Approximated Introduction Date	225	5	3.07 ± 1.70	74	9.89 ± 3.21	32	0.73 ± 0.85	6.76
					Detection Date		13	12.3 ± 1.26	65	10.98 ± 3.28	22	1.29 ± 1.04	7.75
Data Processing Lags Shorten Over Time	Species Reported as Naturalized by 2020	100%	0	Figure 3.3C	Approximated Introduction Date	154	4	2.14 ± 1.12	84	9.83 ± 3.34	27	1.01 ± 0.96	7.33
					Detection Date		11	1.49 ± 1.05	81	9.85 ± 3.24	8	2.95 ± 1.69	8.46
Field Surveying & Data Processing Lags Shorten Over Time	Species Reported as Naturalized by 2020	100%	0	Figure 3.3D	Approximated Introduction Date	74	5	3.25 ± 1.7	90	9.86 ± 3.27	20	1.26 ± 1.14	7.99
					Detection Date		13	2.22 ± 1.38	79	10.99 ± 3.41	8	3.27 ± 2.14	9.26
Field Surveying Lags Lengthen Over Time	Species Reported as Naturalized by 2020	100%	0	Figure 3.3E	Approximated Introduction Date	262	5	3.33 ± 1.75	68	9.97 ± 3.20	40	0.94 ± 0.76	6.44
					Detection Date		0	NA	79	8.89 ± 3.39	21	$0.95 \pm .84$	7.38
Data Processing	Species Reported as	100%	0	Figure 3.3F	Approximated Introduction Date	240	5	2.93 ± 1.53	73	9.96 ± 3.26	36	0.74 ± 0.58	6.62

Lags Lengthen Over Time	Naturalized by 2020				Detection Date		11	1.00 ± 0.87	74	9.83 ± 3.23	16	$1.18 \pm .101$	7.61
Field Surveying & Data Processing Lags Lengthen Over Time	Species Reported as Naturalized by 2020	100%	0	Figure 3.3G	Approximated Introduction Date	213	6	3.05 ± 1.59	74	9.94 ± 3.25	35	0.87 ± 0.68	6.79
					Detection Date		0	NA	88	8.93 ± 3.30	13	0.74 ± 0.52	7.87
All Lags Constant Over Time	Species Vouchered and Identified by 2020	100%	0	Figure 3.4A	Approximated Introduction Date	183	13	1.98 ± 1.14	79	9.95 ± 3.38	30	1.28 ± 0.96	8.17
					Detection Date		5	3.07 ± 1.72	78	9.87 ± 3.24	3	0.72 ± 0.72	7.11
All Lags Constant Over Time	All-Species Publication Compiled in 2000	100%	0	Figure 3.4C	Approximated Introduction Date	375	5	3.52 ± 1.69	58	9.84 ± 3.31	50	0.67 ± 0.61	5.44
					Detection Date		11	1.21 ± 1.07	60	9.95 ± 3.42	29	0.51 ± 0.65	6.24
All Lags Constant Over Time	Species Reported as Naturalized by 2020	85%	107	Figure 3.4D	Approximated Introduction Date	287	4	2.41 ± 1.37	69	8.21 ± 3.17	41	0.68 ± 0.71	5.29
					Detection Date		10	0.74 ± 0.59	73	8.09 ± 3.23	19	0.45 ± 0.46	6.06
All Lags Constant Over Time	Species Reported as	50%	356	Figure 3.4E	Approximated Introduction Date	289	4	2.26 ± 1.05	68	4.83 ± 2.25	41	0.29 ± 0.41	3.13

	Naturalized by 2020				Detection Date		10	0.45 ± 0.33	75	4.69 ± 2.40	16	0.27 ± 0.42	3.55
All Lags Constant Over Time	Species Reported as Naturalized by 2020	Higher Retrieval in Recent Years	258	Figure 3.4F	Approximated Introduction Date	290	4	2.00 ± 0.75	69	6.09 ± 2.44	40	0.58 ± 0.65	4.01
					Detection Date		11	0.42 ± 0.35	72	6.05 ± 2.77	19	0.68 ± 0.54	4.52
All Lags Constant Over Time	Species Reported as Naturalized by 2020	Lower Retrieval in Recent Years	221	Figure 3.4G	Approximated Introduction Date	286	5	3.04 ± 1.31	68	6.83 ± 2.61	42	0.37 ± 0.49	4.32
					Detection Date		11	1.04 ± 0.80	72	6.67 ± 2.77	18	0.34 ± 0.27	4.93

Figures

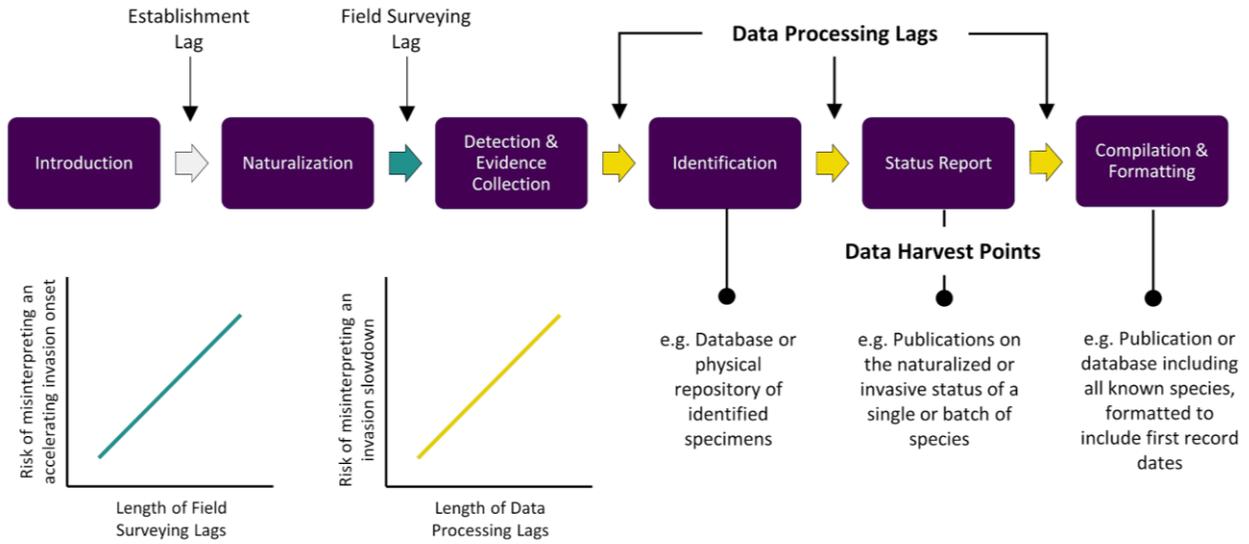


Figure 3.1 Schematic representation of the pipeline for collecting and processing biodiversity data, including time lags (yellow horizontal arrows) that accumulate during the processing of naturalized species data as well as typical points at which the data are harvested for naturalization or invasion trend analyses. The line charts in the bottom left represent a conceptual summary of how the lengthening of field surveying and data processing influence the risk of misinterpreting invasion trends, which are demonstrated in further detail in Figures 2-4.

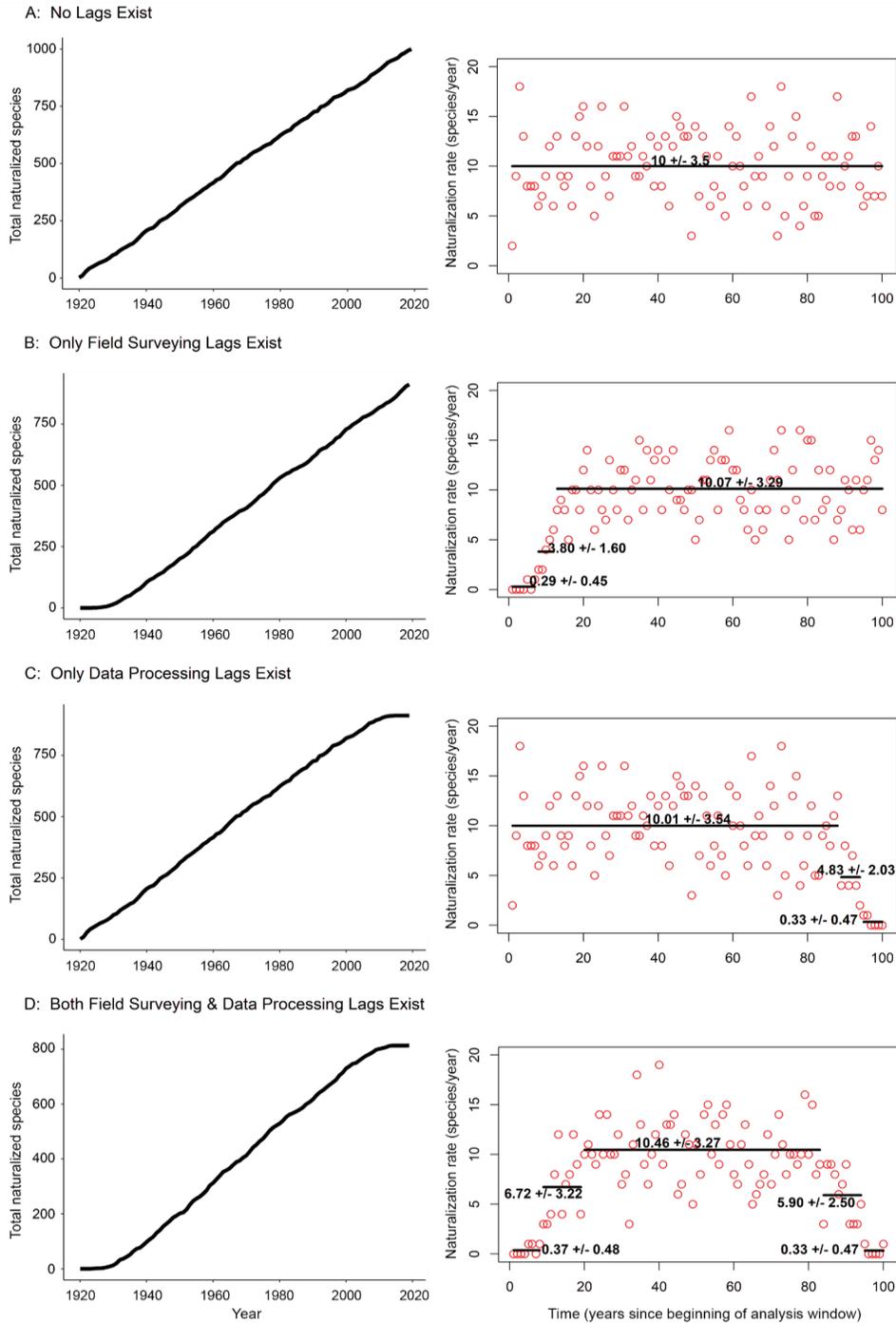
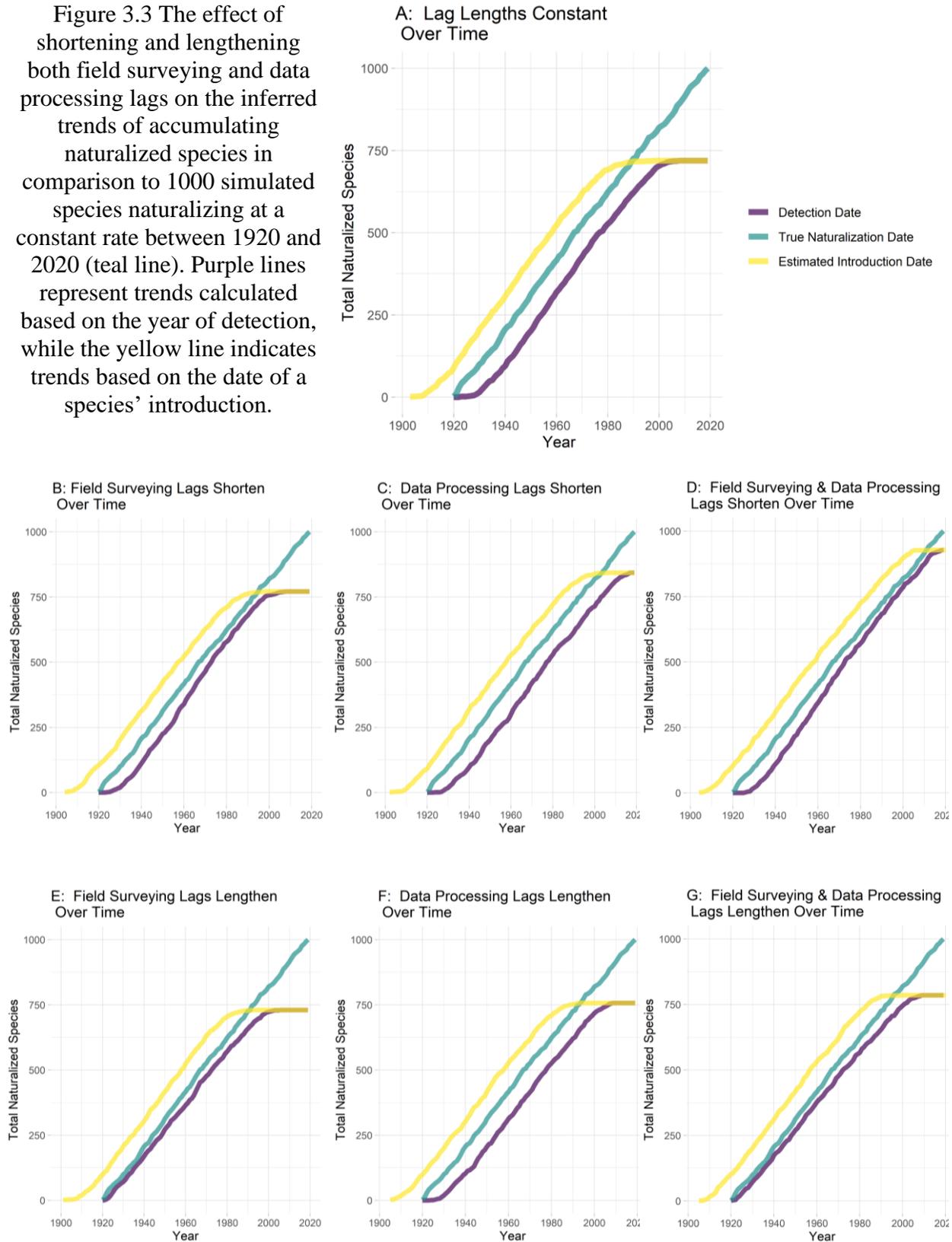
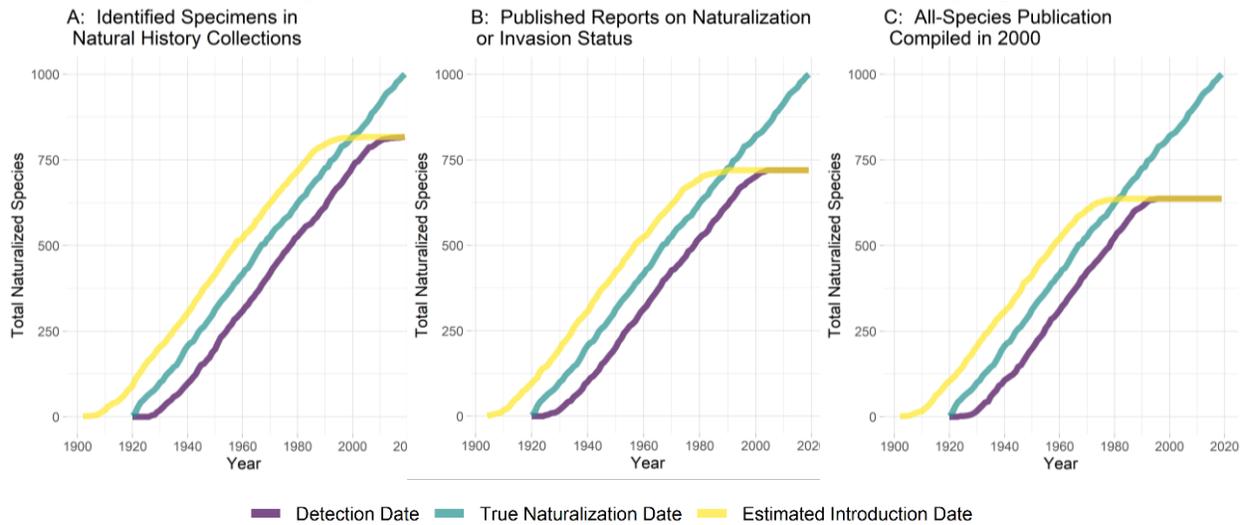


Figure 3.2 A comparison of field surveying and data processing lags showing that field surveying lags affect the interpreted naturalization rate at invasion onset whereas data processing lags affect recent years. Figures in the left column show the accumulation of total naturalized species over time while figures in the right column show the corresponding changepoint analysis conducted on the naturalization rate (red circles = species/year). The length of the horizontal black line represents the length of the trendline segment, while its vertical position indicates its naturalization rate, with numbers denoting the rate +/- the standard deviation.

