A Grain Carried by the Flood:
methods and data for global change ecology amidst a data deluge

Anthony Francesco Cannistra

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
UNIVERSITY OF WASHINGTON
2020

READING COMMITTEE
Lauren Buckley, Chair
Curtis Deutsch
Janneke HilleRisLambers

PROGRAM AUTHORIZED TO OFFER DEGREE:
BIOLOGY

UNIVERSITY OF WASHINGTON

ABSTRACT

A Grain Carried by the Flood: methods and data for global change ecology amidst a data deluge

Anthony Francesco Cannistra

CHAIR OF THE SUPERVISORY COMMITTEE:
Lauren Buckley
Department of Biology

Forecasting the responses of ecological systems to changing environment is a critical area of modern ecology research. An overwhelming amount of openly-available ecological and environmental data is emerging in service of this goal, but often the methodology for producing ecological insight from these heterogeneous data sources is out of reach of standard ecological practice. In this dissertation, I investigate opportunities to use open, heterogeneous ecological data to produce new insight via contributions in modeling methodology, emerging data sources, and global-scale mechanistic analysis. In the first of three chapters, I find that modern nonlinear modeling methods are able to improve range shift predictions made via species traits. Second, I develop a snow cover data product for montane ecological research from an emerging satellite observation platform with unprecedented spatial and temporal resolution. Finally, I contribute testable predictions of phytoplankton physiological responses to marine heatwave events by pairing a globally-distributed observational dataset with empirically-derived thermal reaction norms of fitness. Taken together, these contributions represent both independent discoveries toward more accurate ecological forecasting and the extraordinary potential of an approach to ecological research driven by open ecological and environmental data sources and modern methods.

Table of Contents

Acknowledgements

This dissertation is the product of extraordinary generosity, support, and encouragement from many people and organizations. I'm glad to highlight these here.

Funding Agencies

I gratefully acknowledge the research funding and training opportunities I received from the eScience Institute Integrative Graduate Education and Research Traineeship in Big Data and Data Science, the National Science Foundation Graduate Research Fellowship, and the Earth Science Information Partners Incubator Program. These organizations facilitated my broad exploration and growth by enabling a diverse array of coursework, research projects, and travel opportunities that greatly enriched both my experience as a graduate student and the work in this dissertation, and for those opportunities I am deeply grateful.

My Advisor

With particular gratitude I acknowledge my advisor, Lauren Buckley, for her consistent guidance and support during these past four years. I am grateful for her encouragement and advice in service of both my research and my personal career goals, despite a somewhat meandering path!

My Committee

I am greatly fortunate to have a supportive and inquisitive committee in service of this work. I'm grateful to Janneke Hille Ris Lambers for sharing her expertise on Mount Rainier and plant phenology, and for welcoming me into her lab group meetings. Thank you to Curtis Deutsch for his wide breadth of expertise in providing direction to the projects in this work, especially with regard to the ocean. Thank you to Julian Olden, who brought important conversations about modeling, machine learning, and communicating those results into meetings and my General Exam. I'm grateful for these contributions and all others that each of these individuals have generously provided to this work.

Department and Colleagues

I owe much of my growth and sanity to many individuals from the Department of Biology and the greater University of Washington community. First I must acknowledge my cohort—Romi Ramos Báez, Caroline Cappello, Stuart Graham, Olivia Kosterlitz, Andy Magee, Sage Malingen, and Savannah Olroyd—who are an excellent, supportive, and wildly intelligent group of scientists and friends whose continued support and camaraderie was continuously invaluable in these four years. Thank you to the many folks at the U.W. eScience Institute who have supported my growth as an academic data scientist in myriad ways. I have extraordinary gratitude to Nicoleta Cristea, David Shean, Jessica Lundquist, and the members of the Mountain Hydrology research group for helping me turn my passion for snow, mountains, and remote sensing into a successful research project. Finally, and on behalf of all Biology graduate students, I thank Jennifer Nemhauser for her tireless efforts to make our department a place of unmatched equity, inclusion, and enjoyment for all.

My Family and Friends

The deep gratitude I have for my friends and family in supporting, advising, and encouraging me during the past years is difficult to accurately express. It will barely suffice to say that this dissertation has only happened because of you. Thank you all.

Introduction

Modern ecology has a dual mandate: to broaden and deepen scientific understanding of fundamental processes underlying ecosystem structure and function, and to provide actionable insights to society in the face of drastic global environmental change (Lubchenco, 1998). The latter of the two, sometimes referred to as "action ecology" (Bonilla et al., 2012) relies upon a framework of scientific inquiry that produces "immediate, implementable, and targeted solutions to urgent ecological problems" via "technology-driven, innovative, and aggregative" methodologies (White et al., 2016). The role of the ecologist in this framework is therefore to design rigorous studies that leverage heterogeneous, readily-available, and perhaps imperfectly-collected data sources, with the intent to efficiently generate trustworthy, actionable insight to specifically inform policy and management.

The recent drastic increase in availability and accessibility of data relevant to ecological study serves as the foundation for such work. Readily-accessible information such as ecological observations (e.g. NEON, GBIF), genetic records (e.g GenBank), earth and climate system observations (e.g. NASA EarthData), museum collections (e.g. iDigBio), and other open data sources can be carefully combined by ecologists to efficiently answer questions from organismal to global scales. The potential to unlock actionable insights from these open data, taken together, is extraordinary. While it is worth noting that this "big data" approach to environmental and ecological science has met reasonable critique (e.g. with regard to the potential for distancing between scientist and data, hidden assumptions or bias in heterogeneous data, and the worrisome "data without theory" paradigm; Salmond et al. (2017)), responsible employment of these data and methods is possible and necessary to meet today's global challenges.

To do so effectively, however, requires new methods and tools. The efficient and rigorous analysis of large quantities of heterogeneous environmental and ecological data requires careful data management and modeling practices. Yet the process of transforming "dirty," heterogeneous data into a sound set of records for analysis is often consuming or stretches beyond a given researcher's skill.

The role of this dissertation is to contribute patterns of inquiry, novel tools and data, and synthetic insights gained via the implementation of this sort of "action ecology" in the domain of ecological responses to global environmental change. I do this by leveraging diverse open data sources and methodologies new to ecology to illustrate the potential of this type of research. My intent with this dissertation is to both contribute deeper insights on the processes underlying landscape-scale ecological change as a result of anthropogenic climate change, and also to demonstrate a new paradigm in ecological research, empowered by novel methods and data.

In Chapter 1, I attempt to bring "new life" to existing observations of historical species range shifts to examine whether species traits can be used to predict the magnitude and direction of range shifts under environmental change. I use a combination of previously-published range shift observations and species trait databases to address the long-standing ecological question of whether species' traits can predict range shifts, but without relying upon the assumption that the relationships between trait and range shift are linear, as previous work has. This new analysis relies upon nonlinear machine learning-based modeling methods and emerging model interpretation techniques to interrogate these models for their biological intuition. I show that these nonlinear modeling methods improve our ability to forecast species range shifts given environmental change, suggesting that future work in this domain should consider the nonlinear effects of traits when attempting to use them to forecast species' responses to change.

The second chapter of this work is focused on the generation and validation of a novel data set and toolkit for ecological analysis in environments with seasonal snow cover. The extent and duration of seasonal snow is a strong control on ecological systems,

but currently available tools to measure these patterns are often too spatially coarse, cost prohibitive, or infrequently measured to match the scale of ecological processes or subjects of inquiry. To remedy this, I leverage an emerging optical remote sensing satellite data source from Planet Labs, Inc., to create a daily snow-cover dataset at 3 meter spatial resolution, with the future potential for global coverage. I pair high-resolution airborne lidar data with established machine learning-based computer vision techniques to generate a model for converting daily, high-resolution imagery to this snow cover product. I show that this method produces observations of snow cover that are as accurate as other trusted sources of snow cover data, and provide an open-source implementation of these tools for the scientific community, enabling a wide range of ecological, hydrological, and environmental research.

Finally, I present in my third chapter a study of the effects of marine heatwaves (MHWs), an increasingly frequent climatological anomaly in ocean temperature, on ocean ecosystems at global scale. While previous efforts have succeeded in identifying these anomalies as drivers of particular ecological impacts, these studies have been limited both geographically and taxonomically. I base this study on the intuition that phytoplankton, with their unique biogeochemical and trophic context, can serve as "sentinels" of whole-ecosystem responses to these climatological events. This analysis combines a robust geographically distributed dataset of published phytoplankton thermal physiology data with a global 38-year record of high-resolution sea surface temperature to assess the fitness consequences of marine heatwave events on phytoplankton at planetary scale. I demonstrate that MHW events can cause varying positive and negative phytoplankton fitness consequences, and describe the latitudinal and seasonal patterns that are associated with these responses. This serves as the first globally-distributed analysis of ecological impacts from MHW events, and our findings serve both a framework for forecasting global patterns and as the foundation for further investigation into the ecological responses to future MHW events.

Taken together, these three independent contributions demonstrate the possibility inherent in harnessing the deluge of open data and new methods available to ecologists. I show that this approach leads to both advances in foundational ecological understanding and the production of actionable, relevant, and timely insights to pressing environmental challenges. It is important to note that in demonstrating this approach I do not in any way discredit or disenfranchise the trusted methods and approaches of classical ecology. Much the opposite: the majority of the findings present in these chapters would have been impossible without the careful work of previous ecologists, earth scientists, computer scientists, and others who have contributed to the wealth of data leveraged here; these efforts must continue. The future of "action ecology" relies upon concerted efforts across field and experimental ecology, earth and environmental observation and modeling, statistics and computing, open data production and cataloguing, and public policy to create insights that drive both foundational knowledge and solutions to difficult environmental challenges facing society. Despite being a single effort, the results presented here lay the intellectual and methodological groundwork for further work investigating the effects of global change on ecological systems at varying scales within this emerging paradigm.

References

Bonilla, N. O., Scholl, J., Armstrong, M., Pieri, D., Otero, B., Labrado, A., McPherson, D., and Nelson, M. (2012). Ecological Science and Public Policy: an Intersection of Action Ecology. *Bulletin of the Ecological Society of America*, 93(4):340–346.

Lubchenco, J. (1998). Entering the Century of the Environment: A New Social Contract for Science. *Science*, 279(5350):491–497.

Salmond, J. A., Tadaki, M., and Dickson, M. (2017). Can big data tame a "naughty" world? *The Canadian Geographer / Le Géographe canadien*, 61(1):52–63.

White, R. L., Sutton, A. E., Salguero-Gómez, R., Bray, T. C., Campbell, H., Cieraad, E., Geekiyanage, N., Gherardi, L., Hughes, A. C., Jørgensen, P. S., Poisot, T., DeSoto, L., and Zimmerman, N. (2016). The next generation of action ecology: novel approaches towards global ecological research. *Ecosphere*.

Chapter 1

Improving range shift predictions: enhancing the power of traits

Anthony F. Cannistra, Lauren B. Buckley

SUBMITTED AS A *Statistical Report* TO *Ecology*

Abstract

Accurately predicting species' range shifts in response to environmental change is paramount for understanding ecological processes and global change. In synthetic analyses, traits emerge as significant but weak predictors of species' range shifts across recent climate change. These studies assume linear responses to traits, while detailed empirical work often reveals trait responses that are unimodal and contain thresholds or and other nonlinearities. We hypothesize that the use of linear modeling approaches fails to capture these nonlinearities and therefore may be under-powering traits to predict range shifts. We evaluate the predictive performance of four different machine learning approaches that can capture nonlinear relationships (ridge-regularized linear regression, ridge-regularized kernel regression, support vector regression, and random forests). We apply our models using four multi-decadal range shift datasets in montane plants, montane small mammals, and marine fish. We show that nonlinear approaches perform substantially better than least-squares linear modeling in reproducing historical range shifts. In addition, using novel model observation and interrogation techniques, the trait classes (e.g. dispersal- or diet-related traits) that we identify as primary drivers of model predictions are consistent with expectations. However, disagreements among models in the directionality of trait predictors suggests limits to trait-based statistical predictive frameworks.

Introduction

Species have been responding to recent climate change by tracking their environment in space or time, adapting or acclimating, or facing declines (Parmesan, 2006), but we are largely unable to predict how particular species will respond (Maguire et al., 2015). Extensive documentation of shifts in distribution and seasonal timing (phenology) reveal that responses vary among species markedly in direction and extent even in the same geographic region (Rapacciuolo et al., 2014). Focused empirical studies often succeed in identifying species' traits that govern climate change responses (e.g., Adrian et al. (2006)) and consequently functional approaches are accelerating in climate change ecology (Buckley and Kingsolver, 2012; Urban et al., 2016). However, attempts to use traits to predict the relative magnitude of responses among species generally identify traits that are significant, but often too weak (accounting for $\tilde{1}6\%$ of the among species variation in range shifts, Buckley and Kingsolver (2012)), to confidently project climate change responses (Estrada et al., 2016; MacLean and Beissinger, 2017a). How can we close the discrepancy between traits predicting responses well in some taxonomically focused studies but poorly in broad studies? Are there statistical techniques that would allow us to better generalize the importance of traits in mediating climate change responses?

Addressing such questions is imperative for anticipating and mitigating the biological impacts of climate change. Indeed, traits are already being used to predict species' sensitivity to climate change in vulnerability frameworks (Foden et al., 2013). However, the frameworks remain largely untested and those that have been tested perform poorly in tests (Wheatley et al., 2017). Many attempts to use species' traits to predict the magnitude of their climate change responses rely on linear regression (Buckley and Kingsolver, 2012; MacLean and Beissinger, 2017b), yet detailed empirical studies often reveal non-linear relationships between traits and their function (Stenseth and Mysterud, 2002). Unimodal relationships and thresholds are common. For example, extreme diet specialization may drive a species to track the range shift of a food item (Diamond et al., 2011), but reducing diet specialization only slightly may alleviate the need for a species to track its food. Diet generalization could facilitate species moving to capitalize on newly climatically suitable habitat, yielding a unimodal relationship between diet specialization and the magnitude of range shifts. Likewise, low dispersal ability may prevent a species from tracking its environmental niche (Schloss et al., 2012), but the threshold of dispersal ability that allows species to track their niche may be relatively low. Covariation and trade-offs among traits and differences in the developmental dependencies of traits may also produce nonlinearities (Fitt et al., 2018). Can statistical techniques that allow for non-linear relationships between traits and species' responses improve our predictive ability?

Non-linear modelling techniques have been underutilized in predicting climate change responses (Olden et al., 2008). Standard approaches to capture variable interactions and nonlinearities in linear regressions (such as the explicit inclusion of interacting variables or polynomial expansion) rely on prior knowledge or model selection techniques to determine which variables to select. Other model types, such as machine learning approaches that optimize model parameters, are better suited to capture functional relationships among variables. These models, while offering statistical-robustness and efficiency, can be opaque and rarely afford clear coefficients to inspect when assessing the model's learned correlations.

Here we assess whether machine learning-based models can better use species' traits to hindcast (and subsequently predict) the magnitude and direction of range shifts observed in response to recent climate change. First, we consider whether several models which are able to capture nonlinear relationships can outperform linear models in their hindcasting ability. Second, we use recently developed model interpretability techniques to ask whether model coefficients are consistent across analytical approaches and concur with ecological theory. Novel model inspection approaches can reveal details of model performance, which addresses reasonable concerns about the "black box" nature of many machine learning-based models. We assess model performance using four datasets encompassing a broad taxonomic range. The number of species per dataset ranges from 20 to 176, and range shifts were observed over time spans ranging from 30 to 100+ years. Each dataset was derived from previous evaluations of traits as range shift predictors and consists of a list of focal species, associated species-level traits, and a range shift metric. We examine (1) whether nonlinear methods can improve the ability of traits to describe range shifts compared to linear methods, (2) whether the novel methods identify important traits consistent with significant results from other studies, and (3) whether the directionality of the modeled effect of traits is consistent across model types.

Materials and Methods

Modeling Approach

We applied three classes of learning algorithms: regularized linear regression, kernel-based regression, and tree-based regression. Regularization is a modification to generalized linear regression that limits model complexity to avoid overfitting (Hastie et al., 2009, Section 3.4). Several types of regularization exist; we chose to use a "ridge"-regularized linear model, which imposes a penalty on the magnitude of each learned coefficient. The cumulative effect of this regularization procedure is a set of coefficients which both minimize mean squared error on the training data and prevent overfitting. These coefficients can be interpreted explicitly as with ordinary least squares regression.

While regularization reduces overfitting when compared to a standard least-squares linear fit, regularized linear models are still not able to capture nonlinearities among or interactions between predictive variables. To remedy this, we employ two additional classes of models: kernel-based regression and tree-based regression. A "kernel" is a function that projects a set of input data, often into a high-dimensional space, to allow for the linear "separability" of the data for the purposes of classification or

regression (Hastie et al., 2009, Section 12.3). The Kernel Ridge method employed herein uses a radial basis function (or squared exponential) kernel applied to the training data and fits a ridge-regularized linear model to this transformed input. As a result of this transformation the learned coefficients are not immediately interpretable.

We also evaluate a kernel-based technique known as a support vector machine (SVM). This popular learning method can be formulated for regression, is robust to outliers, and can capture nonlinearities and variable interactions through a similar radial basis function kernel as in the Kernel Ridge approach. Like the Kernel Ridge method, the SVM regressor does not have interpretable coefficients. Finally, we train a random forest regression algorithm to evaluate the performance of tree-based methods. We implement these models in the Python programming language using the scikit-learn software package (Pedregosa et al., 2011), though all analyses can be computed in the R language using available machine learning packages. All code for this project is available on GitHub at `https://github.com/huckleylab/cc_traits`.

For comparison to the original analyses, we apply an ordinary least squares regression model, which assumes linear relationships and no variable interaction. In addition, in order to compare to more traditional nonlinear modeling techniques, we apply a second-degree polynomial transformation with no variable interactions to the input data and use these transformed data to assess both OLS and Ridge-OLS in their ability to capture nonlinear relationships. While the original analyses were mostly conducted in a single-variable framework (that is, to assess the effect size of M different potential predictive variables M models were trained, each model containing only 1 variable), we include all variables in single analyses to enable the models to capture variable interactions. and to follow a common predictive modeling paradigm.

Evaluation

To assess the predictive performance of our models we employ a k-fold cross-validation scheme (Hastie et al., 2009, Section 7.10) combined with a mean squared error metric. This cross-validation technique has been shown to estimate expected description error (Hastie et al., 2009, Section 17.12) by randomly partitioning data points into k subsets, each with N/k members; $k-1$ subsets are used to fit the model (the "training set"), reserving one subset for testing model performance. Since our datasets vary in size considerably, we choose $k = 5$ or $k = 10$ depending on the size of the dataset ($k = 5$ for smaller datasets where $N < 20$). Each of the k subsets is used exactly once for evaluation. A mean squared error (MSE) loss is then computed for the model prediction of range shift magnitude in the reserved test data, which is the sum of squared differences between the model prediction and actual range shift magnitude over all data in the testing set. This process is repeated k times and a mean statistic is computed across the k MSE values that result (known as a "cross-validation mean"). The units of MSE estimates are those of the range shift, so MSE provides a direct assessment of model predictive ability. The same cross-validation mean approach is used to compute average variable importance values to evaluate model drivers, described below.

Model Interpretation

The core of any basic regression analysis is typically an inspection of the significance and direction of the coefficients of a fitted model. However, the kernel methods employed herein (Kernel Ridge, SVM) do not expose any interpretable coefficients. To address this, we utilize the Shapley additive feature value method, proposed in (Lundberg and Lee, 2017). Shapley values are computed for each variable by treating the explanation of a given model's prediction as a model in and of itself–values are computed by training an additive method derived from cooperative game theory to learn each variable's contribution to a model's prediction. Explanations are generated from each prediction in the model training set to identify the most important variables during training and are averaged for each feature across all training examples to generate a whole-model feature importance scale. We perform this procedure for each of the training sets generated by the cross-validation scheme described above to compute average feature importance values. See Lundberg and Lee (2017) for a comprehensive demonstration of this method. To compare the learning techniques, we use either mean regression coefficients (for OLS and Ridge regression), mean Shapley variable importance values (Kernel Ridge and SVM), or mean Gini variable importance values (Random Forest; Breiman (2001)) to rank all variables such that each feature has an "importance ranking" for each of the several regression methods. The nature of this generalized "importance" value is specific to each modeling method and is dependent upon the approach used to derive it as described above. We use the sign of these importance values to compare the directionality of predicted trait

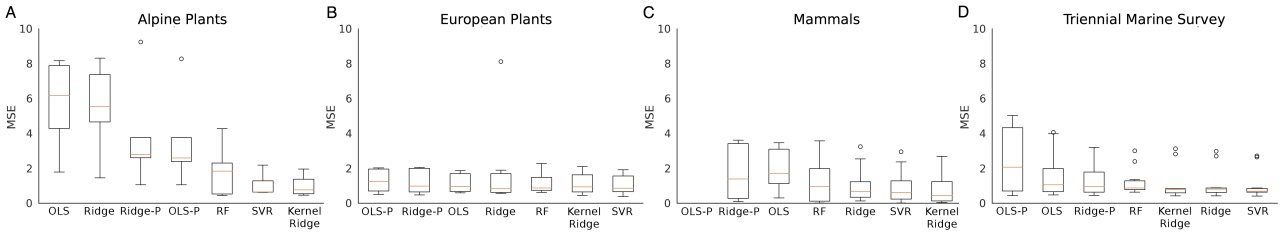

Figure 1.1: The machine learning approaches reduce the mean squared error (MSE, 10-fold cross-validation) of range shift predictions below the MSE of the standard linear regression approach (OLS: ordinary least squares) across three of four datasets (A, C, D). Box represents interquartile range (lower to upper), central lines shows data median, and whiskers represent the range of the data. Support Vector Regression (SVR) and Kernel Ridge models exhibit stronger performance than ridge regularized linear (Ridge) or Random Forest (RF) models across the datasets, and all ML methods outperform OLS and Ridge with polynomial features (OLS-P/Ridge-P) The MSE units correspond to the range shift metric (A-C: m, D: degrees latitude) and thus directly indicate model performance. Note: The range of MSE for the OLS-P model on the Mammals data (Panel C) exceeds the axis range and thus does not appear.

drivers for models other than random forests (with the exception of Gini scores, which are an unsigned information-theoretic measure (Breiman, 2001)). All means are k-fold cross-validated means.

Trait and Range Shift Data

We evaluate our approach independently by replicating analysis across four datasets that (1) repeated historical surveys or conducted continuous surveys along latitudinal or elevational gradients to quantify shifts in northern or upper elevation range boundaries over at least three decades of change and (2) included all surveyed species (i.e. rather than including only species that shifted significantly). The first two data sets are those used by (Angert et al., 2011) to assess the predictive power of traits. These datasets supplement trait data with elevational range shift data for Swiss alpine plants (Holzinger et al. (2008); $N = 139$) and for Western North American small mammals (Moritz et al., 2008). A third database from (Rumpf et al., 2018) consists of elevational range shifts for European montane plants coupled with empirically measured trait data supplemented with data from the TRY Plant Trait Database ((Kattge et al., 2011); https://www.try-db.org and Supplementary Information, Section 2.2) and other databases (e.g. Bjorkman et al., 2018). These two plant databases were chosen in an effort to directly replicate data used in previous range shift studies, despite some overlap among them. The fourth database was created by pairing estimates of latitudinal range shifts from coastal North American marine fish surveys (Pinsky et al., 2013) with functional trait data in Fishbase (https://www.fishbase.org, Froese and Pauly (2010)) and Supplementary Information, Section 2.1). Each dataset includes a directional range shift: negative values indicate shifts downward in elevation (m) for the first three datasets and equatorward in latitude (degrees) for the marine dataset. We remove samples which are missing any traits, one-hot encode categorical traits (i.e., generate one boolean column for each category), and normalize/center the numeric traits to have zero mean and unit norm. After this processing the Swiss plants dataset contains $N = 20$ species and $d = 38$ traits (Table S1); the Yosemite mammal dataset contains $N = 28$ species and $d = 19$ traits (Table S2); the European plants dataset contains $N = 176$ species and $d = 18$ traits (Table S3); the marine fish dataset contains $N = 76$ species and $d = 17$ traits (Table S4).

Results

We find that the machine learning approaches improve model performance over an ordinary least squares (OLS) model baseline. For the two smallest datasets (Swiss alpine plants and Western NA mammals), the four machine learning approaches perform substantially better than the OLS models but similarly to each other. For the other two datasets (European montane plants and marine fish), the performance advantages of the machine learning approaches are less substantial and the four machine learning approaches differ more in their performance. Support Vector Regression (SVR) and Kernel Ridge emerge as the best performing methods across datasets (Figure 1.1), reducing mean error in range shift estimates by an average of 62.8% and 61.6% relative to OLS, respectively (medians of 50.6% and 49.6%, respectively). We focus on results from the Swiss alpine plants dataset (Angert et al., 2011; Holzinger et al., 2008) to demonstrate findings. The initial OLS analysis found that individual predictors accounted for relatively little variance

Figure 1.2: The model approaches select similar traits as important (larger and red = more important) for predicting range shifts across datasets. Traits (described in Appendix S1: Table S1-S4) are listed in order of decreasing mean importance (increasing mean rank, right column) across all methods (except OLS) for each dataset (panels). Model abbreviations are as in Figure 1.1)

Figure 1.3: Models are consistent to varying degrees with regard to the variable importance values and directionality of top-ranking traits. We depict model coefficients (OLS, Ridge), Shapley feature importance values (Kernel Ridge, SVR), or Gini feature importance scores (RF) for top ten traits by rank (Figure 1.2) for each dataset (rows). Model abbreviations are as in Figure 1.1 and traits are as in Figure 1.2. Fractions indicate number of models in agreement with majority regarding directionality of variable contribution to model predictions (RF model is excluded due to the Gini score being a non-directional variable importance measure).

in the extent of the plants' elevational range shifts ($R^2 = 0.05 - 0.18$, (Angert et al., 2011))

In addition to reducing MSE, the traits found to be important predictors in machine learning models correspond to those identified in previous analyses and ecological theory. The best performing models for these data (SVR and Kernel Ridge, lowest MSE: Figure 1.1a) identify dispersal-related traits (e.g., the timing, height, and duration of seed shed along with seed size and dispersal mode for the Alpine and European Plants) as the most important variables in predicting range shift magnitude (Figure 1.2a), which is consistent with our expectation and prior work (Angert et al., 2011). In the original analysis of the European Plants data (Rumpf et al., 2018) an indicator of thermal adaptation ("TemperatureIndicator") was a primary predictor of range shifts; our models selected the same variable as most important (Figure 1.2b). The previous Western NA mammal analysis (Angert et al., 2011; Moritz et al., 2008) identified altitudinal limit as a significant predictor and longevity as a relatively strong, but non-significant, predictor. Our models select those two variables as the most important predictors (Figure 1.2c). Both the initial surveys (Pinsky et al., 2013) and our analysis failed to identify strong trait predictors of marine range shifts and non-linear methods yielded less improvement of MSE than the other datasets. However, habitat traits, such as whether species are pelagic, are top predictors consistent with compilations of individual studies (Poloczanska et al., 2013).

