TESTING THE ABILITY OF DISTRIBUTION MODELS TO MAP RARE PLANT COMMUNITIES IN WATERTON LAKES NATIONAL PARK

CARISA MEY MCGALE
Bachelor of Science, University of Lethbridge, 2022

A thesis submitted to the School of Graduate Studies
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

BIOLOGICAL SCIENCES

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

TESTING THE ABILITY OF DISTRIBUTION MODELS TO MAP RARE PLANT COMMUNITIES IN WATERTON LAKES NATIONAL PARK

CARISA MEY MCGALE

Date of Defence: December 4th, 2025

Dr. J. L. McCune Thesis Supervisor University of Lethbridge Lethbridge, Alberta	Associate Professor	Ph.D.
Dr. P. Achuff Thesis Examination Committee Member	Scientist Emeritus	Ph.D.
Dr. C. Coburn Thesis Examination Committee Member	Professor	Ph.D.
Dr. R. Laird Thesis Examination Committee Member	Professor	Ph.D.
Dr. D. Yevtushenko Chair, Thesis Examination Committee	Professor	Ph.D.

ACKNOWLEDGEMENTS

This research was conducted on Treaty 7 territory, in the lands of the Niitsítapi people of the Kainai, Piikani, and Siksika Nations. I acknowledge the sacred land that is Paahtómahksikimi, and these Nations as the traditional stewards since time immemorial. I am extremely privileged to have conducted my research here.

First and foremost, I am deeply grateful to Dr. Jenny McCune for the opportunity to be part of this project and for being an extraordinary mentor. From our first meeting in 2020 during my undergraduate degree, to working for her across two provinces before my Master's, I am proud to have grown in her lab over so many years. Working with Jenny has opened doors for me that I did not know existed, offering me opportunities and adventures I will cherish forever. Thank you, Jenny, for believing in my abilities, pushing me to be a better scientist, and for your endless patience over these past five years. I am incredibly fortunate have been on the receiving end of your red ink.

Thank you to my committee members, Dr. Peter Achuff, Dr. Craig Coburn, and Dr. Rob Laird, for your time, insight, and thoughtful feedback throughout this journey. A special thank you to Peter for answering my many emails and sharing his expertise on the Waterton dataset and his methodologies. I would also like to express my gratitude for the funding and support received from Parks Canada, the Alberta Conservation Association Grants in Biodiversity, the National Science and Engineering Research Council of Canada, and the University of Lethbridge.

To the members of the McCune lab – thank you for being my community through the highs and lows, and on both sides of the country. I have always felt welcomed and inspired by every mentor, friend, and roommate I have had within this lab. Thank you to

Emma Neigel for encouraging me from my undergraduate thesis through to my Master's, Amy Wiedenfeld for always being down for tea walks and grounding conversations, Kristin Olson for her constant kindness, Kirsty McFadyen for keeping me sane but still silly, Heather Davis for her unwavering optimism, Erik Vilu and Lauren Edison for the many long and enjoyable office chats, and Genoa Alger for all of her help in and outside the field. You have all made this experience richer in ways that go far beyond science.

This project would not have been possible without my superstar field team, who made our unforgettable summer both productive and full of laughter. Thank you to my field assistants, Dani Nadeau and Aiden Lair – your resiliency, patience, and curiosity made all the difference, as did your willingness to be dragged to unseemly places. Lastly, thank you to Nerds Gummy Clusters for carrying us through the toughest of field days.

I would not be where I am (or who I am) without my friends and family. The biggest thank you to my parents for your endless support and generosity, which have allowed me to focus on my studies and the confidence to chase every dream I have ever had. What you have given me is truly priceless. Thank you to my cohort, AWESB girl gang, and everyone I have met along this journey for the encouragement and sense of belonging that have made these years so special. The friendships and connections I have made here have shaped me in ways I'll carry long after this chapter is finished. A final thank you to my friends beyond these walls, for always cheering me on through every new adventure, for listening to me talk through stress or excitement and for sharing in every small victory along the way. Knowing you are always by my side means more to me than words can express. Thank you for reminding me that I am never walking alone.

CONTRIBUTIONS OF AUTHORS

Chapter 2 represents original and unsubmitted work by Carisa McGale (primary author) that is co-authored by Dr. Jenny McCune. Jenny McCune conceived the idea for the study and it was jointly designed by Jenny McCune and Carisa McGale. Carisa McGale compiled the geospatial data, created all species distribution models, conducted field work and plant identifications, analyzed the data and wrote the manuscript. Dr. Jenny McCune contributed to the interpretation of results and editing the manuscript.

ABSTRACT

Effective biodiversity conservation requires understanding the distributions of both individual species and the communities they form. However, few studies have assessed methods for modelling the distributions of ecological communities. I compared two modelling approaches for predicting seven rare plant communities in Waterton Lakes National Park, Alberta, Canada: community distribution models (CDMs), which treat communities as units, and stacked species distribution models (s-SDMs), which model the component species of each community and then assemble these predictions. Consistent with the individualistic view of ecological communities, s-SDMs consistently outperformed CDMs, especially for communities with few occurrence records. Model accuracy was largely unaffected by recent wildfire disturbance, suggesting that these models are just as useful in disturbed landscapes. My findings show that s-SDMs are an effective approach for mapping rare vegetation communities and can help managers locate, monitor, and protect these assemblages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AL	Alpine Ecoregion
AUC	Area Under the (Receiver Operating Characteristic) Curve
CDM	Community Distribution Model
DM	Distribution Model
FP	Foothills Parkland Ecoregion
GLM	Generalized Linear Model
MO	Montane Ecoregion
RF	Random Forest
SDM	Species Distribution Model
s-SDM	Stacked Species Distribution Model
SL	Lower Subalpine Ecoregion
SU	Upper Subalpine Ecoregion
WLNP	Waterton Lakes National Park

Plant Communities

C60	*Populus tremuloides/Amelanchier alnifolia/Heracleum maximum*
C61	*Populus tremuloides/Rubus parviflorus*
C62	*Populus tremuloides/Symphoricarpos occidentalis*
C78	*Picea engelmannii/Equisetum arvense*
C79	*Pinus contorta/Eurybia conspicua-Calamagrostis rubescens*
H25	*Danthonia* spp.*-Festuca campestris-Koeleria macrantha*
H27	*Pseudoroegneria spicata-Festuca campestris*

CHAPTER 1: INTRODUCTION

1.1. Background

Biodiversity encompasses all biological life on Earth (Cardinale et al. 2012; Hooper et al. 2012). We most often think of biodiversity as the number of species present in an area (species richness), but biodiversity consists of diversity at multiple scales, from the genetic diversity between individuals to the diversity of species and communities across the globe (Hortal et al. 2022). Species richness is one aspect of biodiversity that can be measured at different spatial scales. At a regional scale, gamma diversity quantifies the total number of species that are available to colonize local sites (Whittaker 1972). At a smaller and more local scale, alpha diversity measures the total number of species forming an ecological community at these sites (Whittaker 1972). While conservation has traditionally focused on protecting individual species from extinction, thereby maintaining the species richness of a region, there is also diversity among the different assemblages of species that make up a region. This variability in community composition among locations is the beta diversity, and it can provide insight into biodiversity that species richness alone cannot. For example, beta diversity can decline even if species richness remains stable if previously distinct assemblages of species become more similar, a process referred to as biotic homogenization (McKinney & Lockwood 1999). As globalization increases, there is mounting concern that humans are facilitating the homogenization of local communities through the establishment and colonization of exotic and disturbance-tolerant species (Solar et al. 2015; Finderup Nielsen et al. 2019). Although understanding the ecological requirements of individual

species is necessary for their protection, protecting individual species is insufficient to maintain all aspects of biodiversity (Hamilton et al. 2022). Studying communities and their compositions can help us to gain a more complete picture of biodiversity, and guide conservation strategies aimed at preserving this diversity across multiple scales (Wiersma et al. 2005).

1.2. Ecological Communities

Ecological communities are broadly defined as assemblages of species that occur together in space and time (Whittaker 1962). These communities are assembled from a regional species pool: the set of all the species present in a larger region, which is shaped by evolutionary processes such as speciation, extinction, and historical dispersal. Not all species from this pool can establish at each local site, but instead are thought to be "filtered" into these local communities by their tolerances to different ecological constraints (Keddy 1992; HilleRisLambers et al. 2012). Central to this theory is the niche, which is defined as the set of environmental conditions under which a species can survive and reproduce indefinitely (Hutchinson 1957). This range of abiotic conditions a species can tolerate, and therefore occupy, is its fundamental niche. Even if abiotic conditions are suitable, a species must also be suited to tolerate local biotic interactions (e.g. competition and predation), further restricting where a species can persist within its fundamental niche. This combination of abiotic and biotic tolerances constitutes the realized niche of a species, which defines where a species is actually able to persist (Hutchinson 1957). Finally, dispersal ability influences whether a species is able to get to a site where abiotic conditions and biotic interactions allow for a population to persist (Myers & Harms

2009). Therefore, communities are more than a random collection of species but a product of the ecological and evolutionary processes acting upon them.

The definition and classification of ecological communities, and whether their component species are interdependent or independent of one another, has been a long-standing debate amongst ecologists, particularly plant ecologists (Clements 1916; Gleason 1926). Clements (1916) viewed plant communities as discrete "superorganisms" with clear, sharp boundaries. Gleason (1926) believed them to be more fluid "individualistic" assemblages formed by a set of species whose niches just happen to overlap. Although assemblages of species can be grouped into recognizable types, community composition often varies gradually along environmental gradients (McIntosh 1995; Ricklefs 2008). In nature, there are few sharp boundaries where species composition changes completely from one location to another. Even at sharp ecotones, such as the edge of a forest and a meadow, some species are able to persist in both habitats, resulting in gradual change. As a result, the boundaries between two ecological communities are rarely definitive, and any classification we create is a simplification of complex patterns.

Despite this complexity, understanding the distribution of plant communities remains a central pursuit in ecology because of its theoretical and applied importance. Predicting the distribution of communities can provide insight into how biodiversity shifts across environmental space, resilience to disturbance, and responses to global change, while also informing conservation, land management and ecological restoration (Ferrier & Guisan 2006; Guisan & Thuiller 2005; HilleRisLambers et al. 2012). From a theoretical perspective, predicting plant community distributions can help ecologists test hypotheses about the processes that structure communities, such as niche differentiation

and species coexistence. From a practical perspective, predicting plant community distributions can highlight where at-risk communities are most likely to occur and help land managers focus conservation efforts where they are needed (Ferrier & Guisan 2006; Butler et al. 2022). Identifying the factors most influential in structuring communities enables ecologists to better anticipate how community composition may change in response to natural or human disturbances.

Ecologists have developed systematic approaches to describe, measure, and classify communities. Quantitative surveys of species composition and abundance provide the foundation for ordination and clustering methods that quantify how similar or different communities are from one another in terms of their species composition (Legendre & Legendre 2012). Vegetation classification systems, such as the United Kingdom's National Vegetation Classification or the Canadian National Vegetation Classification, are standardized methods to classify and compare communities across a landscape (Jennings et al. 2009; Baldwin et al. 2019). These methods enable ecologists to quantify the changes in community composition through time, relate community patterns to environmental drivers, and develop predictive models of the distributions of different community types.

1.3. Distribution Models

A central goal of ecology is to understand where species can persist across the landscape, which has led to the development of species distribution models (Nix et al. 1977; Nix 1986). Species distribution models (SDMs; also called ecological niche models) use statistical methods to model the relationship between species' occurrences

and the environmental conditions present at these locations (Elith & Leathwick 2009; Franklin 2010). The predictions of these models can then be projected across a region to generate maps of relative habitat suitability or probability of occurrence. SDMs are frequently used to inform conservation planning because of their ability to identify important environmental drivers of species distributions, identify locations of critical habitat, predict the potential spread of invasive species, and identify suitable locations for experimental translocations (Guisan et al. 2013; Guillera-Arroita et al. 2015; Draper et al. 2019). Efforts to test the myriad SDM methodologies have sharply increased in recent years, as researchers recognize the need to establish best practices for producing the most accurate models (Elith et al. 2006; Ferrier & Guisan 2006; Merow et al. 2013; D'Amen et al. 2017).

Despite their utility, SDMs rely on several assumptions that must be carefully considered. They presume that species-environment relationships are relatively stable through time (Guisan & Theurillat 2000), and that the species occurrence data used to build them are representative of the entire ecological niche (Hutchinson 1957; Franklin 1999). SDM outputs can be influenced by sampling bias in the occurrence records and incomplete detection of species, leading to biased predictions of species distributions (Phillips et al. 2008; Newbold et al. 2010). Furthermore, because most SDMs focus on abiotic predictors, they often neglect to account for biotic interactions and dispersal limitations, which can limit the realism of predictions (Dormann et al. 2012). Despite these limitations, extensive research has resulted in the development of best practices for building SDMs (Merow et al. 2013; Araújo et al. 2019; Lee-Yaw et al. 2022) and SDMs have proven highly effective at predicting species occurrences (Anderson 2013; McCune 2016; Radomski et al. 2022; Lee-Yaw et al. 2022)

While distribution modelling has focused mainly on the species-level, ecologists have attempted to apply similar models to predict the geographic ranges of entire communities (Ferrier & Guisan 2006). These models generally take one of two approaches. The first approach is to "assemble first and predict later" (Ferrier & Guisan 2006), modelling a community as a single unit that responds to environmental constraints (Jiménez-Alfaro et al. 2018). This approach aligns with Clements' (1916) "superorganism" view of ecological communities. Alternatively, they can take a "predict first and assemble later" approach (Ferrier & Guisan 2006), in which communities are "deconstructed" into their individual species for modelling and then re-assembled by stacking the predictions for each species (Henderson et al. 2014). This approach aligns more with Gleason's (1926) "individualistic" view of communities.

Community-level distribution models have been used with some success to predict broad vegetation classes (e.g. pine dominated forest, heath woodland) at larger spatial extents (e.g country; Bradter et al. 2011; Jiménez-Alfaro et al. 2018; Simensen et al. 2020; Butler et al. 2022). However, most community level models have focused on predicting species richness rather than composition and are typically built at relatively coarse spatial scales (e.g. 1km x 1km or 10km x 10km grids). The resulting richness maps can highlight broad regional gradients of richness but generalize fine scale processes within each cell, potentially obscuring local scale patterns of species assemblages (Deschamps et al. 2023). At smaller, local scales, accurately predicting community composition might depend on which approach is taken: assemble first and predict later, or predict individual species and then assemble. Both approaches have strengths and weaknesses. Models that predict the distributions of already classified communities capture both the abiotic and biotic factors that influence the distributions of species and

shape community composition but are potentially limited by the availability of occurrence data (Chapman & Purse 2011). In contrast, models that predict each individual species distribution have the advantage of being able to leverage the more abundant individual species data available in public databases but are unable to account for the biotic interactions that influence species persistence and abundance. However, very few studies have compared these two approaches in predicting the same community, and even fewer test the utility of these two approaches using independent survey data (Ferrier & Guisan 2006; D'Amen 2017; Murphy & Smith 2021). As a result, distribution modelling at the community-level has lagged behind species-level SDMs, leaving uncertainties about best practices for building them (Ferrier & Guisan 2006; Barbet-Massin et al. 2018). Resolving these gaps is necessary for community-level models to be applied effectively for advancing ecological theory, and for conservation management and restoration purposes (Guisan et al. 2013; Murphy & Smith et al. 2021).

1.4. Research Objectives

This thesis tested approaches for building community-level distribution models to predict the geographic distributions of seven rare vegetation communities in Waterton Lakes National Park, Alberta, Canada. In this study, we compared the performance of community distribution models (CDMs) and stacked species distribution models (s-SDMs) in their ability to generate accurate models and spatial predictions based on independent survey data. The objectives were:

1. To construct CDMs and s-SDMs, using three separate algorithms, for each rare vegetation community using a set of environmental, topographic, and edaphic predictor variables.
2. To construct the s-SDMs using varying numbers of occurrence records, different sets of species, and different methods of averaging the final stacked predictions.
3. To generate maps that illustrate the predicted distributions of each rare vegetation community.
4. To evaluate model efficacy and utility using independent field survey data and determine which approach performs the best.

CHAPTER 2: TESTING THE ABILITY OF DISTRIBUTION MODELS TO MAP RARE PLANT COMMUNITIES IN WATERTON LAKES NATIONAL PARK

2.1. Abstract

Predicting the distributions of plant communities remains a major challenge in ecology, and it is unclear whether communities are best modelled as whole entities or as the sum of their component species. Community distribution models (CDMs) could better capture plant community patterns by directly modelling plant communities as a whole, but stacked species distribution models (s-SDMs) might outperform them, especially because species-level data is more abundant. Wildfire could reduce the predictive accuracy of distribution models by altering plant community composition and shifting communities to earlier successional stages. We modelled the distribution of seven plant communities of conservation concern in Waterton Lakes National Park, Alberta, Canada. We built CDMs and s-SDMs that varied in algorithm, number of species, number of occurrences included, and the method of stacking. We then tested our models with independent surveys in both recently burned and unburned areas. S-SDMs consistently outperformed CDMs, showing a stronger correlation between model prediction and community similarity to the target community. Interestingly, wildfire did not reduce the utility of distribution models, and they actually performed better in burned areas for some community types. Stacked species distribution models are a practical approach to map the distributions of rare plant communities and help to inform conservation efforts by pinpointing priority areas for targeted field surveys, protection, and management.

2.2. Introduction

Ecologists invest considerable effort in modelling the distributions of plant species, but understanding the distributions of plant communities is equally important. Gaining insight into where particular plant communities occur and the environmental factors that influence their distributions advances ecological theory and supports efforts to manage and protect them. However, generating detailed maps through extensive phytosociological field surveys is time and labour intensive. Vegetation mapping based on classification of aerial photos or satellite imagery is comparatively quick and applicable to large geographic areas but cannot differentiate distinct plant communities within coarser vegetation types (i.e. 'deciduous forest', 'shrubland'). This is because the compositional differences among communities are subtle relative to the spatial and spectral resolutions of these technologies and are frequently obscured by overstory vegetation (Ferrier et al. 2002). Correlative distribution models (DMs) that incorporate both approaches could provide a practical alternative by relating spatial data on environmental and topographic variables to plant community occurrence data from field surveys in order to predict the geographic distributions of individual plant communities (Guisan & Zimmerman 2000; Ferrier et al. 2002; Guisan & Thuiller 2005; Elith et al. 2006). The application of distribution models to individual species (SDMs) is very common, and these models are quite successful at predicting species distributions (e.g. Lee-Yaw et al. 2022); however, fewer studies have used DMs to model the distributions of ecological communities. DMs can be applied to the community level in two ways: by modelling the probability of occurrence of individual component species and stacking the outputs to predict the occurrence of the community (s-SDMs; Ferrier & Guisan 2006) or

by modelling communities as distinct units, where the presence or absence of the community as a whole is used as the input data (CDMs; Butler & Sanderson 2022). It remains unclear which of these methods is the most effective.

The nature of a community is a classic debate in ecology. Clements (1916) viewed communities as entities that respond to environmental conditions as interdependent units, while Gleason (1926) viewed communities as collections of species responding independently to their environments. If Gleason's individualistic view is more accurate, then modelling and stacking individual species distribution models (s-SDMs) might be the most effective method to predict the geographic distribution of a community (Ferrier & Guisan 2006; D'Amen et al. 2017). This approach hinges on niche theory, which posits that each species has a set of environmental conditions under which it can persist (fundamental niche), and its presence reflects both environmental suitability and biotic interactions (realized niche; Gause 1934; Hutchinson 1957; Franklin 1995; Bruno et al. 2003). Therefore, if it is possible to predict where species are, the overlap of their distributions predicts the community composition at any given place. In this sense an ecological community has only a realized niche, dictated by the overlapping realized niches of its characteristic component species (Zimmerman & Kienast 1999). S-SDMs have been used to predict species richness and community composition across a landscape (e.g. Pottier et al. 2013; Calabrese et al. 2014; Zurell et al. 2020). They may also be the most accurate way to forecast future community distributions with climate change, as species are expected to vary in the magnitude and pace of range shifts (Zimmerman & Kienast 1999; Henderson et al 2014). However, communities are defined not only by the presence and absence of particular species, but by patterns of abundance. S-SDMs might overpredict suitable areas because they rely solely on the predicted

presence or absence of individual species (Dubuis et al. 2011). This treats all occurrences equally, including those where species are present in low abundance. By ignoring species abundance patterns, s-SDMs risk predicting locations where a community's component species may co-occur, but not necessarily where the target community exists.

Alternatively, if species within a community are responding to abiotic conditions interdependently, they may be more effectively modelled as units. Zimmerman & Kienast (1999) proposed that using the presence and absence of a community as a whole can capture the biotic interactions that promote or limit specific species associations, rather than simply mapping where abiotic conditions are suitable for all of a community's species. They successfully predicted the occurrence of 34 graminoid communities using climatic, topographic, and edaphic variables. More recently, Bradter et al. (2011) and Butler & Sanderson (2022) modelled the distributions of plant communities with CDMs using the National Vegetation Classification (NVC) of Great Britain. By using records of the entire community assemblage, CDMs implicitly account for both abiotic and biotic factors that shape community composition (Norberg et al. 2019). However, CDMs might be limited by the availability of occurrence data, particularly for rare vegetation communities or in regions lacking detailed field surveys or regions that are largely inaccessible (Radomski et al. 2022). Very few studies have directly compared the ability of CDMs and s-SDMs to estimate plant community distributions. While building community-level DMs is not a new concept (e.g. Kessel 1976), recent reviews emphasize the need for further evaluation of this approach, including direct comparisons of s-SDMs and CDMs at scales relevant to conservation (D'Amen 2017; Murphy & Smith 2021).

It is important to note that the disturbance history of sites within a region could affect the ability of DMs to accurately predict plant community distributions. Community

composition is not only shaped by climate, topography, and soil but also by disturbance history (Johnstone et al 2010; Limb et al. 2018). When disturbance history is excluded as a predictor in a distribution model, predictions may be inaccurate if the community has not yet reached a stable equilibrium state (Ferrier et al. 2002; Pottier et al. 2013). Both anthropogenic disturbances, such as clearcutting, and natural disturbances, such as wildfire, restart the natural process of succession. In areas affected by large disturbances, the observed community might differ from the predicted one because the site is still in early succession and has not reached its potential, late-successional composition. For example, the return of forest communities to pre-fire composition can take decades and the time required varies depending on the environmental conditions (Johnstone et al. 2010; Dawe et al. 2022). Therefore, disturbance history must be considered when assessing DM predictions.

In this study, we compared the accuracy of s-SDMs and CDMs for modelling the distributions of seven rare plant communities in Waterton Lakes National Park (WLNP), Alberta. Parks Canada has conducted vegetation surveys within WLNP that classified the plant communities into discrete vegetation types based on the percent cover of observed species. These surveys provide both community and species-level occurrence data with which to model community distributions. Managers with Parks Canada would like to know the extent and distribution of WLNPs' rare plant assemblages. In addition, the Park experienced a severe wildfire in 2017, providing an opportunity to test model performance in both burned (early successional) and unburned (late successional) areas.

2.3. Methods

2.3.1. Study Area and Plant Communities

Paahtohmaksikimi, or Waterton Lakes National Park (WLNP; 49.0833°N, 113.9167°W), is located in southwestern Alberta in the Canadian Rocky Mountains (Figure 1). It is part of Waterton-Glacier International Peace Park, the world's first transboundary park and a UNESCO World Heritage Site (Parks Canada 2023). Although WLNP is the smallest of Canada's Rocky Mountain national parks (525 km^2), it is home to more than 1,000 vascular plant species (Parks Canada, 2022). There are four distinct ecoregions in the park: the foothills parkland (FP), montane (MO), subalpine (SL = lower subalpine, SU = upper subalpine), and alpine (AL) (Achuff et al. 2002). These ecoregions vary greatly in their climatic conditions, corresponding to a diverse range of plant communities. For example, the foothills parkland ecoregion comprises the lowest altitudes of the park (1280 m – 1500 m above sea level), experiences the warmest and driest climatic conditions (mean annual temperature = 3 °C; growing season precipitation = 377 mm), and contains eleven distinct vegetation communities that range from closed canopy aspen forests to fescue grasslands (Achuff et al. 2002; Natural Regions Committee 2006). In contrast, the alpine ecoregion comprises the park's highest altitudes (2250 m – 2940 m), has cooler and wetter climatic conditions (mean annual temperature = -2.4 °C; growing season precipitation = 472 mm), and contains only four herb-dwarf shrub vegetation communities dominated by mountain avens and lichens (Achuff et al. 2002; Natural Regions Committee 2006).

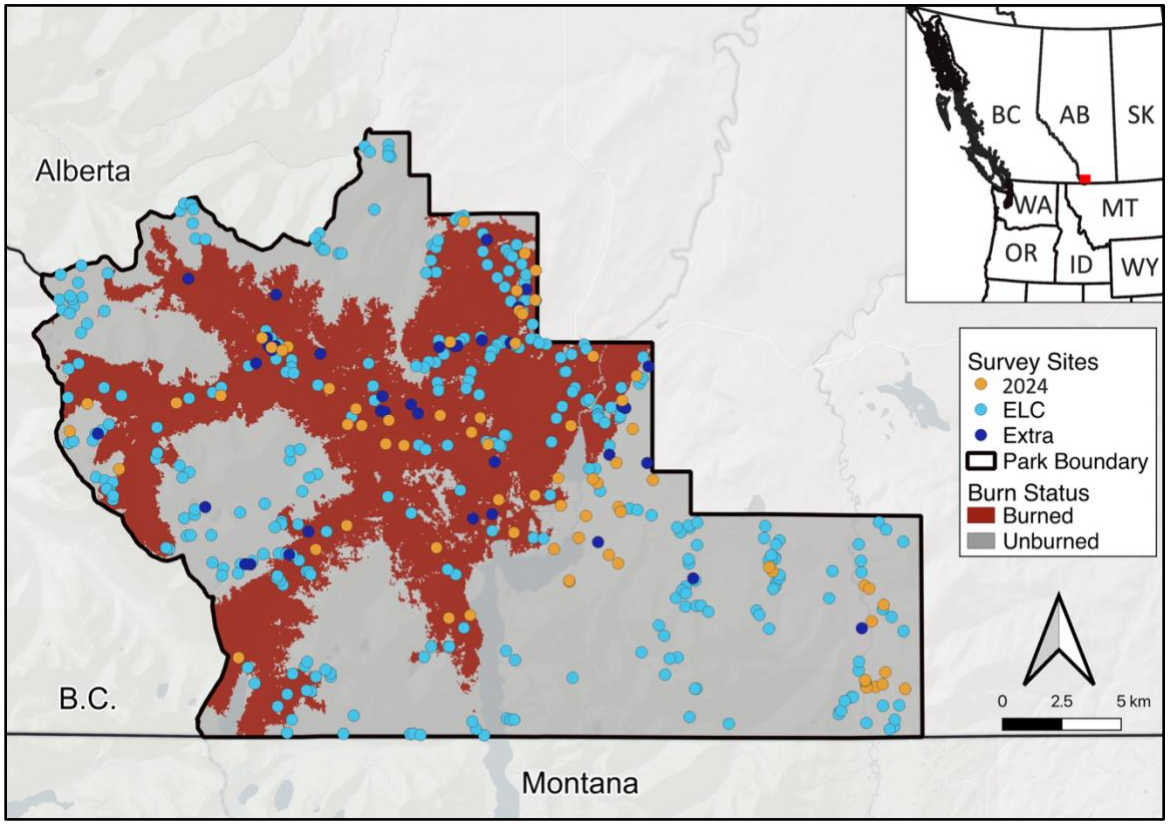

Figure 1. Map of Waterton Lakes National Park, Alberta, Canada showing the occurrence records used to build the CDMs and s-SDMs in this study (n=369: 330 from the Ecological Land Classification plots (ELC; light blue dots) and 39 extra locations provided by Dr. Peter Achuff (Extra; dark blue dots)), the area burned by the 2017 Kenow Wildfire (red), and the 72 independent survey locations from 2024 (orange dots). Inset map shows the location of Waterton (red) in the context of northwestern North America.

From 1994 to 1999, Parks Canada carried out the Ecological Land Classification (ELC) project to create a comprehensive resource documenting and classifying the park's landforms, soils, and vegetation communities (Achuff et al. 2002). Parks Canada established 330 vegetation plots during this project to classify the unique vegetation associations present in WLNP and to collect soils and hydrological information. These relevé plots were placed in homogenous areas representing the predominant vegetation type within map polygons delineated by landform and vegetation characteristics visible on aerial photographs, and they ranged in size from 10 m x 10 m to 20 m x 20 m (Achuff

et al. 2002). Researchers identified every plant species and estimated their percent cover in each plot. They used these data to classify all plots into 45 distinct vegetation communities based primarily on the dominant species of each cover layer (e.g. tree, shrub, herb). Sixteen of these communities are considered special due to their rarity or fragility (Achuff et al. 1997). The ELC was successful in mapping the distribution of broader ecological landscape units across the park based on landform, soil, and vegetation differences (Achuff et al. 1997). However, researchers were unable to map individual vegetation communities, including the 16 designated as special, because distinct communities that occur within the same landscape unit cannot be distinguished from one another based on aerial photographs.

We selected seven of WLNP's 16 special plant communities that were recorded in at least five locations throughout the park, either in ELC plots or at additional locations documented by Dr. Peter Achuff during his tenure as Parks Canada botanist. Previous research shows that five occurrence records can be sufficient to build distribution models (Hernandez et al. 2006; Pearson et al. 2007). The focal plant communities are either closed canopy forest communities or grassland communities (Table 1).

Table 1. Characteristics of the seven special plant communities modelled in this study. NA indicates that a community did not have a dominant species within the specified vegetation layer.

Plant Community Code	Ecoregion	Dominant Tree	Dominant Shrub	Dominant Herb or Dwarf-shrub	Reason for Designation	ELC Plots	Total Presences	Characteristic Species (all)*	Characteristic Species >=5 occurrences^
C60	Foothills Parkland	*Populus tremuloides*	*Amelanchier alnifolia*	*Heracleum maximum*	Non-native plant invasion.	11	20	44	25
C61	Foothills Parkland & Montane	*Populus tremuloides*	*Rubus parviflorus*	NA	Non-native plant invasion.	4	7	22	19
C62	Foothills Parkland & Montane	*Populus tremuloides*	*Symphoricarpos occidentalis*	NA	Non-native plant invasion.	5	5	17	13
C78	Montane & lower Subalpine	*Picea engelmannii*	NA	*Equisetum arvense*	Rarity.	5	6	62	34
C79	Montane & lower Subalpine	*Pinus contorta*	NA	*Eurybia conspicua* & *Calamagrostis rubescens*	Rarity.	5	5	27	21
H25	Foothills Parkland & Montane	NA	NA	*Danthonia spp.* & *Festuca scabrella* & *Koelaria macrantha*	Non-native plant invasion.	12	20	84	57
H27	Foothills Parkland & Montane & Lower Subalpine	NA	NA	*Agropyron spicatum* & *Festuca scabrella*	Non-native plant invasion.	11	11	55	45

* The total number of species mentioned in the ELC plus all statistically significant indicator species.

^ The total number of characteristic species with at least five occurrence records available.

2.3.2. Building Distribution Models

We developed and compared CDMs and s-SDMs to evaluate their ability to predict plant community distributions. To capture potential variation in model predictions due to modelling choices, we built models using different algorithms, species subsets, and stacking methods (Figure 2). We considered also testing joint species distribution models (JSDMs), which model multiple species simultaneously and account for correlations in their occurrences (Hogg, Wang, & Stone 2021). However, JSDMs are computationally demanding (Tobler et al. 2019; Tikhonov et al. 2020) and have not consistently outperformed s-SDMs (e.g. Zurell et al. 2019). Therefore, we focused on comparing CDMs and s-SDMs as contrasting approaches to modelling the distributions of whole plant communities.

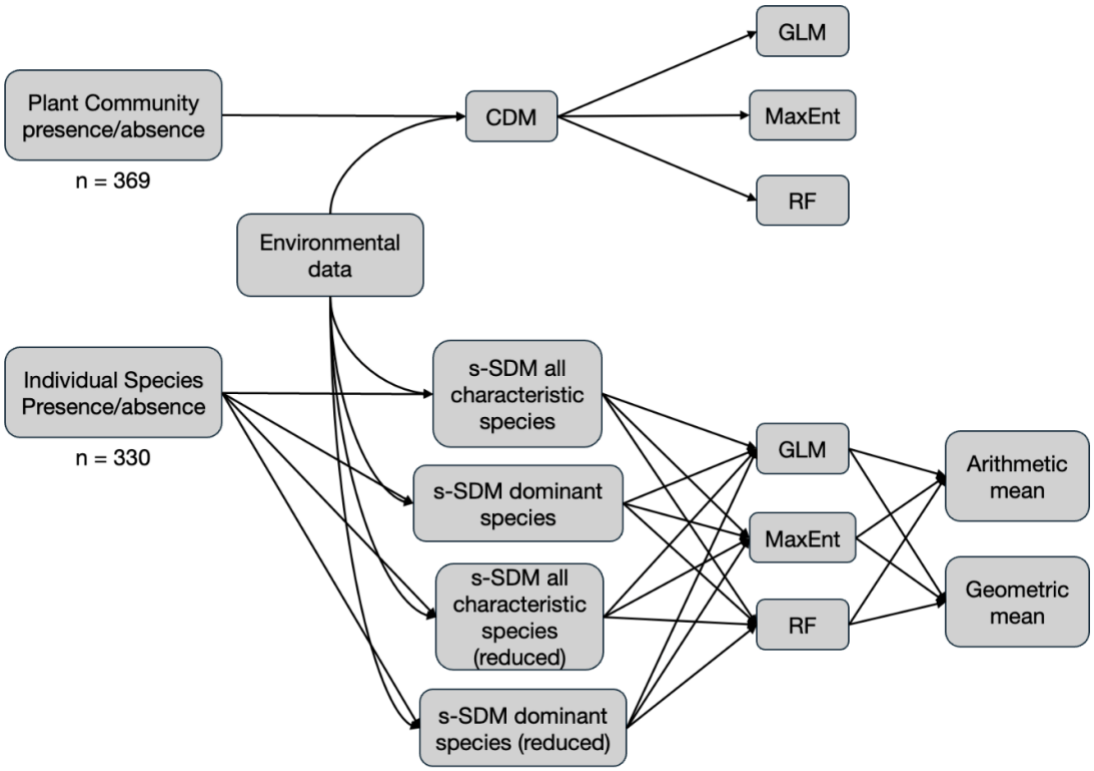

Figure 2. Flow chart outlining the modelling frameworks used to create community distribution models (CDMs) and stacked species distribution models (s-SDMs). We used community or individual species presences/absences in combination with environmental predictors to build each distribution model using three algorithms: generalized linear models (GLMs), MaxEnt, and Random Forests (RF). We built S-SDMs using either all characteristic species or only the dominants, and applied either the arithmetic or the geometric mean to create the final community probability predictions, and used either all available occurrence records or a randomly selected subset equal to the number of occurrences available for the corresponding CDM. Characteristic species refer to those that were a dominant species, listed in the ELC, or were significant indicators of each community as identified by indicator species analysis. Dominant species were species that occurred in the greatest percent cover in these communities and were required by the ELC classification scheme for a plot to be classified as a specific community.

2.3.2.1. Identifying Characteristic Species of Communities

For the CDMs, we designated the ELC plots classified as the target plant community and any supplemental records of each community as presences and all other

ELC plots as absences. For the s-SDMs, it was necessary to decide which species would be stacked. The ELC defined each vegetation community based mainly on the dominant (most abundant) species. We gathered the presence and absence records of each target community's dominant species along with other species mentioned in each community's ELC description (Achuff et al. 2002). Species found in lesser abundances can also be associated with specific communities. These associated species might help to refine the predicted distributions of communities dominated by otherwise common species. Therefore, we also identified indicator species for each target community using Indicator Species Analysis (ISA) in R using the package *indicspecies* (R Core Team Version 4.3.1.; Dufrêne & Legendre 1997). ISA is a technique that identifies species that are significantly more frequent and/or abundant within a defined group compared to other groups (De Cáceres & Legendre, 2009). This yielded 311 indicator species across all seven vegetation communities, with as few as 17 and as many as 84 species for each community. We retained a total set of 146 "characteristic species" – each with >= 5 occurrences – that included dominant species, other species mentioned in community descriptions, and indicator species of each community to build the s-SDMs (Table S1). All occurrence records had a coordinate uncertainty of ~11 m. We built two versions of the stacked models for each vegetation community: one including only the dominant species of each target community, and a second incorporating all characteristic species associated with each target community.

2.3.2.2. Modelling Algorithms

DMs can be built using many different algorithms (Elith et al. 2006; Elith & Leathwick 2009; Aguirre-Gutiérrez et al. 2013; Radomski et al. 2022; Finn et al. 2023). We used three of the most common, and usually among the best performing, algorithms: Generalized Linear Models (GLM), Random Forests (RF), and Maxent (Elith et al. 2006; Cutler et al. 2007; Wisz et al. 2008; Raes & Aguirre-Gutiérrez 2018; Williams et al. 2009; Kaky et al. 2020; Valavi et al. 2022). GLMs have been the predominant algorithm used to develop SDMs with true presence and absence data because of their ability to analyse the typically non-linear relationship between species and their habitats (Austin 1987; Elith & Leathwick 2009). RF is a common machine learning algorithm that uses true presence and absence data and has the ability to handle large numbers of predictor variables without being affected by multicollinearity (Chiaverini et al. 2023). Maxent has been widely used to develop SDMs because of its ability to handle few occurrence records and presence-only data (Hernandez et al. 2006; Sheth et al. 2014). Running Maxent requires decisions about multiple settings, including the 'feature classes' and 'regularization' (Merow et al. 2013), but tuning procedures can be used to choose the best performing combinations based on cross-validation (Phillips & Dudík 2008). For Maxent models, we used a set of randomly selected background points rather than true absence data and tuned all models. We used the same set of environmental predictors for all algorithms, with RF and GLM models using all available predictor layers and Maxent using a subset determined during tuning procedures. We built all models using R (R Core Team Version 4.3.1.) in the *flexsdm* package (Velazco et al. 2022).

2.3.2.3. Environmental Variables

We created all distribution models using a set of 13 biologically relevant environmental predictors known to influence plant species distributions, encompassing climatic, edaphic, hydrologic, and topographic information across WLNP (Table 2; Dubuis et al. 2013; Longcore et al. 2018; Zangiabadi et al. 2021; Butler & Sanderson 2022). We retained these 13 predictors after removing highly correlated, continuous variables from a larger set of predictor variables (i.e. $|r| > 0.70$; Table S2). For highly correlated variable pairs, we retained the variable correlated with the fewest other predictors. Climatic variables are highly correlated with elevation in mountainous regions such as WLNP (Beniston 2006); therefore, we retained only three climatic variables of 11 that we considered for inclusion and did not include elevation as a predictor (Table S2). We resampled all 13 environmental variables to a 50 m spatial resolution and projected them to a North America Albers Equal Area Conic projection in QGIS v.3.36.2 (Table 2). We chose a 50 m cell size to approximate the spatial resolution at which these communities occur, while also providing flexibility during field surveys to adjust plot locations if obstacles were present.

Table 2. Environmental predictor variables used in the model building process for all variations of community and stacked distribution models, for all seven vegetation communities.

