Predicting species distributions with environmental time series data and deep learning

Austin M. Smith[1*], César Capinha[2,3], Andrew M. Kramer[1]

[1]Department of Integrative Biology, University of South Florida, 4202 E. Fowler Ave SCA 110 Tampa, FL 33620.

[2]Centre of Geographical Studies, Institute of Geography and Spatial Planning, University of Lisbon, Rua Branca Edmée Marques, 1600-276 Lisboa, Portugal

[3]Associated Laboratory Terra, Portugal.

[*]Corresponding author email: amsmith11@usf.edu

Author contributions: AMS: Conceptualization, Data Curation, Methodology, Software, Writing – original draft, Writing – review and editing, Visualization. CC: Conceptualization, Methodology, Software, Writing – review and editing. AMK: Conceptualization, Methodology, Writing – review and editing, Visualization, Supervision.

Abstract

Species distribution models (SDMs) are widely used to gain ecological understanding and guide conservation decisions. These models are developed with a wide variety of algorithms - from statistic-based approaches to more recent machine learning algorithms - but one property they all have in common is the use of predictors that strongly simplify the temporal variability of driving factors. On the other hand, recent architectures of deep learning neural networks allow dealing with fully explicit spatiotemporal dynamics and thus fitting SDMs without the need to simplify the temporal and spatial dimension of predictor data. We present a deep learning-based SDM approach that uses time series of spatial data as predictors, and compare it with conventional modelling approaches, using five well known invasive species. Deep learning approaches provided consistently high performing models while also avoiding the use of pre-processed predictor sets, that can obscure relevant aspects of environmental variation.

Keywords: Environmental niche model; WorldClim; Spatiotemporal dynamics; Convolutional neural networks; Residual networks; Inception time; Long short-term memory recurrent neural networks; AutoML

Introduction

Species distribution models (SDMs) have become indispensable tools for predicting the current and future distribution of species. Correlative-based SDMs measure the association between environmental predictors and species observation records in order to determine the probability of occurrence of species or the environmental suitability in new areas (Guisan & Zimmermann 2000; Elith & Leathwick 2009; Peterson *et al.* 2011; Araújo *et al.* 2019). These models can be built with statistic-based approaches and more recently, machine learning algorithms that have the ability to detect patterns in the data for which they were not explicitly programmed to look for (LeCun *et al.* 2015; Zhang & Li 2017; Christin *et al.* 2019). The flexibility and high performance of these latter approaches have made them the standard technique for several types of SDM-related biogeographical studies including estimating habitat suitability, species range expansion or contraction, invasion risk, and species co-occurrences (Pearson & Dawson 2003; Elith *et al.* 2006; Elith & Leathwick 2009; Peterson *et al.* 2011; Norberg *et al.* 2019).

Since their inception, the methods used in SDMs have improved considerably (Araújo *et al.* 2019), with many new techniques being proposed (e.g., Phillips *et al.* 2006; Renner *et al.* 2015; Norberg *et al.* 2019), modelling workflows and reporting protocols being refined and standardized (Thuiller *et al.* 2009; Araújo *et al.* 2019; Feng *et al.* 2019; Zurell *et al.* 2020), and thorough inter-model comparisons of predictive accuracy being performed (Elith *et al.* 2006; Norberg *et al.* 2019). Thus, through this collective research effort and resulting improvements in model performance it is unclear whether there is room for substantial advancement in the performance or implementation of correlative SDMs. In fact, there is an extensive literature pertaining to their methodological limitations (e.g., Pearson & Dawson 2003; Hijmans 2012; Norberg *et al.* 2019; Milanesi *et al.* 2020). Undoubtably, analytical pitfalls are present even with optimized modeling architectures;

however, model quality will always depend on the availability and use of comprehensive distribution data and ecologically informative predictor variables (Pearson & Dawson 2003).

A common aim of SDMs is to identify the set of environmental combinations under which a species can occur and thrive (e.g., Guisan & Zimmermann 2000; Pearson *et al.* 2002; Pearson & Dawson 2003; Booth *et al.* 2014) and SDMs are typically built with temporally invariant summaries of multi-decadal long environmental variation by means of one or a few descriptive statistics, such as the mean and standard deviation (Elith & Leathwick 2009; Thuiller *et al.* 2009; Norberg *et al.* 2019). However, there are several reasons to believe that these pre-processed features of environmental variation, formulated based on expert opinion, are often not optimal from a predictive perspective. For example, although studies have shown that extreme climatic conditions are highly informative for predicting species distributions, these extremes are often not included in readily available datasets of bioclimatic variables and in SDMs (Zimmermann *et al.* 2009; Reside *et al.* 2010; Morán-Ordóñez *et al.* 2018; Stewart *et al.* 2021).

It must also be acknowledged that most, if not all, factors driving species distributions (e.g., climate, land-use) are temporally dynamic. That is, their state varies along time, a property that is often poorly represented, or entirely missing from SDMs. For example, while the distribution of species is often dependent on both the short- and long-term climatological conditions (Pearson & Dawson 2003; Peterson *et al.* 2011; Stewart *et al.* 2021), SDMs rarely account explicitly for temporal variation in these factors. Similarly, predictors representing recent patterns of land use can neglect the legacy of past land use patterns in shaping the observed species distributions (Polaina *et al.* 2019; Chen & Leites 2020). A more subtle common omission in predictors used in SDMs concerns the order in which events take place (Kriticos *et al.* 2012; Tyberghein *et al.* 2012; Assis *et al.* 2018; Title & Bemmels 2018). For example, areas with similar averaged annual

precipitation could have substantially different seasonality in the occurrence and magnitude of this factor (Figure 1), likely resulting in relevant differences in environmental suitability for species. In summary, the high dimensionality of spatial time-series data contains properties that are relevant for species distributions, but that human expertise may be unable to identify and thus represent in temporally invariant predictor sets.