The models employed in this analysis also demonstrate cross-model consistency in identifying trait drivers of range shifts (Figure 1.2). In other words, separate models tend to agree on trait rank, especially in the top 5 traits, suggesting a common effect despite significantly different modeling methodologies (and evaluation strategies). However, there tends to be more agreement among machine learning models than between OLS and machine learning models. In particular, some traits for which thresholds seem likely are less important predictors in OLS than in machine learning models: higher seed shed heights and longer seed shed durations in alpine plants may not lead to more dispersal once thresholds are reached (Figure 1.2a).

Models that assign directionality to a modeled effect (all models except Random Forest), are consistent to varying degrees with regard to the directionality of how traits influence range shifts (Figure 1.3). For example, all models for European plants (except Kernel Ridge) suggest that thermophilic species that release seeds higher have shifted their altitudinal distribution higher in elevation. All models (except SVR) find that mammals with greater longevity and higher altitudinal limits exhibit smaller altitudinal range shifts. In both these examples, the exceptional model suggests the opposite relationship for both traits. Agreement is sometimes strongest within regression (OLS, Ridge) and machine learning (Kernel Ridge, SVR) type models. For example, the machine learning models suggest alpine plants with higher seed shed shift their distribution further, contrary to the findings of regression models. These mixed results suggest limits to the predictive capacity of species' traits (see Discussion). It is important to note that these measures of variable contribution are computed in distinct ways depending on the modeling methodology, and it remains to be seen whether these methods of variable importance are suitable to determine effect directionality.

Discussion

We find that non-linear modeling methods enhance the ability of traits to accurately hindcast observed range shifts. Our findings match biological intuition that biological processes, which respond to environmental conditions and are mediated by species' traits, are rarely linear (Stenseth and Mysterud, 2002). Importantly, we have shown that the novel statistical models identify similar predictor traits consistent with known ecological processes. However, disagreements in the directionality of the relationship between trait values and range shift magnitude suggest limits to trait-based statistical prediction frameworks.

Overall, our analysis offers a mixed outlook for using species' traits in applied predictions, such as analyses of climate change vulnerability. The substantially better predictive performance of non-linear models relative to linear models suggests vulnerability analyses frameworks based on species' traits should be adapted to account for non-linearities. However, disagreements in the directionality of trait predictors and the variability of performance improvement suggest that even non-linear methods for relating traits to climate change responses may have limited predictive accuracy. More mechanistic approaches that describe the processes by which traits mediate fitness and demographic responses to the environment may be required for predictions that require high levels of accuracy (Buckley and Kingsolver, 2012; Urban et al., 2016).

Models that allow for non-linearities should be employed to further reevaluate expectations for how traits govern range shifts. A meta-analysis across

range shift studies found at best moderate support for dispersal ability (body size: 22%, migratory strategy: 10%, movement ability: 50% of studies uphold predicted relationship), reproductive potential (fecundity: 36%, longevity: 60%) and ecological generalization (diet breadth: 27%, habitat breadth: 43%) as predictors of range shift magnitude (MacLean and Beissinger, 2017a). The large gap between expectations and observations highlights the predictive challenges posed by the complex interaction of factors driving range shifts and the need for novel predictive methods. Translating spatial range shifts into metrics of environmental niche tracking (e.g., velocity of climate change, (Loarie et al., 2009)) may also enhance predictive capacity.

Reevaluation with non-linear methods will provide insight into selecting appropriate predictor traits. The availability of trait data has increased substantially since some of our datasets were compiled (e.g. Angert et al., 2011), so refining traits may improve predictive capacity. Still, needs for additional trait data addressing issues such as physiology and evolutionary potential are substantial, and will likely require concerted data collection efforts (Urban et al., 2016). Since species' traits are likely to be phylogenetically conserved, phylogenetic signal in range shifts can be used to assess the potential to use traits to predict range shifts. High phylogenetic signal but weak predictive performance of traits would suggest that improving the traits used as predictors can enhance predictive capacity. The initial analyses that accounted for phylogeny found limited phylogenetic signal in range shifts. We did not account for phylogeny because it is not straightforward to do so in the machine learning models. A recent synthesis of range shift studies (Diamond, 2018) found variable but generally weaker phylogenetic signal in range shifts than in physiological, morphological, and life-history traits. The finding indicates limits to the predictive capacity of traits due to factors such as trait evolution, plasticity, and interactions.

Our methodology—flexible statistical models paired with robust evaluation methods and model interrogation approaches—has been reliably employed across many predictive contexts but has yet to be fully embraced by ecologists. The emerging wealth of publicly available ecological and environmental data, combined with the pressing need for reliable ecological forecasts that are useful in decision-making frameworks, makes this flexible and data-intensive approach a natural fit. Despite promising results, our approach presents several challenges to adoption. Of particular relevance to the ecology community is the lack of traditional statistical techniques to evaluate these methods. Following the machine learning community, we employ k-fold cross-validation to lend statistical robustness to the pertinent evaluative criteria for our models (here, mean squared error). In addition, the use of recent advances in model inspection methods (Shapley values from (Lundberg and Lee, 2017)) represents a necessary departure from the manual inspection and testing of linear model coefficients. As the field of model interpretation grows, ecologists can leverage these developments to verify the ecological processes underpinning the predictions of these unconventional modeling approaches. Understanding and acknowledging these shifts in method evaluation and inspection approaches are critical steps to leveraging these more performant statistical modeling approaches in the ecology community.

The general applicability of this approach should be confirmed by studies including additional taxa, a greater number of species, and using novel model interrogation techniques. Testing predictive models using multiple time periods or regions and simulated species is needed as is investigating the biological underpinnings of limited predictive ability (e.g., trait thresholds, interactions, evolution, and plasticity). Such tests should probe the optimal flexibility of predictive models. This investigation should include mechanistic studies of how traits mediate shifts in both warm and cold range edges. Observed trait-environment relationships should inform non-linear models (e.g., in Bayesian approaches). Further consideration of the environment such as the extent of environmental sampling and environmental novelty is warranted (Maguire et al., 2015). In addition, the approach has the potential to improve predictive accuracy in other ecological domains relevant to policy and decisions making (species distribution modeling, forecasting of ecological carbon flux, etc.). However, as is demonstrated by the range of predictive performance improvement across datasets (e.g. between the Alpine Plants and the European Plants datasets, Figure 1.1), both the biological and analytical limitations to predictive modelling must be considered.

Acknowledgements

We thank contributors to TRY, Fishbase, and the other datasets, and we are grateful for input and assistance from Amy Angert, Malin Pinsky, Ray Huey, and particularly Sabine Rumpf. We thank Steve Beissinger and an anonymous reviewer for constructive input. This work was supported by the National Science Foundation [IGERT DGE-1258485 fellowship to A.F.C., a Graduate Research Fellowship to A.F.C, and DBI-1349865 to L.B.B.].

References

Adrian, R., Wilhelm, S., and Gerten, D. (2006). Life-history traits of lake plankton species may govern their phenological response to climate warming. *Global Change Biology*, 12(4):652–661.

Angert, A. L., Crozier, L. G., Rissler, L. J., Gilman, S. E., Tewksbury, J. J., and Chunco, A. J. (2011). Do species' traits predict recent shifts at expanding range edges? *Ecology Letters*, 14(7):677–689.

Bjorkman, A. D., Myers-Smith, I. H., Elmendorf, S. C., Normand, S., Thomas, H. J. D., Alatalo, J. M., Alexander, H., Anadon-Rosell, A., Angers-Blondin, S., Bai, Y., Baruah, G., te Beest, M., Berner, L., Björk, R. G., Blok, D., Bruelheide, H., Buchwal, A., Buras, A., Carbognani, M., Christie, K., Collier, L. S., Cooper, E. J., Cornelissen, J. H. C., Dickinson, K. J. M., Dullinger, S., Elberling, B., Eskelinen, A., Forbes, B. C., Frei, E. R., Iturrate-Garcia, M., Good, M. K., Grau, O., Green, P., Greve, M., Grogan, P., Haider, S., Hájek, T., Hallinger, M., Happonen, K., Harper, K. A., Heijmans, M. M. P. D., Henry, G. H. R., Hermanutz, L., Hewitt, R. E., Hollister, R. D., Hudson, J., Hülber, K., Iversen, C. M., Jaroszynska, F., Jiménez-Alfaro, B., Johnstone, J., Jorgensen, R. H., Kaarlejärvi, E., Klady, R., Klimešová, J., Korsten, A., Kuleza, S., Kulonen, A., Lamarque, L. J., Lantz, T., Lavalle, A., Lembrechts, J. J., Lévesque, E., Little, C. J., Luoto, M., Macek, P., Mack, M. C., Mathakutha, R., Michelsen, A., Milbau, A., Molau, U., Morgan, J. W., Mörsdorf, M. A., Nabe-Nielsen, J., Nielsen, S. S., Ninot, J. M., Oberbauer, S. F., Olofsson, J., Onipchenko, V. G., Petraglia, A., Pickering, C., Prevéy, J. S., Rixen, C., Rumpf, S. B., Schaepman-Strub, G., Semenchuk, P., Shetti, R., Soudzilovskaia, N. A., Spasojevic, M. J., Speed, J. D. M., Street, L. E., Suding, K., Tape, K. D., Tomaselli, M., Trant, A., Treier, U. A., Tremblay, J.-P., Tremblay, M., Venn, S., Virkkala, A.-M., Vowles, T., Weijers, S., Wilmking, M., Wipf, S., and Zamin, T. (2018). Tundra Trait Team: A database of plant traits spanning the tundra biome. *Global Ecology and Biogeography*, 27(12):1402–1411.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Buckley, L. B. and Kingsolver, J. G. (2012). Functional and Phylogenetic Approaches to Forecasting Species' Responses to Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):205–226.

Diamond, S. E. (2018). Contemporary climate-driven range shifts: Putting evolution back on the table. *Functional Ecology*, 32(7):1652–1665.

Diamond, S. E., Frame, A. M., Martin, R. A., and Buckley, L. B. (2011). Species' traits predict phenological responses to climate change in butterflies. *Ecology*, 92(5):1005–1012.

Estrada, A., Morales-Castilla, I., Caplat, P., and Early, R. (2016). Usefulness of Species Traits in Predicting Range Shifts. *Trends in Ecology & Evolution*, 31(3):190–203.

Fitt, R. N. L., Palmer, S., Hand, C., Travis, J. M. J., and Lancaster, L. T. (2018). Towards an interactive, process-based approach to understanding range shifts: developmental and environmental dependencies matter. *Ecography*, 0(0).

Foden, W. B., Butchart, S. H. M., Stuart, S. N., Vié, J.-C., Akçakaya, H. R., Angulo, A., DeVantier, L. M., Gutsche, A., Turak, E., Cao, L., Donner, S. D., Katariya, V., Bernard, R., Holland, R. A., Hughes, A. F., O'Hanlon, S. E., Garnett, S. T., Şekercioğlu, Ç. H., and Mace, G. M. (2013). Identifying the World's Most Climate Change Vulnerable Species: A Systematic Trait-Based Assessment of all Birds, Amphibians and Corals. *PLOS ONE*, 8(6):1–13.

Froese, R. and Pauly, D. (2010). FishBase.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, volume 27 of *Springer Series in Statistics*. Springer New York, New York, NY.

Holzinger, B., Hülber, K., Camenisch, M., and Grabherr, G. (2008). Changes in plant species richness over the last century in the eastern Swiss Alps: elevational gradient, bedrock effects and migration rates. *Plant Ecology*, 195(2):179–196.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., Cornelissen, J. H. C., Violle, C., Harrison, S. P., Van Bodegom, P. M., Reichstein, M., Enquist, B. J., Soudzilovskaia, N. A., Ackerly, D. D., Anand, M., Atkin, O., Bahn, M., Baker, T. R., Baldocchi, D., Bekker, R., Blanco, C. C., Blonder, B., Bond, W. J., Bradstock, R., Bunker, D. E., Casanoves, F., Cavender-bares, J., Chambers, J. Q., Chapin Iii, F. S., Chave, J., Coomes, D., Cornwell, W. K., Craine, J. M., Dobrin, B. H., Duarte, L., Durka, W., Elser, J., Esser, G., Estiarte, M., Fagan, W. F., Fang, J., Fernández-méndez, F., Fidelis, A., Finegan, B., Flores, O., Ford, H., Frank, D., Freschet, G. T., Fyllas, N. M., Gallagher, R. V., Green, W. A., Gutierrez, a. G., Hickler, T., Higgins, S. I., Hodgson, J. G., Jalili, A., Jansen, S., Joly, C. A., Kerkhoff, a. J., Kirkup, D., Kitajima, K., Kleyer, M., Klotz, S., Knops, J. M. H., Kramer, K., Kühn, I., Kurokawa, H., Laughlin, D., Lee, T. D., Leishman, M., Lens, F., Lenz, T., Lewis, S. L., Lloyd, J., Llusià, J., Louault, F., Ma, S., Mahecha, M. D., Manning, P., Massad, T., Medlyn, B. E., Messier, J., Moles, a. T., Müller, S. C., Nadrowski, K., Naeem, S., Niinemets, Ü., Nöllert, S., Nüske, A., Ogaya, R., Oleksyn, J., Onipchenko, V. G., Onoda, Y., Ordoñez, J., Overbeck, G., Ozinga, W. A., Patiño, S., Paula, S., Pausas, J. G., Peñuelas, J., Phillips, O. L., Pillar, V., Poorter, H., Poorter, L., Poschlod, P., Prinzing, A.,

Proulx, R., Rammig, A., Reinsch, S., Reu, B., Sack, L., Salgado-negret, B., Sardans, J., Shiodera, S., Shipley, B., Siefert, A., Sosinski, E., Soussana, J.-f., Swaine, E., Swenson, N., Thompson, K., Thornton, P., Waldram, M., Weiher, E., White, M., White, S., Wright, S. J., Yguel, B., Zaehle, S., Zanne, a. E., and Wirth, C. (2011). TRY - a global database of plant traits. *Global Change Biology*, 17(9):2905–2935.

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., and Ackerly, D. D. (2009). The velocity of climate change. *Nature*, 462(7276):1052–1055.

Lundberg, S. M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.

MacLean, S. A. and Beissinger, S. R. (2017a). Species' traits as predictors of range shifts under contemporary climate change: A review and meta-analysis. *Global Change Biology*, 23(10):4094–4105.

MacLean, S. A. and Beissinger, S. R. (2017b). Species' traits as predictors of range shifts under contemporary climate change: A review and meta-analysis. *Global Change Biology*, 23(10):4094–4105.

Maguire, K. C., Nieto-Lugilde, D., Fitzpatrick, M. C., Williams, J. W., and Blois, J. L. (2015). Modeling Species and Community Responses to Past, Present, and Future Episodes of Climatic and Ecological Change. *Annual Review of Ecology, Evolution, and Systematics*, 46(1):343–368.

Moritz, C., Patton, J. L., Conroy, C. J., Parra, J. L., White, G. C., and Beissinger, S. R. (2008). Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA. *Science*, 322(5899):261–264.

Olden, J. D., Lawler, J. J., and Poff, N. L. (2008). Machine Learning Methods Without Tears: A Primer for Ecologists. *The Quarterly Review of Biology*, 83(2):171–193.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):637–669.

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., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine Taxa Track Local Climate Velocities. *Science*, 341(6151):1239–1242.

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., Brander, K., Bruno, J. F., Buckley, L. B., Burrows, M. T., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F., Thompson, S. A., and Richardson, A. J. (2013). Global imprint of climate change on marine life. *Nature Climate Change*, 3:919.

Rapacciuolo, G., Maher, S. P., Schneider, A. C., Hammond, T. T., Jabis, M. D., Walsh, R. E., Iknayan, K. J., Walden, G. K., Oldfather, M. F., Ackerly, D. D., and Beissinger, S. R. (2014). Beyond a warming fingerprint: individualistic biogeographic responses to heterogeneous climate change in California. *Global Change Biology*, 20(9):2841–2855.

Rumpf, S. B., Hülber, K., Klonner, G., Moser, D., Schütz, M., Wessely, J., Willner, W., Zimmermann, N. E., and Dullinger, S. (2018). Range dynamics of mountain plants decrease with elevation. *Proceedings of the National Academy of Sciences*, 115(8):1848–1853.

Schloss, C. A., Nuñez, T. A., and Lawler, J. J. (2012). Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences*, 109(22):8606–8611.

Stenseth, N. C. and Mysterud, A. (2002). Climate, changing phenology, and other life history traits: Nonlinearity and match-mismatch to the environment. *Proceedings of the National Academy of Sciences*, 99(21):13379–13381.

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe'er, G., Singer, A., Bridle, J. R., Crozier, L. G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J. J., Holt, R. D., Huth, A., Johst, K., Krug, C. B., Leadley, P. W., Palmer, S. C. F., Pantel, J. H., Schmitz, A., Zollner, P. A., and Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, 353(6304).

Wheatley, C. J., Beale, C. M., Bradbury, R. B., Pearce-Higgins, J. W., Critchlow, R., and Thomas, C. D. (2017). Climate change vulnerability for species-Assessing the assessments. *Global Change Biology*, 23(9):3704–3715.

Appendix 1.1

Name	Type	Description
StorageOrgan	n/a	Storage Organ Presence
ReprModeCode	n/a	Reproductive Mode
flwr_mo_start	phenology	Flowering Start (month)
flwr_mo_end	phenology	Flowering End (month)
flwr_dur_mos	phenology	Flowering Duration (months)
earliest_seed_shed_mo	phenology	Seed Shed Month (earliest)
latest_seed_shed_mo	phenology	Seed Shed Month (latest)
seed_shed_dur_mos	phenology	Seed Shed Duration (months)
SI01_1.0	n/a	Self Compatibility: No
SI01_0.0	n/a	Self Compatibility: Yes
seed_mass_mg	dispersal	Seed Mass (mg)
diaspore_mass_mg	dispersal	Diaspore Mass (mg)
diaspore_min_len_mm	dispersal	Diaspore Minimum Length (mm)
Ave_seed_shed_ht_m	dispersal	Average Seed Shed Height (m)
diaspore_ave_len_mm	dispersal	Average Diaspore Length (mm)
Min_seed_shed_ht_m	dispersal	Minimum Seed Shed Height (m)
Max_seed_shed_ht_m	dispersal	Maximum Seed Shed Height (m)
Ave_seed_shed_ht_m	dispersal	Average Seed Shed Height (m)
nichebreadth_num_flor_zones	ecological	Total Number of Floristic Zones
nichebreadth_amplit_ocean	ecological	Total Number of Oceanic Zones
Bio1_mean_nosyn	climatic	BioClim: Bio1 (mean)
Bio1_std_nosyn	climatic	BioClim: Bio1 (std dev)
Bio1_var_nosyn	climatic	BioClim: Bio1 (var)
Bio1_mean_inclsyn	climatic	BioClim: Bio1 (mean)
Bio1_std_inclsyn	climatic	BioClim: Bio1 (std dev)
Bio1_var_inclsyn	climatic	BioClim: Bio1 (var)
Nbound_lat_GBIF_nosyn	ecological	Northernmost Latitude (GBIF)
GenTime	n/a	Generation Time
Resil	Unknown	Unknown
MaxAlt	ecological	Maximum Altitude
oceanity_ks	ecological	Oceanity: Continental - Oceanic
oceanity_o	ecological	Oceanity: Oceanic
oceanity_os	ecological	Oceanity: Oceanic - Continental
oceanity_sks	ecological	Oceanity: Moderate Continental
oceanity_so	ecological	Oceanity: Moderate Both
oceanity_sos	ecological	Oceanity: Moderate Oceanic
dispersal_mode_animal	dispersal	Animal Dispersal Mode (binary)
dispersal_mode_gravity	dispersal	Gravity Dispersal Mode (binary)
dispersal_mode_water	dispersal	Water Dispersal Mode (binary)
dispersal_mode_wind	dispersal	Wind Dispersal Mode (binary)
BreedSysCode_1.0	dispersal	Breeding: Obligate Outcrossing
BreedSysCode_2.0	dispersal	Breeding: Mostly Outcrossing, Selfing Possible
BreedSysCode_3.0	dispersal	Breeding: Mixed Mating I
BreedSysCode_4.0	dispersal	Breeding: Mixed Mating II

Table 1A.1: Traits of Alpine Plants. Data from Angert et al, 2011 via the BiolFlor Database (Kühn et al, 2004).

Name	Type	Description
Orig_high_limit	ecological	Altitudinal Limit
Litter_size	reproductive	Litter Size (individuals)
Litters_per_yr	reproductive	Litters / year
Young_per_yr	reproductive	Young / year
Mass_g	physical	Mass (g)
Longevity_yrs	physical	Longevity (years)
Bio1_mean	climatic	BioClim: Bio 1 (mean)
Bio1_std	climatic	BioClim: Bio 1 (std dev)
Rangesize_km2	ecological	Range Size (km2)
Food_code	ecological	Food Code: (0: insectivore/herbivore/carnivore, 1: omnivore)
Daily01_0	physical	Daily Rhythm Data – Absent
Daily01_1	physical	Daily Rhythm Data – Present
Annual01_0	physical	Annual Rhythm Data – Absent
Annual01_1	physical	Annual Rhythm Data – Present
Food01_0	physical	Food Preference Data – Absent
Food01_1	physical	Food Preference Data – Present
Daily_rhythm_both	physical	Daily Rhythm – Diurnal + Nocturnal (binary)
Daily_rhythm_diur	physical	Daily Rhythm – Diurnal (binary)
Daily_rhythm_noct	physical	Daily Rhythm – Nocturnal (binary)
Annual_rhythm_fachib	physical	Annual Rhythm – Facultative Hibernation (binary)
Annual_rhythm_nonhib	physical	Annual Rhythm – No Hibernation (binary)
Annual_rhythm_oblhib	physical	Annual Rhythm – Obligate Hibernation (binary)
Food_gran	ecological	Food – Granivore (binary)
Food_herb	ecological	Food – Herivore (binary)
Food_inse	ecological	Food – Insectivore (binary)
Food_omni	ecological	Food – Omnivore (binary)

Table 1A.2: Traits of Mammals. From Angert et al. 2011.

Name	Type	Description
RelativeAbundance	ecological	Species Relative Abundance
TemperatureIndicator	climatic	Temperature Indicator: 1, alpine to nival; 1.5, lower alpine to upper subalpine; 2, subalpine; 2.5, lower subalpine to upper montane; 3, montane; 3.5, lower montane to upper colline; 4, colline.
NutrientIndicator	ecological	Nutrient indicator: 1, very nutrient-poor; 2, nutrient-poor; 3, moderately nutrient-poor to moderately nutrient-rich; 4, nutrient-rich
TerminalVelocity	physical	Shift Velocity
RetInFurCattle	physical	Fur Seed Retention: Cattle
RetInFurSheep	physical	Fur Seed Retention: Sheep
GutSurvival	physical	Seed Gut Survival
SeedReleaseHeight	physical	Seed Release Height
LifeStrategy	ecological	Life Strategy (C-S-R)
LifeSpan	ecological	LifeSpan (yr)
Dominance	ecological	Dominance Index
NoOfVegOffspings	physical	Number Vegetative Offsprings / year
Persistence	physical	Persistence Index
Dispersal	physical	Dispersal Index

Table 1A.3: Traits of European Plants data. From Rumpf et al. 2018. Additional citations in Appendix 1.2.

Name	Type	Description
DepthRangeDeep	ecological	Depth Range
Vulnerability	ecological	FishBase Vulnerability
Length	physical	Length
UsedforAquaculture_commercial	economic	Aquaculture: Commercial Species
UsedforAquaculture_likely future use	economic	Aquaculture: Future Likely
UsedforAquaculture_never/rarely	economic	Aquaculture: Never/Rarely
Electrogenic_Electrosensing only	physical	Electrosensing Only
Electrogenic_no special ability	physical	No Electrosensing
Electrogenic_strongly discharging	physical	Strong Electric Discharge
Electrogenic_weakly discharging	physical	Weak Electric Discharge
DemersPelag_bathydemersal	ecological	Bathydemersal
DemersPelag_benthopelagic	ecological	Benthopelagic
DemersPelag_demersal	ecological	Demersal
DemersPelag_pelagic-neritic	ecological	Pelagic-Neritic
DemersPelag_pelagic-oceanic	ecological	Pelagic-Oceanic
DemersPelag_reef-associated	ecological	Reef-Associated
LTypeMaxM_FL	Physical	Length Type – FL
LTypeMaxM_SL	Physical	Length Type – SL
LTypeMaxM_TL	Physical	Length Type – TL
Fresh_-1	ecological	Freshwater: Present
Fresh_0	ecological	Freshwater: Not Present
Brack_-1	ecological	Brackish: Present
Brack_0	ecological	Brackish: Not Present
Saltwater_-1	ecological	Saltwater: Present

Table 1A.4: Traits of marine organisms from FishBase (Pauly and Froese, 2010). Additional citations in Appendix 1.2

Appendix 1.2

Trait References: TRY Database

Traits referenced from the TRY database, as originally compiled by Rumpf et al., 2018.

