

---

**Applied Data Science master thesis**

**Modelling Vegetation Time-Series in South Africa with LSTM  
networks — A Study Towards Understanding Drought Impacts**



**First examiner:**

Dr. ir. Niko Wanders

**Candidate:**

Daan van der Hoek

**Second examiner:**

Steye Verhoeve MSc.

**Student number:**

0139483

**In cooperation with:**

Paddy O'Brien

July 4, 2025

**Title page image:** (PxHere contributors, 2017)

## Abstract

Droughts pose a serious threat to ecosystems and human well-being by damaging vegetation, their frequency and severity are expected to increase under climate change. To build resilience, it is essential to monitor and anticipate vegetation dynamics using reliable models.

This study explores whether vegetation in Southern Africa, captured through the Enhanced Vegetation Index (EVI), can be effectively modeled using Long Short-Term Memory (LSTM) networks and a compact set of environmental covariates. Given the growing availability of remote sensing data for both vegetation and climatic drivers (such as precipitation, temperature, and drought indices), LSTMs provide a promising approach due to their ability to capture temporal dependencies.

Here we show that the LSTM achieved strong predictive performance with an average  $R^2$  of 0.87 on the test set. Feature ablation and isolation analyses demonstrated that precipitation had the strongest direct signal. However, overall, the features exhibited substantial overlap in the information they carried useful for predicting vegetation dynamics. A spatial correlation between elevation and the importance of short-term drought (SPEI-1) indicates that vegetation response to drought may vary with altitude.

These findings are consistent with prior studies (e.g., Jin et al. (2020)), which found strong associations between EVI and climatic variables. The LSTM's ability to generalize across regions with varying dynamics is particularly promising for ecological forecasting.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Method</b>	<b>5</b>
2.1	Data Preparation . . . . .	5
2.2	Experimental setup . . . . .	8
2.3	Model Evaluation . . . . .	10
2.4	Hyperparameter Tuning . . . . .	11
2.5	Model Interpretation . . . . .	13
<b>3</b>	<b>Results</b>	<b>14</b>
3.1	Model Performance . . . . .	14
3.2	Hyperparameter Tuning . . . . .	17
3.3	Model Training . . . . .	18
3.4	Model Interpretation . . . . .	19
<b>4</b>	<b>Discussion</b>	<b>22</b>
<b>5</b>	<b>Conclusion</b>	<b>24</b>
	<b>Bibliography</b>	<b>26</b>

# 1. Introduction

As climate variability intensifies, understanding the impacts of environmental stressors like droughts becomes ever more critical. Droughts are prolonged periods in which precipitation levels are significantly below average, resulting in an imbalance between water input and water loss in a given area (Wilhite & Glantz, 1985). In addition to a lack of rainfall, droughts are also characterized by increased evaporation, causing further loss of water to the atmosphere (Allen et al., 1998). This combination of decreased precipitation and increased evaporation leads to a persistent moisture deficit (Vicente-Serrano et al., 2010). During droughts, decreased soil moisture can put severe stress on vegetation, sometimes resulting in plant death. Recovery can take months to years, depending on the severity of the drought and the resilience of the ecosystem (Anderegg et al., 2015). Droughts can severely damage ecosystems and negatively impact agricultural yields, thereby threatening food security (Lobell et al., 2011). Furthermore, due to climate change, droughts are expected to occur more frequently and with greater intensity (IPCC, 2023).

This makes it increasingly important to improve adaptive capacity, with the aim of better managing drought impacts and supporting vegetation recovery. To strengthen resilience to environmental extremes, it is essential to monitor and anticipate changes in vegetation, particularly during and after drought periods.

In order to examine vegetation conditions, metrics that can quantify vegetation abundance are highly useful. For this purpose, several vegetation indices have been developed. These are obtained through remote sensing, a technique in which satellites scan the Earth across a range of electromagnetic frequencies to retrieve surface information (Lillesand et al., 2015). The most widely used index is the Normalized Difference Vegetation Index (NDVI), based on the reflective properties of vegetation in the red and near-infrared bands (Rouse et al., 1974). While effective, NDVI has known limitations: it tends to saturate in areas with very dense vegetation and loses accuracy in regions with sparse vegetation. An improved alternative is the Enhanced Vegetation Index (EVI), which reduces sensitivity to atmospheric noise and improves performance in both dense and sparse vegetation regions by incorporating additional spectral bands (Huete et al., 2002). By capturing the temporal dynamics of EVI, it becomes possible to better understand and predict vegetation responses to drought. Therefore, a robust model that can accurately capture these patterns over time is highly valuable. To effectively model the dynamics of EVI, a diverse set of predictor variables is needed. Jin et al. (2020) conducted a study exploring the relationship between EVI and several topographic, climatic, and soil-related variables. They reported strong correlations with temperature, precipitation, and climatic water balance (precipitation minus potential evaporation), and emphasized the importance of interactions between these time-varying variables and static landscape features.

Beyond data availability, choosing a suitable model is essential for accurately capturing the long-term temporal patterns inherent in vegetation dynamics. Historically, process-based

(physical) models have been the standard approach for spatio-temporal modelling. However, advances in machine learning and growing computational power have enabled more flexible data-driven alternatives. In particular, Long Short-Term Memory (LSTM) networks have shown strong performance in learning complex, non-linear temporal patterns. LSTMs build upon the structure of Recurrent Neural Networks (RNNs), which are designed to capture dependencies in time series data. Traditional RNNs often struggle with long-term dependencies due to vanishing gradients, but LSTMs mitigate this by incorporating mechanisms that allow them to selectively retain or discard information over time (Hochreiter & Schmidhuber, 1997). Kratzert et al. (2018) demonstrated the practical value of LSTM models in a hydrological context, by applying them to the rainfall-runoff problem. Their goal was to predict streamflow at the outlet of river basins using meteorological inputs. Remarkably, even without hyperparameter tuning, their LSTM model achieved performance comparable to process-based models and remained robust under extreme conditions.

While improving prediction accuracy is a key goal, interpretability is equally important. For policymakers and ecologists, it is crucial to understand which variables most strongly drive vegetation dynamics. Gaining insight into the underlying drivers of EVI changes helps inform environmental management strategies and enhances scientific understanding. Moreover, the importance of predictor variables can vary across environmental gradients. For instance, in high-altitude areas, elevation or slope may dominate, whereas in drylands, variables related to water availability might be more influential. Understanding these spatial variations in feature importance can provide deeper insight into ecosystem functioning.

To address these questions, recent machine learning research highlights techniques such as feature ablation and SHAP values as valuable tools for interpreting complex models. Ablation importance evaluates the impact of completely removing a feature and retraining the model, offering insight into the model's reliance on specific inputs. SHAP (SHapley Additive exPlanations) values, on the other hand, provide both local and global attributions by quantifying the marginal contribution of each feature to the model's output, and have been successfully applied to deep learning models such as convolutional neural networks (Molnar, 2025; Zhao et al., 2020).

Building on these advancements, the objective of this thesis is: (1) to investigate whether the dynamics of EVI and possible climatic, topographical and soil-related predictors of EVI can be captured in an LSTM model, with the aim of predicting EVI over time; and (2) to use this model to identify the main drivers of EVI and determine whether these differ across geographical regions. Due to limited data availability and computational resources, this study focuses on the geographic region of South Africa and parts of adjacent countries, using data from 2002 to 2022.

The remainder of this thesis is organized as follows: Chapter 2 details the methodology employed and the data is introduced. Chapter 3 presents the results which are discussed in Chapter 4. Finally, Chapter 5 offers the conclusions drawn from the research.

## 2. Method

This chapter starts with the description and the preparation of the data in section 2.1, after which the experimental setup is explained (2.2), containing the splitting of the data, model architecture and the training of the model. The model evaluation procedure is explained in section 2.3, after which the hyperparameter tuning process explained (2.4), and lastly the interpretation of the model will be explained (2.5). The code used along with the data is available in a published GitHub repository (van der Hoek, 2025).

### 2.1 Data Preparation

This study makes use of a range of environmental variables over South Africa, obtained through remote sensing and institutional sources.

#### 2.1.1 Data Description

To capture the spatial dimension of the data, the study area was treated as a grid of pixels, each representing a square region with a resolution of  $0.1^\circ$  latitude/longitude (approximately 10 km at the equator). Although the original datasets came at varying spatial resolutions, all were regridded to this common grid. After removing sea pixels, the final dataset consists of 17,647 land-based pixels.