Figure 3.3 The effect of shortening and lengthening both field surveying and data processing lags on the inferred trends of accumulating naturalized species in comparison to 1000 simulated species naturalizing at a constant rate between 1920 and 2020 (teal line). Purple lines represent trends calculated based on the year of detection, while the yellow line indicates trends based on the date of a species' introduction.



Harvesting Data at Different Points in the Data Processing Pipeline



Differences in Record Retrieval Rate

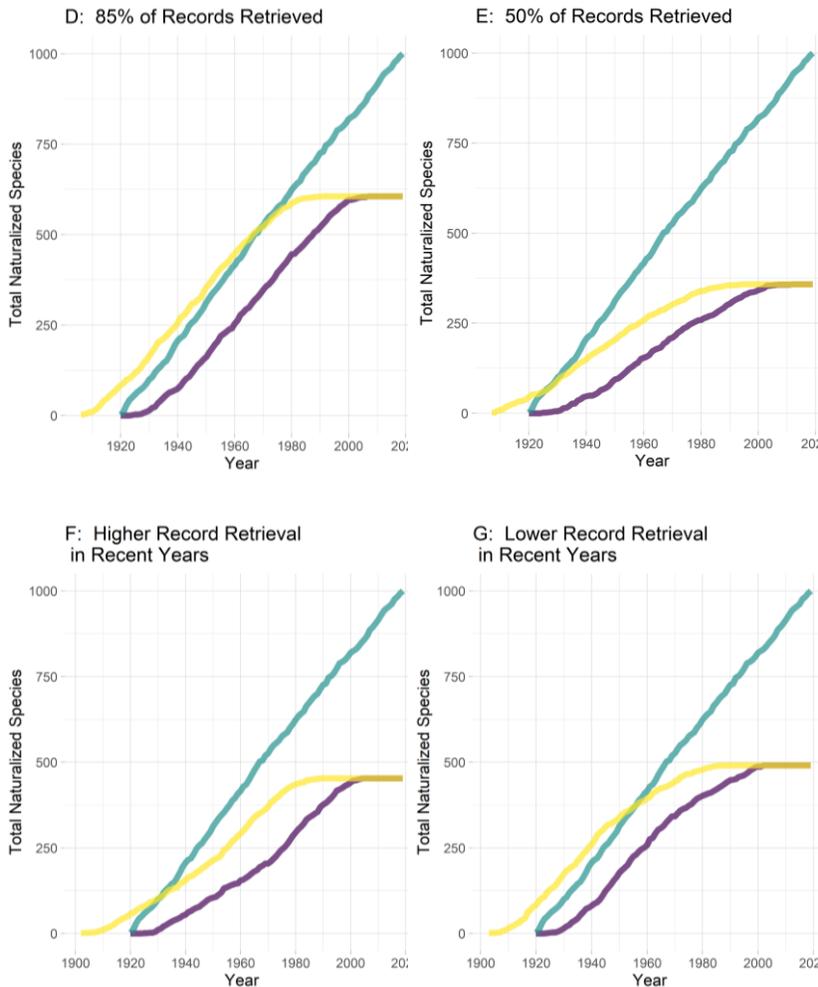
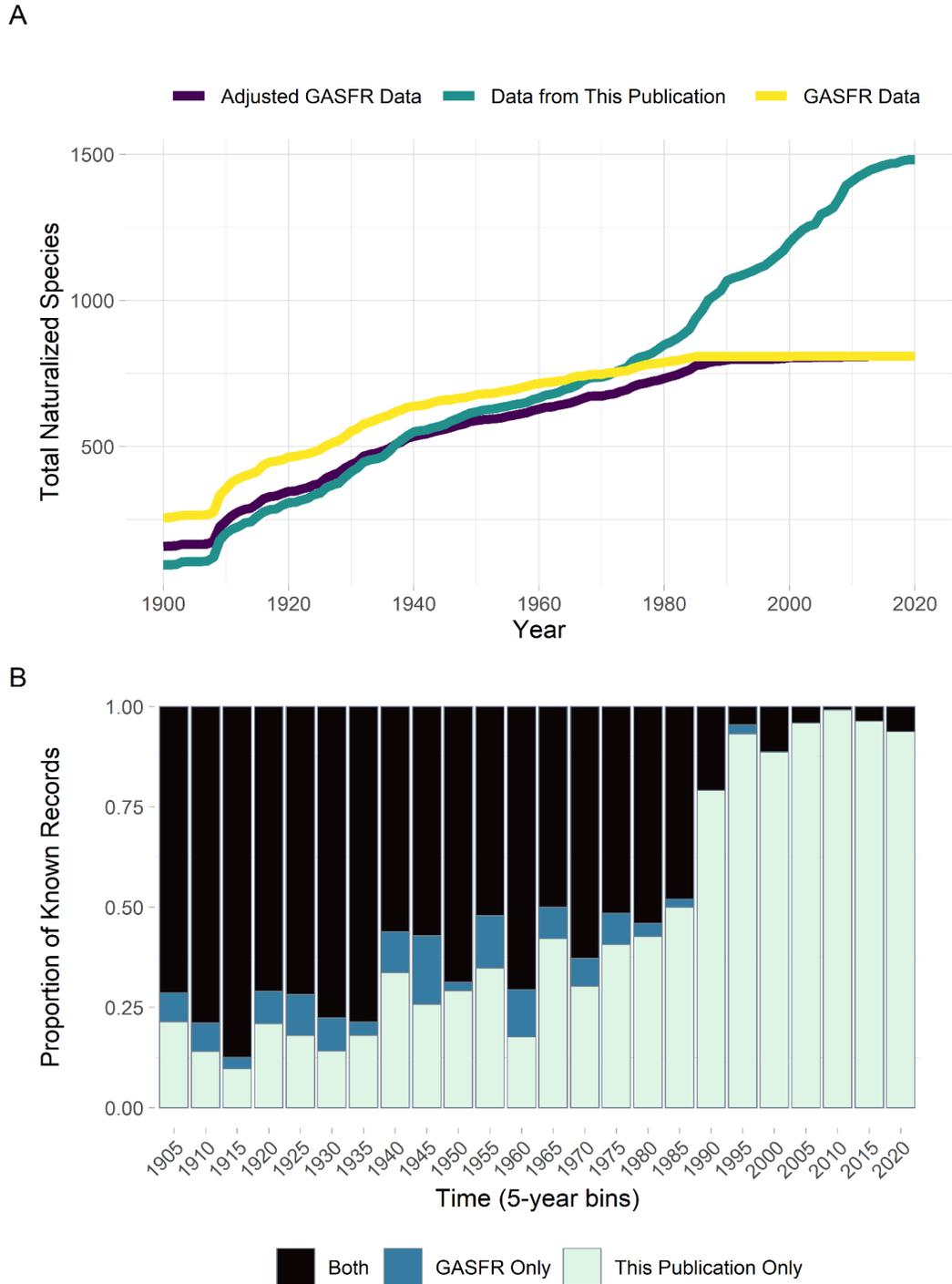


Figure 3.4 The influence variable data collection has on inferred naturalization trendlines, including intentional methodology (how data are harvested) and incomplete sampling. Panels A- C show examples of data being harvested at different points in the data processing pipeline, including identified specimens, naturalization /invasion status reports, and periodically published expert-compiled resources, respectively. Panels D- E represent = 85% and 50% retrieval of published records, respectively, with each species having an equal chance of being retrieved for analysis. Panels F-G represent a temporal sampling bias in which species have an unequal chance of being retrieved, with F = higher retrieval of records in recent decades (50% from 1920-1970, and 85% from 1970-2020), and G = lower retrieval of records in recent decades (85% from 1920-1970, and 50% from 1970-2020).



Chapter 4 – Applying an invasion and risk framework to track non-native island floras and inform management: a case study of challenges and solutions in Hawai‘i

Citation: Brock KC, Daehler CC (2020) Applying an invasion and risk framework to track non-native island floras: a case study of challenges and solutions in Hawai‘i. *NeoBiota* 62: 55–79. <https://doi.org/10.3897/neobiota.62.52764>

Abstract

Islands are plant invasion hotspots, with some having more non-native than native species. Many plants are recent arrivals, leading to concerns that their full spread and impacts are not yet realized. Given that islands host extraordinary numbers of endemic and threatened species, schemes are urgently needed to track the complex, species-rich but data-poor scenarios typical of islands. This study applies the framework by Blackburn et al. (2011) for categorizing invasion stages to Hawai‘i’s non-native plant checklist and identifies potential uses and complications for species tracking and invasion management. Data deficiencies and ambiguities required lumping Blackburn et al.’s categories to align with Hawai‘i’s available data; nonetheless, this coarser categorization describes invasion phases relevant to managers and could provide the basis for an effective tracking system. However, the framework does not accommodate uncertain invasion statuses, which prevents clear categorization of species that exist outside of cultivation but are not definitely naturalized. In response to this obstacle, scores from the Hawai‘i–Pacific Weed Risk Assessment (WRA) are explored to understand their application for predicting naturalization, including standard WRA scores as well as alternative scoring methods. We show that this predictive tool may be a promising supplement to on-the-ground monitoring for data-deficient elements of a flora. Finally, a categorization system for tracking statuses of an entire non-native flora is proposed that requires limited investments in additional data collection while following the rationale of Blackburn et al.’s scheme. This new categorization system may be

used to reveal overall invasion patterns and trends in a region, leading to valuable insights that can inform strategies for biodiversity management and conservation.

Introduction

Oceanic islands have long been considered highly invulnerable, hosting higher ratios of native to non-native richness per area than climatically comparable mainland regions (Lonsdale 1999, Sax and Gaines 2008). Around 40% of island floras are now composed of at least 40% non-native species (Pyšek et al. 2017). These trends are particularly alarming because remote islands are hotspots of endemism, with more than a quarter of all plant species native to islands (Kier et al. 2009). Naturalizations thus far greatly exceed extinctions reported on islands, although many naturalizations are too recent for their impacts to be fully realized, and even historical ones are unlikely to have spread throughout their available habitats yet (Sax and Gaines 2008). These trends indicate a potential “extinction debt” that may be paid against island endemics in the future (Kuussaari et al. 2009, Gilbert and Levine 2013).

Uncertainty of impacts from numerous naturalized species, especially on remote islands with unique endemic taxa, severely complicates management strategies. Conservation decision-making is often based on the precautionary principle, a central concept which proposes that actions taken to prevent known negative consequences should also be applied to scenarios where negative consequences are possible but uncertain (Cooney 2004). In an information-deficient context, the precautionary principle compels conservation biologists to treat all non-native species as if they will inflict negative impacts (a.k.a. “guilty until proven innocent”). Applying this strategy to scenarios where a thousand or more potential invaders exist is unrealistic, as conservation resources are limited; thus, managers are forced to select control targets based on few data and often rely on reports of invasive behavior elsewhere. Although invasion history is

critical for assessing risk, Kueffer et al. (2010) show that problematic species on one island are not necessarily problematic on all, or even most, islands where they are present. Furthermore, invasions of new species without invasive histories continue to be reported worldwide (Seebens et al. 2017). Thus, relying on either a precautionary principle or “bad apple” approach alone could incur high costs, resulting in inefficient and ultimately ineffective conservation of threatened native species and ecosystems.

Given that non-native species are frequently established on islands and that their behavior over time is uncertain, two approaches are needed to provide the basis for evaluating current and future impacts: 1) tracking non-native species along the introduction-naturalization-invasion continuum, and 2) predicting the likelihood of naturalization and invasion when field data are sparse or temporally limited (Wilson et al. 2014, 2018). While the latter has attracted significant attention, the former has been tremendously underappreciated (Hulme 2006), leaving many regions ill-equipped to manage the ever-increasing non-native portion of their floras.

Characterizing the phase and extent of an invasion is critical for assessing control feasibility. Monitoring already-introduced species is also necessary to fine tune predictive tools and catch species that invariably defy expectations. For many regions, the monitoring of non-native species begins with checklists of plants that have naturalized, often from data collected for floristic projects (Pyšek et al. 2004). However, some checklists further categorize species, distinguishing between invasive versus naturalized, old versus newer introductions and/or noting species for which data are deficient (Pyšek et al. 2002, Galasso et al. 2018, Imada 2019). Checklists of cultivated species that are not necessarily naturalized have been compiled for some regions but are rarer (Danihelka et al. 2012). Conservation workers may mistakenly use these checklists as tracking systems, but these lists merely describe snapshots of non-native populations when they

were reported. Such surveys may be incomplete and may be several decades old. Given that many introductions are recent and statuses of these plants may change rapidly, supplementing checklists with additional data to track the invasion process is essential (Wilson et al. 2014).

Nearly three decades' worth of work to characterize the invasion process worldwide and across taxa has culminated in a unified framework proposed by Blackburn et al. (2011). The system proposed, hereon referred to as "the framework", uses 11 categories to describe non-native populations by their phase of invasion, thereby providing a method for tracking statuses. These categories describe a population's progress beginning with transport and introduction through establishment and spread, which correspond to six sequential barriers to invasion success: geography, captivity/cultivation, survival, reproduction, dispersal, and environment. Furthermore, categories also describe between-barrier details (e.g. B1–B3 refers to explicit, limited, and no measures of containment in place, respectively); an inability to breach any one of these barriers amounts to invasion failure (Blackburn et al. 2011). If population statuses are updated frequently enough, this system can be used in conjunction with measures of commonness and distribution to help invasive species managers prioritize targets for control (McGeoch and Latombe 2016, Wilson et al. 2018). Wilson et al. (2014) outlined the framework's application for monitoring the status of invasive trees globally, and it was successfully applied to assess *Acacia* and *Melaleuca* in South Africa, yielding categories on which to base management decisions (Jacobs et al. 2017, Magona et al. 2018). However, no study has assessed the framework's applicability to track an entire non-native flora, including data deficient and cultivated species. Consistent categorization across species and over time for a region allows assessment of invasion trends, which can then be used to suggest broader scale approaches to invasion management.