One way to robustly address these limitations is by calibrating SDMs that consider the full representation of spatial and temporal variability in predictors sets. While this has been largely inaccessible owing to most predictive algorithms requiring tabular-type data, a structure inefficient in representing multidimensional data (Pebesma 2012), more recent deep learning-based models (Botella *et al.* 2018; Christin *et al.* 2019; Deneu *et al.* 2019; Alshahrani *et al.* 2021; Anand *et al.* 2021; Borowiec *et al.* 2021; Huang & Basanta 2021) allow to express this high dimensionality with so called 'tensor-type' data structures. In simple terms, deep learning – a subfield of machine learning – is mainly concerned with the development of models that can automatically process raw, complex, high-dimensional predictor data, and extract useful attributes from it (the so called 'features') without user intervention (LeCun *et al.* 2015; Bengio *et al.* 2021). Recently, Capinha et al. (2021) demonstrated that some deep learning architectures allow using spatial time series data directly as predictors of ecological phenomena, hence overcoming the need of using temporally unvarying, pre-assembled, predictors sets. In addition, SDM practitioners have been also turning to deep learning (Anand *et al.* 2021), but as an alternative algorithms to use in conventional workflows with static predictors (e.g., Botella *et al.* 2018; Benkendorf & Hawkins 2020; Rew *et al.* 2021).

While time-series-based deep learning networks thence have apparent practical and conceptual advantages over conventional approaches in the development of SDMs, to our

knowledge there is no robust comparison of the predictive performance of the two approaches. Hence, we here perform this assessment. Specifically, we use data from a range of taxa and geographical regions and measure the accuracy of predictions of species distributions obtained from deep learning algorithms using spatial time series as predictor variables and from conventional machine learning algorithms using 'standard' predictor variables. We compared and evaluated the results from both approaches to determine whether the theoretical advantages of deep learning with time-series data lead to practical benefits.

Materials and methods

Species data

Our analyses used data for five globally introduced, invasive species: *Cacyreus marshall* (geranium bronze butterfly; Federica *et al.* 2019)*, Harmonia axyridis* (Asian harlequin ladybird; Bidinger *et al.* 2012),), *Myiopsitta monachus* (monk parakeet; Bohl *et al.* 2019), *Pueraria montana* (kudzu; Callen & Miller 2015), and *Sus scrofa* (wild boar; Sales *et al.* 2017). Using data for invasive species matches a common use of SDMs (e.g., Elith & Graham 2009; Gallien *et al.* 2012; Mainali *et al.* 2015; Barbet-Massin *et al.* 2018; Liu *et al.* 2020), while also including common issues of non-stationarity and distribution data quality that may be absent from designed surveys. The number of occurrences and the extent of distribution ranges vary for each species and can be observed in Figure S1.

Ideally, SDMs are built using both presence and absence records; however, most commonly species absences are not recorded. A reasonable alternative is to use pseudo-absences, which are data points pulled from areas lacking recorded occurrences and assumed to represent

125 unsuitable conditions (Phillips *et al.* 2006; Elith & Leathwick 2009; Hijmans *et al.* 2017; Valavi
126 *et al.* 2022). Because SDM results can be sensitive to the spatial extent used for pseudo-absence
127 extraction (VanDerWal *et al.* 2009; Barbet-Massin *et al.* 2012), we performed this extraction from
128 two sampling areas, each representing a distinct assumption. First, as invasive species are often
129 transported widely across the globe (Liu *et al.* 2020), we randomly sampled 10,000 pseudo-
130 absence points worldwide, except in locations where the species is recorded. Alternatively, and
131 considering that the former procedure may overestimate the actual spread of the species (i.e., their
132 absence may be due to dispersal limitations and not to unsuitable conditions), we delimited a
133 second sampling area by creating a 1000-kilometer buffer around each species occurrence point
134 and merging the obtained polygons. By sampling closer to where the species has been recorded,
135 we expect the pseudo-absences to more closely represent areas made available to the species. From
136 these areas we sampled a second set of 10,000 pseudo-absence points, also excluding locations
137 having species records. Importantly, it is common practice to have a significantly larger number
138 of pseudo-absences than presence records; however, having an imbalanced data set can produce
139 inflated statistical scores (Jiménez-Valverde *et al.* 2009; Lunardon *et al.* 2014; Menardi & Torelli
140 2014). Thus, for each species, we used the R package 'ROSE' (Random Over-Sampling
141 Examples; Lunardon *et al.* 2014) to generated a bootstrap-sample of occurrences in each partition
142 to match the number of pseudo-absence points to create a balanced data frame.

143 **Bioclimatic variables**

144 We used the WorldClim (https://www.worldclim.org) database, a collection of high-
145 resolution terrestrial climate and weather data, to build our SDMs. WorldClim (Fick & Hijmans
146 20170 provides both monthly weather maps (i.e., spatiotemporal time-series) spanning several
147 decades, as well as set of 19 static spatial variables summarizing distinct aspects of the long-term

meteorological variation (the so-called 'bioclimatic' variables). These latter variables are derived from a 30-year spatial time series of monthly values of minimum, maximum temperature and precipitation recorded from 1970 to 2000 and are a widely used example of temporally static predictor data being used in conventional SDM approaches (e.g., Pearson & Dawson 2003; Elith *et al.* 2006; Phillips *et al.* 2006; Thuiller *et al.* 2009; Peterson *et al.* 2011; Booth *et al.* 2014; Hijmans *et al.* 2017; Morán-Ordóñez *et al.* 2018; Liu *et al.* 2020). Using this data source allows us to evaluate time-series deep learning networks on an equal footing with conventional methods by using the 30-year spatial time series (i.e., a 360-time step spatial series for each factor) as predictors in the deep-learning models and the 19 static bioclimatic summary variables in conventional algorithms. In addition, previous studies have shown that the inclusion of elevation data provides more accurate predictions (Hof *et al.* 2012; Oke & Thompson 2015; Kiser *et al.* 2022); thus, we also used this data, as provided by WorldClim, in both sets of models. All predictor variables were extracted from 2.5 arc-minute spatial scale raster and polygon layers.

