Universiteit Leiden
The Netherlands

Data Science and AI

Modeling Vegetation Health from Climate Drivers:
Residual-Based Detection of Anomalies

Mariam Gagua

Supervisors:
Dr. Matthijs van Leeuwen & Dr. Lincen Yang

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

01/07/2025

Contents

Abstract

This study models the relationship between climate variables and vegetation health (NDVI) across Europe using Random Forest regression, trained on ERA5-Land climate reanalysis and MODIS NDVI data. While the model achieved strong predictive performance and generalized well across time, the primary focus lies in analyzing residuals, differences between observed and predicted NDVI, and what insights they could reveal beyond predictive accuracy metrics. Results showed that residuals in late summer and early autumn months of the test years were smallest and most stable, aligning with periods of peak vegetation activity. For analyzing the spatial patterns, the study focused on summer months and residuals exceeding 2 $\times$ the standard deviation were flagged as anomalies, highlighting significant model failures. Some regions consistently showed high residuals across years, while others exhibited isolated spikes, reflecting both persistent and event-specific mismatches. These differing patterns hint at a combination of systematic model challenges and context-specific dynamics, warranting further investigation into their underlying causes. A notable example was the 2010 anomaly in western Russia, where large overpredictions aligned with a well-documented heatwave and drought, suggesting a potential breakdown in modeled expectations under extreme climate conditions. These findings highlight the value of residuals not just for model evaluation, but as a lens into unmodeled processes and extreme events. The study concludes that combining predictive modeling with residual-based anomaly detection provides a useful framework for uncovering unexpected vegetation–climate dynamics.

1 Introduction

The relationship between climate variability and vegetation has long been a central topic in environmental research due to its significant implications for food security, sustainability, and ecosystem stability [PLC+19]. Vegetation health serves as a key indicator of how terrestrial ecosystems respond to external pressure, and it is particularly sensitive to fluctuations in climatic variables such as temperature, precipitation, and solar radiation [PVM+05, ZTK+01]. Therefore, analyzing how vegetation responds to climatic variation provides crucial insights into the broader ecological impacts of climate change.

Vegetation health is frequently assessed using satellite-driven vegetation indices, among which the Normalized Difference Vegetation Index (NDVI) is the most widely employed. NDVI measures the difference in reflectance between near-infrared and red wavelengths to infer photosynthetic activity and capture changes in greenness and biomass. It provides a robust representative for plant productivity and is extensively used in operational and scientific context, including agricultural yield forecasting, drought early warning systems, and ecosystem monitoring [PVM+05].

Although the link between NDVI and climate variables is well-established, this relationship is neither static nor uniform [ZTK+01]. It varies considerably by season, geographic regions, land cover type and climate zone. Climatic conditions, in particular temperature and moisture availability, can either promote or inhibit plant growth depending on local conditions. For example, while higher temperatures may promote vegetation growth in temperate climates, characterized by moderate temperatures and sufficient rainfall, they can lead to stress and decline in water-limited regions, where precipitation is low and soil moisture is often insufficient to support sustained plant growth. Additionally, local factors such as soil properties, elevation, and land use can further influence how vegetation responds to climatic inputs. These responses are not only spatially but also temporally

variable, shifting across seasons and years depending on phenological stages or long-term trends [PFY+19].

Understanding how ecosystems respond to ongoing climate change requires close attention to spatial and temporal heterogeneity in vegetation-climate interactions. While many large-scale studies acknowledge these variations, their focus often remains on capturing broad trends or predictive performance [RCVS+19]. These efforts tend to overlook the informative value of model errors, which are typically treated as noise rather than as potential signals of ecological or climatic anomalies.

This thesis proposes that there is valuable insight to be gained from residuals, the differences between observed and predicted NDVI, as they may point to breakdowns in expected relationships and reveal the influence of unmodeled processes or unexpected events. By learning typical spatio-temporal patterns and investigating where the model diverges from them, this study aim to investigate whether meaningful insight can be gained into vegetation sensitivity and environmental dynamics under varying climate conditions.

Europe is selected as the study region due to its broad ecological and climatic diversity, which spans arid, drought-prone regions in the Mediterranean, temperate agricultural zones, boreal forests and mountainous regions in Scandinavia. This diversity allows for testing the model across a wide range of vegetation-climate interactions and enables regional comparisons. Moreover, Europe benefits from being extensively studied and well documented, providing a viable and dependable foundation of data for both modeling and interpretation.

In parallel with advances in remote sensing, environmental monitoring, and growing availability of large-scale datasets, machine learning (ML) has become an increasingly valuable tool in climate and ecological research [KEUR+18]. ML methods, particularly non-linear models, are well suited for modeling interactions between climate and vegetation, which are often shaped by region-specific conditions and evolve over time. Unlike traditional methods based on fixed assumptions, ML models adaptively learn relationships directly from the data, offering greater flexibility in capturing complex patterns and enabling more data-driven insights.

1.1 Research aims and questions

The goal of this thesis is to model the relationship between NDVI and key climate variables across Europe using machine learning, and to identify specific regions and time periods where the modeled relationship deviated from observations, uncovering potential anomalies or unmodeled influences in vegetation-climate dynamics. To that end, the study addresses the following overarching research question: **How do climate variables shape vegetation health across Europe over time and can meaningful insight be gained by analyzing where and when the model's predictions deviate from observed NDVI?**

Three sub-questions guide the analysis:

- **RQ1:** How well can key climate variables predict NDVI across Europe, and which variables contribute most to model performance?
- **RQ2:** In which regions or time periods do the expected relationships between climate variables and NDVI break down?

- **RQ3:** What external factors (e.g., model or data limitations, climate extremes) might explain these deviations?

1.2 Overview

The remainder of thesis is structured as follows:

Chapter 2 introduces the foundational concepts necessary for understanding the study. Chapter 3 reviews related work and situates this research within the broader literature on vegetation–climate modeling. Chapter 4 details the datasets used and the preprocessing steps taken to prepare them for modeling. Chapter 5 presents the core methodology, starting with the construction and evaluation of the baseline machine learning model, followed by an in-depth analysis of residuals to detect and interpret anomalies. Chapter 6 discusses the results in relation to the research questions and concludes the study. Finally, Chapter 7 outlines the study's limitations and suggests ways for future research.

2 Background and concepts

This chapter introduces the key concepts underpinning the study. It begins by describing the Normalized Difference Vegetation Index (NDVI), including how it is derived from satellite reflectance data and what it reveals about vegetation health. Next, it outlines the climate variables considered in the analysis and their ecological relevance. The third subsection addresses spatio-temporal modeling and highlights importance of capturing variation across both space and time. The chapter concludes with a discussion on anomaly detection, focusing on the use of residuals from predictive models as a way to identify unexpected vegetation responses. Together, these concepts form the basis of the thesis and provide the conceptual foundation for the modeling and analysis approach.

2.1 NDVI: What does it tell us?

The Normalized Difference Vegetation Index (NDVI) is one of the most widely used metrics for quantifying vegetation health, biomass and density through remotely sensed data. It is calculated from reflectance measurements in two specific spectral bands: the red (R) and near-infrared (NIR) wavelengths, typically sourced from satellite-based remote sensors [YDO+15].

The NDVI calculation is mathematically expressed as the ratio between the difference and sum of near-infrared and red reflectance values:

$$NDVI = \frac{NIR - R}{NIR + R}$$

This formula uses the distinct spectral characteristics of vegetation: healthy (green) vegetation absorbs visible red light for photosynthesis and reflects a large portion of near-infrared radiation. Conversely, sparse or unhealthy vegetation reflects more of visible light and less near-infrared. Therefore, by combining these bands into a normalized ratio, NDVI provides an indicator of vegetation photosynthetic capacity and overall physical condition. NDVI values range from -1 to 1, where values between 0 and 1 typically correspond to vegetated areas, with higher values indicating denser, healthier vegetation and greater chlorophyll content, while values near 0 or negative are associated with bare soil, water bodies, or urban surfaces.

NDVI has been extensively used for analyzing and monitoring plant phenology — the study of periodic plant life cycle events and timing. Through NDVI variations, researchers can observe the start, duration, and end of growing seasons, as well as evaluate how vegetation responds to seasonal and interannual variability in climate and environmental stresses [RBV+94]. Moreover, research has validated NDVI as a reliable proxy for Net Primary Productivity (NPP), making it a valuable tool for assessing land productivity, ecosystem health and land degradation states [RNH+04].

2.2 Climate variables

Variable	Units	Definition	Relevance
2m Temperature (t2m)	K	Mean air temperature measured two meters above earth's surface;	Regulates photosynthesis, evapotranspiration, and the timing of the growing season by setting thresholds for growth onset and dormancy.
2m Dew Point (d2m)	K	Temperature at which atmospheric moisture condenses into dew at two meters above the surface;	Measure of humidity in the air; Indicates water availability when soil moisture data is unavailable; Relevant for evapotranspiration and plant water stress.
Total Precipitation (tp)	m	The cumulative monthly amount of water (rain and snow) that falls on earth's surface;	Key water input affecting soil moisture availability and drought conditions.
Solar Radiation (ssrd)	Jm^{-2}	Total incoming shortwave solar radiation reaching the earth's surface;	Primary driver of photosynthesis and NDVI variability; Helps explain seasonal and regional patterns in vegetation greenness.
Snow Depth (sde)	m	Monthly average depth of snow cover on the ground;	Influences the timing of spring green-up and the length of the growing season by delaying soil warming.
Leaf Area Index (lai_lv)	m^2m^{-2}	One-half of the total green leaf area per unit ground area for low vegetation types;	Complements NDVI by capturing canopy structure and density.

Table 1: Description and relevance of climate variables used in the study [PLC+19, VSGC+13, NKH+03].

To model vegetation dynamics, six climate-related variables (see Tab. 1) from the ERA5-Land reanalysis dataset [MS19] were selected based on their ecological relevance, availability, and interpretability. These variables capture key environmental conditions that influence plant growth and

productivity, and their inclusion was guided by existing literature on vegetation modeling as well as comparisons with the AgERA5 dataset [(C320], which is tailored for agricultural applications.