The Hawaiian archipelago is an excellent model to assess methods for invasion tracking and prediction as >55% of the total terrestrial vascular flora is comprised of naturalized species (Imada 2012, 2019, Ranker 2016, Price and Wagner 2018). As such, Hawai‘i offers copious examples of both well-established and newly naturalized species to test theoretical and practical aspects of applying invasion frameworks. This includes over 1,600 naturalized and possibly naturalized species, with all but 25 thought to be introduced after 1,778 (European contact), and more than 600 species reported in the last 30 years (Imada 2019). The need to track and mitigate impacts from non-native plants is urgent. The IUCN lists 35% of Hawai‘i’s native flora as threatened, although this is very likely an underestimate as less than half of species have been assessed according to these international criteria (IUCN 2020). Hawai‘i is also an ideal location to evaluate invasion frameworks because an active community of botanists regularly produces reports of new species, contributing to manuals of both flowering plants and ferns that include non-native species (Wagner et al. 1999, Palmer 2003), as well as a periodically updated checklist of naturalized plants (Imada 2019). Moreover, Hawai‘i utilizes a predictive framework known as the Hawai‘i–Pacific Weed Risk Assessment (WRA) that has evaluated over 2,000 species and is used to discourage the planting of high risk plants and identify low risk alternatives, as well as inform managers of potential control targets (Kueffer and Loope 2009). A test of this system indicated that it is 95% successful in predicting major pests that were identified by expert opinion (Daehler et al. 2004). The Hawai‘i–Pacific WRA assigns risk rankings (“Low Risk”, “High Risk”, and an uncertain “Evaluate” category) by calculating a numerical score based on 49 questions about a plant’s biology, which can be divided into 35 questions pertaining to the likelihood that a plant will spread and 14 pertaining to the consequences of their spread (Daehler et al. 2004, Daehler and Virtue 2010). Although weed risk assessment (WRA) is most often

highlighted as a tool to prevent harmful introductions, the resulting WRA data set may also be useful for predicting the progress of already-introduced species along the introduction-naturalization-invasion continuum.

In this study, we assess the viability of applying the framework proposed by Blackburn et al. (2011) to a checklist of non-native plants for the Hawaiian Islands and discuss its applicability for tracking invasions in Hawai‘i. For data-deficient species whose invasion phase cannot be confidently determined from field data, we assess the potential for the Hawai‘i–Pacific WRA to infer the naturalization category to which they might belong. Finally, we propose a modified set of categories based on the framework that are relevant for making management decisions and can accommodate data deficiencies such as those commonly observed in Hawai‘i.

Methods

Aligning the naturalized species checklist

The Bishop Museum’s checklist of naturalized plants (Imada 2019) provided the main list of species for alignment with the population tracking categories outlined in Blackburn et al. (2011). This checklist tallies species that have naturalized or questionably naturalized on any island in the Hawaiian archipelago, including 8 main populated islands (Ni‘ihau, Kaua‘i, O‘ahu, Molokai, Lānai, Maui, Kaho‘olawe, Hawai‘i) ranging in maximum elevation from 380–4,200 m above sea level, and 11 small, sparsely vegetated islands and atolls, ranging from a few to 259 m above sea level (Kure, Midway, Pearl and Hermes, Lisianski, Laysan, Gardner Pinnacles, French Frigate Shoals, Necker, Nihoa, Kaula Rock, Lehua). The checklist sorts non-native species into five categories describing whether they are a Polynesian or post-European introduction, and whether there is uncertainty regarding a species introduction or naturalization status.

We additionally reviewed naturalization reports and herbarium specimen labels, supplementing the checklist by Imada (2019) to include recent naturalizations and recategorizing records that explicitly did not match the definition of naturalized. We define naturalized as non-native plants that survive and reproduce consistently to sustain populations in the wild over many generations without human aid (Richardson et al. 2000, Pyšek et al. 2004), corresponding to categories C3–onwards in the framework (Blackburn et al. 2011). The phrase “in the wild” indicates that a plant is growing outside of captivity or cultivation (not excluding species in human disturbed or urban habitats). Species not meeting these criteria were downgraded from “naturalized” to the “questionably naturalized” category in the checklist of naturalized plants (Imada 2019), defined as “species not confirmed to be naturalized, including present-day adventive and escaped plants, or historical collections not recently vouchered.” Ultimately, the compiled checklist used in our analyses included 1,668 species that have been collected from the wild on at least one of the Hawaiian Islands.

For each island, we summed the number of species in each of the following two categories: “naturalized” (including pre-European introductions by Polynesians), and “questionably naturalized” for the remaining species that could not be clearly categorized after considering available data and criteria in the framework. We then attempted to align species in these two groups, as well as non-naturalized cultivated species in Hawai‘i, with the framework categories. We highlighted specific examples of challenges encountered when applying the framework across a flora and assessed whether the framework is valuable for improving non-native plant species tracking and management in Hawai‘i and elsewhere. Subsequently, we constructed a modified categorization scheme that accommodates the data deficiencies found in Hawai‘i’s non-native flora.

Potential for Weed Risk Assessment scores to predict naturalization

To determine whether the Hawai‘i–Pacific WRA score or components of that score might be useful for inferring the status of “questionably naturalized” species, we examined the distribution of WRA scores among known naturalized species versus non-naturalized cultivated species. We also assessed the correlation between WRA score and number of islands where a species is known to be naturalized. The WRA scores were additionally separated into components related to likelihood of spread and potential consequences of impact, creating two independent scores for each species, following Daehler and Virtue (2010). Among 2,037 WRAs available for analysis (HPWRA 2019), we tested the hypothesis that the full WRA score or the likelihood-of-spread component of the WRA score (hereon referred to as “likelihood-only score”) can predict naturalization by comparing the scores of plants that have naturalized with those that were introduced but have not naturalized. The latter species were identified by cross-referencing a list of plants reported from cultivation assembled by Imada et al. (2000), supplemented with approximately 600 species known from cultivation based on herbarium vouchers. We excluded species that are thought to be very uncommon in cultivation, such as those known from a single collection in a botanical garden, as well as species that are likely held captive in aquaria and have little chance to escape (Staples and Herbst 2005). We also removed possible recent introductions (384 species) first reported in cultivation during the last 20 years because the behavior of these species is uncertain. A previous study found that lag times averaged less than 20 years between first planting and signs of naturalization in Hawai‘i (Daehler 2009; see also Schmidt and Drake 2011). Data on time since first cultivation were too unreliable to investigate longer lag periods, but for a few high-scoring species that had not naturalized, we

checked herbarium records to make sure they were not first recorded slightly earlier (20–40 years ago).

We conducted statistical analyses with the Python library SciPy 1.0 (Virtanen et al. 2020), visualized with seaborn 0.10.0 (Waskom et al. 2020). We used Welch's t -test to determine whether naturalized and non-naturalized species had significantly different WRA scores and likelihood-only scores. Finally, to see if scores are related to a species' potential to spread throughout the archipelago, we compared the WRA and likelihood-only scores to the number of islands naturalized for each plant species. These subsets of the data had non-normal distribution, so we calculated the strength of the correlation with the nonparametric Kendall's τ_b coefficient. The R package mblm (Komsta 2019) enabled us to visualize the linear trend between these variables with the Theil-Sen estimator, which was developed to reflect the strength of Kendall's τ_b (Sen 1968, Wilcox 2010). Kendall's τ_b is rank-based and the associated Theil-Sen estimator is derived from the median slope of pairwise data (Sen 1968). Island counts were restricted to the main Hawaiian Islands minus Ni'ihau, as these are more frequently surveyed and support more variable habitats in comparison to the low elevation islands in the northwestern section of the archipelago (Larrue et al. 2018).

Results

Alignment between the checklist and framework

Cultivated species

More than 7,300 cultivated species have been reported in Hawai'i, although data describing the circumstances of their containment are often lacking so that they cannot be finely categorized according to the framework. Thus, plants that are cultivated in Hawai'i but are not naturalized were roughly aligned to the lumped categories B1 (in captivity or quarantine) and B2

(in cultivation; Figure 4.1). Examples of B1 species may include houseplants that are only planted indoors, which may be less common in Hawai‘i than in temperate areas due to Hawai‘i’s tropical climate. Also falling into this category are valuable orchid species grown by hobbyists in controlled environments such as terraria or greenhouses. We inferred that most common ornamentals and food plants would belong to category B2, where species are planted in suitable conditions with limited intentional measures to restrict their dispersal (e.g. some incidental control or limitation through landscape activities). No species were thought to match categories B3 or C0, although examples from Hawai‘i likely exist that are not readily added to checklists because they are not currently present. For instance, forestry planting records indicate that *Juglans nigra* (black walnut) was planted to assess its suitability as a timber crop, but growth trials determined that its survivability was very low in Hawai‘i (C0) (Nelson and Schubert 1976). Although category B3 is thought to apply to most accidental introductions that are not yet reproducing in the wild (Blackburn et al. 2011), we included contaminants of horticulture in category B2 if some measure of containment is evident (e.g. existing in potting soil imported with an intentionally cultivated plant).

Questionably naturalized species

We considered one hundred eighty species to be “questionably naturalized” at the statewide level, amounting to 342 per-island introductions being classified in this checklist category. Based on our review of Hawai‘i’s records and recommended terminology for conceptualizing plant invasions (Richardson et al. 2000, Pyšek et al. 2004), the “questionably naturalized” category likely contains five distinct classes of species for which there is insufficient data to separate, including: 1) remnants of cultivation, such as long-lived species that were planted some time ago, but where growth of surrounding vegetation masks evidence of it

being a former cultivation site; 2) casual species, where immature or perhaps a few mature individuals originating from cultivated plants may exist outside of cultivation, but for which multiple generations are not produced (i.e. population not self-sustaining); 3) recently introduced invaders that will eventually naturalize, but have not had sufficient time to do so; 4) species that have already naturalized, but only a few individuals have been detected (i.e. current sampling is insufficient to confidently categorize populations that are actually naturalized); and 5) possibly extirpated species, for which historical records indicate that they existed in the wild at one time, but have not been observed for many decades (Figure 4.1).

Remnants from cultivation may arise when homesteads or forestry plots are no longer maintained, making it difficult to determine whether a species is a C1 (existing in the wild but not reproducing) or a mature individual of a naturalized population, especially if no historical planting data exists. An example of C1 species that may appear as naturalized in the field without prior knowledge of their planting history includes approximately 30 species of *Ficus* that were planted on forest reserves without the introduction of their specific pollinator wasp (Skolmen 1980). Casuals would theoretically align to category C2, but in practice, they are particularly difficult to distinguish from recently introduced invaders, as was also noted by Wilson et al. (2014) when describing tree invasions. This uncertainty arises because both categories may initially appear in the field as offspring from a single or few age classes, with the distance from the originally introduced plant dependent on its mode of dispersal (e.g. wind, gravity, or bird-dispersed) rather than its potential to naturalize. One example of a species in Hawai'i that fits the C2 category is *Sequoia sempervirens* (California redwood), for which more than 130,000 individuals were planted before 1960 (Nelson and Schubert 1976, Skolmen 1980). These plantings have been observed producing seedlings, but they have failed to mature beyond

the seedling stage, even after 60+ years, and thus, would likely vanish from the islands if the initial plantings were removed.

Although we currently lack data to distinguish recently introduced invaders from the rest of the “questionably naturalized” group on the checklist, we were conceptually unable to determine which framework category these species would belong to, even if they could be identified. Recently introduced species do not appear to belong to category C3, as they do not yet form self-sustaining populations with multiple generations, but nor do they align with category C2, for which self-sustaining populations will never be formed.

Naturalized species

Of the 1,668 species in our original wild growing non-native checklist, 1,473 are considered naturalized in Hawai‘i. However, these species are not uniformly naturalized across all islands, with no island containing naturalized populations of all these species. Considering each naturalization event separately per island, we counted 4,970 instances by summing the number of naturalizations from all islands. The checklist does not provide information on dispersal and formation of new populations, as needed to distinguish between the last four categories of the framework and thus, the “naturalized” category aligns broadly with C3–E (Figure 4.1). On the other hand, the checklist (Imada 2019) does provide information on the number of islands where each species has naturalized. This information is potentially useful for understanding invasions across island regions; however, the framework does not provide an additional category for species that have naturalized or invaded across multiple geopolitical boundaries. Thus, these species are lumped in the same category with single-island species.

Relationship between WRA score and naturalization

The Hawai'i-Pacific WRA dataset included 828 non-naturalized and 712 naturalized species after questionably naturalized and recently introduced non-naturalized species were removed (Figure 4.2). Differences in likelihood-only scores between naturalized ($\bar{x} = 3.91$, $SD = 4.25$) and non-naturalized ($\bar{x} = 0.23$, $SD = 3.49$) plants were significant (Welch's $t = 18.40$, $df = 1376.54$, $p < 0.001$). Scores were also significantly different between these groups using the standard WRA scoring method (Welch's $t = 27.93$, $df = 1303.27$, $p < 0.001$), but with more separation between the distributions of scores for naturalized ($\bar{x} = 10.41$, $SD = 6.90$) and non-naturalized species ($\bar{x} = 1.60$, $SD = 5.17$) compared to likelihood-only scores. Thus, the standard WRA scoring method differentiates naturalized and non-naturalized species better than likelihood-only scores, making it more useful for inferring the likely status of data deficient species (Daehler and Virtue 2010).

Only 11% of the 436 species with WRA scores less than 1 are naturalized in Hawai'i, which is the upper threshold score used by the Hawai'i-Pacific WRA to designate species as "Low Risk" (Figure 4.2). Lowering the scoring threshold from 0 to -3 for predicting non-naturalized species decreases the error rate by only 1% (to 10% out of 201 species). On the other hand, 78% of the 684 species scoring greater than 6 (the threshold used by the WRA to deem a plant "High Risk") were naturalized. Increasing the threshold score to 12 (including 321 species) is required to increase the representation of naturalized species to 90%.

The WRA score and likelihood-only score were significantly positively correlated with the number of islands on which a plant has naturalized ($p < 0.001$ for both). The trend was weak in both cases, although we observed a more positive correlation for the standard WRA score than the likelihood-only score (Kendall's $\tau_b = 0.27$ versus 0.14, respectively; Figure 4.3).

Predicting the fate of questionably naturalized species

Sixty-three of the 180 “questionably naturalized” species statewide have been assessed by the Hawai‘i–Pacific WRA (Table 4.1), for which we used WRA scores to infer naturalization status. We did not consider island-specific “questionably naturalized” species that have definitely naturalized on another island in the archipelago because these species have already demonstrated the ability to successfully naturalize in Hawai‘i. After removing 8 possibly extirpated species that have not been observed in more than 50 years or whose population declines have been closely monitored (based on herbarium specimens and expert opinion), 55 species remain in the data-deficient “questionably naturalized” category.