Environmental Predictor	Source	Categorical or Continuous	Number of Categories	Original Format	Original Resolution	Value Range & Units	Reference
Aspect	ASTER Global Digital Elevation Model NetCDF V003 (2000–2013)	Continuous	NA	Raster	30 m	0.11 – 360 (°)	NASA et al., 2019
Slope	ASTER Global Digital Elevation Model NetCDF V003 (2000–2013)	Continuous	NA	Raster	30 m	0 – 68.25 (%)	NASA et al., 2019
Sand	International Soil Reference and Information Center	Continuous	NA	Raster	250 m	31 – 57 (%)	Hengl et al., 2017
Silt	International Soil Reference and Information Center	Continuous	NA	Raster	250 m	28 – 53 (%)	Hengl et al., 2017
pH	International Soil Reference and Information Center	Continuous	NA	Raster	250 m	47 – 69	Hengl et al., 2017
Clay	International Soil Reference and Information Center	Continuous	NA	Raster	250 m	7 – 28 (%)	Hengl et al., 2017
Topographic Wetness Index	ASTER Global Digital Elevation Model NetCDF V003 (2000–2013)	Continuous	NA	Raster	30 m	3 – 30	NASA et al., 2019
Normalized Difference Vegetation Index	USGS Landsat 8 Image	Continuous	NA	Raster	30 m	-0.10 – 0.92	EROS, 2020
Net Primary Productivity	USGS Annual MODIS NPP (MOD17A3HGF Version 6.1)	Continuous	NA	Raster	500 m	-981 – 32767	Running & Zhao, 2021
End of Frost-Free Period	Climate NA	Continuous	NA	Raster	800 m	244 – 256 (day of the year)	Wang et al., 2016
Mean Annual Precipitation	Climate NA	Continuous	NA	Raster	800 m	764 – 2426 (mm)	Wang et al., 2016
Mean Annual Radiation	Climate NA	Continuous	NA	Raster	800 m	12 – 19 (MJ $m^{-2}d^{-1}$)	Wang et al., 2016
Waterton NVC Level 5 Map	Parks Canada	Categorical	11	Polygon	NA	NA	Hop et al. 2007

2.3.2.4. Stacked Modelling Choices

A key step in building stacked SDMs is selecting the method used to stack individual models to produce the final cumulative community predictions. Some studies threshold each component SDM to presence/absence before stacking, however this has been shown to overestimate species richness (Dubuis et al. 2011). Others simply sum individual probability predictions (Calabrese 2014). We stacked SDMs in two ways. First, we calculated the arithmetic mean of the predictions of the SDMs for all characteristic species to estimate the likelihood of each community occurring at a given location. However, this approach could yield a high predicted probability of occurrence even if one or more species has a 0% predicted probability of occurrence. To address this, we also created a set of s-SDMs that used the geometric mean, which yields a 0% prediction if any individual component species is predicted to have 0% probability of occurrence, thus excluding locations where any component species are missing and potentially improving the accuracy of community level predictions. We recognized that this approach may be overly conservative, especially when applied to models incorporating all characteristic species, as it requires non-zero predictions for every stacked species.

The number of occurrence records used to build a DM can influences its accuracy (Wisz et al. 2008; Pearson et al. 2007; Wang & Jackson 2023). Vegetation communities were sometimes represented by as few as five occurrences across the park. In contrast, some species had up to 118 presence records. To account for this, we built each s-SDM two ways: (1) using all available species occurrence records ('full model') or (2) using a subset of the total available occurrences for each species corresponding to the total

number of community occurrences available for the CDM ('reduced model'). For each plant community, we selected a random subset of species presence records to align with the CDM's total occurrence count (Table 1). If a species had fewer records than the community threshold, we used all available records.

We generated a total of 27 models (3 CDMs and 24 s-SDMs) for each rare community that represented all possible combinations of modelling choices, for a total of 189 models (Figure 1). We assessed model performance and discrimination using internal cross validation methods to generate threshold-independent and -dependent metrics: the area under the receiver operating characteristic curve (AUC) and the true skills statistic (TSS), respectively. The most used metric of model performance is the AUC, which evaluates the ability of a model to accurately predict presences and absences, where a score of 1 indicates perfect performance and 0.5 or less indicates the model is no better than random chance (Fielding & Bell 1997). We retained all individual species models used in a community's s-SDM regardless of cross-validated accuracy to ensure a standardized set of species used in each model (see Tables S3 through S17 in Appendix 1).

2.3.3. Comparing Performance of Distribution Models

2.3.3.1. Field Surveys

Although many studies rely on the AUC scores derived from internal cross-validation only, DM accuracy based on internal validation is often inflated compared to performance when tested with independent data (Elith & Burgman 2002; Newbold et al. 2010; McCune 2016). We conducted independent field surveys across WLNP to evaluate

our DMs. For each vegetation community, we developed preliminary CDMs and s-SDMs using GLMs that provided an indication of areas with high and low predicted suitability. Using these DMs, we generated a random stratified sample of points to guide site selection and to maximize the range of DM predictions across them. We limited site selection to areas within 1km of a road or trail due to logistical constraints. We stratified grid cells into five categories of predicted probability of occurrence (0-20%, 20-40%, etc.). We used the *sf* package in R to select as many points as possible stratified by probability category across the park with a minimum distance of 50 m from each other (Pebesma & Bivand 2023), resulting in 767 possible survey sites. We chose sites to survey from these 767 sites with the goal of ensuring replication within each category of predicted probabilities, equal representation of burned and unburned sites, as well as accessibility. Although we were unable to design a random stratified survey based on all 189 models, we created a site selection strategy that helped to capture a wide range of DM predictions across all models. We successfully surveyed 72 of these sites across the park, comprising 33 unburned sites and 39 burned sites (Figure 1).

With our field assistants, we navigated to the center of each 50 m x 50 m cell with a Garmin eTrex 20 handheld GPS. We delineated a 20 m x 20 m plot within each cell to remain consistent with the original field sampling protocols used in the ELC (Achuff et al. 2002). We assume that a 20 m x 20 m provides sufficient area to characterize the vegetation community composition within a 50 m x 50 m grid cell (Mueller-Dombois & Ellenberg 1974), while allowing flexibility to shift the plot within the 50 m x 50 m cell if obstacles were present without surveying outside the target area. However, we recognize that some portion of any mismatch between the observed and predicted assemblages

could be due to characteristic species occurring just outside the sampled area (Pottier et al. 2013). At each plot, we recorded the slope aspect and angle using a compass with a built-in clinometer, and elevation using the Garmin GPS unit. We visually estimated percent cover of each vascular plant species across four vegetation layers: (1) herbs, low shrubs, and tree seedlings (<0.5 m in height), (2) tall shrubs and small trees (0.5-2 m), (3) intermediate trees (2-5 m), and (4) large trees (>5 m). We used these vegetation layers to align with the ELC protocol, as they are required to classify each surveyed site into a distinct community using the ELC vegetation key. We then summed across layers to obtain a single cover value per species for analysis.

We lumped some difficult-to-identify species at the genus level, and standardized plant species across all surveys using the Taxonomic Name Resolution Service (*TNRS*) package in R (Table S18).

2.3.3.2. Assessing Model Performance

When assessing the accuracy of SDM predictions, there are two outcomes: a species is either present or absent at a survey site. In contrast, the community present at a site can be measured along a gradient of compositional similarity to a target community. For a DM to be considered useful in predicting community presence and informing targeted surveys, its predictions should be strongly and positively correlated with the compositional similarity of the target community. We measured the average Bray-Curtis similarity between each surveyed plot and the known occurrences of each target community. A Bray-Curtis similarity of 1 indicates two communities share exactly the same species in the same relative abundances, while a value of 0 indicates that the two

communities share no species in common (Equation 1: Bray & Curtis 1957). This similarity index places greater weight on species that are more abundant at a site, and less weight to rare species within the community (Bakker et al. 2023). To reduce the relative influence of the most abundant species, we performed a 4^{th}-root transformation on the ELC and the 2024 survey data prior to calculating the Bray-Curtis similarity. By transforming the data, we were able to account for the presence of less abundant, characteristic species when assessing the similarity of surveyed sites to known target community occurrences. However, we note that the results did not change substantially when using raw values to calculate similarity (data not shown).

$$BC_{ij} = 1 - \frac{\sum_{k=1}^{n} |a_{i,k} - a_{j,k}|}{\sum_{k=1}^{n} (a_{i,k} + a_{j,k})}$$

Equation 1. Equation for the Bray-Curtis Similarity index. The upper portion of the equation calculates the absolute difference in species abundances (*a*) of species between two sites (*i* and *j*) for each species (*k*) and the lower part of the equation calculates the total abundances of all species in both samples, summed across all species (*n*).

Our primary goal was to evaluate which modelling approach, CDMs or s-SDMs, better predicted plant community composition. To do this, we modelled the relationship between each DM's standardized predicted probability of occurrence and the mean Bray-Curtis similarity using generalized linear models (beta family) fit with *glmmTMB* in R (R Core Team Version 4.3.1.; Brooks et al. 2017). We standardized each DM's predictions using the z-score standardization, which scales the mean value to 0 and the standard deviation to 1, to make predictions comparable across DMs. This scaling was necessary because s-SDMs did not always produce predictions spanning the full 0 – 1 range,

frequently producing maximum values less than 1 due to the averaging of many individual SDMs. If the DM is a useful predictor, then we expect the community similarity to increase as the DM prediction increases. To identify the best-performing DM of each type (CDM, s-SDM, or s-SDM with reduced occurrences), we compared the GLMs relating DM prediction to community similarity using Akaike's Information Criterion (AIC). AIC balances goodness of fit against model complexity, with lower AIC scores indicating better model performance and AIC scores with a difference of 2 or greater indicating significantly better models (Akaike 1973). To test whether recent fire disturbance altered the relationship between DM predictions and community similarity, we compared two GLMs for each DM: 1) a model with DM prediction, burn status (burned/unburned), and their interaction as predictors, and 2) a model without the interaction. If fire reduces the predictive accuracy of models, leading to a steeper relationship between DM prediction and community similarity in unburned compared to burned plots, then we expected a significant interaction between burned status and DM predictions. Alternatively, if fire lowers the overall community similarity but we still see a positive relationship between similarity and DM predictions in both burned and unburned areas, we expected significant main effects from the DM predictions but no interaction.

We then compared the explanatory power of the best CDM, s-SDM and reduced s-SDM for each plant community. First, we selected the single best version of each model type based on AIC values of the GLMs relating the average Bray-Curtis similarity to DM predictions. We calculated Ferrari's pseudo-R^2 values for the three best performing DMs

of each vegetation community to estimate the amount of variance in community similarity explained by the DM prediction.

Finally, we used an Analysis of Variance (ANOVA) to test whether the coefficient of the relationship between Bray-Curtis similarity and predicted probability differed among DM types (i.e. CDM, s-SDM, and reduced s-SDM), DM algorithm (GLM, RF, or MaxEnt) community type, or their interactions. When the ANOVA indicated significant differences, we conducted Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc tests to identify which groups differed (Tukey 1949).

In addition to these analyses, we used the ELC's vegetation community key to attempt to classify surveyed plots into one of the 45 defined communities. While our primary goal was to test how well the models guided us towards areas with community composition similar to the target community, we also wanted to determine whether models successfully directed us to the exact target communities. Each community is typically defined by the dominant species within each vegetation layer, and we followed this approach to classify surveys in unburned plots. However, we classified burned plots based on the percent cover of dominant species regardless of vegetation layer. For example, C62 is characterized by trembling aspen (*Populus tremuloides*) in the tree layer (layer 4, trees >5m in height) and thimbleberry (*Rubus parviflorus*) in the herb-dwarf shrub layer (layer 2, herbs). Although trembling aspen would no longer occur in the closed tree layer following the fire, plots where it remained the dominant tree species in lower vegetation layers, along with thimbleberry, would still be classified as C62.

2.3.3.3. Spatial Comparisons

DMs with similar performance scores can differ in patterns of spatial predictions (Syphard & Franklin 2009). In addition to assessing the accuracy of each DM, we examined the spatial predictions of the best performing DMs for each vegetation community. We quantified the spatial agreement of the best CDM, s-SDM, and reduced s-SDM maps for each vegetation community using Spearman's rank correlation tests. Correlation tests are an effective method to compare the spatial congruency of distribution maps made with different algorithms (Syphard & Franklin 2009).

We also compared the frequency of high suitability cells among the best DMs of each vegetation community. We classified a cell as suitable in two ways: (1) if its predicted probability of occurrence exceeded the predicted probability of any cell containing a known occurrence of the target community; or (2) if its predicted Bray-Curtis similarity (based on the GLMs built using independent survey data) exceeded the minimum pairwise Bray-Curtis similarity among all known occurrences of the target community. We then calculated the proportion of cells in WLNP predicted to be suitable for each rare plant community by each model.

2.4. Results

2.4.1. Model Performance and Efficacy

Across all 72 independent surveys, pairwise Bray-Curtis similarity to the target communities ranged from 0 to 0.566. Most DMs showed a positive relationship between predicted probabilities and community similarity (Figure 3, Figure 4, Tables S19 through S25 in Appendix 1), and all three of the top-performing CDM, s-SDM, and reduced s-

SDMs showed a positive relationship between predicted probabilities and community similarity for all seven target communities (Table 3). Ferrari's R^2 values for the GLMs relating predictions of the top-performing DMs to community similarity ranged from 0.115 to 0.536, with the highest values consistently observed for s-SDMs and the lowest for CDMs (Table 4). The three-way interaction between model type, algorithm, and community was not significant ($F = 1.00$, $df = 24$, $p = 0.461$). There was a significant interaction between model type and community ($F = 1.67$, $df = 12$, $p = 0.004$; Figure 5a), where s-SDMs were significantly better than CDMs and reduced s-SDMs in communities C61 and C79 (Table S26). There was also a significant interaction between model type and algorithm ($F = 2.76$, $df = 4\text{-}150$, $p = 0.030$; Figure 5b), where CDMs built with MaxEnt were significantly better than models built with GLMs and reduced s-SDMs built with MaxEnt were significantly better than models built with GLMs and RFs, but algorithm did not significantly influence the performance of the full s-SDMs (Table S27).

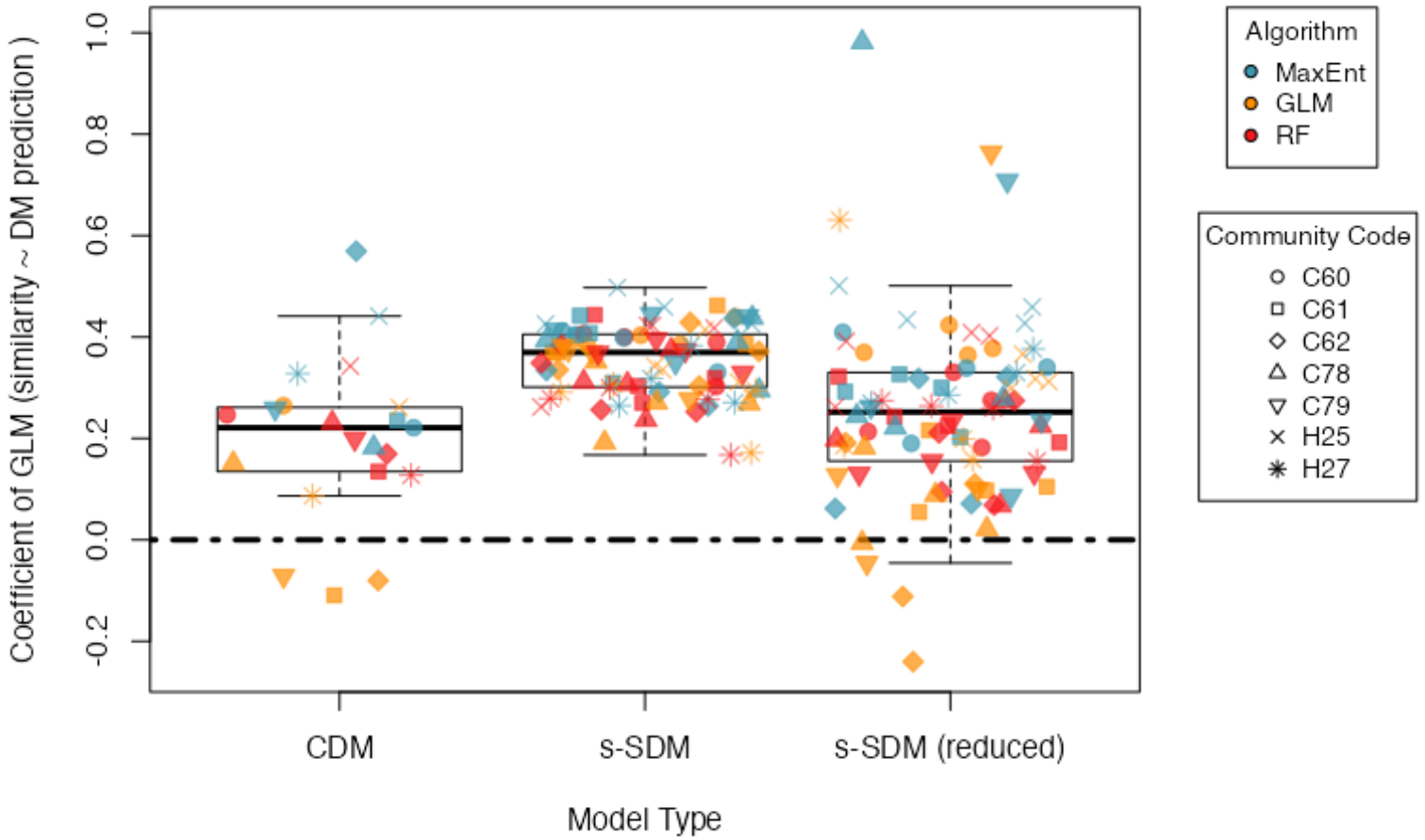

Figure 3. Boxplot of coefficients of GLMs modelling the relationship between average Bray-Curtis similarity (of each surveyed plot to the target vegetation community) and distribution model predictions. Individual model coefficients are shown as jittered points, with colour indicating the algorithm (MaxEnt: blue, GLM: orange, RF: red) and shape representing vegetation community. The dashed line indicates a slope of 0 (no relationship).

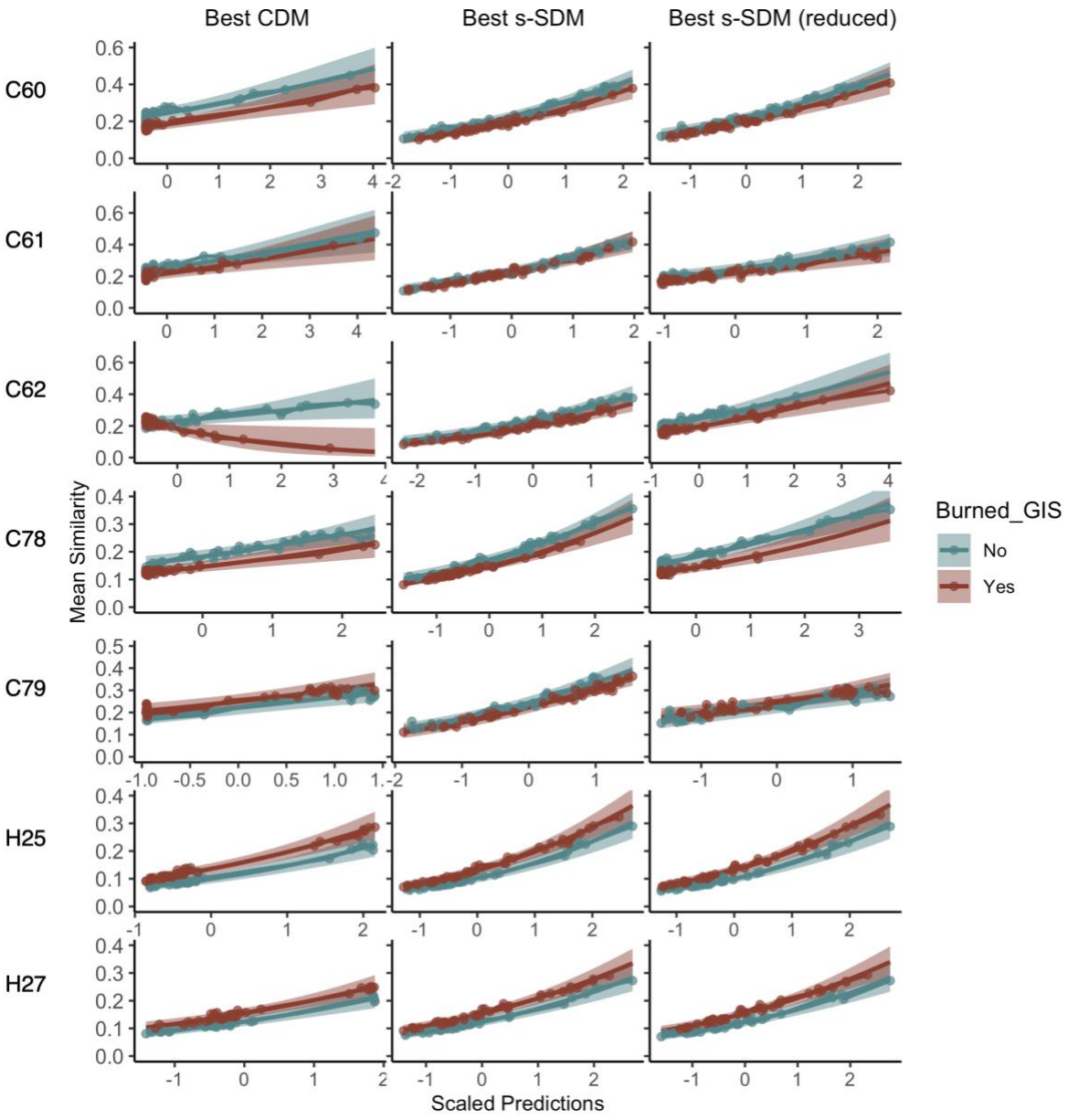

Figure 4. Partial regression plots for the best performing distribution model of each type for each vegetation community, showing the predicted relationship between the average Bray-Curtis similarity of each survey location to the target community and the scaled and transformed distribution model prediction. Actual Bray-Curtis similarity values are shown as points with fitted lines representing unburned (blue) and burned (red) plots, with shaded areas indicating 95% confidence bands.

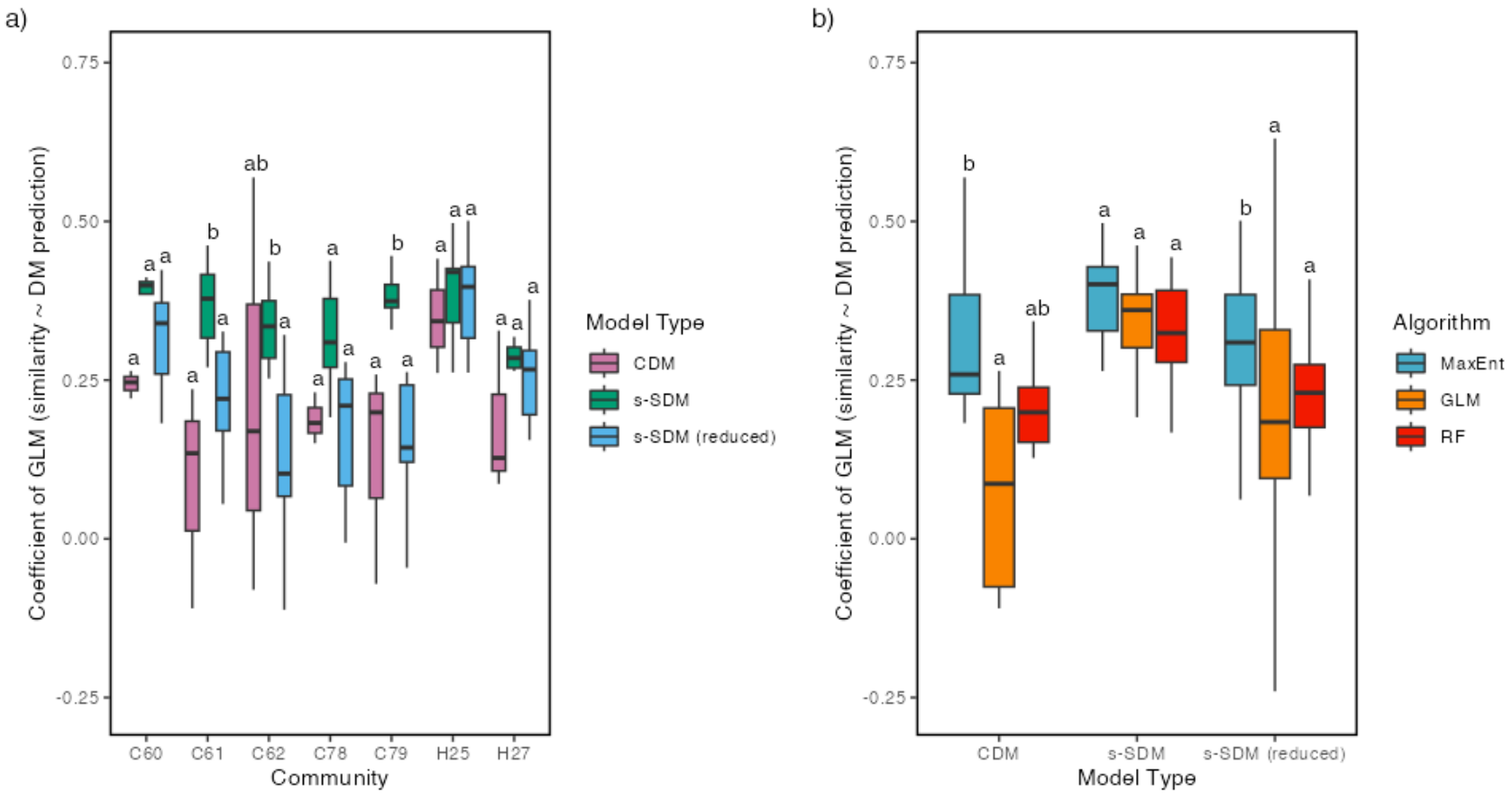

Figure 5. Boxplots showing the interactions between model type and community (a), and model type and algorithm (b) on coefficients of GLM's modelling the relationship between average Bray-Curtis similarity and distribution model predictions. (a) Each box represents the distribution of DM values for a given model type within each community, with colours indicating different model types (CDM: pink, s-SDM: green, s-SDM (reduced): blue). (b) Each box represents the distribution of DM values for a given algorithm within each model type, with colours indicating the algorithm (MaxEnt: blue, GLM: orange, RF: red). Boxes sharing the same letter are not significantly different from one another based on Tukey's HSD post-hoc test ($\alpha = 0.05$). Outliers are hidden.

Table 3. Characteristics of the best performing CDM, s-SDM, and reduced s-SDM for each of the seven vegetation communities based on GLMs relating measured community similarity to the prediction of the DM. For models without a significant interaction between the DM prediction and burn status, interaction coefficients are listed as NA. There was no selection of species for community models, therefore all CDMs are NA in the species column. The best (lowest AIC) DM of each vegetation community is bolded.

Plant Community Code	Model Type	Model Algorithm	Species Included	Arithmetic or Geometric	AIC	DM Coefficient	DM p	Burn Coefficient	Burn p	DM x Burn Coefficient	Ferrari's Pseudo-R^2
C60	CDM	GLM	NA	NA	-140.878	0.26	<0.0001	-0.36	<0.01	NA	0.239
C60	**sSDM**	**GLM**	**All**	**Arithmetic**	**-174.092**	**0.44**	**<0.0001**	**-0.18**	**0.0864**	**NA**	**0.489**
C60	sSDM Red	GLM	All	Arithmetic	-169.488	0.42	<0.0001	-0.15	0.1755	NA	0.460
C61	CDM	MaxEnt	NA	NA	-115.974	0.24	0.0007	-0.19	0.1829	NA	0.144
C61	**sSDM**	**GLM**	**All**	**Arithmetic**	**-153.205**	**0.46**	**<0.0001**	**-0.04**	**0.7085**	**NA**	**0.454**
C61	sSDM Red	MaxEnt	All	Geometric	-124.973	0.33	<0.0001	-0.13	0.2894	NA	0.220
C62	CDM	RF	NA	NA	-120.238	0.17	NA	-0.29	NA	-0.64	0.115
C62	**sSDM**	**MaxEnt**	**All**	**Arithmetic**	**-154.651**	**0.44**	**<0.0001**	**-0.23**	**<0.05**	**NA**	**0.386**
C62	sSDM Red	MaxEnt	All	Geometric	-134.672	0.32	<0.0001	-0.29	<0.05	NA	0.227
C78	CDM	RF	NA	NA	-193.632	0.23	<0.0001	-0.27	<0.05	NA	0.318
C78	**sSDM**	**RF**	**All**	**Arithmetic**	**-226.380**	**0.37**	**<0.0001**	**-0.18**	**0.0631**	**NA**	**0.536**
C78	sSDM Red	MaxEnt	All	Geometric	-206.177	0.28	<0.0001	-0.28	<0.01	NA	0.391
C79	CDM	MaxEnt	NA	NA	-127.397	0.26	0.0001	0.16	0.2100	NA	0.181
C79	**sSDM**	**MaxEnt**	**Dominants**	**Arithmetic**	**-161.141**	**0.45**	**<0.0001**	**-0.13**	**0.2406**	**NA**	**0.472**
C79	sSDM Red	MaxEnt	All	Arithmetic	-128.327	0.26	<0.0001	0.10	0.4376	NA	0.196
H25	CDM	MaxEnt	NA	NA	-212.928	0.44	<0.0001	0.35	<0.01	NA	0.400
H25	sSDM	MaxEnt	All	Arithmetic	-227.763	0.50	<0.0001	0.27	<0.05	NA	0.493
H25	**sSDM Red**	**MaxEnt**	**All**	**Arithmetic**	**-228.780**	**0.50**	**<0.0001**	**0.29**	**<0.05**	**NA**	**0.497**
H27	CDM	MaxEnt	NA	NA	-211.657	0.33	<0.0001	0.24	<0.05	NA	0.370
H27	**sSDM**	**MaxEnt**	**All**	**Arithmetic**	**-230.642**	**0.38**	**<0.0001**	**0.25**	**<0.05**	**NA**	**0.508**
H27	sSDM Red	MaxEnt	All	Arithmetic	-227.453	0.38	<0.0001	0.27	<0.01	NA	0.489

Fourteen of the 72 plots keyed out to the target vegetation communities. These included three occurrences of C60 (*Populus tremuloides/Amelanchier alnifolia/Heracleum maximum*), five occurrences of C61 (*Populus tremuloides/Rubus parviflorus*), four occurrences of C62 (*Populus tremuloides/Symphoricarpos occidentalis*), and two occurrences of C78 (*Picea engelmannii/Equisetum arvense*). The average Bray-Curtis similarity between these classified plots and target ELC plots ranged from 0.285 to 0.432 (C60), 0.264 to 0.432 (C61), 0.257 to 0.417 (C62), and 0.311 to 0.329 (C78; Table 4). The predictions of the DMs at these sites varied by model type, where full s-SDMs typically had the highest predicted values (Table 4). None of the surveyed plots keyed out to C79, H25, or H27.

Table 4. Summary of new occurrences of target vegetation communities with corresponding plot numbers, the community code, predicted suitability values for the best performing model of each model type, and the average Bray-Curtis similarity of new occurrences to target ELC plots of the same community.

Plot Number	Community Code	CDM Predicted Suitability	s-SDM Predicted Suitability	s-SDMr Predicted Suitability	Average Bray-Curtis Similarity to Target
16	C60	0.420	0.630	0.416	0.364
21	C60	0.274	0.651	0.522	0.285
711	C60	0.617	0.659	0.503	0.432
2	C61	0.000	0.279	0.013	0.264
45	C61	0.291	0.680	0.458	0.397
245	C61	0.004	0.363	0.103	0.378
777	C61	0.027	0.731	0.409	0.432
898	C61	0.032	0.368	0.167	0.379
13	C62	0.003	0.687	0.238	0.308
238	C62	0.006	0.456	0.005	0.257
416	C62	0.012	0.571	0.247	0.290
801	C62	0.122	0.688	0.161	0.417
32	C78	0.881	0.646	0.267	0.329
42	C78	0.918	0.377	0.638	0.311

For all vegetation communities, s-SDMs (full or reduced) produced the best AIC scores, whereas CDMs consistently had the worst (Table 3, Table S19 through S25). Across all communities, the full s-SDMs had significantly better performance (AIC score more than 2 units difference) than CDMs and significantly better performance than reduced s-SDMs in all communities except H25 (Table 3). The best reduced s-SDMs had significantly lower AIC than the best CDMs in all communities except C79 (Table 3, Figure 5a). While s-SDMs performed the best across all vegetation communities, for some communities s-SDMs were not significantly better than other model types (Figure 3, Figure 5a).

The best models varied in algorithm, choice of species, and stacking method within and between vegetation communities. While most top-performing models across model types were built with MaxEnt, this was not always the case (Figure 3; Table 3). For example, the top-performing CDM for C78 was built using Random Forests. CDMs and reduced s-SDMs performed significantly better when built with MaxEnt, whereas algorithm choice did not significantly affect the performance of s-SDMs (Figure 5b). All top-performing reduced s-SDMs included all characteristic species, and all but one top-performing full s-SDM also included all characteristic species (Table 3). Most top-performing reduced s-SDMs used the arithmetic mean, whereas all top-performing full s-SDMs used the arithmetic mean (Table 3).

In addition to comparing modelling methodologies, we tested whether DM predictions were affected by recent wildfire disturbance. For most DMs, the slope of the relationship between community similarity in the independent test plots and the DM predictions did not differ in burned versus unburned plots (Figure 4, Tables S19 through

S25). Only one of the top-performing models across all model types and vegetation communities showed a significantly reduced slope in burned plots compared to unburned plots (Table 3, Figure 4). This was the CDM built using RF for vegetation community C62. However, among the other 20 best performing DMs, burn status was a significant predictor of mean community similarity in 11 cases (Table 3). Interestingly, burned sites generally showed lower predicted community similarity than unburned sites in most of the non-grassland communities (Figure 4). However, in grassland communities (H25 and H27) and in two models of community C79, burned plots tended to have greater compositional similarity to the target community than unburned plots at the same DM prediction level (Figure 4).

It is important to note that for some DMs, we were not able to capture the full range of predicted probabilities with our survey plots. In 16% of all DMs, we did not sample any plots in predicted probability ranges greater than 0.5, but in 60% of all models we had at least five plots in probability ranges greater than 0.5.

2.4.2. Spatial Comparisons

Across all model types, spatial predictions generally corresponded with the known occurrences of each target community (Figure S1 through 7 in Appendix I). Areas of higher predicted probability were typically found near known community occurrences, with CDMs showing high suitability only in close proximity to known occurrences. For example, the lowest predicted probability value at any of the known occurrences of C78 and C62 based on the CDMs was 0.91 and 0.94, respectively (Table S28). Although this minimum predicted probability was not consistently this high across all communities, in

six of seven communities CDMs had the highest minimum probability threshold out of the three model types. DMs for communities dominated by *Populus tremuloides* (C60, C61, C62) closely matched their known distributions, showing higher probabilities of occurrence in lower-elevation areas of the park. The full and reduced s-SDMs produced broader predictions, identifying additional areas with non-zero predictions farther from known occurrences but still largely within the ecoregions and areas where these communities are known to occur. The probability thresholds for full and reduced s-SDMs were generally lower than CDMs.

In general, the best performing CDMs and s-SDMs of each vegetation community predict the same areas of high and low suitability across the park (Figure 6). Across all vegetation communities, the spatial predictions between the best full and reduced s-SDMs tend to be the most similar, and the least spatially similar maps tend to be the CDMs and the full s-SDMs (Figure 6, Table 5).

Table 5. Pairwise Spearman's correlation coefficients between the spatial predictions of the best community model (CDM), full stacked species distribution model (s-SDM), and reduced stacked distribution model (s-SDMr) of each vegetation community.

Community Code	CDM:sSDM	CDM:sSDMr	sSDM:sSDMr
C60	0.547	0.527	0.967
C61	0.515	0.798	0.701
C62	0.402	0.187	0.745
C78	0.617	0.471	0.589
C79	0.630	0.919	0.683
H25	0.581	0.565	0.987
H27	0.771	0.835	0.927

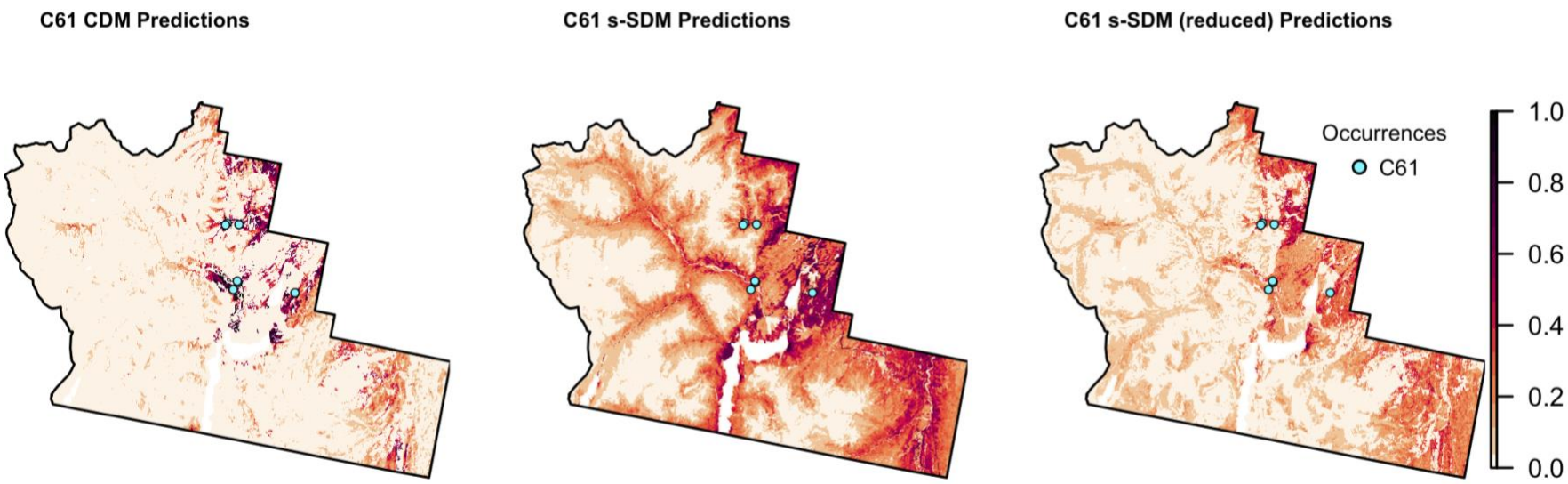

Figure 6. Scaled continuous model predictions for community C61 (*Populus tremuloides/Rubus parviflorus*) by the top performing community distribution model (CDM), stacked species distribution model (s-SDM) and stacked SDM with reduced occurrences (s-SDM reduced). Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C62 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). For maps of the other communities, see Appendix I.

