https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Learning Extreme Vegetation Response to Climate Forcing: A Comparison of Recurrent Neural Network Architectures

Francesco Martinuzzi[1,2,3], Miguel D. Mahecha[1,2,3,4], Gustau Camps-Valls[5], David Montero[2,3], Tristan Williams[5], and Karin Mora[2,3]

[1]Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Leipzig, Germany
[2]Institute for Earth System Science & Remote Sensing, Leipzig University, 04103 Leipzig, Germany
[3]Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany
[4]German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany
[5]Image Processing Laboratory (IPL), Universitat de València, València, Spain

Correspondence: Francesco Martinuzzi martinuzzi@informatik.uni-leipzig.de

Abstract. Vegetation state variables are key indicators of land-atmosphere interactions characterized by long-term trends, seasonal fluctuations, and responses to weather anomalies. This study investigates the potential of neural networks in capturing vegetation state responses, including extreme behavior driven by atmospheric conditions. While machine learning methods, particularly neural networks, have significantly advanced in modeling nonlinear dynamics, it has become standard practice to approach the problem using recurrent architectures capable of capturing nonlinear effects and accommodating both long and short-term memory. We compare four recurrence-based learning models, which differ in their training and architecture: 1) recurrent neural networks (RNNs), 2) long short-term memory-based networks (LSTMs), 3) gated recurrent unit-based networks (GRUs), and 4) echo state networks (ESNs). While our results show minimal quantitative differences in their performances, ESNs exhibit slightly superior results across various metrics. Overall, we show that recurrent network architectures prove generally suitable for vegetation state prediction yet exhibit limitations under extreme conditions. This study highlights the potential of recurrent network architectures for vegetation state prediction, emphasizing the need for further research to address limitations in modeling extreme conditions within ecosystem dynamics.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

1 Introduction

The recent increase in atmospheric CO_2 concentrations only partly reflects anthropogenic emissions, as oceans and land ecosystems contribute to the carbon uptake (Eggleton, 2012; Le Quéré et al., 2018; Canadell et al., 2021). Forests and other terrestrial ecosystems absorb nearly a third of human-made emissions and establish an essential negative feedback within the global carbon cycle (Friedlingstein et al., 2006; Le Quéré et al., 2009). However, during extreme events such as persistent droughts and heatwaves, ecosystems may release more CO_2 into the atmosphere than they absorb due to suppressed photosynthesis (von Buttlar et al., 2018; Sippel et al., 2018). Variations to the frequency and intensity of these events can lead to long-lasting environmental modifications, contributing to positive feedback loops that aggravate climate warming (Reichstein et al., 2013). For example, increases in drought intensities have been consistently linked to excess tree mortality (Grant, 1984; Fensham and Holman, 1999; Liang et al., 2003; Dobbertin et al., 2007), negatively impacting carbon sequestration (Van Mantgem et al., 2009). The frequency, intensity, and duration of extremes over the next few decades are expected to increase compared to previous decades (Seneviratne et al., 2021). Therefore, understanding how vegetation responds to climate drivers becomes crucial in land-atmosphere modeling (Mahecha et al., 2022).

The vegetation response changes over time, showing seasonal patterns and long-term trends (Slayback et al., 2003; Mahecha et al., 2010; De Jong et al., 2011, 2012; Linscheid et al., 2020). This variability is influenced by climate variables such as radiation, temperature, and precipitation, which affect vital biosphere processes such as photosynthesis. These meteorological variables create a range of conditions for the vegetation, from optimal to stressful (Nemani et al., 2003; Seddon et al., 2016). However, the relationship between climate and biosphere involves complex interactions due to the nonlinear response of vegetation to climate drivers (Foley et al., 1998; Zeng et al., 2002; Papagiannopoulou et al., 2017). Furthermore, ecosystems exhibit memory effects (Johnstone et al., 2016; Pappas et al., 2017) that can put their long-term resilience at risk (De Keersmaecker et al., 2016). For instance, extreme heatwaves can negatively impact leaf growth and development that, when coupled with drought conditions, can lead to tree mortality (Teskey et al., 2015). Extreme perturbations can cause irreversible damage (Scheffer et al., 2001) and reduce an ecosystem's resilience (Ghazoul et al., 2015). These factors collectively contribute to the challenge of predicting the vegetation and climate system.

Traditionally, terrestrial biosphere models have played a pivotal role in simulating the impact of climate variability, for example, in land carbon fluxes (Sitch et al., 2003; Krinner et al., 2005). Process-based models are inherently complex and demand substantial computational resources (Watson-Parris, 2021). Despite their robust foundation in physical laws, they sometimes fall short in mirroring the complex dynamics of land-atmosphere interactions accurately (Papale and Valentini, 2003). In response, there has been a growing reliance on machine learning (ML) techniques in Earth sciences (Zhu et al., 2017; Tuia et al., 2023). These methods represent powerful modeling tools, able to find patterns in data that process-based models may not be able to capture. As a consequence, applications of ML models in land-atmosphere interactions are wide-ranging, from local-to-global flux upscaling (Papale et al., 2015; Jung et al., 2020) to the prediction of ecosystem states (Kang et al., 2016; Zhang et al., 2021; Peng et al., 2022). More specifically, recurrent neural networks (RNNs) represent suitable architectures for modeling complex Earth system dynamics (Reichstein et al., 2019; Camps-Valls et al., 2021b) due to their

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

ability to encode nonlinear temporal dependencies (Bengio et al., 1994; LeCun et al., 2015) and capacity to retain information from past inputs (Elman, 1990).

However, RNNs have technical challenges associated with gradient-based training, including the issues of vanishing and exploding gradients, which impede network convergence (Hochreiter, 1998; Pascanu et al., 2013). To tackle these problems, specialized RNN architectures have been developed. Long short-term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) maintain the gradient-based training of the original RNNs while addressing these problems through gating mechanisms. Another architecture, the gated recurrent unit (GRU) (Cho et al., 2014), further refines the LSTM approach, providing comparable results with computational efficiency (Chung et al., 2014). In contrast, echo state networks (ESNs) employ a distinct approach by training only the last layer through linear regression (Jaeger, 2001). The absence of derivatives guarantees non-vanishing or exploding gradients, offering an alternative training solution to gating. The improvements provided by both the gated models and the ESNs allowed the application of these models to different tasks such as rainfall-runoff modeling (Kratzert et al., 2018; Gauch et al., 2021), sea surface temperature estimation (Zhang et al., 2017; Walleshauser and Bollt, 2022) and chaotic systems forecasting (Pathak et al., 2017, 2018; Vlachas et al., 2020; Chattopadhyay et al., 2020; Gauthier et al., 2022) among others.

Expanding on the utility of RNNs, particularly LSTM networks, recent studies have demonstrated their effectiveness in addressing specific challenges within land-atmosphere interactions, modeling land fluxes from meteorological drivers (Reichstein et al., 2018). Further studies reinforced the suitability of RNN approaches for land-atmosphere interactions (Chen et al., 2021). In Besnard et al. (2019), the authors employ LSTMs to predict land fluxes from remote sensing data and climate variables to explore the memory effects of vegetation. Additionally, the work provides a comparison with random forests, showing the better performance of deep learning models in this task. Further explorations in dynamic memory effects with LSTMs have been carried out by Kraft et al. (2019). Given the recent findings on the utility of RNNs and LSTMs in studying dynamics, it is timely to investigate if these tools can accurately predict extreme biosphere dynamics.

Can recurrent neural networks learn the vegetation's extreme responses to climate drivers? Recurrent architectures can embed the dynamics of target systems into their higher dimensional representation (Funahashi and Nakamura, 1993; Hart et al., 2020). When considering the nature of extremes in dynamical systems, they can be understood as specific regions of the phase space (Farazmand and Sapsis, 2019). Combining this with the embedding abilities of RNNs, offers an explanation for the observed efficacy of ESNs and LSTMs in learning extreme events within controlled environments, or "toy models" (Srinivasan et al., 2019; Lellep et al., 2020; Pyragas and Pyragas, 2020; Ray et al., 2021; Meiyazhagan et al., 2021; Pammi et al., 2023). However, the applicability of these findings to land-atmosphere interactions remains unclear. Unlike the systems investigated in previous studies, biosphere dynamics are characterized by stochasticity (Dijkstra, 2013) while their measurements present noise (Merchant et al., 2017). Therefore, exploring how recurrent architectures perform in the specific context of vegetation extremes in ecosystem dynamics is imperative.

To address this question, we investigate the ability of various recurrent architectures to model the response of vegetation states, i.e., spectral reflectance indices, to climate drivers. Vegetation interacts with sunlight, showing specific spectral responses that can be altered during extreme events such as heat waves. Vegetation indices, obtained from the spectral response through

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

linear or nonlinear transformations, are used to quantify these changes, isolating vegetation properties from other influences such as soil background (Zeng et al., 2022); detailed lists of these indices are provided in Zeng et al. (2022) and Montero et al. (2023). Our study focuses on the normalized difference vegetation index (NDVI) (Rouse et al., 1974), which indicates vegetation greenness. To build a model that can accurately predict NDVI responses to climate conditions, we use climate-related variables, such as temperature and precipitation, as inputs.

The goal of this study is threefold: 1) to further solidify the viability of the recurrent neural network approach to model ecosystem dynamics, 2) to investigate whether these models can capture extreme vegetation responses to climate forcing, and finally, 3) to investigate whether a specific RNN architecture is more suited for these tasks. To evaluate the performance of the models, we use metrics such as the normalized root mean squared error (NRMSE) and symmetric mean absolute percentage error (SMAPE) in conjunction with information theory-based measures, namely the entropy-complexity (EC) plots (Rosso et al., 2007). The latter approach quantifies each model's ability to capture the target dynamics. Using the model's residuals, defined as the difference between the prediction and the actual signal, the EC plots return an intuition of the model's ability to capture dynamics beyond the seasonality (Sippel et al., 2016). Additionally, we investigate whether recurrent architectures can effectively capture vegetation responses to extreme events. To our knowledge, no study has compared the performance of recurrent models in the context of extreme events or ecosystem dynamics.

The remainder of this paper is structured as follows. In Sect. 2.1, we present the data we used, including the site selection process and pre-processing steps. Next, in Sect. 2.2, we describe the architecture of the recurrent neural networks and formalize the task. Following this, Sect. 2.3 and Sect. 2.4 give a background on the methods of backpropagation training and echo state networks, respectively. In Sect. 2.5 we describe the procedure to identify extreme events in the NDVI time series. We detail the metrics used in the study in Sect. 2.6. More specifically, in 2.6.1, we introduce NRMSE and SMAPE; in 2.6.2, we illustrate the EC plots; finally, in 2.6.3, we describe the metrics used for evaluating model performance on predicting extreme events. We show our results in Sect. 3. In Sect. 3.1 we compare model performances using NRMSE and SMAPE. Additionally, in Sect. 3.2 we show the results for the EC plots. Finally, we illustrate the models' capability to predict extreme events in 3.3. We draw conclusions and discuss broader implications in section 4.

2 Methods

2.1 Data and Pre-processing

We used optical remote sensing data of forest sites to measure biosphere dynamics, specifically the normalized difference vegetation index (NDVI) (Rouse et al., 1974), which we define as the "target" variable. However, employing NDVI presents drawbacks (Camps-Valls et al., 2021a). Namely, it has a saturating and nonlinear connection to green biomass and responds to greenness rather than the actual plant photosynthesis process. Despite these limitations, this index has been used successfully for various purposes, including, but not limited to, evaluating ecosystem resilience (Yengoh et al., 2015) and tracking the decline of vegetation greenness in the Amazon forests (Hilker et al., 2014). Additionally, NDVI was shown to be a good indicator of vegetation response to extreme climate events (Liu et al., 2013).

