4-19-2022

Invasive Species Identification and Monitoring using Computer Vision, Street View Imagery, and Community Science

Liam Megraw
ltm4654@rit.edu

Invasive Species Identification and Monitoring using Computer Vision, Street View Imagery, and Community Science

by

Liam Megraw

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Environmental Science

Thomas H. Gosnell School of Life Sciences

College of Science

Environmental Science Program

Rochester Institute of Technology

Rochester, NY

April 19, 2022

Committee Approval:

Anna Christina Tyler, PhD — Date
Chair of Committee, Thesis Advisor

Christopher Kanan, PhD — Date
Committee Member

Karl Korfmacher, PhD — Date
Committee Member

Mitchell O'Neill — Date
Committee Member

Table of Contents

Acknowledgements

My greatest thanks go to my advisor, Dr. Christy Tyler, who supported me immeasurably over the past couple of years and shared a wealth of knowledge. She helped me grow incredibly as a writer and a scientist, which made me more thoughtful and contemplative in the process of “telling a story.” Throughout it all, she reassured me at every roadblock. Words fail to express the level of patience, dedication, and care she showed me.

I am also thankful for my committee members, who dedicated time to help guide me to create quality contributions to science (and keep me realistic in the process). To Mitchell O’Neill, who routinely gave feedback and suggestions on my community-science related work, and Dr. Karl Korfmacher who asked tough questions that made me better at substantiating rationales and results. I would also like to thank Dr. Chris Kanan, who explained computer vision concepts to me and guided me in making proper decisions.

Regional managers, coordinators, and especially iMapInvasives Program staff helped make this project what it is and supported me at each step of its creation. To John, who repeatedly helped me with GIS applications and sent me many resources. To Jenn who gave friendly feedback on my concepts, and to Meg who helped me use precise, clear, and concise language. I would also like to thank Hilary, Zack, Matt, Tammara, and Emily-Bell who gave feedback on the interfaces and helped connect me with volunteers.

I have also had friends and colleagues working on this project, without whom this project would have been impossible. Thank you to Arturo and Manoj, who developed the model and annotation interfaces, and to Avery and Hannah, who helped greatly with annotations (and kept it light with jokes along the way).

I would also like to extend thanks to the individuals who volunteered their time to help test the accuracy of the computer vision model in the field. Without them, this project truly could not have been what it was.

Finally, I would like to express gratitude to the New York Department of Environmental Conservation for funding this project under award number DEC 01-C00973GG.

ABSTRACT

Invasive plants present significant challenges for ecosystem integrity, biodiversity preservation, and agricultural production. Continuous surveillance efforts are necessary to detect and effectively respond to emerging infestations. However, monitoring methods currently available each have their own limits, whether due to cost, time, or sampling bias. Computer vision applied to roadside imagery is a previously undeveloped methodological synergy that can support existing monitoring efforts. Further, because roadsides are a vector of spread, they are an ideal pathway to monitor. In this study, we present research and management applications of a dataset generated by a computer vision model for *Phragmites* (Common reed) and knotweed complex species across roadsides in New York State. To better understand spread and inform risk assessment, we examined plant presence in relation to road size, site, and culvert characteristics. Results indicated that *Phragmites* presence decreased with increasing distance from a culvert and was less likely near forested areas and on roadsides with low traffic volume. Results also suggest that the risk of invasion relative to road size is species specific: *Phragmites* was found to occur more frequently on highway ramps and primary roads, while the knotweed complex was found to occur more often on secondary roads. Additionally, we developed a framework in partnership with stakeholders that enables managers and community scientists to verify, interpret and act upon the very large datasets resulting from computer vision models. Two ArcGIS Dashboards, several GIS layers, and web-based forms were created to this end. Through this work, we identified that culverts and highway ramps should be monitored given their role as *Phragmites* hotspots. We also generated a new framework that distills a large amount of new data into a form usable by both professionals

and the public, expanding upon the capacity of existing monitoring workflows. Supplemental tables S1 and S2 detail data sources, filters, and product dependencies.

Chapter 1: Introduction

1.1 Invasive Plant Impacts and Spread

Invasive species are “non-native organism[s] whose introduction causes or is likely to cause economic or environmental harm, or harm to human, animal, or plant health” in the introduced ecosystem (Executive Office of The President, 2016). Invasive species are a global threat to biodiversity (Pysek et al., 2012), can damage agriculture, and can negatively impact ecosystem services (Pejchar & Mooney, 2009). There are over 5,000 invasive plant species in the US, leading to documented damages of over $25 billion per year, while displacing native plant species and altering ecosystem composition and function (Pimentel et al., 2005). Invasive plant impacts may be direct, including community effects or the alteration of nutrient dynamics (Pejchar & Mooney, 2009). Other impacts are indirect, such as the provision of habitat for novel pests or as vectors of disease (Pysek et al., 2012). The cost of these impacts is high; estimated losses from invasive plant impacts on agricultural and rangelands total over $33 billion annually (Pimentel et al., 2005).

All non-native species must overcome certain boundaries or environmental conditions to successfully establish and spread in a new region. At any point, if a limiting factor cannot be overcome, an organism’s propagules will die, and that organism will remain in its original location (Vermeij, 1996). Many factors affect invasive plant spread, stemming from a plant’s innate characteristics, external forces, or the interactions therein. Propagule pressure is an important establishment predictor (Cassey et al., 2018) that is both an evolutionary and human phenomenon, with propagule dispersal and success rate influenced by human activity (e.g., traffic spreading seeds) and/or species-specific life history characteristics (i.e., propagule size, number, and risk-release relationship)

(Stringham & Lockwood, 2021). The likelihood of establishment is also moderated by inter-related factors of the target area that act predominantly regionally, but can also be site-specific and vary greatly among species. Climate, existing biotic communities, land use/degradation, and level of disturbance are several common contributors. Climate change has the capacity to alter abiotic, biotic, and human behavioral factors, thereby accelerating, extending, or shifting the potential range for invasive plants (Hellmann et al., 2008). Thus, it is critical to understand the dynamics of invasive plant distribution and spread in order to minimize damage.

1.2 Invasive Plant Monitoring and Management

Every potential invasive species goes through a sequential process of invasion, which is defined by intervals on the invasion curve - prevention, eradication, containment, and long-term control (Figure 1.1). Each stage corresponds to a larger area infested, greater difficulty of eradication, and higher associated costs. Therefore, identification of new invasions ideally happens early, though it is crucial regardless of invasion stage to enable appropriate action to mitigate impacts (National Invasive Species Council, 2016). Early detection and rapid response (EDRR) is a commonly cited approach to invasive species management (Reaser et al., 2020), in line with attempting to prevent establishment in new areas. With increased detection of invasive plant presence, the benefits are two-fold: landowners aware of a species' presence are twice as likely to enact control efforts (Fischer & Charnley, 2012), and pathways for spread can be contained.

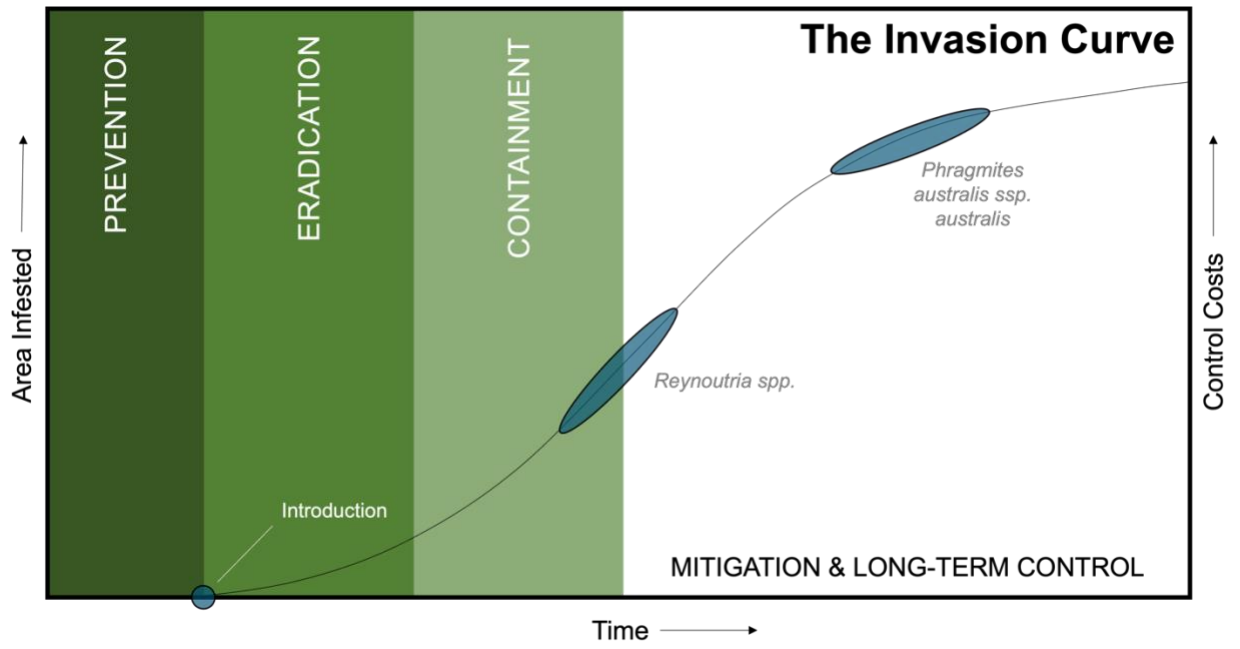

Figure 1.1 *An illustrative invasion curve recreated based on Victorian Government (2010). The curve represents the relationship between area infested and cost of control over time. The resulting management strategies for each stage represent the increasing difficulty of control, from prevention to mitigation and protecting assets. Roughly estimated ranges in New York State for the target species in this study are indicated in dark blue.*

Though monitoring is a necessary task to stem emerging infestations, it can be costly and time-intensive when relying on physical surveying. For example, a subset of nineteen US government agencies spends over $360 million (CPI 2022 dollars) annually – one-eighth of their budget – just on early detection and rapid response (US National Invasive Species Council, 2014). Six additional departments, including the EPA, also have related expenditures not measured by the NISC. These costs and scalability issues can be addressed to some degree with community science, though problems are present with sampling heterogeneity (Dickinson et al., 2010) and variability of data quality (Crall et al., 2011). Aerial or satellite-borne remote sensing methods can address the heterogeneity problem of community science, but imagery often suffers from one of three problems: insufficient spatial resolution (e.g., Olsson et al., 2011; Porter, 2021), insufficient spectral resolution, or high cost of acquisition (Vaz et al., 2018).

A new surveillance method in the form of computer vision applied to Google Street View imagery was proposed by Flores et al. (in prep.). By enabling wide scales of analysis

with limited human review, this new method can leverage and strengthen existing monitoring workflows. The computer vision model applied in this study was created using a convolutional neural network (CNN) architecture developed by Flores et al. (in prep.). Google Street View imagery across New York State (NYS) ranging from 2008 to 2021 was annotated by expert ecologists, then split into training and testing sets (Figure 1.2) (Flores et al., in prep.). The imagery includes all seasons, and varied lighting conditions, though more images were taken in warmer months (Flores et al., in prep.). Final model output takes the form of points with confidence scores that describe the relative probability of a species presence for a specific location.

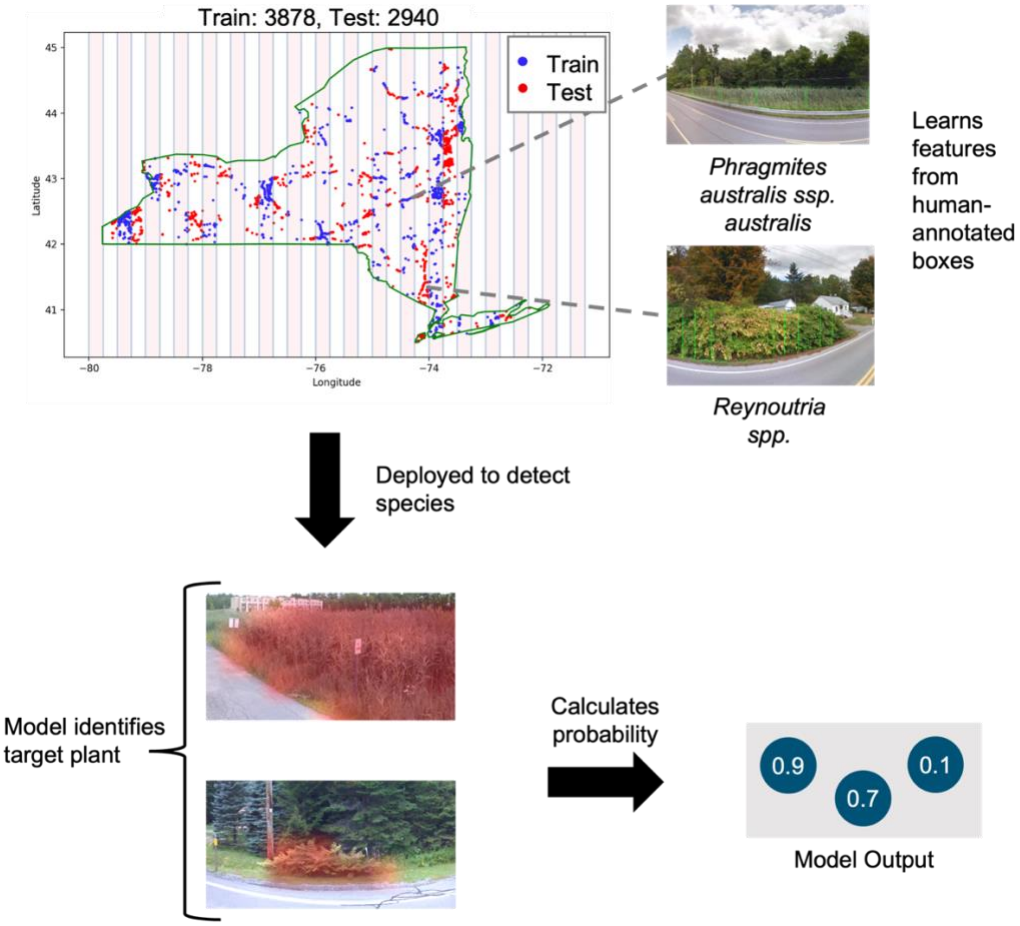

Figure 1.2 Overview *of computer vision model training, testing, and application. Test-train paradigm including location of nearly 7,000 annotated panoramas (reprinted with permission from Flores et al., in prep). Humans annotated bounding boxes for invasive species presence or marked absence in a panorama across New York State. Bounding boxes within training areas (darker stripes) were used for the algorithm to learn. Once trained, the algorithm was applied to the test set of panoramas (lighter stripes). The model can then be applied to a new set of panoramas to calculate presence probability*

The generation of these data sets for target invasive species provides a powerful opportunity to answer important ecological questions about the distribution and spread of roadside invasive plants over large geographic areas and with respect to time. How does the distribution vary with road size and traffic volume? How do documented disturbances lead to enhanced spread? In Chapter 2, I present a case study where I evaluate the distribution of two high impact invasive plants, *Phragmites australis* and knotweed complex species (*Reynoutria japonica*, *R. sachalinensis*, and their hybrid *R. x bohemica*) with respect to the construction of culverts. Culvert construction results in disturbance, and thus we test the hypothesis that such construction leads to enhanced invasion along roadsides adjacent to the culvert relative to roadsides further from culverts.

Simultaneously, such methods for detection of infestations along roadsides will lead to the generation of a massive amount of new data. Management, validation, and use of these data requires development of an intentional pipeline for making data available. To do this, we worked closely with several stakeholders to create a framework and workflow for the incorporation of computer vision results into existing workflows. Chapter 3 describes this process and the resulting interface to be implemented.

In Chapter 4, I summarize this work, highlighting insights and applications. I also suggest future research directions, while discussing data limitations.

Chapter 2: Roadside Presence and Culvert Analysis

2.1 Introduction

2.1.1 Invasive Plant Impacts, Spread, and Monitoring

Across the world, invasive species threaten community integrity and agriculture while disrupting ecosystem services (Pejchar & Mooney, 2009; Pysek et al., 2012). In the US alone, invasive plants led to documented damages of over $25 billion per year, while also displacing native plant species and altering ecosystem composition and function (Pimentel et al., 2005). Identification and control of these invaders is paramount to mitigate harm (National Invasive Species Council, 2003; 2016). Roadside habitats are ideal targets for invasive plant monitoring and control, since they provide disturbed conditions amenable to invasive plant establishment and spread (Christen & Matlack, 2009; Maheu-Giroux & De Blois, 2007), while also increasing propagule dispersal through mechanisms including traffic (Lemke et al., 2019; Taylor et al., 2012) and maintenance activities as proposed by Gelbard & Belnap (2003). Understanding invasion dynamics in these habitats will allow us to better target management efforts. Unfortunately, traditional roadside surveys can be time-consuming and labor intensive. Emerging synergies between tools and data sources – such as computer vision and Street View Imagery (SVI) – provide a fast and cost-effective alternative.

The role of roadsides in invasive plant spread is well-established. Along roadsides, light availability is higher, and stem and canopy densities are lower. Because of this, roadsides act as invasion conduits that allow species to spread along the road axis at higher rates and to larger extents than in non-roadside environments (Christen & Matlack, 2009). By facilitating the geographic expansion of invaders (e.g., Maheu-Giroux & De Blois, 2007; Ward et al. 2020), roads have the potential to increase the scale of impacts

on biodiversity and ecosystem services. In the US, over 80% of land is within 1 km of a road (Riiters & Wickham, 2003), illustrating the scale of invasion risk and the complexity of monitoring. In some instances, road presence and characteristics are better predictors of invasive plant presence than surrounding land use (Joly et al., 2011) or site characteristics (Gelbard & Belnap, 2003). Testing the effect of road size and type have yielded mixed results, with some studies finding differences (e.g., Gelbard & Belnap, 2003; Joly et al., 2011) while others (e.g., Deparis et al., 2020) have not. Spread along road verge habitat (referred to here as the road-adjacent non-forested land) may be exacerbated when combined with construction activities such as for bridges, culverts, or road and shoulder widening.