2.3 Spatio-temporal modeling

Spatio-temporal modeling refers to the analysis of phenomena that vary across both space and time [CW11]. In the context of Earth system science, this approach is particularly valuable because environmental processes, such as vegetation dynamics and climate variability, are inherently embedded within geographic and temporal contexts. Vegetation does not respond uniformly across locations or time periods, and understanding these dynamics requires data and models that capture both dimensions.

In this study, the objective is to model the climate-NDVI relationship across Europe, which evolves with seasonal cycles, interannual variability, and localized conditions. This requires leveraging spatio-temporal data, where each observation is defined by a specific geographic location (latitude and longitude) and time point (monthly resolution). To explicitly capture these dependencies, both spatial and temporal features are directly included as input variables in the model. Latitude and longitude allow the model to learn regional differences in vegetation–climate interactions, reflecting factors like climatic zones or vegetation types. Temporal features, such as the month and year, are included to help the model capture recurring seasonal patterns and long-term trends. Together, these inputs help the model produce region-specific and time-aware predictions.

2.4 Anomaly detection and the role of residuals

In the context of predictive modeling, residuals represent the difference between observed and predicted values.
For a regression model:

$$f : X \to \mathbb{R}$$

the residuals are defined as:

$$Residual = y_{observed} - y_{predicted}$$

In this study, they capture the gap between the actual NDVI and the NDVI predicted by the model based on the input variables. Residuals quantify how much the model's expectation deviates from reality, and therefore are a direct measurement of prediction error. The distribution of residuals helps assess the overall performance and reliability of the model.

Residuals offer insight into the model's ability to generalize the learned climate–NDVI relationship to new data. Small residuals suggest that the model is accurately capturing vegetation responses, while large residuals highlight areas or time period where the model struggles. These deviations are not necessarily noise; they are potentially informative. This way of interpreting residuals builds on the assumption that the model, trained on historical data, captures the dominant and "expected" vegetation responses to climatic conditions. In doing so, it establishes a baseline that reflects how NDVI usually behaves under known climatic conditions. When the model is later applied to new data, large residuals (significant deviations from this baseline) may therefore indicate "unexpected" behavior. Analyzing these residuals can reveal where and when the model underperforms, pointing

to possible weaknesses in the input data, model limitations or external influences not captured in the training process (e.g., domain shift or extreme ecological processes).

Residuals are therefore a useful tool not only for evaluating model performance, but also for investigating the underlying causes of significant prediction errors.

Anomaly detection refers to identification of observations that deviate significantly from an expected pattern or norm [CBK09]. In this context, since large residuals reflect discrepancies from the model's learned baseline, they can serve as indicators of potential anomalies. Thus, residual-based anomaly detection provides a data-driven way to flag unusual vegetation–climate dynamics for further investigation.

Importantly, interpreting residuals as ecological anomalies must be done carefully. As mentioned, high residuals may also stem from model limitations, such as missing predictive variables, poor generalization to unseen conditions, or spatial and temporal heterogeneity not captured by the chosen features. Additionally, data quality issues such as sensor errors or cloud contamination can increase residuals. Therefore, residuals alone cannot confirm the presence of an anomaly; they only serve as signals for further analysis.

3 Related work

As previously mentioned, understanding the relationship between vegetation health and climate variability has long been central to environmental research. A large amount of work has investigated how factors like temperature, precipitation, and solar radiation influence vegetation dynamics, commonly using metrics like the Normalized Difference Vegetation Index (NDVI). Much of this research has employed correlation analysis, linear regression, or statistical decomposition methods to quantify these relationships. While such approaches have provided valuable insights and often account for spatial and seasonal variation, they typically do not use predictive models to define a baseline expectation and then, analyze deviations from it. As a result, model performance is often not interpreted in terms of what prediction errors might reveal about system dynamics or model limitations.

Several recent studies illustrate these limitations. Mehmood et al. (2024) [MRW+24] examine seasonal and regional vegetation health dynamics using NDVI in relation to key climatic variables such as temperature and precipitation. Their study highlights the importance of spatially and temporally explicit monitoring, but the analysis remains mostly correlational and does not explore predictive potential of climate variables, nor systematic interpretation of prediction errors. Pei et al. (2019) [PFY+19] investigate NDVI-climate relationships at different monthly scales in inner Mongolia, revealing lagged effects and sensitivity variations. However, their methodology relies on classical correlation-based time series analysis, without employing predictive models or evaluating model errors. Similarly, Eisfelder et al. (2025) [EUA+25] conduct a comprehensive thirty-year analysis of seasonal NDVI trends across Europe, examining how vegetation dynamics vary across land cover types and biogeographical regions in relation to climatic drivers. The study offers valuable insights into long-term vegetation-climate interactions and spatial variability in trends. However, it adopts a descriptive and statistical trend analysis approach, rather than using predictive modeling frameworks.

More recent efforts have introduced machine learning and signal decomposition methods to

model nonlinear vegetation-climate interactions. Sun et al. (2022) [SLC+22] combined an adaptive decomposition method (ICEEMDAN) with support vector machines to model multiscale patterns and nonlinear trends in NDVI and climate data. While this approach improves forecasting of vegetation dynamics, the focus remains on signal decomposition and time series prediction, rather than on understanding model limitations or spatially mapping where predictive models deviate from expected patterns. Fathollahi et al. (2024) [FWM+24] applied deep learning to forecast global NDVI key climate inputs including temperature, precipitation, and soil moisture. Their model achieved high predictive accuracy and demonstrated the potential of deep learning for large-scale vegetation monitoring. Yet, like many ML-driven environmental studies, the focus is on forecasting rather than detecting where and why predictions go wrong.

This thesis aims to fill a methodological and conceptual gap. Rather than solely focusing on model accuracy, it investigates whether the residuals, differences between observed and predicted NDVI, can serve as meaningful indicators of ecological deviations, model limitations, or unmodeled environmental processes. By combining machine learning with spatially and temporally explicit residual analysis, this study contributes to a growing interest in machine learning for environmental science, and adds a spatially and temporally aware framework for understanding vegetation responses to climate variability, detecting vegetation anomalies and diagnosing model weaknesses.

4 Data and Preprocessing

4.1 Study area

The study area encompasses the European mainland, extending to Great Britain and Iceland. Specifically, the region spans a latitude range from approximately $34°N$ to $67°N$ and a longitude range from $24°W$ to $46°E$. This geographical region encompasses diverse climatic zones.

Europe was selected primarily due to its pronounced climatic and ecological variability, which makes it possible to explore vegetation response to climate conditions across distinct biogeographical areas. These regions include Mediterranean ecosystems, characterized by seasonal droughts and high summer temperatures; temperate zones with moderate temperatures and well-defined growing seasons; and boreal ecosystems in the north, with shorter growing seasons and cooler temperatures. This environmental variability enables meaningful regional comparisons and supports the detection of diverse NDVI–climate relationships, making Europe an ideal testing place for identifying deviations across a range of ecological contexts.

4.2 Datasets

4.2.1 ERA5-Land Climate Reanalysis Data

Climate data for this study was sourced from the ERA5-Land monthly averaged climate reanalysis dataset [MS19], which provides a consistent record of key land surface variables spanning several decades. ERA5-Land is openly accessible through the Copernicus Climate Datastore (CDS) platform. Climate reanalysis data combines past observational data with state-of-the-art modeling approaches to generate continuous and coherent time series of various climate variables.

The specific ERA5-Land dataset chosen for this study provides monthly mean climate values from 1950 onwards at a global scale. Each monthly data point is the average of all hourly values

recorded throughout the entire month. The six climate variables selected were discussed in Section 2.2. The data used in this analysis spans the years 2000-2025. The spatial resolution of ERA5-Land is approximately $0.1° \times 0.1°$ (corresponding to roughly 9 kilometers at the equator). However, due to spherical geometry of the Earth, longitudinal resolution decreases with increasing latitude due to convergence of meridians. A subset of the global ERA5-Land dataset was extracted to match the European study region, as defined by the latitude and longitude boundaries outlined in Section 4.1.

4.2.2 MODIS NDVI Dataset

Vegetation health was analyzed using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's Terra satellite. Specifically, the MOD13A3 version 6.1 Vegetation Indices Monthly L3 Global dataset [Did21] was chosen due to its consistent, reliable and well documented vegetation indices products. The data was openly accessed through NASA's Earthdata platform and downloaded using NASA's Application for Extracting and Exploring Analysis REady Samples (AppEEARS) tool, which allows for tailored spatial and temporal data extraction.

MODIS vegetation indices are derived using atmospherically corrected surface reflectance values recorded in red, near-infrared, and blue spectral bands. NDVI values derived from these bands provide a reliable measure of vegetation canopy greenness, a composite indicator of leaf area, chlorophyll content, and canopy structure.

The selected MOD13A3 NDVI dataset offers monthly composites at a fine spatial resolution of approximately 1km (around $0.008° \times 0.008°$), which is substantially finer than the resolution of ERA5-Land's data. Each monthly NDVI value is a composite derived from the overlapping 16-day MODIS vegetation index products, computed as a weighted average of the best-quality observations within each month. The data used in this study spans from February 2000 to December 2025, closely aligning with the ERA5-Land dataset. Given the temporal mismatch of just one month at the beginning (January 2000 missing is in MODIS dataset), analysis in this study uses data starting from January 2001.

Only the 1 km monthly NDVI variable was employed for the subsequent analysis. Although the dataset also includes a 1 km monthly VI Quality layer, which provides information on the quality and reliability of the corresponding NDVI values, and can be used to assess or filter observations based on data confidence.

4.3 Preprocessing

After selecting and downloading the datasets, several preprocessing steps were carried out to prepare the data for modeling. This section begins by addressing temporal and spatial consistency, detailing how the datasets were aligned to a shared time window and cropped to ensure consistent geographic coverage. Although both datasets are provided at monthly resolution, they differ in spatial resolution, requiring additional steps to match their grids. Thus, second part details the steps of matching the spatial dimensions of the datasets, along with assessing the amount of error introduced by this process to ensure suitability for further analysis.