Based on the aforementioned relationship between WRA scores and naturalization, we find that eleven species have scores <1 , and therefore 89% of these are expected to not naturalize (Figure 4.2; Table 4.1). Reducing the threshold to -3 narrows the pool of questionably naturalized species to only three species and there is almost no change in the expected rate of not naturalizing among the three plant species in this group (90%). Conversely, our data indicate that 78% of the 27 questionably naturalized species with scores greater than 6 (Figure 2) are either naturalized already or will become naturalized. Raising this threshold to 12 identifies a set of only 6 species, 90% of which are likely to become naturalized.

Discussion

Our whole-flora analysis identified 342 cases where naturalization status is currently uncertain, emphasizing the need for effective tracking of non-native populations in the Hawaiian Islands. Hawai‘i’s current checklist (Imada 2019) forms the foundation for a future tracking system, but our attempt to align it to Blackburn et al.’s framework (2011) revealed both benefits and drawbacks of the framework in handling real-world data. Species listed as “questionably

naturalized” proved problematic, and although the Hawai‘i–Pacific WRA may assist with categorizing these species, adjustments to the framework are needed to create a usable system. Based on Hawai‘i’s non-native flora, we propose a 7-category scheme for tracking species’ statuses that largely aligns with the framework (Table 4.2). This tracking system includes the minimal number of status categories we think are critical for informing management decisions while accounting for common data deficiencies and ambiguities.

Ambiguous alignment of questionably naturalized species

The largest obstacle that arose when aligning the species checklist to the framework is that the “questionably naturalized” category contains species on fundamentally different trajectories, from species that have actually died out to those that will very soon become widespread invaders (Figure 4.1). Additional field data are needed to accurately categorize them. Integrating an uncertainty variable into the framework indicating that categories were assigned despite some data deficiencies could avoid pooling species with different trajectories if additional surveys are not possible. However, such a scheme would still not account for species that have been introduced very recently and are just beginning to naturalize. These species do not fit into any of the naturalized categories in the framework (C3–E) because there is not yet proof that they will form self-sustaining populations in the long-term. However, the preceding C2 category implies that these species definitely do not form self-sustaining populations. Unlike species uncertainly aligned to status categories because population data is unknown, additional surveys will not help classify recently introduced invaders because insufficient time has passed for that data to exist. Placing these data-deficient plants in C2 would remove the urgency to eradicate or at least monitor these species even though this is precisely the group that concerns early detection and eradication programs. Contrastingly, assigning these species to the C3

category is not a solution either because, in addition to not meeting the definition of naturalized, it could confound analyses by artificially inflating the number of naturalizations and potentially distract management toward less impactful species.

Hawai‘i’s checklist is not unique in having questionable status categories (Galasso et al. 2018), and similar problems distinguishing between C2 and C3 have been encountered in other applications of the framework to real-world scenarios (Wilson et al. 2014, Robinson et al. 2016). As discussed above, these mismatches sometimes arise from insufficient population data needed to confidently assign categories, which is common because reports of new species are often opportunistic and thorough surveys to delimit entire populations are rarely conducted. However, misalignments also stem from the framework’s conceptually discrete barriers along the invasion continuum, when in reality these boundaries are fuzzy. Moreover, the status of a species will often change over time. For instance, the framework acknowledges that several cycles of reproduction are necessary to form a self-sustaining population, thereby surpassing the reproduction barrier, but cannot specify how many are required. Thus, if a recently introduced species is encountered outside of cultivation and only two different life stages are visible (indicating at least 1 cycle of reproduction after dispersal), the self-sustainability of the population is questionable. Lack of understanding surrounding establishment success has recently been emphasized as an important factor biasing studies on geographical patterns of non-native species richness (Blackburn et al. 2020). Additionally, population sustainability is hard to predict because it depends on numerous interactions between species traits and environmental conditions, and thus should not be assumed (Duncan et al. 2019). The hundreds of species that have been reported as “questionably naturalized” in our analysis confirms that field botanists are often plagued with uncertainty (Magona et al. 2018). This problem emphasizes the need for

status tracking frameworks to address uncertainty, but solutions must assess whether that uncertainty arises from insufficient data about an existing population or from insufficient time to assess a new species' behavior.

WRA scores can assign a likely status for questionably naturalized species

Our data show that the WRA can be a useful tool for predicting naturalization, with scores obtained from the standard WRA scoring method being more able to distinguish naturalized from non-naturalized species than scores from the likelihood questions only (Figure 4.2). This result was surprising because, logically, the likelihood of spread questions should more accurately reflect the propensity for plants to naturalize, whereas impact-related questions seem less relevant to predicting naturalization. The standard WRA was also more strongly correlated with a species' tendency to naturalize across multiple islands relative to the likelihood-only score (Figure 4.3). However, weak correlations for both indicate that other factors besides those accounted for in WRA scores undoubtedly influence the repeated naturalization of these species across islands. For example, a species' ability to naturalize on multiple islands is likely more explained by deliberate or accidental between-island dispersal by humans (breaching the geographic barrier) rather than factors related to biology that are emphasized in the WRA.

Given that WRA scores appear to be more useful than scores derived from the likelihood questions only, the scoring thresholds currently in use to assess risk of weediness by the Hawai'i-Pacific WRA ($> 6 = \text{"High Risk"} , < 1 = \text{"Low Risk"}$) may be sufficient for predicting the likely status of data deficient "questionably naturalized" species (Table 4.1). Considering that 78% of naturalized species with a WRA score over 6 are naturalized, using this threshold to predict naturalization could lead to an error rate of up to 22%. Conversely, 89% of plants with

scores less than 1 were not naturalized, indicating an error rate of up to 11% when predicting that plants will not naturalize. Our ability to assess the WRA as a predictor of naturalization depends on accurate categorization of species in our test data. Undetected naturalizations or species with naturalization potential could exist in our non-naturalized data set (e.g. species whose establishment is limited by urban landscaping), even though rarely cultivated plants and introductions less than 20 years old were removed; this may account for the small skew towards higher scores in the distribution of non-naturalized species contributing to this error rate (Figure 4.2). However, this error rate may be acceptable to invasive species managers who often err on the side of caution when identifying species likely to naturalize.

By applying the Hawai‘i–Pacific WRA thresholds to “questionably naturalized” species (and accepting their associated error rates), we would infer that the 11 species (~20%) with scores less than 1 are unlikely to naturalize and instead belong to the C1 or C2 categories, while 27 (49%) species with scores greater than 6 would likely belong to, or eventually belong to, the C3 category or higher (Table 4.1). This leaves 18 species with scores ranging from 1–6 for which we are unable to infer a status category. Interestingly, scores for species not observed for more than 50 years (“possibly extirpated”) ranged from 3–17; although most were predicted to naturalize, they might have existed as very small populations (making them vulnerable to stochastic extinctions), or they might be naturalized at locations rarely visited by botanists.

Accommodating data deficiencies and management needs

Similar to other inventories that reference the framework when categorizing the status of non-native species (Wilson et al. 2014, Robinson et al. 2016, Henderson and Wilson 2017, Magona et al. 2018, Ansong et al. 2019), we found that the coarse categories typical of species checklists (e.g. cultivated, questionably naturalized and naturalized in Hawai‘i) did not align to

all eleven categories in the framework. Field surveys would be needed to update existing information and, in some cases, collect new types of data, to apply some framework categories. One interpretation of this problem is that the framework's fine categorization scheme cannot be realistically implemented as a tracking system for many regions, especially for developing island nations with limited funds for research and conservation (Russell et al. 2017). However, more status categories than are typically used in species checklists are needed to conduct analyses of patterns and trends to inform invasive species management. Accordingly, an intermediate approach is needed that avoids unrealistic requirements, such as the need for numerous site revisits over an extended time period, and emphasizes the use of information commonly available (e.g. occurrence reports, herbarium specimens, general survey data). To establish a tracking system based on this information, we can superimpose the framework's description of the invasion continuum (Blackburn et al. 2011) onto the checklist categories and the most readily attainable data (Table 4.2). The management goals described in the framework, which include prevention, eradication, containment and mitigation, help delineate a broader status categorization scheme that still informs management decisions.

When naming our proposed status categories in Table 4.2, we avoided the use of alpha- numerics (as are used in the framework) in favor of descriptive terms in order to facilitate an intuitive transition from a basic species checklist to a practical tracking system that is usable by non-specialists. To avoid confusion in terminology, our status category names are consistent with terminology used in the framework as well as other publications concerning communication of biological invasions (Richardson et al. 2000, Pyšek et al. 2004, Blackburn et al. 2011). Groom et al. (2019) provide a list of terms that correspond to each of the 11 categories used in the framework, although we were unable to directly use those terms because our system required

lumping some of the original framework categories. We avoid using the term “invasive” as a status because it is inconsistently applied and often reserved for species that cause harm (Colautti and MacIsaac 2004, Gbedomon et al. 2020) and unlike the biological barriers presented in the framework to describe a species’ journey through the invasion process, harm is context dependent (e.g. impacts to biodiversity versus socio-economic resources) and cannot be conceptualized as a hurdle that must be surpassed (Blackburn et al. 2011).

We combine B1 (measures of containment in place) and B2 (containment limited) into a single category “Contained/Cultivated” because although information about circumstances preventing dispersal can be informative for management, acquiring this data for an entire flora is difficult. Additionally, plants cultivated under strict containment measures are likely far less common than those with limited (or no) attempt to prevent dispersal. We include “Not Self-Sustaining”, aligning to C1–C2, and “Potentially Naturalizing”, referring to recently introduced species that appear to be in the process of naturalizing (but have not yet done so), which is not included as a category in the framework. The adoption of the latter category is important because it provides a list of species in need of careful monitoring and alerts invasive species managers to eradication possibilities without indicating that naturalization has occurred.

Describing the phase of naturalization for entire floras provides a unique challenge because distribution data may be insufficient to determine whether individuals are dispersing, surviving and reproducing at locations beyond introduction sites for numerous species (Blackburn et al. 2011). As a result, many species checklists do not attempt to further categorize species past naturalization (Uludag et al. 2017). On the other hand, data on the spread of species from known introduction sites (e.g. sites of cultivation or accidental seed contamination) may be available for well-surveyed species, especially those that are purposefully monitored by invasive

species control programs. Our solution to this data disparity is that species should be sorted into two categories when data is available, including “Naturalized Where Introduced” and “Naturalized Beyond Introduction Site”, or placed into a more general status category “Naturalized–Unspecified” (C3–E in the framework) if data are insufficient (Table 4.2). We chose to combine C3–D1 within “Naturalized Where Introduced”, including naturalized plants that may or may not disperse beyond their introduction site, but where survival of dispersed offspring is not sufficient (or not yet sufficient) to form new self-sustaining populations. This phenomenon may be observed for plants that are unable to spread beyond their immediate human disturbed surroundings (Rojas-Sandoval and Acevedo-Rodriguez 2015) because they are not adapted to conditions where they are dispersed. This status category contrasts “Naturalized Beyond Introduction Site” (aligning to D2–E), which are known to have spread naturally beyond any possible introduction sites. Species of the former status category are likely to have localized impacts and be easier to contain or eradicate, whereas species in the latter category could potentially already have wide-ranging impacts and require substantial intervention to control or prevent further spread

For the purposes of tracking regional floras, the A category (not transported outside native range) should be adapted to include likely invaders that have not yet arrived in the region of interest, identified via horizon scanning or risk assessment tools. Moreover, the framework does not categorize species that were once present in a region but are now absent and instead describes the invasion continuum as a unidirectional process, with multiple avenues for invasion failure, making it unclear how species can go backward in status. Thus, additional categories would be useful for species that were once wild growing but are now absent, with possible subcategorization according to whether a species disappeared before (C0–C2) or after

naturalization ($\geq C3$) if data is available (Table 4.2). Plant eradication programs exist on most main Hawai‘ian islands, which concentrate on removing species before naturalization or in the early stages of it (Kraus and Duffy 2010). Furthermore, several previously naturalized species that were never targeted by weed control programs have not been observed for numerous decades (Imada 2019) and are presumed to be naturally extirpated. In these examples, it is possible that some species were wrongly reported to be naturalized when in fact they failed to surpass the survival and reproduction barriers necessary to form self-sustaining populations (thus aligning to categories C0–C2), but this is now impossible to ascertain. To avoid confusion, a new category for “No Longer Present” is needed for species that are no longer wild growing and are not present in cultivation (even if data are not available to accurately assess their historic status) in order to identify re-introductions and re-invasions of species that were previously thought to be extirpated (Panetta 2015).

We found that although categories B3 (bypassing cultivation and directly released into the wild) and C0 (the same as B3 but no survival) are useful for conceptualizing barriers to invasion success, they are not practically applicable as status categories (Table 4.2). The reason for this is that plant species encountered in the field are assumed to be surviving (and possibly reproducing), and thus, are assumed to be at least C1. If plants are no longer present when sites are revisited, as would be the case for C0 species which are released in the wild but fail to survive, then field botanists would assume these species have become extirpated or are present in cultivation only (Table 4.2).

Implementation and future directions

Classification schemes used in species checklists, such as the 3 categories used in Hawai‘i’s checklist, may be converted to our system using typically available information while

accommodating more detailed population data when available. To account for uncertainty when assigning categories, we suggest that low, medium and high confidence levels be attached to each status, with guidance available in Supplementary Material 1 following previous examples for categorizing invasive impacts (Hawkins et al. 2015, Wilson et al. 2018). We additionally suggest that, when possible, statuses assigned with low–medium confidence be given a “likely status” using inferential tools, such as species distribution models and risk assessments (Table 4.2, Appendix). The likely status is not intended to be a prediction for a species’ future status but instead deduces its current reality, which is particularly useful for regions with numerous introductions and insufficient field monitoring, where inference may be the only method for estimating a snapshot of invasion statuses for a region’s flora. In our example, analysis of field data for plants in the “questionably naturalized” checklist category can be conducted to reassign as many species as possible into a more informative status category. As we demonstrate here, WRA scores may then be used to assign a likely status of either “Not Self-Sustaining” or one of the naturalized status categories (Table 4.1, 2, Appendix). Nevertheless, it is probable that further analyses and predictive tools will fail to reassign at least a few species. In our example (Table 4.1) we have labeled these unassignable species as having a “Data Deficient” likely status category to incentivize monitoring of these populations.

Assigning species’ statuses along the introduction-naturalization-invasion continuum is an important first step for developing a biodiversity informatics (rather than species-specific) approach to managing invasions and monitoring status changes over time. Changes in spatial extent and population size could further accompany our proposed status tracking system to strengthen assessments of both impacts and control feasibility. For instance, species that are just

beginning to naturalize would be categorized as “Potentially Naturalizing”, a status potentially assumed to be eradicable, but feasibility may be complicated by the presence of numerous introduction sites. Future efforts toward this goal could refer to McGeoch and Latombe (2016), who provide a framework for categorizing commonness and changes in population size.