Campetella, G., Botta-Dukát, Z., Wellstein, C., Canullo, R., Gatto, S., Chelli, S., ... Bartha, S. (2011). Patterns of plant trait–environment relationships along a forest succession chronosequence. Agriculture, Ecosystems & Environment, 145(1), 38–48.

Cerabolini, B., Ceriani, R. M., Caccianiga, M., De Andreis, R., & Raimondi, B. (2003). Seed size, shape and persistence in soil: a test on Italian flora from Alps to Mediterranean coasts. Seed Science Research, 13(1), 75–85.

Choat, B., Jansen, S., Brodribb, T. J., Cochard, H., Delzon, S., Bhaskar, R., ... Hacke, U. G. (2012). Global convergence in the vulnerability of forests to drought. Nature, 491(7426), 752.

Ciocorlan, V. (n.d.). The illustrated flora of Romania, Pteridophyta et Spermatophyta. The reviewed second edition. Ceres Bucuresti Editure, 14–30.

Cornelissen, J. H. C. (1996). An experimental comparison of leaf decomposition rates in a wide range of temperate plant species and types. Journal of Ecology, 573–582.

Cornelissen, J. H. C., Cerabolini, B., Castro-Díez, P., Villar-Salvador, P., Montserrat-Martí, G., Puyravaud, J. P., ... Aerts, R. (2003). Functional traits of woody plants: correspondence of species rankings between field adults and laboratory-grown seedlings? Journal of Vegetation Science, 14(3), 311–322.

Cornelissen, J. H. C., Diez, P. C., & Hunt, R. (1996). Seedling growth, allocation and leaf attributes in a wide range of woody plant species and types. Journal of Ecology, 755–765.

Cornelissen, J. H. C., Quested, H. M., Gwynn-Jones, D., Van Logtestijn, R. S. P., De Beus, M. A. H., Kondratchuk, A., ... Aerts, R. (2004). Leaf digestibility and litter decomposability are related in a wide range of subarctic plant species and types. Functional Ecology, 18(6), 779–786.

Dainese, M., & Bragazza, L. (2012). Plant traits across different habitats of the Italian Alps: a comparative analysis between native and alien species. Alpine Botany, 122(1), 11–21. Diaz, S., Hodgson, J. G., Thompson, K., Cabido, M., Cornelissen, J. H. C. 3, Jalili, A., ... Asri, Y. (2004). The plant traits that drive ecosystems: evidence from three continents. Journal of Vegetation Science, 15(3), 295–304.

Everwand, G., Fry, E. L., Eggers, T., & Manning, P. (2014). Seasonal variation in the capacity for plant trait measures to predict grassland carbon and water fluxes. Ecosystems, 17(6), 1095–1108.

Fischer, M. A., Adler, W., & Oswald, K. (2005). Exkursionsflora für Österreich, Liechtenstein und Südtirol: Bestimmungsbuch für alle in der Republik Österreich, in der Autonomen Provinz Bozen/Südtirol (Italien) und im Fürstentum Liechtenstein wildwachsenden sowie die wichtigsten kultivierten Gefässpflanzen (Farnpflanzen und Samenpflanzen) mit Angaben über ihre Ökologie und Verbreitung. Land Oberösterreich.

Fitter, A. H., & Peat, H. J. (1994). The ecological flora database. Journal of Ecology, 82(2), 415–425.

Fry, E. L., Power, S. A., & Manning, P. (2014). Trait-based classification and manipulation of plant functional groups for biodiversity–ecosystem function experiments. Journal of Vegetation Science, 25(1), 248–261.

Gachet, S., Errol, V., & Tatoni, T. (2005). BASECO: a floristic and ecological database of Mediterranean French flora. Biodiversity & Conservation, 14(4), 1023–1034.

Garnier, E., Lavorel, S., Ansquer, P., Castro, H., Cruz, P., Dolezal, J., ... Golodets, C. (2006). Assessing the effects of land-use change on plant traits, communities and ecosystem functioning in grasslands: a standardized methodology and lessons from an application to 11 European sites. Annals of Botany, 99(5), 967–985.

Green, W. (2009). USDA PLANTS compilation, version 1. NRCS: The PLANTS Database.

Hegi, G. (2008). Illustrierte Flora von Mitteleuropa. Jena, Germany: Weissdorn.

Henrik Bruun, H., Österdahl, S., Moen, J., & Angerbjörn, A. (2005). Distinct patterns in alpine vegetation around dens of the Arctic fox. Ecography, 28(1), 81–87.

Hickler, T. (1999). Plant functional types and community characteristics along environmental gradients on Oland's Great Alvar (Sweden). Sweden: University of Lund.

Hill, M. O., Preston, C. D., & Roy, D. B. (2004). PLANTATT-attributes of British and Irish plants: status, size, life history, geography and habitats. Centre for Ecology & Hydrology.

Hintze, C., Heydel, F., Hoppe, C., Cunze, S., König, A., & Tackenberg, O. (2013). D3: the dispersal and diaspore database–baseline data and statistics on seed dispersal. Perspectives in Plant Ecology, Evolution and Systematics, 15(3), 180–192.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., ... Wirth, C. (2011). TRY - a global database of plant traits. Global Change Biology, 17(9), 2905–2935. http://doi.org/10.1111/j.1365-2486.2011.02451.x

Kew, R. B. G. (2008). Seed information database (SID). Version 7.1.

Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M., ... Klimešová, J. (2008). The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology, 96(6), 1266–1274.

Klimešová, J., & De Bello, F. (2009). CLO-PLA: the database of clonal and bud bank traits of Central European flora §. Journal of Vegetation Science, 20(3), 511–516.

Kühn, I., Durka, W., & Klotz, S. (2004). BiolFlor — a new plant-trait database as a tool for plant invasion ecology. Diversity and Distributions, 10(5-6), 363–365. http://doi.org/10.1111/j.1366-9516.2004.00106.x

Kutschera, L., & Lichtenegger, E. (1982). Wurzelatlas mitteleuropäischer Grünlandpflanzen (Vol. 1). Stuttgart: Gustav Fischer.

Landolt, E., Bäumler, B., Erhardt, A., Hegg, O., Klötzli, F., Lämmler, W., ... Theurillat, J.-P. (2010). Flora indicativa= Ecological inicator values and biological attributes of the flora of Switzerland and the Alps: ökologische Zeigerwerte und biologische Kennzeichen zur Flora der Schweiz und der Alpen. Haupt Verlag.

Medlyn, B. E., Barton, C. V. M., Broadmeadow, M. S. J., Ceulemans, R., De Angelis, P., Forstreuter, M., ... Laitat, E. (2001). Stomatal conductance of forest species after long-term exposure to elevated CO2 concentration: a synthesis. New Phytologist, 149(2), 247–264.

Milla, R., & Reich, P. B. (2011). Multi-trait interactions, not phylogeny, fine-tune leaf size reduction with increasing altitude. Annals of Botany, 107(3), 455–465.

Minden, V., Andratschke, S., Spalke, J., Timmermann, H., & Kleyer, M. (2012). Plant trait–environment relationships in salt marshes: Deviations from predictions by ecological concepts. Perspectives in Plant Ecology, Evolution and Systematics, 14(3), 183–192. http://doi.org/https://doi.org/10.1016/j.ppees.2012.01.002

Moretti, M., & Legg, C. (2009). Combining plant and animal traits to assess community functional responses to disturbance. Ecography, 32(2), 299–309.

Müller-Schneider, P. (1986). Verbreitungsbiologie der Blütenpflanzen Graubündens (Vol. 85). Geobotanisches Institut der ETH, Stiftung Rübel.

Ordonez, J. C., van Bodegom, P. M., Witte, J.-P. M., Bartholomeus, R. P., van Hal, J. R., & Aerts, R. (2009). Plant strategies in relation to resource supply in mesic to wet environments: does theory mirror nature? The American Naturalist, 175(2), 225–239.

Paula, S., Arianoutsou, M., Kazanis, D., Tavsanoglu, Ç., Lloret, F., Buhk, C., ... Rodrigo, A. (2009). Fire-related traits for plant species of the Mediterranean Basin. Ecology, 90(5), 1420.

Pluess, A. R., Schütz, W., & Stöcklin, J. (2005). Seed weight increases with altitude in the Swiss Alps between related species but not among populations of individual species. Oecologia, 144(1), 55–61.

Prentice, I. C., Meng, T., Wang, H., Harrison, S. P., Ni, J., & Wang, G. (2011). Evidence of a universal scaling relationship for leaf CO2 drawdown along an aridity gradient. New Phytologist, 190(1), 169–180.

Quested, H. M., Cornelissen, J. H. C., Press, M. C., Callaghan, T. V, Aerts, R., Trosien, F., ... Jonasson, S. E. (2003). Decomposition of sub-arctic plants with differing nitrogen economies: a functional role for hemiparasites. Ecology, 84(12), 3209–3221.

Römermann, C., Tackenberg, O., & Poschlod, P. (2005). How to predict attachment potential of seeds to sheep and cattle coat from simple morphological seed traits. Oikos, 110(2), 219–230.

Rothmaler, W. (2002). Exkursionsflora von Deutschland (9th ed.). Heidelberg: Spektrum Akademischer.

Schröter, C., Brockmann-Jerosch, M. C., Günthart, A., & Vogler, P. (1926). Das Pflanzenleben der Alpen: eine schilderung der hochgebirgsflora. A. Raustein.

Schweingruber, F. H., & Landolt, W. (2005). The xylem database. Swiss Federal Research Institute WSL Updated.

Spasojevic, M. J., & Suding, K. N. (2012). Inferring community assembly mechanisms from functional diversity patterns: the importance of multiple assembly processes. Journal of Ecology, 100(3), 652–661.

Tackenberg, O. (2001). Methoden zur Bewertung gradueller unterschiede des Ausbreitungspotentials von Pflanzenarten.

Tamme, R., Götzenberger, L., Zobel, M., Bullock, J. M., Hooftman, D. A. P., Kaasik, A., & Pärtel, M. (2014). Predicting species' maximum dispersal distances from simple plant traits. Ecology, 95(2), 505–513.

Thompson, K., Band, S. R., & Hodgson, J. G. (1993). Seed size and shape predict persistence in soil. Functional Ecology, 236–241.

Wirth, C., & Lichstein, J. W. (2009). The Imprint of Species Turnover on Old-Growth Forest Carbon Balances - Insights From a Trait-Based Model of Forest Dynamics BT - Old-Growth Forests: Function, Fate and Value. In C. Wirth, G. Gleixner, & M. Heimann (Eds.), (pp. 81–113). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-92706-8_5

Wohlfahrt, G., Bahn, M., Haubner, E., Horak, I., Michaeler, W., Rottmar, K., ... Cernusca, A. (1999). Interspecific variation of the biochemical limitation to photosynthesis and related leaf traits of 30 species from mountain grassland ecosystems under different land use. Plant, Cell & Environment, 22(10), 1281–1296.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., ... Villar, R. (2004). The worldwide leaf economics spectrum. Nature, 428, 821. Retrieved from https://doi.org/10.1038/nature02403

Trait References: FishBase

References of traits for species in FishBase, supplementing Pinsky et al., 2013.

Allen, M. J., & Smith, G. B. (1988). Atlas and zoogeography of common fishes in the Bering Sea and northeastern Pacific.

Clemens, W. A., & Wilby, G. V. (1961). Fishes of the Pacific coast of Canada (Vol. 68). Fisheries Research Board of Canada Ottawa.

Cohen, D. M., Inada, T., Iwamoto, T., & Scialabba, N. (1990). Gadiform fishes of the world. FAO Fisheries Synopsis, 10(125), I.

Compagno, L. J. V. (1984). Sharks of the world: an annotated and illustrated catalogue of shark species known to date. FAO Fisheries Synopsis.

Cooper, J. A. (1996). Monophyly and intrarelationships of the family Pleuronectidae (Pleuronectiformes), with a revised classification. University of Ottawa (Canada).

Eschmeyer, W. N., & Herald, E. S. (1999). A field guide to Pacific coast fishes: North America. Houghton Mifflin Harcourt.

Lamb, A., & Edgell, P. (1986). Coastal fishes of the Pacific Northwest. Madeira Park, BC: Harbour Publishing Company.

McEachran, J. D., & Dunn, K. A. (1998). Phylogenetic analysis of skates, a morphologically conservative clade of elasmobranchs (Chondrichthyes: Rajidae). Copeia, 271–290.

Nature, C. M. of, Labignan, I., & Waszczuk, H. (1995). Encyclopedia of Canadian fishes. Waterdown Ont.: Canadian Sportfishing Productions.

Nielsen, J. G., Cohen, D. M., Markle, D. F., & Robins, C. R. (1999). FAO species catalogue, v. 18: Ophidiiform fishes of the world (Order Ophidiiformes). An annotated and illustrated catalogue of pearlfishes, cusk-eels, brotulas and other ophidiiform fishes known to date. FAO Fisheries Synopsis.

Page, L. M., & Burr, B. M. (2011). A field guide to freshwater fishes: North America north of Mexico. Houghton Mifflin Harcourt.

Whitehead, P., Nelson, G., & Wongratana, T. (1985). FAO species catalogue. Vol. 7-Clupeiod fishes of the world. An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shad, anchovies and wolf herring. Part 1-Chirocentridae Clupleidae and Pristigasteridae. FAO Fisheries Synopsis, 125.

Chapter 2

High-Resolution CubeSat Imagery and Machine Learning for Detailed Snow-Covered Area

Anthony F. Cannistra, David Shean, Nicoleta Cristea

FOR SUBMISSION TO *Remote Sensing of Environment*

Abstract

Snow cover affects a diverse array of physical, ecological, and societal systems. As such, the development of optical remote sensing techniques to measure snow-covered area (SCA) has enabled progress in a wide variety of research domains. However, in many cases, the spatial and temporal resolutions of currently available remotely sensed SCA products are insufficient to capture SCA evolution at spatial and temporal resolutions relevant to the study of fine-scale spatially heterogeneous phenomena. We developed a convolutional neural network-based method to identify snow covered area using a ~3 meter, ~daily, 4-band optical satellite image dataset with near-global extent from Planet Labs, Inc., a commercial CubeSat developer. By comparing our model performance to high-resolution airborne lidar measurements and other satellite platforms in two North American sites (Sierra Nevada, CA, USA and Rocky Mountains, CO, USA), we show that these emerging image archives have great potential to accurately observe snow-covered area at high spatial and temporal resolutions despite limited radiometric bandwidth and band placement. We achieve average snow classification F-Scores of 0.73 in our training basin and 0.67 in a climatically-distinct out-of-sample basin, suggesting opportunities for model transferability. We also demonstrate shortcomings of these data in forested regions, suggesting avenues for further research. The unparalleled spatial and temporal coverage of CubeSat imagery offers an excellent opportunity for satellite remote sensing of snow, with real implications for ecological and water resource applications.

Introduction

Snow cover is a fundamental planetary surface characteristic that affects a wide range of physical, ecological, and societal systems. The development of optical remote sensing techniques to measure snow-covered area (SCA) over time has enabled progress in a wide variety of research domains. Satellite observations of SCA from sensors such as Landsat Thematic Mapper (TM)/Operational Land Imager (OLI), MODIS Terra, and Sentinel-2 Multispectral Instrument (MSI) have applications in hydrology and water resource management (Andreadis and Lettenmaier, 2006; Schattan et al., 2020), atmospheric science (Fernandes et al., 2009; Painter et al., 2009), ecology (Boelman et al., 2019; Carlson et al., 2015; Dedieu et al., 2016) and in climate adaptation and mitigation (Immerzeel et al., 2009).

In many cases, the spatial and temporal resolutions of currently available remotely sensed SCA products are insufficient to capture SCA evolution at spatial and temporal resolutions relevant to the study of fine-scale spatially heterogeneous phenomena. For example, to study the spatial heterogeneity of snow for accurate runoff prediction in hydrologic modeling, SCA data with high spatial reso-

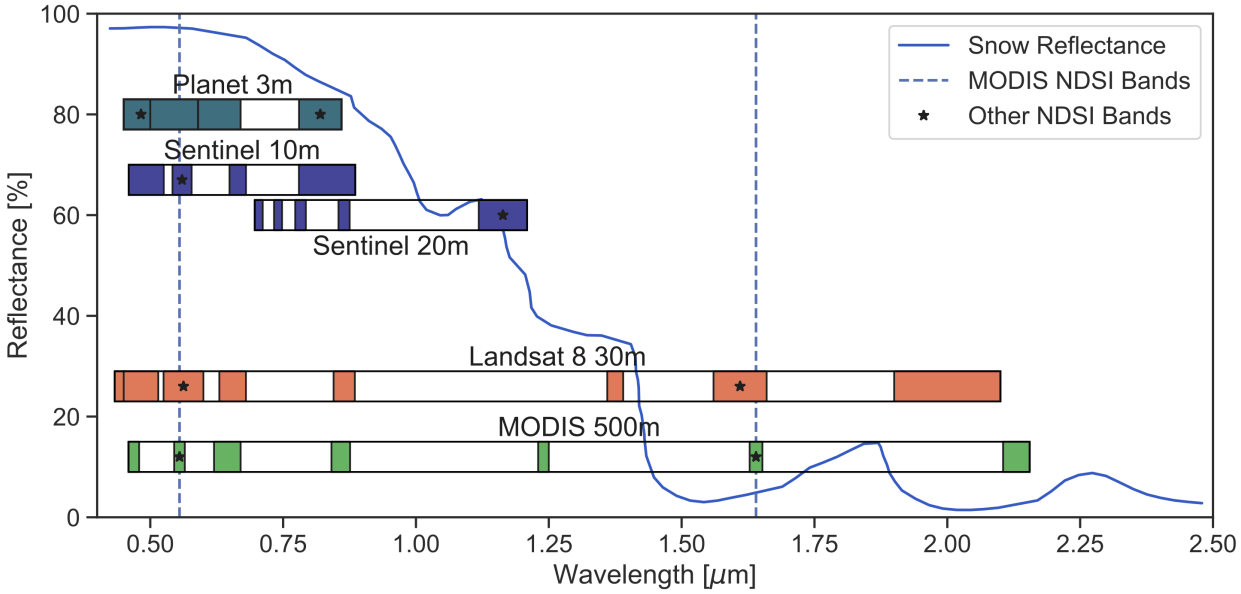

Figure 2.1: Comparison of bands for Earth Observing satellites and reflectance of snow. Planet Labs Inc. PlanetScope (cyan) observes a limited region of the snow reflectance spectrum (from Painter et al., 2009; solid blue line) compared to other optical sensors. Sentinel-2 (dark blue), Landsat 8 (orange), and MODIS (green) platforms observe a broad range of the visible, NIR and SWIR spectrum and thus are able to more readily differentiate snow using radiometric indices such as NDSI (e.g. Hall and Riggs (2007)). Bands used to compute NDSI for each observing platform are marked with asterisks, and MODIS bands are highlighted with dotted blue lines.

lution are needed (Clark et al., 2011; Luce et al., 1999; Lundquist and Dettinger, 2005). In mountain ecosystems, the composition of plant communities and associated phenological events such as flowering and growth vary as a function of snow cover extent and snow duration (Choler, 2005; Ford et al., 2013; Theobald et al., 2017; Venn et al., 2011), which exhibit significant heterogeneity over small spatial and temporal scales (~1-10 m; hours to days during ablation) (Little et al., 1994; Rochefort et al., 1994).

Airborne lidar surveys (ALS) surveys can provide snow observations at fine spatial resolution (e.g., 3 m, Airborne Snow Observatory; Painter et al. (2016)), but are limited to a relatively small spatial extent and are costly to implement, leading to sparse temporal coverage. Comparatively, existing satellite platforms are able to observe snow at a much broader spatial extent but at coarser spatial resolution and variable temporal resolution. Instruments carried by orbiting platforms like Sentinel-2 and Landsat-8, for example, can be used to generate snow observation products (e.g., Normalized Difference Snow Index (NDSI), **?** or fractional snow-covered area [fSCA](Painter et al., 2009)) at 10 or 30 meter resolution, respectively. These platforms have mid-latitude revisit times of between 5 and 16 days. Despite these opportunities, a planetary-scale observational SCA dataset with both high spatial (< 5-10 m) resolution and frequent temporal (~daily) resolution does not yet exist.

Emerging constellations of small imaging satellites (CubeSats) can potentially bridge the temporal and spatial resolution gap between ALS surveys and satellite observations, but methods to derive SCA from these platforms have yet to be studied in detail. Planet Labs, Inc., a CubeSat developer/vendor in California, USA, operates the PlanetScope constellation of over 150 small (3U CubeSat form factor) optical imaging satellites. This constellation offers 4-band multispectral imagery (red [R], green [G], blue [B] and near-infrared [NIR]) at ~3-5 meter resolution with ~daily revisit times across most of the earth's land surface (±81.5 degrees latitude), amounting to a cumulative observational footprint of over 200M km2 day-1 (Planet Labs, Inc., 2019b). However, the spectral bandwidth and available calibration options for these instruments limits the ability of traditional radiometric or spectral analyses for snow. In particular, the visible and shortwave infrared (SWIR) bands typically used for computation of NDSI (e.g. MODIS Terra bands 4 [545 nm – 565 nm] and 6 [1628 nm – 1652 nm] (Riggs and Hall, 2015); Landsat 8 TM bands 2 [450 nm – 515 nm] and 5 [845 nm – 885 nm] (Dozier, 1989)) are not measured by satellites in the

PlanetScope constellation (Figure 2.1).

Despite radiometric quality limitations, the PlanetScope imagery still provides useful snow observations that can be extracted using modern computer vision techniques. Motivated by methodological developments in machine learning on images that have been successful in other domains (e.g. biomedical imaging, Ronneberger et al. (2015)), remote sensing researchers are now applying these techniques to satellite imagery (e.g. for the identification of human structures (Iglovikov et al., 2018) or glacier calving fronts (Mohajerani et al., 2019)). Here we focus on developing and evaluating a convolutional neural network-based approach to enable the delineation of snow-covered area in high-resolution PlanetScope imagery. We use high-resolution, airborne lidar-derived snow cover data to serve as "ground truth" for co-located and contemporaneous Planet imagery, providing a labeled dataset for neural network model training. We train and evaluate the model for well-studied snow sites, and assess both the absolute model performance compared to the ALS ground truth and the relative model performance across particular variables of interest. Here, we develop a convolutional neural network (CNN) approach to derive snow-covered area from the 4-band PlanetScope imagery. We use high-resolution airborne lidar surveys (ALS) in the Sierra Nevada (California, USA) and Rocky Mountains (Colorado, USA) (Painter et al., 2016)) to train the model (Section 2.2) and evaluate model performance relative to ALS and other optical satellite SCA data products.

Data and Methods

This study has three components: imagery and lidar acquisition within the Tuolumne Basin, CA, USA and Upper Gunnison Basin, CO, USA study sites, convolutional netural network classification model development and training, and model performance evaluation. These steps are laid out in detail in this section and schematically described in Figure 2.2.

Study Sites

We considered two study sites for our analysis: the Upper Tuolumne Basin in the Sierra Nevada mountains of California, USA (37.89°N, -119.25°W), and the Gunnison/East River Basin in the Central Rocky Mountains of Colorado, USA (39.08°N, -107.14°W). These sites were selected based on their temporally dense archives of ASO lidar-derived snow depth data and the fact that they span multiple climatological zones (Figure 2.3), which enabled more thorough snowcover model performance evaluation. We leveraged data from the Upper Tuolumne Basin for model training and evaluation, and performed additional evaluation in the Gunnison/East River basin.

Data Acquisition and Processing

PlanetScope Satellite Imagery

Planet Labs, Inc. ("Planet") is a commercial satellite imagery company that operates Earth observing satellite constellations, including the "PlanetScope" (PS) constellation of approximately 130 small ("3U" form factor: 10x10x30 cm) satellites in sun-synchronous orbit. This constellation collects approximately 200M km2 day-1 of optical (red, green, blue, and NIR) land-surface imagery at ~3.7m GSD (at nadir) between ±81.5 degrees latitude, with approximately daily revisit times. With ongoing sensor development and satellite launches, the PlanetScope constellation currently includes multiple generations of instruments, including the Dove and Dove-R satellites (corresponding to "PS2" and "PS2.SD" data products, respectively). These instruments have similar design, with comparable (but not identical) spectral band centers and bandwidths (Table S1). For this study, we use the PS and PS2.SD products interchangeably, and develop a method that can leverage the full PlanetScope archive.

We used the Level-3B PlanetScope "Analytic Ortho Scene" orthorectified multispectral surface reflectance data products. Planet's internal data processing pipeline converts Level-1B top-of-atmosphere radiance (derived by applying sensor darkfield and flat field corrections to raw image data) to Level-3B surface reflectance using near-real-time MODIS data and the 6SV2.1 radiative transfer code (Kotchenova et al., 2008; Planet Labs, Inc., 2019b).

Scene Selection and Acquisition

For model training, we selected archived PlanetScope scenes with spatial and temporal overlap with our airborne lidar "ground truth" snowcover data. We selected images within a 5-7 day window around the lidar acquistion dates to ensure a higher probability of cloud-free imagery for similar snow conditions. PlanetScope images were manually inspected for relative cloud fraction, and images with the best cloud-free coverage and shortest time offset were selected for inclusion into our analysis.