The number of cells predicted suitable by the best DMs varied from 5 cells (0.0000246% of the park, CDM C62) to 172787 cells (0.849% of the park, reduced s-SDM C60; Table S29). The amount of area predicted to be suitable based on the predicted probability threshold was highest for reduced s-SDMs, followed by s-SDMs and then CDMs (Table S28, Figure S8 through S14). For example, for community C61, the amount of suitable area was 0.17%, 0.13%, and 0.0014% for the reduced s-SDM, full s-SDM, and CDM, respectively. On average, the best full s-SDMs predicted 1,244 times more suitable cells than the best CDMs, whereas the reduced s-SDMs predicted 4,013 times more suitable cells. These means are inflated by CDMs for communities with only 5 or 6 cells predicted suitable across the entire park (Table S28). If we again examine C61 as an example, the full s-SDM predicted nine times more suitable cells than the CDM, whereas the reduced s-SDM predicted 12 times more suitable cells than the CDM.

Models predicting the Bray-Curtis similarity across space tended to have areas of high suitability in the same areas that DM predictions did (Figures S15 through S21). Suitability maps based on the minimum pairwise Bray-Curtis threshold generally resembled the suitability predictions using the predicted probability threshold (Figures S22 through S28). The amount of area predicted to be suitable based on the Bray-Curtis similarity thresholds showed the same trends as thresholding using the lowest predicted probability. Generally, less suitable habitat is predicted by the CDMs than the full and reduced s-SDMs and more suitable habitat is predicted by the reduced s-SDMs compared to the full s-SDMs (Table S29). The exception to this is for community C60, where the majority of the park is predicted to be suitable based on the Bray-Curtis threshold (Table S29, Figure S22).

2.5. Discussion

Few studies have tested whether community distribution models or stacked species distribution models are better able to predict the distributions of plant communities. We built models of each type for seven rare plant communities using a variety of algorithms and stacking methods, and tested our models using independent survey data. Our results show that s-SDMs more effectively predict community distributions than CDMs, and that recent wildfire disturbance does not reduce their utility.

Overall, the distribution models showed moderate success in predicting vegetation community occurrences, with the maximum similarity value between survey sites and target communities reaching 0.556. This indicates that the most similar plots shared on average 55% of their species with the target community, in similar abundances. While this value might seem low, pairwise community similarity values among the original ELC plots classified as the same community ranged from 0.110 to 0.690, which indicates that substantial variation in species composition exists even within communities classified as the same vegetation type. Our field surveys documented 14 new occurrences of the seven target vegetation communities. Although 14 occurrences represents only 18% of the total sampling sites, this outcome reflects the study's sampling design, which required surveys across a range of predicted probabilities for multiple communities rather than focusing exclusively on areas of the highest predicted probability of occurrence. Therefore, we demonstrated that DMs are useful for predicting community composition

and effectively guide surveys towards locations more likely to be occupied by the target communities.

2.5.1. Performance of CDMs versus stacked SDMs

Stacked SDMs built with all available species occurrences consistently outperformed the CDMs (average reduction in AIC of 29). The superior performance of s-SDMs might be due entirely to the greater number of presence records available for individual species than for rare communities. Distribution models built with larger sample sizes are more likely to capture the full range of conditions experienced by a species or community, reducing the risk of model overfitting (Elith et al. 2006; Wisz et al. 2008; Boyd et al. 2023). Across communities, there were an average of 1.4 to 13.6 times more species-level records available for the dominant species than the number of corresponding CDM records, depending on the community. The larger number of occurrence records allowed the SDMs to capture a broader range of environmental conditions association with each species. As a result, the predictions of the s-SDMs were less conservative than the CDMs. This is illustrated by the CDMs consistently estimating the suitability of surveyed cells that were classified as the target community at lower values than the s-SDMs (Table 4). These findings align with previous studies showing that DMs underperform when there are limited occurrence records (Ferrier & Guisan 2006; Wisz et al. 2008; Dubuis et al. 201l; Jiménez-Alfaro et al. 2018). For example, Jiménez-Alfaro et al. (2018) found that while CDMs performed reasonably well at large spatial scales, their performance was reduced when fewer occurrence records were included.

It is also possible that s-SDMs performed better than CDMs because they more accurately reflect the way communities are assembled. Stacking individual SDMs might work better because communities are formed, as Gleason argued, as a result of species responding independently to environmental gradients (Gleason 1926; Guisan & Rahbek 2011; Butler et al. 2022). To test whether the performance of s-SDMs was simply due to the use of more occurrence data, we built a set of "reduced" s-SDMs with the number of records for each species no greater than the number of community occurrences. Even with these reduced datasets the s-SDMs still outperformed the CDMs. This superior performance even with reduced data supports the idea that communities are best modelled as the sum of individual species rather than as single cohesive units. However, it is difficult to be sure that our results reflect the underlying processes of community assembly, because the reduced s-SDMs still used more data than the CDMs as multiple species contributed information to the final community prediction. Nevertheless, the implication is that stacking individual species models provides more reliable predictions of a community's distribution, at least when there are relatively few occurrence records of the community as a whole.

While the choice of CDM or s-SDM had the greatest effect on model performance, the DM algorithm also influenced model accuracy. Most of the top performing models across all types were built with MaxEnt, which is consistent with its demonstrated strong performance relative to other algorithms (Elith et al. 2006; Phillips et al. 2006; Wisz et al. 2008). Overall, algorithm choice mattered the most for CDMs and reduced s-SDMs. CDMs built with Maxent outperformed GLMs, whereas reduced s-SDMs built with Maxent outperformed GLMs and RF. In contrast, the performance of

full s-SDMs was not affected by algorithm, most likely because all model types perform well when there are plenty of occurrence records available. Based on these results, we recommend Maxent as the most reliable algorithm choice, especially when occurrence data is limited.

The performance of the s-SDMs was also impacted by the choice of which species to stack and which averaging method to use. When modelling species richness, s-SDM performance can be substantially altered by methodological approaches such as the stacking approach and thresholding (Elith et al. 2006; Dubuis et al. 2011; Calabrese et al. 2014; Norberg et al. 2019; Zurell et al. 2020), but no studies have assessed how different averaging methods influence predictions of community composition. Our study shows that s-SDMs including all component species stacked by taking the arithmetic mean generally achieved the best performance. We suspect this is because averaging species probabilities reduces the influence of low probability values without excluding them, allowing areas of low suitability for a few component species to still be predicted as suitable overall. For reduced s-SDMs, performance was better with the arithmetic mean when occurrence records were relatively abundant (>10 records), whereas the geometric mean improved model fit when data were sparse (<10 records). The geometric mean provides a more conservative stacking approach because it requires non-zero probabilities for all component species. This can be helpful when models are uncertain, but may be overly restrictive when model fits are already strong. In one community (C79), a full s-SDM including only the dominant species with an arithmetic mean performed best. In this case, dominant species can sufficiently define the environmental conditions characterizing a community and adding more species reduces overall model performance.

The AIC of this model was six units lower than that of the best performing model that was built with all species. This reduction in fit when more species were included may reflect the introduction of noise from species with more limited occurrence records, which can lower predictive accuracy (Pearson et al. 2007). Clearly methodological choices can interact with data availability to influence the accuracy of s-SDMs, and models can be tailored to the amount of data available for the target community. When few occurrences are available the geometric mean is the best method, but the arithmetic mean produces the best performance when data is not limited.

Both grassland communities (H25 and H27) and community C79 showed strong correlations between DM predictions and community similarity across all model types, yet I did not encounter these exact communities in the field. For H25 and H27, this may be because several survey plots occurred on or near an ecotone, where the target dominant species were present but the inclusion of species from forested communities prevented their classification as grassland communities. In some cases, this heterogeneity was driven by the encroachment or regeneration of *Populus tremuloides* into the grasslands, shifting the composition of what once may have been a grassland towards a forest assemblage. Encroachment of forests into grasslands has been ongoing for more than a century (Levesque 2005, Stockdale et al. 2019). However, it may also be that even as predicted probability and community similarity increase, these target grasslands simply no longer exist due to dramatic shifts in composition caused by non-native species. Low-elevation grassland communities in the park, including H25 and H27, are threatened by highly competitive exotic species (e.g. *Poa pratensis* and *Phleum pratense*) that outcompete native vegetation (Achuff et al. 2002; Waterton Lakes National Park of

Canada Management Plan 2022). Consistent with this, many of our high probability and high similarity plots contained substantial cover (8-40%) of these invasive grasses, while the target dominants occurred in lower abundances. Thus, models were effective in identifying areas that are environmentally suitable for H25 and H27 to occur but could not result in the classification of these exact communities.

Similarly to the grasslands, we were unable to classify any communities as C79 even though increasing predicted probability coincided with more compositionally similar plots. Despite high similarity, these sites could not be formally classified as C79 because the ELC classification scheme requires *Pinus contorta* dominance. Post-fire regeneration of lodgepole pine has been variable across the park due to drought conditions (Musk et al. 2024; Aspinall et al. 2025) and burned sites with high similarity did not yet have sufficient *P. contorta* to be meet this criterion. Some unburned plots also had high similarity but could not be classified as C79 because other species were also dominating in the under and overstorey along with the target dominants. Although high predicted probabilities did not always correspond to the presence of the exact target community, this outcome is not unexpected. These models are not intended to pinpoint every existing occurrence of a community but to identify environmentally suitable areas where these communities are likely to be found.

2.5.2. Effects of wildfire

We tested whether recent wildfire disturbance reduced the predictive accuracy of the DMs. Interestingly, we found that recent wildfire disturbance did not alter the slope of the relationship between DM predictions and community similarity for the best DMs of

six of the seven vegetation communities in WLNP. This suggests that, in general, the predictions of the best DMs remain just as effective in recently disturbed areas as they are in later-successional stands. This could be a result of the way post-fire succession occurs in this region. Following disturbance, the same species frequently return or resprout, allowing the community to retain many of its original species even though their abundances may differ (Lyon & Stickney 1976). Long-term monitoring studies across the Rocky Mountains, plains, and boreal forests of western North America have also shown that generally, the same species present pre-fire return within a few years post-fire in understory forest communities (Romme et al. 2016; Day et al. 2021; Hamilton & Burton 2023), shrublands (Lantz et al. 2010) and grasslands (Porensky et al. 2018; Hamilton & Burton 2023). For example, a study by Fornwalt et al. (2014) in the Colorado Front Range showed that 90% of pre-fire understorey species returned within five years of severe wildfire. Likewise, Lloren (2021) found that severely burned plots in WLNP retained 37% of their original species on average just three years post-fire, suggesting that many species persisted through disturbance or quickly recolonized. Even after severe wildfire, DMs can be used to model community distributions in the southern Canadian Rockies, but this may not hold for other regions or disturbance types.

We expected that burned plots of a given predicted suitability would have lower similarity to the target vegetation community than unburned sites of equal suitability; however, we found that this was not consistent across vegetation communities. Interestingly, burned plots were on average more compositionally similar to target ELC plots for all the best DMs of grassland communities H25 and H27, and for two of the three best DMs for a lodgepole pine (*Pinus contorta*; C79) community. These

communities may require some level of disturbance, such as wildfire, to maintain the composition recorded in the ELC surveys. The ELC surveys were completed nearly 30 years ago, and during this time other processes were acting upon these communities to alter their compositions, such as woody plant encroachment (Lloren 2021; Wieczorkowski & Lehmann 2022), invasion by non-native species (Gharehaghaji et al. 2019), and the historical effects of grazing (Vild & Rotherham 2021). For example, grassland communities are often maintained by disturbances such as wildfire that help to suppress the encroachment of woody species and reduce the abundance of invasive species (Bond & Keeley 2005; Keeley et al. 2011). In the case of *P. contorta* communities, lodgepole pine requires fire to release the seeds from its serotinous cones, helping to maintain its dominance where recent burns have occurred (Johnstone et al. 2016). Therefore, recent disturbance may not uniformly reduce DM accuracy but depends on the ecology of individual communities.

In only one of the best DMs did we observe an interaction between burn status and DM prediction. For community C62, which is characterized by trembling aspen (*Populus tremuloides*) in the tree layer and western snowberry (*Symphoricarpos occidentalis*) in the shrub layer, there was a positive relationship between similarity and CDM predictions in the unburned sites but a declining trend in the burned sites. In the highest probability burned site, both dominant species occurred at low cover while Saskatoon (*Amelanchier alnifolia*) and Kentucky bluegrass (*Poa pratensis*) dominated. The highest probability sites did have *P. tremuloides* dominating but lacked *S. occidentalis* dominance, with *P. pratensis* and Timothy grass (*Phleum pratense*) dominating instead. Most test sites for this model were found below 0.20 suitability, with

only one burned site between 0.20 and 0.40, which potentially influenced this interaction. Although this was the best CDM based on AIC, this CDM was limited in its ability to predict community composition. This is supported by the fact that we did find four locations of C62 which had all been burned, yet the CDMs predicted probabilities at these locations only ranged between 0.003 and 0.122 (Table 4). This interaction is therefore unlikely to reflect a meaningful interaction but instead is indicative of poor model performance from this CDM, which failed to identify suitable occurrences of the target community.

2.5.3. Spatial Predictions

Overall, the maps showed spatial patterns expected for the target vegetation communities in the region. Deciduous communities were more concentrated in lower-elevation valleys, while grasslands were found in more open areas found across low and high elevations in the park, consistent with our existing knowledge of these communities (Achuff et al. 2002). These patterns suggest that the models generally reflect realistic environmental drivers of community occurrence. However, some discrepancies were evident, where predicted distributions extended more broadly across the park than anticipated. This was especially apparent in s-SDMs, which capture the individual species-environment relationships and produced more generous predictions of occurrence across the park. Because stacked models do not account for biotic interactions or the abundance of species, they can predict suitable conditions for co-occurrence even in locations where those species may not actually form the target community (Calabrese et al. 2014). These over-predictions are not necessarily unrealistic, but rather reflect the

models' focus on predicting environmental suitability rather than their realized occupancy by these communities. These continuous predictions are therefore informative in a conservation context, as they highlight areas that share the environmental conditions that are characteristic of each community which can help to guide field surveys to these most probable locations.

The s-SDMs generally predicted much more area to be suitable than CDMs, although the degree of spatial overlap between model types varied widely among communities. Spearman's correlations between CDMs and s-SDMs ranged from 0.402 to 0.771 (mean = 0.576), while those between CDMs and reduced s-SDMs were more variable, ranging from 0.187 to 0.919 (mean = 0.619). In contrast, Spearman's correlations between full and reduced s-SDMs were consistently high and ranged from 0.589 to 0.987 (mean = 0.857), indicating that reducing the number of occurrence records did not greatly affect the overall spatial predictions. Newly discovered occurrences of the target communities were generally assigned higher predicted suitability by the s-SDMs compared to the CDMs, demonstrating the effectiveness of s-SDMs in identifying suitable locations of these communities. Our findings suggest that s-SDMs may also outperform CDMs in their spatial predictions and may offer more reliable guidance for locating rare plant community occurrences, even though they may be overly generous in the area predicted to be highly suitable.

CDMs predicted much less of the park to be suitable compared to the s-SDMs, likely due to the limited number of occurrence records available to build the CDMs. Models trained with few presences – especially if they are located within a narrow set of similar environmental conditions – tend to overfit those conditions, restricting predictions

to similar environments (Guisan & Thuiller 2005; Araújo & Luoto 2007; Radomski et al. 2022). Consequently, CDMs likely captured an overly restricted environmental niche for most communities, particularly those with fewer than ten occurrences. CDMs most likely underestimate the true range of environments where a community can persist. In contrast, s-SDMs incorporated species-level models built from greater numbers of presences that spanned wider environmental gradients, reflecting the full range of conditions individual species can tolerate. As a result, they predicted greater proportions of the park as suitable, including areas more distant from known occurrences. These predictions likely overestimate true occupancy, but they provide a more inclusive view of the potential environmental conditions each community could occupy based on its component species.

Reduced s-SDMs generally predicted a greater amount of suitable area than CDMs and full s-SDMs. To match the number of community occurrences, we built reduced s-SDMs using a random subset of presence points from across all available presences. Because this random selection often provided an incomplete representation of the environmental conditions a species can occupy, models could either over or underestimate suitable areas. If the randomly selected occurrences represent only part of a species' environmental breadth, the model will underestimate the range of conditions the species can occupy (Thuiller et al. 2004; Chevalier et al. 2021). However, when few occurrences represent the full environmental range of conditions experienced by a species, each occurrence contributes equally to model fitting. As a result, extreme conditions can be predicted to be as suitable as the average conditions a species experiences, producing poorly defined response curves and overgeneralized suitability estimates (Pearson et al. 2004; Thuiller et al. 2004; Moudrý et al. 2023). We expect this

effect led to inflated estimates of suitable area by the reduced s-SDMs. This effect is particularly pronounced for species with broader geographic or environmental ranges, which are inherently more difficult to model accurately than species with more specialized niches (Hernandez et al. 2006; Mateo et al. 2010; Guo et al. 2015). Moreover, stacking models built with small datasets can compound errors from individual species models (e.g. due to sampling bias, ignoring biotic interactions, equilibrium assumptions, etc.), resulting in predicted distributions that are more extensive than are actually suitable (Guisan & Thuiller 2011; Araújo & Luoto 2007; Henderson et al. 2014). The best reduced s-SDMs built with the fewest occurrences generally utilized the geometric mean, potentially helping to constrain these overestimations spatially. In contrast, full s-SDMs consistently used the arithmetic mean and still predicted lower suitable area estimates than their reduced s-SDM counterparts, demonstrating that better calibrated models produced more realistic spatial predictions. When limited occurrence data is available, applying the geometric mean during stacking may help to constrain these overestimations and produce more ecologically realistic results.

The true prevalence of each of these communities in the park is unknown; therefore, we cannot quantify how much the DMs are overestimating or underestimating true distributions. However, the value of the DM predictions lies less in the total area they predict to be suitable than in their ability to guide field surveys to where these communities are most likely to occur. For example, Guisan et al. (2006) showed that SDMs could improve the chance of finding target species up to four times over random surveys. Thus, while the total amount of 'suitable' area should be interpreted cautiously, these models provide an ecologically informed, independently tested foundation for

guiding search efforts and improving our understanding of the communities' true distributions.

2.5.4. Conclusions

Predicting the distributions of plant communities is essential for understanding the processes that shape them and for guiding conservation efforts in the face of accelerating environmental change and disturbance (Guisan et al. 2013; Guillera-Arroita et al. 2015). While species distribution models have been used extensively to map individual species, relatively few studies have evaluated their use for mapping communities (Ferrier & Guisan 2006; Dubuis et al. 2011; D'Amen et al. 2017). Our results demonstrated that stacked species distribution models outperformed community distribution models, illustrating that species-level modelling was more effective in predicting plant community composition, even in recently disturbed landscapes. This finding aligns with an individualistic perspective of community assembly (Gleason 1926), where communities are better modelled as the overlapping occurrences of individualistic species rather than as cohesive superorganisms (Clements 1916). Thus, stacking single-species models provides a practical approach to mapping vegetation communities. Moreover, because s-SDMs rely on individual species occurrences, they also offer practical advantages in data deficient scenarios by allowing researchers to leverage publicly available records from databases such as the Global Biodiversity Information Facility (GBIF). S-SDMs can reliably predict community composition from species-level information, providing a way to predict the distributions of rare or threatened plant communities.

CHAPTER 3: CONCLUSION

Conservation efforts often focus on identifying areas of high species richness, but biodiversity is more than just the number of species present. Effective management also requires understanding community composition and how species co-occur across the landscape, preserving not only the diversity of species but the diversity of communities as well. Predicting the distribution of plant communities is particularly important for rare assemblages with limited occurrences, where conservation decisions depend on knowledge of the composition of communities and where they occur across the landscape.

Predictive models such as SDMs are increasingly used to understand the distributions of species and their assemblages, helping to inform current and future conservation and management (Elith et al. 2006; Henderson et al. 2014). However, our understanding of whether plant communities are best modelled by viewing them as a 'superorganism' or in a more individualistic manner remains limited (Clements 1916; Gleason 1926), particularly for data deficient assemblages in disturbed landscapes (Ferrier & Gusian 2006; D'Amen et al. 2017). Most studies that have modelled communities based on these two views were focused on their utility for mapping broad scale vegetation types or predicting species richness (Henderson et al. 2014; Butler et al. 2022). While some support has been found for both approaches, it is unknown how the effects of data deficiency and disturbance can influence model predictions and evaluation.

This thesis evaluated whether modelling rare plant communities as units (CDMs) or as the sum of their individual species (s-SDMs) more effectively predicted rare plant

communities in a recently disturbed landscape. Consistent with the individualistic view of community assembly, s-SDMs outperformed CDMs in predicting community composition. Although full s-SDMs had the highest predictive performance, s-SDMs built with fewer records still consistently outperformed CDMs. Full s-SDMs were more effective than CDMs particularly for communities with the fewest occurrences available, highlighting the utility of stacked models in predicting the distributions of rare communities. Similarly, performance was only impacted by algorithm choice in CDMs and reduced s-SDMs, where DMs built using MaxEnt outperformed DMs built with Random Forest or Generalized Linear Models. This is in line with findings that MaxEnt is less sensitive to sample size than other algorithms (Hernandez et al. 2006; Phillips et al. 2009).

As disturbance events become more common, it is important to understand whether the performance of distribution models is impacted by successional processes (Ferrier at el. 2002; Johnstone et al. 2010; Pottier et al. 2013). Our study revealed that, for most communities, the relationship between model predictions and community similarity was not affected by a recent wildfire, although similarity tended to be lower in burned sites than in unburned sites with the same predicted suitability. However, in a few communities, burned plots were more compositionally similar to the target assemblages than unburned plots with similar DM predictions, indicating that disturbance is necessary for some communities to maintain their composition over time (Bond & Keely 2005).

Although this study demonstrated greater efficacy of modelling plant communities using a stacked approach compared to a community-unit approach, further studies should compare the performance and predictions of modelling approaches that

explicitly account for biotic interactions, such as joint species distribution models (J-SDMs). Previous studies have directly compared s-SDMs and J-SDMs when predicting species richness, but these comparisons have produced conflicting results regarding which methodology is superior (Norberg et al. 2019; Zurell et al. 2020). Testing J-SDMs with independent data, as done here, would provide a more robust assessment of their ability to predict community composition. Such testing could show whether J-SDMs provide more accurate guidance for conservation planning in data-limited scenarios than stacked approaches, building on my findings to identify the most efficient and accurate method for mapping rare plant communities.

By demonstrating the effectiveness of s-SDMs in predicting plant community distributions, this study provides an approach for informing conservation planning and management decisions. Accurate maps showing where these rare vegetation communities occur within the park can help managers to implement monitoring programs to better conserve them and ensure that planned developments (e.g. new buildings and trails) avoid areas with high probability of rare community occurrence. Moreover, the consistent accuracy of models across burned and unburned areas indicates that they remain reliable tools for identifying priority areas even after disturbance.

References

Achuff, P. L. (1997). Special plant and landscape features of Waterton Lakes National Park, Alberta. Parks Canada, Waterton Lakes National Park, Alberta.

Achuff, P. L., R. L. McNeil, M. L. Coleman, C. Wallis and C. Wershler. (2002). Ecological land classification of Waterton Lakes National Park, Alberta. Vol. I: integrated resource description. Parks Canada, Waterton Park, Alberta. 226 pp.

Aguirre-Gutiérrez, J., Carvalheiro, L. G., Polce, C., van Loon, E. E., Raes, N., Reemer, M., & Biesmeijer, J. C. (2013). Fit-for-purpose: species distribution model performance depends on evaluation criteria -Dutch hoverflies as a case study. *PloS one*, *8*(5), e63708. https://doi.org/10.1371/journal.pone.0063708

Akaike H. *Information theory and an extension of the maximum likelihood principle.* (1973). In: B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Anderson, R. P. (2013). A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York academy of sciences*, *1297*(1), 8-28. https://doi.org/https://doi.org/10.1111/nyas.12264

Araújo, M. B., & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global ecology and biogeography*, *16*(6), 743-753. https://doi.org/10.1111/j.1466-8238.2007.00359.x

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science advances*, *5*(1), eaat4858. https://doi.org/doi:10.1126/sciadv.aat4858

Aspinall, J., Chasmer, L., Coburn, C., & Hopkinson, C. (2025). Post-fire vegetation regeneration during abnormally dry years following severe montane fire. *Forest ecology and management*, *S87*. https://doi.org/10.1016/j.foreco.2025.122750

Austin, M. P. (1987). Models for the analysis of species' response to environmental gradients. *Vegetation*, *69*(1/3), 35-45. https://doi.org/10.1007/BF00038685

Austin, M. P., & Van Niel, K. P. (2011). Improving species distribution models for climate change studies: variable selection and scale. *Journal of biogeography*, *38*(1), 1-8. https://doi.org/10.1111/j.1365-2699.2010.02416.x

Bakker, J. D., Price, J. N., Henning, J. A., Batzer, E. E., Ohlert, T. J., Wainwright, C. E., Adler, P. B., Alberti, J., Arnillas, C. A., Biederman, L. A., Borer, E. T., Brudvig,

L. A., Buckley, Y. M., Bugalho, M. N., Cadotte, M. W., Caldeira, M. C., Catford, J. A., Chen, Q., Crawley, M. J., Daleo, P., Dickman, C. R., Donohue, I., DuPre, M. E., Ebeling, A., Eisenhauer, N., Fay, P. A., Gruner, D. S., Haider, S., Hautier, Y., Jentsch, A., Kirkman, K., Knops, J. M. H., Lannes, L. S., MacDougall, A. S., McCulley, R. L., Mitchell, R. M., Moore, J. L., Morgan, J. W., Mortensen, B., Olde Venterink, H., Peri, P. L., Power, S. A., Prober, S. M., Roscher, C., Sankaran, M., Seabloom, E. W., Smith, M. D., Stevens, C., Sullivan, L. L., Tedder, M., Veen, G. F., Virtanen, R., & Wardle, G. M. (2023). Compositional variation in grassland plant communities. *Ecosphere*, *14*(6). https://doi.org/10.1002/ecs2.4542

Baldwin, K., Chapman, K., Meidiner, D., Uhlig, P., Allen, L., Basquill, S., Faber-Langendoen, D., Flynn, N., Kennedy, C., Mackenzie, W., Major, M., Meades, W., Morneau, C., Saucier, J.-P. (2019). The Canadian National Vegetation Classification: Principles, Methods and Status. Natural Resources Canada, Canadian Forest Service. Information Report GLC-X-23. 162p.

Bando, F. M., Figueiredo, B. R. S., Moi, D. A., Thomaz, S. M., Michelan, T. S., García–Girón, J., Heino, J., Alahuhta, J., Romero, G. Q., & Mormul, R. P. (2023). Invasion by an exotic grass species homogenizes native freshwater plant communities. *The Journal of ecology*, *111*(4), 799-813. https://doi.org/10.1111/1365-2745.14061

Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PloS one*, 13(3), e0193085. https://doi.org/10.1371/journal.pone.0193085

Barnes, K. W., Niemuth, N. D., Vest, J. L., Fields, S. P., Estey, M. E., & Iovanna, R. (2025). Spatial predictions of potentially undisturbed grassland across the conterminous US. *Landscape ecology*, *40*(8), 168. https://doi.org/10.1007/s10980-025-02187-w

Baselga, A., & Araújo, M. B. (2009). Individualistic vs community modelling of species distributions under climate change. *Ecography*, *32*(1), 55-65. https://doi.org/https://doi.org/10.1111/j.1600-0587.2009.05856.x

Beniston, M. (2006). Mountain weather and climate: a general overview and a focus on climatic change in the Alps. *Hydrobiologia*, *562*(1), 3-16. https://doi.org/10.1007/s10750-005-1802-0

Bond, W. J., & Keeley, J. E. (2005). Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems. *Trends in ecology & evolution*, *20*(7), 387-394. https://doi.org/https://doi.org/10.1016/j.tree.2005.04.025

Boyd, R. J., Harvey, M., Roy, D. B., Barber, T., Haysom, K. A., Macadam, C. R., Morris, R. K. A., Palmer, C., Palmer, S., Preston, C. D., Taylor, P., Ward, R., Ball, S. G.,

& Pescott, O. L. (2023). Causal inference and large-scale expert validation shed light on the drivers of SDM accuracy and variance. *Diversity & distributions*, *29*(6), 774-784. https://doi.org/10.1111/ddi.13698

Bradter, U., Thom, T. J., Altringham, J. D., Kunin, W. E., & Benton, T. G. (2011). Prediction of National Vegetation Classification communities in the British uplands using environmental data at multiple spatial scales, aerial images and the classifier random forest. *The Journal of applied ecology*, *48*(4), 1057-1065. https://doi.org/10.1111/j.1365-2664.2011.02010.x

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs*, *27*(4), 326-349. https://doi.org/10.2307/1942268

Brooks, M.E., Kristensen, K., van Benthem, K. J., Magnusson, A., W. Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R journal*, *9*(2), 378-400. doi: 10.32614/RJ-2017-066.

Bruno, J. F., Stachowicz, J. J., Bertness, M. D. (2003). Inclusion of facilitation into ecology theory. *Trends in ecology and evolution*, *18*, 119-125.

Butler, L., & Sanderson, R. A. (2022). National-scale predictions of plant assemblages via community distribution models: leveraging published data to guide future surveys. *The Journal of applied ecology*, *59*(6), 1559-1571. https://doi.org/10.1111/1365-2664.14166

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global ecology and biogeography*, *23*(1), 99-112. https://doi.org/10.1111/geb.12102

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, *486*(7401), 59-67. https://doi.org/10.1038/nature11148

Clements, F. E. (1916). *Plant succession: An analysis of the development of vegetation*. Carnegie Institute of Washington Publication.

Chapman, D. S., & Purse, B. V. (2011). Community versus single-species distribution models for British plants. *Journal of biogeography*, *38*(8), 1524-1535. https://doi.org/10.1111/j.1365-2699.2011.02517.x

Chiaverini, L., Macdonald, D. W., Hearn, A. J., Kaszta, Ż., Ash, E., Bothwell, H. M., Can, Ö. E., Channa, P., Clements, G. R., Haidir, I. A., Kyaw, P. P., Moore, J. H., Rasphone, A., Tan, C. K. W., & Cushman, S. A. (2023). Not seeing the forest for the trees: Generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids. *Ecological informatics*, *75*, 102026. https://doi.org/10.1016/j.ecoinf.2023.102026

Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783-2792. https://doi.org/10.1890/07-0539.1

D'Amen, M., Rahbek, C., Zimmermann, N. E., & Guisan, A. (2017). Spatial predictions at the community level: from current approaches to future frameworks. *Biological reviews of the Cambridge Philosophical Society*, *92*(1), 169-187. https://doi.org/10.1111/brv.12222

Dawe, D. A., Parisien, M.-A., Van Dongen, A., & Whitman, E. (2022). Initial succession after wildfire in dry boreal forests of northwestern North America. *Plant ecology*, *223*(7), 789-809. https://doi.org/10.1007/s11258-022-01237-6

Day, N. J., Carrière, S., & Baltzer, J. (2021). *Pre-fire forest type drives long-term changes in understory plant communities post-fire, Taiga Shield, Northwest Territories* (Report). Forest Ecology and Management. Available online from: https://forestecology.ca/wp-content/uploads/2021/06/Tibbittreport_Final.pdf. Accessed on: September 22nd, 2025

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology*, *90*(12), 3566-3574. https://doi.org/10.1890/08-1823.1

Deschamps, G., Poggiato, G., Brun, P., Galiez, C., & Thuiller, W. (2023). Predict first–assemble later versus assemble first–predict later: revisiting the dilemma for functional biogeography. *Methods in ecology and evolution*, *14*(10), 2680-2696. https://doi.org/https://doi.org/10.1111/2041-210X.14203

Dormann, C. F., et al. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of biogeography*, *39*(12), 2119–2131.

Draper, D., Marques, I., & Iriondo, J. M. (2019). Species distribution models with field validation, a key approach for successful selection of receptor sites in conservation translocations. *Global ecology and conservation*, *19*, e00653. https://doi.org/https://doi.org/10.1016/j.gecco.2019.e00653

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., & Guisan, A. (2011). Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity &*

distributions, *17*(6), 1122-1131. https://doi.org/10.1111/j.1472-4642.2011.00792.x

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., & Guisan, A. (2013). Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of vegetation science*, *24*(4), 593-606. https://doi.org/10.1111/jvs.12002

Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological monographs*, *67*(3), 345. https://doi.org/10.2307/2963459

Elith, J., & Burgman, M. (2002). Predictions and their validation: Rare plants in the Central Highlands, Victoria, Australia. Washington.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., R. Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, *40*(1), 677-697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Earth Resources Observation and Science (EROS) Center. (2020). Landsat 8-9 Operational Land Imager / Thermal Infrared Sensor Level-2, Collection 2 [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9OGBGM6

Esser, L. F., Neves, D., & Jarenkow, J. A. (2025). Species distribution models to help integrate community ecology. *Austral ecology*, *50*(7), e70091. https://doi.org/https://doi.org/10.1111/aec.70091

Feeley, K. J., & Silman, M. R. (2011). Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and distributions*, *17*(6), 1132-1140. https://doi.org/https://doi.org/10.1111/j.1472-4642.2011.00813.x

Ferrier, S., Drielsma, M., Manion, G., & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and conservation*, *11*(12), 2309-2338. https://doi.org/10.1023/A:1021374009951

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *The Journal of applied ecology*, *43*(3), 393-404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, *24*(1), 38-49. https://doi.org/10.1017/S0376892997000088

Finderup Nielsen, T., Sand-Jensen, K., Dornelas, M., & Bruun, H. H. (2019). More is less: net gain in species richness, but biotic homogenization over 140 years. *Ecology letters*, 22(10), 1650-1657. https://doi.org/https://doi.org/10.1111/ele.13361

Finn, K. J., Bergman, J. C., & Lee-Yaw, J. A. (2024). Deciding where to put them: sensitivity tests and independent evaluation are critical when using species distribution models to inform conservation translocations. *The Journal of applied ecology*, *61*(4), 713-732. https://doi.org/10.1111/1365-2664.14616

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS one*, *9*(5), e97122. https://doi.org/10.1371/journal.pone.0097122

Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in physical geography*, *19*(4), 474-499. https://doi.org/10.1177/030913339501900403

Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeography. *Diversity & distributions*, *16*(3), 321-330. https://doi.org/10.1111/j.1472-4642.2010.00641.x

Gause, G. F. (1934). *The Struggle for Existence.* Williams and Wilkins, Baltimore, MD.

Gharehaghaji, M., Kobal, S., Reklau, R., & Minor, E. S. (2019). Management slows down invasion by non-native plants but does not prevent community change over 35 years in urban forests of the Midwestern USA. *Forest ecology and management*, *448*, 424-431. https://doi.org/https://doi.org/10.1016/j.foreco.2019.06.028

Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, *53*(1), 7-26. https://doi.org/10.2307/2479933

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global ecology*

and biogeography, *24*(3), 276-292. https://doi.org/https://doi.org/10.1111/geb.12268

Guisan, A., & Theurillat, J. P. (2000). Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, *30*, 353-384. https://doi.org/10.1127/phyto/30/2000/353

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, *135*(2), 147-186. https://doi.org/10.1016/S0304-3800(00)00354-9

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, *8*(9), 993-1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P., & Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology letters*, *16*(12), 1424-1435. https://doi.org/https://doi.org/10.1111/ele.12189

Guo, C., Lek, S., Ye, S., Li, W., Liu, J., & Li, Z. (2015). Uncertainty in ensemble modelling of large-scale species distribution: effects from species characteristics and model techniques. *Ecological modelling*, *306*, 67-75. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2014.08.002

Hamilton, N. P., & Burton, P. J. (2023). Wildfire disturbance reveals evidence of ecosystem resilience and precariousness in a forest–grassland mosaic. *Ecosphere*, *14*(3), e4460. https://doi.org/10.1002/ecs2.4460

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*, 773–785.

Henderson, E. B., Ohmann, J. L., Gregory, M. J., Roberts, H. M., Zald, H., & Goslee, S. (2014). Species distribution modelling for plant communities: stacked single species or multivariate modelling approaches? *Applied vegetation science*, *17*(3), 516-527. https://doi.org/10.1111/avsc.12085

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m:

global gridded soil information based on machine learning. *PloS one*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., & Mayfield, M. M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annual review of ecology, evolution, and systematics*, *43*(1), 227-248. https://doi.org/10.1146/annurev-ecolsys-110411-160411

Hogg, S. E., Wang, Y., & Stone, L. (2021). Effectiveness of joint species distribution models in the presence of imperfect detection. *Methods in ecology and evolution*, *12*(8), 1458-1474. https://doi.org/https://doi.org/10.1111/2041-210X.13614

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., Gonzalez, A., Duffy, J. E., Gamfeldt, L., & O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105-108. https://doi.org/10.1038/nature11118

Hortal, J., Cabeza, M., Diniz-Filho, J. A. F., von der Heyden, S., Stigall, A. L., & Yeo, D. C. J. (2022). Building a truly diverse biodiversity science. *Biodiversity*, *1*(1), 2. https://doi.org/10.1038/s44185-022-00003-1

Hop, K., M. Reid, J. Dieck, S. Lubinski, and S. Cooper. 2007. U.S. Geological Survey-National Park Service Vegetation Mapping Program: Waterton-Glacier International Peace Park. U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin, August 2007. 131 pp. + Appendixes A-L.

Hutchinson, E. G. (1957). *Cold Spring Harbour Symposia on Quantitative Biology* (Vol. 22). The Biological Laboratory.

Jennings, M. D., Faber-Langendoen, D., Loucks, O. L., Peet, R. K., & Roberts, D. (2009). Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological monographs*, *79*(2), 173-199. https://doi.org/10.1890/07-1804.1

Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta oecologica*, *31*(3), 361-369. https://doi.org/https://doi.org/10.1016/j.actao.2007.02.001

Jiménez-Alfaro, B., Suárez-Seoane, S., Chytrý, M., Hennekens, S. M., Willner, W., Hájek, M., Agrillo, E., Álvarez-Martínez, J. M., Bergamini, A., Brisse, H., Brunet, J., Casella, L., Dítě, D., Font, X., Gillet, F., Hajková, P., Jansen, F., Jandt, U., Kącki, Z., Lenoir, J., Rodwell, J. S., Schaminée, J. H. J., Sekulová, L., Šibík, J., Škvorc, Ž., Tsiripidis, I., & Sveriges, L. (2018). Modelling the distribution and compositional variation of plant communities at the continental scale. *Diversity & distributions*, *24*(7), 978-990. https://doi.org/10.1111/ddi.12736

Johnstone, J. F., Hollingsworth, T. N., Chapin, F. S., III, & Mack, M. C. (2010). Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. *Global change biology*, *16*(4), 1281-1295. https://doi.org/10.1111/j.1365-2486.2009.02051.x

Johnstone, J. F., Allen, C. D., Franklin, J. F., Frelich, L. E., Harvey, B. J., Higuera, P. E., Mack, M. C., Meentemeyer, R. K., Metz, M. R., Perry, G. L. W., Schoennagel, T., & Turner, M. G. (2016). Changing disturbance regimes, ecological memory, and forest resilience. *Frontiers in ecology and the environment*, *14*(7), 369-378. https://doi.org/10.1002/fee.1311

Kaky, E., Nolan, V., Alatawi, A., & Gilbert, F. (2020). A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: a case study with Egyptian medicinal plants. *Ecological informatics*, *60*, 101150. https://doi.org/10.1016/j.ecoinf.2020.101150

Keddy, P. A. (1992). Assembly and response rules: two goals for predictive community ecology. *Journal of vegetation science*, *3*(2), 157-164. https://doi.org/10.2307/3235676

Keeley, J. E., Pausas, J. G., Rundel, P. W., Bond, W. J., & Bradstock, R. A. (2011). Fire as an evolutionary pressure shaping plant traits. *Trends in plant science*, *16*(8), 406-411. https://doi.org/https://doi.org/10.1016/j.tplants.2011.04.002

Kessell, S. R. (1976). Gradient modeling: a new approach to fire modeling and wilderness resource management. *Environmental management*, *1*(1), 39-48. https://go.exlibris.link/18Fq5cpl

Lantz, T. C., Gergel, S. E., & Henry, G. H. R. (2010). Response of green alder (Alnus viridis subsp. fruticosa) patch dynamics and plant community composition to fire and regional temperature in north-western Canada. *Journal of biogeography*, *37*(8), 1597-1610. https://doi.org/10.1111/j.1365-2699.2010.02317.x

Lee-Yaw, J. A., McCune, J. L., Pironon, S., & Sheth, S. N. (2022). Species distribution models rarely predict the biology of real populations. *Ecography*, *2022*(6). https://doi.org/10.1111/ecog.05877

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (3rd English ed., Vol. 24). *Elsevier.* https://doi.org/10.1016/C2010-0-66470-4

Levesque, L. M. (2005). Investigating landscape change and ecological restoration: an integrated approach using historical ecology and GIS in Waterton Lakes National Park, Alberta. MSc Thesis, University of Victoria, Victoria, BC, Canada.