4.3.1 Temporal and spatial alignment

Temporal alignment was an important initial step to ensure MODIS NDVI and ERA5-Land datasets were comparable. The original time ranges of the datasets differed: ERA5-Land spanned from

January 2000 to April 2025, while MODIS NDVI ranged from February 2000 to January 2025. To avoid inconsistencies caused by incomplete or non-overlapping data coverage, the year 2000 was excluded from the analysis. As a result, the datasets were cut to a consistent and complete time period from January 2001 to December 2020. After the temporal alignment, both datasets consisted of monthly measurements of each month, resulting in identical time dimensions.

Despite being extracted over the same general region (Europe), the spatial extents of the two datasets differed slightly. ERA5-Land covered a broader grid, ranging from $30.0°N$ to $72.0°N$ latitude and from $30°W$ to $50°E$ longitude. In contrast, MODIS NDVI data spanned $34.28°N$ to $69.16°N$ in latitude and $24.69°W$ to $46.80°E$ in longitude. To ensure consistent spatial coverage, both datasets were cropped to their shared overlapping region before further processing.

The selected European study region contains numerous water bodies, expected to appear as missing data (NaNs) in both climate and vegetation datasets. Initially, the proportion of missing values appeared unusually high, around 46% in both datasets, which raised concerns about potential data processing errors. However, after generating spatial maps of the data, it was confirmed that the missing values were consistently located over oceans, lakes, and other non-terrestrial surfaces. The fact that the missing areas aligned precisely across both datasets indicates that these values result from intentional data masking of non-land pixels rather than processing errors.

4.3.2 Resolution matching

After temporal and spatial alignment, the next step involved matching the spatial resolutions of the ERA5-Land and MODIS NDVI datasets. The MODIS NDVI dataset has a much finer spatial resolution (0.0083°) compared to ERA5-Land (0.1°). For coherent modeling, both datasets must align spatially and share identical latitude and longitude grid points. This required regridding the NDVI data to match ERA5's coarser resolution.

Spatial regridding is the process of interpolating values from one grid resolution to another. Ideally, spatial regridding would be performed using the conservative regridding method provided by the xESMF library. This method calculates a true area-weighted average by considering all source grid cells (i.e., high-resolution MODIS pixels) that fall within or overlap a target ERA5 grid cell. It accurately accounts for varying pixel sizes due to Earth's curvature and is well-suited for continuous spatial variables like NDVI.

However, applying conservative regridding directly to the full-resolution MODIS dataset proved computationally infeasible due to the large data volume and fine spatial resolution. Therefore, a more efficient alternative was considered: coarsening followed by nearest-neighbor interpolation.

Coarsening is the process of reducing the spatial resolution of a dataset. In this step, the high resolution MODIS dataset was aggregated into a coarser grid by first, dividing the data grid into non-overlapping blocks of 11×11 pixels, and then, for each block, the average NDVI value was calculated by taking the mean of all valid pixels within the block. This reduced the resolution of the dataset from 0.0083° to approximately 0.1°, closely matching that of ERA5-Land. Coarsening heavily reduces the volume of data while preserving general spatial patterns.

Next, to enable direct pixel-wise comparison and modeling alongside ERA5-Land data, the coarsened MODIS grid had to be aligned precisely to the ERA5 spatial grid. This was achieved using nearest-neighbor interpolation: for each ERA5 grid point, the geographically nearest MODIS grid cell was identified, and its NDVI value was assigned to that point. This ensured that each ERA5 climate observation would be paired with a corresponding vegetation value at the same spatial

location. While nearest-neighbor interpolation uses Euclidean distance over lat/lon coordinates and does not account for Earth's curvature (unlike conservative regridding), it offers a substantial gain in computational efficiency by avoiding memory-heavy regridding, and it is expected to provide sufficient spatial accuracy for regional-scale analysis.

The result of these steps is a fully aligned dataset where climate and vegetation variables can be directly compared across the same 0.1° grid, enabling modeling and spatially consistent analysis.

To verify the validity of this simplified method, a comparative analysis was conducted between the coarsened and interpolated dataset and a smaller subset of the data that had been regridded using the more computationally intensive conservative method. The conservative regridding approach was selected as the reference because it is designed to most accurately preserve the original data values by accounting for the spatial distribution of the source grid. The evaluation involved calculating absolute differences between the two outputs on a pixel-by-pixel basis and several validation metrics were used to quantify the agreement and assess the potential loss of accuracy introduced by the simplified approach. These included Mean Absolute Error (MAE), the Pearson correlation coefficient, the maximum observed error, and a visual inspection of the distribution of absolute differences through histogram plots.

The MAE between the two regridded datasets was 0.0249, which is small relative to to the NDVI standard deviation of of the reference NDVI data (0.159), meaning that the more computationally efficient method introduced only minor distortions. Furthermore, the high Pearson correlation coefficient of 0.9688 between the two datasets confirmed that the chosen approach preserves overall spatial patterns of the original high resolution dataset.

However, the maximum absolute difference observed reached approximately 0.5, suggesting the presence of isolated cases where values differ significantly. These discrepancies are likely due to spatial averaging during the coarsening step, particularly in heterogeneous regions where NDVI values vary sharply over short distances, such as mountainous terrain.

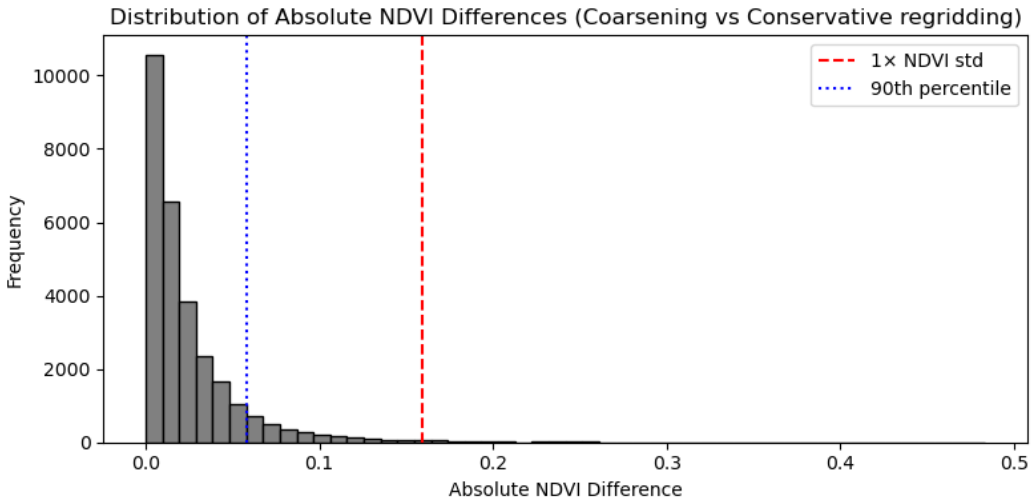


Figure 1: Distribution of absolute NDVI differences between the coarsened + interpolated and conservatively regridded datasets. The blue dashed line marks the 90th percentile, while the red dashed line shows the standard deviation of the reference dataset.

To further assess the distribution of differences between the two methods, a histogram of absolute NDVI differences was generated (see Fig. 1). The distribution reveals that most differences fall well below the NDVI standard deviation threshold (marked in red), indicating that large deviations are not frequent.

Together, these metrics provided a comprehensive assessment of both average and extreme deviations, as well as the overall agreement between the two regridding approaches. This validation provided confidence that the coarsened and interpolated MODIS dataset effectively retains the original spatial patterns while matching the ERA5 resolution, enabling subsequent modeling and analysis.

4.3.3 Final dataset structure

After completing all preprocessing steps, the final dataset was structured for modeling and analysis. Both ERA5-Land and MODIS NDVI cover the period from January 2001 to December 2020, yielding 240 monthly time steps. The spatial extent spans from $34.40°N$ to $67.60°N$ latitude and from $24.30°W$ to $46.40°E$ longitude, resulting in 333 latitude points and 708 longitude points. Spatial resolution of both datasets are matched to a common spatial resolution of $0.1° \times 0.1°$.

For modeling, the final structure of the dataset was obtained by flattening the original multi-dimensional arrays such that each row corresponds to a unique combination of time, latitude, and longitude. After removing samples with missing NDVI values (e.g., over water bodies), the final dataset consisted of approximately 14.6 million spatio-temporal samples, with 12 columns per row.

To make training and evaluation computationally manageable, a random subset of 100,000 samples was drawn separately from the training and testing periods. This sampled data was used to train the baseline Random Forest model, evaluate its predictive performance using metrics such as R^2 and RMSE, and support visual interpretation of model outputs. All model training throughout the study was conducted on this sampled training set. For the residual analysis, however, the full test dataset was utilized to assess spatial and temporal patterns of model error. This allowed for a comprehensive evaluation of residuals across all available spatio-temporal points, ensuring that no potentially meaningful deviations were overlooked due to sampling.

5 Modeling NDVI-climate dynamics

This chapter presents the modeling of the NDVI–climate relationship and the subsequent analysis of model residuals. It begins with the development and evaluation of a baseline model using Random Forest regression, including an assessment of feature importance to understand the contributions of different input variables. The chapter then takes a deeper look into the model residuals, examining their distribution, basic statistics, and identifying specific spatial and temporal patterns where the model failed to accurately predict NDVI. To isolate the most significant errors, a threshold-based approach is used to define and detect anomalous residuals, helping to pinpoint areas and periods with unusually large deviations. Finally, the chapter concludes with an interpretation of the results, discussing potential drivers behind the observed anomalies.

5.1 Baseline model

The primary objective is to build a predictive model that estimates expected NDVI values at specific locations and time points, using concurrent climate variables alongside contextual features such as seasonality and geographic coordinates.

5.1.1 Formalizing the prediction task and set up

The problem is structured as a supervised regression task, where the model learns a function:

$$f : \mathbb{R}^n \to \mathbb{R}$$

such that:

$$E(NDVI_{i,t}) = f(X_{i,t})$$

Where:

- i: spatial location index (latitude and longitude);
- t: time index (monthly resolution);
- $X_{i,t} \in \mathbb{R}^n$: input feature vector for location i and time t;
- $E(NDVI_{i,t}) \in [-1, 1]$: Predicted NDVI value at location i, time t.