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

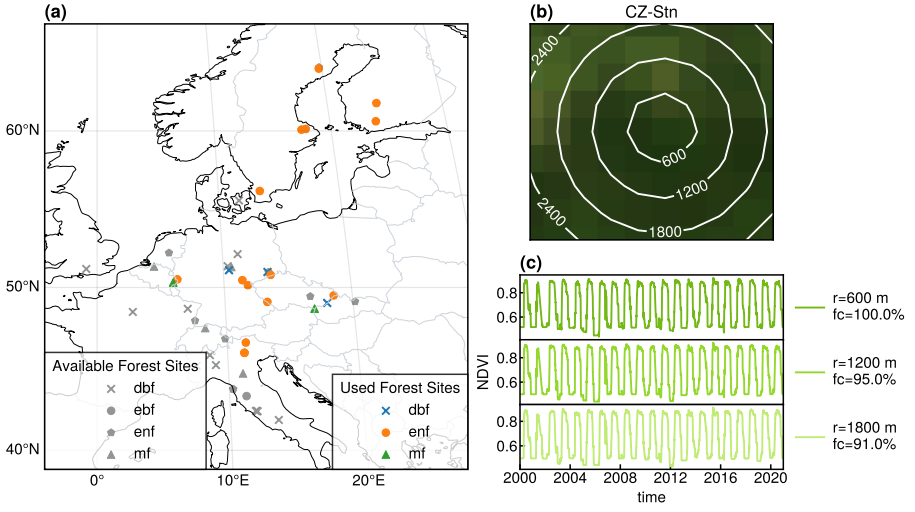

Figure 1. Chosen locations and pre-processing In **(a)**, we show the available and used locations for the study and their forest type. We used locations with more than 80% forest cover, not using the remaining ones depicted in gray in this figure. In **(b)**, we show the process of including additional pixels into the cube for an example location (CZ-Stn). While different radii are shown, we use all the pixels in a single cube. Finally, in **(c)**, we show the NDVI signal corresponding to the mean of the pixels in the area with radius r previously shown for the same example location. Additionally, it also indicates the percentage of the pixels that are flagged as forest (fc).

We used NDVI values obtained from the moderate resolution imaging spectroradiometer (MODIS) in the FluxnetEO dataset v1.0 (Walther et al., 2022). This dataset is a multi-dimensional array of values referred to as a data cube (Mahecha et al., 2020). It comprises a collection of labeled univariate time series at the geographic location of eddy covariance (EC) towers. EC towers are specialized measurement stations designed to capture and quantify land-atmosphere fluxes and meteorological conditions, providing insights into ecosystem health and functionality (Aubinet et al., 2012). The FluxnetEO datasets are gap-filled using measurements from EC towers, thus providing a higher resolution product in time and space compared to the raw MODIS data.

The data cubes in the dataset present a spatial resolution of 500 m and a daily temporal resolution, covering the period of 2000-2020 (inclusive). The cubes span a pixel of size 3 km $\times$ 3 km centered on the EC towers in each location. We further pre-processed the data by averaging the NDVI for each timestep as the mean of all the pixels within the cube, see 1. Additionally, we smoothed the signal using a Savitzky-Golay filter (Savitzky and Golay, 1964; Steinier et al., 1972) with a 7-day window and a polynomial order of 3 (Chen et al., 2004) to eliminate any potential artifacts caused by noise.

We selected study sites based on their forest cover percentage, ensuring over 80% forest cover in each cube, see 1. The forest masks were obtained from the Copernicus Global Land Service (CGLS) product (Buchhorn et al., 2020), which has a resolution of 100 m. These sites represent three different forest types: evergreen broadleaved forests (EBF), mixed forests (MF), and deciduous broadleaved forests (DBF). We chose to study forest sites for their importance to the carbon cycle and because they are the least affected by human influence at a daily time scale. Consequently, we employed the described approach,

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

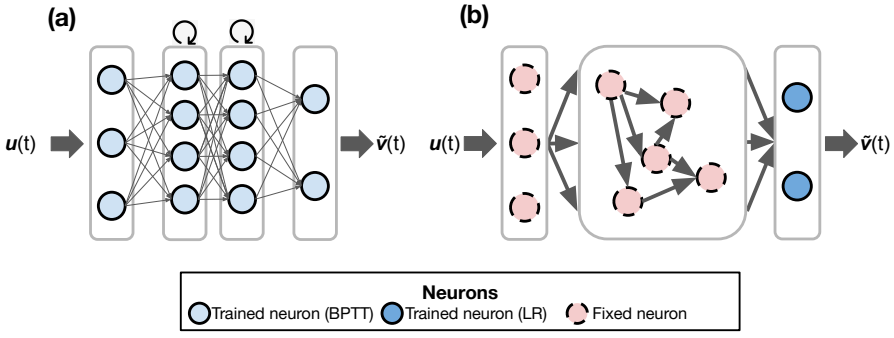

Figure 2. Training methods. Both diagrams illustrate input and predicted data denoted by $\boldsymbol{u}(t)$ and $\tilde{\boldsymbol{v}}(t)$, respectively. Diagram **(a)** presents the training methodology of ESNs. The initial two layers are randomly generated and remain untrained, while only the final layer undergoes one-shot training via linear regression (LR). In contrast, diagram **(b)** portrays the conventional approach used by RNNs, GRUs, and LSTMs, involving backpropagation through time (BPTT). Training encompasses the input layer (comprising three neurons in this instance), stacked recurrent layers (two layers, each with four neurons), and the output layer (two neurons). All these neurons are trained through backpropagation. The circular arrow atop signifies the recurrent layers. The number of neurons and internal recurrent layers serves visualization purposes only.

detailed in Fig. 1, to minimize further imperfections in the vegetation signal caused by human intervention. As a result of this selection criterion, the number of study sites is reduced from 42 to 20.

In this study, we used climate variables as input variables to predict the target variable. Following machine learning terminology, we will refer to them as "features." We selected air temperature (mean, minimum, and maximum), mean sea level pressure, mean global radiation, and precipitation as features. The climate data was obtained from the E-OBS product v26.0e (Cornes et al., 2018). Based on in situ observations, this dataset is spatially interpolated to cover most of the European continent. The spatial resolution is 0.1 degrees (11.1 km $\times$ 11.1 km), and the temporal resolution is daily. The time length of the feature variables is identical to the target variable and spans from 2000 to 2020 (inclusive).

2.2 Approach and Models

We aim to learn the NDVI behavior of forests as a proxy for ecosystem response to climate drivers. We use temperature (mean, minimum, maximum), precipitation, pressure, and radiation. We assume that the knowledge of the target variable is constrained to a certain period, after which only the features are available. Our goal is to build a model that, using those features, can predict the target variable. This is obtained by training the models on the available time interval, which comprises the years 2000-2013 (inclusive) for this study. After the training, the models only use feature variables to predict the NDVI for the remaining period, 2014-2020. Further details for the training setup can be found in Appendix B.

The task can be formalized as follows. We aim to approximate the target variable $\boldsymbol{v}(t) \in \mathbb{R}^{d_v}$ using input data $u(t)$. In the context of this study, we retain a unidimensional approach with $d_v = 1$, given that we have a single target variable, the NDVI.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

We assume that both $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$ are components of the same dynamical system. Under this assumption, the behavior $\boldsymbol{v}(t)$ is influenced by $\boldsymbol{u}(t)$, allowing us to leverage the information from $\boldsymbol{u}(t)$ to estimate $\boldsymbol{v}(t)$. Our setup consists of two sets of data, $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$, which can be measured over a given time period $t \in \{1, 2, \ldots, T\}$. After time T only $\boldsymbol{u}(t)$ remains measurable, and the goal of the recurrent architectures is to create a model efficient in modeling $\boldsymbol{v}(t)$ based solely on the available $\boldsymbol{u}(t)$ data (Lu et al., 2017). In control theory, a model that can predict $\boldsymbol{v}(t)$ based on $\boldsymbol{u}(t)$ data is called "observer" (Hermann and Krener, 1977; Ogata et al., 2010). Despite having different training methods, all the models in this study are built to perform this task.

We consider four different recurrent architectures: recurrent neural networks (RNNs), long short-term memory based networks (LSTMs), gated recurrent unit-based networks (GRUs), and echo state networks (ESNs). Each of these models presents an internal state $\boldsymbol{x}(t) \in \mathbb{R}^{d_x}$, which encodes the temporal dependencies of the input data $\boldsymbol{u}(t) \in \mathbb{R}^{d_u}$, where $d_u = 6$ in our study, representing the dimension of the input data. The internal size d_x is chosen to be $d_x > d_u$. A defining feature of these architectures is the recursive transmission of internal states, facilitating historical data retention as the model progresses through subsequent steps. This evolution of the generic RNN model is given by (Sutskever, 2013)

$$\boldsymbol{x}(t) = H(\boldsymbol{x}(t-1), \boldsymbol{u}(t); \theta), \tag{1}$$

where θ represents the weights and biases of the model, also called parameters, and H represents a generic RNN update function.

2.3 Training - Backpropagation Through Time

Deep learning models like feed-forward neural networks (FFNN) adjust their weights at each training step $t \in \{1, 2, \ldots, T\}$ using a method called backpropagation (BP) (Rumelhart et al., 1986). BP relies on a "loss function" $\mathcal{L}$ given by $\mathcal{L}(\theta) = \sum_{t=1}^{T} \mathcal{L}_t(\theta)$, where θ stands for the network's parameters. Leveraging the loss function, BP minimizes the difference between the model's predicted and actual output by adjusting the model's weights. To do this, BP calculates the gradient of the loss function with respect to each weight and then updates the network weights in a direction that minimizes the loss. One of the most common approaches to minimizing the loss function is stochastic gradient descent (SGD) (Bottou, 2012). However, one of BP's limitations is that it does not account for time dependencies.

For time-dependent models, such as RNNs, LSTMs, and GRUs, handling sequential data poses additional challenges. Backpropagation through time (BPTT) (Rumelhart et al., 1986) is a specialized training method designed for these architectures (Werbos, 1990). Central to BPTT is the notion of "unrolling" the network over time, effectively transforming it into a FFNN where backpropagation can be applied. This allows the model to calculate errors and update weights across the entire sequence, making it possible to capture long-term dependencies and patterns in time series data.

However, applying BPTT across the entire sequence can be computationally intense and time-consuming. Moreover, it often reduces the error to a very small amount by the end of the sequence. To avoid this issue, we use a truncated version of BPTT, limiting the backpropagation to a fixed number of steps, denoted by k (Williams and Zipser, 1995), which is smaller than the total number of training time steps, T (Aicher et al., 2020).

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

This truncated approach ensures efficiency but requires transferring the last hidden state from the truncated section to the initial state in the following sequence. This step maintains some memory and ensures the network's training process continuity.

The final output of all the models considered in this study comes from a feed-forward layer. The parameters of this layer are also trained using BP. The following equation describes the feed-forward layer:

$$\tilde{\boldsymbol{v}}(t) = \sigma(\mathbf{W}^v \boldsymbol{x}(t) + \boldsymbol{b}^v), \tag{2}$$

where $\tilde{\boldsymbol{v}}(t) \in \mathbb{R}^{d_v}$ is the predicted output, $\boldsymbol{x}(t) \in \mathbb{R}^{d_x}$ is the hidden state of the model at time t, $\mathbf{W}^v \in \mathbb{R}^{d_v \times d_x}$ is the weight matrix and $\boldsymbol{b}^v \in \mathbb{R}^{d_x}$ is a bias vector. Additionally, σ represents the activation function. This procedure is illustrated graphically in 2a. Finally, the full details of the models are given in appendix A1 for the RNNs, A2 for the LSTMs, and A3 for the GRUs.

The RNNs, LSTMs, and GRUs have been implemented using the `PyTorch` library (Paszke et al., 2019) accessed through `Skorch` (Tietz et al., 2017).

2.4 Training - Echo State Approach

Echo state networks (ESNs, Jaeger, 2001), along with liquid state machines (Maass et al., 2002), belong to the larger family of reservoir computing (RC) models, based on a shared theoretical background (Verstraeten et al., 2007). The fundamental idea of ESNs is to project the training data into a higher-dimensional, nonlinear system named the "reservoir" through an input layer. This process transforms the input data into vectors called "states." After the data passes through the reservoir, the states are collected. The model is then trained by regressing these states against the target data.

