Effects of Temporal Resolution on Modelling a Vegetation Index with Recurrent Neural Networks

Bachelor Thesis in Computer Science

submitted by

Johannes Schubert
born December 11, 1996 in Eberswalde

written at

Computer Vision Group
Department of Mathematics and Computer Science
Friedrich-Schiller-Universität Jena

in cooperation with

Max-Planck-Institut for Biogeochemistry
07743 Jena
Germany

Supervisor:	Prof. Dr.-Ing. Joachim Denzler
Advisor:	Dr. Martin Jung
Started:	March 12, 2020
Finished:	October 5, 2020

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe. Alle Ausfüüührungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet. Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt und von dieser als Teil einer Prüfungsleistung angenommen.

Die Richtlinien des Lehrstuhls für Examensarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Jena, den 05. Oktober 2020   Johannes Schubert

Überblick

Die Modellierung von Interaktionen zwischen Klima und Vegetation ist in den Biogeowissenschaften ein zentrales Forschungsfeld. Aufgrund einer hohen Komplexität in der Interaktion beider Systeme kommen hier klassisch statistische Ansätze an ihre Grenzen. Ein großes Potential wird hingegen in neuen Ansätzen wie den Recurrent Neural Networks (RNNs) gesehen, die zunehmend auch in den Biogeowissenschaften an Bedeutung gewinnen. Noch fehlen allerdings Erfahrungen, was genau für eine gute Modellierung mit RNNs wichtig wäre. Die vorliegende Arbeit untersucht daher den Aspekt, inwiefern der Einsatz von höher aufgelösten Daten zu einer verbesserten Modellgenerierung führt. Dazu wurden zwei Vegetationsmodelle mithilfe von RNNs erzeugt, die die Wechselwirkung zwischen Klima und Vegetation mithilfe von unterschiedlich zeitlich aufgelösten Klimadaten modellieren. Im Anschluss daran wurde die Leistungsfähigkeit der Modelle miteinander verglichen. Es lässt sich feststellen, dass die Vorhersageleistung des Vegetationsmodells mit höherer zeitlicher Auflösung verbessert werden konnte. Dabei profitierte insbesondere die Vorhersage von Anomalien in Vegetationsdaten. Somit kann davon ausgegangen werden, dass der Einsatz von höher aufgelösten Daten, die Vorhersagefähigkeit von Modellen ökologischer-klimatologischer Beziehungen verbessert.

Abstract

The modeling of interactions between climate and vegetation is a central to research in the field of biogeosciences. Due to a high complexity in the interaction of both systems, classical statistical approaches reach their limits. A great potential, however, is seen in new approaches such as Recurrent Neural Networks (RNNs), which are increasingly gaining importance in the biogeosciences. However, there is still a lack of experience, which would be important for good modeling with RNNs. Therefore, the present work investigates the aspect, to what extent the use of higher-resolution data leads to an improved model generation. For this purpose, two vegetation models were generated using RNNs, which model the interaction between climate and vegetation using climate data with different temporal resolutions. Subsequently, the performance of the models was compared. It was seen that the prediction performance of the vegetation model with higher temporal resolution could be improved. Especially the prediction of anomalies in vegetation data benefited. Thus, it can be assumed that the use of higher-resolution data improves the predictive power of models of ecological-climatological relationships.

Contents

Contents

Acronyms

ANN	Artificial Neural Network
ASHA	Asynchronous Successive Halving Algorithm
LSTM	Long Short-Term Memory model
LSTM-2	LSTM with two layers
LSTM-Bias	LSTM with one layer and its Forget Gate bias initialized to 1
LSTM-Norm	LSTM with one layer and Layer Normalization applied to it
$M_{15\text{-}D}$	Vegetation model generated with $RNN_{15\text{-}D}$
M_B	Baseline model generated with $RNN_{15\text{-}D}$
M_D	Vegetation model generated with RNN_D
Mem	Performance metric to measure the modelling capability of memory effects
MSC	Median seasonal cycle
NDVI	Normalized Difference Vegetation Index
$NDVI_{ANO}$	Anomalies from the MSC of the NDVI
$NDVI_{MSC}$	MSC of the NDVI
$NDVI_{RAW}$	Raw time series of the NDVI
NSE	Nash–Sutcliffe model efficiency coefficient
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
$RNN_{15\text{-}D}$	RNN trained with 15-daily resolved datasets
RNN_D	RNN trained with daily resolved datasets
VI	Vegetation Index

Chapter 1

Introduction

Vegetation dynamics are largely determined by temperature, water and light. They are therefore particularly influenced by climatic conditions. But the vegetation itself also influences the climate system. For example, the composition of the atmosphere is changed by the transfer of water vapor and the exchange of carbon dioxide [54, 72]. These complex interactions between climate and vegetation represent a major challenge for predictions of future climate. They are of high relevance especially against the background of climate change. A better understanding of both components would therefore also enable improved forecasts. Especially the vegetation reaction to past climate variability (so-called memory effects) seem to be of central importance [52, 7].

1.1 Related Work

These memory effects have already been investigated using linear regression models to predict the vegetation state [83, 72, 31]. These studies could contribute to a better understanding of the interaction between climate and biosphere. However, there are indications that linear models may not be sufficient to comprehensively represent vegetation dynamics [54]. Therefore, the investigation of vegetation dynamics requires new approaches to model even these highly nonlinear processes. The success of Recurrent Neural Networks (RNNs) in other disciplines dealing with sequential data, such as Natural Language Processing, suggests that this method may also be relevant for modeling in the biogeoscientific context [65, 86]. First attempts to model memory effects using RNNs have already been made [7, 34]. From the results, it can be concluded that this method can indeed be used to better capture memory effects. However, it is still an open question which key determinants improve the prediction of a model.

1.2 Motivation

It is known from other areas that the success of modeling using deep learning methods often depends on the availability of a large amount of training data. This can be observed, for example, in the ImageNet classification challenge [1]. Here, over one million labelled images are available for training models to achieve or even exceed human performance [25, 36]. A successful transfer to the modeling of ecological and climatological interrelationships could thus be accompanied by a higher resolution of the available input data.

In this thesis such a possibility is examined by comparing the performance of two vegetation state models. For this purpose, two models are generated using RNNs, which were trained with a higher and a lower temporal resolution, respectively. Subsequently, the hypothesis is tested whether the performance of the model actually increases when modeling is performed using higher-resolution features.

1.3 Task Description

Based on Kraft et al. [34], two vegetation models were generated using RNNs. These contained climate data in daily (M_D) and 15-daily (M_{15-D}) resolution. The predictive power of both models was then evaluated and compared in a multi-level evaluation phase. Finally, possible implications and deficits in the process were derived and discussed.

---

[1] http://image-net.org/challenges/LSVRC/

1.4 Overview

The aim of the thesis lies between the disciplines of ecological climatology and machine learning. Due to the interdisciplinary character, a basic introduction to central concepts of these disciplines is given first (chapter 2). In subsection 2.1 the terms vegetation and memory effect are explained and central abiotic factors for vegetation dynamics are outlined. Subsection 2.2 presents concepts that are important for understanding the training process of the RNNs. In chapter 3 the pipline for generating the vegetation models is outlined. Furthermore the metrics that were used for the performance evaluation of the models are introduced. Subsequently, the results of the three-step evaluation procedure are presented and compared (chapter 4). Finally, a concise discussion and classification of the results follows (chapter 5). At the same time, a view on how a better modelling of climate-biosphere interactions could be achieved in the future is given.

Chapter 2

Theoretical Background

2.1 Ecological Climatology

Earth science is a broad term that encompasses all natural sciences that study the earth. More specifically the focus is put onto the interactions of the main components of the earth system: the hydrosphere containing all water on earth, the biosphere the sum of all ecosystems, the atmosphere the layer of gases surrounding the earth and the lithosphere, the outermost shell of the planet. Each of these spheres can be further broken down into more specialized fields [8, pp. 40–50].

The global climate is determined by the interaction of these spheres, the cycles that connect the spheres and the human systems that influence the cycles [8, p. 7]. This subdivision of earth science is researched by the discipline of **climatology**. It investigates the average state of the atmosphere at a given place over time. Therefore the findings are based on long-term observations and modelling of climate and geographical factors [68, p. 3].

Apart from climatology, **ecology** is also of particular importance for this work. It is a branch of biology that is concerned with the study of interactions of organisms among themselves and with their environment.

The union of these two disciplines constitutes the interdisciplinary framework of **ecological climatology** studying the functioning terrestrial ecosystems through their cycling of energy, water, chemical elements and trace gases. Climatology and ecology are seen as components of one another, meaning that atmospheric processes are not only influenced by terrestrial ecosystems but ecosystems influence atmospheric processes as well. This correlation between climate and vegetation was already discovered by Humboldt while he observed that widely separate regions have structurally and functionally similar vegetations if their climates are similar [8, pp. 2–3].

2.1.1 Vegetation

The term vegetation describes the assemblages of plant species forming a continuous and conspicuous plant cover over the landscape with the exception of dry or cold deserts [9, p. 1]. That means, the total of approximately 390 000 plant species with each of them having local populations of individual plants are the ultimate basis of vegetation [13]. In order to study vegetation, it is therefore necessary to investigate the causes of patterns of plant distribution and the relationship of plants to their environment, which is done by the discipline of plant ecology [9, p. 23].

The main **vegetation properties** include [9, pp. 6–7]:

- **species composition**: which plant species build up the local populations

- **structure**: structural patterns which arise from the spatial disposition of different plant species

- physiognomy - the general appearance of vegetation depending on relative abundance of species

- **spatial patterns**: depending on the habitat conditions

- **temporal patterns**: mainly depending on seasonal cycles

The latter two properties are the main one this bachelor thesis will be examining. There are two main **classes of vegetation** - natural and semi-natural, with the distinction being made on the basis on the amount of human influence meaning that semi-natural vegetation is being lightly managed by humans [9, p. 23]. However vegetation can be further classified according to its physiognomic properties, the form and stature of the most conspicuous plants.

The notion that **vegetation change** occurs in regular patterns has been expressed already in 1916 by Clements [14]. For example vegetation develops mostly in a sequential way on bare surfaces. In addition, as part of the normal growth cycle, regular phenological changes in individual plants occur simultaneously in a plant population. This usually occurs in correlation with the seasons. These changes in vegetation are clearly reflected in the time series of the Vegetation Index (VI).

Vegetation index

A VI is a dimensionless radiometric measure that is used as a surrogate measure of the relative intensity and condition of vegetation. There are many vegetation

indices that are functionally equivalent in their information content, with many of these indices using the inverse relationship between red and near-infrared reflectance that is associated with healthy green vegetation. - VIs therefore only reflect the greenness of vegetation [29, p. 325].

A widely used VI is the Normalized Difference Vegetation Index (NDVI) [69]. It is calculated from multispectral remote sensing using the wavelength and the intensity of light reflected from the land surface in the visible and near-infrared spectrum:

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}},$$

where ρ_{nir} is the near-infrared radiant flux and ρ_{red} is the red reflected radiant flux. The formula is derived from the relationship that healthy vegetation reflects a significant amount of light in the near infrared range (and green spectrum) and only a minimal amount in the red (and blue) spectrum. This relationship is due to the fact that in healthy plants, the plant pigment chlorophyll absorbs a significant amount of light from the blue and red spectrum for photosynthesis, whereas the cell structure of the leaves reflects high proportions of light in the near infrared range. In a plant that undergoes senescence in the fall or is exposed to stress, the chlorophyll pigment may disappear and the lack of energy causes the leaf structure to change, reflecting wavelength differently [29, p. 319]. The NDVI produces values ranging from $(-1, +1)$. A high NDVI from 0.6 to 0.9 corresponds to dense, productive vegetation as found in temperate and tropical forest areas with plenty of photosynthetically active plants. Areas with low vegetation such as shrubs and meadows or trees with withering leaves can result in moderate values from 0.2 to 0.5. Areas without vegetation such as barren rock, sand or snow have a very low NDVI of 0.1 or less and clear water scales in the negative value range [82]. When calculating the NDVI, data should be pre-processed to reduce noise effects of clouds, solar altitude and other atmospheric influences. But even after preprocessing, the index is still sensitive to factors such as soil background, especially in areas where vegetation is sparse [29, p. 325].

2.1.2 Main Abiotic Factors Affecting Plant Growth

Since plants are fixed in position, they have to adapt to the conditions of their surrounding environment. These conditions can be living (biotic) or nonliving (abiotic) nature. This section focuses on the latter conditions (see table 2.1) and tries to give a brief overview about them. It is noted that due to the interactions between abiotic

factors with each other, the influence of individual factors on plants often cannot be emphasized. Therefore this is only an analytical separation.

Element		Specific property that varies
solar radiation	light	intensity, duration, periodicity, direction, spectral qualities
	temperature	intensity, duration, periodicity
atmosphere	precipitation	form, frequency, duration
	gases	CO_2,O_2, relative humidity
soil		nutrients, water, morphology

Table 2.1: Main abiotic factors that influence the growth of plants [9, pp. 24–26].

Solar Radiation

Solar radiation determines the light and temperature conditions for plants. Both play a crucial role in plant growth, as light is used as the primary energy source in plants whereas temperature determines the rates of biochemical reactions in plants. Light and temperature are interlinked and influence each other's states on a daily and seasonal basis.

Sunlight   Due to plants fixed position they need to adapt to the different light environments which derive from the interaction of the following characteristics [71, pp. 23–24]:

1. **wavelength of the sunlight** (quality): the biological effectiveness of photosynthesis in plants is highest in the wavelength spectrum from 380 to 710 nm. Blue and red light have the highest effect on plant growth because they stimulate photosynthesis most. Furthermore blue light triggers the vegetative leaf growth while a combination of blue and red light triggers the flowering.

2. **intensity of the sunlight** (quantity): plants adapt to different intensities at all levels of organization (from molecular to morphological). For example trees like the eucalyptus attenuates the intensity of incoming sunlight by changing its growth style and letting it leaves hang.Too little light would create a negative energy balance for the plants meaning that the photosynthesis cannot produce enough energy required by the plant. Up to a certain threshold the production

capacity of photosynthesis increases with a higher intensity of sunlight.Too much light, on the other hand, cannot be used by the plant anymore and could create oxidative stress that can cause extensive cellular damage including damaging the DNA [16].

3. **time period plants have been exposed to light**: continuously changing light conditions cause stress in plants and can hinder further growth. But on the other hand in some plants a consecutive amount of time without light is necessary to start flowering [78].

Air Temperature   Air temperature is the temperature of the atmosphere near the ground that is not influenced by solar radiation, ground heat or heat conduction. It is a crucial abiotic factor for plants because the rates of biochemical reactions in organisms double or triple with $10\,°C$ increase [10, p. 15]. More precisely, fundamental processes such as photosynthesis, respiration, transpiration and seed germination are strongly influenced by it in their functioning. Generally speaking the optimal temperature range for plants, where they are in an active state, is in the range of liquid water. However, this optimal range differs for each species' particular needs (see examples in table 2.2) [71, pp. 45–50]. Temperatures beyond (cold stress) or above (heat stress) the plant's optimal range will determine its fall out from an active state into a state of rigidity, where life occurs only minimally.

	Temperature in °C		
Plant species	Minimum	Optimum	Maximum
temperate deciduous trees and shrubs	-3 to -1	15 to 25	40 to 45
herbaceous (sun) plants	-2 to 0	20 to 30	40 to 50
herbaceous (shade) plants	-2 to 0	10 to 20	$\sim$40

Table 2.2: Temperature range for net photosynthesis at light saturation for three exemplary plant species [9, p. 47].

Atmosphere

The atmosphere supplies the vegetation with CO_2 and O_2, which are needed for photosynthesis and respiration. It can be assumed that the necessary concentrations of these gases are present in the atmosphere in sufficient quantities to allow both processes to occur without problems [9, p. 49]. Precipitation has a critical impact

on vegetation growth. Furthermore it is known that anomalies in precipitation can create a lagged vegetational response of one to two months [11].

Soil

Soil is the outermost part of the earth that is modified by biotic and abiotic processes. It is the water and nutrient storage for plants.

Water   Water is a key element of life because generally speaking it is the foundation or main agent for all biochemical reactions. Furthermore, it is the most limiting abiotic factor to plant growth and productivity and determines the distribution of vegetation. In most plants, leaves and fruits contain about 90 percent water in their cells. This water content must be approximately constant because it creates hydrostatic pressure inside the cells, which in turn is necessary to stiffen the plants' tissue. Furthermore the temperature of a plant is regulated by the evaporation of water from leaf tissues during transpiration. Water also creates pressure for roots to move through the soil and it serves as a primary component in photosynthesis and respiration [71, pp. 277 – 280]. Essentially all the water that plants need is absorbed by their roots through a process called capillary action. Inside the plant the water is passively moved through it. This is due to the negative pressure created by the evaporation of water from the leaves [44].

Nutrients   For their normal functioning plants require many mineral nutrients that stay beneath the soil such as salts, which plants consume as ions. They are primarily absorbed by the root system in a dissolved form known as soil-water solution. The nutrients can be divided into the group of macro-elements and trace-elements. Both are essential for the life of a plant. Macro-elements participate directly in the metabolism (N, S, P), support it indirectly (K) or are required for enzyme reactions. Trace elements, which are mostly heavy metals, are also necessary for enzyme reactions in plants but depend on the species [71, pp. 313–314].

2.1.3 Memory Effects

In climatological research temporal phenomena were always of high significance. Although it was understood since the early 1900s that antecedent events were of great influence for current functioning of plants, research was mostly focussed on concurrent abiotic conditions. Therefore there is still a lack of knowledge explaining

temporal linkages between abiotic conditions and biotic responses. Especially how perturbations such as climate extreme events influence biotic responses on different time scales [52]. These temporal linkages between short-term climate anomalies and vegetation response were named memory effects and are defined as influences that past events have on the present or future responses of an ecosystem to environmental conditions. A memory effect is a stabilizing mechanism in plants to integrate past experiences in their response to current situations and therefore relevant when predicting future changes in vegetation [7].

