Geoscientific
Model Development
Discussions

Predicting Global Terrestrial Biomes with Convolutional Neural Network

Hisashi SATO[1], Takeshi ISE[2]

[1]Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC),
5 Yokohama, 236-0001, JAPAN
[2]Field Science Education and Research Center (FSERC), Kyoto University, Kyoto, 606-8502, JAPAN

Correspondence to: Hisashi SATO (hsatoscb@gmail.com)

Abstract. A biome is a major regional ecological community characterized by distinctive life forms and principal plants.
Many empirical schemes such as the Holdridge Life Zone (HLZ) system have been proposed and implemented to predict the
10 global distribution of terrestrial biomes. Knowledge of physiological climatic limits has been employed to predict biomes,
resulting in more precise simulation, however, this requires different sets of physiological limits for different vegetation
classification schemes. Here, we demonstrate an accurate and practical method to construct empirical models for biome
mapping: A convolutional neural network (CNN) was trained by an observation-based biome map, as well as images
depicting air temperature and precipitation. The trained model accurately simulated a global map of current terrestrial biome
15 distribution. Then, the trained model was applied to climate scenarios toward the end of the 21st century, predicting a
significant shift in global biome distribution with rapid warming trends. Our results demonstrate that the proposed CNN
approach can provide an efficient and objective method to generate preliminary estimations of the impact of climate change
on biome distribution. Moreover, we anticipate that our approach could provide a basis for more general implementations to
build empirical models of other climate-driven categorical phenomena.

20 **1 Introduction**

Terrestrial biomes and climate are among the earliest known ecological concerns, and many empirical schemes have been
proposed to characterize their relationship (Prentice and Leemans, 1990). One of the best known of these schemes is the
Holdridge Life Zone (HLZ) system (Holdridge, 1947), which classifies vegetation distribution using only two independent
variables: the annual mean precipitation and the bio-temperature (i.e., mean of positive air temperature). Due to its simplicity,
25 this scheme has been extensively implemented in numerous studies (Emanuel et al., 1985; Henderson-Sellers, 1991; Lugo et
al., 1999; Monserud and Leemans, 1992; Prentice, 1990).

Despite its relative simplicity, the HLZ scheme accounts well for ecophysiological constraints. This scheme is based on bio-
temperatures, given that plant productivity becomes negligible at temperatures below 0°C. Furthermore, it employs
logarithmic conversions to better depict the relationship between climatic parameters and life zone boundaries, in
30 quantitative recognition of the temperature control of metabolic processes. However, since the HLZ scheme only considers

annual climate means, it cannot account for climatic tolerance (e.g., minimum and maximum temperatures) nor the occurrence and extent of drought seasons, both of which substantially affect biome distribution (Prentice et al., 1992).

Efforts have been made to develop biome-mapping schemes that incorporate these environmental constraints. These implementations are considered to have clear physiological bases (Prentice et al., 1992; Woodward and Williams, 1987), and

35   their predictions simulate present-day distributions of vegetation more accurately than the HLZ scheme. However, an important drawback of this type of approach is that it requires absolute physiological limits for each vegetation type or plant functional type (PFT), for which there is still insufficient comprehensive information, as this cannot be estimated from the geographical distribution of the vegetation (Lavorel et al., 2007). Making matters more difficult, researchers do not share the same classification criteria for terrestrial biomes, and the number of vegetation types or PFTs varies widely from

40   five(Henderson-Sellers, 1991) to almost 100 (Box, 1981), depending on the research purpose and the geographical scale studied. By contrast, empirical approaches like the HLZ scheme do not require detailed physiological data, and thus have the advantage of being easily applicable to any given vegetation classification criteria. Recently, empirical models for biome mapping using various types of environmental data have been developed by employing multinomial logistic regression (Levavasseur et al., 2012; Levavasseur et al., 2013) and machine learning algorithms (Hengl et al., 2018).

45   Here, we demonstrate an accurate and practical method to construct empirical models for operational global biome mapping via a convolutional neural network (CNN) approach. Recently, CNN has been successfully adapted for use in species distribution modelling at regional scales (Benkendorf and Hawkins, 2020; Botella et al., 2018); however, it has not been used to develop global biome models. A CNN is an algorithm for machine learning in which a model learns to conduct classification tasks directly from training data. Model training of a CNN is based on finding patterns in the spatial

50   organization of the training data (typically images) that recognizes its classification well. Unlike other conventional algorithms for machine learning, CNN learns directly from training data without a requirement for manual feature extraction. For example, to account for seasonal variability, previous correlative climate-vegetation models needed to pre-define representative variables. Levavasseur et al. (2013) divided each climatic variable into four "seasonal" predictors by averaging data corresponding 3-month periods (i.e., DJF for winter, MAM for spring, JJA for summer, and SON for fall). By

55   contrast, CNN can automatically extract non-linear seasonal patterns for climatic variables that are relevant in biome classification.

Indeed, Botella et al. (2018) empirically demonstrated that a CNN model performed better at reconstructing species distributions than the popular species distribution modelling method, MAXENT (Phillips et al., 2006). This higher performance was attributed to CNN's efficient use of spatial patterns in environmental variables, which often control species

60   distribution. MAXENT ignores these spatial patterns. A second explanation for the improved performance is that CNN can treat high order interaction effects between input variables, whereas MAXENT, like the majority of other methods, only represents interactions between environmental variables by the products of variable pairs.