More specifically, the ESNs have three layers: an input layer $\mathbf{W}_{\text{in}}$, a reservoir layer $\mathbf{W}$ and an output layer $\mathbf{W}_{\text{out}}$. A distinctive feature of ESNs is that the weights in the input and reservoir layers are fixed. These weights are generated randomly and do not change during training. This approach contrasts with conventional neural networks, where weights are continuously updated during training to reduce errors. For each training time step $t \in \{1, 2, \ldots, T\}$, the hidden states, indicated as $\boldsymbol{x}(t)$, are preserved and accumulated in a matrix $\mathbf{X} \in \mathbb{R}^{d_x \times T}$. Indicated as a "state matrix," this matrix effectively represents the system's dynamics. The last layer of the ESN is obtained through a linear regression operation that uses the states matrix to generate a feed-forward layer, creating the network's output layer:

$$\mathbf{W}_{\text{out}} = \mathbf{Y}^{\text{target}} \mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top + \beta \mathbf{I})^{-1} \tag{3}$$

where $\mathbf{I}$ is the identity matrix and β is a regularization coefficient. The matrix $\mathbf{Y}^{\text{target}} \in \mathbb{R}^{d_v \times T}$ is generated with the desired output $\boldsymbol{v} \in \mathbb{R}_v^d$ stacked column-wise. While layers of recurrent models trained through BPTT can be stacked, ESNs are usually computed from a single inner layer (reservoir).

The construction and training of the ESN allow for the faster computational time of all the proposed models while also solving the vanishing and exploding gradients since no derivatives are taken at any step of the process. An illustration of the ESN approach to training is provided in Fig. 2b. The full details for the ESNs are given in appendix A4. For the implementations of the ESNs, we relayed on the Julia package `ReservoirComputing.jl` (Martinuzzi et al., 2022).

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

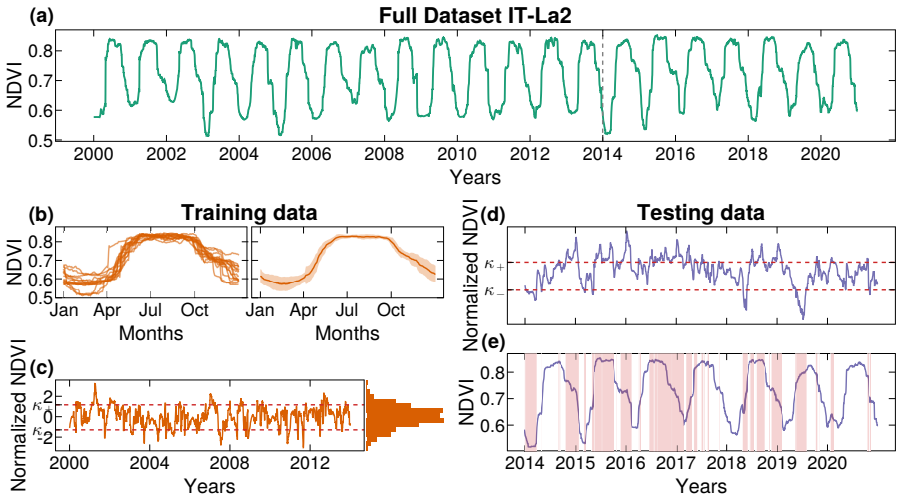

Figure 3. Definition of extreme events. In **(a)**, the NDVI of the dataset is plotted at an example location, IT-La2. The gray dotted line in 2014 represents the separation between training and testing data. In **(b)**, we show each yearly cycle of the training data: on the left are the full signals, and on the right are the mean and standard deviation (SD). In **(c)**, we show the result of the normalization of the training dataset using the mean and SD obtained in the previous step. At this stage, we also define the quantile (fixed at 0.90 in this example) and define the values of the thresholds κ_+ and κ_-. These values are then used in **(d)**, where we identify the extremes in the normalized testing data. The normalization is also done using the mean and SD of the training data. Finally, in **(e)**, the testing data is again shown in its raw form to showcase the extreme response of the vegetation.

2.5 Anomalies and Extremes

In this study, we adopt the approach outlined by Lotsch et al. (2005) to define anomalies based on the seasonal variability observed in the signal. The process described in this section is depicted in Fig. 3. The anomalies at a given time l, denoted by $A(l)$, are defined as follows:

$$A(l) = \frac{s(l) - \bar{s}(l)}{\sigma(l)}. \tag{4}$$

Here, $s(l)$ represents the signal at time l, $\bar{s}(l)$ is the mean of the signal at that time, and $\sigma(l)$ refers to its standard deviation. In this context, the time variable l denotes the specific day of the year, i.e., the third of March, and is determined based on a multi-year mean. The normalization process ensures that the signal exhibits a zero mean, which facilitates identifying extreme events as data points that fall outside a specified distribution range.

After normalizing the data and delineating the anomalies, we characterize extreme events as data points falling outside a specific quantile, chosen to be 0.90, 0.91, …, or 0.99. Based on the selected quantile, we define two threshold parameters κ_+ and κ_- to represent positive and negative extremes, respectively. We determine these parameters individually for each site involved in our study.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

During the training phase, which spans 2000 to 2014, we determine the threshold values κ_+ and κ_-. We apply these threshold values to the test dataset comprising the years from 2014 to 2020 to determine the extreme events in this latter dataset. The normalization of the data in the test datasets is done with the mean and SD values obtained from the training dataset.

2.6 Metrics

2.6.1 General Metrics

In this study, we evaluated the predictive accuracy of our model using two primary metrics: the normalized root mean square error (NRMSE) and the symmetric mean absolute percentage error (SMAPE).

The NRMSE is derived from the root mean square error (RMSE) (Hyndman and Koehler, 2006), and is defined as:

$$\text{NRMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\tilde{v}(t) - v(t))^2}{N}} \frac{1}{v_{\max} - v_{\min}}, \tag{5}$$

where $v(t)$ represents the observed target variable at the t-th observation among a total of N observations, and $\tilde{v}(t)$ is the corresponding model prediction. To facilitate comparisons across different sites, we normalized this metric using the range of the observed data, which is computed as the difference between the maximum $v_{\max}$ and minimum $v_{\min}$ observed values, with $t \in \{1, ..., n\}$.

In addition to NRMSE, we use SMAPE to assess the predictive performance of our model. SMAPE, which is a dimension-agnostic measure, is given by the formula

$$\text{SMAPE} = \frac{100}{n} \sum_{t=1}^{n} \frac{|v(t) - \tilde{v}|}{|v(t)| + |\tilde{v}|}, \tag{6}$$

where n denotes the number of observations, this metric was chosen based on its ability to provide a symmetric measurement of the absolute percentage error, thereby affording a balanced view of the forecast accuracy (Makridakis, 1993; Hyndman and Koehler, 2006).

2.6.2 Entropy-Complexity Plots

In this study, we also use information theory-based quantifiers to analyze the model's residuals, defined as the differences between the model's prediction and the actual measurements. Based on the approach proposed by Sippel et al. (2016), we employ entropy-complexity (EC) plots. These plots return a visual representation of the amount of information still present in the residuals. Higher information content would indicate that the models do not sufficiently approximate the target variable. On the other hand, values of EC closer to white noise would suggest that the models fully reproduce the target variable's behavior. We illustrate this approach, closely following the exposition provided by Sippel et al. (2016).

To generate the EC plots, we consider a metric $\mathcal{H}[P]$ of a probability distribution $P = \{p_i; i = 1, ..., N\}$, with N possible states and with $\sum_{i=1}^{N} p_i = 1$, to quantify the information content of a time series. One such metric is the Shannon entropy $\mathcal{S}[P]$,

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

$$\mathcal{S}[P] = -\sum_{i=1}^{N} p_i \ln[p_i], \tag{7}$$

which is maximized $\mathcal{S}[P_e] = \mathcal{S}_{\max} = \ln N$ for the uniform distribution $P_e = \{p_i = \frac{1}{N}; \forall i = 1, \ldots, N\}$. We can then define the normalized entropy

$$\mathcal{H}[P] = \frac{\mathcal{S}[P]}{\mathcal{S}_{\max}}, \tag{8}$$

which is the first metric used in the EC plots. In addition to the information content of a time series, we are interested in quantifying the complexity $\mathcal{C}[P]$. Following Lopez-Ruiz et al. (1995), we use a definition of complexity $\mathcal{C}[P]$, which is the product of a measure of information, such as entropy $\mathcal{H}[P]$, and a measure of disequilibrium $\mathcal{Q}_J[P, P_e]$.

$$\mathcal{C}[P] = \mathcal{Q}_J[P, P_e]\mathcal{H}[P], \tag{9}$$

In this context, disequilibrium takes the meaning of distance from the uniform distribution of the available states of a system. The definition of disequilibrium $\mathcal{Q}_J[P, P_e]$ makes use of the Jensen-Shannon divergence $\mathcal{J}[P, P_e]$, which quantifies the difference between probability distributions (Grosse et al., 2002), and it is defined as

$$\mathcal{J}[P, P_e] = \mathcal{Q}_0 \left\{ S\left[\frac{P + P_e}{2}\right] - \frac{1}{2}(S[P] - S[P_e]) \right\}, \tag{10}$$

where $\mathcal{Q}_0$ is a normalization constant, (Lamberti et al., 2004; Rosso et al., 2007), chosen such that $\mathcal{Q}_J[P, P_e] \in [0, 1]$.

The computations of entropy and complexity rely on the probability distribution associated with the data. To determine this distribution, we leverage the method proposed by Bandt and Pompe (2002), which analyzes time series data by comparing neighboring values. It involves dividing the data into a set of patterns based on the order of the values and then calculating the probability of each pattern occurring (Rosso and Masoller, 2009). The pattern separation is obtained by embedding the time series in a D dimensional space with a time lag τ. We set D = 6 and $\tau = 1$ as proposed by (Rosso et al., 2007; Sippel et al., 2016). The complexity measure's theoretical upper and lower bounds can now be computed (Martin et al., 2006) and are shown in the plots.

The calculations of the EC plots were performed with the Julia package `ComplexityMeasures.jl` (Haaga and Datseris, 2023) from `DynamicalSystems.jl` (Datseris, 2018).

2.6.3 Extremes as Binary Events

To analyze extreme events, we set thresholds κ_+ and κ_-, as described earlier (in Sect. 2.5) to identify values as either extremes or not extremes. We adopt this binary approach for both the observed data and the data predicted by the models.

Following the method outlined by Hogan and Mason (2011), we classify the outcomes into four categories: hits a, where the model correctly identifies an extreme event; false alarms b, where the model incorrectly flags a value as an extreme event; misses c, where the model fails to identify an extreme event; and correct rejections d, where the model correctly identifies a non-extreme event.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Given n observation points, we employ the following metrics to assess the model's performance on detecting extremes, (Barnes et al., 2009). We define the probability of detection POD as the ratio of correctly identified extreme events and the total number of extreme events, the probability of false detection POFD as the ratio of incorrectly flagged extreme events and all events that are not extreme, the probability of false alarm POFA as the ratio of false alarms and all predicted extreme events, and the proportion correct PC as the ratio of all accurate predictions (both hits and correct rejections) and the total number of observations, given by

$$POD = \frac{a}{a+c},\ POFD = \frac{b}{b+d},\ POFA = \frac{b}{a+b},\ PC = \frac{a+d}{n}, \tag{11}$$

respectively.

These metrics are standard tools for evaluating deep learning models for predicting atmospheric variables such as wind speed (Scheepens et al., 2023) and precipitation (Shi et al., 2015).

3 Results

In the following, we first compare the different models by standard evaluation metrics such as the normalized root mean squared error, NRMSE, and the symmetric mean absolute percentage error, SMAPE (Sect. 3.1). Second, we extend the comparison to information-theoretic quantifiers in the entropy-complexity (EC) plane (Sect. 3.2). We perform this comparison for two cases: (i) the full time series (FS), which encompasses all data points in the year, and (ii) the meteorological growing season (GS), which encompasses the months between May and September only. This division helps us differentiate the models' ability to capture the seasonal cycle, dominated by an oscillation, from their performance in the more stable growing season conditions, where more minor variations are likely harder to be represented. Last, we focus on extreme events and compare the models' performance on extreme events in 3.3.