2.2 Machine Learning in Biogeosciences

To model global interactions between climate soil and vegetation, data from experiments and long-term observations of the earth are used. Since the modelling is highly computer based, the interdisciplinary character of biogeosciences gets extended further by computer science.

   The models can be designed following a theory-based mechanistic, a data-based machine learning or a hybrid approach, a combination of the former two approaches. In mechanistic modelling, researchers first formulate a hypothesis in the form of a simplified mathematical description of the underlying causal mechanisms that give rise to the observed phenomenon. Subsequently, experimental observations are made to test this hypothesis. Machine learning methods do not require a hypothesis in advance, but rather suitable pairs of input and output variables describing the phenomenon to be investigated, a large dataset and a specific learning method that learns complex relationships between input and output data. In both approaches, first a training set is taken from the available data and used to construct and calibrate the model. Then the accuracy of the model is determined by predicting a test set of the dataset [3]. Both paradigms differ qualitatively in that mechanistic models provide the causality missing in machine learning approaches, but their simplified assumptions and extremely specific nature prohibit universal predictions that machine learning approaches can provide. The advantages of one, however, are the disadvantages of the other, which is why both approaches play a role and a symbiotic relationship between the two approaches is sought.

2.2.1 Artificial Neural Networks

An artificial neural network is in its broadest sense a sorted triple (N, V, w) that consists of a finite set of neurons (or units) $N = u_1, u_2, ...$, a finite set of weighted and

directed connections between these neurons $V \subseteq N \times N$ and a function $w : V \to \mathbb{R}$ that defines the weights of the connection between neurons. The behavior of the Artificial Neural Network (ANN) is determined by its weights w [70]. The simplest form is a single-layer network, called perceptron. It contains just an input layer and an output node. By adding more neurons in a layered fashion the input and the output layers get separated by a group of so called hidden layer(s). Different ways of connecting these so-called multi-layer networks lead to different network architectures, such as the feed forward network or the recurrent neural network introduced in section 2.2.2.

Fundamentals on Learning Learning in neural networks is realized by changing the weights of the connections between neurons using rules that can be implemented as an algorithm in a programming language. The process of learning is done during training, where input data are presented to an artificial neural network. There are three main paradigms in how a neural network can be trained: (1) unsupervised learning, where only input patterns are provided to the neural net but no further aid is given. The network itself tries to detect similarities between patterns in the data and clusters them into groups, (2) reinforcement learning, where the network receives feedback on whether it behaves well or not and (3) supervised learning, where pairs of input and appropriate desired output patterns are provided [35]. Supervised learning is the most common form of machine learning and has been used in this thesis as well.

Backpropagation of error Prior to starting the training process a dataset has to be defined. It consists of pairs of an input (x, y) and a desired output $\hat{x}$. A loss function that measures the error between the predicted output $\hat{x}$ and the desired output x is defined. The weights are then adjusted with the help of an optimization algorithm called stochastic gradient descent. It computes a gradient vector for each weight. The gradient indicates by what amount the error would increase or decrease if the weight were increased by a tiny amount. The weights are then adjusted into the opposite direction that the model "descends" towards a local minima of the loss function. The algorithm aims to find the minimum of the loss function, where the error between predicted and desired output is low on average. Stochastic gradient descent is applied iteratively until the average of the loss function stops decreasing. It is stochastic, since in each step only a subset of the data is considered, which gives a noisy estimate. This iterative learning procedure consists of showing a few

examples of input-output pairs to the model, calculating the average error and the resulting gradients and adjusting the weights accordingly. The procedure is known as the backpropagation of error [37].

Hyperparameter Hyperparameters are parameters that determine the network structure and how the network is trained. They can be interpreted as controlling the prior distribution of the resulting model, because primary model parameters like weights are optimized automatically after fixing these parameters [1, p. 125]. They are an important factor in determining the performance of a neural network on any problem. The following provides a (non-exhaustive) list of relevant hyperparameters:

- **activation function**: It transforms the neuron's input and its previous activation state into a new activation state. Mostly the activation function introduces nonlinearity to models which allow ANNs to solve nontrivial problems. A commonly used activation function is the rectified linear activation function $f(x) = max\{0, x\}$ from $[0, \inf)$ or the sigmoid activation function $f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$ ranging from $(0, 1)$ [47].

- **number of hidden layers and units**: These parameters specify the amount of layers and units between the input and output layer

- **dropout**: This is a regularization technique to avoid overfitting and therefore increasing the generalization power of the ANN. The term refers to randomly dropping a unit from the ANN during training. The parameter encodes the percentage of units being chosen [75].

- **number of epochs**: This number describes how often the entire training data is being passed through the ANN. The training data is mostly passed multiple times through the network because the iterative process of gradient descent needs time to fit the model correctly. The number of epochs should be increased up to the point where validation accuracy starts decreasing which is a sign for overfitting. A common way to achieve this is by applying an early stopping algorithm. This algorithm monitors a validation metric and stops the training when no improvement is observed.

- **batch size**: A batch is a subsample of the entire dataset that is given to the network after which parameters are updated. The batch size describes the total number of training examples that are present in a single batch.

- **learning rate**: This parameter controls the step size at each iteration of the backpropagation of error while moving toward a minimum of a loss function. There are also extended stochastic gradient descent-based procedures that adapt the learning rate based on the specific unit such as Adam [32].

Feature preprocessing   It is a common observation that relative values of features have different ranges which leads to the problem of ill-conditioning, where the loss function has the tendency to be more sensitive to some parameters than others. This in turn affects the performance of gradient descent. Preprocessing the input data before helps to let the gradient converge much faster and therefore ensures a better performance [56].

Two feature scaling methods that are mostly used together are mean centering and normalization. Mean centering can be helpful to remove certain types of bias effects. The goal of normalization is to bring the values of the dataset to a common scale without distorting the variation in the value ranges. One common type of normalization is to divide each feature value by its standard deviation (sd). Applying both methods to a dataset standardizes it. That means that each feature is presumed to have been drawn from a standard normal distribution with zero mean and unit variance [1, p. 127].

$$\hat{x} = \frac{x - \text{mean}(x)}{\text{sd}(x)}$$

Cross-Validation   Cross-Validation is a technique for estimating the performance of a predictive model. The technique can help to overcome generalization issues. A classical procedure involves dividing the complete data into q equal segments. One of the segments is then used for testing and the remaining $(q - 1)$ segments are used for training the model. This process is repeated q times by using each of the q segments as a test set. At the end the whole dataset has been predicted [1, p. 180].

2.2.2 Architectures

In machine learning different neural architectures are used. The ones important for this work are introduced in the following.

Feedforward Neural Network

The feedforward neural network was the first ANN. Its architecture is very simple because the information moves only forward, from the input units through the hidden

units and to the output units. There are no cycles or loops in the network. The feedforward network can approximate the function $y = f(x)$ as $\hat{y} = \hat{f}(x; \theta)$ for a given input x with f being the approximated function and θ being the weights that will be learned for the approximation [22, p. 163].

Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a subclass of artificial neural networks that in contrast to feedforward networks perform substantially better on sequential data such as speech or language [37]. This is because they have a feedback mechanism that integrates previous states of the neural network into the current state. RNNs process one element at a time and maintain in its hidden units a ',state vector' that implicitly contains information about the history of all past elements of a sequence. It thus learns the relationship between successive inputs and the decisions made are influenced by what has been learned in the past. They are particularly useful for discovering information in time coded data [80].
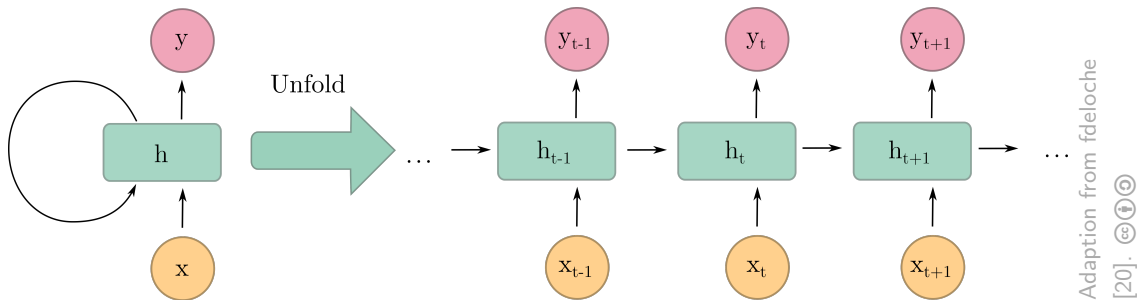
Figure 2.1: A basic RNN (left) with its time-layered representation (right)

The architecture of the networks is characterized by the fact that connections exist from neurons of one layer to neurons of the same or a previous layer. These loops allow the existence of temporal information in the form of a hidden state that changes with each new input [37]. The figure 2.1 illustrates the concatenated structure of an RNN. The hidden state at time t is given by the input vector at time t and the hidden vector at time $(t-1)$.

$$h_t = f(h_{t-1}, x_t)$$

Another function is then used to learn the output probabilities from the hidden states $y_t = g(h_t)$. The implicit assumption of the model is that the time-series has

15

underlying properties that do not change and can be captured in the function $f(\cdot)$ and $g(\cdot)$. Although this is not the case with most real world data, the assumption is helpful for regularization. Backpropagation has to take the sharing of the weights by multiple connections and the temporal length into account when updating the weights during the learning process. That's why this special type is also referred to as backpropagation through time [1, pp. 38–40]. Training of RNNs has been hard because of the resulting high depth of the network. Backpropagation here leads to either exploding or vanishing gradients in earlier layers making updates unstable. But this problem can be solved by using different activation functions such as a ReLU and new training methods such as an adaptive learning rate [1, pp. 28–29].

Even though long-term dependencies can be learned by RNNs, theoretical and empirical evidence shows that information is not stored for very long [6]. This problem got solved by adding an explicit memory to RNNs. These ANNs are called Long Short-Term Memory models (LSTMs) and are described in the following.

Long-Short Term Memory

As mentioned the high number of layers in RNNs gives rise to vanishing or exploding gradients during backpropagation which makes training very unstable. This results out of the successive multiplication of the weight matrices at various time-steps. To address this problem the recurrence equation for the hidden vector which connects different time-steps with each other gets replaced with an explicit memory. The first architecture of an RNN with that feature was proposed by Hochreiter and Schmidhuber in 1997 - the LSTM [27]. The explicit memory - also called memory cell - acts like a leaky gated neuron. The design intends to create a fine-grained control mechanism over the data written into this memory. The memory cell has a connection to itself at the next time-step and copies its content to it depending on gates that decide what information is going to be 'leaked'. The state of the memory is called cell state and is a combination of partial 'forgetting' and 'increment' operations of previous cell states (seen in figure 2.2). These state values are much more persistent and thus have a greater similarity through the memory cell across different temporal layers, which avoids the instability errors RNNs can experience. In other words it is harder for gradients of the next time step to be different from the previous one.

The cell state is controlled by gates that control the input, forget and output operations within the memory [1, p. 292]. In order to determine the hidden state h_t
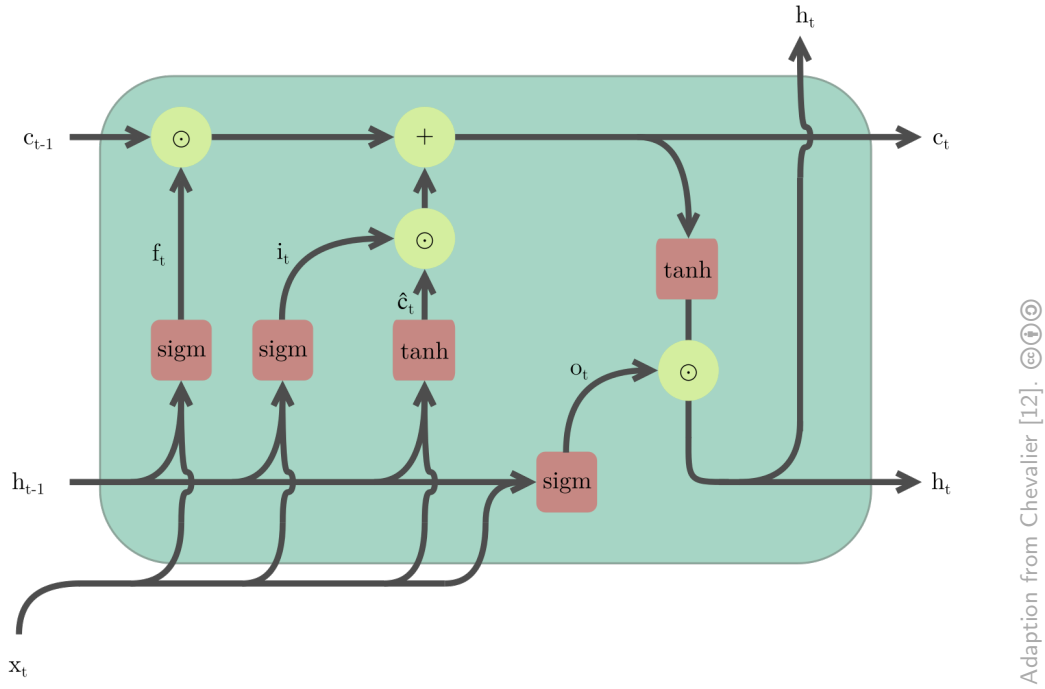
Figure 2.2: Schematic LSTM cell with its different gates.

and the cell state c_t of the memory cell following calculations have to be done. h_t and c_t are both p-dimensional.

First the intermediate variables have to be set up:

$$
\begin{matrix}
\text{Input Gate:} \\
\text{Forget Gate:} \\
\text{Output Gate:} \\
\text{New C.-State:}
\end{matrix}
\begin{bmatrix} i_t \\ f_t \\ o_t \\ \hat{c}_t \end{bmatrix} =
\begin{pmatrix} \text{sigm} \\ \text{sigm} \\ sigm \\ tanh \end{pmatrix}
W \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix},
$$

where x_t refers to the input layer and h_{t-1} to the hidden layer from the previous time step. Both are represented as a column vector and multiplied by the weight matrix W and their gate functions sigmoid (sigm) or tanh operator. In practice biases are also used but that has been omitted for simplicity. The intermediate variables i, f and o are vectors with continuous values between $(0, 1)$ and are referred to input, forget and output gates respectively. Conceptually these vectors can be seen as boolean gates that decide whether (i) to add to a cell state, (ii) to forget a cell state (iii) to allow leakage from the cell state into the hidden state. The boolean

analogy used helps with understanding the different decisions that are being made by the updates in the following. $\hat{c}_t$ contains the newly proposed content for the cell state c_t. The content actually being used in c_t is regulated by the following equation representing the selective forgetting and adding of new information to the long-term memory:

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t,$$

where $\odot$ is used here to denote the element-wise product of vectors. The first part of the equation uses the p forget bits to decide which of the p cell states from the previous time-step to reset to 0. The second part uses p input bits to decide which components of the proposed cell state $\hat{c}_t$ should be added to the new cell state c_t. The additive form of the equation helps as well with overcoming the previously mentioned vanishing gradient problem.The cell-state vector is hence the part of the memory cell that contains a continuously updated long-term memory. The functions decide at each time step (i) which information of the p cell states from the past to reset or (ii) which information of the current input to integrate into its memory. The cell state contains values between $[-1, +1]$ because it is the output of the tanh function. The next equation decides what part of the long-term memory cell state is to be leaked to the hidden state h_t:

$$h_t = o_t \odot \tanh(c_t)$$

The output bits of o_t decide which parts of the cell state should be incorporated into the hidden state [1, pp. 292–295].

Chapter 3

Materials and Methods

3.1 Data

The datasets used are largely consistent with those of Kraft et al. [34], as they cover well the abiotic factors that significantly influence vegetation (see section 2.1). The only exception is the use of the *ERA5* dataset instead of *ERA-interim*, as the former uses an improved prediction system for the reanalysis [26]. That resulted in a total of 6 dynamic and 21 static explanatory variables used to model the vegetation state.[2]

This section is divided into the subsection **Explained Variable** and **Explanatory Variables**. Both terms are commonly used to describe the variable that is to be predicted and the variables that are used for the prediction, respectively.

3.1.1 Explained Variable

VIs are widely used for measuring the abundance, relative intensity and condition of vegetation. For modelling, the most popular index, the NDVI was utilized. The variable was created within the framework of the Global Inventory Monitoring and Modeling System.[3] It is derived from imagery obtained from Advanced Very High Resolution Radiometer instruments [77]. With vegetation change interfering artefacts such as calibration loss or orbital drift were corrected afterwards [49]. Further, gap filling was derived from spline interpolation or from the seasonal profile. The dataset spans the time period from July 1981 to December 2015 with a spatial resolution of 0.083° with 15-daily maximum NDVI values measured ranging from -1 to 1 [77]. The NDVI covers most land regions.

---

[2]All datasets are available in an internal database collection of the Max-Planck-Institute for Biogeochemistry.

[3]https://ecocast.arc.nasa.gov/data/pub/gimms/3g.v1/

3.1.2 Explanatory Variables

The utilized explanatory variables can be analytically divided into two parts. One part of the variables has a temporal dimension with daily resolution and a spatial dimension in the following section referred to as dynamic climate variables. The other part is only spatially resolved and referred to below as static climate variables.

Dynamic climate variables

The following variables were used from the climate reanalysis dataset ERA5:

- the *minimum*, *maximum* and *mean air temperature* at $2\,\mathrm{m}$, measured in Kelvin

- the *surface solar radiation downwards*, measured in $\mathrm{MJ\,m}^{-2}$

- the *relative humidity* in percent

The dataset was released by the European Centre for Medium-Range Weather Forecasts, within the Copernicus Climate Change Service.[4] All variables span the period from 1979 to present with hourly estimates and a spatial resolution of 0.28°.