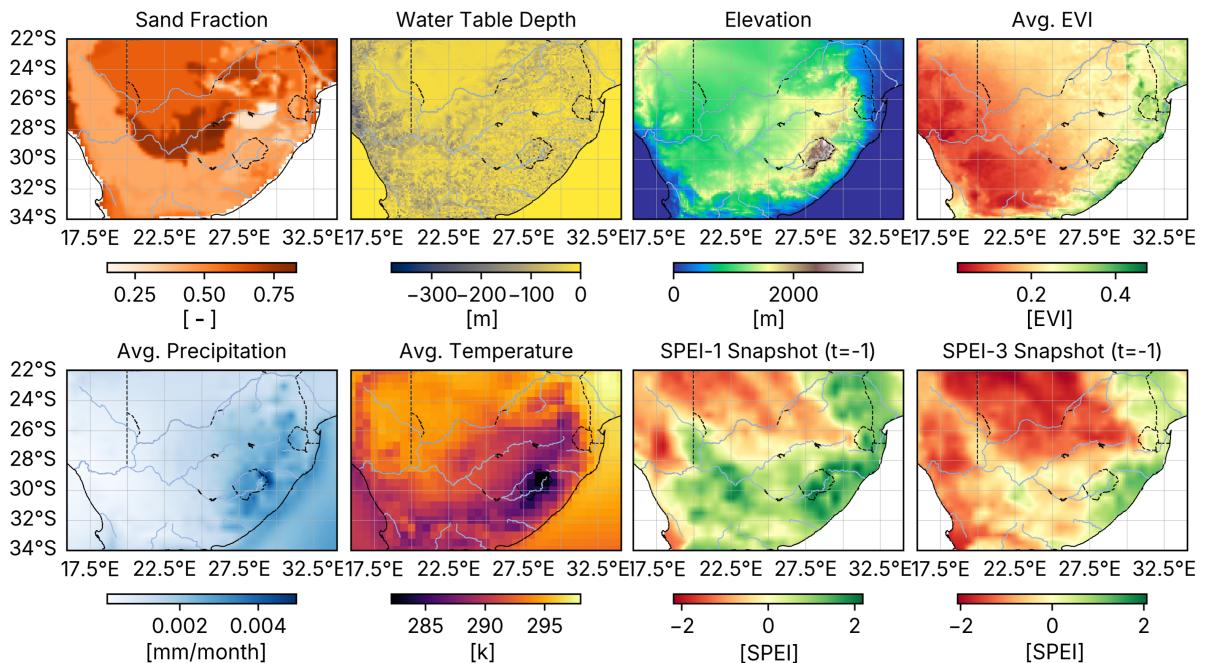
For the entire grid, a set of variables was compiled and classified into two groups: static variables, which do not change over time, and dynamic variables, which are time-dependent and recorded monthly from February 2000 to December 2022, yielding a total of 275 time-steps. Below is a brief description of each variable along with the source:

- **EVI (dynamic):** The Enhanced Vegetation Index (EVI), derived from MODIS data by Didan (2015), serves as the target variable in this study. It quantifies vegetation density and productivity, with values near 0 indicating little or no vegetation, values around 0.25 corresponding to sparse vegetation such as grasslands, and values approaching 0.9 representing dense vegetation like tropical rainforests.
- **Precipitation (dynamic):** Total monthly precipitation measured in millimeters, obtained from the ERA5 reanalysis dataset (Hersbach et al., 2020).
- **Temperature (dynamic):** Average monthly 2-meter air temperature in Kelvin, also sourced from ERA5 (Hersbach et al., 2020). This variable was regridded from a lower resolution.
- **SPEI-1 (dynamic):** The Standardized Precipitation Evaporation Index over a 1-month scale, which reflects short-term anomalies in water balance and is standardized per pixel over time (Ruijsch et al., 2025).
- **SPEI-3 (dynamic):** The 3-month average SPEI, capturing cumulative drought or wetness

over the previous quarter (Ruijsch et al., 2025).

- **Elevation (static)**: Surface elevation in meters above sea level, based on the GMTED2010 dataset by Danielson and Gesch (2011).
- **Sand Fraction (static)**: The proportion of sand in the soil, expressed as a fraction between 0 and 1, obtained from the HWSD dataset by Reynolds et al. (2000).
- **WTD (static)**: Water Table Depth (WTD), the average depth of groundwater from the surface, derived from the global dataset of Fan et al. (2017).

To illustrate the spatial distribution of these variables, Figure 2.1 presents a visual overview. Dynamic variables such as EVI, precipitation, and temperature are shown as long-term means, while SPEI-1 and SPEI-3 are represented using values from the final timestep.



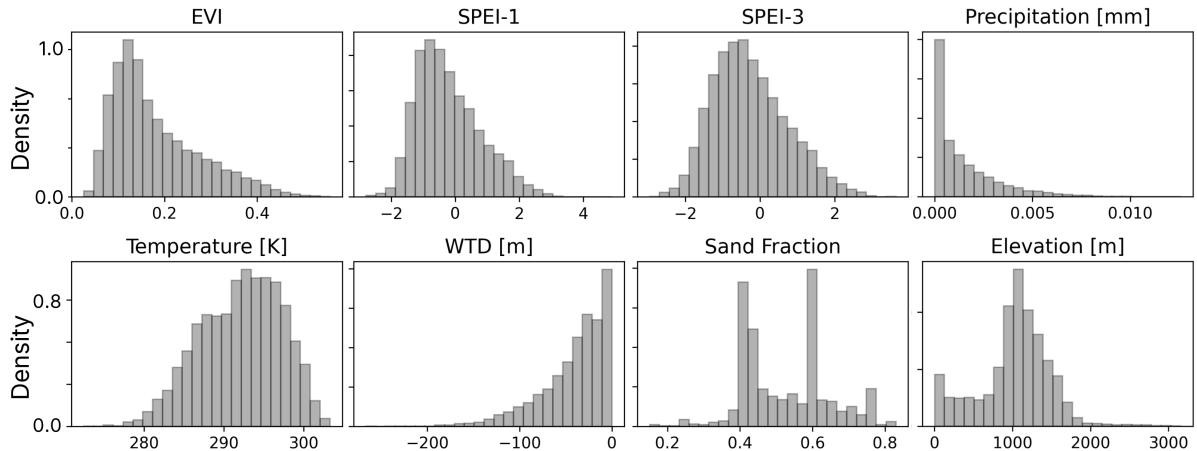
**Figure 2.1:** Overview of the variables used in this study. EVI, precipitation, and temperature are shown as long-term means. SPEI-1 and SPEI-3 are presented as snapshots from the final time-step.

### 2.1.2 Data Preprocessing

To prepare the data for, several preprocessing steps were applied to ensure that the input features were suitable for modelling.

Since the target variable, *EVI*, is continuous, the model employs a linear output activation function. Therefore, for stable training, it is beneficial that all variables are approximately normally distributed, centered (mean = 0), and with a standard deviation of 1.

An initial inspection of the raw variable distributions (Figure 2.2) revealed that EVI, SPEI, and temperature were already approximately normally distributed. In contrast, precipitation, WTD, sand fraction, and elevation have more skewed or complex distributions. In particular, precipitation and WTD showed highly skewed patterns that require non-linear transformations.



**Figure 2.2:** Raw distributions of the variables.

To address this, the following preprocessing steps were applied. First, precipitation and WTD were transformed to reduce skewness:

$$x'_{i,j} = -\log(1 + x_{i,j}) \quad (\text{WTD}) \quad (2.1)$$

$$x'_{i,j,t} = \log(1 + 10,000 \cdot x_{i,j,t}) \quad (\text{Precipitation}) \quad (2.2)$$

Next, the dynamic variables precipitation, temperature, and EVI were standardized *per month*, in order to remove seasonal effects and ensure zero mean and standard deviation equalling one:

$$z_{i,j,t} = \frac{x_{i,j,t} - \mu_{\text{month}(i,j)}}{\sigma_{\text{month}(i,j)}} \quad (2.3)$$

SPEI was already sufficiently normally distributed and did not require further transformation.

Finally, the static variables WTD, elevation, and sand fraction were standardized with Z-score normalization:

$$z_{i,j} = \frac{x_{i,j} - \mu}{\sigma} \quad (2.4)$$

From this point on, EVI, will be referred to as the standardized version.