Limb, R. F., Hovick, T. J., Norland, J. E., & Volk, J. M. (2018). Grassland plant community spatial patterns driven by herbivory intensity. *Agriculture, ecosystems & environment*, *257*, 113-119. https://doi.org/10.1016/j.agee.2018.01.030

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global ecology and biogeography*, *17*(2), 145–151.

Longcore, T., Noujdina, N., & Dixon, P. J. (2018). Landscape modeling of the potential Natural Vegetation of Santa Catalina Island, California. *Western North American naturalist*, *78*(4), 617-632. https://doi.org/10.3398/064.078.0406

Lloren, J. I. (2021). Quantifying plant community change at Waterton Lakes National Park over the past 25 years. MSc Thesis, University of Lethbridge, Lethbridge, AB, Canada. https://go.exlibris.link/ndF9jXbN

Lyon, L. J., and P. F. Stickney. (1976). Early vegetational succession following large Northern Rocky Mountain wildfires, Proceedings Annual [14th] Tall Timbers Fire Ecology Conference and Intermountain Fire Research Council Fire & Land Management Symposium. Missoula, MN. Tall Timbers Research, Inc., Tallahassee, FL. p. 355-373.

Mateo, R. G., Croat, T. B., Felicísimo, Á. M., & Muñoz, J. (2010). Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity & distributions*, *16*(1), 84-94. https://doi.org/10.1111/j.1472-4642.2009.00617.x

McCune, J. L. (2016). Species distribution models predict rare species occurrences despite significant effects of landscape context. *The Journal of applied ecology*, *53*(6), 1871-1879. https://doi.org/10.1111/1365-2664.12702

McIntosh, R. P. (1995). H. A. Gleason's individualistic concept and theory of animal communities: a continuing controversy. *Biological reviews*, *70*(2), 317-357. https://doi.org/https://doi.org/10.1111/j.1469-185X.1995.tb01069.x

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058-1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Moudrý, V., Bazzichetto, M., Remelgado, R., Devillers, R., Lenoir, J., Mateo, R. G., Lembrechts, J. J., Sillero, N., Lecours, V., Cord, A. F., Barták, V., Balej, P., Rocchini, D., Torresani, M., Arenas-Castro, S., Man, M., Prajzlerová, D., Gdulová, K., Prošek, J., Marchetto, E., Zarzo-Arias, A., Gábor, L., Leroy, F., Martini, M., Malavasi, M., Cazzolla Gatti, R., Wild, J., & Šímová, P. (2024). Optimising occurrence data in species distribution models: sample size, positional

uncertainty, and sampling bias matter. *Ecography*, *2024*(12), e07294. https://doi.org/https://doi.org/10.1111/ecog.07294

Mueller-Dombois, D., & Ellenberg, H. (1974). *Aims and Methods of Vegetation Ecology*. https://doi.org/10.2307/213332

Murphy, S. J., & Smith, A. B. (2021). What can community ecologists learn from species distribution models? *Ecosphere*, *12*(12). https://doi.org/10.1002/ecs2.3864

Musk, D., Lloren, J. I., & McCune, J. L. (2024). Tree seedling regeneration in Canada's southern Rocky Mountains: contrasting recently burned and unburned areas. *Northwest science*, *97*(4), 290-305. https://doi.org/10.3955/046.097.0405

Myers, J. A., & Harms, K. E. (2009). Seed arrival, ecological filters, and plant species richness: a meta-analysis. *Ecology letters*, *12*(11), 1250-1260. https://doi.org/10.1111/j.1461-0248.2009.01373.x

NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. (2019). *ASTER Global Digital Elevation Model V003* [Data set]. NASA Land Processes Distributed Active Archive Center. https://doi.org/10.5067/ASTER/ASTGTM.003 Date Accessed: 2024-12-15

Natural Regions Committee. (2006). Natural regions and subregions of Alberta. Compiled by D. J. Downing and W. W. Pettepiece. Government of Alberta, Edmonton, Alberta.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., El Din, S. B., & Gilbert, F. (2010). Testing the accuracy of species distribution models using species records from a new field survey. *Oikos*, *119*(8), 1326-1334. https://doi.org/10.1111/j.1600-0706.2009.18295.x

Nix, H. A., McMahon, J. & Mackenzie, D. (1977). Potential areas of production and the future of pigeon pea and other grain legumes in Australia. In E.S. Wallis & C. Whiteman (Eds.), *The potential for pigeon pea in Australia: Proceedings of Pigeon Pea (Cajanus cajan (L.) Millsp.) Field Day* (pp. 5/1–5/12). University of Queensland, Queensland, Australia.

Nix, H. A. (1986) A biogeographic analysis of Australian elapid snakes. In R. Longmore (Ed.), *Atlas of elapid snakes of Australia: Australian flora and fauna*, (Vol. 7, pp. 4–15). Bureau of Flora and Fauna, Canberra.

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araujo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., Husby, M., Kålås, J. A., Lehikoinen, A., Luoto, M., Mod, H. K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White,

M., Zimmermann, N. E., Gravel, D., Ovaskainen, O., & Sveriges, L. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological monographs*, *89*(3), 1-24. https://doi.org/10.1002/ecm.1370

Parks Canada. (2022). Waterton Lakes National Park of Canada Management Plan. [Online]. Available from: https://parks.canada.ca/pn-np/ab/waterton/info/index/directeur-management/plan-2022

Parks Canada. (2023). Waterton-Glacier International Peace Park. [Online]. Available from: https://parks.canada.ca/culture/spm-whs/sites-canada/sec02l

Pearson, R. G., Dawson, T. P., & Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, *27*(3), 285-298. http://www.jstor.org.uleth.idm.oclc.org/stable/3683610

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of biogeography*, *34*, 102–117.

Pebesma, E., & Bivand, R. (2023). Spatial data science: with applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, *190*(3), 231-259. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications*, *19*(1), 181-197. https://doi.org/10.1890/07-2153.1

Porensky, L. M., Derner, J. D., & Pellatz, D. W. (2018). Plant community responses to historical wildfire in a shrubland–grassland ecotone reveal hybrid disturbance response. *Ecosphere*, *9*(8). https://doi.org/10.1002/ecs2.2363

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C. F., Vittoz, P., & Guisan, A. (2013). The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global ecology and biogeography*, *22*(1), 52-63. https://doi.org/https://doi.org/10.1111/j.1466-8238.2012.00790.x

Raes, N., & Aguirre-Gutiérrez, J. (2018). A modeling framework to estimate and project species distributions in space and time. In C. Hoorn, A. Perrigo, & A. Antonelli (Eds.), *Mountains, climate and biodiversity* (pp. 309–320). Wiley-Blackwell.

Radomski, T., Beamer, D., Babineau, A., Wilson, C., Pechmann, J., & Kozak, K. H. (2022). Finding what you don't know: testing SDM methods for poorly known species. *Diversity & distributions*, *28*(9), 1769-1780. https://doi.org/10.1111/ddi.13536

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/.

Ricklefs, R. E. (2008). Disintegration of the ecological community. *The American naturalist*, *172*(6), 741-750. https://doi.org/10.1086/593002

Romme, W. H., Whitby, T. G., Tinker, D. B., & Turner, M. G. (2016). Deterministic and stochastic processes lead to divergence in plant communities 25 years after the 1988 Yellowstone fires. *Ecological monographs*, *86*(3), 327-351. https://doi.org/10.1002/ecm.1220

Ruggirello, M. J., Bustamante, G., Rodriguez, P., Cruz-Alonso, V., & Soler, R. (2023). Post-fire forest recovery at high latitudes: tree regeneration dominated by fire-adapted, early-seral species increases with latitude. *Annals of forest science.*, *80*(1), 47-47. https://doi.org/10.1186/s13595-023-01213-8

Running, S., & Zhao, M. (2021). *MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061* [Data set]. NASA Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD17A3HGF.061 Date Accessed: 2024-12-15

Sheth, S. N., Jiménez, I., & Angert, A. L. (2014). Identifying the paths leading to variation in geographical range size in western North American monkeyflowers. *Journal of biogeography*, *41*(12), 2344-2356. https://doi.org/10.1111/jbi.12378

Stockdale, C. A., Macdonald, S. E., & Higgs, E. (2019). Forest closure and encroachment at the grassland interface: a century-scale analysis using oblique repeat photography. *Ecosphere*, *10*(6), e02774. https://doi.org/https://doi.org/10.1002/ecs2.2774

Syphard, A. D., & Franklin, J. (2009). Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography*, *32*(6), 907-918. https://doi.org/10.1111/j.1600-0587.2009.05883.x

Thuiller, W., Brotons, L., Araújo, M. B., & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, *27*(2), 165-172. https://doi.org/10.1111/j.0906-7590.2004.03673.x

Tikhonov, G., Duan, L., Abrego, N., Newell, G., White, M., Dunson, D., & Ovaskainen, O. (2020). Computationally efficient joint species distribution modeling of big spatial data. *Ecology*, *101*(2), e02929. https://doi.org/https://doi.org/10.1002/ecy.2929

Tobler, M. W., Kéry, M., Hui, F. K. C., Guillera-Arroita, G., Knaus, P., & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, *100*(8), 1-14. https://doi.org/10.1002/ecy.2754

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*(2), 99-114. https://doi.org/10.2307/3001913

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological monographs*, *92*(1). https://doi.org/10.1002/ecm.1486

Velazco, S. J. E., Rose, M. B., Andrade, A. F. A., Minoli, I., & Franklin, J. (2022). flexsdm: An R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods in ecology and evolution*, *13*(8) 1661-1669. https://doi.org/10.1111/2041-210X.13874

Vild, O., & Douglas Rotherham, I. (2021). Long-term exclosure of sheep-grazing from an ancient wood: Vegetation change after a sixty-year experiment. *Applied vegetation science, 24*(1), e12543. https://doi.org/https://doi.org/10.1111/avsc.12543

Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America*. PloS one, 11*(6), e0156720.

Wang, Y., Wen, S., Farnon Ellwood, M. D., Miller, A. D., & Chu, C. (2018). Temporal effects of disturbance on community composition in simulated stage-structured plant communities. *Ecology and evolution*, *8*(1), 120-127. https://doi.org/10.1002/ece3.3660

Wang, L., & Jackson, D. A. (2023). Effects of sample size, data quality, and species response in environmental space on modeling species distributions. *Landscape ecology, 38*(12), 4009-4031. https://doi.org/10.1007/s10980-023-01771-2

Whittaker, R. H. (1962). *Classification of natural communities* (Vol. 28, No. 1). Botanical Garden. https://go.exlibris.link/Bbtvwwsz

Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon, 21*(3), 213-251. doi:10.2307/1218190

Wieczorkowski, J. D., & Lehmann, C. E. R. (2022). Encroachment diminishes herbaceous plant diversity in grassy ecosystems worldwide. *Global change biology*, *28*(18), 5532-5546. https://doi.org/10.1111/gcb.16300

Wiersma, Y. F., & Urban, D. L. (2005). Beta diversity and nature reserve system design in the Yukon, Canada. *Conservation biology*, *19*(4), 1262-1272. https://doi.org/10.1111/j.1523-1739.2005.00099.x

Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and distributions*, *15*(4), 565-576. https://doi.org/https://doi.org/10.1111/j.1472-4642.2009.00567.x

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & Group, N. P. S. D. W. (2008). Effects of sample size on the performance of species distribution models. *Diversity and distributions*, *14*(5), 763-773. https://doi.org/https://doi.org/10.1111/j.1472-4642.2008.00482.x

Zangiabadi, S., Zaremaivan, H., Brotons, L., Mostafavi, H., & Ranjbar, H. (2021). Using climatic variables alone overestimate climate change impacts on predicting distribution of an endemic species. *PloS one*, *16*(9), e0256918. https://doi.org/10.1371/journal.pone.0256918

Zimmermann, N. E., & Kienast, F. (1999). Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *Journal of vegetation science*, *10*(4), 469-482. https://doi.org/10.2307/3237182

Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of biogeography*, *47*(1), 101-113. https://doi.org/10.1111/jbi.13608

Appendix I: Supplemental Tables and Figures

Table S1. All characteristic species of each of the seven special plant communities with >= five presences, as defined by the Ecological Land Classification of Waterton Lakes National Park and identified through Indicator Species Analysis (ISA).

Community Code	Species
C60	*Actaea rubra, Amelanchier alnifolia, Geranium richardsonii, Heracleum maximum, Lathyrus ochroleucus, Osmorhiza occidentalis, Populus tremuloides, Prunus virginiana, Urtica dioica, Vicia americana, Viola canadensis, Sanicula marilandica, Symphoricarpos occidentalis, Symphoricarpos albus, Eucephalus engelmannii, Prosartes hookeri, Maianthemum racemosum, Thalictrum occidentale, Bromus vulgaris, Elymus glaucus, Bromus inermis, Veraturm viride, Moehringia lateriflora, Calamagrostis canadensis, Streptopus amplexifolius*
C61	*Populus tremuloides, Vicia americana, Viola canadensis, Prunus virginiana, Rosa woodsii, Geranium richardsonii, Amelanchier alnifolia, Osmorhiza occidentalis, Rubus parviflorus, Cornus sericea, Elymus glaucus, Urtica dioica, Symphyotrichum ciliolatum, Berberis repens, Maianthemum stellatum, Symphoricarpos occidentalis, Lathyrus ochroleucus, Sanicula marilandica, Prosartes hookeri*
C62	*Populus tremuloides, Symphoricarpos occidentalis, Prosartes trachycarpum, Berberis repens, Thalictrum occidentale, Eurybia conspicua, Actaea rubra, Rosa acicularis, Bromus vulgaris, Amelanchier alnifolia, Calamagrostis canadensis, Prunus virginiana, Chamaenerion angustifolium*
C78	*Equisetum arvense, Galium triflorum, Habenaria hyperborea, Lonicera involucrata, Picea engelmannii, Packera pseudaurea, Angelica arguta, Geranium richardsonii, Heracleum maximum, Osmorhiza depauperata, Geum macrophyllum, Geum rivale, Rosa acicularis, Petasites sagittatus, Mitella nuda, Taraxacum officinale, Graphephorum cernuum, Lycopodium annotinum, Rubus pubescens, Cornus sericea, Pyrola asarifolia, Bromus vulgaris, Populus trichocarpa, Allium textile, Geum allepicum, Symphyotrichum ciliolatum, Ribes lacustre, Calamagrostis canadensis, Athyrium filix.femina, Maianthemum racemosum, Symphyotrichum subspicatum, Linnaea borealis, Senecio triangularis, Bromus ciliatus*
C79	*Eurybia conspicua, Berberis repens, Calamagrostis rubescens, Clintonia uniflora, Pinus contorta, Maianthemum racemosum, Linnaea borealis, Chamaenerion angustifolium, Symphyotrichum ciliolatum, Chimaphila umbellata, Pyrola asarifolia, Spiraea betulifolia, Cornus canadensis, Lathyrus ochroleucus, Rosa acicularis, Vicia americana, Moehringia lateriflora, Hieracium albiflorum, Vaccinium myrtillus, Alnus crispa, Campanula alaskana*
H25	*Anemone multifida, Pulsatilla nutalliana, Antennaria parvifolia, Balsamorhiza sagittata, Calochortus apiculatus, Cerastium arvense, Danthonia parryi, Primula conjugens, Erigeron caespitosus, Festuca idahoensis, Festuca scabrella, Gaillardia aristata, Galium boreale, Gentianella amarella, Geum triflorum, Hedysarum sulphurescens, Heuchera parvifolia, Koeleria macrantha, Lithospermum ruderale, Lomatium macrocarpum, Lupinus sericeus, Sabulina rubella, Monarda fistulosa, Oxytropis sericea, Oxytropis splendens, Dasiphora fruticosa, Potentilla gracilis, Potentilla pensylvanica, Rhinanthus minor, Sisyrinchium montanum, Eriocoma lemmonii, Anticlea elegans, Elymus trachycaulus, Hypericum perforatum, Danthonia californica, Achillea millefolium, Castilleja occidentalis, Bupleurum americanum, Solidago spathulata, Penstemon confertus, Castilleja lutescens, Geocaulon lividum, Agoseris glauca, Campanula alaskana, Sedum lanceolatum, Silene parryi, Tragopogon dubius, Rosa woodsii, Heterotheca villosa, Carex raynoldsii, Allium cernuum, Geranium viscosissimum, Lilium philadelphicum, Artemisia michauxiana, Arctostaphylos uva.ursi, Lomatium triternatum, Veronica wyomingensis*

Community Code	Species
H27	*Achillea millefolium, Pseudoroegneria spicata, Anemone multifida, Festuca idahoensis, Festuca scabrella, Galium boreale, Lithospermum ruderale, Penstemon confertus, Potentilla gracilis, Campanula alaskana, Lupinus sericeus, Eriocoma lemmonii, Erigeron caespitosus, Allium textile, Antennaria parvifolia, Geranium viscosissimum, Elymus trachycaulus, Arnica sororia, Monarda fistulosa, Phleum pratense, Castilleja lutescens, Comandra umbellata, Linum lewisii, Gaillardia aristata, Koeleria macrantha, Agoseris glauca, Potentilla glandulosa, Fragaria virginiana, Hedysarum sulphurescens, Balsamorhiza sagittata, Geum triflorum, Rosa woodsii, Tragopogon dubius, Astragalus laxmanii, Cirsium hookerianum, Solidago missouriensis, Dasiphora fruticosa, Calochortus apiculatus, Allium cernuum, Hypericum perforatum, Anticlea elegans, Apocynum androsaemifolium, Juncus balticus, Poa cusickii, Poa interior*

Table S2. Pearsons Correlation coefficients for pairwise relationships among all environmental predictor variables* considered for inclusion in model building. We removed highly correlated variables prior to model building.

Variable	Aspe	CLY	CMI	DD0	DD18	DD5	eFFP	EMT	EREF	MAP	MAR	MAT	MSP	MWMT	NDVI	NFFD	NPP	PAS	PH	PPTsm	RADsm	RH	Slope	SLT	SND	Tavesm	Tmaxsm	Tmaxwt	Tminsm	Tminwt	TWI
Aspect	1.00	-0.01	-0.02	-0.02	-0.02	0.02	0.03	0.02	0.02	0.00	-0.01	0.02	-0.01	0.02	-0.10	0.03	0.02	0.00	0.03	-0.01	-0.01	-0.02	-0.04	-0.04	0.03	0.02	0.02	0.03	0.03	0.02	0.01
CLY		1.00	-0.52	-0.52	-0.54	0.56	0.41	0.41	0.55	-0.45	-0.16	0.54	-0.45	0.55	0.33	0.51	-0.17	-0.47	0.69	-0.43	-0.19	-0.61	-0.34	-0.08	-0.17	0.55	0.56	0.56	0.49	0.39	0.03
CMI			1.00	0.96	0.97	-0.98	-0.84	-0.89	-0.96	0.94	0.57	-0.97	0.97	-0.98	-0.40	-0.96	-0.38	0.97	-0.35	0.97	0.66	0.96	0.63	0.40	0.39	-0.98	-0.98	-0.96	-0.95	-0.87	-0.42
DD0				1.00	1.00	-0.99	-0.93	-0.97	-0.98	0.82	0.53	-1.00	0.88	-0.99	-0.34	-0.99	-0.32	0.87	-0.34	0.88	0.61	0.94	0.56	0.39	0.36	-0.99	-0.99	-0.99	-0.98	-0.95	-0.36
DD18					1.00	-1.00	-0.91	-0.95	-0.98	0.85	0.52	-1.00	0.90	-1.00	-0.35	-0.99	-0.33	0.89	-0.36	0.90	0.61	0.96	0.58	0.41	0.37	-1.00	-1.00	-1.00	-1.00	-0.99	-0.93	-0.37
DD5						1.00	0.89	0.92	0.98	-0.87	-0.50	1.00	-0.91	1.00	0.36	0.99	0.33	-0.91	0.38	-0.90	-0.59	-0.97	-0.60	-0.42	-0.37	1.00	0.99	0.98	0.98	0.90	0.38
eFFP							1.00	0.97	0.88	-0.63	-0.50	0.91	-0.73	0.89	0.24	0.94	0.28	-0.70	0.27	-0.74	-0.58	-0.79	-0.44	-0.36	-0.31	0.90	0.88	0.91	0.92	0.98	0.31
EMT								1.00	0.92	-0.71	-0.55	0.95	-0.79	0.93	0.27	0.97	0.32	-0.77	0.27	-0.80	-0.63	-0.83	-0.46	-0.38	-0.32	0.94	0.92	0.94	0.96	1.00	0.33
EREF									1.00	-0.85	-0.49	0.98	-0.89	0.98	0.37	0.97	0.31	-0.90	0.38	-0.89	-0.59	-0.96	-0.56	-0.40	-0.33	0.99	0.96	0.98	0.96	0.90	0.35
MAP										1.00	0.57	-0.85	0.99	-0.87	-0.47	0.99	-0.42	0.99	-0.27	0.98	0.66	0.89	0.67	0.33	0.40	-0.86	-0.87	-0.84	-0.82	-0.68	-0.45
MAR											1.00	-0.51	0.59	-0.51	-0.29	-0.53	-0.29	0.59	-0.08	0.62	0.98	0.44	0.36	0.16	0.24	-0.51	-0.51	-0.51	-0.52	-0.55	-0.29
MAT												1.00	-0.90	1.00	0.35	0.99	0.33	-0.89	0.36	-0.90	-0.61	-0.96	-0.58	-0.41	-0.37	1.00	1.00	1.00	0.99	0.93	0.37
MSP													1.00	-0.91	-0.44	-0.88	-0.42	0.99	-0.27	0.99	0.68	0.91	0.67	0.35	0.41	-0.91	-0.91	-0.89	-0.88	-0.76	-0.46
MWMT														1.00	0.36	0.99	0.34	-0.91	0.37	-0.91	-0.60	-0.96	-0.59	-0.42	-0.37	1.00	1.00	0.99	0.98	0.91	0.38
NDVI															1.00	0.33	-0.04	-0.47	0.11	-0.43	-0.32	-0.40	-0.38	0.21	-0.04	0.36	0.37	0.35	0.33	0.25	0.07
NFFD																1.00	0.34	-0.87	0.33	-0.88	-0.62	-0.92	-0.56	-0.41	-0.37	0.99	0.98	0.98	0.99	0.95	0.38

Table S2. Continued.

Variable	NPP	PAS	PH	PPTsm	RADsm	RH	Slope	SLT	SND	Tavesm	Tmaxsm	Tmaxwt	Tminsm	Tminwt	TWI
Aspect															
CLY															
CMI															
DD0															
DD18															
DD5															
eFFP															
EMT															
EREF															
MAP															
MAR															
MAT															
MSP															
MWMT															
NDVI															
NFFD															
NPP	1.00														
PAS	-0.41	1.00													
PH	-0.33	-0.29	1.00												
PPTsm	-0.43	0.98	-0.28	1.00											
RADsm	-0.34	0.68	-0.10	0.71	1.00										
RH	-0.30	0.92	-0.41	0.89	0.53	1.00									
Slope	-0.38	0.66	-0.07	0.65	0.41	0.62	1.00								
SLT	-0.47	0.34	-0.06	0.36	0.21	0.39	0.27	1.00							
SND	-0.53	0.39	0.23	0.40	0.27	0.36	0.43	0.22	1.00						
Tavesm	0.33	-0.91	0.37	-0.90	-0.61	-0.96	-0.59	-0.41	-0.37	1.00					
Tmaxsm	0.33	-0.91	0.39	-0.91	-0.60	-0.97	-0.59	-0.41	-0.36	1.00	1.00				
Tmaxwt	0.31	-0.89	0.37	-0.88	-0.59	-0.96	-0.58	-0.40	-0.36	0.99	0.99	1.00			
Tminsm	0.34	-0.86	0.31	-0.88	-0.61	-0.91	-0.56	-0.40	-0.38	0.98	0.97	0.97	1.00		
Tminwt	0.31	-0.74	0.27	-0.79	-0.63	-0.80	-0.44	-0.38	-0.31	0.92	0.90	0.92	0.94	1.00	
TWI	0.52	-0.44	-0.13	-0.46	-0.33	-0.36	-0.58	-0.40	-0.45	0.38	0.37	0.36	0.38	0.32	1.00

Table S2. Continued.
* Predictor abbreviations in Table S2 are as follows: MSP – May to September precipitation, CLY – soil clay content, SDN – soil sand content, SLT – soil silt content, DD0 – degree-days below 0°C, DD5 – degree-days above 5°C, DD18 – degree-days below 18°C, NDVI – normalized difference vegetation index, EREF – Hargreaves reference evaporation, RADsm – summer solar radiation, TWI – topographic wetness index, PH – soil pH, Tavesm – summer mean temperature, Tmaxsm – summer mean maximum temperature, Tmaxwt – winter mean maximum temperature, Tminsm – summer mean minimum temperature, Tminwt – winter mean minimum temperature, PAS – annual precipitation as snow, Aspect – slope aspect, Slope – degree slope, MAR – mean annual solar radiation, CMI – Hogg's climate moisture index, eFFP – day of the year the frost-free period ends, EMT – extreme minimum temperature over 30 years, PPTsm – summer precipitation, NFFD – number of frost-free days, MWMT – mean temperature of the warmest month, RH – relative humidity, MAT – mean annual temperature, MAP – mean annual precipitation, NPP – net primary productivity. For full details on climate variables see Wang et al. 2016.

Table S3. Model performance metrics for seven vegetation communities modelled with community distribution models (CDMs) made with GLM, MaxEnt, and RF algorithms. Table reports the number of presences, absences, and background points used in model building, along with the mean AUC and TSS across all model partitions determined through internal cross-validation. True absences were used for GLM and RF models, whereas background points were used in MaxEnt models. MaxEnt parameters for regularization (Regmult) and feature class combinations (Classes) are also shown, as these varied between models.

Community Code	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes*	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
C60	20	349	2000	0.1	2	0.97	0.91	0.97	0.93	0.99	0.95
C61	7	362	2000	0.55	4	0.94	0.93	0.96	0.95	1.00	1.00
C62	5	364	2000	0.1	4	0.78	0.49	0.91	0.91	0.98	0.98
C78	6	363	2000	1	2	0.89	0.69	0.86	0.85	0.86	0.72
C79	5	364	2000	2.8	4	0.78	0.49	0.91	0.91	0.78	0.56
H25	20	349	2000	3	2	0.93	0.78	0.92	0.85	0.99	0.92
H27	11	358	2000	2.45	1	0.91	0.78	0.68	0.60	0.96	0.87

* Feature class combinations used in MaxEnt models coded from 1 to 5, with each successive level including all previous feature types: 1 = Linear, 2 = +Quadratic, 3 = +Hinge, 4 = +Product and 5 = +Threshold

Table S4. Model performance metrics for all component species of community C60 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Actaea rubra	41	289	2000	0.1	5	0.78	0.25	0.71	0.46	0.96	0.77
Amelanchier alnifolia	111	219	2000	2.4	4	0.85	0.54	0.79	0.54	0.98	0.85
Bromus inermis	38	292	2000	1.65	1	0.88	0.48	0.88	0.73	0.95	0.75
Bromus vulgaris	30	300	2000	3	1	0.82	0.43	0.77	0.60	0.96	0.76
Calamagrostis canadensis	24	306	2000	2.35	2	0.87	0.50	0.84	0.66	0.98	0.87
Elymus glaucus	18	312	2000	0.1	3	0.81	0.33	0.84	0.67	0.92	0.61
Eucephalus engelmannii	43	287	2000	3	4	0.79	0.32	0.73	0.52	0.96	0.82
Geranium richardsonii	28	302	2000	0.15	4	0.89	0.45	0.86	0.74	0.98	0.85
Heracleum maximum	59	271	2000	2.95	2	0.79	0.37	0.75	0.51	0.98	0.84
Lathyrus ochroleucus	45	285	2000	0.15	4	0.89	0.53	0.86	0.68	0.99	0.89
Maianthemum racemosum	75	255	2000	0.1	2	0.86	0.59	0.79	0.57	0.99	0.91
Moehringia lateriflora	18	312	2000	2	4	0.76	0.29	0.82	0.69	0.93	0.67
Osmorhiza occidentalis	52	278	2000	1.1	5	0.84	0.41	0.77	0.55	0.99	0.91
Populus tremuloides	53	277	2000	3	4	0.89	0.61	0.81	0.61	0.98	0.87
Prosartes hookeri	7	323	2000	0.15	4	0.96	-0.04	0.83	0.80	0.98	0.98
Prunus virginiana	31	299	2000	2.45	3	0.89	0.51	0.88	0.77	0.98	0.88
Sanicula marilandica	14	316	2000	0.25	4	0.91	0.48	0.93	0.85	0.98	0.92
Streptopus amplexifolius	15	315	2000	3	4	0.85	0.36	0.70	0.52	0.95	0.86
Symphoricarpos albus	26	304	2000	2.25	1	0.86	0.52	0.79	0.61	0.98	0.89
Symphoricarpos occidentalis	51	279	2000	0.3	4	0.90	0.61	0.86	0.69	0.98	0.85
Thalictrum occidentale	139	191	2000	0.85	5	0.83	0.45	0.72	0.41	0.98	0.88
Urtica dioica	17	313	2000	2.9	1	0.89	0.43	0.88	0.78	0.93	0.75
Veratrum viride	74	256	2000	1.6	5	0.75	0.32	0.72	0.47	0.98	0.84
Vicia americana	49	281	2000	0.3	4	0.93	0.69	0.90	0.74	0.99	0.93
Viola canadensis	36	294	2000	0.15	5	0.89	0.56	0.84	0.67	0.99	0.91

Table S5. Model performance metrics for all component species of community C61 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Amelanchier alnifolia	111	219	2000	2.4	4	0.85	0.54	0.79	0.54	0.98	0.85
Berberis repens	63	267	2000	3	1	0.92	0.71	0.83	0.62	0.99	0.92
Cornus sericea	17	313	2000	0.1	5	0.89	0.41	0.87	0.75	0.95	0.78
Elymus glaucus	18	312	2000	0.1	3	0.81	0.33	0.84	0.67	0.92	0.61
Geranium richardsonii	28	302	2000	0.15	4	0.89	0.45	0.86	0.74	0.98	0.85
Lathyrus ochroleucus	45	285	2000	0.15	4	0.89	0.53	0.86	0.68	0.99	0.89
Maianthemum stellata	39	291	2000	3	2	0.79	0.28	0.77	0.52	0.97	0.85
Osmorhiza occidentalis	52	278	2000	1.1	5	0.84	0.41	0.77	0.55	0.99	0.91
Populus tremuloides	53	277	2000	3	4	0.89	0.61	0.81	0.61	0.98	0.87
Prosartes hookeri	7	323	2000	0.15	4	0.96	-0.04	0.83	0.80	0.98	0.98
Prunus virginiana	31	299	2000	2.45	3	0.89	0.51	0.88	0.77	0.98	0.88
Rosa woodsii	49	281	2000	1.7	5	0.89	0.57	0.85	0.69	0.98	0.85
Rubus parviflorus	85	245	2000	2.15	5	0.91	0.62	0.77	0.52	0.99	0.90
Sanicula marilandica	14	316	2000	0.25	4	0.91	0.48	0.93	0.85	0.98	0.92
Symphoricarpos occidentalis	51	279	2000	0.3	4	0.90	0.61	0.86	0.69	0.98	0.85
Symphyotrichum ciliolatum	49	281	2000	3	3	0.78	0.21	0.79	0.57	0.98	0.89
Urtica dioica	17	313	2000	2.9	1	0.89	0.43	0.88	0.78	0.93	0.75
Vicia americana	49	281	2000	0.3	4	0.93	0.69	0.90	0.74	0.99	0.93
Viola canadensis	36	294	2000	0.15	5	0.89	0.56	0.84	0.67	0.99	0.91

Table S6. Model performance metrics for all component species of community C62 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Actaea rubra	41	289	2000	0.1	5	0.78	0.25	0.71	0.46	0.96	0.77
Amelanchier alnifolia	111	219	2000	2.4	4	0.85	0.54	0.79	0.54	0.98	0.85
Berberis repens	63	267	2000	3	1	0.92	0.71	0.83	0.62	0.99	0.92
Bromus vulgaris	30	300	2000	3	1	0.82	0.43	0.77	0.60	0.96	0.76
Calamagrostis canadensis	24	306	2000	2.35	2	0.87	0.50	0.84	0.66	0.98	0.87
Chamaenerion angustifolium	119	211	2000	0.45	5	0.72	0.26	0.70	0.42	0.98	0.88
Eurybia conspicua	82	248	2000	1.1	5	0.87	0.53	0.80	0.57	0.99	0.92
Populus tremuloides	53	277	2000	3	4	0.89	0.61	0.81	0.61	0.98	0.87
Prosartes trachycarpa	42	288	2000	0.1	3	0.88	0.49	0.80	0.61	0.98	0.88
Prunus virginiana	31	299	2000	2.45	3	0.89	0.51	0.88	0.77	0.98	0.88
Rosa acicularis	53	277	2000	0.1	4	0.79	0.25	0.78	0.52	0.96	0.77
Symphoricarpos occidentalis	51	279	2000	0.3	4	0.90	0.61	0.86	0.69	0.98	0.85
Thalictrum occidentale	139	191	2000	0.85	5	0.83	0.45	0.72	0.41	0.98	0.88

Table S7. Model performance metrics for all component species of community C78 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Allium textile	5	325	2000	0.1	2	0.89	0.00	0.87	0.87	0.99	0.98
Angelica arguta	6	324	2000	1.1	1	0.82	-0.07	0.94	0.93	0.86	0.76
Athyrium filix.femina	5	325	2000	0.1	1	0.84	0.00	0.90	0.90	0.90	0.83
Bromus ciliatus	9	321	2000	0.2	3	0.92	0.36	0.90	0.84	0.98	0.92
Bromus vulgaris	30	300	2000	3	1	0.82	0.43	0.77	0.60	0.96	0.76
Calamagrostis canadensis	24	306	2000	2.35	2	0.87	0.50	0.84	0.66	0.98	0.87
Cornus sericea	17	313	2000	0.1	5	0.89	0.41	0.87	0.75	0.95	0.78
Equisetum arvense	27	303	2000	2.3	3	0.83	0.25	0.80	0.64	0.96	0.77
Galium triflorum	39	291	2000	2.1	4	0.84	0.47	0.74	0.57	0.97	0.82
Geranium richardsonii	28	302	2000	0.15	4	0.89	0.45	0.86	0.74	0.98	0.85
Geum allepicum	7	323	2000	0.1	2	0.87	-0.08	0.74	0.72	0.89	0.80
Geum macrophyllum	7	323	2000	0.2	5	0.86	-0.04	0.84	0.80	0.97	0.96
Geum rivale	8	322	2000	0.9	3	0.92	0.41	0.87	0.84	0.96	0.87
Graphephorum cernuum	5	325	2000	0.1	5	0.99	0.00	0.52	0.51	0.97	0.96
Habenaria hyperborea	10	320	2000	2.35	3	0.81	0.16	0.90	0.84	0.91	0.68
Heracleum maximum	59	271	2000	2.95	2	0.79	0.37	0.75	0.51	0.98	0.84
Linnaea borealis	27	303	2000	2	4	0.88	0.46	0.86	0.71	0.98	0.83
Lonicera involucrata	33	297	2000	1	4	0.79	0.26	0.72	0.52	0.98	0.84
Lycopodium annotinum	6	324	2000	0.1	1	0.95	-0.04	0.78	0.78	0.89	0.79
Maianthemum racemosum	75	255	2000	0.1	2	0.86	0.59	0.79	0.57	0.99	0.91
Mitella nuda	5	325	2000	0.45	2	0.89	0.00	0.86	0.86	0.93	0.86
Osmorhiza depauperata	36	294	2000	0.55	5	0.82	0.31	0.76	0.52	0.97	0.84
Packera pseudaurea	11	319	2000	0.1	2	0.83	0.25	0.76	0.67	0.92	0.79
Petasites sagittatus	5	325	2000	0.6	2	0.89	0.00	0.79	0.79	0.99	0.99
Picea engelmannii	119	211	2000	2.05	4	0.85	0.44	0.69	0.39	0.98	0.89
Populus trichocarpa	10	320	2000	0.9	1	0.87	0.28	0.87	0.82	0.95	0.80

Table S7. Continued

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Pyrola asarifolia	22	308	2000	2.95	2	0.83	0.37	0.77	0.61	0.95	0.76
Ribes lacustre	55	275	2000	0.15	5	0.79	0.34	0.73	0.45	0.99	0.87
Rosa acicularis	53	277	2000	0.1	4	0.79	0.25	0.78	0.52	0.96	0.77
Rubus pubescens	7	323	2000	0.1	3	0.89	-0.10	0.89	0.87	0.88	0.80
Senecio triangularis	32	298	2000	2.55	3	0.86	0.37	0.78	0.61	0.97	0.85
Symphyotrichum ciliolatum	49	281	2000	3	3	0.78	0.21	0.79	0.57	0.98	0.89
Symphyotrichum subspicatum	9	321	2000	0.15	1	0.88	0.33	0.59	0.52	0.93	0.73
Taraxacum officinale	34	296	2000	0.15	4	0.82	0.41	0.79	0.60	0.94	0.78