Precipitation, measured in mm, was utilized from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset [5]. The dataset is derived from gauge-, satellite-, and reanalysis-based data. It has been validated on a global scale and global comparisons suggest that MSWEP exhibits more realistic spatial patterns in mean, magnitude, and frequency [5]. The dataset was created by Beck et al. [4] and spans the period from 1979 to 2019 with a 3-hourly temporal and 0.1° spatial resolution.

All dynamic climate variables cover the entire globe.

Static Climate Variables

Available Water Capacity, measured in $\mathrm{mm\,m}^{-1}$ taken from the Harmonized World Soil Database with a spatial resolution of 0.0083° was used [19].[6] This variable indicates the amount of water that can be stored in a soil profile for plant growth [73,

---

[4] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5
[5] http://www.gloh2o.org/
[6] http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/
harmonized-world-soil-database-v12/en/

p. XII].

Water Table Depth, measured in meter by Fan et al. [18] with a spatial resolution of 0.5° was also utilized.[7] The groundwater table depth is globally interpolated from observations.

Further on, *19 Land Cover Fractions* derived from the Moderate Resolution Imaging Spectroradiometer by Friedl et al. [21] of the following classes were employed: Water, Evergreen Needleleaf Forest, Evergreen Broadleaf Forest, Deciduous Needleleaf Forest, Deciduous Broadleaf Forest, Mixed Forest, Closed Shrublands, Open Shrublands, Woody Savannas, Savannas, Grasslands, Permanent Wetlands, Croplands, Urban and Built-up, Cropland/Natural vegetation mosaic, Snow and Ice, Barren or Sparsely Vegetated.[8] In addition to that, the classes Croplands and Croplands/Natural Vegetation mosaic were obtained from Monfreda et al. [46].[9]

3.2 Pipeline

The prediction pipeline gives an overview of the steps conducted to generate the vegetation models. At first, the preprocessing of the datasets is described. Afterwards, the architecture of the RNNs is introduced and their hyperparameter optimization process described. Finally, an overview of the utilized technologies for the implementation is presented.

3.2.1 Preprocessing

First, the resolution of all variables was unified to 0.5° to ensure consistency among them and to reduce noise. In addition, the period January 1982 to December 2015 was selected from all temporally resolved variables, in order to cover the entire period of the explained variable. The preprocessing steps specific for the explained and the explanatory variables are illustrated in the following.

Explained variable Following Kraft et al. [34], only observations were included if the data pixel contained data at more than 80 % of the time steps at aggregation level. Further deserts, defined as pixels with more than 50 % barren and coastal

---

[7]not available online
[8]https://yceo.yale.edu/modis-land-cover-product-mcd12q1
[9]not available online

areas, defined as having 20 % water on it were excluded as well as snow covered regions and water bodies. Afterwards, the explained variable contained only data points in a value range between 0 and 1. Data points derived from interpolation or the seasonal cycle were not removed to provide the model with more training data.

Explanatory variables First, the explanatory variables were standardized with a mean value of 0 and a standard deviation of 1. Next, two datasets were created containing static climate variables and dynamic climate variables in (1) daily resolution and (2) mean-aggregated 15-daily resolution. Note, that the 15-day period used for aggregation is equal to the 15-day measurement period of the NDVI.

3.2.2 Architecture

In order to generate two vegetation models (M_D and M_{15-D}), whose performance was compared, two RNNs were trained on the dataset with daily resolution (RNN_D) and 15-daily resolution (RNN_{15-D}). Based on the information about ANNs in section 2.2, an LSTM formed the basic architecture for both RNNs. The specific architectural design of the RNNs was inspired by the design of Kraft et al. [34], since it produced good results in their work. However, some parts of the models' architecture were changed after an independent model optimization phase. This sections describes, which parts of the RNN's architecture and their hyperparameters' remained the same during the optimization process. Subsequently, the optimized parts are listed and the procedure of hyperparameter optimization is described.

Consistent model architecture

First the input data was passed into a fully connected feed forward network with *two layers* and 128 neurons. Then a dropout of 0.1 was performed as a regularization technique. This was followed by an LSTM architecture that was determined in section 4.1. After the LSTM, another dropout of 0.1 was applied. Ultimately, the prediction was passed through *one fully connected layer* with a *sigmoid activation* function used to map the output of the LSTM to the prediction of the NDVI in the value range between $(0, 1)$.

As the optimization algorithm, the adaptive learning rate optimizer ADAM [33] was utilized since it handles the complex training dynamics of an LSTM better than

plain gradient descent. As the loss function, the mean squared error (MSE) was used with:

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n},$$

where y is the observation, $\hat{y}$ is the prediction and n the batch size. During training, Early Stopping was used if the average validation loss did not decrease further after 6 epochs and the training data was reshuffled at the beginning of every epoch. (Note: Early Stopping was not used during optimization.)

Optimized model parts

Three different *LSTM architectures* were compared in the optimization phase:

- LSTM with two layers (LSTM-2)

  Stacking multiple LSTM layers on top of each other can lead to performance improvements for sequence prediction problems as seen in Graves et al. [23].

- LSTM with one layer and Layer Normalization applied to it (LSTM-Norm)

  The technique normalizes the activities across the neurons in a layer [2], can speed up learning and often improves final performance [85].

- LSTM with one layer and its Forget Gate bias initialized to 1 (LSTM-Bias)

  Initializing the bias of the forget gate of the hidden state to a positive value can lead to performance gains [30].

Furthermore, the *batch size*, the *learning rate* and the *hidden layer size* of the LSTM were determined.

Hyperparameter optimization

At first a batch size was determined that filled up the entire GPU memory. This was done to speed up training process since larger batch sizes result in less noisy gradient estimates and therefore allow SGD-type optimizers to take larger steps, leading to the same convergence as smaller batch sizes would. but in a smaller number of iterations[43].

Next, the Asynchronous Successive Halving Algorithm (ASHA) [40] was used to find a good hyperparameter configuration. The algorithm employs a technique called aggressive early stopping. The technique is based on the assumption that promising hyperparameter configurations tend to score higher, relatively to worse ones, even

early on in the training process and therefore model configurations that perform badly in the beginning can be directly stopped. Due to time constraints, ASHA tested the performance of only 15 randomly selected model configurations for 10 epochs by measuring their validation loss. The reduction factor was set to three, meaning that in every iteration of the algorithm 75 % of bad performing models were no further trained. The model configurations consisted of the LSTM architecture (LSTM-2, LSTM-Norm, LSTM-Bias), the size of its hidden layer $(128, 256, 512)$, and the learning rate $(1 \times 10^{-10}, 1 \times 10^{-10})$. The selection range from which the values were chosen is given in brackets. The model configuration, which achieved the lowest validation loss was selected.

The last step was to confirm the learning rate determined by ASHA with the method described in Smith [74]. For this purpose, the previously determined parameters - except for the learning rate - were fixed. Subsequently, the measured loss occuring at different learning rates between $(1 \times 10^{-8}, 1)$ was plotted against each other. Next, the minimum of the derivative of the resulting function was determined and chosen as the initial learning rate.
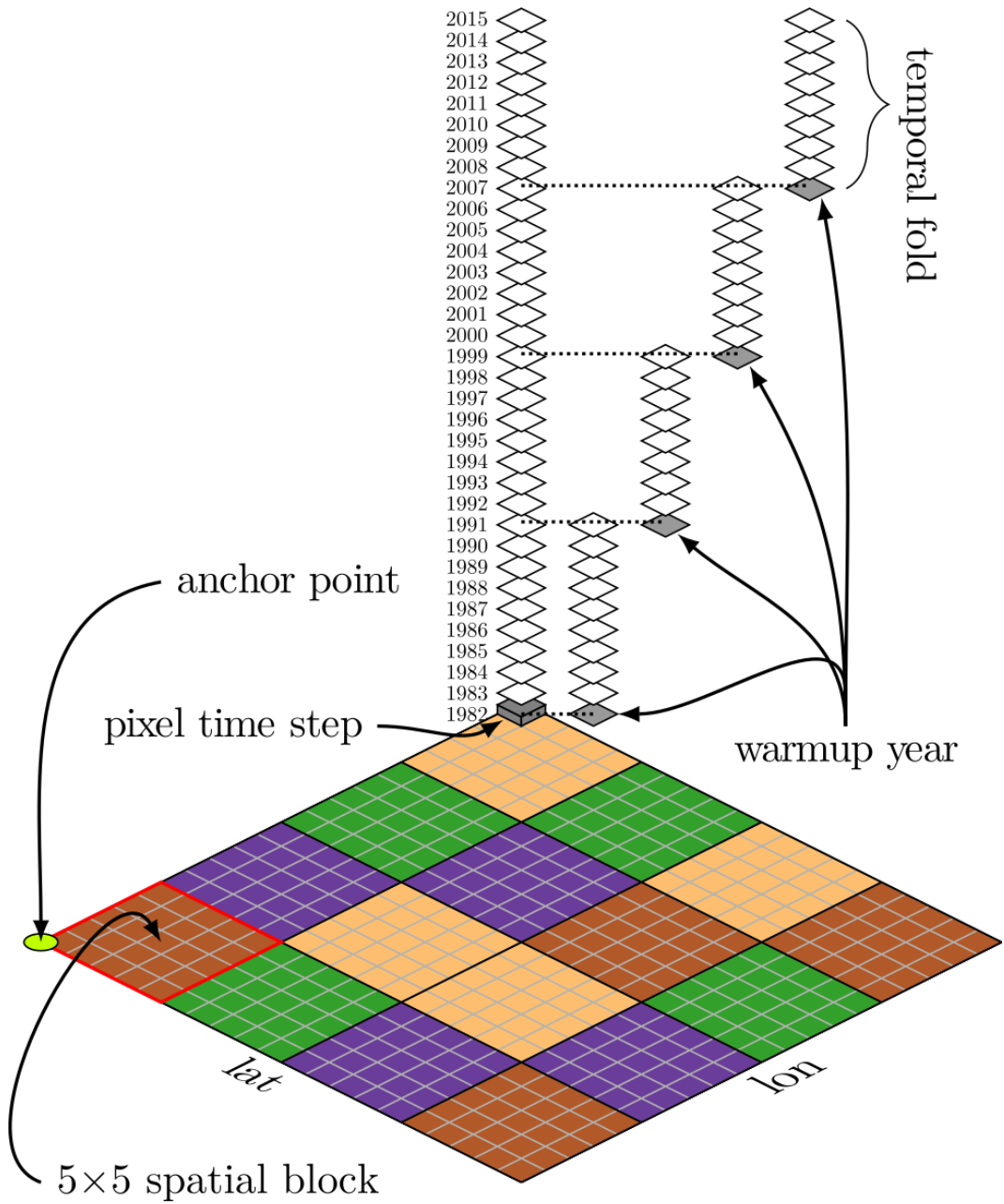
3.2.3 Cross-Validation

There is an autocorrelation in time and space between environmental variables. That means the value pairs of randomly chosen variables at places that are close in time or space are more or less similar than it would be expected for a randomly chosen value pair of these variables [38]. Therefore, when modeling environmental variables, a non-independence of residuals is observed and the model may overfit the data and absorb structured residual variations, that in fact cannot be adequately explained by the utilized input variables. For example, if the model contains residuals resulting from different regional nutrition profiles of the soil, these residuals should not be explained by the input data.

Hence, in contrast to the simple cross-validation strategy outlined in section 2.2.1 and in order to learn in a bias-reduced way, the input data are divided into spatially distinct blocks. The choice of block sizes results in a trade-off between autocorrelation requirements, data limits and computational limits [67]. The chosen implementation aligns with the cross-validation scheme of Kraft et al. [34], where the datasets are divided into spatial and temporal blocks to reduce the autocorrelation between environmental variables. Note that the blocking approach increases the independence between the cross-validation sets, but does not entirely remove it. This scheme was utilized for the prediction of all vegetation models.

Following, is an overview of the implementation of this scheme (see also figure 3.1): spatial blocking was implemented by randomly assigning 5×5 pixel blocks of the world map to one of four spatial folds. This was implemented using a block mask (see figure A.1). The temporal blocking was achieved by splitting the time period from 1982 to 2015 (33 years) into four folds, with the first fold of 10 years and the three remaining folds of 9 years each. These time periods overlap by one year with the next time period. The first year of each time period can be seen as a warm-up period to ensure that the LSTM's memory cell has enough information about potential memory effects encoded before its predictions are used. Each model's training set consisted of two spatial blocks and three time slices to leave as much data as possible for the training of the model. The remaining time slice was used for the validation set and test set, with each of them containing one of the two remaining spatial blocks. This way each fold contained about 73 % of the data for training and 13.5 % for validation purposes. For the test set, 12.6 % of the input data was used. The absolute amount of data points for training, validation and test sets are listed in table 3.1. In total, 16 folds (four folds temporal and four folds spatial blocking) were used for one complete prediction of the NDVI from 1982 to 2015. In order to make predictions more robust, the prediction was repeated five times with different spatial blockings. Thus, each RNN was run a total of 80 times for the generation of its vegetation model.

		Amount of Data Points			
RNN		Training	Validation	Test	Total
	in %	73.8	13.6	12.6	100
RNN_D		186 701 523	34 368 872	31 686 680	252 757 075
$RNN_{15\text{-}D}$		12 267 216	2 258 496	2 082 240	16 607 952

Table 3.1: Valid data points utilized in the training, validation and test set in percent (top row) and in absolute numbers for RNN_D (second row) and $RNN_{15\text{-}D}$ (third row).

Figure 3.1: Spatio-temporal cross-validation scheme: the 4 temporal folds contain 9 years - except for the first fold containing 10 years - with one year overlap to the next fold (warmup year). While the temporal blocking is fixed the temporal subdivision is fixed, the spatial blocking of blocks (of 5×5 pixels) is random. The four four different colors represent one spatial fold each. The entire cross-validation is repeated 5 times with changing anchor points such that the points covered by one block are varying.

3.2.4 Technology

The following list contains all technologies, frameworks and libraries used for the implementation of the pipeline on a Nvidia DGX-1 workstation with four available Tesla V100 GPUs (16GB):

Preprocessing

- xarray 0.16.0 [59] for the preprocessing of the netcdf datasets because of its intuitive handling of multidimensional datasets [28]

Prediction framework

- the Docker Image PyTorch 20.03 by Nvidia [51] with the Python distribution Anaconda with Python 3.6 [60] and numpy 1.18.5 [50]

- from the open source machine learning library PyTorch 1.5.0 [61] the pre-implemented ANN model architectures and the optimization algorithm Adam alongside PyTorch's data loading mechanisms

- PyTorch Lightning 0.8.5 [62], a high-level interface for PyTorch, utilized for structuring the code and to enable an intuitive GPU deployment for the RNNs

Hyperparameter optimization

- the ASHA implementation of the python library RAY 0.8.7 [64]

- PyTorch for implementing the LSTM-2 and LSTM-Bias architecture

- the implentation of the LSTM-Norm architecture from haste 0.4.0 [41]

- from PyTorchLightning its learning rate and batch size finder

Evaluation

- TensorBoard 2.3.0 [76] for logging the training and optimization progress and pandas 0.24.2 [53] for the analysis of the logs

- matplotlib 3.2.1 [42] and xarray for generating plots and maps

The code for the prediction framework was written in Visual Studio Code 1.41.1 [45] via code-server 2.1698 [15]. JupyterLab 1.0.4 [58], was utilized for live-coding during the preprocessing and evaluation step.

3.3 Evaluation Metrics

First, more robust vegetation models M_D and $M_{15\text{-}D}$ were created by aggregating the five model predictions of each models' respective RNN. Afterwards the observation of the NDVI and the models' predictions of it were masked such that only non-gap-filled values i.e. actually measured values were considered during evaluation [57]. Furthermore, the time series of the observation and the predictions was resampled to a monthly resolution.

The raw time series (NDVI_{RAW}) were then split up into their seasonal components (NDVI_{MSC}) and their anomaly components (NDVI_{ANO}). NDVI_{MSC} was computed as the pixel-wise median per month over the whole time series and the anomalies were calculated with $\text{NDVI}_{\text{ANO}} = \text{NDVI}_{\text{RAW}} - \text{NDVI}_{\text{MSC}}$.

In the next step, the predictive performance of M_D and $M_{15\text{-}D}$ was measured for the raw time series and its decomposed components. This was done by calculating the Root Mean Squared Error (RMSE) and the Nash–Sutcliffe model efficiency coefficient (NSE) [48] between the observation and the respective model prediction. The RMSE is a frequently used metric for quantifying the prediction error in terms of the units of the variable calculated by the model:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}},$$

where O_i and P_i represent the sample of size N containing the observations and the model estimate, respectively. It ranges from 0 to ∞, with RMSE = 0 indicating a perfect fit. The NSE is a dimensionless goodness-of-fit indicator ranging from $(-\infty, 1]$. Its relating formula is as follows:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2},$$

where $\overline{O}$ is the mean of the observed values. A NSE of 1 indicates that the model fits perfectly to the observation, while an NSE lower than 0 suggests that the mean of the observed value of the observation is a better predictor than the evaluated model itself [66].

The performance between M_D and $M_{15\text{-}D}$ was compared on different levels:

- **global performance**: it was calculated as the combined RMSE and NSE over all pixels and time-steps.

- **regional performance**: all pixels were pooled over the temporal dimension in order to find out if the models had a diverging performance in specific regions of the world.

- **biome-specific performance**: the maps resulting from the measurement of regional performance were aggregated per hydro-climatic biome (see figure A.2), as defined by [55].

As an additional metric, the modelling capability of memory effects (Mem) as defined by Kraft et al. [34] were calculated between the models with access to time series information (M_{Full}) and a baseline model without information about the temporal context (M_B). The baseline was obtained by training the $RNN_{15\text{-}D}$ with 15-daily aggregated input data, whose time series were randomly permuted. Thus, M_B did only learn instantaneous effects between input and output variables. Mem is defined as the performance improvement between M_{Full} and M_B. It is calculated as the difference between the NSE of $NDVI_{ANO}$ of M_{Full} (Mem_{Full}) and M_B (Mem_B): $Mem = Mem_{Full} - Mem_B$. Mem is measured in %. The idea behind this metric is that the $NDVI_{ANO}$, i.e. the changes in vegetation that deviate from the seasonal cycle, can be interpreted as a memory effect, since these anomalies are mainly caused by short-term and mid-term climatic anomalies (see section 2.1.3). Furthermore, the NSE of $NDVI_{ANO}$ measures the amount of variance in the observation that is found in the prediction. Therefore, a high NSE approximates better the variance between observation and prediction.This is assumed to correlate with a good prediction of memory effects. The difference between Mem_{Full} and Mem_B quantifies the importance of time series information utilized for generating M_{Full} in terms of memory effects.