Each observation corresponds to a unique location-time pair. The input feature vector $X_{i,t}$ a combination of climatic variables, as well as spatial and temporal context.

$$X_{i,t} = \begin{bmatrix} \text{t2m}_{i,t} \\ \text{tp}_{i,t} \\ \text{lai}_{i,t} \\ \text{sd}_{i,t} \\ \text{ssrd}_{i,t} \\ \text{lat}_i \\ \text{lon}_i \\ \text{month\_sin}_t \\ \text{month\_cos}_t \\ \text{year}_t \end{bmatrix}$$

Here, the year was linearly encoded as a feature to captures long-term trends and shifts over time. Latitude and longitude were included as raw features to allow the model to learn geographically varying relationships between climate and vegetation. To better represent seasonal variation, the month variable was encoded using sine and cosine transformations to reflect the cyclical nature of seasonal changes:

$$\text{month}_{\text{sin},t} = \sin\left(2\pi \frac{m_t}{12}\right)$$

$$\text{month}_{\text{cos},t} = \cos\left(2\pi \frac{m_t}{12}\right)$$

where $m_t \in \{1, 2, \ldots, 12\}$ denotes the calendar month at time t.

Unlike simpler linear encoding (e.g January = 1, December = 12), cyclic encoding maps months onto a unit circle. The twelve months are evenly spaced along the circle's circumference, preserving their cyclical nature. This avoids artificial discontinuities and ensures that December and January are close in the feature space, allowing the model to learn smooth seasonal patterns.

Additionally, total precipitation values were originally in meters and were converted to millimeters for interpretability, as well as a one-month lag of precipitation was applied to capture delayed vegetation response to moisture availability.

To assess the model's ability to generalize across time a temporal train-test split was applied. The model was trained on data from January 2001 to December 2006 and tested on data from January 2007 to December 2010. This evaluated the model's ability to predict NDVI in future periods based on past climate conditions. The goal of this split is to test model generalizability, a key criterion for a baseline prediction model.

5.1.2 Model description

A Random Forest Regressor was chosen as the baseline model due to robustness and ability to model nonlinear relationships [Bre01]. Random forests are ensemble models composed of multiple decision trees trained on bootstrapped samples of the data and a random subset of input features (feature subsetting). Each tree makes a prediction, and the forest's final output is average of all tree predictions. At each split in a tree, the model selects the feature and threshold that best reduce the mean squared error (MSE). This allows for the forest to learn flexible and nonparametric mappings from input features to NDVI values. While individual trees are easy to interpret, the ensemble as a whole is less transparent. However, Random Forests still offer useful interpretability through global metrics such as feature importance.

The model was implemented with 100 decision trees (n_estimators=100) for balance between predictive performance and computational cost. A maximum tree depth of 15 (max_depth=15) was set to prevent individual trees from becoming too complex and overfitting to noise. A minimum of 10 samples per leaf node (min_samples_leaf=10) was used for reduced variance. To ensure reproducibility, the random state was fixed (random_state=42), and parallel computation was enabled by using all available processors (n_jobs=-1) to accelerate model training.

5.1.3 Evaluation

The Random Forest model was trained to predict NDVI based on climate and contextual variables, and evaluated on a test set spanning 2007–2010 to assess its generalization capability. Model performance was quantified using two standard metrics: the coefficient of determination (R^2) and root mean squared error (RMSE). On the test set, the model achieved an R^2 of 0.832 and an RMSE of 0.097, indicating strong predictive accuracy despite the complexity and heterogeneity of vegetation–climate interactions.

To evaluate potential overfitting, the model was also assessed on the training data (2001–2006), where it achieved an R^2 of 0.899, giving a performance gap of 0.066. This relatively low difference indicates that the model generalizes well and does not simply memorize training patterns.

A scatter plot comparing predicted versus observed NDVI values (see Fig. 2) provides a visual assessment of model fit. Most points cluster near the 1:1 diagonal, indicating that the model

captures the overall trend reasonably well. However, some scatter remains, particularly in the low-NDVI range, where the model tends to overpredict. These patterns are further explored through a residual-based analysis in the following sections.

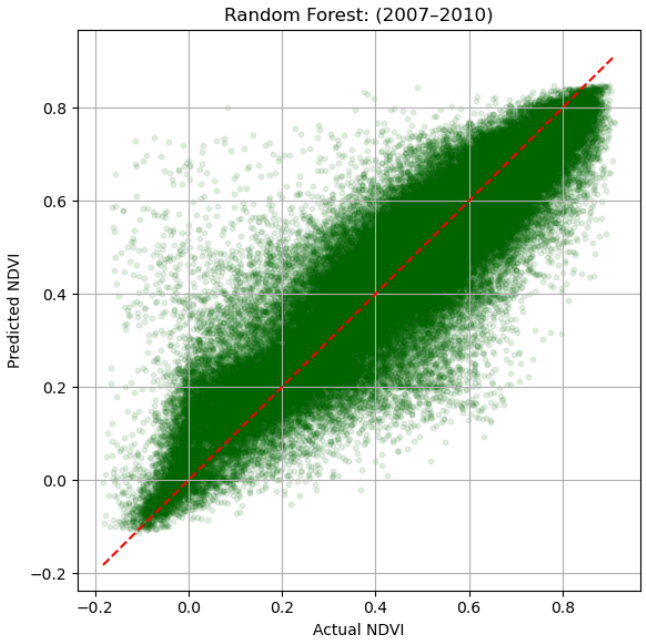


Figure 2: Predicted vs. observed NDVI values for 2007–2010. Points are 100,000 randomly sampled predictions from the test set. The red dashed line represents perfect prediction (1:1 line).

To better understand what drives the model's predictions, feature importance scores were extracted from the trained Random Forest (see Fig. 3). These scores are based on impurity reduction, also known as Gini importance. They reflect the average reduction in mean squared error (MSE) when a feature is used to split a node, aggregated across all trees in the ensemble. Notably, dew point temperature (d2m) emerged as the most influential feature by a wide margin, followed by latitude (lat) and leaf area index (lai_lv). Other variables like temperature (t2m), snow depth (sde) and lagged precipitation (tp_lag1) contributed moderately. Temporal features such as month and year were ranked quite low.

While this method provides a quick estimate of feature relevance, it is known to be biased toward continuous features or those with many unique values. To complement the built-in feature importance, TreeSHAP values were computed for a more robust and grounded interpretation of feature contributions [LL17]. SHAP (SHapley Additive exPlanations) is a unified framework, which attributes a model's prediction to individual input features by evaluating their average marginal contribution across all possible feature combinations. TreeSHAP is a specialized and efficient implementation of SHAP, tailored for tree-based models such as Random Forests. Unlike built-in importance scores, TreeSHAP can quantify the specific contribution of each feature to each individual prediction, offering a more refined understanding of model behavior.

Both Gini importance and TreeSHAP produced largely similar rankings of features, especially for the most influential variables. Dew point consistently emerged as the strongest predictor, followed by latitude and then Leaf Area Index in both methods. Since these approaches evaluate feature

relevance using different principles, agreement among them increases trust and suggests that the model consistently prioritizes the same features, regardless of the interpretation technique used.

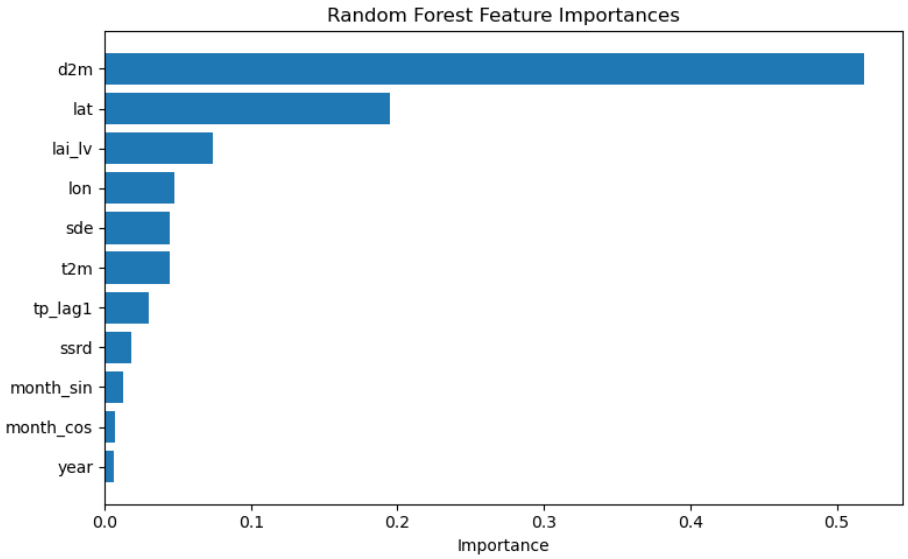


Figure 3: Random Forest feature importances ranked by contribution to reducing prediction error.