We used porder version 0.5.7, an open-source

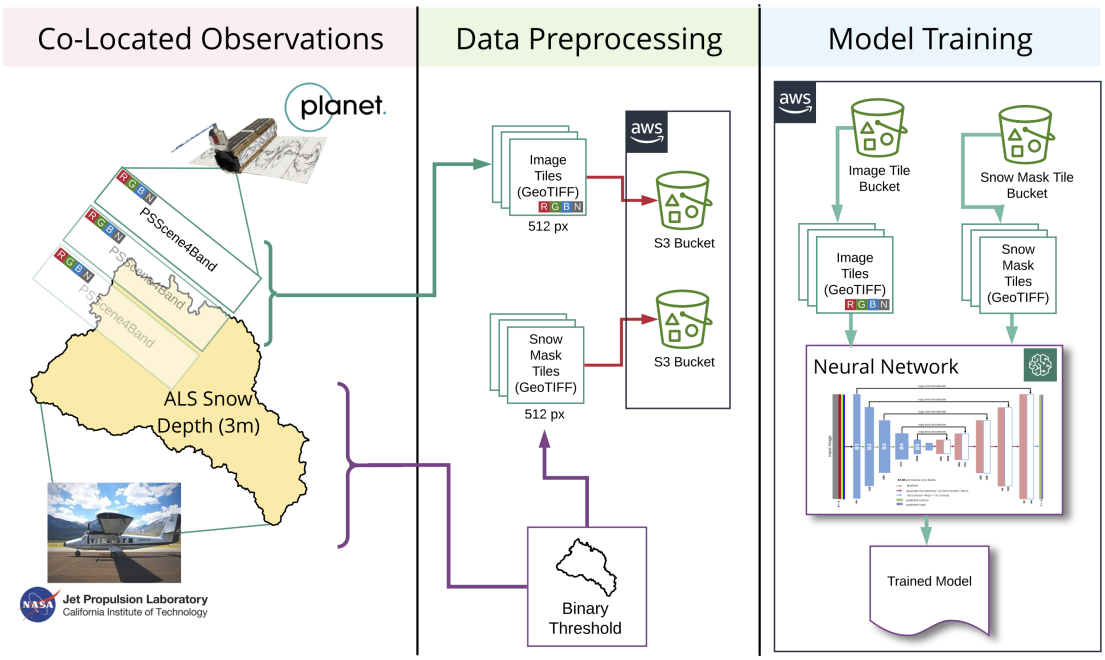

Figure 2.2: Conceptual data processing and model training workflow. Our model training begins with contemporaneous, co-located PlanetScope imagery and ALS snow depth data, which are pre-processed into standardized, gridded GeoTIFFs. A convolutional neural network is trained using Amazon Web Services (AWS) compute infrastructure.

tool (Roy, 2019) for the Planet Orders v2 Application Programming Interface (API) (Planet Labs, Inc., 2019a), to both query the PlanetScope catalog and submit orders. Analytic Ortho Scene assets were queried via the "PSScene4Band" identifier and the "analytic_sr" bundle identifier. We used the Planet Clips API to return only those pixels overlapping our areas of interest (e.g., lidar footprints), both to conserve our imagery quota and reduce data volume.

Airborne lidar

Training and evaluating a neural network model requires high-resolution snowcover classification products to serve as "ground truth" per-pixel labels of snow presence/absence. For this study, we use snowcover masks derived from NASA/JPL Airborne Snow Observatory (ASO) data products (Painter et al., 2016). ASO releases gridded 3-meter snow depth rasters for select watersheds across the western United States with weekly repeat interval from mid-winter through late spring (February – June). These data are available from the National Snow and Ice Data Center's Distributed Data Access Center (`https://nsidc.org/data/ASO_3M_SD`). We generated binary snow classification masks (snow, not snow) from the snow depth products using a threshold of 10 cm (based on ASO accuracy limitations from Currier et al. (2019); Raleigh et al. (2013)). In addition, we derived a tree canopy mask from these data by applying a threshold of 1 m to a canopy height model derived from the ASO ALS survey data. The threshold of 1 m was chosen following from previous efforts demonstrating differential ALS performance in shrub-covered vs. canopy-covered regions when compared to terrestrial surveys (Currier et al., 2019)).

Machine Learning Model

"Machine learning" is a term that is used to describe a set of statistical techniques to build predictive models of an outcome variable from data. Models are "trained" or "fit" to data by selecting a "training" subset of examples from the population of data. These examples are used to derive a predictive relationship with the response variable, and the methodology used to derive this relationship varies for different machine learning approaches. Once fit, models are assessed for their ability to accurately predict response variables given "unseen" samples of data (the "test" subset) which is disjoint from the training set.

In this study we employed a "supervised learning" approach, wherein the presence of the response variable in the data (known as a data "label", in this

Data	Observation Type	Spatial Resolution	Temporal Resolution	Binarization Procedure	Reference
ASO Snow Depth	Airborne lidar	3m	Weekly, during ablation season	Threshold: Depth >10cm	Painter et al. (2016)
Sentinel 2 NDSI	Satellite	10m	5 days	Threshold: NDSI >0.42	Drusch et al. (2012)
Landsat 8 fSCA	Satellite	30m	16 Days	Threshold: fSCA >0.5	U.S. Geological Survey (2018)

Table 2.1: Snow cover datasets used in this study for comparison to PlanetScope-derived snow cover. The "binarization procedure" column describes technique used to derive binary snow cover mask from continuous data fields.

case: lidar-derived snow presence) guides the search for a statistical relationship between the input data ("features") and the response ("label"). Once a supervised learning model is fit using data that contains the response variable, the resulting statistical relationship can be employed to predict the response variable from unlabeled data.

Identifying the spatial extent and categorical classification of regions within images is known as "image segmentation" or "instance segmentation." Identification of snow-covered regions in satellite imagery fits well within this task definition. In our version of the task, the four bands of PlanetScope imagery at each pixel represent the input data to our model (the "features"), and the airborne lidar-derived binary snow presence represent the response variable ("labels").

We employ a convolutional neural network to accomplish this image segmentation task. Neural networks are specific types of machine learning methods designed to extract statistically meaningful linear combinations of input features from data (PlanetScope bands) and model a dependent variable (snow presence/absence) as a nonlinear function of these derived linear combinations (Hastie et al., 2009, Section 11.1). The particular network used in this study is based in the "U-Net" network architecture, a network previously shown to perform well in biomedical image segmentation (Ronneberger et al., 2015), modified by Iglovikov et al. (2018) to perform building detection in satellite remote sensing imagery. We use the resulting network, known as "TernausNetV2", for our satellite image segmentation task. To our knowledge this method has not been applied to the segmentation of snow in satellite imagery.

Model Training

The neural network was trained using data from the Upper Tuolumne Basin site. For training, we paired a given set of binary snow mask tiles from a single ASO collection with the corresponding set of co-located, contemporaneous PS imagery tiles acquired within 3-7 days of the ASO acquisition date (Figure 2.2), and divided this set of image-mask pairs into training and testing subsets via a 70/30% split. Each training effort involved 50 epochs with a batch size of 7 and a learning rate of 2.5×10^{-5}. We leveraged two ASO collection dates (Table 2A.2) in the Tuolumne region by repeating the above procedure for each ASO collection, initializing each subsequent model training procedure with the weights derived from the previous model training run. This allowed the training process to build upon previous training runs, which reduces training time and improves model performance. More technical details of the model training procedure are available in Appendix 1. The open-source implementation of this procedure and all associated data processing workflows are available on GitHub at `https://github.com/acannistra/planet-snowcover`.

Model Performance Evaluation

We use the ASO-derived snow cover mask as "ground truth" for evaluation of our PS model SCA, and compute pixel-based metrics of performance of our PS-derived snow mask with reference to the ASO snow mask. Additionally, to compare against existing SCA datasets, we acquired SCA products from Sentinel-2 and Landsat-8 that were both contemporaneous and co-located with the PlanetScope and ASO lidar snow masks described above for each study site. We computed normalized difference snow index (NDSI) from Sentinel-2 bands 3 (10 m) and 11 (nearest-neighbor resampled to 10 m) as:

$$\text{NDSI} = \frac{\text{B3} - \text{B11}}{\text{B3} + \text{B11}}$$

and generated snowcover masks using an NDSI threshold of 0.42 (Drusch et al., 2012). We used 30-m Landsat 8 fSCA products available from the

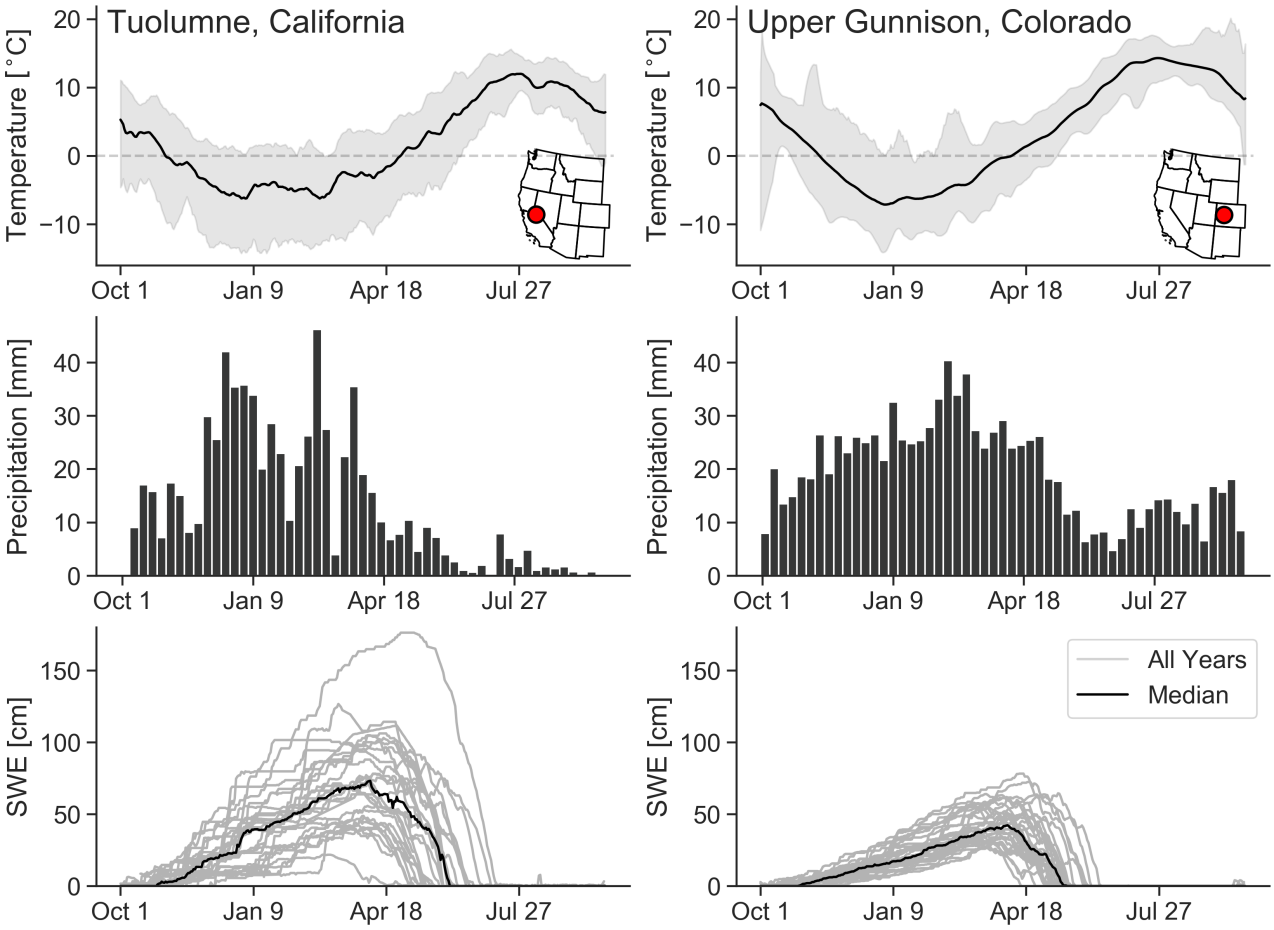

Figure 2.3: The Tuolumne (California, USA) and Upper Gunnison (Colorado, USA) watersheds are members of distinct climatological zones. Primarily within a maritime climate, the Tuolumne region (left) exhibits higher and more variable snow water equivalent accumulation (bottom; data from individual years shown in light gray, daily climatological median as dark black) when compared to the Gunnison site (right). While temperature patterns are similar (top; daily climatological means for period of record, 10-day rolling average), precipitation patterns (middle; climatological mean of daily total precipitation, summarized weekly) exhibit more variability at the Tuolumne site (left). Tuolumne temperature, precipitation, and SWE observations are derived from a corrected time series at a California Department of Water Resources snow pillow site at Dana Meadows (SWE: 1980-2015; Temp. + Precip.: 2002-2015; see [22]). Gunnison observations were obtained from the United States Department of Agriculture National Resources Conservation Service SNOTEL site at North Lost Trail, CO (1985-2019; site No. 669; `https://nrcs.usda.gov`).

United States Geological Survey (U.S. Geological Survey, 2018), and generated snowcover masks using an fSCA threshold of 0.5. A summary of all SCA products used in this study is presented in Table 2.1.

For each snowcover dataset (e.g. CubeSat SCA, Landsat-8 fSCA, and Sentinel-2 NDSI) we computed precision, recall, F-score and balanced accuracy with reference to contemporaneous airborne lidar-derived snow cover (see Table 2.1).

Precision is the percentage of snow classifications predicted by our model that are also snow classifications in the compared dataset:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2.1)$$

Recall is the percentage of true snow classifications predicted by our model that are also true snow classifications in the compared dataset

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2.2)$$

F-Score or F1 score is the harmonic mean of precision and recall:

$$\text{FScore} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.3)$$

Balanced accuracy normalizes the true positive and true negative predictions by the number of true

positive and true negative samples to allow for a less biased assessment of accuracy given the relative accuracy of each prediction type:

$$\text{Balanced Accuracy} = \frac{\text{True Positive Rate} + \text{True Negative Rate}}{2} \quad (2.4)$$

Within-region assessment and Out-of-region transferability

We created an assessment regime using the above metrics to assess the model performance within the training region (Upper Tuolumne, CA) and compared it to model performance outside of the training region (Gunnison, CO). This strategy to allowed us to determine the transferability of a given trained model to a geographically/climatically distinct region. To do this we identified PS imagery that overlapped with an ASO "ground truth" snow cover mask in each basin, excluding any individual image with an overlap footprint of less than 2M pixels (roughly 18 km2). In assessing Tuolumne (within-region) performance, we selected a subset of PS imagery via the 30% "test set" identified during the model training procedure which overlapped with a single ASO collection (`ASO_3M_SD_USCATE_20180528`, see Table 2A.2 and Figure 2.4). This ensured that the chosen imagery was not part of the model training procedure. For the Gunnison (out-of-region) assessment, we selected a set of PS images that overlapped a single ASO collection (`ASO_3M_SD_USCOGE_20180524`, see Table 2A.3). We then used the trained model from the Tuolumne basin to predict snowcover for these Gunnison images, and computed the mean and standard deviation of each metric outlined above.

Comparison with other satellite datasets

We compared our PS model results with co-located snowcover products generated from other satellite datasets (Table 2.1) acquired within 5-15 days of the ASO observation acquisition date. We applied nearest-neighbor resampling to the ASO snow cover rasters to match the native spatial resolution of each satellite snow cover dataset. Means and standard deviations of each metric described above were then computed with reference to the contemporaneous binarized ASO collection.

Performance under tree canopy

We also examined the effect of tree canopy cover on our PS model performance. In particular, we evaluated the ability of high-resolution PlanetScope imagery to identify snow in canopy-covered regions versus uncovered regions, and compared this performance to the snowcover products from other satellite sources. We applied the tree canopy mask described above to PS-based model predictions and each of the alternative SCA datasets and computed snow identification performance as above. We performed this analysis for all images included in earlier performance analyses at both the Tuolumne and Gunnison sites. Percentage of tree canopy coverage in these images was computed to be between 22-55%.

Results

We divide the evaluation of our snow classification model into three assessments. We first show that the PS snow classification model trained using data from the Upper Tuolumne Basin, California, USA produces improved snow classification results when compared with other satellite sources. Second, we find comparable performance when assessing the same model (fitted with Tuolumne data) using out-of-region data from the Gunnison River/East River Basin, Colorado, USA. Finally, we demonstrate mixed model performance under tree canopies in both the Tuolumne and the Upper Gunnison basins when compared to other SCA products.

Within-Region Model Performance (Tuolumne, CA)

We used out-of-sample data from Tuolumne to compare our PS model output to the reference ASO snow cover, and snowcover products derived from Sentinel-2 NDSI and Landsat 8 fSCA across four metrics of performance (Figure 2.5C). The PS model consistently outperforms the other, lower-resolution satellite products, though the standard deviation computed across test images do overlap between the PS model statistics and those of other products. For example, the PS model produces results with a mean F-score of 0.73 (S.D. = 0.12), compared to 0.64 (Landsat 8, S.D. = 0.15) and 0.63 (Sentinel-2, S.D. = 0.14). We also find differences in balanced accuracy, with a mean score of 0.82 (S.D. = 0.07) for our PS model, compared to 0.75 (Landsat 8, S.D. = 0.07) and 0.75 (Sentinel-2, S.D. = 0.13). See Figure 2.5 and Table 2A.4.

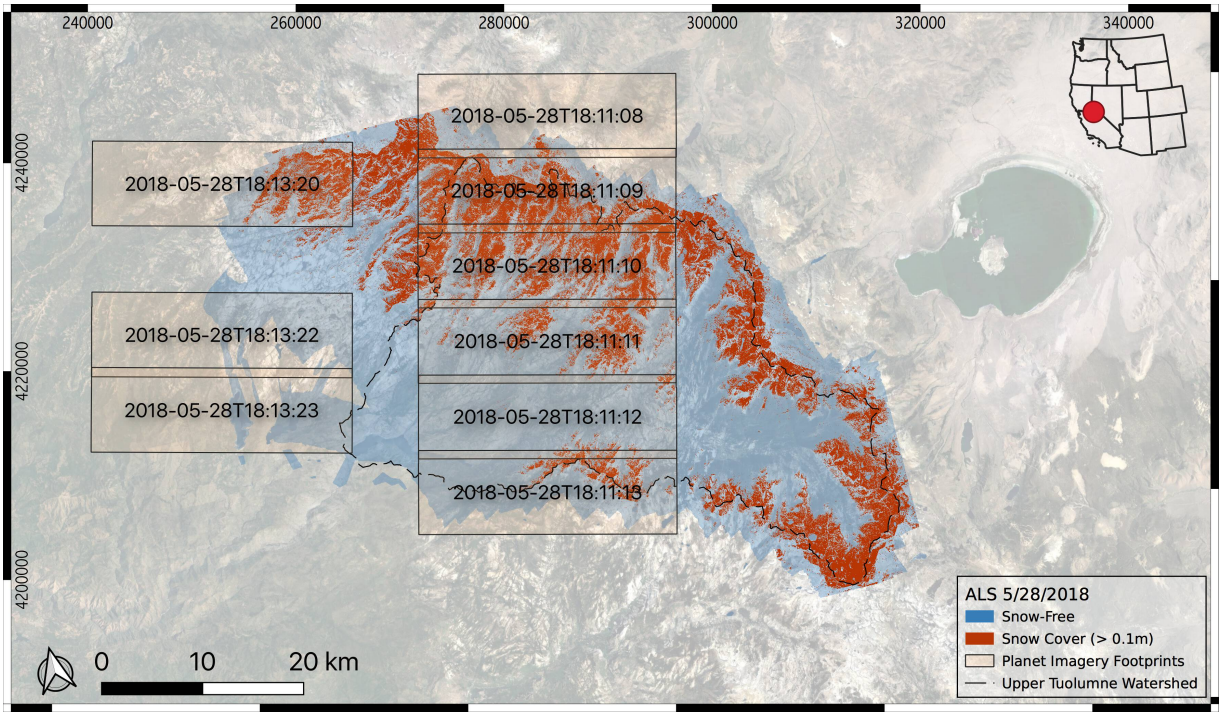

Figure 2.4: A subset of the cloud-free May 28, 2018 PlanetScope scenes (boxes) and the contemporaneous May 28, 2018 ALS data (`ASO_3M_SD_USCATE_20180528`) for Tuolumne site. Red pixels were snow-covered in the ALS "ground truth" data, while blue pixels were identified as snow-free. Both datasets were used for model training and within-region evaluation. See Table 2A.2 for a full list of the PS scenes used for training.

Out of region Model Performance (Gunnison/East River Basin, CO)

When evaluating performance using out-of-region PS imagery from the Gunnison/East River Basin, CO, the Tuolumne-trained model exhibits comparable performance to Landsat 8 fSCA- and Sentinel-2 NDSI-derived snow masks across three of four classification metrics (precision differs; Figure 2.6C). For example, balanced accuracy for our PS Model was 0.75 (S.D. = 0.08), comparable to 0.77 (Landsat 8, S.D. = 0.10) and 0.76 (Sentinel-2, S.D. = 0.10); similarly, F-score for the PS Model was 0.67 (S.D. = 0.18), comparable to 0.73 (Landsat 8, S.D. = 0.19) and 0.68 (Sentinel-2, S.D. = 0.23). Figure 2.7 provides an illustration of a 2 km by 2 km segment of the data used to compute these metrics, highlighting the improved performance of PS-based snow masks over other methods. In addition, we observe that the presence of tree canopy (e.g. vegetation with height greater than 1 m from a lidar-derived canopy height model) causes all four performance metrics to decline in this single scene (Figure 2.6, panel D). Complete metrics are available in Tables 2A.5 and 2A.7.

Model performance beneath tree canopy

For the Gunnison site, the PS model performance across three of four metrics (Recall, F-Score, and Balanced Accuracy) was lower for canopy-covered areas compared to both canopy-free areas and mean performance across entire images (Figure 2.8A). Similar reduced accuracy under canopy was also observed in the 10 m Sentinel 2 SCA data product (Figure 2.8B), but was considerably less prominent in the 30 m Landsat 8 fSCA data product (Figure 2.8C). Full metrics are available in Table 2A.7.

Discussion

We demonstrated, through evaluation across two climatically distinct sites, that PlanetScope imagery can offer reliable snowcover classification at high spatial (~3 m) and temporal (~1-2 d) resolution (Figure 2.5; Figure 2.6C). Our convolutional neural network methodology enables SCA identification from PlanetScope data through deep learning and pattern recognition. This approach leverages all available PlanetScope bands and proved successful to exploit the RGB and NIR band placements that posed

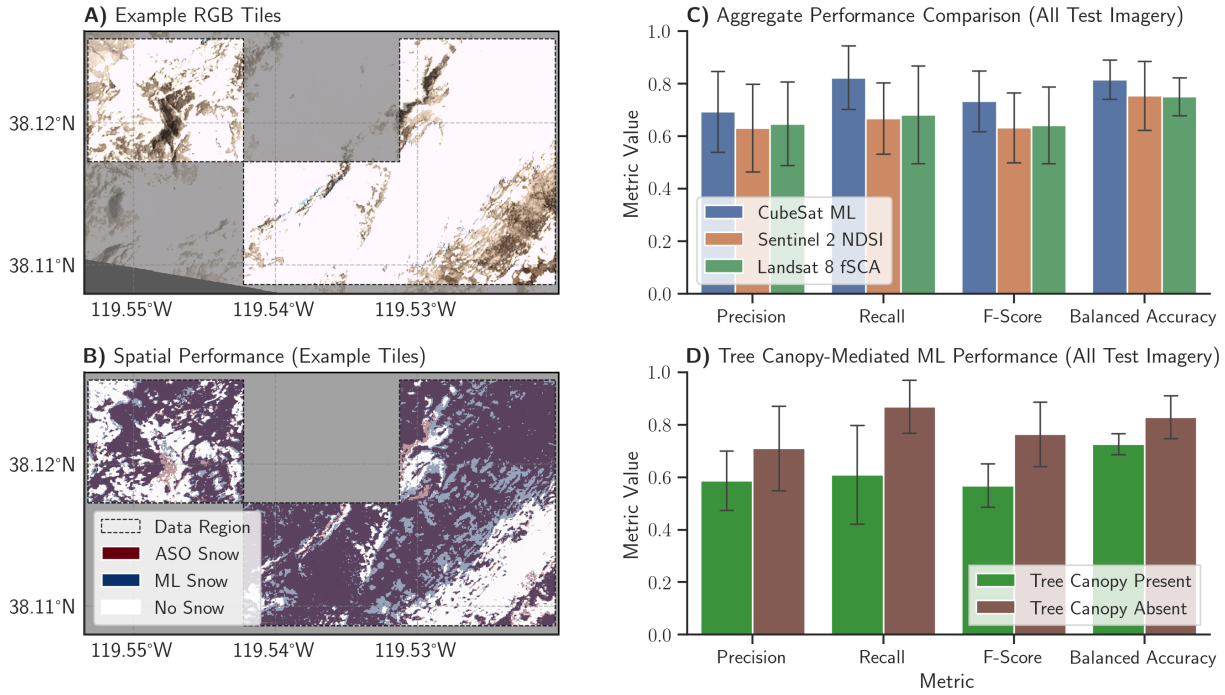

Figure 2.5: Metrics of SCA segmentation performance, Tuolumne, CA. **A)** Example out-of-sample ("test") RGB PS tiles; **B)** binary snow classification from PS model (blue) and ASO (red). Overlap is shown in darker hue. Gaps in ASO snow-covered area are due to the threshold value of 0.1 m during the binarization of ASO snow depth. **C)** Aggregate performance comparison across all Tuolumne out-of-sample imagery, compared to Sentinel-2 and Landsat 8 snow classifications. Bars indicate standard deviation across imagery. PS model snow masks perform as well or better than other classifications. **D)** Mean tree canopy-mediated model performance of PS model-derived snow cover, aggregated across all imagery. Bars indicate standard deviation. Canopy cover decreases model performance across all four metrics. Full statistics available in Tables 2A.4 and 2A.6. Includes copyrighted material of Planet Labs, Inc. All Rights Reserved.

challenges in deriving SCA via traditional radiometric indices (Figure 2.1). This contribution represents both a novel application of an existing method (convolutional neural networks) to the detection of snow in remotely sensed optical imagery and a detailed assessment of PlanetScope imagery for snow remote sensing applications.

