
Meta-modeling large-scale spatial data using Convolutional Neural Networks

Kiri A. Stern

Département de Sciences Biologiques
Université de Montréal
1375 Avenue Thérèse-Lavoie-Roux, Montréal, QC H2V 0B3
kiri.stern@umontreal.ca

Timothée Poisot

Département de Sciences Biologiques
Université de Montréal
1375 Avenue Thérèse-Lavoie-Roux, Montréal, QC H2V 0B3
timothee.poisot@umontreal.ca

Abstract

Species connectivity models play an important role in ecological research and biodiversity assessment. Unfortunately, simulations of connectivity models are typically slow, therefore preventing the rapid iteration and updates of models when evaluating different scenarios.

In this pilot study, we present the proof of concept of utilizing Deep Learning methodologies as a novel approach in ecology for significantly reducing the prediction rate of species connectivity models.

1 Introduction and motivation

Ecology is the study of interactions between biotic and abiotic factors (Pocock, M., 2019). These relationships provide insight for proper maintenance of resource and energy allocations for the conservation of healthy ecosystem biodiversity (Mohanty, S., 2017). Humans benefit from biologically diverse systems as biodiversity is paramount for the proper functioning of natural ecosystems and the services they provide mankind (Rapport, D., 1998; Schlapfer, F., 1999; Lohbeck, *et al.*, 2016).

Species connectivity modeling is an integral component of ecological research; applications include, but are not limited to: assessing biodiversity, species and habitat management/restoration/conservation/protection, forecasting the potential effects of environmental and climate change, predicting invasive species, and testing ecological theories (McRae, *et al.*, 2008; Rudnik, *et al.*, 2012). Connectivity models apply graph theory, network theory, and/or circuit theory to quantify connectivity properties with the scale of the network to map species movements (*i.e.* connectivity) across spatial dimensions (Rayfield, *et al.*, 2011).

Species connectivity is considered one of the most important indices for biodiversity (Ayram, *et al.*, 2015; Ng, *et al.*, 2013; UNEP, 2019). It is defined as the strength of possible paths between points (Newman, M., 2018), or alternatively, the flux of ecological matter between patches (Hillman, *et al.*, 2018). Unfortunately, predicting species connectivity in response to altered habitat due to land use/fragmentation and climate change is often difficult because of the lack of data and slow calculations (Isaac, *et al.*, 2020).

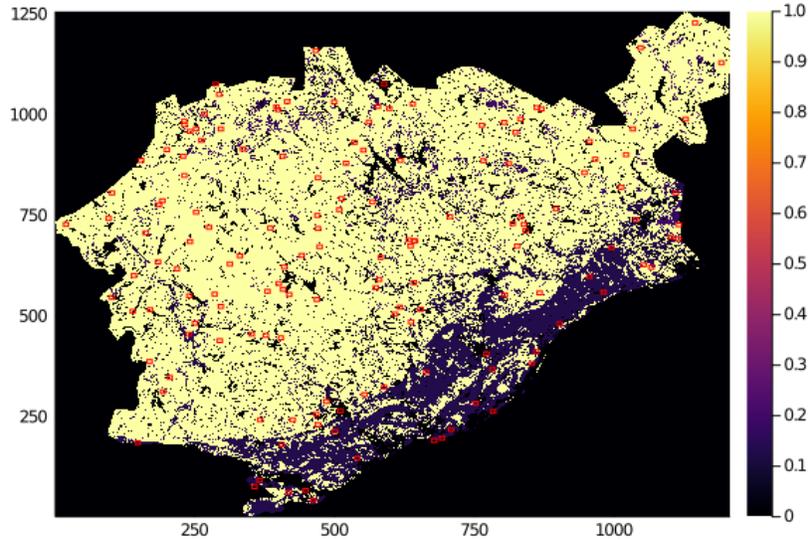


Figure 1: **Sampling points.** 150 randomly selected 9x9 sampling points used for model training and validation

Convolutional Neural Networks (CNNs) are a form of Deep Learning (DL) architecture used on grid-like structured data. As the name implies, CNNs apply a convolutional operation to identify patterns or local features from large data sets to obtain information (Goodfellow, *et al.*, 2017).

