2021

REGIONAL IMPACTS OF INVASIVE SPECIES AND CLIMATE CHANGE ON BLACK ASH WETLANDS

Joseph Shannon
Michigan Technological University, jpshanno@mtu.edu

REGIONAL IMPACTS OF INVASIVE SPECIES AND CLIMATE CHANGE ON BLACK ASH WETLANDS

By

Joseph Shannon

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Forest Science

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Forest Science.

College of Forest Resources and Environmental Sciences

Dissertation Co-Advisor:	*Fengjing Liu*
Dissertation Co-Advisor:	*Randall Kolka*
Committee Member:	*Amy Marcarelli*
Committee Member:	*Laura Bourgeau-Chavez*
College Dean:	*Andrew Storer*

Table of Contents

Author Contribution Statement

Chapter 2 has been submitted to the journal Hydrological Processes. The experiment was designed, performed, analyzed, and written by Joe Shannon. Fengjing Liu provided valuable feedback on experimental design, analysis, and writing. Randy Kolka and Matthew Van Grinsven provided valuable feedback and review of the analysis and writing.

Acknowledgements

I want to thank my wife Danielle Shannon for making it possible to complete this dissertation. Without her support to me directly, and her extra effort in caring for family it would have been impossible to complete this dissertation. She also shared her knowledge of evaluating and communicating climate change impacts on forests, greatly improving Chapter 3.

I would like to thank my advisor Fengjing Liu. His feedback on the design and analysis of my research and his tireless efforts reviewing drafts resulted in better analyses and clearer communication.

I would like to thank my co-advisor Randy Kolka and committee members Amy Marcarelli and Laura Bourgeau-Chavez for valuable conversations and components of my comprehensive exams that helped to shape and improve my research.

I would like to thank Andrew Verdin for discussion of synthetic weather generators and code examples.

I would like to thank the EPA Great Lakes Restoration Initiative, the Ecosystem Science Center, the Graduate Student Government, and the MTU Graduate School Finishing Fellowship program for providing the funding to carry out and present this research.

Abstract

For more than a decade intensive research on the ecohydrology of black ash wetland ecosystems has been performed to understand these systems before they are drastically altered by the invasive species, emerald ash borer (EAB). In that time there has been little research aimed at the scale and persistence of the alterations. The scale and persistence will be dependent not only on the immediate impacts of EAB, but also on vegetative response, the true extent of black ash wetlands on the landscape, and the compounding influence of a changing climate. Three distinct but related research articles are presented here to demonstrate a method for moderate resolution mapping of black ash across its entire range, understand the relative impacts of EAB and climate change on probable future wetland conditions, and a methodological study to quantify and reduce uncertainty around water level measurements that underpin much of our understanding in these systems. At a higher level, this research is intended to serve as a bridge between study-site level changes and the spatial and temporal extent of those changes, opening new research questions to better understand these relatively rapid shifts in regional forested wetlands.

1 Introduction

My goal for the research presented in this dissertation is to evaluate the regional extent of and resilience to a mass disturbance of one ecosystem type facing multiple interacting pressures. Forested wetlands are important features on the landscape and like other wetlands provide a range of ecosystem services from water quality and quantity regulation to wildlife habitat. Of specific focus here are black ash (*Fraxinus nigra* Marsh) wetlands, which are undergoing a massive transition as the invasive insect emerald ash borer (*Agrilus planipennis* Fairmaire, EAB) continues its 20-year expansion in North America. In addition to the presumed loss of the primary canopy component, these forests, like other wetlands around the world, will be impacted by a changing climate. Wetland function and form are dependent upon their hydroperiod so changes in precipitation and evaporative demand will be expected to impose lasting changes within these ecosystems. Black ash wetlands now represent an ecosystem impacted by two drivers that have overlapping effects and can be expected to severely disrupt or alter wetland function across the entire range of the ecosystem. How these effects will influence one another and whether they will coalesce or cancel out cannot be understood from any existing field studies or reviews. Nor is the full extent of these changes known as there are no moderate to high resolution maps of black ash across its entire range.

My dissertation provides insight into the scale of the regional impacts of EAB and climate change on black ash wetlands. The goal of each chapter is to provide both directly applicable results for managers and researchers, and to demonstrate novel approaches that can be applied in a broader range of settings. I have focused my efforts on a better understanding of the extent of black ash on the landscape and on how EAB and climate change will affect the future of hydrology in these wetlands. These two questions can 1) be combined with each other and with previous research to provide a better understanding of the long-term, regional scope of EAB, 2) serve as a basis for expanding existing and future research, and 3) demonstrate useful approaches to evaluate other large-scale and interacting disturbances. The questions of spatial extent and future impacts require incorporating uncertainty as a first-order result, introduced from the requisite modeling, as well as from the measurements used to construct those models. After identifying a source of uncertainty and potential error in a common water level monitoring approach used to collect my field data, I incorporated a study on pressure transducer accuracy and correction into my research program. This work proved essential in developing the wetland water level models used to simulate potential future conditions. In the summer of 2020, the first EAB infestation on the Ottawa National Forest was identified near the wetlands that served as the basis for my research. The results from the simulated post-EAB conditions in those wetlands, including those presented in this dissertation, will provide a guideline for future EAB-related research in the region.

I proposed and designed my research program with the explicit goal of pushing our understanding of the impact of EAB in black ash wetlands into the wider landscape and future conditions. The majority of existing work on black ash wetlands, is focused on within wetland processes (e.g., vegetation response, nutrient cycling, seedling survival, water levels). These studies have had three main benefits: provide a better understanding

of a poorly studied ecosystem before it is irrecoverably altered, estimate some of the anticipated impacts of the loss of black ash on these systems, and provide guidance to land managers looking to mitigate negative impacts.

Attempts have been made to estimate the total scale of the impact using USFS Forest Inventory and Analysis data, and these have provided useful insights into the relative magnitude of some changes (Iverson et al., 2008). To completely understand the scale of the disturbance we need a better idea of the total amount of black ash in the region and its distribution. My first chapter is a mapping project to identify black ash across its bi-national range. The approach I developed combines unsupervised and supervised techniques to create a classifier chain. Relying on existing forest inventory data and moderate-resolution satellite imagery the classifier presented identifies black ash from over 100 other possible dominant species across its wide range. The results are roughly equivalent to previous work over smaller geographic ranges with less phonologic and species diversity. One could expect that with the final map it will be possible to easily scale up previous work and begin to get quantities for a whole range of ecosystem impacts. That approach will work for components of black ash systems that do not vary significantly and are contained within a wetland unit (such as biomass of black ash in the region). However, for effects that can vary significantly or are exported from the wetland, more research is needed to determine the range of propagation, the amount of variation between systems, and potential network effects. Developing a better understanding of where black ash is on the landscape is critical to future research that can answer some of the questions about changes to the interconnected hydrologic network to which black ash wetlands belong. It will also deliver immediate effects at placing an order of magnitude on the scale of expected impacts, and bracket worst- and best-case planning scenarios.

Just as the results of Chapter 1 provide a platform for quantifying the total regional impact of changes to black ash wetlands, the second chapter provides a platform for my own hydrologic research and the research of the wider hydrology community. Wetland, groundwater, and stream stage monitoring often rely on water levels monitored through pressure transducers. These instruments provide a continuous series that can be recorded at nearly any timestep, enabling a wide range of derivative analyses from seasonal hydrographs to stormflow response to sub-daily estimates of evapotranspiration, or as in Chapter 3 development of wetland water level models. Each of these derivative analyses, Chapter 3 included, relies on the raw data from water level monitoring to draw conclusions, determine future research questions, and drive management decisions. There are, however, scenarios where environmental factors, specifically temperature gradients, can cause errors or a decrease in precision in these records. There has been some published work describing these errors, potential causes, and deployment or correction solutions, but these have been primarily for individual deployments or instruments. I developed a laboratory experiment to systematically compare correction equations and the uncertainty they add across monitoring scenarios and multiple pressure transducer pairs. My results and the case study presented give clear evidence of the magnitude of the problem. This work also demonstrates how the correction equations can be used to increase measurement precision. While increasing precision enables finer scale research

questions, increasing the gross accuracy of water level records can remove spurious relationships and reduce noise that can mask meaningful signals.

My final chapter built directly on the improved water level records created using the results of my second chapter. In Chapter 3, I used statistical wetland models and synthetic weather series to simulate the probable future conditions of previously studied black ash wetlands. The impacts of EAB on northern black ash wetlands has only been studied for less than a decade, so that even with artificially implemented EAB impacts there has been little time to understand the expected long-term change in wetland conditions. The persistence of any change in this system is twofold: the first is how wetland vegetation conditions change during recovery or management and how this affects hydrologic functionality, and the second is how climate change will force a potentially new hydrologic regime onto the system, which will influence, and be influenced by, wetland vegetation conditions. By simulating wetlands under the interactive effects of different vegetation and climate regimes, it was possible to look not just at probable future conditions, but also the relative importance of each change to those conditions. The result of that separation enables further research and management decisions to consider the persistence of the impact of EAB in these wetlands.

Each of the three chapters in this dissertation focus on different subject areas but relate back to the magnitude and drivers of change in forested wetlands. The results of each are also presented with uncertainty as a key component. The first and third chapters represent the types of results we would expect to contain uncertainty. Researchers would never claim to be certain about the classification of a forest species across the landscape, or about what future conditions will prevail in 80 years. The second chapter highlights that research questions do not need to span regions or reach into the future to have uncertain answers. Ecohydrology focuses on interconnected systems and all the inputs, outputs, and relationships in a system can rarely be known, much less measured. Knowing that certain answers are almost always outside of the scope of the question, I intend the results presented in this dissertation to also spur new questions around regional changes driven by an invasive species and a changing climate.

Note: The work in this dissertation spanned the period of the COVID-19 pandemic when a curtailment of field and laboratory research changed plans for many researchers. The lack of ability to sample in the spring and part of the summer 2020 season and dry conditions leading to few samples in the 2019 field season hampered expanding some portions of the existing datasets. While some of the initial research planned for this dissertation was disrupted, but I am fortunate and grateful that so much of this work could build on previously collected data.

2 Mapping Black Ash across the Species' Range using a Classifier Chain

2.1 Abstract

Unlike many other North American ash species that are located in mixed forests, black ash (*Fraxinus nigra* Marsh) is a dominant canopy species in forested wetlands. Black ash is threatened by the invasive insect emerald ash borer (*Agrilus planipennis* Fairmaire) and there have been numerous site-scale study results obtained over the previous 12 years. In the binational Great Lakes watershed and eastward, black ash wetlands are an important component on the landscape, but existing mapping efforts for black ash are generally too coarse for field identification, or confined in geographic extent to enable allocation of restoration resources. Accurate mapping at this scale will require ground-truth data from across the range of black ash and computationally intensive image processing and classification over a broad geographic extent. Pre-existing measurement plots from forest inventory systems were used in place of purpose-collected ground truth data. To account for the wide range of potential tree species across a large area I created a set of hierarchical species-label cohorts that did or did not contain black ash. The hierarchical cohorts were then individually modeled using random forest classifiers and each potential combinations of models was tested as a separate classifier chain. The models were trained and selected with a bias for user's accuracy, penalizing false positives more than false negatives. Across the entire range of black ash the classifier achieved training producer's and user's accuracy of 57% and 85%, and hold-out test data producer's and user's accuracy of 29% and 65%. These accuracies are comparable to other, state-level, efforts at mapping black ash. Finally, I identify potential improvements to test or implement in preparation for the creation of a final range-level black ash cover map.

2.2 Introduction

Forested wetlands are important features on the landscape that play a valuable role in hydrology and habitat at the local and watershed scales (Burkett and Kusler, 2000). Forested wetlands is a general term to refer to a wide range of ecosystems that vary in structure, function, and characteristics (Mader, 1991). The variety of structure, composition, and hydrology within the catch-all term 'forested wetlands' means that mapping efforts can generally be divided into two categories: general forested wetland identification, and targeted forest-type mapping. More general efforts of mapping forested wetlands are works that classify and map broad functional groups of forested wetlands, such as the National Wetland Inventory (NWI) and the work mapping peatland ecosystems in Bourgeau-Chavez et al (2017). Some efforts, such as the NWI have been considered a conservative estimate of forested wetlands as detecting wetland hydroperiods under a forest canopy is difficult (Lang et al., 2015; Wu, 2018). When a specific class of forested wetland is of interest a more narrowly scoped mapping effort is usually required to build upon general forested wetland mapping efforts. In the last 20 years, forested wetlands with high proportions of black ash (*Fraxinus nigra* Marshall) in their canopies have been of interest to managers and remote sensing researchers because of the invasive insect Emerald Ash Borer (*Agrilus planipennis* Fairmaire, EAB).

Black ash is a hardwood wetland species that is a common component of northern forested wetlands in the Great Lakes region and east through Ontario, New England, and the Canadian Maritime provinces (Figure 1). These wetlands are currently estimated to cover 1 million ha (Hofmeister et al, in preparation) through the region. In 2002 the new invasive species EAB was found in southeast Michigan, USA (Haack et al., 2002). EAB results in extremely high mortality in all species of ash native to North America and has spread rapidly after first being identified (Herms and McCullough, 2014). The expansion of EAB has now reached areas of the region where black ash is more prominent on the landscape and is the dominant canopy component in many forested wetlands, and drastic changes in canopy cover are expected as a result. Researchers and managers have been working to understand the impacts that can be expected following the loss of black ash and what potential tactics can be used to retain the services these wetlands provide (Bolton et al., 2018; Davis et al., 2019; Diamond et al., 2018; Looney et al., 2015; Shannon et al., 2018; Van Grinsven et al., 2017).

The ongoing extensive impact on black ash wetlands will require prioritization of limited management resources (time, money, growing stocks, etc.) to identify and respond to changes from EAB and functional loss of black ash. Understanding the prevalence and range of any resource within an area is a necessary step on the path to effective management. There is currently no consistent spatial information regarding the presence or absence of black ash across its entire range. One potential reason is that wet operating conditions and low product value made black ash a relatively unimportant commercial timber species, possibly contributing to a lack of directed research or mapping of this species. With the need to consider similar management questions across the entire range of black ash over a short period, developing a clearer picture of black ash on the landscape is a critical need.

Some characteristics of black ash wetlands may be useful in mapping efforts. Black ash has a relatively unique phenology (Ahlgren, 1957; Isaacson et al., 2012) within its range, with a late start of season and an early senescence. Additionally, black ash is extremely tolerant of wet conditions that persist into the growing season with surface flooding and saturated soils as late as August in some locations, with rapid fall rebound (Van Grinsven et al., 2017). This combination of potentially identifiable signals and management demand has led to recent efforts at mapping black ash across sub-range footprints including at the sub-state (Engelstad et al., 2019) and state level (Host et al., 2020). The most important predictor for the models in Engelstad et al. (2019) was compound topographic index derived from 20-m LiDAR data available in the state of Minnesota. The importance of this predictor restricts their approach to areas with publicly available LiDAR imagery, which prevents the extension of this model to the entire range of black ash. Engelstad et al. (2019) and other efforts at mapping black ash (e.g., Wolter et al., 1995) have utilized the phenology of black ash as signal for classifying images. Additionally, Host et al. (2020) combined a phenological model with physical topography data.

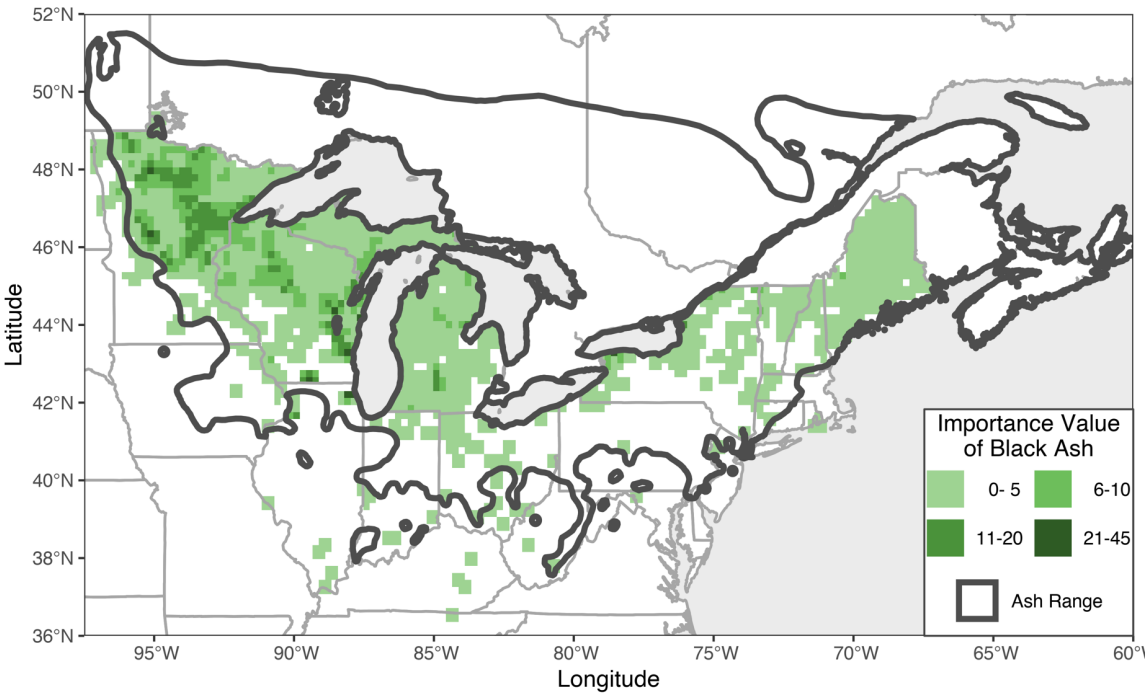


Figure 1. The USFS Forest Inventory and Analysis-derived importance value of black ash on the landscape in the United States. The importance value of a species shows is a measure of how often a species occurs on the landscape and how prevalent the species is within plots where it is present (Prasad and Iverson, 2003). Also shown is Little's species range (see text in 2.2) for black ash.

2.2.1 Optical and Radar Imagery

Optical imagery has a long history of use in remote sensing and mapping, and programs such as the Landsat satellites provide a long, relatively stable dataset. This dataset, and other lower resolution datasets (MODIS), have been used in many mapping efforts and also to monitor and classify land-use change (Loveland and Dwyer, 2012; Rocchio et al., 2018). Optical imagery benefits from the long period of record, the relative ease of interpretation of these images, and a knowledge base derived from decades of previous research and application. The major challenge in using optical imagery is the limited availability of cloud-free images, which can be particularly problematic in some regions (Lang et al., 2015). This limitation can make it impossible to use narrow date windows and multi-date methods in classification problems if cloud-free images are not available for the period of use. One practical approach to reducing this impact is to increase the revisit frequency, reducing the amount of time between images, increasing the chance of favorable conditions within a given window. For example, the ESA Sentinel-2 satellite program has a revisit period of 5 days compared to 16 days for Landsat 8. Related to masking by clouds is that optical/IR images can only be used to map soil-moisture (relative to surrounding areas), and inundation in forested wetlands only during periods of senescence due to the forest canopy (Lang et al., 2015; Wu, 2018).

To overcome the challenge of limited data due to lack of visibility it is possible to turn to active sensors that send and measure received signals. The biggest advantage of radar frequencies is that their long wavelengths can penetrate cloud cover and potentially provide consistent images of any region, however some formats are more difficult to process and interpret (Lang et al., 2015). Synthetic aperture radar (SAR) can be used to map the soil moisture and presence of standing water in open (C-HH/L-HH) and closed canopy (L-HH) wetlands (Bourgeau-Chavez et al., 2009; Lang et al., 2015; Wu, 2018). Additionally, SAR can also be informative regarding the overall biomass (HV) of vegetation present in the wetlands (Bourgeau-Chavez et al., 2017). The longer wavelength of L-Band SAR can pass through a vegetative canopy with minimal interaction/scatter and strike the soil or water below, with water causing highly specular reflection of the long wavelength. These differing interactions lead to a general pattern of return signal strength in the absence of a forest canopy: dry conditions lead to absorption and scattering of the signal and a relatively weak return, as soil water increases the amount of backscatter detected increases, and as water pools and the surface becomes inundated the returned signal strength sharply drops off as specular reflection leads to all/most of the signal being reflected away from the satellite (Bourgeau-Chavez et al., 2009). This trend is clear in the HH and VV polarization but is stronger in the HH polarization as the VV polarization is more sensitive to surface roughness (including water ripples, floating debris, and emergent vegetation) (Ahern et al., 2018). When a forest canopy is present the sharp drop in signal return is mitigated by a double bounce effect where the specular reflection from the surface water is redirected back in the direction of the satellite. This signal is again stronger in the HH polarization because of the decreased sensitivity to surface roughness, which extends from the water surface to the plant surface for the VV polarization.

2.2.2 Ancillary Predictors

Other contiguous landscape information can be used to supplement imagery, including continuous (slope, annual rainfall, etc) or categorical (US EPA ecoregions, USDA hardiness zones, existing landcover maps) (Ploton et al., 2020). Specifically, black ash often exists in expansive, low relief landforms (e.g., glacial lake beds) and in small depressions within more varied landforms (e.g., glacial moraines). Both of these features are identifiable via remotely-sensed terrain measures such as slope, concavity, and convexity (Wu, 2018). Many categorical contiguous spatial products are available that are derived from field observations, modeled outputs, or remotely sensed attributes. For example, Little's species range maps (Little and Viereck, 1971) can provide a hard or soft filter for presence of a species by setting probability of a species to 0% or some other low value outside of the stated range. Little's species range maps for individual tree species represent the current distribution of a species and were created from a combination of forest inventory and biogeographic variables.

While not the same type of remote sensing mapping efforts presented above, it is important to mention other efforts such as the United States Tree Atlas (Iverson et al., 2008). These products make use of a range of biogeographic and climatic predictors to model the landscape suitability for a given species, predicting their habitat suitability and

prevalence on the landscape. This type of modeling effort can be used to directly augment remote sensing classification as in Iverson et al (2008) or to set reasonable boundaries on species presence, serving as a type of hurdle model before a species can be predicted on the landscape (Stehman and Foody, 2019). This approach can be extended so that in place of biogeographic predictors, other classification and mapping products are used as predictors or as post-classification corrections, relying on an authoritative source to differentiate between covers of confusion (Bourgeau-Chavez et al., 2013).

2.2.3 Regional-Scale Mapping Efforts

Regional mapping efforts for any purpose present a specific set of challenges, some of which are magnified depending on the purpose of the mapping exercise. Generally, as the mapped area increases in size, availability of unified datasets decreases, confounding effects increase (e.g., phenology differences within a species), the number competing classes increases, and computational cost increases (Gorelick et al., 2017). As a result of the increased computation with increased range, historically the resolution of input datasets and processed products was coarser than desired, though the magnitude of this impact shrinks with advances in software and hardware. Additionally, validation of maps at this scale must be certain to consider the wide spatial extent to be evaluated. Final map classifications and errors could be correlated with other biogeographic trends (Dormann et al., 2007). Across large regions training data may be more sparsely located within the mapping region presenting less chance for autocorrelation to play a role in the validation models, or clustered within the region presenting a greater chance of auto-correlation impact (Ploton et al., 2020; Roberts et al., 2017).

Identifying a single species across a large geographic area can be generally framed as one of two types of classification problems: a large multiclass problem or a single large binary problem. Binary classifications at this scale will often result in at least one class that has high variability based on pooling of multiple land classes into a single target class. There are scenarios where this may be acceptable or even desirable. For example, if creating a binary hardwood/conifer classifier in northern forests, a relatively good indicator would be the difference between summer and winter greenness or NDVI, as all conifers except tamarack would be easily differentiated from all hardwoods, regardless of species. A good binary classification problem is one where one class has a distinct feature that is lacking in all other classes, and it is possible to define some threshold to identify that feature. It can be difficult to find a single feature that has little enough variation among all pooled classes such that a threshold can be identified. In these cases, it is not possible to use a binary classification and a full land cover classification is better suited. Importantly, binary classifications and binary maps are not entirely synonymous. Multiclass supervised or unsupervised classifiers can be reduced to binary maps after the initial classification. Rather than attempt to find a feature that distinctly identifies a single class, this approach uses a larger set of features to identify a set of classes that includes the target binary class. The advantage to this approach is that it allows for finer differentiation between classes while still identifying the class of interest. This is similar to the approach in Bourgeau-Chavez (2009) Lake St Clair case study and in the land cover change approaches discussed in Stehman and Foody (Stehman and Foody, 2019,

sec. 4.5). When assessing the accuracy of the final maps, relative to full land cover type classifications, binary classifications are easier to assess for accuracy (only a 2x2 confusion matrix).

The goal of this research is to produce a classification tool for accurately identifying black ash to be applied across the entire bi-national species range. Such a classifier should be able to handle the correlated multiclass structure of forest inventory data making use of widely available predictors. Due to data availability the classifier will be developed and tested against data from the United States using predictors that are available in both nations.

2.3 Materials and Methods

2.3.1 Study Area

The study area for this mapping effort is bounded by the species range of black ash, roughly a rectangular region stretching from Minnesota to the Atlantic Ocean in a band of latitudes that covers the North American Great Lakes (Figure 1). The region includes 16 American states and 6 Canadian provinces, totaling 220.8 million hectares (Figure 1). Within this region there are over 100 common tree species, and phenology spans a wide range for leaf-out and senescence. Across that area in the United States, black ash is known to be prevalent on the landscape in Minnesota, northern Wisconsin, Michigan's Upper Peninsula, and Maine (Figure 1). Similar metrics are not available from Canada, but other mapping efforts show that black ash is a common species along the same latitudinal band defined by the northern and southern boundaries of the Great Lakes (Beaudoin et al., 2014).

2.3.2 Forest Inventory and Forest Masks

Forest inventory data were solicited from a variety of sources across the range of black ash. Spatially explicit forest inventory plots were provided by St. Louis County, MN Land and Minerals Department, the Superior National Forest, and the St. Regis Mohawk Tribe Environment Division (Table 1). Data-sharing agreements were reached with the Wisconsin Department of Natural Resources and USDA US Forest Service (USFS), granting access to generally confidential plot coordinates for their state and national forest inventory plots. The USFS Forest Inventory and Analysis (FIA) program performs an annual survey panel on approximately 20% of permanent inventory plots within each US state (Bechtold and Patterson, 2015). Based on the sampling grid there is approximately one inventory plot for every 2400 hectares (Gray et al., 2012). Each plot consists of four 7.32-meter radius subplots arranged with one centered at the plot center, and the other three located at bearings of 0º, 120º and 240º 36.6 meters from the plot center (Bechtold and Patterson, 2015). The Wisconsin Continuous Forest Inventory (WICFI) plots are located on Wisconsin State Forest Land. WICFI subplots are the same size as FIA subplots, but there are only two plots, with one centered on the main plot coordinates and one located 36.6 meters away on a bearing of 180º. For both the FIA and WICFI the individual subplots were used as separate observation in training and testing wetland models. All provided WICFI points were used in this analysis, but the spatial and

temporal extent of USFS FIA plots required some subset of plots to be selected. All FIA plots with black ash present as a live, dominant or co-dominant canopy tree in the FIA database (Bechtold and Patterson, 2015) were selected for use in modeling. To create a manageable size set of plots, plots without any black ash stems were filtered to retain only plots within the footprint of selected Landsat images (see below). As a preliminary filter means to match the prediction step in generating a final map, non-forested inventory plots were removed from the dataset using the Hansen forest cover map (Hansen et al., 2013) as an authoritative source of forest/non-forest classification.

Basal area for each tree was calculated using the recorded diameter at breast height. Because the size and design of the sample points varied by source (Table 1), tree basal area was expressed as basal area per unit area. This conversion allowed for basal area to be directly compared as sums within species and plots to determine the total basal area for a species within a plot, and the proportion of plot basal area represented by each species present. Some additional filtering of inventory plots was necessary solely for processing considerations. Each of the forest inventory plots had the date of the inventory recorded and no inventory plots measured prior to 2010 were used. Prior to the random selection of non-ash FIA plots (above), plots were filtered to exclude those located in Landsat scene footprints that contained less than 4 plots.

Three response features were calculated from the forest inventory data for model training, testing, and tuning. The first was the proportion of plot basal area attributed to black ash stems (Y_{prop}). The second was simply a binary indicator of whether black ash was the dominant species by basal area on a plot ($Y_{binary} = Y_{prop} \geq 0.5$). The final response was the dominant species on each plot by proportion of basal area ($Y_{species}$). Details of the application of each response metric can be found in Sections 2.3.4.3 and 2.3.4.4.

Table 1. Number and design of forest inventory plots used to train predictive models by source. Pecentage of total ash plots may not sum to 100% due to rounding.

Source	Plot Design	Number in Dataset	Percent of Plots with >=50% Basal Area as Ash
FIA	Nested Subplots	107,469	47.40%
WICFI	Nested Subplots	5,567	3.25%
St. Regis Mohawk	Fixed area (square)	7	0.06%
Superior NF	Variable Radius	698	6.24%
St. Louis County	Variable Radius	5,074	43.06%
	Total	**118,815**	

2.3.3 Remotely Sensed Data

Optical imagery was retrieved from the Landsat program using Landsat 5 and Landsat 8. Landsat 7 was excluded from this analysis because of the known scan-line errors, which would have resulted in high bandwidth and processing costs for significantly less gain than Landsat5 & Landsat8 imagery (Scaramuzza and Barsi, 2005). Landsat metadata and images were retrieved via the USGS EarthExplorer Machine-to-Machine API. For each forest inventory plot 5 previous years of available images were retrieved, and images were filtered to include only those with less than 90% cloud cover. At the time of download the most consistent imagery available across the region (including both nations) was Level 1 top-of-atmosphere (TOA) images. To use these images, the encoded data were converted from digital number (DN) to TOA reflectance using the rescaling factors in the image metadata applied as described by the USGS Landsat program documentation (USGS, 2020). Table 2 shows the Landsat bands that were used for this analysis. Landsat 8 bands 1 and 11 were not used because there is no Landsat 5 equivalent. Additional information can be retrieved from optical imagery though the calculation of indexes based on the relative magnitude of signal between various bands. I calculated both vegetation and moisture indices, the description of which are in Table 3.

Table 2. Landsat 5 and 8 bands tested as potential predictors for black ash wetlands.

Common Name	Landsat 5 Band	Landsat 8 Band
blue	1	2
green	2	3
red	3	4
near infrared (NIR)	4	5
shortwave infrared 1 (SWIR1)	5	6
shortwave infrared 2 (SWIR2)	7	7
thermal infrared 1 (TIRS1)	6	10

Start of growing season (SOS), end of growing season (EOS), and day of year of maximum NDVI (MAXT) were extracted from eMODIS phenological date rasters for each plot/Landsat image combination (Jenkerson et al., 2010). For each plot ‘day of season’ and ‘days from peak’ were calculated as the difference between Landsat acquisition date and the start of season and MAXT, respectively. Any day of the season between SOS and EOS was labeled as ‘growing season’ and days outside that range as ‘dormant season’. The growing season was further refined to early, mid, and late using the half-way points between SOS and MAXT and MAXT and EOS as the breakpoints.

Synthetic aperture radar L-band data with HH polarization acquired by ALOS PALSAR were downloaded from the Alaska Satellite Facility (ASF) (ASF DAAC, 2015). The ALOS PALSAR sensor collected between 2006 and 2011 and all images used in this analysis fall between March 31st and November 30th in those years. In total 5,634

Table 3. Vegetation and moisture indices calculated as potential predictors of black ash stands.

Index	Name	Description
NDVI	Normalized difference vegetation index (Masek et al., 2006; Vermote et al., 2016)	Measure of vegetation density and health
EVI	Enhanced vegetation index (Masek et al., 2006; Vermote et al., 2016)	Similar to NDVI but better handles exposed soil and atmospheric conditions
NDMI	Normalized difference moisture index (Wilson and Sader, 2002)	Measure of vegetation moisture content
$MNDWI_M$	Modified Normalized difference water index (McFeeters, 1996)	Provides a measure of open water
$MNDWI_G$	Modified Normalized difference water index (Gao, 1996)	Measure of vegetation moisture content, and is sensitive to surface water
NBR2	Normalized burn ratio 2 (Masek et al., 2006; Vermote et al., 2016)	Measure of water sensitivity in plants, associated with other moisture indices using similar wavelength bands (Ji et al., 2011)

PALSAR images were collected under fine-beam single polarization (10m resolution), fine-beam dual polarization (20m resolution), and polarimetric (30m resolution) modes. All downloaded images were high-resolution radiometric terrain corrected (RTC) products provided by ASF. RTC removes the impact of terrain on SAR backscatter measurements and of terrain influences that lead to geolocation errors. The images were masked to retain only pixels with incidence angles between 30 and 55º to avoid the effect of incidence angle on forest canopy penetration (Lang et al., 2008). A 3x3 focal median filter was applied to each PALSAR image to remove speckle inherent in SAR imagery (Bourgeau-Chavez et al., 2013). The median filter and subsequent processing was performed on the RTC-generated gamma-nought values.

To incorporate topographic information about the location of black ash wetlands, data from the NASADEM product was downloaded through the same EE machine-to-machine API. NASADEM provides continuous, global 30-m slope, aspect, and curvature data derived from the shuttle radar topography mission (USGS EROS, 1994). From this set of data slope and profile and plan curvatures (profc and planc) were extracted, where profile and plan curvature represent curvature parallel to and perpendicular to the slope, respectively.

For each of the above Landsat bands and indices and the median-filtered HH SAR returns (collectively referred to as signals) a variety of additional potential predictive features

were developed. These included the mean, minimum, maximum, and standard deviation of each signal for the entire year, or the eMODIS phenological periods of the growing season, the dormant season, the early, mid, and late growing season, and the peak growing season (MAXT ± 30 days). Due to the specific phenology of black ash the same values were calculated from images collected in September. The slope, profc, and planc were all used without further processing.

Eight additional predictors were calculated to create potential masking predictors following the decision tree thresholds in Wolter et al. (1995). These were the differences in mean mid-growing season NDVI/EVI and mean September NDVI/EVI to capture the early senescence of black ash (mask_frni_ndvi, mask_frni_evi). The differences between mean mid- and early- growing season NDVI/EVI to capture the early leaf-out of trembling aspen (*Populus tremuloides* Michx) (mask_potr5_ndvi, mask_potr5_evi). The difference between the mean growing and dormant season NDVI and EVIs to distinguish coniferous species (mask_conifer_ndvi, mask_conifer_evi). The difference between the mean mid- and late- growing season red band to distinguish red maple (*Acer rubrum* L.). Trembling aspen and red maple were selected because they both can grow in similar wet conditions. Finally, the differences between mean mid- and late- and mid- and early- HH PALSAR returns were created to distinguish areas that remain wetter throughout the growing season (d_late_mid_hh, d_early_mid_hh).

2.3.4 Modeling Approach

2.3.4.1 Spatial Cross-Validation

Evaluating spatial classification requires special care relative to other types of predictive models. Spatial data need to be blocked to avoid spatial autocorrelation transferring information between nearby training and test data points (Roberts et al., 2017). One approach is to split the study area into grid cells and perform dataset splitting on the cells rather than on the points. Creating cells or blocks of test and train points effectively isolates the transfer of information between the test and training sets for any points away from the grid boundaries. Using a 50 km grid over the range of black ash, I split the data into an 80/20 train/test dataset. Test grid cells were determined by randomly sampling cells and summing the cumulative number of ash plots in the selected cells (using Y_{binary}) until 20% of the ash plots were assigned to the test set. In this way the 80/20 split stratifies by the ash plots and not just the number of grid cells (Figure 2). The result was 30% of the grid cells were considered test data and 70% are in the training data while the ash/non-ash plots were both split as 20% testing and 80% training. The blockCV R package (Valavi et al., 2019) was used on the training plots to create 10 folds for CV training and tuning of models. The grid size is set smaller for the fold to allow for permutations of data within the larger training/test grids.

All potential predictors had robust scalar (centering using the median and applying a unit scale transformation) and Yeo Johnson (a power transformation to reduce predictor skew) transformations applied. These transformations were fit to the predictor values in the training dataset only and later applied to the testing dataset.

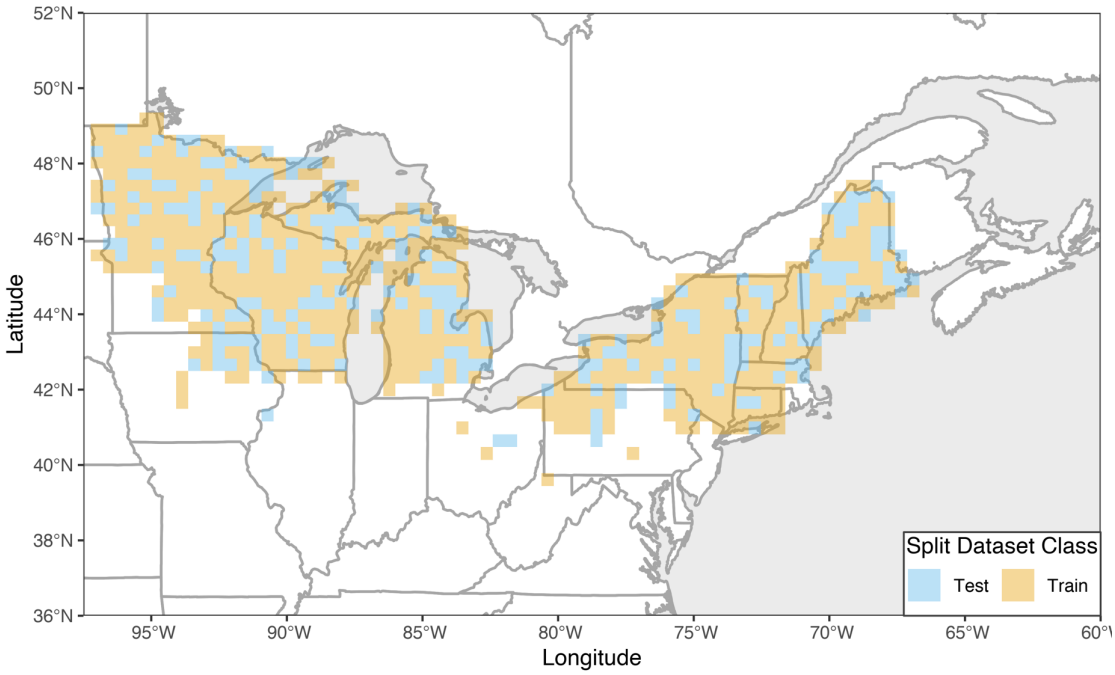


Figure 2. Distribution of spatially-blocked cross-validation data splits. Each cell is 50 x 50 km and cells were randomly sampled as 'test' data until the plots within the cells contained 20% of the labeled ash plots.

2.3.4.2 Predictor Selection

The above procedure generates summary statistics for multiple signals and indices for multiple phenological periods, which has a multiplicative effect on the number of potential predictors (*final number of potential predictors* = 422). Within a smaller footprint (either geographically or by some other filter (as opposed to 'all forested land within the range of black ash'), useful predictors can often be selected directly. But across 99 species and over 200 million hectares it is difficult to know what features may differentiate black ash from another species that is phenologically or hydrologically similar, and how predictor importance may vary across the mapping area. We chose to reduce the number of potential important features through a series of scoring steps. For each model fit and tested (see below) 23 predictors were chosen from the entire pool. The topographic predictors (slope, planc, profc) were included in each model. All other predictors were scored using the mutual information classifier, providing a measure of binary classification of each individual predictor (Kuhn and Johnson, 2013). The predictors were sorted by score and the top 20 predictors were chosen in sequence. After each predictor was selected (starting with the best single predictor) all other predictors that had a correlation of ≥ 0.8 with the selected one were removed from consideration regardless of their mutual information classifier score. This process was repeated, adding the top scoring remaining predictor to the selection until 20 predictors were selected. Twenty was chosen as an initial threshold to limit the number of uninformative predictors fed to the classifiers, based on the results of predictor importance value in Section 2.4.3 this number can be increased or decreased in a final fitting. Correlation was retrieved

from a 422 x 422 correlation matrix created using Spearman's ranked R^2 all pairs of predictors. In this way the most informative predictors were selected while excluding predictors that were strongly co-linear. Removing co-linear predictors was especially important given the wide pool of candidate predictors all created from a similar base of optical imagery data. No imputation was performed to fill missing data. Points with any missing predictors were removed from the analysis and predictors with more than 10% of their data points missing were removed from consideration before scoring.