3.1 Comparison of Prediction Accuracy

We present our results in Table 1, showing that the ESN outperformed all other models for both the FS and GS signals, closely followed by the LSTM. While the GRU exhibited comparable results to the LSTM for the FS signal, the differences became more pronounced in the GS signal. In contrast, the RNN consistently delivered the least favorable results across the board and exhibited the highest standard deviation (SD) among the analyzed models. In Fig. A1, we include the comparison of these metrics per site, which shows great variation in the models' performance across different sites.

The performance rankings of the models remained consistent when transitioning from the FS signal to the GS signal. Notably, the SMAPE metric indicated increased accuracy for all models, while the NRMSE metric suggested decreased accuracy. Similarly, we observe a reduction in the SD for the SMAPE but an increase in the NRMSE. This increase is very noticeable in the GRU and RNN models.

The models generally exhibited similar performance, with the ESN yielding slightly better results and the RNN demonstrating the least accurate forecasts. The gated methods showed similar performances. While these metrics provide an initial picture

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

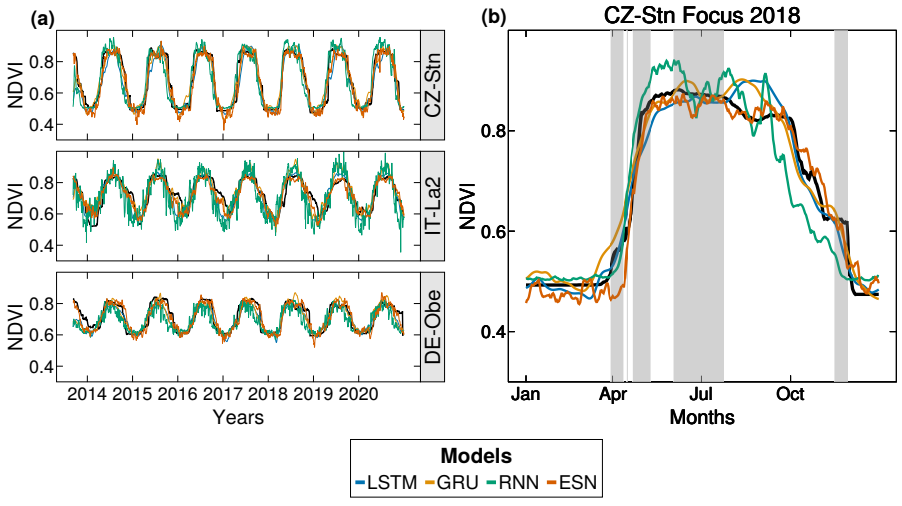

Figure 4. Time series and predictions for selected locations. This figure illustrates the predictive performance of four distinct recurrent architectures at specified NDVI time-series locations. The baseline is shown in black, while the predictions only use a singular run from a set of 50 per model. Panel **(a)** delineates the results obtained at three selected locations. 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. It is pertinent to underline that the predictions generated by the RNNs carry considerable noise, thereby affecting the interpretations drawn from the entropy-complexity plots.

of the models' performance in the task, the results do not indicate a clear best performer among the models. These similarities in the results' metrics emphasize the need for further exploration using alternative evaluation tools.

Table 1. Accuracy of the models. The table illustrates the performance of the four different models across various scenarios. "FS" represents the "full season," representing the full data set used for predictions without isolating specific events or months. "GS" denotes the "growing season," which encompasses the peak summer months in addition to May and September. Models with the highest accuracy in each category are highlighted using bold text. The arrows pointing down (↓) near the metric's name indicate that smaller values represent higher accuracy.

		LSTM	GRU	RNN	ESN
FS	SMAPE ↓	5.97 ± 1.84	6.19 ± 2.23	7.86 ± 2.69	$\mathbf{4.77} \pm 1.14$
	NRMSE ↓	$0.172 \pm 3.23 \cdot 10^{-2}$	$0.170 \pm 3.68 \cdot 10^{-2}$	$0.218 \pm 4.24 \cdot 10^{-2}$	$\mathbf{0.135} \pm 3.29 \cdot 10^{-2}$
GS	SMAPE ↓	4.57 ± 1.25	4.84 ± 1.65	6.57 ± 1.89	$\mathbf{3.14} \pm 0.82$
	NRMSE ↓	$0.222 \pm 7.04 \cdot 10^{-2}$	$0.229 \pm 9.18 \cdot 10^{-2}$	$0.297 \pm 10.61 \cdot 10^{-2}$	$\mathbf{0.153} \pm 5.19 \cdot 10^{-2}$

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

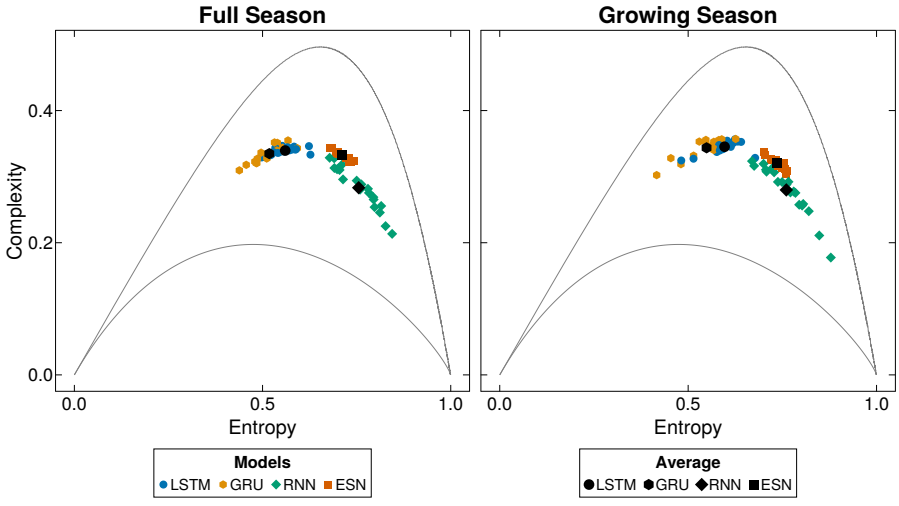

Figure 5. Entropy-complexity curves of model's residuals. The entropy-complexity plots are computed from the residuals of each model at each location (color). The mean of each model's performance over all locations is shown in black. These plots visualize the amount of information and complexity left in the residuals. Ideal values would reside in the lower left corner, symbolizing white noise in the residuals. The gray lines denote the theoretical upper and lower bounds.

3.2 Comparison of the Entropy-Complexity

Drawing from information theory, we use entropy-complexity (EC) plots (introduced in Sec 2.6.2) to examine the residuals, defined as the differences between the model's predictions and actual measurements. Residuals convey valuable information about a model's performance. In an ideal scenario, where a model perfectly represents system dynamics, these residuals should resemble white noise and position in the lower right corner of the EC plane (Sippel et al., 2016).

Figure 5 shows the EC plots of the models' residuals. Additionally, we show the mean of these metrics per model architecture across all sites. We find that the residuals cluster by model across all locations, with minimal overlap between each model's clusters. The LSTM and GRU models occupy the curve's ascending left side, indicating the presence of more structure in the residuals. In contrast, the ESN and, to a greater extent, the RNN are positioned closer to the white noise region, implying less structure in the residuals.

Comparing the FS signal to the GS signal shows no apparent differences in the results, underscoring the consistency of model performance. Based on the positioning of RNNs in the EC plots, we would expect a substantially improved performance of these models compared to the other model architectures. However, inspecting the predicted time series suggests a different explanation: the RNN model predictions shows large variability (Fig. 4, which obfuscates the underlying dynamics. Because the EC plots capture the resulting residuals' noise, one can misinterpret the positioning of RNNs as favorable results.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

3.3 Extreme Events

In this section, we use the metrics introduced in Sect. 2.6.3 to evaluate the models' ability to capture extreme vegetation dynamics. Fig. 6a shows a distinct division in the probability of detection (POD): the LSTM and ESN models exhibit low values, indicating an inability to detect extreme events. The RNN and GRU models show higher values, indicating better performance. However, the POD values are generally low for all models. Furthermore, all models display some level of overlap in the confidence bands. Finally, it is possible to observe a worsening of the results with increased quantiles. Figure 6b shows a lower probability of false detection (POFD) of ESNs, indicating that they perform better at avoiding the prediction of extremes when there are none compared to the other models. Conversely, the RNN, which predicts noisy behavior, as shown in Sect. 3.1, exhibits higher POFD values. The probability of false alarm (POFA; Fig. 6c) follows a similar pattern, with the ESN outperforming the other models. The GRU and LSTM display more comparable behaviors, while the RNN lags behind, showing the least favorable performance among all models. The displayed metric consistently leans towards the higher end of the potential value range. Like in the case of the POD, the values of the POFA get worse with each increase of the quantiles. Finally, Fig. 6d shows the probability of correctness (PC). The ESN leads the graph despite showing the lowest POD. Following closely are the LSTM and GRU, which exhibit similar performance. The RNN maintains the poorest performance across all metrics, emphasizing its limitations in capturing extreme vegetation dynamics.

To examine the prediction performance of extreme NDVI reduction during the summer season, we focus on June, July, and August using only κ_{-}, showcased in Fig. 7. In Fig. 7a, we see the probability of detection (POD) for different models. The RNN performs best, outperforming the LSTM, GRU, and ESN, delivering similar results. Figure 7b shows the probability of false detection (POFD). Here, the RNN performs the worst, while the ESN stands out with the best values and the lowest standard deviation (SD) among all models. The gating-based models, LSTM and GRU, perform almost identically. We find similar performances among models in Fig. 7c. The ESN demonstrates a decrease in the probability of false alarm (POFA) values for the 0.92 quantiles. Any small gaps in performance observed for lower quantiles are effectively bridged at higher values. Finally, Fig. 7d illustrates the ESN as the top-performing model, exhibiting the lowest SD among the models. The gated models, LSTM and GRU, display similar performances, while the RNN ranks as the poorest-performing model in this analysis of NDVI summertime decreases.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

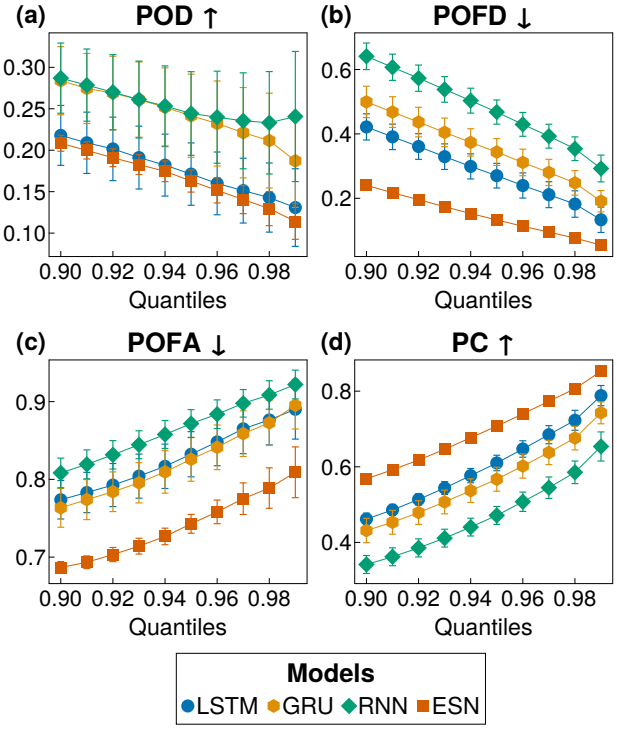

Figure 6. Extremes as binary events. The results of quantiles spanning from 0.90 to 0.99 are shown. All values are expressed as percentages. In **(a)**, we show the detection percentage (POD), indicating how many extreme events have been detected. In **(b)**, we show the percentage of false detection (POFD). In **(c)** we depict the probability of false alarms (POFA). Finally, in **(d)** we show the proportion correct (PC). The error bars represent the standard deviation over 50 simulations with different initializations for each model. The arrows (up ↑, and down ↓) by the metric's name indicate the direction of the optimal values.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

