Review

Not peer-reviewed version

From Bioclimatic Envelopes to Machine Learning: A Journey Through the History, Present, and Future of Species Distribution Modeling With Practical Tips for Use and Notes to Bryophytes

Lukáš Číhal *

Posted Date: 14 April 2023

doi: 10.20944/preprints202304.0367.v1

Keywords: species distribution modeling; machine learning; MaxEnt; bryophytes; climatic change

Review

From Bioclimatic Envelopes to Machine Learning: A Journey through the History, Present, and Future of Species Distribution Modeling with Practical Tips for Use and Notes to Bryophytes

Lukáš Číhal

Silesian Museum, Nádražní okruh 31, CZ-746 01 Opava, Czech Republic; cihal@szm.cz

Abstract: Species distribution modeling (SDM) has come a long way since its inception. Starting as simple bioclimatic envelope models based on expert knowledge, species distribution models (SDMs) have evolved into complex and sophisticated models that incorporate multiple sources of data and machine learning algorithms. Today, SDMs play a crucial role in addressing pressing conservation and management issues, including the impacts of climate change on species ranges and the assessment of species vulnerability to extinction. In this article, we will embark on a journey through the history, present, and future of SDM, exploring its evolution from bioclimatic envelopes to machine learning. We will also provide practical tips on how to use SDMs effectively and discuss the exciting future developments in this field. Whether you are a seasoned SDM expert or new to this field, this article will provide valuable insights into the exciting world of SDM. By exploring the rich history and current state of the field, we hope to shed light on the tremendous potential of SDM for improving our understanding of the distribution of species in a changing world.

Keywords: species distribution modeling; machine learning; MaxEnt; bryophytes; climatic change

---

1. Introduction

SDM is a powerful tool in ecology that can be used to predict the theoretical distribution of the species based on environmental variables. It involves creating a statistical model that relates the known distribution of a species to environmental variables. By using these relationships, the model can then be used to predict the theoretical distribution map or a probability map that indicates the potential habitat suitability in areas where no data on its occurrence exists. But the output of an SDM can take different forms, depending on the modeling approach and the software used.

There are also other terms in ecological research for SDM, such as bioclimatic [1], or species niche modeling, habitat suitability modeling [2] habitat suitability modeling [3], ecological niche modeling [4,5], and environmental niche modeling [6], bioclimatic envelope modeling [7,8], or species-habitat models [9].

These terms are often used interchangeably with SDM, and the choice of terminology will depend on the specific research question, the data available, and the context of the study. In our review, we are using the terms SDM (for species distribution modeling) and SDMs (for species distribution models) as they are commonly known and used.

We have many applications in ecology and conservation for SDMs, including identifying areas of high conservation value [10], understanding the potential impacts of environmental change on biodiversity [11], or predicting the spread of invasive species [12]. Among others, SDMs can also be used to inform management decisions [13], such as selecting sites for protected areas [14] or planning for species reintroductions [15]. There are many other applications, but this is more than enough for our idea.

The development of SDM has been driven by advancements in both statistical modeling techniques and the availability of environmental data. With the growth of open-access data repositories and the development of new software tools, SDM has become increasingly accessible and is now widely used in ecological research and management. In this context, the aim of SDM is to provide a framework for understanding the complex relationships between species and their environment.

SDMs typically use statistical or machine learning algorithms to model the relationship between species occurrences and environmental variables. Input data includes species occurrence records and environmental variables such as climate, topography, land use, soil characteristics, and others [16].

The basic outputs of SDMs are usually maps, which are valuable tools for understanding and predicting the distribution of species. The field of SDM has a rich history and over the past several decades has evolved dramatically, with the development of new statistical methods and the availability of large datasets of species occurrences and environmental variables [17].

2. A brief history of SDM

The history of SDM can be traced back to the late 19th century, however, it was not until the late 20th century that the field of SDM began to gain momentum, with the development of new statistical methods and the availability of large datasets of species occurrences and environmental variables [18].

The first SDMs predicted the potential distribution of a species based on its tolerance to certain environmental variables such as temperature and rainfall, but these models were largely based on expert knowledge and lacked the rigor and accuracy of modern SDMs [19].

One of the earliest works in this field was carried out by Robert H. Whittaker in 1956 [20], but the very basics of SDM were already established in earlier research that characterized biological patterns in relation to geographical and environmental factors for example by Grinnell 1904 (The origin and distribution of the chest-nut-backed chickadee), Murray 1866 (The geographical distribution of mammals) or Schimper 1903 (Plant-geography upon a physiological basis) [21].

Another very important part of the history of SDM was the formulation of Hutchinson's concept of the niche, which is a fundamental and important concept in ecology and SDM [22]. G. E. Hutchinson formulated his view of the niche in the 1950s, building on the earlier work of J. Grinnell and Elton [23]. Hutchinson defined the niche as a multidimensional hypervolume in environmental space, where each dimension represents an environmental factor or resource that can limit the distribution or abundance of a species [24]. This view of the niche has since become a cornerstone of ecological theory and has been applied in various fields, including SDM and it is Hutchinson’s view of the niche that provides the most insight and utility for SDMs [25]. This concept is fundamental to the development of SDMs, which aim to identify the environmental variables that best explain the observed occurrence or abundance of a species and to predict its potential distribution across a geographic area.