Time series deep learning

Currently, there are several existing deep learning architectures able to deal with temporal variability in predictor data (Box 1.) However, it is generally difficult to know a priori which type of architecture will work best for each task (He *et al.* 2021). As an alternative, 'Mcfly' (Van Kuppevelt et al. 2020), a TensorFlow (https://www.tensorflow.org) wrapper package incorporates automatized model assembly and learning procedures (i.e., AutoML; He *et al.* 2021); this assembles a collection of candidate models of different architectures with randomly defined hyperparameters, which are then trained, tuned, and tested to determine the most suitable model architecture and parametrization for the data being modeled. Mcfly offers four deep learning architectures able to use time series as predictor variables: convolutional neural networks (CNN);

171 Residual networks, (ResNet); Inception Time (IT); and long short-term memory recurrent neural
172 networks (LSTM) (Box 1.). We use the AutoML feature of this package to generate five randomly
173 assembled candidate models of each architecture (i.e., 20 total candidate models) and applied them
174 to each species and buffer combination.

175 A standard way to evaluate the predictive accuracy of SDMs is to measure its ability to
176 predict part of the distribution data left aside from model calibration (Guisan & Zimmermann
177 2000; Norberg *et al.* 2019). As such, we partitioned our data into three sets; training data used for
178 learning or calibrating a model; validation data for evaluating a model's performance during
179 training or hyperparameter tuning; and a test set to evaluate the final model. Training, validation,
180 and testing sets were 70%, 15%, 15% of the total dataset respectively. To reduce the computational
181 costs in candidate model selection phase, we trained each candidate model using a random subset
182 of ~50% of the training data during four calibration epochs, allowing an initial tuning of
183 hyperparameters. Each model was then evaluated using classification accuracy on the validation
184 data**.** The candidate model (out of the initial 20) with the highest accuracy score, was then trained
185 a model with the full training set during 50 epochs. The predictive accuracy of the model in each
186 epoch was evaluated using the validation data and measured with the area under the receiver
187 operating curve (AUC). The AUC is a favored metric in SDM studies since it is independent of
188 class prevalence (McPherson *et al.* 2004) and is threshold independent as opposed to metrics based
189 on a single, often, arbitrary threshold (Brotons *et al.* 2004; Shabani *et al.* 2018). Finally, for each
190 species and type of pseudo-absence data, we retained the epoch model with best AUC as our final
191 model and evaluated it with the test data.

192 All deep learning models were built in Python Ver. 3.7.1 with the Keras Ver. 2.2.4
193 (https://keras.io) and TensorFlow Version 2.5 (https://www.tensorflow.org) deep learning suites,

and the Mcfly Ver. 3.1.0 (Van Kuppevelt et al. 2020). For further review, model code and data can be found at Zenodo.org (DOI: 10.5281/zenodo.7255395), which should also allow non-experts to easily implement these methods.

Conventional machine learning

We used several geospatial packages for conventional 'static' species distribution modeling in R (Ver. 4.0.3) , most notably 'dismo' (Hijmans *et al.* 2017). The 19 bioclimatic predictors were used for training and testing these models and are extracted from the raw time series data used in the deep learning models. We used three popular and typically well performing machine learning algorithms: Gradient Boosting Machines (GBM; e.g., Friedman 2001); Maximum Entropy (MaxEnt; Phillips *et al.* 2006; Elith *et al.* 2011); and Random Forest (RF; Breiman 2001; Cutler *et al.* 2007).

We used the same training and testing samples as for the deep learning models to build and test the static models. Note that, unlike the deep leaning networks, the conventional algorithms use internal validation metrics, and thus it is not necessary to manually define validation datasets. For both deep learning and conventional models, we created global prediction maps. These maps are ranked gradient plots that represent variance and spatial uncertainty (Guisan & Zimmermann 2000), which is often more important in studies pertaining to invasive species (Bidinger *et al.* 2012; Sales *et al.* 2017; Barbet-Massin *et al.* 2018; Liu *et al.* 2020). We compared the deep learning predictions to the conventional predictions to measure model dissimilarity (Figure 2) by subtracting the raster pixel values of the conventional methods from those in deep learning rasters. Value of 0 indicated model agreement, positive values as areas where deep learning showed stronger suitability prediction, and negative values as areas where the conventional algorithm favored.

Results

There were no clear patterns regarding which candidate deep learning architectures were selected for each species and pseudo-absence extraction strategy (Table 1). LSTMs were selected the most frequently (4), followed by CNN (3), IT (2), and ResNet (1). LSTM were selected for two of five species, with CNN, IT, and ResNet being selected once for models using pseudo-absences extracted from 1000km buffers. Both LSTM and CNN were selected twice, and IT selected once for models using pseudo-absences extracted at the global scale. *Myiopsitta monachus* was the only species where both models were built using the same architecture, LSTM.

The time-series deep learning models using pseudo-absences extracted at the global scale scored an average AUC of 0.977 across all five species, with the lowest AUC measured for the wild boar model (0.925) and highest being the kudzu model (0.994) (Figure 3A; Table 1). The performances of these models were consistently equal to or second to the best conventional method, with a small average deviance from the top scoring model of 0.002. For the conventional models, RF scored the best for three species (*Cacyreus marshall*; *Pueraria montana*; and *Sus scrofa*) and MaxEnt for the other two species. GBM scored the lowest for all five species but only saw an average deviance of 0.014 and all species-model combinations scored an AUC above 0.9. When using pseudo-absences extracted from 1000km buffers, all algorithms achieved lower, but still very good AUC values (average AUC across all models = 0.932; Figure 3B). The deep learning models scored AUC values above 0.9 for all five species, with only RF doing the same. Deep learning models scored the best for *Cacyreus marshall*, *Harmonia axyridis*, and *Sus scrofa* while RF and Maxent scored the best for *Pueraria montana* and *Myiopsitta monachus* respectively, though DL placed second in both cases. GBM consistently placed the lowest among the algorithms for all five species with an average deviance of 0.032 from the best scoring algorithm.

Each model produced a global prediction raster which ranked location suitability from 0 - 1 (Figures S2:S11). Deep learning models predicted higher suitability scores (0.8 -1) around occurrence points and a smoother gradient between them. MaxEnt and RF models most frequently showed higher suitability scores at more isolated locations, generally near the recorded occurrences. GBM models were the most inconsistent as they showed some degree of suitability on all parts of the globe. For all five species, models built with global background points showed more conservative ranges where models using background points pulled from a 1000km buffer showed a more expanded potential range.

