May 21, 2024

Nonlinear Processes in Geophysics
Manuscript ID: EGUSPHERE-2023-2368

Dear Referee 1,

We thank you for taking the time to read our manuscript and for the helpful comments posted. This letter details the changes we made to address the points raised.

We highlight the referee comments in *blue italics*. We provide our replies to the comments and the changes we made to the manuscript. Additionally, we showcase the manuscript additions and alterations in **bold font**.

1. *Title. Considering that the models applied to all NDVI time series and extreme cases are analyzed for performance comparison, I would suggest removing "extreme" from the title to better reflect the broader application of the model*

   While we wanted the title to reflect a specific section of the analysis, given its novelty in the field, we now see how the wording could be misleading. We suggest the following alternative title:

   Learning Vegetation Response to Climate Drivers and Extremes with Recurrent Neural Networks

   While the new title maintains the word "extreme," we shifted the focus to underline the study's broader scope.

2. Section 2.1. The models are trained separately for each site. Is it feasible to train a universal model with all the stations by incorporating some static features (such as elevation, latitude, longitude, etc.,)? There have been studies showing the LSTM model benefits from diverse training datasets. How might this approach impact conclusions, especially regarding whether gated models benefit more from augmented training data given their complex architectures?

   Thank you for the interesting discussion point. It is feasible to train a global model over all the investigated locations even without incorporating static features. We presume that the global model would show slightly worse overall performances compared to what we showed in the paper. This is likely due to the complexity that arises from the much larger feature space. Proper hyperparameter search and optimization would be computationally expensive and time-consuming. Consequently, performances might not be optimal for all locations. While NRMSE and SMAPE could be similar, entropy complexity plots and extremes would show a decreased performance.

   There are two reasons why additional static features would not be very helpful in this setting:

   - We used 20 locations in the study. Therefore, static variables such as elevation, latitude, and longitude do not provide meaningful insights into the model given that each point is a unique feature.
   - Our task is time series prediction in known locations. Static variables are helpful in situations in which this task is performed in unseen locations [1].

   Considering a scenario in which we have access to a much larger set of locations, the conclusions of the previous section would still apply. The proposed global model would likely perform better with static variables if given a sufficient number of locations. The additional information would come at the cost of an additional increase in the dimensionality of the feature space, incurring an increase in the complexity of the training. Finally, to the best of our knowledge, we are not familiar with studies using ESNs as a global model, or incorporating static variables.

We appreciate your interest in understanding our modeling choices. Our study focuses on the local response of vegetation to climate drivers. To more closely investigate the models' ability in this specific setting, we used local models instead of a global approach. Thus, a global approach was beyond the scope of our study.

3. *DL features. The input features need to be clarified, particularly mean temperature and sea level pressure. Are these area means over the European continent, or are they in-situ measurements? If the latter, how do the gridded dataset correlate with in-situ NDVI values? Additionally, clarify the input time window size for the model.*

   We obtained the mean temperature and sea level pressure data from the E-OBS dataset v26.0e, as mentioned in our manuscript in Section 2.1 Line 136. These climate variables are not area means over the entire European continent; they are based on in-situ measurements spatially interpolated to cover most of the European continent, as detailed in Line 137. The interpolation is twofold: (1) deterministic modeling is employed to capture the long-term spatial trend in the data for daily values. (2) Subsequently, stochastic interpolation, utilizing a Gaussian Random Field (GRF) simulation, is applied to the residuals from this model to generate the daily ensemble [2]. In this instance, the ensemble refers to an approach in which the parameters are varied within a range, the mean of which is the 'best guess.' This allows us to generate a more robust gridded dataset than a classical interpolation approach with a resolution. The E-OBS dataset has a spatial resolution of 0.1 degrees (approximately 11.1 km × 11.1 km), which enables us to approximate the climate conditions at a regional scale around each study site.

   Regarding the correlation of these gridded dataset variables with in-situ NDVI values, the NDVI data was obtained from the moderate resolution imaging spectroradiometer (MODIS) in the FluxnetEO dataset v1.0, which provides high-resolution temporal and spatial data specifically for the locations of eddy covariance (EC) towers. The spatial resolution of our NDVI data cubes is 500 m, which is significantly finer than the E-OBS climate data. We employed a spatial aggregation method to ensure a meaningful comparison and correlation between these datasets. This method involved averaging the NDVI data over the whole spatial dimension available in the cube (a 3 km × 3 km area centered on the EC tower). This process is illustrated in Lines 123-125, as well as in Figure 1b and 1c. This preprocessing created a more direct and relevant comparison between the climate variables and the NDVI measurements at each study site.