The development continued slowly and in the 1980s was still constrained by a lack of data and computational power. The approaches used were relatively simplistic compared to the more sophisticated models used today. However, this early work laid the foundation for the development of more advanced techniques, which have been used to study species distributions in more detail and to make predictions about how they may be affected by future environmental changes.

In the 1990s, SDM continued to evolve with several key developments as GIS (geographic information system) whose widespread adoption was a game-changer for SDM [19]. GIS allowed researchers to better visualize and analyze spatial data, and to incorporate information on landscape structure and connectivity into their models. In the 1990s, there was a growing recognition of the importance of non-linear relationships between species and environmental variables [26] and this led to the faster development of modeling techniques, such as artificial neural networks, that could better capture these complex relationships. Researchers in the 1990s also began to explore the idea of using multiple modeling techniques, or "ensemble modeling," to improve the accuracy and robustness of SDMs. This approach is still widely used today.

After the year 2000, SDM continued to evolve and improve in several ways. For example, machine learning algorithms, such as random forests and support vector machines, became widely

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2023 doi:10.20944/preprints202304.0367.v1

used in the 2000s and beyond. Also, the increasing availability of geospatial data and computational resources led to a rapid expansion of SDMs [18]. The increasing availability of large-scale environmental and biodiversity data, as well as the emergence of citizen science initiatives, has enabled researchers to build more detailed and comprehensive models of species distributions (see also [27]).

With growing concerns about climate change, there has been a renewed emphasis on understanding how environmental factors, such as temperature and precipitation, affect species distributions. This has led to the development of new modeling techniques that can account for future climate scenarios. While earlier SDMs often focused on a single or limited number of environmental factors, researchers in the 2000s and beyond have increasingly sought to incorporate multiple factors, including biotic interactions and human impacts, into their models (see also [18]).

The development of open-source software and data resources, such as The R Project for Statistical Computing [28] and GBIF [29], has made it easier for researchers to develop and share SDMs, facilitating collaboration and accelerating progress in the field. Also, the development of new statistical methods, such as generalized linear models (GLMs) and maximum entropy (MaxEnt), revolutionized the field of SDM [30]. These methods allowed for more sophisticated models that could incorporate multiple environmental variables and account for non-linear relationships between species distributions and the environment.

In conclusion, SDM is a field that has evolved significantly over the past several decades, and it continues to play an important role in our understanding of species-environment relationships. With the increasing availability of data and advances in computational power, it is likely that this field will continue to expand and evolve in the coming years.

3. Advantages and disadvantages of SDM

SDM has several advantages, but also some limitations, we will discuss both in more detail.

3.1. Advantages of SDM

In general, SDMs allow researchers to gain a better understanding of the factors that influence the distribution of species. This understanding is crucial for making informed decisions about conservation and management strategies [31].

It can be also used to make predictions about the future distribution of species in response to environmental changes, such as climate change, habitat loss, and land use changes which can be used to identify areas that are likely to be important for conservation in the future [32].

Models can be developed using large amounts of environmental data, which is often more cost-effective and efficient to obtain than data on the presence and absence of species [33]. This makes it a valuable tool for species conservation and management efforts. They can be used to improve the monitoring and management of species by providing information on the distribution and relative importance of different habitats for different species. This information can be used to prioritize conservation efforts and allocate resources more effectively.

SDMs can also inform policy decisions by providing a better understanding of the distribution and potential impacts of species and their habitats. This can be used to develop and implement conservation and management policies that are effective and scientifically sound. By predicting the potential distribution of species, SDMs can support conservation and management efforts by identifying areas of high conservation priority and areas that may be at risk from future environmental changes. See [34].

They also provide insight into the relationships between species and their environments, which can improve our understanding of the drivers of biodiversity and help to inform conservation and management decisions. By combining historical occurrence data with environmental data, SDMs can be also used to make predictions about the potential distribution of species, even in regions where data are lacking. And last but not least SDMs can be used to explore the potential impacts of future climate change on species distributions, helping to inform adaptation and mitigation efforts [35].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2023 doi:10.20944/preprints202304.0367.v1

Overall by bringing together data from a variety of sources, SDM provides a means of synthesizing information about species distributions, which can help to fill gaps in our knowledge and improve our understanding of biodiversity patterns and processes.

3.2. Disadvantages of SDM

There are also some disadvantages that need to be considered when using SDM for ecological and conservation purposes. One of the biggest challenges in SDM is obtaining high-quality, reliable data, because models are only as good as the data they are based on, and data biases or inaccuracies can negatively impact the results of the model. This count especially for rare, elusive, and cryptic species [36].

SDMs also typically focus on single species and ignore the complex interactions between species and their environment. This can result in models that are oversimplified and do not accurately reflect the real-world relationships between species and their environment [37].

Another problem is that there is no one "best" model for SDM, and the choice of model depends on the specific goals and data available for a given study. This can make it difficult to compare results across studies and to determine the most appropriate model for a given situation. But broadly speaking, nonparametric approaches with the capability of controlling for model complexity outperformed traditional regression methods, with MaxEnt and boosted regression trees still among the top performing models [38].