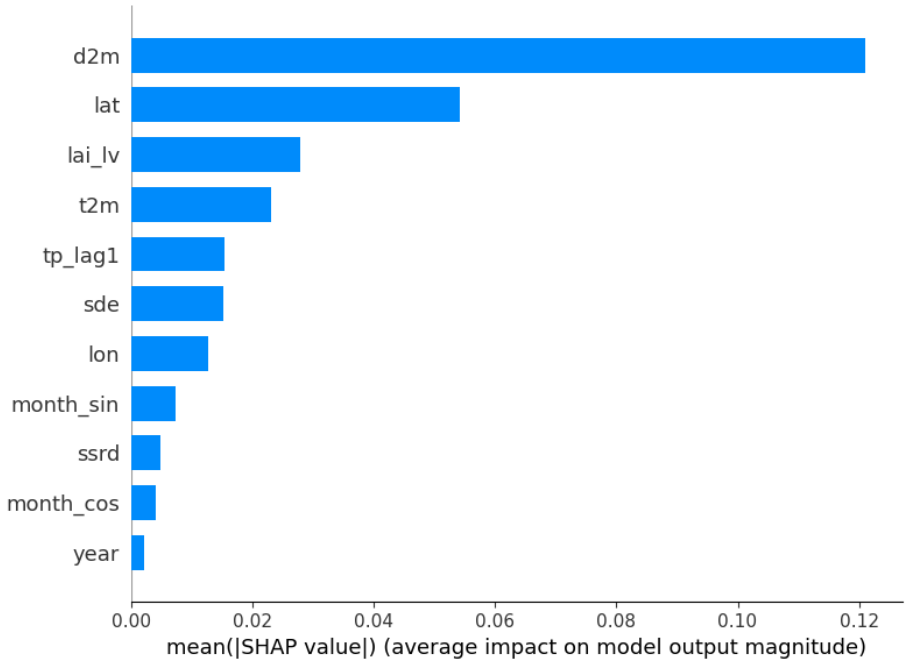


Figure 4: SHAP summary plot showing the average absolute impact of each input feature on the model's NDVI predictions. Features are ranked by their overall contribution magnitude across all test samples.

5.2 Anomaly detection via residuals

5.2.1 Residual statistics and distribution

To identify when and where the model deviated from the expected climate-NDVI relationships, residuals were computed as the difference between observed NDVI values and the NDVI values predicted by the Random Forest model:

$$Residual_{i,t} = NDVI_{observed;i,t} - NDVI_{predicted;i,t}$$

where i and t index spatial location and time, respectively.

A residual near zero indicates a good model fit. A large positive residual suggests underprediction — vegetation was greener than expected under given climate conditions, while a large negative residual indicates overprediction — the model overestimated greenness, predicting higher vegetation health than observed.

While the model was trained on a random sample of 100,000 spatio-temporal points to ensure computational efficiency, all subsequent residual analyses were conducted on the full test dataset, focusing on the test period (2007-2010).

Basic statistics give an overview of the residual distribution and help assess model performance:

- Mean: 0.0017;
- Mean Absolute Error: 0.0704;
- Standard deviation (σ): 0.0976;
- 10th percentile value -0.1098;
- 90th percentile: 0.1089.

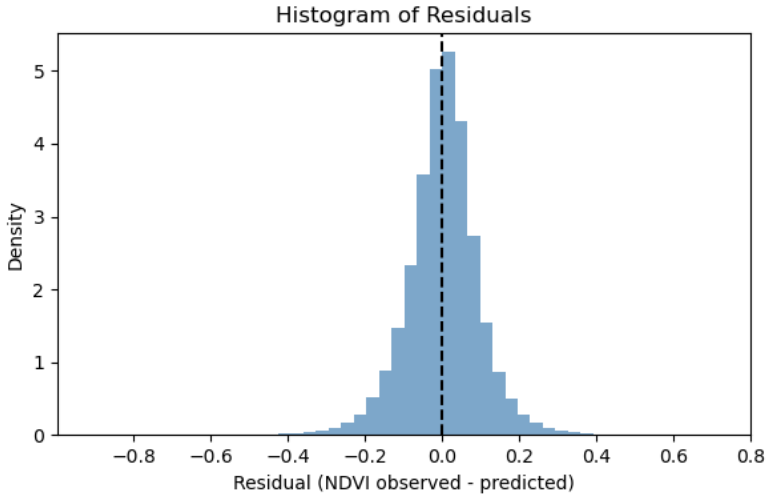


Figure 5: Distribution of residuals (NDVI observed – predicted) for test period 2007–2010.

The residual mean (0.0017), being signed, indicates that over- and underpredictions nearly balance out on average, but this can obscure the true magnitude of individual errors. To address

this, the mean absolute error (MAE = 0.0704) provides a more representative measure of typical prediction error magnitude, regardless of direction. The standard deviation reflects overall spread, while the 10th and 90th percentiles further characterize the range of most residuals. Consistent with these values, the majority of residuals lie within ±0.1 NDVI units.

The histogram (see Fig 5) shows the distribution that is nearly symmetric and bell shaped, centered around zero, though with slightly heavier tails. Around 94.9% of residuals lie within $\pm 2 \times \sigma$. A longer left tail suggests a skew towards negative residuals, overprediction in certain regions or time periods.

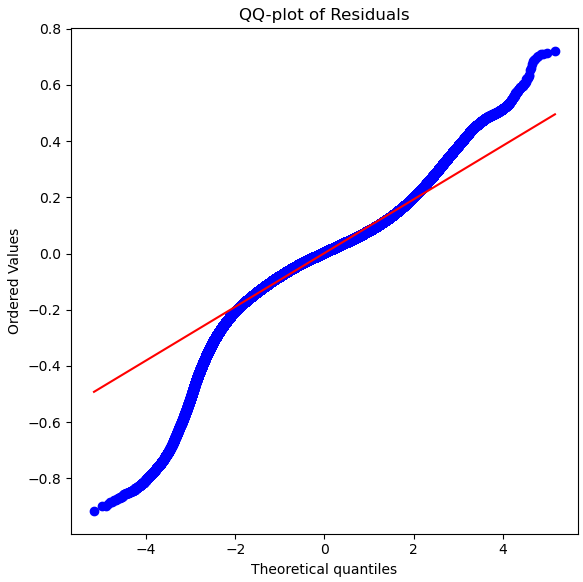


Figure 6: Q-Q plot of residuals versus theoretical normal distribution. The deviation from the red line at both tails indicates heavy-tailed behavior.

A quantile-quantile (Q-Q) plot was used to compare the empirical residuals to a theoretical normal distribution (see Fig 6). Although residuals from Random Forest models are not expected to be Gaussian, the Q-Q plot still serves as a useful diagnostic tool to visualize the spread and extremity of residuals. In this case, the central portion (around 0) of the distribution aligns well with the expectation, but deviations in the tails confirm non-normality. The 'S' shape suggests fat-tails, meaning that the model produces large errors, both under- and overpredictions, more frequently than would be expected under Gaussian assumptions.

5.2.2 Spatio-temporal residual patterns

To explore temporal variation in model performance, monthly mean residuals were calculated for the full test period (2007-2010) and time series was plotted (see Fig 7). Each point represents the average residual across all grid cell for a single month.

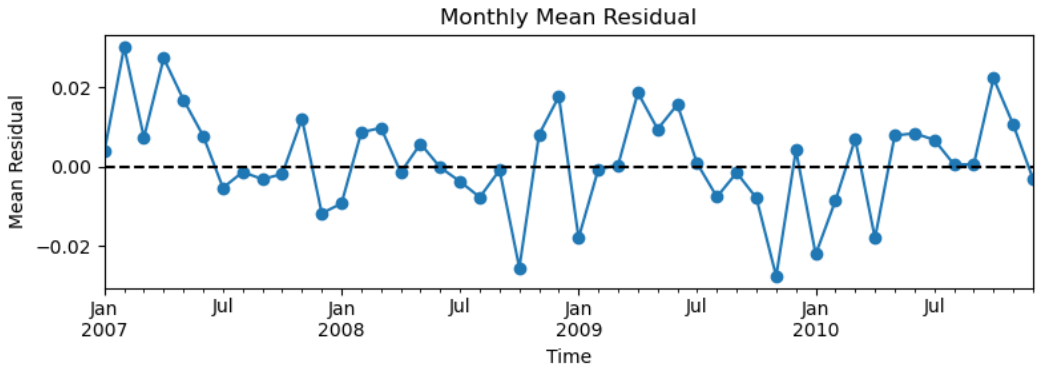


Figure 7: Monthly mean residuals (NDVI observed – predicted) from January 2007 to December 2010. Positive values indicate underprediction; negative values indicate overprediction.

Although overall residual mean remains close to zero, several patterns emerge. Notable peaks in model underprediction of NDVI occur in February and April of 2007, as well as April 2009, while overpredictions appear in October 2008, January of both 2009 and 2010, and November 2009. These deviations may indicate that the model's performance varies with season and year, possibly due to challenges in capturing certain vegetation–climate dynamics during specific periods. To quantify this, mean of absolute residuals and standard deviations across seasons and months were calculated (see Tab. 2 and Tab. 3)

The seasonal and monthly means reinforce the interpretation derived from the monthly plot. The means were calculated with absolute residuals and reflect the average magnitude of prediction errors, regardless of direction (over- or underpredictions), while the standard deviation indicates how variable the residuals are across pixels. Seasonal values are aggregated over all months belonging to each season, and monthly values are computed across the same calendar month for each year in the test period (2007–2010).

Residuals are consistently lowest during late summer and early autumn, with July, August, and September, showing the smallest mean absolute errors (e.g., July: 0.0042, August: 0.0043) and standard deviations. This suggests the model performs best during periods of peak vegetation activity and relative climatic stability. In contrast, residuals are higher and more variable in spring and late autumn particularly in April (Mean = 0.0164, Std = 0.0204) and October–November, suggesting difficulty in modeling vegetation transitions, such as greening or vegetation decline. Winter months also show elevated errors, possibly due to snow cover, low sun angles, and mixed surface conditions, which complicate both NDVI retrieval and modeling.