To address the issues of slow computation of connectivity simulations, we adopted a CNN model architecture as a new method to quickly and accurately predict species connectivity. The rapid iteration and updates of the model is particularly useful for providing recommendations to stakeholders and/or conservationists when evaluating different scenarios.

2 Datasets

The model inputs two spatial resolutions, the Resistance and Origin layers, and returns one spatial resolution, the Connectivity. Resistance layers are defined as the addition of spatial features that could determine a species distribution (Lecours-Tessier, *et al.*, 2020). Spatial features can include topography, human structures, and land composition, all of which has an impact on the energetic cost required to travel across a patch of land. The Origin layer, for simplicity, is consistent across all species and represents the likelihood of a species occurring in an area. The output Connectivity layer is the result of the combination of the Resistance and Origin layers.

All three spatial layers for the black bear, *Ursus americanus*, dataset covering the Laurentian region of Quebec, Canada, were taken from Lecours-Tessier *et al.* (2020) “omnidirectional and omnifunctional connectivity analyses with a diverse species pool”.

3 Model Setup, Architecture, and Training

3.1 Model Setup

Figure1 presents the one hundred and fifty 9x9 samples that were randomly selected from inside the data bounds of the original 1255x1206 images. Samples were then partitioned into training, validation and testing datasets. The validation set was procured from the training set by shuffling 10% of the data. Training and validation layers were then grouped by 32 into minibatches which were passed through the model during training. Once trained, each 9x9 predicted connectivity images were concatenated together to produce the final full 1242x1206 connectivity map.

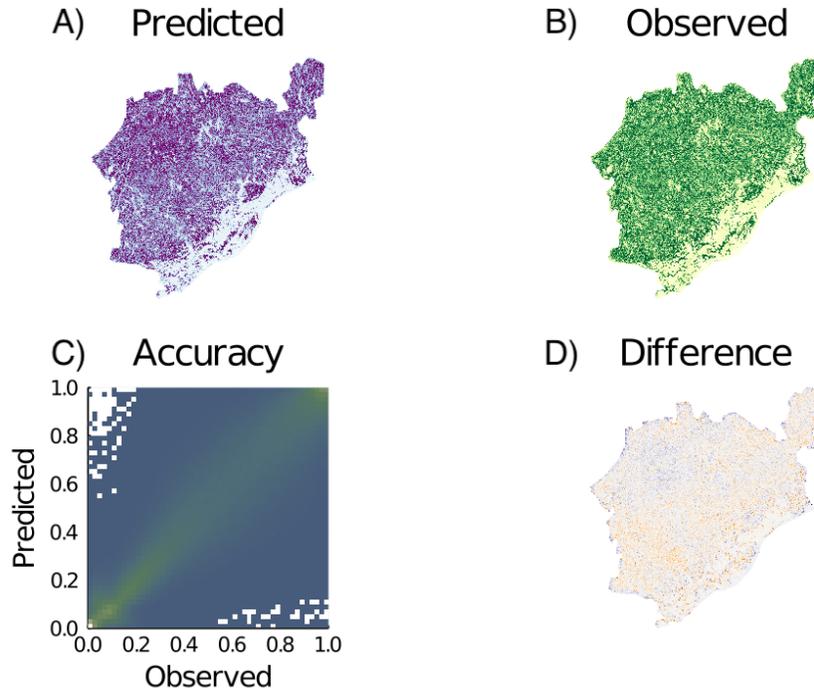


Figure 2: **Model on black bear dataset.** **A)** Black bear connectivity map predicted by the CNN model. **B)** Corresponding connectivity map produced by Omniscape. **C)** Histogram of the model's accuracy. **D)** Difference between predicted values and observed values

3.2 Model Architecture

The model consists of nine layers. The first layer applies a 3x3 convolution filter to a 4D input of size 9x9x2x32 (where 9x9 is the size of the image; the Resistance and Origin layers make up the 2 channels; and minibatched into groups of 32). The 16-channel output is then activated with the ReLU activation function. A 2x2 Maxpooling window makes up the second layer. The third, fourth, fifth, and sixth layers consist of another 3x3 convolutional filter, ReLU activation function, and 2x2 Maxpool window respectively; while the seventh layer flattens the 3D tensor to a 2D one, which is suitable for dense layer and training. The eighth layer consists of a fully connected (*i.e.* Dense) layer that is then reshaped back into a 4D object to match the dimensions of the validation set for comparison (9x9x1x32).