So while SDMs can provide useful predictions, they are not always accurate, particularly in areas with limited data or complex environmental relationships. It is important to evaluate the predictive accuracy of models and to use caution when making decisions based on their results. SDMs are also limited by the spatial scale at which they are constructed, and their results may not be transferable to different spatial scales [39]. This can limit their utility for making predictions at different scales and for evaluating the effects of global change on species distributions.

Despite everything mentioned, SDM can provide valuable insights into species distributions, but it is important to understand their limitations and use them in conjunction with other methods to make informed decisions. For more information about challenges (disadvantages) in SDM see also [40].

4. General knowledge and basic assumptions needed for SDM

General knowledge is critical because it helps in selecting appropriate data, methods, and assumptions, which are all essential for producing accurate and reliable SDM results. For example, having knowledge about the ecological requirements of a species can help in selecting appropriate environmental variables for the model. Similarly, understanding the limitations of the modeling algorithms can help in interpreting the results and avoiding potential biases or errors.

4.1. Basic assumption for SDM

- Biology and ecology: A strong understanding of biology and ecology is important for understanding the relationships between species and their environment. This includes knowledge of the life history, biology, and ecology of the species of interest, as well as the interactions between species and their environment.
- Geographical information systems (GIS): GIS is a key tool for creating and analyzing SDMs. To fully understand SDM, you should learn the basics of GIS, including how to create and manipulate maps, how to work with spatial data, and how to use GIS software.
- Statistics and machine learning: SDMs are typically based on statistical and machine learning techniques. To understand the methods used in SDM, you should have a solid understanding of basic statistics and be familiar with common machine learning algorithms, such as regression and classification models.
- Environmental data: Environmental data, such as climate, topography, and land use data, are critical inputs for SDMs. To fully understand SDM, you should learn how to obtain, process, and analyze environmental data.

- Programming: Many SDMs are implemented using computer programs, such as R or Python. To understand how these models are constructed and how to use them, you should have a basic understanding of programming and be familiar with at least one programming language.

So to fully understand SDM, you should have a solid foundation in biology, ecology, GIS, statistics, machine learning, environmental data processing, and programming. Additionally, you should be familiar with current research in the field and be able to critically evaluate the results of SDMs.

5. Modeling overview

There are a number of approaches that evolve according to the type of data, the type of questions, and the advantages of individual methods, here we will take a closer look at some of them as well as at different data types.

5.1. Presence-only and presence-absence data

There are two main types of data used in SDMs, presence-only data, and presence-absence data. Presence-only data are data that only provide information on where a species has been observed, but not where it has not been observed. This can include data from science projects, museum specimens, or other sources where species presence can be recorded, but absence cannot (see also [41]). Models that use only presence data are called presence-only. Presence-absence data, on the other hand, provide information on where a species is present as well as where it is absent. This can be obtained through field surveys or by analyzing historical data. Models that use both presence and absence data are called presence-absence SDMs.

The main difference between presence-only and presence-absence SDMs is the type of data used to train the model. Presence-only models rely solely on the presence data and try to infer the environmental conditions that are associated with those observations. This can result in biased models, as the model may overpredict the distribution of the species in areas with similar environmental conditions where the species has not been observed.

Presence-absence models, on the other hand, have both presence and absence data and can estimate both the environmental conditions that are associated with the presence and the conditions that are associated with absence. This can result in more accurate models, as they account for the true absence of the species in certain areas.

In summary, while both presence-only and presence-absence SDMs are used to predict the geographic distribution of a species, the main difference between them is the type of data used to train the model. Presence-absence models tend to be more accurate because they account for both presence and absence data, while presence-only models may be biased due to the lack of absence data. On the other hand, presence-only data can also offer several advantages in SDM, particularly in situations where presence-absence data are not available or are difficult to collect or because of flexibility in modeling, because they can be used in a range of modeling frameworks, including MaxEnt, random forest models and others, which can account for variation in sampling effort and imperfect detection (see also [42]).

5.2. Model types

5.2.1. Correlative models

Correlative models are statistical models that aim to identify the relationships between the presence or absence of a species and environmental variables, without necessarily incorporating the underlying ecological or physiological processes that drive these relationships. See also [43].

Some of the most known correlative models types include:

- Maximum Entropy (MaxEnt): MaxEnt is a machine learning algorithm that is widely used in SDM. It works by finding the distribution that is most likely to occur given the environmental

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2023 doi:10.20944/preprints202304.0367.v1

variables and the known occurrence data for a species. The algorithm has been shown to perform well in a variety of situations and is well-suited for dealing with complex data.

- Generalized Linear Models (GLMs): GLMs are a family of statistical models that are widely used in SDM. They are simple to implement, fast, and flexible, and can be used to model a wide range of relationships between species and environmental variables.
- Boosted Regression Trees (BRTs): BRTs are a type of machine learning algorithm that uses a series of decision trees to make predictions. BRTs have been shown to perform well in a variety of situations, and are well-suited for dealing with complex and non-linear relationships between species and environmental variables.
- Artificial Neural Networks (ANNs): ANNs are a type of machine learning algorithm that is modeled after the structure and function of the human brain. ANNs are capable of dealing with complex and non-linear relationships between species and environmental variables.
- Random Forest (RFs): RFs are a type of machine learning algorithm that is based on the decision tree algorithm. RFs use multiple decision trees to make predictions and are well-suited for dealing with complex and non-linear relationships between species and environmental variables.