Table 2: Seasonal mean and standard deviation of monthly absolute NDVI residuals (2007–2010)

Season	**Mean Residual**	**Std. Dev.**
Winter	0.0115	0.0149
Spring	0.0108	0.0113
Summer	0.0055	0.0072
Autumn	0.0102	0.0145

Table 3: Monthly mean and standard deviation of monthly absolute NDVI residuals (2007–2010)

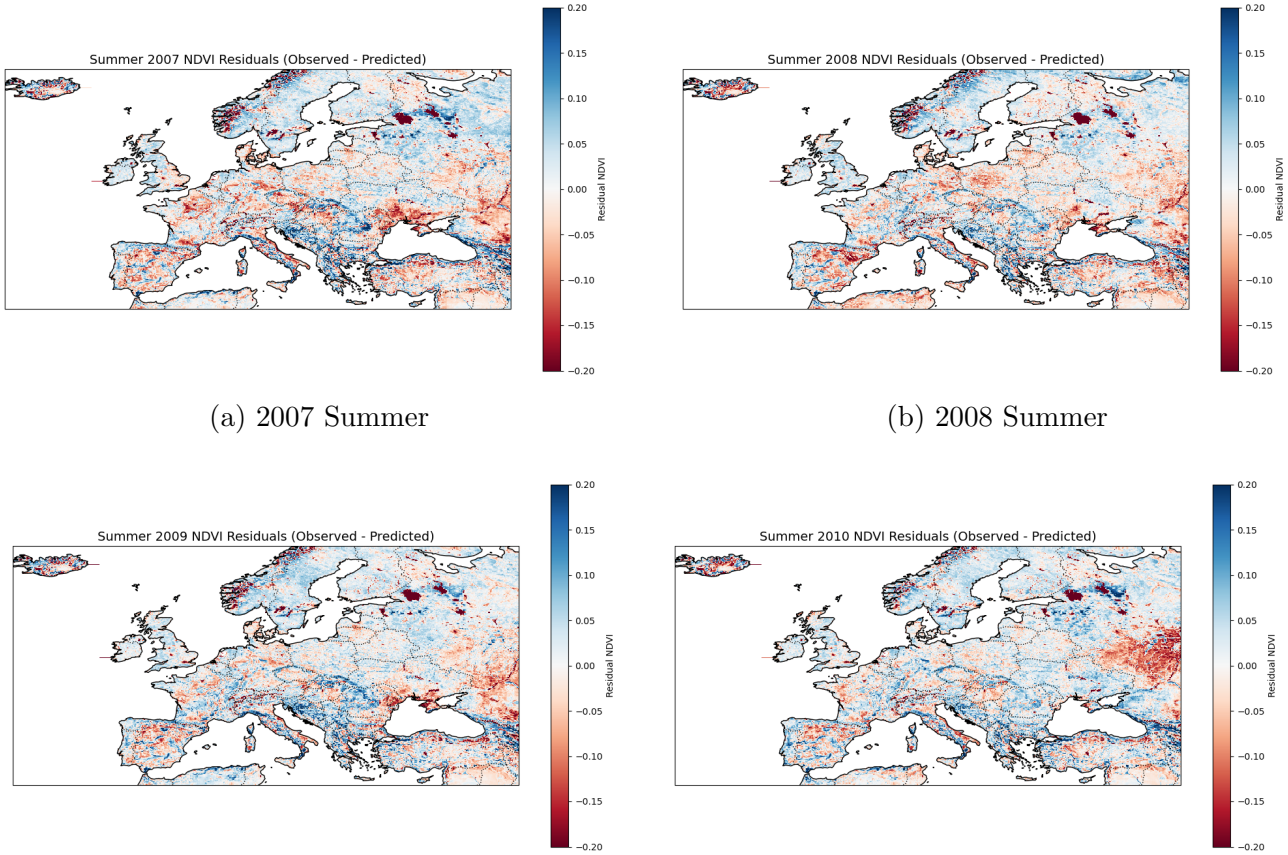
Month	**Mean Residual**	**Std. Dev.**
January	0.0133	0.0115
February	0.0120	0.0167
March	0.0061	0.0040
April	0.0164	0.0204
May	0.0100	0.0048
June	0.0080	0.0065
July	0.0042	0.0054
August	0.0043	0.0043
September	0.0015	0.0015
October	0.0144	0.0198
November	0.0146	0.0190
December	0.0092	0.0125

Residual analysis showed that some of the lowest mean errors and standard deviations occurred during late summer months (notably July and August), suggesting that model performance may be relatively more stable during this period, perhaps due to more stable vegetation conditions and fewer sources of noise in both the data and the modeling process. While this observation alone was not conclusive, it contributed to the decision to focus subsequent analysis on the summer months. Specifically, only June through August were used. This choice was also motivated by practical considerations, as visualizing and interpreting residual patterns across all months would be less feasible and potentially less informative. Focusing on a shorter, climatically stable period made spatial patterns easier to interpret.

To investigate spatial patterns of model performance, mean residual maps were generated for summer months of each test year (2007-2010) (see Fig. 8). These heatmaps show the mean residual per pixel and capture deviations between predicted and observed NDVI across Europe.

Red areas indicate overpredictions, while blue areas indicate underpredictions. The color scale is symmetric and capped at ± 0.20 NDVI units to highlight strong residuals. Visual patterns suggest recurring overpredictions in some mountainous areas (e.g., Norway and the Alps) and underpredictions along several coastal regions, such as the Balkans. While these patterns are not universally consistent and residuals occur across all landscapes, they hint at systematic model biases potentially tied to complex terrain or coastal dynamics. In mountainous regions, mispredictions may result from unresolved topographic effects on vegetation phenology, microclimates, or limitations in how NDVI responds to steep terrain. Mispredictions in coastal areas could stem from edge effects, mixed land pixels, or unique coastal vegetation dynamics not captured by the current predictors.

Some residual patterns appear consistently across multiple years, suggesting stable model biases potentially linked to unresolved spatial features. Others vary in location and intensity over time, indicating year-specific deviations that could reflect changing climate conditions or transient ecological processes.



(a) 2007 Summer

(b) 2008 Summer

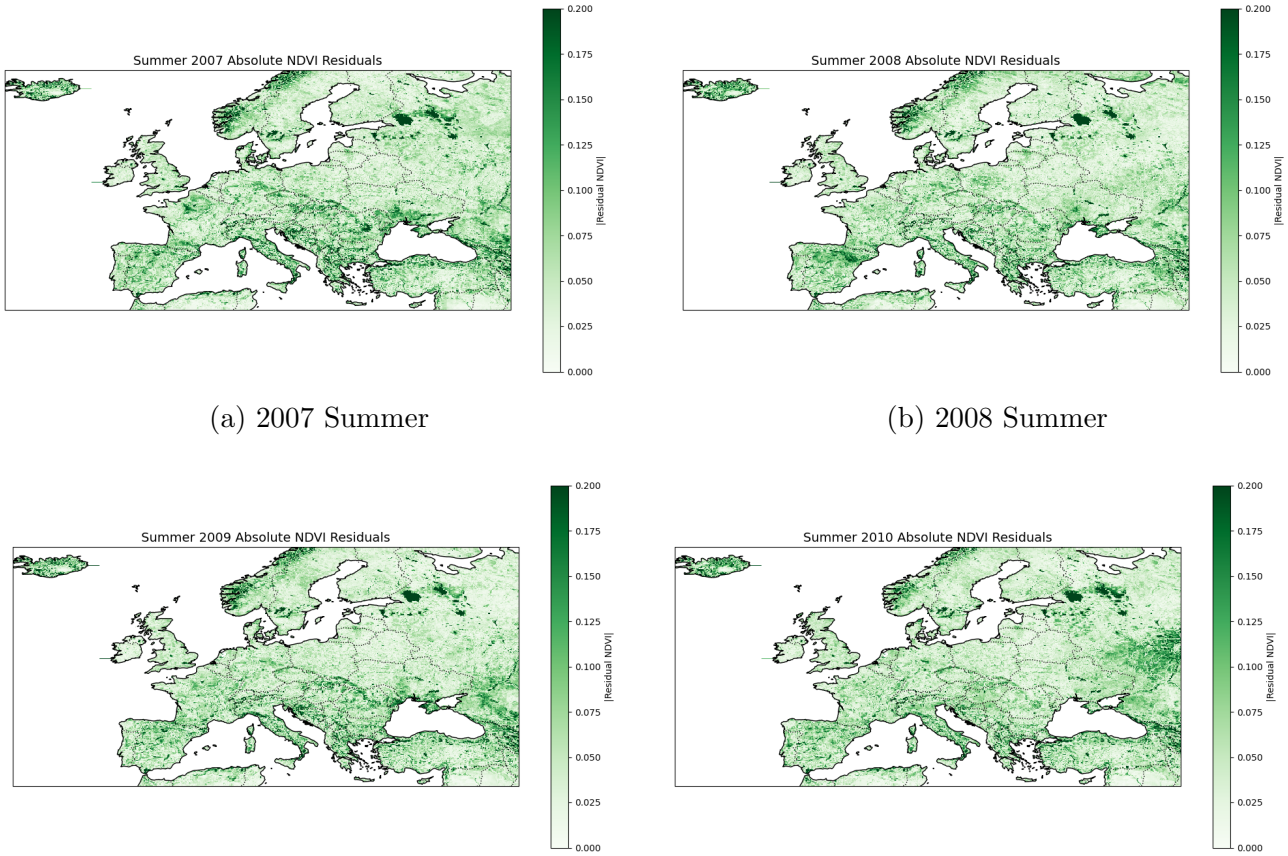
(c) 2009 Summer

(d) 2010 Summer

Figure 8: Mean residuals per pixel during Summer (June-August) for 2007, 2008, 2009 and 2010; capturing both overpredictions and underpredictions.

To complement the signed residual analysis, absolute mean residuals were also computed (see Fig. 9). These plots help avoid the risk of cancellation between over- and underpredictions when aggregating over time, since they take the absolute value of the residual. They provide a clearer view of the magnitude of model errors.

Interestingly, the absolute residuals follow a similar seasonal and spatial pattern to the signed ones, which gives confidence in the robustness of the observed error in the signed mean residual maps. For instance, in 2010, the western Russian region continues to show elevated residuals in both signed and absolute maps. This could suggest that this anomaly is not an artifact of averaging, but reflects a substantial deviation in NDVI behavior.



(a) 2007 Summer

(b) 2008 Summer

(c) 2009 Summer

(d) 2010 Summer

Figure 9: Mean absolute residuals per pixel during Summer (June-August) for 2007, 2008, 2009 and 2010.

5.2.3 Anomalous residuals

To shift from general residual analysis toward identifying specific breakdowns in model performance, this study defines and analyzes 'anomalous' residuals, those that exceed a set threshold. Rather than interpreting every residual equally, this step aims to isolate only the most extreme and potentially meaningful mismatches between observed and predicted NDVI values.