Here, we trained a CNN with a biome distribution map based on observed data and graphs depicting climate data. While previous applications of CNN for vegetation classification have been based on maps of environmental variables (such as

Geoscientific
Model Development
Discussions

65   climate and topology), we follow Ise and Oba (2019) and Ise and Oba (2020), who demonstrated that climatic conditions, represented using graphical images, would be used as training data for CNN models. After evaluating the accuracy of the biome map reconstructed by this method, we applied the trained CNN to climatic scenarios toward the end of the 21st century, and predicted significant shifts in global biome distribution with rapid warming trends.

2 Methods

2.1 Data

For training the CNN model, we employed potential land cover types and the monthly climate information from the ISLSCP2 (Ramankutty and Foley, 2010) and CRU TS4.00 (Harris and Jones, 2017) datasets, respectively. Both datasets have a 0.5° global surface grid resolution. The ISLSCP2 dataset is an observation-based biome map, which classifies the global land surface into 15 vegetation types (Fig. 1a). The CRU TS4.00 is based on an archive of climatic conditions

75   observed in more than 4,000 weather stations distributed worldwide. Climatic conditions between 1971 and 1980 were selected for CNN training since this time period is just before the beginning of a clear global warming trend (Rood, 2015), and the number of meteorological stations that contributed to the dataset remained relatively stable (Harris et al., 2014).

In machine learning experiments, a fraction of the training data is typically divided randomly into two subsets, of which one is used for model training, and the other is then used to validate the trained model. This study used the CRU TS4.0 climate

80   data as training data, which was generated by interpolating data from weather stations, meaning that values in each grid are not independent of those in nearby grids. Under these circumstances, validation using the typical procedures described above would risk overfitting (i.e., training the model too closely or exactly to a particular set of data, thereby creating a model that may fail to fit additional data or reliably predict future observations) (Leinweber, 2007). Therefore, other climate datasets were used for validating the trained model: NCEP/NCAR reanalysis, and the HadGEM2-ES and MIROC-ESM datasets.

85   Notably, the nature of these three datasets is different from that of the CRU TS4.00; the NCEP/NCAR consists of reanalysis data that incorporates observed and weather model output data, while the other two datasets were derived only from climate models. Details of these climate datasets are available in Table S1. To be consistent with the training data, the spatial resolutions of the validation data were linearly interpolated to a 0.5° grid mesh, and climatic conditions from 1971 to 1980 were employed.

90   In this study, the accuracy when the model was applied to the training climate dataset (i.e., the CRU dataset) is referred to as the "training accuracy", which shows how well the model was trained to extract common features of each category from images. The accuracy for the validation climate dataset (i.e., the NCEP/NCAR reanalysis, Had2GEM-ES, and MIROC-ESM datasets) is referred to as the "test accuracy", which shows how the model is robust against independent input data.

Geoscientific
Model Development
Discussions

2.2 Visualization of climate data for machine learning

95 We graphically represented the standardized air temperature and precipitation data on a grid using R statistical computing
software version 3.3.3 (R-Core-Team, 2018). These images will be referred to hereafter as visualized climatic environments
(VCEs). For efficient machine learning, climate data were standardized prior to visualization. The -20–30°C monthly mean
air temperature range and 8–400 mm/month precipitation range were log-transformed to 0.01–1.00. Values below and above
these ranges were respectively treated as 0.00 and 1.00. To evaluate how seasonality of climate regulates the biome, we also
100 conducted CNN training with annual mean air temperature and annual precipitation. For this analysis, an annual mean bio-
temperature range of 0–30°C and an annual precipitation range of 80–4000 mm/year were used. Here, bio-temperature was
defined as the mean of positive daily air temperatures. Using the annual mean bio-temperature and annual precipitation, we
first evaluated how different representations of the VCEs influenced the training and found no major differences (Table S2),
and hence the most compact VCE with the smallest computation time requirement, the RGB colour tile, was used for this
105 entire study.

In the VCE of the RGB colour tile, up to three climate variables can be represented by RGB channels. To find the optimal
combination of climatic variables, we systematically evaluated the model performance of 14 combinations of climatic
variable experiments for both annual and monthly means (Tables S3 and S4, respectively). Downward shortwave radiation
and humidity were added for this evaluation, as all of the climate datasets contain these. Generally, training accuracy
110 increases with the number of climatic variables; however, the test accuracy does not increase further after two climatic
variables. This suggests that models with three climatic variables are at risk of overfitting. Amongst the models of annual
and monthly means of climatic variables, the model with monthly mean air temperature and monthly precipitation had the
highest test accuracy. Therefore, models that combined air temperature (bio-temperature for the model of annual mean
climate) and precipitation were employed for the entire study.

115 We also evaluated the influences of different transformations of climatic variables (Table S5) and assignment patterns of air
temperature and precipitation to RGB colour channels of the VCE (Table S6) on the resulting accuracy. Based on these
evaluations, we settled on models with a combination of air temperature (bio-temperature for the model of annual mean
climate) and precipitation, both of which are log transformed, and assigned to the blue and red channels, respectively, of the
colour tile VCE representation. Examples of VCEs of annual mean climate and monthly mean climate are shown in figures
120 S1 and S2, respectively.