Correlative models have several advantages for SDM, including their ease of use and their ability to identify important environmental variables that influence species distribution. However, it is important to recognize that correlative models do not necessarily capture the underlying ecological or physiological processes that drive these relationships.

5.2.2. Mechanistic models

Mechanistic models are another type of model used for SDM. Unlike correlative models, mechanistic models incorporate the underlying ecological or physiological processes that drive the relationship between the presence or absence of a species and environmental variables [44].

There are several types of mechanistic models:

- Ecophysiological models incorporate physiological processes, such as energy balance and water balance, to predict how environmental factors affect a distribution.
- Biophysical models simulate the physical processes that determine a species' distribution, such as energy and mass transfer, and can be used to predict how changes in environmental conditions affect a distribution.
- Agent-based models simulate the behavior of individual organisms and can be used to predict how the interactions between individuals and the environment affect a distribution.
- Population models simulate population dynamics, including birth and death rates, dispersal, and competition, and can be used to predict how changes in environmental conditions affect a distribution.
- Hybrid models combine elements of different mechanistic approaches to provide a more comprehensive understanding of distribution.

Mechanistic SDMs simulate the different processes that determine species distributions and use these simulations to predict the distribution of species under different environmental conditions. They can provide a more mechanistic understanding of species distributions and can be used to explore how species may respond to environmental change. However, mechanistic SDMs can be more complex and data-intensive than correlative SDMs, and they may require detailed knowledge of a species' biology that is not always available. As a result, they are often used in combination with other modeling approaches to provide a more comprehensive understanding of species distributions.

Rather than identifying one or the other as 'better,' we suggest that researchers take great care to use the method best-suited to each specific research question, and be conscious of the weaknesses of any method, such that inappropriate interpretations are avoided [45] (see also [46]).

5.3. MaxEnt

For a better description, we have chosen MaxEnt as one of the most popular methods, because it ranks among the most widespread algorithms for many reasons and has been widely used in ecology and conservation biology. It is a powerful tool for understanding the factors that drive

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2023 doi:10.20944/preprints202304.0367.v1

species distribution and studies have shown that it is a useful tool for predicting the potential distribution of species, particularly when compared to other SDM algorithms [38].

5.3.1. Strengths of MaxEnt

MaxEnt has several strengths that make it a popular choice for SDM. For example, it is well-suited for handling small sample sizes, which are common in species distribution data. It can also handle complex interactions between environmental variables, such as non-linear relationships or high-order interactions, allowing it to capture the complex ecological processes that determine species distribution. MaxEnt provides measures of variable importance, allowing researchers to identify the environmental factors that are most important in determining a species distribution. And last but not least, with a user-friendly interface and a range of software packages available for different platforms MaxEnt is easy to use [47].

5.3.2. Weaknesses of MaxEnt

Although MaxEnt is a great tool, like all methods, even MaxEnt has certain limitations and weaknesses that one should be aware of. Specifically, MaxEnt has a tendency to overfit the training data, especially when the number of parameters is high [48]. This means that the model may perform well on the training data, but poorly on new, unseen data. Also MaxEnt can be computationally intensive, especially for large datasets. This can make the training process slow and limit the ability to scale the method to big data. MaxEnt models can be complex and difficult to interpret, especially when there are many features. This can make it challenging to understand the relationships between predictors and responses and to make meaningful inferences about the data. Despite these limitations, MaxEnt remains a popular and powerful machine learning technique and is widely used in many applications due to its ability to handle large datasets and its robust performance.

6. The general methodology of the SDM

A general methodology provides a standardized approach to SDM, which allows researchers to follow a set of guidelines and procedures to produce consistent and comparable results across different studies and provide numerous benefits, including standardization, reproducibility, efficiency, and consistency. This would improve the overall quality and reliability of SDM results and contribute to a better understanding of species distributions and ecology.

Note that SDM is a complex and rapidly evolving field, and the specific methodology used may vary depending on the research question, data availability, and the goals of the modeling exercise. For more information, see also [49–51].

6.1. Typical steps in SDM methodology

- Data collection: Gather species occurrence data and environmental variables that may influence the species distribution. Species occurrence data can be obtained from sources such as museum collections, field surveys, or citizen science programs. Environmental variables may include bioclimatic data, topographical information, land use/land cover data, and others.
- Data preparation: Clean and process the data, ensuring that it is suitable for modeling. This may involve filtering out unreliable or inconsistent data, transforming variables to match the scale of the model, and handling missing data.
- Model development: Select and apply a statistical model to the data. Commonly used models include generalized linear models (GLMs), generalized additive models (GAMs), and machine learning algorithms such as random forests and MaxEnt. The choice of model will depend on the type of data, the research question, and the complexity of the relationship between species and environmental variables.
- Model evaluation: Evaluate the performance of the model using techniques such as cross-validation, ROC curves, and area under the curve (AUC) metrics or others. This step helps to determine the accuracy and reliability of the model.