3.3 Model Training

The model is trained using the ADAM optimizer (set at 0.001) and prints out the model's performance against the validation set after each iteration. Performance is calculated using a simple Least Square Errors (L2 Loss) function divided by the actual value and augmented by adding Gaussian random noise to images to make it more robust. The training is set for 200 epochs, however will quit out if the model attains 99.9% accuracy or when the best accuracy has not improved after 5 epochs following a decrease in learning rate. The parameters for the best run are saved out to a Binary JSON (.BSON) file and model performance is printed out and visually verified by plotting the heatmap and scatter plot of the connectivity values versus the model's output.

4 Results

Figure 2 represents the prediction of the trained model applied on the black bear species data. The model obtained a 98.49% accuracy after 58 epochs and took 1071.88 seconds while running on a single CPU. Figure 2A represents the output of the model when applied to the black bear data; it is the predicted connectivity. Figure 2B is the original connectivity map, which was used for the supervised

learning and is what the model is trying to predict. This was generated using Omniscap. Figure 2C is a histogram of the accuracy of the model showing how well the connectivity values predicted by the model are compared to what we would expect to observe. Figure 2D is the difference between Figure 2B and 2A, where dark yellow pixels indicate areas of highly overestimated connectivity and dark blue pixels denote highly underestimated connectivity.

5 Discussion

Omniscap is a program which utilizes a moving window algorithm applied to circuit theory for the omnidirectional modeling of species connectivity (Tessier, *et al.*, 2020; McRae, *et al.*, 2016). While Omniscap has established itself as a leading platform for connectivity analysis (Conservation Biology Institute, 2020), simulations can take up to an hour or more on CPU and approximately 15 minutes when run on Compute Canada remote clusters (Daphnée Lecours-Tessier, personal communication, September 24, 2020).

Based on the results, we can conclude that, despite the need for some improvements, the use of a CNN model was highly successful in minimizing the run time required to predict species connectivity, all while maintaining high accuracy. The accuracy is considered to be very good as seen in Figure 2C, where most data points lay along the diagonal. The fact that it is not a perfect linear fit is, in fact, desirable as that indicates the model is not over-fitting.

Limitations of the model include: the ability to generalize between species (results not shown) and difficulty estimating areas of very high connectivity. The model's architecture and/or the low number of sampling points used for training and validation (only 150) can possibly explain these constraints. Future architectures should investigate the model's performance when trained and tested on different species individually. The architecture which has reasonable accuracy on each species should be the one used when generalizing with multiple species. The small sample size could be another contributing factor in the model's poor ability to predict areas of high connectivity. It was chosen to minimize run time during the creation and optimization of the model.

It was observed, when exploring various model architectures, that a layer ending with a small number of channels allowed for a better prediction of the low connectivity values, while a layer outputting a large number of channels improved predictions for high connectivity values. However, when combining small and large channels, the model performed sub-optimally. Different variations of channel sizes should be explored more extensively. Although, it is possible that areas of very high connectivity are not representative of a species' true dispersal patterns and that the model, in fact, reduces these artifacts and generates a more realistic species distribution map.

Another minor improvement would be to utilize the properties of a sliding window to smooth out the connectivity maps and produce less pixelated images. Additionally, running the simulations on GPU will significantly further reduce computation time and could easily be implemented given Julia's ML-specific package, Flux.jl, which includes functions to easily convert models and data to and from the GPU if available.

6 Conclusion and Broader Impact

Species connectivity modelling is an important component in ecological management. The adoption of DL techniques to map species connectivity has demonstrated its potential as a successful method in modelling large scale spatial data. For the first attempt in implementing a CNN to reduce computation time and accurately predict species connectivity outcomes, our model has shown to be effective, despite its need for some improvements, as described in the Discussion section.

Decreasing the run time it takes to simulate species connectivity allows for a quicker analysis in conservation efforts, predicting the potential spread of invasive species, assessing biodiversity and ecosystem health, and more. Rapid iterations and updates of the model can allow stakeholders, conservationists, epidemiologists, etc. to assess multiple scenarios in a short amount of time. However, the need for a completely reliable model is integral in the success of using CNNs for rapidly determining connectivity models. Sacrificing too much accuracy for decreased computation time can lead to unintended consequences when making critical environmental management decisions. However, ultimately, the novel implementation of DL methodologies applied in ecology has demonstrated itself

as a powerful tool. Further explorations in using CNNs for connectivity models and the ability to generalize across species should be further investigated.