Chapter 4

Results

4.1 Hyperparameters

This section gives a short overview of the optimization process. Table 4.1 lists the final hyperparameter configurations that were selected for the training of RNN_D and $RNN_{15\text{-}D}$, and thus also for the prediction of the vegetation model M_D or $M_{15\text{-}D}$ and M_B.

Determination of Batch Size In the first step of the optimization phase the batch size finder determined 83 and 682 as the batch size that filled up the entire GPU for RNN_D and for $RNN_{15\text{-}D}$ respectively. To assure that no memory allocation error would occur during training, the batch size was slightly decreased to 80 (RNN_D) and 680 ($RNN_{15\text{-}D}$) filling roughly 15 GB GPU memory. With this batch size one training epoch consisted of approximately 380 (RNN_D) and 45 iterations ($RNN_{15\text{-}D}$).

Hyperparameter Search The following hyperparameter search with ASHA generally went as planned, except that three randomly generated configurations lead to NaN losses for $RNN_{15\text{-}D}$. That resulted in only twelve configurations being effectively evaluated for this RNN. The search revealed that both RNNs did not show a high sensitivity to different hyperparameter configurations except for the model architecture.

The best performing configuration for RNN_D - with an average validation loss of 4.44×10^{-3} - was found to be the LSTM-2 architecture with a hidden layer size of 256 and a learning rate of 3.9×10^{-4}. In the case of $RNN_{15\text{-}D}$ the configuration with the highest performance in the tests was the LSTM-2 architecture as well with a hidden layer size of 512 and a learning rate of 5.5×10^{-4}. The model had an average validation loss of 4.56×10^{-3}. The results of the second best configurations can be found in table B.1.

Learning Rate Finder To reassure that the randomly chosen learning rate was a good fit for both RNNs the learning rate finder was deployed and different learning rates between $(1 \times 10^{-8}, 1)$ tested (see B.1). The optimal learning rate resulted in being close to the ones already determined by ASHA: 5.25×10^{-4} for RNN_D and 4.37×10^{-4} for $RNN_{15\text{-}D}$.

RNN	Batch Size	Learning Rate in 10^{-4}	LSTM	
			Architecture	Hidden Units
RNN_D	80	4.37	LSTM-2	256
$RNN_{15\text{-}D}$	680	5.25	LSTM-2	512

Table 4.1: Final hyperparameter configurations of both RNNs utilized for predicting the vegetation models.

4.2 Performance

4.2.1 Training Progress

Since the model generated with RNN_D contained about 15 times more data points than the models generated with $RNN_{15\text{-}D}$, the training times differed substantially. The prediction of M_D needed a total of 27.5 GPU days and for $M_{15\text{-}D}$ and M_B 7.5 GPU days. In order to optimize the GPU usage, the training was conducted in two phases: Phase one contained, the prediction of $M_{15\text{-}D}$ and M_B with $RNN_{15\text{-}D}$ and phase two the prediction of M_D with RNN_D. For all RNN runs the mean validation loss was slightly higher than the mean train loss indicating a good fit for all models for their respective datasets. A plot of the average training and validation progress is found in figure B.2. The mean validation and test loss differed only marginally for all runs, with a slight worsening (1.8 %) for $M_{15\text{-}D}$, a slight improvement (0.8 %) for M_D and remaining unchanged for M_B (see table 4.2). As seen in table 4.2 The lowest average test loss was achieved for the training of M_D with 3.87×10^{-3}. A scatter plot containing the test losses for all runs is found in figure B.3.

Model	Training Times		Number of		Loss in 10×10^{-3}		
	in h/fold	in d/block	epochs	steps	Training	Validation	Test
M_D	8.2	5.5	40.8	10 429	2.58	3.9	3.87
$M_{15\text{-}D}$	2.3	1.5	51.7	1571	3.14	4.17	4.1
M_B	2.4	1.6	54.1	1641	5.03	5.79	5.79

Table 4.2: Average training progress metrics over 80 runs, that have been conducted to generate the models M_D, $M_{15\text{-}D}$ and M_B.

4.2.2 Evaluation

The performance evaluation was conducted on different levels for the vegetation models. At first, an exemplary time series decomposition is demonstrated to give a first impression of the different performances between M_D, $M_{15\text{-}D}$ and M_B. This is followed by a short analysis of the global performance metrics and a more detailed evaluation of the regional an more specifically biome-specific performance differences between M_D and $M_{15\text{-}D}$.

Time-Series Decomposition

For the time series decomposition a 50km grid of the Thuringian Forest of the year 1990 was selected. Figure 4.1 contains this decomposition of all models column-wise for M_D, $M_{15\text{-}D}$ and M_B, respectively. Furthermore, the time series of the observation (grey) is drawn into each plot to give a visual impression of the predictive performance of each model. The first row of each column contains the raw time series and row two and three its decompositions. The text box on the top right of each graph contains the RMSE and NSE for the respective model and time series of the exemplary data point.

Visually comparing the raw time series (solid line), it shows that M_D slightly outperforms $M_{15\text{-}D}$ and M_B by approximating the bumps in the function better. In contrast the Median seasonal cycle (MSC) (dashed line) was approximated well by all models, starting with a steep rise in February that continued until May. The sequence was followed by a slow flattening until October where the decrease started to accelerate. The only exception was the sudden rise of the NDVI in October. This rise was incorrectly approximated by all models however M_D and $M_{15\text{-}D}$ did adapt better to this deviation than M_B did. The better performance of M_D and

---

[9]The year was selected by chance.

$M_{15\text{-}D}$ compared to M_B was again reflected in the anomalies (dotted line), whereby M_D showed the best fit with a NSE of 0.31. Nevertheless, it is apparent that the predictive accuracy of the models differs, especially for the anomalies. This is also evident when looking at the spatially-aggregated time series of the full time range in figure B.4.
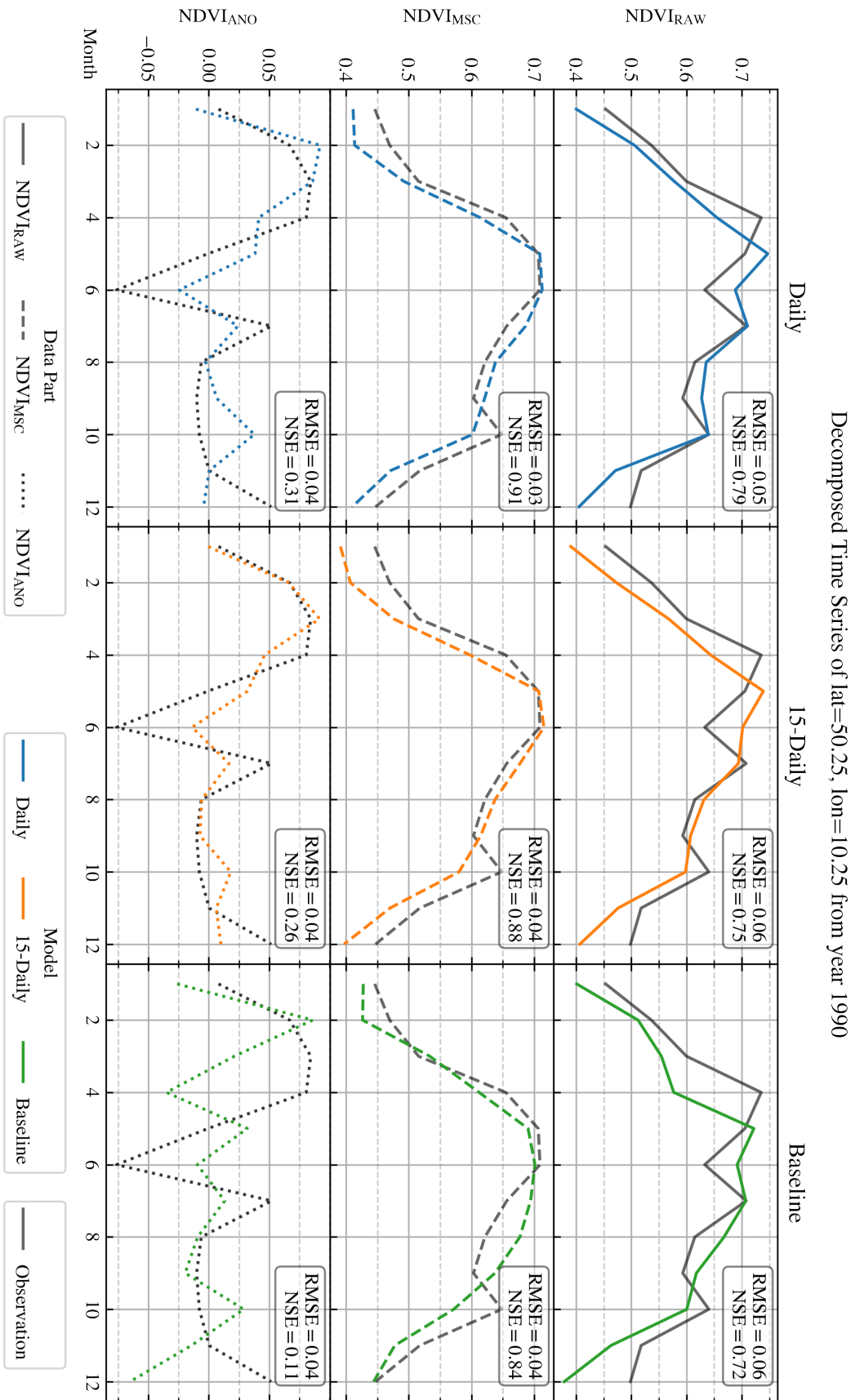
Figure 4.1: Exemplary time series decomposition of the NDVI of the year 1990 for a 50 km grid of the Thuringian Forest at 50.25°N, 10.25°S.

Global Metrics

The better performance of M_D observed in the previous section was found on the global level as well. M_D achieved an RMSE of 0.052 compared to model M_{15-D} with an RMSE of 0.054. This is an error reduction of 4 %. The NSE increased by roughly 0.4 % from 0.950 to 0.954 from M_{15-D} to M_D (see table 4.3).

Looking at the decomposed time parts between M_D and M_{15-D}, major improvements in the prediction performance were seen primarily for the anomalies with regards to the NSE (about 38 %) and the RMSE (2.5 %). For the median seasonal cycle the NSE improved only marginally by 0.3 % and the RMSE the rmse improved by 5.7 %.

As expected, the baseline model performed worst for all metrics. Further, the performance improvements between M_B and the daily and 15-daily model occurred in the same manner as observed between M_{15-D} to M_D: The NSE increased significantly (about 57.9 % and 14.5 %) for the anomalies and the RMSE decreased (about 26.9 % and 22.6 %) for the median seasonal cycle (from M_B to M_D and from M_B to M_{15-D} respectively).

Metrics	Time Series	Models			increase in %		
		M_D	M_{15-D}	M_B	$M_{15-D} \rightarrow M_D$	$M_B \rightarrow M_{15-D}$	$M_B \rightarrow M_D$
	$NDVI_{ANO}$	0.158	0.115	−0.100	37.8	215	258
NSE	$NDVI_{MSC}$	0.977	0.974	0.957	0.3	1.8	2.1
	$NDVI_{RAW}$	0.954	0.950	0.926	0.4	2.5	2.9
	$NDVI_{ANO}$	0.039	0.040	0.044	−2.5	−10.3	−12.5
RMSE	$NDVI_{MSC}$	0.037	0.039	0.050	−5.7	−22.6	−26.9
	$NDVI_{RAW}$	0.052	0.054	0.066	−4	−17.3	−20.6

Table 4.3: Global model performance in terms of NSE and RMSE for M_D, M_{15-D}, M_B for for all time series (column 3 to 5). Improvement of model prediction in percent from M_{15-D} to M_D and M_B to M_{15-D} or M_D, respectively (column 6 to 8).

Regional Performance

In order to investigate how the performance differences between the vegetation models observed at the global level were expressed regionally, the performance metrics were aggregated over the time dimension and then analysed. This way a map could be plotted that contained the mean performance of each pixel in terms of RMSE and NSE.

Looking at figure 4.2, the performance of M_D and $M_{15\text{-}D}$ (two upper maps) as well as its difference (bottom map) for $NDVI_{RAW}$ is illustrated. Both upper maps show that the RMSE is distributed homogeneously over the entire measuring area. Regions with a low RMSE can be found mostly in arid regions, such as areas bordering on the Sahara or Western Australia, probably due to low vegetation signal. The bottom map of difference between M_D and $M_{15\text{-}D}$ indicates through its blue shading that M_D achieved overall a lower RMSE. Especially between 0° to 20°S and Eastern Europe.

Figure 4.3 contains three maps with the performance map for the NSE of the $NDVI_{RAW}$. As seen in the upper two maps the NSE was in general high in the northern temperate and boreal regions, eastern South America, as well as Steppe and Savanna ecosystems of Africa. All these regions have a distinct seasonal NDVI signal. In contrast rainforests and dry regions showed lower NSE values. This can be attributed to the fact that the influence of the seasonal cycle is only marginal in these regions and therefore prediction errors take a larger effect due to lower overall variance of the MSC. The difference shaded in red (bottom map) indicates that main improvements of the NSE were found in tropical regions in South America and Africa and in the north of Australia.

Comparing the RMSE for the decomposed parts of the time series (see figure B.5, bottom two maps) that the main improvements from $M_{15\text{-}D}$ to M_D were found for the MSC (map bottom left). Nevertheless, the red dots on the map indicate that in some cases the predictions got slightly worse. For the anomalies (map bottom right), main improvements were found in northern temperate regions.

Major improvements from $M_{15\text{-}D}$ to M_D in terms of NSE (see figure B.6) were found for for $NDVI_{MSC}$ (map bottom left) in the southern hemisphere (the north of South America, Central Africa and Australia) and for $NDVI_{ANO}$ (map bottom right) in Europe, Central Africa and the eastern part of North America, which are the same regions on the map that improved the RMSE of $NDVI_{ANO}$ as well.

Next, Mem, the performance metric to measure the modelling capability of memory effects was determined for both models. Mem was 25 % for M_D and 21.9 % for

$M_{15\text{-}D}$. The improvement of 3.1 % between $M_{15\text{-}D}$ and M_D indicate that memory effects in vegetation change were potentially better modelled with M_D. Looking at the map of differences between both (see figure 4.4) the prediction improved especially in boreal regions and the subtropical and transitional regimes.

Biome-specific Performance

Figure 4.5 and table B.2 give insight about the performance of the models in different hydroclimatic biomes. In the following, the number in brackets indicates the respective biome(s). The mean RMSE for M_D and $M_{15\text{-}D}$ lies in the range from 0.04 (6) to 0.06 (4). In terms of NSE all boreal regions (1-4) achieve a high NSE of over 0.9 followed by mid-latitude regions (7,8) with an NSE of around 0.8. The lowest NSE is seen in the subtropical water driven (6) and tropical biomes (11) with around 0.4.

Comparing M_D to $M_{15\text{-}D}$ it is found that the biggest improvements between both models are located in the tropics (10) with the NSE improving 12.5 % and the RMSE improving 5.8 % from $M_{15\text{-}D}$ to M_D followed by the transitional energy-driven regime (11) and the subtropical water-driven regime (6). The violin plots indicate that the variance is highest in these regions considering all models. Furthermore, it is notable that the biggest performance improvement between $M_{15\text{-}D}$ and M_D are found in the same biomes, whose predictions improved most from M_B to $M_{15\text{-}D}$. The lowest improvements with respect to the NSE and RMSE were found to be in the boreal biomes (1-4) from M_B to $M_{15\text{-}D}$ as well as from $M_{15\text{-}D}$ to wM_D.