### 2.1.3 Feature correlations

To assess the relationship between the variables in the data before modelling, Pearson's correlation coefficients were calculated between every pair of variables. The resulting correlation matrix is shown in Table 2.1. The highest correlation with the target variable EVI is SPEI-3 with 0.427, suggesting a higher SPEI-3 results in a higher EVI. The same applied for SPEI-1 and precipitation, but with lower correlations of 0.256 and 0.212 respectively. Temperature has a negative correlation with EVI, suggesting that a higher temperature indicated a lower EVI. The values between the covariates all relatively high with absolute correlations of at least 0.366 up to 0.655.

**Table 2.1:** Correlation matrix of dynamic variables and the target variable.

	SPEI-3	SPEI-1	Prec.	Temp.	EVI
<b>SPEI-3</b>		0.655	0.504	-0.466	0.427
<b>SPEI-1</b>	0.655		0.845	-0.503	0.256
<b>Prec.</b>	0.504	0.845		-0.366	0.212
<b>Temp.</b>	-0.466	-0.503	-0.366		-0.237
<b>EVI</b>	0.427	0.256	0.212	-0.237	

## 2.2 Experimental setup

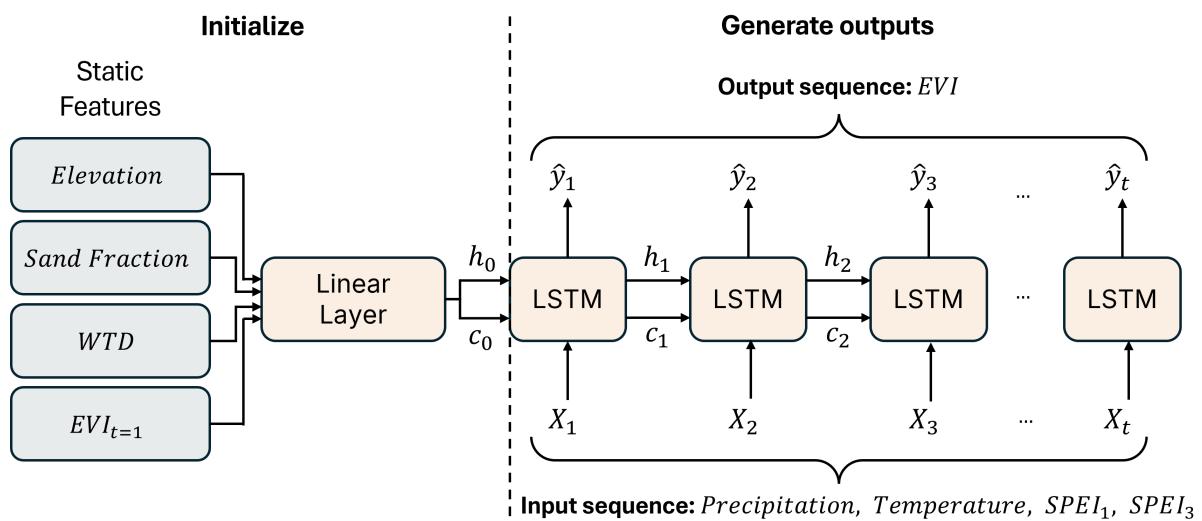
### 2.2.1 Data Splitting

Due to the large availability of pixels (17k) and the limited availability of timesamples (275), a spatial data split was chosen for this study. The full dataset was split up into a train, validation, and test dataset. Where the validation dataset is used for early stopping and hyperparameter tuning and the test set is the final unseen data for final evaluation. The 3-way split was done random. The influence of split ratio on performance will be assessed in the hyperparameter tuning section.

### 2.2.2 Model Architecture

The goal of the LSTM model is to predict the full sequence of EVI values of a location over time, given the static variables and the full sequence of dynamic variables. The LSTM is a type of Recurrent Neural Network (RNN), this means time-series inputs are fed into the model sequentially, resulting in a predicted sequence. What makes the network recurrent, is that the state of the model before accepting a new input, changes based on the previous inputs. The LSTM has a more sophisticated layout with a hidden state, cell state, and several gates, allowing it to maintain a long-term memory. A full explanation and mathematical foundation of the LSTM is provided by Van Houdt et al. (2020). The LSTM hidden state is often initialized with zero or random values. This creates problems for the first predictions, since often the hidden state needs several timestamps to activate, causing decreased performance for the first couple of timestamps. To address this problem, the static features are used to initialize the hidden state

of the LSTM. The goal is to incorporate four static features (elevation, sand fraction, water table depth, and the first EVI value) into the initial hidden state. The first EVI value is included to provide additional context at the start of the sequence. These features are passed through a linear layer, which serves as a simple neural network, where every slot in the hidden state gets a weight for every static feature and a bias, to map the static features to the initial hidden state. This approach is intended to mitigate the LSTM's warm-up phase. After initialization, the prediction process begins: at each time step, the dynamic features (precipitation, temperature, SPEI-1, and SPEI-3) are put into the LSTM. The model predicts the value for the next time-step and updates the hidden and cell states, which are carried forward to the following step. This process continues until the entire sequence has been processed. An overview of the complete model architecture is shown in Figure 2.3.



**Figure 2.3:** LSTM architecture for EVI prediction: static features initialize the hidden state via a linear layer, then dynamic inputs are fed sequentially to generate the EVI time series. The LSTM block can consist of multiple stacked LSTM cells.

### 2.2.3 Training

In order to let the model learn from the data, data is fed through the model, after which the error (loss) is calculated, and back-propagated to update model parameters (weights) based on the loss. A maximum of 120 epochs was imposed for computational reasons, with early stopping (patience = 10 epochs) applied to the validation loss. Training used the Adam optimizer and mean squared error (MSE) as the loss function:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (2.5)$$

A dynamic batch-size scheme was employed: a batch size of 1 corresponded to a single complete sequence (its four static features plus all time-step inputs), while larger batch sizes grouped multiple independent sequences per weight update. The optimal batch size will be

obtained through hyperparameter tuning. A warm-up period of the first  $n$  time steps was excluded from loss computation and weight updates to allow the hidden state to stabilize before training signals were applied. Applying a loss penalty to the first few predictions during training may be ineffective, as the LSTM requires sufficient historical context to produce accurate outputs. The length of this warm-up was treated as a tunable hyperparameter.

## 2.3 Model Evaluation

After training of the model, performance will be assessed by feeding the static and dynamic variables of the unseen data through the model and comparing the outputs with the actual EVI values. Model performance will be evaluated using multiple metrics, and by analysing both spatially and temporal error distributions, to assess how well it captures variation across space and time.

### 2.3.1 Evaluation Metrics

Model performance was assessed using the following metrics:

- **Mean Squared Error (MSE):** Retained as the training loss, MSE (eq. 2.5) penalizes larger errors more heavily but is expressed in squared units of EVI, which can reduce interpretability.
- **Root Mean Squared Error (RMSE):** By taking the square root of MSE, RMSE returns the error to the original EVI units, facilitating direct comparison with observed values:

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (2.6)$$

- **Mean Absolute Error (MAE):** MAE measures the average magnitude of errors without squaring, making it less sensitive to outliers, but less penalizing for big errors:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| \quad (2.7)$$

- **Coefficient of Determination ( $R^2$ ) and Explained Variance (EV):** Both metrics range from 0 to 1 and quantify the fraction of variance in EVI accounted for by the model (0 = predicting the mean, 1 = perfect fit). Explained Variance does not penalize systematic bias, whereas  $R^2$  does; a substantial gap between  $R^2$  and EV therefore indicates consistent over- or under-prediction. Both values are independent from the unit of the prediction,

making interpretation very intuitive.

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.8)$$

$$EV = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \quad (2.9)$$

### 2.3.2 Cross-Validation

To assess the robustness and stability of the model, a repeated train–test splitting procedure was performed. Despite the large dataset reducing sampling variance, we conducted 10 independent splits. For each split, the model was trained on the training portion and evaluated on the held-out test portion. We then computed the mean and standard deviation of each evaluation metric across the 10 runs. A low standard deviation indicates that the model’s performance is stable and robust to different data partitions.