For both sets of models, deep learning ranked suitable locations higher than MaxEnt or RF even when model predictions agreed, while deep learning and GBM models showed a high degree of dissimilarity (Figures S12:S21). When comparing the results using point from different spatial extents, deep learning models showed the most agreement between predictions, while conventional methods were more variable (Figure S22:S26).

Discussion

To our knowledge, this is the first study comparing SDMs built from spatial time series based deep-learning models to conventional machine learning algorithms. Deep learning can address several conceptual concerns seen in conventional modeling methods, especially potential human-mediated omissions, or bias in selection of features to represent in the predictor dataset. These methods also allow the model to account for temporal environmental variation, something known to be biologically relevant and not well-represented in pre-assembled sets of bioclimatic variables (Morán-Ordóñez et al., 2018, Reside et al., 2010, Stewart et al., 2021, Zimmermann et al., 2009). Here we show the performance of randomly generated deep learning models was

similar to or exceeded conventional static methods. Our results extend the findings from previous studies (Botella *et al.* 2018; Benkendorf & Hawkins 2020; Anand *et al.* 2021; Capinha *et al.* 2021; Rew *et al.* 2021) showing that deep learning is an extremely powerful approach, and suggest that these models should be increasingly considered for species distribution modeling, especially with the use of spatial time series data.

Each of the four deep learning architectures were selected for at least one species and pseudo-absence combination, with LSTM and CNN being the most commonly selected. Considering that we use a relatively small sample of candidate models (n = 20) for each species and type of pseudo-absence extraction strategy, there are likely to be significant gains in performance by understanding the strengths of different architectures and their components and doing more intensive optimization of hyper-parameters. Like conventional ANNs, the convergence of training and validation AUCs and model accuracy are dependent on the number of training epochs performed (Fawaz *et al.* 2019; Benkendorf & Hawkins 2020; Capinha *et al.* 2021). While our models ran 50 epochs, further analysis is recommended on the effect this has on time series DL-SDMs.

We used a simple random subset of the data to select among candidate architectures of deep learning models and to determine the maximum number of epochs that the models should train for (i.e., the number of times the full set of training data passes through the network, adjusting model weights). This approach still prevails in the field (Araújo et al., 2019) and here has produced robust final models (e.g., when comparing its predictive performance to most of the conventional models); however, it might also be worth exploring data splitting strategies that more carefully consider the non-independence of data points, as caused for instance by spatial autocorrelation (Araújo et al., 2019; Ploton et al., 2020), including the spatial cross-validation method (Elith &

Leathwick 2009; Hijmans 2012; Wenger & Olden 2012; Roberts *et al.* 2017; Ploton *et al.* 2020). Thus, one challenge that deserves particular attention is that of model overfitting. Given the high number of trainable parameters found in deep learning models, particular care should be taken to prevent overfitting (Li *et al.* 2019; Benkendorf & Hawkins 2020). Under these circumstances, the data splitting strategies used to inform the model about 'the reality it should search for' takes on significant relevance.

Measuring the differences in spatial predictions between suitability maps resulting from different algorithms trained on the same data is one way to assess predictive uncertainty (Kearney *et al.* 2010; Beale & Lennon 2012; Iturbide *et al.* 2018). In this study, we compared prediction by calculating the difference between time series deep learning maps to the conventional algorithms, and by comparing the predictions across two different spatial extents. Our results show that time series deep learning models generally predicted an area of relatively high suitability intermediate to the range of some conventional algorithms. Also, deep learning models were the most similar across both spatial extents, which can be a common problem in conventional workflows. However, we note that determining which model is not a trivial process but requires extensive validation assessment to remove uncertainty (Huang *et al.* 2018; Iturbide *et al.* 2018; Norberg *et al.* 2019; Grimmett *et al.* 2020). Converting to binary maps can simplify the decision process, but only when an appropriate threshold is implemented, which is equally difficult to evaluate (Liu *et al.* 2005; Jiménez-Valverde & Lobo 2007; Liu *et al.* 2009). Thus, further review of these methods should be explored, including whether it makes sense to make ensemble models between conventional and deep learning predictions, or between deep learning architectures.

Despite the results obtained, it is unlikely that time-series models will always outperform conventional approaches. For example, in cases where deep learning models performed best, the

difference was generally small to the best performing conventional model. This is likely to be the common case for large-scale distribution patterns, that respond strongly to general, long-term patterns of climate (Pearson & Dawson 2003), which tend to be well represented in the existing repositories of pre-processed climatic features. Instead, we argue that the capacity of time-series deep learning models for considering a higher dimensionality of possibly relevant features makes them better equipped to make accurate predictions under a wider diversity of settings, spanning from distributions shaped by the simpler patterns of climate to those resulting from an intricate web of relationships involving complex spatial and temporal dynamics of multiple factors.

In addition, although time-series deep learning models could bring substantial benefits to SDMs, there are also several other limitations worth addressing, though most of these are also faced by conventional approaches. These include spatial autocorrelation (F. Dormann *et al.* 2007), extrapolation errors (Liu *et al.* 2020), data imbalance (Jiménez-Valverde *et al.* 2009), or sampling bias (Fithian *et al.* 2015). This assumes particular significance if the predictions are aimed for new regions or time periods, i.e., are to be 'transferred' (e.g., Yates *et al.* 2018; Liu *et al.* 2020; Taheri *et al.* 2021). The challenges imposed by these issues are inherent to the early stage of using these models for species distribution modelling. Hence, we expect that, as these issues become increasingly explored, the hurdles they cause will become resolved to some extent − in a similar manner to what occurred for conventional models. However, it is important not to underestimate the potential complexity of these tasks, particularly given the higher dimensionality of the predictor data that are now involved (for example, our conventional static machine learning models were given ~380,000 measurements to process i.e., 19 BIO variables × ~20,000 instances, where the time-series DL models were fed ~28,800,000 i.e., 4 time series × ~20,000 instances × 360 time steps.

Conclusion

We have described and demonstrated conceptual and practical benefits of deep learning models for predicting species distributions. The capacity of these models to automatically identify relevant features from, high-dimensional, temporal data reduces reliance on human supervision in the definition of relevant environmental features to include in the models. This can advance the field by providing robust predictions even when a priori knowledge about the features that are most influential in shaping species distributions is limited, while matching the performance of existing methods in well understood systems. However, several challenges still need to be addressed before the full potential of these models becomes realized. We hope to facilitate and encourage ecological modellers to explore, test and help overcome the limitations of deep learning models to advance understanding of species distributions, much in the way that conventional models were explored and improved over the last twenty years.