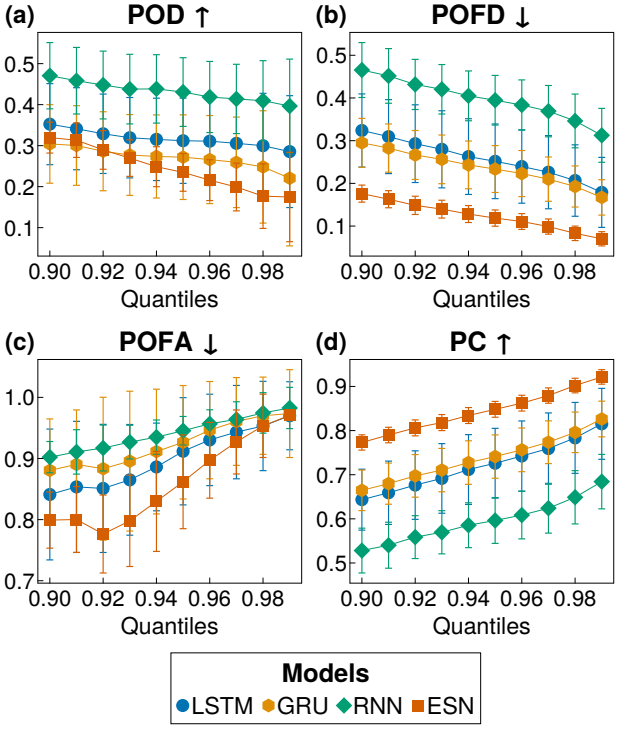

Figure 7. Negative extreme events for summer months. Only the negative extremes during the summer period are considered here. The results of quantiles spanning from 0.90 to 0.99 are shown. All values are expressed as percentages. In **(a)**, we show the detection percentage (POD), indicating how many extreme events have been detected. In **(b)**, we show the percentage of false detection (POFD). In **(c)** we depict the probability of false alarms (POFA). Finally, **(d)** we shown the proportion correct (PC). The error bars represent the standard deviation over 50 simulations with different initializations for each model. The arrows (up ↑, and down ↓) by the metric's name indicate the direction of the optimal values.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

4 Discussion

In the following section, we discuss the results, starting in 4.1 with the performance of the models and their comparison in modeling vegetation greenness in response to climate drivers. Subsequently, in 4.2, we analyze their performance in learning the extremes in the normalized vegetation index (NDVI) time series. Finally, we discuss the limitations of this study and highlight future research directions in 4.3.

4.1 Model's Performance and Comparison

Similar to previous investigations (Benson et al., 2023), this work underlines the ability of recurrent neural networks to model vegetation greenness. This is highlighted by the good values of NRMSE and SMAPE obtained by the ESNs and LSTMs. Comparing models, ESNs perform better in standard measures, but their NRMSE and SMAPE standard deviations (SD) overlap with those of LSTMs and GRUs. Furthermore, RNNs ranked last in this comparison. Based on these metrics, gated architectures and ESNs appear almost equally effective, with ESNs having a slight advantage. The performance varies considerably across sites, as illustrated in Fig. A1, and can explain the overlap in the standard deviation. Despite site specific training, this difference in performances still persists, and it could be the subject of future studies.

To extend the comparison, we also provide a comparison of entropy-complexity (EC) plots. The EC plots showed the presence of fewer leftover signals in the residuals of the ESNs' predictions. The LSTMs followed, with the GRUs showing the worst performance. Contrary to the previous metrics, RNNs seem to outperform all other models in the EC plots. However, examining their predictions shows this is not the case due to the high noise in RNN predictions, distorting the EC plot insights. Thus, our results illustrate that it is crucial not to rely solely on a single set of metrics when evaluating results in similar studies.

Existing studies center on creating a single global model (Kraft et al., 2019; Diaconu et al., 2022). However, we opted to train individual models for each site. Our goal was to examine how effectively each architecture could capture the specific dynamics of an ecosystem without giving the models any additional contextual information. This approach emphasizes the models' inherent capability to understand each ecosystem's unique attributes.

4.2 Do the Models Capture Extremes?

Our results for the analysis of vegetation extreme responses to climate drivers using daily data showed that none of the models notably outperformed the others in all the metrics considered. The ESNs performed well in avoiding false alarms but failed to obtain a high accuracy for predicting actual extremes, performing similarly to LSTMs. ESNs also showed the lowest standard deviation among all the models considered. RNNs performed worst among all metrics analyzed. Overall, all the models showed poor results for this task. When we only considered negative extremes in the summer season, the results showed again that the recurrent architectures could not correctly capture extremes in vegetation responses to climate forcing. The ESNs performed best in avoiding false extremes and the overall accuracy of the task's execution, while the RNNs showed the worst prediction performance.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Motivated by recent ESN applications to data from laboratory experiments (Pammi et al., 2023), we expected ESNs to outperform the other models in the extreme prediction task. Instead, the results showed that, while the ESN provided good results, they were not outstanding compared to the BPTT gated models. Additionally, none of the models showed good results, contrary to other applications in the Earth sciences showcasing the successful application of recurrent models to extremes (Frame et al., 2022).

4.3 Limitation and Future Directions

The study is subject to a few limitations. The most crucial limitation pertains to data quality. The provided NDVI data undergoes gap-filling, employing state-of-the-art techniques; nevertheless, certain artifacts persist in the final output. Another limitation arises from the availability of daily data solely spanning the last 21 years. This constraint hampers the training of machine learning models, notorious for their dependency on abundant data points for optimal performance. Marcolongo et al. (2022) show that a deep learning approach is suitable for detecting extreme vegetation responses. Using simulations, the study had access to 100,000 years of daily data, indicating that if enough data is available, neural networks can properly model extremes. Furthermore, the metrics we used were shown to be sensitive to factors like noise. Using various metric sets, we identified limitations in methods like the EC plots. For instance, the noise in the RNN's prediction was incorrectly attributed to residual noise, distorting the outcomes. Moreover, standard metrics like NRMSE and SMAPE provide a narrowed view. This is evident when comparing ESNs and LSTMs: even though they had similar scores with these metrics, their distinctions were clear in the EC plots.

ESN and LSTM demonstrated comparable performance in modeling vegetation greenness. However, these models exhibited limited predictive capability for extreme events. Spatial information has proven helpful in similar investigations (Requena-Mesa et al., 2021; Diaconu et al., 2022; Kladny et al., 2022; Robin et al., 2022; Benson et al., 2023) and it could be beneficial to explore the extent to which it contributes to extreme conditions. Furthermore, the models used could be tailored more to the task. In (Bonavita et al., 2023), the authors point out the limitations in using the mean-squared-error loss function, a practice still widely diffused in the field and adopted in this paper. Furthermore, Rudy and Sapsis (2023) show the superior performance of loss functions based on weighting and relative entropy compared to other loss functions in learning extreme events. Finally, further investigations could benefit from deeper explorations of ESNs. With their user-friendly nature, straightforward hyper-parameter tuning, and fast training, they stand out as a robust modeling method, able to compete with top-tier deep learning architectures such as LSTMs.

5 Conclusions

We compared the performance of recurrent neural networks in modeling biosphere dynamics, i.e., vegetation states, in response to climate drivers. Using daily data, we assessed the effectiveness of these network architectures in capturing extreme anomalies within these vegetation dynamics. To discern variations in performance across different scenarios, we employed various metrics such as normalized root mean square error and symmetric mean absolute percentage error paired with information theory

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

quantifiers. Our findings revealed that ESNs and LSTMs performed similarly for most analyzed metrics, indicating that no single model outperformed others. Additionally, all the models under investigation failed to model the extreme responses of the vegetation. This work highlights the necessity to continue refining and developing specialized models that can more adeptly capture extreme vegetation responses to climate factors.

Code availability. The code for this study is available at https://github.com/MartinuzziFrancesco/rnn-ndvi

Data availability. The data used in this study is available online:

- E-OBS dataset (Cornes et al., 2018) at www.ecad.eu/download/ensembles/download.php
- FluxnetEO dataset (Walther et al., 2022) at https://meta.icos-cp.eu/collections/tEAkpU6UduMMONrFyym5-tUW

Appendix A: Recurrent Neural Network Details

A1 Recurrent Neural Networks

The most basic version of RNN was proposed by Elman (Elman, 1990). The equations used to obtain the hidden state $\boldsymbol{x}(t) \in \mathbb{R}^{d_x}$ can be described as follows:

$$\boldsymbol{x}(t) = \sigma(\mathbf{W}_{\text{in}}^{x}\boldsymbol{u}(t) + \mathbf{W}^{x}\boldsymbol{x}(t-1) + \boldsymbol{b}^{x}) \tag{A1}$$

where σ is the activation function, $\mathbf{W}_{\text{in}}^{x} \in \mathbb{R}^{d_x \times d_u}$ and $\mathbf{W}^{x} \in \mathbb{R}^{d_x \times d_x}$ are the weight matrices and $\boldsymbol{b}^{x} \in \mathbb{R}_{x}^{d}$ is a bias vector. In addition, $\boldsymbol{u}(t) \in \mathbb{R}_{u}^{d}$ is the input vector at time t.

The main problem of this model is the vanishing and exploding gradient due to the multiple calculations of the gradient during backpropagation through time (Werbos, 1988).

A2 Long Short Term Memory

For LSTMs the hidden state $\boldsymbol{x}(t) \in \mathbb{R}^{d_u}$ is obtained as follows (Hochreiter and Schmidhuber, 1997)

$$\boldsymbol{f}(t) = \sigma_g(\mathbf{W}_{\text{in}}^{f}\boldsymbol{u}(t) + \mathbf{W}^{f}\boldsymbol{x}(t-1) + \boldsymbol{b}^{f}) \tag{A2}$$

$$\boldsymbol{i}(t) = \sigma_g(\mathbf{W}_{\text{in}}^{i}\boldsymbol{u}(t) + \mathbf{W}^{i}\boldsymbol{x}(t-1) + \boldsymbol{b}^{i}) \tag{A3}$$

$$\boldsymbol{o}(t) = \sigma_g(\mathbf{W}_{\text{in}}^{o}\boldsymbol{u}(t) + \mathbf{W}^{o}\boldsymbol{x}(t-1) + \boldsymbol{b}^{o}) \tag{A4}$$

$$\tilde{\boldsymbol{c}}(t) = \tanh(\mathbf{W}_{in}^{c}\boldsymbol{u}(t) + \mathbf{W}^{c}\boldsymbol{x}(t-1) + \boldsymbol{b}^{c}) \tag{A5}$$

$$\boldsymbol{c}(t) = \boldsymbol{f}(t) \odot \boldsymbol{c}(t-1) + \boldsymbol{i}(t) \odot \tilde{\boldsymbol{c}}(t) \tag{A6}$$

$$\boldsymbol{x}(t) = \boldsymbol{o}(t) \odot \sigma_x(\tilde{\boldsymbol{c}}(t)) \tag{A7}$$

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

where $\boldsymbol{f}(t)$ is the *forget* gate, $\boldsymbol{i}(t)$ is the *input* gate and $\boldsymbol{o}(t)$ is the *output* gate. The activation functions σ_g are usually set to be sigmoid, and σ_x is set to be the hyperbolic tangent. However, it can be set to unity for some variations of the model (e.g., "peephole" LSTM, Gers and Schmidhuber, 2000). The matrices $\mathbf{W}_{\text{in}}^{j} \in \mathbb{R}^{d_x \times d_u}$ and $\mathbf{W}^{j} \in \mathbb{R}^{d_x \times d_x}$ for $j \in \{f, i, o, c\}$ are the weight matrices while the vectors $\boldsymbol{b}^{j} \in \mathbb{R}^{d_x}$ for $j \in \{f, i, o, c\}$ are bias vectors. The vector $\boldsymbol{u}(t) \in \mathbb{R}^{d_u}$ represents the input vector at time t.