Tracking impacts alongside statuses and population extent would be especially useful, and could be achieved by integrating the proposed IUCN Environmental Impact Classification for Alien Taxa (EICAT; Blackburn et al. 2014, Hawkins et al. 2015) and its socio-economic equivalent (SEICAT; Bacher et al. 2018).

Our analyses allude to possible uses of the WRA beyond the novel use we describe here as well as its original goal of identifying potential weeds. High scoring species that have failed to naturalize are priorities for monitoring and prevention (particularly multi-island introductions), and could be compared with similarly-scored species that have naturalized, possibly identifying important traits or conditions inhibiting species otherwise prone to invade. Conversely, investigating low scoring species that have naturalized could reveal possible sources of error during prediction, suggesting location-specific contexts that promote invasion (e.g. remote island ecosystems with low native diversity). While we show that WRAs can help assign naturalized statuses for data deficient species, future studies could investigate avenues for inferring other positions along the invasion continuum (Appendix). For instance, identifying extirpations is challenging because data necessary to establish absence is often insufficient and hence could be made easier with inferential tools. Guidance on declaring extinctions and extirpations for native species is available (IUCN 2017), but a framework that applies these principles to non-native eradications and natural extirpations is sorely needed.

Conclusions

Our attempt to apply the framework by Blackburn et al. (2011) to Hawai‘i revealed its limitations as a tracking system for entire non-native floras. Specifically, the framework does not address species that have uncertain statuses, which are common surrounding the early stages of naturalization. However, our findings indicate that this issue can be resolved for many species by predicting a likely status using WRA scores. We therefore demonstrate that species tracking systems and predictive tools like WRA should be integrated to strengthen their ability to inform management of already-introduced plant species. We further propose a system for tracking entire floras that accommodates real-world data while retaining categories relevant to invasive plant managers. Our system is particularly useful for invader-rich floras, such as those common on islands, where a mix of on-the-ground information and methods to address data gaps is necessary to solve real-world biosecurity dilemmas.

References

- Ansong M, Pergl J, Essl F, Hejda M, van Kleunen M, Randall R, Pyšek P (2019) Naturalized and invasive alien flora of Ghana. *Biological Invasions* 21: 669–683. <https://doi.org/10.1007/s10530-018-1860-7>
- Bacher S, Blackburn TM, Essl F, Genovesi P, Heikkilä J, Jeschke JM, Jones G, Keller R, Kenis M, Kueffer C, Martinou AF, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Roy HE, Saul W-C, Scalera R, Vilà M, Wilson JR, Kumschick S (2018) Socio-economic impact classification of alien taxa (SEICAT). *Methods in Ecology and Evolution* 9: 159–168. <https://doi.org/10.1111/2041-210X.12844>
- Blackburn TM, Cassey P, Duncan RP (2020) Colonization pressure: a second null model for invasion biology. *Biological Invasions* 22: 1221–1233. <https://doi.org/10.1007/s10530->

- Blackburn TM, Essl F, Evans T, Hulme PE, Jeschke JM, Kuhn I, Kumschick S, Markova Z, Mrugala A, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Ricciardi A, Richardson DM, Sendek A, Vila M, Wilson JRU, Winter M, Genovesi P, Bacher S (2014) A unified classification of alien species based on the magnitude of their environmental impacts. *PLoS Biology* 12: 11. <https://doi.org/10.1371/journal.pbio.1001850>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Colautti RI, MacIsaac HJ (2004) A neutral terminology to define “invasive” species. *Diversity and Distributions* 10: 135–141. <https://doi.org/10.1111/j.1366-9516.2004.00061.x>
- Cooney R (2004) The Precautionary Principle in Biodiversity Conservation and Natural Resource Management: An Issues Paper for Policy-Makers, Researchers and Practitioners. In: IUCN Policy and Global Change Series No. 2. IUCN: Gland, Switzerland, Cambridge, UK, 51.
- Daehler CC, Denslow JS, Ansari S, Kuo H-C (2004) A risk-assessment system for screening out invasive pest plants from Hawai‘i and other pacific islands. *Conservation Biology* 18: 360–368. <https://doi.org/10.1111/j.1523-1739.2004.00066.x>
- Daehler CC, Virtue JG (2010) Likelihood and consequences: reframing the Australian weed risk assessment to reflect a standard model of risk. *Plant Protection Quarterly* 25: 52–55.
- Danihelka J, Chrtek Jr. J, Kaplan Z (2012) Checklist of vascular plants of the Czech Republic. *Preslia* 84: 647–811. <http://www.preslia.cz/P123Danihelka.pdf>
- Duncan RP, Cassey P, Pigot AL, Blackburn TM (2019) A general model for alien species richness. *Biological Invasions* 21: 2665–2677. <https://doi.org/10.1007/s10530-019->

- Galasso G, Conti F, Peruzzi L, Ardenghi NMG, Banfi E, Celesti-Grapow L, Albano A, Alessandrini A, Bacchetta G, Ballelli S, Mazzanti MB, Barberis G, Bernardo L, Blasi C, Bouvet D, Bovio M, Cecchi L, Guacchio ED, Domina G, Fascetti S, Gallo L, Gubellini L, Guiggi A, Iamónico D, Iberite M, Jiménez-Mejías P, Lattanzi E, Marchetti D, Martinetto E, Masin RR, Medagli P, Passalacqua NG, Peccenini S, Pennesi R, Pierini B, Podda L, Poldini L, Prosser F, Raimondo FM, Roma-Marzio F, Rosati L, Santangelo A, Scoppola A, Scortegagna S, Selvaggi A, Selvi F, Soldano A, Stinca A, Wagensommer RP, Wilhalm T, Bartolucci F (2018) An updated checklist of the vascular flora alien to Italy. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 152: 556–592. <https://doi.org/10.1080/11263504.2018.1441197>
- Gbedomon RC, Salako VK, Schlaepfer MA (2020) Diverse views among scientists on non-native species. *NeoBiota* 54: 49–69. <https://doi.org/10.3897/neobiota.54.38741>
- Gilbert B, Levine JM (2013) Plant invasions and extinction debts. *Proceedings of the National Academy of Sciences* 110: 1744–1749. <https://doi.org/10.1073/pnas.1212375110>
- Groom Q, Desmet P, Reyserhove L, Adriaens T, Oldoni D, Vanderhoeven S, Baskauf SJ, Chapman A, McGeoch M, Walls R, Wiczorek J, Wilson J, Zermoglio PF, Simpson A (2019) Improving Darwin Core for research and management of alien species. *Biodiversity Information Science and Standards* 3: e38084. <https://doi.org/10.3897/biss.3.38084>
- HPWRA (2019). Hawai‘i-Pacific Weed Risk Assessment. <https://www.hpwra.org>. Accessed October 2019.
- Hawkins CL, Bacher S, Essl F, Hulme PE, Jeschke JM, Kühn I, Kumschick S, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Vilà M, Wilson JRU, Genovesi P, Blackburn TM (2015) Framework and guidelines for implementing the proposed IUCN Environmental Impact Classification for Alien Taxa (EICAT). *Diversity and Distributions* 21: 1360–1363. <https://doi.org/10.1111/ddi.12379>

- Henderson L, Wilson JRU (2017) Changes in the composition and distribution of alien plants in South Africa: An update from the Southern African Plant Invaders Atlas. *Bothalia* 47: 26. <https://doi.org/10.4102/abc.v47i2.2172>
- Hulme PE (2006) Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology* 43: 835–847. <https://doi.org/10.1111/j.1365-2664.2006.01227.x>
- Imada CT (2012) Hawaiian native and naturalized vascular plant checklist. Bishop Museum Technical Report 60: 29 pp. + 27 appendices.
- Imada CT (2019) Hawaiian naturalized vascular plant checklist. Bishop Museum Technical Report 69: 23 pp.
- Imada CT, Staples GW, Herbst DR (2000) Annotated checklist of cultivated plants of Hawai‘i, Bernice Pauahi Bishop Museum. Retrieved from <http://www2.bishopmuseum.org/HBS/botany/cultivatedplants>. Accessed November 2018.
- IUCN (2020) The IUCN Red List of Threatened Species. Version 2020-1. <https://www.iucnredlist.org>. Accessed June 2020.
- Jacobs L, Richardson D, Lepschi B, Wilson J (2017) Quantifying errors and omissions in alien species lists: The introduction status of *Melaleuca* species in South Africa as a case study. *NeoBiota* 32: 89–105. <https://doi.org/10.3897/neobiota.32.9842>
- Kier G, Kreft H, Lee TM, Jetz W, Ibsch PL, Nowicki C, Mutke J, Barthlott W (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 9322–9327. <https://doi.org/10.1073/pnas.0810306106>
- Komsta L (2019) mblm: Median-Based Linear Models. R package version 0.12.1. <https://CRAN.R-project.org/package=mblm>

- Kraus F, Duffy DC (2010) A successful model from Hawai'i for rapid response to invasive species. *Journal for Nature Conservation* 18: 135–141.
<https://doi.org/10.1016/j.jnc.2009.07.001>
- Kueffer C, Daehler CC, Torres-Santana CW, Lavergne C, Meyer JY, Otto R, Silva L (2010) A global comparison of plant invasions on oceanic islands. *Perspectives in Plant Ecology Evolution and Systematics* 12: 145–161. <https://doi.org/10.1016/j.ppees.2009.06.002>
- Kueffer C, Loope L (2009) Prevention, early detection and containment of invasive, nonnative plants in the Hawaiian Islands: current efforts and needs. Honolulu (HI): Pacific Cooperative Studies Unit, University of Hawai'i at Manoa, Department of Botany. PCSU Technical Report, 166. <http://hdl.handle.net/10125/26984>
- Kuussaari M, Bommarco R, Heikkinen RK, Helm A, Krauss J, Lindborg R, Öckinger E, Pärtel M, Pino J, Rodà F, Stefanescu C, Teder T, Zobel M, Steffan-Dewenter I (2009) Extinction debt: a challenge for biodiversity conservation. *Trends in Ecology & Evolution* 24: 564–571. <https://doi.org/10.1016/j.tree.2009.04.011>
- Larrue S, Butaud JF, Daehler CC, Ballet S, Chadeyron J, Oyono R (2018) Persistence at the final stage of volcanic island ontogeny: Abiotic predictors explain native plant species richness on 111 remote Pacific atolls. *Ecology and Evolution* 8: 12208–12220.
<https://doi.org/10.1002/ece3.4680>
- Lonsdale WM (1999) Global Patterns of Plant Invasions and the Concept of Invasibility. *Ecology* 80: 1522–1536. [https://doi.org/10.1890/0012-9658\(1999\)080\[1522:GPOPIA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1999)080[1522:GPOPIA]2.0.CO;2)
- Magona N, Richardson DM, Roux JLL, Kritzing-Klopper S, Wilson JRU (2018) Even well-studied groups of alien species might be poorly inventoried: Australian *Acacia* species in South Africa as a case study. *NeoBiota* 39: 1–29.
<https://doi.org/10.3897/neobiota.39.23135>
- McGeoch MA, Latombe G (2016) Characterizing common and range expanding species. *Journal*

of Biogeography 43: 217–228. <https://doi.org/10.1111/jbi.12642>

Nelson RE, Schubert TH (1976) Adaptability of selected tree species planted in Hawai‘i forests. U.S. Department of Agriculture Forest Service Resource Bulletin, Berkeley, California, 22 pp.

Palmer DD (2003) Hawai‘i’s ferns and fern allies. Honolulu, University of Hawai‘i Press. ISBN: 978-0824833473

Panetta FD (2015) Weed eradication feasibility: lessons of the 21st century. *Weed Research* 55: 226–238. <https://doi.org/10.1111/wre.12136>

Price JP, Wagner WL (2018) Origins of the Hawaiian flora: Phylogenies and biogeography reveal patterns of long-distance dispersal. *Journal of Systematics and Evolution* 56: 600–620. <https://doi.org/10.1111/jse.12465>

Pyšek P, Sádlo J, Mandák B (2002) Catalogue of alien plants of the Czech Republic. *Preslia* 74: 97–186.

Pyšek P, Richardson DM, Rejmanek M, Webster GL, Williamson M, Kirschner J (2004) Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon* 53: 131–143. <https://doi.org/10.2307/4135498>

Pyšek P, Pergl J, Essl F, Lenzner B, Dawson W, Kreft H, Weigelt P, Winter M, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cardenas D, Cardenas-Toro J, Castano N, Chacon E, Chatelain C, Dullinger S, Ebel AL, Figueiredo E, Fuentes N, Genovesi P, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Maurel N, Meerman J, Morozova O, Moser D, Nickrent D, Nowak PM, Pagad S, Patzelt A, Pelter PB, Seebens H, Shu WS, Thomas J, Velayos M, Weber E, Wieringa JJ, Baptiste MP, van Kleunen M (2017) Naturalized alien flora of the world: species diversity, taxonomic and phylogenetic patterns, geographic distribution and global hotspots of plant invasion. *Preslia* 89: 203–274. <https://doi.org/10.23855/preslia.2017.203>

- Ranker TA (2016) What do we know about Hawaiian ferns and lycophytes? *Journal of Systematics and Evolution* 54: 626–637. <https://doi.org/10.1111/jse.12213>
- Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions* 6: 93–107. <https://doi.org/10.1046/j.1472-4642.2000.00083.x>
- Robinson TB, Alexander ME, Simon CA, Griffiths CL, Peters K, Sibanda S, Miza S, Groenewald B, Majiedt P, Sink KJ (2016) Lost in translation? Standardising the terminology used in marine invasion biology and updating South African alien species lists. *African Journal of Marine Science* 38: 129–140. <https://doi.org/10.2989/1814232X.2016.1163292>
- Rojas-Sandoval J, Acevedo-Rodriguez P (2015) Naturalization and invasion of alien plants in Puerto Rico and the Virgin Islands. *Biological Invasions* 17: 149–163. <https://doi.org/10.1007/s10530-014-0712-3>
- Russell JC, Meyer JY, Holmes ND, Pagad S (2017) Invasive alien species on islands: impacts, distribution, interactions and management. *Environmental Conservation* 44: 359–370. <https://doi.org/10.1017/s0376892917000297>
- Sax DF, Gaines SD (2008) Species invasions and extinction: The future of native biodiversity on islands. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11490–11497. <https://doi.org/10.1073/pnas.0802290105>
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* 63 (324): 1379–1389. <https://doi.org/10.2307/2285891>.
- Skolmen RG (1980) *Plantings on the forest reserves of Hawai'i 1910-1960*. Honolulu, USA
- Staples G, Herbst DR (2005) *A Tropical Garden Flora: Plants Cultivated in the Hawaiian Islands and Other Tropical Places*. Honolulu, Bishop Museum Press. 908 pp. ISBN: 1581780397
- Uludag A, Aksoy N, Yazlık A, Arslan ZF, Yazmış E, Uremis I, Cossu TA, Groom Q, Pergl J,

- Pyšek P, Brundu G (2017) Alien flora of Turkey: checklist, taxonomic composition and ecological attributes. *NeoBiota* 35: 61–85. <https://doi.org/10.3897/neobiota.35.12460>
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, Vanderplas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0 Contributors (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17: 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wagner WL, Herbst DR, Sohmer SH (1999) *Manual of the Flowering Plants of Hawai‘i*, revised ed. Bishop Museum Special Publication 97. Honolulu, University of Hawai‘i Press. 1918 pp. ISBN: 978-0824821661
- Waskom M, Botvinnik O, Ostblom J, Lukauskas S, Hobson P, Gelbart M, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J, de Ruiter J, Pye C, Hoyer S, Vanderplas J, Villalba S, Kunter G, Quintero E, Bachant P, Martin M, Meyer K, Swain C, Miles A, Brunner T, O’Kane D, Yarkoni T, Williams ML, Evans C (2020) mwaskom/seaborn: v0.10.1. Zenodo. <https://doi.org/10.5281/zenodo.3629446>
- Wilcox RR (2010) Theil-Sen estimator. In: *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy* (2nd edn). Berlin, Springer-Verlag, pp. 207–210. ISBN: 978-1441955241.
- Wilson JRU, Faulkner KT, Rahlao SJ, Richardson DM, Zengeya TA, Wilgen BW van (2018) Indicators for monitoring biological invasions at a national level. *Journal of Applied Ecology* 55: 2612–2620. <https://doi.org/10.1111/1365-2664.13251>
- Wilson JRU, Caplat P, Dickie IA, Hui C, Maxwell BD, Nuñez MA, Pauchard A, Rejmánek M, Richardson DM, Robertson MP, Spear D, Webber BL, van Wilgen BW, Zenni RD (2014) A standardized set of metrics to assess and monitor tree invasions. *Biological Invasions* 16: 535–551. <https://doi.org/10.1007/s10530-013-0605-x>

Tables

Table 4.1 Species in the “Questionably Naturalized” checklist category alongside their likely status category derived from WRA scores and time since last observation.