2.3.4.3 Classifier Chain

To deal with the noisiness of a single binary classification and the high number of species in a multiclass classification, I choose to create a chain classifier, breaking down the classification of black ash into a series of smaller hierarchical binary problems (Read et al., 2021). Some classification groups may be easy to identify (deciduous vs coniferous), but others are less straight-forward. For instance, should species with similar phenologies or those in similar hydrologic settings be grouped together? Creating classification groups and models that were likely to perform well on binary classification required identifying groups of species that were similar given the predictors I generated. A pairwise species distance matrix was developed using the median Euclidean distance between each pair of species observations from full set of predictors for the training data. Potential species groups were created using hierarchical clustering on the median distance between each pair of species. Hierarchical clustering is an unsupervised classification technique that seeks to minimize some measure of dissimilarity (or maximize a measure of similarity) within each cluster. I used the agglomerative Ward variance minimizing algorithm which starts within individual species groups and recursively combines groups into clusters (species in my case), aiming to minimize the variance within all clusters. The resulting set clusters can be considered cohorts of similar species and can be represented on a dendrogram showing the hierarchical clustering structure of the species (Figure 3). At each branch point on the dendrogram a binary model could be constructed to classify the species on either side of the branch. For example the classifier at the left-most point in Figure 3 would be primarily separating coniferous and deciduous species. In this way it is possible to create a chain of binary classifiers that differentiate between various cohorts of species at each step. The orange line in Figure 3 represents the path of decisions required to reach the black ash 'leaf' on the dendrogram.

Black ash is nested within 9 clusters that range from black ash (with 98 species in the non-ash cohort) to all potential dominant species. Each of those clusters represents a binary decision between the set of species which contains black ash (S_{pres}) and the set of species that does not contain black ash (S_{abs}) (Table 4). Constructing a classifier chain using this structure becomes an exercise of identifying the optimal path through the 8 decision points to get to black ash. The decision points are hierarchical and can be lumped with decision points above or below them to create classifier chains that ranged from 1 to 8 models long. For each decision point a model was fit to all possible hierarchical cohorts described by that decision point. For example, the second decision point would have two associated models, the first would assume that decision point 1 was

modeled separately and S_{pres} and S_{abs} did not include the species from the previous S_{abs} cohort (S_{abs-1}). The second associated model assumes that decision point 1 was skipped, and correspondingly S_{abs} would include the species from S_{abs-1} and S_{abs-2}. The ability to pool decision points and species cohorts is important to find the optimal classifier chain. Each classifier added to the chain increases uncertainty, the possibility of misclassification, and the potential for overfitting the final model to the training data. Selection of the final set of classifiers in the chain is a trade-off between those costs and accurately identifying black ash.

After considering all decision points, and all potential pooling points, there are 36 potential models to describe all decisions. Predictors were evaluated independently for each of the 36 models as described in 2.3.4.2, resulting in 36 independent and varying sets of 23 predictors, one for each decision point model. For each of these 36 models, a random forest classifier was trained using 10-fold cross-validation to avoid overfitting. The folds used in cross-validation were spatially-blocked as in Spatial Cross-Validation. Random forest classifiers create a series of decision trees with each one working with a randomly selected subset of predictors to avoid overfitting the model to the training data. Cross-fold validation involves splitting the data into 10 partitions and using 9 of these to train the model and 1 to evaluate performance. The model fitting is 10 times repeated until each partition has been used as the testing data for evaluating performance. Each random forests model was then trained to classify the respective S_{abs} and S_{pres} for the decision point. During training the models excluded previously classified species, effectively assuming perfect classification at each previous step in the potential chain. With the 36 fit models I was able to evaluate the 256 combinations of classifiers that could result in a binary classification of black ash as the final step Figure 3. Scoring each chain was done by predicting on the training dataset for each 'link' in the chain of classifiers. The final classification was determined by taking the product of the probability of ash predictions for each observation from each of the models in the chain. If the product of the probabilities was greater than 0.5 then the prediction was labeled as ash. These chain-level final predictions were scored against Y_{binary} using the $F_{0.5}$-metric, which is the harmonic mean of the precision and recall with more weight given to the precision of the predictions over the recall. Precision is the positive predictive value of the model and describes how many of the predicted ash values are truly ash (how many predicted ash plots were true ash) and corresponds to the user's accuracy of a map. Recall is the true positive rate of the model and describes the ratio of true predictions of ash presence to the total number of observed ash plots (how many true ash plots were correctly predicted) and corresponds to the producer's accuracy of a map. Weighting the model score towards precision has the effect of biasing model selection towards more conservative predictions of black ash. The model is selected towards labeling low-confidence black ash predictions as non-ash to have fewer false positive predictions. The final chain was selected as the combination of models with the highest $F_{0.5}$ score for the combined prediction of black ash.

2.3.4.4 Prediction Threshold

For very imbalanced classification problems (Table 1), classifying points using a probability threshold of 0.5 may result in poor model performance (Kuhn and Johnson, 2013). To determine the optimal classification threshold, a series of thresholds can be used to create predicted Y_{binary} from predicted Y_{prop}, and each of these classifications is scored. The threshold with the best training performance is then used in the final predictions. For single-species classification problems, labeling observed inventory plots by dominant species can be performed in a similar manner. For each inventory plot, the binary response of ash/non-ash is applied if the proportion of basal area as black ash is greater than a threshold basal area proportion (Wolter and Townsend, 2011). We performed a grid search on possible probability and basal area proportion thresholds over the training data probability predictions, scoring each combination between 0.05 and 0.95 in increments of 0.01 (Figure 4). These thresholds were again evaluated with the $F_{0.5}$ score, giving heavier weight to models with higher precision scores.

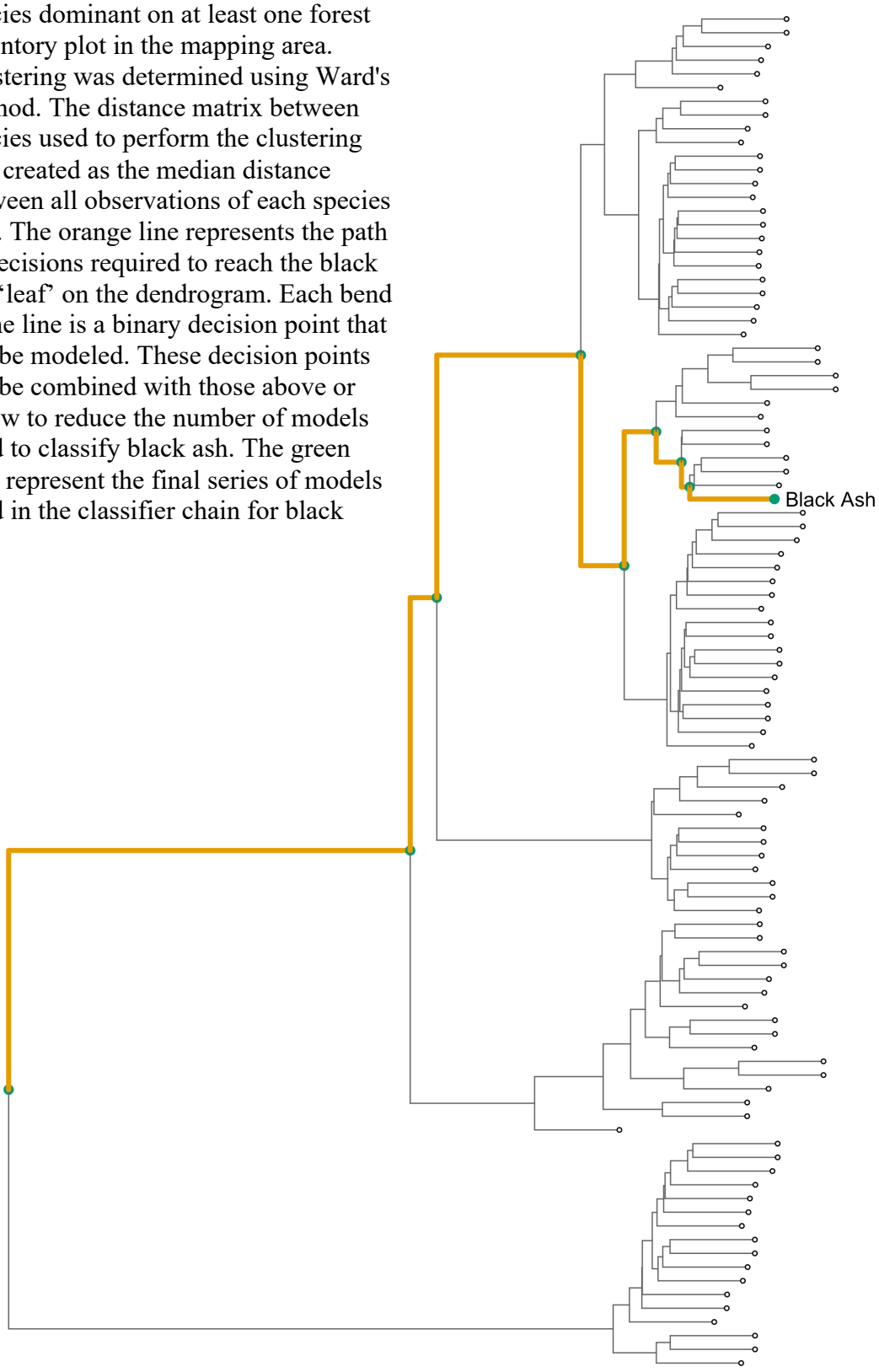
2.3.5 Software

Analysis was performed in both R (R Core Team, 2019) and Python (Van Rossum and Drake, 2009) making use of the tidyverse packages (Wickham et al., 2019), data.table (Dowle and Srinivasan, 2019), terra (Hijmans, 2021), sf (Pebesma, 2018), pandas (Wes McKinney, 2010), scikit-learn (Pedregosa et al., 2011), numpy (Harris et al., 2020), and scipy (Virtanen et al., 2020).

Table 4. The balance of species for each potential model in the classification chain for black ash. This balance represents only a single classification model structure. During chain classifier training each evaluated step removed species already considered 'classified' in previous chain classifiers.

Model ID	Number of Species in each binary	
	Black Ash Present	Black ash Not Present
1	82	17
2	70	29
3	57	42
4	23	76
5	18	81
6	15	84
7	4	95
8	3	96
9	1	98

Figure 3. Hierarchical clustering of all species dominant on at least one forest inventory plot in the mapping area. Clustering was determined using Ward's method. The distance matrix between species used to perform the clustering was created as the median distance between all observations of each species pair. The orange line represents the path of decisions required to reach the black ash 'leaf' on the dendrogram. Each bend in the line is a binary decision point that can be modeled. These decision points can be combined with those above or below to reduce the number of models used to classify black ash. The green dots represent the final series of models used in the classifier chain for black ash.



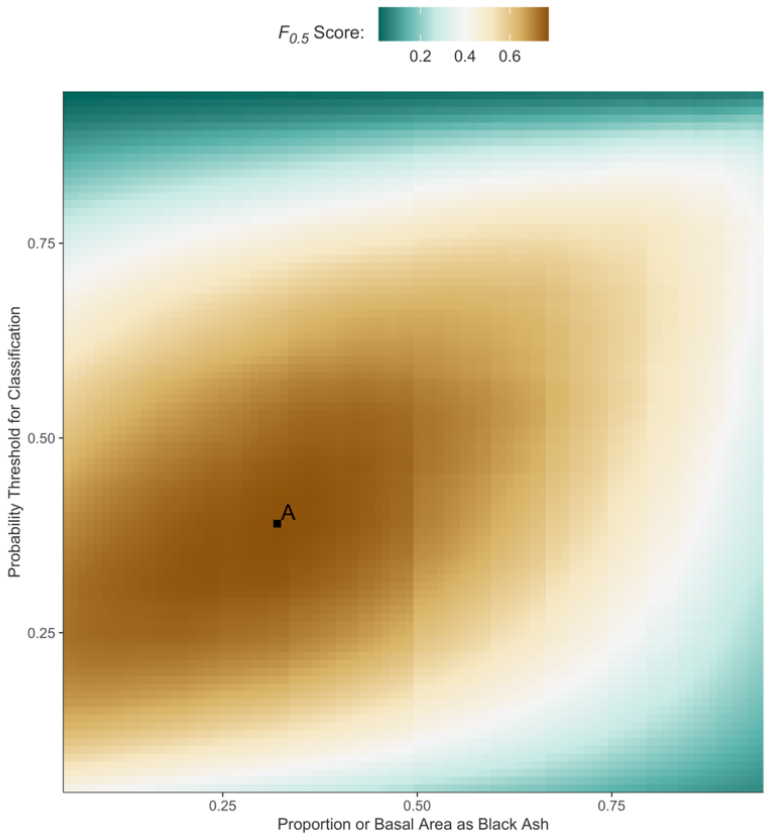


Figure 4. Heatmap showing the prediction $F_{0.5}$ score across a grid-search of potential probability and proportion of basal area thresholds. Point `A` shows the location of optimal model performance.

2.4 Results

2.4.1 Classifier Structure and Accuracy

The best performing classifier chain consisted of 3 models: 1, 7, 9 (Table 4, Figure 5). The final model $F_{0.5}$ during chain selection (trained on Y_{binary}) was 0.77. When scoring the classifiers $F_{0.5}$ increased sharply from a classifier of length 1 to one of length 3, where it reached a peak, and then began an increasingly rapid descent after length 4 (Figure 5). Model 1 contains 16 coniferous species and one genus-level hardwood classification in S_{abs} (Figure 6). Model 7 distinguished black ash and its cohorts (n = 4) from an additional 78 species, the large majority of hardwood species, and model 9 classified black ash from the remaining 3 species. After performing the grid search to identify the optimal classification thresholds for predicted probability and proportion of black ash, these values were set to 0.39 predicted probability and 32% of the plot basal area as black ash (Figure 4). When this classifier chain was used to predict the holdout test data, the $F_{0.5}$ score was 0.52. The most common species misclassified as a black ash stand in the full dataset was quaking aspen (*Populus tremuloides*, Michx.) and the next most common false positive (10.31%) was green ash (*Fraxinus pennsylvanica*, Marshall) (Table 5).

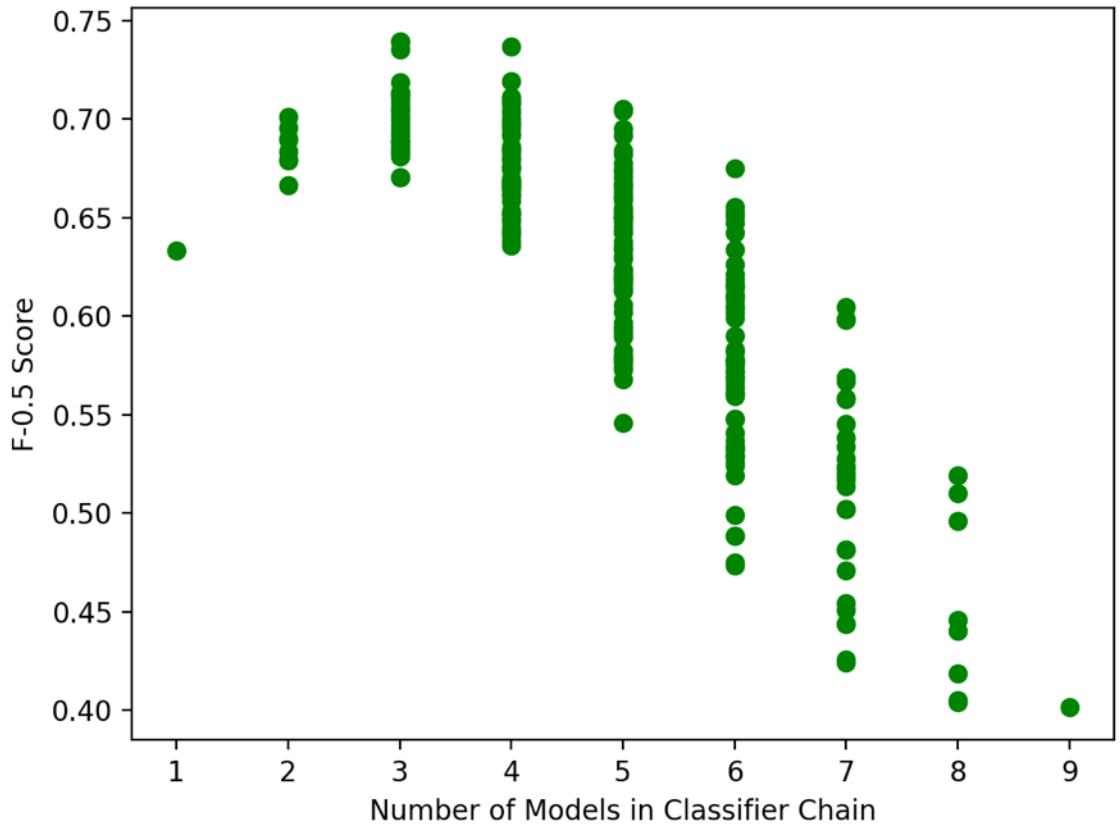


Figure 5. $F_{0.5}$ score for all possible classifier chains plotted as a function of number of models in the classifier chain.

Confusion matrices (Story and Congalton, 1986) for both the training and testing data sets are shown in Table 6. These results show that using the $F_{0.5}$ metric for performance evaluation has, as expected, optimized for precision (User's Accuracy), with higher precision in both the training and testing data. Optimizing for precision comes at the cost of recall (Producer's Accuracy), which was lower in both datasets. The final classifications are conservative, in that they are only likely to classify ash if there is high confidence that ash exists on the ground. The overall accuracy for both datasets is similar and is an improvement over the base proportion of observed black ash plots (94.64%), indicating the model is an improvement over random assignment.

Table 5. Absolute count and proportion of false positives (misclassified as black ash) by dominant plot species.

Species	Number of Plots Labeled Black Ash	Percent of Total False Positives
quaking aspen	148	17.75%
green ash	86	10.31%
red maple	85	10.19%
paper birch	81	9.71%
American elm	56	6.71%
American basswood	49	5.88%
balsam poplar	42	5.04%
other species (< 5% each)	287	34.41%

Table 6. Confusion matrix summary for classification performance for the model training and model testing datasets.

Training Data:

	Actual Presence	Actual Absence	*Total*	User's Accuracy	
Predicted Presence	3281	580	*3861*	**84.98%**	*Precision*
Predicted Absence	2506	76215	*78721*	**96.82%**	
Total	*5787*	*76795*			
Producer's Accuracy	**56.70%**	**99.24%**		**96.26%**	*Overall Acc.*
	Recall				

Testing Data:

	Actual Presence	Actual Absence	*Total*	**User's Accuracy**	
Predicted Presence	448	240	*688*	**65.12%**	*Precision*
Predicted Absence	1088	34457	*35545*	**96.94%**	
Total	*1536*	*34697*			
Producer's Accuracy	**29.17%**	**99.31%**		**96.33%**	*Overall Acc.*
	Recall				

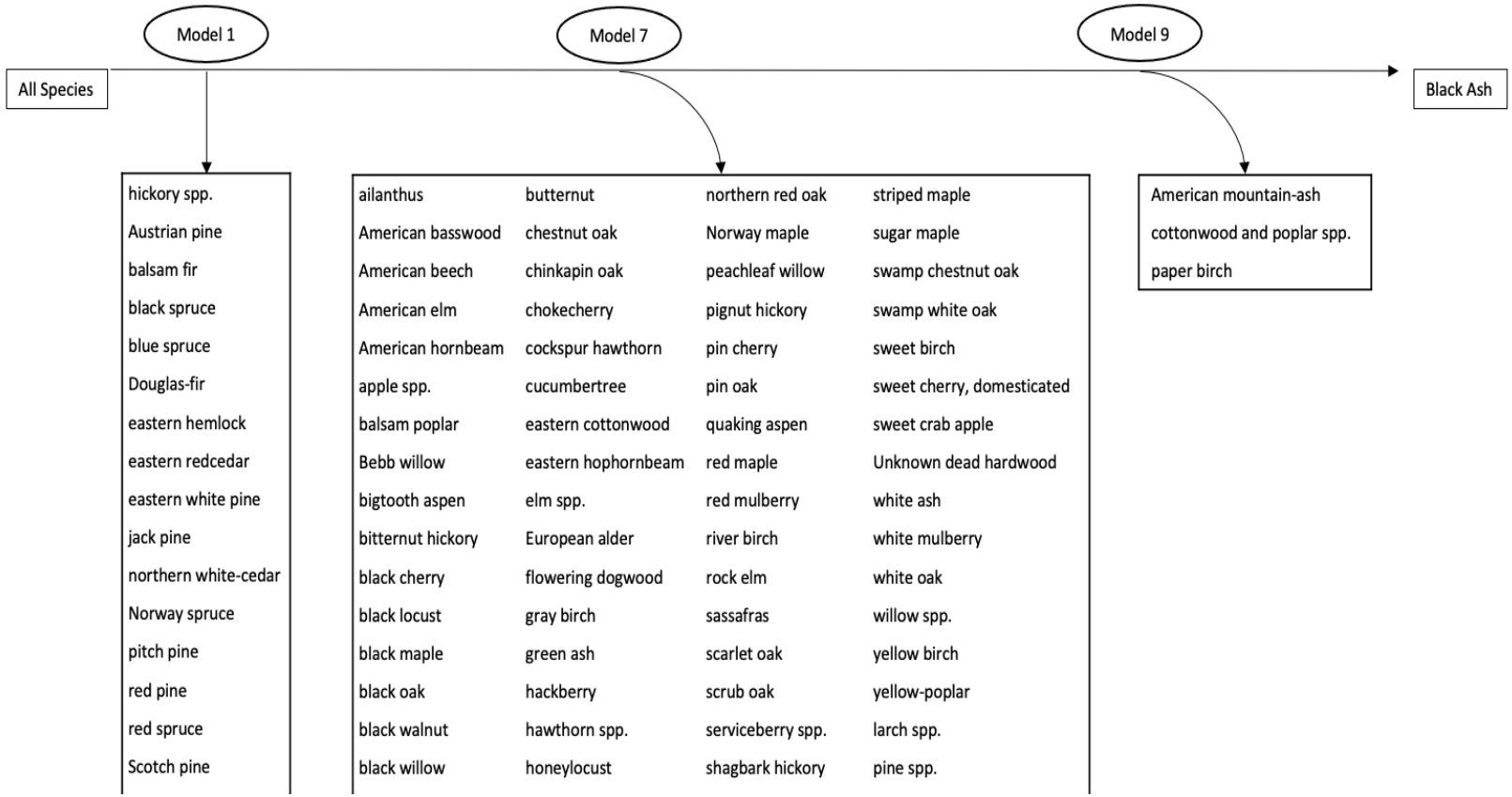


Figure 6. Final set of models chosen to classify black ash. Each model represents a binary choice between the group of species that contains black ash and all unclassified non-black ash species.

2.4.2 Spatial Patterns in Accuracy

In the United States black ash is most prevalent in the upper Great Lakes states (Figure 1). While black ash has areas of prevalence in northern New England, it is rarely found with such high importance on the landscape outside of Central and Northern Minnesota, Northern Wisconsin. Our predictions follow the same pattern, with ash classification occurring primarily in Central and Northern Minnesota, Northern Wisconsin, and the western Upper Peninsula of Michigan (Figure 7A). Our greatest area of predicted ash density is within Minnesota. Less frequent ash predictions were present in Michigan's lower peninsula, New York, and northern New England. Like the overall presence of ash, the relative rates of false positives and true positives are lower outside of the upper Great Lake region (Figure 7). User's and producer's accuracy are both lower in the northeastern states (Figure 8), which may suggest that there is some regional variation in black ash classification signature that is not being captured. The high density of black ash plots in the upper Great Lakes results in a classifier that is slightly biased to predicting black ash based on the characteristics of that region. The greatest area of false positive classifications also occurred in the upper Great Lakes region, primarily in Northern Minnesota (Figure 7C). When summarized by region, user's and producer's accuracy showed the highest values in Minnesota and Wisconsin and lower performance in other states (Figure 8), which is partly a function of how many more ash plots are located in Minnesota and Wisconsin.

2.4.3 Predictor Importance

As described, selected predictors were allowed to vary between models, and for each model the top 20 predictors were chosen, in addition to slope, plan (planc), and profile (profc). In total the final chain used 47 unique predictors, with 28 predictors being used in only a single model. The top performing features in model 1 were the annual standard deviation of NDVI and the EVI conifer mask described above (Table 7). For both models 7 and 9 the top performing predictors were moisture-related indices during September including mean NDMI, minimum $MNDWI_G$, and mean NBR (Table 7). Many of the top 20 predictors still had relatively low importance values and no easily interpretable reason for inclusion. None of the selected predictors were derived from PALSAR data, and the slope and curvature metrics were the least important features in all three models.

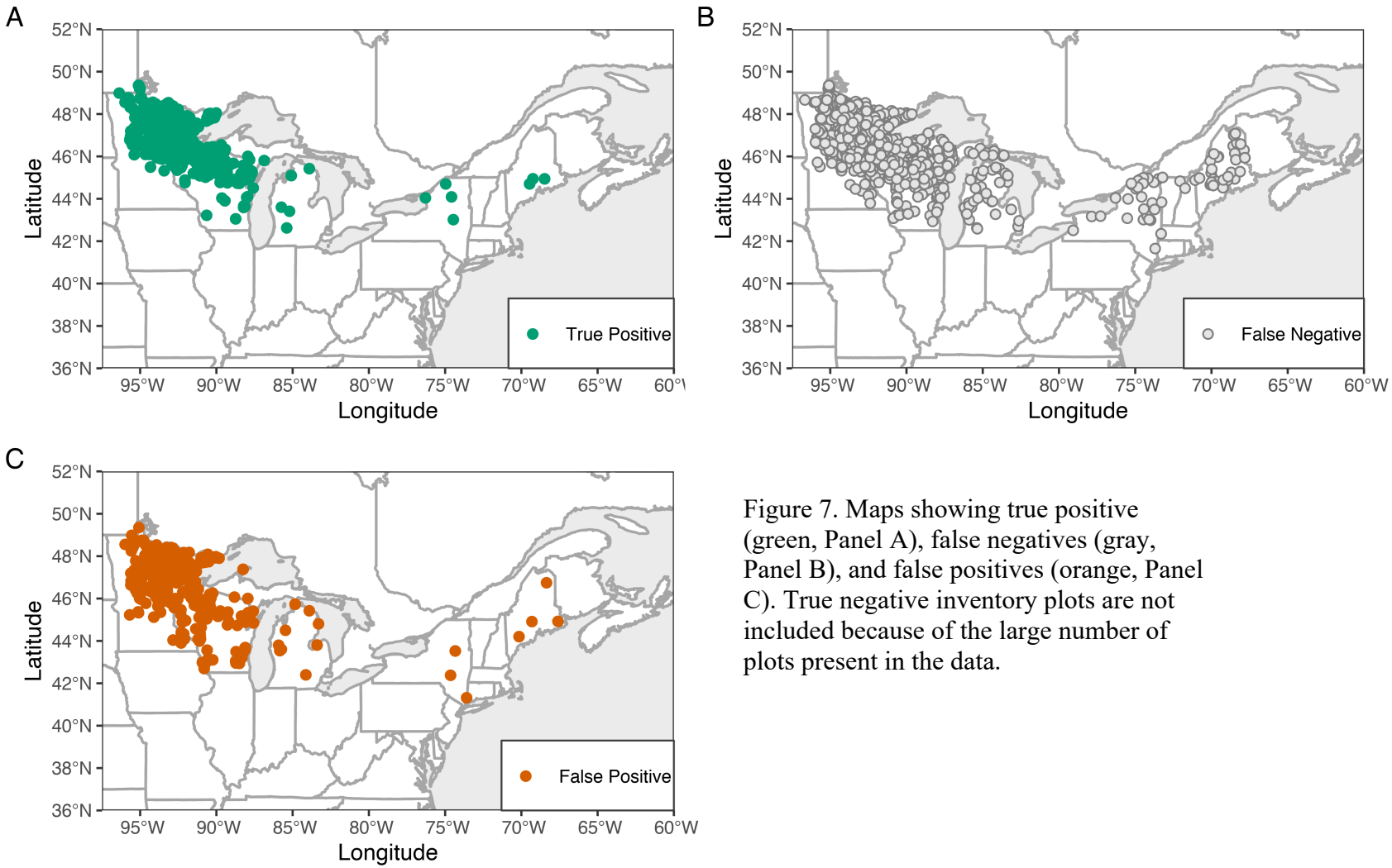


Figure 7. Maps showing true positive (green, Panel A), false negatives (gray, Panel B), and false positives (orange, Panel C). True negative inventory plots are not included because of the large number of plots present in the data.

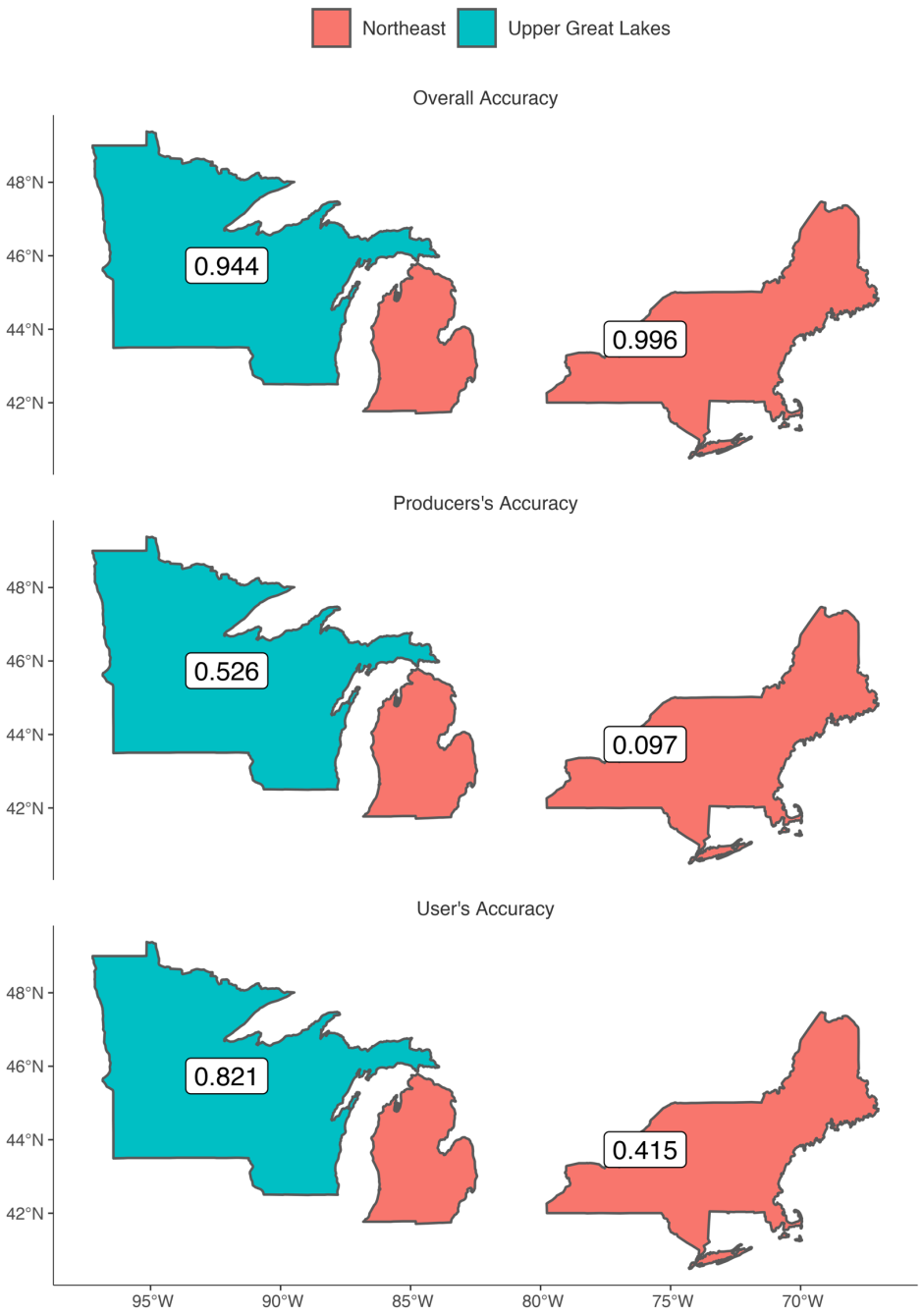


Figure 8. Regional summaries of the full dataset (training and test) for Overall, Producer's, and User's Accuracy in black ash classification. The response of overall accuracy is dominated by the proportion of black ash plots in each region, with fewer black ash plots present (as in the highlighted eastern states), overall accuracy can be very high because even a 100% false negative rate would have little influence if black ash represents only 2% of the plots.

Table 7. Feature importance for the 23 predictors selected for each of the chosen models. A higher feature importance indicates that a predictor contributed more to classification in the model.

Rank	Model 1 Predictors	Model 7 Predictors	Model 9 Predictors
1	Std. Dev. Annual NDVI (0.18)	Mean Sept. NDMI (0.14)	Min. Sept. $MNDWI_G$ (0.11)
2	EVI Conifer Mask (0.13)	Min. Sept. $MNDWI_G$ (0.12)	Mean Sept. NBR2 (0.10)
3	Mean Annual SWIR1 (0.09)	Min. Early Growing Season NIR (0.08)	Mean Sept. NDMI (0.10)
4	Std. Dev. Annual NIR (0.07)	Min. Peak Summer Red (0.06)	Max. Dormant SWIR1 (0.08)
5	Mean Dormant $MNDWI_G$ (0.07)	Min. Peak Summer Blue (0.05)	Min. Sept. NBR2 (0.07)
6	Mean Peak Summer $MNDWI_M$ (0.05)	Min. Late Growing Season NIR (0.04)	Max. Early Growing Season SWIR2 (0.07)
7	Max. Early Growing Season SWIR2 (0.05)	Max. Sept. NBR2 (0.04)	Min. Early Growing Season NIR (0.05)
8	Min. Late Growing Season $MNDWI_G$ (0.04)	Min. Peak Growing Season Blue (0.04)	Min. Peak Summer Red (0.05)
9	Max. Annual NDVI (0.04)	Min. Annual Red (0.04)	Min. Annual Red (0.04)
10	Max. Mid. Growing Season NIR (0.03)	Min. Peak Summer NIR (0.04)	Min. Peak Growing Season Blue (0.03)
11	Mean Mid. Growing Season $MNDWI_M$ (0.03)	Min. Annual Blue (0.03)	Min. Sept. Green (0.03)
12	Min. Sept. SWIR1 (0.03)	Min. Dormant NIR (0.03)	Min. Peak Summer Blue (0.03)
13	Max. Dormant SWIR1 (0.02)	Min. Sept. SWIR1 (0.03)	Min. Annual Blue (0.03)
14	Min. Early Growing Season SWIR1	Min. Peak Summer Green	Min. Dormant Red

	(0.02)	(0.03)	(0.03)
	Max. Peak Summer NBR2	Min. Sept. Green	Min. Sept. NIR
15	(0.02)	(0.03)	(0.03)
	Max. Growing Season SWIR1	Min. Sept. Blue	Min. Early Growing Season Green
16	(0.02)	(0.03)	(0.02)
	Max. Late Growing Season NIR	Min. Peak Summer SWIR2	Max. Sept. NBR2
17	(0.02)	(0.03)	(0.02)
	Max. Peak Summer SWIR1	Min. Sept. Red	Min. Sept. Red
18	(0.02)	(0.03)	(0.02)
	Max. Sept. SWIR1	Min. Dormant Blue	Min. Sept. Blue
19	(0.02)	(0.03)	(0.02)
	Max. Mid. Growing Season SWIR1	Min. Mid. Growing Season Blue	Min. Dormant Blue
20	(0.02)	(0.03)	(0.02)
	slope	slope	slope
21	(0.01)	(0.02)	(0.02)
	profc	planc	planc
22	(0.01)	(0.02)	(0.01)
	planc	profc	profc
23	(0.01)	(0.02)	(0.01)

2.5 Discussion

2.5.1 Binary Classification through a Chain Classifier

Classifying a single species across a broad landscape presents a set of challenges related to the number of alternative classes, and an expected increase in variance within each of those classes as the geographic range increases. In this research I presented a chain classifier to break the 99 classes into a series of binary classification problems. The classification chain is similar to the decision tree classification approach described in this region as early as 1995 (Wolter et al., 1995). In that work the authors used multiple images from a single footprint and created a decision tree classifier built from phenological knowledge of species within the footprint. This manual approach to differentiating species becomes more difficult as the number of species increases and as the phenological timing varies significantly within a larger mapping area. To account for the added complexity, I used an adaptive approach to the classifier length and species cohorts within each classifier step. Rather than attempt to define cohorts of species whose behavior could be expected to remain fixed geographically and phenologically, the potential species groups were created using hierarchical clustering on a distance matrix between all points within each group. However, for each set of species cohorts used there is a probability of misclassification during training and application. As the number of classifiers in the chain increases, the total probability of misclassification also increases. This explains the relationship in Figure 5, where the longest classifier chains showed the worst overall performance in classifying black ash, while with very short chains (1-2 classifiers), the species cohorts have too much internal variance to provide good binary classification. Depending on the exact classification links chosen, the trade-off between internal variance from larger cohorts, and increased probability of misclassification from more classifiers seems to result in maximum performance in the 3-4 classier chains. The power of this adaptive approach can be seen in the first two classifiers in the chain. The first classifier is the clearest split between forest species, evergreen and deciduous, and would probably be the first step in any chain classification approach. The second model in the chain selects many, but not all, remaining hardwoods and separates them from the cohort containing black ash (Figure 6). It is unlikely that any exploratory analysis of the data would have revealed this particular set of species. In fact, manual splitting may have purposefully included species from this list in the ash cohort, such as species that frequently occur with black ash (red maple and elms (*Ulmus spp*.)). Just as manual classification may have misplaced these species, we saw these other wet-site species were primary sources of classifier confusion (Table 5). The straightforward interpretation of both the species cohorts and predictors of high importance supports the use of this approach.

The nature of a chain classifier means that classification errors in the early steps propagate through to later steps. This was one factor motivating the decision to score classification chains using a metric weighted towards precision. Precision-weighted scores favor classifiers that have fewer false positives, rather than treating false positives and false negatives as equal. This is an important feature because it mimics the assumptions used to generate the species cohort training data. When each potential

classifier was trained it was implicitly assumed that all previous classifiers had 100% accuracy and no species from the previous non-ash cohort were included in the training set. This assumption only went into the training of the models. During chain selection and prediction, the final classifications were the product of all the probability predictions for the presence of ash. By weighting chain selection towards higher precision predictions, the training conditions were more faithfully recreated. Taking the product of the individual model predictions contributed to the increased uncertainty and poor performance in longer chains (Figure 5).

2.5.2 User's and Producer's Accuracy

Selection of the $F_{0.5}$ precision-weighted metric for model selection and threshold setting results in weighting the mapping product towards user's accuracy over producer's accuracy. Producer's accuracy is analogous to the recall of a model, what percentage of the true ash plots were labeled as ash. User's accuracy is analogous to the precision of a model, what percentage of plots classified as ash were observed as ash. In general, optimizing for one value will result in a decrease in the other value, especially in very imbalanced datasets such as this. Both measures are useful as modeling metrics, but the choice of which to optimize is a function of the goal of the mapping project. Classification highly optimized for producer's accuracy will have higher rates of false positives. These maps will be more likely to capture the full extent of the class of interest, but the increase in false positives means there will be lower confidence about positive labels in the final product. Targeting increases in user's accuracy will reduce the occurrence of false positives but increase the number of false negatives. In practice this means fewer areas of ash will be mapped, but there will be higher confidence about the ground truth of those classifications.

From a land manager's perspective dealing with EAB, there are practical implications for how a map will be used depending on if it was optimized for user's or producer's accuracy. A map with higher producer's accuracy will represent something more aligned with the 'worst-case scenario.' Such a map would target how much ash may be on the landscape, and what its general distribution is over the landscape or within certain watersheds. A map with a higher user's accuracy represents a higher precision map, where not all ash is included, but what is on the map is very likely to exist on the ground. This map would be ideal for determining where to focus inventory and restoration resources to not waste travel and fieldwork visiting false positives. I chose to optimize my predictions for the latter case to enable better allocation of resources for preemptive entry or responsive management to EAB. However, this choice does come at the expense of increasing false negatives and missing potential restoration sites or underestimating cumulative impact. Ideally a map product such as this could be interacted with to show a range of user's and producer's accuracy to provide both advantages.