2.3 Training of the CNN model

The LeNet (Lecun et al., 1998), which is the world's first CNN, was employed for this study. A description of the CNN
model, its parameter settings, and computer environment are available in Supplementary Information 1. To train the CNN
model, ten VCEs corresponding to years 1971–1980 were generated for each grid using the CRU data, resulting in 572,640
125 VCEs (i.e., 10 × 57,264 grids). These VCEs were assigned to 15 categories according to the observation-based biome of the

4

grid, and the CNN model was trained to determine biomes from the VCEs. The numbers of training VCEs for each biome ranged from 4,490 (comprising temperate broadleaf evergreen forest/woodland areas) to 91,740 (comprising evergreen/deciduous mixed forest area). The training was conducted for each of the annual and monthly sets of VCEs, and their computation times for training completion were 109 and 132 minutes, respectively.

2.4 Validation of the trained model

To validate the trained CNN model, a VCE of the average climate conditions from 1971 to 1980 was obtained for each grid and each validation climate dataset. These VCEs were applied to the trained CNN model and were classified by their most plausible biome. It took roughly 8 minutes to complete the VCE classification (i.e., 57,264 in total) for each climate dataset. Then, the computed biome distributions were validated by quantitative comparison with the observation-based biome map of ISLSCP2.

For comparing the differences and similarities between two biome maps, cross-tabulation matrices were obtained for each comparison. Tables S7 and S8 show cross-tabulation matrices of training accuracies as examples. Using these matrices, the differences between the two biome maps were separated into two components: quantity disagreement and allocation disagreement (Pontius and Millones, 2011). Here, a quantity disagreement indicates a discrepancy between the proportions of the categories (i.e., the biome), while an allocation disagreement indicates a discrepancy in the spatial allocation of the categories under a given set of category proportions in the reference and comparison maps.

The use of one particular climatic dataset for training and three different climatic datasets for validation introduces a source of arbitrary error. To examine the dependency of climatic datasets for training and reconstructing performances, an experiment was performed wherein training and reconstruction of the same biome map was conducted using all combinations of the four historical climatic datasets, and then the reconstructive accuracies were compared.

Ten years of climate data may be insufficient to accurately train the model. We therefore conducted a sensitivity test in which performance was compared among models trained on monthly climate data averaged over 10-year (1971–1980; control), 20-year (1961–1980), and 30-year (1951–1980) periods. Validation datasets for each model were averaged over the same periods as the training data.

We used different climate datasets for training and validating the models to avoid overfitting that may be caused by dependencies in values among nearby grids in the training data (CRU TS4.0). To assess the effects of overfitting, we compared performance among four models that differed with respect to the grain size of training data. Nearby grid cells (0.5°) of the CRU dataset were aggregated by one of four grain sizes: 1×1 (0.5°), 2×2 (1.0°), 4×4 (2.0°), and 8×8 (4.0°). For each grain size group, 70% of grains were randomly selected for model training, and the remaining were assigned to validation. Validation with coarser grains should be less impacted by overfitting. In addition to the extent of overfitting, grain size may also influence training efficiency, because a coarser grain may skew the allocation ratio of minor biomes between training and validation sub-groups, especially when these biomes have clumped distributions. To assess this possibility, validation was also conducted using other climate datasets.

Geoscientific
Model Development
Discussions

2.5 Application of the CNN model to future climate scenarios

160   Following validation, the CNN model trained with monthly mean climate data was used to predict future biome distribution maps by applying climate scenarios for the 21st century. These predictions were conducted in combinations of two GCMs (i.e., MIROC-ESM and HadGEM2-ES) and two representative concentration pathways (RCPs; i.e., RCP2.6 and RCP8.5). These RCPs represent the atmospheric greenhouse gas (GHG) concentration forecasts adopted by the IPCC for its 5th Assessment Report (AR5) in 2014. RCP2.6 assumes that global annual GHG emissions will peak between 2010 and 2020

165   and decline substantially afterwards. By contrast, RCP8.5 assumes that emissions will continue to rise throughout the 21st century. The scenarios RCP2.6 and RCP8.5, respectively project that atmospheric CO_2 could reach 421 ppm and 936 ppm by the end of the 21st century (IPCC, 2013).

3 Results and Discussion

3.1 Reconstruction of the current biome distribution with the CNN model

170   A comparison of the training accuracies between the annual climate model and the monthly climate model demonstrated that simulation of some biomes largely depended on climate seasonality (Fig. 1). Besides the most plausible biome, the CNN outputs its certainty. Geographical distribution of the certainty clearly shows considering seasonality improves the certainty except Northern parts of South America and African continents where no apparent seasonality exists (Fig. S3). These results are consistent with Prentice et al. (1992), demonstrating that global biome distribution is under substantial controls of

175   climatic tolerance and the occurrence and extent of drought seasons. In fact, seasonality significantly improved the average training accuracies from 3.5% to 61.9% for tropical deciduous forests, 0.4% to 54.8% for temperate broadleaf evergreen forests, and 24.5% to 79.0% for boreal deciduous forests (Tables S7 and S8). The same pattern can be observed in test accuracy comparisons (Fig. 2, Fig. S3), although temperate broadleaf evergreen and boreal deciduous forests were largely absent from Had2GEM-ES and MIROC-ESM, respectively. These absences would be due to differences in the reconstructed

180   current climate among datasets (Fig. S4). Overall, for all climatic datasets examined, better training and test accuracies were consistently obtained in CNN models trained with monthly mean climate data than in those trained with annual mean climate data (Fig 3). Thus, the CNN model trained with monthly mean climate data was used for analysis with the climate scenarios in the 21st century.