## 2.4 Hyperparameter Tuning

The model as explained in subchapter 2.2.2, comes with several hyperparameters regarding the architecture and the training, usually having a severe impact performance of the model. Therefore, it is necessary to find the set of hyperparameters that will give the optimal performance. This section will first aim to explain the hyperparameters and the search space, and then the optimization strategy.

### 2.4.1 Hyperparameters Tuned

The hyperparameters affect several aspects of the modelling pipeline: they influence the model architecture, the training process, and the data splitting strategy. An overview of the tuned hyperparameters is displayed in table 2.2. To examine the influence of using more data to train and validate the model, the train and validation ratio were treated as hyperparameters, these 2 ratio’s also determine the test ratio  $test_{ratio} = 1 - (train_{ratio} + val_{ratio})$ . To ensure the test ratio is at least 30%, the train and validation ratio range from 0.0 to 0.4, and 0.3 respectively.

The hidden state size reflects the amount of the capacity the LSTM has to process information and "remember" over time. A bigger hidden state will mean the memory can hold bigger patterns, but it also makes the model more complex and harder to train by introducing more model parameters. To give a wide range, values from 16 to 256 will be examined with all powers of 2 in between. The other parameter of the model architecture is the number of LSTM layers, which indicates how many LSTM cells are stacked on top of each other for each time step. More layers enables the model to perform more non-linear transformations allowing it to model more complex relationships. Like the hidden state size, this comes with the cost of

more parameters, making training harder.

Regarding the training, the learning rate and the batch size are intertwined. The batch size corresponds to how many pixels are used for a single weight update, and the learning rate is a number that is multiplied with the gradient, determining how quick or slow the model learns. Therefore, a larger batch size also allows a higher learning rate because larger batches produce more stable gradient estimates, reducing the risk of divergence when using a higher learning rate. To not restrict the learning capacities of the model, wide ranges were chosen as seen in table 2.2. The dropout ratio controls the relative amount of hidden state units to be randomly set at zero before being fed to the next LSTM layer during training, which prevents overfitting. Finally, the number of warm-up steps is the number of time step predictions that are not used for training.

**Table 2.2:** Hyperparameter search space for the LSTM model, detailing each hyperparameter's type, sampling scale, and distribution or range.

Hyperparameter	Type	Scale	Distribution/Range
Train ratio	float	uniform	[0.0...0.4]
Validation ratio	float	uniform	[0.0...0.3]
Hidden state size	integer	categorical	[16, 32, 64, 128, 256]
Number of layers	integer	uniform	[1...4]
Dropout ratio	float	uniform	[0.0...0.4]
Learning Rate	float	log	[ $10^{-4}$ ... $10^{-1}$ ]
Batch size	integer	categorical	[32, 64, 128, 256, 512]
Warm-up steps	integer	uniform	[0...25]

A hyperparameter trial consists of building a model with a specific hidden state size and number of layers, then training it using a defined dropout rate, learning rate, batch size, and number of warm-up steps. The model is trained and validated according to a chosen train-validation split. During the hyperparameter tuning, the modelling is treated as a black-box, where the inputs are the hyperparameters and the output is the validation loss.

#### 2.4.2 Tuning Method

A traditional, well known, and often used method for hyperparameter tuning is grid-search, which needs a defined search-space, and then does an exhaustive search by picking all possible combinations of the Hyperparameters within the search space and computing the loss. This is however very computationally expensive and inefficient, mainly because the number of possible combinations can be very high with a larger number of hyperparameters and a lot of suboptimal combinations are tested. A more efficient approach, the Optuna hyperparameter tuning framework (Akiba et al., 2019), builds a probabilistic model of which sets of hyperparameters are likely to perform well. This is constantly updated after each run and that way the model searches efficiently. To implement it, ranges need to be defined per hyperparameter, and a number of trials, which is the number of times the model will be trained.

## 2.5 Model Interpretation

To evaluate which variables are the driving factors of EVI, several analyses are performed to identify which variables have the greatest predictive capability. And, to examine whether these influences differ across geographical regions, the static variables (Elevation, Water Table Depth, and Sand Fraction) are also taken into account.

### 2.5.1 Feature Ablation Analysis

To assess feature performance, the model will be retrained and evaluated using different subsets of the input features. The performance of each feature set will then be compared to the baseline configuration that includes all features. For each feature, two types of experiments will be conducted: exclusion and isolation. In the exclusion setting, one feature is left out while the model is trained on the remaining three. This allows us to assess the relative contribution of that feature. In the isolation setting, the model is trained using only the feature in question, enabling an evaluation of its standalone predictive power. These experiments provide insight into how much of the variance can be explained by each feature individually. They also inform about the cruciality of a feature. If omitting a feature leads to a large increase in error, it likely contains important information not captured by the others.

As a next step, the analysis will explore whether geographic variation influences the importance of specific input features. To this end, correlations between static geographic features and the performance difference of isolated dynamic features (relative to the baseline) will be computed. A significant or meaningful correlation would suggest that the importance of certain dynamic features depends on underlying geographic characteristics.

### 2.5.2 SHAP Value Analysis

To give a broader picture, a SHapley Additive exPlanations (SHAP) value analysis is performed, which does not require retraining the model, but instead uses the trained model and input data with the aim to explain the influence of input variables on a certain prediction. It works by starting with a real set of input features for a data point. Instead of perturbing the input using completely random values, which may lead to unrealistic combinations of features and consequently meaningless model outputs, it uses background data to generate more realistic input samples. A full mathematical explanation of SHAP values is provided by Lundberg and Lee (2017).

SHAP values are calculated at the level of individual data points. To achieve the total influence per feature on the EVI, the absolute SHAP values are first aggregated over time, and then aggregated over space, allowing assessment the overall importance of each feature.

## 3. Results

The results of this study are presented in several sections. First, a general analysis of model performance is provided in Section 3.1. This is followed by the outcomes of the hyperparameter tuning in Section 3.2, and an overview of the training results in Section 3.3. Finally, Section 3.4 offers an interpretation of the model using feature ablation and SHAP values, placing the findings within the context of existing literature.

### 3.1 Model Performance

Cross-validated performance at the global dataset level is presented in Section 3.1.1, and a more in-depth analysis of test set performance is presented in Section 3.1.2.

#### 3.1.1 Cross-Validation

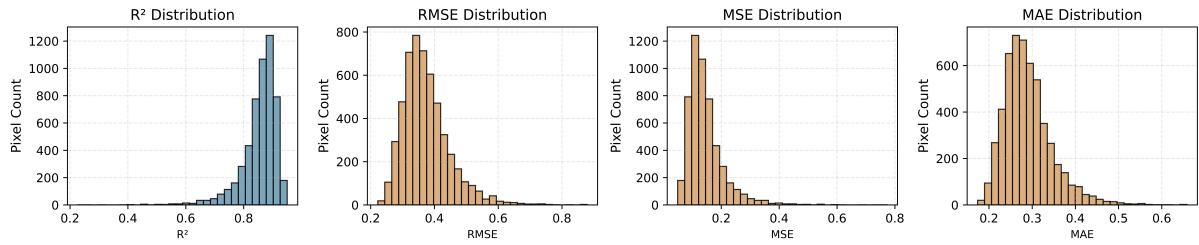
To ensure that the model generalizes well to unseen data, a cross-validation procedure was performed of which the results are shown in Table 3.1. The mean squared error (MSE) on the training and validation sets is 0.10 and 0.13, respectively. This indicates no signs of overfitting, as the performance gap between the two is minimal. The test set yields a mean absolute error (MAE) of 0.27, corresponding to an average prediction error of approximately 0.27 standard units of EVI. The Explained Variance scores are nearly identical to the  $R^2$  scores across all datasets, indicating that the model produces unbiased predictions without systematic over- or underestimation. Furthermore, the standard deviations across folds are consistently low (rounded to 0.01 for all metrics), highlighting that model performance is highly robust and not driven by chance or favourable splits of the data. An important nuance is that the EVI values have been standardized per month, meaning the model is effectively predicting anomalies rather than raw values. As the seasonal component has been removed and the data is scaled by the monthly standard deviation, a value of 1 corresponds to a larger change in EVI for pixels with higher anomaly variability. Consequently, the  $R^2$  reflects the proportion of anomaly variance explained by the model, and the MAE indicates the average prediction error in units of standard deviations of the anomaly.

**Table 3.1:** Cross-validated model performance on the train, validation and test datasets, reported as mean  $\pm \sigma$ , using a 10-fold 3-way random spatial split cross-validation strategy.