There are several approached to identifying anomalies in a residual distribution. In this study, a fixed statistical threshold based on standard deviation (σ) of the residuals is adopted. Specifically, a residual is defined as anomalous if:

$$|Residual| > 2 \times \sigma$$

Given that the empirical standard deviation of the residuals is approximately 0.09, this results in an anomalous threshold of about ± 0.18. Any prediction error beyond this margin is flagged as an anomaly.

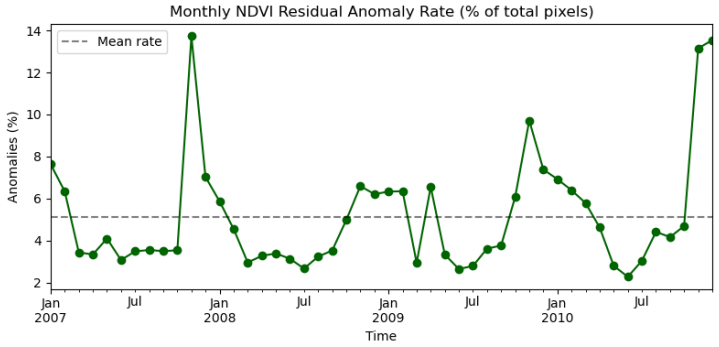


Figure 10: Monthly NDVI Residual Anomaly Rate as a percentage of total valid pixels per month. Anomalies are defined based on the absolute residual exceeding a threshold, regardless of direction (over- and underpredictions). The dashed line indicates the mean anomaly rate over the test period.

To assess how often and when these anomalies occur, for each month in the test period (2007–2010), the percentage of anomalous residuals is computed by first, flagging as anomalies all pixels with residuals exceeding ±2 standard deviations from the mean, and then, for each month, the number of anomaly-flagged pixels is divided by the total number of valid pixels, yielding the monthly anomaly rate as a percentage. This helps reveal seasonal or monthly peaks in model deviations. The results are shown in Fig. 10. A seasonal pattern is somewhat visible again: the model is generally more accurate during summer seasons. Peaks in anomaly rates mostly align with the periods of low vegetation cover. Notable peaks are 2007 November, 2009 November, 2010 November and December, each reaching well above the average anomaly rate.

While temporal trends help evaluate how model performance varies across seasons, spatial patterns of anomaly frequency reveal where the model repeatedly fails. For this, anomaly flags were summed for each pixel across all summer months from 2007 to 2010. A pixel's value in the resulting map (Fig. 11) indicates how many times the absolute residual at that location exceeded the anomaly threshold, regardless of direction (over- or underprediction). This highlights areas with frequent high-magnitude errors, displaying spatial patterns of consistent model failures.

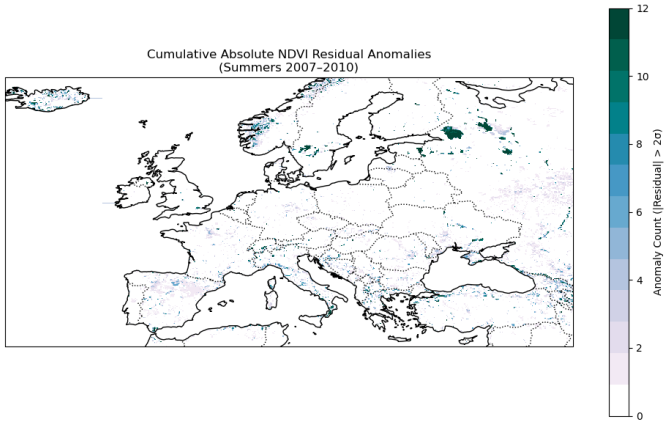
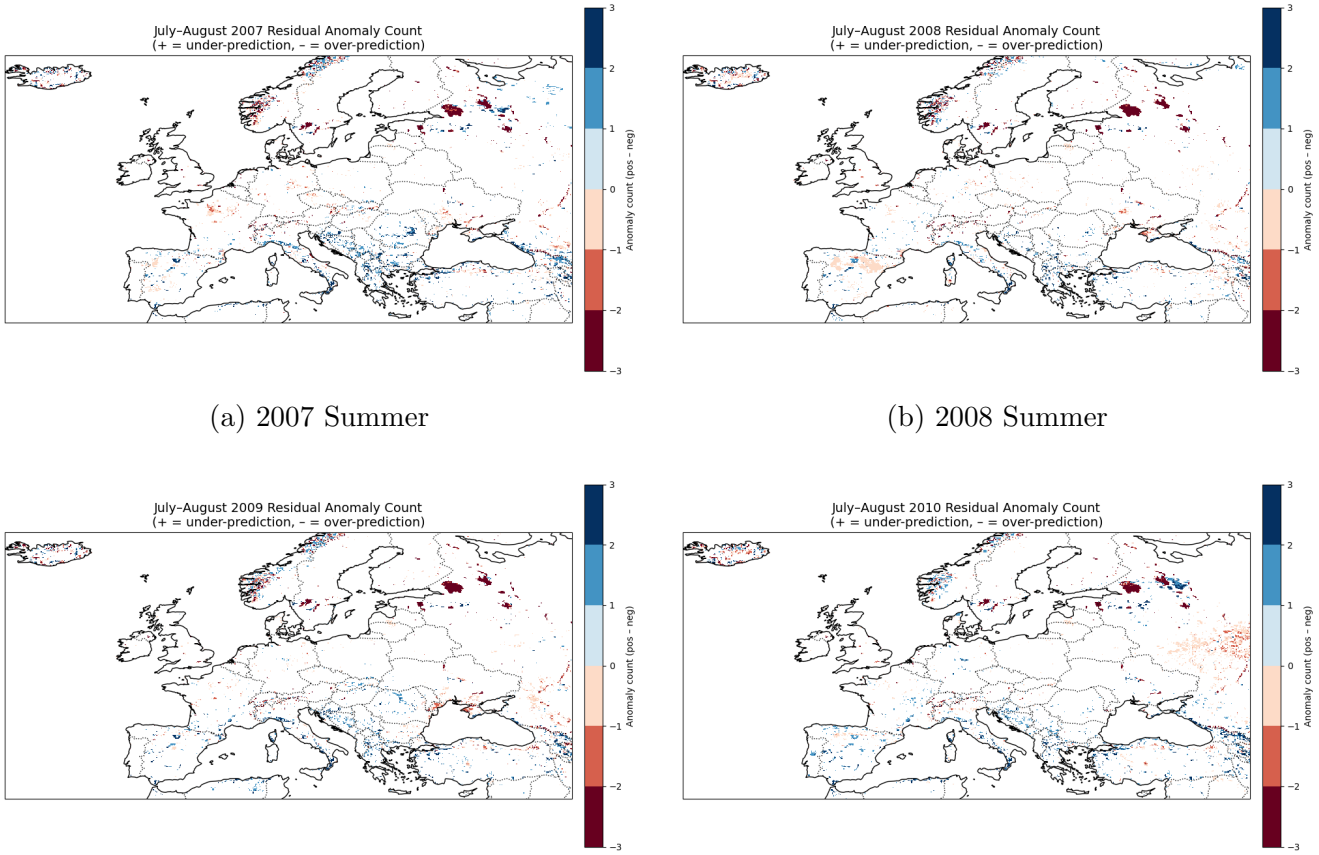


Figure 11: Cumulative count of absolute residual anomalies per pixel during summer months (June–August), aggregated over 2007–2010. Only residuals exceeding the $\pm 2\sigma$ threshold are included. The color indicates how many summer months each pixel was flagged as anomalous.



(a) 2007 Summer

(b) 2008 Summer

(c) 2009 Summer

(d) 2010 Summer

Figure 12: Residual anomaly counts per pixel for summer 2007, 2008, 2009, 2010. Colors indicate how many times (out of three summer months per year) each pixel exceeded the anomaly threshold, with the sign denoting whether NDVI was underpredicted (positive) or overpredicted (negative).

While the aggregated anomaly map is useful for identifying areas where the model consistently performs poorly during summer months, it does not reveal the year-to-year variation in model performance. To address this, separate anomaly maps for each individual year were produced (see Fig. 12). These maps display the residual anomaly counts per pixel for summer months, allowing for observation of interannual shifts in model behavior. The spatial patterns in these maps resemble those seen in the earlier summer residual heatmaps, but here they specifically highlight instances where prediction errors exceed a defined threshold. It is important to note that because these year-specific maps retain the sign of the residuals, over- and underpredictions occurring in different summer months at the same location may cancel each other out, potentially obscuring the full frequency of anomalous behavior at that pixel.

5.3 Interpretation of results

5.3.1 Insights from the baseline model

The baseline model shows that a relatively simple machine learning approach, using only a handful of climate and contextual variables, can account for a substantial portion of NDVI variability across

Europe. The high R^2 and low RMSE indicate that much of the spatial and temporal variability in NDVI can be captured by climate variables and basic spatio-temporal features. This supports the notion that vegetation activity, as measured by NDVI, is highly climate-driven at broad scales.

Notably, the scatter plot exposes frequent overprediction of NDVI. This "upper-left wedge" of the plot probably corresponds to winter, late autumn and early spring months, where the model interprets normal climate conditions as suitable for growth, but NDVI drops due to dormancy, snow cover, or low solar angles, factors that are either unmodeled or poorly represented in the input variables. There are also visible cases where the model predicted low NDVI, but observed NDVI was higher (underpredictions), but much less pronounced.

The importance of dew point temperature (d2m) tells us that near-surface moisture availability is the dominant predictor of vegetation greenness in this setting. This outcome is not just a reflection of its raw predictive power, but also of how much unique information d2m provides that cannot be substituted by any other variables. If d2m were to be removed from the model, temperature would likely take its place as the primary predictor. And in this case, temperature (t2m) is ranked quite low, because d2m captures near-surface moisture and is strongly correlated with relative humidity and indirectly with precipitation and temperature. So, the model leverages it as a kind of proxy for vegetation water stress, especially during growing seasons. The relatively low importance assigned to temperature (t2m) is therefore not a sign that temperature is irrelevant to NDVI, but more so that, in the presence of d2m, temperature adds little extra predictive value. It could also imply that drought signals are embedded in this variable (d2m) and the model could be capturing vegetation stress indirectly through it without explicitly modeling drought. This highlights an important limitation in interpreting Random Forest feature importances: they are conditional on the presence of other features.