For all combinations of CNN models and climatic data, the allocation disagreement was much larger than the quantity

185   disagreement: while the allocation disagreement ranged from 0.227 to 0.392, the quantity disagreement varied from 0.037 to 0.200 (Fig. 3). The larger allocation disagreement can be explained by the tendency of observation-based biome distributions to be fragmented over areas with similar climatic conditions (Fig. 1a), while model reconstructed biome distributions have more continuous structures (Figs. 1b, 1c, and 2) (For example, Australian continent). The probability of the most plausible biome tend to be lower for these fragmented regions (Fig. S3), suggesting these regions have climatic conditions suitable for

190   multiple potential biomes. The lower quantity disagreement demonstrated that the CNN model reconstructed the fraction of the global biome composition under the current climatic conditions well. As the main purpose of this research is to develop an empirical model of climatic controls on biome distribution, this would indicate that the reconstructions of biome maps with the CNN models are actually much more accurate for their particular purpose than implied by the accuracies found from the simple map comparison.

195   Table S9 compared the dependence of reconstruction accuracy on combinations of climate datasets for training and test climate datasets. Accuracies were higher and less variable when the climate dataset for training and testing were identical (0.701–0.734), compared to when these datasets were different (0.394–0.559). These results suggest that uncertainty in historical climate reconstruction is a larger source of failure in reconstructing biome distribution than the dependency of training on a particular climate dataset. Hence, this result also demonstrates the robustness of our approach.

200   No major trends were observed in test accuracies in the sensitivity test, which compared performance among models trained using monthly climate averaged over 10-, 20-, and 30-year periods, (Table S10). This indicates that climate data averaged over a 10-year period are sufficient for model training. However, long-term climatic conditions are important in controlling biome distribution via extreme climates, which may cause complete reorganization of systems and communities (e.g., Breshears et al., 2005) and may provide important opportunities for, and constraints to, plant recruitment. Considering the

205   effects of extreme climates in the model would be an interesting topic for future study.

Grain size of the training and validation data did not result in noticeable differences in training and test accuracies, with the exception of the CRU dataset (Table S11), demonstrating that the influence of grain size on training efficiency is negligible. In contrast, test accuracies of the CRU dataset were lower at coarser grain sizes, at 80.4%, 78.2%, 76.1%, and 72.2% for the 1×1, 2×2, 4×4, and 8×8 grain sizes, respectively. These results suggest that dependencies in values among nearby grids

210   in the CRU dataset resulted in overfitting. However, the effect of overfitting appears to have been much smaller than that of systematic differences among climate datasets (Fig. S4); irrespective of grain size, test efficiencies of the CRU dataset were least 19.5% higher than those of other datasets. Therefore, our validation method underestimates the actual performance of the models, and performance is much better than we demonstrated in this manuscript.

3.2 Prediction of biome distribution with the CNN model

215   The applications of the CNN model to the climate scenarios predicted a significant shift in global biome distributions (Fig. 4) and area coverages (Fig. S5) under rapid warming trends (Figs. S6 and S7). For both GCM outputs, more intense biome shifts were predicted for RCP8.5 than for RCP2.6, but the shift trends remained consistent. The most visible change was the expansion of temperate forests over boreal forests in both North America and Eurasia. Boreal and cold vegetation shrank and its composition changed; tundra areas gave way to boreal forests, while boreal evergreen forests became confined to a

220   narrow strip at higher latitudes. Tropical vegetation remained relatively unchanged, but nearly all tropical deciduous forests in the southern hemisphere were substituted by savanna, which coincided with a reduction in annual precipitation (Figs. S6 and S7).

Given the uncertainty of the climatic predictions derived from the ESMs and RCP scenarios, our analysis of the climate change effect only indicates the potential for considerable changes in biome distribution at the end of the 21st century.

225 Besides, changes in expected biome, which is an equilibrium state of vegetation coverage, are not always accompanied by immediate changes in actual vegetation. In fact, these time lags can be very long (i.e., decades to millennia) because the adjustment of vegetation to new climate conditions entails a series of plant population dynamics processes, such as seed dispersal, establishment, competition against other existing plants, and reproduction (Sato and Ise, 2012). Our study cannot infer such transient changes in vegetation, however, current process base approaches are also not a reliable option for

230 reconstructing plant population dynamic processes at the global scale; biome map predictions under common changing climate scenarios differ significantly from state-of-the-art dynamic global vegetation models (DGVMs) (Pugh et al., 2020). Hence, empirical and top down approaches, like our simulation, should still have an important role to play in approximate mapping of biomes under changing climatic conditions.

3.3 Limitations and future directions of our approach

235 There are two types of approach to mapping biomes: the correlative climate-vegetation approach and process-based approach (Notaro et al., 2012; Yates et al., 2009). We employed the former, which has advantages and disadvantages compared to the latter. An advantage of the correlative approach is that it is relatively straightforward and may be rapidly applied to different climate change scenarios. Indeed, models using the correlative approach are a common tool for predicting the impacts of climate change on biodiversity for conservation planning, because they can be easily used to simultaneously assess large

240 numbers of species (e.g., Thomas et al., 2004).

An important disadvantage of the correlative method is that extrapolating current correlations between climate and biome distributions into the future may lead to seriously biased predictions; strong performance in the present climate does not guarantee similar performance under a new set of climatic conditions that may occur in the future. However, neither Had2GEM-ES (Fig. S3f and Fig. S8a, b) nor MIROC-ESM (Fig. S3h, and Fig. S8c, d) shows apparent expansions of biome

245 uncertainty in projected climatic conditions at the end of the 21st century. This may suggest outside the environmental space of the training data is not conspicuous at the global scale. For quantifying methodological uncertainty might also result from comparison of performances between correlative and process-based models in 'unsuitable' outside the environmental space of the training data (Yates et al., 2009).