For the biome-aggregated metrics for the time series decomposition (see figure B.7) it is striking that for M_D and $M_{15\text{-}D}$ the MSC in boreal regions (1-3) has a similar NSE of around 0.98 and an RMSE of around 0.03. Further, the metrics for the MSC diverge between M_B and both other models strongly in the tropics (1), subtropics (6,9) and transitional regimes (5,11). The RMSE for $NDVI_{ANO}$ is similar for all models but slightly higher in boreal regions (1-4) and the subtropical energy-driven biome (9). The highest prediction of anomalies was reached in boreal regions (1-4) and the transitional water driven regime (5), with (1) improving by 70 % from 0.14 to 0.24 (NSE) and by 6 % from 0.046 to 0.043 (RMSE) from $M_{15\text{-}D}$ to M_D. The distribution of M_D and $M_{15\text{-}D}$ was found to be similar for all biomes for the decomposed and raw time series.
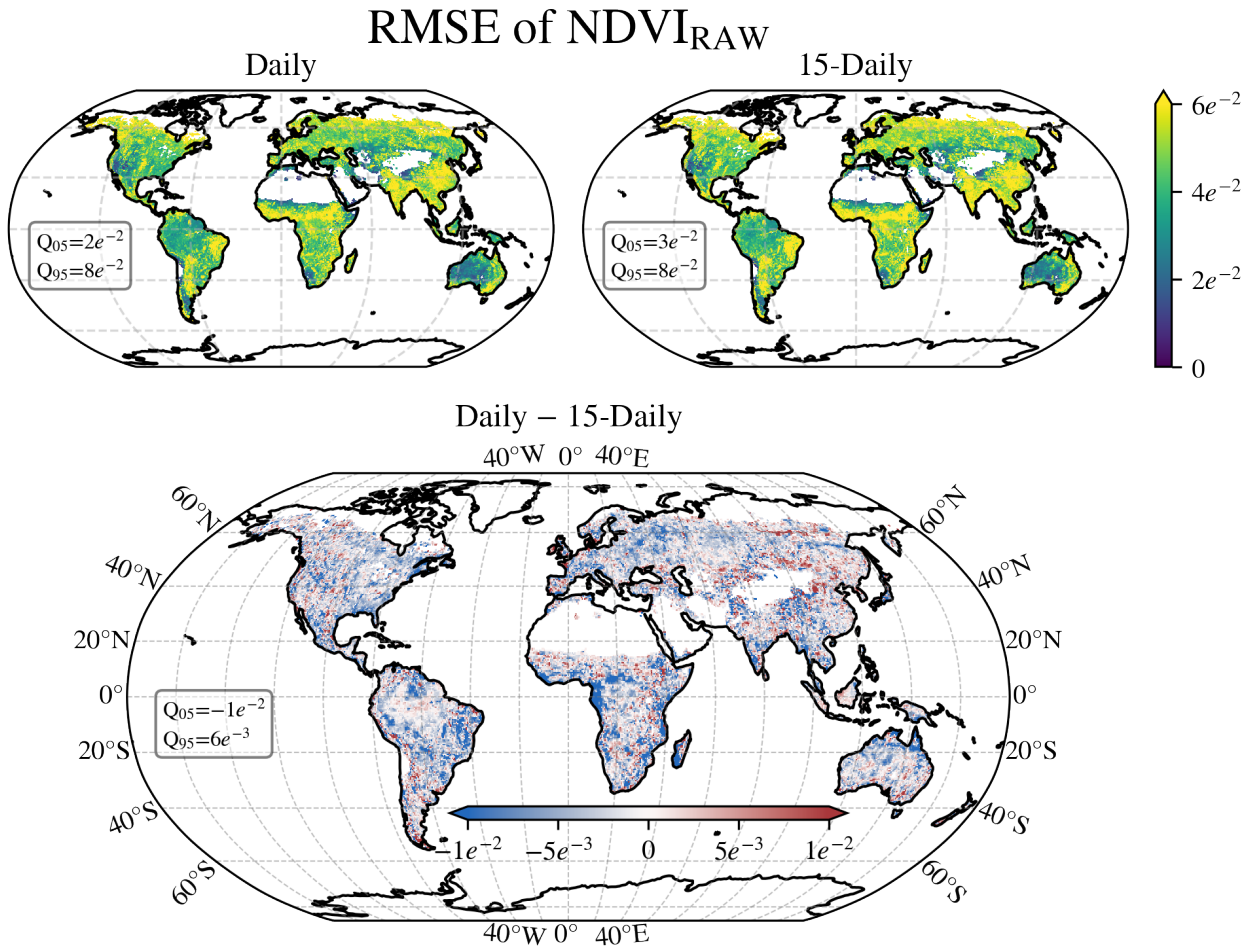
RMSE of NDVI$_{\text{RAW}}$

Daily

$Q_{05}=2e^{-2}$
$Q_{95}=8e^{-2}$

15-Daily

$Q_{05}=3e^{-2}$
$Q_{95}=8e^{-2}$

$6e^{-2}$

$4e^{-2}$

$2e^{-2}$

0

Daily $-$ 15-Daily

40°W 0° 40°E

60°N

40°N

20°N

0°

$Q_{05}=-1e^{-2}$
$Q_{95}=6e^{-3}$

20°S

40°S

60°S

60°N

40°N

20°N

0°

20°S

40°S

60°S

$-1e^{-2}$ $-5e^{-3}$ 0 $5e^{-3}$ $1e^{-2}$

40°W 0° 40°E

Figure 4.2: Performance map in terms of RMSE for NDVI$_{\text{RAW}}$: the two upper maps show the mean RMSE of M$_{\text{D}}$ and M$_{\text{15-D}}$ for the aggregated raw time series. Both maps are anchored in the same value range between 0 and 0.06. Darker [lighter] values indicate places with a lower [higher] RMSE. The bottom map is the map of difference between M$_{\text{D}}$ and M$_{\text{15-D}}$. The color bar is centered around 0 between anchor points -0.01 and 0.01. Red [blue] indicates that the prediction of the M$_{\text{D}}$ [M$_{\text{15-D}}$] is better. The boxes in the left part of the maps show the value range of the 5th and 95th quantile of the data.
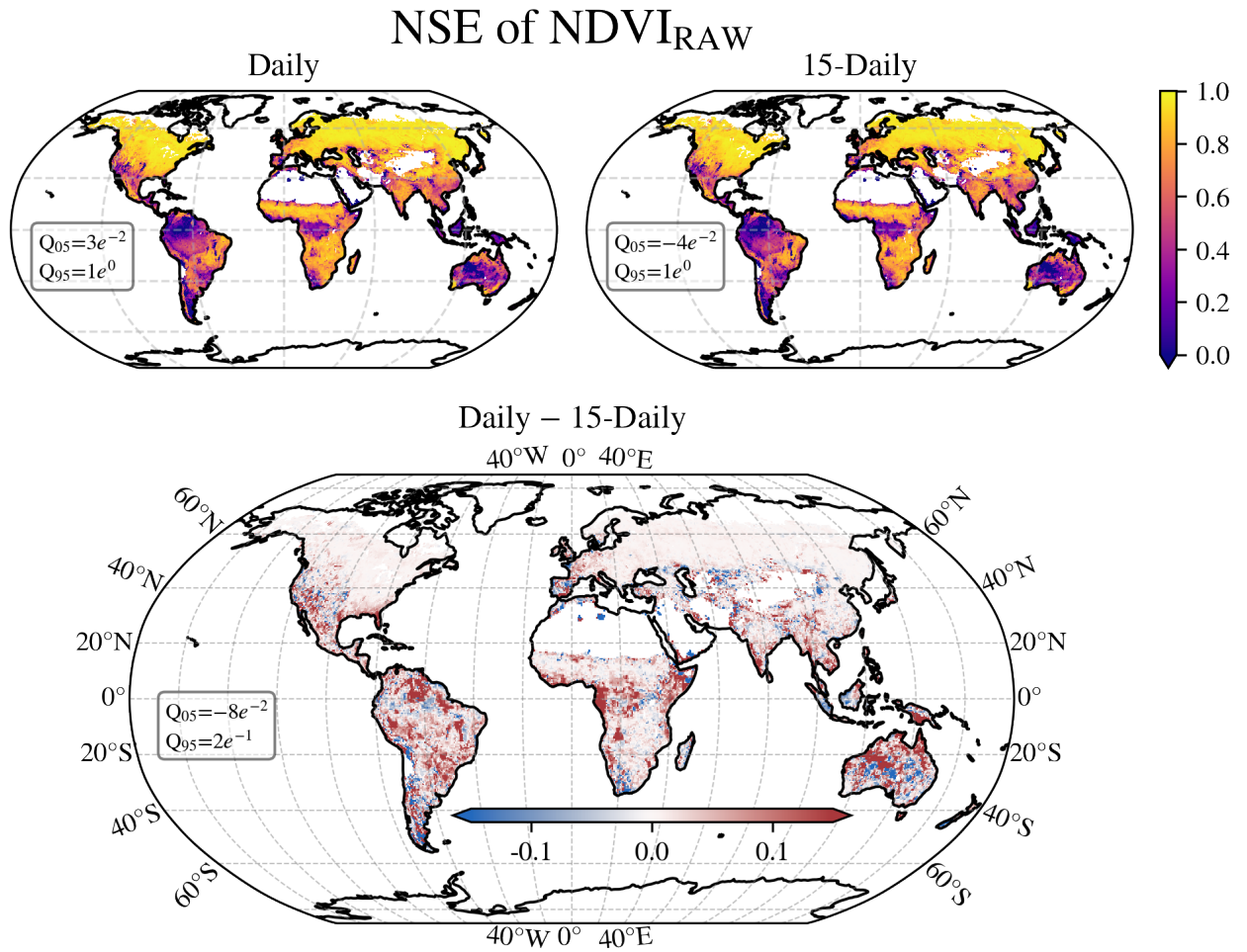
NSE of NDVI$_{RAW}$

Figure 4.3: Performance map in terms of NSE for NDVI$_{RAW}$: the two upper maps show the mean NSE of M$_D$ and M$_{15-D}$ for the aggregated raw time series. Both maps are anchored in the same value range between 0 and 1. Darker [lighter] values indicate places with a lower [higher] NSE. The lower map is the map of difference between M$_D$ and M$_{15-D}$. The color bar is centered around 0 between anchor points -0.15 and 0.15. Red [blue] indicates that the prediction of the M$_D$ [M$_{15-D}$] is better. The boxes in the left part of the maps show the value range of the 5th and 95th quantile of the data.
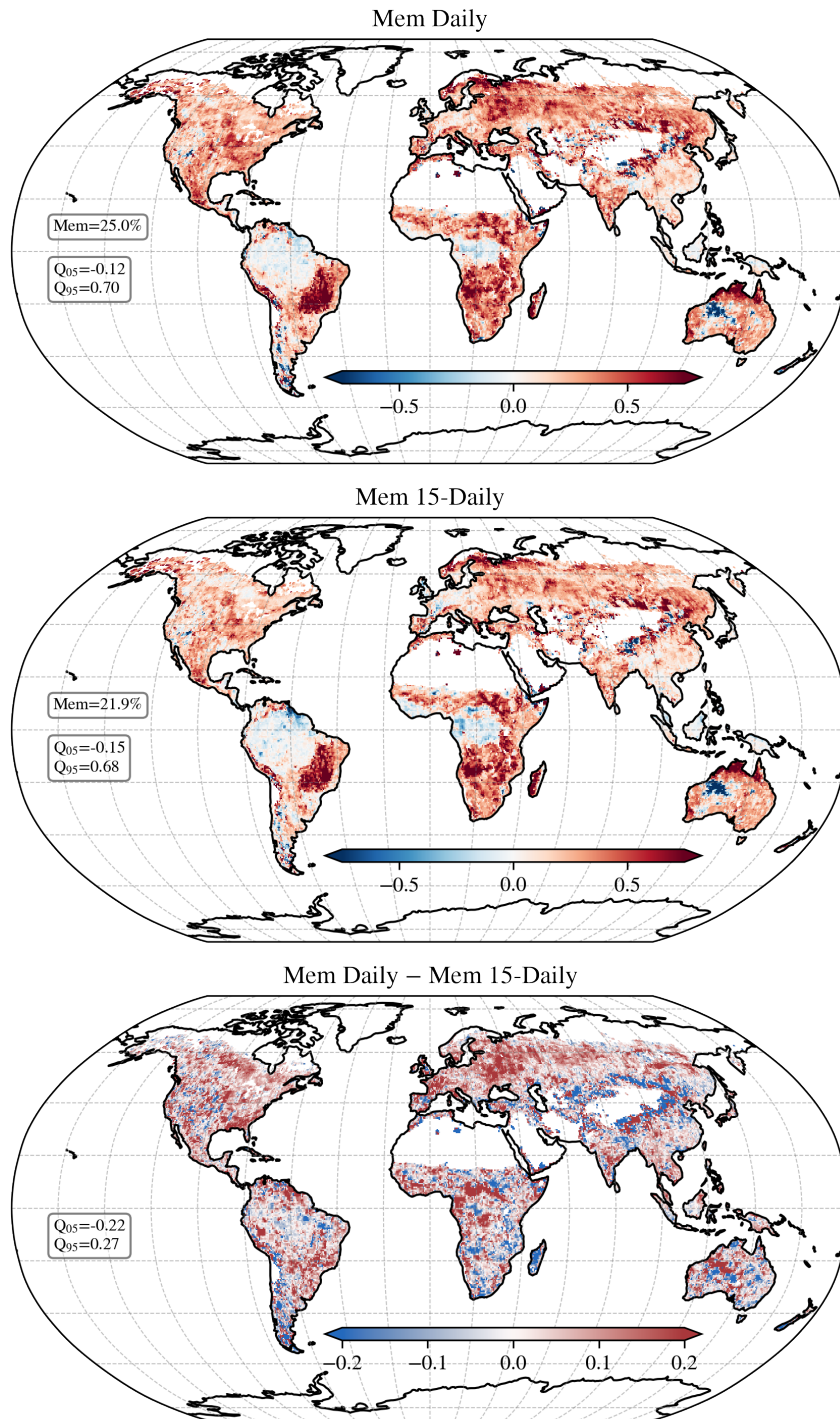
Figure 4.4: Performance maps in terms of Mem: the upper and middle map show the Mem of M_D and $M_{15\text{-}D}$ respectively. Both maps are centered around 0 between anchor points -0.75 and 0.75. Red shades indicate places where the memory effects are better modelled. The bottom map shows the spatial differences in modelling Mem between M_D and $M_{15\text{-}D}$. The color bar is centered around 0 between anchor points -0.2 and 0.2. Red [blue] indicates spots, where memory effects are more predicted by M_D [$M_{15\text{-}D}$]. The boxes in the left part of the maps show the value range of the 5th and 95th quantile of the data. Furthermore, for the upper and middle map the globally aggregated Memory Effect is stated as 'Mem'.
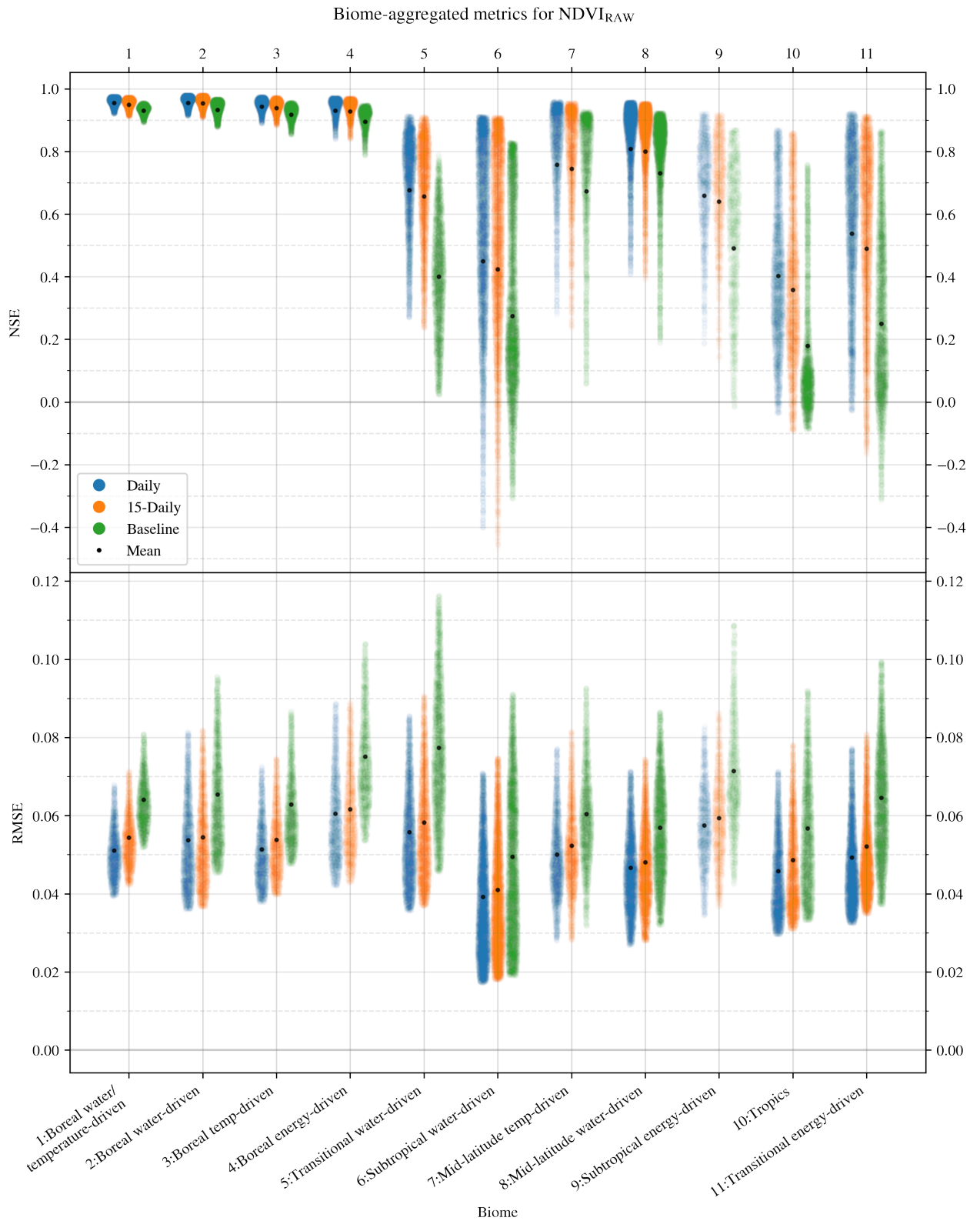
Figure 4.5: Biome-specific performance in terms of RMSE and NSE for NDVI$_{\text{RAW}}$: the violin plots show the per-biome aggregated NSE (upper graph) and RMSE (lower graph) for M$_{\text{D}}$ (blue) M$_{\text{15-D}}$ (orange) and M$_{\text{B}}$ (green) for the raw time series. The violin indicates the distribution of data points between Q_{05} and Q_{95}. The mean of the distribution is marked with a black dot.

Chapter 5

Discussion

5.1 Model Performance

The temporally and spatially pooled improvement in performance from $M_{15\text{-}D}$ to M_D for the prediction of $NDVI_{RAW}$ was 0.4% in terms of NSE and 4% in terms of RMSE. Although this improvement may seem small, the decomposition shows that especially the prediction of the anomalies with regards to the NSE show a large improvement of 37.8% and in the RMSE by 4%. Compared to M_B, there is even an improvement of 258% (NSE) and 26.9% (RMSE). Further on, the maps in figure 4.3 indicate that these NSE improvements happened especially in regions, where both models achieved a low NSE (Tropics, Australia). There is no noticeable improvement in the prediction of the $NDVI_{MSC}$ in terms of NSE, which may be due to the fact that both models already achieve very high NSE values of over 0.97. Nevertheless, the 5.7% enhancement in RMSE from $M_{15\text{-}D}$ to M_D suggests that the observed MSC approximation could be improved. These observations can be taken as an indication that the prediction of memory effects has indeed improved qualitatively.