Metric	Train	Validation	Test
MSE	$0.10 \pm 0.01$	$0.13 \pm 0.01$	$0.13 \pm 0.01$
RMSE	$0.32 \pm 0.01$	$0.36 \pm 0.01$	$0.36 \pm 0.01$
MAE	$0.24 \pm 0.01$	$0.28 \pm 0.01$	$0.27 \pm 0.01$
$R^2$	$0.90 \pm 0.01$	$0.87 \pm 0.01$	$0.87 \pm 0.01$
Explained Variance	$0.90 \pm 0.01$	$0.87 \pm 0.01$	$0.87 \pm 0.01$

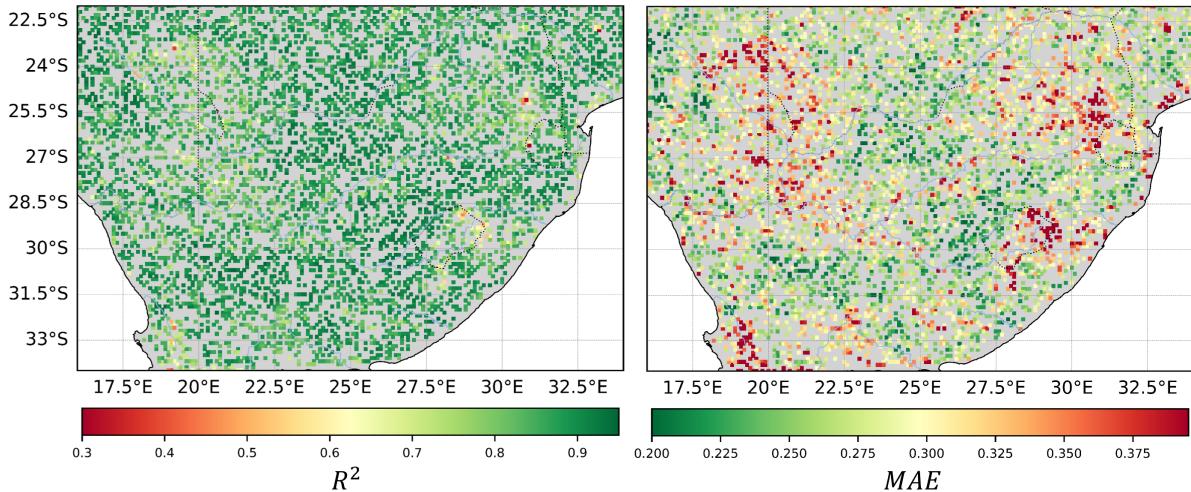
### 3.1.2 Test Set Performance

To further assess the performance of the model on unseen data, an extensive evaluation was performed on the test set. Figure 3.1 shows the distribution of the prediction accuracy across all pixels in the test set, in all metrics, a skewed distribution is visible, with a tail of lower performing locations. The histogram of  $R^2$  values demonstrates that the model can predict EVI with an  $R^2$  of 80% or higher for the vast majority of pixels. Only a small number of locations exhibit lower performance, with very few falling below 60%. A similar pattern is observed for the other performance metrics in the histograms.



**Figure 3.1:** Distribution of model performance across all test set pixels, shown for different metrics.

The spatial distribution of performance offers further insights into which regions in Southern Africa are well or poorly predicted. The  $R^2$  map (left panel of Figure 3.2) shows that most locations achieve high explained variance (around 0.8), with only a few areas displaying lower performance. The MAE map (right panel of Figure 3.2) reveals a similar spatial pattern, with slightly more pronounced clusters of lower performance.



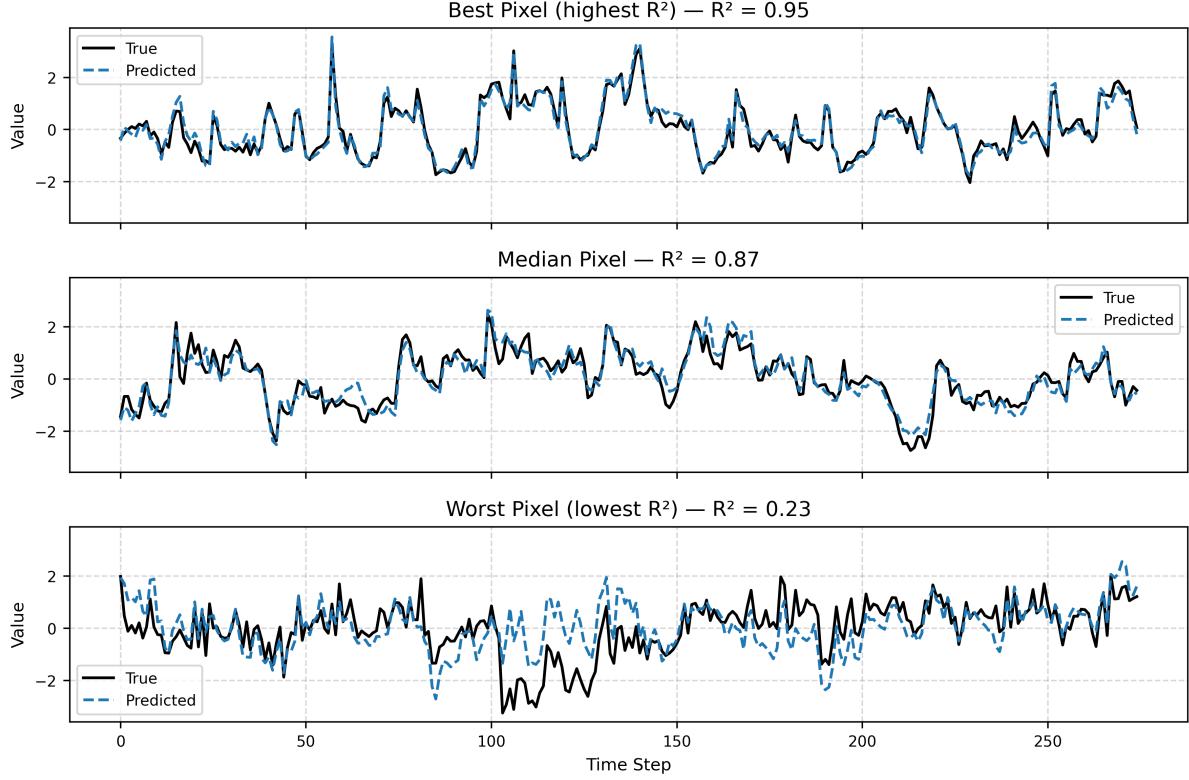
**Figure 3.2:** Spatial distribution of model performance across South Africa. Left:  $R^2$  per location. Right: MAE per location.

The temporal dynamics of model performance were also analyzed by selecting the best, median, and worst performing pixels and plotting their time series (Figure 3.3). The top panel shows the best performing pixel that achieved 0.95  $R^2$ , which exhibits near-perfect predictions with only minor errors. Due to the visible high variance in the true EVI values for this pixel, the relative prediction errors remain small, resulting in an even higher  $R^2$ . The middle panel

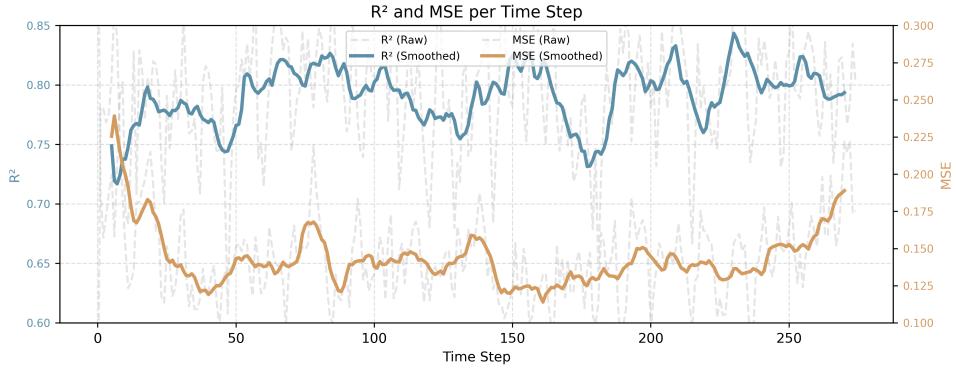
## Results

---

shows the median performing pixel, with an  $R^2$  of 0.87. Here, the model predictions follow the true values closely, with only minor deviations at certain time steps. The bottom panel presents the worst performing pixel (with an  $R^2$  of 0.23), where larger systematic errors occur, particularly between timesteps 100 and 150, during which the model consistently over-predicts. In addition, the low variance of this pixel outside the mispredicted period also contributes to the lower  $R^2$ .