   Finally, regarding the input time window size for the model, the NDVI data and the climate variables (air temperature, mean sea level pressure, global radiation, and precipitation) have a daily temporal resolution and span the time period of 2000 to 2020. For our machine learning model, we used these daily measurements as input features without embeddings or fixed lookback windows, as detailed in Line 149, allowing the model to capture the day-to-day variations under the climate conditions and their impact on the NDVI measurements.

4. *Figure2, I believe subfigure a is for conventional RNN, LSTM, and b is for ESN. The figure caption is mismatched.*

   Thank you for pointing out this oversight, which is now corrected. The new caption refers to the schematics as a) backpropagation-based recurrent neural networks and b) echo state networks. We propose the following new caption:

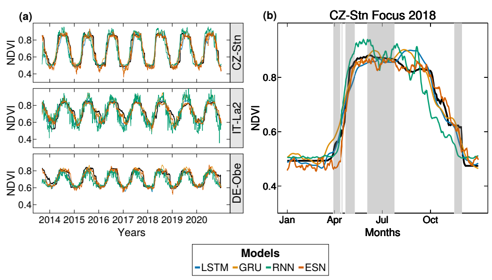
   Diagram (a) presents the conventional approach used by RNNs, GRUs, and LSTMs, involving backpropagation through time (BPTT). In contrast, diagram (b) portrays the training methodology of ESNs. The initial two layers are randomly generated and

remain untrained, while only the final layer undergoes a one-shot training via linear regression (LR)

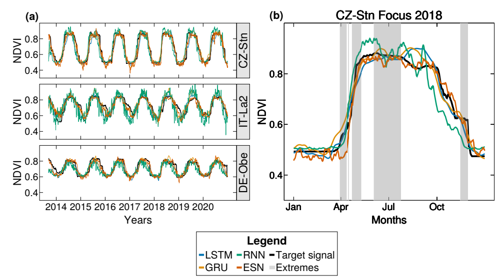
5. *Figure4, is the black line for the real/target value? I would suggest adding it in the legend.*

   Thank you for pointing this out; we have updated the figure's legend to indicate that the black line represents the real/target values. We have also added a description for the greyed-out areas in the legend, which correspond to extreme events as identified by the procedure outlined in section 2.5. Furthermore, we clarified the figure caption, which is as follows:

   The target time series is shown in black, while the predictions only use a singular run from a set of 50 per model. Panel (a) delineates the results obtained at three selected locations: Germany, Italy, and Czech Republic (CZ-Stn). Subsequently, panel (b) offers a magnified view of the outcomes at the CZ-Stn location in 2018, highlighting the extremes defined by a 90% threshold using a greyed-out area.



(a) Old figure 4

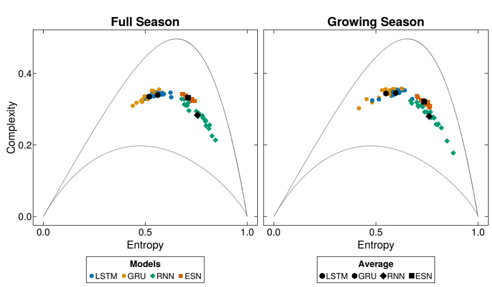


(b) Updated figure 4

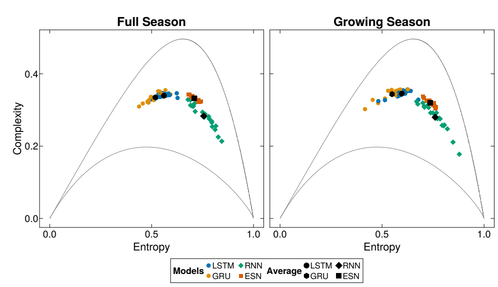
Figure 1: Side-by-side comparison of Figure 4 after the suggested changes.

6. *Figure5, the legend color in the right panel is missing. Please revise it for completeness.*

   The double legend in Figure 5 refers to both panels, with the color legend describing the entropy and complexity measures per 50 runs per model and the black legend describing the mean of the respective models. Your comment made us realize that this is somewhat misleading. We have added a more precise, unified legend to fix this problem. Figure 2 shows the differences between the old and the new Figure 5. We thank the reviewer for pointing out a potential source of confusion.