Family	Species Name	WRA score	WRA rating	Likely Status*
Fabaceae	<i>Lespedeza cuneata</i> (Dum. Cours.) G.Don	17	High Risk	Extirpated
Apiaceae	<i>Eryngium foetidum</i> L.	15	High Risk	Naturalized–Unspecified
Asteraceae	<i>Tithonia rotundifolia</i> (Mill.) S.F.Blake	15	High Risk	Naturalized–Unspecified
Menispermaceae	<i>Stephania japonica</i> (Thunb.) Miers	13	High Risk	Naturalized–Unspecified
Tamaricaceae	<i>Tamarix aphylla</i> (L.) H.Karst.	13	High Risk	Naturalized–Unspecified
Cannabaceae	<i>Cannabis sativa subsp. indica</i> (Lam.) E.Small & Conquist	12	High Risk	Naturalized–Unspecified
Bromeliaceae	<i>Tillandsia usneoides</i> (L.) L.	12	High Risk	Naturalized–Unspecified
Asparagaceae	<i>Asparagus falcatus</i> L.	11	High Risk	Naturalized–Unspecified
Poaceae	<i>Lamarckia aurea</i> (L.) Moench	11	High Risk	Naturalized–Unspecified
Poaceae	<i>Panicum virgatum</i> L.	11	High Risk	Extirpated
Acanthaceae	<i>Barleria lupulina</i> Lindl.	10	High Risk	Naturalized–Unspecified
Begoniaceae	<i>Begonia nelumbiifolia</i> Schlttdl & Cham.	10	High Risk	Naturalized–Unspecified
Combretaceae	<i>Quisqualis indica</i> L.	10	High Risk	Naturalized–Unspecified
Urticaceae	<i>Laportea aestuans</i> (L.) Chew	10	High Risk	Naturalized–Unspecified
Sapindaceae	<i>Allophylus cobbe</i> (L.) Raeusch.	9	High Risk	Naturalized–Unspecified

Iridaceae	<i>Sisyrinchium rosulatum</i> E.P.Bicknell	9	High Risk	Extirpated
Orchidaceae	<i>Vanilla planifolia</i> Jacks.	9	High Risk	Naturalized– Unspecified
Rhamnaceae	<i>Ziziphus mauritiana</i> Lam.	9	High Risk	Naturalized– Unspecified
Fabaceae	<i>Acacia retinodes</i> Schldl.	8	High Risk	Naturalized– Unspecified
Apocynaceae	<i>Allamanda schottii</i> Pohl	8	High Risk	Naturalized– Unspecified
Primulaceae	<i>Ardisia virens</i> Kurz	8	High Risk	Naturalized– Unspecified
Poaceae	<i>Echinochloa esculenta</i> (A.Braun) H.Scholz	8	High Risk	Extirpated
Vitaceae	<i>Tetrastigma voinieranum</i> (Baltet) Pierre ex Gagnep.	8	High Risk	Naturalized– Unspecified
Fabaceae	<i>Acacia robusta</i> Burch. <i>subsp. clavigera</i> (E.Mey.) Brenan	7	High Risk	Naturalized– Unspecified
Bromeliaceae	<i>Aechmea bracteata</i> (Sw.) Griseb.	7	High Risk	Naturalized– Unspecified
Primulaceae	<i>Ardisia sieboldii</i> Miq.	7	High Risk	Naturalized– Unspecified
Scrophulariaceae	<i>Buddleja paniculata</i> Wall.	7	High Risk	Naturalized– Unspecified
Poaceae	<i>Cenchrus elegans</i> (Hassk.) Veldk.	7	High Risk	Naturalized– Unspecified
Poaceae	<i>Melinis nerviglumis</i> (Franch.) Zizka	7	High Risk	Naturalized– Unspecified
Marcgraviaceae	<i>Norantea guianensis</i> (Aubl.)	7	High Risk	Naturalized– Unspecified
Pinaceae	<i>Pinus pinaster</i> Aiton	7	High Risk	Naturalized– Unspecified
Apocynaceae	<i>Acokanthera schimperi</i> (A.DC.) Schweinf.	6	Evaluate	Data Deficient
Polygonaceae	<i>Homalocladium platycladum</i> (F.Muell.) L.H.Bailey	6	Evaluate	Data Deficient

Fabaceae	<i>Platymiscium stipulare</i> Benth.	6	Evaluate	Data Deficient
Lamiaceae	<i>Clerodendrum myricoides</i> (Hochst.) Vatke	6	High Risk	Data Deficient
Bignoniaceae	<i>Markhamia lutea</i> (Benth.) K.Schum.	5	High Risk	Data Deficient
Plantaginaceae	<i>Maurandya antirrhiniflora</i> Humb. & Bonpl. ex Willd.	5	Evaluate	Extirpated
Aizoaceae	<i>Mesembryanthemum cordifolium</i> L.f.	5	Low Risk	Data Deficient
Fabaceae	<i>Parkia timoriana</i> (DC.) Merr.	5	Evaluate	Data Deficient
Asteraceae	<i>Coreopsis tinctoria</i> Nutt.	4.5	High Risk	Extirpated
Euphorbiaceae	<i>Synadenium grantii</i> Hook.f.	4	Low Risk	Data Deficient
Cactaceae	<i>Peniocereus hirschtianus</i> (K.Schum.) D.R.Hunt	4	Evaluate	Data Deficient
Bignoniaceae	<i>Radermachera sinica</i> (Hance) Hemsl.	4	Evaluate	Data Deficient
Moraceae	<i>Antiaris toxicaria</i> Lesch.	3	Evaluate	Data Deficient
Annonaceae	<i>Cananga odorata</i> (Lam.) Hook.f & Thoms	3	Low Risk	Data Deficient
Euphorbiaceae	<i>Euphorbia albomarginata</i> Torr. & A.Gray	3	Low Risk	Extirpated
Pinaceae	<i>Pinus jeffreyi</i> A.Murray bis	3	Evaluate	Data Deficient
Solanaceae	<i>Solandra maxima</i> (Sessé & Moç.) P.S.Green	3	Evaluate	Data Deficient
Moraceae	<i>Ficus pumila</i> L.	2	Low Risk	Data Deficient
Fabaceae	<i>Sesbania grandiflora</i> L. Pers.	2	Low Risk	Data Deficient
Myrtaceae	<i>Eucalyptus pulchella</i> Desf.	1	Evaluate	Data Deficient
Plantaginaceae	<i>Linaria purpurea</i> (L.) Mill.	1	High Risk	Extirpated
Anacardiaceae	<i>Anacardium occidentale</i> L.	0	Low Risk	Not Self-Sustaining
Araliaceae	<i>Plerandra elegantissima</i> (Veitch ex Mast.) Lowry, G.M.Plunkett & Frodin	0	Low Risk	Not Self-Sustaining
Podocarpaceae	<i>Afrocarpus mannii</i> (Hook.f.) C.N.Page	-1	Low Risk	Not Self-Sustaining

Marantaceae	<i>Calathea zebrina</i> (Hort. ex Bosse)	-1	Low Risk	Not Self-Sustaining
Boraginaceae	<i>Cordia sebestena</i> L.	-1	Low Risk	Not Self-Sustaining
Fabaceae	<i>Delonix regia</i> (Bojer ex Hook.) Raf.)	-1	Low Risk	Not Self-Sustaining
Ebenaceae	<i>Diospyros blancoi</i> A.DC.	-1	Low Risk	Not Self-Sustaining
Myrtaceae	<i>Melaleuca styphelioides</i> (Sol. ex Gaertn.) Sm.	-2	Low Risk	Not Self-Sustaining
Apocynaceae	<i>Beaumontia multiflora</i> Teijsm. & Binn.	-4	Low Risk	Not Self-Sustaining
Magnoliaceae	<i>Magnolia champaca</i> (L.) Baill. ex Pierre	-5	Low Risk	Not Self-Sustaining
Malvaceae	<i>Pachira aquatica</i> Aubl.	-6	Low Risk	Not Self-Sustaining

*Species that have not been observed outside of cultivation for over 50 years are assigned a predicted status of “Extirpated”, and in all cases, they are known from less than 3 herbarium vouchers. Species having WRA scores > 6 and < 1 were assigned “Naturalized–Unspecified” and “Not Self-Sustaining”, respectively, while species with scores from 1–6 were assigned “Data Deficient” (Table 2).

Table 4.2 Practical adaptation of the framework requiring minimal additional data collection beyond species checklists.

Categories as per Blackburn et al. (2011), with labels in brackets as per Groom et al. (2019)	Proposed Status Tracking Categories		Category Description
A–Not transported beyond limits of native range.	Not Present		No history of introduction or previously introduced (e.g. for cultivation or forestry), but never wild-growing and no longer present.
Not Included	No Longer Present *	Establishment Failure	Previously wild growing at one time but was purposefully or naturally removed before self-sustaining populations formed (naturalization).
		Extirpated	Previously forming self-sustaining populations (naturalized) at one time, but no longer existing through purposeful (eradication) or natural means.
B1 (captive) –Individuals transported beyond limits of native range, and in captivity or quarantine (i.e. individuals provided with conditions suitable for them, but explicit measures of containment are in place)	Contained / Cultivated		Existing in cultivation or somehow contained. Includes accidental soil contaminants that are contained within pots or aquaria alongside purposefully cultivated species.
B2 (cultivated) –Individuals transported beyond limits of native range, and in cultivation (i.e. individuals provided with conditions suitable for them but explicit measures to prevent dispersal are limited at best)			
B3 (released) –Individuals transported beyond limits of native range, and directly released into novel environment	Not Included		N/A
C0 (failing) –Individuals released into the wild (i.e. outside of captivity or cultivation) in location where introduced, but incapable of surviving for a significant period	Reappropriated into "Establishment Failure" Above		

C1 (casual) —Individuals surviving in the wild (i.e. outside of captivity or cultivation) in location where introduced, no reproduction	Not Self-Sustaining		Plants surviving in the wild with sufficient evidence suggesting that offspring, if produced, do not contribute to a self-sustaining population.
C2 (reproducing) —Individuals surviving in the wild in location where introduced, reproduction occurring, but population not self-sustaining			
Not Included	Potentially Naturalizing		Plants apparently surviving and reproducing in the wild, but insufficient time has passed to determine if a self-replacing population exists.
C3 (established) —Individuals surviving in the wild in location where introduced, reproduction occurring, and population self-sustaining	Naturalized - Unspecified*	Naturalized Where Introduced	Plants that form self-sustaining populations without human intervention (e.g. cultivation) but have not dispersed a significant distance from their point of introduction.
D1 (colonising) —Self-sustaining population in the wild, with individuals surviving a significant distance from the original point of introduction			
D2 (invasive) —Self-sustaining population in the wild, with individuals surviving and reproducing a significant distance from the original point of introduction		Naturalized Beyond Introduction Site	Plants that form self-sustaining populations without human intervention (e.g. cultivation) and have dispersed and established a significant distance from their point of introduction.
E (widespread Invasive) —Fully invasive species, with individuals dispersing, surviving and reproducing at multiple sites across a greater or lesser spectrum of habitats and extent of occurrence			

*Use if data to assign subcategories is insufficient

Figures

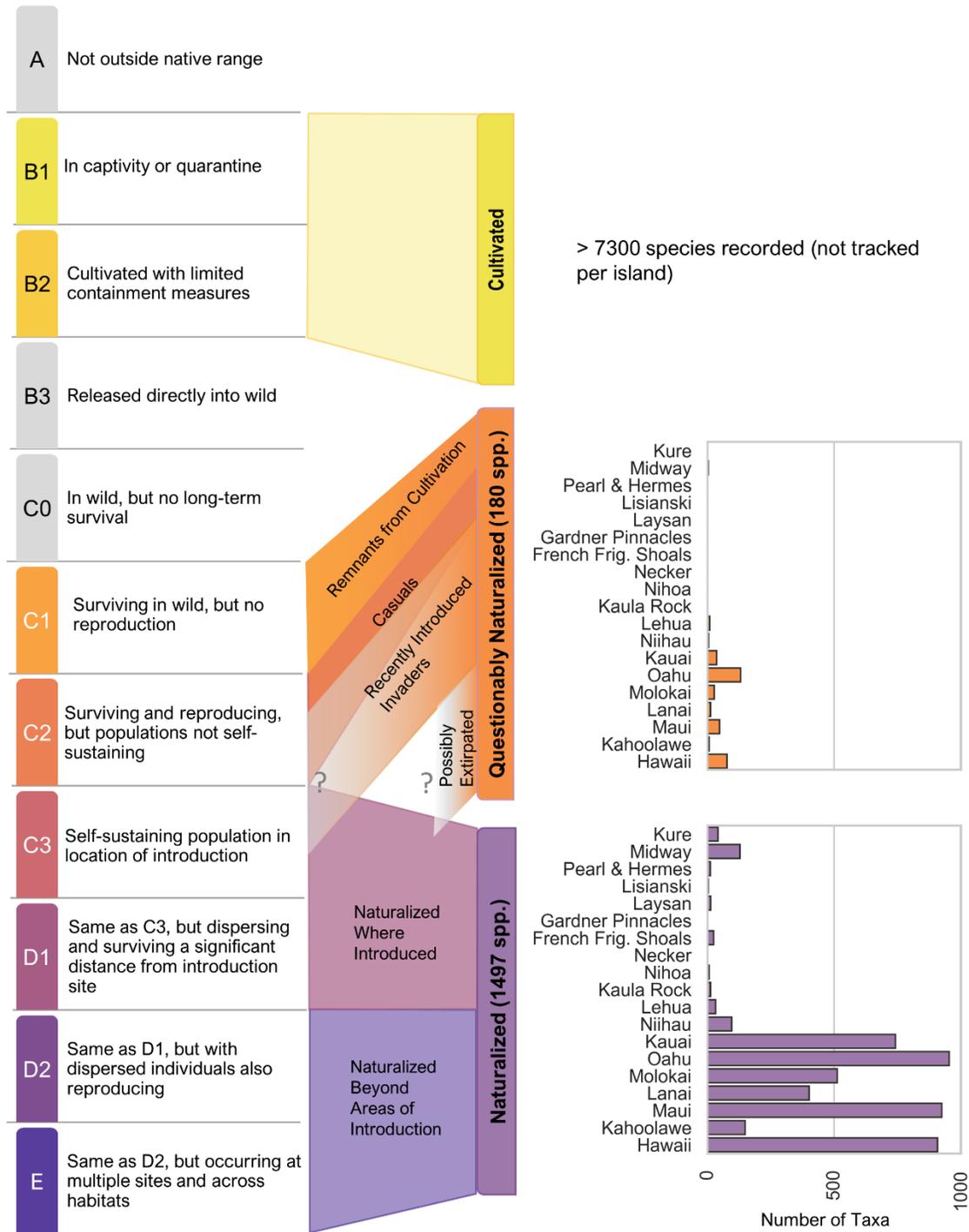


Figure 4.1 Alignment of Blackburn et al.'s (2011) scheme with categories in Hawai'i's non-native species checklist. Bar graphs depict counts of species that are considered "Naturalized" (purple) and "Questionably Naturalized" (orange) in Hawai'i per island, with statewide totals next to checklist status names. Colored polygons represent alignment or misalignment with the recommended categories. Full descriptions of categories are available in Table 4.2.