Consider the role of construction as a vector for invasive species introduction. Though likely unintentional, dispersal can result from purposeful planting of ornamental species (e.g., Deparis et al., 2022) or accidental spread from other areas (e.g., Barlow et al., 2017). Corcos et al. (2020) found that roadside bare soil did not re-vegetate to pre-disturbance levels even after a growing season, while Ailstock et al. (2001) found that *Phragmites australis* (common reed) seeds only established on unvegetated, non-burned soils. These results suggest that construction could create openings for invasive plant recruitment. Depending on road type and weather, seeds on vehicles can be transported for over 200 kilometers (Taylor et al., 2012), so the possibility of dispersal remains even for relatively small construction projects. Together, one would expect increased traffic and equipment from off-site to increase invasive plant propagule pressure. Further, when fill soil is used in construction, it may be contaminated with seeds, rhizomes, or other plant fragments.

Existing regulations may be insufficient to prevent spread from construction activities. The US Plant Protection Act does require inspection of fill soil to prevent noxious weed introduction, but only if it is from other nations, its territories, or areas under federal quarantine (Title 7 U.S.C. § 7701). Interstate movement of “plant products” can be regulated by the same Act, but regulation of intra-state transit is left to the states themselves (Title 7 U.S.C. § 7712). In contrast, the Canadian Ministry of Environment dictates that parties receiving soil must create and follow a management plan to prevent the introduction of invasive species (Ministry of Environment, Conservation and Parks, 2019). Before policy decisions are made, the role of construction should be evaluated further.

Given the role of roadsides in invasive plant spread, it is worthwhile to pursue efficient methods of roadside detection. Roadside detection allows us to identify infestations earlier and monitor those that are ongoing, helping management in myriad ways. An emerging technologic application with promise in this regard is computer vision, which can be defined broadly as both the discipline and practice of using algorithms to automate image interpretation tasks (Huang, 1996). Combined with deep learning techniques, computer vision can be used for invasive species identification with potential to overcome limitations of current monitoring methods. Applied on Street View Imagery (SVI), there is exciting potential for both low-cost invasive plant detection and ongoing monitoring over large spatial scales; Google’s imagery alone allows up to ten million miles (over 16 million km) of roadside habitat to be assessed globally (Escobar, 2019). Further, car-level imagery allows for identification of understory plants and emerging infestations that would otherwise remain undetected using at-nadir aerial and satellite imagery.

SVI has been used in studies to compare human observation/assessment results in the field versus in virtual drive-throughs (e.g., Griew et al., 2013; Rousselet et al., 2013), with two (Deus et al., 2016; Kotowska et al., 2021) directly examining SVI's utility for invasive plant identification. In each case, authors found high agreement between methods. Another study examined the feasibility of using unsupervised classification methods with SVI to identify a set of invasive plants that include our target species, though results were not promising (Connell, 2015). Computer vision has also been applied to SVI for tasks including estimating demographics (Gebru et al., 2017) and the inventory of urban trees (Wegner et al., 2016). However, to our knowledge, computer vision and deep learning have not yet been applied to SVI for the purpose of invasive plant identification. Given limited time, ability, and resources, managers must prioritize where to monitor and intervene such that mitigation efforts are optimal. Understanding the factors that allow introduction and facilitate spread are necessary components to a more comprehensive approach when managing invasive species.

In New York State, US, there are at least 284 terrestrial invasive plants (NY iMapInvasives, 2021). Of these, *Phragmites australis ssp. australis* and the knotweed complex *(Figure 2.1)* present significant ecological and economic challenges. These invasive plants are ideal target species to study when examining roadside habitat conditions and maintenance activities. Both do well in wet areas (such as roadside ditches or culverts) and tolerate high salt concentrations (Keller, 2000; Rouifed et al., 2012). Further, each can establish via seed or from plant fragments (Albert et al., 2015; Bram & McNair, 2004) that might be present in contaminated fill soil.

Phragmites australis (common reed) has three recognized subspecies, with both the invasive European lineage *P. australis ssp. australis* (hereafter referred to as *Phragmites*), and the native North American lineage *P. ssp. americanus* present in the Eastern US (Saltonstall, 2002; Saltonstall, 2003). While the native lineage has existed for tens of thousands of years in North America, the non-native subspecies were likely introduced sometime after the 1800s (Saltonstall, 2002). *Phragmites* reproduces both sexually and asexually (Albert, et al., 2015), though primarily via rhizomes, and has expanded across the US since its introduction (Saltonstall, 2002). Upon invasion, *Phragmites* forms tall, dense monocultures that reduce native alpha diversity (Keller, 2000). The primary mechanism for takeover is density-dependent resource competition (Uddin & Robinson, 2017) with a secondary mechanism facilitated by allelopathy (Rudrappa, 2009; Uddin et al., 2014; Uddin & Robinson, 2017; Weidenhamer et al., 2013). Because non-native *Phragmites* can tolerate higher salt concentrations than its native counterpart (Vasquez et al., 2005), roadside conditions are amenable to establishment and spread; roadside growth may also present visibility concerns due to its tall growth pattern.

Reynoutria japonica (Japanese knotweed) is one of several in the invasive knotweed complex (*Reynoutria japonica*, *R. sachalinensis*, and their hybrid *R. x bohemica*) that have similar morphology, growth characteristics, and impacts. *R. japonica* is another species introduced as an ornamental in the late 1800s, originally from Asia. It proliferates in wet conditions and is tolerant of salinity up to seawater concentration (Rouifed et al., 2012), allowing it to colonize disturbed areas such as roadsides or culverts. It forms dense, often clonal vegetation stands that can cut native species density

in half and reduce species richness by 2-5 times (Aguilera et al., 2010; Wilson et al. 2017). *R. japonica* exhibits high seed germination rates in addition to extensive growth via rhizomes (Bram & McNair, 2004). The primary mechanism for takeover is reducing light availability, by plants that may reach 2 m tall and develop biomass 2-5 times that of neighboring communities (Aguilera et al., 2010).

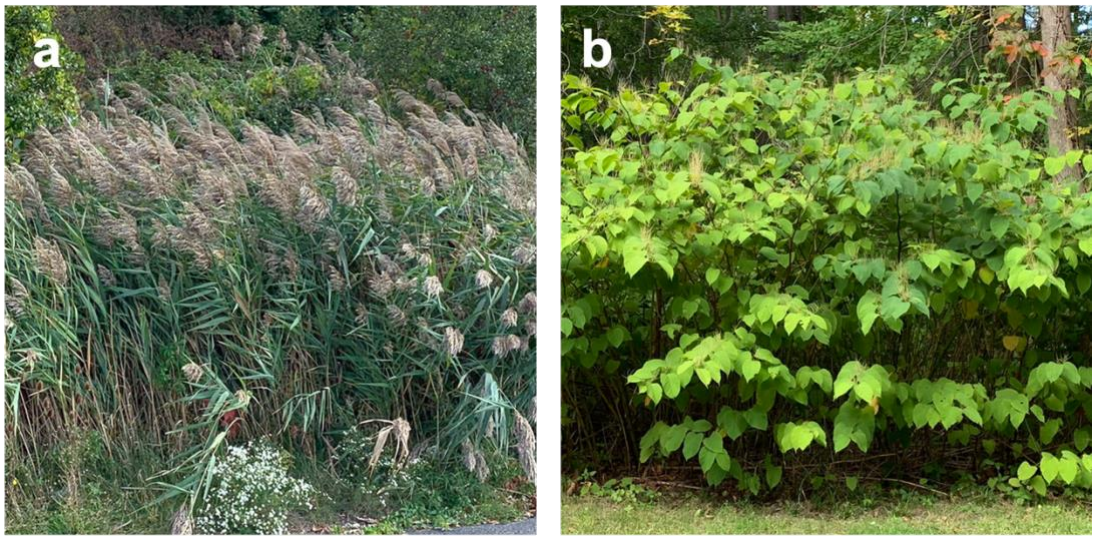

Figure 2.1 *Example images of the two invasive plants examined in this study. The species are a)* Phragmites australis ssp. australis *(common reed) and b) a member of the knotweed complex,* Reynoutria x bohemica

With our species of interest, we do not know to what extent spread is being impacted by road characteristics or habitat conditions. Computer vision enables us to generate a large amount of data that can be used to examine road, road verge, and regional characteristics. By doing so, we can gain a more accurate perception of spread risk. Therefore, in this study, we use data from a convolutional neural network applied to SVI (Flores et al., in prep.) to identify two high-priority invasive plants in New York State, US.

The objectives of this study were to examine potential reasons for the presence of the knotweed complex and *Phragmites* both within and between culvert sites. We hypothesized that the size of culvert construction footprints (width and number of lanes crossed) and age would be positively related to invasive plant presence, given larger disturbance and longer establishment times. We also hypothesized that roads with higher traffic volumes would have increased invasive plant presence due to increased wind-dispersed propagule spread. Similarly, we hypothesized that roads with larger verges (using lanes as a proxy) would be associated with increased plant presence due to providing a larger area for establishment. Finally, we had three hypotheses regarding land cover types. We expected that adjacent developed area, through disturbance and propagule pressure, would be positively related to plant presence. We also expected agricultural areas to be positively related to presence, given that higher nutrient loads increase growth and competitive ability of our target plants (Kettenring & Whigham, 2018; Parepa et al., 2019; Uddin & Robinson, 2018). Finally, adjacent forest cover and associated shading was expected to be negatively related to *Phragmites* presence (Kettenring & Whigham 2018; Li et al. 2011) but not for the knotweed complex due to shade tolerance and dominance in understories (Wilson et al., 2017).

2.2 Methods

To evaluate the relationships among the presence of *Phragmites* or knotweed complex species and road characteristics, we generated the predicted presence/absence of each species using a computer vision algorithm applied to Google Street View roadside imagery for New York (NY), excluding New York City and Long Island since those areas

had no culverts within the date range selected. We then generated a geospatial dataset for road culverts (width, built date), road characteristics (number of lanes, traffic volume), and generalized land cover (agriculture, developed, forest) for comparison. The presence of each species along a transect running parallel to the road away from each culvert was evaluated in combination with the other characteristics using a binary logistic model to determine the potential interactions among culvert construction, road type and the infestation by target species.

2.2.1 Geospatial Data Acquisition and Pre-Processing

External datasets were accessed from several sources. Large culverts (between 1.5 and 6.1 m) built before 2019 (latest available at time of analysis) were downloaded from the New York State (NYS) GIS Clearinghouse. Estimated average annualized daily traffic (AADT) volume data calculated in 2019 were also downloaded from the NYS GIS Clearinghouse. These data contain 2019 estimates based on actual traffic counts from between 2002-2019. For simplicity, the 2019 estimate was used instead of traffic counts at the time of culvert construction. Per-county 2021 road data were downloaded from the US Census FTP website. Land cover for 2016 (latest at time of analysis) from the Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD LULC) was downloaded via Google Earth Engine.

Computer vision model prediction point data for the two plants were provided by Flores et al. (in prep) for panoramas captured by the Google Street View program between 2011 and 2018. These data take the form of panorama locations with a model-generated confidence score for each target species. If these scores are to be interpreted

as predictions of presence and absence, confidence scores below a specific value ("threshold") must be considered absences. At lower thresholds, the model is more *sensitive*, meaning a higher proportion of plants present in the field are captured by the model. The trade-off is a frequent occurrence of false positives. The threshold for each species was chosen to minimize the mean false detection rate (i.e., the proportion of predicted presences that are false) across four test sets. This is equivalent to maximizing mean precision. The threshold used for *Phragmites* was 0.95 while for the knotweed complex it was 0.98. The corresponding mean false discovery rate for *Phragmites* is 5.5% and 4.5% for the knotweed complex. These rates are estimated based on validation sets of expert-generated annotations for species presence and absence (Flores et al., in prep.). When the model was deployed on representative sampling zones, comparisons between onscreen and field validation methods yielded high (>90%) agreement (Flores et al., in prep.).

2.2.2 Culvert and Road Selection

A subset of culverts was selected from the entire set of culverts built in the study area of NY after 2006 based on the following criteria. Processing was split between UTM zones 17N and 18N for more accurate distance calculations when geodesic computation was not available. A total of 34 culverts selected in 17N and 93 culverts selected in 18N (Figure 2.2). We set an initial search distance of 13 m as the maximum distance expected between culverts and Street View panoramas, given GPS accuracies and panorama capture interval (line c in Figure 2.3). Of the 374 culverts built in 2007 or later, 175 were within 13 m of model data and were selected. Any culverts farther than 13 m but less than

1 km away from model data were manually reviewed to visually assign them to a road, if possible. Of the remaining, 176 had no model data within 1 km and thus were not considered. Culverts within two km of another culvert on a road with the same NYS Department of Transportation identification number were also manually reviewed and removed. This was done to avoid confounding effects from overlapping culvert transects. In two instances, there were duplicate points recorded for the same culvert, so the duplicates farther from model data were removed.

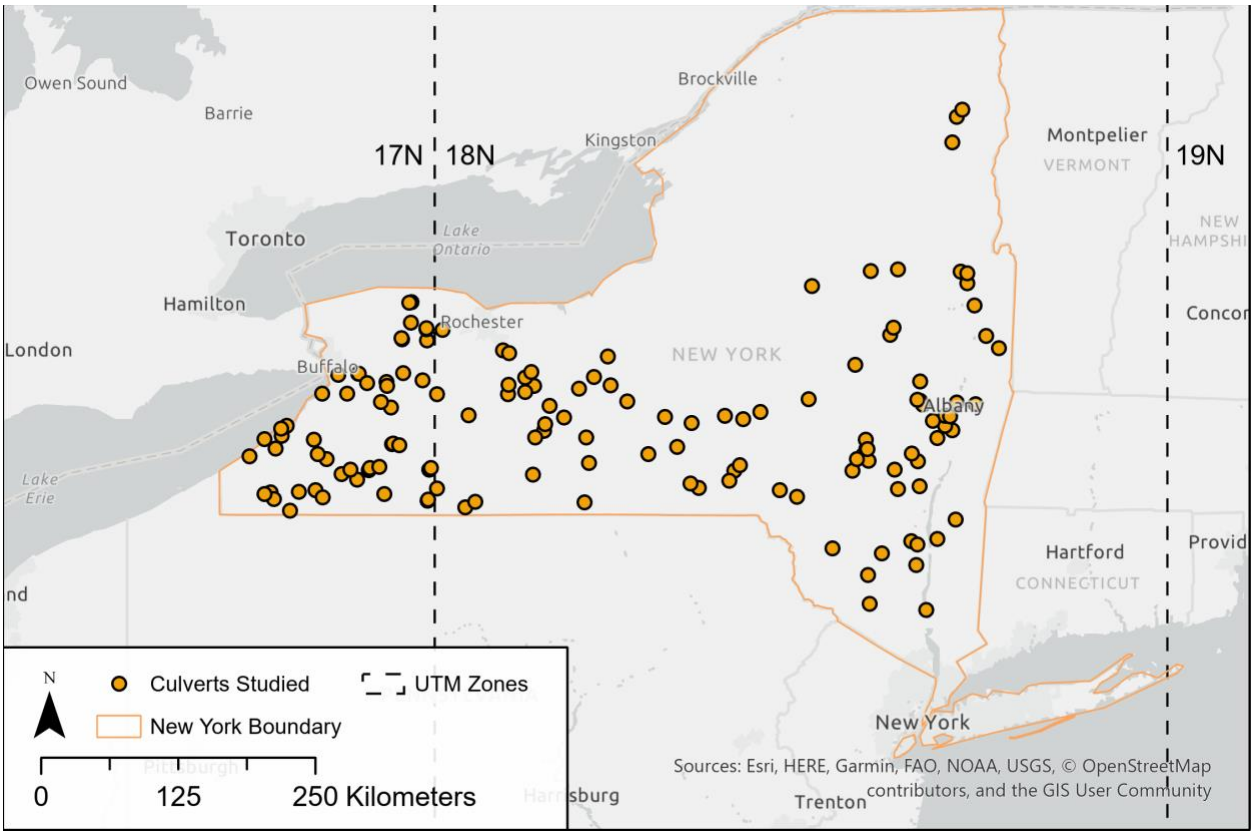

Figure 2.2 *Locations of the 127 culverts examined. All culverts were built in 2007 or later and were farther than 2 km from another culvert.*

The presence of invasive plants was evaluated along two transects running bidirectionally away from one another along the primary road associated with each culvert (Figure 2.3). To generate the transects, roads in the AADT dataset were manually

reviewed and selected to account for irregularly split features when making transects. To do this, AADT roads were intersected with 1 km buffers around the culvert features. Then every culvert was manually reviewed, selecting all contiguous roads with the same NYS Department of Transportation identification number. In one case where a road ended at a “T” intersection with neither road sharing the same NYS Department of Transportation identification number, a direction was chosen via random number generation in Microsoft Excel v16.59. In one case (culvert C081161), the culvert lay between roads on a highway, so we selected the road closer to the ESRI Service Area network used for transect point assignment (explained in the next section).