References

1.
Alshahrani, H.M., Al-Wesabi, F.N., Al Duhayyim, M., Nemri, N., Kadry, S. & Alqaralleh, B.A. (2021). An automated deep learning based satellite imagery analysis for ecology management. *Ecological Informatics*, 66, 101452.

2.
Anand, A., Pandey, M.K., Srivastava, P.K., Gupta, A. & Khan, M.L. (2021). Integrating Multi-Sensors Data for Species Distribution Mapping Using Deep Learning and Envelope Models. *Remote Sensing*, 13, 3284.

3.
Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R. *et al.* (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5, eaat4858.

4.
Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A. & De Clerck, O. (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27, 277-284.

5.
Barbet-Massin, M., Rome, Q., Villemant, C. & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PloS one*, 13, e0193085.

6.
Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in ecology and evolution*, 3, 327-338.

7.
Beale, C.M. & Lennon, J.J. (2012). Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 247-258.

8.
Bengio, Y., Lecun, Y. & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM*, 64, 58-65.

9.
Benkendorf, D.J. & Hawkins, C.P. (2020). Effects of sample size and network depth on a deep learning approach to species distribution modeling. *Ecological Informatics*, 60, 101137.

10.
Bidinger, K., Lötters, S., Rödder, D. & Veith, M. (2012). Species distribution models for the alien invasive Asian Harlequin ladybird (Harmonia axyridis). *Journal of Applied Entomology*, 136, 109-123.

11.
Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford university press.

12.
Bohl, C.L., Kass, J.M. & Anderson, R.P. (2019). A new null model approach to quantify performance and significance for ecological niche models of species distributions. *Journal of Biogeography*, 46, 1101-1111.

13.

Booth, T.H., Nix, H.A., Busby, J.R. & Hutchinson, M.F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, 20, 1-9.

14.

Borowiec, M.L., Dikow, R.B., Frandsen, P.B., McKeeken, A., Valentini, G. & White, A.E. (2021). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*.

15.

Botella, C., Joly, A., Bonnet, P., Monestiez, P. & Munoz, F. (2018). A deep learning approach to species distribution modelling. In: *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Springer, pp. 169-199.

16.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

17.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27, 437-448.

18.

Callen, S.T. & Miller, A.J. (2015). Signatures of niche conservatism and niche shift in the North American kudzu (Pueraria montana) invasion. *Diversity and Distributions*, 21, 853-863.

19.

Capinha, C., Ceia-Hasse, A., Kramer, A.M. & Meijer, C. (2021). Deep learning for supervised classification of temporal data in ecology. *Ecological Informatics*, 61, 101252.

20.

Chen, X. & Leites, L. (2020). The importance of land-use legacies for modeling present-day species distributions. *Landscape Ecology*, 35, 2759-2775.

21.

Christin, S., Hervet, É. & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10, 1632-1644.

22.

Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

23.

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. *et al.* (2007). Random forests for classification in ecology. *Ecology*, 88, 2783-2792.

24.

Deneu, B., Servajean, M., Botella, C. & Joly, A. (2019). Evaluation of Deep Species Distribution Models Using Environment and Co-occurrences. Springer International Publishing Cham, pp. 213-225.

25.

Elith, J. & Graham, C.H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32, 66-77.

26.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A. *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

27.
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, 677-697.
28.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, 17, 43-57.
29.
F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G. *et al.* (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609-628.
30.
Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L. & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33, 917-963.
31.
Federica, P., Francesca, M., Cristiana, C., Silvia, G., Ramona, V., Sara, C. *et al.* (2019). From Africa to the Alps: risk assessment on an invasion by Cacyreus marshalli (Butler, 1898). *Journal of Insect Conservation*, 23, 279-288.
32.
Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C. & Papeş, M. (2019). A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution*, 3, 1382-1395.
33.
Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37, 4302-4315.
34.
Fithian, W., Elith, J., Hastie, T. & Keith, D.A. (2015). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6, 424-438.
35.
Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
36.
Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E. & Thuiller, W. (2012). Invasive species distribution models–how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, 21, 1126-1136.
37.
Grimmett, L., Whitsed, R. & Horta, A. (2020). Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. *Ecological Modelling*, 431, 109194.
38.
Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, 135, 147-186.
39.
He, X., Zhao, K. & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622.

40.
Hijmans, R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93, 679-688.
41.
Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. & Hijmans, M.R.J. (2017). Package 'dismo'. *Circles*, 9, 1-68.
42.
Hof, A.R., Jansson, R. & Nilsson, C. (2012). The usefulness of elevation as a predictor variable in species distribution modelling. *Ecological Modelling*, 246, 86-90.
43.
Huang, Q., Fleming, C.H., Robb, B., Lothspeich, A. & Songer, M. (2018). How different are species distribution model predictions?—Application of a new measure of dissimilarity and level of significance to giant panda Ailuropoda melanoleuca. *Ecological Informatics*, 46, 114-124.
44.
Huang, Y.-P. & Basanta, H. (2021). Recognition of Endemic Bird Species Using Deep Learning Models. *IEEE Access*, 9, 102975-102984.
45.
Iturbide, M., Bedia, J. & Gutiérrez, J.M. (2018). Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, 166, 19-29.
46.
Jiménez-Valverde, A., Lobo, J. & Hortal, J. (2009). The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, 10, 196-205.
47.
Jiménez-Valverde, A. & Lobo, J.M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta oecologica*, 31, 361-369.
48.
Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010). Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation letters*, 3, 203-213.
49.
Kiser, A.H., Cummings, K.S., Tiemann, J.S., Smith, C.H., Johnson, N.A., Lopez, R.R. *et al.* (2022). Using a multi-model ensemble approach to determine biodiversity hotspots with limited occurrence data in understudied areas: An example using freshwater mussels in México. *Ecology and Evolution*, 12, e8909.
50.
Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. *et al.* (2012). CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3, 53-64.
51.
LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *nature*, 521, 436-444.
52.