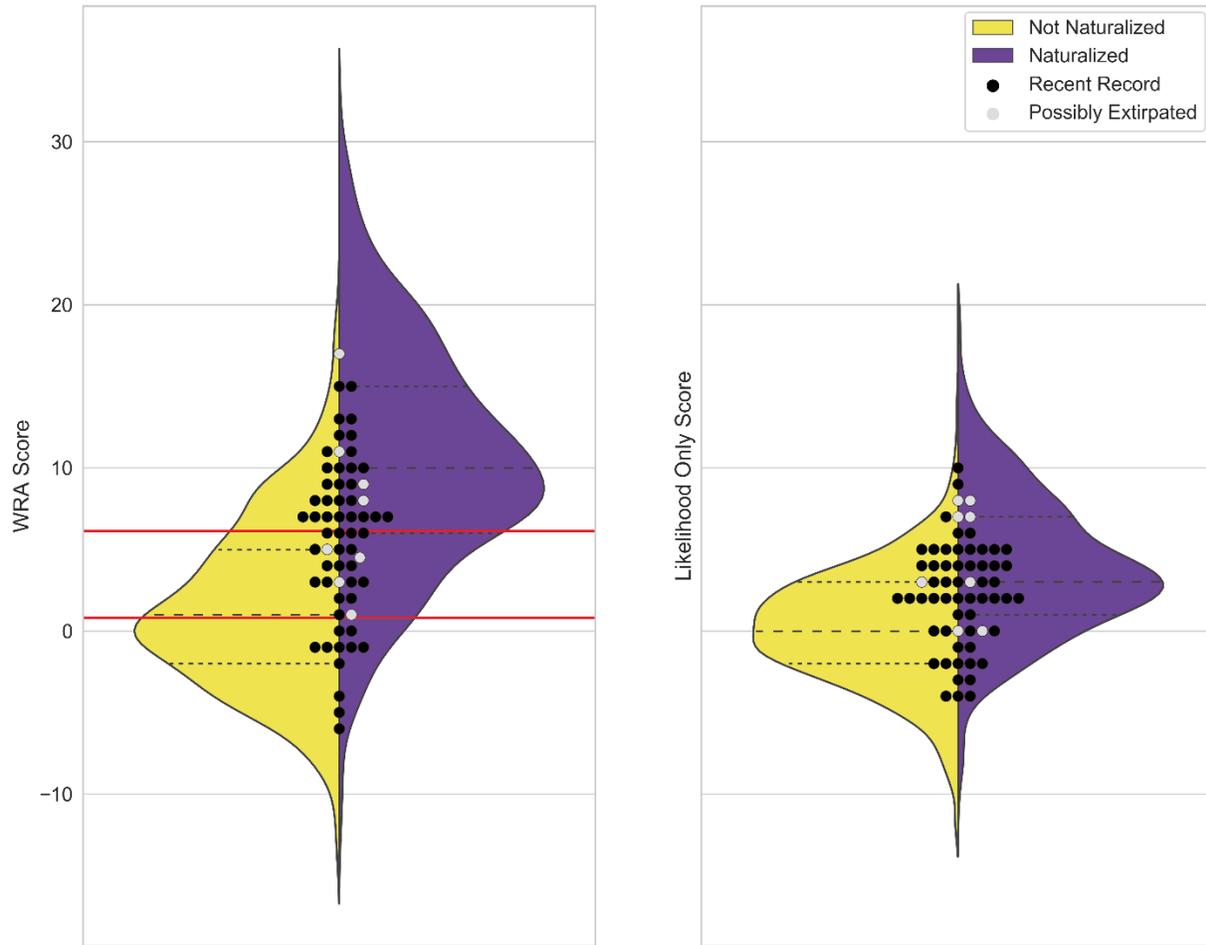


Figure 4.2 Distribution of non-naturalized (yellow) and naturalized (purple) scores based on WRA (left) and likelihood-only scoring (right). Dotted lines represent the lower quartile, median, and upper quartile while thick solid red lines represent the scoring threshold used by the Hawai‘i–Pacific WRA for designating species as high risk (> 6) and low risk (< 1). Dots represent scores for “Questionably Naturalized” species; grey = possibly extirpated, black = recently observed.

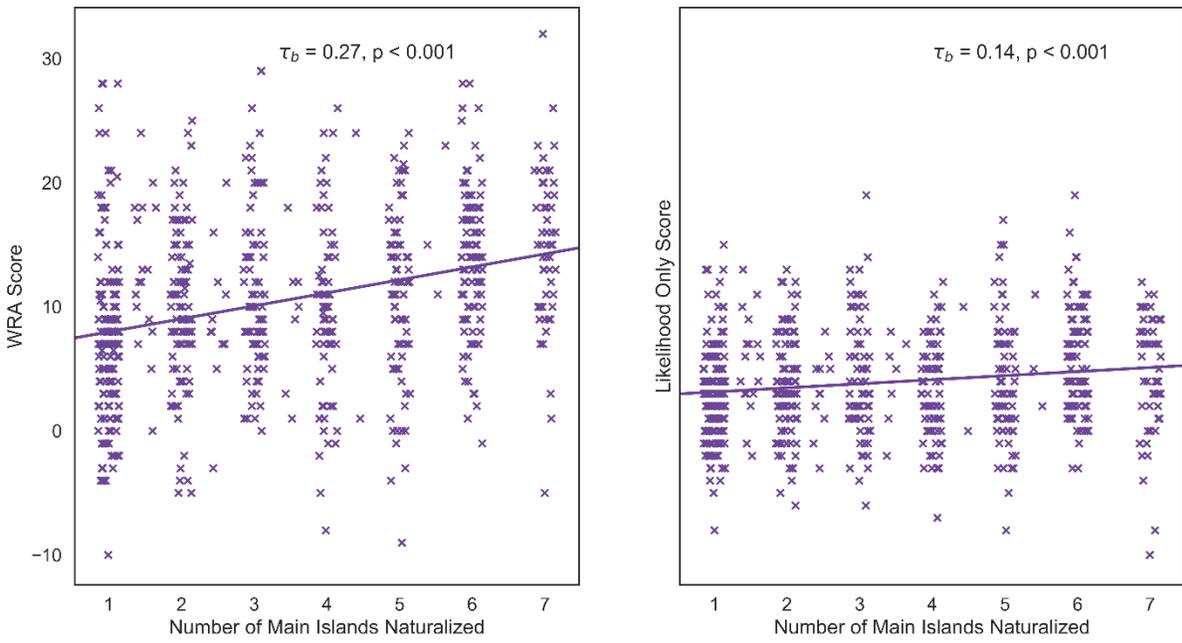


Figure 4.3 Relationship between the number of Hawaiian Islands naturalized with WRA (left) and likelihood-only scores (right). The fitted Theil-Sen estimator line is shown with Kendall's τ_b and associated p -value.

Chapter 5 – Conclusion

The aim of this work was to help prevent invasions by fundamentally recognizing them as a biodiversity problem requiring a global perspective despite implementation of mitigation strategies at smaller scales. By analyzing the diversity and temporal aspects of a large non-native flora originating from across the globe, I was able to uncover patterns that may be useful for the implementation of prevention strategies, both in Hawai‘i and other similar areas. The detection of these patterns emphasized the value of biodiversity data for developing indicators relevant to invasive species policy, and thus, I also focused on the practical application of these data to measure and track non-native biodiversity change. It is my hope that applying the methodologies outlined in this dissertation not only increases the accuracy of policy-relevant indicators, but also helps others who seek to understand invasions in the context of global biodiversity change.

In the second chapter, I sought to characterize the non-native flora of Hawai‘i, describing its diversity and origins to understand which kinds of species have naturalized and how they arrived. I also examined how the human-related factors relating to their arrival have changed, and whether those factors affected the rate at which certain plants have naturalized. The data showed that horticulture and landscaping is a major introduction pathway for naturalizing plants in current-day Hawai‘i, but that the relative importance of these introduction pathways can vary significantly in the span of a single human lifetime. This suggests that prevention strategies will need to adapt over time to remain effective.

The identification of the horticultural industry as the most important introduction pathway for new plant invaders poses a great challenge for biosecurity planning due to the fact that all introductions are purposeful. While strategies that target accidental introductions can use methods that do not discriminate between species (e.g., UV sterilization in the ballast tanks of

transoceanic cargo ships), the prevention of potentially invasive species that are introduced on purpose, must rely on numerous species-specific risk assessments (Callow and Callow 2011, Braga et al. 2020). Rather than creating a long “blacklist” of species that are not allowed to be imported, one avenue to reduce invasion risk and the number of individual risk assessments required is to apply a “whitelist” approach similar to that implemented in New Zealand, which instead curates a list of low risk plants (Hulme 2020). However, major legislative hurdles exist in developing a whitelist strategy for Hawai‘i. The Plant Protection Act, signed into federal law in 2000, gives the Administrator of the Animal and Plant Health Inspection Service (APHIS) the authority to regulate interstate movement of any plant using a blacklist approach, and forbids individual states from imposing regulations that supersede this authority (Appleby 2005). Thus, any changes to plant prevention policies must be executed at the federal level.

However, opportunities still exist to prevent ornamental plant invasions in Hawai‘i, including formal efforts to regulate inter-island movement (i.e. a rapid petition process for noxious weed listing) and voluntary measures that can be implemented by nursery and landscape business owners. Invasive species committees, which exist on all of the main islands, have close relationships with the community and have succeeded in managing invasions via voluntary agreements (Kraus and Duffy 2010). Hawai‘i and Kauai islands’ invasive species programs have developed a program called “Plant Pono” to endorse nursery and landscaping businesses that only sell native plants or non-invasive species that have received a low risk rating by the Hawai‘i-Pacific Weed Risk Assessment (Daehler et al. 2004). Currently, business owners benefit from endorsement alone, which may attract customers that are well-educated on invasive species issues. However, the state government could aid in the promotion of this program by allowing

some form of tax credit for endorsed businesses, which could provide further incentive for nurseries to import and sell only native or non-invasive plants.

In addition to providing guidance for policy-makers, the temporal trends uncovered for naturalizing plants in Hawai'i contribute to a developing theory in the invasion biology literature about the rate at which we should expect non-native species to accumulate (Turbelin et al. 2017). Earlier aspects of this theory focused on the human element of transporting species, hypothesizing that the rate of naturalizations will increase as globalized economies continue to expand or possibly decrease in areas that have witnessed increased awareness of impacts and improved invasive species policies (Meyerson and Mooney 2007, Essl et al. 2011). Recent work builds on the theory around economic expansion by pointing out that the number of new non-native species that can be introduced is theoretically finite (as the number of species on the planet is also finite), and that the rate at which species will continue to accumulate is dependent on the number of species in a source pool and whether new source pools will continue to be accessed (Seebens et al. 2018). So far, source pools have been defined in a geographic context, such that economic expansion may allow access to a new source pool by doing business in a new geographic location for the first time, which provides a new pathway for species exchange between the two regions (Seebens et al. 2018). However, my study did not show an increase in species originating from certain geographic regions over time or from a greater diversity of regions. Instead, I found that variation in naturalization rates was strongly correlated with fluctuations in the likely reason for species introductions, which are influenced by dominant socio-economic activities. This finding suggests that for purposefully introduced species, a source pool may be conceptualized as the number of species in trade for a specific purpose (e.g. ornamental plants), such that the development of new economic revenue streams could access

new pools of species circulating in trade around the globe. If this hypothesis is generally supported, then access to new source pools would increase with economic diversification or shifts in the dominant economy, while economic expansion into new geographic regions may play a comparatively minor role.

The rapid climate matching assessment for the entire non-native flora completed in chapter 2 revealed that many non-native species are apparently suited to upper-elevation climates currently dominated by remnant native ecosystems. This finding yields potentially useful suggestions for protected area management, which is timely given recent calls to protect half of Earth by 2050 (Wilson 2016, Pimm et al. 2018). I found that species currently associated with warmer, low-elevation climates in Hawai‘i may be equally at home in cooler, higher elevation climates, and that managers should focus on actions that reduce opportunities for higher elevation establishment, such as minimizing disturbance and introductions around near high-value areas (Catford et al. 2012). Additionally, the high number of non-native species that apparently overlap in climate preference with endangered native species reveals Hawai‘i as an ideal system for understanding and predicting cumulative impacts from multiple invaders on endangered species.

Current ecosystem-level studies often focus on the impacts of just a few species with notorious invasion histories. This narrow focus is often out of necessity because impacts are difficult to predict and quantifying their effect on endangered ecosystems is an expensive and data intensive task (Blackburn et al. 2014, Dick et al. 2017). However, as invasive species become more numerous, management plans must account for cumulative impacts from multiple invaders while maximizing control success on small budgets. One important question that remains unanswered is whether many low–moderate impact invaders can cause the same damage

as a few high impact invaders. For instance, research into the near-extinction of the American chestnut (*Castanea dentata*) revealed that three pathogens were responsible, with each one infecting and killing the tree in a different part of its range (Holmes et al. 2009). Similarly, multiple non-native plants may be required to cause the extinction of an endangered plant, even if each non-native is not successful in all parts of the endangered species' native range (Daehler 2003). Insight into this problem has the potential to overhaul management strategies by focusing on the context dependency of potential invasions (i.e. what resources do we have to lose?) instead of prioritizing actions based on species-specific impact scores with less regard for the types of resources they might impact. Hawai'i hosts numerous naturalized and endangered plants with small ranges, which facilitates studies to quantify and compare ecological niche breadth and functional traits for native and non-native species. My work provides a rough categorization of climate preferences for all Hawai'i's naturalized plants, providing a starting point for selecting non-native species for future ecological niche studies based on ones that may prefer similar climates to native species.