**Figure 3.3:** Time series of the best, median, and worst performing pixels in the test set. True values and model predictions are shown.



**Figure 3.4:** Model performance over time. Smoothed  $R^2$  and MSE per timestep on the test set.

Figure 3.4 shows the evolution of model performance over time, with smoothed  $R^2$  and MSE values per timestep. The  $R^2$  line shows some temporal variation, with periods of higher and lower prediction accuracy. The MSE starts at a relatively high value (above 0.225) and gradually decreases as the model accumulates more temporal context, stabilizing around timestep

50. This indicates that the LSTM model requires approximately 50 months of historical information to achieve optimal accuracy, suggesting that EVI dynamics depend on temporal dependencies spanning roughly four years.

## 3.2 Hyperparameter Tuning

To find the optimal hyperparameters for the model, the Optuna algorithm was run over 151 trials. Figure 3.5 shows scatter plots of the loss versus each hyperparameter. Since the objective is to minimize the loss, the lowest points in each plot represent the best-performing configurations. It is important to note that these plots do not show a simple average trend, poor-performing combinations result in scattered higher loss values. Instead, the lower edge of each scatter plot shows how good the model can perform for each value of the hyperparameter.

Although some randomness is present, a clear trend is visible for the hidden state size: larger hidden sizes consistently allow lower loss values. While the trend appears linear, the x-axis is logarithmic, meaning the improvement is in fact logarithmic. This result is expected, as a larger hidden state provides more flexibility and greater memory capacity to the model. The results suggest that an even larger hidden size (e.g. 512) could further improve performance. However, this was not pursued, as the number of model parameters would increase substantially. Since the optimization objective does not penalize model size, Optuna naturally favours larger hidden sizes. Therefore, a hidden size of 256 was selected as it provided a good balance between performance and computational feasibility within the available resources.

The number of stacked LSTM layers shows a clear optimum at three layers, although differences in performance between configurations are relatively small. Too few layers limit the model's capacity to capture non-linear transformations, while too many layers make training more difficult. The results suggest that some degree of non-linearity is present in the data, but not to an extreme extent. Based on this finding, a three-layer architecture was selected.

For hyperparameters related to training, the dropout ratio does not have a strong influence on performance overall. However, higher dropout values tend to result in better validation loss. This is likely because dropout helps to prevent overfitting. Since the upper limit tested was 0.4 (which also resulted in the best performance) it is likely that even higher dropout ratios could have yielded slightly better results.

The learning rate appears to have found an optimum, as the lowest loss is observed around intermediate values. For batch size, a clear trend emerges: smaller batch sizes tend to result in better performance. This may be due to the high variance in the data, which makes it harder for the model to generalize when using large batches. Since the lower limit of batch sizes tested was 32, it is possible that even smaller batch sizes could have led to further improvements. Based on these results, a batch size of 32 and a learning rate of  $4.19 \times 10^{-3}$  were selected.

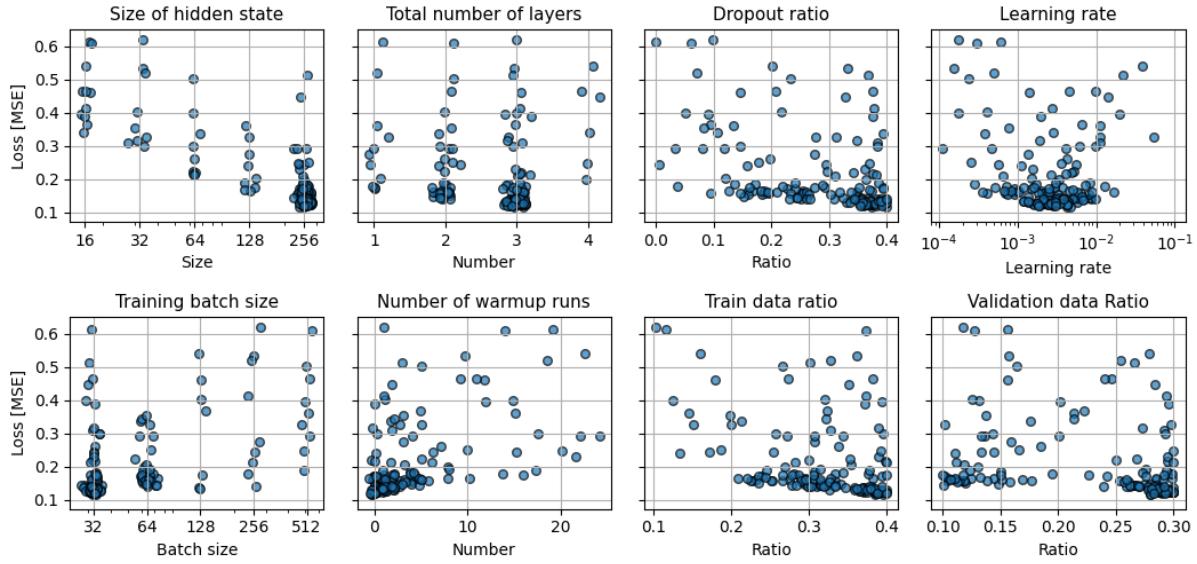
The number of warm-up steps shows a trend where lower values result in better performance, with the best results achieved at the lower limit of 0. Therefore, a value of 0 warm-up steps was selected.

For both the validation and training split ratios, higher values generally lead to better performance. It is positive to see that the model benefits from more data, although the effect is

## Results

---

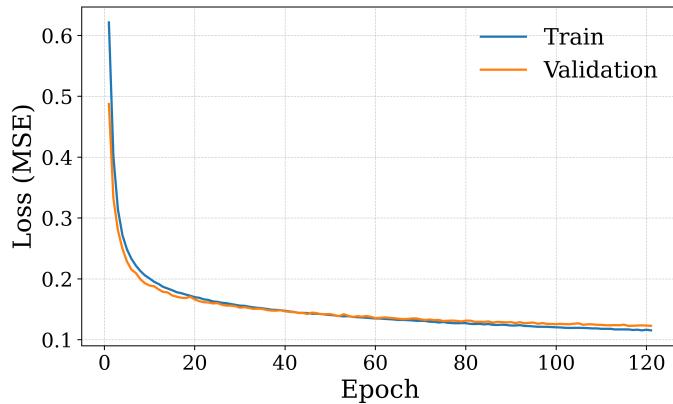
not very large, even with smaller training ratios, performance remains relatively stable. Since sufficient data is available, a validation ratio of 0.3 and a training ratio of 0.4 were selected.



**Figure 3.5:** Scatter plots of model loss versus each hyperparameter, based on 151 Optuna trials. The plots illustrate the relationship between hyperparameter values and model performance across the search space.

### 3.3 Model Training

The final model was trained using the Adam optimizer and the MSE as the loss function, with a batch size of 32 and a learning rate of  $4.19 \times 10^{-3}$ . Training was performed for a maximum of 120 epochs, with early stopping based on validation loss and a patience of 10 epochs. Training dynamics were stable, as shown in Figure 3.6, with both training and validation loss converging after approximately 70 epochs. No signs of overfitting were observed, as the gap between training and validation loss remained small throughout training. The relatively large number of epochs required for convergence is likely due to the complexity of the model architecture (multiple LSTM layers with large hidden states), combined with the inherent noise and variability in the environmental data and the spatio-temporal nature of the task.



**Figure 3.6:** Training and validation loss curves over epochs.

## 3.4 Model Interpretation

In this section, the results of model interpretation methods are examined. In Section 3.4.1, the importance of the input features is evaluated using both ablation analysis and SHAP values. Section 3.4.2 further explores the role of static features to assess how feature importance varies across geographical regions.

### 3.4.1 Feature importance

To evaluate the effects of individual input variables on the EVI, a feature ablation analysis was performed, using exclusion and isolation.