(a) Old figure 5



(b) Updated figure 5

Figure 2: Side-by-side comparison of Figure 5 after the suggested changes.

7. *Table 1. Clarify whether the standard deviation is derived from the 20 selected sites.*

   The standard deviation is indeed derived from the study sites. We have revised the legend to state that explicitly. Thank you for pointing this out. We added the following sentence to the caption:

The means are calculated across 50 runs for each of the 20 different sites, and then the mean of these 20 location-derived means is shown. Similarly, the standard deviations are calculated for each location across 50 runs and then averaged to provide the overall standard deviation for each metric.

8. *I would suggest adding a comparison of training and inference speed of the selected networks for a more comprehensive evaluation.*

   We acknowledge the importance of assessing the performance of neural network architectures in terms of speed. However, due to the models being implemented in different programming languages (Julia for the echo state networks and Python for the backpropagation-based approaches), a direct speed comparison is not representative of differences in the models' performance. In fact, in [3], we show that the Julia implementation of echo state networks offers faster computational times than existing Python packages. Additionally, existing literature (for example [4, 5, 6, 7]) provides extensive comparisons of these architectures' speeds. We have added the following introduction to Appendix A in which we detail existing works comparing these architectures:

   In this appendix, we provide the mathematical details behind the models used in this study. Starting from A1, we introduce the simple version of the RNN, which provides the basic equation for all following models. Subsequently, in A2 and A3, we illustrate the gated approaches of LSTMs and GRUs, respectively. Finally, in A4, we showcase the ESNs. While the general approach for all the models shows some similarities, different constructions exhibit variations in complexity [8, 9] and training speed [4, 5].

We look forward to hearing from you.
Sincerely and on behalf of all authors,

Francesco Martinuzzi

Leipzig University
Center for Scalable Data Analytics and Artificial Intelligence,
Institute for Earth System Science and Remote Sensing, and
Remote Sensing Centre for Earth System Research

References

[1] Basil Kraft, Martin Jung, Marco Körner, Christian Requena Mesa, José Cortés, and Markus Reichstein. Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in big Data*, 2:31, 2019.

[2] Richard C. Cornes, Gerard van der Schrier, Else J. M. van den Besselaar, and Philip D. Jones. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409, 2018.

[3] Francesco Martinuzzi, Chris Rackauckas, Anas Abdelrehim, Miguel D Mahecha, and Karin Mora. Reservoircomputing. jl: an efficient and modular library for reservoir computing models. *The Journal of Machine Learning Research*, 23(1):13093–13100, 2022.

[4] Luca Cerina, Marco D Santambrogio, Giuseppe Franco, Claudio Gallicchio, and Alessio Micheli. Efficient embedded machine learning applications using echo state networks. In *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 1299–1302. IEEE, 2020.

[5] Chenxi Sun, Moxian Song, Derun Cai, Baofeng Zhang, Shenda Hong, and Hongyan Li. A systematic review of echo state networks from design to application. *IEEE Transactions on Artificial Intelligence*, 5(1):23–37, 2022.

[6] Joschka Herteux, Christoph Räth, Amine Baha, Giulia Martini, and Duccio Piovani. Forecasting trends in food security: a reservoir computing approach. *arXiv preprint arXiv:2312.00626*, 2023.

[7] Xingshang Li, Fanjun Li, Shoujing Zheng, and Qianwen Liu. Nox concentration prediction with a flexible cascaded echo-state network in a cement clinker calcination system. *IEEE Transactions on Industrial Informatics*, 2024.

[8] Pedro J Freire, Yevhenii Osadchuk, Bernhard Spinnler, Antonio Napoli, Wolfgang Schairer, Nelson Costa, Jaroslaw E Prilepsky, and Sergei K Turitsyn. Performance versus complexity study of neural network equalizers in coherent optical systems. *Journal of Lightwave Technology*, 39(19):6085–6096, 2021.

[9] Pedro Freire, Sasipim Srivallapanondh, Bernhard Spinnler, Antonio Napoli, Nelson Costa, Jaroslaw E Prilepsky, and Sergei K Turitsyn. Computational complexity optimization of neural network-based equalizers in digital signal processing: A comprehensive approach. *Journal of Lightwave Technology*, 2024.