2.5.3 Landscape-Scale Spatial Autocorrelation

Not presented above was a previous attempt to better isolate phenological and species shifts throughout the region using a set of spatially relative 'neighborhood' features. Each of the original Landsat and PALSAR bands and indices, and the September, annual, peak,

and growing season summary features were compared to other plots within 100 km to determine their distance from the mean and maximum of nearby plot measures and the z-score of the plot was calculated relative to the nearby plots. These predictors were added to the 422 described above and used to calculate the same type of species-level distance matrix and species-cohort clusters. With the increased predictors the result was 12 rather than 9 clusters containing ash. Models to classify these potential cohorts were developed as above, resulting in 12 independent and possibly varying sets of 23 predictors, one for each model. Each model was then trained to classify the respective S_{abs} and S_{pres}. With the now 66 fit models I was able to evaluate 2,048 combinations of classifiers that could result in a binary classification of black ash. A classifier of length 4 was chosen as the best chain classifier. To avoid misinterpretation as my true results, I will not recreate confusion matrices here. Results were similar in magnitude to the results in Table 6 with a user's accuracy of 61.70% and a producer's accuracy of 26.66% for the withheld dataset. Examination of the spatial distribution of predictions and errors revealed information exchange between the spatially blocked training and testing datasets, and landscape-level autocorrelation of black ash prediction (Figure 9).

The clustered nature of black ash's landscape prevalence (sparse across much of its range with areas of expansive landscape coverage) presented a different risk in classification. The implementation of spatial blocking for selecting training and testing data was intended to reduce the effect of spatial autocorrelation on the results. As designed the blocking used would prevent the influence of training and testing data being near-immediate neighbors. The clustered distribution of black ash combined with the use of neighborhood-scale predictors reduced the effectiveness of spatial blocking. Even with the spatial blocking the classifier over-valued the importance of the neighborhood predictors (not presented, but almost all chosen features were neighborhood metrics). Most black ash plots were in the upper Great Lakes region, and most of those plots had high similarity with their neighborhood plots due to the clustering of black ash in that region. Rather than the neighborhood predictors being used to differentiate isolated pockets of ash from surrounding non-ash (as was the design intent), the classifier learned that when neighborhood plots are very similar, they must be black ash. This effect was further compounded by information exchange between spatial blocking (neighborhoods could span multiple spatial blocks). Landscape-level spatial autocorrelation could potentially explain why no ash was predicted outside of the upper Great Lakes region, and why the false positives were also concentrated in the same region (Figure 9).

The clustering of predictions and false positives in one region is not completely conclusive of auto-correlation driving ash prediction. With a 20% recall rate (amount of labeled ash that was classified as ash), we would expect predictions to be concentrated around high densities of observed black ash plots. Additionally anecdotal spot-checks of false positives in the upper Great Lakes region indicate that some subplots of CFI inventory plots observed as 'not ash' were nested within plots containing ash observations. It is likely that at least some of these false positives are the result of a single subplot's species composition not being representative of the forest composition over the entire pixel's coverage. This is not unexpected as the subplots have a radius of only 7.32 m and are not co-located with pixel centers. Again, this type of false positive should be

expected in areas where CFI plots are most likely to be observed as ash, the upper Great Lakes.

The main results presented in this study were a refit of each of the modeling steps excluding neighborhood predictors from consideration. The results presented here provide much stronger evidence that the initial black ash classifier chain was likely the result of landscape-level spatial autocorrelation. A known effect of spatial autocorrelation is that classifiers score higher on performance metrics than their true performance (Ploton et al., 2020). This type of error can be subtle as model performance is not necessarily so good as to raise suspicion, and predictions match expectations (high ash predictions in areas with high ash prevalence). The classifier chain without neighborhood predictors was not subject to information exchange between spatial blocks and so represents a more independently trained and evaluated model.

2.5.4 Predictor Selection

The predictors showing the most importance for each model have easily interpretable impacts on the model. In model 1 the annual standard deviation of NDVI and the EVI conifer mask both highlight inventory plots where these growing season and dormant season vegetation indices show less variation than other plots. This is a clearly differentiating signal between evergreen and deciduous species, which was the cohort division between the two species cohorts. The most important predictors for models 7 and 9 are moisture-based indices during the September period. The September mean NDMI was the first and third most important predictors for models 7 and 9, respectively. While the September minimum $MNDWI_G$ was the second and first most important predictors for models 7 and 9, respectively. The NDMI and $MNDWI_G$ are both indicators of vegetation moisture content and $MNDWI_G$ is sensitive to standing water (Gao, 1996). Their inclusion in these models is likely capturing the earlier senescence of black ash relative to many other hardwood species and fall water level rebound. The fact that both indices are included would suggest that they are not well correlated, as my prediction selection procedure removed colinear predictors, and selecting one as indicative of vegetation moisture would have resulted in a loss of important information. The other top predictor in model 9 was NBR2, which can highlight moisture in vegetation and is also sensitive to surface water, as it is closely related to the land surface water index (Ji et al., 2011), again potentially highlighting early senescence or water level rebound. The heavy weighting of moisture-based indices in models 7 and 9 likely contributed to high prevalence of wet-site species in the false positive plots (Table 5). Unexpectedly none of the models relied on the MODIS-adjusted phenological day of season variables, suggesting that the phenology of black ash or the wetland hydroperiod is relatively consistent across the range of training data.

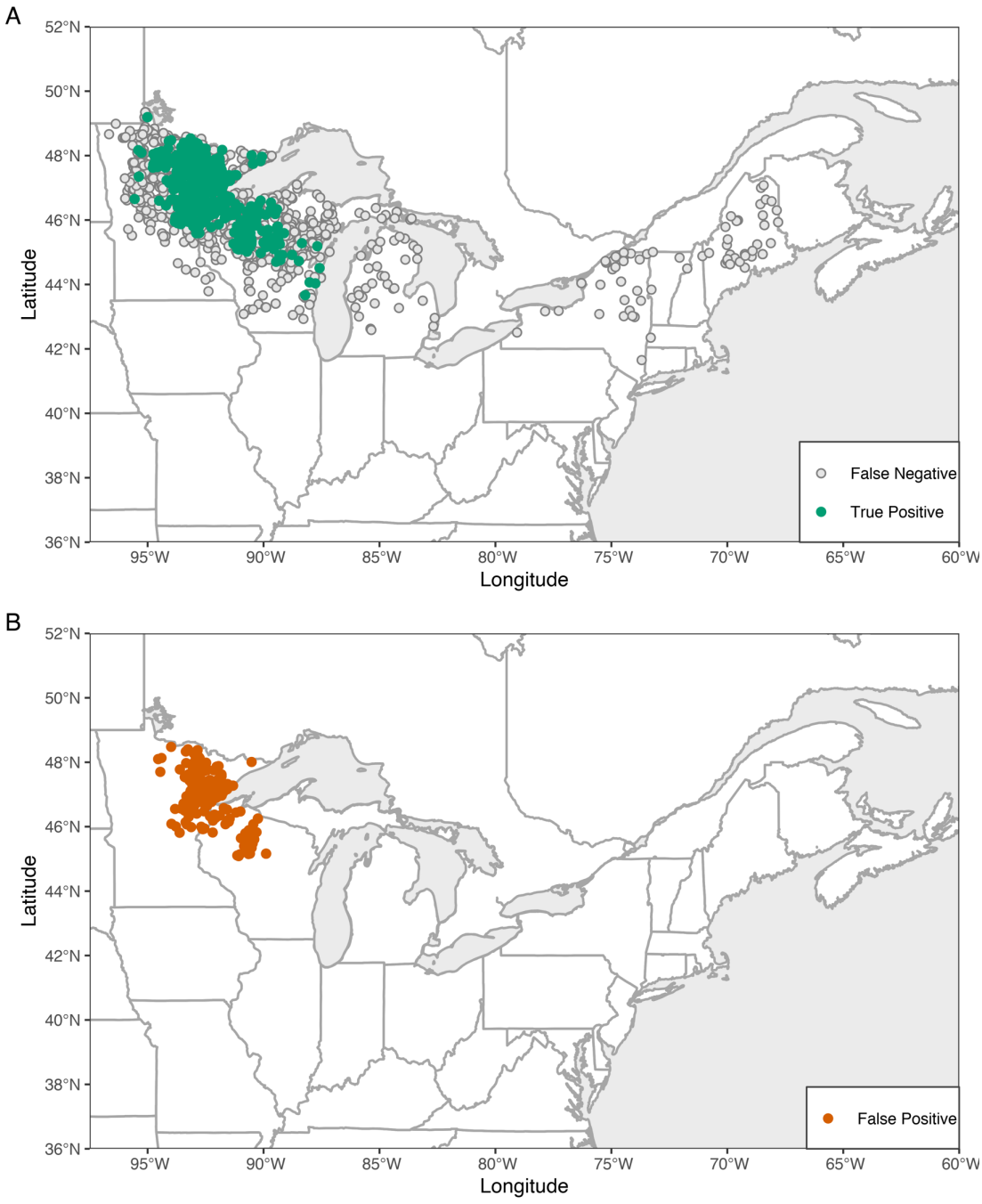


Figure 9. Observed and predicted black ash using classifiers suffering from spatial auto-correlation showing true positive (top panel, green), false negative (top panel, gray), and false positive (bottom panel, orange) classifications of black ash.

2.5.5 Forest Inventory Plots as Training Data

This project seeks to use existing forest inventory data to train a mapping classifier across a large geographic area. Using existing forest inventory data was an approach that enabled a larger and more geographically diverse set of training data than could have otherwise been collected. Using field data not purposely collected for remote sensing can lead to four challenges: the data are not collected at the scale of the minimum mapping unit; spatial accuracy of sample location is not always a primary concern; the quality/quantity of information collected may not match project requirements (e.g., likely no pictures for confirmation, or target classes must be inferred from other collected information); and the samples may not be drawn from a rigorous sampling design. Using well-designed forest inventories for training data limits concerns around data quality and rigorous sampling design. Forest inventories are generally statistically robust sampling designs because they are attempts to accurately represent the landscape through a series of discrete sample points (Bechtold and Patterson, 2015), and I do not need to infer any target classes (species) because they are inherent in the nature of the data collected. However, the design of field data collection, the spatial accuracy, and the interaction between these two issues are concerns for using existing forest inventory data in this project. These two issues interact because the smaller ground sample is likely represented within a single pixel. Significant error in geolocating the points may mean that a different pixel is entirely associated with the ground information, and there is no guarantee that even within the correct pixel the ground sample represents the entirety of the pixel. The largest source of inventory plots, the USFS FIA, samples ~0.067 ha in four circular plots. Coarser safeguards for spatial accuracy were implemented by comparing sample point locations to an existing forest mask to discard sample data that have obviously erroneous coordinates. However, there is no guarantee of correct location information for any individual inventory plot used without visiting the field site and recreating the inventory data for comparison.

The data in Table 8 provides an opportunity to evaluate the impact of differing forest inventory systems on classification accuracy. The table presents the percentage of plots from each data source for total, ash, and non-ash plots. Each grouping of plots is compared to a corresponding error type. Total proportion of plots is compared to total error, ash plots are compared to false negatives, and non-ash plots are compared to false positives. These comparisons highlight whether different inventory systems and plot designs contribute more or less error that would be expected. The Superior National Forest and St. Louis County inventory data were both collected using variable radius plots. In this type of inventory plot the limiting distance from plot center for a tree to be included in the inventory plot is a proportional function of its diameter. These inventory sources contributed more misclassifications than expected based on the total proportion of plots they represent. Specifically, Table 8B and C show that these plots were more likely to be classified as ash (higher than expected rate of false positive and lower than expected rate of false negatives). This may be a result of smaller effective plot diameters in stands with small mean diameters. The larger fixed-radius plot size of the USFS FIA data appears to have had the opposite effect with an increase in false negatives and a decrease in false positives relative to expected proportions.

2.5.6 Comparison to Other Efforts

Previous efforts at mapping black ash and other co-occurring species have occurred in Minnesota and Wisconsin where the species is more common on the landscape (Baumann et al., 2017; Engelstad et al., 2019; Host et al., 2020; Wolter et al., 1995; Wolter and Townsend, 2011). Wolter et al. (1995) presented a decision tree approach to classifying hardwoods in northern Wisconsin. Their black ash classification was dependent on the early fall senescence, an early example of using this signal to classify black ash. The accuracy results presented are for a black ash-wet conifer forest type, which is not directly comparable to this work because there are a range of possible dominant species mapped to the category containing black ash. The work in Host et al. (2020) and Engelstad et al. (2019) are the most directly comparable to this study. Each study was a classification of black ash within Minnesota over a Landsat scene footprint (Engelstad) and the whole state (Host). Engelstad used a similar classification tree approach and found that the most important predictor was a LiDAR-derived compound topographic index followed by the tasseled-cap disturbance index, differenced NBR, and growing season NDMI. The final product had a user's accuracy of 75.3% and a producer's accuracy of 95.3%. The results exceed those presented here, but classification was done over only a single Landsat scene footprint, reducing the phenologic and species diversity present in the mapping region.

Host et al. (2020) worked across the state of Minnesota and used harmonic curves fit to multi-year time series of NDVI. Results were evaluated on USFS FIA plots in Minnesota and presented for two ash species combined and at the species level. Overall user's accuracy for the presence of black or green ash was 32% and the producer's accuracy was 64%. User's accuracy was not provided at the species level, but producer's accuracy for black ash was 72%.

The work in Wolter and Townsend (2011) was not directly focused on mapping black ash, but included it as one of the species of interest. The paper presented results confined to a single Landsat scene footprint using principal-least squares regression on purpose-collected field data in northern Minnesota to model species-level basal area. They did produce dominant species estimates and for black ash presented a producer's accuracy of 79% and a user's accuracy of 92%. These results demonstrate the value of ground data that was collected for remote sensing purposes. Their dominant class cover maps also used a data-derived threshold for determining proportion of basal area necessary to be considered a black ash canopy. Their threshold (39%) was comparable to the threshold set in the final step of my classifier (32%), though more work would be necessary to determine if this basal area threshold represents a change in canopy spectral signal or texture indicative of black ash.

Of these efforts only Host et al. (2020) approaches a large scale, and the classification work there may be influenced by the same phenomenon described in Landscape-Scale Spatial Autocorrelation. Seasonal NDVI amplitude and NDVI median would be expected to have a latitudinal gradient within Minnesota which corresponds to the importance

Table 8. Percentage of plot allocation by inventory data source for all plots (A), ash plots (B), and non-ash plots (C). For each grouping of plots (non-ash, ash, all) an associated percentage of misclassified plots is presented as well as the difference between the percentage of plots and the percentage of misclassified plots.

A) All Plots and Any Errors

Source	**Percentage of Total Plots**	**Percentage of All Errors**	**Relative Difference**
US Forest Service FIA	90.45%	63.14%	-27.31%
Wisconsin Forest Inventory	4.69%	4.25%	-0.43%
St. Regis-Mohawk	0.01%	0.09%	0.09%
Superior National Forest	0.59%	4.39%	3.80%
St. Louis County	4.27%	28.12%	23.85%

B) Ash Plots and False Negatives

Source	**Percentage of Total Ash Plots**	**Percentage of False Negatives**	**Relative Difference**
US Forest Service FIA	47.39%	60.40%	13.00%
Wisconsin Forest Inventory	3.25%	4.18%	0.93%
St. Regis-Mohawk	0.05%	0.11%	0.06%
Superior National Forest	6.24%	4.61%	-1.63%
St. Louis County	43.06%	30.70%	-12.36%

C) Non-Ash Plots and False Positives

Source	**Percentage of Total Non-Ash Plots**	**Percentage of False Positives**	**Relative Difference**
US Forest Service FIA	93.24%	74.70%	-18.54%
Wisconsin Forest Inventory	4.78%	4.56%	-0.22%
St. Regis-Mohawk	0.00%	0.00%	0.00%
Superior National Forest	0.22%	3.48%	3.26%
St. Louis County	1.76%	17.27%	15.51%

value gradient of black ash within the state. A spatial assessment of accuracy was not presented by the authors, so it is unclear how the risks of spatial autocorrelation or north-south phenological gradients within MN were evaluated or addressed. It may have been possible to use an approach that was shown to correct for latitudinal effects on black ash phenological gradients. In Wisconsin the approach of dynamic time-warping, which adjusts Landsat image dates based on MODIS-derived phenology was shown to more accurately capture the distinct end-of-season signature of black ash (Baumann et al., 2017).

2.6 Potential Improvements and Future Work

A critical analysis of this research shows areas where inputs or processing could be updated to improve results, and where future work on this topic may wish to focus. The response metric of plots classified as ash could have made use of more information available in the FIA database. The initial labeling in this study assumed that a stand was black ash if the total basal area of the stand was more than 50% of black ash. However, the FIA dataset also describes portions of the plot as conditions, and this information could have been used to more accurately determine if not just the basal area of the plot was primarily black ash, but if that canopy condition constituted the majority of the ground area of the plot. As an example, a plot with a small portion under black ash and the vast majority under advanced poplar regeneration could still have been labeled as black ash because the majority of canopy-tree basal area would have been black ash. Further labeling the plots based on the expected classification threshold (32%) may have resulted in a better final classification. The classifiers were trained using a 50% threshold to label inventory plots as black ash, but the final threshold setting and the work in Wolter and Townsend (2011) both indicated that a lower proportion of black ash may create a strong enough signal for identification.

The Landsat imagery used in this study were from the Top-of-Atmosphere (TOA) collection and were not corrected to represent surface reflectance. Correction for atmospheric distortion is a complicated process (Vermote et al., 2016) and was outside the realm of this research. The TOA collection was used because the final research product will apply the developed classifiers to the entire range of black ash, covering portions of the United States and Canada, and surface-reflectance data were not readily available for the Canadian portion of the range, excluding it from consideration in the models. However, since the model work presented was completed, Landsat surface-reflectance data has become readily available for the Canadian portion of the range of black ash and could provide a valuable improvement over TOA images (Vermote et al., 2016). There were no significant contributions from the raw or differenced PALSAR data and the forced inclusion of topographic metrics did not contribute significantly to any of the models. While the PALSAR data was not coincident with the inventory work, it was still expected to identify areas that were wet due to geomorphology. These areas should not have changed in the few years between acquisition of the imagery and the inventory data. The topographic features did not include any additional derived indices (such as CTI from Englestad et al. (2019) and Host et al. (2020)) and was at 30 m resolution. It is

difficult to say with certainty which of these factors contributed to their lack of predictive power.

Future work on this classification is intended to lead to a publicly available map covering the entire range of black ash. The first step towards that is to test the performance of models fit using only the top performing predictors. If results are consistent with those presented here, processing could be done more rapidly and easily deployed on a platform such as Google Earth Engine (Gorelick et al., 2017). Aside from rapid deployment making use of an online platform rather than a static product would allow users to set their own confidence threshold, meaning the map could provide both a worst-case scenario of potential ash loss as well as a high-confidence estimate of ash stand locations. Though likely to not be incorporated into this version of the final map, it would be useful to explore weighting each individual classifier's contribution to the final prediction using the accuracy of that model. In this way models with lower certainty would be less likely to be selected for the final chain or would contribute less to the predictions.

2.7 Conclusion

This research demonstrates how unsupervised (hierarchical clustering) and supervised (random forests) machine learning approaches can be combined to classify a single relatively rare species over a large geographic area. The result is an adaptive classifier chain that finds the optimal set of multi-species cohorts to distinguish black ash from 99 other species over the entire United States portion of its range. The final 3-link classifier chain has logically sound species groupings and primary predictors. Initially most conifers are removed using predictors related to the annual variation in NDVI/EVI. Next most hardwood species are removed by comparing metrics of vegetation moisture content, and finally the remaining hardwood species are filtered out using vegetation moisture content and an index sensitive to the presence of water. The classifiers were optimized for user's accuracy to provide a tool for focusing restoration resources rather than evaluating worst case scenarios and provided a user's accuracy of 65% on withheld data with no obvious spatial trends in classification errors. Additionally, this research highlights the potential risk of spatial autocorrelation in land cover classifiers. While future work could potentially improve on this model, the most pressing next steps are to produce classifications across the entire species range of black ash.

2.8 References

Ahern, F., Brisco, B., Murnaghan, K., Lancaster, P., Atwood, D.K., 2018. Insights Into Polarimetric Processing for Wetlands From Backscatter Modeling and Multi-Incidence Radarsat-2 Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 3040–3050. https://doi.org/10.1109/JSTARS.2018.2850155

Ahlgren, C.E., 1957. Phenological observations of nineteen native tree species in northeastern Minnesota. Ecology 38, 622–628.

ASF DAAC, 2015. ALOS PALSAR_Radiometric_Terrain_Corrected_high_res; Includes Material © JAXA/METI 2007.

Baumann, M., Ozdogan, M., Richardson, A.D., Radeloff, V.C., 2017. Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. Int. J. Appl. Earth Obs. Geoinformation 54, 72–83. https://doi.org/10.1016/j.jag.2016.09.005

Beaudoin, A., Bernier, P.Y., Guindon, L., Villemaire, P., Guo, X.J., Stinson, G., Bergeron, T., Magnussen, S., Hall, R.J., 2014. Mapping attributes of Canada's forests at moderate resolution through kNN and MODIS imagery. Can. J. For. Res. 44, 521–532. https://doi.org/10.1139/cjfr-2013-0401

Bechtold, W.A., Patterson, P.L., 2015. The Enhanced Forest Inventory and Analysis Program National Sampling Design and Estimation Procedures (No. SRS-GTR-80). U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC. https://doi.org/10.2737/SRS-GTR-80

Bolton, N., Shannon, J., Davis, J., Grinsven, M., Noh, N., Schooler, S., Kolka, R., Pypker, T., Wagenbrenner, J., 2018. Methods to Improve Survival and Growth of Planted Alternative Species Seedlings in Black Ash Ecosystems Threatened by Emerald Ash Borer. Forests 9, 146. https://doi.org/10.3390/f9030146

Bourgeau-Chavez, L., Riordan, K., B., R., Miller, N., Nowels, M., 2009. Improving Wetland Characterization with Multi-Sensor, Multi-Temporal SAR and Optical/Infrared Data Fusion, in: Jedlovec, G. (Ed.), Advances in Geoscience and Remote Sensing. InTech. https://doi.org/10.5772/8327

Bourgeau-Chavez, L.L., Endres, S., Powell, R., Battaglia, M.J., Benscoter, B., Turetsky, M., Kasischke, E.S., Banda, E., 2017. Mapping boreal peatland ecosystem types from multitemporal radar and optical satellite imagery. Can. J. For. Res. 47, 545–559. https://doi.org/10.1139/cjfr-2016-0192

Bourgeau-Chavez, L.L., Kowalski, K.P., Carlson Mazur, M.L., Scarbrough, K.A., Powell, R.B., Brooks, C.N., Huberty, B., Jenkins, L.K., Banda, E.C., Galbraith, D.M., Laubach, Z.M., Riordan, K., 2013. Mapping invasive Phragmites australis in the coastal Great Lakes with ALOS PALSAR satellite imagery for decision support. J. Gt. Lakes Res. 39, 65–77. https://doi.org/10.1016/j.jglr.2012.11.001

Burkett, V., Kusler, J., 2000. CLIMATE CHANGE: POTENTIAL IMPACTS AND INTERACTIONS IN WETLANDS OF THE UNITED STATES [1]. JAWRA J. Am. Water Resour. Assoc. 36, 313–320. https://doi.org/10.1111/j.1752-1688.2000.tb04270.x

Davis, J.C., Shannon, J.P., Van Grinsven, M.J., Bolton, N.W., Wagenbrenner, J.W., Kolka, R.K., Pypker, T.G., 2019. Nitrogen cycling responses to simulated emerald ash borer infestation in Fraxinus nigra-dominated wetlands. Biogeochemistry 145, 275–294. https://doi.org/10.1007/s10533-019-00604-2

Diamond, J.S., McLaughlin, D.L., Slesak, R.A., D'Amato, A.W., Palik, B.J., 2018. Forested versus herbaceous wetlands: Can management mitigate ecohydrologic regime

shifts from invasive emerald ash borer? J. Environ. Manage. 222, 436–446. https://doi.org/10.1016/j.jenvman.2018.05.082

Dormann, C., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dowle, M., Srinivasan, A., 2019. Data.Table: Extension of `data.Frame` (Manual).

Engelstad, P.S., Falkowski, M.J., D'Amato, A.W., Slesak, R.A., Palik, B.J., Domke, G.M., Russell, M.B., 2019. Mapping black ash dominated stands using geospatial and forest inventory data in northern Minnesota, USA. Can. J. For. Res. 49, 892–902. https://doi.org/10.1139/cjfr-2018-0481

Gao, B.-C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-Scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gray, A.N., Brandeis, T.J., Shaw, J.D., McWilliams, W.H., Miles, D., 2012. Forest Inventory and Analysis Database of the United States of America (FIA) 8.

Haack, R.A., Jendek, E., Liu, H., Marchant, K.R., Petrice, T.R., Poland, T.M., Ye, H., Lansing, E., 2002. The Emerald Ash Borer: A New Exotic Pest in North America 5.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853. https://doi.org/10.1126/science.1244693

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Herms, D.A., McCullough, D.G., 2014. Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. Annu. Rev. Entomol. 59, 13–30. https://doi.org/10.1146/annurev-ento-011613-162051

Hijmans, R.J., 2021. terra: Spatial data analysis (manual).

Host, T.K., Russell, M.B., Windmuller-Campione, M.A., Slesak, R.A., Knight, J.F., 2020. Ash Presence and Abundance Derived from Composite Landsat and Sentinel-2 Time Series and Lidar Surface Models in Minnesota, USA. Remote Sens. 12, 1341. https://doi.org/10.3390/rs12081341

Isaacson, B.N., Serbin, S.P., Townsend, P.A., 2012. Detection of relative differences in phenology of forest species using Landsat and MODIS. Landsc. Ecol. 27, 529–543. https://doi.org/10.1007/s10980-012-9703-x

Iverson, L., Prasad, A.M., Matthews, S.N., Peters, M., 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. For. Ecol. Manag. 254, 390–406. https://doi.org/10.1016/j.foreco.2007.07.023

Jenkerson, C.B., Maiersperger, T., Schmidt, G., 2010. eMODIS: A User-Friendly Data Source (Report No. 2010–1055), Open-File Report. https://doi.org/10.3133/ofr20101055

Ji, L., Zhang, L., Wylie, B.K., Rover, J., 2011. On the terminology of the spectral vegetation index (NIR − SWIR)/(NIR + SWIR). Int. J. Remote Sens. 32, 6901–6909. https://doi.org/10.1080/01431161.2010.510811

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3

Lang, M., Bourgeau-Chavez, L., Tiner, R., Klemas, V., 2015. Advances in Remotely Sensed Data and Techniques for Wetland Mapping and Monitoring, in: Tiner, R., Lang, M., Klemas, V. (Eds.), Remote Sensing of Wetlands. CRC Press, pp. 79–116. https://doi.org/10.1201/b18210-7

Lang, M.W., Townsend, P.A., Kasischke, E.S., 2008. Influence of incidence angle on detecting flooded forests using C-HH synthetic aperture radar data. Remote Sens. Environ. 112, 3898–3907.

Little, E.L., Viereck, L.A., 1971. Atlas of united states trees:(no. 1146). Conifers and important hardwoods, by EL little, jr. US Government Printing Office.

Looney, C.E., D'Amato, A.W., Palik, B.J., Slesak, R.A., 2015. Overstory treatment and planting season affect survival of replacement tree species in emerald ash borer threatened *Fraxinus Nigra* forests in minnesota, USA. Can. J. For. Res. 45, 1728–1738. https://doi.org/10.1139/cjfr-2015-0129

Loveland, T.R., Dwyer, J.L., 2012. Landsat: Building a strong future. Remote Sens. Environ. 122, 22–29. https://doi.org/10.1016/j.rse.2011.09.022

Mader, S.F., 1991. Forested wetlands classification and mapping: a literature review. NCASI Tech. Bull. USA.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat Surface Reflectance Dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72. https://doi.org/10.1109/LGRS.2005.857030

McFeeters, S.K., 1996. The use of the Normalized Difference Water index (NDW) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Pebesma, E., 2018. Simple features for r: Standardized support for spatial vector data. R J. 10, 439–446. https://doi.org/10.32614/RJ-2018-009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540. https://doi.org/10.1038/s41467-020-18321-y

Prasad, A.M., Iverson, L.R., 2003. Little's range and FIA importance value database for 135 eastern US tree species.

R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.

Read, J., Pfahringer, B., Holmes, G., Frank, E., 2021. Classifier Chains: A Review and Perspectives. J. Artif. Intell. Res. 70, 683–718. https://doi.org/10.1613/jair.1.12376

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881

Rocchio, L.E.P., Connot, P., Young, S., Ramsayer, K., Owen, L., Bouchard, M., Barnes, C., 2018. Landsat benefiting society for fifty years (Report).

Scaramuzza, P., Barsi, J., 2005. Landsat 7 scan line corrector-off gap-filled product development, in: Proceeding of Pecora. pp. 23–27.

Shannon, J., Van Grinsven, M., Davis, J., Bolton, N., Noh, N., Pypker, T., Kolka, R., 2018. Water Level Controls on Sap Flux of Canopy Species in Black Ash Wetlands. Forests 9, 147. https://doi.org/10.3390/f9030147

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199. https://doi.org/10.1016/j.rse.2019.05.018

Story, M., Congalton, R., 1986. Accuracy Assessment: A User's Perspective. REmote Sens. Brief 52, 397–399.

USGS, 2020. Using the USGS Landsat Level-1 Data Product [WWW Document]. URL https://www.usgs.gov/core-science-systems/nli/landsat/using-usgs-landsat-level-1-data-product (accessed 10.27.21).

USGS EROS, 1994. Shuttle Radar Topography Mission 1 Arc-Second Global Digital Elevation Model. https://doi.org/10.5066/F7PR7TFT

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232.

Van Grinsven, M.J., Shannon, J.P., Davis, J.C., Bolton, N.W., Wagenbrenner, J.W., Kolka, R.K., Pypker, T.G., 2017. Source water contributions and hydrologic responses to simulated emerald ash borer infestations in depressional black ash wetlands. Ecohydrol. Ecosyst. Land Water Process Interact. Ecohydrogeomorphology 10, e1862. https://doi.org/10.1002/eco.1862

Van Rossum, G., Drake, F.L., 2009. Python 3 reference manual. CreateSpace, Scotts Valley, CA.

van Zyl, J.J., 2011. Synthetic aperture radar polarimetry. John Wiley & Sons.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wes McKinney, 2010. Data Structures for Statistical Computing in Python, in: van der Walt, S., Jarrod Millman (Eds.), Proceedings of the 9th Python in Science Conference. pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Wilson, E.H., Sader, S.A., 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sens. Environ. 80, 385–396. https://doi.org/10.1016/S0034-4257(01)00318-2

Wolter, P.T., Mladenoff, D.J., Host, G.E., Crow, T.R., 1995. Improved Forest Classification in the Northern Lake States Using multi-temporal landsat imagery. Photogramm. Eng. Remote Sens. 61, 1129–1143.

Wolter, P.T., Townsend, P.A., 2011. Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. Remote Sens. Environ. 115, 671–691. https://doi.org/10.1016/j.rse.2010.10.010

Wu, Q., 2018. GIS and Remote Sensing Applications in Wetland Mapping and Monitoring, in: Comprehensive Geographic Information Systems. Elsevier, pp. 140–157. https://doi.org/10.1016/B978-0-12-409548-9.10460-9

3 Magnitude, Consequences, and Correction of Temperature-Derived Errors for Absolute Pressure Transducers under Common Monitoring Scenarios[1]

3.1 Abstract

Continuous water level monitoring using absolute pressure transducers with onboard datalogging is common practice in hydrologic studies. While there has been some discussion and study of temperature-derived error (TDE), there has not been a systematic evaluation of the problem. We sought to answer three questions: 1) are current best-practices enough to avoid these errors, 2) can laboratory correction be used to correct field data from varying conditions, and 3) what is the scale of the additional uncertainty of the correction procedure? We evaluated the magnitude of such errors under laboratory conditions that mimicked common monitoring scenarios. Using field data, we also demonstrated the impact of TDEs on calculated daily mean water level and diurnal signal decomposition to estimate evapotranspiration (ET). To address instrument and model uncertainty, we fit 1000 possible correction models using a double-bootstrap approach. Correction models fit expected error as a function of water and air temperature and rate of change of air temperature. TDEs were a significant source of error, resulting in recorded data outside of manufacturer-stated instrument uncertainty, with 45% of bootstrap models showing significant but small TDEs under best-practice deployment. Correction equations did introduce additional error, often on a much smaller scale than instrument uncertainty. When tested against a validation dataset, correction equations effectively reduced total measurement uncertainty below instrument uncertainty by up to 65%. The effects of TDEs on case-study field data resulted in 56% of daily mean values outside of instrument error bounds (errors: -1.5–4.2 cm). Our results suggest that a single laboratory correction equation can be used across monitoring scenarios, though we suggest matching deployment conditions as closely as practical. Identification and correction of TDEs are essential to avoid misguided conclusions, downstream analysis, and water resources management.

3.2 Introduction

Hydrologic monitoring for research and management, including monitoring wells, piezometers, and stream gauges, often utilizes submerged pressure transducers to measure water levels. These instruments are easy to deploy, require minimum access, and provide long-term continuous records in an easily analyzed format via on-board datalogging. They have been developed in two primary configurations: absolute (unvented) and gauge (vented). Absolute transducers measure pressure relative to a fixed pressure chamber on one side of a flexible membrane. Gauge transducers measure pressure across a similar membrane with one chamber connected to the atmosphere via a

[1] The material contained in this chapter has been submitted to the journal Hydrological Processes

vent tube. The vent tube in a gauge transducer directly compensates for atmospheric pressure atop the water column. Absolute transducers require a secondary reading of atmospheric pressure for a post-monitoring correction. While the approach for both pressure transducers is simply stated, previous work has concluded that "measurement of the water level…is not simple" (Liu and Higgins, 2015, p. 72). Earlier guidance documents for monitoring water levels with pressure transducers warned that "these systems commonly are not adequately supported by quality-assurance procedures" (Freeman et al. 2004, p. 1). Freeman et al. (2004) goes on to recognize the effects of temperature on pressure transducer function from theoretical and practical standpoints. Even in a closed tank, the relatively straightforward task of measuring water level may be confounded by propagating instrument uncertainty through each step required to arrive at final water levels (Tamari and Aguilar-Chávez, 2010).

Apart from known instrument error, additional uncertainty caused by air and water temperatures, and the rates of temperature change have been theorized and observed in water level measurements. Freeman et al. (2004) highlighted water temperature gradients and rapid temperature fluctuations as potential sources of error, while Cain, Davis, Loheide II, and Butler (2004) described air temperature-derived artifacts within gauge transducers. Within absolute transducers, artifacts were shown to arise from the temperature of the instrument (Moore, Vasconelos, Zech, & Soares, 2016), as well as from differing thermal regimes between barometric and water pressure transducers (Cuevas, Matias Calvo, Little, & Pino, 2010; McLaughlin and Cohen, 2011). Liu and Higgins (2015) and Moore et al. (2016) went on to suggest that all transducers be tested for inaccuracies over the range of temperature measurements and individual correction equations be developed to apply to collected data. That conclusion is supported by others who have found that air temperature-induced errors can be as high as 19% in streamflow monitoring (Cuevas et al., 2010) or 1.5 cm day^{-1} in groundwater monitoring (McLaughlin and Cohen, 2011).

To preempt these errors for absolute transducers, it is recommended as 'best practice' deployment of barometric and water level pressure transducers to place both in similar thermal regimes (i.e. within a dry well at the same depth) (McLaughlin and Cohen, 2011). However, this type of deployment is not always possible due to cost or setting. In other cases, previous deployments may have not recognized the potential magnitude of error, but still represent important data sources. We would expect that different monitoring objectives (*e.g.*, groundwater or river stage height) may be impacted differently with the magnitude of error controlled by the thermal range of the water and air pressure monitoring. To date, studies on these types of errors have been focused on individual deployment scenarios without an attempt to compare the magnitude of errors under various common monitoring scenarios. In this study we defined four such scenarios which are a cross of two factors: thermal setting and deployment practice. We grouped thermal setting into stable (groundwater) and variable (surface water) conditions. Deployment practice was considered as 'best practice' or 'temperature-difference-biased' (transducers deployed without regard to similar thermal settings). The four scenarios were then best-practice groundwater (GW_{best}), temperature-biased groundwater (GW_{bias}), best-practice surface water (SW_{best}), and temperature-biased surface water (SW_{bias}).

Without a systematic comparison between these scenarios, it is difficult to determine the primary source of error, the potential magnitude of error, and the appropriate mitigation or correction. A systematic evaluation can also show if 'best practice' is enough to eliminate temperature-induced errors.

Based on the above theoretical and experimental results, the potential sources of error for absolute transducers are water temperature, air temperature, and the rate of change for both water and air temperatures. Apparent error introduced by temperature differential is likely the response of individual transducers to out-of-phase variation in air and water temperatures. While a vent tube that directly links air and water temperature can be subject to temperature differentials along the length of the tube (Liu and Higgins, 2015), there is no analogous physical process for absolute transducers. This suggests that temperature differentials between loggers are not a direct driver of error for absolute transducers. Asynchronous variation in air and water temperatures can result in a stronger gradient masking the effect of a weaker gradient or aligned gradients amplifying errors. The sequential correction shown in Figure 1 demonstrates water level measurement error based on air temperature (Figure 1A), water temperature (Figure 1B), and water temperature after removing the air temperature error (Figure 1C). There are two possible approaches to the final correction. One is to develop and apply univariate corrections to a reference pressure for each transducer. The second is to develop a single multivariate correction to a fixed water depth for each pair of air and water transducers. The single multivariate correction presents a simpler alternative because it can be performed based only on a known water level without requiring a pressure chamber or other suitable reference pressure.

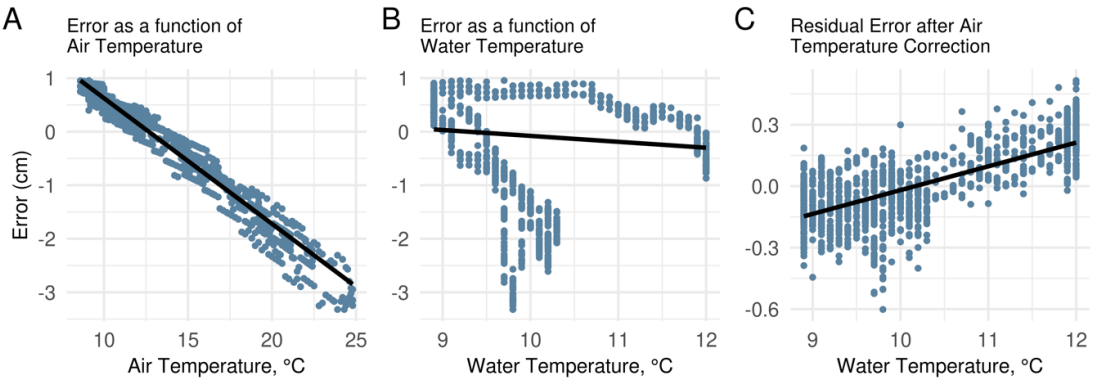


Figure 1. Error is driven by a combination of individual transducer errors, which can result in masking by the stronger trend. Panel A shows water level error as a function of air temperature, illustrating a strong linear trend. Panel B shows a dichotomous relationship between water level error and water temperature relative to the expected linear relationship. Panel C is the relationship between error and water temperature after the influence of air temperature is removed via OLS linear regression. Data taken from a barometric pressure transducer 1066019 and water pressure transducer 1062452 during the $SW_{bias\text{-}2}$ period.

Implied in the consideration of temperature bias above is that there is additional, quantifiable error that should be considered in the estimate of total measurement uncertainty. Without accounting for total uncertainty, an analysis built on water levels may lead to skewed results as the initial inaccuracies compound. For example, many studies have relied on the diurnal signal in well water levels and streamflow to estimate evapotranspiration and groundwater flow, but temperature derived artifacts could also drive diurnal fluctuation (Carlson Mazur, Wiley, & Wilcox, 2014; Kirchner, 2009; Loheide II, Butler, and Gorelick, 2005; McLaughlin and Cohen, 2014; Watras, Rubsam, & Buffam, 2017; White, 1932). Some researchers may recognize the potential for errors and propose using a rolling average or a daily mean to smooth out and cancel the errors. However, if there is a strong gradient in any source of error (e.g. Figure 1A), the result will be a systematic error that cannot be removed using a smoothing approach developed for random errors. Therefore, quantifying and correcting temperature-induced errors is essential to ensure that the observed data trends are real and not entirely or partially artifacts. While there are some attempts to provide that validation through high-frequency manual readings (Cuevas et al., 2010; Gribovszki, Kalicz, & Szilágyi, 2013), it is often impractical or impossible to capture the range of potential temperature and water level values, leaving laboratory-derived correction equations as the best alternative, as recommended by Moore et al. (2016) and Liu and Higgins (2015).