Table S8. Model performance metrics for all component species of community C79 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Alnus crispa	35	295	2000	3	3	0.86	0.52	0.75	0.59	0.98	0.89
Berberis repens	63	267	2000	3	1	0.92	0.71	0.83	0.62	0.99	0.92
Calamagrostis rubescens	46	284	2000	2.9	4	0.85	0.52	0.81	0.65	0.99	0.90
Campanula alaskana	59	271	2000	0.95	4	0.78	0.27	0.71	0.43	0.97	0.82
Chamaenerion angustifolium	119	211	2000	0.45	5	0.72	0.26	0.70	0.42	0.98	0.88
Chimaphila umbellata	36	294	2000	0.65	5	0.89	0.48	0.80	0.63	0.98	0.88
Clintonia uniflora	53	277	2000	1.8	3	0.97	0.76	0.85	0.68	0.99	0.92
Cornus canadensis	10	320	2000	2.3	1	0.87	0.27	0.94	0.92	0.97	0.95
Eurybia conspicua	82	248	2000	1.1	5	0.87	0.53	0.80	0.57	0.99	0.92
Hieracium albiflorum	24	306	2000	2.6	4	0.83	0.36	0.69	0.54	0.97	0.89
Lathyrus ochroleucus	45	285	2000	0.15	4	0.89	0.53	0.86	0.68	0.99	0.89
Linnaea borealis	27	303	2000	2	4	0.88	0.46	0.86	0.71	0.98	0.83
Maianthemum racemosa	75	255	2000	0.1	2	0.86	0.59	0.79	0.57	0.99	0.91
Moehringia lateriflora	18	312	2000	2	4	0.76	0.29	0.82	0.69	0.93	0.67
Pinus contorta	76	254	2000	3	4	0.87	0.47	0.79	0.56	0.99	0.89
Pyrola asarifolia	22	308	2000	2.95	2	0.83	0.37	0.77	0.61	0.95	0.76
Rosa acicularis	53	277	2000	0.1	4	0.79	0.25	0.78	0.52	0.96	0.77
Spiraea lucida	118	212	2000	2.65	3	0.89	0.60	0.76	0.50	0.98	0.87
Symphyotrichum ciliolatum	49	281	2000	3	3	0.78	0.21	0.79	0.57	0.98	0.89
Vaccinium myrtillus	75	255	2000	0.8	4	0.85	0.49	0.78	0.53	0.98	0.88
Vicia americana	49	281	2000	0.3	4	0.93	0.69	0.90	0.74	0.99	0.93

Table S9. Model performance metrics for all component species of community H25 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Achillea millefolium	88	242	2000	2.9	1	0.84	0.46	0.80	0.53	0.96	0.84
Agoseris glauca	33	297	2000	2.05	1	0.78	0.33	0.80	0.63	0.96	0.78
Allium cernuum	41	289	2000	0.5	1	0.86	0.42	0.79	0.60	0.96	0.82
Anemone multifida	53	277	2000	0.25	2	0.87	0.45	0.81	0.58	0.97	0.84
Antennaria parvifolia	33	297	2000	2.1	1	0.83	0.29	0.82	0.63	0.97	0.84
Anticlea elegans	39	291	2000	0.75	4	0.84	0.43	0.78	0.59	0.93	0.70
Arctostaphylos uva.ursi	44	286	2000	2.95	2	0.81	0.37	0.78	0.55	0.96	0.78
Artemisia michauxiana	6	324	2000	0.1	5	0.79	-0.07	0.68	0.63	0.82	0.64
Balsamorhiza sagittata	15	315	2000	1.65	5	0.87	0.37	0.82	0.71	0.94	0.86
Bupleurum americanum	11	319	2000	0.1	4	0.91	0.14	0.96	0.96	0.99	0.98
Calochortus apiculatus	31	299	2000	1.6	2	0.86	0.50	0.83	0.70	0.97	0.80
Campanula alaskana	59	271	2000	0.95	4	0.78	0.27	0.71	0.43	0.97	0.82
Carex raynoldsii	5	325	2000	1.5	4	0.93	-0.01	0.84	0.84	0.89	0.82
Castilleja lutescens	23	307	2000	0.65	1	0.88	0.55	0.86	0.73	0.99	0.94
Castilleja occidentalis	10	320	2000	0.55	1	0.76	0.21	0.63	0.51	0.77	0.22
Cerastium arvense	25	305	2000	0.45	2	0.86	0.44	0.83	0.72	0.93	0.71
Danthonia californica	6	324	2000	0.1	4	0.76	-0.08	0.69	0.66	0.88	0.87
Danthonia parryi	11	319	2000	0.55	1	0.93	0.28	0.92	0.86	1.00	0.99
Dasiphora fruticosa	75	255	2000	0.6	5	0.81	0.41	0.81	0.57	0.98	0.88
Elymys trachycaulus	15	315	2000	2.9	1	0.90	0.44	0.89	0.83	0.93	0.74
Erigeron caespitosus	21	309	2000	0.1	2	0.96	0.71	0.92	0.84	0.99	0.97
Eriocoma lemmonii	15	315	2000	1.95	2	0.94	0.54	0.91	0.84	0.99	0.96
Festuca campestris	43	287	2000	0.1	2	0.87	0.43	0.85	0.66	0.98	0.85
Festuca idahoensis	35	295	2000	2.25	1	0.86	0.26	0.87	0.75	0.98	0.81
Gaillardia aristata	38	292	2000	0.95	3	0.91	0.50	0.91	0.76	0.98	0.92
Galium boreale	117	213	2000	2.2	1	0.85	0.49	0.80	0.55	0.98	0.86
Gentianella amarella	9	321	2000	0.2	2	0.86	0.30	0.70	0.63	0.94	0.78
Geocaulon lividum	6	324	2000	0.1	2	0.88	-0.12	0.71	0.69	0.74	0.47

Table S9. Continued.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Geranium viscosissimum	40	290	2000	2.75	1	0.87	0.51	0.88	0.74	0.98	0.84
Geum triflorum	9	321	2000	0.1	1	0.86	0.28	0.93	0.86	0.99	0.98
Hedysarum sulphurescens	77	253	2000	0.7	5	0.79	0.31	0.78	0.53	0.97	0.82
Heterotheca villosa	7	323	2000	0.8	3	0.89	0.00	0.86	0.85	0.89	0.82
Heuchera parvifolia	6	324	2000	2.25	2	0.90	-0.05	0.93	0.93	0.91	0.91
Hypericum perforatum	15	315	2000	0.1	1	0.91	0.46	0.80	0.69	0.94	0.86
Koeleria macrantha	24	306	2000	2.6	4	0.93	0.58	0.86	0.79	0.97	0.92
Lilium philadelphicum	5	325	2000	0.8	1	0.87	-0.01	0.77	0.77	0.96	0.96
Lithospermum ruderale	31	299	2000	2	1	0.89	0.50	0.87	0.72	0.99	0.90
Lomatium macrocarpum	6	324	2000	0.1	5	0.73	-0.06	0.94	0.94	0.94	0.94
Lomatium triternatum	22	308	2000	2.45	1	0.88	0.46	0.85	0.69	0.94	0.79
Lupinus sericeus	50	280	2000	0.25	4	0.87	0.53	0.87	0.67	0.98	0.87
Monarda fistulosa	30	300	2000	2.15	3	0.89	0.52	0.88	0.73	0.98	0.87
Oxytropis sericea	23	307	2000	0.55	3	0.90	0.53	0.85	0.70	0.99	0.93
Oxytropis splendens	17	313	2000	2.25	1	0.87	0.38	0.79	0.72	0.96	0.76
Penstemon confertus	37	293	2000	0.1	2	0.88	0.48	0.83	0.66	0.98	0.90
Potentilla gracilis	31	299	2000	3	1	0.93	0.61	0.89	0.80	0.99	0.92
Potentilla pensylvanica	8	322	2000	0.4	3	0.92	0.40	0.85	0.83	1.00	0.99
Primula conjugens	12	318	2000	0.55	3	0.90	0.44	0.76	0.68	0.91	0.75
Pulsatilla nuttalliana	22	308	2000	0.6	2	0.92	0.59	0.88	0.73	0.98	0.84
Rhinanthus minor	12	318	2000	3	1	0.95	0.53	0.91	0.84	0.96	0.84
Rosa woodsii	49	281	2000	1.7	5	0.89	0.57	0.85	0.69	0.98	0.85
Sabulina rubella	21	309	2000	0.7	2	0.78	0.27	0.70	0.51	0.90	0.61
Sedum lanceolatum	31	299	2000	2.55	2	0.87	0.52	0.82	0.65	0.97	0.81
Silene parryi	15	315	2000	1.65	1	0.86	0.40	0.75	0.63	0.97	0.82
Sisyrinchium montanum	14	316	2000	0.1	1	0.88	0.41	0.89	0.81	0.96	0.83
Solidago spathulata	13	317	2000	3	2	0.91	0.44	0.75	0.63	0.97	0.82
Tragopogon dubius	6	324	2000	0.8	1	0.68	-0.01	0.97	0.96	0.97	0.97
Veronica wyomingensis	14	316	2000	0.65	1	0.84	0.34	0.80	0.71	0.95	0.74

Table S10. Model performance metrics for all component species of community H27 included in the stacked species distribution models (s-SDMs) built with all species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Achillea millefolium	88	242	2000	2.9	1	0.84	0.46	0.80	0.53	0.96	0.84
Agoseris glauca	33	297	2000	2.05	1	0.78	0.33	0.80	0.63	0.96	0.78
Allium cernuum	41	289	2000	0.5	1	0.86	0.42	0.79	0.60	0.96	0.82
Allium textile	5	325	2000	0.1	2	0.89	0.00	0.87	0.87	0.99	0.98
Anemone multifida	53	277	2000	0.25	2	0.87	0.45	0.81	0.58	0.97	0.84
Antennaria parvifolia	33	297	2000	2.1	1	0.83	0.29	0.82	0.63	0.97	0.84
Anticlea elegans	39	291	2000	0.75	4	0.84	0.43	0.78	0.59	0.93	0.70
Apocynum androsaemifolium	6	324	2000	0.5	1	0.78	-0.01	0.85	0.83	0.87	0.79
Arnica sororia	6	324	2000	2.9	4	0.83	0.00	0.94	0.94	0.96	0.95
Astragalus striatus	6	324	2000	0.1	1	0.94	0.00	0.94	0.94	0.95	0.95
Balsamorhiza sagittata	15	315	2000	1.65	5	0.87	0.37	0.82	0.71	0.94	0.86
Calochortus apiculatus	31	299	2000	1.6	2	0.86	0.50	0.83	0.70	0.97	0.80
Campanula alaskana	59	271	2000	0.95	4	0.78	0.27	0.71	0.43	0.97	0.82
Castilleja lutescens	23	307	2000	0.65	1	0.88	0.55	0.86	0.73	0.99	0.94
Cirsium hookerianum	14	316	2000	0.1	2	0.78	0.27	0.62	0.48	0.89	0.65
Comandra umbellata	10	320	2000	0.1	4	0.89	0.34	0.83	0.74	0.94	0.84
Dasiphora fruticosa	75	255	2000	0.6	5	0.81	0.41	0.81	0.57	0.98	0.88
Elymys trachycaulus	15	315	2000	2.9	1	0.90	0.44	0.89	0.83	0.93	0.74
Erigeron caespitosus	21	309	2000	0.1	2	0.96	0.71	0.92	0.84	0.99	0.97
Eriocoma lemmonii	15	315	2000	1.95	2	0.94	0.54	0.91	0.84	0.99	0.96
Festuca campestris	43	287	2000	0.1	2	0.87	0.43	0.85	0.66	0.98	0.85
Festuca idahoensis	35	295	2000	2.25	1	0.86	0.26	0.87	0.75	0.98	0.81
Fragaria virginiana	104	226	2000	2.25	1	0.77	0.36	0.76	0.47	0.98	0.85
Gaillardia aristata	38	292	2000	0.95	3	0.91	0.50	0.91	0.76	0.98	0.92

Table S10. Continued.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Galium boreale	117	213	2000	2.2	1	0.85	0.49	0.80	0.55	0.98	0.86
Geranium viscosissimum	40	290	2000	2.75	1	0.87	0.51	0.88	0.74	0.98	0.84
Geum triflorum	9	321	2000	0.1	1	0.86	0.28	0.93	0.86	0.99	0.98
Hedysarum sulphurescens	77	253	2000	0.7	5	0.79	0.31	0.78	0.53	0.97	0.82
Hypericum perforatum	15	315	2000	0.1	1	0.91	0.46	0.80	0.69	0.94	0.86
Juncus balticus	11	319	2000	0.1	2	0.77	0.21	0.98	0.98	0.99	0.99
Koeleria macrantha	24	306	2000	2.6	4	0.93	0.58	0.86	0.79	0.97	0.92
Linum lewisii	8	322	2000	0.85	4	0.95	0.37	0.91	0.87	0.97	0.89
Lithospermum ruderale	31	299	2000	2	1	0.89	0.50	0.87	0.72	0.99	0.90
Lupinus sericeus	50	280	2000	0.25	4	0.87	0.53	0.87	0.67	0.98	0.87
Monarda fistulosa	30	300	2000	2.15	3	0.89	0.52	0.88	0.73	0.98	0.87
Penstemon confertus	37	293	2000	0.1	2	0.88	0.48	0.83	0.66	0.98	0.90
Phleum pratense	37	293	2000	0.2	3	0.93	0.66	0.86	0.69	0.98	0.81
Poa cusickii	7	323	2000	0.15	5	0.80	-0.04	0.81	0.73	0.90	0.83
Poa interior	8	322	2000	0.85	1	0.84	0.24	0.64	0.58	0.91	0.73
Potentilla glandulosa	10	320	2000	0.9	5	0.71	0.04	0.63	0.60	0.83	0.52
Potentilla gracilis	31	299	2000	3	1	0.93	0.61	0.89	0.80	0.99	0.92
Pseudoroegneria spicata	33	297	2000	0.65	5	0.86	0.41	0.88	0.77	0.99	0.87
Rosa woodsii	49	281	2000	1.7	5	0.89	0.57	0.85	0.69	0.98	0.85
Solidago missouriensis	7	323	2000	3	4	0.86	-0.04	0.88	0.85	1.00	1.00
Tragopogon dubius	6	324	2000	0.8	1	0.68	-0.01	0.97	0.96	0.97	0.97

Table S11. Model performance metrics for all component species of community C60 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Actaea rubra	20	289	2000	0.95	4	0.89	0.50	0.76	0.58	0.97	0.80
Amelanchier alnifolia	20	219	2000	3	1	0.83	0.41	0.74	0.63	0.98	0.85
Bromus inermis	20	292	2000	0.6	2	0.94	0.63	0.84	0.69	0.98	0.85
Bromus vulgaris	20	300	2000	2.7	4	0.86	0.55	0.80	0.70	0.99	0.89
Calamagrostis canadensis	20	306	2000	0.85	3	0.83	0.42	0.81	0.70	0.95	0.75
Elymus glaucus	18	312	2000	0.1	3	0.81	0.33	0.84	0.67	0.91	0.61
Eucephalus engelmannii	20	287	2000	0.6	5	0.83	0.41	0.71	0.54	0.93	0.64
Geranium richardsonii	20	302	2000	0.35	2	0.94	0.68	0.92	0.84	0.98	0.91
Heracleum maximum	20	271	2000	0.1	5	0.81	0.31	0.62	0.42	0.97	0.87
Lathyrus ochroleucus	20	285	2000	2.6	4	0.91	0.48	0.81	0.64	0.99	0.91
Maianthemum racemosa	20	255	2000	0.1	4	0.95	0.64	0.79	0.62	0.98	0.90
Moehringia lateriflora	18	312	2000	2	4	0.76	0.29	0.82	0.69	0.94	0.68
Osmorhiza occidentalis	20	278	2000	0.35	4	0.88	0.39	0.80	0.65	0.97	0.82
Populus tremuloides	20	277	2000	1.15	3	0.85	0.36	0.78	0.64	0.95	0.78
Prosartes hookeri	7	323	2000	0.15	4	0.96	-0.04	0.83	0.80	0.98	0.98
Prunus virginiana	20	299	2000	0.1	4	0.92	0.50	0.90	0.82	0.98	0.89
Sanicula marilandica	14	316	2000	0.25	4	0.91	0.48	0.93	0.85	0.98	0.92
Streptopus amplexifolius	15	315	2000	3	4	0.85	0.36	0.70	0.52	0.96	0.85
Symphoricarpos albus	20	304	2000	1.95	1	0.90	0.56	0.80	0.70	0.97	0.84
Symphoricarpos occidentalis	20	279	2000	0.1	3	0.93	0.64	0.81	0.69	1.00	0.99
Thalictrum occidentale	20	191	2000	0.1	3	0.89	0.50	0.71	0.55	0.95	0.77
Urtica dioica	17	313	2000	2.9	1	0.89	0.43	0.88	0.78	0.94	0.76
Veratrum viride	20	256	2000	2.1	3	0.79	0.34	0.72	0.53	0.98	0.86
Vicia americana	20	281	2000	0.2	4	0.90	0.51	0.81	0.71	0.95	0.79
Viola canadensis	20	294	2000	0.3	4	0.93	0.62	0.80	0.71	0.99	0.88

Table S12. Model performance metrics for all component species of community C61 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Amelanchier alnifolia	7	219	2000	2.95	4	0.92	-0.02	0.66	0.66	0.86	0.67
Berberis repens	7	267	2000	0.35	5	0.92	-0.02	0.68	0.64	0.89	0.80
Cornus sericea	7	313	2000	2	3	0.88	0.00	0.78	0.74	1.00	1.00
Elymus glaucus	7	312	2000	1.5	1	0.84	-0.09	0.80	0.79	0.83	0.66
Geranium richardsonii	7	302	2000	1.7	2	0.90	-0.03	0.79	0.77	0.92	0.91
Lathyrus ochroleucus	7	285	2000	1.55	2	0.88	-0.01	0.82	0.81	0.97	0.97
Maianthemum stellata	7	291	2000	3	4	0.90	-0.09	0.63	0.61	0.92	0.92
Osmorhiza occidentalis	7	278	2000	2.95	2	0.81	-0.03	0.71	0.70	0.82	0.65
Populus tremuloides	7	277	2000	2.05	1	0.88	-0.01	0.88	0.84	0.98	0.98
Prosartes hookeri	7	323	2000	0.1	4	0.96	-0.04	0.88	0.84	0.98	0.98
Prunus virginiana	7	299	2000	2.2	1	0.86	-0.13	0.93	0.89	1.00	1.00
Rosa woodsii	7	281	2000	1.05	2	0.72	-0.01	0.90	0.86	0.93	0.92
Rubus parviflorus	7	245	2000	0.45	3	0.84	-0.07	0.69	0.66	1.00	1.00
Sanicula marilandica	7	316	2000	3	2	0.94	-0.01	0.81	0.80	0.97	0.96
Symphoricarpos occidentalis	7	279	2000	0.15	5	0.74	0.00	0.55	0.51	0.91	0.83
Symphyotrichum ciliolatum	7	281	2000	2.45	3	0.78	-0.10	0.84	0.80	0.94	0.94
Urtica dioica	7	313	2000	3	1	0.93	-0.02	0.67	0.65	1.00	1.00
Vicia americana	7	281	2000	2.65	2	0.87	-0.02	0.83	0.83	0.92	0.91
Viola canadensis	7	294	2000	0.15	1	0.82	-0.03	0.68	0.63	0.98	0.98

Table S13. Model performance metrics for all component species of community C62 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Actaea rubra	5	289	2000	0.2	5	0.82	-0.08	0.74	0.74	0.98	0.98
Amelanchier alnifolia	5	219	2000	0.15	5	0.94	0.00	0.67	0.67	0.83	0.66
Berberis repens	5	267	2000	0.15	4	0.93	-0.01	0.83	0.83	0.98	0.98
Bromus vulgaris	5	300	2000	0.8	5	0.77	-0.06	0.78	0.78	1.00	1.00
Calamagrostis canadensis	5	306	2000	2	1	0.75	-0.15	0.95	0.95	0.90	0.83
Chamaenerion angustifolium	5	211	2000	2.95	1	0.58	-0.21	0.89	0.89	0.92	0.92
Eurybia conspicua	5	248	2000	0.15	5	0.87	-0.02	0.52	0.52	0.94	0.93
Populus tremuloides	5	277	2000	3	4	0.86	-0.03	0.89	0.89	0.78	0.60
Prosartes trachycarpa	5	288	2000	2.7	3	0.85	-0.05	0.89	0.88	0.90	0.82
Prunus virginiana	5	299	2000	0.15	1	0.82	-0.01	0.70	0.70	0.95	0.95
Rosa acicularis	5	277	2000	0.3	2	0.88	0.00	0.74	0.74	0.91	0.90
Symphoricarpos occidentalis	5	308	2000	2.75	2	0.86	-0.03	0.42	0.42	0.90	0.82
Thalictrum occidentale	5	297	2000	0.1	5	0.88	-0.01	0.54	0.54	0.78	0.63

Table S14. Model performance metrics for all component species of community C78 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Allium textile	5	325	2000	0.1	2	0.89	0.00	0.87	0.87	0.98	0.98
Angelica arguta	6	324	2000	1.6	1	0.82	-0.07	0.93	0.92	0.86	0.77
Athyrium filix.femina	5	325	2000	0.1	1	0.84	0.00	0.90	0.90	0.87	0.73
Bromus ciliatus	6	321	2000	2.3	3	0.74	-0.05	0.82	0.79	0.91	0.83
Bromus vulgaris	6	300	2000	0.1	1	0.88	-0.02	0.78	0.76	0.99	0.98
Calamagrostis canadensis	6	306	2000	3	1	0.78	-0.07	0.64	0.63	0.97	0.97
Cornus sericea	6	313	2000	0.75	2	0.93	-0.02	0.84	0.82	1.00	1.00
Equisetum arvense	6	303	2000	0.15	4	1.00	0.00	0.62	0.60	1.00	1.00
Galium triflorum	6	291	2000	0.1	4	0.85	-0.15	0.71	0.71	0.88	0.80
Geranium richardsonii	6	302	2000	0.9	2	0.97	-0.03	0.85	0.85	0.97	0.97
Geum allepicum	6	323	2000	0.1	2	0.72	-0.02	0.82	0.78	0.88	0.69
Geum macrophyllum	6	323	2000	0.25	2	0.78	-0.05	0.74	0.73	0.92	0.91
Geum rivale	6	322	2000	0.1	2	0.80	-0.03	0.80	0.80	0.94	0.93
Graphephorum cernuum	5	325	2000	0.1	3	0.99	0.00	0.73	0.73	0.92	0.84
Habenaria hyperborea	6	320	2000	0.1	3	0.90	-0.05	0.79	0.79	0.92	0.85
Heracleum maximum	6	271	2000	0.15	5	0.79	-0.09	0.81	0.81	0.91	0.81
Linnaea borealis	6	303	2000	2.05	1	0.99	0.00	0.79	0.75	1.00	1.00
Lonicera involucrata	6	297	2000	0.1	4	0.79	-0.09	0.81	0.80	0.84	0.69
Lycopodium annotinum	6	324	2000	0.25	4	0.95	-0.04	0.80	0.76	0.86	0.69
Maianthemum racemosum	6	255	2000	0.1	3	0.80	-0.11	0.58	0.52	0.91	0.83
Mitella nuda	5	325	2000	0.15	2	0.89	0.00	0.86	0.86	0.93	0.86
Osmorhiza depauperata	6	294	2000	0.8	1	0.87	-0.02	0.63	0.62	0.82	0.72
Packera pseudaurea	6	319	2000	2.05	1	0.89	-0.11	0.63	0.60	0.95	0.95
Petasites sagittatus	5	325	2000	2.3	2	0.89	0.00	0.54	0.53	0.99	0.99
Picea engelmannii	6	211	2000	1.8	2	0.82	-0.02	0.85	0.82	0.90	0.82
Populus trichocarpa	6	320	2000	0.3	2	0.86	-0.03	0.73	0.69	0.88	0.78
Pyrola asarifolia	6	308	2000	0.15	5	0.77	-0.07	0.75	0.75	0.87	0.78

Table S14. Continued.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Ribes lacustre	6	275	2000	0.15	5	0.79	-0.10	0.67	0.63	0.88	0.87
Rosa acicularis	6	277	2000	0.1	4	0.69	-0.14	0.69	0.66	0.89	0.81
Rubus pubescens	6	323	2000	0.25	3	0.73	-0.08	0.76	0.74	0.87	0.68
Senecio triangularis	6	298	2000	0.35	1	0.70	-0.15	0.67	0.65	0.90	0.83
Symphyotrichum ciliolatum	6	281	2000	0.1	3	0.92	-0.02	0.84	0.80	1.00	1.00
Symphyotrichum subspicatum	6	321	2000	0.1	1	0.89	-0.01	0.84	0.81	0.91	0.83
Taraxacum officinale	6	296	2000	2.75	4	0.88	-0.01	0.93	0.93	0.97	0.96

Table S15. Model performance metrics for all component species of community C79 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Alnus crispa	5	295	2000	1.35	1	0.88	-0.01	0.56	0.56	0.82	0.72
Berberis repens	5	267	2000	0.15	4	0.93	-0.01	0.83	0.83	0.98	0.98
Calamagrostis rubescens	5	284	2000	0.15	4	0.88	-0.01	0.83	0.83	0.65	0.45
Campanula alaskana	5	271	2000	0.1	2	0.62	-0.14	0.58	0.57	0.89	0.81
Chamaenerion angustifolium	5	211	2000	2.95	1	0.58	-0.21	0.89	0.89	0.92	0.92
Chimaphila umbellata	5	294	2000	0.3	1	0.94	0.00	0.81	0.81	0.86	0.77
Clintonia uniflora	5	277	2000	0.95	1	0.88	0.00	0.80	0.80	0.86	0.76
Cornus canadensis	5	320	2000	1.65	4	0.87	-0.04	0.96	0.96	0.93	0.92
Eurybia conspicua	5	248	2000	0.15	5	0.87	-0.02	0.52	0.52	0.94	0.93
Hieracium albiflorum	5	306	2000	0.1	3	0.74	-0.22	0.71	0.71	0.70	0.46
Lathyrus ochroleucus	5	285	2000	2.7	4	0.92	-0.01	0.79	0.79	0.91	0.91
Linnaea borealis	5	303	2000	0.1	2	0.77	0.00	0.76	0.76	0.85	0.78
Maianthemum racemosa	5	289	2000	1.35	1	0.82	-0.08	0.71	0.71	0.89	0.81
Moehringia lateriflora	5	312	2000	1.5	3	0.78	-0.10	0.72	0.72	0.85	0.71
Pinus contorta	5	254	2000	0.1	3	0.85	-0.15	0.69	0.69	0.91	0.90
Pyrola asarifolia	5	308	2000	2.95	4	0.87	-0.01	0.74	0.74	0.97	0.96
Rosa acicularis	5	277	2000	0.3	2	0.88	0.00	0.74	0.74	0.91	0.90
Spiraea lucida	5	325	2000	0.15	3	0.73	-0.01	0.75	0.75	0.90	0.71
Symphyotrichum ciliolatum	5	281	2000	0.15	3	0.82	-0.02	0.94	0.93	0.92	0.85
Vaccinium myrtillus	5	324	2000	2.3	2	0.77	-0.13	0.83	0.83	0.86	0.77
Vicia americana	5	248	2000	0.3	1	0.87	-0.02	0.87	0.87	0.91	0.83

Table S16. Model performance metrics for all component species of community H25 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Achillea millefolium	20	242	2000	0.1	2	0.90	0.50	0.78	0.60	0.95	0.76
Agoseris glauca	20	297	2000	0.3	4	0.73	0.22	0.74	0.57	0.95	0.74
Allium cernuum	20	289	2000	2.35	1	0.89	0.43	0.80	0.68	0.98	0.83
Anemone multifida	20	277	2000	0.35	4	0.84	0.22	0.78	0.62	0.94	0.80
Antennaria parvifolia	20	297	2000	2.8	1	0.86	0.41	0.76	0.61	0.95	0.83
Anticlea elegans	20	291	2000	1.7	2	0.79	0.28	0.74	0.58	0.96	0.85
Arctostaphylos uva.ursi	20	286	2000	0.35	1	0.82	0.25	0.78	0.61	0.92	0.73
Artemisia michauxiana	6	324	2000	0.1	5	0.79	-0.07	0.68	0.63	0.81	0.55
Balsamorhiza sagittata	15	315	2000	1.65	5	0.87	0.37	0.82	0.71	0.95	0.86
Bupleurum americanum	11	319	2000	0.4	4	0.91	0.14	0.98	0.97	0.99	0.98
Calochortus apiculatus	20	299	2000	0.5	1	0.85	0.47	0.79	0.65	0.97	0.86
Campanula alaskana	20	271	2000	0.9	1	0.84	0.42	0.79	0.67	0.94	0.73
Carex raynoldsii	5	325	2000	1.4	4	0.93	-0.01	0.84	0.84	0.94	0.92
Castilleja lutescens	20	307	2000	1.3	3	0.92	0.59	0.81	0.70	0.99	0.95
Castilleja occidentalis	10	320	2000	0.55	1	0.76	0.21	0.63	0.51	0.80	0.27
Cerastium arvense	20	305	2000	0.65	2	0.85	0.40	0.76	0.62	0.98	0.86
Danthonia californica	6	324	2000	0.1	4	0.76	-0.08	0.69	0.66	0.88	0.87
Danthonia parryi	11	319	2000	0.55	1	0.88	0.28	0.92	0.86	0.99	0.99
Dasiphora fruticosa	20	255	2000	0.8	5	0.85	0.42	0.68	0.55	0.97	0.87
Elymys trachycaulus	15	315	2000	2.7	1	0.90	0.44	0.89	0.84	0.94	0.74
Erigeron caespitosus	20	309	2000	0.25	3	0.97	0.69	0.94	0.89	0.99	0.95
Eriocoma lemmonii	15	315	2000	2.1	2	0.94	0.54	0.91	0.85	0.99	0.97
Festuca campestris	20	287	2000	1	3	0.85	0.36	0.80	0.65	0.91	0.68
Festuca idahoensis	20	295	2000	2.5	4	0.91	0.52	0.75	0.64	0.98	0.87
Gaillardia aristata	20	292	2000	1.75	1	0.93	0.58	0.88	0.77	0.99	0.95
Galium boreale	20	213	2000	0.3	1	0.92	0.57	0.76	0.61	0.95	0.83
Gentianella amarella	9	321	2000	0.1	1	0.86	0.30	0.69	0.64	0.93	0.72
Geocaulon lividum	6	324	2000	0.1	2	0.88	-0.12	0.71	0.69	0.78	0.51

Table S16. Continued.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Geranium viscosissimum	20	290	2000	0.75	1	0.86	0.38	0.83	0.68	0.98	0.83
Geum triflorum	9	321	2000	0.1	1	0.86	0.28	0.93	0.86	0.99	0.98
Hedysarum sulphurescens	20	253	2000	0.2	5	0.78	0.25	0.80	0.66	0.96	0.80
Heterotheca villosa	7	323	2000	0.85	3	0.89	-0.01	0.87	0.86	0.88	0.80
Heuchera parvifolia	6	324	2000	2.25	2	0.90	-0.05	0.93	0.93	0.93	0.92
Hypericum perforatum	15	315	2000	0.15	2	0.91	0.46	0.81	0.70	0.94	0.86
Koeleria macrantha	20	306	2000	2.3	4	0.95	0.63	0.85	0.78	0.98	0.91
Lilium philadelphicum	5	325	2000	1	1	0.87	-0.01	0.77	0.77	0.92	0.85
Lithospermum ruderale	20	299	2000	2.75	4	0.87	0.46	0.85	0.72	0.97	0.83
Lomatium macrocarpum	6	324	2000	0.5	4	0.73	-0.06	0.94	0.94	0.93	0.93
Lomatium triternatum	20	308	2000	2.15	2	0.89	0.47	0.86	0.72	0.96	0.84
Lupinus sericeus	20	280	2000	2.15	4	0.91	0.55	0.79	0.65	0.98	0.93
Monarda fistulosa	20	300	2000	1.1	2	0.90	0.52	0.87	0.72	0.98	0.87
Oxytropis sericea	20	307	2000	0.15	5	0.87	0.35	0.77	0.60	0.97	0.85
Oxytropis splendens	17	313	2000	2.25	1	0.87	0.38	0.79	0.72	0.96	0.78
Penstemon confertus	20	293	2000	2.85	1	0.80	0.21	0.81	0.68	0.97	0.86
Potentilla gracilis	20	299	2000	0.55	1	0.97	0.67	0.90	0.83	0.97	0.87
Potentilla pensylvanica	8	322	2000	0.25	3	0.92	0.40	0.88	0.86	0.99	0.99
Primula conjugens	12	318	2000	0.6	3	0.90	0.44	0.75	0.67	0.93	0.79
Pulsatilla nuttalliana	20	308	2000	1.15	2	0.89	0.53	0.87	0.76	0.96	0.84
Rhinanthus minor	12	318	2000	3	1	0.95	0.53	0.91	0.84	0.97	0.84
Rosa woodsii	20	281	2000	0.9	4	0.82	0.33	0.82	0.69	0.98	0.80
Sabulina rubella	20	309	2000	0.1	2	0.82	0.34	0.73	0.56	0.95	0.80
Sedum lanceolatum	20	299	2000	1.6	2	0.84	0.33	0.83	0.71	0.95	0.73
Silene parryi	15	315	2000	0.8	1	0.86	0.40	0.76	0.62	0.97	0.85
Sisyrinchium montanum	14	316	2000	0.1	1	0.88	0.41	0.89	0.81	0.96	0.79
Solidago spathulata	13	317	2000	3	2	0.91	0.44	0.75	0.63	0.97	0.81
Tragopogon dubius	6	324	2000	0.1	1	0.68	-0.01	0.96	0.96	0.97	0.97
Veronica wyomingensis	14	316	2000	0.65	1	0.84	0.37	0.80	0.69	0.95	0.74

Table S17. Model performance metrics for all component species of community H27 included in the stacked species distribution models (s-SDMs) built with a reduced set of species occurrence records available from the ELC. Columns as described in Table S3.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Achillea millefolium	11	242	2000	2.5	1	0.85	0.30	0.91	0.87	0.94	0.86
Agoseris glauca	11	297	2000	3	4	0.73	0.08	0.67	0.57	0.95	0.81
Allium cernuum	11	289	2000	1.15	1	0.87	0.35	0.69	0.59	0.96	0.91
Allium textile	5	325	2000	0.1	2	0.89	0.00	0.87	0.87	0.98	0.98
Anemone multifida	11	277	2000	0.4	3	0.81	0.23	0.84	0.81	0.87	0.66
Antennaria parvifolia	11	297	2000	2.05	4	0.78	0.10	0.72	0.57	0.85	0.59
Anticlea elegans	11	291	2000	1.65	2	0.83	0.21	0.81	0.75	0.91	0.71
Apocynum androsaemifolium	6	324	2000	1.55	4	0.78	-0.01	0.84	0.83	0.86	0.77
Arnica sororia	6	324	2000	2.8	4	0.83	0.00	0.95	0.95	0.93	0.92
Astragalus striatus	6	324	2000	0.1	1	0.94	0.00	0.94	0.94	0.93	0.93
Balsamorhiza sagittata	11	315	2000	2.45	3	0.93	0.33	0.89	0.83	1.00	1.00
Calochortus apiculatus	11	299	2000	0.45	1	0.88	0.27	0.80	0.73	0.97	0.90
Campanula alaskana	11	271	2000	0.1	4	0.75	0.14	0.56	0.48	0.95	0.80
Castilleja lutescens	11	307	2000	2.8	1	0.86	0.23	0.80	0.74	0.97	0.93
Cirsium hookerianum	11	316	2000	2.3	4	0.75	0.15	0.71	0.62	0.86	0.61
Comandra umbellata	10	320	2000	0.1	4	0.89	0.34	0.83	0.72	0.93	0.85
Dasiphora fruticosa	11	255	2000	0.35	4	0.98	0.47	0.75	0.64	0.88	0.70
Elymys trachycaulus	11	315	2000	0.25	1	0.91	0.34	0.85	0.77	0.98	0.92
Erigeron caespitosus	11	309	2000	1.4	4	0.83	0.28	0.93	0.90	0.99	0.98
Eriocoma lemmonii	11	315	2000	0.5	2	0.88	0.23	0.92	0.85	1.00	1.00
Festuca campestris	11	287	2000	1.45	3	0.94	0.26	0.88	0.83	0.94	0.87
Festuca idahoensis	11	295	2000	2.05	3	0.79	0.21	0.65	0.56	0.96	0.87
Fragaria virginiana	11	226	2000	1.7	4	0.89	0.28	0.87	0.79	0.91	0.71
Gaillardia aristata	11	292	2000	1.4	2	0.95	0.32	0.89	0.84	0.95	0.82
Galium boreale	11	213	2000	0.1	5	0.81	0.20	0.79	0.67	0.94	0.86
Geranium viscosissimum	11	290	2000	0.25	5	0.92	0.33	0.79	0.68	0.97	0.90
Geum triflorum	9	321	2000	0.1	1	0.86	0.28	0.93	0.86	0.99	0.98
Hedysarum sulphurescens	11	253	2000	0.4	4	0.92	0.35	0.84	0.71	0.90	0.62
Hypericum perforatum	11	315	2000	1.05	3	0.91	0.40	0.76	0.66	1.00	1.00

Table S17. Continued.

Species	Number of presences	True Absences	Background Points	MaxEnt Regmult	MaxEnt Classes	GLM AUC mean	GLM TSS mean	MaxEnt AUC mean	MaxEnt TSS mean	RF AUC mean	RF TSS mean
Juncus balticus	11	319	2000	0.1	2	0.77	0.21	0.98	0.97	0.99	0.99
Koeleria macrantha	11	306	2000	0.55	1	0.93	0.34	0.83	0.70	0.99	0.99
Linum lewisii	8	322	2000	1.05	1	0.95	0.37	0.91	0.87	0.94	0.87
Lithospermum ruderale	11	299	2000	1.4	3	0.88	0.34	0.77	0.74	0.95	0.88
Lupinus sericeus	11	280	2000	1.85	4	0.90	0.27	0.69	0.61	0.86	0.67
Monarda fistulosa	11	300	2000	0.1	5	0.92	0.35	0.78	0.70	0.98	0.91
Penstemon confertus	11	293	2000	1.85	4	0.85	0.27	0.76	0.64	0.99	0.97
Phleum pratense	11	293	2000	0.1	4	0.89	0.29	0.90	0.84	0.96	0.82
Poa cusickii	7	323	2000	0.1	5	0.80	-0.04	0.83	0.75	0.90	0.82
Poa interior	8	322	2000	0.85	1	0.84	0.24	0.64	0.58	0.93	0.72
Potentilla glandulosa	10	320	2000	0.85	5	0.71	0.04	0.63	0.60	0.85	0.52
Potentilla gracilis	11	299	2000	1	1	0.94	0.19	0.97	0.94	0.98	0.97
Pseudoroegneria spicata	11	297	2000	0.25	2	0.83	0.25	0.70	0.62	0.87	0.69
Rosa woodsii	11	281	2000	0.1	5	0.93	0.28	0.80	0.69	0.99	0.98
Solidago missouriensis	7	323	2000	3	4	0.86	-0.04	0.88	0.85	0.99	0.99
Tragopogon dubius	6	324	2000	1.75	2	0.68	-0.01	0.96	0.95	0.97	0.97

Table S18. Species identified to genus level due to difficulties in differentiating between them in the field or from samples.

Genus	Entries included in dataset prior to lumping	Reason for lumping
Antennaria	*alpina* *anaphaloides* *lanata* *media* *neglecta* *umbrinella*	Without flowers, species are very similar in morphology, making field identification difficult.
Danthonia	*californica* *intermedia* *parryi*	*Danthonia spp*. requires a dissecting scope for identification and even then can be difficult to tell apart due to minute features. Additionally, *Danthonia spp*. is listed as the dominant of community H25, indicating that multiple *Danthonia spp.* could be present.
Prosartes	*hookeri* *trachycarpa*	These two species are very similar in morphology and are difficult to identify without fruits.
Rosa	*acicularis* *woodsii*	These two species are very similar in morphology and can be found in the same habitat, making their identification difficult.
Thalictrum	*occidentalis* *venulosum*	These two species are very similar in morphology and can be found in the same habitat, making their identification difficult without fruits present.