5.2 Biome-Specific Memory Effects

The performance of both models with regard to Mem could be reproduced with similar results as seen in Kraft et al. (see figure 4.4). According to this metric, the greatest influence of memory effects was measured in subtropical and transitional ecosystems. These ecosystems comprise the arid and semi-arid regions of the world, i.e. regions with particularly limited water resources [55]. Even though plants have adapted to these water limitations their growth is still strongly dependent on soil moisture. The utilized explanatory variables, especially precipitation, water table depth, solar radiation and temperature have a great influence on soil moisture. Thus, it may be assumed that the state of soil moisture is implicitly modelled in the

memory of the LSTM. Another possible explanation for the particularly high measurement of memory effects in these regions would be, that there is little vegetation and therefore little changes in the NDVI. Wrong predictions of M_B thus produce larger deviations from the observations, which leads to erroneously recognized memory effects due to the way how memory effects are calculated. Furthermore it was observed that the variance of the predictions in terms of NSE was especially high in these biomes (5,6,9,11 in figure 4.5). Additionally, in most of these biomes (6,9,11), neither $M_{15\text{-}D}$ nor M_D is able to achieve an NSE well above 0 for the anomalies (see figure B.7). Consequently, memory effects measured in these regions should be considered with caution.

The strongest prediction improvement in memory effects of up to 70% (1) between $M_{15\text{-}D}$ and M_D were found in boreal biomes (1-4 in figure B.7). These memory effects could be related to the process of thermal acclimation of plants, which can be observed in boreal biomes. Thermal acclimation is the mechanism of adapting the optimal temperature for photosynthesis not to instantaneous temperatures but to previous temperature curves (ranging from days to weeks) [81]. Therefore, the momentary photosynthesis performance and thus the greenness of the vegetation depends on previous temperature processes. If, for example, a sudden cold spell occurs, the plants would need a few days to adapt their photosynthesis performance to the new temperature. During this time their photosynthesis performance would be reduced and the vegetation would appear less green. The higher resolution of the temperature variables of M_D may have made the model more capable of connecting the anomalies in the temperature profile with the correct prediction of the NDVI.

A further observation was, that the biome-specific model performance in terms of NSE between $NDVI_{ANO}$ and $NDVI_{MSC}$ is highly correlated with an R^2 of 81.9%: looking at figure B.7 the biggest performance gains in NSE from $M_{15\text{-}D}$ to M_D for the prediction of $NDVI_{ANO}$ were found in these regions that had the a very high NSE with a low variance for the prediction of $NDVI_{MSC}$ (1-5). A good prediction of the median seasonal cycle is thus possibly the condition for a good prediction of anomalies in biomes.

5.3 Limitations and Possible Enhancements

The general prediction capability of the models is limited by the following factors:

- **noise**: the datasets used have inherent noise and biases, which can be attributed to the sensor which collects the data, the changing observation conditions and the different methods of data processing.

- **complexity of modelled processes**: vegetation processes are very complex and depend on other abiotic factors, such as the nutrient supply and moist of the soil and biotic factors, neither of which were directly conveyed to the model.

- **autocorrelation of climatological variables**: the cross-validation strategy used has an inherent bias-variance trade-off, which means that although we counteract the problem of overfitting to residuals that cannot be explained by the input variables, the model becomes less adaptable to local characteristics and mostly learns from global memory effect pattern instead.

Furthermore, the prediction of the memory effects are limited by vegetation anomalies which are less robust to this noise than the prediction of the median seasonal cycle. Therefore the true effect of climate anomalies on vegetation anomalies is believed to be larger.

A potential improvement in modelling climate biosphere interactions could be achieved by deploying more advanced RNN architectures such as the Transformer [79]. This assumption is based on the observation that in the domain of natural language processing classical LSTMs - as the ones deployed in this study - have been replaced by Transformer models due to their superior performance [17, 63]. Since in both domains - natural language processing and biogeosciences - deal with sequential data, it is quite possible that the improvements can be transferred to biogeosciences as well (see [84]).

During the optimization phase of this work it was noted that the hyperparameter sensitivity of the deployed RNNs was not very high. Nevertheless, it should be emphasized that a proper choice of hyperparameters is important when deploying deep neural networks. The reason being is that a good hyperparameter is likely to speed up the learning process and enhance the ANNs performance. Key parameters that should be considered during any hyperparameter optimization are the optimization

algorithm, the learning rate, the batch size and the hyperparameters related to the model design such as the number of hidden layers or the width of hidden layers. In order, to increase the probability of finding a an optimal configuration during hyperparameter optimization, an attempt should be made to test a reasonably high amount of different configurations. Since this process is computationally and timely expensive it is advisable to parallelize the process by deploying ASHA.[39]. For more information on Optimization, refer to Yu et al. [87].

5.4 Conclusion

In this thesis I showed that the prediction function of M_D improved compared to M_{15-D}. The improvement for that can be attributed to the higher resolution of the data, since the model performed better on all levels during evaluation. It can therefore be assumed that the use of higher-temporal-resolution data leads to qualitatively better modeling results of ecological-climatological relationships if in theory a higher resolution can have a significant influence on the interactions to be modeled. This would also be in line with observations in the area of natural language processing, where larger text corpora improve the predictive power of models [24]. However, it should be noted that a higher data resolution is always associated with a higher computing effort. Which data resolution to choose should therefore depend on the respective targets. In addition, modeling approaches in the field of biogeosciences are always subject to fundamental limitations, due to noise in utilized datasets and the spatio-temporal cross-validation scheme, used to reduce the autocorrelation between climatological and ecological variables.

The qualitative improvement of the model also allowed better detection of memory effects in boreal regions. A possible explanation for these memory effects could be that they result from the process of thermal acclimation of plants. This could be used as a starting point for more detailed investigation on the phenomenon in boreal regions.

Chapter 6

Acknowledgments

I would like to thank Basil Kraft for his comments and assistance in implementing the modeling pipeline and data analysis. Furthermore I would like to thank Sujan Koirala, Jacob Nelson and Çağlar Küçük for the pleasant and collegial working atmosphere in the office and Uli Weber for providing and preprocessing the datasets used in this thesis. My sincere thanks also go to Martin Jung, Joachim Denzler and Markus Reichstein, who gave me the opportunity to join the institute first as an intern and later to write my bachelor thesis. My final thanks goes to my family and friends who supported me during the time of writing with valuable comment suggestions which gave me an inspiration to improve the quality and for their caring work during the time of writing.

Appendix A

Materials and Methods

A.1 Cross-Validation Scheme
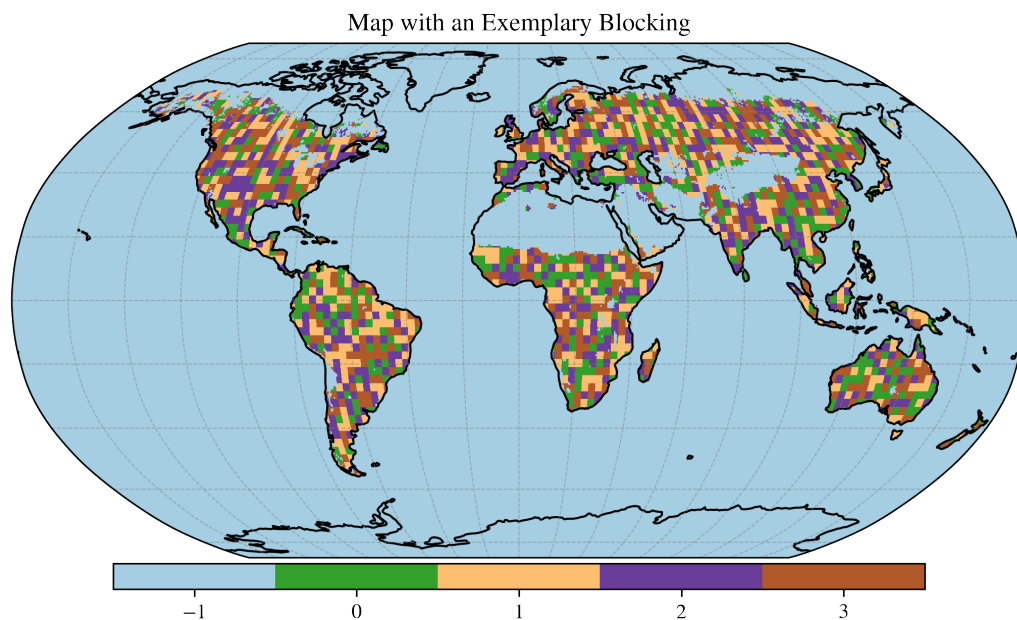
Figure A.1: Map with an exemplary blocking: this map has been utilized for implementing the spatial cross-validation. Color -1 marks places on the map that were masked and therefore not used. The colors 0 to 3 mark randomly generated blocks that were assigned to a training, validation or test set during the prediction of vegetation models.

A.2 Evaluation Metrics

Overview over Hydro-climatic Biomes
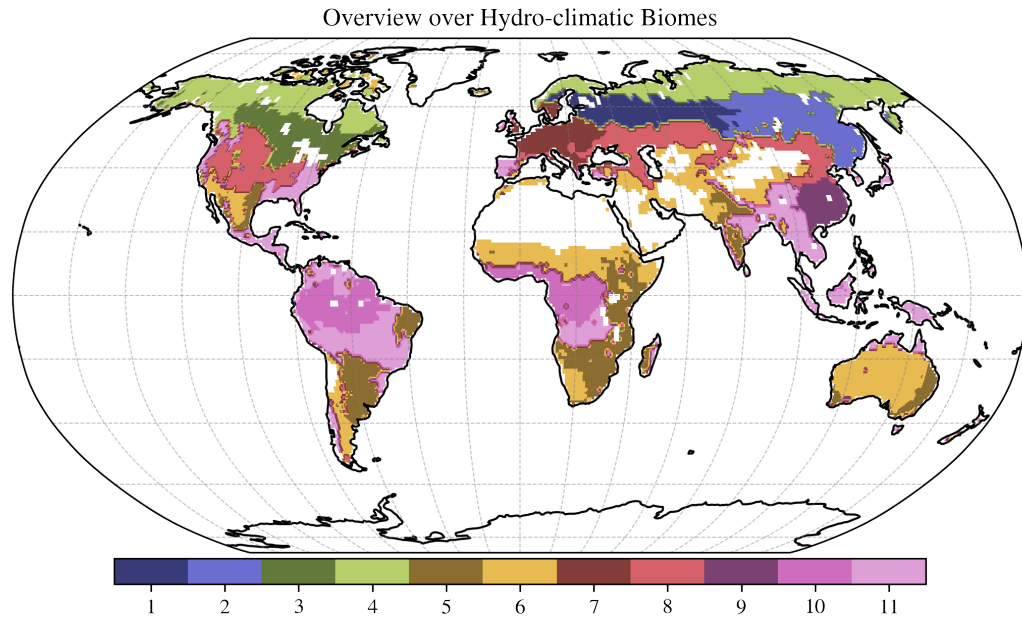
Figure A.2: Overview over hydro-climatic biomes as defined by [55]: 1:Boreal water/temperature-driven, 2:Boreal water-driven, 3:Boreal temp-driven, 4:Boreal energy-driven, 5:Transitional water-driven, 6:Subtropical water-driven, 7:Mid-latitude temp-driven, 8:Mid-latitude water-driven, 9:Subtropical energy-driven, 10:Tropics, 11:Transitional energy-driven

Appendix B

Results

B.1 Hyperparameters

RNN	Validation Loss	Learning Rate	LSTM	
	in 10^{-3}		Architecture	Hidden Units
RNN_D	4.86	4.22	LSTM-Bias	128
$\text{RNN}_\text{15-D}$	4.76	1.01	LSTM-2	256

Table B.1: Second best hyperparameter configuration of both RNNs with respect to the least validation error; determined by ASHA.
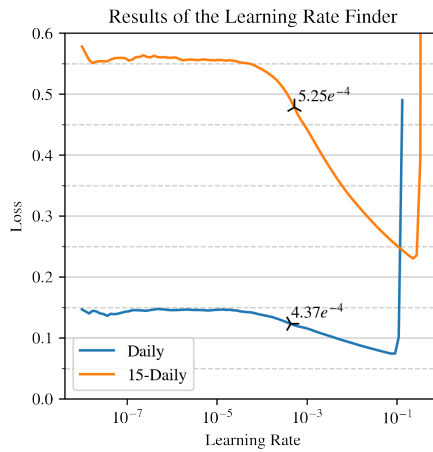
Figure B.1: Results learning rate finder: the two functions are derived from plotting learning rates against their resulting mean squared error loss for RNN_D (blue) and $\text{RNN}_\text{15-D}$ (orange). The marked points indicate the optimal learning rate that has been chosen for the respective model.

B.2 Training Progress
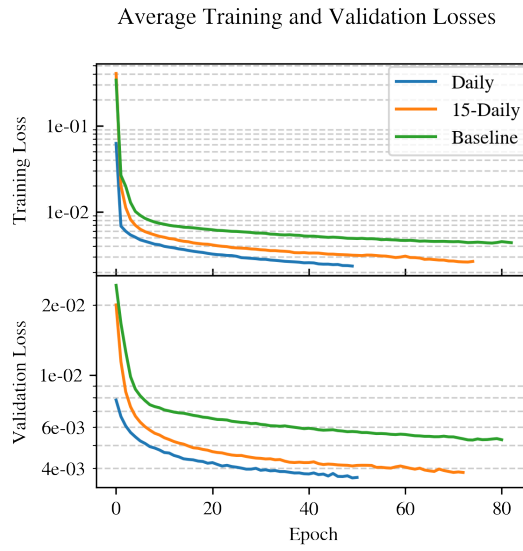
Figure B.2: Average training and validation losses: the upper [lower] graph contains a plot of the average train [average validation] loss over all runs against the number of epochs for M_D (blue), $M_{15\text{-}D}$ (orange) and M_B (green).
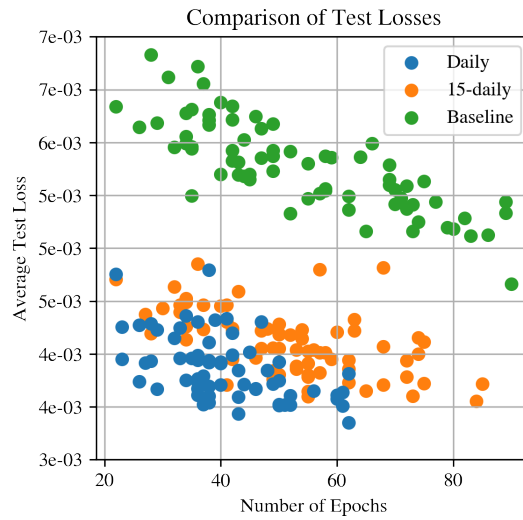
Figure B.3: Comparison of test losses: the scatter plot contains the test loss for all model runs of M_D (blue) $M_{15\text{-}D}$ (orange) and M_B (green) after finishing training at the respective number of epochs.