A second disadvantage of the correlative approach is that it cannot infer impacts of elevated atmospheric CO_2 on biome

250 distribution. An increase in CO_2 may favour forests over grasslands due to the advantage that C3 plants may gain over C4 plants under such conditions (Bond et al., 2003). Notably, palaeoecological studies have demonstrated that C4 ecosystems were more extensive during the last glacial maximum and decreased in abundance following deglaciation in response to increased atmospheric CO_2 concentrations (Ehleringer et al., 1997).

DGVMs, which use process-based approaches, may facilitate the identification of areas where elevated CO_2 may affect

255 biome distribution under projected climates. Indeed, the third phase of the Inter-sectoral Impact Model Inter-comparison

Project, now in progress (Warszawski et al., 2014), includes a sensitivity test for CO2 in which biome distribution is compared between scenarios of both climate and CO2 change, and scenarios of climate change only. We should note, however, that even for current state-of-the-art process-based models, incorporating effects of elevated CO2 is not straightforward due to their complexity; effects appear to be taxon specific, to interact strongly with soil type and climate,

260 and to be highly dependent on nitrogen availability (Korner, 2003; Spinnler et al., 2002).

We must also keep in mind that the correlative climate-vegetation approach ignores feedbacks between vegetation and climate, which are known to influence vegetation distribution at equilibrium (Pitman, 2003), as well as present-day species distributions that are not in equilibrium with present-day climates (e.g., Woodward, 1990).

This study only considers biome distribution at the 0.5-degree scale. At this scale, climate can be regarded as the dominant

265 factor that determines vegetation composition, and hence correlative climate-vegetation approach fits well in identifying vegetation distribution. However, at more local scales, topography, soil type, and fine-scale biotic and abiotic interactions (e.g., habitat structure, fire, storms) become increasingly important (Willis and Whittaker, 2002). One possible extension of our study is integrating these factors, acting at different spatial scales, into a hierarchical modelling framework (Pearson and Dawson, 2003).

270 Our study adopted the LeNet architecture implementation, which has six hidden layers, to create CNN models. Botella et al. (2018) found that a deep network (six hidden layers) outperformed a shallow network (one hidden layer) for building species distribution models; however, Benkendorf and Hawkins (2020) found that using more than two hidden layers was of no benefit, and argued that the usefulness of deeper networks depends on the size of the training dataset. Therefore, carefully selecting the approximate complexity of architecture implementation may improve model accuracy. Also, the default

275 parameters in NVIDIA DIGITS 6.0 remained largely unchanged. Our approach was kept relatively simple to demonstrate the robustness of our concept; however, further improvements to the scheme could be explored by selecting other implementation architectures and systematically testing the effect of parameter modulation.

4 Conclusion

Regardless of the limitations discussed above, this study provides an efficient and practical method for generating

280 preliminary estimations of the potentially dramatic impact of climate change on biome distributions. Since this method is simply an application of image classification AI, it demands much less technical skill and computer resources compared to other modern techniques such as those evaluated by Levavasseur et al. (2012), Levavasseur et al. (2013), and Hengl et al. (2018), for example. This method could also be applied to building empirical models of other climate-driven phenomena such as cropping systems and the spread of vector-borne diseases, and hence has potential to be a de facto standard for

285 building empirical models across a range of research and application fields.

Data availability

All data required to reproduce the analyses described herein are publicly available at the following URL/DOI: https://doi.org/10.5281/zenodo.4401233.

Author contributions

290 H.S. conceived and conducted the experiments. H.S. and T.I. analysed the results. H.S. wrote the manuscript. H.S. and T.I. reviewed the manuscript.

References

300 Benkendorf, D. J. and Hawkins, C. P.: Effects of sample size and network depth on a deep learning approach to species distribution modeling, Ecological Informatics, 60, 10.1016/j.ecoinf.2020.101137, 2020.

Bond, W. J., Midgley, G. F., and Woodward, F. I.: The importance of low atmospheric CO2 and fire in promoting the spread of grasslands and savannas, Global Change Biol., 9, 973-982, 10.1046/j.1365-2486.2003.00577.x, 2003.

Botella, C., Joly, A., Bonnet, P., Monestiez, P., and Munoz, F.: A Deep Learning Approach to Species Distribution

305 Modelling, in: Multimedia Tools and Applications for Environmental & Biodiversity Informatics, 169-199, 10.1007/978-3-319-76445-0_10, 2018.

Box, E. O.: Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography, Tasks for Vegetation Science, 1, Springer Netherlands, 10.1007/978-94-009-8680-0, 1981.

Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., Romme, W. H., Kastens, J. H., Floyd, M.

310 L., Belnap, J., Anderson, J. J., Myers, O. B., and Meyer, C. W.: Regional vegetation die-off in response to global-change-type drought, Proceedings of the National Academy of Sciences of the United States of America, 102, 15144-15148, 10.1073/pnas.0505734102, 2005.

Ehleringer, J. R., Cerling, T. E., and Helliker, B. R.: C4 Photosynthesis, Atmospheric CO2, and Climate, Oecologia, 112, 285-299, 1997.