In this study we present a lab experiment to isolate and correct the separate impacts of transducer responses to water and air temperatures using absolute transducers. Our experiment is designed as a systematic comparison of errors under the four scenarios described above (GW_{best}, GW_{bias}, SW_{best}, SW_{bias}). We compare the corrected and uncorrected water levels on a test set of laboratory data as well as a case study using field data from an ongoing hydrologic project. The comparison between corrected and uncorrected laboratory and field data highlights the impacts on derived analyses and future conclusions. Finally, we calculate the propagated uncertainty for common absolute transducer deployment conditions and consider whether or not temperature correction can be used to increase measurement precision beyond the manufacturer-stated error.

3.3 Data and Methods

3.3.1 Data Collection & Preparation

We conducted a five-day experiment recording constant water levels using absolute pressure transducers in four deployments, which represent a fully crossed design with stable and variable temperatures and similar and dissimilar air and water thermal regimes (Table 1). A replication of SW_{bias} condition ($SW_{bias\text{-}1}$ and $SW_{bias\text{-}2}$) was collected as a training set for developing new measurement uncertainty bounds (see details below). Eighteen pressure transducers of varying ages and models, all manufactured by Solinst, Ltd. (Georgetown, ON, Canada), were used to record water levels (Table 2). Two transducers, both Solinst Levellogger Juniors, were used to monitor barometric pressure. All instruments contained an absolute pressure transducer, temperature sensor, and on-board data logging. These transducers also provide internal temperature compensation of absolute pressure measurements.

Table 1. Description of experimental conditions. Best practice describes air and water pressure transducers deployed under similar thermal regimes. Temperature-differential Biased indicates that barometric and water pressure transducers were not deployed under similar thermal regimes. All experimental periods were conducted over five consecutive days.

Scenario	Monitoring Analog	Deployment Description	Temperature	Air/Water Temperature Regime
GW_{best}	Groundwater	Best Practice	Stable	Similar
GW_{bias}	Groundwater	Temperature-differential Biased	Stable	Dissimilar
SW_{best}	Surface water	Best Practice	Variable	Similar
SW_{bias-1}, SW_{bias-2}[1]	Surface water	Temperature-differential Biased	Variable	Dissimilar

[1] SW_{bias} conditions were repeated and SW_{bias-2} data were used to generate an independent correction model tested using data from all other periods.

Table 2. Summary of pressure transducer model type, measurement resolution, and associated errors expressed as standard deviation, derived from manufacturer-stated 99% instrument accuracy. The lower portion of the table lists each transducer serial number under its instrument model.

Model	Number of Loggers	Pressure Resolution (cm)	Pressure Error (cm)	Temperature Resolution (°C)	Temperature Error (°C)
LT_Jr	8	$1.4x10^{-1}$	$2.15x10^{-1}$	$1.0x10^{-1}$	$4.30x10^{-2}$
LT_EDGE	2	$1.2x10^{-4}$	$4.30x10^{-2}$	$3.0x10^{-2}$	$2.15x10^{-2}$
LT_EDGE_JR	10	$1.4x10^{-1}$	$2.15x10^{-1}$	$1.0x10^{-1}$	$4.30x10^{-2}$

LT_Jr		LT_EDGE		LT_EDGE_JR	
1033239	1062534	2013939	2025928	2030899	2064739
1062452	1065861			2059683	2064745
1062520	1066016			2064734	2069158
1062528	1066019			2064737	2100561
				2064738	2104452

Water level transducers were suspended from a brace across a large container with the transducers' zero points aligned (Figure 2). The transducers were covered with a fixed depth of water and a layer of canola oil was added to prevent evaporation over the length of the experiment. Water level was measured at the start and end of the experimental period to assure no evaporation had occurred. To create the best-practice condition of thermally similar air and water environments, we placed a weighted, open-topped cylinder in the center of the container. During the two periods with synchronous (similar)

thermal regimes, the two barometric transducers were suspended in the cylinder below the level of the water surface. The entire apparatus was covered loosely with aluminum foil to prevent direct wind effects.

Best-practices and temperature-differential bias experiments were each performed twice under conditions designed to mimic common measurement targets. For stable thermal regimes, the experimental apparatus was placed in a non-climate-controlled underground room to minimize temperature variations. Varying thermal regime periods were conducted outdoors, out of direct solar radiation. For both SW_{bias} periods, the barometric transducers were placed under a black radiation shield to magnify temperature differentials and purposely induce errors from rapid temperature fluctuations such as those observed by Liu and Higgins (2015). Correction models (see below) were developed individually for each experimental period. To test correction model performance against novel data collected under different conditions, the final model coefficients developed from data collected during the $SW_{bias\text{-}2}$ experimental period were used to predict corrections for all other experimental periods. This approach allows for evaluating the common scenario of a single lab-derived correction being applied to monitoring situations that differ from lab conditions.

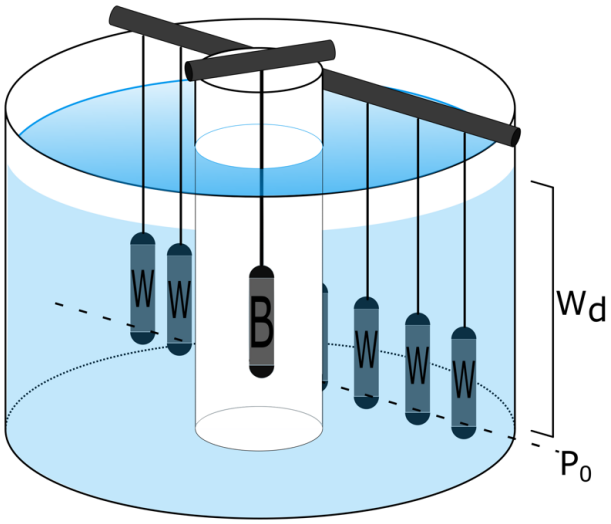


Figure 2. Diagram showing arrangement of pressure transducers within experimental apparatus (reduced number shown for clarity), where B represents a barometric transducer and W represents a water transducer. P_0 marks the water pressure transducer zero point and W_d denotes the water depth over the zero point of the transducers. During the study, two barometric transducers were deployed in the central tube in best-practice study scenarios. During all study periods, 18 water transducers were suspended from two cross braces.

Transducers were deployed and data retrieved using Solinst Levellogger v. 4.4.0 software. Transducers were set to record at 1-minute intervals with time synchronized to computer time at launch. To avoid disrupting the experimental apparatus, the five experimental periods were run serially without retrieving data from the transducers. Experimental conditions were manipulated by relocating the barometric transducers (bias vs best practices) or the entire experimental apparatus (GW vs SW).

3.3.2 Water Level Correction

All of the sensors used in this study are designed to measure pressure at a relatively constant temperature and utilize a constant water density of 1000 kg m^{-3} to convert pressure to depth (Personal Communication, Solinst). To account for the larger range of water temperatures observed in this study, the recorded pressure was adjusted to account for the variable density of water.

$$P_c = \frac{P * \rho_o}{\rho_T} \quad (1)$$

where ρ_T is the density of water (kg m^{-3}) as a function of water temperature taken from Tanaka, Davis, Oeuto, & Bignell (2001, Eq. 1), ρ_o is the assumed density of water (1000 kg m^{-3}), and P and P_c are the observed and density-corrected pressure readings in units of depth (cm for this study). Hereinafter, all references to transducer pressure refer to the density-compensated pressure.

Actual depth of water (W_d) was measured separately for each transducer from the water surface to the transducer zero point (Figure 2). For each water and air transducer pair (n = 36), the measured absolute water pressure (P_w) minus the measured absolute air pressure (P_a) determined the raw recorded water level ($WL_{raw} = P_w - P_a$). Water level error (ϵ_L) was then calculated as WL_{raw} minus W_d.

A multivariate linear model was fit between the response variable water level error (ϵ_L) and the explanatory variables air temperature (T_a, ^oC), water temperature (T_w, ^oC), and air temperature gradient (ΔT_a, $0.01^oC\ min^{-1}$) for each pair of water and air transducers using ordinary least squares (OLS). Due to the small change in air temperature at the minute timestep, the unit of ΔT_a is set as 0.01°C min^{-1} to place it on the same scale as ϵ_L, T_a, and T_w. The 36 models (one each per water and air pressure transducer pair) have the form

$$\epsilon_L = \beta_0 + \beta_a T_a + \beta_w T_w + \beta_{\Delta T_a} \Delta \quad (2)$$

where β_0 is the intercept term and β_a, β_w, and $\beta_{\Delta Ta}$ represent regression coefficients for respective variables. Figure 1C illustrates how the scale and direction of one source of error (air temperature 1A) masks the form of another source of error (water temperature, 1B), supporting the use of a single multivariate correction such as Equation 2.

To account for the uncertainty of instrument error in the temperature correction models, Equation 2 was adapted to Equation 3 below and was fit repeatedly using a two-stage parametric bootstrapping approach (Buonaccorsi, 2010). The two-stage bootstrap first

draws a sample of the observed data with replacement (a single data point can be included more than one time) to create a data set of the same size as the observations. For each sampled coordinate ($\epsilon_L, T_a, T_w, \Delta T_a$), additional random error is added to each observation, drawn from a normal distribution centered on zero with standard deviation equal to the known instrument error for each variable (Table 2). This second stage allows the observations to vary within the manufacturer-stated uncertainty. Each perturbation will affect the fit of the correction model. The dual bootstrap procedure was repeated 1000 times for each pair of transducers. This procedure generates 1000 sets of potential model coefficients, allowing for 1000 sets of correction predictions that can be used to evaluate the variability of correction models when instrument error is incorporated. The final output is visualized in Figure 3, where we show coefficient estimates of an OLS model without accounting for instrument error and the distribution of 1000 bootstrap-generated coefficients, both from a single transducer pair.

$$\epsilon'_L = \epsilon_L + N(0, \sigma_{L'}); \sigma_{L'} = \sqrt{2\sigma_L^2}$$

$$T'_a = T_a + N(0, \sigma_T)$$

$$T'_w = T_w + N(0, \sigma_T)$$

$$\Delta T'_a = \Delta T_a + N(0, \sigma_{T'}); \sigma_{T'} = \sqrt{2\sigma_T^2}$$

$$\epsilon'_L = \beta_0 + \beta_a T'_a + \beta_w T'_w + \beta_{\Delta T_a} \Delta T'_a \quad (3)$$

where T_a', T_w', $\mathbf{\Delta}T_a'$ denote bootstrap samples plus instrument error. The bootstrap procedure was carried out independently within each of the five experiments. For each bootstrap model, a backward elimination stepwise approach was used. This approach removes any slope parameters not significant at $\alpha = 0.05$ and iteratively refits until only intercept and any significant predictors remain. To determine the full range of error (sum of correction error and instrument error), final predicted error values ($\hat{\epsilon}_L$) were drawn from a normal distribution centered on the expected value of ϵ'_L (Equation 3) with a standard deviation equivalent to the residual standard error of the respective bootstrap model. This procedure resulted in 1000 estimates of ϵ'_L (Equation 3), representing instrument error, with each perturbed by the addition of random noise where the scale was determined by the standard error of the respective bootstrap model ($\sigma_{\epsilon'_L}$), representing correction error. The entire error estimate can be presented as $\hat{\epsilon}_L$ (Equation 4).

$$\hat{\epsilon}_{L,i} \sim N\left(\epsilon'_L, \sigma_{\epsilon'_L}\right) \quad (4)$$

where $\hat{\epsilon}_{L,i}$ is the predicted error (or inversely the predicted correction required) for the *i*th record. For each sample record, we used the median of the 1,000 $\hat{\epsilon}_{L_i}$values as the final predicted error within an experiment and took the 0.025 and 0.975 quantiles of the same 1,000 $\hat{\epsilon}_{L_i}$ values as the upper and lower prediction intervals (Figure 4). For analysis and

plotting, constant transducer offsets were removed from both WL_{raw} and WL_{corr} by subtracting the mean of ϵ_L and $\hat{\epsilon}_L$ from WL_{raw} and WL_{corr}, respectively, centering the estimates on the true water depth.

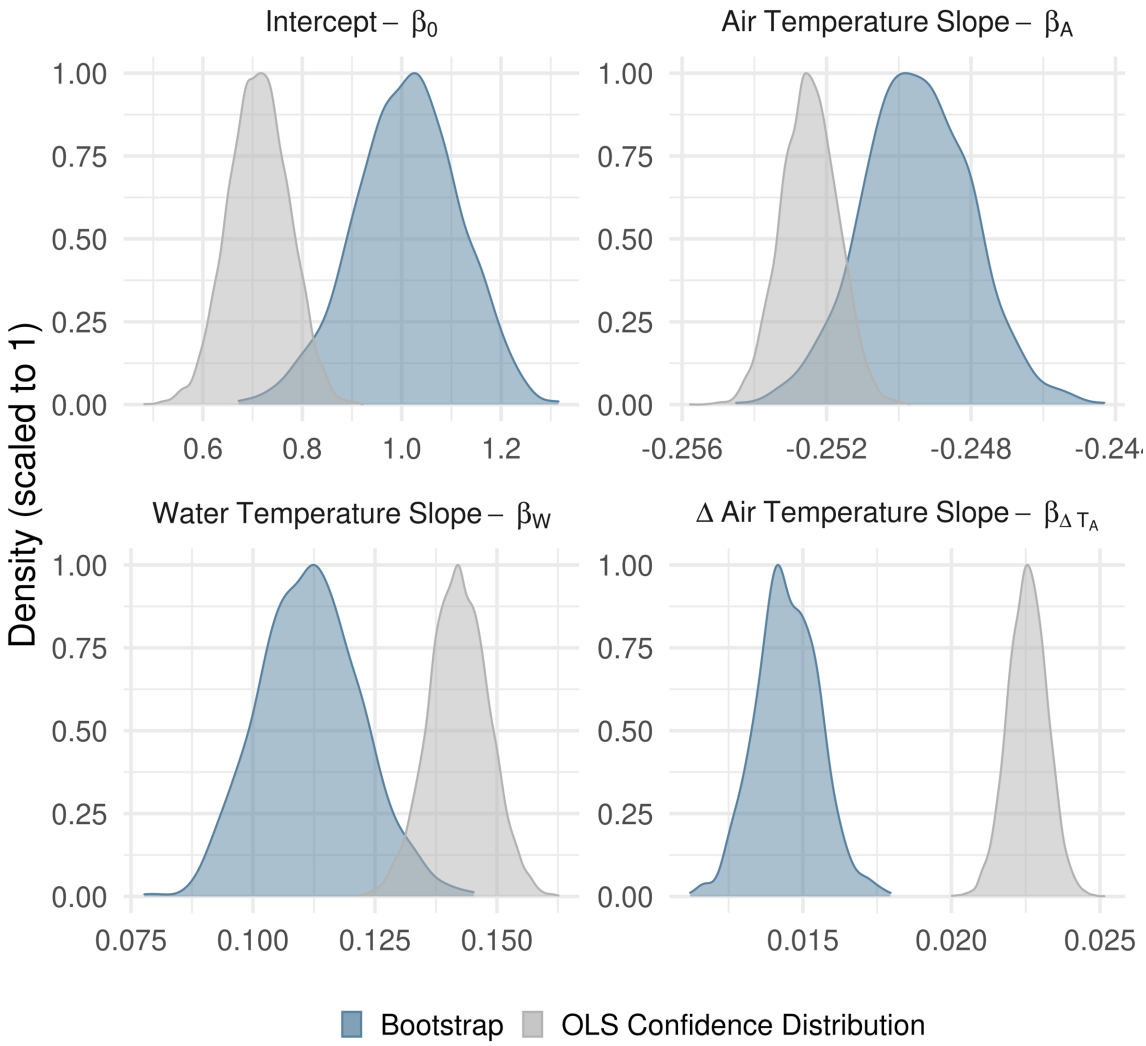


Figure 3. A comparison of correction model coefficients for a single transducer pair (chosen at random, Barometric SN: 1066019, Water SN: 2069158) as derived from ordinary least squares (OLS) regression and dual bootstrap analysis. The bootstrap density curve represents the distribution of all 1000 bootstrap models. The OLS density curve represents the confidence distribution of the fitted coefficients without accounting for instrument errors (~N(μ, σ) where μ = best estimate of coefficient and σ = standard error of coefficient estimate).

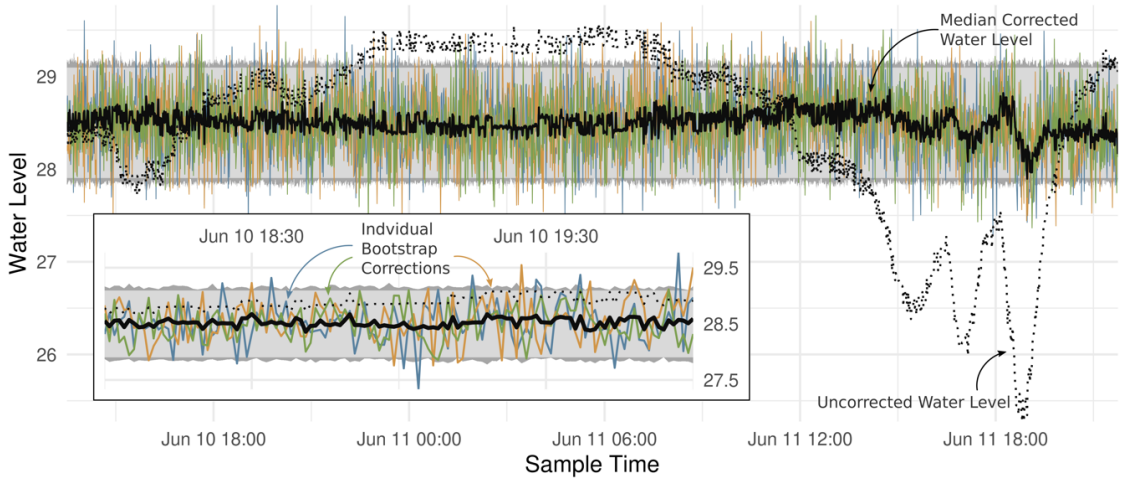


Figure 4. A visualization of the development of a single water-barometric transducer correction model to correct water level. The known water depth for this example is ~28.5 cm. The dotted line represents the raw water level from a pressure transducer pair, which can be seen to vary greatly from a constant water level. Each of the three colored lines represent a single instance of bootstrap model (a random selection of 3 out of 1000 correction models fit) and the solid black line is the median response from all 1000 models. Gray bands show the 95% confidence bands derived from the instrument error as stated by the manufacturer (light gray) and propagated (manufacturer-stated instrument error plus additional error from model uncertainty) error (dark gray). The expanded section shows the same series for a two-hour period for clarity.

3.3.3 Uncertainty Analysis

The impact of temperature-derived errors on water level measurements was evaluated by comparing instrument error, propagated error, and measured error of corrected data. Within each experimental period, the magnitudes of instrument error and propagated error from the bootstrap prediction bounds were compared using a *t*-test. Differences between experimental periods were compared using a one-way ANOVA and with Tukey Honest Significant Differences.

We believe correction to be a necessary step to reduce bias in water level measurements, but additional uncertainty introduced from correction equations could lead to higher variance around the final measurement value. To evaluate the bias-variance trade-off, WL_{corr} was calculated for the $SW_{bias\text{-}1}$ using the correction equations from $SW_{bias\text{-}2}$. New accuracy bounds for the period were computed as $\mu_{\hat{\epsilon}_L} \pm 1.96 * \sigma_{\hat{\epsilon}_L}$, where $\mu_{\hat{\epsilon}_L}$ is the mean of the predicted errors and $\sigma_{\hat{\epsilon}_L}$ is the standard deviation of predicted errors as in Equation 4. The new accuracy bounds were then tested against manufacturer-stated instrument error bounds using a paired *t*-test. As $\hat{\epsilon}_L$ varies for each record in a time series based on the scale of correction, there could be a very high *n* for comparisons of derived and

manufacturer-stated error bounds. To avoid artificially inflating significance through high n, all t-tests were performed between the fixed instrument error band and the mean of the propagated error band to hold $n = 36$ within each experiment.

3.3.4 Correction Evaluation

To evaluate correction across conditions, the final models generated for the $SW_{bias\text{-}2}$ period experiment were used to predict $\hat{\epsilon}_L$ for all other experimental periods. This period represents the least controlled experimental conditions, which may represent the only available compensation equation development conditions. For each water-air transducer pair, the $SW_{bias\text{-}2}$–derived correction model was used to predict the error ($\hat{\epsilon}_L$), which was in turn used to correct the recorded water levels such that $WL_{corr} = WL_{raw} - \hat{\epsilon}_L$. Correction effectiveness was evaluated based on two criteria: number of recorded points that fell outside of the range (out of range, OOR) of combined instrument errors ($n_{\epsilon,i}$) and the root mean squared error (RMSE) from the known water depth. Effect of correction was tested as the pairwise difference in estimated marginal means of the RMSE for corrected and uncorrected data within a given experimental period. Additionally, estimated marginal means of the corrected data RMSE were compared across experimental periods.

3.3.5 Case Study

We applied the proposed correction protocol to observed wetland water level data to highlight propagation of uncorrected errors into subsequent analyses. The data for this case study comes from a forested headwater depressional wetland (0.81 ha) in the western Upper Peninsula of Michigan, USA. The data were collected during the ice-free season of 2018. Transducer 1066016 was deployed in a 5.08-cm internal diameter (I.D.) monitoring well and transducer 1065861 was deployed in a sheltered open-air area with both transducers recording synchronous 15-minute data (Van Grinsven et al., 2017). In the described deployment scenarios from above, this would correspond to temperature-differential biased groundwater monitoring (GW_{bias}). We will illustrate the impact of temperature bias on two common analyses: mean daily water level and evapotranspiration (ET) derived from diurnal water level variation. For the former analysis, we calculated mean uncorrected and corrected water levels on a daily timestep. For the latter analysis, we followed methods built on White (1932).

White (1932) proposed an approach for decomposing daily groundwater level records into evapotranspiration (ET) and groundwater recharge. The approach involves estimating net inflow rate (G_{in}) during a recharge period, typically the pre-dawn hours. This rate and the change in water level between recharge periods on subsequent days can be used to determine ET. Converting water level fluctuation to surface ET also requires accounting for the specific yield of the substrate surrounding the well. Loheide II (2008) and Loheide II *et al.* (2005) expanded this work to improve estimates of specific yield and account for variation in groundwater flow with water level change. The White and Loheide II works focused on groundwater-phreatophyte interactions, but the method has been extended to use in wetland settings (Diamond, McLaughlin, Slesak, & D'Amato,

2018; McLaughlin and Cohen, 2014; Telander et al., 2015). In these settings, Loheide's readily available specific yield was expanded to ecosystem specific yield (E_{Sy}) to acknowledge the influence of factors beyond the substrate surrounding the well (McLaughlin and Cohen, 2014). We applied the adjusted White methodology to the 15-minute wetland well data to calculate daily ET for the raw and corrected water level records (ET_{raw} and ET_{corr}, respectively).

Our estimates of E_{Sy} were derived as the ratio between the change in water level and the change in water availability (rainfall plus snowmelt less PET) during the drawdown period (start of water level records to minimum observed water level). This is an inverse analog of the rain-to-rise ratio used in other studies to estimate E_{Sy} (McLaughlin and Cohen, 2014). We used this analogous approach because our site is influenced by two factors known to obfuscate a clear rain-to-rise E_{Sy} relationship: surface water exchange (Watras et al., 2017) and additional hydrologic cycles at longer wavelengths (e.g. snowmelt and seasonal inundation patterns) (Zhu, Young, Healey, & Jasoni, 2011). E_{Sy} estimates were then fit to an asymptotic function with water level as the independent variable using iterated re-weighted least squares non-linear regression (Maechler et al., 2020). These models allow for continuous estimation of E_{Sy}. Curves for E_{Sy} were derived separately using the raw and corrected data. The respective E_{Sy} functions were then used to separate the raw and corrected diurnal fluctuations into G_{in} and ET. For validation of the correction procedure, we compared the ET_{raw} and ET_{corr} values to potential evapotranspiration (PET) calculated using the modified Hargreaves-Samani equation (Hargreaves and Allen, 2003) with data from nearby meteorological stations. For analysis, only dry days were used to avoid the complicating factors of interception, suppressed ET, and variable water level increase. We defined dry days as days with daily rainfall or snowmelt of less than 1 mm in the preceding two-day period.

3.3.6 Software

All data processing and analysis were performed in R (R Core Team 2019) using the packages *data.table* (Dowle and Srinivasan, 2019) for data manipulation, and *ggplot2* (Wickham, 2016) and *patchwork* (Pedersen, 2019) for visualization. The *drake* (Landau, 2018), *rmarkdown* (Allaire et al., 2020; Xie, Allaire, & Grolemund, 2018), *rticles* (Allaire, Xie, Foundation, & Wickham, 2019), and *rbbt* (Dunnington and Wiernik, 2020) packages were used to generate a reproducible project and publication.

3.4 Results

3.4.1 Uncertainty Analysis

If the instrument error was assumed as the only source of error, the 95% confidence band (CI) is calculated by combining the instrument errors stated in Table 2, where $CI = 2 * Z_{0.95} * \sqrt{(\sigma_1^2 + \sigma_2^2)}$. The resulting error was 0.859 cm for the LT_EDGE water pressure transducers and 1.91 cm for the LT_Jr and LT_EDGE_JR water pressure transducers. When measurement errors and temperature effects were propagated to represent full uncertainty, every experimental period showed a significant increase in uncertainty bounds. The magnitude of the increase varied among experimental treatments from 0.042

Table 3. Mean Propagated and instrument errors for all 36 transducer-pair models by experimental period. Difference and standard error show the change in confidence band size when all errors are incorporated and tested for significance using a paired *t*-test within each experimental period (denoted by '*', $\alpha = 0.05$). Difference in propagated errors between experimental periods was tested using a one-way ANOVA.

Experimental Period	**Mean Propagated Error**	**Mean Instrument Error**	**Difference +/- SE**	**Significance of Propagated Error Between Groups** †
GW_{best}	1.20	1.15	0.042 ± 0.002*	a
GW_{bias}	1.20	1.15	0.047 ± 0.001*	a
SW_{best}	1.23	1.15	0.071 ± 0.008*	a
$SW_{bias\text{-}1}$	1.39	1.15	0.23 ± 0.02*	b
$SW_{bias\text{-}2}$	1.63	1.15	0.48 ± 0.03*	c

* Indicates the propagated and instrument errors were significantly different at $\alpha = 0.05$.

† Letters indicate significance grouping where periods with the same letter are not significantly different

cm for GW_{best} to 0.48 cm for $SW_{bias\text{-}2}$ (Table 3). Significant differences were observed in propagated errors between experimental periods. Experimental period $SW_{bias\text{-}2}$ had significantly larger uncertainty than $SW_{bias\text{-}1}$, which was significantly larger than the remaining three periods. This variability corresponds to higher variability in air temperature and change in air temperature during this experimental period (Figure S2).

Due to large discrepancies in correction model accuracy between the barometric transducers, we compared the propagated errors and the instrument errors separately for each set of 18 models (Table S2). For transducer 1066019, the propagated errors were found to be significantly less than the instrument error (mean decrease in 95% confidence bounds = 0.67 cm). Transducer 1065861 was found to have propagated errors that were significantly higher than instrument error (mean increase in 95% confidence bounds = 0.21 cm).

3.4.2 Correction Effectiveness

The model form described in Equation 2 showed a good fit to the $SW_{bias\text{-}2}$ data with bootstrapped estimates of adjusted R^2 ranging from 0.89 to 0.98. The bootstrap predictions of final WL_{corr} had RMSE values ranging from 0.32 to 0.51 cm, which includes both instrument error and correction model uncertainty (Table S1). For reference, observed (WL_{raw}) RMSE for that period ranged from 1.53 to 2.60 cm relative to known water depth. When the $SW_{bias\text{-}2}$ model was used to predict other experimental periods, the RMSEs of raw and corrected water level errors relative to true water level showed differences of -0.61–0.04, where a negative value indicates a smaller RMSE following correction (Table 4). The reduction in RMSE was greatest for the for $SW_{bias\text{-}1}$ period and along with the GW_{bias} period was significant at the $\alpha = 0.05$ level. The SW_{best}

period showed a significant increase in RMSE following correction. We also tested model performance after excluding two transducers (water transducers 2025928 and 2030899) from the analysis based on their status as outliers (Figure 5). With the two outlier transducers removed, the increase in RMSE in the SW_{best} period became a decrease of 0.02 and was no longer significant. When the RMSE of WL_{corr} was compared to instrument error bounds using a *t*-test with $\alpha = 0.05$ and $H_A: \hat{\epsilon}_L - \epsilon_{instrument} < 0$, we found RMSE of WL_{corr} was significantly less than instrument error for all experimental periods.

Prior to correction, 0.00 to 56.92% of measured points fell outside of the 95% confidence band defined by the combined stated instrument errors of the barometric and water pressure transducers, with all periods except $SW_{bias\text{-}1}$ showing less than 1% of points outside of the error range (Table 4). Following the correction, only 6.17% of $SW_{bias\text{-}1}$ points fell outside of that same confidence band. A small increase of points outside of the band was observed for SW_{best} (2.48%) and GW_{best} (< 0.005%) following correction. When the interval is expanded to include instrument uncertainty and the errors associated with temperature correction, only $SW_{bias\text{-}1}$ showed any raw data points (35.40%) outside of the expanded error band. Following correction this fell to 0.36% of points for $SW_{bias\text{-}1}$ but increased from 0.00 to 0.32% for SW_{best}. All corrected points that fell outside of the instrument and propagated error bands were associated with the two abnormal transducers mentioned above.

Table 4. Root–mean–squared error of predicted water level and percentage of observations of raw and corrected water levels falling outside of instrument accuracy range and propagated error range. RMSE and propagated error were determined using the final model developed using the $SW_{bias\text{-}2}$ dataset to test a single correction equation across monitoring scenarios. Mean values of the 36 transducer pairs reported.

		Percentage of Observed or Corrected Measurements Outside 95% Error Range			
		Instrument Error		**Propagated Error**	
Experiment	**RMSE of WL_{corr}†**	**WL_{raw}**	**WL_{corr}**	**WL_{raw}**	**WL_{corr}**
GW_{best}	0.08 (-0.00)	0.00	0.00	0.00	0.00
GW_{bias}	0.09 (-0.06)	0.36	0.00	0.00	0.00
SW_{best}	0.18 (0.04)‡	0.13	2.48	0.00	0.32‡
$SW_{bias\text{-}1}$	0.24 (-0.61)	56.92	6.17	35.40	0.36

* Indicates that the change in RMSE was significantly different at $\alpha = 0.05$.
† Values in parentheses show the difference in RMSE (Δ RMSE) between observed (WL_{raw}) water level error and corrected water level error (WL_{corr}) using the $SW_{bias\text{-}2}$ models, where a negative value indicates a reduction in RMSE.
‡ Transducers 2025928 and 2030899 showed non-standard behaviour (discussed in text). When they are removed from the analysis Δ RMSE = 0.02 (not significant at $\alpha = 0.05$) and no corrected values fell outside of the propagated error range.

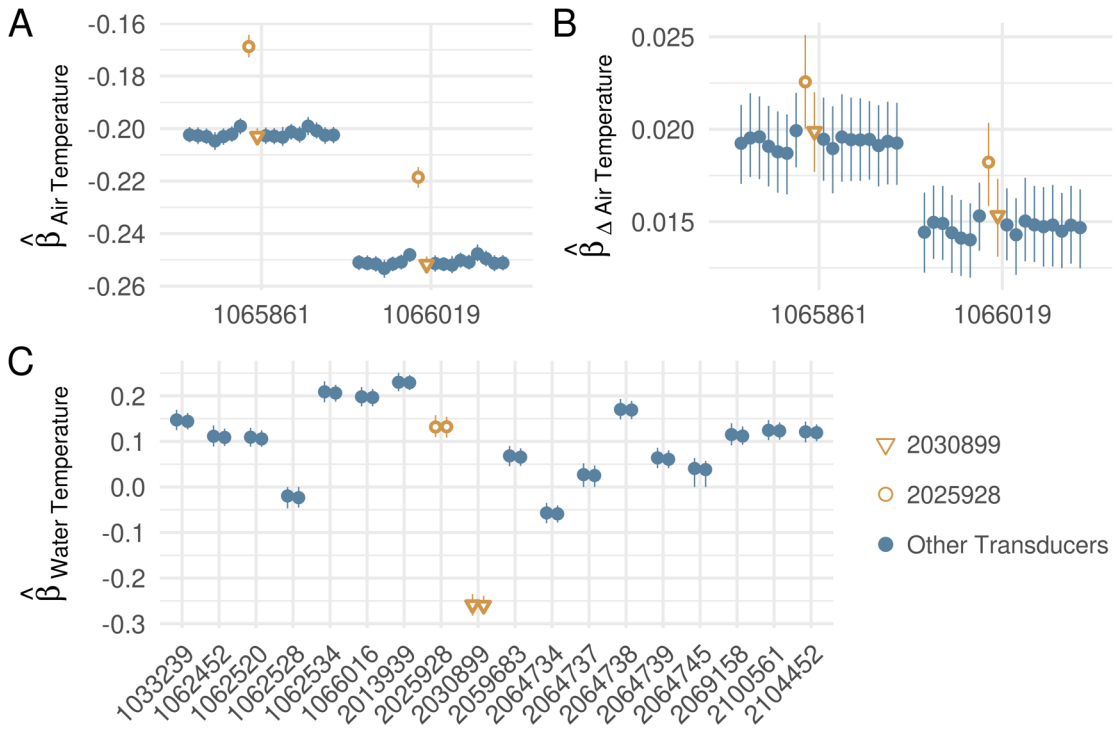


Figure 5. Mean and 95% confidence interval for correction model coefficients for the 1000 bootstrap models for the $SW_{bias\text{-}2}$ period. Panels A and C show coefficients grouping by barometric transducer and panel B shows models grouping in pairs for each water-barometric transducer pair. Water transducers 2025928 (hollow circle) and 2030899 (hollow inverted triangle) are highlighted to illustrate disagreement with other models for one or more coefficients.

3.4.3 Case Study

For much of the case study period, the overall trend in water levels tracks well between the raw and corrected levels (Figure 6A). Differences become apparent when we look at the daily mean water level (Figure 6B) and the derived ET estimates (Figure 6C). When mean WL_{raw} and WL_{corr} are compared, 92 days (55.8%) were found to have temperature corrections that exceeded combined instrument error (Figure 6B). Daily mean errors were greatest in the fall, and smallest during the summer months, ranging from a 1.5 cm underestimate to a 4.2 cm overestimate of mean daily water level.

Errors at the sub-daily range were also found when comparing WL_{raw} and WL_{corr}. The two days shown in Figure 6C have daily mean water level errors within the expected instrument error but the estimates of G_{in} and ET derived from these diurnal fluctuations are qualitatively different upon visual inspection. The uncorrected data shows higher inflow and corresponding higher ET even when the days show similar raw and corrected water level change. During the drawdown period used to calculate E_{Sy}, the temperature-

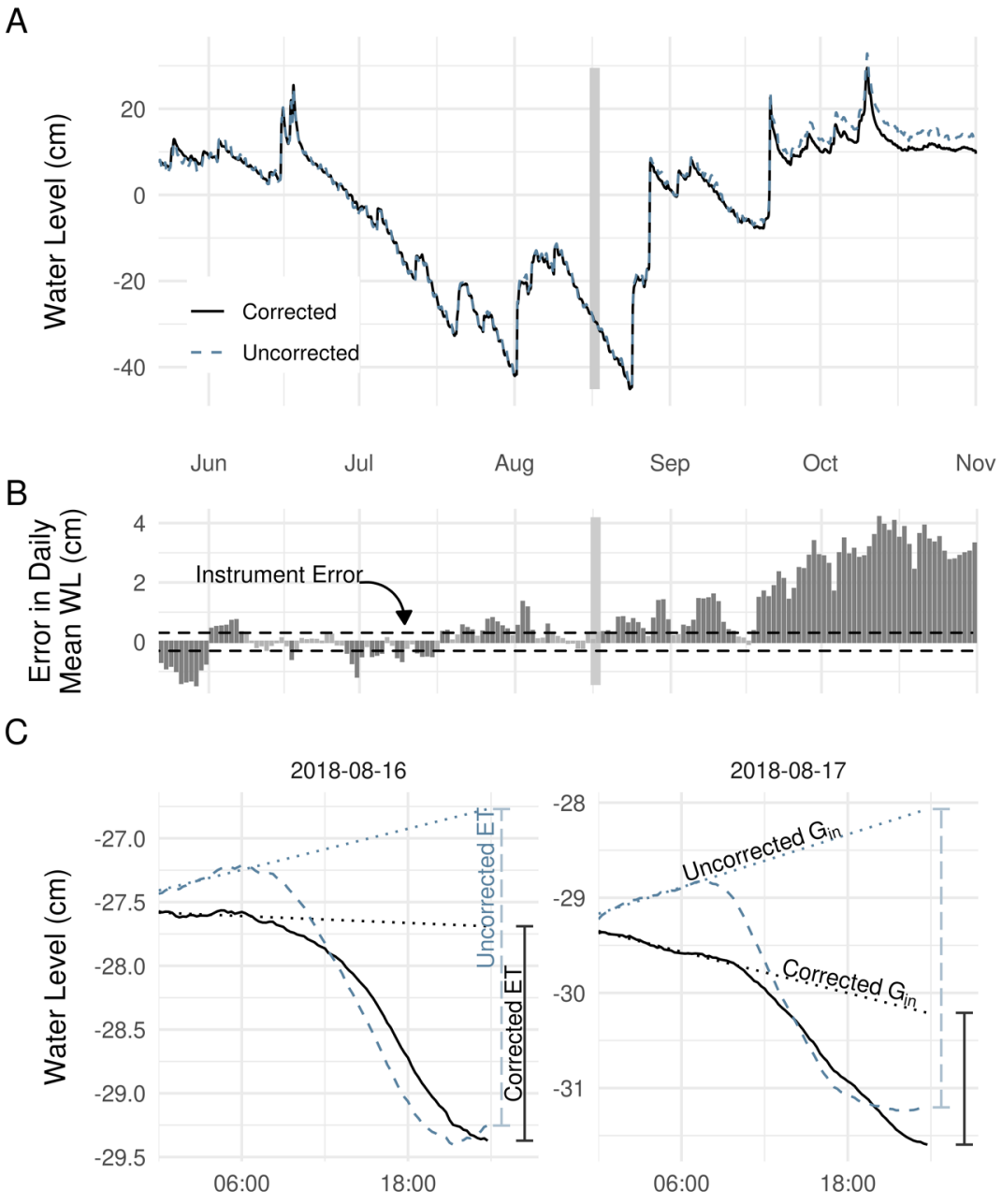


Figure 6. A comparison of corrected and uncorrected wetland water levels. Panel A shows the entire growing season at 15-minute interval. Panel B shows the error in daily mean water level. Panel B dashed lines represent manufacturer-stated instrument error and darker columns indicate errors greater than combined instrument error. The vertical shaded bar in Panels A and B highlight the two days shown in Panel C. Panel C shows details of two diurnal water level cycles and the decomposed components of the White (1932) method: inflow (G_{in}) and evapotranspiration (ET). In panel C the solid line represents corrected water level, the dashed blue line represents uncorrected water levels and the dotted lines illustrate the basic White methodology of estimating G_{in}.

