ARPHA Conference Abstracts 8: e149201
doi: 10.3897/aca.8.e149201

OPEN ACCESS

Conference Abstract

Recurrent neural network emulation of time-evolving agroecosystem model outputs: a framework for efficient emulator-based calibration

Viivi Liisa Aakula[‡], Istem Fer[‡], Julius Vira[‡]

‡ Finnish Meteorological Institute, Helsinki, Finland

Corresponding author: Viivi Liisa Aakula (viivi.aakula@fmi.fi)

Received: 10 Feb 2025 | Published: 28 May 2025

Citation: Aakula V, Fer I, Vira J (2025) Recurrent neural network emulation of time-evolving agroecosystem model outputs: a framework for efficient emulator-based calibration. ARPHA Conference Abstracts 8: e149201. https://doi.org/10.3897/aca.8.e149201

Abstract

Process-based models are widely used in studying agroecosystem dynamics, with computational representations simulating the interactions within the studied system. However, due to the complexity of simulating detailed biophysical processes, process-based models are often computationally inefficient when running large-scale and iterative simulations, required for parameter calibration. With increasing spatiotemporal scales, the computational requirements grow significantly.

Calibration of process-based models is necessary for adjusting the model parameter values to reflect the characteristics of a specific environment, enhancing the model accuracy (Wallach et al. 2021). Bayesian calibration is a well-established practice in aligning ecosystem models with observed data, with the Markov Chain Monte Carlo (MCMC) approach being a widely adopted method. However, the iterative sampling of MCMC methods can require tens of thousands to hundreds of millions of simulations to explore the parameter space effectively and achieve convergence. This computational burden increases significantly with large spatiotemporal scales.

© Aakula V et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

One approach to address the inefficiency of process-based models is to build computationally lighter emulators of these models. This means building a surrogate model to approximate the process-based model by employing machine learning methods and using the emulator in heavy or iterative computations such as calibration, in place of the actual model. In addition to reduced computational demands, emulators allow scalability to large data sets, for example high-resolution spatiotemporal data, and efficient exploration of different scenarios, such as running thousands of simulations for uncertainty quantification.

Emulation of time-dependent model outputs allows learning the temporal dynamics of ecosystem processes. Recurrent neural networks have proven efficient in agroecosystem modelling (Liu et al. 2022, Zou et al. 2024) for reproducing temporal dependencies while other machine learning have limitations in learning. The well-established recurrent neural network Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) is designed to learn both short- and longer-term dependencies, and has been used in building emulators to model time-dependent outputs (Mohammed et al. 2022, Qi et al. 2022). However, applications of LSTM networks in emulation of time-dependent agroecosystem model outputs have been limited.

We apply the LSTM for building an emulator for the process-based model BASGRA (Höglind et al. 2016, Höglind et al. 2020) for managed grasslands and finally use the emulator in calibrating the process model parameters against eddy covariance flux measurements. Our objective was to build an emulator capable to predict sequences of 53 weekly values of leaf area index (LAI), net primary production (NPP), gross primary production (GPP), harvest carbon flux and soil moisture, based on weekly meteorological forcings and model parameters, given as input to the network. The model parameters on plant and soil properties were chosen based on a sensitivity analysis, and the meteorological forcing data was obtained from the ERA5 (Hersbach et al. 2023) dataset.

In addition to LSTM networks, feed-forward neural networks (FNNs) were used in further learning representations of the LSTM outputs and shaping the final output to the desired form. To determine an architecture for the network, we applied the Tree-structured Parzen Estimator (TPE) (Bergstra et al. 2011), a computationally efficient Bayesian optimization algorithm, to find the best performing network from a predefined hyperparameter space. Each tested hyperparameter set was validated with 5-fold cross validation with data from a large area of Europe, with separate temporal data sets for each fold. Finally, the hyperparameters yielding to most accurate results were chosen for the emulator.

The final, optimized emulator explained over 95% of the variation of the process-based model for all the emulated features, which demonstrates a high accuracy of the emulator, and its applicability in various modelling tasks in place of the actual model.