A3 Gated Recurrent Units

The equations to obtain the hidden state $\boldsymbol{x}(t) \in \mathbb{R}^{d_x}$ for GRUs are defined as follows (Cho et al., 2014):

$$\boldsymbol{r}(t) = \sigma(\mathbf{W}_{\text{in}}^{r}\boldsymbol{u}(t) + \mathbf{W}^{r}\boldsymbol{x}(t-1) + \boldsymbol{b}^{r}) \tag{A8}$$

$$\boldsymbol{z}(t) = \sigma(\mathbf{W}_{\text{in}}^{z}\boldsymbol{u}(t) + \mathbf{W}^{z}\boldsymbol{x}(t-1) + \boldsymbol{b}^{z}) \tag{A9}$$

$$\tilde{\boldsymbol{x}}(t) = \tanh(\mathbf{W}_{in}^{x}\boldsymbol{u}(t) + \mathbf{W}^{x}(\boldsymbol{r}(t) \odot \boldsymbol{x}(t-1)) + \boldsymbol{b}) \tag{A10}$$

$$\boldsymbol{x}(t) = \boldsymbol{z}(t) \odot \boldsymbol{x}(t-1) + (1 - \boldsymbol{z}(t)) \odot \tilde{\boldsymbol{x}}(t) \tag{A11}$$

where $\boldsymbol{r}(t)$ is the *reset* gate, $\boldsymbol{z}(t)$ is the *update* gate and $\boldsymbol{u}(t) \in \mathbb{R}^{d_u}$ is the input signal. The activation functions σ are set to be sigmoid. As in the LSTM, the matrices $\mathbf{W}_{\text{in}}^{j} \in \mathbb{R}^{d_x \times d_u}$ and $\mathbf{W}^{j} \in \mathbb{R}^{d_x \times d_x}$ for $j \in \{r, z, x\}$ are the weight matrices while the vectors $\boldsymbol{b}^{j} \in \mathbb{R}^{d_x}$ for $j \in \{r, z, x\}$ are bias vectors.

A4 Echo State Networks

The hidden state $\boldsymbol{x}(t) \in \mathbb{R}^{d_x}$ for the ESN is defined as (Jaeger, 2001):

$$\boldsymbol{x}(t) = (1-\alpha)\boldsymbol{x}(t-1) + \alpha\tanh(\mathbf{W}_{\text{in}}\boldsymbol{u}(t) + \mathbf{W}\boldsymbol{x}(t-1)), \tag{A12}$$

where α is the leaky coefficient and $\boldsymbol{u}(t) \in \mathbb{R}^{d_u}$ is the input data. Similarly as before the matrices $\mathbf{W}_{\text{in}} \in \mathbb{R}^{d_x \times d_u}$ and $\mathbf{W} \in \mathbb{R}^{d_x \times d_x}$ are the weights matrices, with the difference that this time these matrices do not undergo training or change. Since they are kept fixed, the initialization of these matrices also plays a role in predicting the model. The standard choices are to create $\mathbf{W}_{\text{in}}$, also called *input* matrix, as a dense matrix with weights randomly sampled from a uniform distribution in the range $[-\sigma, \sigma]$. The weight matrix $\mathbf{W}$ is usually referred to as the *reservoir* matrix, and it is usually built from an Erd "os–Rényi graph configuration. This matrix shows a high sparsity, usually in the $1 - 10\%$ range, and its values are also randomly sampled from a uniform distribution $\in [-1, 1]$. Subsequently, the matrix is scaled to obtain a chosen spectral radius $\rho(\mathbf{W})$. The values of the spectral radius, size of the matrix, and its sparsity are the main hyperparameters for ESN models.

The output is obtained through a linear feed-forward layer:

$$\tilde{\boldsymbol{v}}(t) = \mathbf{W}_{\text{out}}\boldsymbol{x}(t), \tag{A13}$$

where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{d_v \times d_x}$ is the *output* matrix. This matrix is the only one whose weights undergo training. Unlike the models illustrated before, this training is not done using BPTT but simple linear regression. During the training phase, all the inputs

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

$\boldsymbol{u}(t) \in \mathbb{R}^{d_u}$ are passed through the ESN, and the respective expansions (hidden states) are saved column-wise in a *states* matrix $\mathbf{X} \in \mathbb{R}^{d_x \times d_T}$ where T is the length of the training set $t = 1, ..., T$. In a similar way, the matrix $\mathbf{Y}^{\text{target}} \in \mathbb{R}^{d_v \times d_T}$ is built with the desired output $\boldsymbol{v} \in \mathbb{R}^d_v$ is stacked column-wise. This way, the output layer can be obtained using ridge regression with the following closed form:

$$\mathbf{W}_{\text{out}} = \mathbf{Y}^{\text{target}}\mathbf{X}^{\text{T}}(\mathbf{X}\mathbf{X}^{\text{T}} + \beta\mathbf{I})^{-1}, \tag{A14}$$

where $\mathbf{I}$ is the identity matrix and β is a regularization coefficient.

Appendix B: Computational Details

B1 Experimental Settings

All the figures in this paper (with the exception of Fig. 1) have been obtained using the Julia language (Bezanson et al., 2017) package `Makie.jl` (Danisch and Krumbiegel, 2021). All the models have been optimized using grid search. For each optimal set of hyperparameters, 100 different runs have been carried out for each model and site. The 50 best-performing of these runs have been used for the results shown here. Unless specified otherwise, the results showcased are mean and standard deviation over these 50 runs per location, over all locations.

B2 Details of the models

In the proposed models based on TBPTT, the parameters are optimized using the stochastic optimization method called Adam (Kingma and Ba, 2014), while dropout (Srivastava et al., 2014) is used to protect against over-fitting. Additionally, early stopping is employed to halt training when the validation loss has not changed, with a patience factor of 50 epochs. The weights θ_W are initialized using the technique proposed in (Glorot and Bengio, 2010), drawing from a uniform distribution. The biases θ_b are initialized by drawing values from a uniform distribution.

Different layers of recurrent networks are also stacked on top of each other, providing a deeper model. The number of layer and weights per layer is also optimized. All the hyper-parameters undergo optimization using grid search using the root mean square error (RMSE) as a guiding measure. Split temporal cross-validation is used (Cerqueira et al., 2020).

B3 Grid Search Parameters

Table A1 provides the parameters used for the grid search in this study. The values are either given as a list, divided by a comma (a, b, c), or as an interval, separated by a colon $(start : step : stop)$. In the first case, the values indicated are the values used. In the second case, the values used are between the first and last, with a step size indicated by the second value.

B1 Site Comparison

Here, we provide a comparison of the models' performance across the study sites.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Table A1. ESNs hyperparameters grid search values.

ESN			
Sparsity	Ridge coefficient	Leaky coeff.	Radius
$0.01 : 0.01 : 0.1$	$1.0 \times 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$	$0.5 : 0.05 : 1.0$	$0.9 : 0.05 : 1.5$

Table B1. RNN/LSTM/GRU hyperparameters grid search values.

RNN/LSTM/GRU			
Learning rate	Hidden dimension	Num. layers	Dropout
$1.0 \times 10^{-3}, 10^{-4}, 10^{-5}$	$32, 64, 128$	$2, 3, 4$	$0.1 : 0.1 : 0.4$

Author contributions. FM and KM conceptualized the work, and FM carried out the simulations. KM, MM, GCV, and TW provided suggestions for the analysis. KM and MM supervised the work. DM formulated the data pre-processing pipeline, while FM implemented it. FM wrote the manuscript with contributions from all authors.

Competing interests. The authors declare that there are no competing interests.

Acknowledgements. We thank Sophia Walther for explaining the FluxnetEO data set in detail. This research was supported by grants from the European Space Agency and ESA (AI4Science - Deep Extremes and Deep Earth System Data Lab). FM and MDM acknowledge the financial support from the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI. KM and MDM acknowledge support from the Saxon State Ministry for Science, Culture and Tourism (SMWK project 232171353). DM and MDM acknowledge support from the "Digital Forest" project, Ministry of Lower-Saxony for Science and Culture (MWK) via the program Niedersächsisches Vorab (ZN3679); MDM acknowledges support from the German Aerospace Center, DLR (ML4Earth). We thank the European Union for funding XAIDA via Horizon 2020 grant no. 101003469.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

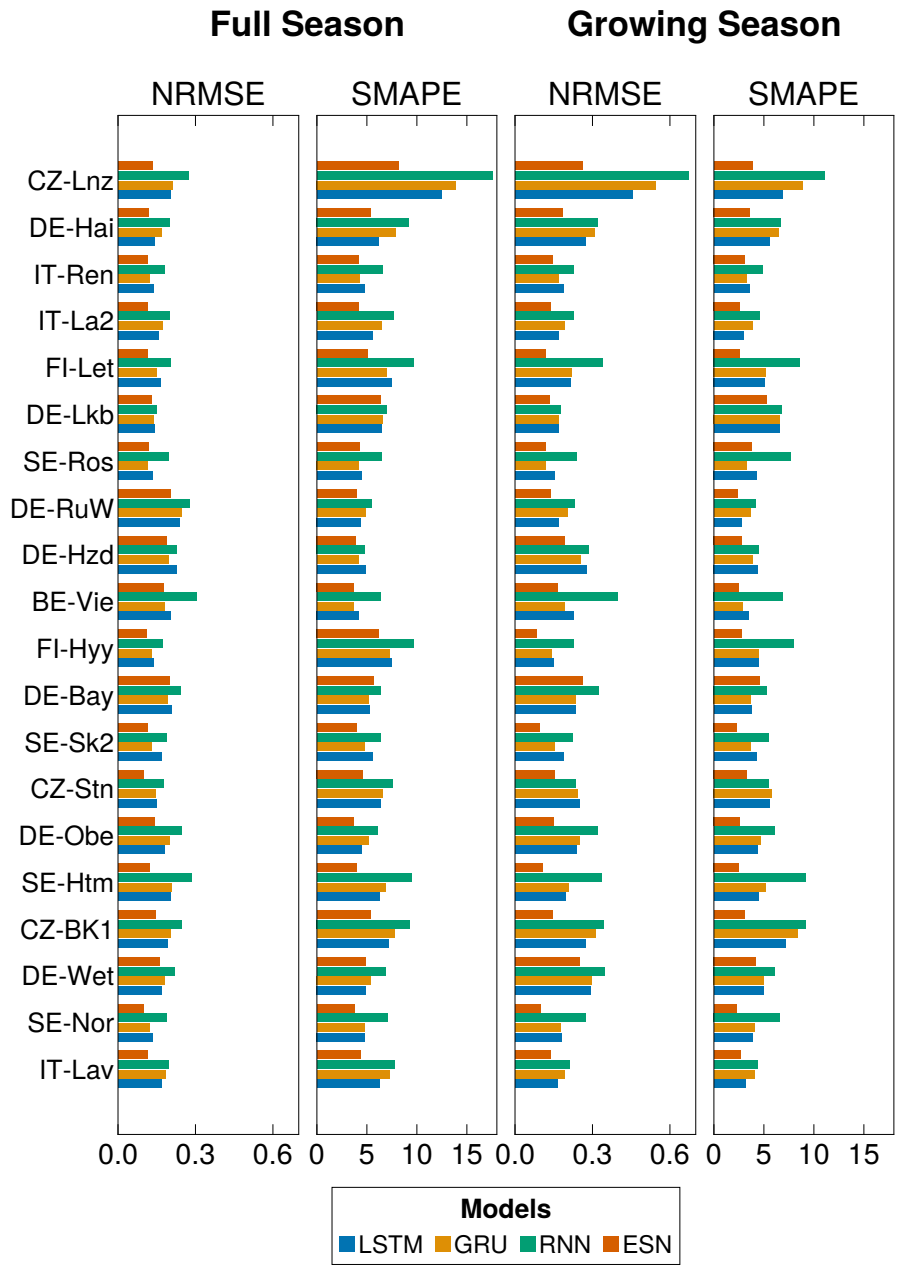

Figure A1. Mean results across locations. We show the metrics for all the analyzed locations and all models. Full season refers to the use of the full dataset for the results. Growing season indicates that we only utilized the months between May and September (included). The figure shows the mean of 50 runs per location.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

References

Aicher, C., Foti, N. J., and Fox, E. B.: Adaptively truncating backpropagation through time to control gradient bias, in: Uncertainty in Artificial Intelligence, pp. 799–808, PMLR, 2020.