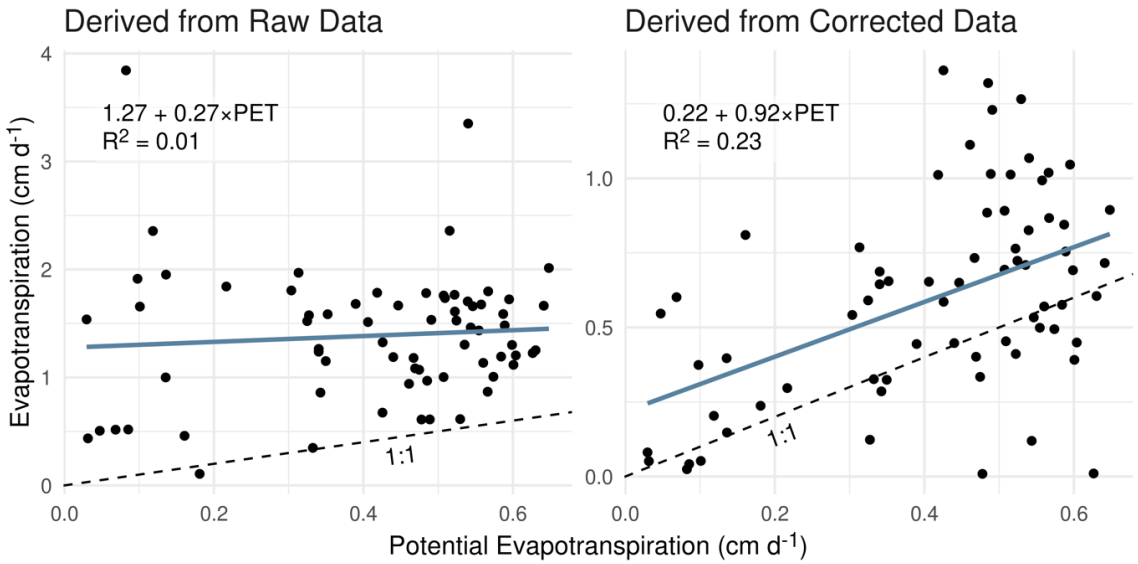


Figure 7. Linear regression fit between potential evapotranspiration and White (1932) method estimates of evapotranspiration. The estimates from the raw data showed a significant intercept but not a significant slope. The estimates from the corrected data showed a significant intercept and slope. The dashed line shows a 1:1 relationship to represent perfect estimation of *PET*.

induced error was generally small (Figure 6B) resulting in similar E_{Sy} functions (Figure S1). While there is no validation data to compare estimated inflow to actual inflow, we can compare estimated *ET* to *PET*. ET_{raw} showed negligible correlation with *PET* (R^2 = 0.078), while ET_{corr} showed moderate correlation with *PET* (R^2 = 0.48). Fitting an OLS linear regression between estimated *ET* and *PET* showed no significant slope between ET_{raw} and *PET*, but a significant slope, close to 1 for ET_{corr} and *PET* (Figure 7).

3.5 Discussion

3.5.1 Temperature-Induced Bias and Correction Equations

Our results show that it is possible to correct water-depth derived from a pair of water and barometric absolute pressure transducers without the need to reference to an external data source. Any transducer correction approach must be a trade-off between convenience, absolute accuracy, and uncertainty. The described approach to correction achieves accuracy that corrects up to 89% of erroneous data to within the bounds of combined instrument error and up to 99% to within the fully propagated error bounds (Table 4, percentage of raw out-of-range points within range following correction). Because it does not rely on external pressure readings, it can be performed at a monitoring location with little to no extra equipment. By removing any dependency on

external instrumentation, we also avoid a tertiary source of error, minimizing the total uncertainty of the corrected water level measurements.

Importantly, the presented approach captures the individual transducer corrections. One possibility of the dual transducer model is that the model would be fit to some artifact of the interaction between the transducers, rather than each transducer being represented in the air and water temperature coefficients. Figure 5 shows that the air temperature coefficients were grouped by barometric transducer (two distinct regions of overlapping ranges), and the water temperature coefficients were grouped by water transducer (18 distinct regions of overlap of the two models representing each water transducer). Clear transducer differentiation also supports the assertion above that water level errors using absolute pressure transducers are introduced through independent water and air temperature errors and not temperature differences. Correcting for water level error using temperature difference would disregard the differential transducer responses. Based on the scale of the coefficients for air temperature and water temperature, these errors would likely be small in this study. The error would be directly proportional to the sensitivity of each transducer to temperature and would likely vary between manufacturers and instruments.

Both Cain et al. (2004) and Liu and Higgins (2015) recommend lab-based correction equations be developed for all deployed transducers. Building upon that, we have demonstrated that it is possible to develop a single lab-based correction with a structure suitable for varied deployment environments. Our correction model developed on $SW_{bias\text{-}2}$ performed well correcting other experimental periods. We were also able to reduce the total uncertainty for an independent dataset below instrument error using our $SW_{bias\text{-}2}$ correction equations. However, we recommend developing a correction equation under conditions similar to field conditions whenever possible to ensure accurate correction. The degraded predictive performance of $SW_{bias\text{-}2}$ models under SW_{best} conditions relative to all other periods highlights this point (Table 4). Under SW_{best} conditions we observed a small but significant increase in RMSE for corrected water levels. We believe that the inaccuracies have two potential causes at their root. The first is that the SW_{best} conditions had higher temperatures than any of the other experimental periods. Models developed on $SW_{bias\text{-}2}$ data were therefore extrapolating when predicting under SW_{best} conditions. The second possibility is the internal temperature compensation of the pressure transducers. Freeman et al. (2004) anticipated that isolating and correcting temperature-driven errors would be more difficult in transducers with internal compensation. The data used to train our correction models were purposefully collected under the extreme conditions ($SW_{bias\text{-}2}$), with large, rapid temperature fluctuations in both the air and water temperature readings. Under these conditions the internal temperature compensation algorithm is likely overwhelmed. Under conditions with slower temperature fluctuations (SW_{best}), the correction models may overcorrect the data because the internal temperature compensation has already dampened the temperature-induced errors. Identifying extrapolation and internal-compensation driven errors illustrates the importance of using an independent dataset to validate correction models.

Model performance in this study was highly dependent on individual transducers. Based on the results of the uncertainty analysis, barometric transducer 1065861 appears to be more sensitive to air temperature and high rates of temperature change than transducer 1066019 as evidenced by the difference in results of propagated error bounds (Table S2). This led to larger errors at the beginning and end of the $SW_{bias\text{-}1}$ period when rates of temperature change were greatest (Figure S2). Transducer 2030899 shows larger errors and along with 2025928 accounts for all of the increase in OOR measurements in the SW_{best} period. Further investigation into this logger suggests that during the SW_{best} period the direction of the relationship between water temperature and error reversed relative to all other experimental periods (Figure S3). Transducer 2064734 showed the same effect but to a lesser extent. At this time, it is impossible to tell if this is due to the high temperatures during this period, differential internal compensation among transducers, or some other error such as a trapped air bubble against the transducer membrane. The experimental and modeling approach laid out above can, and should, be used to adjust the correction equations for individual transducers such as 2030899 and 2064734. That scale of tuning is outside of the scope of this study, which is instead focused on commonalities in correction that can be applied across instruments and manufacturers and the role of temperature-derived errors in total uncertainty. The above discussion highlights an advantage of simultaneous correction of multiple instruments. By simultaneously including multiple pairs of barometric and water pressure transducers, we are able to identify and isolate additional errors specific to the barometric or water transducer.

3.5.2 The Role of Uncertainty in Water Level Measurements

We found that under certain conditions water level errors exceeded instrument accuracy bounds (Figures 4 & 6, Table 4). As expected from previous work (McLaughlin and Cohen, 2011), experimental periods showed the largest errors when air and water transducers were under separate thermal regimes. Our results indicate that when deploying transducers to variable temperature environments, similar air and water thermal regimes (best practices) are enough to prevent significant increases in uncertainty relative to stable temperature environments. Though total error is likely to be greater than stated instrument error, monitoring of small streams, surface waters, and other water sources with diurnal temperature signals is not inherently more uncertain than groundwater or other stable-temperature environments when best practices are observed.

The observed errors, both within and outside of instrument accuracy bounds, were correlated with environmental conditions. This correlation is especially vexing because the points that are most likely to lie outside of the error bounds occur simultaneously with periods of hydrologic interest. As the errors are associated with higher temperatures and high rates of temperature change, the errors will often co-occur with changes in flow following snowmelt or rain events or periods of peak evapotranspiration. To confidently attribute diurnal or seasonal signals to true variation rather than artificial errors requires additional validation. One source of validation is using additional instrumentation that is not susceptible to the same error source, such as frequent manual measurements of stage/level (e.g., Cuevas et al., 2010; Gribovszki et al., 2013). Where this is not practical or possible, we have shown that correction equations can quantify the additional

uncertainty of environmentally-induced errors and rule them out as the cause of the signal of interest.

Even when transducers are deployed under 'best practices', temperature-induced errors will often occur. The magnitude of the error may be similar to instrument error and thus may not be apparent. Our results show that even if not apparent, this error should be considered because it represents systematic bias rather than a random error. Of the 1000 models fit per experimental period, 55.1% of GW_{best}, 2.2% of GW_{bias}, and 28.8% of SW_{best} final models were intercept-only. This trend shows that under increasingly ideal conditions more models showed no significant relationship with air temperature, water temperature, or rate of change of air temperature. However, even under the most stable monitoring scenarios (GW_{best}) almost half of all fit bootstrap models showed significant correlation between error and environmental variables. The error associated with these intercept-only models still provides important information regarding true water level accuracy and they are included in the propagated error estimates. The intercept-only models are likely not devoid of temperature derived errors, and SW_{best} and GW_{bias} show slightly, but significantly higher σ_{mod} than other models ($\sigma_{intercept\ models} - \sigma_{slope\ models}$: GW_{best}: -0.005 (p = 0.42); GW_{bias}: -0.012 (p = 0.02); SW_{best}: -0.010 (p = 0.010)). This small increase in σ_{mod} is likely the result of unmodeled temperature-driven errors. As temperature artifacts were present under all conditions, it is not unexpected to find that true uncertainty was greater than instrument error under all conditions. Taken together it is likely that even under ideal conditions thermal artifacts will affect measurement accuracy in a systematic way and correction is recommended.

Increases in measurement error can have important implications for research quantifying small changes or weak diurnal (or other cyclic) signals. Observations that fall outside of, but close to the stated instrument accuracy range, may not be outside of the range of full uncertainty. Using a validation dataset collected under the same conditions, we demonstrated above that we could quantify the error of newly corrected observations and increase final measurement precision. In our study our best performing correction models showed a 65% reduction in the uncertainty range, from 1.19 cm to 0.42 cm (Table S2). A reduction of this magnitude is not insignificant in the scale of diurnal signals often studied.

3.5.3 Case Study

For both analyses performed in the case study we found significant bias introduced due to temperature-induced errors in WL_{raw}. The purpose of water level records in monitoring and research is to enable further analysis. Unaccounted errors in instrument records have measurable impacts, which could lead to erroneous conclusions and potentially to real-world management decisions that are at best misguided and at worst harmful. The presence of detectable error is dependent on the scale of the analysis performed. In Figure 6, the two highlighted days in panel C show approximately a 1.7-fold difference in ET estimates, representing estimation errors of 170% and 166% of ET_{corr}. In Panel B we see that the daily mean error on these days was within the range of expected instrument error. Quantifiable errors may be more likely to be detected at finer scales (sub-daily) than at

coarser scales (daily) as aggregation smooths out evenly distributed errors. But during periods where the errors are systematically positive or negative no level of smoothing will bring temperature-induced errors within the range of instrument error. The errors observed in the fall are one such period for this case study (Figure 6A).

We observed the largest errors in the fall after wetland water levels had rebounded to the surface. This is due to a combination of seasonal air and water temperature trends. This period had the lowest air temperatures, and the highest water temperatures (not shown). Figure 1A & C show that these conditions led to the largest errors in our laboratory experiments. As the air temperatures decline in the fall and wetland water temperatures increase due to groundwater inputs, the temperature bias increases. Errors in the spring would likely be large, but somewhat smaller than fall errors with low air temperatures and low water temperatures due to snowmelt. Our observed daily mean error was of similar magnitude to the 1.5 cm increase in diurnal variation observed by McLaughlin and Cohen (2011). Our estimates of ET error were much larger than the 19% increase in streamflow diurnal fluctuation presented in Cuevas *et al.* (2010), though *ET* is often a smaller component of the water budget in such a setting so that larger percentage errors can be expected.

There were two potential sources of error when calculating *ET* using the modified White method: the magnitude and shape of the diurnal fluctuations and the determination of E_{Sy}. We have concluded that the poor correlation between ET_{raw} and *PET* is the result of the diurnal water level fluctuations. We did not see evidence that the E_{Sy} calculations played a role as the final E_{Sy} functions were similar for the raw and corrected datasets (Supplemental Figure S1). It is important to note that while ET_{corr} showed a moderate correlation with *PET* and a significant slope near 1, the R^2 of the OLS model was low. This is in line with other White-method separation studies where low R^2 values are observed. Soylu, Lenters, Istanbulluoglu, & Loheide (2012) found R^2 values of less than 0.2 for the White method and 0.33–0.40 for their proposed sine-wave based method. McLaughlin and Cohen (2014) did not report R^2, but their *ET*/*PET* index showed high variability, suggesting low R^2. Even with direct eddy flux measurement of *ET*, Lafleur, Hember, Admiral, & Roulet (2005) found an R^2 as low as 0.56 between *ET* and *PET* in a Canadian shrub wetland. Based on other studies, our wetland could be expected to have poor performance for the White Method. Nearby surface water dynamics and connectivity to other surface water sources have been shown to have high levels of interference with determining E_{Sy} or White-method *ET* (Watras et al., 2017; Zhu et al., 2011).

3.5.4 Uncertainty and Transducer Type

This study focused on absolute pressure transducers, but some of the findings presented here could inform similar work on gauge pressure transducers. Cain et al. (2004) found that temperature changes within the vent-tube of gauge pressure transducers introduce water level artifacts, which was explored further by Liu and Higgins (2015) in multi-day laboratory and long-term field experiments. Their findings showed that errors can be introduced due to water temperature or rapid fluctuations in air temperature. Their

laboratory experiment found 1.4 mm $°C^{-1}$ error as water temperature changed (with constant air temperature). These corrections corresponded to errors of up to 0.7 cm in their field experiments, with at least some of that error due to the rate of temperature change affecting the connection to the atmosphere via the vent tube. They also highlight the impact of rate of temperature change, with more rapid gradients inducing larger water level errors, which agrees with the results from our study. Agreement between the two logger types is expected as both contain internal electronics and membranes that must equilibrate to the new temperatures (Freeman et al., 2004). The methodology for developing correction equations presented here could be easily adapted to gauge pressure transducers.

3.6 Conclusions

We presented a laboratory experiment demonstrating that thermal bias can increase measurement uncertainty beyond instrument error. In the case of absolute pressure transducers, the added uncertainty is not the result of temperature difference between transducers, but is a combination of individual transducer biases. These biases can impact uncertainty across commonly used deployment scenarios for hydrologic monitoring. Even under best deployment practices, individual logger response to similar thermal regimes will result in increased uncertainty, and can result in measurable water level errors. Transducer and transducer-pair specific correction equations can be used to reduce errors and fully quantify measurement uncertainty. In addition, laboratory-derived correction can be used to reduce final uncertainty below stated instrument error, providing a mechanism for higher resolution analyses. We provide the following recommendations for addressing and correcting temperature-derived errors.

- For each transducer, check for *all* potential error-inducing responses to temperature and rates of change in temperature.
- Even under best practice deployments thermal error should be considered. This is especially true if monitored water sources vary in temperature, e.g. surface water and shallow groundwater.
- Correction equations should be developed under conditions that capture the full range of temperatures likely to be monitored.
- To maximize correction benefits, collect separate validation datasets not used to derive equations. These datasets can be used to create new and narrower uncertainty estimates.
- When possible, correct multiple transducers simultaneously to better isolate problematic instruments via cross-reference. If that is not possible, compare to local weather station(s) for validation of your observations.

This study focused on trend correction rather than offset. Often systematic offsets are corrected through manual calibration measurements during deployment. If deployed conditions will not allow for manual calibration measurements, the above approaches are applicable, but offsets must be carefully calculated.

3.7 References

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2020. *Rmarkdown: Dynamic Documents for R*. Manual.

Allaire, JJ, Yihui Xie, R Foundation, Hadley Wickham, Journal of Statistical Software, Ramnath Vaidyanathan, Association for Computing Machinery, et al. 2019. *Rticles: Article Formats for R Markdown*. Manual.

Buonaccorsi, John P. 2010. *Measurement Error: Models, Methods, and Applications*. Interdisciplinary Statistics Series. Boca Raton, Florida: Chapman and Hall/CRC.

Cain, Samual F., III, Gregory A. Davis, Steven P. Loheide II, and James J Butler Jr. 2004. "Noise in Pressure Transducer Readings Produced by Variations in Solar Radiation." *Ground Water* 42 (6): 939–44. https://doi.org/10.1111/j.1745-6584.2004.t01-12-.x.

Carlson Mazur, Martha L., Michael J. Wiley, and Douglas A. Wilcox. 2014. "Estimating Evapotranspiration and Groundwater Flow from Water-Table Fluctuations for a General Wetland Scenario" *Ecohydrology* 7 (2): 378–90. https://doi.org/10.1002/eco.1356.

Cuevas, Jaime, Matías Calvo, Christian Little, Mario Pino, and Paul Dassori. 2010. "Are Diurnal Fluctuations in Streamflow Real?" *Journal of Hydrology and Hydromechanics* 58 (3). https://doi.org/10.2478/v10098-010-0014-0.

Diamond, Jacob S., Daniel L. McLaughlin, Robert A. Slesak, Anthony W. D'Amato, and Brian J. Palik. 2018. "Forested Versus Herbaceous Wetlands: Can Management Mitigate Ecohydrologic Regime Shifts from Invasive Emerald Ash Borer?" *Journal of Environmental Management* 222 (September): 436–46. https://doi.org/10.1016/j.jenvman.2018.05.082.

Dowle, Matt, and Arun Srinivasan. 2019. *Data.Table: Extension of 'data.Frame'*. Manual.

Dunnington, Dewey, and Brenton M. Wiernik. 2020. *Rbbt: R Interface to the Better BiBTex Zotero Connector*. Manual.

Freeman, Lawrence A, Michael C Carpenter, Donald O Rosenberry, Joseph P Rousseau, Randy Unger, and John S McLean. 2004. "Use of Submersible Pressure Transducers in Water-Resources Investigations." In *Book 8: Instrumentation*, A3. Reston, VA: US Geological Survey.

Gribovszki, Zoltán, Péter Kalicz, and József Szilágyi. 2013. "Does the Accuracy of Fine-Scale Water Level Measurements by Vented Pressure Transducers Permit for Diurnal Evapotranspiration Estimation?" *Journal of Hydrology* 488 (April): 166–69. https://doi.org/10.1016/j.jhydrol.2013.03.001.

Hargreaves, George H., and Richard G. Allen. 2003. “History and Evaluation of Hargreaves Evapotranspiration Equation.” *Journal of Irrigation and Drainage Engineering* 129 (1): 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53).

Kirchner, James W. 2009. “Catchments as Simple Dynamical Systems: Catchment Characterization, Rainfall-Runoff Modeling, and Doing Hydrology Backward” *Water Resources Research* 45 (2). https://doi.org/10.1029/2008WR006912.

Lafleur, Peter M., Robbie A. Hember, Stuart W. Admiral, and Nigel T. Roulet. 2005. “Annual and Seasonal Variability in Evapotranspiration and Water Table at a Shrub-Covered Bog in Southern Ontario, Canada.” *Hydrological Processes* 19 (18): 3533–50. https://doi.org/10.1002/hyp.5842.

Landau, William Michael. 2018. “The Drake R Package: A Pipeline Toolkit for Reproducibility and High-Performance Computing.” *Journal of Open Source Software* 3 (21).

Liu, Z., and C. W. Higgins. 2015. “Does Temperature Affect the Accuracy of Vented Pressure Transducer in Fine-Scale Water Level Measurement?” *Geoscientific Instrumentation, Methods and Data Systems* 4 (1): 65–73. https://doi.org/10.5194/gi-4-65-2015.

Loheide II, Steven P. 2008. “A Method for Estimating Subdaily Evapotranspiration of Shallow Groundwater Using Diurnal Water Table Fluctuations.” *Ecohydrology* 1 (1): 59–66. https://doi.org/10.1002/eco.7.

Loheide II, Steven P., James J. Butler, and Steven M. Gorelick. 2005. “Estimation of Groundwater Consumption by Phreatophytes Using Diurnal Water Table Fluctuations: A Saturated-Unsaturated Flow Assessment” *Water Resources Research* 41 (7). https://doi.org/10.1029/2005WR003942.

Maechler, Martin, Peter Rousseeuw, Christophe Croux, Valentin Todorov, Andreas Ruckstuhl, Matias Salibian-Barrera, Tobias Verbeke, Manuel Koller, Eduardo L. T. Conceicao, and Maria Anna di Palma. 2020. *Robustbase: Basic Robust Statistics*. Manual.

McLaughlin, Daniel L., and Matthew J. Cohen. 2011. “Thermal Artifacts in Measurements of Fine-Scale Water Level Variation” *Water Resources Research* 47 (9). https://doi.org/10.1029/2010WR010288.

———. 2014. “Ecosystem Specific Yield for Estimating Evapotranspiration and Groundwater Exchange from Diel Surface Water Variation” *Hydrological Processes* 28 (3): 1495–1506. https://doi.org/10.1002/hyp.9672.

Moore, Mitchell F., Jose G. Vasconcelos, Wesley C. Zech, and Elis P. Soares. 2016. “A Procedure for Resolving Thermal Artifacts in Pressure Transducers.” *Flow Measurement*

and Instrumentation 52 (December): 219–26. https://doi.org/10.1016/j.flowmeasinst.2016.10.010.

Pedersen, Thomas Lin. 2019. *Patchwork: The Composer of Plots*. Manual.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Soylu, Mehmet Evren, John D. Lenters, Erkan Istanbulluoglu, and Steven P. Loheide. 2012. "On Evapotranspiration and Shallow Groundwater Fluctuations: A Fourier-Based Improvement to the White Method" *Water Resources Research* 48 (6). https://doi.org/10.1029/2011WR010964.

Tamari, Serge, and Ariosto Aguilar-Chávez. 2010. "Probando Transductores de Presión Sumergibles Para Monitorear El Nivel Del Agua En Tanques." *Tecnología Y Ciencias Del Agua* 1 (3): 71–88.

Tanaka, M, G Girard, R Davis, A Peuto, and N Bignell. 2001. "Recommended Table for the Density of Water Between 0 C and 40 C Based on Recent Experimental Reports." *Metrologia* 38 (4): 301–9. https://doi.org/10.1088/0026-1394/38/4/3.

Telander, Andrew C., Robert A. Slesak, Anthony W. D'Amato, Brian J. Palik, Kenneth N. Brooks, and Christian F. Lenhart. 2015. "Sap Flow of Black Ash in Wetland Forests of Northern Minnesota, USA: Hydrologic Implications of Tree Mortality Due to Emerald Ash Borer." *Agricultural and Forest Meteorology* 206 (June): 4–11. https://doi.org/10.1016/j.agrformet.2015.02.019.

Van Grinsven, Matthew J., Joseph P. Shannon, Joshua C. Davis, Nicholas W. Bolton, Joseph W. Wagenbrenner, Randall K. Kolka, and Thomas G. Pypker. 2017. "Source Water Contributions and Hydrologic Responses to Simulated Emerald Ash Borer Infestations in Depressional Black Ash Wetlands." *Ecohydrology* 10 (7): e1862. https://doi.org/10.1002/eco.1862.

Watras, C. J., K. A. Morrison, J. L. Rubsam, and I. Buffam. 2017. "Estimates of Evapotranspiration from Contrasting Wisconsin Peatlands Based on Diel Water Table Oscillations." *Ecohydrology* 10 (4): e1834. https://doi.org/10.1002/eco.1834.

White, Walter Noy. 1932. *A Method of Estimating Ground-Water Supplies Based on Discharge by Plants and Evaporation from Soil: Results of Investigations in Escalante Valley, Utah*. Vol. 659. US Government Printing Office.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman and Hall/CRC.

Zhu, Jianting, Michael Young, John Healey, Richard Jasoni, and John Osterberg. 2011. “Interference of River Level Changes on Riparian Zone Evapotranspiration Estimates from Diurnal Groundwater Level Fluctuations.” *Journal of Hydrology* 403 (3-4): 381–89. https://doi.org/10.1016/j.jhydrol.2011.04.016.

4 Black Ash Wetland Response to Future Climate Conditions

4.1 Abstract

Wetlands around the globe are already being impacted by changing temperature and precipitation patterns. Simultaneously black ash forested wetlands are expected to lose much of their overstory canopy due to the invasive insect Emerald Ash Borer (EAB). Field experiments and modeling efforts have provided information on species tolerance of post-EAB conditions and future climate adapted species. No studies have yet examined the interaction of the loss of ash and future climate scenarios on wetland conditions. I developed daily water level models for three wetland canopy conditions: ash forest, non-ash forest, and herbaceous/shrub. Modeled simulations were evaluated under current climate conditions and under two future climate scenarios representing warmer/drier or hotter/wetter conditions. For each canopy-climate combination, 10,000 annual synthetic weather sequences were used as inputs to the water level models. Weather sequences were generated using stochastic weather generators conditioned on 30 years of daily projections using localized constructed analog-downscaled global climate model outputs. Simulated wetland hydrology remained highly variable based on seasonal precipitation and evaporative demand. I compared to the probability of exceeding certain water level thresholds, which represent water level drawdown, surface inundation, and stream-network connectivity. As conditions become drier and warmer, inundation and connectivity become less common, and drawdown increases. Our results present the total impact of the interaction of vegetation and climate change as well as the individual impact for each driver. For non-ash forests, wetland drawdown is increased under future climate conditions compared to ash forests and herbaceous/shrub cover. Simulated non-ash forests were assumed to be less flood adapted than black ash, simulated with lower transpiration rates under inundated conditions. Our simulations suggest persistently wetter wetland conditions under these forests. Replacing black ash forests with non-ash forests has as the same magnitude impact on future hydrology as simulated climate change impacts under both scenarios evaluated. The impact of transitioning to herbaceous/shrub cover has smaller magnitude impact than that of climate change. Both alternative vegetation covers led to wetter conditions relative to black ash, while both potential climate scenarios resulted in drier conditions. Our results show that the interaction of climate change and EAB could result in a canceling effect with each change mitigating the severity of the other. These findings provide insight into the interaction of two common threats to forest ecosystems and can be used to guide management approaches on the landscape.

4.2 Introduction

Shifts in temperature and precipitation (timing, frequency, and quantity) associated with climate change are already impacting global wetland ecology and will intensify into the future (Burkett and Kusler 2000; Moomaw et al. 2018). In general, the Great Lakes region of North America is expected to see declining summer precipitation while total annual precipitation stays stable or increases (Byun and Hamlet 2018; Hayhoe et al.

2010). These projected changes result in increased precipitation in the spring and winter months. Winter precipitation will also experience a phase change, seeing a reduction in the proportion of precipitation as snow, increased occurrence of rain-on-snow melt events in the winter and spring, and earlier snowmelt timing (Notaro, Bennington, and Vavrus 2015). Local conditions will vary and our study area in the western Upper Peninsula of Michigan, USA is projected to see a smaller than regional decline or even slight increases in summer precipitation in numerous downscaling scenarios across multiple GCMs (Byun and Hamlet 2018). Taken together, the shift in precipitation timing, the reduction of total snowfall, and an earlier onset of snowmelt will reshape the annual hydrologic budget. These changes will alter wetland hydroperiods, with water availability increasing in the winter and spring and decreasing in the summer and early fall. The impact of reduced or steady precipitation in the summer will be compounded by increased evaporative demand (Byun and Hamlet 2018). Summer temperatures are expected to increase 2-8 °C (Byun and Hamlet 2018; Hayhoe et al. 2010), and without a commensurate increase in summer precipitation it will lead to more frequent drought conditions and increased water stress on ecosystems.

Black ash (*Fraxinus nigra* Marsh) is an important hardwood component of many forested wetlands in the northern United States and southern Canada. Wetlands with a large or majority black ash component face the same climate change impacts as other wetlands in the region. They also face loss of the major canopy species due to an invasive insect, Emerald Ash Borer (*Agrilus planipennis* Fairmaire, EAB). Emerald ash borer was first found in the United States in southeast Michigan in 2002 (Haack et al. 2002). It is known to infest and cause high mortality in all ash native to North America (Herms and McCullough 2014). As of this writing it is present in 35 US states and 5 Canadian provinces and has invaded the western Upper Peninsula of Michigan in 2020. High mortality and the importance of ash in regional forested wetlands has led to research on the impacts of their loss and strategies to mitigate those impacts. Iverson et al. (2016) evaluated the potential replacement species for black ash in the context of habitat availability, species migration, and replacement species susceptibility to climate change impact, providing a useful resource for climate-informed species replacements for black ash. However, black ash grows in a range of geomorphic settings and local site conditions can have a strong influence on hydrology, forest structure, and plant community among wetlands (Kolka et al. 2018). Considering both Iverson et al. (2016) and Kolka et al. (2018), we can identify four necessary components to evaluate mitigation efforts in black ash wetlands: 1) climate-informed species selection, 2) species tolerance of local site conditions, 3) site conditions following EAB infestation, and 4) site conditions in a future climate.

Previous and ongoing work in Michigan and Minnesota has assessed mitigation in light of three of the four components of the combined impact of climate change and EAB on black ash wetlands. Researchers in Michigan (Bolton et al. 2018), Minnesota (Looney et al. 2015), and Wisconsin (Bolton et al. 2018), planted seedlings to evaluate potential canopy species at the wetland rather than landscape level. The plantings in Bolton et al. (2018) and Looney et al. (2015) took place under simulated EAB infestation, where the seedlings were subjected to adverse conditions due to the hydrologic impact of the loss of

black ash (Slesak et al. 2014; Van Grinsven et al. 2017), as well as the increased competition from herbaceous growth under increased light conditions (Looney et al. 2016, 2017; Davis et al. 2017).

The researchers addressed three of the four components necessary for mitigation efforts in black ash wetlands: 1) some planted species at the northern edge of their current range and evaluated the impact 2) under present-day site conditions and 3) under simulated-EAB site conditions. What is not possible in field studies is to address the interaction of EAB and climate change on site conditions. Just as we have seen EAB impacts cascade through black ash ecosystems affecting hydrology, plant communities, and nutrient cycling, we can expect the future climate-driven changes to hydrology result in similar cascades. Focusing on the future hydrologic characteristics of these wetlands will inform us on the impacts to other functions, as hydrology is the critical control on wetland ecosystems (Brinson 1993). In order to understand future hydrologic conditions, we have developed wetland water level models and evaluated potential changes in water levels under future climate scenarios.

Simulated daily time-step weather conditions are available for various future climate scenarios to serve as model inputs, providing realizations of 30-year climate periods. However, previous studies have demonstrated high-interannual variation in seasonal wetland water level drawdown and rebound in black ash wetlands (Van Grinsven et al. 2017). Estimating the behavior of a highly variable system with 30 years of daily climate projects could lead to biased or high-variance results, making it difficult to draw proper conclusions about the combined impacts of climate change and EAB. A 30-year period provides a representation of normal climate variation, but the weather sequences underlying a certain climate can take many more forms than the observed (or simulated) 30 annual weather sequences. Therefore, simulation studies of alternative scenarios or future conditions require more weather sequences to quantify the most likely response and the distribution of possible responses. Stochastic weather generators (SWGs) provide a tool for generating synthetic time series of weather that simulate conditions under an observed or projected climatology (Wilks and Wilby 1999). SWGs have also been used as a downscaling technique for global or regional climate models (Wilks 2012; Verdin et al. 2015, 2019). One class of SWG known as Richardson (Wilks 2012) are built from parametric representations fit to climate conditions using multiple generalized linear models and varied statistical distributions. This approach can be implemented using any observed (or simulated) weather series that define particular climatic conditions. This characteristic makes them well suited to compare multiple climate scenarios without the overhead cost of executing complete GCM or regional climate model (RCM) runs.

We have performed simulation experiments combining observed wetland hydrology and synthetic weather sequences to quantify the interactions of EAB and future climate scenarios. We developed wetland hydrology models for current canopy conditions (ash-dominated forested wetlands), non-forested conditions (herbaceous and shrub/scrub wetlands), and potential future canopy conditions (forested wetlands under current co-dominant species). We evaluated each class of model under two potential future climate scenarios for the end of the 21st century (2070-2099). The two future scenarios are

defined by a less sensitive GCM (generally projects smaller magnitude regional climate impacts) under the moderate representative concentration pathway (RCP 4.5), and a more sensitive GCM (projects larger magnitude regional climate impacts) under the business-as-usual representative concentration pathway (RCP 8.5) (van Vuuren et al. 2011). The two-scenario 'bookend' approach provides a range of potential future conditions as opposed to a multi-model ensemble approach which masks some of the uncertainty in potential future conditions (Swanston et al. 2016). For an unbiased comparison between current and future conditions both GCMs are used to evaluate the wetland models under historic (1980-2009) climate conditions.

We expect that the interaction of hydrologic impacts of EAB and climate change will result in a tempering of the two individual impacts in this region. While simulation of post-EAB conditions have led to increased water levels and reduced drawdown rates in growing season, future climate conditions in the region will result in reduced water availability during that same period. These two opposing drivers should result in some moderation to both impacts.

4.3 Methods

4.3.1 Study Sites & Data

Wetland water levels measured from 2012 to 2020 at fourteen black ash-dominated wetlands in the western Upper Peninsula of Michigan, USA were used to develop and evaluate our wetland hydrologic models (Figure 1; Table 1). The wetlands range in size from 0.29 to 1.19 ha and 30–80% of the basal area consists of black ash with histosol soils over a confining layer located at an average depth of 118.8 cm (Davis et al. 2017; Van Grinsven et al. 2017). The region has average minimum and maximum annual temperatures of -11.3 °C and 18.2 °C and an average annual precipitation of 101 cm for the climate period of 1980-2009 at the Bergland Dam (46°35'13"N, 89°32'51"W) meteorological station (Arguez et al. 2012). Wetland water levels were continuously monitored in 2" inner-diameter driven wells from 2012-2020 and logged every fifteen minutes using Solinst Levellogger Junior pressure transducers (Solinst, Ontario, CA), with more details available in Van Grinsven et al. (2017). Barometric compensation was performed using data from Solinst Levellogger Junior pressure transducers deployed at a subset of study wetlands. Compensated water levels were corrected for temperature differentials as in Chapter 3. Following an initial control period of two growing seasons, two-thirds of the wetlands in the study were treated to simulate the impacts of an EAB infestation. At one-third of the sites all ash trees greater than 1" in diameter were girdled and at the other half of the treatment sites all ash trees greater than 1" in diameter were hand-felled and left on site.

Daily precipitation and daily minimum/maximum temperatures from existing meteorological stations were used as inputs to the wetland water level models described below. From these input drivers we derived daily solar radiation, PET, precipitation as snowfall and rain, and snowmelt. Precipitation records were retrieved from the National

Table 1. Wetland treatment size for field-study sites used to develop wetland water level models.

Wetland	Treatment	Area (ha)
009	Ash Cut	1.19
053	Ash Cut	0.82
077	Ash Cut	0.60
119	Girdle	0.33
139	Ash Cut	1.54
140	Girdle	0.61
151	Girdle	0.28
156	Ash Cut	0.35

Centers for Environmental Information Hourly Precipitation Dataset ("Hourly Precipitation Data (HPD) Network, Version 2.R2" 2021) using the stations USC00201088 (Bruce Crossing, MI), USC00204328 (Kenton, MI), USC00475352 (Mercer Ranger Station, WI), USC00206215 (Ontonagon, MI), USC00476398 (Park Falls, WI), USC00476518 (Phelps, WI), USC00476939 (Rainbow Reservoir Lake, Tomahawk, WI), USC00477140 (Rice Reservoir, Tomahawk, WI), and USC00208680 (Watersmeet Fish Hatchery) and summed to daily values. Daily minimum and maximum temperatures were taken from the Global Historical Climatology Network (GHCND) dataset (Menne, Durre, Korzeniewski, et al. 2012; Menne, Durre, Vose, et al. 2012) for the same stations, which were collocated stations for HPD and GHCND. Where data were missing, values were filled using inverse distance weighting with data retrieved from the Mesowest (Horel et al. 2002) network stations BPLM4 (Baraga Plains, MI), KTNM4 (Kenton, MI), PIEM4 (Pelkie, MI), WKFM4 (Wakefield, MI), WMTM4 (Watersmeet, MI). Solar radiation data were also retrieved from the listed Mesowest stations. The solar radiation and temperature data were used to fit the Bristow-Campbell method coefficients to estimate solar radiation from latitude and daily temperature range using the PIEM4 Mesowest site (Bristow and Campbell 1984; Bojanowski 2016). Bristow-Campbell solar radiation was used to calculate potential evapotranspiration (PET) via the modified Hargreaves-Samani equation (Hargreaves and Allen 2003). Precipitation was partitioned into snowfall, rain, and snowmelt inputs using the CemaNeige snow accounting routine (SAR) (Valéry, Andréassian, and Perrin 2014). The CemaNeige SAR is temperature-index based and accounts for accumulation, snowpack cold content, and snowmelt through a thermal state weighting coefficient and a degree-day melt coefficient. The CemaNeige SAR is implemented in the R package airGR (Coron et al. 2017).

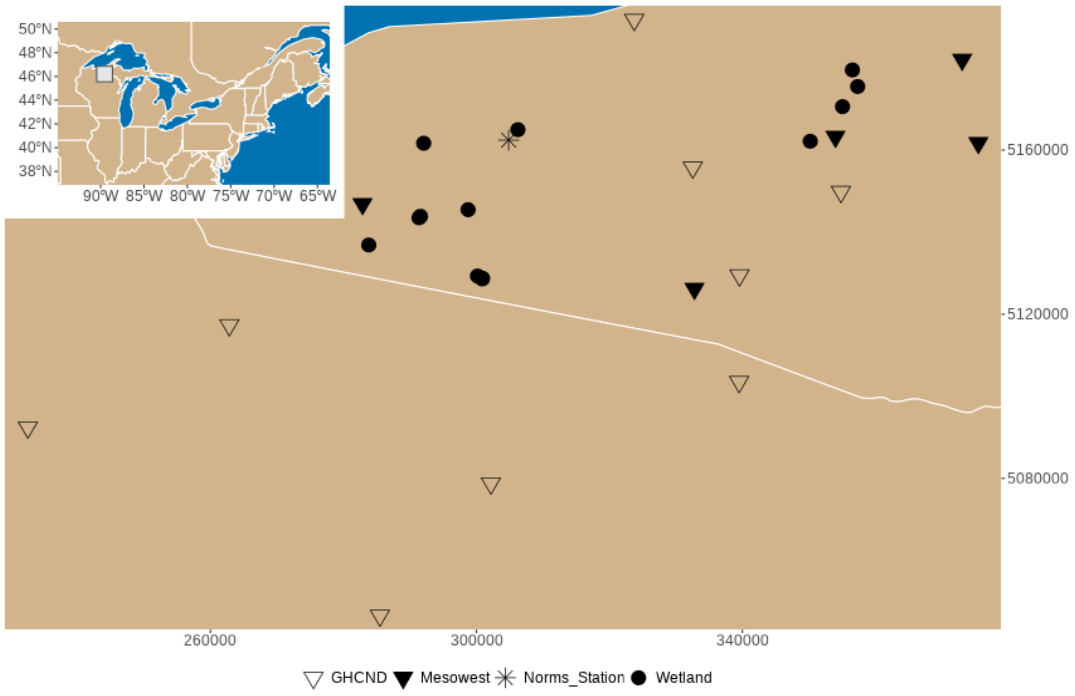


Figure 1. Map of study site and meteorological stations of Global Historic Climate Network – Daily (GHCND), Mesowest, and Norms Station used for data retrieval. Coordinates are in meters, UTM Zone 16N.

4.3.2 Wetland Hydrology Models

The link between hydrology drivers (rainfall, snowmelt, PET) and daily water level response has been shown to be non-linear and varying with stage (Loheide II 2008; White 1932). In wetland systems the relationship between the driver magnitude and response magnitude has been termed the ecosystem specific yield (E_{Sy}) (McLaughlin and Cohen 2014) (see Section 3.2.5 in Chapter 3 on Pressure Transducer Uncertainty for more information about the development of E_{Sy} over time). E_{Sy} has previously been empirically derived as the ratio between precipitation inputs and water level rise ($\frac{P}{\Delta WL}$) (McLaughlin and Cohen 2014). The relationship between empirical E_{Sy} and water level can then be modeled to provide a continuous estimate of E_{Sy}. Models fit to the E_{Sy}~Water Level relationship include exponential (McLaughlin and Cohen 2014; Watras et al. 2017), quadratic (McLaughlin and Cohen 2014), and step-wise regression (McLaughlin and Cohen 2014). Identifying E_{Sy} using the rainfall-rise ratio can be unworkable in the presence of confounding hydrologic variation such as surface water connectivity and low-frequency seasonal changes in water availability (Watras et al. 2017; Zhu et al. 2011). Both of these factors were present in our study wetlands and no clear relationship existed to model E_{Sy} with the above model forms, requiring an alternative approach.