- Model application: Use the model to make predictions about the distribution of the species in regions where data are lacking or in future scenarios. This can involve mapping the predicted distribution, estimating the potential range of the species, or exploring the effects of environmental changes on the distribution.
- Model interpretation: Interpret the results of the model, including the relative importance of different environmental variables, the accuracy of the model predictions, and the implications of the results.

7. SDM and climate change

Correlative species distribution models (SDMs) are widely utilized in conservation biogeography, particularly in spatially explicit biogeographic models. These models are highly valued due to their ability to forecast potential range shifts of species and communities in response to climate change. As a result, SDMs can serve as a valuable tool for informing and guiding conservation management planning, including the development of collaborative transboundary conservation frameworks. It's no surprise that SDMs are the most popular method in this field [52].

SDM has become increasingly important in recent years, particularly in the context of climate change. As the earth's climate changes, species are likely to experience shifts in their geographic range and habitat suitability and SDM can be used to model and predict these changes, which is essential for understanding the potential impacts of climate change on biodiversity and informing conservation and management decisions.

Climate change is expected to have profound effects on the distribution and abundance of species. Changes in temperature and precipitation patterns, as well as extreme weather events, are likely to result in shifts in the geographic range and habitat suitability of many species [53]. For example, some species may experience range contractions or even extinction, while others may expand their range into new areas. These changes can have cascading effects on ecosystem dynamics and ecosystem services, such as pollination and pest control [54]. SDM can be used to predict how species distributions are likely to shift under different climate scenarios [55]. This can help to identify areas of high conservation value, areas where species are likely to be at risk, and areas where invasive species may become more of a threat. SDM can also be used to inform conservation planning and management, such as identifying corridors or refugia for species that may be particularly vulnerable to climate change (see also [56]).

SDM can be used in relation to climate change in several ways. For example, for predicting the potential distribution of species under different climate scenarios, identifying areas that are likely to be suitable for species under future climate conditions, or analyzing the rate and direction of species range shifts in response to changing climate conditions. We can also identify areas that may serve as potential refugia for species during times of environmental change and assess the vulnerability of species to climate change, based on their current distribution, environmental requirements, and sensitivity to climate.

8. Notes to bryophytes

Bryophytes are a group of small, non-vascular plants that include mosses, liverworts, and hornworts. They are found in almost every terrestrial environment and play important roles in regulating the global carbon cycle and maintaining ecosystem stability. Especially, bryophytes play a crucial role in carbon sequestration, which is the process of removing carbon dioxide from the atmosphere and storing it in plant tissues, but as global warming continues, bryophyte growth and productivity may decline due to changes in temperature, precipitation, and moisture availability, potentially leading to a reduction in the capacity of these plants to sequester carbon [57].

Global warming and climate change also have significant impacts on them [58]. As bryophytes are highly sensitive to changes in temperature and moisture levels [59], one of the most significant effects of global warming on bryophytes is the alteration of their distribution patterns [60]. As temperatures rise and precipitation patterns change, bryophytes can be forced to move to higher elevations or latitudes in search of suitable habitats [61].

 doi:10.20944/preprints202304.0367.v1

Another impact of global warming on bryophytes can be also in the alteration of their physiology, because even though bryophytes are known to have high levels of desiccation tolerance [62], as temperatures rise, bryophytes are likely to experience increased rates of evapotranspiration, leading to decreased water availability and increased susceptibility to desiccation stress [63].

All this means that bryophytes are highly vulnerable to the impacts of global warming and climate change, which are likely to alter their distribution patterns, physiology, and ecological roles. Therefore, it is essential to monitor and conserve bryophyte populations to ensure their continued survival and their contribution to global ecosystem stability. For those purposes, SDM can be used to investigate the environmental factors that influence the distribution of bryophyte species and to predict how these species may respond to future changes in climate or land use.

But there are also some challenges in modeling bryophytes. For example, one challenge of modeling bryophyte distribution is their small size and patchy distribution, which can make it difficult to accurately map their distribution at a large scale. Additionally, bryophyte species often have complex ecological requirements, such as specific microhabitats or associations with certain tree species, and that may be difficult to capture in an SDM. Despite these (and others challenges), SDMs have been used successfully to study the distribution of bryophytes in a variety of ecosystems.

9. About future directions in SDM

SDM has made significant advancements over the past few decades, and it continues to be an active area of research in ecology. With the ongoing development of new modeling techniques and the availability of increasingly sophisticated data, there are many potential future directions of SDM.

One potential future direction in SDM is the integration of more complex environmental variables and spatially explicit models. As the availability of high-resolution environmental data increases, there is a growing need to incorporate more complex environmental interactions into SDM. Additionally, spatially explicit models that incorporate spatial autocorrelation and dispersal dynamics can improve the accuracy of SDM by accounting for the spatial dependencies and movements of species.