Table S19. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community C60 based on GLMs relating measured community similarity to the prediction of the DM. DM names list the model algorithm used (GLM, RF, or MaxEnt), the stacking method used for s-SDMs (a = arithmetic or g = geometric), and whether all or the dominant species were included in the s-SDM (a = all or d = dominants), and whether they were reduced or not (red). For models without a significant interaction between the DM prediction and burn status, interaction coefficients are listed as NA. For models with a significant interaction the DM coefficient is the coefficient of the group (burned or unburned) with the highest value. The best model of each type is bolded.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	**C60_glm**	**-1.125**	**0.089**	**0.26**	**0.055**	**<0.0001**	**-0.36**	**0.126**	**0.0051**	**NA**	**NA**	**0.5227**	**-140.878**
CDM	C60max	-1.126	0.092	0.22	0.057	<0.001	-0.35	0.130	0.0089	NA	NA	0.2092	-135.498
CDM	C60rf	-1.166	0.093	0.25	0.060	<0.001	-0.28	0.130	0.0348	NA	NA	0.8214	-137.400
sSDM	**C60_aa_glm**	**-1.252**	**0.077**	**0.44**	**0.052**	**<0.0001**	**-0.18**	**0.106**	**0.0864**	**NA**	**NA**	**0.9447**	**-174.092**
sSDM	C60_aa_max_tune	-1.184	0.078	0.40	0.054	<0.0001	-0.30	0.108	0.0076	NA	NA	0.5167	-166.127
sSDM	C60_aa_rf	-1.314	0.089	0.41	0.060	<0.0001	-0.04	0.125	0.7249	NA	NA	0.6603	-159.056
sSDM	C60_ad_glm	-1.179	0.079	0.40	0.056	<0.0001	-0.30	0.110	0.0070	NA	NA	0.8503	-164.142
sSDM	C60_ad_max_tune	-1.191	0.080	0.41	0.062	<0.0001	-0.29	0.111	0.0115	NA	NA	0.3881	-161.735
sSDM	C60_ad_rf	-1.267	0.087	0.40	0.063	<0.0001	-0.14	0.120	0.2627	NA	NA	0.9128	-157.570
sSDM	C60_ga_glm	-1.236	0.084	0.39	0.055	<0.0001	-0.20	0.116	0.0907	NA	NA	0.7154	-159.363
sSDM	C60_ga_max_tune	-1.192	0.086	0.33	0.057	<0.0001	-0.26	0.121	0.0345	NA	NA	0.1690	-149.606
sSDM	C60_ga_rf	-1.243	0.097	0.30	0.068	<0.0001	-0.15	0.137	0.2707	NA	NA	0.5566	-140.669
sSDM	C60_gd_glm	-1.189	0.082	0.39	0.058	<0.0001	-0.28	0.114	0.0150	NA	NA	0.7959	-158.566
sSDM	C60_gd_max_tune	-1.191	0.082	0.40	0.064	<0.0001	-0.28	0.113	0.0146	NA	NA	0.3314	-158.420
sSDM	C60_gd_rf	-1.283	0.090	0.39	0.063	<0.0001	-0.10	0.125	0.4260	NA	NA	0.9192	-154.963
sSDMr	**C60_aa_glm_red**	**-1.265**	**0.079**	**0.42**	**0.052**	**<0.0001**	**-0.15**	**0.110**	**0.1755**	**NA**	**NA**	**0.6384**	**-169.488**
sSDMr	C60_aa_max_tune_red	-1.204	0.078	0.41	0.053	<0.0001	-0.26	0.108	0.0186	NA	NA	0.9474	-168.005
sSDMr	C60_aa_rf_red	-1.273	0.096	0.33	0.065	<0.0001	-0.10	0.136	0.4793	NA	NA	0.6516	-145.300
sSDMr	C60_ad_glm_red	-1.129	0.080	0.36	0.052	<0.0001	-0.38	0.112	0.0011	NA	NA	0.7517	-160.194
sSDMr	C60_ad_max_tune_red	-1.130	0.085	0.34	0.060	<0.0001	-0.37	0.118	0.0026	NA	NA	0.7175	-151.365
sSDMr	C60_ad_rf_red	-1.174	0.097	0.21	0.066	<0.01	-0.25	0.137	0.0678	NA	NA	0.8781	-133.082
sSDMr	C60_ga_glm_red	-1.236	0.084	0.38	0.054	<0.0001	-0.19	0.117	0.1010	NA	NA	0.7372	-158.295
sSDMr	C60_ga_max_tune_red	-1.217	0.087	0.34	0.056	<0.0001	-0.21	0.122	0.0856	NA	NA	0.9869	-151.761

Table S19. Continued.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
sSDMr	C60_ga_rf_red	-1.234	0.100	0.28	0.069	<0.001	-0.16	0.142	0.2645	NA	NA	0.8205	-137.442
sSDMr	C60_gd_glm_red	-1.123	0.081	0.37	0.055	<0.0001	-0.40	0.114	0.0007	NA	NA	0.8784	-157.638
sSDMr	C60_gd_max_tune_red	-1.060	0.096	0.19	0.064	<0.01	-0.46	0.137	0.0013	NA	NA	0.7509	-131.439
sSDMr	C60_gd_rf_red	-1.162	0.099	0.18	0.067	<0.01	-0.27	0.139	0.0570	NA	NA	0.4401	-130.423

Table S20. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community C61 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	C61_glm	-1.015	0.108	-0.11	0.074	0.1124	-0.31	0.150	0.040	NA	NA	0.0646	-107.046
CDM	**C61max**	**-1.093**	**0.103**	**0.24**	**0.064**	**<0.001**	**-0.19**	**0.141**	**0.183**	**NA**	**NA**	**0.5437**	**-115.974**
CDM	C61rf	-1.080	0.108	0.13	0.070	0.0665	-0.20	0.150	0.189	NA	NA	0.7708	-107.895
sSDM	**C61_aa_glm**	**-1.207**	**0.085**	**0.46**	**0.057**	**<0.0001**	**-0.04**	**0.116**	**0.709**	**NA**	**NA**	**0.9151**	**-153.205**
sSDM	C61_aa_max_tune	-1.136	0.089	0.40	0.061	<0.0001	-0.16	0.122	0.202	NA	NA	0.3961	-139.817
sSDM	C61_aa_rf	-1.283	0.096	0.44	0.065	<0.0001	0.12	0.134	0.385	NA	NA	0.9071	-141.205
sSDM	C61_ad_glm	-1.102	0.090	0.39	0.061	<0.0001	-0.21	0.124	0.092	NA	NA	0.8425	-137.554
sSDM	C61_ad_max_tune	-1.052	0.084	0.44	0.065	<0.0001	-0.33	0.117	0.007	NA	NA	0.3376	-145.121
sSDM	C61_ad_rf	-1.115	0.098	0.32	0.067	<0.0001	-0.17	0.134	0.215	NA	NA	0.6746	-125.396
sSDM	C61_ga_glm	-1.165	0.095	0.37	0.063	<0.0001	-0.10	0.131	0.451	NA	NA	0.5285	-131.742
sSDM	C61_ga_max_tune	-1.136	0.099	0.31	0.065	<0.0001	-0.13	0.136	0.343	NA	NA	0.3280	-123.412
sSDM	C61_ga_rf	-1.185	0.107	0.30	0.076	<0.001	-0.04	0.149	0.807	NA	NA	0.4501	-119.118
sSDM	C61_gd_glm	-1.105	0.093	0.37	0.062	<0.0001	-0.20	0.127	0.118	NA	NA	0.5898	-133.997
sSDM	C61_gd_max_tune	-1.032	0.087	0.41	0.065	<0.0001	-0.35	0.122	0.005	NA	NA	0.5299	-138.741
sSDM	C61_gd_rf	-1.095	0.101	0.27	0.067	<0.001	-0.19	0.138	0.167	NA	NA	0.9206	-119.358
sSDMr	C61_aa_glm_red	-1.085	0.112	0.10	0.073	0.1885	-0.18	0.158	0.250	NA	NA	0.0785	-106.256
sSDMr	C61_aa_max_tune_red	-1.116	0.100	0.30	0.072	<0.0001	-0.17	0.137	0.228	NA	NA	0.4020	-120.770
sSDMr	C61_aa_rf_red	-1.201	0.106	0.32	0.070	<0.0001	-0.01	0.147	0.940	NA	NA	0.9774	-122.477
sSDMr	C61_ad_glm_red	-1.047	0.107	0.10	0.070	0.1447	-0.26	0.147	0.086	NA	NA	0.7465	-106.654
sSDMr	C61_ad_max_tune_red	-1.073	0.104	0.20	0.071	<0.01	-0.22	0.143	0.124	NA	NA	0.2262	-112.323
sSDMr	C61_ad_rf_red	-1.109	0.105	0.23	0.070	<0.01	-0.16	0.145	0.269	NA	NA	0.8310	-113.835
sSDMr	C61_ga_glm_red	-1.112	0.106	0.22	0.067	<0.01	-0.15	0.147	0.297	NA	NA	0.8892	-113.508
sSDMr	**C61_ga_max_tune_red**	**-1.134**	**0.098**	**0.33**	**0.066**	**<0.0001**	**-0.14**	**0.134**	**0.289**	**NA**	**NA**	**0.1902**	**-124.973**
sSDMr	C61_ga_rf_red	-1.156	0.110	0.24	0.075	<0.01	-0.08	0.154	0.622	NA	NA	0.6878	-113.907
sSDMr	C61_gd_glm_red	-1.045	0.108	0.05	0.068	0.4268	-0.25	0.148	0.090	NA	NA	0.5574	-105.158
sSDMr	C61_gd_max_tune_red	-0.979	0.099	0.29	0.069	<0.0001	-0.42	0.143	0.004	NA	NA	0.1561	-119.875
sSDMr	C61_gd_rf_red	-1.109	0.108	0.19	0.071	<0.01	-0.15	0.150	0.312	NA	NA	0.9671	-111.240

Table S21. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community C62 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	C62_glm	-1.122	0.107	-0.08	0.072	0.2484	-0.25	0.149	0.093	NA	NA	0.1789	-113.341
CDM	C62max	-1.119	0.104	0.57	0.210	NA	-0.25	0.142	NA	-0.58	0.224	0.0150	-116.275
CDM	**C62rf**	**-1.203**	**0.105**	**0.17**	**0.078**	**NA**	**-0.29**	**0.160**	**NA**	**-0.64**	**0.234**	**0.0016**	**-120.238**
sSDM	C62_aa_glm	-1.246	0.087	0.43	0.059	<0.0001	-0.10	0.118	0.379	NA	NA	0.5033	-154.314
sSDM	**C62_aa_max_tune**	**-1.180**	**0.084**	**0.44**	**0.060**	**<0.0001**	**-0.23**	**0.114**	**0.045**	**NA**	**NA**	**0.8306**	**-154.651**
sSDM	C62_aa_rf	-1.305	0.098	0.38	0.065	<0.0001	0.03	0.134	0.835	NA	NA	0.5010	-142.063
sSDM	C62_ad_glm	-1.258	0.099	0.34	0.067	<0.0001	-0.05	0.135	0.710	NA	NA	0.7116	-133.875
sSDM	C62_ad_max_tune	-1.206	0.099	0.29	0.067	<0.0001	-0.14	0.135	0.319	NA	NA	0.5121	-128.878
sSDM	C62_ad_rf	-1.286	0.111	0.26	0.077	<0.01	0.03	0.157	0.860	NA	NA	0.6059	-122.288
sSDM	C62_ga_glm	-1.263	0.095	0.37	0.062	<0.0001	-0.06	0.130	0.667	NA	NA	0.6096	-140.531
sSDM	C62_ga_max_tune	-1.208	0.096	0.33	0.064	<0.0001	-0.15	0.130	0.262	NA	NA	0.9746	-135.169
sSDM	C62_ga_rf	-1.293	0.100	0.35	0.066	<0.0001	0.02	0.139	0.908	NA	NA	0.6978	-136.186
sSDM	C62_gd_glm	-1.236	0.101	0.30	0.068	<0.0001	-0.08	0.138	0.542	NA	NA	0.7894	-129.398
sSDM	C62_gd_max_tune	-1.195	0.101	0.26	0.068	<0.001	-0.15	0.137	0.274	NA	NA	0.5260	-125.542
sSDM	C62_gd_rf	-1.283	0.111	0.25	0.077	<0.01	0.02	0.158	0.882	NA	NA	0.5519	-121.979
sSDMr	C62_aa_glm_red	-1.113	0.106	-0.12	0.090	NA	-0.21	0.146	NA	0.31	0.142	0.0325	-114.580
sSDMr	C62_aa_max_tune_red	-1.212	0.099	0.32	0.069	<0.0001	-0.14	0.134	0.315	NA	NA	0.9418	-131.267
sSDMr	C62_aa_rf_red	-1.270	0.107	0.27	0.068	<0.001	-0.01	0.149	0.949	NA	NA	0.5269	-125.871
sSDMr	C62_ad_glm_red	-1.127	0.106	-0.11	0.075	0.1221	-0.25	0.146	0.094	NA	NA	0.1766	-114.399
sSDMr	C62_ad_max_tune_red	-1.149	0.107	0.07	0.073	0.3275	-0.20	0.147	0.167	NA	NA	0.4366	-112.968
sSDMr	C62_ad_rf_red	-1.167	0.112	0.07	0.078	0.3856	-0.17	0.157	0.282	NA	NA	0.8700	-112.762
sSDMr	C62_ga_glm_red	-1.166	0.107	0.11	0.064	0.0994	-0.17	0.148	0.240	NA	NA	0.0616	-114.724
sSDMr	**C62_ga_max_tune_red**	**-1.129**	**0.093**	**0.32**	**0.060**	**<0.0001**	**-0.29**	**0.130**	**0.032**	**NA**	**NA**	**0.2385**	**-134.672**
sSDMr	C62_ga_rf_red	-1.229	0.109	0.21	0.071	<0.01	-0.07	0.151	0.626	NA	NA	0.9600	-119.860
sSDMr	C62_gd_glm_red	-1.176	0.106	-0.24	0.162	0.0405	-0.18	0.143	0.216	NA	NA	0.0644	-116.205
sSDMr	C62_gd_max_tune_red	-1.149	0.107	0.06	0.073	0.3968	-0.20	0.147	0.172	NA	NA	0.5370	-112.727
sSDMr	C62_gd_rf_red	-1.181	0.112	0.09	0.077	0.2269	-0.15	0.158	0.353	NA	NA	0.7150	-113.469

Table S22. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community C78 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	C78_glm	-1.424	0.079	0.15	0.047	<0.01	-0.39	0.114	0.0010	NA	NA	0.4760	-186.676
CDM	C78max	-1.418	0.077	0.18	0.052	<0.001	-0.41	0.110	0.0004	NA	NA	0.8335	-189.014
CDM	**C78rf**	**-1.500**	**0.080**	**0.23**	**0.053**	**<0.0001**	**-0.27**	**0.117**	**0.0239**	**NA**	**NA**	**0.3208**	**-193.632**
sSDM	C78_aa_glm	-1.541	0.066	0.35	0.045	<0.0001	-0.23	0.095	0.0164	NA	NA	0.7091	-222.488
sSDM	C78_aa_max_tune	-1.498	0.067	0.44	0.065	NA	-0.36	0.092	NA	-0.21	0.091	0.0207	-221.422
sSDM	**C78_aa_rf**	**-1.575**	**0.066**	**0.37**	**0.045**	**<0.0001**	**-0.18**	**0.095**	**0.0631**	**NA**	**NA**	**0.2027**	**-226.380**
sSDM	C78_ad_glm	-1.424	0.068	0.27	0.046	<0.0001	-0.43	0.098	0.0000	NA	NA	0.0717	-206.309
sSDM	C78_ad_max_tune	-1.455	0.070	0.39	0.072	NA	-0.40	0.099	NA	-0.22	0.099	0.0274	-207.225
sSDM	C78_ad_rf	-1.470	0.067	0.31	0.046	<0.0001	-0.36	0.096	0.0004	NA	NA	0.7971	-214.146
sSDM	C78_ga_glm	-1.444	0.076	0.19	0.044	<0.001	-0.36	0.111	0.0017	NA	NA	0.0519	-193.308
sSDM	C78_ga_max_tune	-1.479	0.069	0.29	0.044	<0.0001	-0.34	0.099	0.0011	NA	NA	0.6006	-210.831
sSDM	C78_ga_rf	-1.510	0.083	0.24	0.055	<0.0001	-0.25	0.122	0.0413	NA	NA	0.5805	-193.784
sSDM	C78_gd_glm	-1.502	0.075	0.27	0.049	<0.0001	-0.28	0.109	0.0142	NA	NA	0.6257	-202.910
sSDM	C78_gd_max_tune	-1.457	0.070	0.39	0.071	NA	-0.40	0.099	NA	-0.22	0.098	0.0260	-207.656
sSDM	C78_gd_rf	-1.524	0.072	0.31	0.047	<0.0001	-0.25	0.104	0.0196	NA	NA	0.8037	-210.946
sSDMr	C78_aa_glm_red	-1.385	0.083	0.02	0.053	0.6909	-0.45	0.122	0.0004	NA	NA	0.0845	-178.331
sSDMr	C78_aa_max_tune_red	-1.423	0.073	0.24	0.055	<0.0001	-0.42	0.105	0.0001	NA	NA	0.2778	-195.897
sSDMr	C78_aa_rf_red	-1.510	0.076	0.27	0.050	<0.0001	-0.26	0.110	0.0187	NA	NA	0.7435	-201.432
sSDMr	C78_ad_glm_red	-1.377	0.081	-0.01	0.058	0.9166	-0.47	0.117	0.0002	NA	NA	0.1046	-178.184
sSDMr	C78_ad_max_tune_red	-1.387	0.072	0.22	0.049	<0.0001	-0.48	0.105	0.0000	NA	NA	0.9954	-195.834
sSDMr	C78_ad_rf_red	-1.446	0.077	0.20	0.051	<0.001	-0.37	0.111	0.0015	NA	NA	0.2926	-190.671
sSDMr	C78_ga_glm_red	-1.449	0.078	0.18	0.046	<0.001	-0.35	0.114	0.0028	NA	NA	0.1620	-191.255
sSDMr	**C78_ga_max_tune_red**	**-1.503**	**0.073**	**0.28**	**0.046**	**<0.0001**	**-0.28**	**0.106**	**0.0097**	**NA**	**NA**	**0.8424**	**-206.177**
sSDMr	C78_ga_rf_red	-1.391	0.081	0.07	0.055	0.2377	-0.44	0.118	0.0003	NA	NA	0.4282	-179.567
sSDMr	C78_gd_glm_red	-1.370	0.080	0.09	0.049	0.1057	-0.49	0.116	0.0001	NA	NA	0.1349	-180.790
sSDMr	C78_gd_max_tune_red	-1.378	0.078	-0.04	0.057	NA	-0.38	0.115	NA	1.02	0.373	0.0134	-182.499
sSDMr	C78_gd_rf_red	-1.450	0.074	0.22	0.047	<0.0001	-0.37	0.107	0.0010	NA	NA	0.2886	-195.792

Table S23. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community C79 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	C79_glm	-1.185	0.105	-0.07	0.068	0.2835	0.06	0.143	0.699	NA	NA	0.0578	-113.677
CDM	**C79max**	**-1.258**	**0.097**	**0.26**	**0.064**	**<0.001**	**0.16**	**0.128**	**0.210**	**NA**	**NA**	**0.8522**	**-127.397**
CDM	C79rf	-1.281	0.103	0.20	0.065	<0.01	0.22	0.139	0.118	NA	NA	0.5743	-120.996
sSDM	C79_aa_glm	-1.284	0.087	0.37	0.057	<0.0001	0.18	0.114	0.110	NA	NA	0.8772	-147.242
sSDM	C79_aa_max_tune	-1.184	0.079	0.41	0.056	<0.0001	-0.02	0.107	0.876	NA	NA	0.1290	-156.405
sSDM	C79_aa_rf	-1.318	0.086	0.40	0.056	<0.0001	0.23	0.113	0.039	NA	NA	0.5682	-151.588
sSDM	C79_ad_glm	-1.239	0.085	0.37	0.057	<0.0001	0.10	0.112	0.389	NA	NA	0.6579	-147.756
sSDM	**C79_ad_max_tune**	**-1.131**	**0.076**	**0.45**	**0.056**	**<0.0001**	**-0.13**	**0.106**	**0.241**	**NA**	**NA**	**0.1265**	**-161.141**
sSDM	C79_ad_rf	-1.246	0.086	0.37	0.057	<0.0001	0.10	0.113	0.358	NA	NA	0.5770	-146.634
sSDM	C79_ga_glm	-1.313	0.100	0.28	0.064	<0.0001	0.26	0.134	0.051	NA	NA	0.8251	-128.964
sSDM	C79_ga_max_tune	-1.225	0.085	0.35	0.056	<0.0001	0.08	0.114	0.492	NA	NA	0.1316	-143.962
sSDM	C79_ga_rf	-1.314	0.094	0.33	0.062	<0.0001	0.25	0.125	0.046	NA	NA	0.3378	-136.225
sSDM	C79_gd_glm	-1.238	0.083	0.38	0.055	<0.0001	0.09	0.110	0.416	NA	NA	0.3755	-149.722
sSDM	C79_gd_max_tune	-1.136	0.077	0.44	0.056	<0.0001	-0.11	0.107	0.285	NA	NA	0.1413	-160.188
sSDM	C79_gd_rf	-1.249	0.086	0.37	0.056	<0.0001	0.11	0.113	0.329	NA	NA	0.5715	-146.064
sSDMr	C79_aa_glm_red	-1.237	0.105	0.10	0.066	0.1365	0.15	0.140	0.296	NA	NA	0.2661	-114.743
sSDMr	**C79_aa_max_tune_red**	**-1.224**	**0.095**	**0.26**	**0.063**	**<0.0001**	**0.10**	**0.126**	**0.438**	**NA**	**NA**	**0.8085**	**-128.327**
sSDMr	C79_aa_rf_red	-1.296	0.102	0.23	0.063	<0.001	0.24	0.137	0.080	NA	NA	0.5741	-124.534
sSDMr	C79_ad_glm_red	-1.209	0.104	-0.05	0.066	0.4877	0.10	0.138	0.469	NA	NA	0.1633	-113.008
sSDMr	C79_ad_max_tune_red	-1.190	0.104	0.09	0.064	0.1892	0.06	0.140	0.655	NA	NA	0.2313	-114.251
sSDMr	C79_ad_rf_red	-1.226	0.103	0.13	0.067	0.0536	0.12	0.136	0.369	NA	NA	0.8720	-116.251
sSDMr	C79_ga_glm_red	-1.248	0.105	0.13	0.063	0.0572	0.16	0.141	0.242	NA	NA	0.6540	-116.143
sSDMr	C79_ga_max_tune_red	-1.221	0.096	0.24	0.060	<0.001	0.10	0.128	0.436	NA	NA	0.0925	-125.778
sSDMr	C79_ga_rf_red	-1.246	0.105	0.13	0.066	0.0600	0.16	0.140	0.253	NA	NA	0.8840	-116.063
sSDMr	C79_gd_glm_red	-1.017	0.106	0.76	0.210	NA	-0.13	0.137	NA	-0.67	0.219	0.0035	-123.589
sSDMr	C79_gd_max_tune_red	-1.101	0.104	0.71	0.236	NA	-0.03	0.136	NA	-0.61	0.244	0.0172	-120.433
sSDMr	C79_gd_rf_red	-1.204	0.101	0.16	0.064	0.0196	0.08	0.134	0.552	NA	NA	0.6002	-117.972

Table S24. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community H25 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	H25_glm	-2.108	0.117	0.26	0.060	<0.001	0.28	0.148	0.057	0.02	NA	0.8963	-184.728
CDM	**H25max**	**-2.198**	**0.100**	**0.44**	**0.053**	**<0.0001**	**0.35**	**0.125**	**0.006**	**-0.01**	**NA**	**0.9066**	**-212.928**
CDM	H25rf	-2.135	0.109	0.34	0.056	<0.0001	0.30	0.138	0.030	-0.06	NA	0.6798	-197.541
sSDM	H25_aa_glm	-2.151	0.104	0.41	0.057	<0.0001	0.28	0.130	0.032	0.05	NA	0.6698	-205.850
sSDM	**H25_aa_max_tune**	**-2.174**	**0.091**	**0.50**	**0.049**	**<0.0001**	**0.27**	**0.115**	**0.020**	**0.03**	**NA**	**0.7919**	**-227.763**
sSDM	H25_aa_rf	-2.198	0.107	0.42	0.061	<0.0001	0.37	0.131	0.006	0.14	NA	0.2632	-204.280
sSDM	H25_ad_glm	-2.163	0.112	0.34	0.057	<0.0001	0.34	0.139	0.016	0.11	NA	0.3246	-195.165
sSDM	H25_ad_max_tune	-2.126	0.097	0.46	0.051	<0.0001	0.21	0.123	0.094	0.03	NA	0.7469	-217.436
sSDM	H25_ad_rf	-2.154	0.104	0.42	0.058	<0.0001	0.28	0.130	0.031	0.03	NA	0.7972	-206.531
sSDM	H25_ga_glm	-2.124	0.112	0.31	0.055	<0.0001	0.29	0.141	0.042	-0.03	NA	0.8051	-192.392
sSDM	H25_ga_max_tune	-2.156	0.101	0.43	0.051	<0.0001	0.28	0.127	0.028	0.00	NA	0.9639	-211.405
sSDM	H25_ga_rf	-2.196	0.124	0.26	0.071	<0.001	0.44	0.154	0.005	0.10	NA	0.4850	-181.393
sSDM	H25_gd_glm	-2.135	0.111	0.34	0.055	<0.0001	0.29	0.139	0.036	0.01	NA	0.9097	-195.203
sSDM	H25_gd_max_tune	-2.153	0.101	0.43	0.051	<0.0001	0.27	0.126	0.032	0.03	NA	0.7589	-212.041
sSDM	H25_gd_rf	-2.161	0.102	0.42	0.055	<0.0001	0.30	0.128	0.021	-0.02	NA	0.8936	-209.434
sSDMr	H25_aa_glm_red	-2.150	0.110	0.37	0.059	<0.0001	0.31	0.137	0.026	0.07	NA	0.5662	-197.674
sSDMr	**H25_aa_max_tune_red**	**-2.188**	**0.091**	**0.50**	**0.049**	**<0.0001**	**0.29**	**0.114**	**0.011**	**0.04**	**NA**	**0.6987**	**-228.780**
sSDMr	H25_aa_rf_red	-2.206	0.110	0.39	0.061	<0.0001	0.39	0.135	0.004	0.16	NA	0.1791	-200.609
sSDMr	H25_ad_glm_red	-2.158	0.115	0.31	0.058	<0.0001	0.34	0.143	0.017	0.13	NA	0.2671	-190.646
sSDMr	H25_ad_max_tune_red	-2.105	0.097	0.46	0.053	<0.0001	0.17	0.125	0.172	0.02	NA	0.8664	-216.645
sSDMr	H25_ad_rf_red	-2.163	0.104	0.40	0.056	<0.0001	0.31	0.131	0.018	-0.01	NA	0.9457	-205.092
sSDMr	H25_ga_glm_red	-2.128	0.113	0.30	0.054	<0.0001	0.30	0.143	0.035	-0.01	NA	0.9470	-190.859
sSDMr	H25_ga_max_tune_red	-2.159	0.100	0.43	0.051	<0.0001	0.28	0.126	0.026	0.01	NA	0.8998	-212.120
sSDMr	H25_ga_rf_red	-2.196	0.124	0.26	0.071	<0.001	0.44	0.154	0.005	0.10	NA	0.4846	-181.374
sSDMr	H25_gd_glm_red	-2.131	0.112	0.32	0.055	<0.0001	0.30	0.141	0.038	0.02	NA	0.8480	-192.608
sSDMr	H25_gd_max_tune_red	-2.133	0.100	0.43	0.052	<0.0001	0.23	0.127	0.067	0.02	NA	0.8445	-212.354
sSDMr	H25_gd_rf_red	-2.181	0.103	0.41	0.055	<0.0001	0.34	0.129	0.009	0.00	NA	0.9784	-207.059

Table S25. Characteristics of all CDMs, s-SDMs, and reduced s-SDMs for community H27 based on GLMs relating measured community similarity to the prediction of the DM. Columns as described in Table S19.

Model	DM Name	Intercept	Intercept SE	DM Coefficient	DM SE	DM p	Burn Coefficient	Burn SE	Burn p	DM x Burn Coefficient	DM x Burn SE	DM x Burn p	AIC
CDM	H27_glm	-1.945	0.101	0.09	0.059	0.1550	0.31	0.128	0.016	0.13	NA	0.2700	-182.083
CDM	**H27max**	**-1.946**	**0.083**	**0.33**	**0.049**	**<0.0001**	**0.24**	**0.107**	**0.023**	**0.01**	**NA**	**0.8970**	**-211.657**
CDM	H27rf	-1.962	0.100	0.13	0.061	0.0484	0.34	0.128	0.009	0.06	NA	0.6115	-183.956
sDM	H27_aa_glm	-1.966	0.084	0.31	0.048	<0.0001	0.29	0.108	0.008	-0.02	NA	0.8027	-209.751
sSDM	**H27_aa_max_tune**	**-1.960**	**0.074**	**0.38**	**0.042**	**<0.0001**	**0.25**	**0.095**	**0.011**	**0.00**	**NA**	**0.9572**	**-230.642**
sSDM	H27_aa_rf	-1.992	0.088	0.30	0.052	<0.0001	0.34	0.111	0.003	0.10	NA	0.3454	-206.054
sSDM	H27_ad_glm	-1.920	0.085	0.29	0.047	<0.0001	0.22	0.111	0.054	-0.13	NA	0.1938	-206.788
sSDM	H27_ad_max_tune	-1.912	0.088	0.27	0.049	<0.0001	0.21	0.115	0.075	-0.06	NA	0.5402	-202.224
sSDM	H27_ad_rf	-1.952	0.090	0.28	0.054	<0.0001	0.27	0.114	0.018	0.08	NA	0.4510	-200.839
sSDM	H27_ga_glm	-1.958	0.096	0.17	0.047	<0.01	0.32	0.123	0.010	0.07	NA	0.4777	-189.671
sSDM	H27_ga_max_tune	-1.965	0.083	0.32	0.044	<0.0001	0.28	0.105	0.008	0.02	NA	0.7781	-214.096
sSDM	H27_ga_rf	-1.930	0.096	-0.19	0.112	NA	0.30	0.123	NA	0.36	0.133	0.0063	-185.692
sSDM	H27_gd_glm	-1.920	0.085	0.29	0.047	<0.0001	0.21	0.110	0.055	-0.13	NA	0.1934	-207.677
sSDM	H27_gd_max_tune	-1.910	0.089	0.26	0.049	<0.0001	0.21	0.115	0.077	-0.06	NA	0.5371	-201.443
sSDM	H27_gd_rf	-1.952	0.089	0.28	0.053	<0.0001	0.27	0.114	0.017	0.09	NA	0.4255	-201.365
sSDMr	H27_aa_glm_red	-1.966	0.097	0.19	0.054	<0.01	0.33	0.122	0.008	0.20	NA	0.0656	-190.250
sSDMr	**H27_aa_max_tune_red**	**-1.970**	**0.076**	**0.38**	**0.044**	**<0.0001**	**0.27**	**0.097**	**0.007**	**0.05**	**NA**	**0.5991**	**-227.453**
sSDMr	H27_aa_rf_red	-1.997	0.092	0.26	0.053	<0.0001	0.36	0.116	0.002	0.14	NA	0.2026	-199.210
sSDMr	H27_ad_glm_red	-1.913	0.094	0.20	0.051	<0.001	0.23	0.122	0.060	-0.02	NA	0.8292	-191.943
sSDMr	H27_ad_max_tune_red	-1.909	0.081	0.33	0.046	<0.0001	0.18	0.106	0.092	-0.07	NA	0.4888	-215.634
sSDMr	H27_ad_rf_red	-1.966	0.090	0.27	0.053	<0.0001	0.30	0.114	0.009	0.03	NA	0.7816	-201.024
sSDMr	H27_ga_glm_red	-1.951	0.097	0.16	0.048	<0.01	0.31	0.124	0.013	0.01	NA	0.8931	-187.871
sSDMr	H27_ga_max_tune_red	-1.984	0.089	0.27	0.045	<0.0001	0.34	0.113	0.003	0.13	NA	0.1672	-204.533
sSDMr	H27_ga_rf_red	-1.932	0.098	-0.12	0.100	NA	0.30	0.126	NA	0.27	0.128	0.0327	-182.754
sSDMr	H27_gd_glm_red	-1.790	0.104	0.63	0.221	NA	0.11	0.130	NA	-0.50	0.228	0.0430	-189.312
sSDMr	H27_gd_max_tune_red	-1.915	0.086	0.29	0.047	<0.0001	0.21	0.112	0.067	-0.06	NA	0.5482	-206.453
sSDMr	H27_gd_rf_red	-1.972	0.091	0.26	0.053	<0.0001	0.32	0.115	0.007	0.03	NA	0.7571	-199.345

Table S26. Tukey's HSD post-hoc results for pairwise comparisons between model types (CDM, s-SDM, and s-SDM reduced) within each community, showing estimated differences, standard errors, degrees of freedom, *t*-ratios, and *p*-values. Significant differences are bolded.

Community Code	Contrast	Estimate	SE	df	t.ratio	p.value
C60	CDM-sSDM	-0.144	0.076	150	-1.906	0.1405
C60	CDM-sSDMr	-0.074	0.076	150	-0.977	0.5927
C60	sSDM-sSDMr	0.070	0.048	150	1.469	0.3085
C61	**CDM-sSDM**	**-0.287**	**0.076**	**150**	**-3.794**	**<0.001**
C61	CDM-sSDMr	-0.128	0.076	150	-1.689	0.2129
C61	**sSDM-sSDMr**	**0.159**	**0.048**	**150**	**3.328**	**<0.01**
C62	CDM-sSDM	-0.115	0.076	150	-1.515	0.287
C62	CDM-sSDMr	0.105	0.076	150	1.391	0.3481
C62	**sSDM-sSDMr**	**0.220**	**0.048**	**150**	**4.594**	**<.0001**
C78	CDM-sSDM	-0.131	0.076	150	-1.734	0.1958
C78	CDM-sSDMr	-0.043	0.076	150	-0.573	0.8348
C78	sSDM-sSDMr	0.088	0.048	150	1.836	0.1613
C79	**CDM-sSDM**	**-0.248**	**0.076**	**150**	**-3.279**	**<0.01**
C79	CDM-sSDMr	-0.112	0.076	150	-1.48	0.3033
C79	**sSDM-sSDMr**	**0.136**	**0.048**	**150**	**2.845**	**<0.05**
H25	CDM-sSDM	-0.046	0.076	150	-0.612	0.8138
H25	CDM-sSDMr	-0.033	0.076	150	-0.44	0.8988
H25	sSDM-sSDMr	0.013	0.048	150	0.271	0.9602
H27	CDM-sSDM	-0.096	0.076	150	-1.274	0.4123
H27	CDM-sSDMr	-0.102	0.076	150	-1.346	0.3723
H27	sSDM-sSDMr	-0.005	0.048	150	-0.114	0.9929

Table S27. Tukey's HSD post-hoc results for pairwise comparisons among algorithms (GLM, MaxEnt, RF) within each model type (CDM, s-SDM, and s-SDM reduced), showing estimated differences, standard errors, degrees of freedom, *t*-ratios, and *p*-values for the interaction between model type and algorithm. Significant differences are bolded.

Model Type	Contrast	Estimate	SE	df	t.ratio	p.value
CDM	**GLM-MaxEnt**	**-0.248**	**0.063**	**150**	**-3.956**	**<0.001**
CDM	GLM-RF	-0.135	0.063	150	-2.164	0.0808
CDM	MaxEnt-RF	0.112	0.063	150	1.792	0.1756
s-SDM	GLM-MaxEnt	-0.038	0.031	150	-1.198	0.4560
s-SDM	GLM-RF	0.010	0.031	150	0.311	0.9479
s-SDM	MaxEnt-RF	0.047	0.031	150	1.51	0.2892
s-SDMr	**GLM-MaxEnt**	**-0.129**	**0.031**	**150**	**-4.124**	**<0.001**
s-SDMr	GLM-RF	-0.029	0.031	150	-0.921	0.6279
s-SDMr	**MaxEnt-RF**	**0.100**	**0.031**	**150**	**3.203**	**<0.01**

Table S28. Counts and proportions of the park predicted to be suitable according to the best performing model of each type (CDM, s-SDM, and s-SDM reduced) for the seven target vegetation communities in Waterton Lakes National Park (number of cells = 203,470). Predicted probability thresholds were defined as the lowest probability among cells corresponding to known occurrences of each community in each model.

Community Code	CDM Predicted Probability Threshold	CDM Cells Predicted Suitable	CDM Proportion of Area Suitable	s-SDM Predicted Probability Threshold	s-SDM Cells Predicted Suitable	s-SDM Proportion of Area Suitable	s-SDMr Predicted Probability Threshold	s-SDMr Cells Predicted Suitable	s-SDMr Proportion of Area Suitable
C60	0.031	26712	0.131	0.184	79482	0.391	0.007	172787	0.849
C61	0.677	2913	0.014	0.351	26786	0.132	0.162	34367	0.169
C62	0.942	5	0.000	0.645	7749	0.038	0.020	100168	0.492
C78	0.910	6	0.000	0.257	42839	0.211	0.020	48209	0.237
C79	0.756	23852	0.117	0.721	37024	0.182	0.453	41184	0.202
H25	0.046	93786	0.461	0.153	94409	0.464	0.136	83526	0.411
H27	0.257	119927	0.589	0.130	99768	0.490	0.156	102113	0.502

Table S29. Counts and proportions of cells predicted to be suitable across the best performing model of each model type (CDM, s-SDM, and s-SDM reduced) for the seven target vegetation communities in Waterton Lakes National Park (number of cells = 203,470). Bray-Curtis similarity thresholds were defined as the lowest pairwise similarity between two known ELC communities classified as the same type.