315   Emanuel, W. R., Shugart, H. H., and Stevenson, M. P.: Climatic-change and the broad-scale distribution of terrestrial ecosystem complexes, Clim. Change, 7, 29-43, 1985.

Harris, I. C. and Jones, P. D.: CRU TS4.00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2015), doi:10.5285/edf8febfdaad48abb2cbaf7d7e846a86, 2017.

320   Harris, I. C., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - The CRU TS3.10 dataset, Int. J. Climatol., 34, 623-642, 10.1002/joc.3711, 2014.

Henderson-Sellers, A.: Developing an Interactive Biosphere for Global Climate Models, Vegetatio, 91, 149-166, 10.1007/bf00036054, 1991.

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., and Prentice, I. C.: Global mapping of potential natural
325   vegetation: an assessment of machine learning algorithms for estimating land potential, Peerj, 6, 10.7717/peerj.5457, 2018.

Holdridge, L. R.: Determination of world plant formations from simple climatic data, Science, 105, 367-368, 10.1126/science.105.2727.367, 1947.

IPCC: Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change, Cambridge, United Kingdom and New York, NY, USA2013.

330   Ise, T. and Oba, Y.: Forecasting Climatic Trends Using Neural Networks: An Experimental Study Using Global Historical Data, Frontiers in Robotics and AI, 6, 10.3389/frobt.2019.00032, 2019.

Ise, T. and Oba, Y.: VARENN: graphical representation of periodic data and application to climate studies, npj Climate and Atmospheric Science, 3, 10.1038/s41612-020-0129-x, 2020.

Korner, C.: Ecological impacts of atmospheric CO2 enrichment on terrestrial ecosystems, Philosophical Transactions of the
335   Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 361, 2023-2041, 10.1098/rsta.2003.1241, 2003.

Lavorel, S., Díaz, S., Cornelissen, H., Garnier, E., Harrison, S. P., McIntyre, S., Pausas, J. G., Pérez-Harguindeguy, N., and Urcelay, C.: Plant functional types: are we getting any closer to the Holy Grail?, in: Terrestrial Ecosystems in a Changing world, edited by: Canadell, J. G., Pataki, D. E., and Pitelda, L. F., Springer-Verlag, Heidelberg, 171-186, 2007.

340   LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86, 10.1109/5.726791, 1998.

Leinweber, D. J.: Stupid data miner tricks, 2007.

Levavasseur, G., Vrac, M., Roche, D. M., and Paillard, D.: Statistical modelling of a new global potential vegetation distribution, Environmental Research Letters, 7, 10.1088/1748-9326/7/4/044019, 2012.

345   Levavasseur, G., Vrac, M., Roche, D. M., Paillard, D., and Guiot, J.: An objective methodology for potential vegetation reconstruction constrained by climate, Global Planet. Change, 104, 7-22, 10.1016/j.gloplacha.2013.01.008, 2013.

Geoscientific
Model Development
Discussions

Lugo, A. E., Brown, S. L., Dodson, R., Smith, T. S., and Shugart, H. H.: The Holdridge life zones of the conterminous United States in relation to ecosystem mapping, J. Biogeogr., 26, 1025-1038, 10.1046/j.1365-2699.1999.00329.x, 1999.

Monserud, R. A. and Leemans, R.: Comparing global vegetation maps with the Kappa-statistic, Ecol. Model., 62, 275-293,
350   10.1016/0304-3800(92)90003-w, 1992.

Notaro, M., Mauss, A., and Williams, J. W.: Projected vegetation changes for the American Southwest: combined dynamic modeling and bioclimatic-envelope approach, Ecol. Appl., 22, 1365-1388, 10.1890/11-1269.1, 2012.

Pearson, R. G. and Dawson, T. P.: Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?, Global Ecol. Biogeogr., 12, 361-371, 10.1046/j.1466-822X.2003.00042.x, 2003.

355   Phillips, S. J., Anderson, R. P., and Schapire, R. E.: Maximum entropy modeling of species geographic distributions, Ecol. Model., 190, 231-259, 10.1016/j.ecolmodel.2005.03.026, 2006.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, Int. J. Climatol., 23, 479-510, 2003.

Pontius, R. G. and Millones, M.: Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for
360   Accuracy Assessment, Int. J. Remote Sens., 32, 4407-4429, DOI: 10.1080/01431161.2011.552923, 2011.

Prentice, I. C. and Leemans, R.: Pattern and process and the dynamics of forest structure - a simulation approach, J. Ecol., 78, 340-355, 1990.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M.: A global biome model based on plant physiology and dominance, soil properties and climate, J. Biogeogr., 19, 117-134, 1992.

365   Prentice, K. C.: Bioclimatic Distribution of Vegetation for General-Circulation Model Studies, Journal of Geophysical Research-Atmospheres, 95, 11811-11830, 10.1029/JD095iD08p11811, 1990.

Pugh, T. A. M., Rademacher, T. T., Shafer, S. L., Steinkamp, J., Barichivich, J., Beckage, B., Haverd, V., Harper, A., Heinke, J., Nishina, K., Rammig, A., Sato, H., Arneth, A., Hantson, S., Hickler, T., Kautz, M., Quesada, B., Smith, B., and Thonicke, K.: Understanding the uncertainty in global forest carbon turnover, Biogeosciences, 17, 3961-3989, 10.5194/bg-17-3961-
370   2020, 2020.

R-Core-Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing [code], 2018.