B.3 Evaluation

B.3.1 Time-Series Decomposition
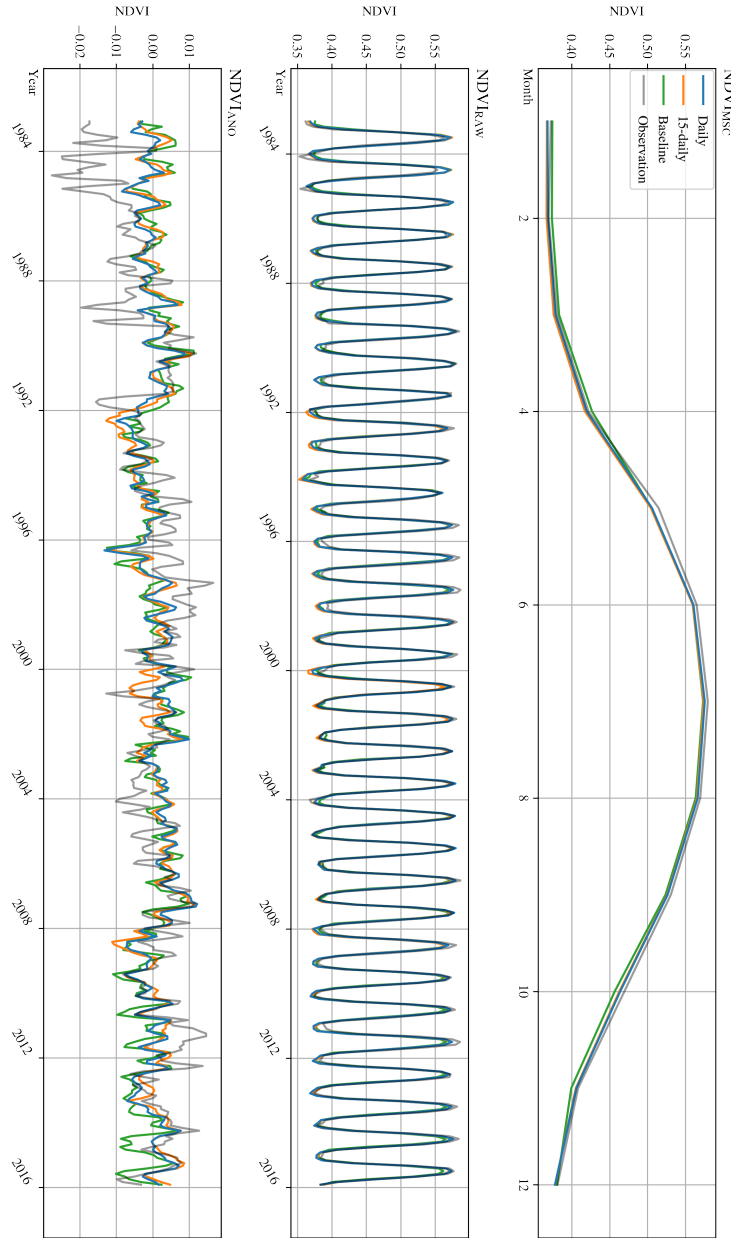
Figure B.4: Spatially-aggregated time series from 1984 to 2016: the raw time series (middle) with its median seasonal cycle (top) and its anomalies (bottom) for M_D (blue), M_{15-D} (orange), M_B (green) and the observation (grey).
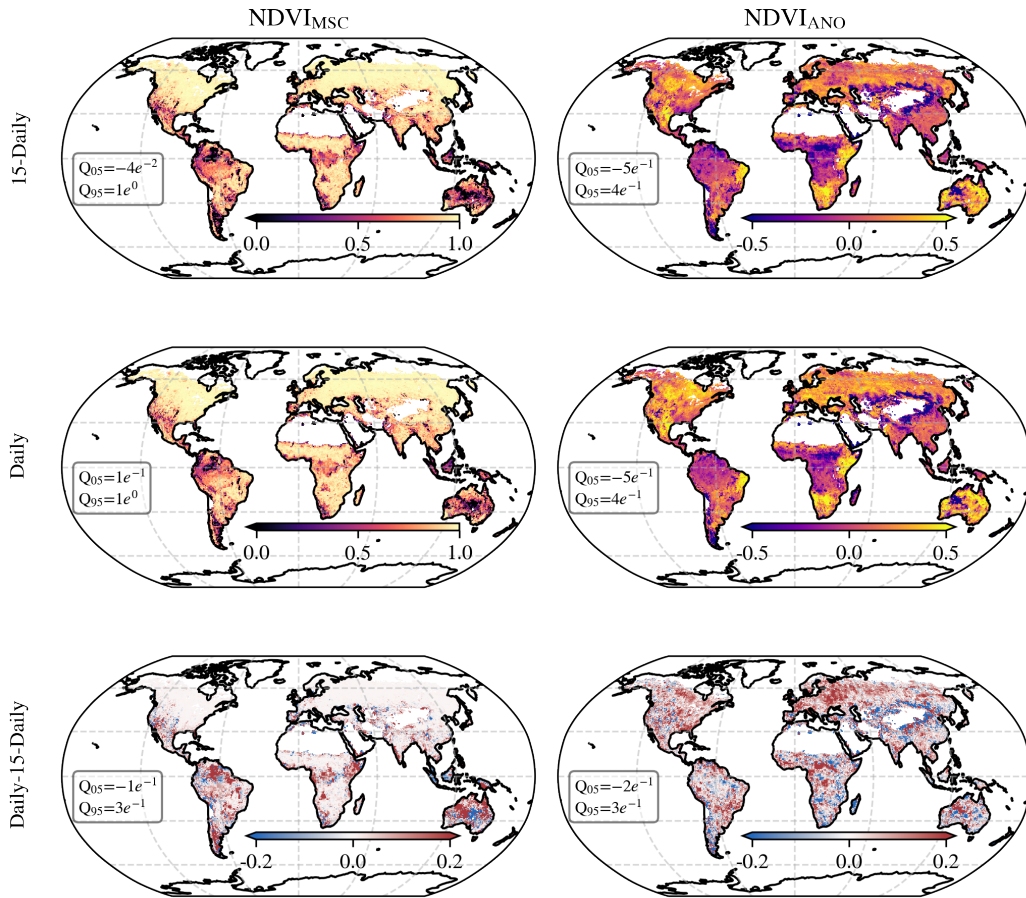
B.3.2 Regional Performance
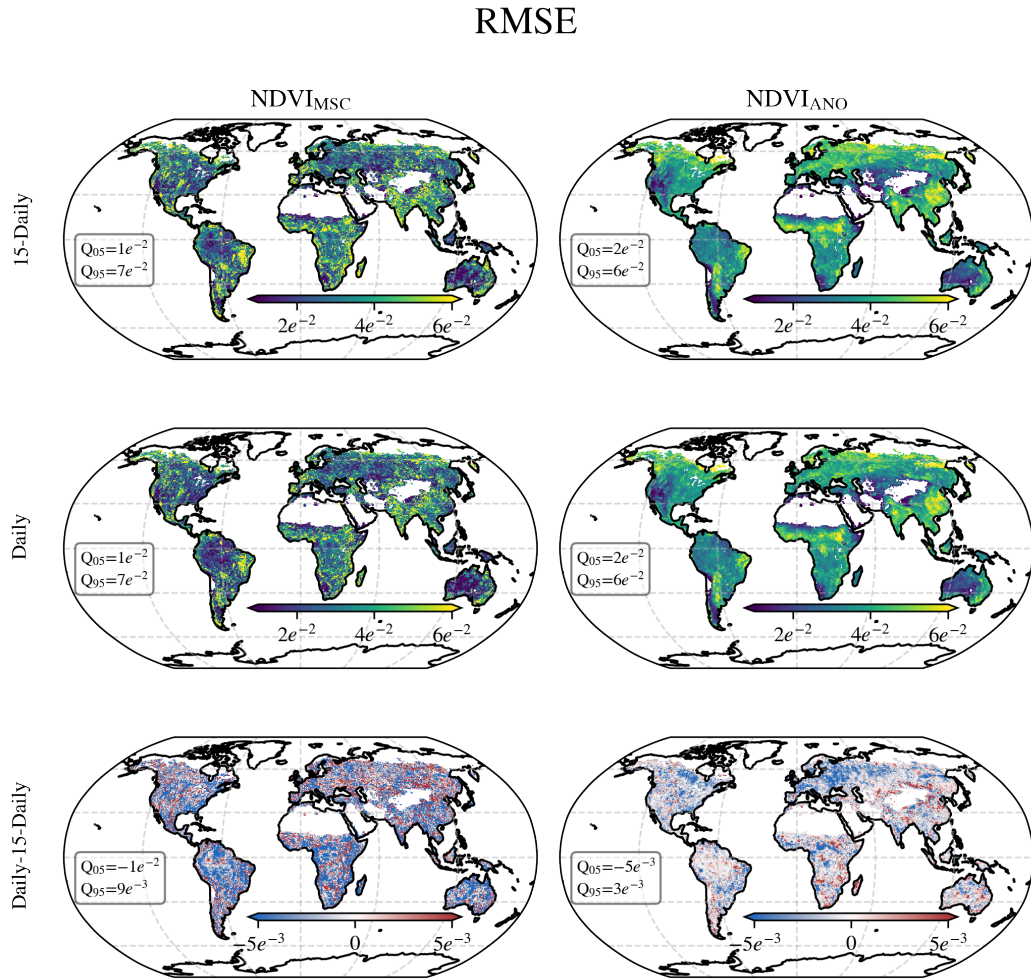
RMSE

Figure B.5: Performance maps in terms of RMSE for the MSC (left column) and the anomalies (right column): the upper four maps are anchored in the same value range between 0.01 to 0.06 to allow a visual comparison. Darker [lighter] values indicate places with a lower [higher] RMSE. The lower two maps indicate the difference between M_D and $M_{15\text{-}D}$. The color bar of both maps is centered around 0 between anchor points -0.005 and 0.005. Blue [red] indicates that the prediction of the M_D [$M_{15\text{-}D}$] is better. The boxes in the left part of the maps show the value range of the 5th and 95th quantile of the data.

NSE

Figure B.6: Performance maps in terms of NSE for the MSC (left column) and the anomalies (right column): the two left maps for NDVI$_{MSC}$ are anchored in the value range between 0 and 1. The two right maps for NDVI$_{ANO}$ are centered around 0 and anchored in the value range between -0.5 and 0.5. Lighter [darker] values indicate places with a higher [lower] NSE. Note: Only the two maps below each other can be compared visually. The lower two maps indicate the difference between M$_D$ and M$_{15\text{-}D}$. The color bar of both maps is centered around 0 between anchor points -0.2 and 0.2. Red [blue] indicates that the prediction of the M$_D$ [M$_{15\text{-}D}$] is better. The boxes in the left part of the maps show the value range of the 5th and 95th quantile of the data.

B.3.3 Biome-Specific Performance

Biome	Metrics	Models			increase in %		
		M_D	$M_{15\text{-}D}$	M_B	$M_{15\text{-}D} \rightarrow M_D$	$M_B \rightarrow M_{15\text{-}D}$	$M_B \rightarrow M_D$
1	NSE	0.96	0.95	0.93	0.6	2.7	2.1
	RMSE	0.051	0.054	0.064	−6.1	−20.3	−15.1
2	NSE	0.96	0.95	0.93	0.2	2.4	2.2
	RMSE	0.054	0.054	0.065	−1.5	−17.9	−16.7
3	NSE	0.94	0.94	0.92	0.6	2.9	2.3
	RMSE	0.051	0.054	0.063	−4.6	−18.2	−14.3
4	NSE	0.93	0.93	0.89	0.3	4.0	3.7
	RMSE	0.061	0.062	0.075	−1.9	−19.4	−17.9
5	NSE	0.68	0.66	0.4	3.1	69.0	64.0
	RMSE	0.056	0.058	0.077	−4.2	−27.8	−24.6
6	NSE	0.45	0.42	0.27	6.2	63.8	54.3
	RMSE	0.039	0.041	0.049	−4.4	−20.7	−17.1
7	NSE	0.76	0.74	0.67	1.7	12.6	10.7
	RMSE	0.05	0.052	0.06	−4.3	−17.1	−13.3
8	NSE	0.81	0.8	0.73	1.0	10.7	9.5
	RMSE	0.047	0.048	0.057	−2.9	−18.0	−15.6
9	NSE	0.66	0.64	0.49	3.0	34.4	30.6
	RMSE	0.057	0.059	0.071	−3.2	−19.5	−16.8
10	NSE	0.4	0.36	0.18	12.5	123.7	98.9
	RMSE	0.046	0.049	0.057	−5.8	−19.2	−14.3
11	NSE	0.54	0.49	0.25	9.8	115.2	96.1
	RMSE	0.049	0.052	0.065	−5.3	−23.7	−19.4

Table B.2: Per-biome aggregated model performance in terms of NSE and RMSE of models M_D, $M_{15\text{-}D}$, M_B (column 3 to 5) for $NDVI_{RAW}$. Improvement of model prediction in percent from $M_{15\text{-}D}$ to M_D and M_B to $M_{15\text{-}D}$ or M_D, respectively (column 6 to 8).
1:Boreal water/temperature-driven, 2:Boreal water-driven, 3:Boreal temp-driven, 4:Boreal energy-driven, 5:Transitional water-driven, 6:Subtropical water-driven, 7:Mid-latitude temp-driven, 8:Mid-latitude water-driven, 9:Subtropical energy-driven, 10:Tropics, 11:Transitional energy-driven
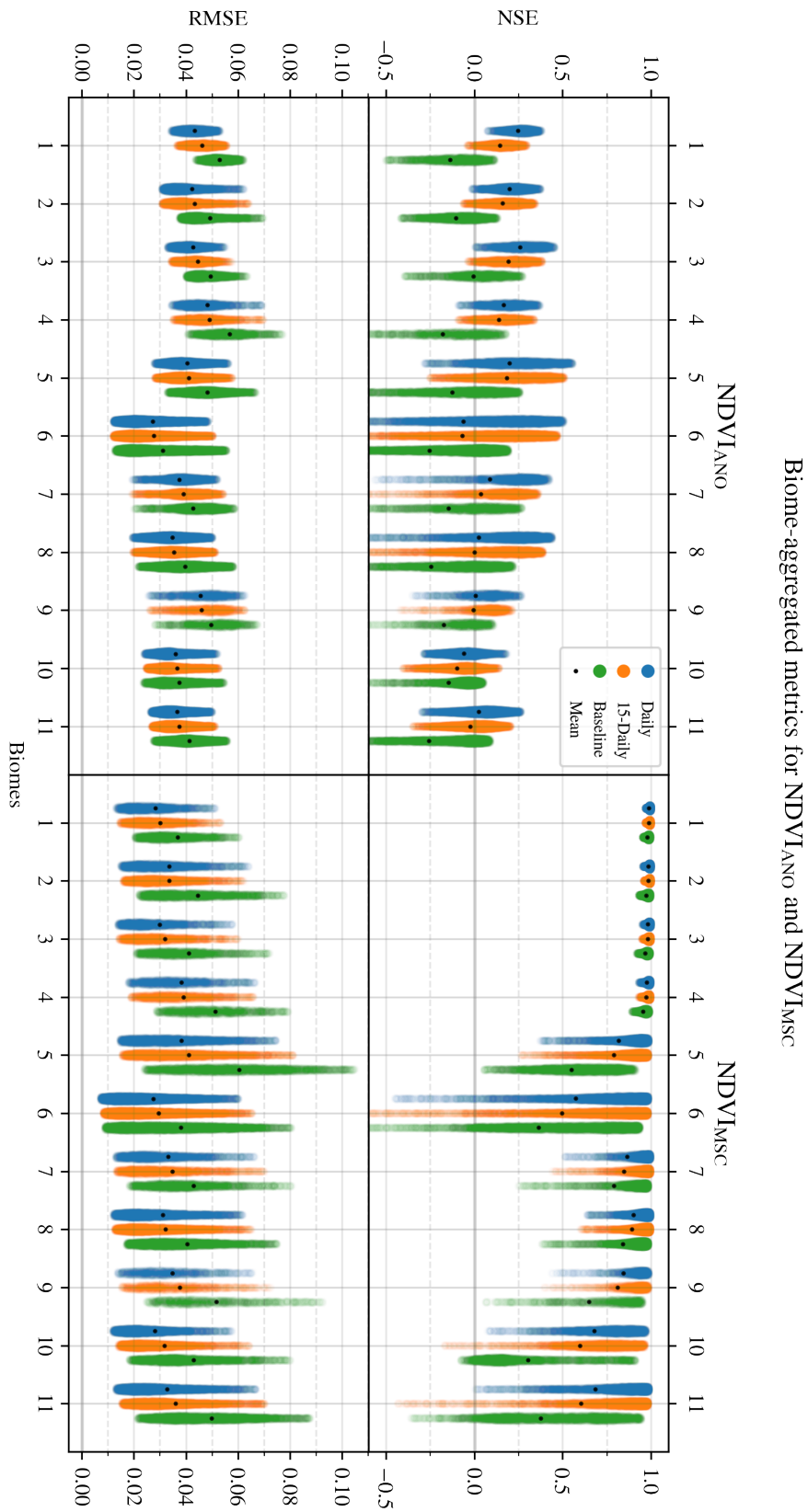
Figure B.7: Biome-specific performance for NDVI$_{\mathrm{ANO}}$ (first column) and NDVI$_{\mathrm{MSC}}$ (right column) in terms of NSE (upper graphs) and RMSE (lower graphs) for M$_{\mathrm{D}}$ (blue) M$_{\mathrm{15\text{-}D}}$ (orange) and M$_{\mathrm{B}}$ (green). The violin indicates the distribution of data points between Q_{05} and Q_{95}. The mean of the distribution is marked with a black dot.

Bibliography

[1] C. C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018. ISBN: 978-3-319-94462-3. DOI: `10.1007/978-3-319-94463-0`. URL: `https://www.springer.com/gp/book/9783319944623` (visited on 08/06/2020).

[2] J. L. Ba, J. R. Kiros, and G. E. Hinton. *Layer Normalization*. July 21, 2016. arXiv: `1607.06450 [cs, stat]`. URL: `http://arxiv.org/abs/1607.06450` (visited on 09/07/2020).

[3] R. E. Baker et al. "Mechanistic Models versus Machine Learning, a Fight Worth Fighting for the Biological Community?" In: *Biology letters* 14.5 (5 2018). DOI: `10.1098/rsbl.2017.0660`. eprint: `29769297`.

[4] H. E. Beck et al. "Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling". In: *Hydrology and Earth System Sciences* 21.12 (12 2017), pp. 6201–6217. DOI: `10.5194/hess-21-6201-2017`.

[5] H. E. Beck et al. *Global-Scale Evaluation of 23 Precipitation Datasets Using Gaugeobservations and Hydrological Modeling*. preprint. Global hydrology/Instruments and observation techniques, Aug. 14, 2017. DOI: `10.5194/hess-2017-508`. URL: `https://www.hydrol-earth-syst-sci-discuss.net/hess-2017-508/hess-2017-508.pdf` (visited on 07/13/2020).

[6] Y. Bengio, P. Simard, and P. Frasconi. "Learning Long-Term Dependencies with Gradient Descent Is Difficult". In: *IEEE Transactions on Neural Networks* 5.2 (Mar. 1994), pp. 157–166. ISSN: 1941-0093. DOI: `10.1109/72.279181`.

[7] S. Besnard et al. "Memory Effects of Climate and Vegetation Affecting Net Ecosystem CO2 Fluxes in Global Forests". In: *PLOS ONE* 14.2 (Feb. 6, 2019), e0211510. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0211510`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211510` (visited on 07/30/2020).

[8] G. B. Bonan. *Ecological Climatology: Concepts and Applications*. Third edition. New York, NY, USA: Cambridge University Press, 2016. 692 pp. ISBN: 978-1-107-04377-0 978-1-107-61905-0.

[9] C. J. Burrows. *Processes of Vegetation Change*. London: Unwin Hyman, 1990. 551 pp. ISBN: 978-0-04-580013-1 978-0-04-580012-4.

[10] G. S. Campbell and J. M. Norman. *Introduction to Environmental Biophysics*. 2nd ed. New York: Springer, 1998. 286 pp. ISBN: 978-0-387-94937-6.

[11]   Z. Chen, W. Wang, and J. Fu. "Vegetation Response to Precipitation Anomalies under Different Climatic and Biogeographical Conditions in China". In: *Scientific Reports* 10.1 (1 Jan. 21, 2020), p. 830. ISSN: 2045-2322. DOI: 10.1038/s41598-020-57910-1. URL: https://www.nature.com/articles/s41598-020-57910-1 (visited on 09/25/2020).