Community Code	CDM Bray-Curtis Similarity Threshold	CDM Cells Predicted Suitable	CDM Proportion of Suitable Area	s-SDM Bray-Curtis Similarity Threshold	s-SDM Cells Predicted Suitable	s-SDM Proportion of Area Suitable	s-SDMr Bray-Curtis Similarity Threshold	s-SDMr Cells Predicted Suitable	s-SDMr Proportion of Area Suitable
C60	0.120	203470	1.000	0.120	141245	0.694	0.120	158435	0.120
C61	0.240	20281	0.100	0.240	21636	0.106	0.240	27462	0.240
C62	0.322	567	0.003	0.322	3639	0.018	0.322	4153	0.322
C78	0.240	2565	0.013	0.240	3411	0.017	0.240	2171	0.240
C79	0.359	0	0.000	0.359	2376	0.012	0.359	55	0.359
H25	0.163	13751	0.068	0.163	14544	0.071	0.163	14256	0.163
H27	0.110	137196	0.674	0.110	81833	0.402	0.110	91677	0.110

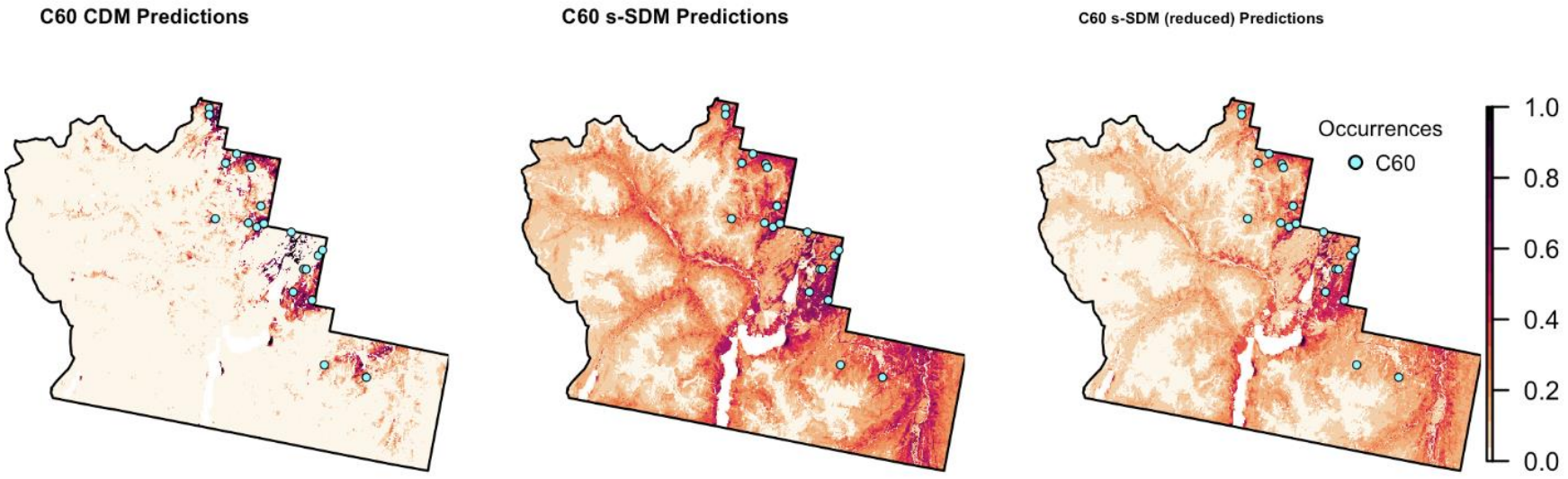

Figure S1. Scaled predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C60. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C60 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

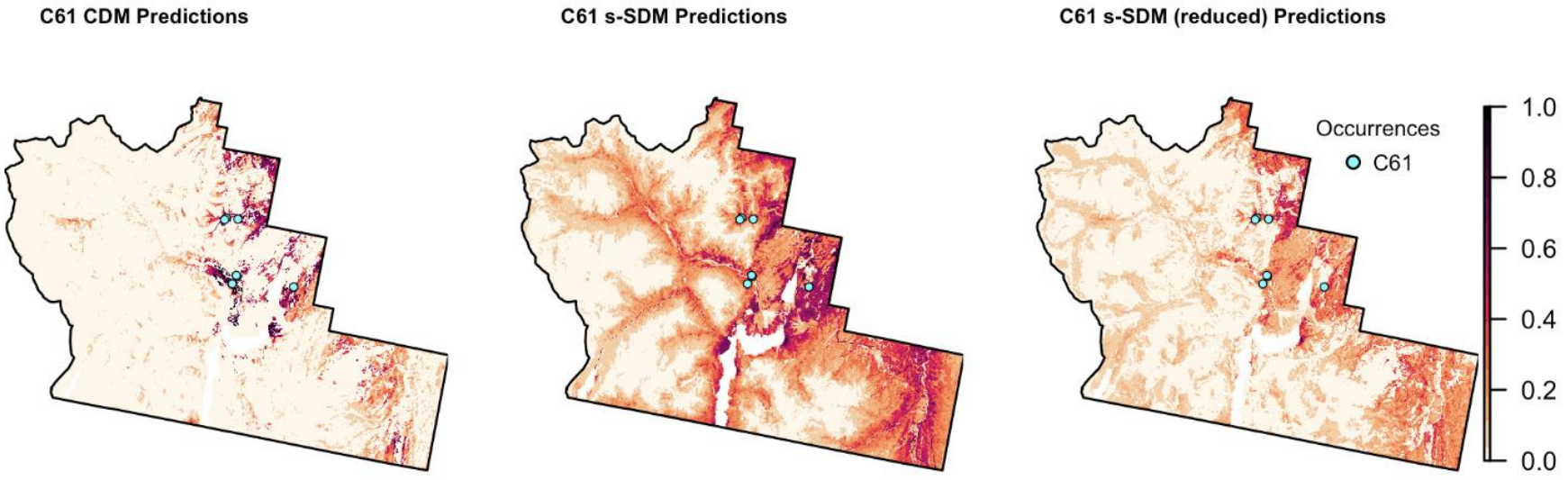

Figure S2. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C61. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C61 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

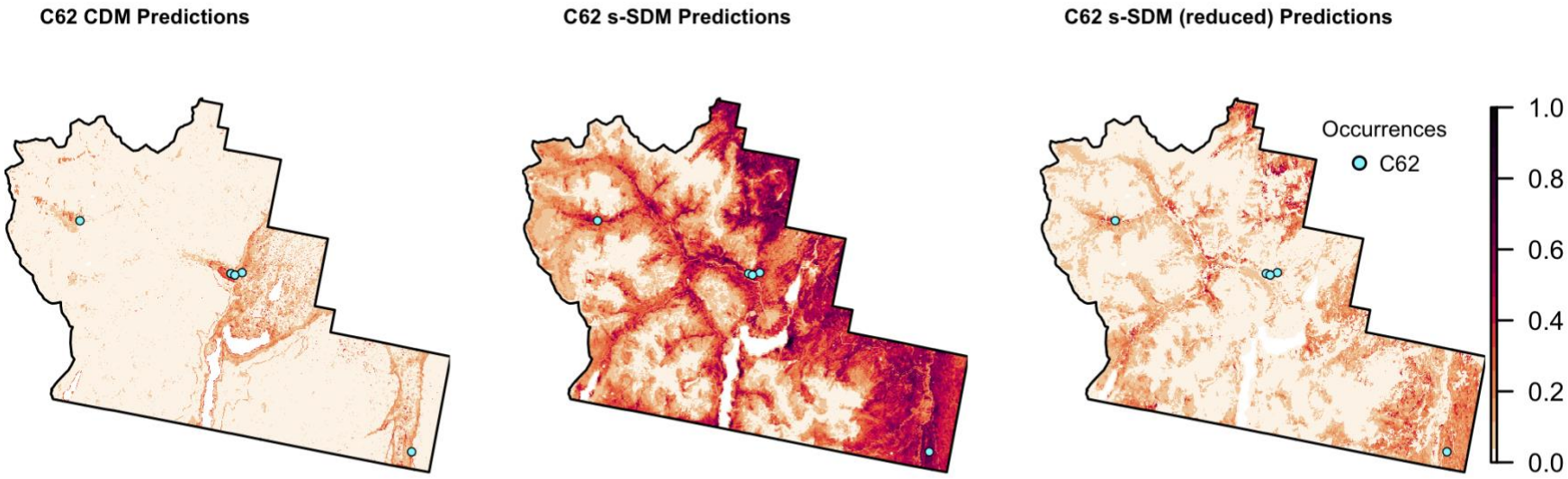

Figure S3. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C62. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C62 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

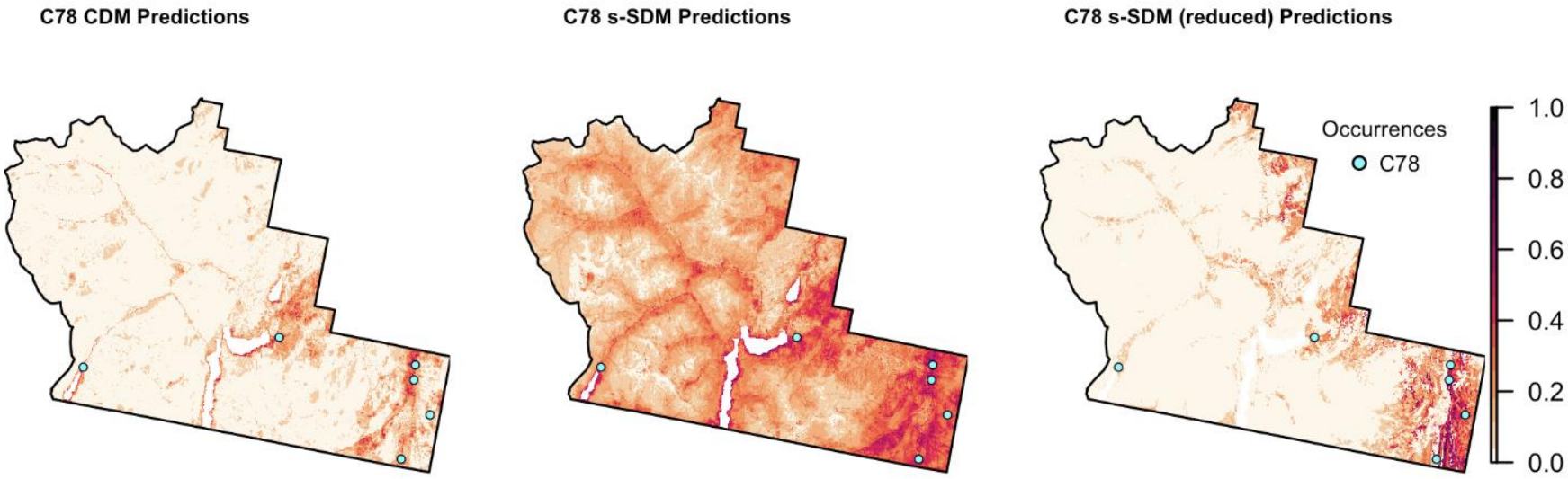

Figure S4. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C78. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C78 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

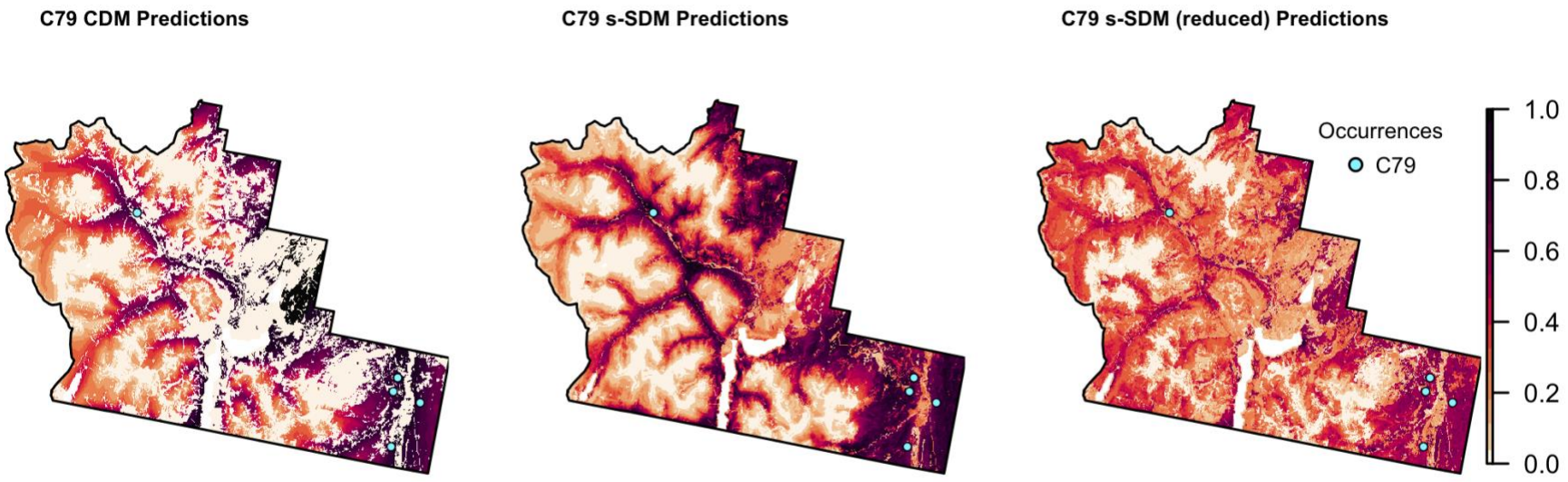

Figure S5. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C79. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community C79 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

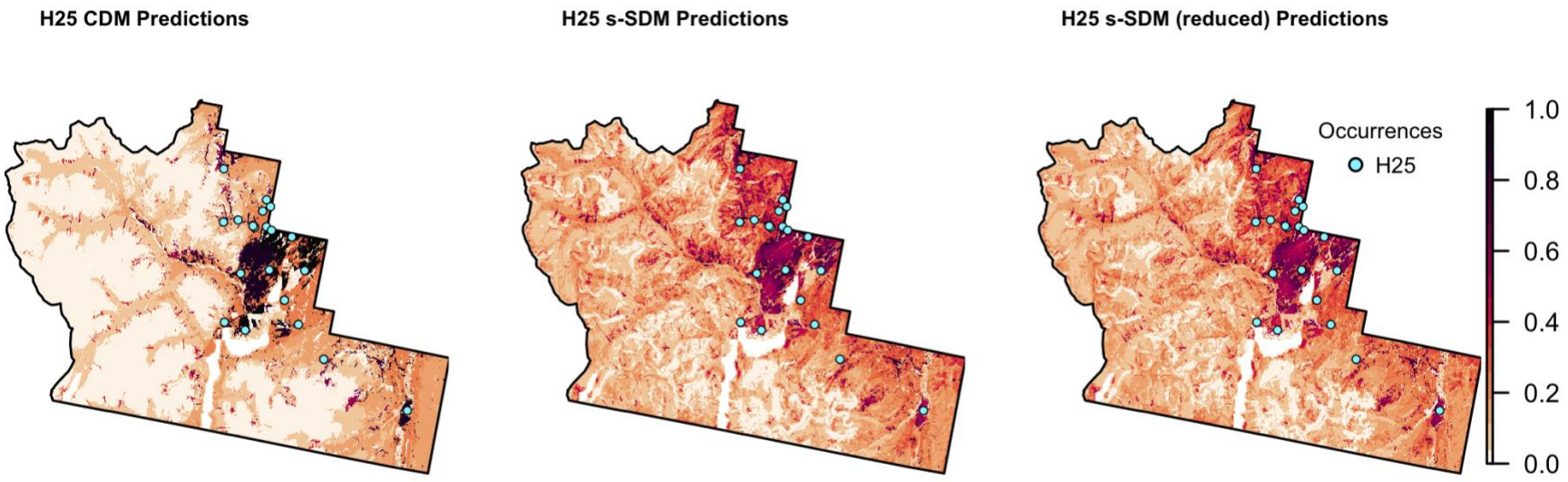

Figure S6. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community H25. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community H25 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

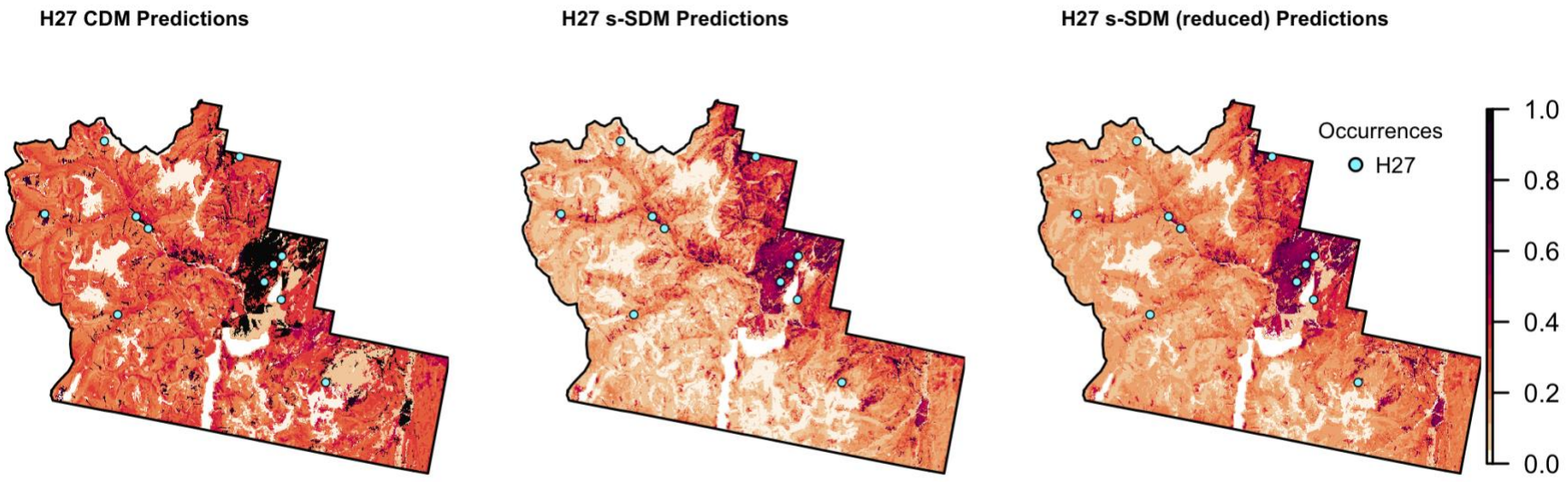

Figure S7. Predicted probability distributions by the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community H27. Higher likelihood of occurrence is shown with darker colours (values closer to 1) and lower probability of occurrence is shown with lighter colours (values closer to 0). Previously known occurrences classified as community H27 from the ELC or additional known presences used to build the CDMs are shown on each map (blue points). Specifics of each best model (algorithm, species included, stacking method, etc.) can be found in Table 3.

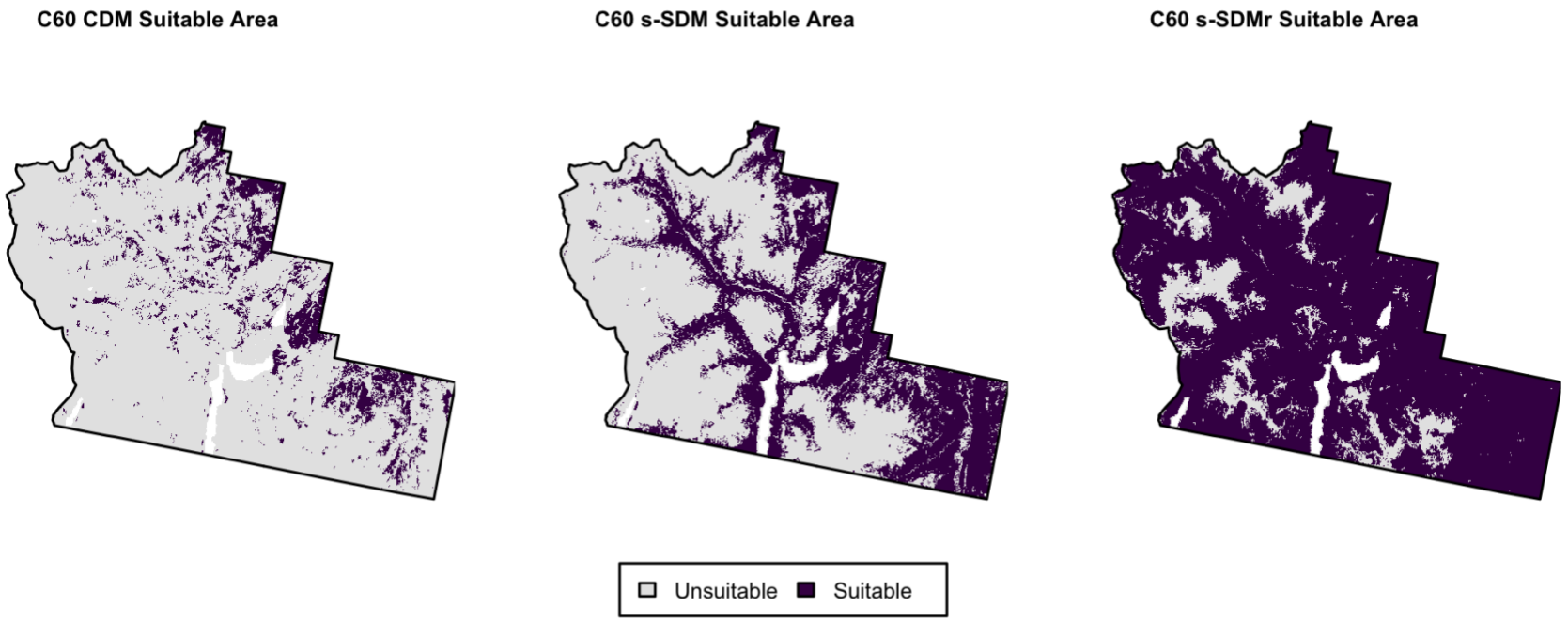

Figure S8. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C60. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known C60 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

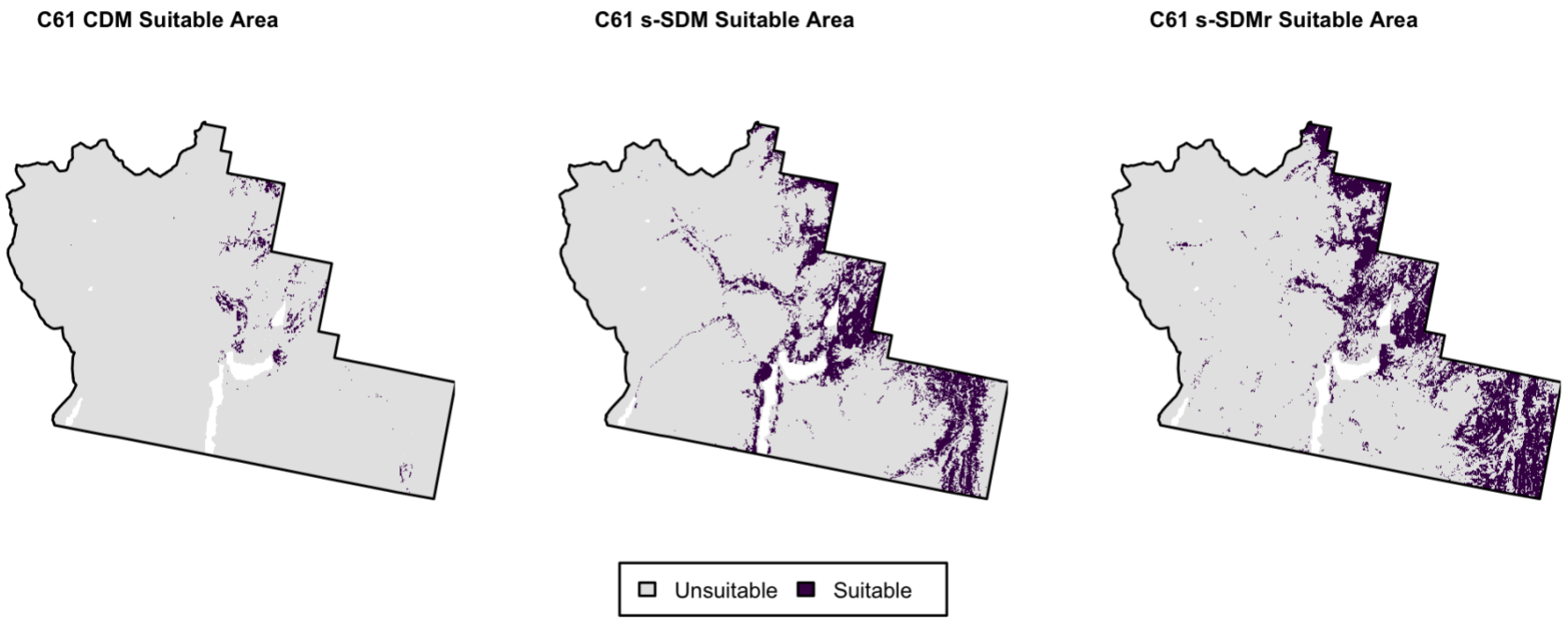

Figure S9. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C61. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known C61 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

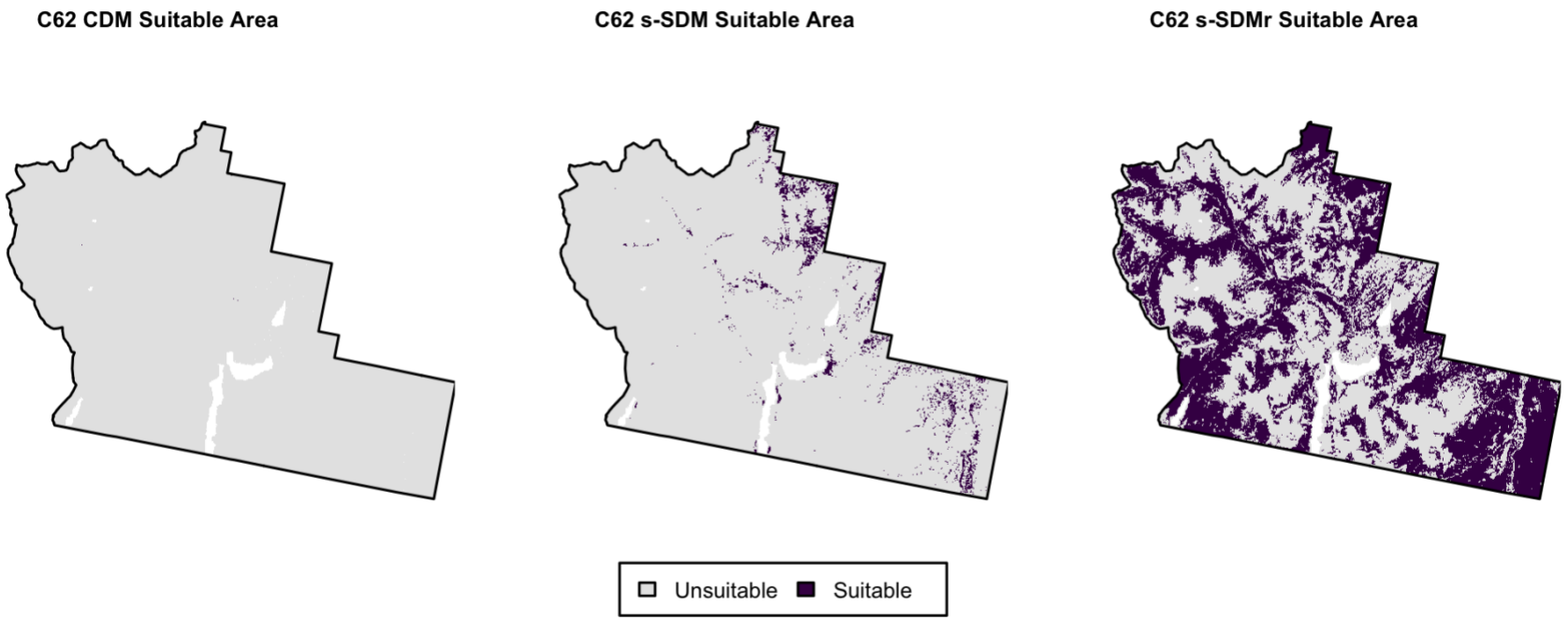

Figure S10. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C62. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known C62 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

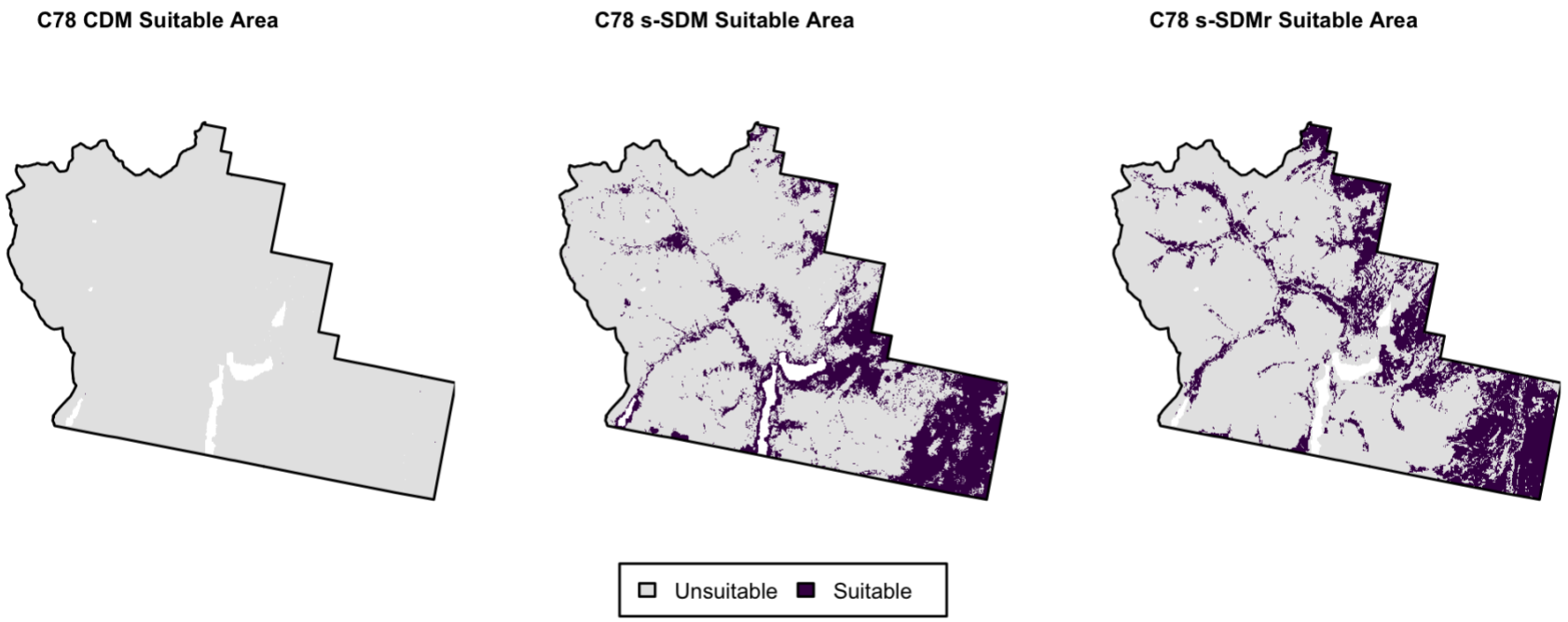

Figure S11. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C78. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known C78 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

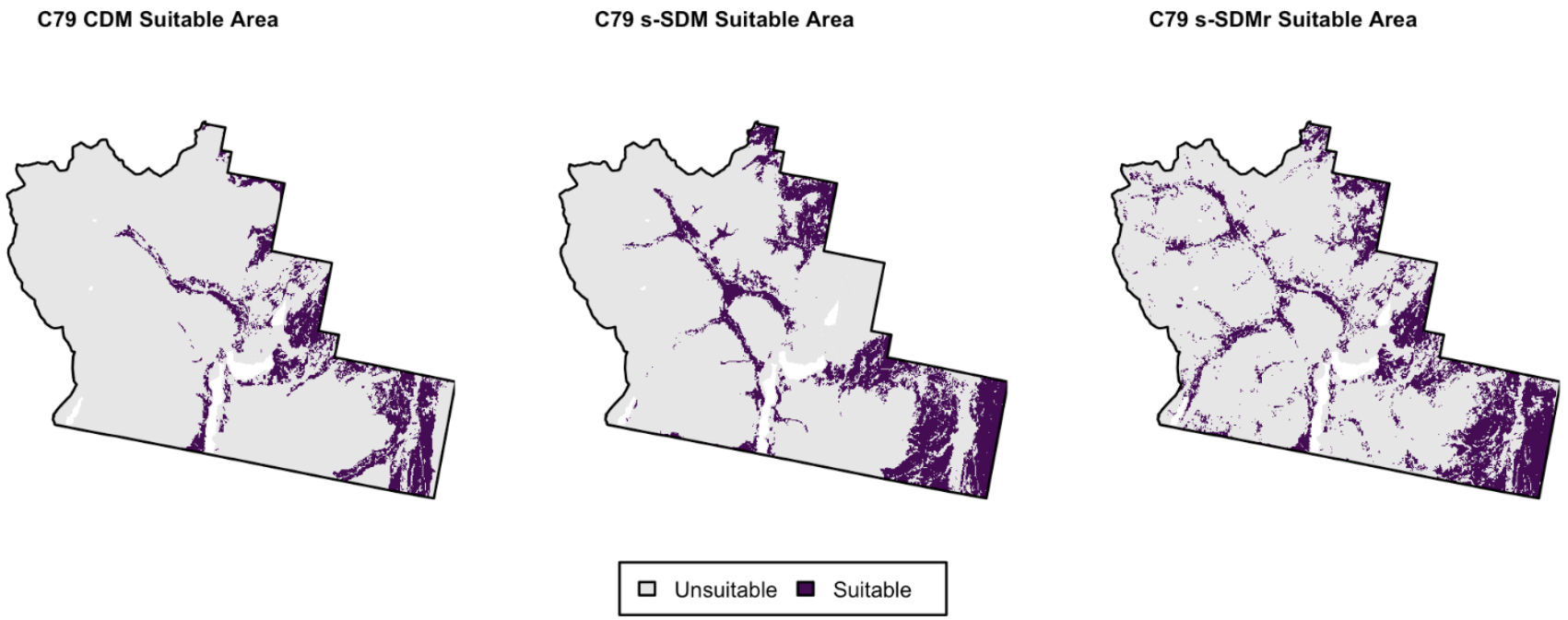

Figure S12. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community C79. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known C79 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

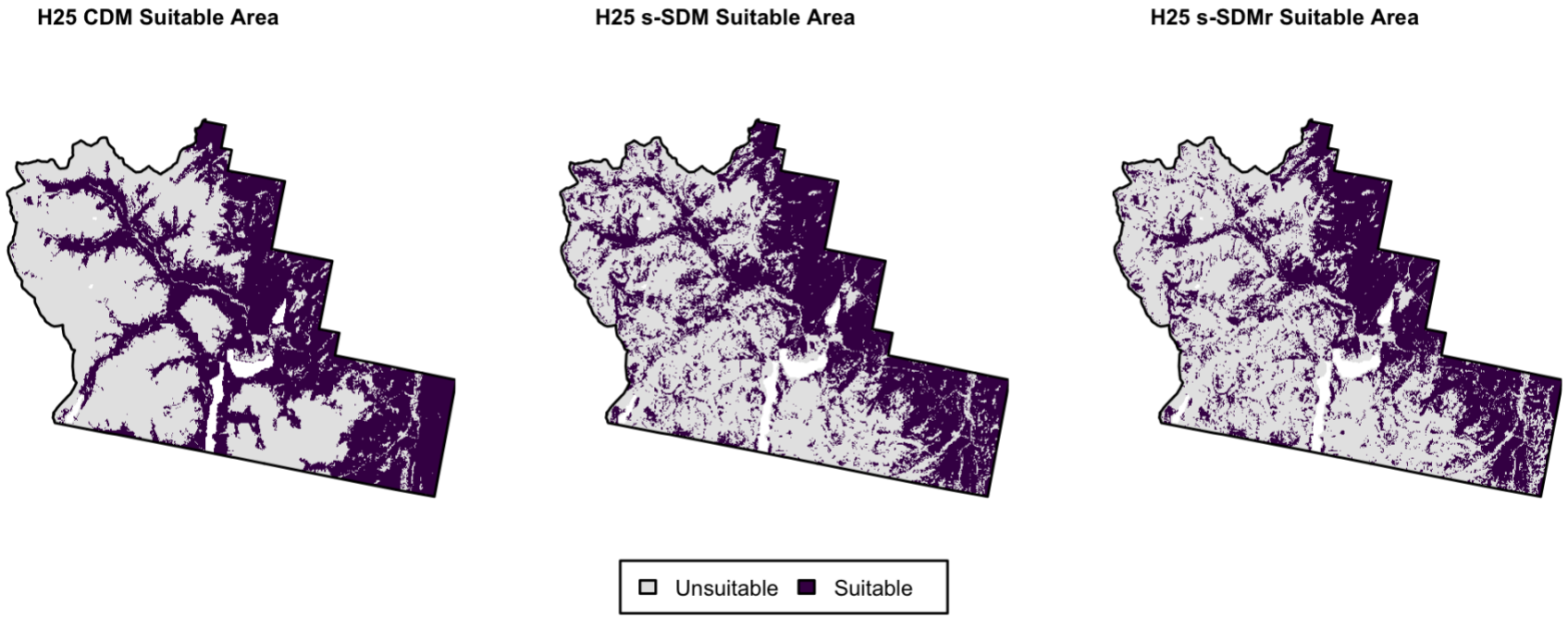

Figure S13. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community H25. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known H25 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

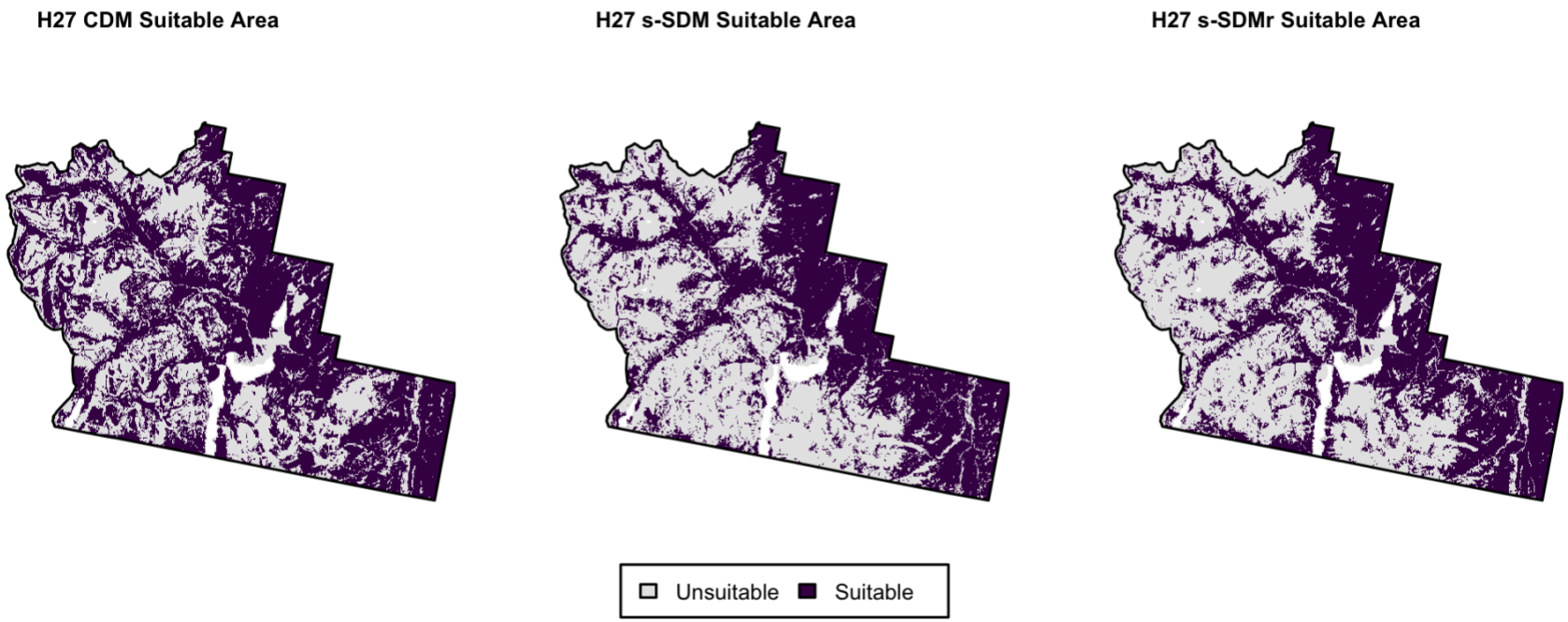

Figure S14. Maps showing predicted suitable and unsuitable areas of occurrence for the best model of each type (CDM, s-SDM, and s-SDM (reduced)) for community H27. Suitability was defined using a model-specific threshold equal to the lowest predicted probability value at a known H27 occurrence. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S28.