Aubinet, M., Vesala, T., and Papale, D.: Eddy covariance: a practical guide to measurement and data analysis, Springer Science & Business Media, 2012.

Bandt, C. and Pompe, B.: Permutation entropy: a natural complexity measure for time series, Physical review letters, 88, 174 102, 2002.

Barnes, L. R., Schultz, D. M., Gruntfest, E. C., Hayden, M. H., and Benight, C. C.: Corrigendum: False alarm rate or false alarm ratio?, Weather and Forecasting, 24, 1452–1454, 2009.

Bengio, Y., Simard, P., and Frasconi, P.: Learning long-term dependencies with gradient descent is difficult, IEEE transactions on neural networks, 5, 157–166, 1994.

Benson, V., Requena-Mesa, C., Robin, C., Alonso, L., Cortés, J., Gao, Z., Linscheid, N., Weynants, M., and Reichstein, M.: Forecasting localized weather impacts on vegetation as seen from space with meteo-guided video prediction, arXiv preprint arXiv:2303.16198, 2023.

Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. W., Dutrieux, L. P., Gans, F., et al.: Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests, PloS one, 14, e0211 510, 2019.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B.: Julia: A fresh approach to numerical computing, SIAM review, 59, 65–98, 2017.

Bonavita, M., Schneider, R., Arcucci, R., Chantry, M., Chrust, M., Geer, A., Le Saux, B., and Vitolo, C.: 2022 ECMWF-ESA workshop report: current status, progress and opportunities in machine learning for Earth System observation and prediction, npj Climate and Atmospheric Science, 6, 87, 2023.

Bottou, L.: Stochastic gradient descent tricks, in: Neural Networks: Tricks of the Trade: Second Edition, pp. 421–436, Springer, 2012.

Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe, https://doi.org/10.5281/ZENODO.3939050, 2020.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., et al.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, eabc7447, 2021a.

Camps-Valls, G., Tuia, D., Zhu, X. X., and Reichstein, M.: Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences, John Wiley & Sons, 2021b.

Canadell, J., Monteiro, P., Costa, M., Cotrim da Cunha, L., Cox, P., Eliseev, A., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global Carbon and other Biogeochemical Cycles and Feedbacks, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., p. 673–816, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.007, 2021.

Cerqueira, V., Torgo, L., and Mozetič, I.: Evaluating time series forecasting models: An empirical study on performance estimation methods, Machine Learning, 109, 1997–2028, 2020.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Chattopadhyay, A., Hassanzadeh, P., and Subramanian, D.: Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: Reservoir computing, artificial neural network, and long short-term memory network, Nonlinear Processes in Geophysics, 27, 373–389, 2020.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L.: A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter, Remote sensing of Environment, 91, 332–344, 2004.

Chen, Z., Liu, H., Xu, C., Wu, X., Liang, B., Cao, J., and Chen, D.: Modeling vegetation greenness and its climate sensitivity with deep-learning technology, Ecology and Evolution, 11, 7335–7345, 2021.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078, 2014.

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

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, accessed: 01/12/2022, 2018.

Danisch, S. and Krumbiegel, J.: Makie.jl: Flexible high-performance data visualization for Julia, Journal of Open Source Software, 6, 3349, https://doi.org/10.21105/joss.03349, 2021.

Datseris, G.: DynamicalSystems.jl: A Julia software library for chaos and nonlinear dynamics, Journal of Open Source Software, 3, 598, https://doi.org/10.21105/joss.00598, 2018.

De Jong, R., de Bruin, S., de Wit, A., Schaepman, M. E., and Dent, D. L.: Analysis of monotonic greening and browning trends from global NDVI time-series, Remote Sensing of Environment, 115, 692–702, 2011.

De Jong, R., Verbesselt, J., Schaepman, M. E., and De Bruin, S.: Trend changes in global greening and browning: contribution of short-term trends to longer-term change, Global Change Biology, 18, 642–655, 2012.

De Keersmaecker, W., van Rooijen, N., Lhermitte, S., Tits, L., Schaminée, J., Coppin, P., Honnay, O., and Somers, B.: Species-rich semi-natural grasslands have a higher resistance but a lower resilience than intensively managed agricultural grasslands in response to climate anomalies, Journal of Applied Ecology, 53, 430–439, https://doi.org/10.1111/1365-2664.12595, 2016.

Diaconu, C.-A., Saha, S., Günnemann, S., and Zhu, X. X.: Understanding the Role of Weather Data for Earth Surface Forecasting using a ConvLSTM-based Model, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1362–1371, 2022.

Dijkstra, H. A.: Nonlinear climate dynamics, Cambridge University Press, 2013.

Dobbertin, M., Wermelinger, B., Bigler, C., Bürgi, M., Carron, M., Forster, B., Gimmi, U., Rigling, A., et al.: Linking increasing drought stress to Scots pine mortality and bark beetle infestations, The scientific world journal, 7, 231–239, 2007.

Eggleton, T.: A short introduction to climate change, Cambridge University Press, 2012.

Elman, J. L.: Finding structure in time, Cognitive science, 14, 179–211, 1990.

Farazmand, M. and Sapsis, T. P.: Extreme events: Mechanisms and prediction, Applied Mechanics Reviews, 71, 050 801, 2019.

Fensham, R. and Holman, J.: Temporal and spatial patterns in drought-related tree dieback in Australian savanna, Journal of Applied Ecology, pp. 1035–1050, 1999.

Foley, J. A., Levis, S., Prentice, I. C., Pollard, D., and Thompson, S. L.: Coupling dynamic models of climate and vegetation, Global change biology, 4, 561–579, 1998.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S.: Deep learning rainfall–runoff predictions of extreme events, Hydrology and Earth System Sciences, 26, 3377–3392, 2022.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al.: Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison, Journal of climate, 19, 3337–3353, 2006.

Funahashi, K.-i. and Nakamura, Y.: Approximation of dynamical systems by continuous time recurrent neural networks, Neural networks, 6, 801–806, 1993.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., and Hochreiter, S.: Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network, Hydrology and Earth System Sciences, 25, 2045–2062, 2021.

Gauthier, D. J., Fischer, I., and Röhm, A.: Learning unseen coexisting attractors, Chaos: An Interdisciplinary Journal of Nonlinear Science, 32, 113 107, https://aip.scitation.org/doi/full/10.1063/5.0116784, 2022.

Gers, F. and Schmidhuber, J.: Recurrent nets that time and count, in: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, IEEE, https://doi.org/10.1109/ijcnn.2000.861302, 2000.

Ghazoul, J., Burivalova, Z., Garcia-Ulloa, J., and King, L. A.: Conceptualizing forest degradation, Trends in ecology & evolution, 30, 622–632, 2015.

Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, edited by Teh, Y. W. and Titterington, M., vol. 9 of *Proceedings of Machine Learning Research*, pp. 249–256, PMLR, Chia Laguna Resort, Sardinia, Italy, https://proceedings.mlr.press/v9/glorot10a.html, 2010.

Grant, P. J.: Drought effect on high-altitude forests, Ruahine range, North Island, New Zealand, New Zealand journal of botany, 22, 15–27, 1984.

Grosse, I., Bernaola-Galván, P., Carpena, P., Román-Roldán, R., Oliver, J., and Stanley, H. E.: Analysis of symbolic sequences using the Jensen-Shannon divergence, Physical Review E, 65, 041 905, 2002.

Haaga, K. A. and Datseris, G.: JuliaDynamics/ComplexityMeasures.jl: v2.7.2, https://doi.org/10.5281/zenodo.7862020, 2023.

Hart, A., Hook, J., and Dawes, J.: Embedding and approximation theorems for echo state networks, Neural Networks, 128, 234–247, 2020.

Hermann, R. and Krener, A.: Nonlinear controllability and observability, IEEE Transactions on Automatic Control, 22, 728–740, https://doi.org/10.1109/TAC.1977.1101601, 1977.

Hilker, T., Lyapustin, A. I., Tucker, C. J., Hall, F. G., Myneni, R. B., Wang, Y., Bi, J., Mendes de Moura, Y., and Sellers, P. J.: Vegetation dynamics and rainfall sensitivity of the Amazon, Proceedings of the National Academy of Sciences, 111, 16 041–16 046, 2014.

Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 6, 107–116, 1998.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Hogan, R. J. and Mason, I. B.: Deterministic Forecasts of Binary Events, https://doi.org/10.1002/9781119960003.ch3, 2011.

Hyndman, R. J. and Koehler, A. B.: Another look at measures of forecast accuracy, International Journal of Forecasting, 22, 679–688, https://doi.org/10.1016/j.ijforecast.2006.03.001, 2006.

Jaeger, H.: The "echo state" approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148, 13, 2001.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Johnstone, J. F., Allen, C. D., Franklin, J. F., Frelich, L. E., Harvey, B. J., Higuera, P. E., Mack, M. C., Meentemeyer, R. K., Metz, M. R., Perry, G. L., et al.: Changing disturbance regimes, ecological memory, and forest resilience, Frontiers in Ecology and the Environment, 14, 369–378, 2016.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020.

Kang, L., Di, L., Deng, M., Yu, E., and Xu, Y.: Forecasting vegetation index based on vegetation-meteorological factor interactions with artificial neural network, in: 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 1–6, IEEE, 2016.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kladny, K.-R. W., Milanta, M., Mraz, O., Hufkens, K., and Stocker, B. D.: Deep learning for satellite image forecasting of vegetation greenness, bioRxiv, pp. 2022–08, 2022.

Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., and Reichstein, M.: Identifying dynamic memory effects on vegetation state using recurrent neural networks, Frontiers in big Data, 2, 31, https://www.frontiersin.org/articles/10.3389/fdata.2019.00031/full, 2019.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using long short-term memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, 2018.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, 2005.

Lamberti, P. W., Martin, M., Plastino, A., and Rosso, O.: Intensive entropic non-triviality measure, Physica A: Statistical Mechanics and its Applications, 334, 119–131, 2004.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., et al.: Trends in the sources and sinks of carbon dioxide, Nature geoscience, 2, 831–836, 2009.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., et al.: Global carbon budget 2018, Earth System Science Data, 10, 2141–2194, 2018.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, nature, 521, 436–444, 2015.

Lellep, M., Prexl, J., Linkmann, M., and Eckhardt, B.: Using machine learning to predict extreme events in the Hénon map, Chaos: An Interdisciplinary Journal of Nonlinear Science, 30, 013 113, 2020.

Liang, E., Shao, X., Kong, Z., and Lin, J.: The extreme drought in the 1920s and its effect on tree growth deduced from tree ring analysis: a case study in North China, Annals of Forest Science, 60, 145–152, 2003.

Linscheid, N., Estupinan-Suarez, L. M., Brenning, A., Carvalhais, N., Cremer, F., Gans, F., Rammig, A., Reichstein, M., Sierra, C. A., and Mahecha, M. D.: Towards a global understanding of vegetation–climate dynamics at multiple timescales, Biogeosciences, 17, 945–962, 2020.

Liu, G., Liu, H., and Yin, Y.: Global patterns of NDVI-indicated vegetation extremes and their sensitivity to climate extremes, Environmental Research Letters, 8, 025 009, 2013.

Lopez-Ruiz, R., Mancini, H. L., and Calbet, X.: A statistical measure of complexity, Physics letters A, 209, 321–326, 1995.

Lotsch, A., Friedl, M. A., Anderson, B. T., and Tucker, C. J.: Response of terrestrial ecosystems to recent Northern Hemispheric drought, Geophysical Research Letters, 32, 2005.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brockett, R., and Ott, E.: Reservoir observers: Model-free inference of unmeasured variables in chaotic systems, Chaos: An Interdisciplinary Journal of Nonlinear Science, 27, 041 102, 2017.

Maass, W., Natschläger, T., and Markram, H.: Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural computation, 14, 2531–2560, 2002.