Comparison to current SCA Approaches

Previous efforts to measure SCA at similarly high resolutions have relied upon the statistical or dynamical downscaling of coarser radiometrically derived observations to fractional SCA products, such those derived from 500 m daily MODIS (Cristea et al., 2017; Painter et al., 2009) or 30 m, ~weekly Landsat (Painter et al., 2009). While effective, these approaches are ultimately limited by the methodological and technical assumptions inherent in downscaling coarse observational data. For example, while downscaling MODIS observations is a reliable approach in most cases, observational factors such as sensor viewing angle and forest presence have been shown to significantly degrade fSCA quality and ultimately to affect the downscaled SCA product (Cristea and Lundquist, 2016; Rittger et al., 2019). Fractional SCA derived from sensors with large field of view such as MODIS can be affected by viewing angle and resampling strategy, with errors as large as 50% over forested areas at the end of MODIS scan line (Xin et al., 2012). These errors will be smaller in PlanetScope data. PlanetScope images are released with nadir view angles of less than 5 degrees and fewer than 20% saturated pixels. Here we have demonstrated the effectiveness of a SCA retrieval method at native ~3.7 meter GSD from daily, near-nadir PlanetScope observations.

Comparisons of PS-derived SCA with other high-resolution optical satellite SCA data products (e.g. Figure 2.5 and Figure 2.6C) show that our method offers an accurate observational SCA technique. This finding is also promising for emerging multi-sensor fusion methodologies. Combining multisource and multitemporal remote sensing data with machine learning models has been suggested as a suitable method to overcome limitations of individual

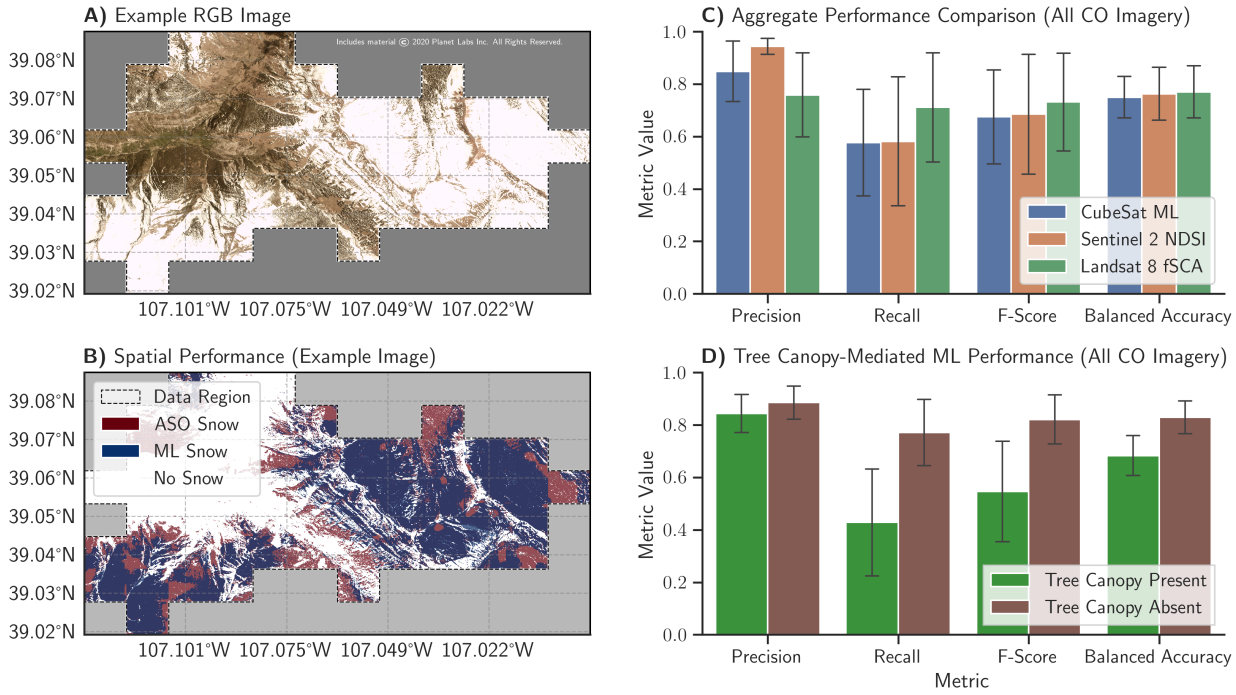

Figure 2.6: Out-of-region performance of Tuolumne-trained snow-covered area identification model in the Gunnison River Basin, Colorado, USA. **A)** Sample PS image. **B)** Classification results. Our model exhibits systematic misclassification behavior, demonstrated here in a selected image (Planet Scene ID#: `20180524_172637_0f2d`, Panel A). Regions of the selected image that contain snow (Panel B; red coloring) are mis-classified as being snow-free by our model (Panel B; absence of blue coloring). **C)** Comparison with other satellite data. The PS model shows comparable performance to SCA derived from Sentinel-2 and Landsat 8 platforms across four metrics with reference to ASO ground truth (Panel C; Bars represent mean metric values across multiple distinct images covering a single ASO collection, error bars show standard deviation). **D)** Performance under canopy. Model performance metrics decline in the presence of tree canopy cover (canopy height $>$ 1 m, Panel D) as determined by an ASO-derived canopy height model. Bars show means; error bars indicate standard deviation across all Gunnison imagery.

data sources (e.g. resolution, measurement type) (Ghamisi et al., 2019). For example, the global PlanetScope archive of daily imagery is an avenue for sourcing high-resolution imagery which could be further refined by other sensing methodologies with improved spectral bandwidth and band placement. Our methodology, when applied to this fused data product, may demonstrate further improved SCA segmentation performance while maintaining desirable spatial and temporal resolution characteristics. The advantages of using high-resolution snow data are once again shown in Figure 2.7, where finer-scale patterns of snow distribution can be observed in PlanetScope-derived SCA relative to coarser products. The ability to observe these fine-scale patterns is a principal advantage of PS-derived data for scientific applications.

Utility of High-Resolution Snow Cover

With the potential to be applied to the global catalog of PlanetScope imagery, our method is well-poised to serve as a useful tool across myriad scientific domains. For example, in studies of alpine ecology, snow disappearance date (SDD) has been shown to be a strong control on the regimes of ecological disturbance that mediate plant diversity and species persistence in heterogeneous alpine environments (Choler, 2005). Changing SDD has also been shown to cause reassembly in mountain wildflower communities (Theobald et al., 2017). However, due to the intensive nature of observing snow cover in the field, such ecological insights are necessarily accompanied with caveats regarding the accuracy of observed SDD or geographic scope. As such, high resolution SCA, both in space and time, could serve to both improve the accuracy of ecological studies reliant on SDD observations and allow for existing geographically distributed ecological datasets to be used in understanding the relationship of SDD to other ecological variables. Hydrologic modeling is another domain with a demonstrated need for this approach; the spatially heterogeneous distribution of snow can be more accurately observed with higher-resolution

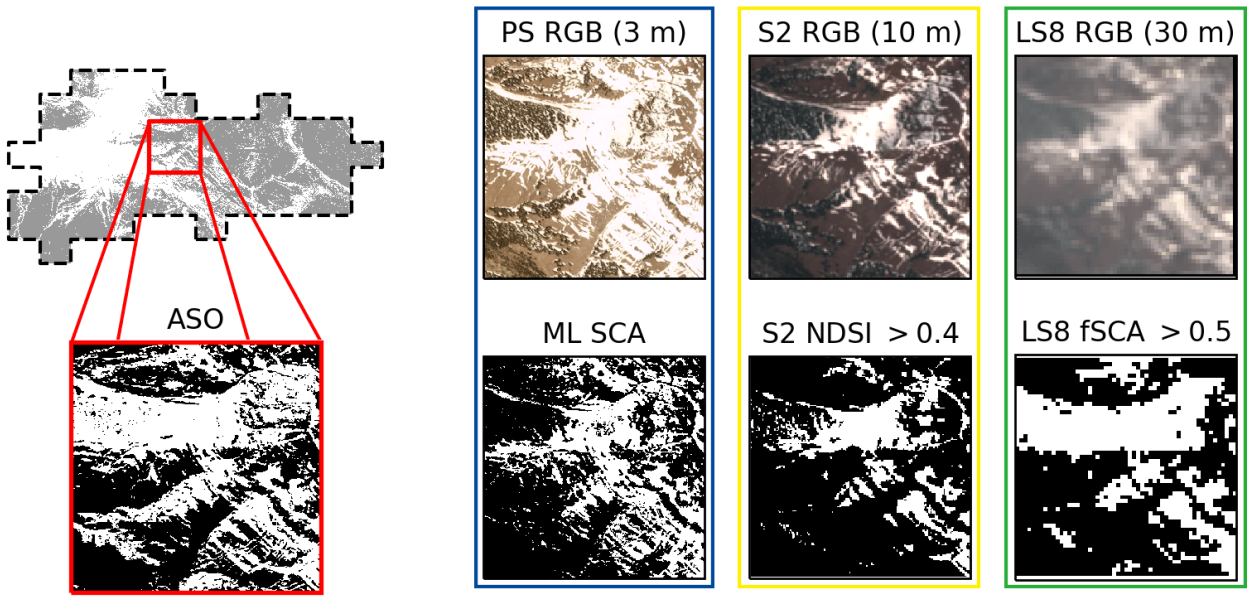

Figure 2.7: Fine-scale patterns of snow distribution can be observed in 3 m PlanetScope derived SCA in contrast to coarser observations. This figure compares contemporaneous RGB observations and derived SCA for three optical satellite platforms (PlanetScope, Sentinel 2, and Landsat 8, see Table 1) to lidar-derived SCA within a 2 km by 2 km region selected from a single PS Scene (Planet Scene ID# `20180524_172637_0f2d`) within the Upper Gunnison River basin domain, depicted by dotted-line region and in Figure 2.6. Includes copyrighted material of Planet Labs, Inc. All Rights Reserved.

SCA, with applications in predicting streamflow timing (Lundquist and Dettinger, 2005) and validating hydrologic model parameterizations (Clark et al., 2011; Luce et al., 1999).

Observed Limitations

Similar to other studies focused on optical remote sensing of snow, tree canopy cover presents a challenge for our methodology. This is also illustrated in Figure 2.7, where PlanetScope-derived SCA fails to identify snow cover in canopy-covered regions containing snow (as seen from ASO and PS RGB), but other radiometric approaches (e.g. Sentinel-2 NDSI and Landsat 8 fSCA) are more successful. More generally, the performance difference between canopy-present and canopy-absent regions observed in PS-derived snow masks is substantial (Figure 2.5D; Figure 2.6D; Figure 2.8A), but compares to the difference observed in the Sentinel-2 NDSI-derived SCA product when examined at the Gunnison site (Figure 2.8B). Though Landsat 8 fSCA performance exhibits a more muted decline in response to canopy presence (again at Gunnison site; Figure 2.8C), previous studies have shown that both Landsat and MODIS experience degraded snow observation performance in tree-covered and tree-shadowed areas (Kane et al., 2008; Raleigh et al., 2013). In addition, though we do not formally investigate the impact on cloud cover on snow identification performance, clouds are likely to present a challenge to SCA identification for our PS model, either via obfuscation or as a source for false-positives. This is especially true in comparison to other optical platforms due to the limited 4-band nature of PS imagery.

Importantly, we do not consider snow identification beneath tree canopies to be a useful metric of the utility of this method, as canopy obfuscation puts a hard technical upper-bound on possible performance in these regions. However, the ability of our methodology to identify snow in forested regions and in forest gaps is an important issue for further inquiry. Several techniques have been proposed to improve snow mapping in vegetated areas and among forest gaps. These include corrections using canopy reflectance models (Klein et al., 1998), observation-based approaches (Rittger et al., 2019), new radiometric indices (Wang et al., 2015), as well as combination of data from multiple sensors (Dressler et al., 2006; Durand et al., 2008; Raleigh et al., 2013). Although the current iteration of our neural network-based method does not yet achieve desirable performance in mapping snow in forested or cloudy areas, this methodology holds promise. A major contributor to the desirable performance of the convolutional "U-Net" architecture employed here is an underlying ability to "learn" meaningful spatial patterns from training data (Ronneberger et al., 2015), a quality

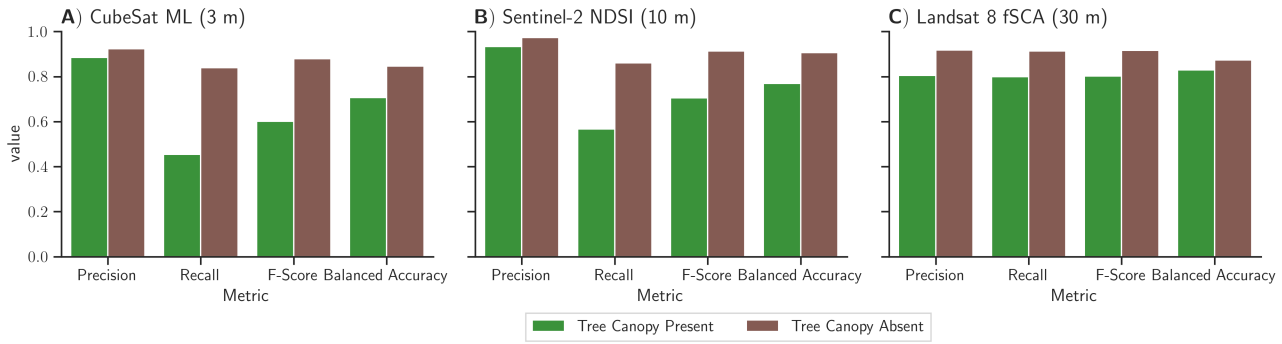

Figure 2.8: Effects of tree canopy on SCA identification performance for a single Gunnison, CO PS scene (Percent Tree Canopy: 55%). Panels show impact of tree canopy on snow classification performance with A) the CubeSat (PS) model B) Sentinel-2 NDSI, and C) Landsat 8 fSCA for a single PlanetScope image (ID#: `20180524_172637_0f2d`, Figure 2.6A) in the Upper Gunnison Basin, CO. Finer resolution SCA products (3 m CubeSat ML and 10 m Sentinel 2 NDSI, panels A and B) exhibit more pronounced performance differences between canopy-covered areas (green bars) and canopy-free areas (brown bars) than a coarser observational product (Landsat 8), where the discrepancies are less pronounced.

that will perhaps improve canopy/cloud differentiation and improve SCA performance for these cases. To leverage this potential opportunity, future work could incorporate forest canopy cover or cloud detection algorithms during the model training process, potentially allowing the model to differentiate canopy/cloud from non-obstructed snow, improving performance; however, we hypothesize that utilizing this methodology to enhance SCA segmentation performance among and between tree canopies may limit the geographic extensibility of a given model due to regional differences in forest structure.

PlanetScope imagery offers a novel lens into snow dynamics via an unique combination of spatial resolution and high revisit frequency. However, there are important technical considerations to acknowledge when utilizing these relatively new observational data. First, when compared to rigorously calibrated platforms such as MODIS or Landsat 8 OLI, imagery from Dove satellites (that comprise the PlanetScope fleet) exhibits radiometric calibration errors that can impact any analyses reliant upon them. Recent research has investigated methods to re-calibrate PlanetScope imagery using coarser Landsat observations (Houborg and McCabe, 2016). Furthermore, two sensor types ("instruments") are present in the PlanetScope constellation ("PS2" and "PS2.SD") resulting from ongoing sensor development and satellite launches. These instruments have comparable (but not identical) spectral band centers and bandwidths (Table S1). The inconsistency of the bandwidth and band centers in these PlanetScope data may affect the ability of our modeling methodology to identify snow. In addition, preliminary work in generating digital elevation models (DEMs) from PlanetScope imagery has revealed limitations in georeferencing accuracy, especially in remote or mountainous regions, potentially leading to offsets and RGB vs NIR band misalignment Bhushan et al. (2020). These shortcomings are unlikely to be significant in applications such as snow cover identification, but future work to assess the magnitudes of error in these data is warranted.

Future Work

This study demonstrates the continued utility of publicly available snow depth measurements derived from airborne lidar. Without the accurate snow depth products from the Airborne Snow Observatory, we would lack high-resolution "ground truth" data necessary for model training. Some opportunities for future work exist regarding the optimal approach to creating a binary snow mask from these snow depth data, and to integrate ASO spectrometer data for improved classification. We chose to use a threshold of 10 cm to convert the ASO snow depth field into a binary snow mask based on currently available field-validated assessments of ASO accuracy (see Methods). However, it is likely that this choice had a significant impact on our results; qualitative examination of PlanetScope-derived SCA compared to actual contemporaneous imagery reveals that our method identifies snow-covered regions that in many cases are "no-snow" regions in the ASO-derived snow masks. This suggests that our model's identification of shallow/seasonal snowpacks may in some cases be incorrectly counted as a misclassification, lowering performance metrics. Future work should investigate the choice of threshold to determine the impacts of this choice on model performance.

The potential for our method to extract high-

resolution snow cover over time across regions without ground-truth data is an important facet of our work. Given the global observational footprint of the PlanetScope imagery and the relative sparsity of repeat lidar snow depth observations, determining the skill of a model trained in a lidar-rich region and in other regions without lidar snow depth data is an important metric. In our assessment of model performance in the Upper Gunnison River Basin, CO (Figure 2.6C) we observed good model performance compared to contemporaneous ALS observations when using the Tuolumne-based model to identify SCA. While it is necessary to take into account the above caveats regarding tree canopy structure differences across sites, this result suggests transferability potential—models trained in ALS-rich regions can be used in other regions to identify snow covered area without re-training the model in those new regions. This finding also suggests that the large but spatially constrained archive of existing ALS SCA data can be leveraged to create a model, or set of models, which is able to provide 3 meter SCA observations across a much larger spatial footprint.

Conclusions

We developed a method to identify snow covered area using PlanetScope CubeSat imagery, a ~3 meter, ~daily, 4-band optical image dataset with ~global extent. Our method performs comparatively to other state-of-the-art remotely-sensed measures of snow cover (F-Score: 0.73, compared to 0.64 and 0.63 for Landsat 8 fSCA and Sentinel-2 NDSI). We show that these emerging image archives have great potential to accurately observe snow-covered area at high spatial and temporal resolutions despite limited radiometric bandwidth and band placement. We used a machine-learning based image segmentation approach, powered by a convolutional neural network modeling framework, to meet the challenge of producing a snow-covered area product from these data. Metrics of snow classification performance at our primary training and assessment site (the Upper Tuolumne Basin, CA, USA) show levels of classification accuracy comparable both to high-resolution airborne lidar surveys (ASO) and snow classifications derived from other platforms (Sentinel-2 and Landsat 8). The transferability potential of our model was examined via validation in a climatically and geographically distinct basin (Upper Gunnison, CO, USA), where we found slightly lower performance levels when compared to the model training basin. Finally, we demonstrate our model's diminished performance as a result of tree canopy cover in both study sites, suggesting the potential for future work in assessing the utility of the PS archive in identifying snow covered area in vegetated regions. The unparalleled spatial and temporal coverage of CubeSat imagery offers an excellent opportunity for satellite remote sensing of snow, with real implications for ecological and water resource applications.

Appendix 1: Cyberinfrastructure & Model Training

Training a neural network is a computationally demanding task requiring access to large quantities of data and specialized hardware. In particular, computers with large memory and graphics processing units (GPUs) greatly shorten the training time for our model and enable quicker experimentation. In addition, the large volume of both airborne lidar and satellite imagery data co-located with our study sites required us to have access to large data storage facilities. For these reasons we chose the compute and storage resources provided by Amazon Web Services (AWS), a commercial cloud service provider, to enable our training procedure.

Our implementation of the training procedure is based in the Python programming language (v.3.5; Python Software Foundation, https://www.python.org) using PyTorch (Paszke et al., 2019), and is a heavily modified version of the "robosat.pink" software, an open-source set of command-line tools to enable machine learning with satellite imagery via the TernausNetV2 image segmentation network (Courtin and Hofmann, 2019; Iglovikov et al., 2018). The original software in this package was developed for three band remote sensing imagery and as such was not able to leverage multispectral data. We modified the package to enable the use of any N-band multispectral imagery data product and to allow for the use of cloud-based data storage and computation infrastructure.

To allow for quicker experimentation and simpler reproducibility, we packaged the training code, dependencies, and other software for our neural network implementation into a platform-agnostic computational working environment (or "container") via Docker (Merkel, 2014). We used the AWS "SageMaker" service to manage the training of our network, which greatly simplified experimentation with different network parameterizations and datasets. We chose the "`p2_xlarge`" AWS EC2 instance type for our training, as it afforded sufficient memory and

graphics processing units for the training task.

Once acquired, imagery and airborne lidar-derived snow masks were stored as single or multi-band GeoTIFF files Open Geospatial Consortium (2019) in AWS Simple Storage Service (S3) "buckets" to enable access by further processing tools. To enable co-registration of the snow mask data with imagery data and produce standardized "data units" required by neural network training, we divided the raw imagery and snow mask data into 512 by 512 pixel images, or "tiles," derived from a standardized global grid. We used the Spherical Web Mercator Spherical Tile standard (sometimes referred to as the "slippy map" tile standard due to their employment in interactive mapping applications) to define the grid of tiles OpenStreetMap (2019), and use the "mercantile," "rasterio," and "rio-tiler" open-source software packages to enable gridding and storage of these images Mapbox (2019); Vincent (2019). The spherical Web Mercator tile standard assigns a unique spatially explicit identifier to each 512x512 pixel image tile, which can then be used to align imagery tiles and snow mask tiles (e.g. the tiles "snow/1/2/3.tif" and "image/1/2/3.tif" have identical spatial extent). The resulting tiles were of similar spatial resolution to the input data (2.4 m), differing slightly due to resampling required to match the Spherical Web Mercator standard. These tiles were stored as GeoTIFF files in AWS S3 buckets tagged with their image or ASO collection identifiers and dates of collection. This preprocessing effort was completed via Jupyter notebooks (Kluyver et al., 2016) on AWS Elastic Compute Cloud (EC2) compute instances (see Figure 2A.1).

Figures and tables begin next page...

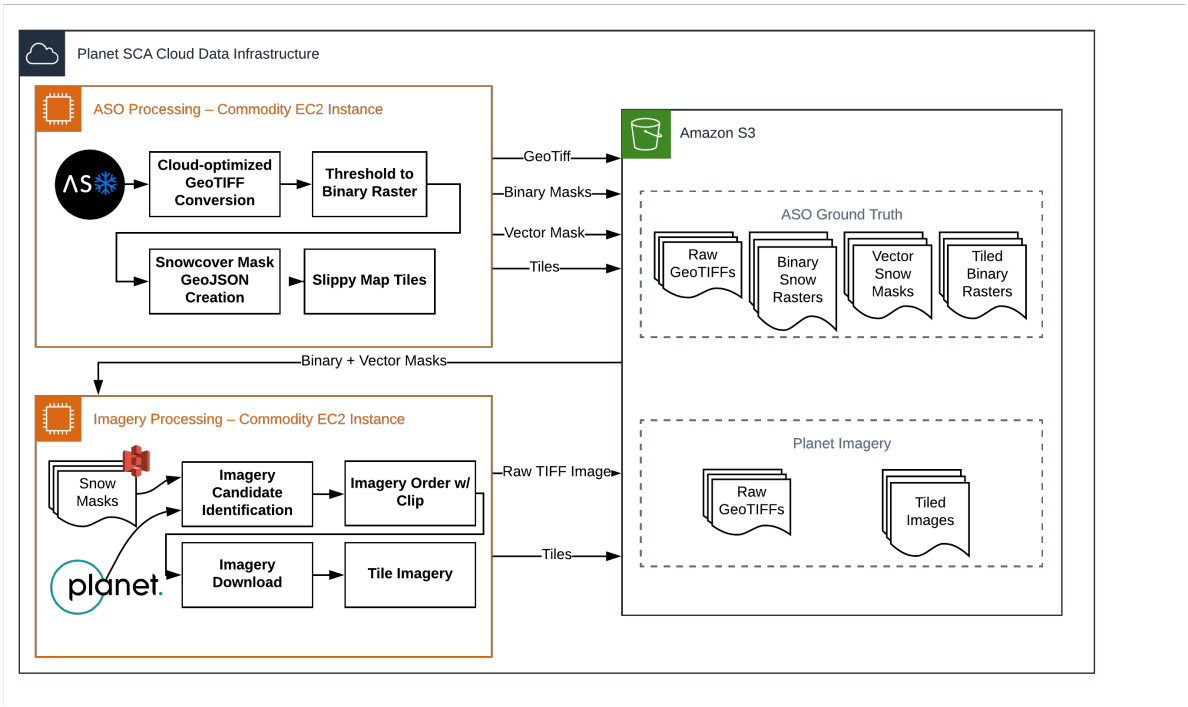

Figure 2A.1: Schematic of data preprocessing procedure for co-located Planet Labs Inc. satellite imagery and Airborne Snow Observatory snow mask data via Amazon Web Services cloud infrastructure.

	PS2	PS2.SD
Blue:	455 - 515 nm	464 - 517 nm
Green:	500 - 590 nm	547 - 585 nm
Red:	590 - 670 nm	650 - 682 nm
NIR:	780 - 860 nm	846 - 888 nm

Table 2A.1: Spectral bandwidth of PS2 and PS2.SD instruments within the PlanetScope constellation, (Planet Labs, Inc., 2019a).

ASO/Ground Truth ID	ASO Basin ID	ASO Basin Name	Planet Scene ID
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_180846_1002
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_180847_1002
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181108_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181109_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181110_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181111_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181112_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181113_1025
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181319_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181320_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181322_1005
ASO_3M_SD_USCATE_20180528	USCATE	TOLUMNE + CHERRY/ELEANOR	20180528_181323_1005
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181447_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181448_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181450_0f32
ASO_3M_SD_USCASJ_20180601	USCASJ	SAN JOAQUIN MAIN FORK	20180601_181451_0f32

Table 2A.2: Data used for training the Tuolumne regional model. Two ASO observations in the Greater Tuolumne region (`ASO_3M_SD_USCATE_20180528` and `ASO_3M_SD_USCASJ_20180601`) are overlapped with contemporaneous Planet Labs imagery for training. These data together represent over 368M labeled ~3.7 m pixels with a footprint of approximately 5,000 km^2.