Another direction for the future of SDM is the development of more comprehensive and dynamic models that can account for multiple factors simultaneously. For example, future SDMs may more frequently incorporate not only climate and environmental variables but also biotic interactions, such as competition, predation, and mutualism, as well as demographic processes, such as dispersal, population growth, and mortality. Such models would provide a more holistic understanding of species distributions and the factors that influence them.

Also, the development of more user-friendly and accessible software packages and data repositories is an important direction for the future of SDM, because many of the current SDM tools are complex and require advanced technical skills to use, which can limit their accessibility to non-experts. By developing more user-friendly interfaces and open-source software, researchers can increase the accessibility of SDM and promote the use of the technique in a wider range of research and management applications.

The future of SDM is likely to involve the integration of more complex environmental variables, the development of more comprehensive and dynamic models, and the creation of more user-friendly software and data repositories. These future directions in SDM have the potential to improve the accuracy and utility of SDM for understanding and managing species distributions in a rapidly changing world. There are many other directions of development, but that would be a separate article (see also [64]).

10. Conclusion

In conclusion, SDM is an essential tool for understanding the potential impacts of environmental change on biodiversity. It has been used to provide insights into the past and present distribution of species and to predict their potential distribution in the future. SDM is among others a valuable tool for identifying areas of high conservation value, areas where species are likely to be at risk, and areas where invasive species may become more of a threat.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2023 doi:10.20944/preprints202304.0367.v1

In the past, the SDM research has seen significant advancements, with the development of new modeling techniques, the availability of increasingly sophisticated data, and the growth of scientific understanding of ecological processes. The presence of the SDM is characterized by the growing need for effective conservation and management strategies in the face of environmental change. The future of SDM is exciting, and there are many potential avenues for development. These include the integration of more complex environmental variables and spatially explicit models, the development of more comprehensive and dynamic models, and the creation of more user-friendly software and data repositories. These developments will allow SDM to provide even more accurate and informative predictions of how species distributions are likely to shift under different environmental scenarios. By continuing to advance the science of SDM, researchers, and practitioners can better understand the impacts of environmental change on biodiversity.

Funding: This work was financially supported by the Ministry of Culture of the Czech Republic by institutional financing of long-term conceptual development of the research institution (Silesian Museum, MK000100595).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Franklin, J. Species distribution models in conservation biogeography: developments and challenges. *Diversity and distributions* **2013**, *19*(10), 1217–1223.
2. Franklin, J. Mapping species distributions: spatial inference and prediction; Cambridge University Press: Cambridge, United states of America, 2010.
3. Hirzel, A. H.; Le Lay, G. Habitat suitability modelling and niche theory. *Journal of applied ecology* **2008**, *45*(5), 1372–1381.
4. Peterson, A. T.; Soberón, J. Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza Conservação* **2012**, *10*(2), 102–107.
5. Valencia–Rodríguez, D.; Jiménez–Segura, L.; Rogéliz, C. A.; Parra, J. L. Ecological niche modeling as an effective tool to predict the distribution of freshwater organisms: The case of the Sabaleta Brycon henni (Eigenmann, 1913). *PLoS One* **2021**, *16*(3), e0247876.
6. Warren, D. L.; Glor, R. E.; Turelli, M. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **2010**, *33*(3), 607–611.
7. Araújo, M. B.; Peterson, A. T. Uses and misuses of bioclimatic envelope modeling. *Ecology* **2012**, 93(7), 1527–1539.
8. Heikkinen, R. K.; Luoto, M.; Araújo, M. B.; Virkkala, R.; Thuiller, W.; Sykes, M. T. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography* **2006**, *30*(6), 751–777.
9. Iverson, L. R.; Prasad, A. M.; Matthews, S. N.; Peters, M. P. Lessons learned while integrating habitat, dispersal, disturbance, and life–history traits into species habitat models under climate change. *Ecosystems* **2011**, *14*, 1005–1020.
10. Wilson, C. D.; Roberts, D.; Reid, N. Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using Margaritifera margaritifera (L.). *Biological Conservation* **2011**, *144*(2), 821–829.
11. Jones, M. C.; Dye, S. R.; Fernandes, J. A.; Frölicher, T. L.; Pinnegar, J. K.; Warren, R.; Cheung, W. W. Predicting the impact of climate change on threatened species in UK waters. *PloS one* **2013**, *8*(1), e54216.
12. Plášek, V.; Číhal, L.; Müller, F.; Smoczyk, M.; Marková, I.; Fialová, L. Quo Vadis, Orthotrichum pulchellum? A Journey of Epiphytic Moss across the European Continent. *Plants* **2022**, *11*(20), 2669.
13. Sofaer, H. R.; Jarnevich, C. S.; Pearse, I. S.; Smyth, R. L.; Auer, S.; Cook, G. L.; ... Hamilton, H. Development and delivery of species distribution models to inform decision–making. *BioScience* **2019**, *69*(7), 544–557.
14. Fajardo, J.; Lessmann, J.; Bonaccorso, E.; Devenish, C.; Munoz, J. Combined use of systematic conservation planning, species distribution modelling, and con–nectivity analysis reveals severe conservation gaps in a megadiverse country (Peru). *PloS one* **2014**, *9*(12), e114367.