Li, H., Li, J., Guan, X., Liang, B., Lai, Y. & Luo, X. (2019). Research on overfitting of deep learning. In: *2019 15th International Conference on Computational Intelligence and Security (CIS)*. IEEE, pp. 78-81.

53.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.

54.

Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: a review. In: *Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia*, p. 4247.

55.

Liu, C., Wolter, C., Xian, W. & Jeschke, J.M. (2020). Species distribution models have limited spatial transferability for invasive species. *Ecology Letters*, 23, 1682-1692.

56.

Lunardon, N., Menardi, G. & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *R journal*, 6.

57.

Mainali, K.P., Warren, D.L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G. *et al.* (2015). Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global change biology*, 21, 4464-4480.

58.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of applied ecology*, 41, 811-823.

59.

Menardi, G. & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28, 92-122.

60.

Milanesi, P., Della Rocca, F. & Robinson, R.A. (2020). Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. *Ecology and evolution*, 10, 1087-1092.

61.

Morán-Ordóñez, A., Briscoe, N.J. & Wintle, B.A. (2018). Modelling species responses to extreme weather provides new insights into constraints on range and likely climate change impacts for Australian mammals. *Ecography*, 41, 308-320.

62.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J. *et al.* (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89, e01370.

63.

Oke, O.A. & Thompson, K.A. (2015). Distribution models for mountain plant species: the value of elevation. *Ecological Modelling*, 301, 72-77.

64.

Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography*, 12, 361-371.

65.
Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P. (2002). SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological modelling*, 154, 289-300.
66.
Pebesma, E. (2012). spacetime: Spatio-temporal data in R. *Journal of statistical software*, 51, 1-30.
67.
Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. *et al.* (2011). *Ecological niches and geographic distributions (MPB-49)*. Princeton University Press.
68.
Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190, 231-259.
69.
Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V. *et al.* (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications*, 11, 1-11.
70.
Polaina, E., González-Suárez, M. & Revilla, E. (2019). The legacy of past human land use in current patterns of mammal distribution. *Ecography*, 42, 1623-1635.
71.
Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J. *et al.* (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6, 366-379.
72.
Reside, A.E., VanDerWal, J.J., Kutt, A.S. & Perkins, G.C. (2010). Weather, not climate, defines distributions of vagile bird species. *PloS one*, 5, e13569.
73.
Rew, J., Cho, Y. & Hwang, E. (2021). A robust prediction model for species distribution using bagging ensembles with deep neural networks. *Remote Sensing*, 13, 1495.
74.
Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G. *et al.* (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913-929.
75.
Sales, L.P., Ribeiro, B.R., Hayward, M.W., Paglia, A., Passamani, M. & Loyola, R. (2017). Niche conservatism and the invasive potential of the wild boar. *Journal of Animal Ecology*, 86, 1214-1223.
76.
Shabani, F., Kumar, L. & Ahmadi, M. (2018). Assessing accuracy methods of species distribution models: AUC, specificity, sensitivity and the true skill statistic. *Global Journal of Human Social Science*, 18, 6-18.
77.
Stewart, S., Elith, J., Fedrigo, M., Kasel, S., Roxburgh, S., Bennett, L. *et al.* (2021). Climate extreme variables generated using monthly time-series data improve predicted distributions of plant species. *Ecography*, 44, 626-639.

78.
Taheri, S., Naimi, B., Rahbek, C. & Araújo, M.B. (2021). Improvements in reports of species redistribution under climate change are required. *Science Advances*, 7, eabe1110.
79.
Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369-373.
80.
Title, P.O. & Bemmels, J.B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, 41, 291-307.
81.
Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global ecology and biogeography*, 21, 272-281.
82.
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, e01486.
83.
VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological modelling*, 220, 589-594.
84.
Wenger, S.J. & Olden, J.D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3, 260-267.
85.
Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S. *et al.* (2018). Outstanding challenges in the transferability of ecological models. *Trends in ecology & evolution*, 33, 790-802.
86.
Zhang, J. & Li, S. (2017). A Review of Machine Learning Based Species' Distribution Modelling. In: *2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*. IEEE, pp. 199-206.
87.
Zimmermann, N.E., Yoccoz, N.G., Edwards, T.C., Meier, E.S., Thuiller, W., Guisan, A. *et al.* (2009). Climatic extremes improve predictions of spatial patterns of tree species. *Proceedings of the National Academy of Sciences*, 106, 19723-19728.
88.
Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J. *et al.* (2020). A standard protocol for reporting species distribution models. *Ecography*, 43, 1261-1277.

TABLE 1 Summary of deep learning model selection and performance statistics.

Global

Species	Architecture	Accuracy	Best epoch	Validation AUC	Test AUC
Cacyreus marshall	CNN	0.968	14	0.988	0.986
Harmonia axyridis	LSTM	0.933	2	0.989	0.988
Myiopsitta monachus	LSTM	0.963	38	0.997	0.991
Pueraria montana	CNN	0.979	3	0.995	0.994
Sus scrofa	IT	0.786	6	0.92	0.925

1000km Buffer

Species	Architecture	Accuracy	Best epoch	Validation AUC	Test AUC
Cacyreus marshall	LSTM	0.843	46	0.923	0.924
Harmonia axyridis	LSTM	0.863	14	0.943	0.952
Myiopsitta monachus	CNN	0.889	50	0.966	0.962
Pueraria montana	IT	0.908	2	0.971	0.971
Sus scrofa	ResNet	0.731	26	0.901	0.908