Ramankutty, N. and Foley, J. A.: ISLSCP II Potential Natural Vegetation Cover, doi:10.3334/ORNLDAAC/961, 2010.

Rood, R. B.: Let's call it: 30 years of above average temperatures means the climate has changed:
375   https://theconversation.com/lets-call-it-30-years-of-above-average-temperatures-means-the-climate-has-changed-36175,   last access: 23 April.

Sato, H. and Ise, T.: Effect of plant dynamic processes on African vegetation responses to climate change: Analysis using the spatially explicit individual-based dynamic global vegetation model (SEIB-DGVM), J. Geophys. Res-Biogeo., 117, 10.1029/2012jg002056, 2012.

380   Spinnler, D., Egh, P., and Korner, C.: Four-year growth dynamics of beech-spruce model ecosystems under CO2 enrichment
       on two different forest soils, Trees-Structure and Function, 16, 423-436, 10.1007/s00468-002-0179-1, 2002.

       Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira,
       M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M.
       A., Peterson, A. T., Phillips, O. L., and Williams, S. E.: Extinction risk from climate change, Nature, 427, 145-148,

385   10.1038/nature02121, 2004.

       Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model
       Intercomparison Project (ISI-MIP): project framework, Proc Natl Acad Sci U S A, 111, 3228-3232,
       10.1073/pnas.1312330110, 2014.

       Willis, K. J. and Whittaker, R. J.: Ecology - Species diversity - Scale matters, Science, 295, 1245-+,

390   10.1126/science.1067335, 2002.

       Woodward, F. I.: The Impact of Low-Temperatures in Controlling the Geographical-Distribution of Plants, Philos. T. Roy.
       Soc. B, 326, 585-593, 10.1098/rstb.1990.0033, 1990.

       Woodward, F. I. and Williams, B. G.: Climate and Plant-Distribution at Global and Local Scales, Vegetatio, 69, 189-197,
       10.1007/bf00038700, 1987.

395   Yates, C. J., Elith, J., Latimer, A. M., Le Maitre, D., Midgley, G. F., Schurr, F. M., and West, A. G.: Projecting climate
       change impacts on species distributions in megadiverse South African Cape and Southwest Australian Floristic Regions:
       Opportunities and challenges, Austral Ecol., 35, 374-391, 10.1111/j.1442-9993.2009.02044.x, 2009.
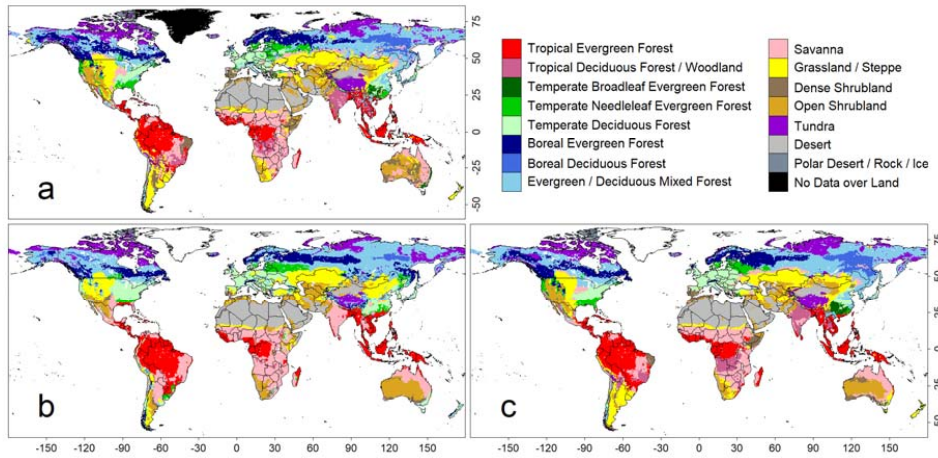
Figure 1: Comparison of global biome distributions used to evaluate the training accuracies of the convolutional neural network (CNN) model. (a) An observation-based biome map of the ISLSCP2. (b) Biome map derived from the CNN model that was trained with images of annual mean climate. (c) Biome map derived from a CNN model trained with images of monthly mean climate.