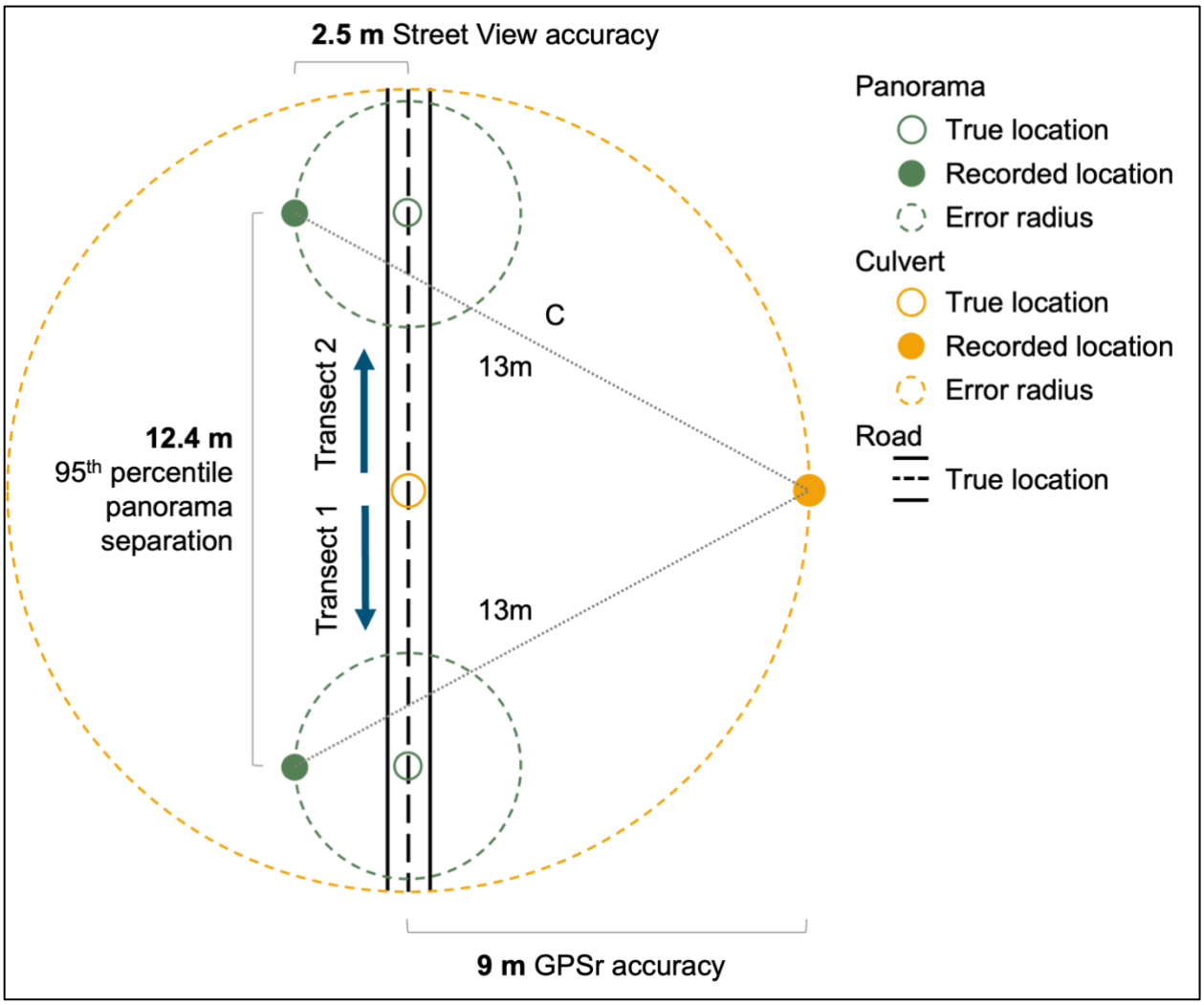

Figure 2.3 *Illustration transects and of the farthest scenario used to derive maximum distance threshold between panorama and culvert locations. The 95th percentile of distance between panoramas for the dataset is 12.4 m. Google stipulates that at minimum, 50% of locations recorded should be within 2.5 m of their true location (Google, 2018). The culvert data do not have metadata regarding the GPS receiver used to capture points, so we can assume the 95th percentile of horizontal accuracy of 9 m reported by the US Department of Defense (2008). In this scenario, both panorama locations would be selected.*

2.2.3 Creating Transect Points and Joining Data

Accuracy of the road network was estimated to appropriately select model data and culvert points. The metadata for the AADT dataset did not indicate a road geometry source or accuracy estimates, so features were compared to the US Census road dataset. A total of 31% of AADT line features shared a segment with census roads and

97% intersected, indicating a partial match. The most recent study available for US Census road feature accuracy was conducted by Zandbergen, et al. (2011) for the 2009 roads dataset, which indicated that 90% of Census road features should be within 11.3 m of their true location.

A search radius of 15 m between transect points and model data was determined by road accuracy, panorama accuracy, and panorama interval. Given that transect points were located along the AADT line features, 90% of transects should be within 11.3 m of the true location of a road (as described above). Google requires that 50% of recorded panorama locations be within 2.5 m the true location of a road (Google, 2018). After calculating near distance in ArcGIS Pro v2.5.2 (Esri, 2020) and identifying percentiles in Microsoft Excel v16.59, it was determined that 95% of panoramas are within 12.4 m of one another. Assuming the farthest locations possible, most transect points along the census roads should have at least one panorama within 15 m (following the same logic as presented in Figure 2.3). This 15 m search distance dictated that the first and closest transect point could be no closer than 30 m to the culvert.

Transect points were created at 30, 60, 120, 240, 480, and 960 m with culverts representing distance 0. These distances were chosen to capture a range of distances with more granularity closer to the culvert. Transect point location assignment was calculated using the Service Area analysis tool. This Service Area approach was chosen instead of radial buffers to correctly assign transect distances on curved roads and sharp turns. The default ESRI roads were used as the input network. The travel parameters selected were “Driving Distance” and “Not Using Time”, since it is irrelevant to consider time to reach destination constrained by speed limits and traffic conditions for plant

spread. All Network Analyst restrictions on road movement were removed, since factors such as road closure at time of analysis were not relevant for invasive plant spread. U-turns were not allowed, preventing undesired duplicate transect points. The following output geometry attributes were used: high precision, overlap, and disks. Output geometry was not simplified, and a 25 m polygon trim distance was used. These two options avoided odd angles that would have distorted distance assignment. Transect points were then created by intersecting the resulting Service Area polygons with the manually verified AATV road features (Figure 2.4a). Transect points were merged with culvert points for following analyses and are hereafter simply referred to as “transect points”.

Computer vision model-predicted presences for each species were then joined to transect points. To do this, thresholded model data within 15 m for each species were spatially joined to transect points (Figure 2.4b). For any culverts manually added into the dataset, the closest point on a road parallel to the culvert was used as the center for the 15 m search distance. There was only one culvert manually added that had *Phragmites* present under this condition. A total of 41 transect points in zone 18N did not have model data within 15 m and were removed, as was one point in zone 17N.

There were some discrepancies in transect point generation. In zone 18N, ten additional transect points were generated than should be expected based on number of culverts and transect points per culvert. In zone 17N, 11 additional were generated. These are likely due to the AADT road data geometry and how the Service Area analysis operates. These errors were not removed, since it was difficult to numerically determine

which points were indeed errors and which were correct. Note that an overpass near culvert C081161 caused extra transect points to be added, so they were removed.

Generalized land cover types including forested, developed, and agricultural land were included as site characteristics for analysis (Figure 2.4c). There is evidence that *Phragmites* exhibits some shade intolerance (Kettenring & Whigham 2018; Li et al. 2011), which suggests that forest cover may be a predictor of presence. Thus, mixed forest, coniferous forest, and deciduous forest were combined for the forest class when modelling. Developed (medium and high intensity) were also considered, given a previous study (Maheu-Giroux & de Blois 2007) found numerically higher colonization rates in industrial and commercial areas compared with others. Agricultural (hay/pasture and cultivated crop) areas were also included because of the increased growth and competition rates of *Phragmites* in nutrient-rich settings (Uddin & Robinson, 2018). Each of these types were isolated by masking the NLCD 2016 raster, intersected with 30 m transect point buffers, and had their areas summarized using a spatial join.

Finally, road and culvert attributes were added to all transect points. Culvert point attributes were spatially joined to a 1 km buffer and then spatially joined to transect points within that buffer. The near tool was used to calculate distances between culverts and their assigned road features to identify the maximum distance between the two. Road attributes were then spatially joined to transect points when within the maximum geodesic distance identified (Figure 2.4d).

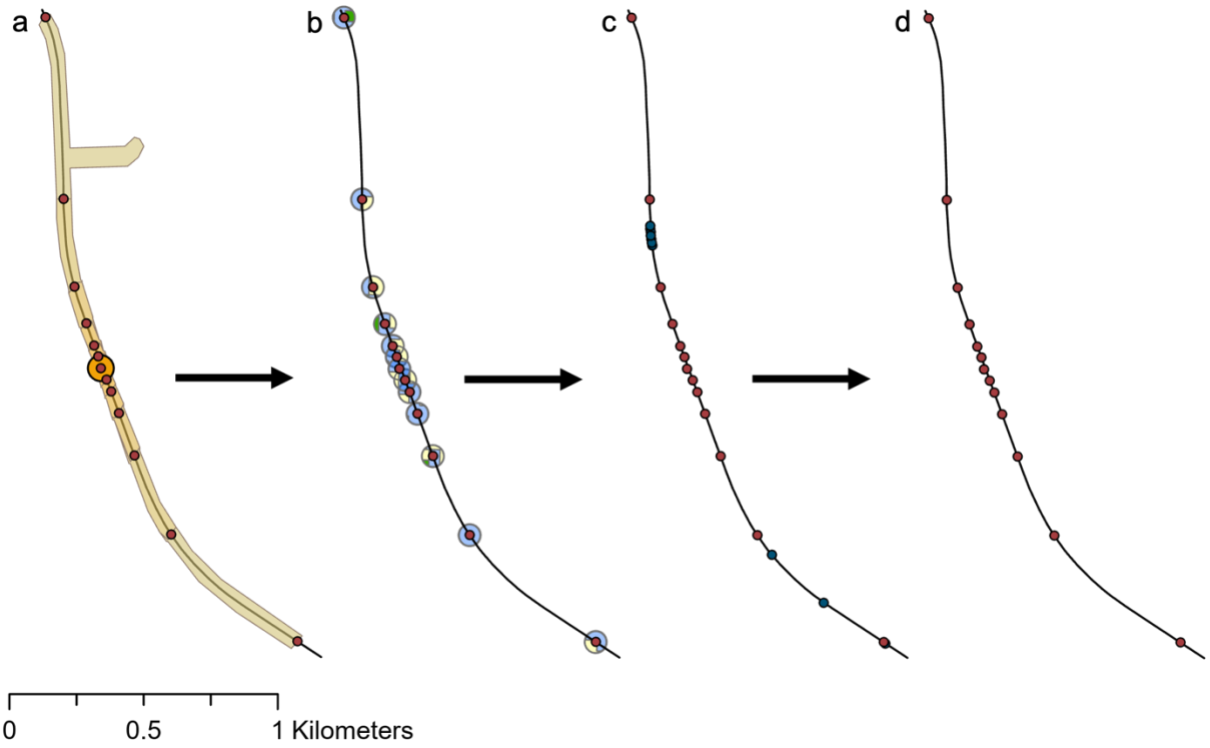

Figure 2.4 *Broad steps in the process used to generate transect points and join attributes. In (a) ArcGIS Service Areas were intersected with road features to create transect points. In (b) computer vision model-predicted presence data within 15 m of points were joined to transect points. In (c) generalized land cover areas were summarized to 30 m point buffers. Finally, in (d) road and culvert attributes were joined to transect points.*

A binary logistic model was performed in RStudio 2021.09.1 Build 372 (RStudio Team, 2021) using the base R package to predict *Phragmites* and knotweed complex presence based on road, site, and culvert characteristics. The variance inflation factor calculated using the car package (Fox & Weisberg, 2019) resulted in no values above two, indicating no severe multicollinearity between predictor variables. Cook's distance was calculated using the base R package and indicated numerous outliers for the logistic model above (4/n), where n is the number of observations. For the *Phragmites* model, this outlier set included every transect point with a computer vision predicted presence plus three additional transect points. For the knotweed complex model, the outlier set only included every transect point with a computer vision predicted presence. Each of these

data points strongly affect the model fit. Since this is a binary model with less than 5% of records being presences, one would expect presences to be outliers of the dataset and therefore strongly influence the model fit. The sample size requirement was satisfied for the variables included in the refined model for *Phragmites* that only included significant variables.

2.2.4 Entire Dataset Road Size Analysis

A follow-up analysis was performed to evaluate road types (primary, secondary, tertiary, highway ramp) as predictors of plant presence, since nearly all transect points in the culvert analysis were present on two-lane roads. While the culvert analysis used over 8,000 panoramas, the 1 km buffers for the 127 culverts studied contained computer vision predictions for 94,219 panoramas. Of these, 1,325 were farther away than 15 m from a road and thus excluded since road type could not be easily identifiable. This 15 m threshold is the same as described in section 2.2.3, accounting for panorama capture interval, GPSr accuracy, and US Census road feature accuracy. These panoramas came from scenarios such as when road features did not exist for the roads where panoramas were captured, or from isolated photospheres that were not captured along roadsides.

Each of these codes was assigned to a road type based on the metadata descriptions. Road MTFCC codes were first spatially joined to panorama locations. Primary roads (MTFCC = S1100) are major roads such as interstates, secondary roads (S1200) are usually state or county highways, ramps (S1630) are roads that enable access to primary roads, and tertiary roads (S1400, S1640, S1740, S1780) include all

remaining roads intended for automobile traffic (US Census Bureau, 2021a). There were no alleys (S1730) in the road dataset studied.

Spatial autocorrelation in these data can artificially inflate significance so this phenomenon was examined. Spatial autocorrelation essentially makes observations "pseudo replicates" of one another, which violates independent sampling requirements of many statistical tests (including chi square) (Legendre, 1993). Logically, an observation close to a known plant presence is also likely to be a plant presence, given how plants disperse across the landscape. This phenomenon was identified by varying the distance that defined a sufficiently far (i.e., independent) observation. To do this, presence data for *Phragmites* were spatially joined to various grid scales (100, 200, and 300 m) and analyzed using the Cluster and Outlier (Anselin Local Moran's I) tool in ArcGIS Pro v2.5.2. Then, a 2 km Euclidean distance neighborhood was used to include all grid cells within culvert buffers, with inverse distance as the spatial relationship to mirror plant dispersal. The false discovery rate correction was applied, to adjust for the over 2500 cells being analyzed. All other options were left as default. Ultimately, a 250 m aggregation size was selected as it was sufficiently large to prevent any significant spatial correlation at $p<0.05$. The same resolution was assumed for the knotweed complex, then knotweed complex presences were joined to the grid cells.

An additional road size analysis was performed to identify the effects of road type on plant presence. When grid cells with both presences and absences were included, grid cell presence counts were heavily skewed right and median presences were zero for both plants and all four road types. Therefore, ANOVAs and Kruskal-Wallis tests could not be used. As such, a chi-square test was performed on raw frequencies in

RStudio 2021.09.1 Build 372 (RStudio Team, 2021) using the base R package. It is important to note that when using these raw frequencies, any significance found may be overstated because of the spatial autocorrelation that remained unaccounted for (Legendre, 1993).

2.3 Results and Discussion

2.3.1 Culverts and Transect Points

A total of 52 of New York's 62 counties had culverts built in 2007 or later. None were in New York City. Of the 52, 38 counties had panoramas that overlapped with culvert locations. One of the 38 counties only had culverts where transects overlapped with those of a nearby culvert, so it was not sampled. Of the 127 culverts studied, the vast majority were from two-lane roads (Figure 2.5a). All ten culverts predicted by the computer vision model to have *Phragmites* at the location of the culvert had *Phragmites* visible in at least one year of SVI per manual inspection. Notably, only one culvert built in 2018 was sampled (Figure 2.5b). The land area sampling distributions (Figure 2.5f-h) are somewhat jagged, likely due to buffers overlapping between culvert and the 30 m transect points.

Based on our method, one would expect 127 transect points at distance zero (i.e., the culverts themselves), and 254 points for each of the transect distances. However, the true number of transect points analyzed differed across the culverts. The effect of the 21 extra transect points generated by our method and 42 transect points removed are illustrated by the distribution in Figure 2.5e. Because of these errors and missing data points, the distance of 480 m is slightly oversampled (by 4 points), while the distances

between 30 m and 240 m are slightly under sampled (by 8 – 11 points). These are likely from instances where the AADT road data had spurs present, or loops in the ESRI network caused the Service Area analysis to generate an additional point on the transect road. Sampling could be improved by selecting only one transect direction and removing artifacts from the Service Area analysis.