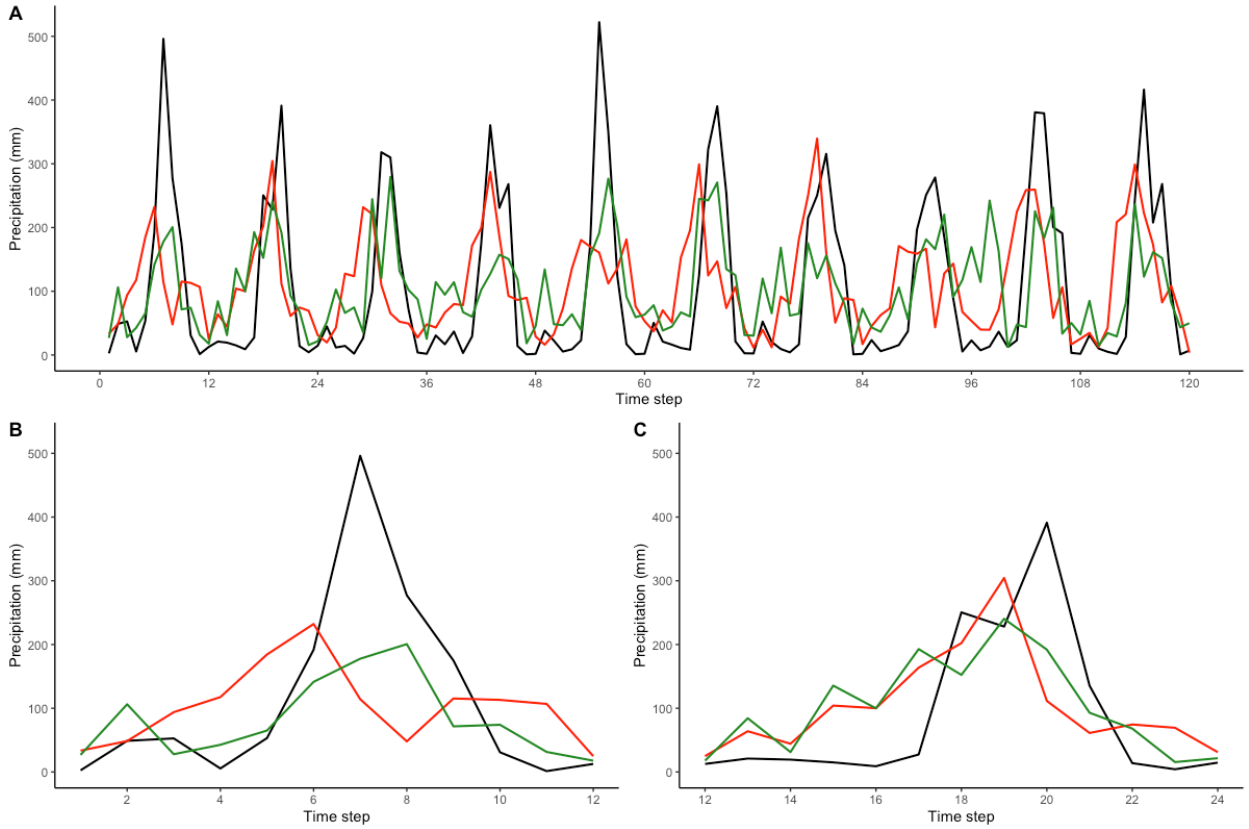

FIGURE 1 A.) Precipitation data from the 1990 - 2000 for three locations that fall along the same lines of latitude. The locations include Tampa, FL, US (green line), Nainapur, Uttar Pradesh, India (black line), and Songtao Miao Autonomous County, Tongren, Guizhou, China (red line). These points were extracted from WorldClim BIO12, which measured averaged annual precipitation. All three points have roughly the same pixel value of 1260 mm. B.) precipitation for the year 1990 and C.) 1991. Note, precipitation not only differs spatial, but also temporally.

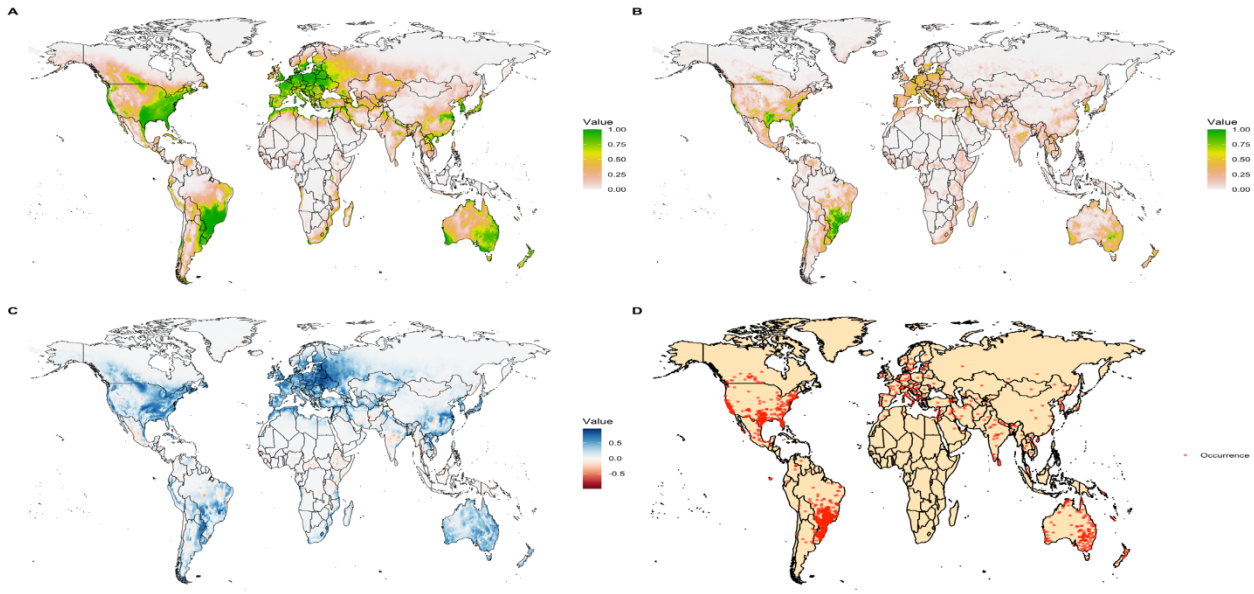

FIGURE 2 Visual example of global predictions for global *Sus scrofa* models, with both A.) Deep Learning and B.) Random Forest, the best scoring conventional method. C.) Difference measure of the two models, which is produced by subtracting the pixel values of Random Forest from Deep Leaning. Value of 0 indicates model agreement, where positive values (blue scale) are areas where Deep Learning has a stronger suitability prediction negative value (red scale) are areas where Random Forest favor. D.) The raw occurrence data used in the modeling process.