Table 3.2 gives an overview of the ablation analysis, with on the left side the exclusion part, on the right side the isolation part, and at the bottom the baseline. Looking at the  $R^2$  of the exclusion part, the values come relatively close to the baseline, with values ranging around 0.83 or 0.84, while the baseline is 0.87. Since there are no excluded features that lead to a very significant drop in performance, the model is not very dependent on a single feature. Apparently, if a feature is excluded, the model can use information from other features to fill this gap. Table 2.1 also shows the correlations of all variables used, including the correlations between the covariates. There, we can see that the correlations are relatively high. This helps explain why the model does not rely heavily on individual features, as the features contain overlapping information due to their correlation. Looking at other metrics in the exclusion part of Table 3.2, the same pattern is presented, but for MSE and RMSE, it is reversed as expected, since higher values there indicate a greater error. A closer look reveals *SPEI-3*, *Precipitation*, and *SPEI-1* have virtually the same value. However, temperature has a value that stands out with a higher  $R^2$ . This could indicate that *Temperature*, carries the least unique information. Another explanation could be the lower spatial resolution of this variable as seen in Figure 2.1.

**Table 3.2:** Ablation analysis results: comparing exclusion (all-but-one) and isolation (only-one).

Feature	Exclusion				Isolation			
	$R^2$	EV	MSE	RMSE	$R^2$	EV	MSE	RMSE
SPEI-1	0.831	0.831	0.169	0.411	0.744	0.747	0.256	0.506
SPEI-3	0.832	0.832	0.168	0.410	0.829	0.829	0.171	0.414
Precipitation	0.832	0.832	0.168	0.410	0.841	0.841	0.159	0.399
Temperature	0.840	0.840	0.160	0.400	0.790	0.791	0.210	0.458
<b>Baseline (All)</b>	0.871	0.871	0.129	0.359				

Looking at the right side of Table 3.2, we see the isolation part. This part focuses on how well an individual feature can predict EVI on its own. Compared to the exclusion part, we see more variation between the features. It is interesting to note that precipitation alone can already explain 84% of the variance, while the baseline explains 87%. This means that precipitation is a very strong predictor. SPEI-3 also scores highly with 81%, followed by temperature at 79%, and SPEI-1 with 74%. The fact that individual features reach prediction scores relatively close the baseline, may be due to correlations between the features.

In the exclusion setting, all the features except temperature seem to have the same im-

portance, while in the isolation setting, precipitation has the highest performance. This likely means that temperature lacks some information (which is present in other features), while precipitation individually carries the most information relevant to predict EVI.

In addition to the ablation analysis, a SHAP value analysis was also performed. Table 3.3 shows the results. Each feature has an associated value, and to improve interpretability, these values have been normalized as fractions summing to 1. Looking at the raw SHAP values, they are all quite low. This might seem strange at first, but it is likely because SHAP values only capture the direct effect, which is, the effect at a single time point. The historical influence is not included. So, the direct contribution at the same timestamp appears very small according to SHAP values, which is consistent with expectations because as shown in Section 3.1.2, historical data (possibly going back up to four years) can impact current predictions. Notably, precipitation emerges as the most influential predictor, consistent with its dominance in the isolation analysis and, Temperature emerges as the least important feature, which is consistent with the Exclusion analysis. However, the results of the SHAP value analysis should be interpreted with caution, as the direct effect alone may not fully reflect their overall contribution to the model's predictions, since only direct effects are accounted for.

Variable	Raw SHAP	Normalized SHAP
Precipitation	0.00131	0.469
SPEI-1	0.00090	0.323
SPEI-3	0.00039	0.140
Temperature	0.00019	0.068

**Table 3.3:** Dynamic feature importance: raw and normalized (sum=1) SHAP values. Aggregated over time and space

One final remark about feature importance is that, due to the large overlap in information among the features, no single feature stands out as particularly influential or crucial. All features contribute to explaining a significant portion of the variance in EVI. So, there is neither a crucial feature nor an irrelevant one, each plays a meaningful role.

### 3.4.2 Correlation with static features

To investigate whether there is a relationship between geographic conditions and the importance of input variables, the difference between the isolation analysis scores and the baseline was calculated for each location. Then, Pearson correlation coefficients were computed between these differences and the static features: elevation, groundwater depth, and sand fraction. The results can be seen in Table 3.4. Since Pearson's correlation assumes a linear relationship, scatterplots were also analysed to check for non-linear relationships, but no clear patterns were observed.

Overall, few strong correlations emerged with the only notable one between elevation and SPEI-1. This correlation suggests that higher elevation is associated with greater importance of SPEI-1. Since SPEI-1 reflects short-term drought conditions, this may indicate that short-term drought becomes more relevant for predicting vegetation at higher altitudes. The reasons behind certain feature importances may also be linked to other geographic conditions that are

not captured by the three static variables used here. A more extensive analysis using extra (derived) features would be needed to explore this further.

Static Feature	Precipitation	SPEI-1	SPEI-3	Temperature
elevation	0.061	<u>0.291</u>	0.090	0.097
wtd	0.021	0.065	0.009	0.063
sand	-0.005	0.078	-0.045	0.027

**Table 3.4:** Correlations (R) of performance  $\Delta$  vs static features ( $R^2$  and MSE), underline for meaningful correlations.

## 4. Discussion

Reflecting on the findings, it is encouraging to see that EVI can be predicted with high accuracy using a compact set of environmental variables. The results align well with previous literature, such as Jin et al. (2020), which found strong associations between EVI and climatic factors like temperature, precipitation, and water balance. The model confirms that these dynamic drivers are indeed sufficient to explain a large portion of vegetation dynamics, 87% of the anomaly variance on average. To potentially further improve performance, incorporating additional environmental predictors could be valuable. For example, Galvão et al. (2023) suggest that solar radiation influences vegetation dynamics. Including this variable in the LSTM model as an additional dynamic predictor could therefore enhance its predictive power.

The LSTM architecture proved effective in capturing temporal dependencies, mainly because correlations between inputs and EVI anomalies were not always strong (Table 2.1), implying that much of the predictive power arises from learned temporal patterns and non-linear relationships.

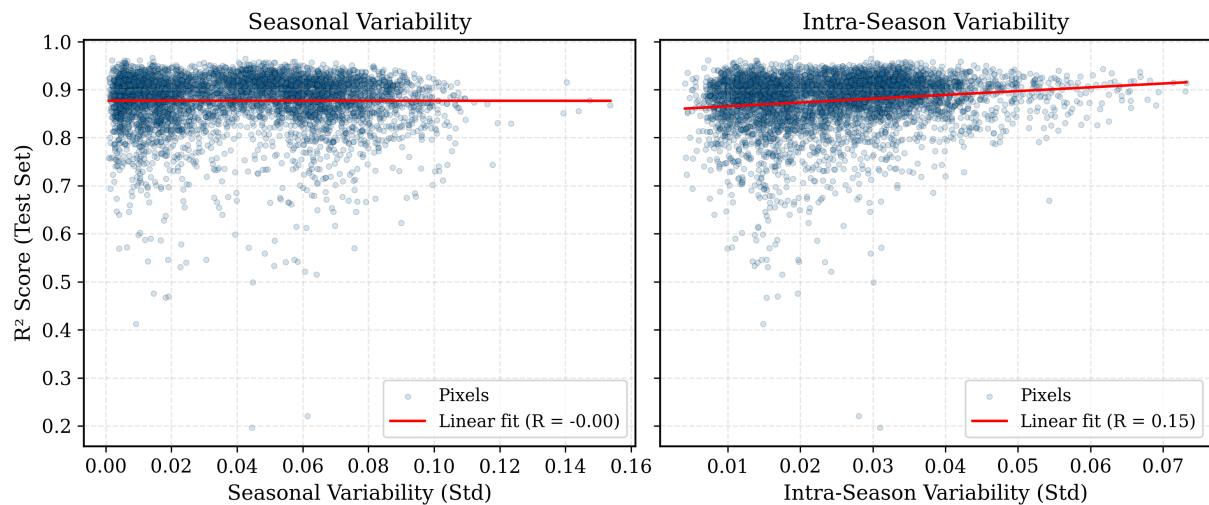
Notably, a single LSTM model was able to generalize across different locations within South Africa. This suggests that, despite regional variability, shared climate–vegetation relationships exist that can be learned. However, this also raises questions about transferability. South Africa has relatively low levels of human landscape modification, unlike regions such as the Netherlands (Theobald et al., 2020). It remains uncertain whether the same model would generalize well to areas with stronger human influences. Future work should evaluate the model’s transferability, ideally by applying it to contrasting regions and using spatial data splits to minimize leakage between training and validation sets. Also, since a random spatial split was used, neighboring pixels can have similar vegetation dynamics, there’s a risk that a validation pixel may closely resemble a nearby training pixel, potentially leading to optimistic estimates of generalization performance. Explicit spatial splits or spatial cross-validation schemes may help address this issue.