Mahecha, M. D., Fürst, L. M., Gobron, N., and Lange, H.: Identifying multiple spatiotemporal patterns: A refined view on terrestrial photosynthetic activity, Pattern Recognition Letters, 31, 2309–2317, 2010.

Mahecha, M. D., Gans, F., Brandt, G., Christiansen, R., Cornell, S. E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., et al.: Earth system data cubes unravel global multivariate dynamics, Earth System Dynamics, 11, 201–234, 2020.

Mahecha, M. D., Bastos, A., Bohn, F. J., Eisenhauer, N., Feilhauer, H., Hartmann, H., Hickler, T., Kalesse-Los, H., Migliavacca, M., Otto, F. E., et al.: Biodiversity loss and climate extremes—study the feedbacks, Nature, 612, 30–32, 2022.

Makridakis, S.: Accuracy measures: theoretical and practical concerns, International journal of forecasting, 9, 527–529, 1993.

Marcolongo, A., Vladymyrov, M., Lienert, S., Peleg, N., Haug, S., and Zscheischler, J.: Predicting years with extremely low gross primary production from daily weather data using Convolutional Neural Networks, Environmental Data Science, 1, e2, 2022.

Martin, M., Plastino, A., and Rosso, O.: Generalized statistical complexity measures: Geometrical and analytical properties, Physica A: Statistical Mechanics and its Applications, 369, 439–462, 2006.

Martinuzzi, F., Rackauckas, C., Abdelrehim, A., Mahecha, M. D., and Mora, K.: ReservoirComputing. jl: An Efficient and Modular Library for Reservoir Computing Models, Journal of Machine Learning Research, 23, 1–8, 2022.

Meiyazhagan, J., Sudharsan, S., and Senthilvelan, M.: Model-free prediction of emergence of extreme events in a parametrically driven nonlinear dynamical system by deep learning, The European Physical Journal B, 94, 156, 2021.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., De Leeuw, G., et al.: Uncertainty information in climate data records from Earth observation, Earth System Science Data, 9, 511–527, 2017.

Montero, D., Aybar, C., Mahecha, M. D., Martinuzzi, F., Söchting, M., and Wieneke, S.: A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research, Scientific Data, 10, 197, 2023.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999, science, 300, 1560–1563, 2003.

Ogata, K. et al.: Modern control engineering, vol. 5, Prentice hall Upper Saddle River, NJ, 2010.

Pammi, V. A., Clerc, M. G., Coulibaly, S., and Barbay, S.: Extreme Events Prediction from Nonlocal Partial Information in a Spatiotemporally Chaotic Microcavity Laser, Phys. Rev. Lett., 130, 223 801, https://doi.org/10.1103/PhysRevLett.130.223801, 2023.

Papagiannopoulou, C., Miralles, D. G., Decubber, S., Demuzere, M., Verhoest, N. E., Dorigo, W. A., and Waegeman, W.: A non-linear Granger-causality framework to investigate climate–vegetation dynamics, Geoscientific Model Development, 10, 1945–1960, 2017.

Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Global Change Biology, 9, 525–535, 2003.

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M. D., Margolis, H., et al.: Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks, Journal of Geophysical Research: Biogeosciences, 120, 1941–1957, 2015.

Pappas, C., Mahecha, M. D., Frank, D. C., Babst, F., and Koutsoyiannis, D.: Ecosystem functioning is enveloped by hydrometeorological variability, Nature ecology & evolution, 1, 1263–1270, 2017.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Pascanu, R., Mikolov, T., and Bengio, Y.: On the difficulty of training recurrent neural networks, in: International conference on machine learning, pp. 1310–1318, Pmlr, 2013.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems, 32, 2019.

Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., and Ott, E.: Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data, Chaos: An Interdisciplinary Journal of Nonlinear Science, 27, 121 102, https://aip.scitation.org/doi/full/10.1063/1.5010300, 2017.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, Physical review letters, 120, 024 102, https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.024102, 2018.

Peng, Q., Li, X., Shen, R., He, B., Chen, X., Peng, Y., and Yuan, W.: How well can we predict vegetation growth through the coming growing season?, Science of Remote Sensing, 5, 100 043, 2022.

Pyragas, V. and Pyragas, K.: Using reservoir computer to predict and prevent extreme events, Physics Letters A, 384, 126 591, 2020.

Ray, A., Chakraborty, T., and Ghosh, D.: Optimized ensemble deep learning framework for scalable forecasting of dynamics containing extreme events, Chaos: An Interdisciplinary Journal of Nonlinear Science, 31, 111 105, 2021.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., et al.: Climate extremes and the carbon cycle, Nature, 500, 287–295, 2013.

Reichstein, M., Besnard, S., Carvalhais, N., Gans, F., Jung, M., Kraft, B., and Mahecha, M.: Modelling landsurface time-series with recurrent neural nets, in: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 7640–7643, IEEE, 2018.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, 2019.

Requena-Mesa, C., Benson, V., Reichstein, M., Runge, J., and Denzler, J.: EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task., in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1132–1142, 2021.

Robin, C., Requena-Mesa, C., Benson, V., Alonso, L., Poehls, J., Carvalhais, N., and Reichstein, M.: Learning to forecast vegetation greenness at fine resolution over Africa with ConvLSTMs, arXiv preprint arXiv:2210.13648, 2022.

Rosso, O. A. and Masoller, C.: Detecting and quantifying temporal correlations in stochastic resonance via information theory measures, The European Physical Journal B, 69, 37–43, 2009.

Rosso, O. A., Larrondo, H., Martin, M. T., Plastino, A., and Fuentes, M. A.: Distinguishing noise from chaos, Physical review letters, 99, 154 102, 2007.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al.: Monitoring vegetation systems in the Great Plains with ERTS, NASA Spec. Publ, 351, 309, 1974.

Rudy, S. H. and Sapsis, T. P.: Output-weighted and relative entropy loss functions for deep learning precursors of extreme events, Physica D: Nonlinear Phenomena, 443, 133 570, 2023.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, nature, 323, 533–536, 1986.

Savitzky, A. and Golay, M. J.: Smoothing and differentiation of data by simplified least squares procedures., Analytical chemistry, 36, 1627–1639, 1964.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Scheepens, D. R., Schicker, I., Hlaváčková-Schindler, K., and Plant, C.: Adapting a deep convolutional RNN model with imbalanced regression loss for improved spatio-temporal forecasting of extreme wind speed events in the short to medium range, Geoscientific Model Development, 16, 251–270, https://doi.org/10.5194/gmd-16-251-2023, 2023.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B.: Catastrophic shifts in ecosystems, Nature, 413, 591–596, 2001.

Seddon, A. W., Macias-Fauria, M., Long, P. R., Benz, D., and Willis, K. J.: Sensitivity of global terrestrial ecosystems to climate variability, Nature, 531, 229–232, 2016.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., p. 1513–1766, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.013, 2021.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, Advances in neural information processing systems, 28, 2015.

Sippel, S., Lange, H., Mahecha, M. D., Hauhs, M., Bodesheim, P., Kaminski, T., Gans, F., and Rosso, O. A.: Diagnosing the dynamics of observed and simulated ecosystem gross primary productivity with time causal information theory quantifiers, PloS one, 11, e0164 960, 2016.

Sippel, S., Reichstein, M., Ma, X., Mahecha, M. D., Lange, H., Flach, M., and Frank, D.: Drought, heat, and the carbon cycle: a review, Current Climate Change Reports, 4, 266–286, 2018.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., et al.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global change biology, 9, 161–185, 2003.

Slayback, D. A., Pinzon, J. E., Los, S. O., and Tucker, C. J.: Northern hemisphere photosynthetic trends 1982–99, Global Change Biology, 9, 1–15, 2003.

Srinivasan, P. A., Guastoni, L., Azizpour, H., Schlatter, P., and Vinuesa, R.: Predictions of turbulent shear flows using deep neural networks, Physical Review Fluids, 4, 054 603, 2019.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15, 1929–1958, 2014.

Steinier, J., Termonia, Y., and Deltour, J.: Smoothing and differentiation of data by simplified least square procedure, Analytical chemistry, 44, 1906–1909, 1972.

Sutskever, I.: Training recurrent neural networks, University of Toronto Toronto, ON, Canada, 2013.

Teskey, R., Wertin, T., Bauweraerts, I., Ameye, M., McGuire, M. A., and Steppe, K.: Responses of tree species to heat waves and extreme heat events, Plant, cell & environment, 38, 1699–1712, 2015.

Tietz, M., Fan, T. J., Nouri, D., Bossan, B., and skorch Developers: skorch: A scikit-learn compatible neural network library that wraps PyTorch, https://skorch.readthedocs.io/en/stable/, 2017.

Tuia, D., Schindler, K., Demir, B., Camps-Valls, G., Zhu, X. X., Kochupillai, M., Džeroski, S., van Rijn, J. N., Hoos, H. H., Del Frate, F., et al.: Artificial intelligence to advance Earth observation: a perspective, arXiv preprint arXiv:2305.08413, 2023.

https://doi.org/10.5194/egusphere-2023-2368
Preprint. Discussion started: 17 October 2023

Van Mantgem, P. J., Stephenson, N. L., Byrne, J. C., Daniels, L. D., Franklin, J. F., Fulé, P. Z., Harmon, M. E., Larson, A. J., Smith, J. M., Taylor, A. H., et al.: Widespread increase of tree mortality rates in the western United States, Science, 323, 521–524, 2009.

Verstraeten, D., Schrauwen, B., d'Haene, M., and Stroobandt, D.: An experimental unification of reservoir computing methods, Neural networks, 20, 391–403, 2007.

Vlachas, P., Pathak, J., Hunt, B., Sapsis, T., Girvan, M., Ott, E., and Koumoutsakos, P.: Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics, Neural Networks, 126, 191–217, https://doi.org/https://doi.org/10.1016/j.neunet.2020.02.016, 2020.

von Buttlar, J., Zscheischler, J., Rammig, A., Sippel, S., Reichstein, M., Knohl, A., Jung, M., Menzer, O., Arain, M. A., Buchmann, N., et al.: Impacts of droughts and extreme-temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones, Biogeosciences, 15, 1293–1318, 2018.

Walleshauser, B. and Bollt, E.: Predicting sea surface temperatures with coupled reservoir computers, Nonlinear Processes in Geophysics, 29, 255–264, 2022.

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., Carvalhais, N., Ermida, S. L., Brümmer, C., Schrader, F., et al.: A view from space on global flux towers by MODIS and Landsat: the FluxnetEO data set, Biogeosciences, 19, 2805–2840, 2022.

Watson-Parris, D.: Machine learning for weather and climate are worlds apart, Philosophical Transactions of the Royal Society A, 379, 20200 098, 2021.

Werbos, P. J.: Generalization of backpropagation with application to a recurrent gas market model, Neural networks, 1, 339–356, 1988.

Werbos, P. J.: Backpropagation through time: what it does and how to do it, Proceedings of the IEEE, 78, 1550–1560, 1990.

Williams, R. J. and Zipser, D.: Gradient-based learning algorithms for recurrent, Backpropagation: Theory, architectures, and applications, 433, 17, 1995.

Yengoh, G. T., Dent, D., Olsson, L., Tengberg, A. E., and Tucker III, C. J.: Use of the Normalized Difference Vegetation Index (NDVI) to assess Land degradation at multiple scales: current status, future trends, and practical considerations, Springer, 2015.

Zeng, N., Hales, K., and Neelin, J. D.: Nonlinear dynamics in a coupled vegetation–atmosphere system and implications for desert–forest gradient, Journal of Climate, 15, 3474–3487, 2002.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J. M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., et al.: Optical vegetation indices for monitoring terrestrial ecosystems globally, Nature Reviews Earth & Environment, 3, 477–493, 2022.

Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X.: Prediction of sea surface temperature using long short-term memory, IEEE geoscience and remote sensing letters, 14, 1745–1749, 2017.

Zhang, Z., Xin, Q., and Li, W.: Machine Learning-Based Modeling of Vegetation Leaf Area Index and Gross Primary Productivity Across North America and Comparison With a Process-Based Model, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 802, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002802, 2021.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources, IEEE geoscience and remote sensing magazine, 5, 8–36, 2017.