ASO/Ground Truth ID	ASO Basin ID	ASO Basin Name	Planet Scene ID
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172142_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172143_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172144_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172145_103d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172326_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172327_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172329_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172330_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172331_0f51
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172634_0f2d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172635_0f2d
ASO_3M_SD_USCOGE_20180524	USCOGE	GUNNISON – EAST RIVER	20180524_172637_0f2d

Table 2A.3: Data used for model evaluation in Upper Gunnison, CO basin. One ASO observation (`ASO_3M_SD_USCOGE_20180524`) was used to identify contemporaneous PlanetScope imagery.

SCA	Metric	Mean	Std. Dev.
CubeSat ML	Balanced Accuracy	0.82	0.07
	F-Score	0.73	0.12
	Precision	0.69	0.15
	Recall	0.82	0.12
Landsat 8 fSCA	Balanced Accuracy	0.75	0.07
	F-Score	0.64	0.15
	Precision	0.65	0.16
	Recall	0.68	0.19
Sentinel 2 NDSI	Balanced Accuracy	0.75	0.13
	F-Score	0.63	0.13
	Precision	0.63	0.17
	Recall	0.67	0.14

Table 2A.4: Metrics of Snow Covered Area products in the **Tuolumne basin** as compared to ASO-derived "ground truth." Mean and Standard deviation are computed across 6 regions defined by PlanetScope scene boundaries. Only data not included in model training procedure are used in evaluation.

SCA	Metric	Mean	Std. Dev.
CubeSat ML	Balanced Accuracy	0.75	0.08
	F-Score	0.67	0.18
	Precision	0.85	0.12
	Recall	0.58	0.20
Landsat 8 fSCA	Balanced Accuracy	0.77	0.10
	F-Score	0.73	0.19
	Precision	0.76	0.16
	Recall	0.71	0.21
Sentinel 2 NDSI	Balanced Accuracy	0.76	0.10
	F-Score	0.68	0.23
	Precision	0.94	0.03
	Recall	0.58	0.25

Table 2A.5: Metrics of Snow Covered Area products in the **Upper Gunnison River basin** as compared to ASO-derived "ground truth." Mean and Standard deviation are computed across 12 regions defined by PlanetScope scene boundaries.

SCA	Veg. Type	Metric	Mean	Std. Dev.
CubeSat ML	No Vegetation	Balanced Accuracy	0.83	0.08
		F-Score	0.76	0.12
		Precision	0.71	0.16
		Recall	0.87	0.10
	Vegetation Only	Balanced Accuracy	0.73	0.04
		F-Score	0.57	0.08
		Precision	0.59	0.11
		Recall	0.61	0.19
Landsat 8 fSCA	No Vegetation	Balanced Accuracy	0.75	0.08
		F-Score	0.65	0.17
		Precision	0.69	0.18
		Recall	0.67	0.20
	Vegetation Only	Balanced Accuracy	0.76	0.06
		F-Score	0.59	0.06
		Precision	0.52	0.08
		Recall	0.72	0.11
Sentinel-2 NDSI	No Vegetation	Balanced Accuracy	0.75	0.13
		F-Score	0.66	0.15
		Precision	0.68	0.19
		Recall	0.67	0.14
	Vegetation Only	Balanced Accuracy	0.73	0.13
		F-Score	0.50	0.07
		Precision	0.45	0.10
		Recall	0.61	0.14

Table 2A.6: Metrics of classification performance of snow covered area products, including our PS Model, in vegetated and unvegetated regions of the **Tuolumne basin**. "Vegetation Only" corresponds to pixels containing vegetation > 1 m as derived from the ASO Canopy Height Model. Mean and Standard deviation are computed across 6 regions defined by PlanetScope scene boundaries. Only data not included in model training procedure are used in evaluation.

SCA	Veg. Type	Metric	Mean	Std. Dev.
CubeSat ML	No Vegetation	Balanced Accuracy	0.83	0.06
		F-Score	0.82	0.09
		Precision	0.88	0.06
		Recall	0.77	0.13
	Vegetation Only	Balanced Accuracy	0.68	0.08
		F-Score	0.55	0.19
		Precision	0.84	0.07
		Recall	0.43	0.20
Landsat 8 fSCA	No Vegetation	Balanced Accuracy	0.78	0.12
		F-Score	0.75	0.23
		Precision	0.78	0.20
		Recall	0.73	0.25
	Vegetation Only	Balanced Accuracy	0.75	0.09
		F-Score	0.70	0.17
		Precision	0.72	0.15
		Recall	0.68	0.19
Sentinel-2 NDSI	No Vegetation	Balanced Accuracy	0.81	0.10
		F-Score	0.77	0.21
		Precision	0.94	0.06
		Recall	0.69	0.24
	Vegetation Only	Balanced Accuracy	0.72	0.09
		F-Score	0.60	0.23
		Precision	0.94	0.02
		Recall	0.48	0.24

Table 2A.7: Metrics of classification performance of snow covered area products, including our PS Model, in vegetated and unvegetated regions of the **Upper Gunnison River basin**. "Vegetation Only" corresponds to pixels containing vegetation $>$ 1 m as derived from the ASO Canopy Height Model. Mean and Standard deviation are computed across 12 regions defined by PlanetScope scene boundaries.

References

Andreadis, K. M. and Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29(6):872–886.

Bhushan, S., Shean, D., Alexandrov, O., and Henderson, S. (2020). Automated tools to derive short-term glacier velocity from high-resolution commercial satellite imagery. *Earth and Space Science Open Archive*.

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., Bohrer, G., Brinkman, T. J., Cosgrove, C. L., Eitel, J. U. H., Hebblewhite, M., Kimball, J. S., LaPoint, S., Nolin, A. W., Pedersen, S. H., Prugh, L. R., Reinking, A. K., and Vierling, L. A. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, 14(1):010401.

Carlson, B. Z., Choler, P., Renaud, J., Dedieu, J.-P., and Thuiller, W. (2015). Modelling snow cover duration improves predictions of functional and taxonomic diversity for alpine plant communities. *Annals of Botany*, 116(6):1023–1034.

Choler, P. (2005). Consistent shifts in Alpine plant traits along a mesotopographical gradient. *Arctic Antarctic and Alpine Research*, 37(4):444–453.

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Hreinsson, E. O., and Woods, R. A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, 47:W07539.

Courtin, O. and Hofmann, D. J. (2019). RoboSat.pink Computer Vision framework for GeoSpatial Imagery.

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRis-Lambers, J., and Lundquist, J. D. (2017). An evaluation of terrain-based downscaling of fractional snow covered area data sets based on LiDAR-derived snow data and orthoimagery. *Water Resources Research*, 53(8):6802–6820.

Cristea, N. C. and Lundquist, J. D. (2016). An evaluation of terrain-based downscaling of MODIS-based fractional snow covered area datasets over the Tuolumne River, CA based on lidar-derived snow data. page 4, Seattle, WA.

Currier, W. R., Pflug, J., Mazzotti, G., Jonas, T., Deems, J. S., Bormann, K. J., Painter, T. H., Hiemstra, C. A., Gelvin, A., Uhlmann, Z., Spaete, L., Glenn, N. F., and Lundquist, J. D. (2019). Comparing Aerial Lidar Observations With Terrestrial Lidar and Snow-Probe Transects From NASA's 2017 SnowEx Campaign. *Water Resources Research*, 55(7):6285–6294.

Dedieu, J.-P., Carlson, B. Z., Bigot, S., Sirguey, P., Vionnet, V., and Choler, P. (2016). On the Importance of High-Resolution Time Series of Optical Imagery for Quantifying the Effects of Snow Cover Duration on Alpine Plant Habitat. *Remote Sensing*, 8(6):481.

Dozier, J. (1989). Spectral signature of alpine snow cover from the landsat thematic mapper. *Remote Sensing of Environment*, 28:9–22.

Dressler, K. A., Leavesley, G. H., Bales, R. C., and Fassnacht, S. R. (2006). Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. *Hydrological Processes*, 20(4):673–688.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120:25–36.

Durand, M., Molotch, N. P., and Margulis, S. A. (2008). Merging complementary remote sensing datasets in the context of snow water equivalent reconstruction. *Remote Sensing of Environment*, 112(3):1212–1225.

Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., and Hall, A. (2009). Controls on Northern Hemisphere snow albedo feedback quantified using satellite Earth observations. *Geophysical Research Letters*, 36(21).

Ford, K. R., Ettinger, A. K., Lundquist, J. D., Raleigh, M. S., and Hille Ris Lambers, J. (2013). Spatial Heterogeneity in Ecologically Important Climate Variables at Coarse and Fine Scales in a High-Snow Mountain Landscape. *PLoS ONE*, 8(6):e65008.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P. M., and Benediktsson, J. A. (2019). Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 7(1):6–39.

Hall, D. K. and Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12):1534–1547.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, volume 27 of *Springer Series in Statistics*. Springer New York, New York, NY.

Houborg, R. and McCabe, M. F. (2016). High-Resolution NDVI from Planet's Constellation of Earth Observing Nano-Satellites: A New Data Source for Precision Agriculture. *Remote Sensing*, 8(9):768.

Iglovikov, V. I., Seferbekov, S., Buslaev, A. V., and Shvets, A. (2018). TernausNetV2: Fully Convolutional Network for Instance Segmentation. *arXiv:1806.00844 [cs]*.

Immerzeel, W. W., Droogers, P., de Jong, S. M., and Bierkens, M. F. P. (2009). Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing. *Remote Sensing of Environment*, 113(1):40–49.

Kane, V. R., Gillespie, A. R., McGaughey, R., Lutz, J. A., Ceder, K., and Franklin, J. F. (2008). Interpretation and topographic compensation of conifer canopy self-shadowing. *Remote Sensing of Environment*, 112(10):3820–3832.

Klein, A. G., Hall, D. K., and Riggs, G. A. (1998). Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrological Processes*, 12(10-11):1723–1744.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J. B., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., and al, e. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows.

Kotchenova, S. Y., Vermote, E. F., Levy, R., and Lyapustin, A. (2008). Radiative transfer codes for atmospheric correction and aerosol retrieval: intercomparison study. *Applied Optics*, 47(13):2215.

Little, R., Peterson, D., and Conquest, L. (1994). Regeneration of Sub-Alpine Fir (Abies lasiocarpa) Following Fire - Effects of Climate and Other Factors. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 24(5):934–944.

Luce, C. H., Tarboton, D. G., and Cooley, K. R. (1999). Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model. *Hydrological Processes*, 13(12-13):1921–1933.

Lundquist, J. D. and Dettinger, M. D. (2005). How snowpack heterogeneity affects diurnal streamflow timing. *Water Resources Research*, 41(5):W05007.

Mapbox, I. (2019). Mercantile. https://github.com/mapbox/mercantile.

Merkel, D. (2014). Docker. *Linux Journal*. doi:abs/10.5555/2600239.2600241.

Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E. (2019). Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study. *Remote Sensing*, 11(1):74.

Open Geospatial Consortium (2019). GeoTIFF Standard.

OpenStreetMap (2019). Slippy Map Tilenames.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., and Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, 184:139–152.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4):868–879.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc.

Planet Labs, Inc. (2019a). Planet Developer Resource Center.

Planet Labs, Inc. (2019b). Planet Imagery Product Specifications.

Raleigh, M. S., Rittger, K., Moore, C. E., Henn, B., Lutz, J. A., and Lundquist, J. D. (2013). Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada. *Remote Sensing of Environment*, 128:44–57.

Riggs, G. A. and Hall, D. K. (2015). MODIS Snow Products Collection 6 User Guide. Technical report.

Rittger, K., Raleigh, M. S., Dozier, J., Hill, A. F., Lutz, J. A., and Painter, T. H. (2019). Canopy Adjustment and Improved Cloud Detection for Remotely Sensed Snow Cover Mapping. *Water Resources Research*, 55(n/a).

Rochefort, R. M., Little, R. L., Woodward, A., and Peterson, D. L. (1994). Changes in sub-alpine tree distribution in western North America: a review of climatic and other causal factors. *The Holocene*, 4(1):89–100.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*.

Roy, S. (2019). samapriya/porder: porder: Simple CLI for Planet ordersV2 API. doi:10.5281/zenodo.3547667.

Schattan, P., Schwaizer, G., Schöber, J., and Achleitner, S. (2020). The complementary value of cosmic-ray neutron sensing and snow covered area products for snow hydrological modelling. *Remote Sensing of Environment*, 239:111603.

Theobald, E. J., Breckheimer, I., and HilleRisLambers, J. (2017). Climate drives phenological reassembly of a mountain wildflower meadow community. *Ecology*, 98(11):2799–2812.

U.S. Geological Survey (2018). Collection-1 Landsat Level-3 Fractional Snow Covered Area (FSCA) Science Product.

Venn, S. E., Green, K., Pickering, C. M., and Morgan, J. W. (2011). Using plant functional traits to explain community composition across a strong environmental filter in Australian alpine snowpatches. *Plant Ecology*, 212(9):1491–1499.

Vincent, S. (2019). rio-tiler. https://github.com/cogeotiff/rio-tiler.

Wang, X.-Y., Wang, J., Jiang, Z.-Y., Li, H.-Y., and Hao, X.-H. (2015). An Effective Method for Snow-Cover Mapping of Dense Coniferous Forests in the Upper Heihe River Basin Using Landsat Operational Land Imager Data. *Remote Sensing*, 7(12):17246–17257.

Xin, Q., Woodcock, C. E., Liu, J., Tan, B., Melloh, R. A., and Davis, R. E. (2012). View angle effects on MODIS snow mapping in forests. *Remote Sensing of Environment*, 118:50–59.

Chapter 3

Seasonal and latitudinal effects of marine heatwaves on phytoplankton

Anthony F. Cannistra, Lauren B. Buckley

Abstract

Marine heatwave events (MHWs), defined as discrete, anomalously warm periods of ocean temperature, are projected to occur more frequently and with greater intensity in the coming decades as a result of a warming climate. Recent significant MHWs have been linked to severe ecological consequences, including population declines due to increased mortality and changes in community composition. However, the geographic and taxonomic extent of these prior studies has been limited. In this study we examine the effects of MHW events on phytoplankton fitness across the global ocean via a 38-year satellite record of sea surface temperature in an effort to elucidate global patterns of ecological impacts from MHWs. We use a robust geographically distributed data set of thermal reaction norms of phytoplankton growth rate to link MHW temperature anomalies to fitness across a broad latitudinal and taxonomic range, and assess the size and directionality of these effects. We find that MHW events cause both positive and negative fitness consequences in phytoplankton with significant seasonal and latitudinal patterns. We estimate that MHW events tend to increase fitness during cold periods and decrease fitness during warmer periods. We find differential responses in our estimates that are dependent on latitude, with polar regions experiencing higher-magnitude, largely positive MHW fitness responses in contrast to tropical and subtropical regions, where more frequently negative fitness consequences were estimated. The unique biogeochemical and trophic context of phytoplankton may allow for the framing of phytoplankton fitness consequences as context for whole-ecosystem responses to these climatological events. These findings serve as the first globally-distributed analysis of ecological impacts from MHW events, and suggest that the ecological impacts of marine heatwaves will be widespread, with detrimental impacts focused in warm and low-latitude regions.

Introduction

Emerging developments in the physical sciences with regard to changing climate warrant careful examination by ecologists to determine whether our understanding of interacting ecological phenomena will be significantly affected. One such recent development in oceanography has been the study of climatological anomalies in sea surface temperature known as "marine heatwaves" (MHWs). Defined generally, MHWs are prolonged, discrete, anomalously warm periods in sea surface temperature at a particular location. Recent work has further standardized this definition to refer to periods of five or more days with temperatures greater than the 90th percentile, based on a historical climatology (Hobday et al., 2016). This standard definition, extended from prior work in atmospheric heatwaves, was developed partly in an effort to allow studies of the ecological effects of extreme climate events (ECEs) to proceed in a standardized way in marine ecosystems and to allow the comparison of similar studies in different regions at different times.

MHWs represent a class of climatological

anomaly hereafter referred to as an extreme climate event (ECE, see Intergovernmental Panel on Climate Change (2018)). The relative increase in frequency, intensity, and duration of ECEs is an established characteristic of anthropogenic climate change with significant demonstrated ecological impacts (Easterling et al., 2000; Ummenhofer and Meehl, 2017; van de Pol et al., 2017). A recent review of 519 observational studies of ecological responses to extreme events revealed widespread ecological changes in response to these events, with over 100 cases of a >25% population decline and 31 cases of local extirpation of a species; negative responses were characterized for 57% of studies examined, with some studies exhibiting recovery at decadal timescales (Maxwell et al., 2019). At ecosystem scale, extreme terrestrial events (e.g. heatwaves, precipitation anomalies) have led to significant changes in the carbon cycle of select systems like forests and grasslands, potentially resulting in decreased carbon storage and widespread plant die-off (Reichstein et al., 2013; Ummenhofer and Meehl, 2017).

Several case studies of the ecological effects of MHWs have emerged, mostly focused on individual events. For example, the 2015/16 Tasman Sea heatwave was shown to cause increased mortality in oyster, abalone, and salmon populations (Oliver et al., 2017), and the 2013-2015 heatwave in the northeast Pacific was associated with novel community compositions, an increase in warm-water copepod species in the Northern California region, and increased mortality of sea lions, whales, and sea birds (Cavole et al., 2016; Frölicher and Laufkötter, 2018). Some studies have attempted to generalize the effect of anomalous sea surface temperature events spatially: a 2015-16 pan-tropical coral bleaching event was shown to be caused primarily by these anomalies (Hughes et al., 2017). However, the foci of these and similar MHW studies are single events, and short time periods, leaving questions remaining about the long-term implications of extreme events (Bailey and van de Pol, 2016; van de Pol et al., 2017). Spatially and temporally integrative studies of the effects of ECEs are warranted for this reason.

In both marine and terrestrial systems, extreme climatological events can be mechanistically linked to organismal and ecological impacts via physiology. Foundational work to examine the causal drivers of ecological responses to climate warming in marine ecosystems has revealed physiological bases for these responses. For example, global patterns of phytoplankton diversity have been shown to be driven by temperature and environment (Righetti et al., 2019; Thomas et al., 2012), and differences in physiological plasticity and acclimation were shown to differentially buffer thermal safety in extreme environments in several taxa (Morley et al., 2019).

In this work we examine whether thermal physiology—in particular, thermal performance curves (TPCs, see Huey and Stevenson (1979)) of population growth rate in phytoplankton—can broadly link MHW events to ecological impacts. As noted, most evidence of the ecological impacts of MHW events is limited geographically and taxonomically (see Smale et al. (2019)). Thermal performance curves of population growth rate can be used to represent the reaction norm of fitness as a function of environment for a given organism generally. As such, we leverage phytoplankton TPC data as a systematic way to examine the potential ecological effects of MHW events for many related species across broad spatial scales. We use phytoplankton as our focal taxonomic group due to both their significant biogeochemical and trophic roles and the availability of robust, empirically-derived thermal performance curves of growth across myriad species from Thomas et al. (2012, 2016), who have extensively demonstrated the biogeographic patterns of thermal adaptation in phytoplankton. The significant biogeochemical role of phytoplankton in the global climate system (Falkowski, 1994) suggests important questions regarding the effect of MHW events on phytoplankton-mediated biogeochemical fluxes. Prior work has demonstrated that TPCs of functional traits of diatoms (such as growth rate, cell size, protein synthesis, and nutrient uptake) are a critical link between environment (temperature) and the biogeochemical role of phytoplankton (e.g. as major drivers of nutrient cycles) in their environment, motivating our choice to leverage TPCs to assess MHW impacts (Baker et al., 2016). In addition, the primary production role of phytoplankton in their trophic context suggests that phytoplankton responses to MHWs could be extrapolated to higher trophic levels. The precise nature of this extrapolation is difficult to precisely determine, but previous work has shown that changes in trophic dynamics as a result of warming events are controlled by temperature dependent effects on herbivory, for example, which is relevant to assessments of net productivity (O'Connor et al., 2011).

We seek to examine whether patterns exist in the physiological responses of phytoplankton to MHW events. We base this study on a 38-year record of global, high-resolution sea surface temperature data to detect and characterize MHW events, and leverage empirically-derived thermal performance curves of growth rate across a variety of globally-distributed

species to examine differences in phytoplankton performance as a function of MHWs. We examine whether latitudinal and seasonal patterns exist in MHW responses via a statistical modeling approach. We also attempt to validate our findings using remotely-sensed measurements of Chlorophyll-a. While TPCs of population growth rate may not necessarily be the finest physiological measure with which to assess whole-organism responses to extreme climate events (e.g. due to interacting effects of reproduction, nutrient availability, seasonality, etc.; see Baker et al. (2016)), we are able to take advantage of globally-distributed, empirically-derived TPC data for this study, which could illuminate global patterns and suggest future avenues for further research.

Methods and Data

Marine Heatwave Detection

To detect marine heatwaves, and to enable subsequent linkage of environmental observations to thermal performance, we employed the Optimal Interpolation Sea Surface Temperature (OISST, Version 2) product from the National Oceanic and Atmospheric Administration, acquired via the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from `https://psl.noaa.gov/`. These data are high-resolution (1/4 degree) daily sea surface temperature measurements from 1981-present, derived by interpolating measurements from various observational platforms (satellites, ships, buoys, oceanographic floats), accounting for bias in each platform (Banzon et al., 2016; Reynolds et al., 2007). Access, storage and subsequent analysis of these data was completed in the Python programming language (v. 3.7.6; Python Software Foundation, `http://www.python.org`) via the xarray library Hoyer and Hamman (2017). Computational and data storage resources were facilitated by Amazon Web Services (AWS; `https://aws.amazon.com`).

Actual detection of marine heatwaves (MHWs) was performed using the xarray (Hoyer and Hamman, 2017) and marineHeatWaves Python packages (Hobday et al., 2016); `https://github.com/ecjoliver/marineHeatWaves`). We performed this detection with default settings at 1/4 degree grid-cell granularity using daily OISST records as input to the detection code across a 38-year period (1981 - 2019). Each MHW was described by the metrics outlined in Table 3.1 (p. 48). The software to compute these detections, and to complete all subsequent analyses, is available at `https://github.com/huckleylab/phyto_mhw`.

Plankton Physiology

Empirically Derived Thermal Performance Curves

We used thermal performance curves (TPCs) from Thomas et al. (2012) and Thomas et al. (2016) to quantify the effect of marine heatwaves on the thermal physiology of phytoplankton. These data contained parameterizations of a population growth rate thermal reaction norm function for 474 geographically-distinct isolates of estuarine ($N = 43$), freshwater ($N = 214$), and marine ($N = 206$) phytoplankton species. We subsequently refer to these reaction norms of population growth rate and TPCs interchangeably. Growth rate TPCs for each isolate were of the form

$$g(T) = ae^{bt}\left[\left(\frac{z-T}{\omega/2}\right)^2\right] \tag{3.1}$$

where niche width (e.g. the range of temperatures over which growth rate is positive) is represented by ω, and z determines the location of the maximum of the quadratic portion of this function. When $b = 0$, this value is identical to the temperature at which a species achieves its maximum growth rate. However, when b is non-zero, the maximum value of Eq. 3.1 falls above (or potentially below, $b < 0$), and is found via numerical optimization. A maximum-likelihood (MLE) curve fitting approach was used to derive these parameters such that mean growth rate at a given temperature T followed the equation:

$$\mu = g(T) + \mathcal{N}(0, \sigma^2) \tag{3.2}$$

This MLE approach was applied to a data set of laboratory-derived empirical growth rate measurements for each isolate from the literature by Thomas et al. Only growth rate measurements fitting a standard set of experimental procedures were used (grown under constant conditions [to eliminate time-dependent effects of acclimation], salinity between 30-40 ppt, light levels greater than or equal to 100 microeinsteins$\cdot m^2 \cdot s^{-1}$, not experimentally nutrient-limited, day length $\geq$ 10h). TPC parameters were fit to these data for each isolate using a maximum likelihood numerical approximation approach (see Thomas et al. (2016, 2012) for data selection details). For this study we selected only marine isolates with "high-quality" TPCs (as determined from

Metric	Notation	Definition	Units
Climatological Mean	T_m	$T_m(j) = \sum_{y=y_s}^{y_e} \sum_{d=j-5}^{j+5} \frac{T(y,d)}{11(y_e-y_s+1)}$ where j is day of year, y_s and y_e are the start and end of the climatological base period, and T is the daily SST on day d of year y.	°C
90th Percentile Threshold	T_{90}	$T_{90}(j) = P_{90}(X)$ where P_{90} is the 90th percentile and $P_{90}(X)$ where $X = \{T(y,d) \mid y_s \leq y \leq y_e, j-5 \leq d \leq j+5\}$	°C
Start Date	t_s	t where $T(t) > T_{90}(j)$ and $T(t-1) < T_{90}(j)$	days
End Date	t_e	t where $t_e > t_s$, $T(t) < T_{90}(j)$, and $T(t-1) > T_{90}(j)$	days
Duration	D	$D = t_e - t_s$ (For MHWs, $t_e - t_s \geq 5$ and gap ≤ 2 days)	days
Maximum Intensity	$i_{\max}$	$i_{\max} = \max(T(t) - T_m(j))$	°C
Mean Intensity	i_{mean}	$i_{\text{mean}} = \overline{T(t) - T_m(j)}$	°C
Cumulative Intensity	i_{cum}	$i_{\text{cum}} = \int_{t_s}^{t_e-1} (T(t) - T_m(j))\, dt$	°C days

Table 3.1: Characteristic metrics of marine heatwaves from Hobday et al. (2016). T represents daily OISST data, and is written both as a function of time t (e.g. $T(t)$) and as a function of year y and day-of-year d (e.g. $T(y,d)$).

the distribution of $T_{\min}$, $T_{\max}$, and ω; see Thomas et al. (2016)). Our final TPC data set contained $N = 89$ curves for isolates distributed between latitudes 74.83°S and 76.28°N (see Figure 3.1). These data encompassed individual species isolates across 8 groups, 41 genera, and 48 individual species.

Biogeographic Physiological Performance Analysis

We performed several analyses to elucidate the effects of marine heatwave events on phytoplankton physiology.