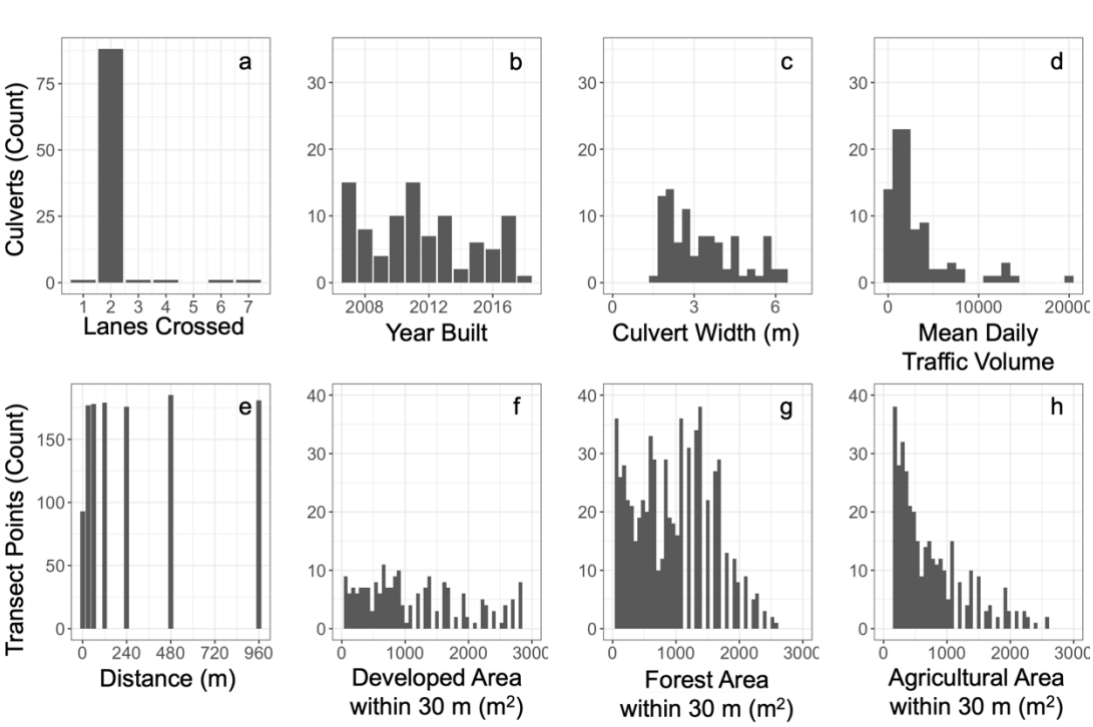

Figure 2.5 *Distribution of road and culvert characteristics in relation to culvert count (a-d) and site characteristics by transect point count (e-h).*

2.3.2 Predictors of Presence in Transects

Of the eight variables considered in the binary logistic regression as predictors, three were significant at $p<0.05$ for *Phragmites* presence, while none were for the knotweed complex. The most notable result for the knotweed complex was traffic volume ($p=0.09$). It is likely that sample size was insufficient to establish relationships for the knotweed complex, given it was only present in 34 transect points. For *Phragmites,*

distance to culvert, traffic volume, and forest area were significant (Table 2.1). Distance to culvert ($p<0.03$) and forest area ($p<0.05$) were inversely related to likelihood of presence, while traffic was positively related to the likelihood of presence. Insignificant variables were then excluded from the final model (Table 2.2). One parameter of the binary logistic regression is a series of odds ratios, which can be interpreted as power functions modeled by r^n, where r is the ratio and n is the number of increases of a specific unit (or alternatively, the power function in Figure 2.6b). The likelihood of finding *Phragmites* 1 km away from a culvert, for example, would be 63% lower (0.999^1e3) than the likelihood of finding it *at* the culvert itself (Table 2.2). These results match the true trends observed (Figure 2.6a), however the modeled relationship is influenced to some degree by the variable number of transect distances sampled. This could be addressed by sampling only one transect per culvert instead of two. When comparing coefficients in the final model (Table 2.2), distance to culvert is still the factor most influential on likelihood of presence, being 10-fold that of traffic volume and nearly 2-fold that of forest area.

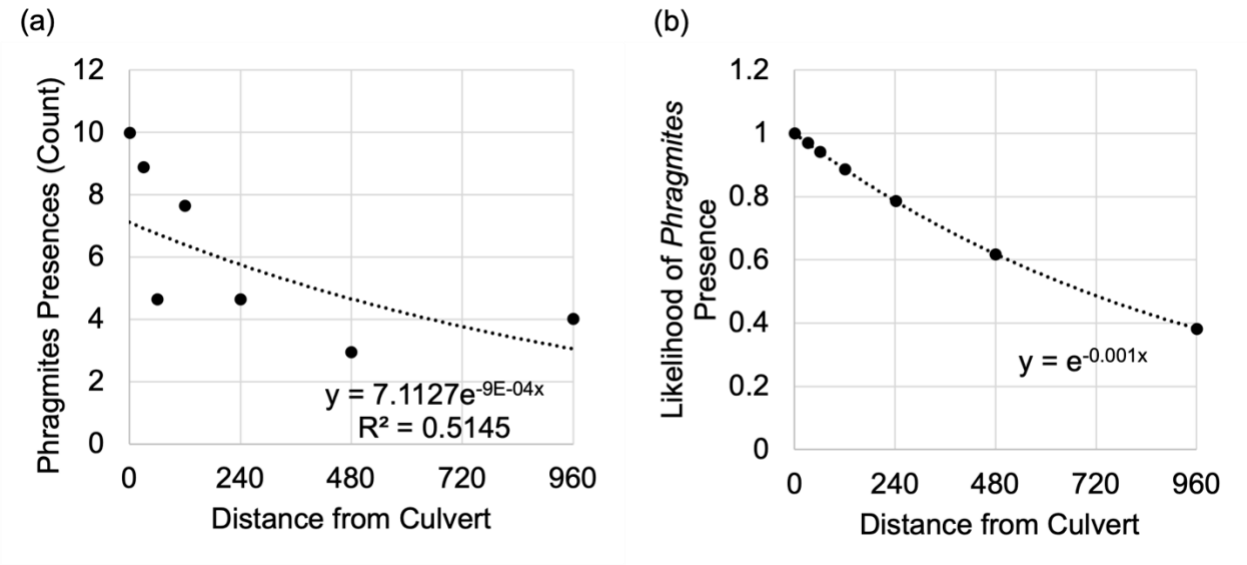

Figure 2.6 *Plots of the (a) actual and (b) modeled presence as a function of distance from culvert. To account for variance in transect distances sampled, presence values were scaled for (a), dividing the number of computer vision predicted presences by x/127, where x is the number of transect points at some distance, d. The modeled presence likelihood (b) is from the refined binary logistic model.*

These results support the hypothesis that culverts act as "hotspots" for *Phragmites* presence, though there are still at least three potential mechanisms explaining why this is the case. The first could indeed be that culvert construction itself is responsible, either by aiding spread via creation of a disturbed patch of soil or fill soil directly introducing *Phragmites* propagules. With shale gas pads in Pennsylvania, US, it was found that construction resulted in the introduction of an array of invasive species, likely from traffic spreading propagules from surrounding areas to newly disturbed ones (Barlow et al., 2017). There are a couple of key differences in the construction phase of shale gas pads and culverts examined here. Namely, shale gas pads have multiple components and associated infrastructure (e.g., pipelines) required for their operation (Costa et al., 2018), which would logically result in larger construction footprints and a higher volume of construction-related traffic compared to culverts. Although the disturbance and propagule pressure are likely greater, trends were expected to be similar with culvert construction

just at a much lower magnitude. If the expected trend were the case, the size of the construction footprint should create a larger disturbance (Corcos et al., 2020), increasing the likelihood of establishment given sufficient propagule pressure. Contrary to this expectation, culvert width was not a significant predictor. It is possible that the differences in culvert sizes examined here were not large enough to identify a difference due to footprint size. Further, if culverts themselves were acting as an introduction event, one would also expect that older culverts would increase the likelihood of presence in the transect given a longer time to spread outward following an introduction, if the introduction occurred close in time to the initial construction. However, this was not the case as culvert age was not a significant predictor of presence.

The second mechanism could be that conditions near culverts are amenable to establishment from intersecting streams sources, either at the time of construction or after. Notably, Maheu-Giroux & de Blois (2007) found that *Phragmites* first invaded transportation corridors (roads & railways), then spread outward. When riparian habitats did not intersect road and railroad corridors, they were about 6% as invaded by *Phragmites* compared to riparian habitats that *did* intersect these corridors (Maheu-Giroux & de Blois, 2007). Thus, it appears linear wetlands (including roadsides) are acting as a primary spread vector while *Phragmites* is spreading along intersecting riparian zones to a lesser degree.

The third mechanism could be that conditions near the culvert are the most amenable to *Phragmites* establishment compared to the rest of the roadside habitat, given their affinity for wet conditions (Keller, 2000; Maheu-Giroux & de Blois 2007). However, if the second case were true, one would expect the decay in likelihood to be

more extreme (i.e., there would be a high likelihood near the culvert which would then immediately drop off). Additionally, Maheu-Giroux & de Blois (2007) found that riparian habitats intersecting railroads and roads were about 60% as invaded as linear wetlands (road, railroad, agricultural ditch, and riparian habitat) overall, which suggests that these riparian habitats are less amenable to *Phragmites* establishment than other anthropogenic linear wetlands.

Table 2.1 *Summary of initial binary logistic model of Phragmites presence. Significance is denoted by * (at p<0.05) and *** (at p<0.001).*

Variable	β (Estimate)	S.E.	Odds Ratio	z	p	
Distance	-0.00100	0.00046	0.9990	-2.18	**0.02950**	*
Lane Count	-0.0692	0.1450	0.93	-0.48	0.63	
Culvert Width	0.0089	0.0940	1.01	0.09	0.92	
Built Year	0.0561	0.0375	1.06	1.50	0.13	
Mean Daily Traffic Volume	0.00010	0.00003	1.0001	3.55	**0.00038**	***
Developed Area	0.0003	0.0002	1.00	1.54	0.12	
Forest Area	-0.00052	0.00035	0.9995	-1.98	**0.04718**	*
Agricultural Area	-0.0004	0.0004	1.00	-0.98	0.33	
(Intercept)	-115.64	75.343				

Table 2.2 *Summary of refined binary logistic model of Phragmites presence.*

Variable	β (Estimate)	S.E.	Odds Ratio	z	*p*
Distance	-0.00101	0.00045	0.9990	-2.24	0.02524
Mean Daily Traffic Volume	0.00010	0.00003	1.0001	3.88	0.00010
Forest Area	-0.00056	0.00024	0.9994	-2.34	0.01905
(Intercept)	-2.96300	0.21748			

These results also partially supported the hypothesis that road characteristics influence *Phragmites* presence. Traffic positively influenced presence. For every 1,000

cars travelling on a road in a day, the likelihood of *Phragmites* presence increases by nearly 11% (Figure 2.7). This relationship aligned with expectations, agreeing with previous studies that examined traffic's influence on seed dispersal in invasive plant species (e.g., Lemke et al., 2018). Lane count was not a significant predictor of plant presence, contrary to expectations. It is possible the effect of lane size went unnoticed since nearly all culverts sampled were located on two-lane roads (Figure 2.5a).

Similarly, hypotheses about adjacent landcover were partially supported. Based on model results, a 100 m^2 increase in forested area within 30m results in about a 5% lower likelihood of finding *Phragmites* (Figure 2.7). *Phragmites* exhibits shade intolerance in microcosm settings (Kettenring & Whigham 2018; Li et al. 2011), which is likely the mechanism behind nearby forest cover lowering the likelihood of finding *Phragmites.* In contrast to expectations, developed area was not a predictor of presence. Maheu-Giroux & De Blois (2007) found much higher colonization rates in commercial areas and slightly higher colonization in industrial areas compared with others, especially as infestations became widespread. One possible explanation is an insufficiently sampled development gradient, given the relatively small buffer chosen and roads often being classified as the selected developed land cover types. It is also possible that the omission of culverts within 2 km of one another biased sampling against more highly developed areas.

Also of note is that agricultural areas were not significant predictors of presence. Like many invasive plants, *Phragmites* grows especially well in nitrogen-rich environments (Kettenring & Whigham, 2018) and increases in competitive ability under fertilized conditions (Uddin & Robinson, 2018). Thus, one would expect nutrient loads from adjacent agricultural drainage ditches to promote *Phragmites* growth. It is possible

that this is another issue stemming from the choice of scale, since one study measuring proximity in a continuous manner found both agriculture and nitrogen concentration to be influential when predicting presence (Carlson-Mazur et al., 2014). On the aggregate, proximity to agriculture was still less influential than proximity to developed land, road density, and topographic roughness (Carlson-Mazur et al., 2014). The relative importance of developed land compared to agricultural land is also supported by Maheu-Giroux & De Blois (2007), since they found that agricultural areas on the aggregate were less invaded than developed landcover types. With these two studies combined, it appears both factors are important, but their relative importance depends on relative abundances of each.

There are also factors not considered here that may help to explain the presence, introduction, and spread of *Phragmites* and the knotweed complex. Sensitivity to buffer size should be tested for agricultural and developed land covers to identify if larger buffers better represent dynamics occurring at larger scales. Broad habitat characteristics such as those modeled by eco-regions should be examined, since they have been shown to influence both invasive plant presence (Carlson-Mazur et al., 2014) and richness (Ward et al., 2020).

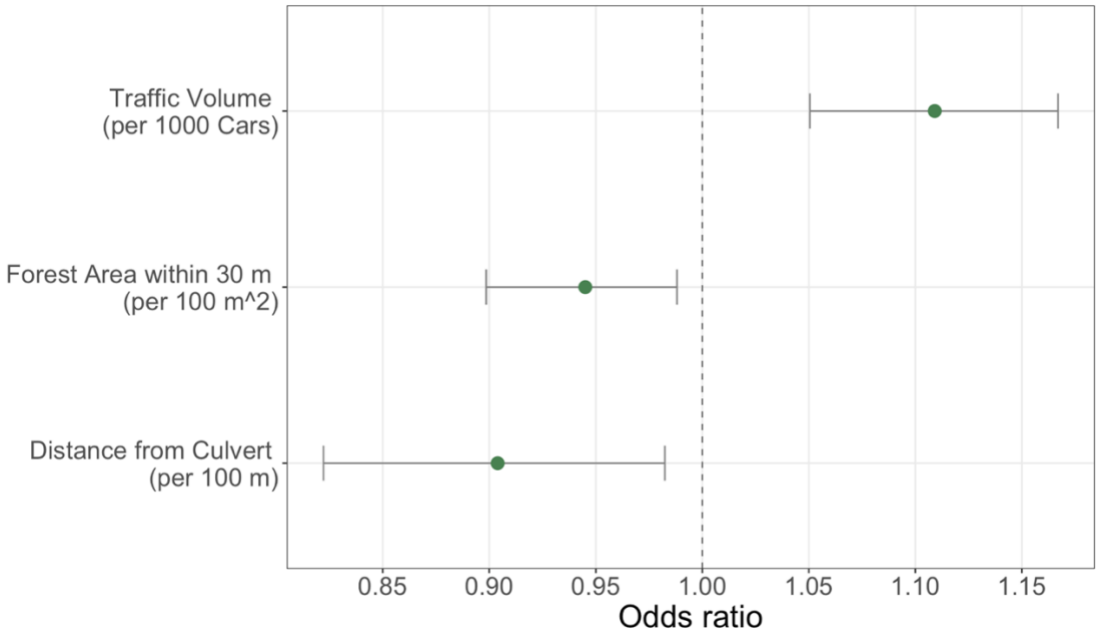

Figure 2.7 *Plot of odds for Phragmites presence based on a refined binary logistic model. Error bars represent 95% CI.*

2.3.3 Road Size Analysis

The hypothesis of larger roads increasing the likelihood of invasive plant presence was tentatively supported with *Phragmites*. Larger roads were expected to be predictive of presence since amenable conditions are provided by larger verges (Christen & Matlack, 2009) and larger roads have been found to greatly increase the likelihood of plant presence previously (Joly et al., 2011). *Phragmites* was much more likely to be found on ramps and somewhat more likely to be found on primary roads than expected, while it was less likely to be found on secondary and tertiary roads (table 2.3) (χ^2= 3480.3, df = 3, p-value < 0.0001). It is possible that large interchanges create ideal open areas for stands of *Phragmites* to establish. Further, with a relatively high density of impervious surfaces, water runoff and its associated salts (e.g., Denby et al., 2016) may accumulate, creating wet conditions that may enable *Phragmites* to outcompete other species more effectively given its salt tolerance (Albert et al., 2015;

Vasquez et al., 2005). These conditions may be especially pronounced at cloverleaf interchanges, where rings of roads surround depressions.

The larger road hypothesis was tentatively rejected with the knotweed complex. The knotweed complex was less likely to be found on primary roads and slightly less likely on tertiary roads than would be expected but was found on secondary roads and ramps more than expected (table 2.3) (χ^2= 62.505, df = 3, p-value = <0.0001). Again, it is important to note that all significance reported here is somewhat exaggerated given spatial autocorrelation (Legendre, 1993).

Table 2.3 *Chi-square results for raw frequency presences. Frequencies were calculated from the entire dataset of 94,219 points within 15 m of roads and inside 1 km culvert buffers. Absence frequencies are excluded from this table.*

	Phragmites			**Knotweed Complex**		
χ^2	3480.3			62.505		
df	3			3		
p-value	< 0.0001			<0.0001		
	Expected	Observed	Residual	Expected	Observed	Residual
Primary	472	620	6.81	135	62	-6.29
Secondary	1100	840	-8.01	317	389	4.06
Tertiary	618	267	-14.1	176	167	-0.736
Ramp	71.5	541	55.5	20.4	31	2.33

2.3.4 Insights for Monitoring and Risk of Spread Modeling

Considering the results presented here, there is potential to improve current risk modeling and targeting of monitoring efforts. In New York for example, managers have access to a spatial prioritization model developed by Shappell et al. (2016) that considers ecologically significant, protected areas, and risk of spread factors (NYNHP, 2016a). The component weighted the highest in the risk of spread component is a

landscape condition assessment, which models the ecological effects of various stressors including land use and development types (Feldmann & Howard, 2013). The model assumes that rates of road-parallel spread risk is equal for primary and secondary roads. There is reason to evaluate traffic volume as a potential addition into risk of spread modeling; results here suggest that traffic facilitates *Phragmites* spread, consistent with other studies of the same phenomenon with different species (e.g., Lemke et al., 2018). Road type also appears to play a role, though it seems to be species-specific. As such, it warrants study with additional species, but perhaps cannot be broadly applied to a general risk model. While culverts may not function as hotspots for all species, their association with increased *Phragmites* presence warrants consideration as an additional place to prioritize monitoring. This is especially prudent to consider since culverts connect road and stream networks together.

2.4 Conclusions

Here we presented an examination of road, road verge, and culvert characteristics in relation to invasive plant presence by applying a novel form of data generated from computer vision and roadside imagery. Using results from over 8,000 roadside panoramas across 127 culverts in 37 counties within New York State, we determined that traffic volume, forested area, and culvert presence significantly influence the likelihood of *Phragmites* presence. By adding an additional 80,000 panoramas, we determined that *Phragmites* was much more likely to be found on ramps and somewhat more likely to be found on primary roads, while the knotweed complex was more likely to be found on secondary roads and ramps.

Given the results of this study and others, it remains unclear precisely *why* culverts are acting as hotspots. Culvert age and construction footprint were not related to presence, and presence of stream feature does not seem to be indicative of *Phragmites* presence (Maheu-Giroux & de Blois, 2007). Therefore, a link between construction of culverts and introduction of *Phragmites* cannot be made conclusively. The dataset used here did not include model data from SVI revisits of the same location, though multiple dates of imagery may be obtained in the future to update the analysis. Increased temporal resolution of computer vision model data would help in identifying how culvert construction and introduction dates compare.