References

- Ayram, C. A., Mendoza, M. E., Etter, A., & Salicrup, D. R. (2015) Habitat connectivity in biodiversity conservation. *Progress in Physical Geography: Earth and Environment*, **40**(1): 7-37. doi:10.1177/0309133315598713
- Conservation Biology Institute (2020) A Climate Adaptation Conservation Planning Database for North America. *AdaptWest* Retrieved September 30, 2020. <https://adaptwest.databasin.org/pages/climate-connectivity-priorities-omniscape>
- Goodfellow, I., Bengio, Y., & Courville, A. (2017) 1 Introduction. In *Deep learning* pp. 1–12. essay, MIT Press.
- Hillman, J. R., Lundquist, C. J., & Thrush, S. F. (2018) The Challenges Associated With Connectivity in Ecosystem Processes. *Frontiers in Marine Science*, **5** doi:10.3389/fmars.2018.00364
- Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., . . . & O’Hara, R. B. (2020) Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*, **35**(1): 56-67. doi:10.1016/j.tree.2019.08.006
- Lohbeck, M., Bongers, F., Martinez-Ramos, M., & Poorter, L. (2016) The importance of biodiversity and dominance for multiple ecosystem functions in a human-modified tropical landscape. *Ecology*, **97**(10): 2772-2779. doi:10.1002/ecy.1499
- McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**: 2712-2724. doi:10.1890/07-1861.1
- McRae, Brad & Popper, Ken & Jones, Aaron & Schindel, Michael & Buttrick, S. & Hall, Kimberly & Unnasch, Robert & Platt, Jim. (2016) Conserving Nature’s Stage: Mapping Omnidirectional Connectivity for Resilient Terrestrial Landscapes in the Pacific Northwest. 10.13140/RG.2.1.4158.6166.
- Mohanty, S. (2017) The Importance of Studying Ecology. *Research & Reviews: Journal of Ecology and Environmental Sciences*. e-ISSN:2347-7830 & p-ISSN:2347-7822
- Newman, M. E. J. (2019) 6.12 Independent paths, connectivity, and cut sets. In *Networks* pp. 142–144. essay, Oxford University Press
- Ng, C. N., Xie, Y. J., & Yu, X. J. (2013) Integrating landscape connectivity into the evaluation of ecosystem services for biodiversity conservation and its implications for landscape planning. *Applied Geography*, **42**: 1-12. doi:10.1016/j.apgeog.2013.04.015
- Pocock, M. (2019) Why ecology matters. *British Ecological Society* Retrieved September 23, 2020, from <https://besjournals.onlinelibrary.wiley.com/hub/whyecologymatters>
- Rapport, D. (1998) Assessing ecosystem health. *Trends in Ecology & Evolution*, **13**(10): 397-402. doi:10.1016/s0169-5347(98)01449-9
- Rayfield B, Fortin MJ, & Fall A. (2011) Connectivity for conservation: a framework to classify network measures. *Ecology*. **92**(4):847-58. doi: 10.1890/09-2190.1. PMID: 21661548.
- Rudnick, D., Beier, P., Cushman, S., Dieffenbach, F., Epps, C.W., Gerber, L., Hartter, J., Jenness, J., Kintsch, J., Merenlender, A.M., Perkle, R.M., Preziosi, D.V., Ryan, S.J., & S. C. Trombulak. (2012) The Role of Landscape Connectivity in Planning and Implementing Conservation and Restoration Priorities. *Issues in Ecology*, **16**. Ecological Society of America. Washington, DC.
- Schlapfer, F. (1999) Expert Estimates about Effects of Biodiversity on Ecosystem Processes and Services. *Oikos*, **84**(2): 346. doi:10.2307/354673
- Tessier, D. L., Maranger, R., & Poisot, T. (2020) Omnidirectional and omnifunctional connectivity analyses with a diverse species pool. <https://doi.org/10.1101/2020.02.03.932095>
- UNEP (2019). Frontiers 2018/19 Emerging Issues of Environmental Concern. *United Nations Environment Programme, Nairobi*. ISBN:978-92-807-3737-0