Characterizing Phytoplankton MHW Responses

For each phytoplankton isolate ($N = 89$), we computed a thermal performance (growth rate) time series by applying the TPC associated with each isolate to the 38-year (1981-2019) daily OISST record from the grid cell nearest to the given isolate's isolation location. We refer to this location as "the isolate's location," and refer to all MHW events at this location as "experienced by" the organism represented by each isolate. Using co-located and contemporaneous MHW detections described above, we filtered this time series to contain only data that overlaps MHW events. For each MHW event i experienced by isolate k, we computed metrics of physiological performance defined in Table 3.

To control for differences in maximum growth rate magnitude among the isolates, we scaled each isolate's computed growth rate from a possible growth rate range of $[-P_k(T_{\text{opt},k}), P_k(T_{\text{opt},k})]$ to the range $[-1, 1]$ via a linear scaling as below, where P_k is the thermal reaction norm of growth rate for isolate k:

$$P_{i,j}(t)_{\text{scaled}} = \frac{(P_{i,j}(t) - (-P_k(T_{\text{opt},k}))) (1 - (-1))}{(P_k(T_{\text{opt},k}) - (-P_k(T_{\text{opt},k})))} \times -1 \quad (3.3)$$

We used this scaled performance defined above to examine whether seasonal or latitudinal patterns exist in phytoplankton thermal responses to MHWs. To assess latitudinal patterns, we first averaged the scaled $\delta P_{i,j}$ metric for each MHW event in the grid cell nearest to each isolate. We then aggregated these means for each isolate by computing a mean and standard deviation across all MHW events experienced in each isolate's grid cell across the 38-year record period. We also aggregated these data seasonally (astronomical seasons, e.g. Spring: 20 March – 19 June; Summer: 20 June – 21 September, Fall: 22 September – 20 December; Winter: 21 December – 19 March) by computing means and standard deviations for MHWs experienced within each season for each isolate. During our seasonal aggregation, an off-

Metric	Notation	Definition	Units
Isolate Performance	P_k	$P_k(T) = g_k(T)$ where g_k is the thermal reaction norm of population growth rate for isolate k (Eq. 3.2).	Growth Rate
Mean Performance	$P_{k,m}$	$P_{k,m}(j) = P_k(T_m(j))$ where j is day of year and $T_m(j)$ is the climatological mean temperature at day j (see Table 3.1).	Growth Rate
MHW Performance	$P_{k,i}$	$P_{k,i}(t) = P_k(T(t))$ for $t_{s,i} \leq t < t_{e,i}$. P_k is the thermal performance curve for isolate k. t_s and t_i are the start and end of MHW event i, respectively.	Growth Rate
Performance Difference	$\Delta P_{k,i}$	$\Delta P_{k,i}(t) = P_{k,i}(t) - P_{k,m}(d)$ where d is the day of year at time t	Growth Rate

Table 3.2: Physiological performance metrics associated with marine heatwave event i for a given phytoplankton isolate k. See Table 3.1 for complete definitions of MHW metrics.

set of 180 days was applied to MHW start, end, and peak dates for all isolates with locations in the southern hemisphere (e.g. latitudes 90°S–0°) to allow for comparison with northern hemisphere isolates.

Statistical Analyses

We sought to determine whether any significant explanatory relationships existed among latitude, season, and phytoplankton fitness responses to MHWs. To do this we used the mean scaled performance difference ($P_{i,j}(t)_{\text{scaled}}$ see Eq. 3 and Table 3) of each isolate during each MHW event detected at the location nearest to the isolate's collection location, as above. Because each thermal performance curve used to compute these performance ratios represents a single geographically-distinct phytoplankton isolate with unique fitness responses, we consider each of these isolates to be contributing variation to our observations irrespective of the phenomena we seek to examine. To account for this variation in response among isolates we used a linear mixed effects model with isolate as a random effect, and with fixed effects of the absolute value of latitude, factor(season), local yearly mean sea surface temperature (to incorporate any effect of "warm" and "cool" years), and all interacting terms. In addition, to examine whether "stronger" MHW events revealed more significant relationships, we performed two fits of this linear mixed effects model: one containing all MHW events for all isolates, and another containing only those MHW events corresponding to a 1.5°C increase over the climatological mean temperature (e.g. $i_{\text{mean}} > 1.5$°C, see Table 3.1). This threshold was chosen due to its biological, climatological, and public-policy significance as the minimum achievable magnitude of warming given current climate and economic trajectories as outlined in reports such as the IPCC "Global Warming of 1.5C" report (IPCC 2018). Models were fit using the R programming language (R Core Team, 2019) via the lme4 (Bates et al., 2015, p. 4) and lmerTest (Kuznetsova et al., 2017) packages.

Validation

In an effort to validate the predicted effects of MHW events on phytoplankton fitness we leveraged remotely-sensed measurements of chlorophyll-a as a proxy for phytoplankton presence. Chlorophyll-a (Chl-a) is a pigment present in phytoplankton, and has well-studied optical characteristics that have long been employed to assess phytoplankton abundance in coastal and ocean waters (e.g. Yentsch, 1960). Our Chl-a data were obtained via the Ocean and Land Color Instrument (OLCI) onboard the Sentinel-3A/B satellites, an orbiting platform jointly operated by the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The OLCI is a push-broom imaging spectrometer with 21 spectral bands, each measured at 300m/pixel. Together, the Sentinel 3A and 3B satellites provide less than two day equatorial revisit times. The Sentinel-3 mission provides OLCI-derived chlorophyll-a observations as a Level-2 data product via the OC4Me algorithm (Fletcher and Agency, 2012; O'Reilly et al., 1998).

To compare chlorophyll-a concentrations to phytoplankton performance during MHW events, we

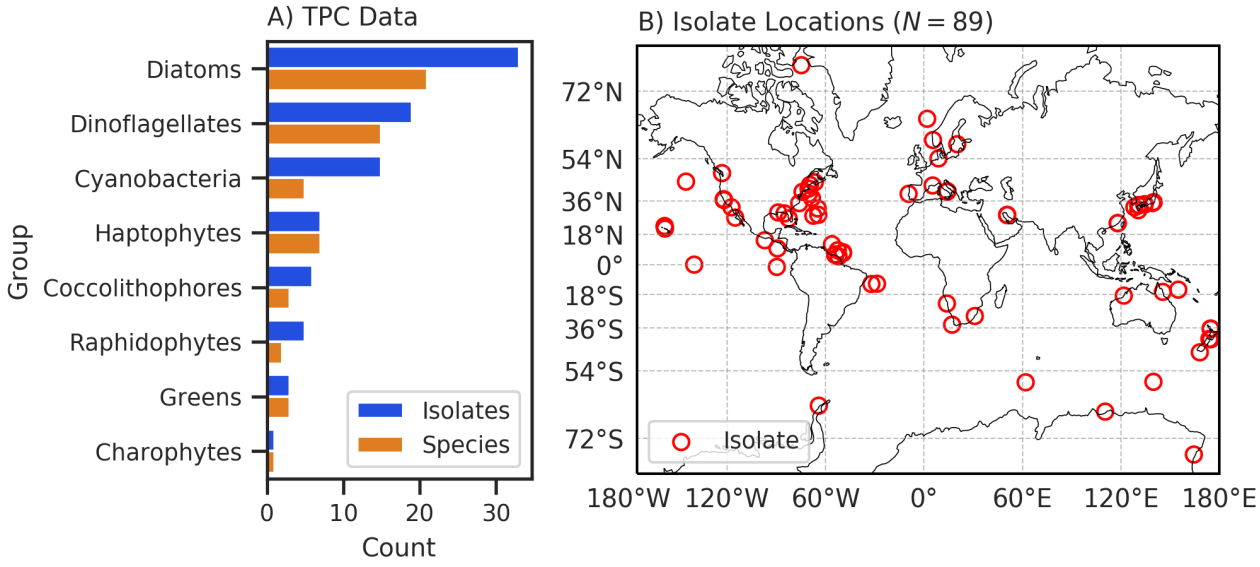

Figure 3.1: Group distribution (A) and geographic locations (B) of phytoplankton isolates ($N = 89$) with associated thermal performance curves of growth. Data are from Thomas et al. (2012, 2016) and were limited to only those observations collected in marine regions and with TPC parameters matching expected distributions.

identified a single phytoplankton isolate (*Pseudo-nitzschia granii*, a diatom) with a collection location in the open ocean (northeast Pacific, approx. 45°N, 145°W; to avoid the detrimental effects of turbidity in coastal waters). We then selected a two-year time period containing both MHW periods and non-MHW periods to allow for seasonal variation in chlorophyll concentrations. We obtained 846 OLCI Level 2 scenes at reduced resolution (Product Type: "`OL_2_WRR___`", 1.2 km/pixel, to ease computational burden) overlapping the chosen isolate collection location during this chosen time period (2016-2017; data from 2018 not available at time of analysis due to an unknown error). After removing cloudy scenes via a Level 2 cloud mask provided with each scene 205 scenes were available for analysis. We selected pixels within a circular buffer (20 km radius) around the isolation location and computed the mean chlorophyll-a concentration among these pixels. To account for the potential for time lag in phytoplankton (Chl-a) responses to temperature changes, we apply a 1-day lag when comparing isolate performance and chlorophyll-a concentrations such that each chlorophyll-a observation is paired with a phytoplankton performance measurement from one day prior to the Chl-a measurement.

Results

Opposing Patterns in MHW Physiological Responses

We observed several patterns of phytoplankton physiological responses to marine heatwave events. As illustrated in Figure 3.2, we found that MHW events can precipitate both positive fitness responses (e.g. $\Delta P_{k,i} > 0$) and negative fitness responses ($\Delta P_{k,i} < 0$) of varying magnitudes. In this example, two heatwaves of similar mean intensity (3.57°C vs 3.77°C) cause physiological responses that differ in both magnitude and direction (mean performance differences of 0.14 and -0.05, respectively). These fitness responses additionally have a spatial component, largely aligned with the geographic footprint of the thermal anomaly.

Seasonal and Latitudinal Patterns

Both seasonal and latitudinal patterns in phytoplankton physiological performance in response to MHWs were evident in our results. By aggregating all MHW events experienced by each isolate (Figure 3.3), we observed a general pattern of positive fitness responses in the temperate regions and poles, with considerably more mixed responses in the tropics and subtropics. These mixed responses appear to vary considerably by season, with summer events

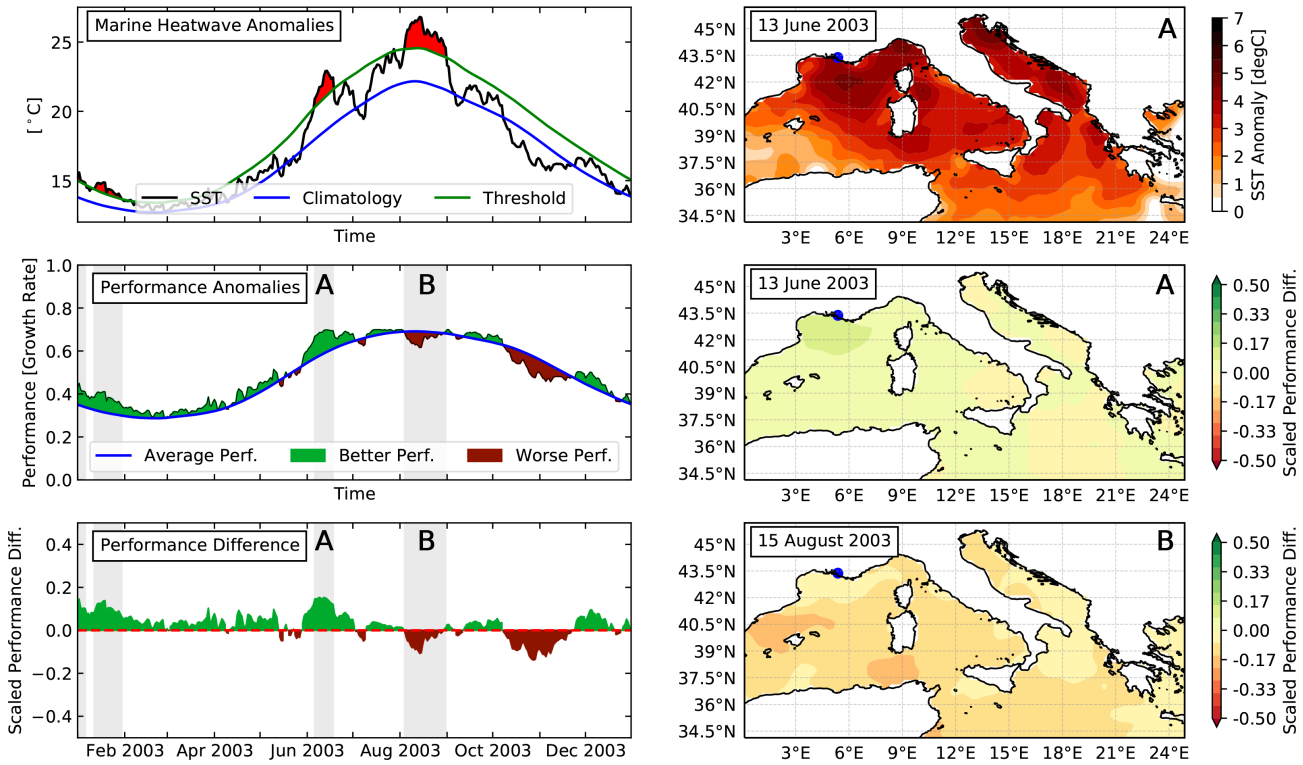

Figure 3.2: Two heatwaves during the 2003 Mediterranean thermal anomaly period illustrating opposing fitness responses in a single phytoplankton isolate dependent on seasonal timing (*Prorocentrum minimum*, a dinoflagellate). Left column depicts oceanographic (Marine Heatwave Anomalies) and thermal-fitness (Performance Anomalies, Performance Difference) manifestations of two MHW events (June and August; labeled **A** and **B**) between Jan 2003–Jan 2004; right column depicts the spatial distribution of (top to bottom) SST Anomaly, scaled performance difference of MHW event **A**, and scaled performance difference of MHW event **B**.

often causing more negative fitness responses than winter or spring events, which themselves exhibit improvements in fitness.

Characterization of Seasonal and Latitudinal Relationships

Statistical modeling revealed significant seasonal patterns in phytoplankton physiological responses to MHW events that interact with latitude. Among the two linear mixed effects models, both models contained significant effects of latitude and MHW-induced fitness difference ($p < 0.001$, Figure 3.4, Table 3A1, p. 60). Significant and contrasting effects of season on fitness difference were also found, with summer events causing negative fitness effects ($p < 0.001$) and winter events causing positive fitness effects ($p < 0.001$). Both models also demonstrate significant interacting relationships between latitude and season. In particular, winter MHW events show positive fitness responses with increasing latitude (e.g. toward the polar regions), in contrast to summer events, which trend toward negative, higher-magnitude fitness responses with increasing latitude (Figure 3.5, Table 3A1). (Significant coefficients between latitude and season also were found in seasonal models see Table 3A2, p. 61).

Validation

During the 2016-2017 validation period, *Pseudo-nitzschia granii* experienced three marine heatwave events in 2016 and none in 2017. The 2016 events occurred August-September, led to negative fitness responses, and are correlated with a decrease in Chl-a concentrations (Figure 3.5). In contrast, during the same period in 2017, a lack of MHW events led to minimal deviations in performance from the climatological mean and no apparent trend in Chl-a concentration changes. When compared directly, there was a weak positive relationships between scaled Pseudo-nitzschia granii performance difference and Chlorophyll-a concentrations (Figure 3.6; OLS linear fit, $p = 0.02$; $R^2 = 0.03$, Pearson correlation = 0.17). Chl-a concentrations were highlight variable during non-MHW conditions, making it difficult to confidently detect fitness responses to MHWs. How-

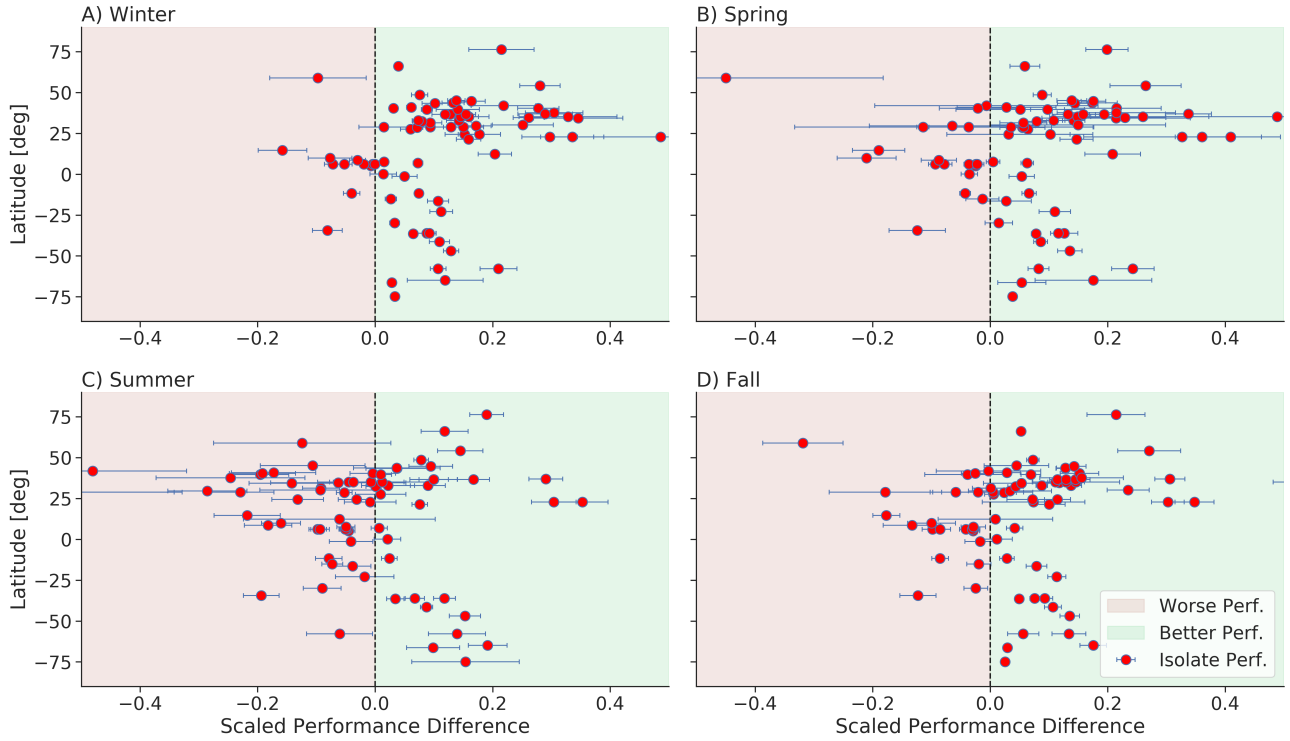

Figure 3.3: Aggregated scaled performance difference of phytoplankton isolates ($N = 86$) during marine heatwave events ($N = 5299$). We show mean and standard deviation of the aggregate (mean) performance difference for all MHW events during 38 years of the OISST record, computed by independently evaluating physiological performance of each isolate during MHW events detected at the location nearest to the isolate collection location.

ever, Chl-a conditions are quite consistently low during periods of predicted low performance associated with MHW events (Figure 3.7).

Discussion

In this first attempt to mechanistically generalize the ecological impacts of marine heatwave events via phytoplankton fitness responses, our findings suggest avenues to understanding the underlying mechanisms driving whole-ecosystem responses to these extreme climate events. We have observed latitudinal and seasonal patterns in phytoplankton fitness responses that, when placed in biogeochemical and trophic context, may enhance our ability to forecast large-scale ecological responses.

Latitudinal Patterns in Responses

Our findings align generally with existing observations of phytoplankton thermal performance. In particular, we identify enhanced vulnerability to MHW events at tropical and subtropical latitudes (Figure 3.3). This finding corresponds with other work demonstrating that phytoplankton thermal niche width is narrowest in the tropics, suggesting a higher vulnerability to extremes in those regions (Thomas et al., 2016). In addition, the coupling of this narrow niche breadth with the small range of temperature variability in tropical and subtropical region suggests that small increases in temperature nearer at or above T_{opt} could lead to significant fitness responses. We also observe increases in fitness in the polar regions, suggesting that phytoplankton populations in these regions are potentially less vulnerable to MHW events than those in the tropics. Marine species range limits have been shown to generally correspond with limits predicted via physiological thermal constraints at both high and low latitudes, which suggests that abiotic constraints may play a role in phytoplankton fitness across latitudinal gradients Pinsky et al. (2020); Sunday et al. (2012). However, caution must be taken to avoid over-extrapolating these latitudinal and seasonal patterns, as the patterns of impact of extreme events have been shown to be highly variable globally (Buckley and Huey, 2016). More information about the context of these thermal physiological shifts, discussed below, is necessary to make stronger claims about the patterns of ecological MHW responses.

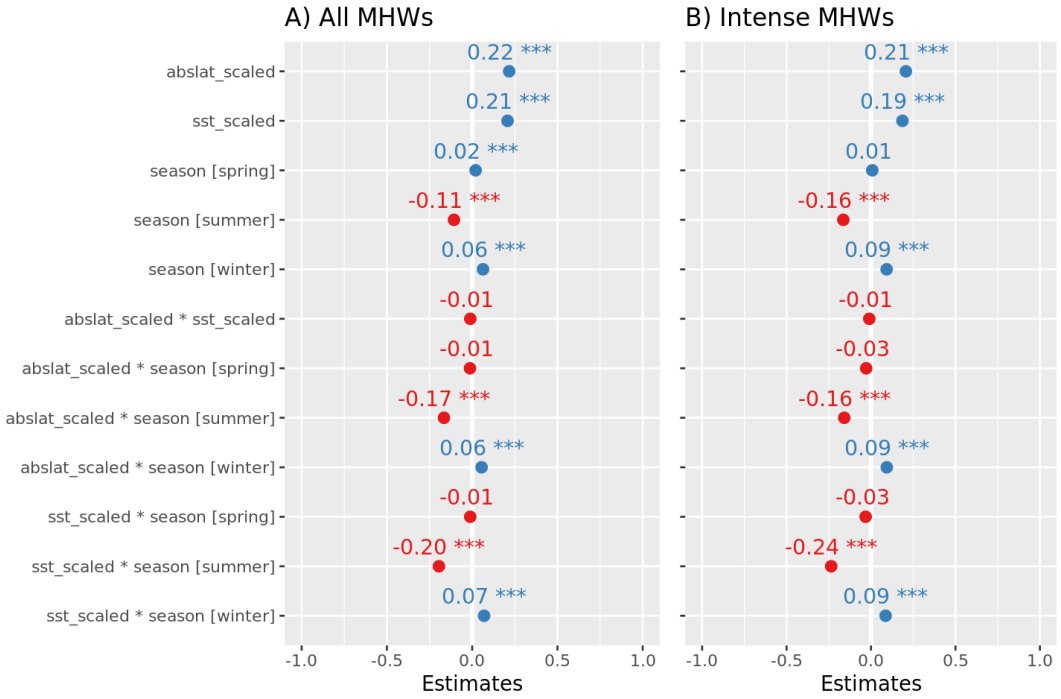

Figure 3.4: Linear mixed effects model coefficients for two models of scaled phytoplankton performance difference during marine heatwave events. The first model (panel A) contains all MHW events experienced by all isolates ($N = 5299$), and the second model contains only those during intense MHWs ($i_{\text{mean}} > 1.5°$ C, $N = 2721$). "absolat_scaled" = Scaled Absolute Latitude; "sst_scaled" = scaled annual mean sea surface temperature.

Ecological Context

We have shown here that phytoplankton experience a range of fitness responses to marine heatwave events, from positive to negative, with varying degrees of magnitude (Figures 3.2 and 3.3). This wide breadth suggests a similarly broad range of whole-ecosystem responses to MHW events due to the basal tropic context of phytoplankton, though the precise nature of these responses is likely to vary considerably. The primary basis we have for extrapolating these temperature-mediated growth rate changes to ecological responses is theoretical: prior modeling work has shown negative or negligible dependence of herbivore and plant abundance on temperature, despite positive effects of temperature on growth rates. In addition, sudden changes in trophic stability were observed under particular ecological and environmental conditions, potentially leading to herbivore extinction despite increased primary production (O'Connor et al., 2011). These effects are both primarily associated with resource limitation, a primary driver of organismal fitness discussed below. Relatedly, changes in the biogeochemical role of phytoplankton in the global carbon cycle could be extrapolated from these results. MHW-mediated changes in phytoplankton growth likely alter net primary production, leading to subsequently altered rates of nutrient uptake. Taken together at global scale, these changes could be significant enough to change carbon uptake regimes as MHW events become longer and more frequent (Oliver et al., 2018).