These results can still be used to improve existing spread risk modeling for more accurate estimation of the conditions influencing spread. Results also suggest that culverts and highway ramps are locations that warrant additional study, monitoring, and assessment of preventative measures to efficiently contain the spread of certain invasive plants.

The work here presented a case study of the analytical possibilities of using computer vision data from roadside imagery. There are myriad applications for this data, given the wide spatial and temporal scales available for study. These future applications could include the direct examination of spread over time, or abundance mapping on the road network at large.

Chapter 3: Computer Vision Monitoring Framework

3.1 Introduction

3.1.1 Invasive Plant Monitoring

In the US, invasive plants have led to documented damages of over $25 billion per year, while causing estimated losses on agricultural and rangelands of over $33 billion annually (Pimentel et al., 2005). To mitigate or prevent harm from these plants, identification and control is crucial (National Invasive Species Council, 2003; 2016). A widely recommended framework for invasive species management is early detection and rapid response (EDRR), a part of which is constant monitoring for new infestations (Reaser et al., 2020). Unfortunately, monitoring is expensive and time-consuming. New applications of technologies such as computer vision promise to solve both problems, though they pose their own challenges with the generation of massive amounts of data. Currently, professional field survey, aerial or satellite-based remote sensing, and community science are the primary monitoring methods used, but each come with their own limitations.

Professional field study of invasive plants is a common method of both monitoring and research. Notably, as of September 2011, field observational studies comprised nearly half of all studies on invasive species (Lowry et al., 2012). Field methods targeted for use by professionals, such as the New York Rapid Assessment Model (NYRAM) developed by the New York Natural Heritage Program (NYNHP) require both plot- and transect-based surveys of invasive plant abundance and cover (Shappel et al., 2016). Compared to surveying on foot, car surveys are a method that can be used to monitor invasive plants more quickly, easily, and cost effectively (e.g., Catry et al., 2015; Shuster et al., 2005). However, both methods not only require time of professionals for travel and survey, but also monetary compensation and lodging costs.

Community or citizen science is an approach that leverages non-professionals to enable low-cost data collection at both large and site-specific scales (Larson et al., 2020). Community science programs are numerous and increasing (Bois et al., 2011). In NY, the NYNHP's iMapInvasives Program exists to facilitate data management between professionals and community members to inform management decisions. Data from these programs are known to vary in quality due to effort or knowledge (Crall et al., 2011) and in spatial distribution due to over-sampling of residential areas (Dickinson et al., 2010). The iMapInvasives Program sources many records from professionals, and non-professional record submissions are professionally verified to address data quality (NYNHP, 2020). Total reports from community scientists likely still underestimate the true number of invasive plants, since volunteers are much more prone to false negatives than false positives (Crall et al., 2011). Even with programs such as iMapInvasives that source from both professionals and community scientists, many times these professional surveys are only occurring where funded projects are ongoing, so underestimation and spatially heterogenous reporting are likely to remain problematic.

Aerial and satellite-based remote sensing can address the heterogeneity problem of professional field surveys to a certain extent while also eliminating the community science sampling bias. Widely available multispectral satellite imagery such as Landsat-8 or Sentinel-2 is often not of high enough spatial resolution to detect emerging patches of plants (e.g., Olsson et al., 2011). Even freely available sub-meter aerial multispectral data such as from the National Agricultural Imagery Program (NAIP) can have an insufficient resolution. Porter (2021) performed supervised classification on both NAIP and high-resolution drone imagery, finding that patches of their target grassland invasive plant frequently went undetected in NAIP. Widely available Hyperspectral imagery such as Hyperion is better for discerning invasive plants from other species but is currently spatially insufficient for early detection (e.g., Ramsey et al., 2005). Hyperspectral mages can be captured at higher spatial resolutions, but this is impractical for large areas due to cost (Vaz et al., 2018). With all these platforms, understory detection remains a challenge regardless of spatial or spectral resolution.

3.1.2 Leveraging New Technologies for Monitoring

Given the cost, scale, technical, or labor issues present with current monitoring methods, alternative solutions are needed. One solution can be found in a novel approach that combines computer vision, ground-level imagery of roadsides, and community science. Computer vision is a discipline that uses algorithms to enable a computer to identify objects (e.g., plants) in an image (Huang, 1996). Though several companies have their own roadside imagery, Google alone has taken over ten million miles (over 16 million km) of 360-degree panoramic Street View Imagery (SVI)

(Escobar, 2019). Previously, SVI has been used successfully when assessing the feasibility of “virtual field studies” (e.g., Kotowska et al. 2021; Deus et al. 2016; Griew et al 2013; Rousselet et al. 2013) and unsuccessfully for invasive plant identification using unsupervised classification methods (Connell, 2015). Computer vision has been used for multiple applications in an ecological context, ranging from identifying wildlife in photos from many sources (Berger-Wolf et al., 2017) to inventorying urban trees present in SVI (Wegner et al., 2016).

A proof-of-concept conducted by Flores et al. (in prep.) applied computer vision and deep learning on SVI for invasive plant identification. In that study, a computer vision model was created for five case study plants: the invasive knotweed complex (*Reynoutria japonica*, *R. sachalinensis*, and *R. x bohemica*), *Phragmites australis ssp. australis* (common reed, hereafter referred to as *Phragmites*), *Lythrum salicaria* (purple loosestrife), *Ailanthus altissima* (tree-of-heaven), and *Pastinaca sativa* (wild parsnip). They demonstrate the large spatial scales of monitoring (including of the understory) possible when leveraging computer vision and SVI. At maximized sensitivity, the overall false negative rates were below 6% for both *Phragmites* and the knotweed complex, but had corresponding overall false discovery rates (i.e., proportion of panoramas incorrectly classified as presences) of 60% for knotweed and 39% for *Phragmites* (Flores et al., in prep.). However, field verification of screen-validated positive detections yielded very high agreement (>90%), where false positives were mostly (69%) attributed to mowing (Flores et al., in prep.). These outcomes suggest that on-screen validation may be required for positive confirmation of records, but field verification, which is far more time and resource intensive, may not necessary. Thus, to achieve highly accurate

detection, positive model outcomes must be combined with on-screen validation of panoramas in order to remove false positives.

On-screen validation is beneficial given concerns with safety and legality. Many model-generated predictions are along highways, where cars pass by at speeds of 90-105 kph (N.Y. Comp. Codes R. & Regs. tit. 21 § 103.2; N.Y. VAT tit. 7, art. 30, §1180b). Additionally, there were legal concerns both with stopping adjacent to private property and along interstate and state highways, where vehicle stopping and standing are prohibited (N.Y. VAT tit. 7, art. 32, §1202).

The verification of predictions derived from computer vision is one of many projects for which community scientists can volunteer. Though many projects utilize physical on-site verification, there are numerous examples of exclusively virtual community science projects. For example, one project existed to facilitate validation of land classification via satellite images (Brovelli et al. 2015 as reviewed in Fritz et al. 2017), while another similar project aimed at a different audience incorporated change detection scenes (Fritz et al. 2017).

Because these computer vision models result in a massive amount of new data at very fine resolutions that must be properly managed, verified, and used, a new system was developed to harness the distributed power of the community science network and more rapidly put results into the hands of managers. We created this pipeline by addressing the following objectives:

1. create a framework and set of recommendations for integrating AI into existing monitoring workflows;
2. create public GIS layers of all species predictions;

3. develop method and interface for validation; and
4. disseminate results to stakeholders.

Throughout monthly discussions over an 18-month period with PRISM managers, community science program staff, and invasive species experts at the NYNHP, a prioritization and verification framework was developed using *Phragmites* and the invasive knotweed complex as case study plants. Within this framework, two web-based ArcGIS Dashboard interfaces were created to present results and encourage community scientist participation (Figure 3.1). Prioritization analyses were embedded in these interfaces to help identify under-reported areas and guide human verification effort. To make results useable by administrators, managers, and community scientists, we created maintenance, access, and tutorial documents. We also held multiple webinars to make managers aware of the newly created tools at their disposal. Upon completion, all data products, documentation, and both interfaces were sent to be hosted by the NYNHP.

The proposed monitoring workflow process (Figure 3.1 Items 1-4) can be conceptualized in multiple “tiers” of review. First, the computer vision model reviews all panoramas, resulting in a subset of predictions that are “leads in need of on-screen validation” by community scientists. Next, community scientists verify these presences and submit an iMapInvasives record. If the species is not found, a “not observed” record is submitted and immediately incorporated into the iMapInvasives database. If the community scientist believes the record *is* found, a presence record is submitted for confirmation or rejection by the preexisting iMapInvasives network’s taxonomic experts

(New York iMapInvasives, n.d.) before being integrated into the iMapInvasives database.

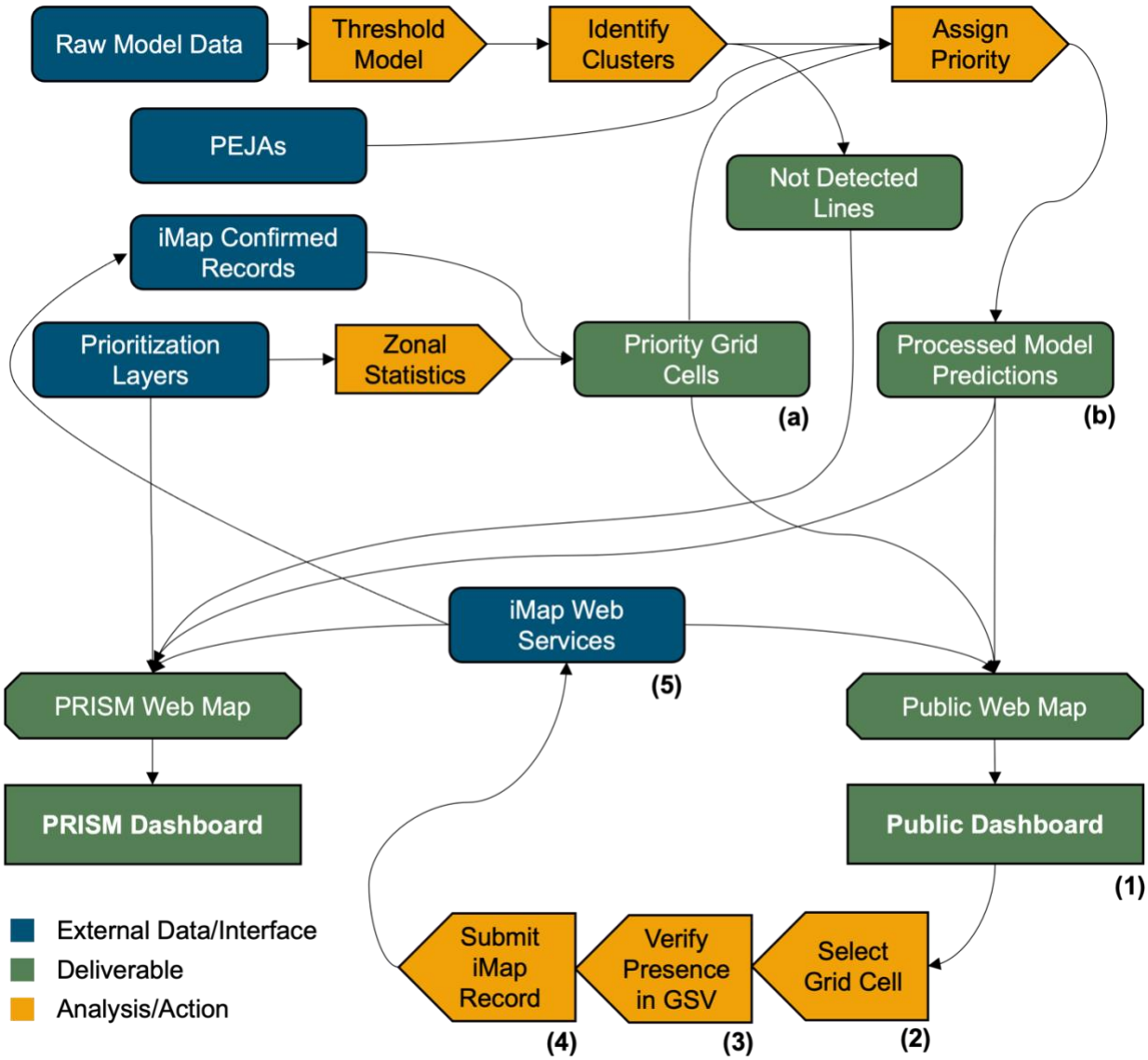

Figure 3.1 *An overview of the constituent components and flow of data for the dashboard interfaces and associated facilitation products. Priority grid cells (a) and Potential Environmental Justice Areas (PEJAs) are a prerequisite for model processing (b), and are also displayed directly in the Public Dashboard (1). In the Public Dashboard, users can select a grid cell for review (2), verify the presence in Google Street View (GSV) (3), and then submit a record to the iMapInvasives database (4). Data are hosted and visible as "unconfirmed" records that are then confirmed or denied by professionals (5). Whether unconfirmed or confirmed, these records are visible in the Partnership for Regional Invasive Species Management (PRISM) Dashboard.*

3.2 Interface Development

3.2.1 Geospatial Data Acquisition and Processing

Computer vision model prediction point data for *Phragmites* and the invasive knotweed complex were provided by Flores et al. (in prep.) then pre-processed via thresholding. These data included results for over 300,000 panoramas in Broome County, NY (Figure 3.2) captured between August 2011 and November 2018, at a median geodesic distance of 2.4 m between panorama centers. Broome County was chosen as a candidate region for these analyses given the varied land covers, road types, ecologically significant, and potential environmental justice areas present. In broad terms, the computer vision model "looks for" patterns recognizable based on training data, then calculates a confidence score for each species it was trained on. The output data can then be accessed in a tabular format where each row contains panorama metadata and confidence scores for all species. Model output was processed by thresholding these confidence scores, where the "threshold" is a selected range of probability values that are chosen based on desired performance criteria. This performance criteria includes true positive (presence) and negative (absence) rates, and/or false positive and negative rates. A useful guide for performance criteria is a set that facilitates early detection and rapid response (EDRR) approach, widely considered a core tenet of invasive species management (U.S. Department of the Interior, 2016). To this end, it is appropriate to "cast a wide net" by choosing a threshold that maximizes sensitivity level (i.e., the highest detection rate) across multiple test datasets. The threshold chosen for *Phragmites* was 0.45 while for the knotweed complex it was 0.12, using the model version "gentle-snowball-307." This criterion presents a maximally sensitive set of results that could then

be filtered to higher confidence scores associated with lower false positive rates. This choice also allowed higher confidence scores to be set for community scientists, proposed in stakeholder meetings to potentially help with volunteer confidence and retention.

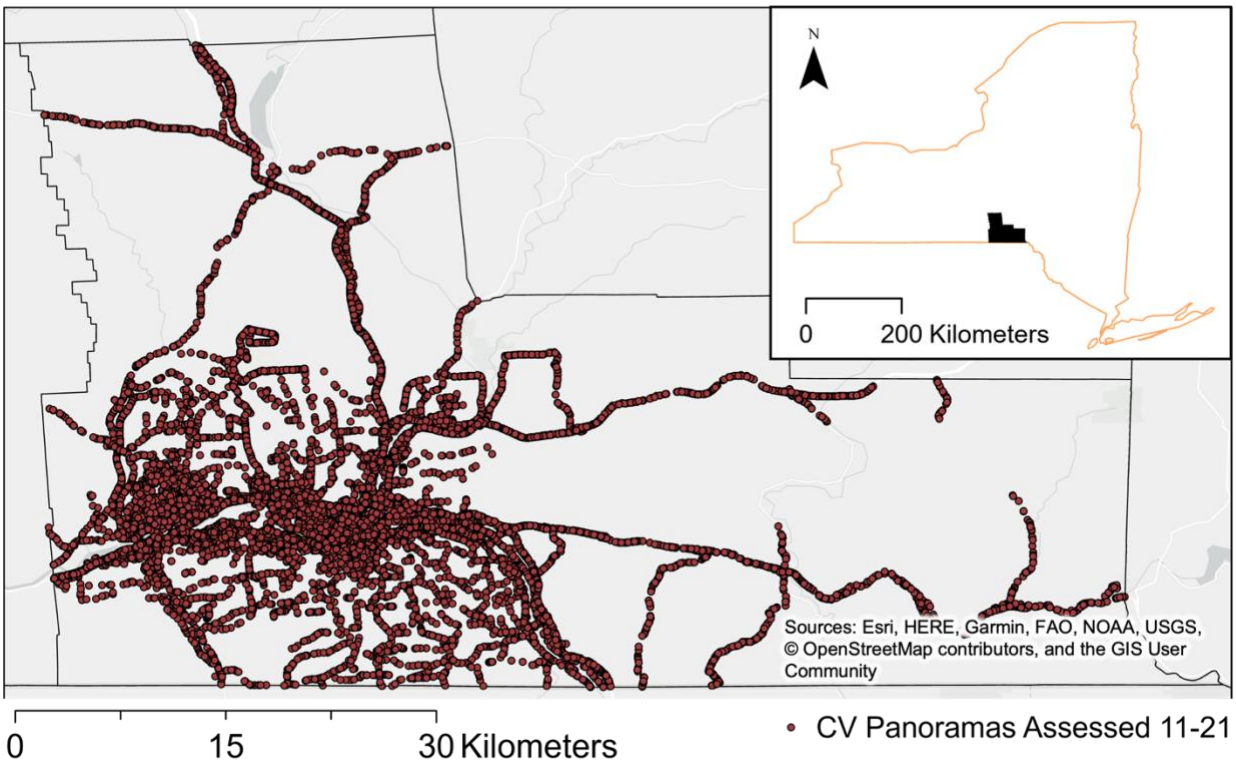

Figure 3.2 *Map of Google Street View panoramas in Broome County, New York assessed by the computer vision model from Flores et al. (in prep.) and used for analysis.*

After thresholding, clustering was a prerequisite for assigning spatial priority to presence predictions. Cluster assignment was completed via the DBSCAN method in ArcGIS Pro v2.5.2 (Esri, 2020). The minimum cluster number was set to two, while a maximum distance of 12 m between predictions defined a cluster. This distance is the 95th percentile of geodesic separation distance and can be conceptualized as the threshold for what is considered "contiguous." Percentiles were derived using Excel v16.59 with the output of the Near tool in ArcGIS Pro. The 95th percentile was chosen

instead of the maximum since some panoramas were up to 60 m apart, either due to being isolated “photospheres” or gaps in imagery data. The Near tool was used again to assign the linear ID of the closest US Census roads, to help separate out clusters.