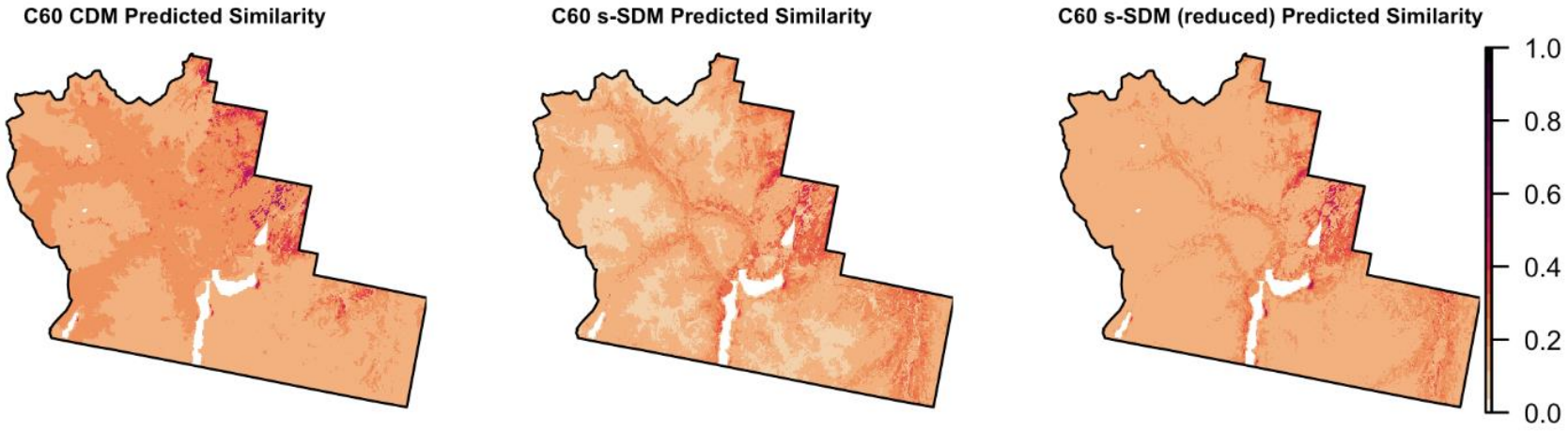

Figure S15. Maps showing predicted Bray-Curtis similarity across the park for community C60 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known C60 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

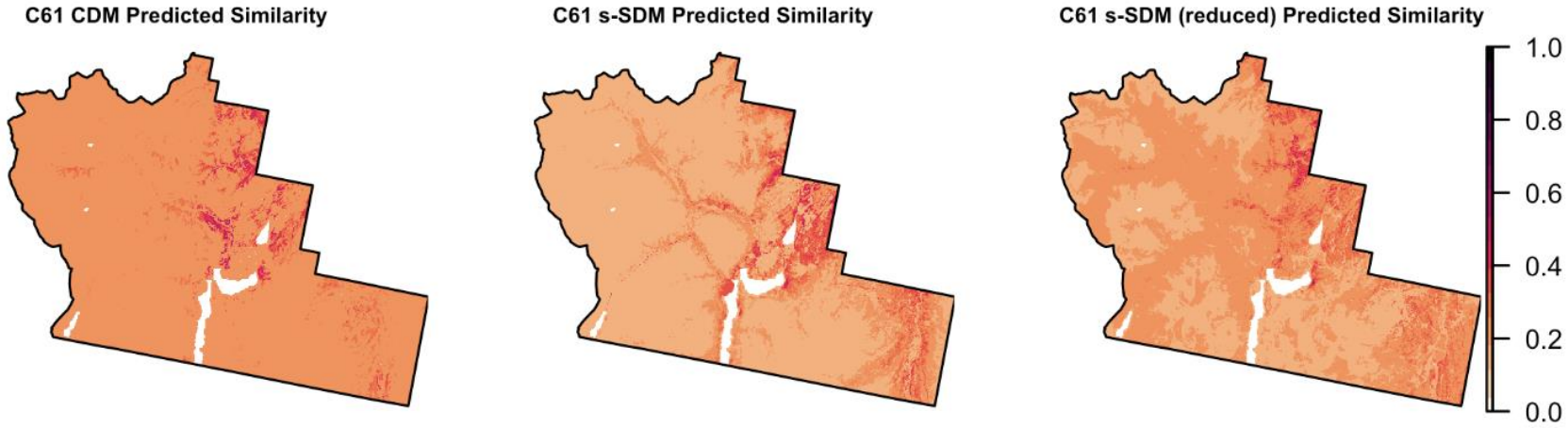

Figure S16. Maps showing predicted Bray-Curtis similarity across the park for community C61 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known C61 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

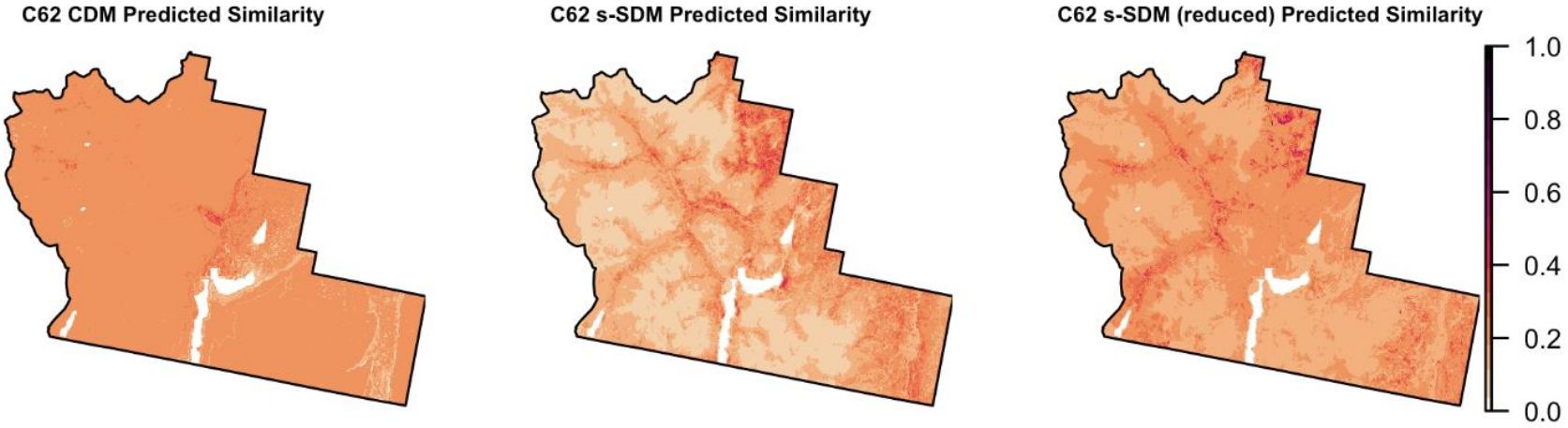

Figure S17. Maps showing predicted Bray-Curtis similarity across the park for community C62 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known C62 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

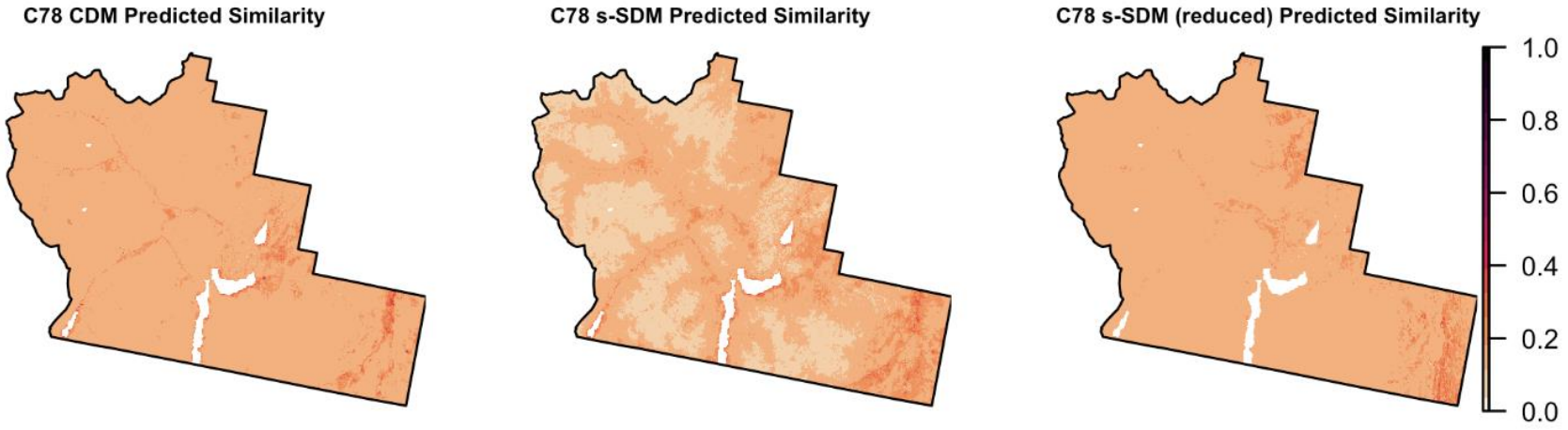

Figure S18. Maps showing predicted Bray-Curtis similarity across the park for community C78 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known C78 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

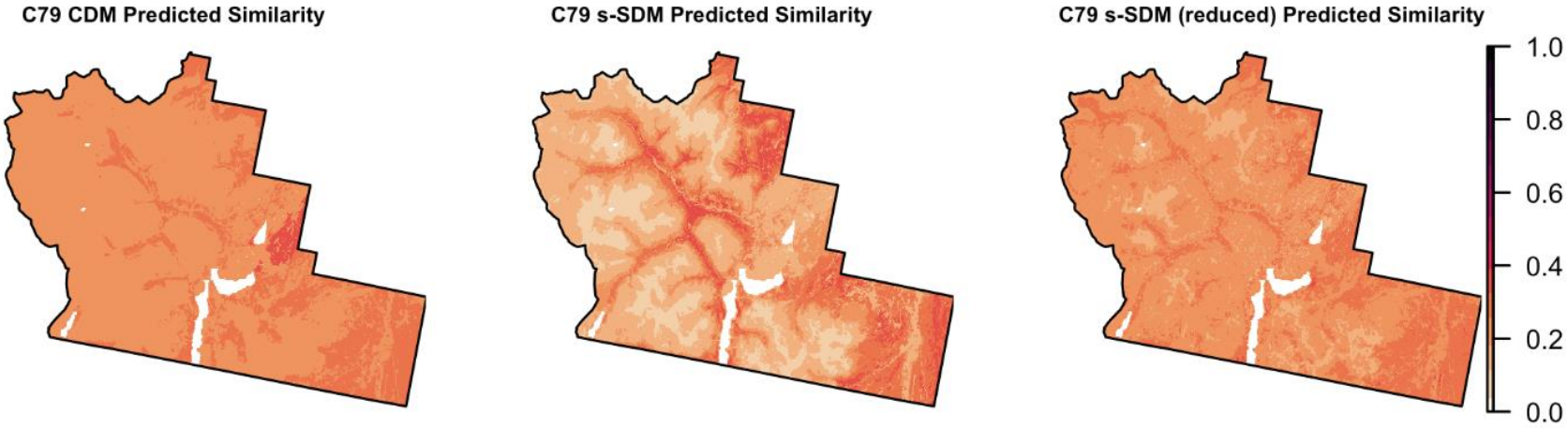

Figure S19. Maps showing predicted Bray-Curtis similarity across the park for community C79 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known C79 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

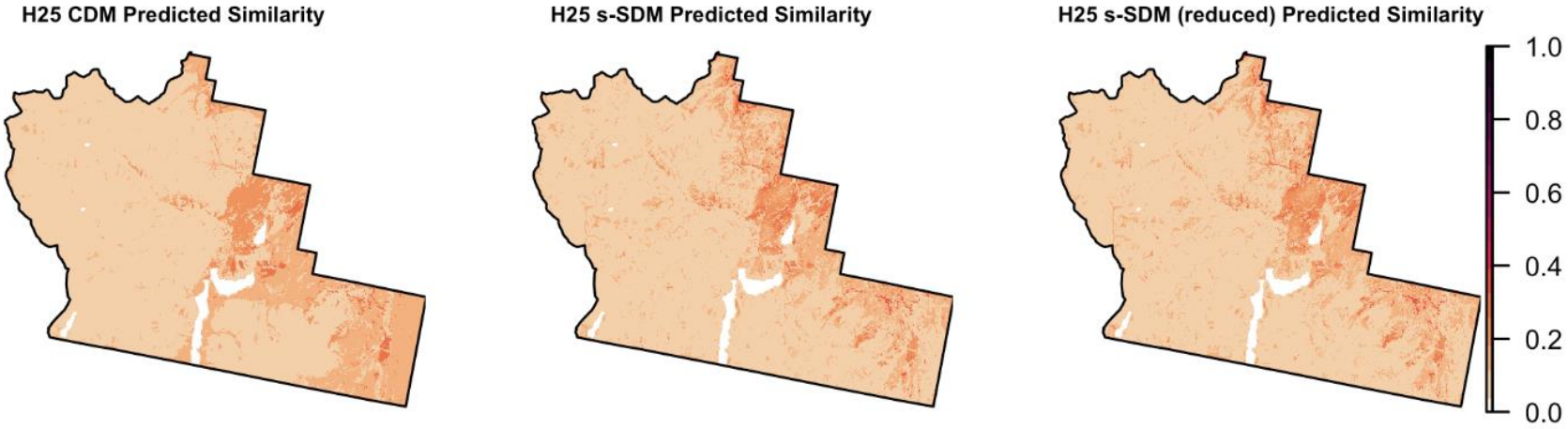

Figure S20. Maps showing predicted Bray-Curtis similarity across the park for community H25 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known H25 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

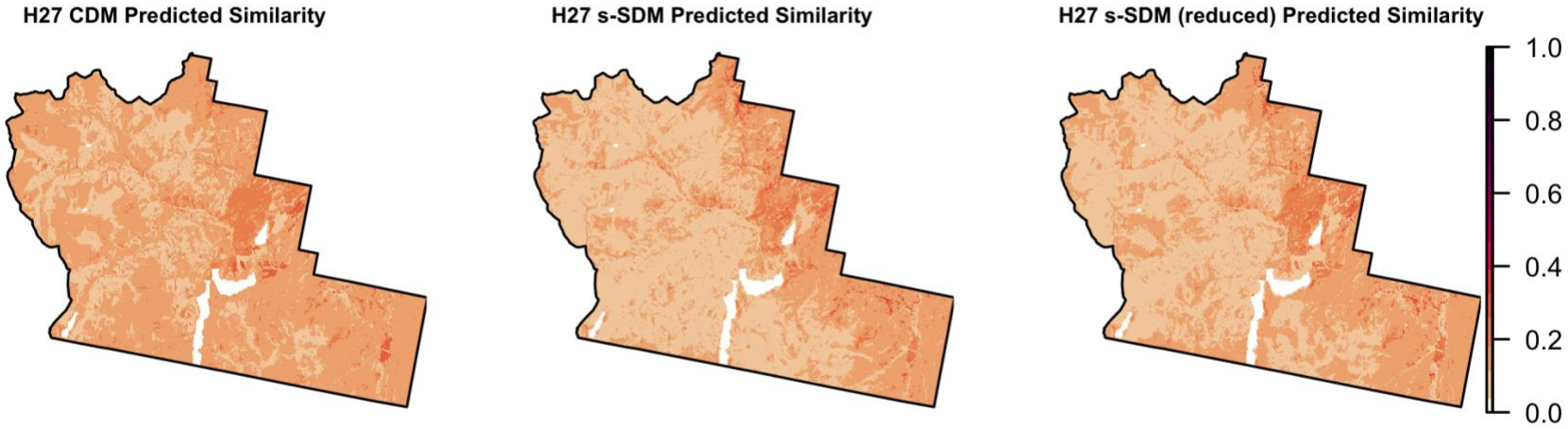

Figure S21. Maps showing predicted Bray-Curtis similarity across the park for community H27 using the best model of each type (CDM, s-SDM, s-SDM (reduced)). Predictions were derived from generalized linear models (GLMs) relating community similarity to each DMs suitability predictions. Darker areas represent higher predicted similarity to known H27 communities (values closer to 1), while lighter areas represent lower similarity (values closer to 0).

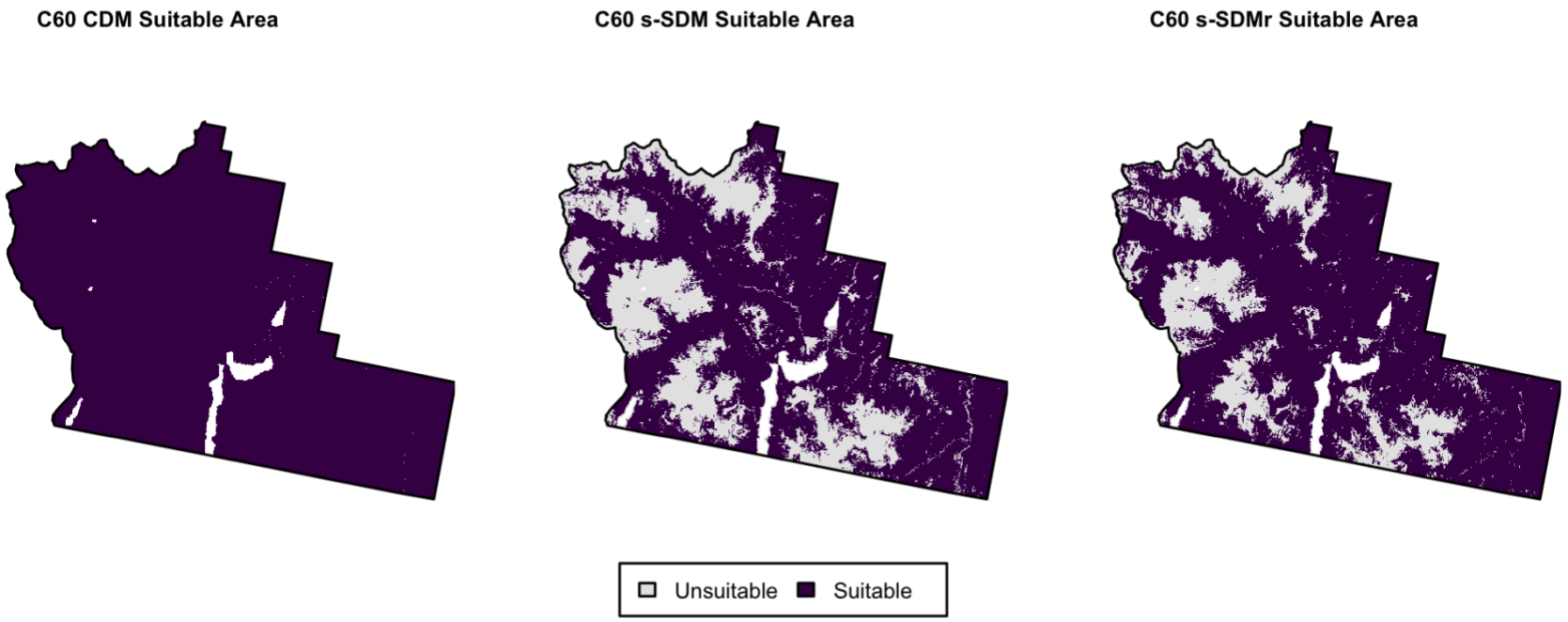

Figure S22. Maps showing predicted suitable and unsuitable areas for community C60 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as C60 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

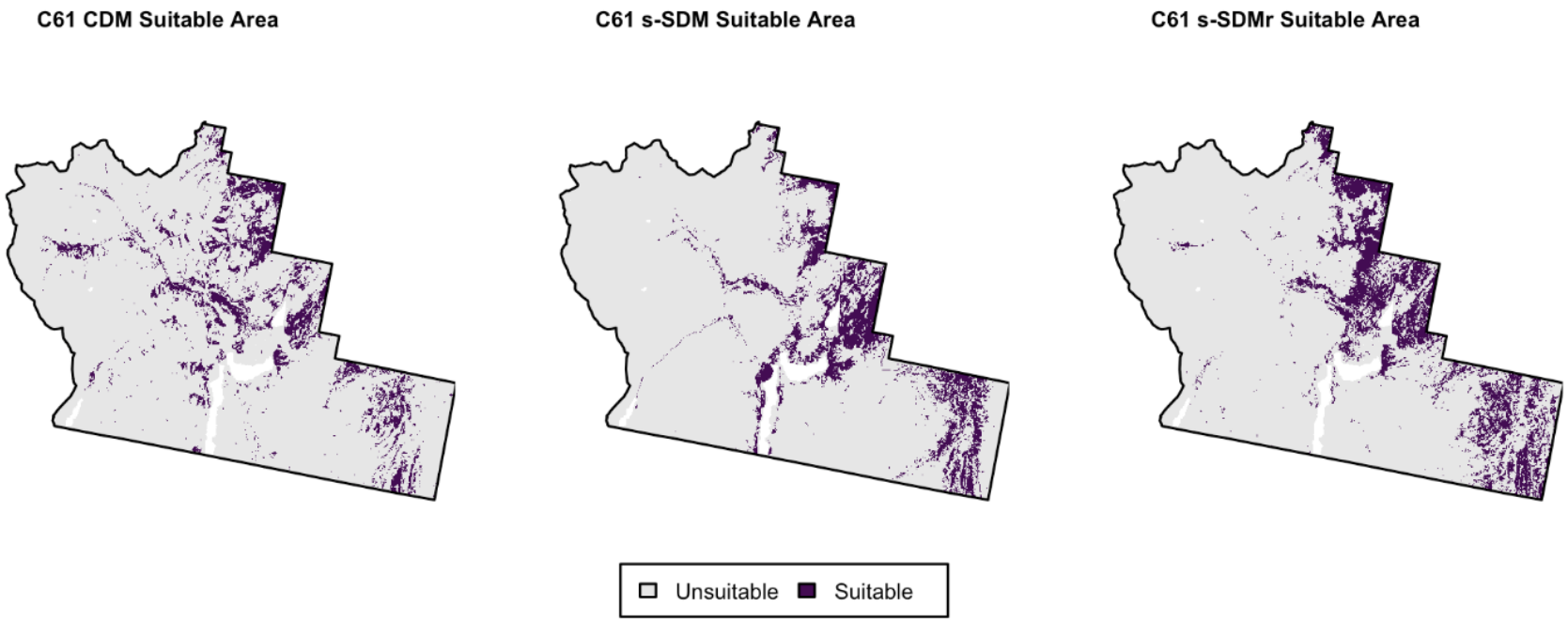

Figure S23. Maps showing predicted suitable and unsuitable areas for community C61 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as C61 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

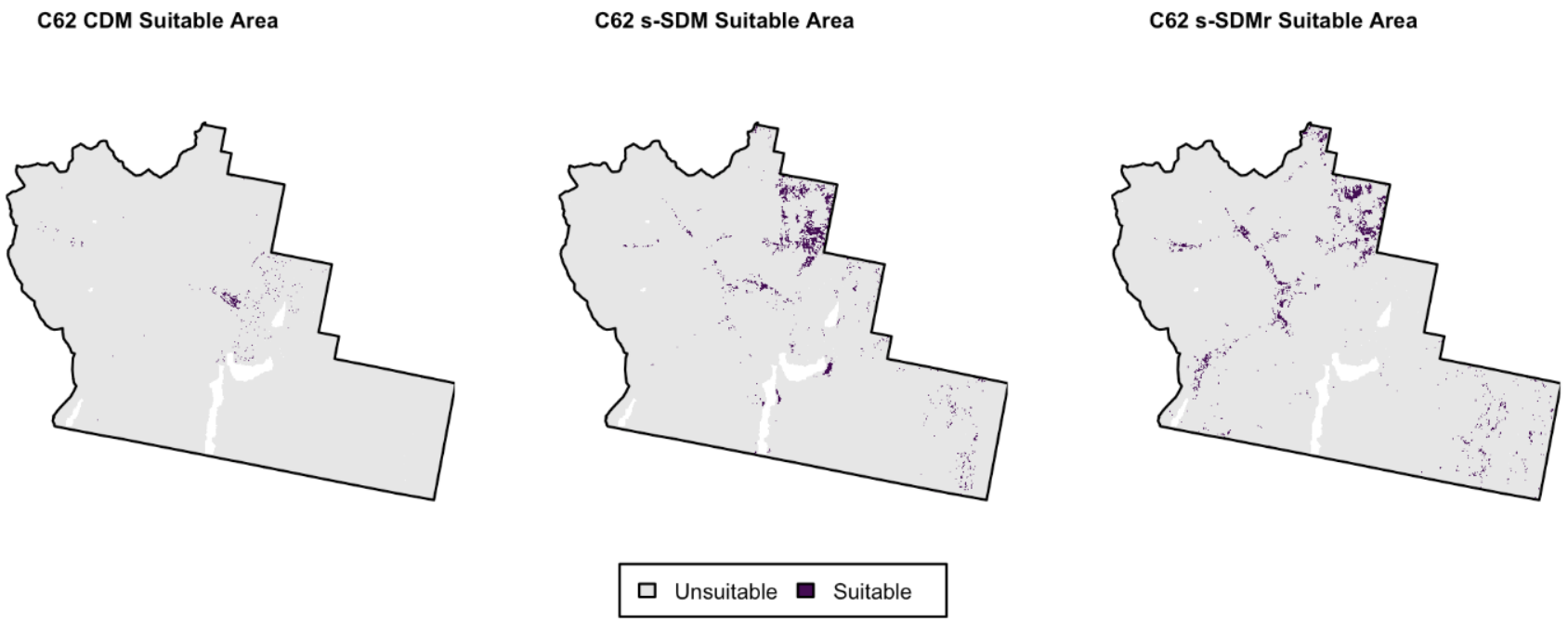

Figure S24. Maps showing predicted suitable and unsuitable areas for community C62 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as C62 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

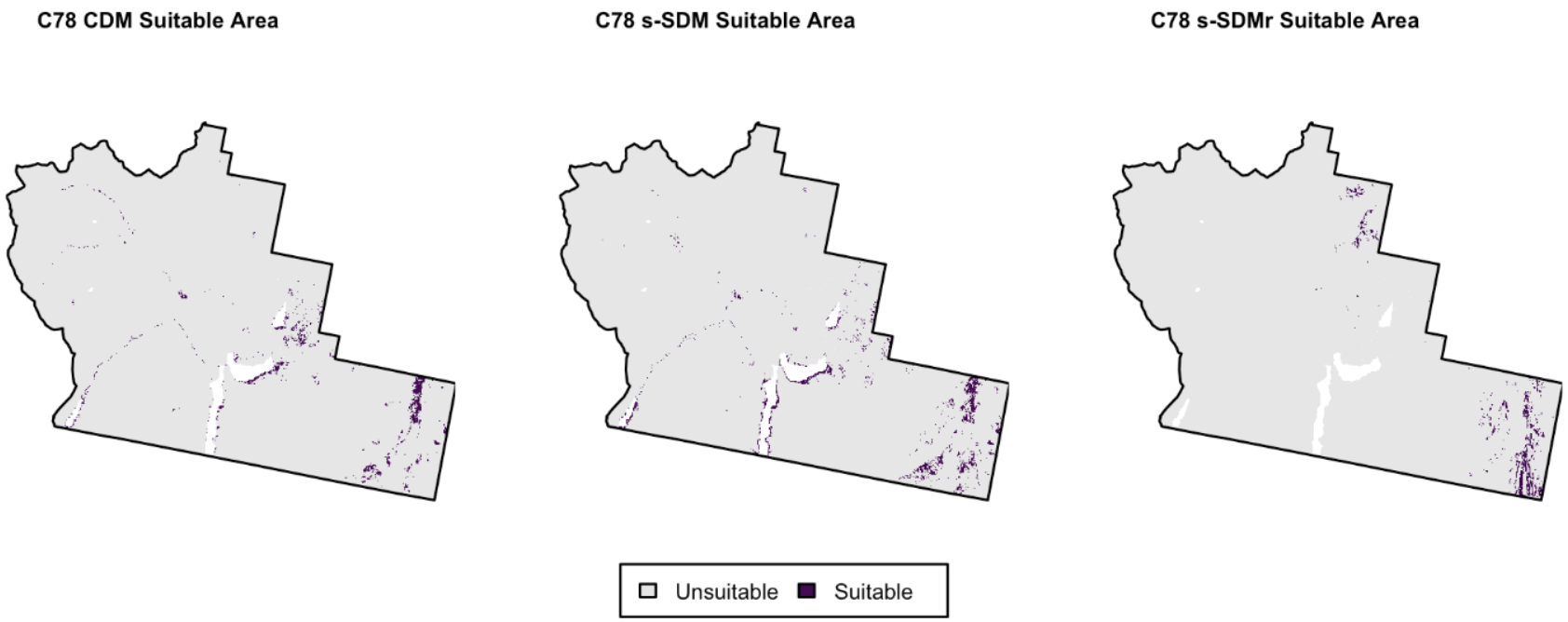

Figure S25. Maps showing predicted suitable and unsuitable areas for community C78 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as C78 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

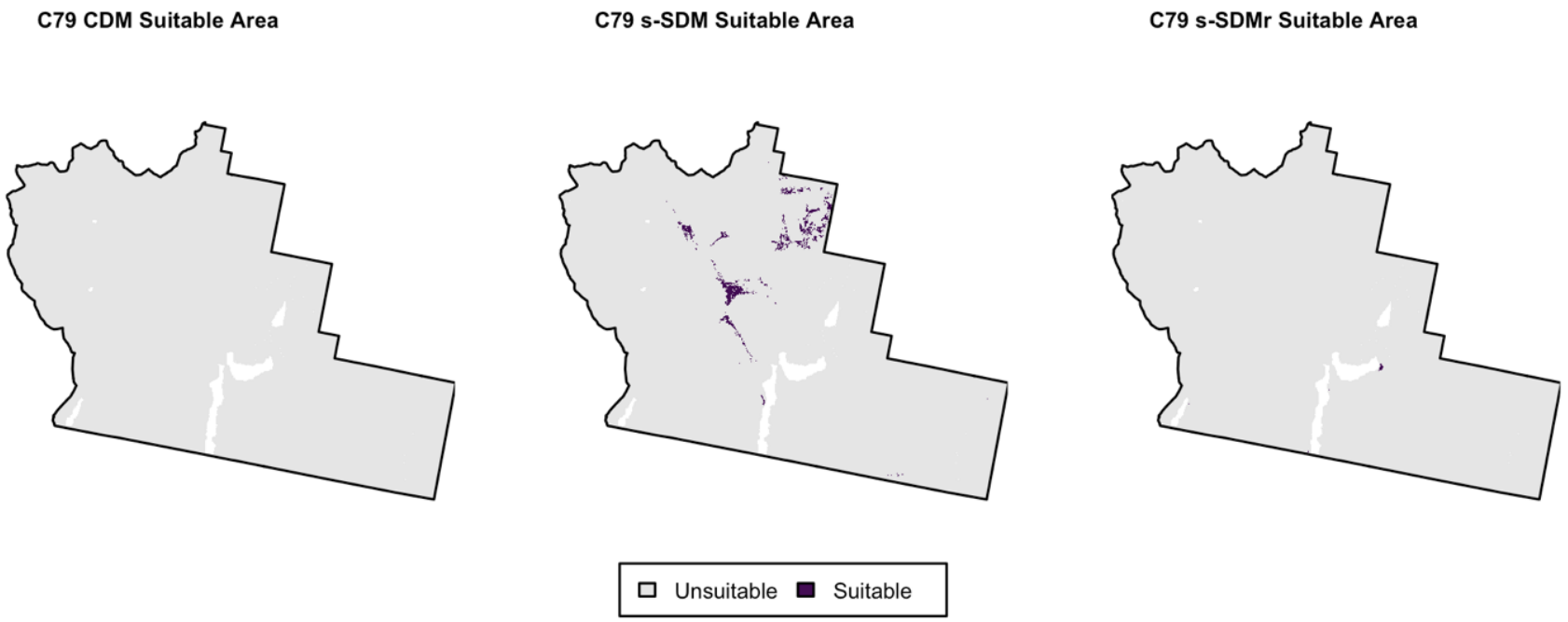

Figure S26. Maps showing predicted suitable and unsuitable areas for community C79 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as C79 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

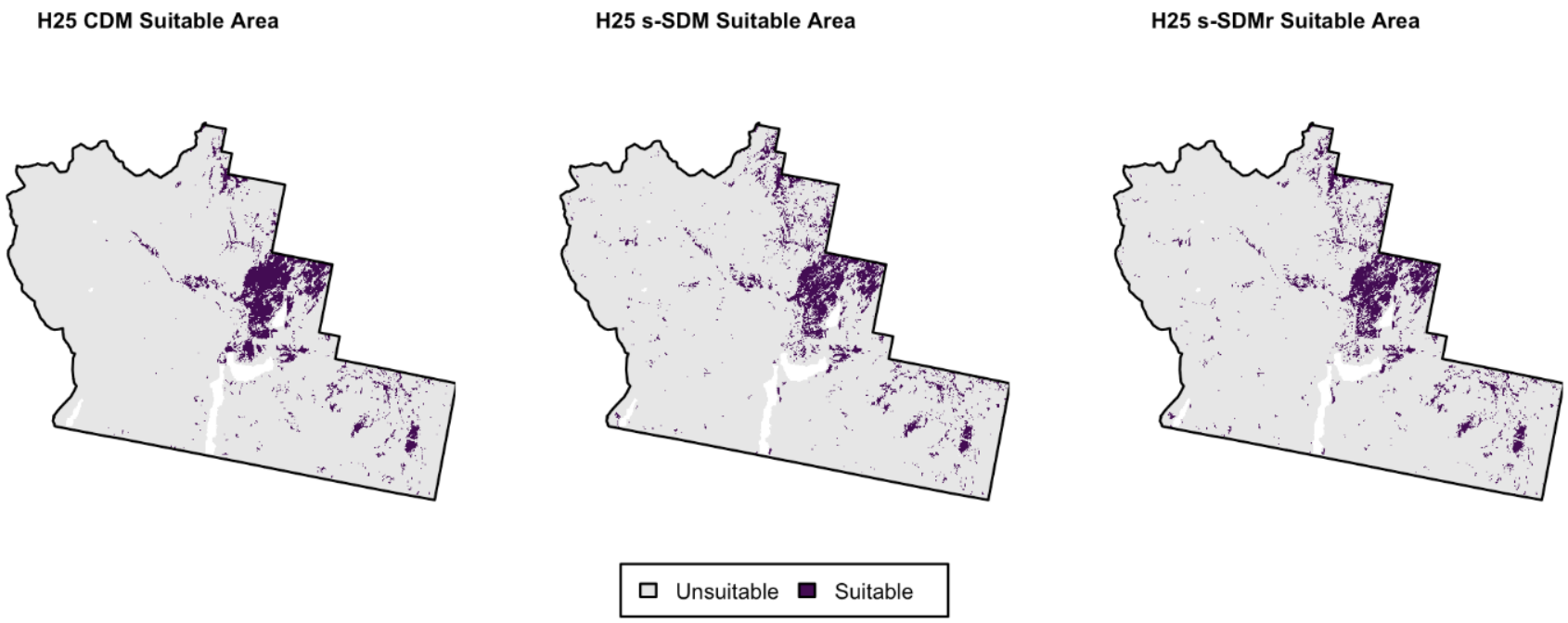

Figure S27. Maps showing predicted suitable and unsuitable areas for community H25 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as H25 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.

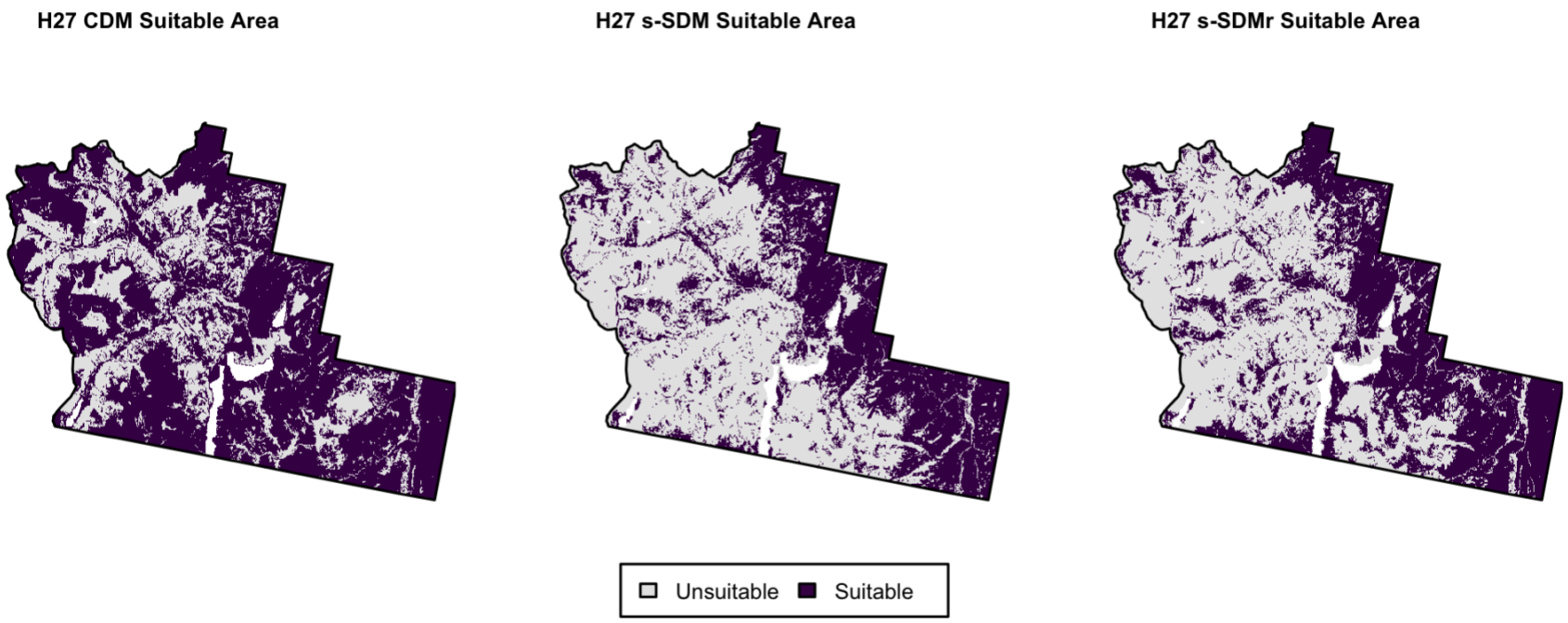

Figure S28. Maps showing predicted suitable and unsuitable areas for community H27 based on the best model of each type (CDM, s-SDM, and s-SDM (reduced)). Suitability was defined using the minimum pairwise Bray-Curtis similarity value between known ELC plots classified as H27 and applied across all three models. Areas above this threshold are shown as suitable (dark purple) and those below as unsuitable (gray). Threshold information and proportions of predicted suitable area for each map can be found in Table S29.