In the third chapter I tackled a common issue that emerges when performing analyses on biodiversity datasets—the existence of bias arising from the collection and processing of data by multiple people (Hughes et al. 2021). Specifically, I illuminated the importance of time lags that accumulate during the processing of first record data for naturalized species and how they can interact with common data compilation tactics to produce misleading invasion trend lines. This was achieved using simulated data because little data exists on the lengths of these lags and how they might vary over time, among taxonomic groups, and between geographic regions (Zenetos et al. 2019). However, further analysis of a case study comparing data collection techniques by

two different researchers provided convincing real-world evidence that these factors can have a significant impact on how invasion trends are interpreted.

The order of chapters in this dissertation reflects my true discovery process regarding invasion trends: I analyzed trends of naturalizing plants in Hawai‘i before I explored a conceptual framework for biases that might underlie these trends, when ideally, these discoveries would take place in the opposite order. However, I show that using herbarium specimens and dates of actual naturalization (as opposed to dates that align more closely with introduction time) decreases the length of the false plateau period at the end of the trendline. Thus, by using these data sources to explore naturalization trends in Hawai‘i, I was using the best available data, although this realization became clearer in hindsight.

Under-discussed sources of sampling bias like those outlined in chapter three are indicative of growing pains associated with the expanding use of large biodiversity datasets (Hughes et al. 2021). Given the urgent need to understand trends at broad scales, many more analyses are likely to include data that are unfamiliar to the person performing the analysis, which may result in biased conclusions. More work is needed to standardize and annotate these data to increase their useability by invasive species researchers and policy-makers, with a specific focus on data that are integrated into indicators intended to be compared amongst regions and over time (Vicente et al. 2021).

One potential avenue to accelerate our understanding of best practices for measuring and predicting invasive species trends is to borrow lessons discussed in the literature for native species. For example, unlike the literature on invasions, a rich conversation exists on sampling biases that arise when examining trends in the discovery of new native species—a process that is similar in many ways to discovering first records of non-native species (Costello et al. 2013,

Mora et al. 2013, Moura et al. 2018). Additionally, the recommendations that arose from these discussions are similar to those I proposed, including investments in taxonomic resources and platforms that allow easy access to data (Costello et al. 2015). The parallels that exist for both native and non-native species further emphasizes the need to understand factors that underlie efficient (and inefficient) data processing, and that these factors may indicate a region's ability to process both native and non-native species data.

While chapter three sought to conceptualize how biases might be introduced during data processing and analysis, chapter four sought to address ambiguities originating from field observations. Together, these two chapters help address an outstanding need in invasive species research and management to understand the minimum infrastructure needed to monitor invasive species. Chapter four specifically endeavored to develop a system for tracking the status of non-native plants after their arrival. This was done by analyzing Hawai‘i’s naturalized species checklist through the lens of the most frequently cited conceptual framework in invasion biology, which categorizes a species’ advancement from introduction to invasion (Blackburn et al. 2011, Wilson et al. 2020). I discovered that hundreds of species were considered “questionably naturalized”, and that despite the common assumption that these are species just beginning to naturalize, this pool contained species on drastically different population trajectories. This unclear categorization poses a problem for management because species that are unlikely to form self-sustaining populations are lumped together with emerging invaders that pose a significant risk, which makes choosing eradication targets difficult. I attempted to solve this problem by analyzing data collected from the Hawai‘i-Pacific Weed Risk Assessment (Daehler et al. 2004), ultimately creating a novel method to address uncertainties in biodiversity monitoring by integrating predictive tools.

These findings exemplify the difficult (but not insurmountable) nature of translating field observations for hundreds of species into a format that is useful to researchers and managers. Nonetheless, a finer categorization of statuses along the introduction-invasion continuum is likely to have a large impact on our understanding of factors that influence impacts and large-scale patterns of non-native species richness. The vast majority of trends described in the field of invasion biology use datasets of species broadly defined as naturalized, with no information on which ones cause major impacts or species that have failed to invade (Diez et al. 2009). These data are needed to fine-tune predictive tools so that we can distinguish between species that are likely to do extensive damage from those that merely form self-sustaining populations.

Despite over 30 years of research to understand the reasons why species invade and how they inflict damage, the field of invasion biology is yet to uncover a unified theory to address these knowledge gaps (Heger et al. 2021). Although several patterns and commonalities exist among invasive species, there are often many exceptions, which prevents the simple application of generalizations to aid management. For example, Enders et al (2020) summarize no less than 39 well-recognized invasion theories, with no theory supported by 100% of studies. I propose that biodiversity informatics can fill these gaps. In addition to searching for unifying patterns, we should acknowledge that invasions are inherently complex and ungeneralizable, and that advances in data science have made it possible to track the outcome of non-native species introductions like never before. Taxonomic data, like species checklists, can be linked to ecological and geographical data to understand how species-specific traits interact with location-specific factors to threaten valued resources. In this manner, we can make data-driven management decisions and fine-tune the predictive tools that are necessary to prevent invasions.

References

- Appleby AP (2005) A history of weed control in the United States and Canada—a sequel. *Weed Science* 53: 762–768. <https://doi.org/10.1614/WS-04-210.1>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Blackburn TM, Essl F, Evans T, Hulme PE, Jeschke JM, Kuhn I, Kumschick S, Markova Z, Mrugala A, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Ricciardi A, Richardson DM, Sendek A, Vila M, Wilson JRU, Winter M, Genovesi P, Bacher S (2014) A Unified Classification of Alien Species Based on the Magnitude of their Environmental Impacts. *Plos Biology* 12: 11. <https://doi.org/10.1371/journal.pbio.1001850>
- Braga C, Hunsucker K, Gardner H, Swain G (2020) A novel design to investigate the impacts of UV exposure on marine biofouling. *Applied Ocean Research* 101: 102226. <https://doi.org/10.1016/j.apor.2020.102226>
- Callow JA, Callow ME (2011) Trends in the development of environmentally friendly fouling-resistant marine coatings. *Nature Communications* 2: 244. <https://doi.org/10.1038/ncomms1251>
- Catford JA, Daehler CC, Murphy HT, Sheppard AW, Hardesty BD, Westcott DA, Rejmanek M, Bellingham PJ, Pergl J, Horvitz CC, Hulme PE (2012) The intermediate disturbance hypothesis and plant invasions: Implications for species richness and management. *Perspectives in Plant Ecology Evolution and Systematics* 14: 231–241. <https://doi.org/10.1016/j.ppees.2011.12.002>
- Costello MJ, May RM, Stork NE (2013) Can We Name Earth's Species Before They Go Extinct? *Science* 339: 413–416. <https://doi.org/10.1126/science.1230318>
- Costello MJ, Vanhoorne B, Appeltans W (2015) Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology* 29: 1094–1099. <https://doi.org/10.1111/cobi.12496>
- Daehler CC (2003) Performance Comparisons of Co-Occurring Native and Alien Invasive

Plants: Implications for Conservation and Restoration. *Annual Review of Ecology, Evolution, and Systematics* 34: 183–211.

<https://doi.org/10.1146/annurev.ecolsys.34.011802.132403>

Daehler CC, Denslow JS, Ansari S, Kuo H-C (2004) A Risk-Assessment System for Screening Out Invasive Pest Plants from Hawai'i and Other Pacific Islands. *Conservation Biology* 18: 360–368. <https://doi.org/10.1111/j.1523-1739.2004.00066.x>

Dick JTA, Laverty C, Lennon JJ, Barrios-O'Neill D, Mensink PJ, Britton JR, Médoc V, Boets P, Alexander ME, Taylor NG, Dunn AM, Hatcher MJ, Rosewarne PJ, Crookes S, MacIsaac HJ, Xu M, Ricciardi A, Wasserman RJ, Ellender BR, Weyl OLF, Lucy FE, Banks PB, Dodd JA, MacNeil C, Penk MR, Aldridge DC, Caffrey JM (2017) Invader Relative Impact Potential: a new metric to understand and predict the ecological impacts of existing, emerging and future invasive alien species. *Methods in Ecology and Evolution*: 1259–1267. [https://doi.org/10.1111/1365-2664.12849@10.1111/\(ISSN\)2041-210X.INSP_2017](https://doi.org/10.1111/1365-2664.12849@10.1111/(ISSN)2041-210X.INSP_2017)

Diez JM, Williams PA, Randall RP, Sullivan JJ, Hulme PE, Duncan RP (2009) Learning from failures: testing broad taxonomic hypotheses about plant naturalization. *Ecology Letters* 12: 1174–1183. <https://doi.org/10.1111/j.1461-0248.2009.01376.x>

Enders M, Havemann F, Ruland F, Bernard-Verdier M, Catford JA, Gómez-Aparicio L, Haider S, Heger T, Kueffer C, Kühn I, Meyerson LA, Musseau C, Novoa A, Ricciardi A, Sagouis A, Schittko C, Strayer DL, Vilà M, Essl F, Hulme PE, Kleunen M van, Kumschick S, Lockwood JL, Mabey AL, McGeoch MA, Palma E, Pyšek P, Saul W-C, Yannelli FA, Jeschke JM (2020) A conceptual map of invasion biology: Integrating hypotheses into a consensus network. *Global Ecology and Biogeography* 29: 978–991. <https://doi.org/10.1111/geb.13082>

Essl F, Dullinger S, Rabitsch W, Hulme PE, Hülber K, Jarošík V, Kleinbauer I, Krausmann F, Kühn I, Nentwig W, Vilà M, Genovesi P, Gherardi F, Desprez-Loustau M-L, Roques A, Pyšek P (2011) Socioeconomic legacy yields an invasion debt. *Proceedings of the National Academy of Sciences* 108: 203–207. <https://doi.org/10.1073/pnas.1011728108>

Heger T, Jeschke JM, Kollmann J (2021) Some reflections on current invasion science and

- perspectives for an exciting future. *NeoBiota* 68: 79–100.
<https://doi.org/10.3897/neobiota.68.68997>
- Holmes TP, Aukema JE, Holle BV, Liebhold A, Sills E (2009) Economic Impacts of Invasive Species in Forests. *Annals of the New York Academy of Sciences* 1162: 18–38.
<https://doi.org/10.1111/j.1749-6632.2009.04446.x>
- Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, Yang Q, Zhu C, Qiao H (2021) Sampling biases shape our view of the natural world. *Ecography* 44: 1259–1269.
<https://doi.org/10.1111/ecog.05926>
- Hulme PE (2020) Plant invasions in New Zealand: global lessons in prevention, eradication and control. *Biological Invasions* 22: 1539–1562. <https://doi.org/10.1007/s10530-020-02224-6>
- Kraus F, Duffy DC (2010) A successful model from Hawai‘i for rapid response to invasive species. *Journal for Nature Conservation* 18: 135–141.
<https://doi.org/10.1016/j.jnc.2009.07.001>
- Meyerson LA, Mooney HA (2007) Invasive alien species in an era of globalization. *Frontiers in Ecology and the Environment* 5: 199–208. [https://doi.org/10.1890/1540-9295\(2007\)5\[199:IASIAE\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[199:IASIAE]2.0.CO;2)
- Mora C, Rollo A, Tittensor DP (2013) Comment on “Can We Name Earth’s Species Before They Go Extinct?” *Science* 341: 237–237. <https://doi.org/10.1126/science.1237254>
- Moura MR, Costa HC, Peixoto MA, Carvalho ALG, Santana DJ, Vasconcelos HL (2018) Geographical and socioeconomic determinants of species discovery trends in a biodiversity hotspot. *Biological Conservation* 220: 237–244.
<https://doi.org/10.1016/j.biocon.2018.01.024>
- Pimm SL, Jenkins CN, Li BV (2018) How to protect half of Earth to ensure it protects sufficient biodiversity. *Science Advances* 4: eaat2616. <https://doi.org/10.1126/sciadv.aat2616>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pysek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guenard B, Jager H, Kartesz J, Kenis M, Kuhn I, Lenzner B,

- Liebholt AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Stajerova K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Turbelin AJ, Malamud BD, Francis RA (2017) Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography* 26: 78–92. <https://doi.org/10.1111/geb.12517>
- Vicente J, Vaz A, Roigé Valiente M, Winter M, Clarke D, Lenzner B, McGeoch M (2021) Existing indicators do not adequately monitor progress towards meeting invasive alien species targets. <https://doi.org/10.32942/osf.io/mn52e>
- Wilson EO (2016) *Half-earth: our planet's fight for life*. WW Norton & Company.
- Wilson JRU, Datta A, Hirsch H, Keet J-H, Mbobo T, Nkuna KV, Nsikani MM, Pyšek P, Richardson DM, Zengeya TA, Kumschick S (2020) Is invasion science moving towards agreed standards? The influence of selected frameworks. *NeoBiota* 62: 569–590. <https://doi.org/10.3897/neobiota.62.53243>
- Zenetos A, Gratsia E, Cardoso A-C, Tsiamis K (2019) Time lags in reporting of biological invasions: the case of Mediterranean Sea. *Mediterranean Marine Science* 20: 469–475. <https://doi.org/10.12681/mms.20716>

Appendix

Part 1—Guidance on status confidence rating

These categories are adapted from Hawkins et al. (2015).

*Note: these levels reflect the confidence of a plant's status at the time it was assessed.

Confidence Level	Determination
High (About 90% certainty)	Status is based on direct observational evidence <i>and</i> Status is based on thorough surveying or multiple occurrence records <i>and</i> Data sources are recent, reliable, of good quality and are not controversial or contradictory
Medium (About 65-75% certainty)	Status is based on direct observational evidence but some information is inferred <i>and/or</i> Status is based on partial surveying or an insufficient number of occurrence records <i>and/or</i> Data sources may be somewhat out of date, questionable in their reliability and quality, and/or somewhat ambiguous or contradictory
Low (About 35% certainty)	Status is based entirely, or almost entirely, on inferred information <i>and/or</i> Status is based on insufficient surveying or a single-few occurrence records <i>and/or</i> Data sources do not exist, or are outdated and of very poor quality

Part 2—Adapting a Regional Checklist to Track Invasion Statuses

Original Checklist

Species
Species A
Species B
Species C
Species D
Species E
Species F
Species G
Species H
Species I

Categorize species and assign confidence levels based on a variety of data sources about the plant in the location/region examined



Example data sources:
existing checklist categories, reports, surveys, herbarium labels, photographs

Tracking Scheme: Field-based Status

Status	Confidence Level
Potentially Naturalizing*	High
Not Self-Sustaining	Low
Naturalized Where Introduced	High
Naturalized—Unspecified	Low
Not Self-Sustaining	High
Naturalized—Unspecified	Medium
Naturalized Beyond Introduction Site	High
Extirpated	Low
Naturalized—Unspecified	Low

* High priority for Eradication Assessment

Infer likely status of species with low-medium confidence levels using data that is not location specific



Example inferential tools:
weed risk assessments, species distribution models, dispersal kernels, seedbank longevity, detection probability models

Tracking Scheme: Inferred Likely Status

Likely Status
NA
Not Self-Sustaining
NA
Not Self-Sustaining
NA
Naturalized—Unspecified
NA
Data deficient**
Data deficient**

** Highest priority for follow up surveys