Model predictions *below* the thresholds described above are considered “not detected” records. These points were selected for each species, then buffered by the same geodesic distance as clusters (12 m). The resulting buffers were then intersected with the US Census roads to output roads where a specific species was not detected by the model.

Several data products were accessed and displayed directly on the ArcGIS Dashboards. The NYNHP spatial prioritization layer was accessed as an ArcGIS Online map service from the NYS Department of Environmental Conservation. Confirmed, unconfirmed, and not-detected layers for *Phragmites* and the knotweed complex were accessed as ArcGIS Online map service from iMapInvasives. All data products sourced are described in greater detail in Table S1.

3.2.2 Spatial Prioritization

Stakeholder consensus was used as the primary basis for determining priorities. Through these discussions, three main priorities were decided upon. The first was that areas with high scores on the NYNHP spatial prioritization model should be ranked with higher importance. This spatial prioritization model was developed by Shappell et al. (2016) and includes three factors that are aggregated into a comprehensive score: ecologically significant areas, high priority preservation areas, and risk of spread (NYNHP, 2016a). The second priority was that model predictions in areas with data gaps should be verified first since presence or absence in those areas is unknown. The final

priority was that smaller, more recent predictions should be investigated first, since they would be expected to be easier to eradicate if able to do so. To help achieve these goals, a two-dimensional binary classification (i.e., four categories) was devised (Table 3.1).

To reach the first prioritization goal, the spatial priority dimension of a point prioritization framework was partially prepared by selecting grid cells. Grids for UTM zones 17N and 18N at a 1 km^2 size were sourced from the National Geospatial Agency website. The invasive species spatial prioritization model created by the NYNHP in 2016 was directly provided by the NYNHP. Zonal statistics were completed in ArcGIS Pro v2.5.2 (Esri, 2020) on an integer version of the NYNHP comprehensive score raster to assign the mean comprehensive score to each 1 km^2 grid cell. Then, the 90^{th} percentile of the grid cells' mean comprehensive score was identified for each PRISM jurisdiction in Microsoft Excel v16.59. Any cells containing model data that were below the 90^{th} percentile of the mean comprehensive score in a specific PRISM jurisdiction were excluded.

To reach the second goal, a second component of the spatial priority dimension was considered by removing areas with existing reports. All species presence reports were downloaded from the iMapInvasives website on November 6, 2021 for four knotweeds and *Phragmites australis ssp. australis*. The knotweed complex (*R. japonica*, *R. sachalinensis*, their hybrid *R. x bohemica, and "knotweed unspecified"*) was chosen for two reasons: morphologic similarities exist such that the model does not differentiate between them, and for management purposes they are treated the same. Grid cells were removed from the set if the species predicted by the model exactly matched existing iMapInvasives reports in the cell.

A third component was added when assigning the spatial priority dimension of the point prioritization framework to incorporate social and economic considerations. Potential environmental justice areas (PEJAs) as of 2021 were sourced from the NYS Department of Environmental Conservation website. PEJAs are US Census block groups defined as having a population with 22.82% below the federal poverty level, and/or are comprised of 52.42% (urban) or 26.28% (rural) minoritized groups (New York State Department of Environmental Conservation, 2021a). Any model-predicted points within PEJAs or cells above the 90th percentile of mean comprehensive score were considered to have the first dimension of spatial high priority. Any points that did not match these criteria were spatial low priority.

To reach the third and final prioritization goal, the “emerging” dimension of the point prioritization was assigned. This dimension was created to help with rapid response. The first condition to be considered “emerging” was that the model-predicted presences were on a panorama from within the past three growing seasons. The second was that the model-predicted presence must not belong to a cluster, which acts as a proxy for infestation size. If both conditions are met, a prediction is considered “emerging.” If not, the prediction is considered “not emerging.” The combination of all of these designations can be seen in Table 3.1.

Finally, for additional filtering functionality, model predictions had attributes added. Taking the output for each species/complex separately, the geodesic distance to the nearest iMapInvasives record (at the time of analysis) of the same target species/complex was calculated for every presence prediction.

Table 3.1 *Mutually exclusive decision criteria for assigning priority levels to model predictions. The mean comprehensive score refers to the numeric spatial prioritization model. All priority levels were included in the manager dashboard, but only priorities 1-3 were displayed in the public dashboard.*

	Single presence, w/in last 3 growing seasons	Cluster, and/or older than 3 growing seasons
1 km^2 grid's mean comprehensive score >=90th percentile for PRISM jurisdiction **or** within PEJA	Emerging: Spatial high priority (P1)	Not emerging: Spatial high priority (P3)
1 km^2 grid's mean comprehensive score <90th percentile for PRISM jurisdiction	Emerging: Spatial low priority (P2)	Not emerging: Spatial low priority (P4)

3.2.3 Public Interface

An ArcGIS Dashboard interface was developed (Figure 3.3) to include individual species layers with the goal of targeting citizen scientist verification efforts to high-priority computer vision model predictions. Predictions in designated priority levels 1-3 (Table 3.1) were included in the public interface. Confirmed, unconfirmed, and not-detected layers filtered on project submissions for the two target species are also displayed. Any predictions on private roads (i.e., where the US Census-designated field MTFCC = S1740) were removed for privacy concerns.

Gamification was implemented when creating this Dashboard to encourage greater participation. Deterding et al. (2011) define gamification as the "use of game design elements in non-game contexts." Gamification has been implemented previously through displaying leaderboards in online land cover validation and change detection interfaces (Fritz et al. 2017). Another example comes from the iNaturalist program (https://www.inaturalist.org), which presents leaderboards for picture submissions of macro-organisms while simultaneously facilitating biodiversity research. The companion

of iNaturalist, Seek, adds challenges and badges to the submission process (iNaturalist, 2020).

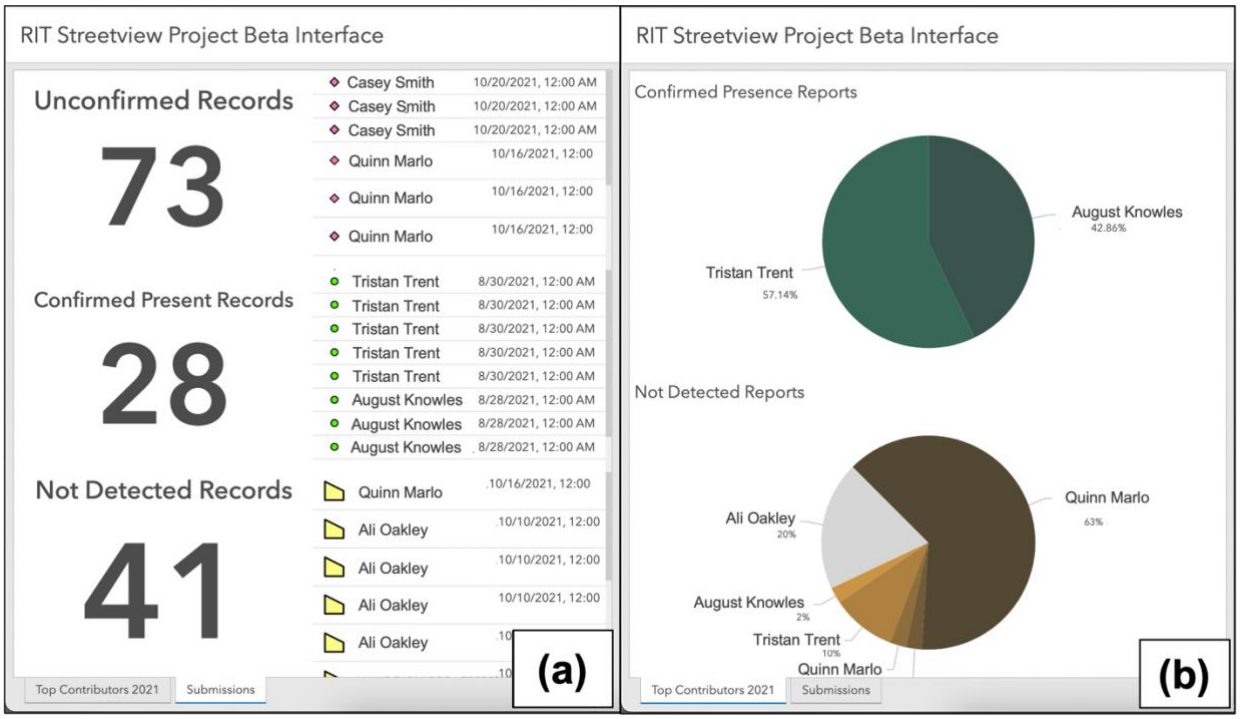

Figure 3.3 *Example showing gamified components of the Public Dashboard (with names changed). In one pane, submission tallies and recent submissions per-type are shown (a). Another pane (b) shows charts of top contributors for each category.*

In this Dashboard, gamified elements include top contributor charts and recent contributor lists. Community scientists can filter submitted records by PRISM jurisdiction or for the whole state, and by target species. Records displayed on the map, tallies, charts, and lists all update accordingly. These features enable community scientists to see the magnitude of their contribution and how it compares with others.

The 1 km^2 grid cells described in section 3.2.2 were displayed on the public interface to enable community scientists to sign-up for a verification area. Grid cells can have three statuses: unverified, in progress, and verified. Community scientists can browse the map and filter grid cells based on completion status, reducing irrelevant

information. A web based Survey123 form was created in ArcGIS Survey123 Connect v3.13.251 (Esri, 2021) and linked within grid cell pop-ups on the dashboard. When a volunteer would like to sign up for a cell, they would mark it as “in progress” on the form.

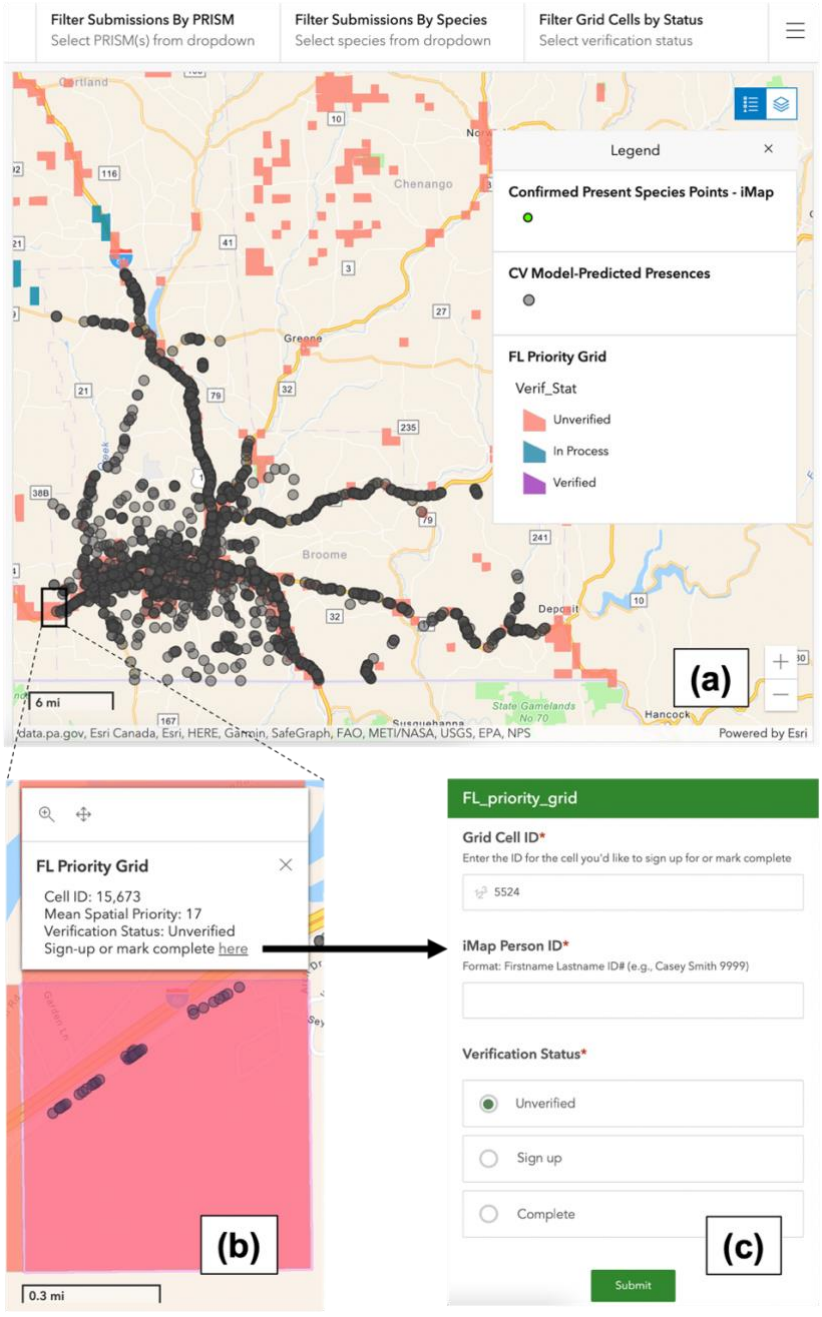

Figure 3.4 *Example of Public Dashboard grid selection process.* The map view (a) shows *verification grid cells* and *model presence* predictions and a top ribbon with *spatial and attribute filters. Individual grid cells (b) link to a form (c) for signing up and marking complete.*

After signing up for a cell, individuals verify model predictions *within* the cell and create iMapInvasives records. Individuals can access individual model prediction points, then view linked panoramas with model-predicted presences from a link in a pop-up. If a prediction is thought to be correct by the reviewer, they enter a “remotely sensed” presence report with a screenshot of the panorama onto the iMapInvasives web submission form. If the reviewer believes the prediction is false, the same process can be followed, where a not-detected record is submitted instead of a presence. Once all predictions in a cell are completed, the community scientist marks a cell as verified, locking its status from public editing. Every record submission is hosted and visible as an “unconfirmed” record that can then be confirmed or denied by professionals.

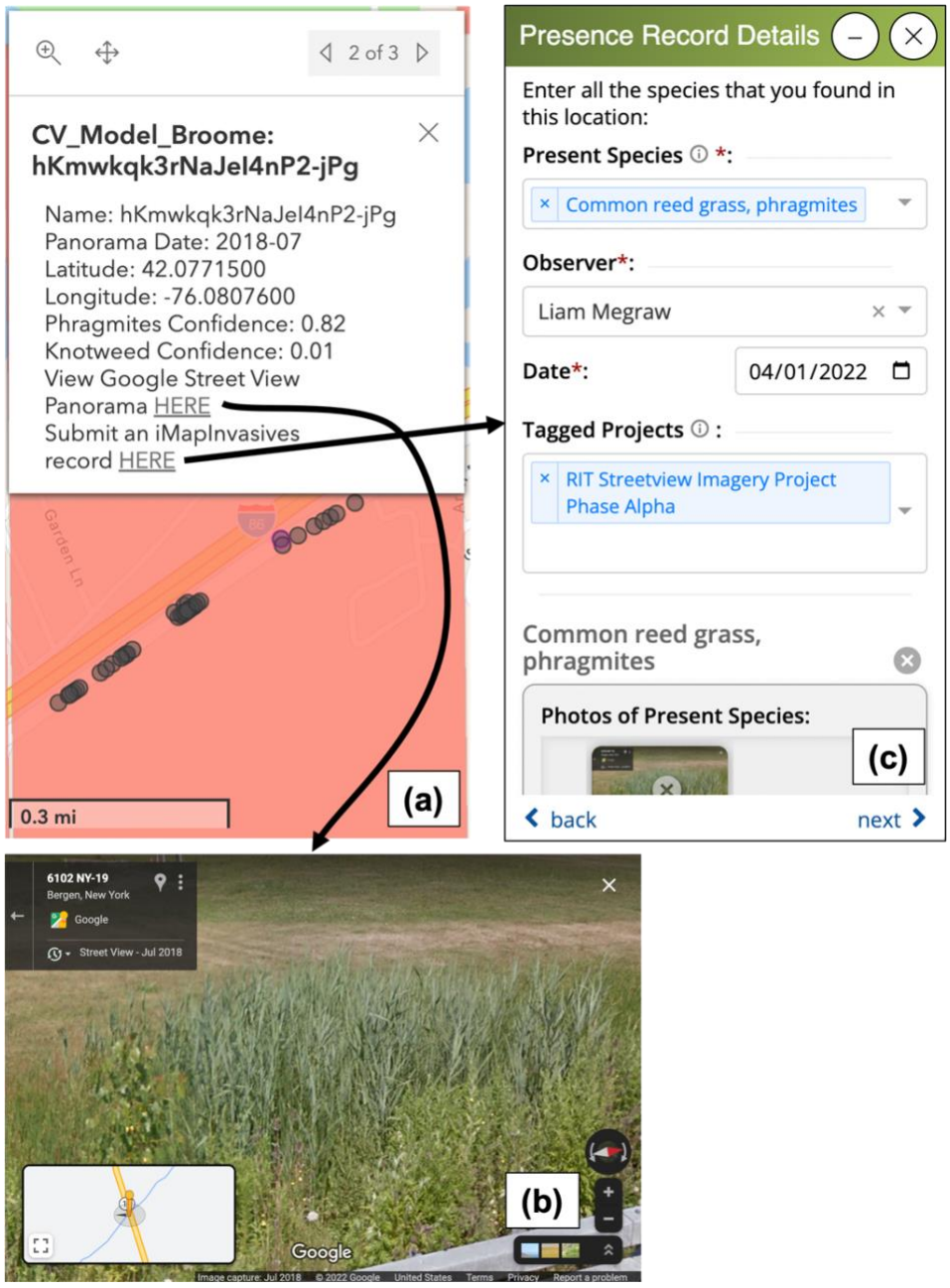

Figure 3.5 *Example of Public Dashboard review and verification process. After signing up for a grid cell, users select individual points (a) to visually examine the associated panorama (b). After review, the iMapInvasives link can be followed to submit a record (c).*

3.2.4 Manager Interface and Data Products

A second ArcGIS Online Dashboard interface was built for PRISM staff to view records submitted by community scientists, to help prioritize areas of intervention, and to highlight areas with a surveillance need.