15. Miranda, E. B.; Menezes, J. F. S.; Farias, C. C.; Munn, C.; Peres, C. A. Species distribution modeling reveals strongholds and potential reintroduction areas for the world's largest eagle. *PloS one* **2019**, *14*(5), e0216323.
16. Deneu, B.; Servajean, M.; Botella, C.; Joly, A. Evaluation of deep species distribution models using environment and co–occurrences. In *Experimental IR Meets Multilinguality*, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10 (pp. 213–225). Springer International Publishing.
17. Elith, J.; H. Graham, C.; P. Anderson, R.; Dudík, M.; Ferrier, S.; Guisan, A.; ... E. Zimmermann, N. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*(2), 129–151.
18. Elith, J.; Leathwick, J. R. Species distribution models: ecological expla–nation and prediction across space and time. *Annual review of ecology, evolution, and systematics* **2009**, *40*, 677–697.
19. Guisan, A.; Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecological modelling* **2000**, *135*(2–3), 147–186.
20. Whittaker, R. H. Vegetation of the great smoky mountains. *Ecological Monographs* **1956**, *26*(1), 2–80.
21. Elith, J.; Leathwick, J. R. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics* **2009**, *40*, 677–697.
22. Soberón, J. M. Niche and area of distribution modeling: a population ecology perspective. *Ecography* **2010**, *33*(1), 159–167.
23. Colwell, R. K.; Rangel, T. F. Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences* **2009**, *106*(supplement_2), 19651–19658.
24. Takola, E.; Schielzeth, H. Hutchinson's ecological niche for individuals. *Biology & Philosophy* **2022**, *37*(4), 25.
25. Elith, J.; Franklin, J. Species distribution modeling. In Encyclopedia of Biodiversity: Second Edition; Levin, SA.; Eds.; Elsevier inc.: Amsterdam, Netherlands, 2013; pp. 692–705.
26. Lek, S.; Delacoste, M.; Baran, P.; Dimopoulos, I.; Lauga, J.; Aulagnier, S. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological modelling* **1996**, *90*(1), 39–52.
27. Feldman, M. J.; Imbeau, L.; Marchand, P.; Mazerolle, M. J.; Darveau, M.; Fenton, N. J. Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS One* **2011**, *16*(3), e0234587.
28. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing 2022, Vienna, Austria. URL https://www.R–project.org/ (accessed 13. 4. 2023).
29. GBIF Secretariat: GBIF Backbone Taxonomy. https://doi.org/10.15468/39omei Accessed via https://www.gbif.org/species/5284517 (accessed 13. 4. 2023).
30. Phillips, S. J.; Anderson, R. P.; Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecological modelling* **2006**, *190*(3–4), 231–259.
31. Guisan, A.; Tingley, R.; Baumgartner, J. B.; Naujokaitis-Lewis, I.; Sutcliffe, P. R.; Tulloch, A. I.; ... Buckley, Y. M. Predicting species distributions for conserva–tion decisions. *Ecology letters* **2013**, *16*(12), 1424–1435.
32. Kaky, E.; Nolan, V.; Alatawi, A.; Gilbert, F. A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics* **2020**, *60*, 101150.
33. Isaac, N. J.; Jarzyna, M. A.; Keil, P.; Dambly, L. I.; Boersch–Supan, P. H.; Browning, E.; ... O'Hara, R. B. Data integration for large-scale models of species distributions. *Trends in ecology & evolution* **2020**, *35*(1), 56–67.
34. Číhal, L.; Kaláb, O. Species distribution models for critically endangered liverworts (Bryophyta) from the Czech Republic: a guide to future survey expeditions. *Acta Musei Silesiae, Scientiae Naturales* **2017**, *66*(2), 101–110.
35. Biber, M. F., Voskamp, A., Hof, C. Potential effects of future climate change on global reptile distributions and diversity. *Global Ecology and Biogeography* **2023**, *32*(4), 519–534
36. Aubry, K. B.; Raley, C. M.; McKelvey, K. S. The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS One* **2017**, *12*(6), e0179152.
37. Godsoe, W.; Harmon, L. J. How do species interactions affect species distribution models? *Ecography* **2012**, *35*(9), 811–820.
38. Valavi, R.; Guillera-Arroita, G.; Lahoz-Monfort, J. J.; Elith, J. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs* **2022**, *92*(1), e01486.