As an alternative we fit an inverse analog to $\frac{P}{\Delta WL}$, deriving E_{Sy} from the ratio of cumulative water availability and water level. We began by fitting a quadratic curve to

the relationship between a year-to-date water availability index from the beginning of the growing season to the point of minimum wetland water level (Figure 2), where

$$WA_{YTD} = P_{YTD} + M_{YTD} - PET_{YTD},$$

Eq. 1

WA_{YTD}, P_{YTD}, M_{YTD}, and PET_{YTD} are year-to-date water availability, rainfall, snow melt, and potential evapotranspiration, respectively. In developing this index, it is assumed that $PET_{YTD} = cAET_{YTD}$, where c is an unvarying constant representing a stable relationship between *PET* and *AET*. For this portion of the water year these sites are not water-limited (Van Grinsven, 2017). Empirical E_{Sy} was then considered as the first derivative of the quadratic curve, which has the form $\frac{\Delta\ WL}{\Delta\ WA}$. This approach resulted in an asymptotic relationship between E_{Sy} and water level, suggesting agreement with the exponential forms in McLaughlin and Cohen (2014) and Watras et al. (2017). Defining E_{Sy} as $\frac{\Delta\ WL}{\Delta\ WA}$ provided additional information to fit the relationship between E_{Sy} and water level by allowing the use of days without rainfall in the development of E_{Sy}. Models for E_{Sy} were fit using the year with the greatest water level drawdown for each wetland to capture the widest range of E_{Sy} variation. Each wetland had the same basic form of E_{Sy} function. The functions were fit using a single hierarchical model with each of the coefficients allowed to vary independently within sites. Models were fit using the brms package in R (Bürkner 2017, 2018). The asymptotic structure of the E_{Sy} function requires that a lower bound be placed on E_{Sy} predictions to avoid values less than or equal to zero. The minimum value of E_{Sy} was allowed to vary by wetland and was fit along with other model parameters as described below.

Wetland water levels were simulated on a continuous basis using daily inputs. For each daily step, wetland water level change was determined as a function of the current water level, daily rainfall (*R*), daily *PET*, daily snowmelt (*M*), and estimated streamflow (*Q*) (Equation 2; Table 1). Maximum wetland water levels were estimated from the the control period as the mode of the wetland water level record. Streamflow was assumed to occur whenever wetland water levels were at or above the maximum water level. This is similar to the approach in McLaughlin et al. (2019) where wetland surface water connectivity was determined from wetland water level records. Daily water level was used to estimate $\hat{E}_{Sy}$, which served as a multiplier for water level response to rainfall, *PET*, and snowmelt components. In addition, each driver (*R*, *PET*, *M*, *Q*) had a coefficient (fitting described below). Snowmelt and precipitation were also fit with a first order autoregressive filter to simulate slow flow contributions to wetland water levels. Within a single day only the larger of *PET* or *P* was used as driver of water level change, accounting for suppressed transpiration from wet leaves.

Training data for wetland water level models was selected for each wetland as the year with the greatest water level drawdown within the control period (Table 3). The remaining years of data in the control-period years were used for wetland model

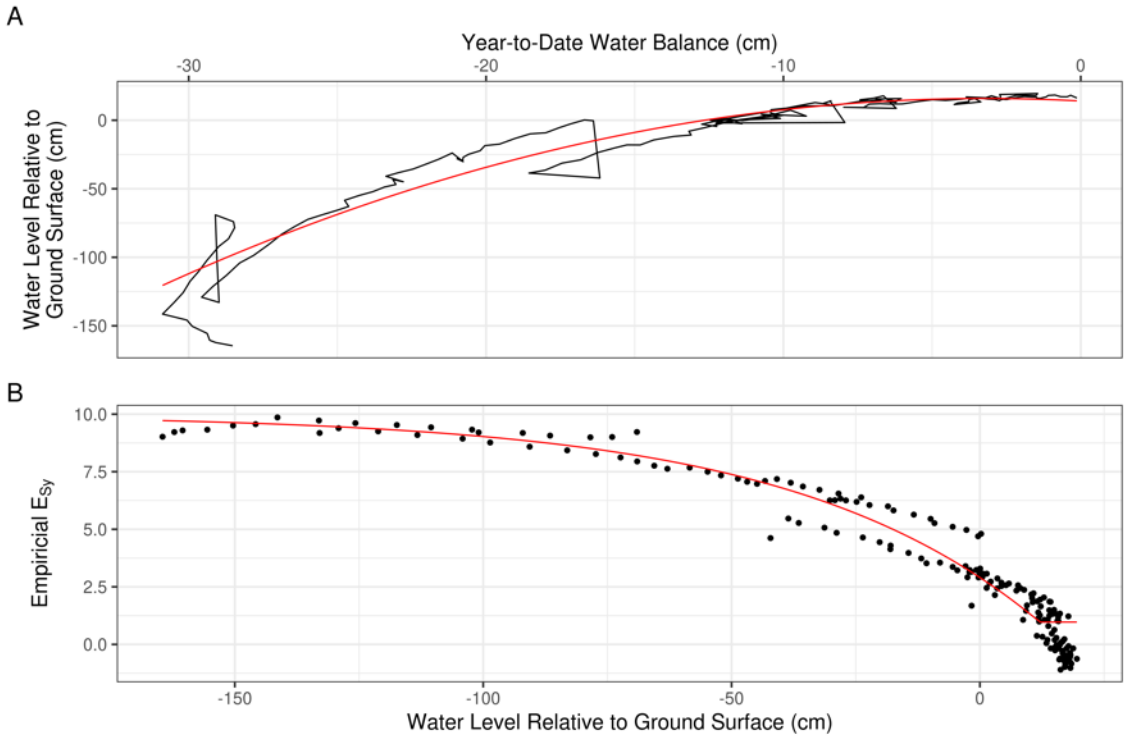


Figure 2. Derivation of ecosystem specific yield for a single study wetland. Panel A shows the quadratic relationship between wetland water level and water balance (total liquid inputs minus potential evapotranspiration) for the drawdown period. Panel B shows the relationship between the derivative of the fitted line from Panel A, representing an empirical estimate of E_{Sy} and wetland water level.

Table 2. Parameter notation and description for wetland water level models.

Parameter	**Definition**
t	Daily time step
WL	Wetland water level (relative to ground surface)
WL_{max}	Maximum wetland water level (relative to ground surface)
M, β_M	Snowmelt and snowmelt coefficient
Q, β_Q	Streamflow and streamflow coefficient
R, β_R	Rainfall and rainfall coefficient
PET, $\beta_P ET$	Potential evapotranspiration and potential evapotranspiration coefficient
$f_E Sy$	Estimate of ecosystem specific yield
ϕ_M	Autoregressive coefficient for snowmelt
ϕ_R	Autoregressive coefficient for rainfall
a, b, c	Fitted coefficients for asymptotic regression

evaluation. Wetland parameters were estimated using the L-BFGS-B algorithm via the R optim() function (R Core Team 2019) minimizing the weighted root-mean square error of the predictions. Weights were applied asymmetrically where days with observed water levels greater than or equal to WL_{max} were assigned the original equal weight (1/n). As observed water levels decreased below WL_{max} weights increased as the square of the difference between observed water level and WL_{max} increased. This weighting structure gave increased weight to the drawdown period of the hydroperiod, which was otherwise dominated by surface water conditions. To model impacted wetlands that have not recovered to a fully forested state, the observed treatment data from the same site were used to reparameterize the control condition model. Treatment-period training data were selected in the same manner described above to again maximize water level drawdown in the training set. The wetland models were reparameterized for only $\hat{\beta}_{PET}$, $\hat{\beta}_{R}$ terms. Finally, we simulated reforested black ash wetlands under one set of potential future forest composition. For the future forest conditions, we assumed a mix of the current co-dominant species, red maple and yellow birch, would become established with similar stand basal area. Future forested simulations were performed using the same parameters as the control period with a reduction of $\hat{\beta}_{PET}$. The reduction in $\hat{\beta}_{PET}$ was a function of water level and current proportion of site basal area as black ash:

$$\hat{\beta}_{PET-Adj} = \hat{\beta}_{PET} * [(1 - BA_{ash}) + BA_{ash} * [1.45077 - 0.05869 * (WL - WL_{max})]^{-1}] \quad \text{Eq. 2}$$

This equation is derived from previous work showing a water level-dependent difference in sap flux between black ash and current co-dominant species (Shannon et al. 2018).

Because individual models were trained for each wetland, there are no training data available for non-forested conditions under the field-study control sites. Therefore, we fit wetland water level models to only 8 of the field-study wetlands: the 8 treatment (girdle and ash cut) wetlands. Comparisons of the combined and separate impacts of EAB and climate change were performed against modeled baselines rather than field-study observed conditions. Limiting the number of sites and comparing to modeled future simulated baselines with each alternative vegetation condition ensures that comparisons are not biased by the number of sites within each group. Control conditions are represented by the modeled baselines, avoiding the potential of identifying model artifacts as significant when comparing observed and modeled results.

Table 3. Years of data used for model training and testing for each wetland model.

Site	Model Training Year	Model Testing Year
009	2012	2019
053	2013	2019
077	2012	2015
119	2012	2018
139	2012	2018
140	2012	2019
151	2012	2015
156	2012	2015

4.3.3 Future Climate Conditions

Wetland models were run under simulated current (1980-2009) and future (2070-2099) climate conditions. Future climate forcing was evaluated under the RCP 4.5 and RCP 8.5 representative concentration (of greenhouse gas) pathway scenarios. As described in van Vuuren et al. (2011) RCP 4.5 represents moderate future climate forcing while RCP 8.5 represents a future forcing under increasing growth in GHG emissions. Daily climate was projected for all scenario-period combinations by the General Fluid Dynamics Lab Coupled Model (GFDL-CM3) and National Center for Atmospheric Research Community Earth System Model (CCSM4) global climate models (GCMS) and downscaled by the localized constructed analogs (LOCA) (Pierce, Cayan, and Thrasher 2014; Gent et al. 2011; Griffies et al. 2011). LOCA downscaled data were retrieved from the downscaled CMIP3 and CMIP5 Climate and Hydrology Projections archive (http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/).

Performance of GCMs in the North American Great Lake region varies mostly due to the unaccounted regional climate impact of the Great Lakes (Notaro, Bennington, and Vavrus 2015; Rood and Briley 2018). GCM model selection was guided by performance and sensitivity (magnitude of change under a given emissions scenario relative to other GCMs) of the models in the Great Lakes region (Byun and Hamlet 2018). Both GFDL-CM3 and CCSM4 were found to be well performing models for the Great Lakes region, and CCSM4 and GFDL-CM3 represent models that show less and more sensitive responses to given RCP forcing, respectively. Inspection of LOCA daily simulations of GFDL-CM3 and CCSM4 for the 1980-2009 period show excellent alignment with observed conditions during the same period (Figure 3). There appears to be fewer dry days in the LOCA dataset relative to the observed dataset. Monthly and seasonal precipitation totals showed good agreement between the LOCA and observed datasets and no bias-correction was performed. The two future climate scenarios used for simulations were the CCSM4 under RCP 4.5 and GFDL-CM3 under RCP 8.5, respectively referred to as the warm & dry and hot & wet scenarios based on projections for the study area.

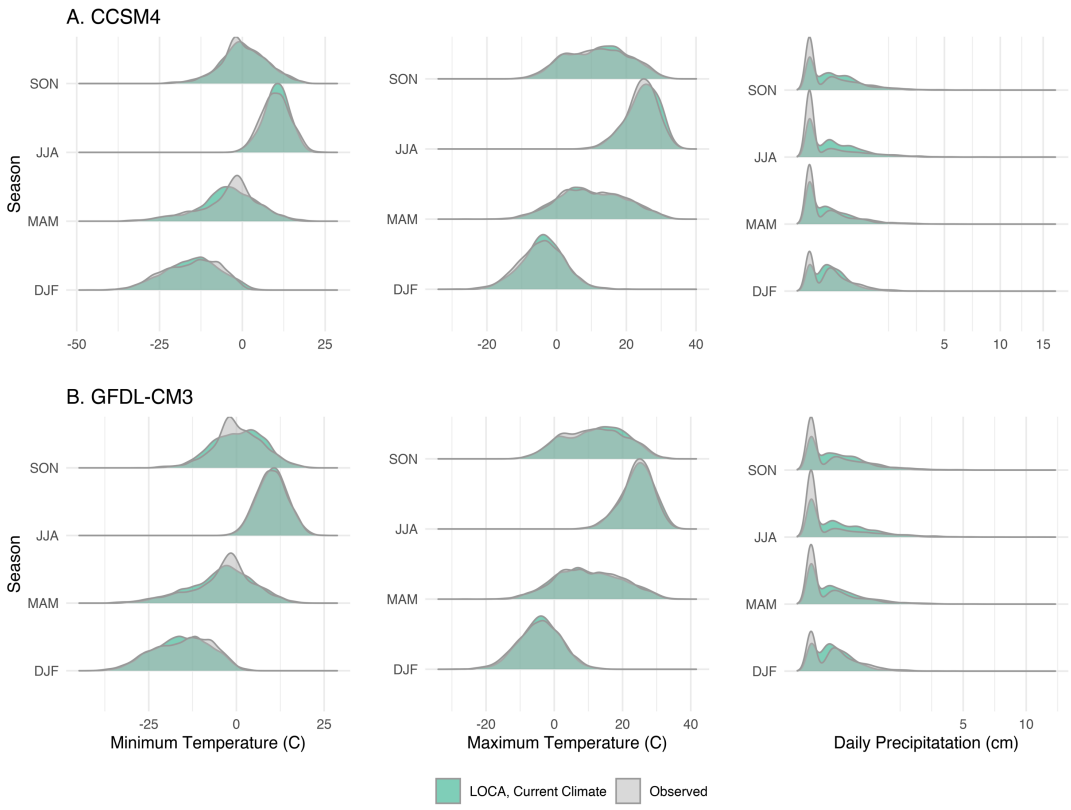


Figure 3. Comparison of LOCA daily projections and observed values for the climate normal period (1980-2009) from the GFDL-CM3 (A-C) and CCSM4 (D-F) global climate models.

4.3.4 Stochastic Weather Generator

The objective of projecting wetland conditions under future climate scenarios is to understand the expected response, and the range of possible responses. The 30-year periods of daily projections above provide a representation of the normal climate variation under each scenario (Baddour and Kontongomde 2007). Quantifying the range of weather patterns under future climate scenarios and the corresponding wetland responses is best performed with many more years of simulation using stochastic processes (Wilks 2012). While the LOCA downscaled-GCM data represents a model of physical processes, stochastic weather generators (SWGs) are statistical models used to generate additional weather series that are drawn from the statistical distributions the original modeled data. We followed the generalized linear model (GLM) based SWG described in Verdin et al. (2015), which is a type of the more general Richardson SWG (Richardson 1981).

Our SWG consists of four GLM models to simulate precipitation occurrence, precipitation amount, minimum daily temperature, and maximum daily temperature. In the form, minimum and maximum daily temperatures are fit using separate models with the same model form for each station of interest. Each daily value is predicted using the

previous day's minimum and maximum temperatures, a pair of Fourier harmonics, and seasonal mean minimum and maximum temperatures. To simulate the daily variation inherent within a climate season additional error is added to the predicted daily minimum and maximum temperatures. The additional error is drawn from a multivariate skew-normal distribution conditioned on precipitation occurrence. This approach captures the correlation between daily minimum and maximum temperatures and between daily temperatures and precipitation occurrence. Precipitation occurrence is modeled using the logistic family with a *probit* link function. Occurrence is considered as a first order Markov process dependent upon the previous day's precipitation occurrence (Wilks and Wilby 1999). Occurrence models also include a pair of Fourier harmonics to capture any seasonality in precipitation patterns, and the seasonal totals of precipitation for the observed period as covariates. Precipitation amounts are drawn from a gamma distribution defined by shape and scale parameters. The shape parameter is extracted from a gamma-family GLM with a log link fit to the observed precipitation amounts. The predictors in the model are a pair of Fourier harmonics and the seasonal precipitation totals. The scale of gamma distribution describing the distribution of storm sizes is a function of the GLM gamma-family shape and the predictors from GLM fit to precipitation amounts. This approach allows the storm distribution to vary temporally and capture seasonality in not only total precipitation but also the size of the storms. When seasonal means and totals are included as predictors within the GLM predictors it is said that the SWG is conditioned on the observed climate. When applying a conditioned SWG to a new climate projection it is assumed that the distribution of the precipitation amounts will remain constant under the new seasonally-defined climate scenario (Wilks 2012). We removed this assumption from our analysis by developing separate SWGs for each model-scenario combination using the LOCA daily projections for GLM fitting.

The work in Verdin et al. (2015) shows that these SWGs can capture and simulate the spatial correlation in temperature and precipitation. We did not include spatial correlation in our SWGs because of the relatively small size of the study area and the fact that each wetland is considered an independent system. For each model-scenario combination the LOCA-downscaled projected daily data from the coordinates of the Bergland Dam meteorological station were used to fit the SWG GLMs. Each model-scenario SWG was used to simulate 10,000 individual annual synthetic weather series with seasonal conditioning drawn randomly from the 30 years of projection data. Conditioning the models on individual years of observed data is intended to increase interannual (or inter-simulation) variability, which can otherwise be limited in SWGs (Wilks and Wilby 1999).

4.3.5 Data Analysis

Tests for statistical significance are unperformed and unreported with a single exception. This research is entirely dependent on simulation data drawing conclusions from modeled wetland water levels driven by synthetic weather series generated from parametric descriptions of LOCA downscaled GCM projections. Reporting statistical significance and *p*-values would provide a false sense of certainty and dichotomy to the future climate and wetland conditions. We report our results by contrasting the probability of observing

certain wetland conditions under different vegetation types and climate scenarios and account for the uncertainty inherent from this type of simulation work. As an exception, the performance of water level models are evaluated using a mixed-effects linear model to demonstrate no statistical difference between observed and modeled water levels.

4.3.5.1 Total, EAB, and Climate Impacts

The distribution of modeled wetland water levels was evaluated for each future vegetation and climate scenario combination. The 10,000 simulations for each combination were summarized by simulated day of year to compute the median and the 67% highest density credible intervals (HDCIs) (McElreath 2020) for each combination. HDCIs are potentially asymmetric intervals that contain the stated proportion of all observations. They differ from confidence intervals providing information about the most probable range for results rather than identifying the range of values that would be expected to contain the true mean. This approach was used to quantify future conditions and baseline scenarios for comparison and benchmarking. Modeled baselines provide consistency between the baseline, or control, period and alternative vegetation cover and climate scenarios, which is critical for drawing meaningful conclusions. Apart from the advantage of consistency, modeled baselines can also provide a flexible tool to answer more questions about the drivers of the observed changes. Different baselines can be computed from the simulations by combining the six vegetation-climate combinations. To evaluate total impact of EAB and climate change, future non-black ash conditions were compared to black ash under the current climate. Alternative baseline comparisons include comparing each vegetation cover to itself under alternative climate scenarios, and comparing vegetation covers to each other within a climate scenario. These two alternative baselines allowed us to attribute how much of the observed total impact was attributable to each driver.

4.3.5.2 Critical Ecohydrological Thresholds

Observed water levels within these wetlands show high interannual variation under field control and treatment conditions. Rather than considering this low-signal, high-variance variable, we defined critical ecohydrological thresholds (*CEHTs*) as a measure of impact. *CEHTs* were set to capture wetland **connectivity** to the downstream hydrologic network via streamflow and subsurface flow, **inundation** when wetland water levels were near the soil surface with surface water likely in microtopographic hollows, and **drawdown** when wetland water levels dropped far below the surface of the wetland (Table 4). Wetland water levels were compared to these thresholds and the number of days that a wetland was above or below a given threshold could be used to calculate the probability of occurrence of that *CEHT*. We chose to calculate these probabilities at the monthly scale, though they could be computed for other time scales from daily to annually.

4.3.5.3 Wetland Model Performance

Wetland water level model performance was evaluated using the retained independent testing datasets. We calculated R^2 as a metric for the relationship between observed and modeled values. R^2 cannot be used to evaluate the accuracy of the models because consistent over- or under-predictions can still result in high R^2 values (Krause, Boyle, and

Bäse 2005). The median error of daily modeled water level was used to measure model bias and the root median squared error (RMedSE) of the daily modeled water level to assess overall predictive accuracy. RMedSE was used in place of RMSE because we expect some outlier errors where regionally-informed rainfall records do not match observed wetland rainfall. Relative RMedSE was calculated as the RMedSE relative to the observed annual range in daily wetland water levels to provide additional context on the scale of the errors. In addition to the overall accuracy of the models we want the models to accurately capture the probability of occurrence of the *CEHT*s. We tested the predicted probabilities of inundation, connectivity, and drawdown to the observed probabilities of the same conditions using a linear mixed effects model with dependent variable probability of occurrence, population effects of observed/predicted and wetland status, and a group effect for wetland (Lenth 2021). Summary and statistical tests were performed using estimated marginal means (Bates et al. 2015).

Table 4. Water level thresholds and definition of three conditions considered as critical ecohydrological thresholds. The range from 10-50 cm below the wetland surface is a high probability occurrence under all vegetation and climate conditions, revealing little about the impact of either driver. The occurrence of these thresholds was used to compare the impacts of EAB and climate change on the future of black ash wetland hydrology.

Name	Water Level Threshold	Definition
Connectivity	Within or above 5 cm below maximum site water level	Contributing water downstream
Inundation	Within or above 10 cm below wetland surface	Surface flooding or soil saturation
Drawdown	At or below 50 cm below wetland surface	Dry surface and non-saturated soils

4.4 Results

4.4.1 SWG Performance

Synthetic weather generator structure and fitting can be evaluated by comparing generated weather sequences with the underlying observed/simulated climate data. The summary statistics of the sequences should align well with underlying data and the distribution of observed values should closely match the shape of the underlying data (Gregory, Wigley, and Jones 1993). Table 5 shows seasonal and climate model scenario summaries of LOCA and SWG variables. Values are reported as the mean and a 95% confidence interval (1.96 times standard deviations) around the mean. The percent of outliers were calculated using the intervals defined by the LOCA dataset. When comparing all of the simulated SWG data (equivalent to 10,000 years) the mean values and interquartile ranges of daily minimum and maximum temperature and precipitation align well with the LOCA-generated daily values (Table 5). Fewer total monthly precipitation outliers were observed for the SWG synthetic series than the LOCA climatology. Minimum and maximum temperatures showed an increase in outliers in the synthetic weather series for JJA (June, July, August) and SON (September, October, November) seasons, and a decrease in outliers for the DJF (December, January,

February) and MAM (March, April, May) seasons. The difference in outliers was small and may be partially explained by the scale of the two datasets. The LOCA data represents 30 years of data compared to the 10,000 years of synthetic weather series. To compare the SWG and LOCA values as climates, we randomly sampled 30 sets of 30-year SWG data from the observed 10,000 synthetic series. The distribution of these samples was compared to the distribution of LOCA values in Figure 4. Figure 4A shows good agreement for all seasons for daily minimum and maximum temperature. The distribution of simulated monthly precipitation totals is shown in Figure 4B. Compared to minimum and maximum temperature, monthly precipitation totals show greater variability among the 30 sets of 30-year synthetic weather series. All three variables show that the synthetic weather climates cluster around the LOCA value as a mean.

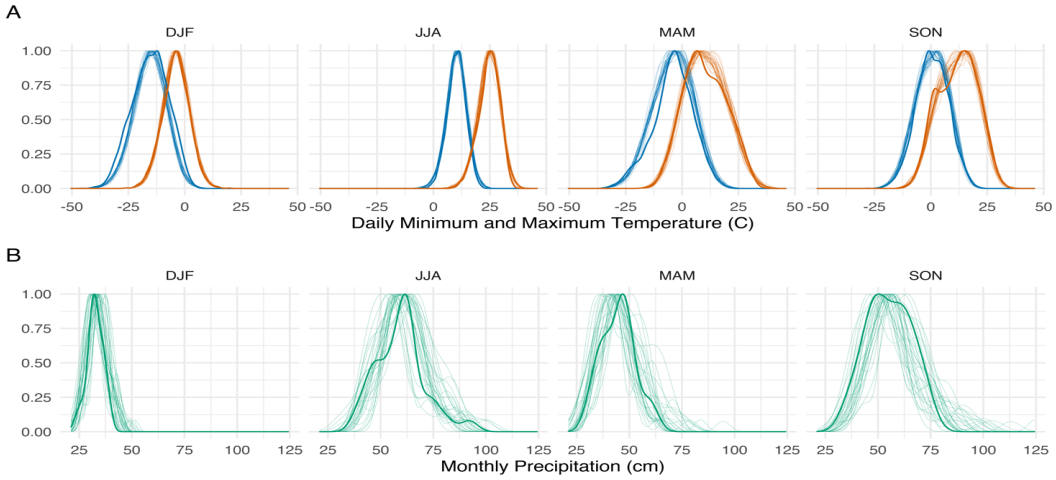


Figure 4. Density plots of 30 year climatology of LOCA generated data (bold), and 30 random samples of 30-year climatology periods from the full 10,000 years of SWG synthetic weather series (light).

4.4.2 Wetland Water Level Model Performance

One wetland, 119, had all of the withheld training years with R^2 below 0.6, which was chosen as a threshold for unsatisfactory model results based on work by Moriasi et al. (2015) (Figure 5). Models at this wetland did not show noticeably higher rates of bias or error relative to the other wetlands. Site 119 is a closed basin with no surface outlet and was shown to have different source water characteristics from other sites, showing much less connectivity to groundwater sources (Van Grinsven et al. 2017). Table 6 summarizes the wetland model metrics by site status, presenting the minimum, maximum, mean, and median values seen across all site-year combinations. Mean and median model performance was the similar between site statuses for all model metrics (Table 6). After excluding wetland 119, R^2 values ranged from 0.44–0.95. The models had an overall negative bias, which indicates drier wetland conditions than were observed (Table 6). Both the median rRMedSE is moderate for both Control and Treated condition models (~10%), but notable outliers were present (wetland 053) (Table 6).

Table 5. Comparison of 30 years of LOCA-generated climate values and 10,000-year SWG simulated weather series. Value are presented as mean with the 95% confidence interval in parentheses (± 1.96*SD). Percentage of outliers is the number of each value type that falls outside of the interval defined by ± 1.96*SD for the LOCA-generated values.

	Season	LOCA	SWG
Minimum Temperature	*CCSM4*		
	DJF	-15.11 (-31.25, 1.04); 4.21% outliers	-15.14 (-29.70, -0.57); 2.96% outliers
	MAM	-4.22 (-21.68, 13.24); 6.34% outliers	-4.22 (-21.40, 12.96); 4.46% outliers
	JJA	10.26 (2.69, 17.83); 5.14% outliers	10.23 (1.62, 18.85); 8.48% outliers
	SON	0.57 (-13.08, 14.21); 5.64% outliers	0.56 (-13.51, 14.63); 5.30% outliers
	GFDL-CM3		
	DJF	-15.48 (-31.03, 0.07); 3.84% outliers	-15.60 (-29.43, -1.76); 2.73% outliers
	MAM	-4.02 (-21.73, 13.69); 5.18% outliers	-3.98 (-21.97, 14.01); 4.87% outliers
	JJA	10.27 (2.03, 18.51); 4.93% outliers	10.29 (1.30, 19.28); 7.27% outliers
	SON	0.73 (-13.62, 15.09); 4.43% outliers	0.63 (-14.30, 15.56); 5.10% outliers
Maximum Temperature	*CCSM4*		
	DJF	-3.84 (-15.80, 8.11); 5.58% outliers	-3.74 (-15.38, 7.91); 4.52% outliers
	MAM	10.27 (-8.24, 28.78); 3.70% outliers	10.34 (-7.84, 28.53); 3.53% outliers
	JJA	24.80 (15.71, 33.89); 4.78% outliers	24.82 (14.81, 34.84); 7.40% outliers
	SON	11.81 (-5.10, 28.72); 2.78% outliers	11.78 (-4.78, 28.34); 3.20% outliers
	GFDL-CM3		
	DJF	-4.16 (-15.87, 7.54); 4.80% outliers	-4.10 (-15.51, 7.31); 4.54% outliers
	MAM	10.20 (-8.53, 28.93); 3.48% outliers	10.23 (-8.59, 29.06); 4.25% outliers
	JJA	24.51 (15.37, 33.65); 4.82% outliers	24.56 (15.04, 34.07); 5.89% outliers
	SON	11.98 (-5.38, 29.34); 2.82% outliers	12.07 (-4.86, 29.00); 2.96% outliers
Precipitation	*CCSM4*		
	DJF	0.19 (-0.33, 0.70); 4.80% outliers	0.19 (-0.29, 0.68); 4.67% outliers
	MAM	0.24 (-0.63, 1.10); 5.04% outliers	0.24 (-0.50, 0.98); 3.91% outliers
	JJA	0.32 (-0.90, 1.53); 4.13% outliers	0.32 (-0.62, 1.26); 3.19% outliers
	SON	0.31 (-0.71, 1.33); 4.65% outliers	0.32 (-0.60, 1.25); 4.19% outliers
	GFDL-CM3		
	DJF	0.18 (-0.32, 0.67); 4.77% outliers	0.18 (-0.28, 0.65); 4.70% outliers
	MAM	0.25 (-0.66, 1.16); 5.25% outliers	0.25 (-0.52, 1.02); 3.76% outliers
	JJA	0.33 (-0.84, 1.50); 4.38% outliers	0.34 (-0.64, 1.32); 3.82% outliers
	SON	0.29 (-0.61, 1.20); 4.84% outliers	0.31 (-0.58, 1.19); 4.84% outliers

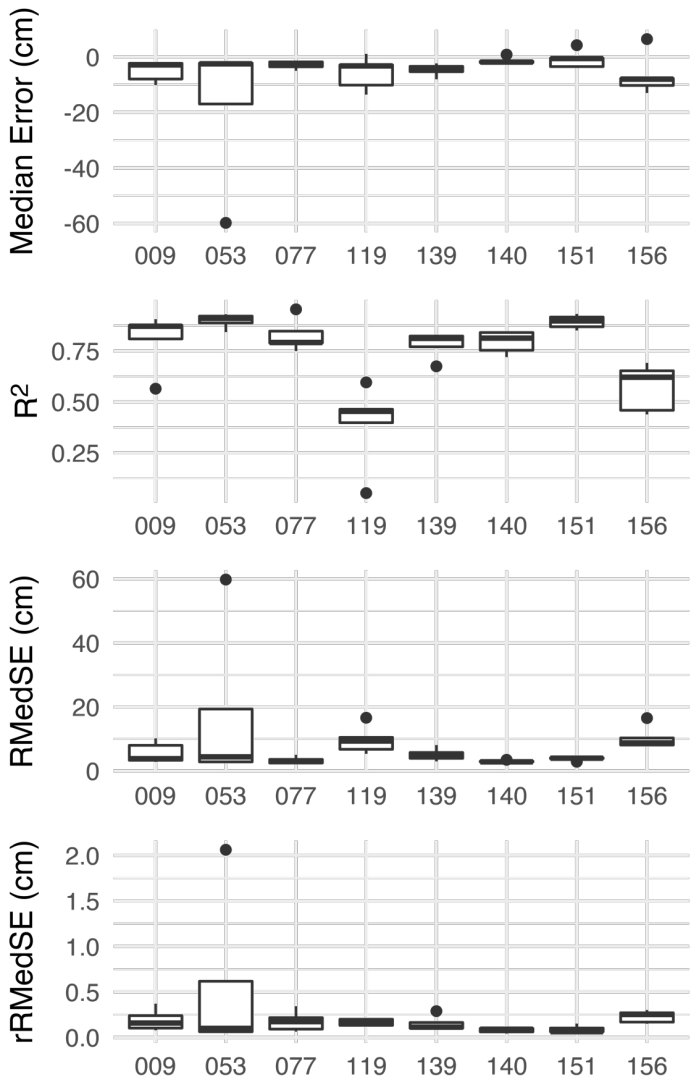


Figure 5. Wetland model performance metrics. Metrics were calculated for each site-year combination and are presented within sites. RMedSE and rRMedSE are root median squared error and root median squared error relative to the annual range of daily wetland water levels within that year.

When compared to observed wetland water levels, modeled wetland water levels showed no significant difference in probability of occurrence of *CEHT*s (Table 7). The range of site-level probability of each level of interest were similar between modeled and observed data (Figure 6). Though not significant, these results show a slight systematic bias towards drier conditions in the modeled wetland water levels. All remaining comparisons of current conditions to combinations of future vegetation conditions and climate scenarios will use modeled current conditions. This comparison ensures that the described model bias does not impact results or conclusions, but it does create a disconnect between observed conditions and expected changes. It also sets equal sample sizes between the control and future conditions.

Table 6. Minimum, maximum, and median values of wetland water level model metrics, comparing observed and modeled wetland water levels. Metrics were evaluated on withheld test data for each combination of site and year of data. Prior to calculating these metrics one sites with incompatible hydrology for the model structure were removed from the analysis.

Model Metric	Site Condition	Minimum	Maximum	Mean	Median
R^2	Control	0.44	0.95	0.71	0.72
	Treated	0.05	0.93	0.75	0.81
Median Error (cm)	Control	-13.61	0.79	-4.67	-2.01
	Treated	-59.79	6.37	-5.38	-2.94
RMedSE	Control	2.48	16.61	7.46	4.11
	Treated	2.37	59.79	7.04	4.26
rRMedSE	Control	0.04	0.21	0.12	0.14
	Treated	0.06	2.06	0.22	0.13

Table 7. Tests comparing predicted probability of occurrence for critical ecohydrological thresholds (*CEHT*). Reported values show the seasonal probability of daily water level exceeding the respective *CEHT*. Control and treated conditions refer to **observed** control and treated conditions, respectively, or **predicted** modeled black ash and modeled non-forested conditions. All data presented derived from only the withheld test period for each wetland.

	Status	Observed	Predicted
Connectivity	Control	0.72 (0.52, 0.91)	0.73 (0.55, 0.92)
	Treated	0.74 (0.65, 0.84)	0.71 (0.58, 0.83)
Drawdown	Control	0.00 (-0.05, 0.05)	0.02 (-0.05, 0.09)
	Treated	0.00 (-0.02, 0.02)	0.02 (-0.03, 0.08)
Inundation	Control	0.90 (0.80, 1.01)	0.90 (0.78, 1.02)
	Treated	0.92 (0.84, 1.01)	0.87 (0.77, 0.96)

4.4.3 Wetland Water Levels

Wetland water levels under future climate conditions were highly variable under both scenarios and both vegetation types (Figure 7). Figure 7 shows the change from the modeled baseline of probability that wetland water levels will be above or below the *CEHT*s defined above. The reported probability is the proportion of observations where daily water level surpasses each threshold out of all modeled daily water levels in a month (~240,000). The median probability and a range representing 67% of all observations (highest density continuous intervals, HDCI) are shown as point and line estimates for the warm & dry and hot & wet sensitive future climate scenarios. Modeled black ash conditions are also reported as the median and 67% HDCI, represented as a crossbar and shaded area.

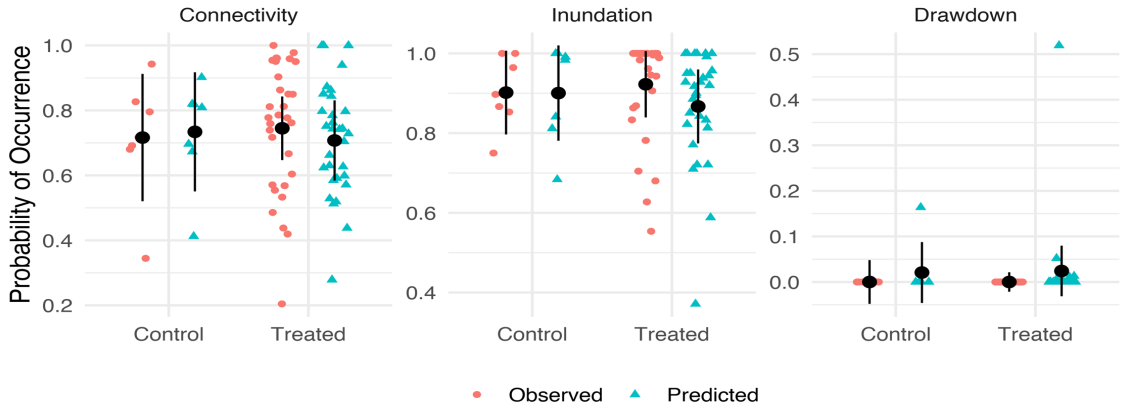


Figure 6. Tests comparing predicted probability of occurrence for critical ecohydrological thresholds (*CEHT*). - Reported values show the seasonal probability of daily water level exceeding the respective *CEHT*. Control and treated conditions refer to **observed** control and treated conditions or **predicted** modeled black ash and modeled non-forested conditions. Black point and error bar represent estimated marginal mean and its 95% confidence interval. Individual points represent data observed or simulated for one year at one site. All data presented derived only from the withheld test period for each wetland.

For non-forested conditions both future climate scenarios showed a decrease in probability of connectivity relative to current wetland conditions (Figure 7A). The warm & dry climate scenario consistently showed a lower median probability of connectivity and was more likely to have simulated conditions outside of the HDCI of current conditions. The hot & wet climate scenario showed greatest agreement with current conditions in July and August. Future forested conditions were more likely to show connectivity in July and August under the more sensitive climate scenario, and only slightly less likely than current conditions under the warm & dry climate scenario (Figure 7B). Connectivity to the larger hydrologic network was much less likely to occur under the hot & wet climate scenario and both vegetation conditions in May (Figure 7A & B). Under both vegetation conditions the warm & dry climate scenario resulted in a lower probability of connectivity relative to the hot & wet climate scenario. While both vegetation conditions showed significant overlap between the warm & dry climate conditions and control conditions, the non-forested conditions tended to skew towards current drier years (lower probability of connectivity) and future forested conditions skewed towards current wetter years (high probability of connectivity).

Inundation results in Figure 7C & D closely resemble the patterns observed in the connectivity results. Probabilities of inundation and connectivity are not exclusive. This is in line with fit model parameters that showed maximum sustained water levels were above the surface for almost every wetland and treatment combination. Generally, inundated conditions would be less prevalent under future non-forested conditions than the future-forested conditions. For non-forested conditions there was little overlap between baseline conditions and the warm & dry climate scenario results (Figure 7C). This was contrary to forested conditions where the warm & dry climate scenario results

showed more overlap with baseline conditions, and the hot & wet scenario varied by month relative to baseline conditions. Future-forested conditions showed increased probability for inundation in July, August, and September and both vegetation conditions showed lower probability for inundation in June (Figures 7C and 7D).

The probability that water levels will drop to less than 50 cm below the wetland soil surface agree with the connectivity and inundation results (Figure 7E & F). Under both climate scenarios the future forest conditions showed a lower probability of drawdown than under baseline conditions, with almost no overlap with current conditions in July, August, and September (Figure 7F). Drawdown events were very unlikely with near zero probability of occurring under future forested conditions under both climate scenarios. Probability of drawdown in non-forested conditions increased in both magnitude and variability in August and September under the warm & dry future climate scenario, and increased in variability in the same period under the hot & wet future climate scenario. Differences from current conditions were greatest in August under the warm & dry climate scenario (Figure 7E).