The Dashboard interface has three main components: the top ribbon, the left sidebar with tabs, and the main map (Figure 3.4). On the map, model predictions are displayed as a heatmap at lower zoom levels, transitioning to color-coded points at higher zoom levels. Model-predicted absences are also displayed as contrasting orange lines. The first tab on the left sidebar allows managers to see tallies within indicators of model predictions broken down by the four priority levels. The second tab on the left sidebar enables managers to view summaries of record submissions, similarly to the public interface. The top bar allows managers to filter by various attributes, including management area (PRISM jurisdiction) boundaries, target species, and distance between existing reports and model-predicted presences. When filtered, model-predicted presences, model-predicted absences, and iMapInvasives records all update accordingly in the various dashboard elements.

Additional components in the top-right let managers select their desired base map, show the legend, or set layer visibility. Within layer visibility, managers have the option to show the NYNHP spatial prioritization maps, including comprehensive score, risk of spread, protected areas, and areas of ecological significance. Since all model data are displayed, managers can use these additional layers to assess the importance of individual predictions with deeper information than the prediction priority levels provide alone.

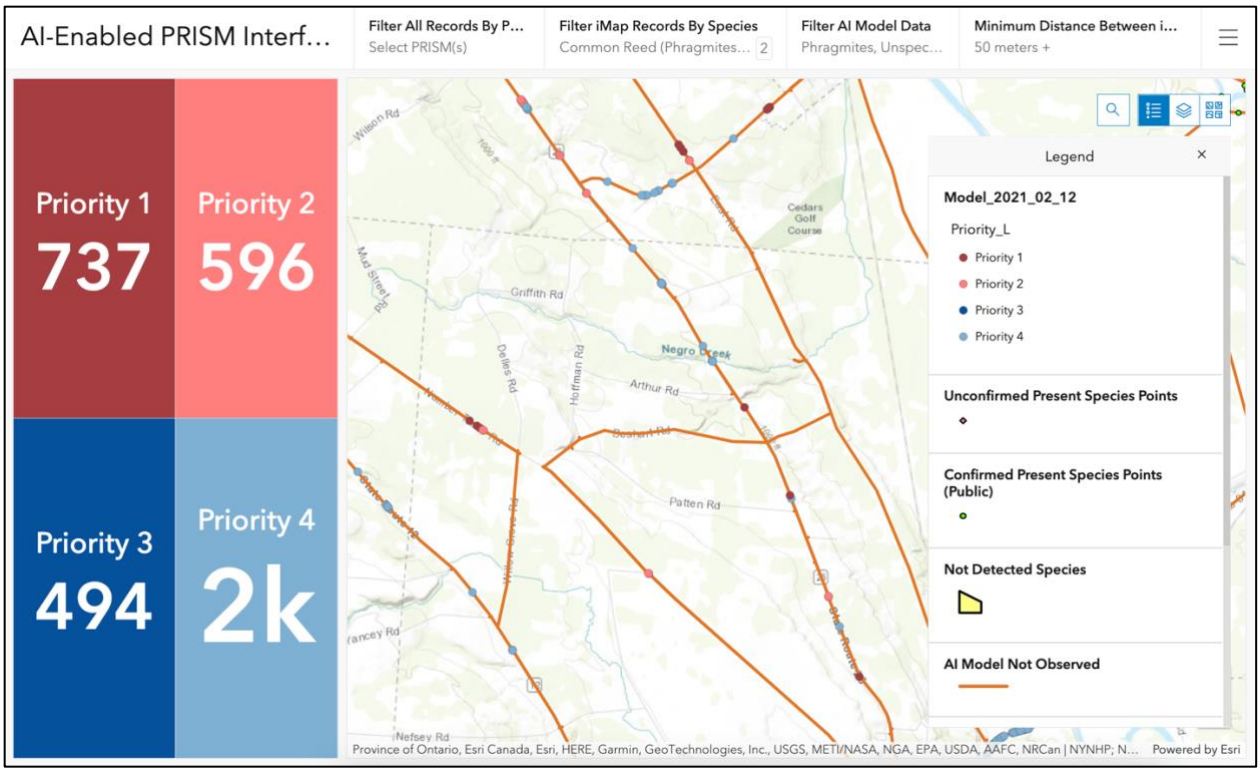

Figure 3.6 *Example of the manager Dashboard showing model prediction points, tallies on the left pane, filters on the top ribbon, and legend pop-up on the right.*

3.3 Limitations and Future Work

3.3.1 Spatial Prioritization

The multi-leveled spatial prioritization framework proposed here offers managers an at-a-glance look that encompasses multiple dimensions simultaneously. Given the inability to manage *all* invasive species, prioritization frameworks such as the one here are necessary. Myriad considerations are built into the prioritization framework, including spread risk, ecological significance, social dimensions, infestation size, and recentness of discovery.

It is notable though that the emerging dimension of the spatial prioritization scheme could warrant further examination. There is potential to adjust the timeframe of what is considered “recent” for the purposes of the “emerging” definition, in line with the

conceptual relationship of spread and cost (Figure 1.1). As built, the “emerging” definition dictates that computer vision model-predicted presences must be from the past three growing seasons. This choice was made to highlight locations where “rapid” response could be possibxle. However, it is important to note that what is considered “rapid” for invasive species detection is often not defined explicitly in literature (Reaser et al., 2020), and in a case where “rapid” was defined, the recommendation was within two years (Lodge, 2006), though that recommendation is for *eradication* of a species newly introduced on a national scale. Therefore, further investigation into risk and spread could be useful to help define what a “rapid” timeframe might be for management response.

The approach of using the NYNHP comprehensive score model when determining high-priority grid cells warrants review. The current model (Shappell et al., 2016) does not differ in risk of spread between primary and secondary roads. A study from Joly et al., (2011) suggests that primary roads may have faster spread rates, as do roads with higher traffic volumes (Lemke et al., 2019). Additionally, managers noted in discussion that even though they are aware of the prioritization model, it is often not utilized because of site-specific project funding or not being optimally calibrated for large, highly important areas such as the Adirondack Park.

3.3.2 Public Interface

The Public Dashboard developed enables large amounts of new data to be verified in an organized, prioritized way. By integrating with existing tools and platforms, the workflow proposed here is an extension of current capacity that strengthens early

detection and ongoing monitoring capabilities. Simultaneously, the verification workflow supported by the Dashboard makes community science accessible to more people that may not have the ability or transportation to physically participate in other community science efforts.

There are several opportunities for refinement of the Public Dashboard. Within the workflow that community scientists complete, the steps can be somewhat cumbersome. The user must manually navigate to the iMapInvasives web interface and enter the same information many times, including the observation method, project, and date of panorama. Auto-generated URLs or another method to directly link and pre-fill iMapInvasives form components would save time and streamline the verification workflow.

The grid cell sign-up process could be completely removed if spatial filtering was incorporated for model-predicted presences. As of the Dashboard's creation, the grid cell sign-up process was enacted to visually show where community science verification effort is or will be occurring. With spatial filtering functionality, model predictions would be filtered out almost instantaneously upon a record being submitted sufficiently close to it. The benefit of this proposed modification is that it would save community scientists time and ensure the map is up to date independent of user input.

3.3.3 Manager Interface

Widespread monitoring is a vital aspect of invasive species control (Reaser et al., 2020). This interface enables both ongoing and early monitoring of invasive plants at multiple scales and use-cases. Consider a conservation or natural resources professional working on the ground at one to several sites. This professional would have a scope of work where a finer scale of data is preferable. The professional might want to know if

there are any previously un-discovered invasive plant presences either within their site or in its buffer. In this scenario, a lower model confidence threshold and short distance to existing reports could be used, and all priority levels could be of interest. On a larger scale, consider a professional coordinating monitoring and control efforts across multiple counties. At this scale, there are invasive plants seemingly *everywhere*, so it is beneficial to have prioritization options. The coordinator might be interested in asset-based protection, prioritizing funding and efforts towards highly confident, previously undetected presences that would threaten areas of high ecological significance. To this end, stricter filtering could be used to locate areas in need of survey. These cases are but two examples of possible applications and questions that can be answered when using this Dashboard, though there are likely many other use cases that vary in both scale and goal.

The Manager Dashboard does have multiple limitations that could be improved upon. Perhaps the largest limitation is the difficulty of accessing roadside imagery for generating results as new panoramas are captured, given both the technical knowledge needed access imagery and the paywall for its acquisition. Additional filters could be beneficial for managers, especially for computer vision model attributes such as detected or not detected date. Currently if a manager wants to identify gaps in current reporting, the distance filter is the only item that can be used. A more explicit visualization of reporting gaps could be useful, which could take the form of a comparison between existing reports and model-predicted reports in a gridded format.

3.4 Conclusions

Here we presented a framework for incorporating computer vision into current invasive plant monitoring workflows in a cost- and time-effective manner. An ArcGIS Dashboard created for use by community scientists acts a platform to enable the verification of computer vision presence predictions. Map products created here distill a massive amount of data into a comprehensive format that aids in the prioritization of both professional and community scientist monitoring efforts. Further, we show how computer vision and Street View Imagery can greatly enhance the scale of current surveillance capacity. A second ArcGIS Dashboard displays these data in a way that is applicable to managers of varying scopes and areas of work within terrestrial invasive plant management. This enables the monitoring of very large areas of roadsides, a known vector for invasive plant spread, with little investment of time for field monitoring and allows for prioritization of regional efforts towards control and eradication.

Even though the two target plants of this study are already widespread at regional scales, there is value in their continued monitoring to assess continued spread into new regions with a changing climate and the efficacy of control efforts. Multiple recaptures of SVI allow for study before and after roadside treatment efforts, whether mechanical, chemical, or biological. Notably, a psyllid has been approved for release to biologically control the knotweed complex (New York Invasive Species Research Institute, 2020), while *Phragmites* has a biocontrol agent in review (Blossey, et al. 2019). By leveraging data from before and after the release of these biocontrol agents, their impacts and spread may be quantifiable, dependent on spatio-temporal overlap of SVI. There is even potential for expansion of this framework, since two of the other species (*Ailanthus*

altissima and *Lythrum salicaria)* that Flores et al. (in prep.) developed a computer vision model for have their own biocontrol agents either disseminated or in review.

The applications developed here act as a proof-of-concept that can be expanded to other species. Flores et al. (in prep.) show that while excellent identification performance for *Phragmites* and the knotweed complex required thousands of training and testing panoramas to be annotated, a useful model can be created with only hundreds. Given the relatively short annotation time, it is entirely feasible to create computer vision models for additional plant species. If a species is not widely present in a certain geography but is present in an area with similar image background characteristics (e.g., a similar ecoregion), there is potential for model creation. An important caveat to note is that while SVI vastly increases the spatial and temporal scales of monitoring that is possible, the data are not homogenous so *comprehensive* detection especially for emerging species is not possible. When complete, additional species can be incorporated to the interfaces and verification workflow easily, simply by updating filters. Any analyses presented here can be replicated with ease for additional species or PRISMs, given that the data utilized here is publicly accessible or already hosted by NYS. There is even potential to expand the geographic scope of this framework to other states in the US or countries, by replicating the spatial prioritization work by Shappell et al. (2016) and the New York Natural Heritage Program (2016a), the computer vision model by Flores et al. (in prep.), and the environmental justice designations of the NYS DEC (New York State Department of Environmental Conservation, 2021a).

Chapter 4: Conclusions

4. Conclusions

Through this study, we showed ecological and management applications of a novel form of data. Computer vision combined with deep learning and roadside imagery greatly enhances our ability to monitor and gain valuable insights from invasive species at wide spatial and temporal scales. This holds true despite the spatial and temporal heterogeneity of roadside imagery.

Results presented here help to target interventions, enabling more efficient invasive species management efforts. We used computer vision generated results for thousands of panoramas to examine characteristics influencing spread and found that culverts and traffic are two meaningful additional factors to consider for future risk modeling, monitoring, and study. Combined with previous study on road-stream connections (e.g., Maheu-Giroux & de Blois, 2007), there is reason to believe culverts are acting as a connection between road and riparian spread. Highway ramps and primary roads may also be facilitating spread, given the higher rates of *Phragmites* occurrence. Therefore, targeting safe control efforts in these areas may be ideal to slow the spread of *Phragmites*. For the knotweed complex, examining why it appears to be present more frequently on secondary roads and ramps would be worthwhile.

We also showed one way to make computer vision generated data accessible and useful to managers through the creation of an organizing framework. This framework condenses vast amounts of data into a more interpretable format that weaves into and augments existing monitoring capacities. The interfaces developed from this project and its partners enable an entirely new, powerful set of hundreds of thousands of datapoints to be made useful by managers. Further, these interfaces spatially prioritize computer

vision results, allow for various scales of work, and give managers customized control of how to view data for decision-making.

While roadside imagery can address limitations and biases of other monitoring methods, it is important to acknowledge the limitations and biases of roadside imagery itself. Google's SVI is captured in a manner that is heterogeneous in spatial and temporal distribution. While there are revisits of the same location, not all locations are revisited with the same frequency. For example, the Adirondack Park, a region with large natural protected areas in northeastern NY, has numerically fewer years of imagery than urban areas such as Buffalo, Rochester, or Albany (Flores et al., in prep.). When there are revisits, these revisits may also be at inconsistent intervals and/or seasonally biased (Flores et al., in prep). There are even roads without SVI captured, easily visible in Google's web map interface (Google, n.d. a). Together, these attributes cause spatial and temporal gaps to be present in this data source. Further, because Google publishes revisit locations at a county resolution and only for the current year (Google, n.d. b), it may be hard to plan how to use SVI for specific, targeted locations. In this sense, the data that can be gleaned from SVI is somewhat "opportunistic" in nature.

When looking towards future applications of these methods and data, the cost of SVI (though low) is worth consideration. There is a $5.60 – $7.00 charge per 1,000 panoramas depending on volume ordered (Google, 2022). Considering the median geodesic distance of panoramas being 2.4 m apart in Broome County, NY, regional datasets can have tens if not hundreds of thousands of panoramas. One potential cost-saving measure could include purchasing panoramas at lower intervals since infestations are often visible from several panoramas. Flores et al. (in prep.) also designed several

computer vision models to be constrained to flowering dates of the target plant, so careful selection of capture dates could also make SVI more cost-effective.

The framework and analyses presented here act as launching points with myriad potential future applications. With computer vision model predictions for larger spatial extents, results can be shared that are relevant to more managers, and existing reporting biases can be examined. Larger spatial scales also enable ecological analyses that are more representative of all roadside conditions, without the potential bias of selecting locations near streams as was the case in this study. With historical panoramas available, temporal resolution could be increased. A higher temporal resolution would allow for identifying and comparing spread rate with various road, land use, and habitat characteristics. Pre- and post-construction imagery could provide insight into the specific timing of culvert introductory events to examine fill soil more directly as an introductory medium. There is also potential to monitor biocontrol efficacy, for both the target plants studied here and others brought into the detection pipeline.

References

Aguilera, A. G., Alpert, P., Dukes, J. S., & Harrington, R. (2010). Impacts of the invasive plant Fallopia japonica (Houtt.) on plant communities and ecosystem processes. *Biological Invasions*, *12*(5), 1243–1252. https://doi.org/10.1007/s10530-009-9543-z

Ailstock, S., Norman, M., & Bushmann, P. (2001). Control and effects upon biodiversity in freshwater nontidal wetlands. *Restoration Ecology 9*(1). https://doi.org/10.1007/s00442-014-3188

Albert, A., Brisson, J., Belzile, F., Turgeon, J., & Lavoie, C. (2015). Strategies for a successful plant invasion: The reproduction of Phragmites australis in north-eastern North America. *Journal of Ecology*, *103*(6), 1529–1537. https://doi.org/10.1111/1365-2745.12473

Barlow, K. M., Mortensen, D. A., Drohan, P. J., & Averill, K. M. (2017). Unconventional gas development facilitates plant invasions. *Journal of Environmental Management*, 202, 208–216. https://doi.org/10.1016/j.jenvman.2017.07.005

Basic rule and maximum limits, N.Y. VAT Title 7, Article 30, §1180b (2016).

Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. v., Holmberg, J. A., Parham, J., Menon, S., Crall, J., van Oast, J., Kiciman, E., & Joppa, L. (2017). *Wildbook: Crowdsourcing, computer vision, and data science for conservation*. http://arxiv.org/abs/1710.08880

Bois, S. T., Silander, J. A., & Mehrhoff, L. J. (2011). Invasive plant atlas of new England: The role of citizens in the science of invasive alien species detection. *BioScience*, *61*(10), 763–770. https://doi.org/10.1525/bio.2011.61.10.6

Bram, M. R., & McNair, J. N. (2004). Seed germinability and its seasonal onset of Japanese knotweed (Polygonum cuspidatum). *Weed Science*, *52*(5), 759–767. https://doi.org/10.1614/p2002-053

Carlson Mazur, M. L., Kowalski, K. P., & Galbraith, D. (2014). Assessment of suitable habitat for Phragmites australis (common reed) in the Great Lakes coastal zone. *Aquatic Invasions*, *9*(1), 1–19. https://doi.org/10.3391/ai.2014.9.1.01

Cassey, P., Delean, S., Lockwood, J. L., Sadowski, J. S., & Blackburn, T. M. (2018). Dissecting the null model for biological invasions: A meta-analysis of the propagule pressure effect. *PLoS Biology*, *16*(4), 1–16. https://doi.org/10.1371/journal.pbio.2005987

Christen, D. C., & Matlack, G. R. (2009). The habitat and conduit functions of roads in the spread of three invasive plant species. *Biological Invasions*, *11*(2), 453–465. https://doi.org/10.1007/s10530-008-9262-x

Connell, P. J. (2015). *Using oblique imagery to identify invasive plant species in using oblique imagery to identify invasive plant species in Onondaga County, NY : a feasibility study Onondaga County, NY [Undergraduate honor's thesis].* https://digitalcommons.esf.edu/honors/99

Corcos, D., Nascimbene, J., Campesan, M., Donadello, D., Segat, V., & Marini, L. (2020). Establishment dynamics of native and exotic plants after disturbance along roadsides. *Applied Vegetation Science*, *23*(2), 277–284. https://doi.org/10.1111/avsc.12481

Costa, D., Neto, B., Danko, A. S., & Fiúza, A. (2018). Life Cycle Assessment of a shale gas exploration and exploitation project in the province of Burgos, Spain. *Science*

of the Total Environment, *645*, 130–145. https://doi.org/10.1016/j.scitotenv.2018.07.085

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, *4*(6), 433–442. https://doi.org/10.1111/j.1755-263X.2011.00196.x

Denby, B. R., Ketzel, M., Ellermann, T., Stojiljkovic, A., Kupiainen, K., Niemi, J. v., Norman, M., Johansson, C., Gustafsson, M., Blomqvist, G., Janhäll, S., & Sundvor, I. (2016). Road salt emissions: A comparison of measurements and modelling using the NORTRIP road dust emission model. *Atmospheric Environment*, *141*, 508–522. https://doi.org/10.1016/j.atmosenv.2016.07.027

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 9–15. https://doi.org/10.1145/2181037.2181040

Deus, E., Silva, J. S., Catry, F. X., Rocha, M., & Moreira, F. (2015). Google Street View as an alternative method to car surveys in large-scale vegetation assessments. *Environmental Monitoring and Assessment*, *188*(10). https://doi.org/10.1007/s10661-016-5555-1

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Escobar, T. (2019, Dec. 13). Google Maps 101: how imagery powers our map. https://www.blog.google/products/maps/google-maps-101-how-imagery-powers-our-map/

Esri. (2021). ArcGIS survey123 connect (3.13.251) [Computer software]. Redlands, CA. https://www.esri.com/en-us/arcgis/products/arcgis-survey123/overview

Esri. (2020). ArcGIS pro (Version 2.5.2) [Computer software]. Redlands, CA. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Flores, A., Kanan, C., Tyler, A., Acharya, M., & Megraw, L. (in prep.). *Using artificial intelligence on street view imagery to detect five key invasive plant species in New York State.* Unpublished manuscript.

Fox, J. & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fritz, S., See, L., & Brovelli, M. (2017). *Motivating and Sustaining Participation in VGI.* In Fritz S., See L., Foody G., Mooney P., Olteanu-Raimond A., Fonte C., et al. (Eds.), Mapping and the Citizen Sensor (pp. 93-118). London: Ubiquity Press. http://www.jstor.org.ezproxy.rit.edu/stable/j.ctv3t5qzc.8

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), 13108–13113. https://doi.org/10.1073/pnas.1700035114

Gelbard, J. L., & Belnap, J. (2003). Roads as conduits for exotic plant invasions in a semiarid landscape. *Conservation Biology*, *17*(2), 420–432. http://www.jstor.org/stable/3095361
Google. (2018, August 2). "Street View ready (pro grade)" specifications. https://developers.google.com/streetview/ready/specs-prograde
Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., & Wilkinson, P. (2013). *Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity*. 1–7.
Hellmann, J. J., Byers, J. E., Bierwagen, B. G., & Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. *Conservation Biology*, 22(3), 534–543. https://doi.org/10.1111/j.1523-1739.2008.00951.x
Huang, T. S. (1996). *Computer vision: Evolution and promise* (C. Vandoni, Ed.). CERN. https://doi.org/10.5170%2FCERN-1996-008.21
Joly, M., Bertrand, P., Gbangou, R. Y., White, M. C., Dubé, J., & Lavoie, C. (2011). Paving the way for invasive species: Road type and the spread of Common ragweed (Ambrosia artemisiifolia). *Environmental Management*, *48*(3), 514–522. https://doi.org/10.1007/s00267-011-9711-7
Keller, B. E. M. (2000). Plant diversity in Lythrum, Phragmites, and Typha marshes, Massachusetts, U.S.A. *Wetlands Ecology and Management*, *8*(6), 391–401. https://doi.org/10.1023/A:1026505817409
Kettenring, K. M., & Whigham, D. F. (2018). The Role of Propagule Type, Resource Availability, and Seed Source in Phragmites Invasion in Chesapeake Bay Wetlands. *Wetlands*, *38*(6), 1259–1268. https://doi.org/10.1007/s13157-018-1034-5
Kotowska, D., Pärt, T., & Żmihorski, M. (2021). Evaluating Google Street View for tracking invasive alien plants along roads. *Ecological Indicators*, *121*(May 2020). https://doi.org/10.1016/j.ecolind.2020.107020
Larson, E. R., Graham, B. M., Achury, R., Coon, J. J., Daniels, M. K., Gambrell, D. K., Jonasen, K. L., King, G. D., LaRacuente, N., Perrin-Stowe, T. I. N., Reed, E. M., Rice, C. J., Ruzi, S. A., Thairu, M. W., Wilson, J. C., & Suarez, A. v. (2020). From eDNA to citizen science: emerging tools for the early detection of invasive species. *Frontiers in Ecology and the Environment*, 194–202. https://doi.org/10.1002/fee.2162
Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*(6), 1659–1673.
Lemke, A., Kowarik, I., & von der Lippe, M. (2019). How traffic facilitates population expansion of invasive species along roads: The case of common ragweed in Germany. *Journal of Applied Ecology*, *56*(2), 413–422. https://doi.org/10.1111/1365-2664.13287
Li, F., Li, Y., Qin, H., & Xie, Y. (2011). Plant distribution can be reflected by the different growth and morphological responses to water level and shade in two emergent macrophyte seedlings in the Sanjiang Plain. *Aquatic Ecology*, *45*(1), 89–97. https://doi.org/10.1007/s10452-010-9334-8
Lowry, E., Rollinson, E. J., Laybourn, A. J., Scott, T. E., Aiello-Lammens, M. E., Gray, S. M., Mickley, J., & Gurevitch, J. (2013). Biological invasions: A field synopsis,

systematic review, and database of the literature. *Ecology and Evolution*, *3*(1), 182–196. https://doi.org/10.1002/ece3.431

Maheu-Giroux, M., & de Blois, S. (2007). Landscape ecology of Phragmites australis invasion in networks of linear wetlands. *Landscape Ecology*, *22*(2), 285–301. https://doi.org/10.1007/s10980-006-9024-z

Ministry of the Environment, Conservation and Parks. (2019, April 4). Management of Excess Soil - A Guide for Best Management Practices. *Queen's Printer for Ontario.* https://www.ontario.ca/page/management-excess-soil-guide-best-management-practices

National Geospatial Agency. (n.d.). Military Grid Reference System 1 km polygon [Geospatial data set]. https://earth-info.nga.mil/index.php?dir=coordsys&action=mgrs-1km-polyline-dloads

National Invasive Species Council. (2003). General guidelines for the establishment and evaluation of invasive species early detection and rapid response systems. Version 1. 1-16. https://www.doi.gov/sites/doi.gov/files/migrated/invasivespecies/edrr/upload/Guidelines-for-Early-Detection-Rapid-Response.pdf

National Invasive Species Council. (2014). Fiscal year 2013 interagency Invasive species performance budget. 2.

National Invasive Species Council. (2016). Management plan: 2016–2018. 1-35. https://www.doi.gov/sites/doi.gov/files/uploads/2016-2018-nisc-management-plan.pdf

New York Invasive Species Research Institute. (2020). Knotweed biocontrol released in NYS. Cornell University. Retrieved April 27, 2022 from http://www.nyisri.org/2020/07/knotweed-biocontrol-released-in-nys/

New York State Department of Environmental Conservation. (2021a). Office of Environmental Justice PEJA metadata. Retrieved May 26, 2021 from http://gis.ny.gov/gisdata/metadata/nysdec.PEJA.xml

New York State Department of Environmental Conservation. (2021b). *Potential environmental justice areas* [Geospatial data set]. New York State GIS Clearinghouse. https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1273

New York iMapInvasives. (2021). *Terrestrial invasive plant species list for New York State* [Data set]. NatureServe. Retrieved April 20, 2021 from https://www.nyimapinvasives.org/public-map

New York iMapInvasives. (2021). *Confirmed present species records* [Geospatial data set]. NatureServe. Retrieved November 6, 2021 from https://www.nyimapinvasives.org/public-map

New York iMapInvasives. (n.d.). Confirming records – PRISM staff, agency staff, natural resource professionals. Retrieved April 3, 2022, from https://www.nyimapinvasives.org/_files/ugd/a36481_0a50745e981041c4ba4628f1dbb89550.pdf

New York Natural Heritage Program. (2016a). *Invasive species spatial prioritization layer components.* Retrieved January 22, 2021 from https://tinyurl.com/y57ft8w4

New York Natural Heritage Program. (2016b). *New York iMapinvasives invasive species spatial prioritization - comprehensive score* [Geospatial data set]. https://arcg.is/15Xaiq0

New York Office of Information Technology Services. Average annualized daily traffic volume. [Dataset]. https://data.ny.gov/Transportation/Annual-Average-Daily-Traffic-AADT-Beginning-1977/6amx-2pbv

Olsson, A. D., van Leeuwen, W. J. D., & Marsh, S. E. (2011). Feasibility of invasive grass detection in a desertscrub community using hyperspectral field measurements and landsat TM imagery. *Remote Sensing*, *3*(10), 2283–2304. https://doi.org/10.3390/rs3102283

Parepa, M., Kahmen, A., Werner, R. A., Fischer, M., & Bossdorf, O. (2019). Invasive knotweed has greater nitrogen-use efficiency than native plants: evidence from a 15N pulse-chasing experiment. *Oecologia*, *191*(2), 389–396. https://doi.org/10.1007/s00442-019-04490-1

Pejchar, L., & Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends in Ecology and Evolution*, *24*(9), 497–504. https://doi.org/10.1016/j.tree.2009.03.016

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*. https://doi.org/10.1016/j.ecolecon.2004.10.002

Plant Protection Act, H.R 2559, 106th Con. (2000).

Porter, K. L. (2021). *Detection of Sericea Lespedeza (Lespedeza cuneata) infestations using small unmanned aerial system imagery*. [master's thesis, Emporia State University]. ProQuest Dissertations and Theses Global.

Pyšek, P., Jarošík, V., Hulme, P. E., Pergl, J., Hejda, M., Schaffner, U., & Vilà, M. (2012). A global assessment of invasive plant impacts on resident species, communities and ecosystems: The interaction of impact measures, invading species' traits and environment. *Global Change Biology*, *18*(5), 1725–1737. https://doi.org/10.1111/j.1365-2486.2011.02636.x

Ramsey, E., Rangoonwala, A., Nelson, G., & Ehrlich, R. (2005). Mapping the invasive species, Chinese tallow, with EO1 satellite Hyperion hyperspectral image data and relating tallow occurrences to a classified Landsat Thematic Mapper land cover map. *International Journal of Remote Sensing*, *26*(8), 1637–1657. https://doi.org/10.1080/01431160512331326701

Reaser, J. K., Burgiel, S. W., Kirkey, J., Brantley, K. A., Veatch, S. D., & Burgos-Rodríguez, J. (2020). The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. In *Biological Invasions* (Vol. 22, Issue 1). Springer. https://doi.org/10.1007/s10530-019-02156-w

Riitters, K. H., & Wickham, J. D. (2003). How far to the nearest road? *Frontiers in Ecology and the Environment*, *1*(3), 125–129. https://doi.org/10.2307/3867984

Rouifed, S., Byczek, C., Laffray, D., & Piola, F. (2012). Invasive Knotweeds are Highly Tolerant to Salt Stress. *Environmental Management*, *50*(6), 1027–1034. https://doi.org/10.1007/s00267-012-9934-2

RStudio Team (2021). RStudio: Integrated development environment for R (Version 2021.9.1.372) [Computer software]. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Rudrappa, T., Choi, Y. S., Levia, D. F., Legates, D. R., Lee, K. H., & Bais, H. P. (2009). Phragmites australis root secreted phytotoxin undergoes photo-degradation to

execute severe phytotoxicity. *Plant Signaling and Behavior*, *4*(6), 506–513. https://doi.org/10.4161/psb.4.6.8698
Saltonstall, K. (2002). Cryptic invasion by a non-native genotype of the common reed, Phragmites australis, into North America. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(4), 2445–2449. https://doi.org/10.1073/pnas.032477999
Saltonstall, K. (2003). Microsatellite variation within and among North American lineages of Phragmites australis. *Molecular Ecology*, *12*(7), 1689–1702. https://doi.org/10.1046/j.1365-294X.2003.01849.x
Shappell, L., Feldmann, L., Spencer, E., and Howard, T. (2016). New York State wetland condition assessment. EPA Wetland Program Development Grant. Final Report. *New York Natural Heritage Program*, Albany, NY. http://www.wetlandsforum.org/NYNHP_NYRAM_Ver4.2_Oct2016_Distro.pdf
Speed limits, N.Y. Comp. Codes R. & Regs. tit. 21 § 103.2 (n.d.).
Stringham, O. C., & Lockwood, J. L. (2021). Managing propagule pressure to prevent invasive species establishments: propagule size, number, and risk-release curve. *Ecological Applications*, 0–2. https://doi.org/10.1002/eap.2314
Stopping, standing or parking prohibited in specified places, N.Y. VAT tit. 7, art. 32, §1202 (2016).
Taylor, K., Brummer, T., Taper, M. L., Wing, A., & Rew, L. J. (2012). Human-mediated long-distance dispersal: An empirical evaluation of seed dispersal by vehicles. *Diversity and Distributions*, *18*(9), 942–951. https://doi.org/10.1111/j.1472-4642.2012.00926.x
Uddin, M. N., & Robinson, R. W. (2017). Allelopathy and resource competition: the effects of Phragmites australis invasion in plant communities. *Botanical Studies*, *58*(1). https://doi.org/10.1186/s40529-017-0183-9
Uddin, M. N., Robinson, R. W., & Caridi, D. (2014). Phytotoxicity induced by Phragmites australis: An assessment of phenotypic and physiological parameters involved in germination process and growth of receptor plant. *Journal of Plant Interactions*, *9*(1), 338–353. https://doi.org/10.1080/17429145.2013.835879
Uddin, M. N., & Robinson, R. W. (2018). Can nutrient enrichment influence the invasion of Phragmites australis? *Science of the Total Environment*, *613–614*, 1449–1459. https://doi.org/10.1016/j.scitotenv.2017.06.131
US Census Bureau. (2021a). 2021 MAF/TIGER feature class codes [Metadata]. https://www2.census.gov/geo/pdfs/reference/mtfccs2021.pdf
US Census Bureau. (2021b). *TIGER/lline road shapefiles* [Geospatial data set]. https://www2.census.gov/geo/tiger/TIGER2021/ROADS/
US Census Bureau. (2016a). *TIGER/line county boundary shapefiles* [Geospatial data set]. https://www2.census.gov/geo/tiger/TIGER2016/COUNTY/
US Census Bureau. (2016b). *TIGER/line state boundary shapefiles* [Geospatial data set]. https://www2.census.gov/geo/tiger/TIGER2016/STATE/
US Department of Defense. (2008). *Global positioning system standard positioning service performance standard.* https://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf
US Department of the Interior. (2016). Safeguarding America's lands and waters from invasive species: A national framework for early detection and rapid response,

Washington D.C. https://www.doi.gov/sites/doi.gov/files/National%20EDRR%20Framework.pdf
Vasquez, E. A., Glenn, E. P., Brown, J. J., Guntenspergen, G. R., & Nelson, S. G. (2005). Salt tolerance underlies the cryptic invasion of North American salt marshes by an introduced haplotype of the common reed Phragmites australis (Poaceae). *Marine Ecology Progress Series*, *298*(May 2020), 1–8. https://doi.org/10.3354/meps298001
Vaz, A. S., Alcaraz-Segura, D., Campos, J. C., Vicente, J. R., & Honrado, J. P. (2018). Managing plant invasions through the lens of remote sensing: A review of progress and the way forward. *Science of the Total Environment*, *642*, 1328–1339. https://doi.org/10.1016/j.scitotenv.2018.06.134
Vermeij, G. J. (1996). An agenda for invasion biology. Biological Conservation, 78(1–2), 3–9. https://doi.org/10.1016/0006-3207(96)00013-4
Victorian Government. (2010). *Invasive plants and animals policy framework.* DPI Victoria, Melbourne.
Weidenhamer, J. D., Li, M., Allman, J., Bergosh, R. G., & Posner, M. (2013). Evidence Does not Support a Role for Gallic Acid in Phragmites australis Invasion Success. *Journal of Chemical Ecology*, *39*(2), 323–332. https://doi.org/10.1007/s10886-013-0242-y
Wilson, M. J., Freundlich, A. E., & Martine, C. T. (2017). Understory dominance and the new climax: Impacts of Japanese knotweed (Fallopia japonica) invasion on native plant diversity and recruitment in a riparian woodland. *Biodiversity Data Journal*, *5*. https://doi.org/10.3897/BDJ.5.e20577
Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., & Xian, G. (2018). *United States national land cover database 2016* [Geospatial data set]. Multi-Resolution Land Characteristics Consortium. https://www.mrlc.gov/national-land-cover-database-nlcd-2016
Zandbergen, P. A., Ignizio, D. A., & Lenzer, K. E. (2011). Positional Accuracy of TIGER 2000 and 2009 Road Networks. *Transactions in GIS*, *15*(4), 495–519. https://doi.org/10.1111/j.1467-9671.2011.01277.x