Additionally, while EVI was standardized during preprocessing, no back-transformation was applied to interpret predictions in absolute terms. This simplifies the modeling and allows for anomaly-focused interpretation, but future work might benefit from recovering absolute vegetation levels, especially when linking outputs to ecological thresholds or policy targets. A potential downside is that strong seasonality combined with low intra-season variability may lead to exaggerated standardized errors, which might not be as meaningful when viewed in terms of real EVI values.

Figure 4.1 shows how model performance relates to vegetation variability across locations. The standard deviation of the twelve monthly means represents how much variation there is between seasons, while the average of the twelve monthly standard deviations reflects how much variation occurs within each season. The left plot shows the relationship between sea-

sonal variability and performance. There is no clear trend, and both the regression line and correlation coefficient suggest no association. On the right, the relationship between intra-seasonal variability and performance shows a very subtle positive correlation. While the correlation is low, it is interesting to note that all low-performing outliers are found in locations with relatively low intra-seasonal variability. One possible conclusion is that higher intra-seasonal variability helps the model perform better, which would make sense since smaller real-world differences are inherently harder to predict than larger ones.

Trend was not explicitly examined in this analysis. The standardization technique used assumes stationarity. Future research could explore whether alternative standardization methods might be more appropriate, particularly if trends are present in the EVI data.



**Figure 4.1:** Relationship Between Vegetation Variability and LSTM Performance, split up into seasonal variability (Std of all monthly means) and intra-season variability (Average of all monthly Std's).

Using the model and findings of this research, several paths can be explored in future research. The model can be applied to specific locations by modifying input variables to simulate alternative climate situations, such as milder droughts. This can help explore how vegetation might respond differently under such conditions and offer insights into strategies that support faster recovery.

If forecasts of the dynamic input variables are available, the model can be used to predict future vegetation states. This could support early detection of vegetation stress. Additionally, testing the model's ability to generalize across time, for example by using a temporal split, would be useful to assess its robustness in predicting time periods it has not seen before.

## 5. Conclusion

This study was set out to determine whether vegetation dynamics in South Africa, captured through the Enhanced Vegetation Index (EVI), could be effectively modelled using Long Short-Term Memory (LSTM) networks, what the main drivers of EVI are, and to what extent different environmental drivers influence these dynamics.

The LSTM model demonstrated strong predictive capabilities, achieving an average  $R^2$  of 0.87 on the test set, confirming that complex temporal patterns of vegetation can be captured using a relatively compact set of input variables. These included both dynamic climate variables (precipitation, temperature, and SPEI) and static landscape features (elevation, groundwater depth, and sand fraction). The model showed robust performance across different data splits and most locations.

Feature importance analyses revealed that no single dynamic variable was solely responsible for EVI variability. Instead, several features contributed meaningfully and often interchangeably due to their inter-correlations. The ablation analysis suggests that precipitation contains the most predictive information, and temperature provides the least unique information, since most of it is captured in the other features. The feature importance did not remain constant across space: short-term drought signals (SPEI-1) were found to be more predictive at higher elevations, hinting at interactions between elevation and climate sensitivity of vegetation.

Altogether, this study demonstrates that environmental predictors contain sufficient information to accurately model vegetation dynamics, and that LSTM networks are capable of effectively extracting and leveraging this information. These findings are consistent with previous research, such as Jin et al. (2020), who identified strong relationships between EVI and key climatic and topographic variables, including temperature, precipitation, and climatic water balance. In addition, the analysis suggests that the relative importance of certain drivers, such as short-term drought indicators, may depend on local environmental conditions.

## References

- PxHere contributors. (2017). South african savanna landscape [Image licensed under CC0 (Public Domain) via PxHere. Accessed 2025-07-01]. <https://pxhere.com/en/photo/1130295>
- Jin, Y., Feng, X., Friedl, M. A., Myers, J. A., Zhu, Z., & Davidson, E. A. (2020). Topographic, soil, and climate drivers of drought sensitivity in forests. *Scientific Reports*, 10, 11529. <https://doi.org/10.1038/s41598-020-75273-5>
- Wilhite, D. A., & Glantz, M. H. (1985). Understanding: The drought phenomenon: The role of definitions. *Water International*, 10(3), 111–120. <https://doi.org/10.1080/02508068508686328>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements*. Food; Agriculture Organization of the United Nations.
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696–1718. <https://doi.org/10.1175/2009jcli2909.1>
- Anderegg, W. R. L., Hicke, J. A., Fisher, R. A., Allen, C. D., Aukema, J., Bentz, B., Hood, S., Lichstein, J. W., Macalady, A. K., McDowell, N., Pan, Y., Raffa, K., Sala, A., Shaw, J. D., Stephenson, N. L., Tague, C., & Zeppel, M. (2015). Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytologist*, 208(3), 674–683. <https://doi.org/10.1111/nph.13477>
- Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333(6042), 616–620. <https://doi.org/10.1126/science.1204531>
- on Climate Change (IPCC), I. P. (2023, June). *Climate change 2021 – the physical science basis: Working group i contribution to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). Wiley.
- Rouse, J., Haas, R., Schell, J., & Deering, D. (1974). Monitoring vegetation systems in the great plains with erts. *Third Earth Resources Technology Satellite-1 Symposium*, 1, 309–317.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sensing of Environment*, 83(1-2), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Molnar, C. (2025). *Interpretable machine learning: A guide for making black box models explainable* (3rd ed.). <https://christophm.github.io/interpretable-ml-book>
- Zhao, W., Joshi, T., Nair, V. N., & Sudjianto, A. (2020). Shap values for explaining cnn-based text classification models. <https://doi.org/10.48550/ARXIV.2008.11825>
- van der Hoek, D. (2025, July). *Da-vanderhoek/pixel-lstm-forecasting: Added data* (Version v1.1). Zenodo. <https://doi.org/10.5281/zenodo.15805575>

## Bibliography

---

- Didan, K. (2015). Mod13a2 modis/terra vegetation indices 16-day l3 global 1km sin grid v006. <https://doi.org/10.5067/MODIS/MOD13A2.006>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Ruijsch, D., Mourik, J. v., Biemans, H., Hauswirth, S. M., & Wanders, N. (2025). Thrive or wither: Exploring the impacts of multi-year droughts on vegetation. <https://doi.org/10.22541/essoar.174534298.81830573/v1>
- Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (gmted2010)* (tech. rep. No. Open-File Report 2011–1073). U.S. Geological Survey. <https://pubs.usgs.gov/of/2011/1073/>
- Reynolds, C. A., Jackson, T. J., & Rawls, W. J. (2000). Estimating soil water-holding capacities by linking the food and agriculture organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, 36(12), 3653–3662. <https://doi.org/10.1029/2000wr900130>
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth [Dataset includes global water table depth (WTD) simulations. Data accessed via: <https://wci.earth2observe.eu/thredds/catalog/usc/root-depth/catalog.html>]. *Proceedings of the National Academy of Sciences*, 114(40), 10572–10577. <https://doi.org/10.1073/pnas.1712381114>
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53. <https://doi.org/10.1007/s10462-020-09838-1>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyper-parameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, 2623–2631. <https://doi.org/10.1145/3292500.3330701>
- Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. <https://arxiv.org/abs/1705.07874>
- Galvão, L. S., Arlanche Petri, C., & Dalagnol, R. (2023). Coupled effects of solar illumination and phenology on vegetation index determination: An analysis over the amazonian forests using the superdove satellite constellation. *GIScience amp; Remote Sensing*, 61(1). <https://doi.org/10.1080/15481603.2023.2290354>
- Theobald, D. M., Kennedy, C., Chen, B., Oakleaf, J., Baruch-Mordo, S., & Kiesecker, J. (2020). Earth transformed: Detailed mapping of global human modification from 1990 to 2017. *Earth System Science Data*, 12(3), 1953–1972. <https://doi.org/10.5194/essd-12-1953-2020>