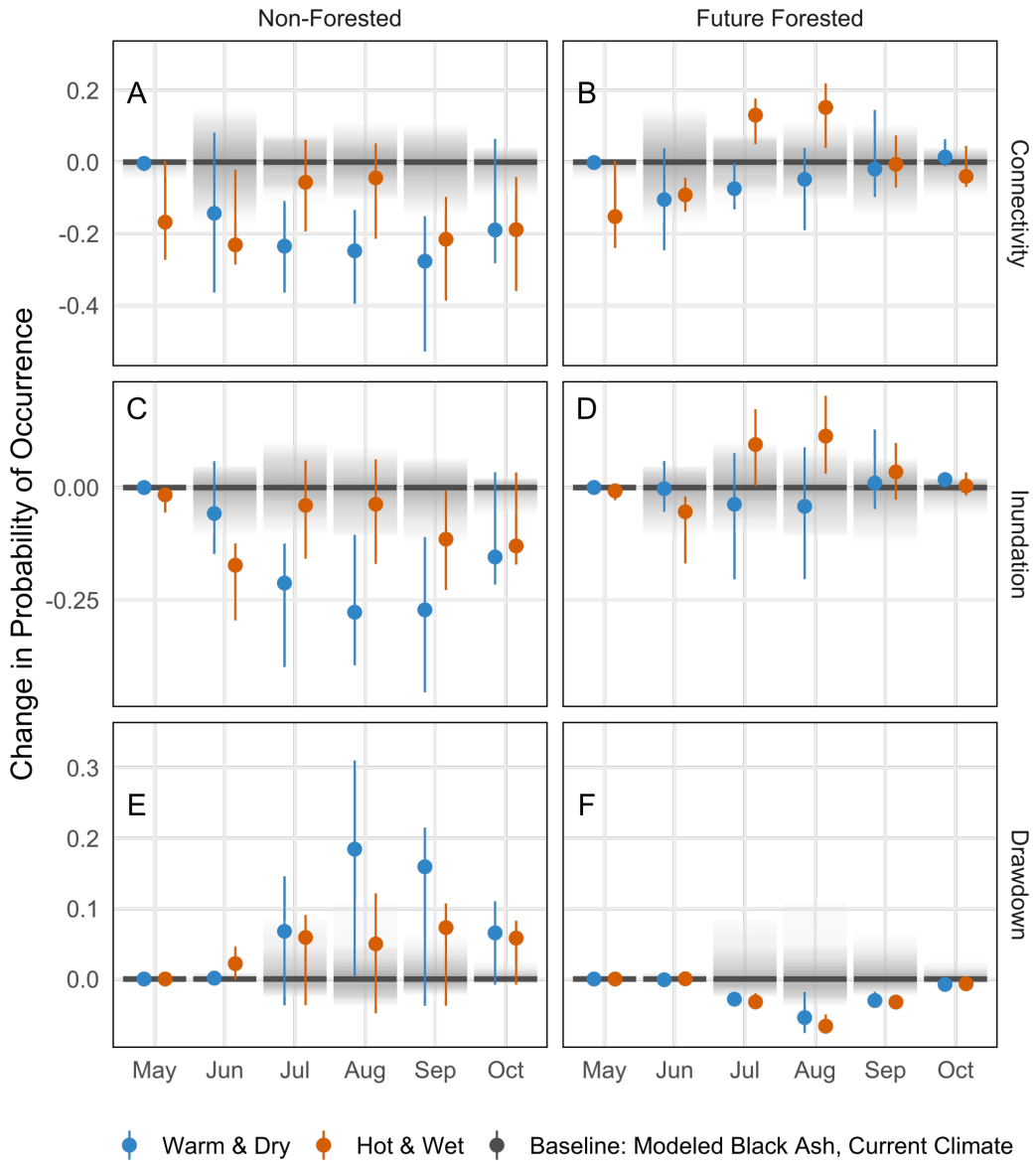


Figure 7. Change in probability of water levels above (connectivity, inundation) or below (drawdown) ecohydrologically significant thresholds under future forest and climate scenarios relative to baseline of black ash under a current climate. Probability of occurrence is the proportion of days in each month that simulated water levels reached exceeded each threshold. Values are reported as the median (point/bar) and the bounds of a 67% highest density continuous interval (an interval, potentially asymmetric, that contains 67% of simulated values). Blue and orange colors represent warm & dry and hot & wet future climate scenarios. Gray shaded area represents 67% HDCI of model simulations for control black ash conditions under the current climate (baseline condition) and the black crossbar represents the baseline monthly median probability. Some conditions and time periods showed very little variation (e.g. baseline conditions in May panels E and F) making shading and range bars indistinguishable.

4.5 Discussion

4.5.1 SWG

Our SWG performed well at simulating weather series that conformed to LOCA-downscaled future climate scenarios (Table 5, Figure 4). Precipitation, however, showed more inter-simulation variability than minimum or maximum temperature. This result is understandable when we consider that in the study region precipitation is only weakly seasonal, while minimum and maximum temperature have strong seasonal signals. There may be a slight trend in under-estimating the range of observed future climate conditions (Table 5). This trend is likely a result of reversion to the seasonal mean due to a lack of available external weather drivers. Under future climate scenarios extreme events are expected to be more commonplace (Peltier et al., 2018; Vavrus and Behnke, 2014) , and these results suggest event size could be better parameterized by increasing the probability of extreme events through selection of distribution choices and modeling framework (Verdin et al. 2019). This shortcoming of the SWGs should not affect our findings because our objective is not to model response to extreme events, but overall probability of *CEHT* water levels. These water levels will be affected by extreme precipitation events but the systems do not have infinite storage and more extreme events can be expected to pulse through the system more quickly (Vavrus and Behnke, 2014). The quicker the pulse passes through the system the smaller the impact on overall probability of wetland water levels exceeding *CEHTs.*

We chose to use an SWG approach that relied on only a single estimate of climate within the study area. Alternatively, regional variation in climate and weather patterns could be incorporated by using additional meteorological stations in the development of the SWG. A spatially-explicit SWG approach takes into account the covariance of regional meteorological stations and has been shown to faithfully capture larger regional trends in climate and weather (Verdin et al. 2015). Within our study region there are no clear weather or climate gradients and so we assumed that each study wetland was within a single climatic unit. This results in a much simpler SWG-construction process, but the resulting SWG cannot be used outside of the study region without first refitting on data from a different meteorological station(s). To simulate black ash wetland response on a regional basis would require addressing known temperature and precipitation gradients through the range of black ash. In addition to using only a single spatial representation of future climate, we chose to use individual years of simulation. The direct impact of this decision would miss long-term droughts and wet spells. We deliberately chose to use single-year simulations because of the length of available training data. During our study the region was in the process of transitioning from a relatively dry period to a relatively wetter period (Figure 8) Our training data are primarily drawn from the drier portion of the study period. Without longer-periods of training data we do not feel that we can capture both the intra- and inter-annual dynamics of wetland hydrology. If we assumed all 10,000 years of synthetic weather series were contiguous weather records, the longest period of drought years (annual precipitation less than annual PET) was 7 and 9 years for the warm & dry and hot & wet climate scenarios, respectively. By using single-year simulations we are implicitly assuming that fall and winter (SON and DJF) precipitation

is high enough to recharge the local hydrology driving these wetlands, restarting the cycle at or near maximum wetland water levels. When we consider only fall and winter precipitation we found 0%, 8.7%, and 10.1% of synthetic weather series had totals below the tenth percentile of observed data, warm & dry, and hot & wet climate scenarios, respectively. Unsurprisingly close to 10% of the synthetic weather winters were drier than the tenth percentile of the LOCA downscaled values. Synthetic (and LOCA) winters were wetter than observed winters, supporting the assumption that dormant-season precipitation would be enough to recharge wetland water levels.

Figure 8. Regional cumulative water availability for the period from November 1, 2005 through October 21, 2020. Water availability was calculated as the difference between rainfall plus melt and potential evapotranspiration.

4.5.2 Ecosystem Specific Yield

The approach used to calculate E_{Sy}, relating the magnitude of water level change to the drivers of that change, was an alternative to those presented in previous research. We developed this alternative based on unsuccessful attempts to implement E_{Sy} equations through the rainfall/rise method described in McLaughlin and Cohen (2014). Directly proving the theoretical basis of this approach is outside the scope of this study, and would be best performed with specifically-designed experimental or simulation work. Our research has however provided empirical evidence for the application of our approach. Section 3.3.5 in Chapter 3 demonstrates that *ET* estimates derived using this approach perform comparably to other implementations for calculating E_{Sy} (see White method details in the chapter mentioned above). Those results showed that calculated *ET* was correlated with daily *PET* estimates from nearby meteorological stations. To extend that comparison we examine how well a simple regression between daily water level change and adjusted or unadjusted hydrologic drivers (*PET*, *R*, *M*) perform. Figure 9 shows the results of a quantile regression of pooled data (present here for discussion purposes not used elsewhere in this study) explaining daily change in wetland water level as a function

of *PET, P* and *M*. We found that *pseudo-*R^2 of the model improved from 0.53 to 0.63. The difference between *pseudo-*R^2 are large enough to be of note, but more importantly we can see that water level drawdown is extremely underestimated by the unadjusted model (Figure 9A and 9B). Our results show that periods with the highest drawdown are not correctly modeled using the raw inputs. These periods of high drawdown are predicted when the E_{Sy} adjusted hydrologic inputs are used. Periods of high drawdown generally occurred mid-season when water levels were farther below the surface. Our approach agrees with others (McLaughlin and Cohen 2014; Watras et al. 2017) who found that the impact of E_{Sy} increases as water levels drawdown and that flooded conditions approach an E_{Sy} of 1. This relationship results in the largest adjustment to inputs during these mid-season periods, explaining the differences observed between panels A and B in Figure 9. These results together with those in Section 3.3.5 in Chapter 3 provide empirical evidence that this alternative formulation captures the variation of E_{Sy} with wetlands and can be used in scenarios where the rainfall-runoff ratio approach is masked by outside influences such as low-frequency hydrologic signals (Zhu et al. (2011)) and surface water connectivity (Watras et al. 2017).

4.5.3 Wetland Water Level Models

Wetland water level model performance was adequate for the objective of this study, which was to observe changes in the hydrographs of black ash wetlands and compare the probability of occurrence for *CEHT*s. Overall the retained models showed good correlation between observed and modeled water levels (Figure 5A, Table 6). We found that 91% of site-year combinations outside of site 119 had R^2 values above 0.6 for the model test period. RMedSE and relative RMedSE had similar patterns among sites (Figure 5). This suggests that from both absolute and relative standpoint, the wetland models as developed capture dynamic wetland water levels more accurately than static wetland water levels. Daily percent bias estimates (PBIAS) would be helpful to evaluate error relative to observed daily water level. However, wetland water levels cross 0 so that the relative error approaches infinity as water levels rise or fall towards the surface.

Importantly for our conclusions the wetland models performed well in predicting the probability of occurrence of *CEHT* levels. Table 7 and Figure 6 show how probability of occurrence of connectivity, inundation, and drawdown were not found to be different between the observed and modeled data in the test periods. They also show that there is no non-significant systematic model bias (positive or negative) in the probabilities. Although no systematic model bias was identified for *CEHT* analysis, Table 6 does indicate a bias towards drier simulations when daily water level measurements are compared. Therefore, the use a modeled control baseline is still an important safeguard against drawing conclusions based on potential from model artefacts. Comparing the systematically-biased results could exaggerate or mask real expectation of wetter or drier wetland conditions.

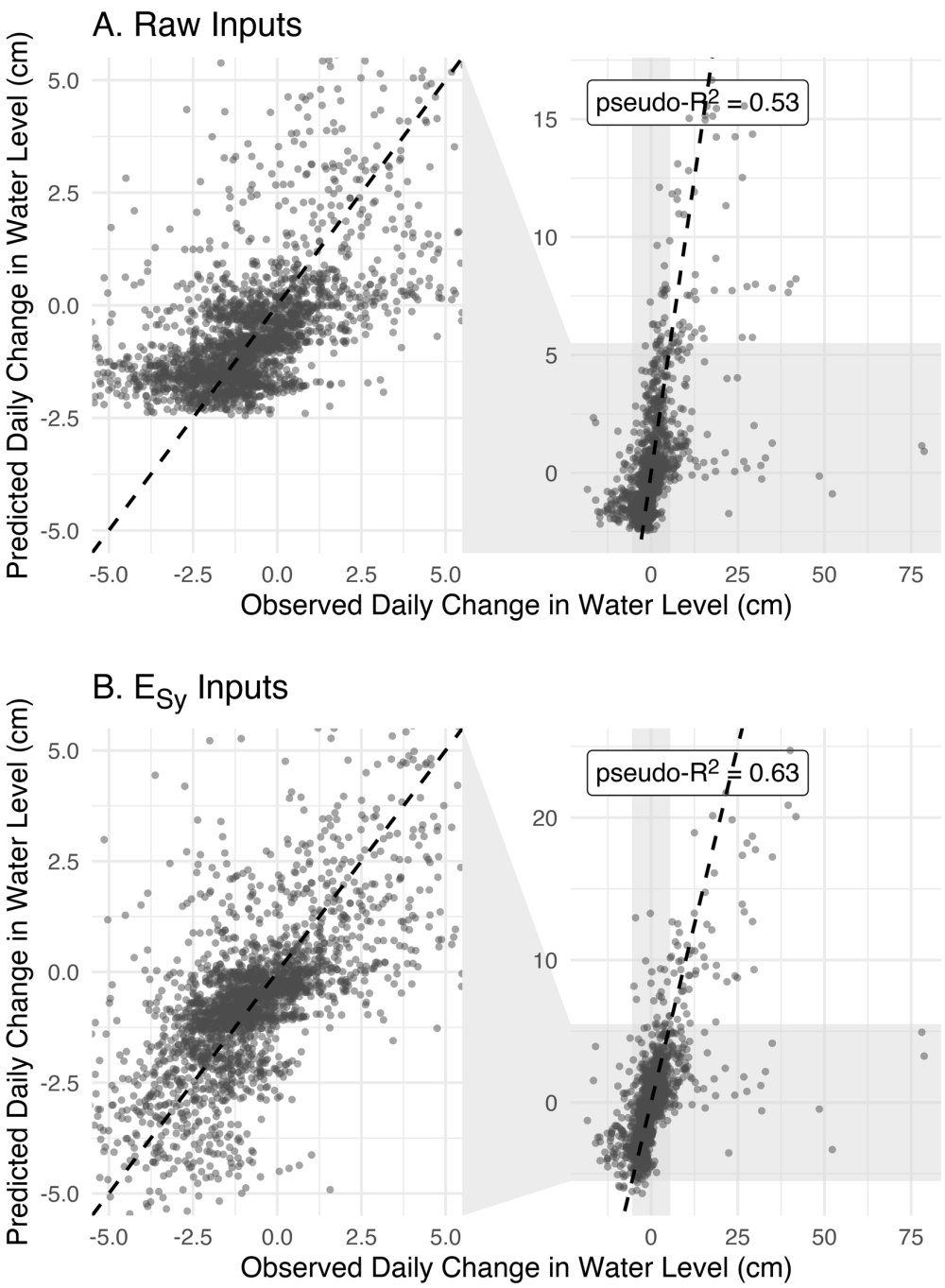


Figure 9. Predictions of daily change in water level using a quantile regression with observed daily water level as the dependent variable and *PET*, *rainfall*, and *snowmelt* as the independent variables. The model was fit two times, once with raw hydrologic inputs (A) and once with inputs adjusted by ecosystem specific yield (B). The right panels show the full dataset while the left panels zoom in on an interval [-5, 5] for both the observed and predicted water levels. The dashed line shows a 1:1 line.

The next step for these models is to fit them using a hierarchical Bayesian approach, which will allow propagation of the model uncertainty into the final predictions. This should improve model fit by incorporating information from all sites into a single model for each vegetation condition while allowing model parameters to vary between sites. This approach will also increase the available data for each vegetation condition when fitting the models as well as the overall available data. With three hierarchical models, the field-study control sites can be incorporated without resulting in unbalanced vegetation-conditions comparisons. If model performance does not significantly improve with that approach more, or potentially all years of data can be used to parameterize the wetland models, rather than the current single year of data for each period. Using most or all of the data for model parameterization precludes the options of evaluating model performance by cross-validation. Kozak and Kozak (2003) provide an alternative view that replaces cross-validation with lack-of-fit metrics, which evaluate model predictions for violation of model assumptions (e.g., non-random error). Their main conclusion was that cross-validation increased the variance of model estimates potentially resulting in poor model choice.

4.5.4 Future Hydrologic Conditions

The importance of these simulation results is to help relate potential future conditions to current conditions in dry, wet, or 'normal' years. This allows researchers and managers to answer questions similar to: What will these sites look like in a 'normal' year in the future? How will these sites respond to wet/dry conditions in the future? For example, we could expect that during a wet year under the hot & wet future climate scenario a non-forested wetland on these sites would have a similar probability of surface inundation as a black ash wetland does today (Figure 7C). In general, black ash wetlands that remain forested under potentially less inundation-adapted species can be expected to have wetter conditions under hot & wet future climate scenarios, and under warm & dry future climate scenarios are more likely to experience conditions similar to today. Conclusions could be reached for each combination of climate and vegetation scenario and this information can and should be used to influence management approaches and timing in responding to EAB in black ash wetlands.

The combined impact of EAB and climate is shown as the difference between future climate conditions (blue and orange) and black ash conditions under historical climate (shaded gray) in Figure 7. We can draw additional information about the individual effects of future climate and vegetative cover on wetlands by altering how we calculate the baseline. In Figure 7 the baseline is set as simulated black ash forests under current climate conditions. To isolate the effects of each climate scenario of wetland hydrology we can compare the response of a vegetative cover under future climate conditions against its response under current climate conditions. Conversely, we can evaluate the impact of vegetative cover by comparing each vegetative cover to simulated black ash forests under current climate conditions. Figures 10 and 11 show the probability of occurrence of each *CEHT* under conditions where the vegetative cover and climate, respectively, are held constant. Figures 11 & 10 present other comparisons using the same symbology structure as Figure 7. Table 8 presents the change in probability of

occurrence for each critical water level and alternative vegetative condition that can be attributed to EAB or a warm & dry or hot & wet sensitive future climate scenario.

We found that EAB impacts lead to slightly wetter conditions under non-forested conditions and the present climate. This agrees with observed water level response analysis in Van Grinsven et al. (2017)), where absolute water level response to simulated post-EAB conditions resulted in wetter conditions and significantly lower growing season drawdown rates. A modeled alternate forest composition (Future Forested) shows that EAB impact alone would lead to dramatically wetter conditions relative to black ash forests. This result is expected as we modeled our future forested conditions based on previous work showing that co-dominant hardwoods had significantly lower seasonal transpiration estimates than black ash (Shannon et al. (2017)). Under both alternative vegetative conditions, we found that the isolated impact of future climate scenarios would lead to much drier sites. The probability of surface water connectivity or inundation occurring are dramatically decreased under both climate scenarios. The probability of wetland water levels dropping below 50 cm was increased under future climate scenarios and non-forested conditions, but much lower probability of occurring under future-forested conditions. The probability of each *CEHT* is not symmetrical around the median, indicating for instance that the impact of EAB on Connectivity is more likely to be wetter conditions than looking at the median alone would suggest (Figures 6 and 11).

Our hypothesis that the impacts of EAB and climate change would counteract each other is partially supported by our results. The dynamics of wetland water levels are complex and climate change and invasive species can each have major consequences (Burkett and Kusler 2000; Moomaw et al. 2018). We can summarize our findings by saying we expect wetlands that transition from black ash forests to non-forested sites to become drier, primarily due to the effects of increased summer evaporative demand (Figures 7, 12 & 13). Our simulations saw less summer precipitation under the warm & dry scenario, while under the hot & wet scenario we observed an increase in large storms (Figure 13). If management or natural regeneration leads to establishment of future forests with lower site transpiration rates, these wetlands can be expected to have the occurrence of high water conditions remain stable or increase in frequency, while dry conditions (water levels more than 50 cm below the surface) become more rare (Figures 7, 12 & 13). Our results show that the impact of climate change in this region will lead to consistently drier wetland conditions. Changes to vegetative cover can amplify or counteract hydrologic alterations attributable to climate change.

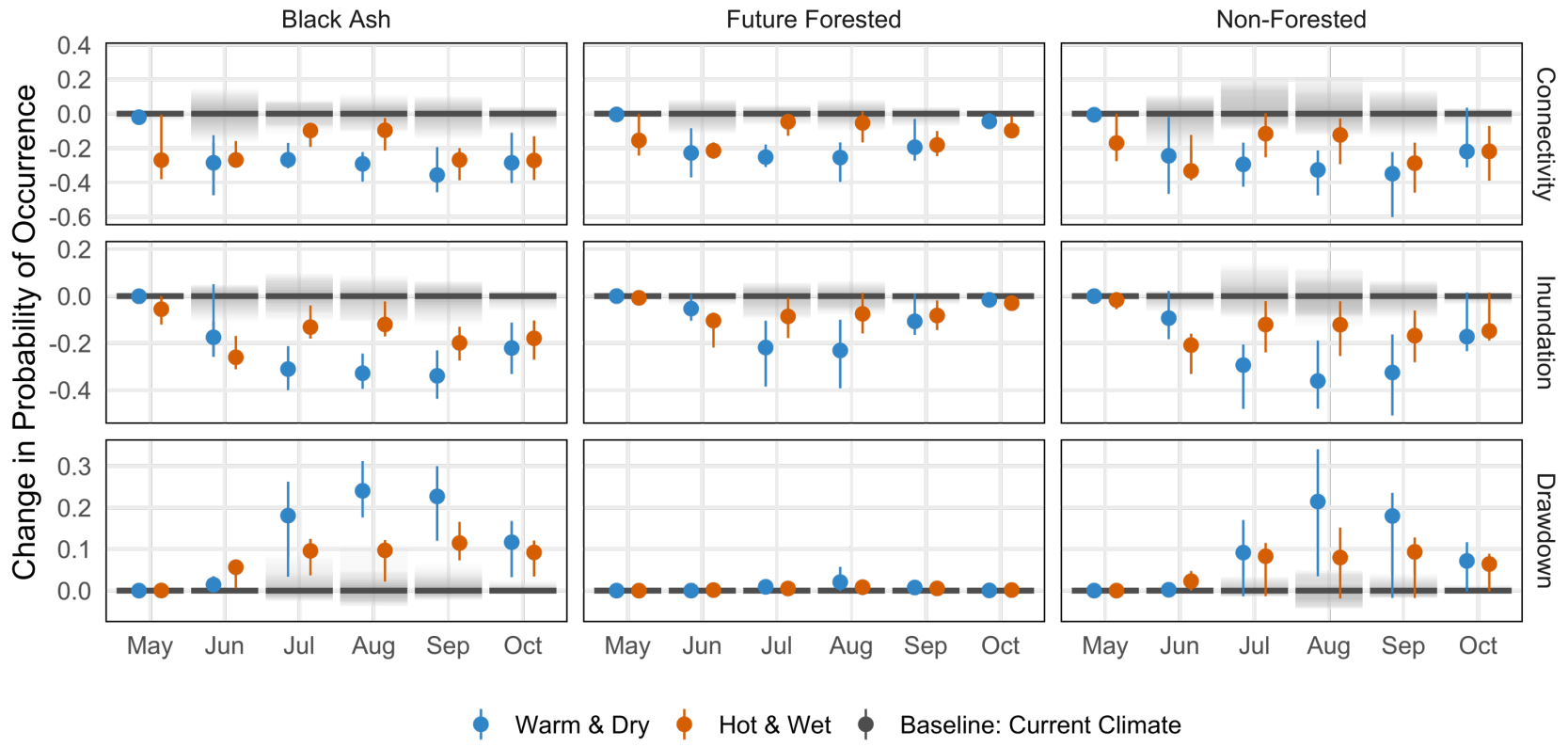


Figure 10. Change in probability of occurrence for critical ecohydrological wetland water levels under combinations of vegetative cover and climate conditions relative to the respective vegetative cover under current climate baseline. Each column compares simulated wetland response within a fixed vegetation condition across multiple climate scenarios. Values are reported as the median (point/bar) and the bounds of a 67% highest density continuous interval (an interval, potentially asymmetric, that contains 67% of simulated values).

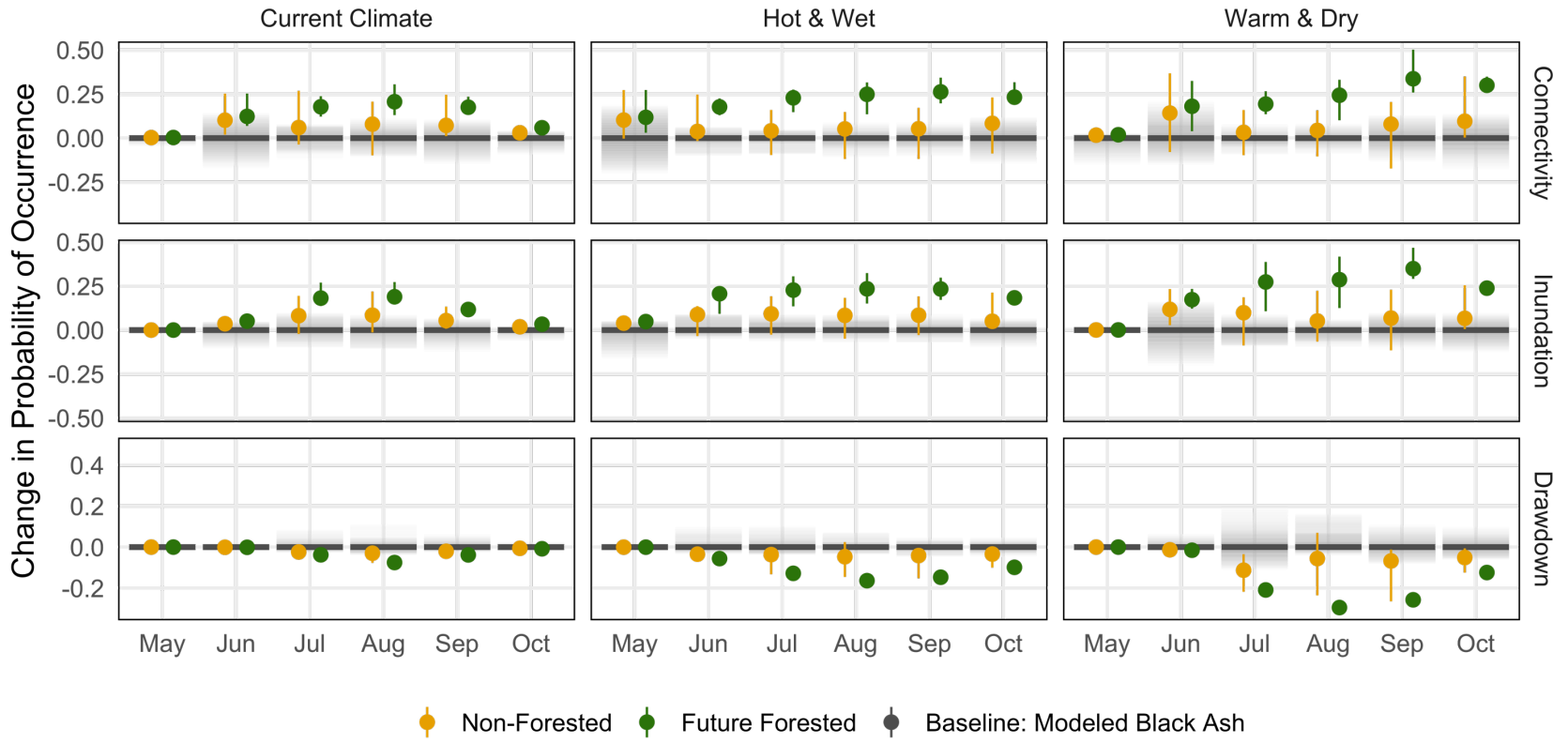


Figure 11. Change in probability of occurrence for critical ecohydrological wetland water levels under combinations of vegetative cover and climate conditions relative to the respective climate scenario under black ash cover. Each column compares simulated wetland response within a fixed climate scenario across multiple potential vegetative covers. Differences within a column represent the impact of EAB and potential management decisions. Values are reported as the median (point/bar) and the bounds of a 67% highest density continuous interval (an interval, potentially asymmetric, that contains 67% of simulated values).

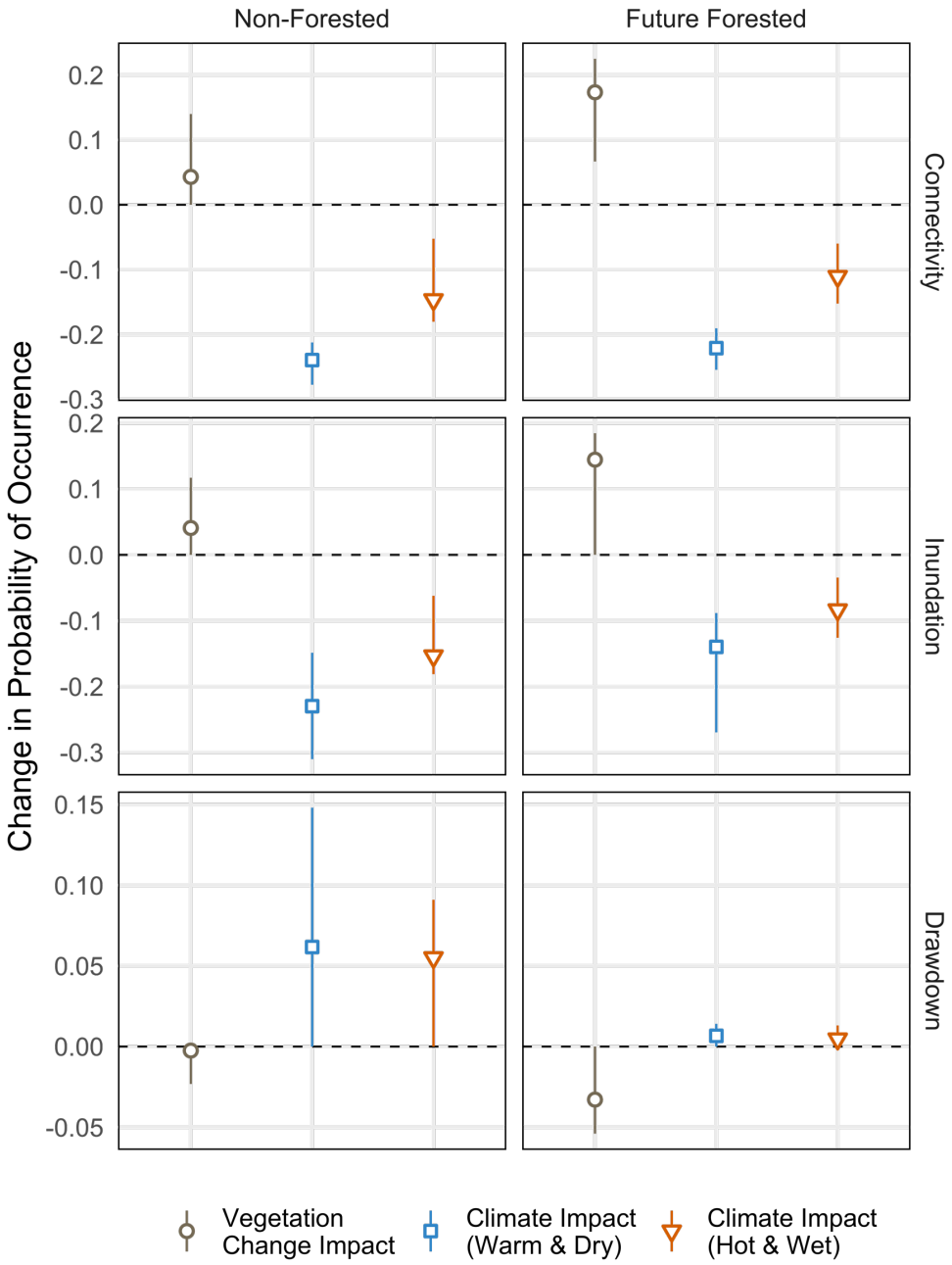


Figure 12. Expected change in probability of occurrence for critical ecohydrological water levels that can be attributed to future climate scenarios or emerald ash borer (EAB) impact and management response to EAB. Vegetation impact values represent the difference between simulated black ash forests and each alternative vegetative condition under current climate conditions. Climate impacts represent the difference between each alternative vegetative condition under the current climate and future climate (warm & dry or hot & wet) scenarios. Differences are calculated using the daily values from July through September, highlighting growing season impacts. Values are reported as the median (point) and the bounds of a 67% highest density continuous interval (an interval, potentially asymmetric, that contains 67% of simulated values).

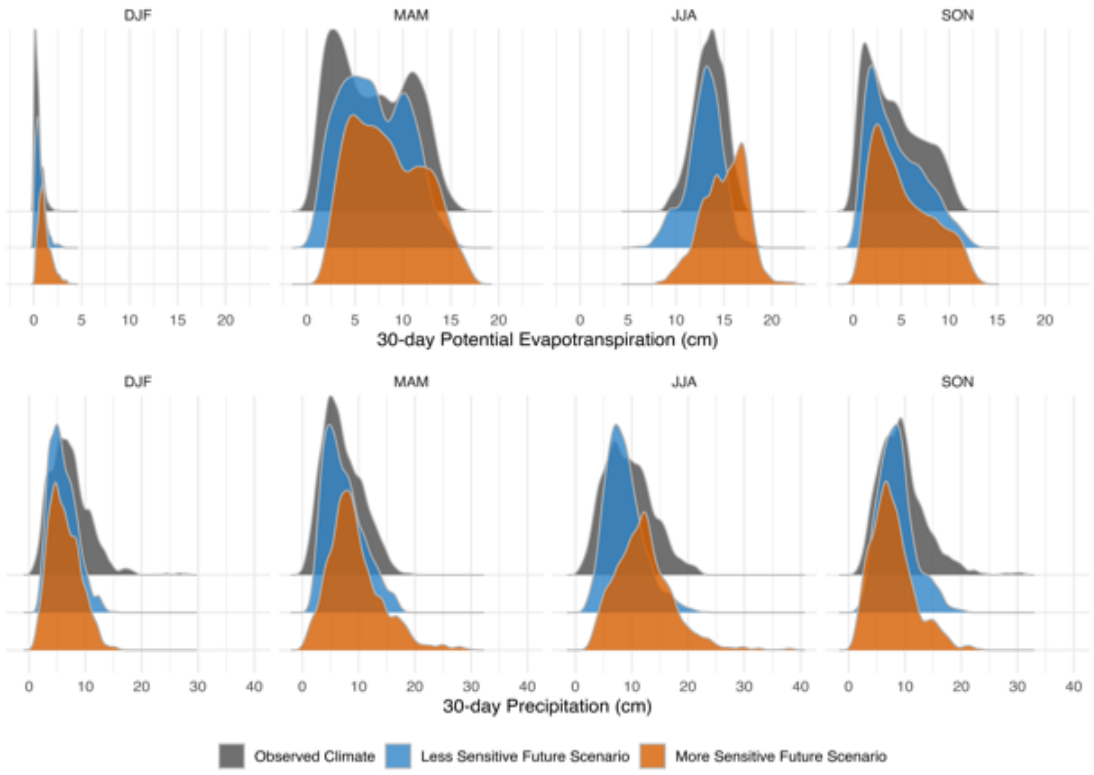


Figure 13. Distribution of 30-day total precipitation and potential evapotranspiration presented by climate season and climate scenario. All data present is for the Bergland Dam, MI, USA GHCND station. Observed climate is from 1980-2009, future climate is from LOCA-downscaled climate scenarios for the period from 2070-2099. Potential evapotranspiration was calculated from observed and GCM outputs using the Hargreaves-Samani equation (Hargreaves and Allen 2003).

4.5.5 Drivers of Future Wetland Hydrologic Conditions

4.5.5.1 Future Climate

The future climate scenarios used in this study present two alternative future scenarios. The warm & dry scenario summer is projected to be slightly drier than observed conditions, with similar *PET* (Table 8). When summer water deficit is calculated as the *P-PET* we see that with a given year conditions will show a decrease in water availability. The hot & wet scenario is a much wetter projection than both the observed climate and future climate (Table 8). There is little to no overlap in the 67% HDCIs for summer precipitation between the hot & wet and other scenarios. However, a commensurate increase in summer evaporative demand may lead to only slightly wetter summer conditions under the hot & wet scenario. Spring and early summer water levels are currently strongly influenced by the snowfall and melt regimes (Van Grinsven et al. 2017). Our simulations show that snowmelt was reduced in quantity and shifted earlier in the year for both future scenarios, with more dramatic impacts under the hot & wet

scenario. This shift will likely have the result of increasing water deficit on the landscape relative to the values presented in Table 8.

Table 8. Summer (JJA) precipitation, potential evapotranspiration, and water deficit (P - PET) for the climate scenarios in this study. Future climate data were taken from LOCA downscaled daily values, and observed climate was taken from observed conditions at Bergland Dam, MI, USA from 1980-2009. Reported values are the median and bounds of the 67% highest density continuous interval of seasonal total precipitation by observed or projection year.

Climate	Precipitation (cm)	Potential Evapotranspiration (cm)	Water Deficit (cm)
Observed Climate	29.93 (22.20, 32.75)	41.24 (38.20, 43.13)	-11.12 (-19.88, -6.95)
Warm & dry Future Scenario	26.26 (21.86, 30.73)	40.14 (37.93, 42.07)	-13.23 (-21.38, -10.99)
Hot & wet Future Scenario	36.09 (29.82, 40.46)	46.60 (42.56, 48.23)	-10.85 (-18.18, -2.78)

4.5.5.2 Interaction of AET and E_{Sy}

It is expected that vegetation conditions with lower transpiration rates have lower rates of water level drawdown and therefore increased water levels. In these wetlands this effect is compounded by the dynamics of E_{Sy}. E_{Sy} increases in magnitude as water levels decline (Figure 2), which may be a result of reduced soil pore space and wetland cross-sectional area so that a smaller volume of change results in a large water level difference (McLaughlin and Cohen 2014). The effect is that *PET* results in smaller changes to wetland water levels under wet conditions than dry conditions. In this way the impact of EAB may result in a feedback loop in which water levels remain elevated:

1. *AET* begins to draw down water levels, E_{Sy} is low,
2. Water levels decline slowly because of reduced *AET* and low E_{Sy},
3. As *PET* increases to mid-season peak, E_{Sy} and *AET* are both lower than under black ash,
4. High *PET* impact on water levels are reduced by lower *AET* and sustained low E_{Sy}

Under the current black ash conditions E_{Sy} increases more rapidly due to higher black ash *AET*. The increase in E_{Sy} accelerates the impact of higher *AET*, creating faster and larger declines in wetland water level throughout the growing season (Figure 14). This feedback loop may partially explain the persistence of hydrologic impact following EAB disturbance observed in Michigan and Minnesota (Diamond et al. 2018; Kolka et al. 2018). In effect, the impact of EAB may shock these systems into an alternative stable state of elevated water levels (Scheffer and Carpenter 2003). Our results indicate that

future climate scenarios will likely have a large enough impact to again shock these systems out of a stable state. We cannot say from these simulations whether the wetlands will reach a new stable state under future climate conditions or what that state would be.

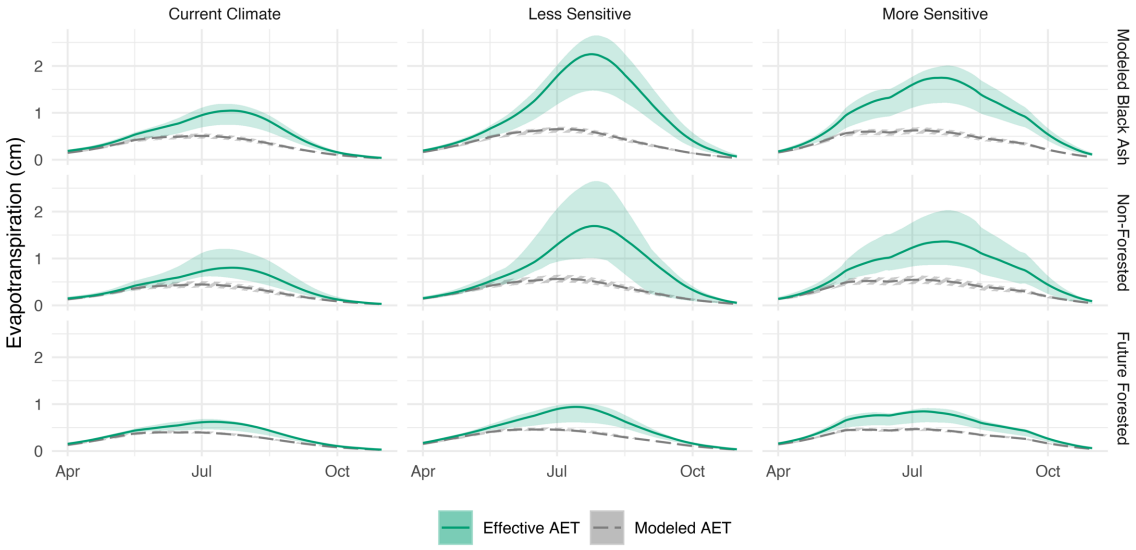


Figure 14. Modeled and effective actual evapotranspiration under various climate scenarios and vegetation conditions. Modeled AET is the synthetic PET multiplied by the model coefficient used in the relevant wetland model. Effective AET is the modeled AET times the value of E_{Sy} predicted from the contemporaneous water level. The difference between the two values demonstrates the two-stage impact of reduced ET, where AET and E_{Sy} are reduced.

4.5.5.3 Reduced Evaporation and Non-Canopy Transpiration

We observed wetter conditions under our Future Forest simulations relative to both Black Ash and Non-Forested simulations under all climate scenarios (Figures 6, 11). Two factors are likely contributing to this result. The first is that our future forest composition is known to have lower transpiration rates than the existing black ash canopy (Shannon et al. 2018). As a model choice we opted to assume that the next most dominant canopy species would be the likely replacement canopy species. The future forest we modeled is a narrow range of potential future forest compositions. Natural regeneration or planting efforts may lead to future forested species composition that more closely matches the evaporative potential of the current black ash canopy. Secondly, the future forested conditions may have a lower total *AET* than the non-forested conditions. In black ash wetlands in Minnesota, simulated post-EAB conditions (girdled and standing ash) were found to have higher water levels than site where the ash stems were harvested and removed (Diamond et al. 2018). The authors attributed the result to limited *AET* leading to reduced solar energy and wind-driven boundary layer mixing due to the still standing stems. Our conditions have a notable difference from that study in that we are assuming there are living trees on site. However, a closed canopy would have the same effect of reducing open water evaporation and understory transpiration. With sufficiently reduced

canopy transpiration the effect of certain forest compositions could result in reduced overall *AET*.