14
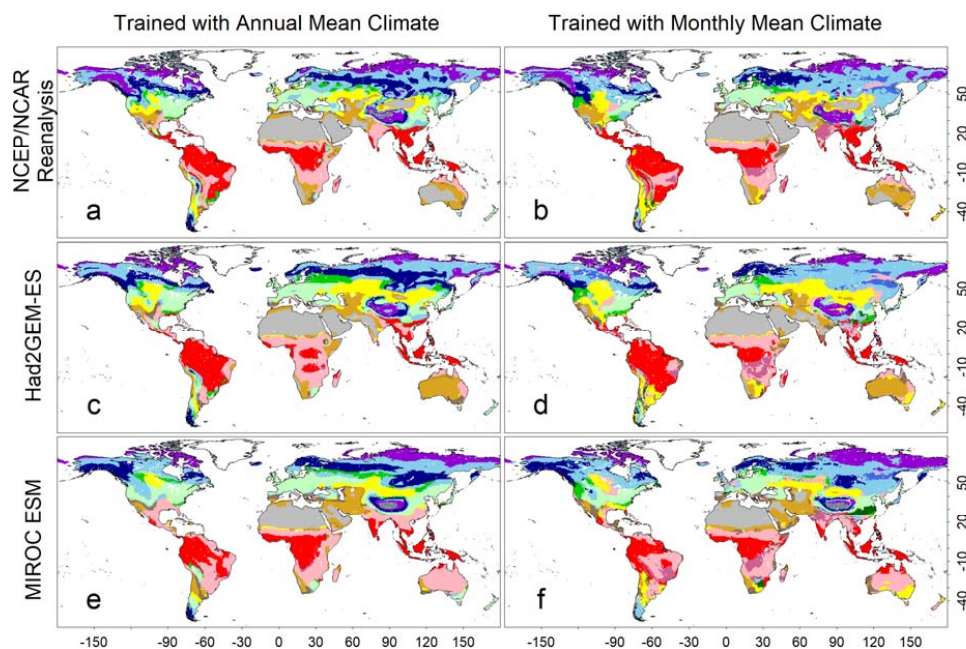
Figure 2: Test accuracies representing how the trained CNN models simulate a biome map with climatic conditions spanning from 1971 to 1980. (a, c, e) Biome map generated by the CNN model that was trained with annual mean climate images from the CRU dataset. (b, d, f) Biome map generated by the CNN model that was trained by monthly mean climate images from the CRU dataset. Three climatic datasets, which were not involved during the training process, were employed to generate these maps. (a, b) NCEP/NCAR reanalysis data; (c, d) output of the Had2GEM-ES dataset; and (e, f) output of the MIROC-ESM dataset. Colour definitions are available in Figure 1.

15

Geoscientific
Model Development
Discussions
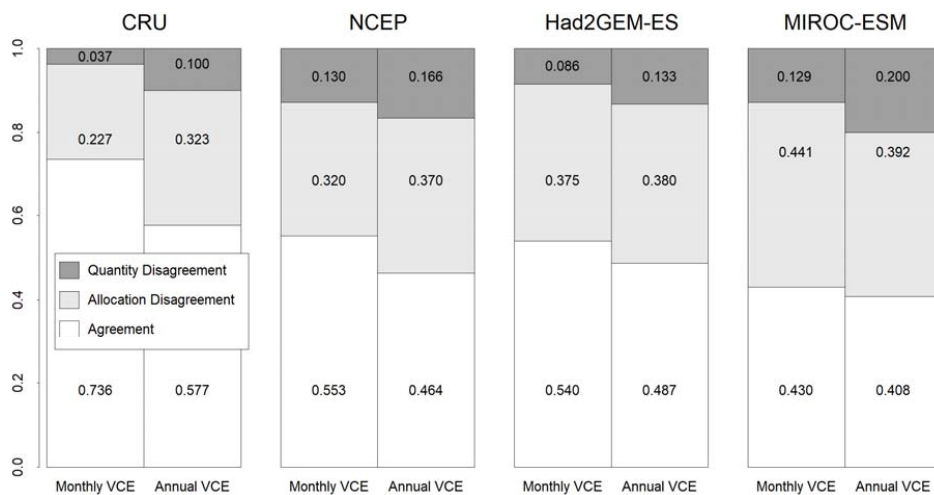
Figure 3: Fractions of agreement and disagreement between observation-based biome map and simulated biome maps trained by monthly mean climate or annual mean climate from CRU climate data spanning from 1971 to 1980. These CNN models were adapted to one of the four climatic datasets (CRU, NCEP, Had2GEM-ES, and MIROC-ESM) spanning the same period of the training data. The fraction of agreement of the CRU corresponds to the training accuracy, while that of other climate data corresponds to the test accuracy.

415
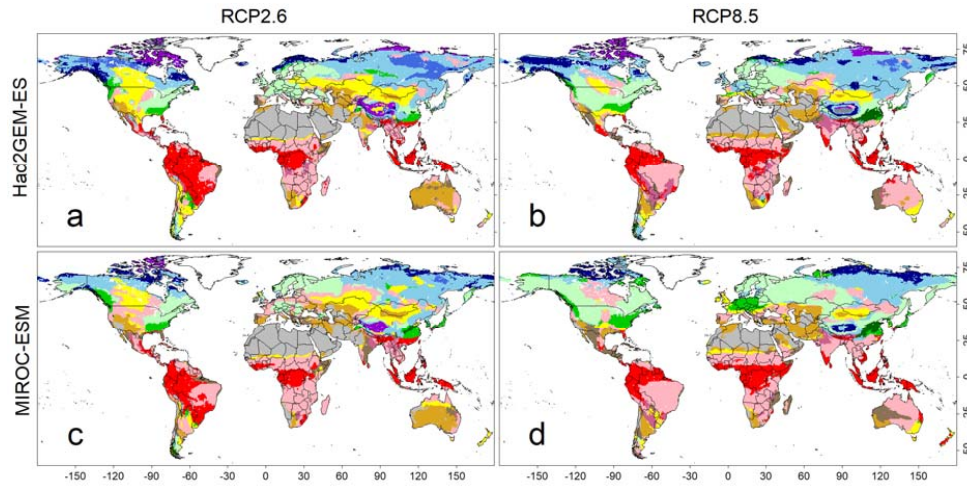
Figure 4: Fractions of agreement and disagreement between observation-based biome map and simulated biome maps trained by monthly mean climate or annual mean climate from CRU climate data spanning from 1971 to 1980. These CNN models were adapted to one of the four climatic datasets (CRU, NCEP, Had2GEM-ES, and MIROC-ESM) spanning the same period of the training data. The fraction of agreement of the CRU corresponds to the training accuracy, while that of other climate data corresponds to the test accuracy.