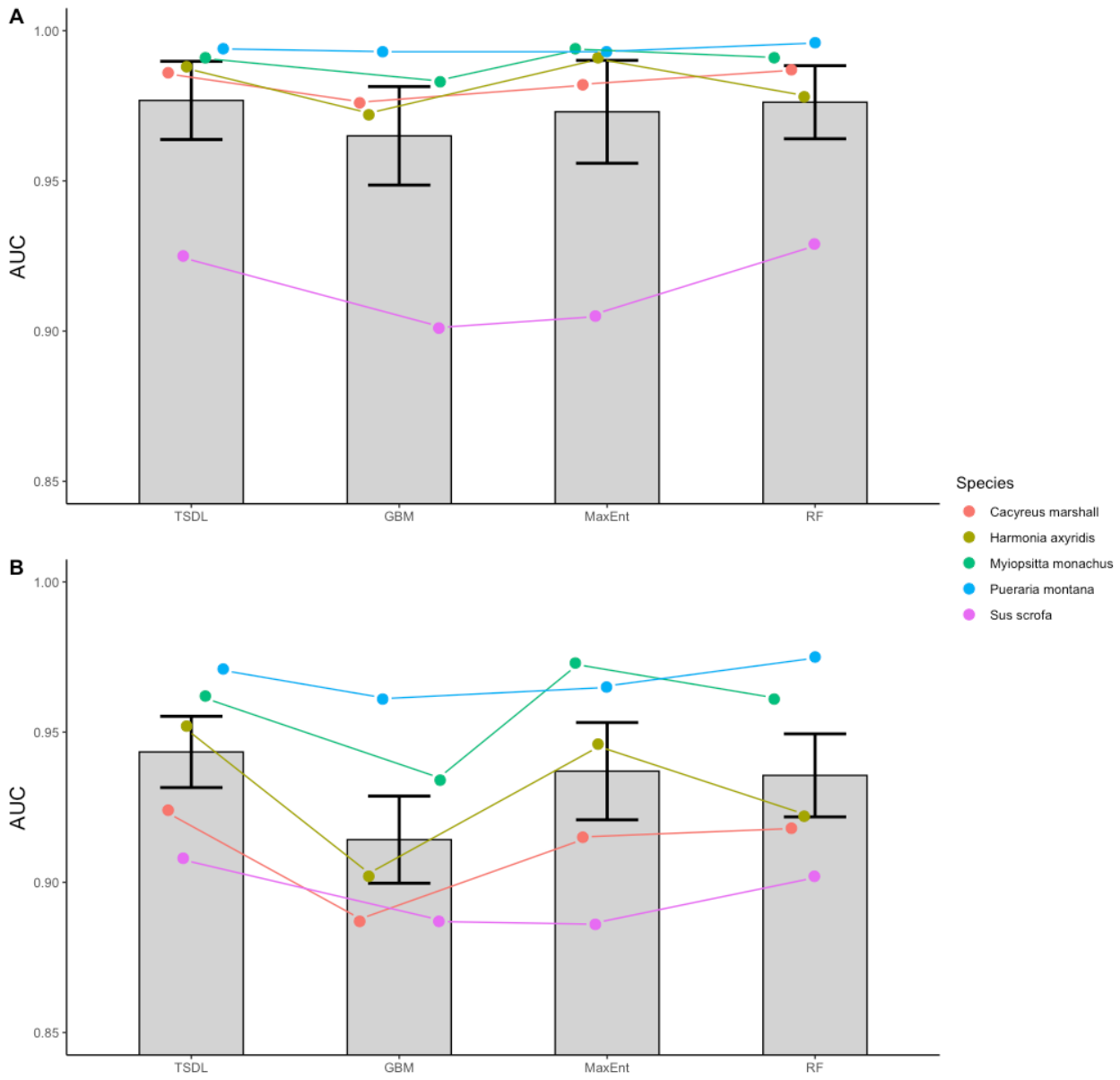

FIGURE 3 A comparison of time series deep learning (TSDL) to gradient boosting machine (GBM), maximum entropy (MaxEnt), and random forest (RF) built using static bioclimatic variables. Models were built from both A.) a global spatial extent and B.) a reduced, 1000-kilometer extent. Gray bars represent the average AUC across all species and include standard error bars.

BOX 1 Summary of deep learning architectures.

Deep learning is a complex network of processing levels known as artificial neural networks (ANNs) meant to replicate the functions of the human brain. ANNs are characterized by several stacked layers of information processing units or 'neurons', which are capable of transforming input data into simpler features. Traditional ANNs, like most modern algorithms, are constructed from pre-selected features and follow a feed-forward process; each iteration or epoch of the learning processes all the input data first, recalibrates internal hyperparameters, and then attempts to optimize until statistical convergence (Bishop 1995). Deep learning neural networks, on the other hand, often consists of multiple layers of ANNs processing at each internal layer.

Convolutional neural networks (CNNs) are one of two dominant categories of deep learning architectures. As the name implies, CNNs refer to convolutional layers; in essence, internal filter functions of varying length convolve with patches of the time series data to measure how much these represent features of presumed relevance. The filtered features are then processed in rectification and pooling layers, which transform data and reduce feature dimensionality for further analysis. The procedure can be replicated along stacked layers, resulting in a hierarchy of increasingly complex features. The final processing layer is a fully connected network that resembles a conventional ANN and is where classification outputs are generated. To optimize the learning objectives, different layers are added to the standard CNN protocol. For example, Residual networks (ResNets) are CNNs with additional process known as Residual Blocks which allow networks to perform both a feed-forward process and data processed in layers several steps ahead simultaneously. Likewise, Inception Time networks are a

sort of hybrid with standard CNNs and ResNet such that they use components of CNN convolutions and ResNet Residual Blocks, but in parallel.

The second category of architectures is recurrent neural networks (RNN), which were specifically designed to analyze sequential data (Fawaz *et al.* 2019). These models incorporate feedback loops, which allow models to learn from their own predictions, as well as the general trends in each layer. Standard RNNs tend to prioritize short-term signals over long-term trends (Chung *et al.* 2014). Therefore, the inclusion of "gated-units", or algorithmic protocols that determine if networks should remember or forget information, were incorporated, thus forming the basis of LSTM models.