The contrast between the Non-Forested and Future-Forested simulated hydrology suggests an important management tool. Iverson et al (2016) and the work from Bolton (2018) and Looney (2015) laid out a framework and results for evaluating potential replacement species considering site conditions of black ash wetlands. The results presented here show the opportunity for management decisions that take into account the impact of future vegetation on site hydrologic conditions. The general trend of drier future conditions can be to some degree counteracted by management for cover with lower evapotranspiration rates. Drier conditions are not always preferable across the landscape and this tactic could be used to retain water on the landscape, creating refugia of standing water or cool moist soils for flora and fauna. Although much of the region is expected to have drier summers our localized study area is expected to have wetter summers. Our results can still be applied to the broader region. The impact of climate change in other areas of the region would become even more pronounced. Vegetation conditions would still amplify or counteract climate impacts, and in some areas may become even more important to providing wet or moist refugia on the landscape.

4.6 Future Research

It should be stated that these models and simulations cannot capture the full range of potential changes from EAB, other potential invasive species, and climate change. The following is a non-exhaustive list of processes and assumptions that this work could not address and warrant further study.

- Watras (2014) found that there are decadal scale oscillations of water levels within the Great Lakes and inland lakes in the Great Lakes Region. These can be expected to impact wetland water levels similarly.
- While our results showed a smaller and earlier snowmelt pulses we did not attempt to model the more complex dynamics of rain-on-snow events and a compressed melting period reducing recharge to the intermediate groundwater sources shown to feed these wetlands (Van Grinsven et al. 2017).
- Our single-year simulations do not account for the potential effects of multi-year drought or for periods of significantly reduced fall/winter precipitation on wetland water level rebound.
- The response of each species and plant community will show individual climate change responses that are not captured in these models and may be non-linear (Short 2016).
- Beyond the potential for unforeseen vegetation responses to future climates is the uncertainty of future climate conditions. The authors of the fourth national climate assessment have highlighted that early climate prediction have under-predicted contemporary shifts in response to climate change. Unfortunately, this means that our models built on those simulations may be underestimating the magnitude of future changes.

- Recently, EAB was discovered near the wetlands used to develop water level models in this study. It is important to study this natural infestation in black ash wetlands and compare it to these and previous results.

Wetlands provide a range of ecological services from water quality and stormflow retention to carbon storage, all of which stand to be impacted by climate change (Moomaw et al. 2018). We did not attempt to quantify how changes in wetland hydrology will influence the other services. By highlighting the magnitude and drivers of hydrologic change in these systems we hope to spur future research.

4.7 Conclusions

Our research has shown that changes in evaporative demand and precipitation regimes will likely result in drier conditions on what are now black ash wetlands. In addition, the functional loss of black ash due to EAB presents challenges and opportunities for the future of these wetlands. A large extent of forested wetlands may require intervention to retain desired benefits or features they currently provide. These interventions can be used to drive sites towards wetter conditions, retaining water on the landscape that may otherwise be lost under a changing climate.

To implement long-term management objectives in wetland and forest management requires a deep understanding of the systems. We have shown that as water levels decline, the impact of each additional driver increases due to changes in E_{Sy}. Changes in *AET* can interact with E_{Sy} to create feedback loops underscoring that knowledge of species adaptation to wet conditions and capacity to respond quickly to drying conditions is critical for projected future site conditions. The transition from black ash to alternative vegetative cover can counteract or amplify the impacts of future conditions with higher evaporative demand. The relationship between evaporative demand and E_{Sy} can amplify the effect of wetland water level drawdown.

4.8 References

Arguez, Anthony, Imke Durre, Scott Applequist, Russell S. Vose, Michael F. Squires, Xungang Yin, Richard R. Heim, and Timothy W. Owen. 2012. “NOAA’s 1981 U.S. Climate Normals: An Overview.” *Bulletin of the American Meteorological Society* 93 (11): 1687–97. https://doi.org/10.1175/BAMS-D-11-00197.1.

Baddour, Omar, and Hama Kontongomde, eds. 2007. “THE ROLE OF CLIMATOLOGICAL NORMALS IN A CHANGING CLIMATE.” WCDMP-No. 61, WMO-TD No. 1377. Geneva: World Meteorological Organization.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bojanowski, Jedrzej S. 2016. “Sirad: Functions for Calculating Daily Solar Radiation and Evapotranspiration.” https://CRAN.R-project.org/package=sirad.

Bolton, Nicholas, Joseph Shannon, Joshua Davis, Matthew Grinsven, Nam Noh, Shon Schooler, Randall Kolka, Thomas Pypker, and Joseph Wagenbrenner. 2018. “Methods to Improve Survival and Growth of Planted Alternative Species Seedlings in Black Ash Ecosystems Threatened by Emerald Ash Borer.” *Forests* 9 (3): 146. https://doi.org/10.3390/f9030146.

Brinson, Mark M. 1993. “A Hydrogeomorphic Classification for Wetlands.” *Wetlands Research Program Technical Report WRP-DE-4* WRP-DE-4 (August): 101. https://doi.org/10.2134/agronj2001.931131x.

Bristow, Keith L., and Gaylon S. Campbell. 1984. “On the Relationship Between Incoming Solar Radiation and Daily Maximum and Minimum Temperature.” *Agricultural and Forest Meteorology* 31 (2): 159–66. https://doi.org/10.1016/0168-1923(84)90017-0.

Burkett, Virginia, and Jon Kusler. 2000. “CLIMATE CHANGE: POTENTIAL IMPACTS AND INTERACTIONS IN WETLANDS OF THE UNITED STATES [1].” *JAWRA Journal of the American Water Resources Association* 36 (2): 313–20. https://doi.org/10.1111/j.1752-1688.2000.tb04270.x.

Bürkner, Paul-Christian. 2017. “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

———. 2018. “Advanced Bayesian Multilevel Modeling with the R Package brms.” *The R Journal* 10 (1): 395–411. https://doi.org/10.32614/RJ-2018-017.

Byun, Kyuhyun, and Alan F. Hamlet. 2018. “Projected Changes in Future Climate over the Midwest and Great Lakes Region Using Downscaled Cmip5 Ensembles: PROJECTED CLIMATE CHANGES OVER THE MIDWEST AND GREAT LAKES REGION.” *International Journal of Climatology* 38 (April): e531–53. https://doi.org/10.1002/joc.5388.

Coron, L., G. Thirel, O. Delaigue, C. Perrin, and V. Andréassian. 2017. “The Suite of Lumped GR Hydrological Models in an R Package.” *Environmental Modelling & Software* 94 (August): 166–71. https://doi.org/10.1016/j.envsoft.2017.05.002.

Davis, Joshua C., Joseph P. Shannon, Nicholas W. Bolton, Randall K. Kolka, and Thomas G. Pypker. 2017. “Vegetation Responses to Simulated Emerald Ash Borer Infestation in *Fraxinus Nigra* Dominated Wetlands of Upper Michigan, USA.” *Canadian Journal of Forest Research* 47 (3): 319–30. https://doi.org/10.1139/cjfr-2016-0105.

Diamond, Jacob S., Daniel L. McLaughlin, Robert A. Slesak, Anthony W. D’Amato, and Brian J. Palik. 2018. “Forested Versus Herbaceous Wetlands: Can Management Mitigate Ecohydrologic Regime Shifts from Invasive Emerald Ash Borer?” *Journal of Environmental Management* 222 (September): 436–46. https://doi.org/10.1016/j.jenvman.2018.05.082.

Gent, Peter R., Gokhan Danabasoglu, Leo J. Donner, Marika M. Holland, Elizabeth C. Hunke, Steve R. Jayne, David M. Lawrence, et al. 2011. “The Community Climate System Model Version 4.” *Journal of Climate* 24 (19): 4973–91. https://doi.org/10.1175/2011JCLI4083.1.

Gregory, Jonathan M, TML Wigley, and PD Jones. 1993. “Application of Markov Models to Area-Average Daily Precipitation Series and Interannual Variability in Seasonal Totals.” *Climate Dynamics* 8 (6): 299–310.

Griffies, Stephen M., Michael Winton, Leo J. Donner, Larry W. Horowitz, Stephanie M. Downes, Riccardo Farneti, Anand Gnanadesikan, et al. 2011. “The GFDL Cm3 Coupled Climate Model: Characteristics of the Ocean and Sea Ice Simulations.” *Journal of Climate* 24 (13): 3520–44. https://doi.org/10.1175/2011JCLI3964.1.

Haack, Robert A, Eduard Jendek, Houping Liu, Kenneth R Marchant, Toby R Petrice, Therese M Poland, Hui Ye, and East Lansing. 2002. “The Emerald Ash Borer: A New Exotic Pest in North America,” 5.

Hargreaves, George H., and Richard G. Allen. 2003. “History and Evaluation of Hargreaves Evapotranspiration Equation.” *Journal of Irrigation and Drainage Engineering* 129 (1): 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53).

Hayhoe, Katharine, Jeff VanDorn, Thomas Croley, Nicole Schlegal, and Donald Wuebbles. 2010. “Regional Climate Change Projections for Chicago and the US Great Lakes.” *Journal of Great Lakes Research* 36 (January): 7–21. https://doi.org/10.1016/j.jglr.2010.03.012.

Hegewisch, K. C., J. T. Abatzoglou, O. Chegwidden, and B. Nijissen. 2021. “’Climate Mapper’ Web Tool.” Climate Toolbox. https://climatetoolbox.org.

Herms, Daniel A., and Deborah G. McCullough. 2014. “Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management.” *Annual Review of Entomology* 59 (1): 13–30. https://doi.org/10.1146/annurev-ento-011613-162051.

Horel, John, Michael Splitt, L Dunn, J Pechmann, B White, C Ciliberti, S Lazarus, J Slemmer, D Zaff, and J Burks. 2002. “Mesowest: Cooperative Mesonets in the Western United States.” *Bulletin of the American Meteorological Society* 83 (2): 211–26.

“Hourly Precipitation Data (HPD) Network, Version 2.R2.” 2021. NOAA National Centers for Environmental Information.

Iverson, Louis, Kathleen S. Knight, Anantha Prasad, Daniel A. Herms, Stephen Matthews, Matthew Peters, Annemarie Smith, Diane M. Hartzler, Robert Long, and John Almendinger. 2016. “Potential Species Replacements for Black Ash (Fraxinus Nigra) at the Confluence of Two Threats: Emerald Ash Borer and a Changing Climate.” *Ecosystems* 19 (2): 248–70. https://doi.org/10.1007/s10021-015-9929-y.

Kolka, Randall, Anthony D'Amato, Joseph Wagenbrenner, Robert Slesak, Thomas Pypker, Melissa Youngquist, Alexis Grinde, and Brian Palik. 2018. "Review of Ecosystem Level Impacts of Emerald Ash Borer on Black Ash Wetlands: What Does the Future Hold?" *Forests* 9 (4): 179. https://doi.org/10.3390/f9040179.

Kozak, Antal, and Robert Kozak. 2003. "Does Cross Validation Provide Additional Information in the Evaluation of Regression Models?" 33: 12.

Krause, P., D. P. Boyle, and F. Bäse. 2005. "Comparison of Different Efficiency Criteria for Hydrological Model Assessment." *Advances in Geosciences* 5 (December): 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

Lenth, Russell V. 2021. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. https://github.com/rvlenth/emmeans.

Loheide II, Steven P. 2008. "A Method for Estimating Subdaily Evapotranspiration of Shallow Groundwater Using Diurnal Water Table Fluctuations." *Ecohydrology* 1 (1): 59–66. https://doi.org/10.1002/eco.7.

Looney, Christopher E., Anthony W. D'Amato, Shawn Fraver, Brian J. Palik, and Michael R. Reinikainen. 2016. "Examining the Influences of Tree-to-Tree Competition and Climate on Size-Growth Relationships in Hydric, Multi-Aged Fraxinus Nigra Stands." *Forest Ecology and Management* 375 (September): 238–48. https://doi.org/10.1016/j.foreco.2016.05.050.

Looney, Christopher E., Anthony W. D'Amato, Brian J. Palik, and Robert A. Slesak. 2015. "Overstory Treatment and Planting Season Affect Survival of Replacement Tree Species in Emerald Ash Borer Threatened *Fraxinus Nigra* Forests in Minnesota, USA." *Canadian Journal of Forest Research* 45 (12): 1728–38. https://doi.org/10.1139/cjfr-2015-0129.

Looney, Christopher E., Anthony W. D'Amato, Brian J. Palik, Robert A. Slesak, and Mitchell A. Slater. 2017. "The Response of Fraxinus Nigra Forest Ground-Layer Vegetation to Emulated Emerald Ash Borer Mortality and Management Strategies in Northern Minnesota, USA." *Forest Ecology and Management* 389 (April): 352–63. https://doi.org/10.1016/j.foreco.2016.12.028.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC.

McLaughlin, Daniel L., and Matthew J. Cohen. 2014. "Ecosystem Specific Yield for Estimating Evapotranspiration and Groundwater Exchange from Diel Surface Water Variation: ECOSYSTEM S $_{Y}$ FOR ESTIMATING ET AND GROUNDWATER EXCHANGE." *Hydrological Processes* 28 (3): 1495–1506. https://doi.org/10.1002/hyp.9672.

McLaughlin, Daniel L., Jacob S. Diamond, Carlos Quintero, James Heffernan, and Matthew J. Cohen. 2019. “Wetland Connectivity Thresholds and Flow Dynamics From Stage Measurements.” *Water Resources Research* 55 (7): 6018–32. https://doi.org/10.1029/2018WR024652.

Menne, Matthew J., Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, et al. 2012. “Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.28-Upd-2021031419.” http://doi.org/10.7289/V5D21VHZ.

Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. 2012. “An Overview of the Global Historical Climatology Network-Daily Database.” *Journal of Atmospheric and Oceanic Technology* 29 (7): 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1.

Moomaw, William R., G. L. Chmura, Gillian T. Davies, C. M. Finlayson, B. A. Middleton, Susan M. Natali, J. E. Perry, N. Roulet, and Ariana E. Sutton-Grier. 2018. “Wetlands In a Changing Climate: Science, Policy and Management.” *Wetlands* 38 (2): 183–205. https://doi.org/10.1007/s13157-018-1023-8.

Moriasi, D. N., M. W. Gitau, N. Pai, and P. Duggaupati. 2015. “Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria.” *Transactions of the ASABE* 58 (6): 1763–85. https://doi.org/10.13031/trans.58.10715.

Notaro, Michael, Val Bennington, and Steve Vavrus. 2015. “Dynamically Downscaled Projections of Lake-Effect Snow in the Great Lakes Basin.” *Journal of Climate* 28 (4): 1661–84. https://doi.org/10.1175/JCLI-D-14-00467.1.

Peltier, W.R., d’Orgeville, M., Erler, A.R., Xie, F., 2018. Uncertainty in Future Summer Precipitation in the Laurentian Great Lakes Basin: Dynamical Downscaling and the Influence of Continental-Scale Processes on Regional Climate Change. Journal of Climate 31, 2651–2673. https://doi.org/10.1175/JCLI-D-17-0416.1

Pierce, David W., Daniel R. Cayan, and Bridget L. Thrasher. 2014. “Statistical Downscaling Using Localized Constructed Analogs (LOCA).” *Journal of Hydrometeorology* 15 (6): 2558–85. https://doi.org/10.1175/JHM-D-14-0082.1.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, C. W. 1981. “Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation.” *Water Resources Research* 17 (1): 182–90. https://doi.org/10.1029/WR017i001p00182.

Rood, Richard, and Laura Briley. 2018. “Great Lakes Ensemble April 2017 - April 2018 Progress Report.” Great Lakes Integrated Sciences + Assessments. http://glisa.umich.edu/media/files/Ensemble_Progress_Report_April_2018.pdf.

Scheffer, Marten, and Stephen R. Carpenter. 2003. “Catastrophic Regime Shifts in Ecosystems: Linking Theory to Observation.” *Trends in Ecology & Evolution* 18 (12): 648–56. https://doi.org/10.1016/j.tree.2003.09.002.

Shannon, Joseph, Joshua Davis, Matthew Van Grinsven, Nicholas Bolton, Nam Jin Noh, Thomas Pypker, Randall Kolka, and Joseph Wagenbrenner. 2017. “Water Level Controls on Transpiration of Co-Dominant Species in Black Ash Wetlands.” Poster Presentation. Duluth, MN, USA.

Shannon, Joseph, Matthew Van Grinsven, Joshua Davis, Nicholas Bolton, Nam Noh, Thomas Pypker, and Randall Kolka. 2018. “Water Level Controls on Sap Flux of Canopy Species in Black Ash Wetlands.” *Forests* 9 (3): 147. https://doi.org/10.3390/f9030147.

Short, Frederick T. 2016. “Impacts of Climate Change on Submerged and Emergent Wetland Plants.” *Aquatic Botany*, 15.

Slesak, Robert A., Christian F. Lenhart, Kenneth N. Brooks, Anthony W. D’Amato, and Brian J. Palik. 2014. “Water Table Response to Harvesting and Simulated Emerald Ash Borer Mortality in Black Ash Wetlands in Minnesota, USA.” *Canadian Journal of Forest Research* 44 (8): 961–68. https://doi.org/10.1139/cjfr-2014-0111.

Swanston, Christopher W., Maria K. Janowiak, Leslie A. Brandt, Patricia R. Butler, Stephen D. Handler, P. Danielle Shannon, Abigail Derby Lewis, et al. 2016. “Forest Adaptation Resources: Climate Change Tools and Approaches for Land Managers. 2nd Ed.” NRS-GTR-87-2. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-GTR-87-2.

Valéry, Audrey, Vazken Andréassian, and Charles Perrin. 2014. “‘As Simple as Possible but Not Simpler’: What Is Useful in a Temperature-Based Snow-Accounting Routine? Part 1 Comparison of Six Snow Accounting Routines on 380 Catchments.” *Journal of Hydrology* 517 (September): 1166–75. https://doi.org/10.1016/j.jhydrol.2014.04.059.

van Vuuren, Detlef P., Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C. Hurtt, et al. 2011. “The Representative Concentration Pathways: An Overview.” *Climatic Change* 109 (1-2): 5–31. https://doi.org/10.1007/s10584-011-0148-z.

Van Grinsven, Matthew J., Joseph P. Shannon, Joshua C. Davis, Nicholas W. Bolton, Joseph W. Wagenbrenner, Randall K. Kolka, and Thomas G. Pypker. 2017. “Source Water Contributions and Hydrologic Responses to Simulated Emerald Ash Borer Infestations in Depressional Black Ash Wetlands.” *Ecohydrology* 10 (7): e1862. https://doi.org/10.1002/eco.1862.

Vavrus, S.J., Behnke, R.J., 2014. A comparison of projected future precipitation in Wisconsin using global and downscaled climate model simulations: Implications for public health. International Journal of Climatology 34, 3106–3124. https://doi.org/10.1002/joc.3897

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, and Richard W. Katz. 2015. "Coupled Stochastic Weather Generation Using Spatial and Generalized Linear Models." *Stochastic Environmental Research and Risk Assessment* 29 (2): 347–56. https://doi.org/10.1007/s00477-014-0911-6.

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, Guillermo Podestá, and Federico Bert. 2019. "BayGEN: A Bayesian Space-Time Stochastic Weather Generator." *Water Resources Research* 55 (4): 2900–2915. https://doi.org/10.1029/2017WR022473.

Watras, C. J., K. A. Morrison, J. L. Rubsam, and I. Buffam. 2017. "Estimates of Evapotranspiration from Contrasting Wisconsin Peatlands Based on Diel Water Table Oscillations." *Ecohydrology* 10 (4): e1834. https://doi.org/10.1002/eco.1834.

Watras, C. J., J. S. Read, K. D. Holman, Z. Liu, Y.-Y. Song, A. J. Watras, S. Morgan, and E. H. Stanley. 2014. "Decadal Oscillation of Lakes and Aquifers in the Upper Great Lakes Region of North America: Hydroclimatic Implications: DECADAL WATER LEVEL OSCILLATION." *Geophysical Research Letters* 41 (2): 456–62. https://doi.org/10.1002/2013GL058679.

White, Walter Noy. 1932. *A Method of Estimating Ground-Water Supplies Based on Discharge by Plants and Evaporation from Soil: Results of Investigations in Escalante Valley, Utah*. Vol. 659. US Government Printing Office.

Wilks, D S, and R L Wilby. 1999. "The Weather Generation Game: A Review of Stochastic Weather Models." *Progress in Physical Geography* 23 (3): 329–57. https://doi.org/10.1177/030913339902300302.

Wilks, Daniel S. 2012. "Stochastic Weather Generators for Climate-Change Downscaling, Part II: Multivariable and Spatially Coherent Multisite Downscaling: Stochastic Weather Generators for Climate-Change Downscaling." *Wiley Interdisciplinary Reviews: Climate Change* 3 (3): 267–78. https://doi.org/10.1002/wcc.167.

Zhu, Jianting, Michael Young, John Healey, Richard Jasoni, and John Osterberg. 2011. "Interference of River Level Changes on Riparian Zone Evapotranspiration Estimates from Diurnal Groundwater Level Fluctuations." *Journal of Hydrology* 403 (3-4): 381–89. https://doi.org/10.1016/j.jhydrol.2011.04.016.

A Supplemental Data for Chapter 2

Table S1. Model coefficients and RMSE for all 36 transducer pairs used in the study. Coefficients are reported as the mean ± 1 standard error of the coefficient estimates from all 1000 bootstrap model fits. RMSE is the mean RMSE for all 1000 bootstrap model fits.

Water Transducer	Barometric Transducer	Intercept	Slopes (°C)			RMSE (cm)
			Air Temperature	Water Temperature	Temperature Gradient	
2013939	1066019	6.975 ± 0.003	-0.248 ± 0.000	0.229 ± 0.000	0.015 ± 0.000	0.343
2013939	1065861	9.256 ± 0.004	-0.199 ± 0.000	0.230 ± 0.000	0.020 ± 0.000	0.388
2025928	1066019	4.272 ± 0.004	-0.219 ± 0.000	0.132 ± 0.000	0.018 ± 0.000	0.464
2025928	1065861	6.568 ± 0.004	-0.169 ± 0.000	0.132 ± 0.000	0.023 ± 0.000	0.484
2059683	1066019	1.624 ± 0.003	-0.251 ± 0.000	0.065 ± 0.000	0.015 ± 0.000	0.397
2059683	1065861	3.887 ± 0.004	-0.203 ± 0.000	0.068 ± 0.000	0.019 ± 0.000	0.439
2064737	1066019	-0.712± 0.005	-0.252 ± 0.000	0.025 ± 0.000	0.015 ± 0.000	0.396
2064737	1065861	1.561 ± 0.005	-0.203 ± 0.000	0.027 ± 0.001	0.020 ± 0.000	0.442
2064745	1066019	1.085 ± 0.004	-0.248 ± 0.000	0.038 ± 0.000	0.015 ± 0.000	0.402
2064745	1065861	3.353 ± 0.005	-0.199 ± 0.000	0.041 ± 0.000	0.019 ± 0.000	0.44
2030899	1066019	-2.254 ± 0.003	-0.252 ± 0.000	-0.260 ± 0.000	0.015 ± 0.000	0.406
2030899	1065861	0.026 ± 0.003	-0.203 ± 0.000	-0.259 ± 0.000	0.020 ± 0.000	0.447
2064734	1066019	-0.147 ± 0.003	-0.252 ± 0.000	-0.059 ± 0.000	0.014 ± 0.000	0.393
2064734	1065861	2.126 ± 0.004	-0.203 ± 0.000	-0.057 ± 0.000	0.019 ± 0.000	0.437
2100561	1066019	-0.079 ± 0.003	-0.251 ± 0.000	0.123 ± 0.000	0.015 ± 0.000	0.398
2100561	1065861	2.201 ± 0.004	-0.203 ± 0.000	0.124 ± 0.000	0.019 ± 0.000	0.443
2069158	1066019	1.010 ± 0.003	-0.249 ± 0.000	0.112 ± 0.000	0.014 ± 0.000	0.394
2069158	1065861	3.270 ± 0.004	-0.201 ± 0.000	0.115 ± 0.000	0.019 ± 0.000	0.441
2064738	1066019	0.243 ± 0.003	-0.250 ± 0.000	0.169 ± 0.000	0.015 ± 0.000	0.397
2064738	1065861	2.523 ± 0.004	-0.201 ± 0.000	0.170 ± 0.000	0.019 ± 0.000	0.44

2064739	1066019	1.217 ± 0.003	-0.251 ± 0.000	0.061 ± 0.000	0.015 ± 0.000	0.393
2064739	1065861	3.478 ± 0.004	-0.202 ± 0.000	0.064 ± 0.000	0.019 ± 0.000	0.436
2104452	1066019	-0.788 ± 0.003	-0.251 ± 0.000	0.119 ± 0.000	0.015 ± 0.000	0.395
2104452	1065861	1.485 ± 0.004	-0.202 ± 0.000	0.121 ± 0.000	0.019 ± 0.000	0.438
1062452	1066019	1.218 ± 0.003	-0.251 ± 0.000	0.109 ± 0.000	0.015 ± 0.000	0.395
1062452	1065861	3.486 ± 0.004	-0.203 ± 0.000	0.111 ± 0.000	0.020 ± 0.000	0.441
1033239	1066019	0.290 ± 0.003	-0.251 ± 0.000	0.144 ± 0.000	0.014 ± 0.000	0.393
1033239	1065861	2.551 ± 0.004	-0.202 ± 0.000	0.147 ± 0.000	0.019 ± 0.000	0.439
1062534	1066019	-0.291 ± 0.003	-0.252 ± 0.000	0.206 ± 0.000	0.014 ± 0.000	0.394
1062534	1065861	1.975 ± 0.004	-0.203 ± 0.000	0.209 ± 0.000	0.019 ± 0.000	0.437
1062520	1066019	3.230 ± 0.003	-0.252 ± 0.000	0.106 ± 0.000	0.015 ± 0.000	0.393
1062520	1065861	5.483 ± 0.004	-0.203 ± 0.000	0.109 ± 0.000	0.020 ± 0.000	0.436
1062528	1066019	4.740 ± 0.004	-0.253 ± 0.000	-0.023 ± 0.000	0.014 ± 0.000	0.39
1062528	1065861	6.999 ± 0.005	-0.205 ± 0.000	-0.020 ± 0.001	0.019 ± 0.000	0.433
1066016	1066019	2.807 ± 0.003	-0.251 ± 0.000	0.196 ± 0.000	0.014 ± 0.000	0.393
1066016	1065861	5.083 ± 0.004	-0.202 ± 0.000	0.198 ± 0.000	0.019 ± 0.000	0.435

Table S2. Measurement error ranges for all 36 transducer pairs used in the study as 95% confidence bands. Instrument error derived from the manufacturer stated instrument accuracy for each of the water and barometric pressure transducers. Propagated error is the final error that incorporates both the instrument error and the mean of the uncertainty introduced from the correction models.

Water Transducer	Barometric Transducer	Propagated Error (cm)	Instrument Error (cm)
1033239	1065861	1.326	1.191
1033239	1066019	0.457	1.191
1062452	1065861	1.392	1.191
1062452	1066019	0.445	1.191
1062520	1065861	1.407	1.191
1062520	1066019	0.427	1.191
1062528	1065861	1.38	1.191
1062528	1066019	0.435	1.191
1062534	1065861	1.448	1.191
1062534	1066019	0.421	1.191
1066016	1065861	1.391	1.191
1066016	1066019	0.424	1.191
2013939	1065861	1.372	0.859
2013939	1066019	0.481	0.859
2025928	1065861	1.015	0.859
2025928	1066019	0.899	0.859
2030899	1065861	1.344	1.191
2030899	1066019	0.522	1.191
2059683	1065861	1.34	1.191
2059683	1066019	0.468	1.191
2064734	1065861	1.376	1.191
2064734	1066019	0.486	1.191
2064737	1065861	1.368	1.191
2064737	1066019	0.477	1.191
2064738	1065861	1.335	1.191
2064738	1066019	0.483	1.191
2064739	1065861	1.382	1.191
2064739	1066019	0.456	1.191
2064745	1065861	1.514	1.191
2064745	1066019	0.481	1.191
2069158	1065861	1.408	1.191
2069158	1066019	0.453	1.191
2100561	1065861	1.377	1.191
2100561	1066019	0.478	1.191
2104452	1065861	1.356	1.191
2104452	1066019	0.49	1.191

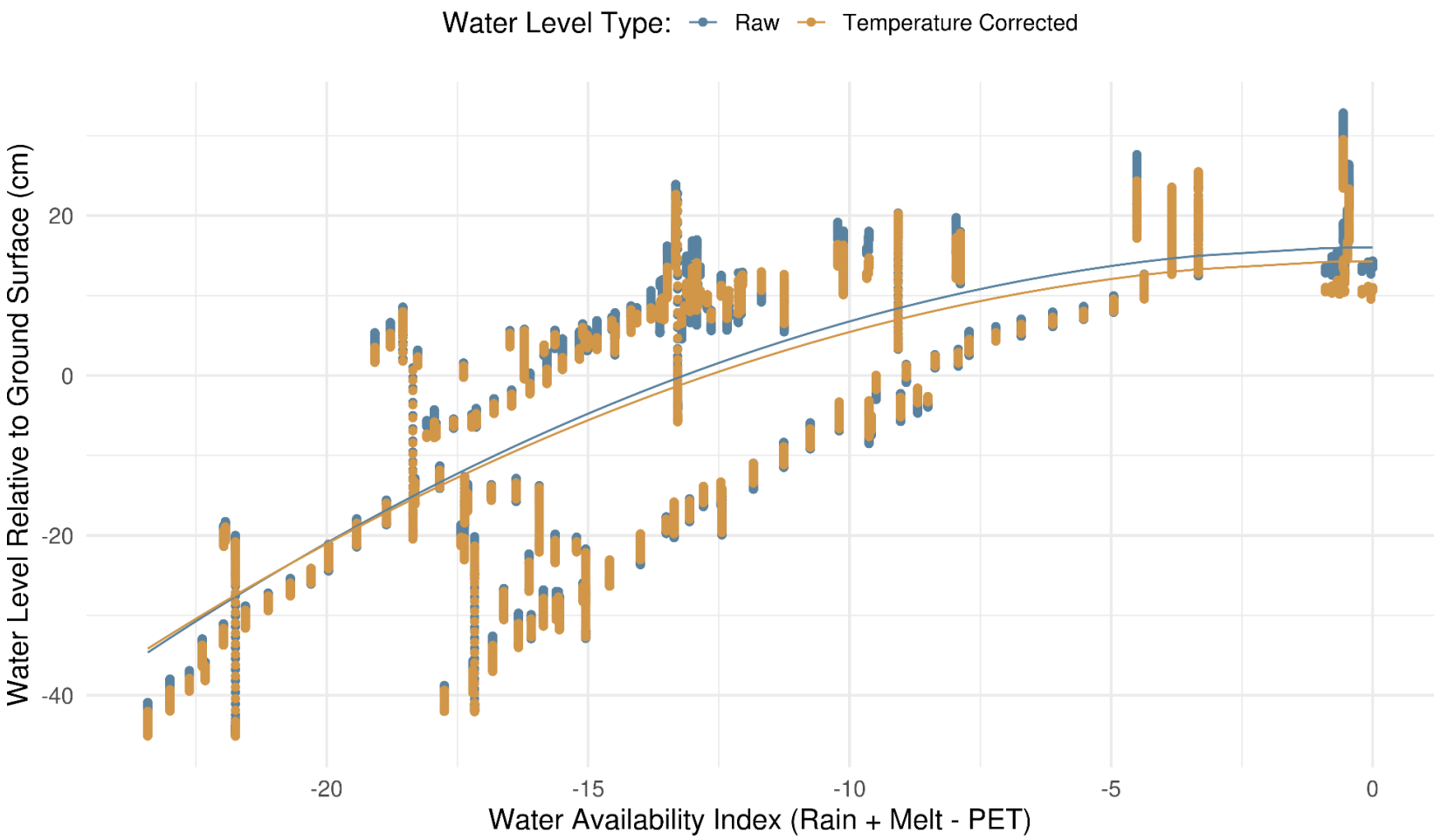


Figure S1. Observed ecosystem specific yield (E_{Sy}), points, and fitted functions to estimate continuous E_{Sy} for both raw and temperature-corrected data using median coefficients for the appropriate pressure transducer pair model.

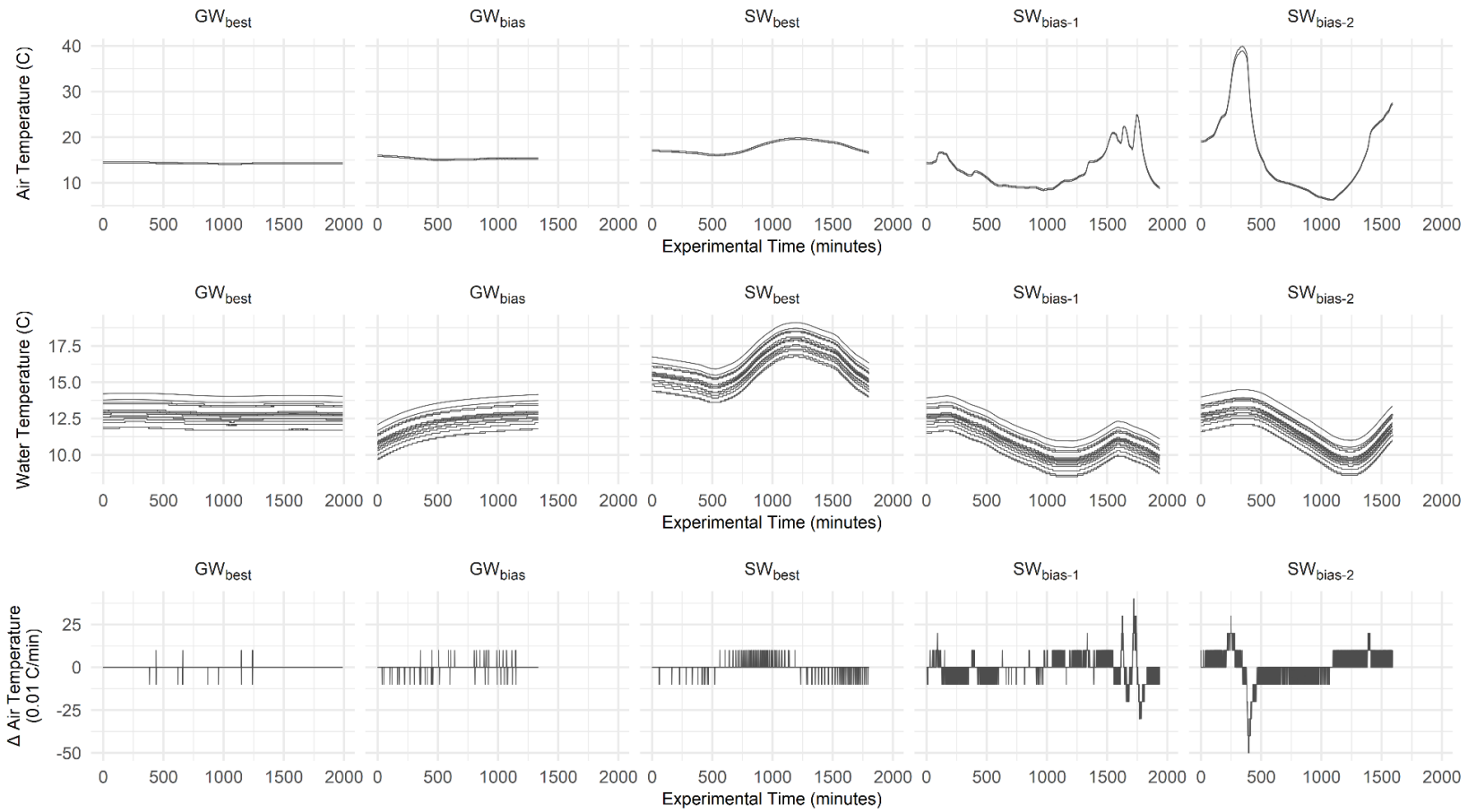


Figure S2. Environmental conditions (air and water temperature, and rate of change of air temperature) over each experimental period as recorded by the individual transducers in the study.

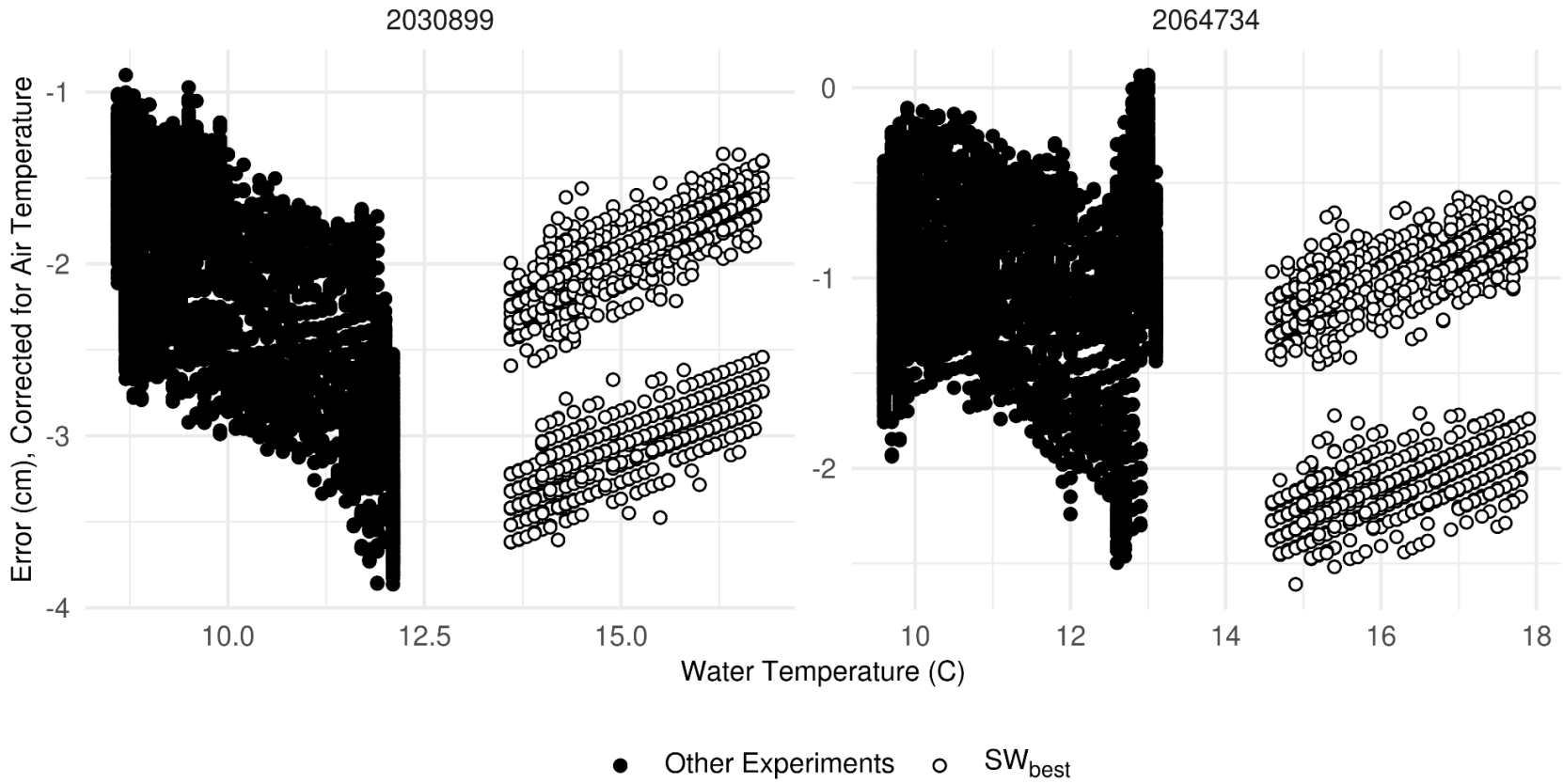


Figure S3. Water-temperature driven error for two water pressure transducers. Points are differentiated according to experimental period and show conditions under which the relationship between water temperature and error reverses.