39. Liu, C.; Wolter, C.; Xian, W.; Jeschke, J. M. Species distribution models have limited spatial transferability for invasive species. *Ecology Letters* **2020**, *23*(11), 1682–1692.
40. Araújo, M. B.; Guisan, A. Five (or so) challenges for species distribution modelling. *Journal of biogeography* **2006**, *33*(10), 1677–1688.
41. Konowalik, K.; Nosol, A. Evaluation metrics and validation of presence–only species distribution models based on distributional maps with varying coverage. *Scientific Reports* **2021**, *11*(1), 1–15.
42. Morera-Pujol, V.; Mostert, P. S.; Murphy, K. J.; Burkitt, T.; Coad, B.; McMahon, B. J.; ... Ciuti, S. Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance. *Ecography* **2023**, *2023*(2), e06451.
43. Miller, J. Species distribution modeling. *Geography Compass* **2010**, *4*(6), 490–509.
44. Kearney, M. R.; Wintle, B. A.; Porter, W. P. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation letters* **2010**, *3*(3), 203–213.
45. Peterson, A. T.; Papeş, M.; Soberón, J. Mechanistic and correlative models of ecological niches. *European Journal of Ecology* **2015**, *1*(2), 28–38
46. Evans, T. G.; Diamond, S. E.; Kelly, M. W. Mechanistic species distribution modelling as a link between physiology and conservation. *Conservation physiology* **2015**, *3*(1).
47. Elith, J.; Phillips, S. J.; Hastie, T.; Dudík, M.; Chee, Y. E.; Yates, C. J. A statistical explanation of MaxEnt for ecologists. *Diversity and distributions* **2011**, *17*(1), 43–57.
48. Morales, N. S.; Fernández, I. C.; Baca–González, V. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? *A systematic review. PeerJ* **2017**, *5*, e3093.
49. Beery, S.; Cole, E.; Parker, J.; Perona, P.; Winner, K. Species distribution modeling for machine learning practitioners: A review. In ACM SIGCAS conference on computing and sustainable societies 2021, 329–348.
50. Zurell, D.; Franklin, J.; König, C.; Bouchet, P. J.; Dormann, C. F.; Elith, J.; ... Merow, C. A standard protocol for reporting species distribution models. *Ecography* **2020**, *43*(9), 1261–1277.
51. Bryn, A.; Bekkby, T.; Rinde, E.; Gundersen, H.; Halvorsen, R. Reliability in distribution modeling – A synthesis and step–by–step guidelines for improved practice. Frontiers in *Ecology and Evolution* **2021**, 830.
52. Blair, M. E.; Le, M. D.; Xu, M. Species distribution modeling to inform transboundary species conservation and management under climate change: Promise and pitfalls. *Frontiers of Biogeography* **2022**, *14*(1).
53. Erauskin–Extramiana, M.; Arrizabalaga, H.; Cabre, A.; Coelho, R.; Rosa, D.; Ibaibarriaga, L.; Chust, G. Are shifts in species distribution triggered by cli–mate change? A swordfish case study. Deep Sea Research Part II: *Tropical Studies in Oceanography* **2020**, *175*, 104666.
54. Hughes, L. Climate change impacts on species interactions: assessing the threat of cascading extinctions. Saving a Million Species: Extinction Risk from Climate Change 2012, 337–359.
55. Zhu, Y.; Wei, W.; Li, H.; Wang, B.; Yang, X.; Liu, Y. Modelling the potential distribution and shifts of three varieties of Stipa tianschanica in the eastern Eurasian Steppe under multiple climate change scenarios. *Global ecology and conservation* **2018**, *16*, e00501.
56. Zhang, Y.; Clauzel, C.; Li, J.; Xue, Y.; Zhang, Y.; Wu, G.; ... Li, D. Identifying refugia and corridors under climate change conditions for the Sichuan snub-nosed monkey (Rhinopithecus roxellana) in Hubei Province, China. *Ecology and Evolution* **2019**, *9*(4), 1680–1690.
57. Douma, J. C.; van Wijk, M. T.; Lang, S. I.; Shaver, G. R. The contribution of mosses to the carbon and water exchange of arctic ecosystems: quantification and relationships with system properties. *Plant, cell environment* **2007**, *30*(10), 1205–1215.
58. Turetsky, M. R. The role of bryophytes in carbon and nitrogen cycling. *The Bryologist* **2003**, *106*(3), 395–409.
59. Gignac, L. D. Bryophytes as indicators of climate change. *The bryologist* **2001**, *104*(3), 410–420.
60. Zanatta, F.; Engler, R.; Collart, F.; Broennimann, O.; Mateo, R. G.; Papp, B.; ... Vanderpoorten, A. Bryophytes are predicted to lag behind future climate change despite their high dispersal capacities. *Nature Communications* **2020**, *11*(1), 5601.
61. Steinbauer, M. J.; Grytnes, J. A.; Jurasinski, G.; Kulonen, A.; Lenoir, J.; Pauli, H.; ... Wipf, S. Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature* **2018**, *556*(7700), 231–234.
62. Proctor, M. C.; Oliver, M. J.; Wood, A. J.; Alpert, P.; Stark, L. R.; Cleavitt, N. L.; Mishler, B. D. Desiccation–tolerance in bryophytes: a review. *The bryologist* **2007**, *110*(4), 595–621.
63. Vitt, D. H.; Crandall–Stotler, B.; Wood, A. Bryophytes: survival in a dry world through tolerance and avoidance. Plant ecology and evolution in harsh environments. In. Rajakaruna, N., Boyd, R., Harris, T.

 doi:10.20944/preprints202304.0367.v1

(eds.) *Plant Ecology and Evolution in Harsh Environments*. Nova Publishers: New York, United states of America, 2014, 267–295.

64. Qazi, A. W.; Saqib, Z.; Zaman–ul–Haq, M. Trends in species distribution modelling in context of rare and endemic plants: A systematic review. *Ecological Processes* **2022**, *11*(1), 1–11.