[12]   G. Chevalier. *The Long Short-Term Memory (LSTM)*. May 16, 2018. URL: https://commons.wikimedia.org/wiki/File:The_LSTM_cell.png (visited on 10/03/2020).

[13]   M. J. M. Christenhusz and J. W. Byng. "The Number of Known Plants Species in the World and Its Annual Increase". In: *Phytotaxa* 261.3 (3 May 20, 2016), pp. 201–217. ISSN: 1179-3163. DOI: 10.11646/phytotaxa.261.3.1. URL: https://www.biotaxa.org/Phytotaxa/article/view/phytotaxa.261.3.1 (visited on 07/24/2020).

[14]   F. E. (E. Clements. *Plant Succession; an Analysis of the Development of Vegetation*. In collab. with Cornell University Library. Washington, Carnegie Institution of Washington, 1916. 658 pp. URL: http://archive.org/details/cu31924000531818 (visited on 07/24/2020).

[15]   Coder. *Code-Server*. Coder, Aug. 8, 2020. URL: https://github.com/cdr/code-server (visited on 08/08/2020).

[16]   V. Demidchik. "Mechanisms of Oxidative Stress in Plants: From Classical Chemistry to Cell Biology". In: *Environmental and Experimental Botany* 109 (Jan. 1, 2015), pp. 212–228. ISSN: 0098-8472. DOI: 10.1016/j.envexpbot.2014.06.021. URL: http://www.sciencedirect.com/science/article/pii/S0098847214001750 (visited on 07/22/2020).

[17]   J. Devlin and M.-W. Chang. *Open Sourcing BERT: State-of-the-Art Pre-Training for Natural Language Processing*. Nov. 2, 2018. URL: http://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html (visited on 09/29/2020).

[18]   Y. Fan, H. Li, and G. Miguez-Macho. "Global Patterns of Groundwater Table Depth". In: *Science* 339.6122 (Feb. 22, 2013), pp. 940–943. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1229881. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.1229881 (visited on 08/15/2020).

[19]   FAO/IIASA/ISRIC/ISS-CAS/JRC. *Harmonized World Soil Database (Version 1.1)*. FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2009. 43 pp.

[20]   fdeloche. *A One-Unit Recurrent Neural Network (RNN)*. June 19, 2017. URL: https://commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg (visited on 10/03/2020).

[21]   M. A. Friedl et al. "MODIS Collection 5 Global Land Cover: Algorithm Re-
finements and Characterization of New Datasets". In: *Remote Sensing of En-
vironment* 114.1 (Jan. 15, 2010), pp. 168–182. ISSN: 0034-4257. DOI: 10.1016/
j.rse.2009.08.016. URL: http://www.sciencedirect.com/science/
article/pii/S0034425709002673 (visited on 07/24/2020).

[22]   I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Cambridge, Mas-
sachusetts: The MIT Press, Nov. 18, 2016. 800 pp. ISBN: 978-0-262-03561-3.

[23]   A. Graves, A.-R. Mohamed, and G. Hinton. *Speech Recognition with Deep
Recurrent Neural Networks*. Mar. 22, 2013. arXiv: 1303.5778 [cs]. URL: http:
//arxiv.org/abs/1303.5778 (visited on 09/04/2020).

[24]   A. Halevy, P. Norvig, and F. Pereira. "The Unreasonable Effectiveness of
Data". In: *IEEE INTELLIGENT SYSTEMS* (2009), p. 5.

[25]   K. He et al. *Deep Residual Learning for Image Recognition*. Dec. 10, 2015.
arXiv: 1512.03385 [cs]. URL: http://arxiv.org/abs/1512.03385 (visited
on 09/18/2020).

[26]   H. Hersbach et al. "The ERA5 Global Reanalysis". In: *Quarterly Journal of
the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. ISSN: 1477-
870X. DOI: 10.1002/qj.3803. URL: https://rmets.onlinelibrary.wiley.
com/doi/abs/10.1002/qj.3803 (visited on 08/15/2020).

[27]   S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". In: *Neural
Computation* 9.8 (Nov. 1, 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.
1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.
9.8.1735 (visited on 08/07/2020).

[28]   S. Hoyer and J. J. Hamman. "Xarray: N-D Labeled Arrays and Datasets in
Python". In: *Journal of Open Research Software* 5 (Apr. 5, 2017), p. 10. ISSN:
2049-9647. DOI: 10.5334/jors.148. URL: http://openresearchsoftware.
metajnl.com/articles/10.5334/jors.148/ (visited on 08/08/2020).

[29]   J. R. Jensen. *Introductory Digital Image Processing: A Remote Sensing Per-
spective*. 4th Edition. Pearson, Sept. 10, 2015. 656 pp.

[30]   R. Jozefowicz, W. Zaremba, and I. Sutskever. "An Empirical Exploration of
Recurrent Network Architectures". In: *International Conference on Machine
Learning*. International Conference on Machine Learning. PMLR, June 1, 2015,
pp. 2342–2350. URL: http://proceedings.mlr.press/v37/jozefowicz15.
html (visited on 09/07/2020).

[31]   W. D. Keersmaecker et al. "A Model Quantifying Global Vegetation Re-
sistance and Resilience to Short-Term Climate Anomalies and Their Rela-
tionship with Vegetation Cover". In: *Global Ecology and Biogeography* 24.5
(2015), pp. 539–548. ISSN: 1466-8238. DOI: 10.1111/geb.12279. URL: https:
//onlinelibrary.wiley.com/doi/abs/10.1111/geb.12279 (visited on
07/30/2020).

[32] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization.* Jan. 29, 2017. arXiv: 1412.6980 [cs]. URL: http://arxiv.org/abs/1412.6980 (visited on 08/10/2020).

[33] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization.* Jan. 29, 2017. arXiv: 1412.6980 [cs]. URL: http://arxiv.org/abs/1412.6980 (visited on 09/03/2020).

[34] B. Kraft et al. "Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks". In: *Frontiers in Big Data* 2 (2019). ISSN: 2624-909X. DOI: 10.3389/fdata.2019.00031. URL: https://www.frontiersin.org/articles/10.3389/fdata.2019.00031/full#h3 (visited on 08/15/2020).

[35] D. Kriesel. *A Brief Introduction to Neural Networks.* 2007. URL: available%20at%20http://www.dkriesel.com.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25.* Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (visited on 09/18/2020).

[37] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning". In: *Nature* 521.7553 (May 28, 2015), pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. pmid: 26017442.

[38] P. Legendre. "Spatial Autocorrelation: Trouble or New Paradigm?" In: *Ecology* 74.6 (1993), pp. 1659–1673. ISSN: 1939-9170. DOI: 10.2307/1939924. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1939924 (visited on 09/18/2020).

[39] L. Li. *Massively Parallel Hyperparameter Optimization.* Dec. 12, 2018. URL: https://blog.ml.cmu.edu/2018/12/12/massively-parallel-hyperparameter-optimization/ (visited on 09/29/2020).

[40] L. Li et al. *A System for Massively Parallel Hyperparameter Tuning.* Mar. 15, 2020. arXiv: 1810.05934 [cs, stat]. URL: http://arxiv.org/abs/1810.05934 (visited on 09/06/2020).

[41] LMNT. *Haste.* LMNT, Inc., Sept. 2, 2020. URL: https://github.com/lmnt-com/haste (visited on 09/06/2020).

[42] Matplotlib Developers. *Matplotlib.* Matplotlib Developers, Oct. 3, 2020. URL: https://github.com/matplotlib/matplotlib (visited on 10/03/2020).

[43] S. McCandlish et al. *An Empirical Model of Large-Batch Training.* Dec. 14, 2018. arXiv: 1812.06162 [cs, stat]. URL: http://arxiv.org/abs/1812.06162 (visited on 09/12/2020).

[44] A. J. McElrone et al. *Water Uptake and Transport in Vascular Plants.* 2013. URL: https://www.nature.com/scitable/knowledge/library/water-uptake-and-transport-in-vascular-plants-103016037/# (visited on 07/29/2020).

[45] Microsoft. *Vscode.* Microsoft, Aug. 8, 2020. URL: https://github.com/microsoft/vscode (visited on 08/08/2020).

[46] C. Monfreda, N. Ramankutty, and J. A. Foley. "Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000". In: *Global Biogeochemical Cycles* 22.1 (2008). ISSN: 1944-9224. DOI: 10.1029/2007GB002947. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007GB002947 (visited on 08/15/2020).

[47] V. Nair and G. E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning.* ICML'10. Madison, WI, USA: Omnipress, June 21, 2010, pp. 807–814. ISBN: 978-1-60558-907-7.

[48] J. E. Nash and J. V. Sutcliffe. "River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles". In: *Journal of Hydrology* 10.3 (Apr. 1, 1970), pp. 282–290. ISSN: 0022-1694. DOI: 10.1016/0022-1694(70)90255-6. URL: http://www.sciencedirect.com/science/article/pii/0022169470902556 (visited on 09/19/2020).

[49] *NDVI: Normalized Difference Vegetation Index-3rd Generation: NASA/GFSC GIMMS — NCAR - Climate Data Guide.* June 22, 2020. URL: https://climatedataguide.ucar.edu/climate-data/ndvi-normalized-difference-vegetation-index-3rd-generation-nasagfsc-gimms (visited on 06/22/2020).

[50] NumPy Developers. *NumPy.* NumPy Developers, Oct. 3, 2020. URL: https://github.com/numpy/numpy (visited on 10/03/2020).

[51] Nvidia. *PyTorch Container.* 2020. URL: https://docs.nvidia.com/deeplearning/frameworks/pdf/PyTorch-Release-Notes.pdf.

[52] K. Ogle et al. "Quantifying Ecological Memory in Plant and Ecosystem Processes". In: *Ecology Letters* 18.3 (2015), pp. 221–235. ISSN: 1461-0248. DOI: 10.1111/ele.12399. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12399 (visited on 07/30/2020).

[53] pandas Developers. *Pandas.* pandas, Oct. 3, 2020. URL: https://github.com/pandas-dev/pandas (visited on 10/03/2020).

[54] C. Papagiannopoulou et al. "A Non-Linear Granger-Causality Framework to Investigate Climate–Vegetation Dynamics". In: *Geoscientific Model Development* 10.5 (May 17, 2017), pp. 1945–1960. ISSN: 1991-959X. DOI: 10.5194/gmd-10-1945-2017. URL: https://gmd.copernicus.org/articles/10/1945/2017/ (visited on 09/10/2020).

[55] C. Papagiannopoulou et al. *Global Hydro-Climatic Biomes Identified via Multi-Task Learning*. preprint. Earth and Space Science Informatics, Apr. 25, 2018. DOI: 10.5194/gmd-2018-92. URL: https://gmd.copernicus.org/preprints/gmd-2018-92/gmd-2018-92.pdf (visited on 10/04/2020).

[56] G. Petneházi. *Recurrent Neural Networks for Time Series Forecasting*. Dec. 31, 2018. arXiv: 1901.00069 [cs, stat]. URL: http://arxiv.org/abs/1901.00069 (visited on 08/08/2020).

[57] J. E. Pinzon and C. J. Tucker. "A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series". In: *Remote Sensing* 6.8 (8 Aug. 2014), pp. 6929–6960. DOI: 10.3390/rs6086929. URL: https://www.mdpi.com/2072-4292/6/8/6929 (visited on 07/19/2020).

[58] Project Jupyter. *Jupyterlab*. JupyterLab, Sept. 6, 2020. URL: https://github.com/jupyterlab/jupyterlab (visited on 09/06/2020).

[59] pydata. *Xarray*. Python for Data, Sept. 13, 2020. URL: https://github.com/pydata/xarray (visited on 09/13/2020).

[60] Python Software Foundation. *Python Language Reference, Version 3.6*. URL: available%20at%20https://www.python.org/ (visited on 08/08/2020).

[61] PyTorch. *PyTorch Documentation*. 2020. URL: available%20at%20https://www.pytorch.org (visited on 08/10/2020).

[62] PyTorchLightning. *Pytorch-Lightning*. Pytorch Lightning, Aug. 10, 2020. URL: https://github.com/PyTorchLightning/pytorch-lightning (visited on 08/10/2020).

[63] A. Radford et al. "Language Models Are Unsupervised Multitask Learners". In: (2019), p. 24.

[64] ray-project. *Ray*. ray-project, Sept. 6, 2020. URL: https://github.com/ray-project/ray (visited on 09/06/2020).

[65] M. Reichstein et al. "Deep Learning and Process Understanding for Data-Driven Earth System Science". In: *Nature* 566.7743 (7743 2019), pp. 195–204. DOI: 10.1038/s41586-019-0912-1. eprint: 30760912.

[66] A. Ritter and R. Muñoz-Carpena. "Performance Evaluation of Hydrological Models: Statistical Significance for Reducing Subjectivity in Goodness-of-Fit Assessments". In: *Journal of Hydrology* 480 (Feb. 14, 2013), pp. 33–45. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2012.12.004. URL: http://www.sciencedirect.com/science/article/pii/S0022169412010608 (visited on 09/19/2020).

[67] D. R. Roberts et al. "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure". In: *Ecography* 40.8 (2017), pp. 913–929. ISSN: 1600-0587. DOI: 10.1111/ecog.02881. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02881 (visited on 09/18/2020).

[68] R. V. Rohli and A. J. Vega. *Climatology*. 4 edition. Burlington, Massachusetts: Jones & Bartlett Learning, June 22, 2017. 418 pp. ISBN: 978-1-284-11998-5.

[69] J. W. Rouse Jr. et al. "Monitoring Vegetation Systems in the Great Plains with Erts". In: *NASA Special Publication* 351 (1974), p. 309. URL: http://adsabs.harvard.edu/abs/1974NASSP.351..309R (visited on 09/14/2020).

[70] J. Schmidhuber. "Deep Learning in Neural Networks: An Overview". In: *Neural Networks* 61 (Jan. 2015), pp. 85–117. ISSN: 08936080. DOI: 10.1016/j.neunet.2014.09.003. arXiv: 1404.7828. URL: http://arxiv.org/abs/1404.7828 (visited on 08/05/2020).

[71] E.-D. Schulze, E. Beck, and K. Müller-Hohenstein. *Plant Ecology*. Springer Science & Business Media, Feb. 18, 2005. 716 pp. ISBN: 978-3-540-20833-4. Google Books: rDo8hLWtWzgC.

[72] A. W. R. Seddon et al. "Sensitivity of Global Terrestrial Ecosystems to Climate Variability". In: *Nature* 531.7593 (7593 Mar. 2016), pp. 229–232. ISSN: 1476-4687. DOI: 10.1038/nature16986. URL: https://www.nature.com/articles/nature16986 (visited on 09/10/2020).

[73] T. F. Shaxson and R. G. Barber. *Optimizing Soil Moisture for Plant Production: The Significance of Soil Porosity*. In collab. with F. a. A. O. of the United Nations. FAO Soils Bulletin 79. Rome: Food and Agriculture Organization of the United Nations, 2003. 107 pp. ISBN: 978-92-5-104944-0.

[74] L. N. Smith. *Cyclical Learning Rates for Training Neural Networks*. Apr. 4, 2017. arXiv: 1506.01186 [cs]. URL: http://arxiv.org/abs/1506.01186 (visited on 09/06/2020).

[75] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html (visited on 08/10/2020).

[76] TensorFlow. *TensorBoard*. tensorflow, Sept. 14, 2020. URL: https://github.com/tensorflow/tensorboard (visited on 09/14/2020).

[77] "AVHRR NDVI3g". In: (2014). Ed. by C. Tucker and U. Weber.

[78] A. M. VanDerZanden. *Environmental Factors Affecting Plant Growth — OSU Extension Service*. Jan. 2008. URL: https://extension.oregonstate.edu/gardening/techniques/environmental-factors-affecting-plant-growth (visited on 07/22/2020).

[79] A. Vaswani et al. *Attention Is All You Need*. Dec. 5, 2017. arXiv: 1706.03762 [cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 09/29/2020).

[80] M. Venkatachalam. *Recurrent Neural Networks*. 2019. URL: https://towardsdatascience.com/recurrent-neural-networks-d4642c9bc7ce (visited on 10/09/2019).

[81] D. A. Way and W. Yamori. "Thermal Acclimation of Photosynthesis: On the Importance of Adjusting Our Definitions and Accounting for Thermal Acclimation of Respiration". In: *Photosynthesis Research* 119.1 (Feb. 1, 2014), pp. 89–100. ISSN: 1573-5079. DOI: 10.1007/s11120-013-9873-7. URL: https://doi.org/10.1007/s11120-013-9873-7 (visited on 09/29/2020).

[82] Weier, John, Herring, David. *Measuring Vegetation (NDVI & EVI)*. Aug. 30, 2000. URL: https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_1.php (visited on 10/08/2019).

[83] D. Wu et al. "Time-Lag Effects of Global Vegetation Responses to Climate Change". In: *Global Change Biology* 21.9 (2015), pp. 3520–3531. ISSN: 1365-2486. DOI: 10.1111/gcb.12945. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12945 (visited on 09/10/2020).

[84] N. Wu et al. *Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case*. Jan. 22, 2020. arXiv: 2001.08317 [cs, stat]. URL: http://arxiv.org/abs/2001.08317 (visited on 09/30/2020).

[85] J. Xu et al. *Understanding and Improving Layer Normalization*. Nov. 16, 2019. arXiv: 1911.07013 [cs, stat]. URL: http://arxiv.org/abs/1911.07013 (visited on 09/07/2020).

[86] W. Yin et al. *Comparative Study of CNN and RNN for Natural Language Processing*. Feb. 7, 2017. arXiv: 1702.01923 [cs]. URL: http://arxiv.org/abs/1702.01923 (visited on 10/05/2020).

[87] T. Yu and H. Zhu. *Hyper-Parameter Optimization: A Review of Algorithms and Applications*. Mar. 12, 2020. arXiv: 2003.05689 [cs, stat]. URL: http://arxiv.org/abs/2003.05689 (visited on 09/30/2020).

List of Figures

List of Tables