Finally, we used the emulator in calibration against GPP data from three grassland sites across Finland with a Hamiltonian Monte Carlo algorithm which makes use of the differentiability of the emulator. We tested multiple calibration scenarios, using different subsets of the data as calibration data and validation data. Across all calibration setups,

the calibrated emulator outperformed the prior emulator in GPP prediction accuracy, while also informing the predictive performance of the unobserved output parameters (leaf area index and harvest yield).

The trained emulator demonstrated accurate approximation of the process-based model and successful application for calibration. The network was able to learn simulated relationships between soil properties, meteorological drivers and carbon dynamics of the ecosystem. Moreover, the emulator building framework would generally be applicable for other similar agroecosystem models, as it incorporates the components needed for producing time series output from sequentially dependent and static inputs.

Keywords

Agroecosystem models; Emulation; Recurrent neural networks; Long short-term memory (LSTM); Bayesian calibration

Presenting author

Viivi Aakula

Presented at

ORAL

Hosting institution

Finnish Meteorological Institute

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Bergstra J, Bardenet R, Bengio Y, Kegl B (2011) Algorithms for Hyper-Parameter Optimization. In: Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger KQ (Eds) Advances in Neural Information Processing Systems. 24. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf
- Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J (2023)

ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/cds.adbb2d47

- Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Computation 9 (8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
- Höglind M, Van Oijen M, Cameron D, Persson T (2016) Process-based simulation of growth and overwintering of grassland using the BASGRA model. Ecological Modelling 335: 1-15. https://doi.org/10.1016/j.ecolmodel.2016.04.024
- Höglind M, Cameron D, Persson T, Huang X, van Oijen M (2020) BASGRA_N: A model for grassland productivity, quality and greenhouse gas balance. Ecological Modelling 417 https://doi.org/10.1016/j.ecolmodel.2019.108925
- Liu Q, Yang M, Mohammadi K, Song D, Bi J, Wang G (2022) Machine Learning Crop Yield Models Based on Meteorological Features and Comparison with a Process-Based Model. Artificial Intelligence for the Earth Systems 1 (4). https://doi.org/10.1175/aies-d-22-0002.1
- Mohammed H, Michel Tornyeviadzi H, Seidu R (2022) Emulating process-based water quality modelling in water source reservoirs using machine learning. Journal of Hydrology 609 https://doi.org/10.1016/j.jhydrol.2022.127675
- Qi S, He M, Bai Z, Ding Z, Sandhu P, Zhou Y, Namadi P, Tom B, Hoang R, Anderson J (2022) Multi-Location Emulation of a Process-Based Salinity Model Using Machine Learning. Water 14 (13). https://doi.org/10.3390/w14132030
- Wallach D, Palosuo T, Thorburn P, Hochman Z, Gourdain E, Andrianasolo F, Asseng S, Basso B, Buis S, Crout N, Dibari C, Dumont B, Ferrise R, Gaiser T, Garcia C, Gayler S, Ghahramani A, Hiremath S, Hoek S, Horan H, Hoogenboom G, Huang M, Jabloun M, Jansson P, Jing Q, Justes E, Kersebaum KC, Klosterhalfen A, Launay M, Lewan E, Luo Q, Maestrini B, Mielenz H, Moriondo M, Nariman Zadeh H, Padovan G, Olesen JE, Poyda A, Priesack E, Pullens JWM, Qian B, Schütze N, Shelia V, Souissi A, Specka X, Srivastava AK, Stella T, Streck T, Trombi G, Wallor E, Wang J, Weber TD, Weihermüller L, de Wit A, Wöhling T, Xiao L, Zhao C, Zhu Y, Seidel S (2021) The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. Environmental Modelling & Software 145 https://doi.org/10.1016/j.envsoft.2021.105206
- Zou H, Chen J, Li X, Abraha M, Zhao X, Tang J (2024) Modeling net ecosystem exchange of CO2 with gated recurrent unit neural networks. Agricultural and Forest Meteorology 350 https://doi.org/10.1016/j.agrformet.2024.109985