Validation of Performance Effects

We observed weak support for our hypothesis that phytoplankton thermal responses to MHW events would be detectable in remotely-sensed Chlorophyll-a concentrations in the open ocean. Though our findings show modest declines in Chl-a concentrations during MHW events (Figures 3.6, 3.7), we cannot conclude that these observed effects are as a result of the events themselves. Our validation study was limited to a single location across two years of observations due to the relative recency of Sentinel-3 data availability (2016), a data gap in 2018 preventing our access to these observations, and our desire to use an isolate in the open ocean, of which there are few. The high variability of Chlorophyll-a concentrations (due to factors unstudied here such as nutrient availability or ocean currents) necessitates a considerably larger set of observations, evenly distributed across time, nutrient regimes, and other controls, to allow for the attribution of detrimental fitness responses to

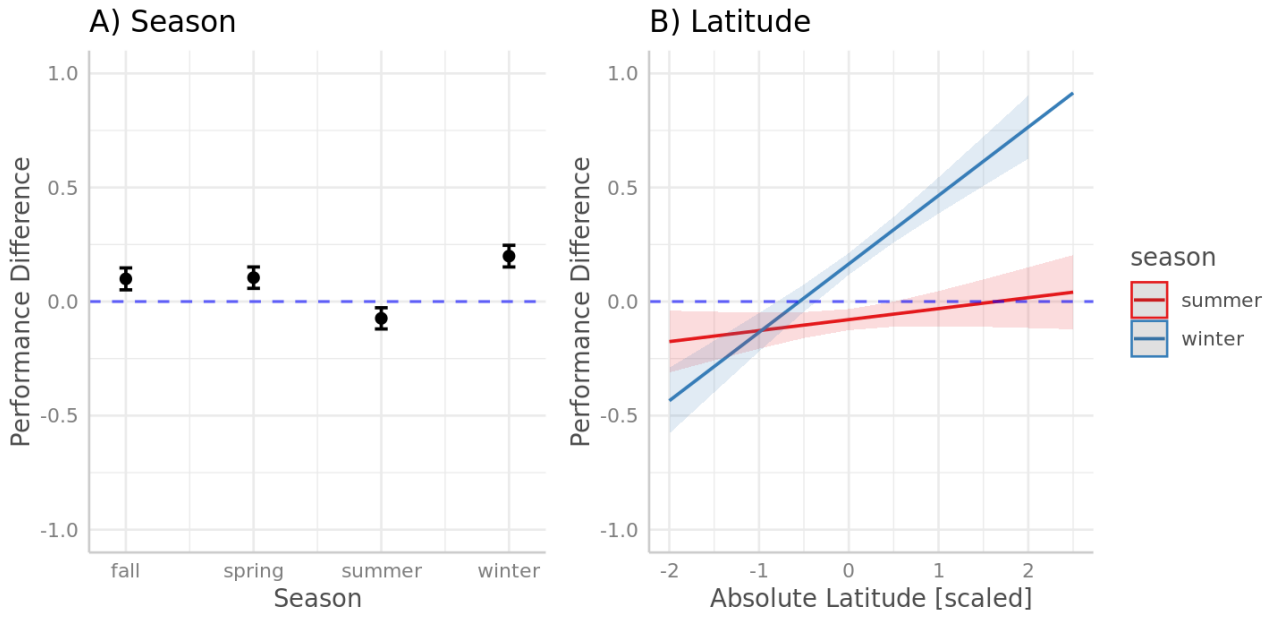

Figure 3.5: Marginal effects of season and latitude on scaled phytoplankton performance difference during MHW events for linear mixed effects model of intense heatwaves ($i_{\text{mean}} > 1.5°\text{C}$). Summer and Winter responses are significant ($p < 0.001$, Table 3A1, p. 60), as are latitude*summer and latitude*winter interaction terms ($p < 0.001$; Table 3A1).

changes in Chl-a concentrations. As discussed below, thermal physiology is one of many facets contributing to phytoplankton fitness reponses to environmental change, and these other confounding factors likely influence our ability to observe performance changes via remotely-sensed Chl-a concentrations.

Observed Limitations & Future Work

This study has several important limitations to consider, which are opportunities for future work. We detail these in Table 3.3 and in the following section. We have made the fundamental assumption that the thermal reaction norms of phytoplankton growth rate represent a reliable mechanism with which to assess temperature-dependent fitness of these organisms in an ecologically-relevant manner. While others have successfully demonstrated the soundness of this assumption (e.g. Buckley and Huey, 2016; Thomas et al., 2016, 2012), it is important to acknowledge the possible effects of other physiological, environmental, and ecological influences on phytoplankton fitness and population persistence in order to fully assess the impacts of MHWs on these organisms and whole ecosystems, and we attempt to do so below. See Sinclair et al. (2016) for a more complete review of the simplifying assumptions associated with using TPCs to predict fitness responses to environmental variability.

For example, the quantity and type of available nutrients is a primary and significant environmental control over phytoplankton fitness, and is likely to vary negatively with temperature (Falkowski et al., 1998) and with changes in biogeochemical cycles due to climate change (Hutchins and Boyd, 2016). The relationship of temperature and nutrient availability in particular has been shown to decrease T_{opt} for phytoplankton species, thus making phytoplankton species considerably more vulnerable to warming (Thomas et al., 2017). Thus, the effect of MHW events may shift "down," leading to more frequent negative fitness responses of higher magnitude and fewer, lower-magnitude positive fitness responses than observed here. Further study is warranted to both develop more general mechanistic understandings of phytoplankton fitness dependence on temperature and nutrients and to apply these models to MHW events. However we expect that the general latitudinal and seasonal patterns of phytoplankton responses will remain consistent with our findings.

Additionally, it is prudent to consider the possibility of physiological plasticity and the evolution of thermal tolerance in phytoplankton as potentially contrasting or compounding influences on MHW response. For example, our observation of the significant seasonal effects of phytoplankton MHW responses depends upon an assumption of no plasticity among thermal responses between seasons. While relatively little is known about seasonal plasticity in temperature-mediated responses to growth

Limitation	Type	Notes
Unobserved variability in water column temperature	Environmental Data	The NOAA OI Sea Surface Temperature data used in this study is representative of the first vertical 0.5 m of the water column. As phytoplankton occupy the majority of the photic zone (~200 m), these temperature measurements are only relevant for a portion of the phytoplankton population. However, due to factors like wave mixing, the general trend of warming or cooling responses across the photic zone are likely to correlate with measured temperatures.
Phytoplankton adaptive evolution	Physiology	Phytoplankton thermal optima are well-correlated with local environmental conditions (Thomas et al., 2012), suggesting an evolutionary process of adaptation. Any future extrapolation must consider the effect of selection pressures leading to evolution of thermal tolerance traits.
Extrapolation of constant-conditions TPCs	Physiology	The phytoplankton thermal reaction norm parameterizations used herein are derived from laboratory experiments where phytoplankton temperature responses were examined under otherwise constant conditions. However, phytoplankton may exhibit plasticity in growth rate responses to temperature depending on other abiotic
Geographically disparate role of temperature in describing phytoplankton richness patterns	Physiology	Righetti et al., (2019) show that our core assumption of temperature as the most important driver of phytoplankton fitness is likely only relevant in particular regions (e.g. the 60% of the ocean with mean annual temperatures $>$ 19°C). They suggest that environmental *variability* may be a strong determinant of species richness patterns and therefore individual species growth patterns.
Latitudinal patterns of nutrient availability	Physiology	Strong latitudinal patterns exist in nutrient availability, particularly at mid- to high-latitudes, which are likely to interact with temperature to affect the latitudinal patterns of MHW responses observed here.
Narrow seasonal validation window	Validation Data	The Sentinel-3A/B satellites, launched in 2016, provide a relatively short temporal catalog for validation. As such, we present a comparison across two years' worth of data, containing only a single MHW event. Further investigation should examine whether Chlorophyll-a is correlated to physiological responses in other seasons, which can likely be accomplished in other geographic locations.

Table 3.3: Explanations of limitations of this study to be addressed in future work. We categorize these limitations by the domain of the study they are pertinent to.

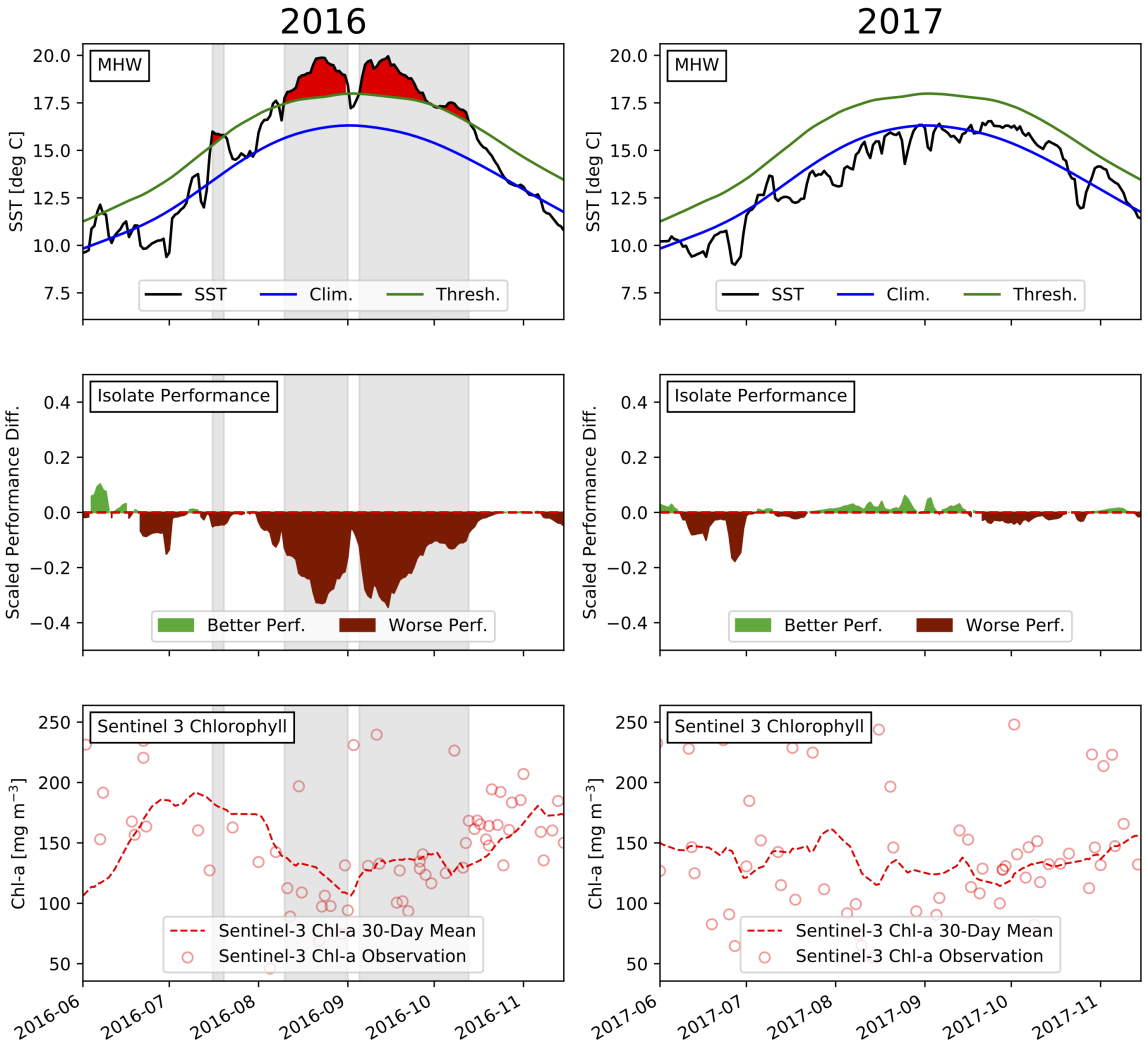

Figure 3.6: Remotely-sensed chlorophyll-a validation of *Pseudo-nitzschia granii* performance in the northeast Pacific (45 N, 145 W) during MHW events and non-MHW periods. Sentinel-3 estimates of Chlorophyll-a are derived from the OLCI instrument via the OC4Me algorithm; only cloud-free pixels within a circular buffer of radius 20 km around the isolation location are used in analysis. MHW regions are highlighted in bright red (top panel) and gray shading (all panels).

among phytoplankton, previous work in other taxa has demonstrated no apparent relationship between latitude or thermal seasonality and plasticity in heat tolerance, but some plasticity in cold tolerance exists, with more pronounced plastic effects observed in marine taxa (Gunderson and Stillman, 2015). We hypothesize that ocean regions with high environmental variability contain phytoplankton species that exhibit a similar plasticity regime, potentially altering the general pattern of seasonal responses. In addition, the interacting effects of evolution, environmental variability, and plasticity are important to consider, as the magnitude of seasonal variability can change the impact of plasticity on the rate and direction of trait evolution (Kingsolver and Buckley, 2017). Phytoplankton physiological traits are well-adapted to local conditions (Thomas et al., 2016), and some modeling attempts have demonstrated environmentally-driven variation in the genetic capacity for thermal adaptation among phytoplankton groups (Huertas et al., 2011; Thomas et al., 2012), but much remains unknown about the precise evolutionary dynamics of thermal tolerance in phytoplankton and should be subject to further investigation.

In summary, we have for the first time identified marine heatwave events as significant drivers of physiological fitness responses in phytoplankton, with potential ecosystem-level effects. While the precise nature of these effects remains uncertain, the global ecological response to MHW events will likely exhibit latitudinal and seasonal patterns due to the dif-

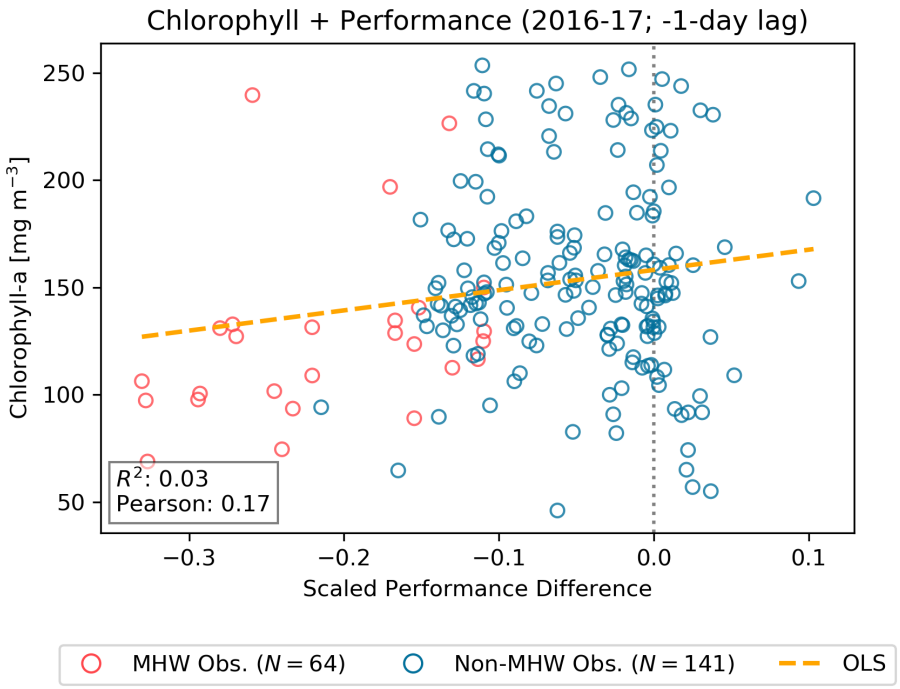

Figure 3.7: Comparison of scaled *Pseudo-nitzschia granii* performance difference and remotely-sensed chlorophyll-a concentrations. Performance difference values from one day prior to chlorophyll-a concentration measurements are shown to account for lag time in phytoplankton response to temperature. OLS performance difference coefficient is weakly significant ($p = 0.02$).

ferential responses of phytoplankton across latitude and season. We hope that these findings will lay the groundwork for assessing the ecological impacts of marine heatwave events into the future.

References

Bailey, L. D. and van de Pol, M. (2016). Tackling extremes: challenges for ecological and evolutionary research on extreme climatic events. *Journal of Animal Ecology*, 85(1):85–96.

Baker, K. G., Robinson, C. M., Radford, D. T., McInnes, A. S., Evenhuis, C., and Doblin, M. A. (2016). Thermal Performance Curves of Functional Traits Aid Understanding of Thermally Induced Changes in Diatom-Mediated Biogeochemical Fluxes. *Frontiers in Marine Science*, 3.

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., and Hankins, W. (2016). A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth System Science Data*, 8(1):165–176.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Buckley, L. B. and Huey, R. B. (2016). Temperature extremes: geographic patterns, recent changes, and implications for organismal vulnerabilities. *Global Change Biology*, 22(12):3829–3842.

Cavole, L., Demko, A., Diner, R., Giddings, A., Koester, I., Pagniello, C., Paulsen, M.-L., Ramirez-Valdez, A., Schwenck, S., Yen, N., Zill, M., and Franks, P. (2016). Biological Impacts of the 2013–2015 Warm-Water Anomaly in the Northeast Pacific: Winners, Losers, and the Future. *Oceanography*, 29(2).

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000). Climate Extremes: Observations, Modeling, and Impacts. *Science*, 289(5487):2068–2074.

Falkowski, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis Research*, 39(3):235–258.

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, 281(5374):200–206.

Fletcher, K. and Agency, E. S., editors (2012). *Sentinel-3: ESA's global land and ocean mission for GMES operational services*. Number 1322/3 in ESA SP. ESA Communications, Noordwijk.

Frölicher, T. L. and Laufkötter, C. (2018). Emerging risks from marine heat waves. *Nature Communications*, 9(1):1–4.

Gunderson, A. R. and Stillman, J. H. (2015). Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society B: Biological Sciences*, 282(1808):20150401.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Feng, M., Holbrook, N. J., Moore, P. J., Scannell, H. A., Sen Gupta, A., and Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, 141:227–238.

Hoyer, S. and Hamman, J. J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, 5:10.

Huertas, I. E., Rouco, M., López-Rodas, V., and Costas, E. (2011). Warming will affect phytoplankton differently: evidence through a mechanistic approach. *Proceedings of the Royal Society B: Biological Sciences*, 278(1724):3534–3543.

Huey, R. B. and Stevenson, R. D. (1979). Integrating Thermal Physiology and Ecology of Ectotherms: A Discussion of Approaches. *Integrative and Comparative Biology*, 19(1):357–366.

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., Babcock, R. C., Beger, M., Bellwood, D. R., Berkelmans, R., Bridge, T. C., Butler, I. R., Byrne, M., Cantin, N. E., Comeau, S., Connolly, S. R., Cumming, G. S., Dalton, S. J., Diaz-Pulido, G., Eakin, C. M., Figueira, W. F., Gilmour, J. P., Harrison, H. B., Heron, S. F., Hoey, A. S., Hobbs, J.-P. A., Hoogenboom, M. O., Kennedy, E. V., Kuo, C.-y., Lough, J. M., Lowe, R. J., Liu, G., McCulloch, M. T., Malcolm, H. A., McWilliam, M. J., Pandolfi, J. M., Pears, R. J., Pratchett, M. S., Schoepf, V., Simpson, T., Skirving, W. J., Sommer, B., Torda, G., Wachenfeld, D. R., Willis, B. L., and Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645):373–377.

Hutchins, D. A. and Boyd, P. W. (2016). Marine phytoplankton and the changing ocean iron cycle. *Nature Climate Change*, 6(12):1072–1079.

Intergovernmental Panel on Climate Change (2018). *Global warming of 1.5° C*.

Kingsolver, J. G. and Buckley, L. B. (2017). Evolution of plasticity and adaptive responses to climate change along climate gradients. *Proceedings of the Royal Society B: Biological Sciences*, 284(1860):20170386.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13):1–26.

Maxwell, S. L., Butt, N., Maron, M., McAlpine, C. A., Chapman, S., Ullmann, A., Segan, D. B., and Watson, J. E. M. (2019). Conservation implications of ecological responses to extreme weather and climate events. *Diversity and Distributions*, 25(4):613–625.

Morley, S. A., Peck, L. S., Sunday, J. M., Heiser, S., and Bates, A. E. (2019). Physiological acclimation and persistence of ectothermic species under extreme heat events. *Global Ecology and Biogeography*, 28(7):1018–1037.

O'Connor, M. I., Gilbert, B., and Brown, C. J. (2011). Theoretical Predictions for How Temperature Affects the Dynamics of Interacting Herbivores and Plants. *The American Naturalist*, 178(5):626–638.

Oliver, E. C. J., Benthuysen, J. A., Bindoff, N. L., Hobday, A. J., Holbrook, N. J., Mundy, C. N., and Perkins-Kirkpatrick, S. E. (2017). The unprecedented 2015/16 Tasman Sea marine heatwave. *Nature Communications*, 8(1):1–12.

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng, M., Gupta, A. S., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Straub, S. C., and Wernberg, T. (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications*, 9(1):1–12.

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans*, 103(C11):24937–24953.

Pinsky, M. L., Selden, R. L., and Kitchel, Z. J. (2020). Climate-Driven Shifts in Marine Species Ranges: Scaling from Organisms to Communities. *Annual Review of Marine Science*, 12(1).

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

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., and Wattenbach, M. (2013). Climate extremes and the carbon cycle. *Nature*, 500(7462):287–295.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, 20(22):5473–5496.

Righetti, D., Vogt, M., Gruber, N., Psomas, A., and Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, 5(5):eaau6253.

Sinclair, B. J., Marshall, K. E., Sewell, M. A., Levesque, D. L., Willett, C. S., Slotsbo, S., Dong, Y., Harley, C. D. G., Marshall, D. J., Helmuth, B. S., and Huey, R. B. (2016). Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? *Ecology Letters*, 19(11):1372–1385. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.12686.

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuysen, J. A., Donat, M. G., Feng, M., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Gupta, A. S., Payne, B. L., and Moore, P. J. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change*, 9(4):306–312.

Sunday, J. M., Bates, A. E., and Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, 2(9):686–690.

Thomas, M. K., Aranguren-Gassis, M., Kremer, C. T., Gould, M. R., Anderson, K., Klausmeier, C. A., and Litchman, E. (2017). Temperature–nutrient interactions exacerbate sensitivity to warming in phytoplankton. *Global Change Biology*, 23(8):3269–3280.

Thomas, M. K., Kremer, C. T., Klausmeier, C. A., and Litchman, E. (2012). A Global Pattern of Thermal Adaptation in Marine Phytoplankton. *Science*, 338(6110):1085–1088.

Thomas, M. K., Kremer, C. T., and Litchman, E. (2016). Environment and evolutionary history determine the global biogeography of phytoplankton temperature traits: Phytoplankton temperature trait biogeography. *Global Ecology and Biogeography*, 25(1):75–86.

Ummenhofer, C. C. and Meehl, G. A. (2017). Extreme weather and climate events with ecological relevance: a review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1723).

van de Pol, M., Jenouvrier, S., Cornelissen, J. H. C., and Visser, M. E. (2017). Behavioural, ecological and evolutionary responses to extreme climatic events: challenges and directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1723).

Yentsch, C. S. (1960). The influence of phytoplankton pigments on the colour of sea water. *Deep Sea Research (1953)*, 7(1):1–9.

Appendix

Tables on next page...

	Performance Diff. [all events]				**Performance Diff. [intense events]**			
Predictors	*Estimates*	*CI*	*Statistic*	*p*	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	0.06	0.01 – 0.10	2.64	**0.008**	0.08	0.04 – 0.13	3.44	**0.001**
abslat_scaled	0.22	0.16 – 0.27	7.85	**<0.001**	0.21	0.14 – 0.27	5.94	**<0.001**
sst_scaled	0.21	0.16 – 0.25	8.58	**<0.001**	0.19	0.12 – 0.25	5.70	**<0.001**
season [spring]	0.02	0.01 – 0.03	4.21	**<0.001**	0.01	-0.01 – 0.02	0.91	0.365
season [summer]	-0.11	-0.12 – -0.10	-23.75	**<0.001**	-0.16	-0.18 – -0.15	-21.27	**<0.001**
season [winter]	0.06	0.05 – 0.07	12.85	**<0.001**	0.09	0.08 – 0.11	10.44	**<0.001**
abslat_scaled * sst_scaled	-0.01	-0.04 – 0.02	-0.81	0.421	-0.01	-0.04 – 0.02	-0.60	0.551
abslat_scaled * season [spring]	-0.01	-0.04 – 0.01	-1.10	0.272	-0.03	-0.07 – 0.01	-1.48	0.140
abslat_scaled * season [summer]	-0.17	-0.19 – -0.14	-14.84	**<0.001**	-0.16	-0.19 – -0.12	-8.61	**<0.001**
abslat_scaled * season [winter]	0.06	0.03 – 0.08	4.56	**<0.001**	0.09	0.05 – 0.13	4.39	**<0.001**
sst_scaled * season [spring]	-0.01	-0.03 – 0.01	-0.96	0.335	-0.03	-0.07 – 0.01	-1.64	0.101
sst_scaled * season [summer]	-0.20	-0.22 – -0.17	-17.46	**<0.001**	-0.24	-0.27 – -0.20	-12.95	**<0.001**
sst_scaled * season [winter]	0.07	0.05 – 0.09	5.76	**<0.001**	0.09	0.05 – 0.13	4.12	**<0.001**
Random Effects								
σ^2		0.01				0.01		
τ_{00}		0.02_{isolate}				0.02_{isolate}		
ICC		0.66				0.63		
N		75_{isolate}				69_{isolate}		
Observations		5299				2721		
Marginal R^2 / Conditional R^2		0.249 / 0.746				0.314 / 0.744		

Table 3A1: Coefficients of two linear mixed models of phytoplankton performance ratio during marine heatwave events.

	Winter Only				**Spring Only**			
Predictors	*Estimates*	*CI*	*Statistic*	*p*	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	0	-0.10 – 0.09	-0.02	0.986	-0.02	-0.14 – 0.10	-0.37	0.714
abslat	0.01	0.00 – 0.01	3.69	**<0.001**	0	0.00 – 0.01	2.26	**0.024**
sst_scaled	-0.05	-0.11 – 0.02	-1.45	0.147	-0.01	-0.10 – 0.08	-0.32	0.747
abslat * sst_scaled	0	0.00 – 0.00	3.69	**<0.001**	0	-0.00 – 0.00	1.58	0.115
	Summer Only				**Autumn Only**			
Predictors	*Estimates*	*CI*	*Statistic*	*p*	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	-0.18	-0.30 – -0.07	-3.12	**0.002**	-0.03	-0.13 – 0.08	-0.46	0.642
abslat	0	0.00 – 0.01	2.03	**0.042**	0	0.00 – 0.01	2.07	**0.039**
sst_scaled	0.11	0.02 – 0.19	2.55	**0.011**	-0.03	-0.11 – 0.05	-0.71	0.478
abslat * sst_scaled	0	-0.00 – -0.00	-2.3	**0.022**	0	-0.00 – 0.00	1.68	0.093
Random Effects	Winter	Spring	Summer	Autumn				
σ^2	0	0.01	0.01	0				
τ_{00}	0.02_{isolate}	0.02_{isolate}	0.03_{isolate}	0.02_{isolate}				
ICC	0.94	0.74	0.86	0.88				
N	75_{isolate}	75_{isolate}	75_{isolate}	75_{isolate}				
Observations	1141	1363	2003	792				
Marginal R^2 / Conditional R^2	0.192 / 0.955	0.080 / 0.756	0.085 / 0.870	0.104 / 0.890				

Table 3A2: Comparison of seasonal models of scaled performance difference during MHW events. In all models a significant relationship exists between season and absolute latitude (abslat).