With the addition of TreeSHAP analysis alongside the built-in Random Forest feature importance, the dominance of dew point temperature (d2m) as the top predictor remains consistent and further validated. Both methods agree that near-surface moisture availability plays a crucial role in explaining vegetation greenness across Europe. The low importance of temporal variables of month and year is also consistent across both methods. However, it's worth mentioning that month was encoded using two cyclic components (sine and cosine), which could dilute the perceived impact of each component individually, even if together they capture seasonal effects. Additionally, much of the seasonal and interannual variation is already embedded within the climate variables, making the standalone contribution of such calendar-based variables minimal in both methods.

5.3.2 When did the model fail?

The residuals provide insight not just into model accuracy, but also into where and when the model misrepresents NDVI values. The histogram of residuals shows a relatively symmetric, bell-shaped distribution centered around zero, indicating that most errors are small and unbiased on average. Despite strong overall performance, the slightly left-skewed residual distribution suggests a mild tendency to overestimate vegetation greenness. The QQ plot reveals fat tails, meaning that both extreme over and under predictions happen more often than expected under a Gaussian assumption. While these deviations are informative, they are not entirely unexpected. In complex interactions such as the vegetation-climate one, residuals are rarely normally distributed due to the presence of nonlinear dynamics and unmodeled factors.

The monthly residual time series shows that, while residuals generally remain close to zero, several

months stand out with notable peaks, particularly in February, April, October and November. These patterns suggest that the model struggles more during transitional periods. The underpredictions in spring likely reflect the model's inability to fully capture the rapid green-up that happens with certain phenological triggers such as day length. In contrast, winter and autumn residuals could be due to mixed surface conditions, snow cover, or low sun angles, which complicate NDVI retrieval and may lead the model to default toward average values. The monthly residual time series is based on signed residuals and reflects the direction of prediction errors. The over- and underpredictions could cancel each other out, since they are aggregated over space. This makes it useful for identifying broad temporal patterns but potentially hides the true magnitude of model errors.

To further quantify model performance across time, seasonal and monthly mean absolute residuals and their standard deviations were computed. These values avoid the cancellation effect of signed residuals and instead reflect the true magnitude of errors. The results confirm the pattern observed in the time series: summer and early autumn months consistently show the lowest error magnitudes and variability, while other seasons, especially winter, exhibit higher mean errors and standard deviations. This indicates that the model performs most consistently and reliably in months of July to September, while transitional and cold seasons introduce greater uncertainty.

To extend the residual analysis and better capture extreme deviations, a monthly anomaly rate plot was produced based on the percentage of test set pixels with residuals exceeding ± 2 standard deviations. This plot revealed a strong seasonal trend, with anomaly rates peaking during late autumn and winter months, such as November 2007, November 2009, and December 2010. These periods are typically marked by more variable surface conditions and reduced vegetation activity, which can complicate NDVI interpretation and contribute to higher model uncertainty. By flagging unusually large count of prediction errors, the anomaly rate plot complements the residual statistics and supports the interpretation that the model performance varies systematically across seasons.

5.3.3 Where did the model fail?

The spatial heatmaps of residuals during the summer months (2007–2010) provide important insight into where the model fails. Two types of residual maps were produced: signed residuals (which retain directionality and whether NDVI was over- or underpredicted) and absolute residuals (which reflect only the magnitude of errors). Both types show highly similar spatial structures, reinforcing the reliability of the observed patterns.

Across all years, mountainous regions of Norway exhibit persistent patterns of both high positive and negative residuals. These are likely linked to snow cover, steep terrain effects, or low solar angles, all of which can startle NDVI interpretation and modeling. Southern Europe, especially parts of Spain, Italy, and Greece, displays consistently high residuals in summer, both over- and underpredictions. This likely reflects the model's difficulty in accurately capturing vegetation dynamics in drought-prone Mediterranean areas, where variability in water availability and land cover can lead to abrupt and nonlinear NDVI responses during dry periods in summer.

The persistence of certain residual patterns across multiple years, such as overpredictions in mountainous regions and underpredictions along coastal zones, points to structural limitations in the model. In mountainous areas, residuals may reflect unresolved topographic effects, microclimatic variability, or how NDVI behaves over steep terrain. Along coastlines, edge effects, mixed land–water pixels, or unique coastal vegetation dynamics may not be well captured by the current feature set. These spatial patterns are also evident in the training set predictions,indicating that they likely

stem from inherent mismatches between the predictors and the real-world complexity of these landscapes, rather than pure generalization failure.

The 2010 residual map, both the signed and absolute versions, stands out with large overpredictions across western Russia, consistent with the well-documented extreme heatwave and drought that occurred that summer [aD11]. In both representations, residuals are unusually high, indicating a consistent failure of the model to capture the actual vegetation response. The signed maps reveal that these are primarily overpredictions, where the model expected higher NDVI than observed, while the absolute maps reinforce the magnitude of the error without the risk of directional cancellation.

A potential cause of model failures in certain years or regions is domain shift, when the test data falls outside the distribution of conditions the model was trained on. Since the model was trained on a limited historical window (2001–2006), it may not have encountered the full range of climatic extremes or vegetation responses present in later years. This mismatch can lead to a breakdown in the learned climate–vegetation relationships, resulting in large residuals. Such predictive failures can point to ecologically significant deviations, whether due to extreme events, emerging trends, or limitations in model representation.

Additional visualizations were used to assess the model's performance in detecting significant mismatches. A cumulative map of absolute NDVI residual anomalies across the four summer seasons (2007–2010) revealed where model errors repeatedly exceeded the anomaly threshold ($|residual| > 2 \times \sigma$). This plot highlights persistent regions of high error magnitude with darker colors indicating more frequent failures. When viewed by year, the anomaly count maps capture the frequency and direction of anomalies per pixel, showing both systematic and transient breakdowns. Several consistent spatial patterns emerge across the four years. Mountainous Norway shows recurring clusters of anomalies. Southern Europe, particularly Spain, Italy, and the Balkans, also show recurrent anomalies, but these include both under- and overpredictions. In Central and Eastern Europe, the anomaly signal varies more year to year, but certain regions (e.g., western Russia, Ukraine) show clear spikes, especially in 2010, aligning with the extreme heatwave and drought. These results were also observed previously in residual heatmaps.

6 Conclusion

This thesis aimed to model the relationship between vegetation health, measured by NDVI, and key climatic variables across Europe, using machine learning. Beyond predictive performance, the study set out to uncover whether residuals, differences between observed and predicted NDVI, could reveal breakdowns in expected climate–vegetation relationships and provide insight into unmodeled processes or unexpected events.

Using Random Forest regression, a baseline model was built that achieved strong generalization performance across the test set. This demonstrates that key climate variables can predict NDVI with high accuracy across diverse European regions, addressing the first research question. Dew point temperature (d2m) emerged as the most important predictor, which highlights the dominant role of near-surface moisture in shaping vegetation activity.

To answer the second research question of where and when these modeled relationships failed, residual analysis was conducted. The results showed that the model tends to overpredict NDVI

during early spring, winter and late autumn months, likely due to unresolved influences such as snow cover, low sun angles, or rapid phenological transitions. Shifting focus to the summer months, when NDVI is typically more stable and easier to interpret, allowed for clearer detection of model failures. Aggregated spatial maps highlighted regions where the model frequently produced high-magnitude residuals, including parts of mountainous terrain and coastal areas. In specific years, such as 2010, notable residual anomalies appeared across western Russia, suggesting that the model struggled to generalize under extreme or previously unseen climate conditions.

These deviations may reflect domain shifts or external influences not captured by the model, such as extreme heatwaves or droughts, offering a possible explanation for the breakdowns observed and addressing the third research question. While it is not possible to definitively attribute all large residuals to specific causes, cases like the 2010 overpredictions in western Russia align with known climate extremes and suggest that such anomalies could reflect real ecological disturbances rather than mere model error.

In summary, this thesis not only shows that vegetation health across Europe can be effectively modeled using climate data, but also that residuals offer valuable signals. By analyzing when and where model predictions fail, the study demonstrates how machine learning models can be used as diagnostic tools to identify potential anomalies in vegetation–climate dynamics.

7 Limitations and future work

While this study demonstrates the utility of machine learning for modeling vegetation–climate interactions and detecting their breakdowns, several limitations constrain the robustness, interpretability, and generalizability of the findings.

One key limitation stems from the input features and preprocessing pipeline. Although ERA5 variables like dew point temperature, precipitation, and radiation are core climatic drivers, important factors such as soil moisture, vegetation type, and land management are not included. The lack of spatially explicit variables like land cover, elevation, or slope likely contributes to systematic model biases observed in heterogeneous areas, such as agricultural zones, coasts, or mountainous regions, where vegetation responses vary more sharply over space. Residual analysis further revealed that high model errors are not restricted to unseen (test) data: certain training regions also show consistently poor fits. This suggests that the model struggles even when fitting familiar conditions, possibly due to missing explanatory variables or structural limitations in the model itself. Addressing these issues may require expanding the feature set to include topographic and land use data, or adopting models better equipped to learn fine-scale spatial variation. Additionally, while MODIS and ERA5 were carefully regridded and aligned, the coarsening and interpolation procedures may introduce minor spatial artifacts. Although validation confirmed minimal distortions, any averaging over complex landscapes could obscure fine-scale vegetation responses, particularly near land–water boundaries or abrupt terrain changes.

The use of Random Forest regression offered strong performance and a degree of interpretability via feature importances. However, these models are not inherently designed to capture spatio-temporal dependencies. For example, neighboring pixels or adjacent months are treated as independent, limiting the model's ability to learn dynamic vegetation trajectories or spatially continuous phenomena. Future studies could explore architectures that incorporate explicit spatial and temporal structure, such as convolutional neural networks (CNNs), to better represent ecosystem memory,

climate lags, and cross-boundary vegetation responses.

A further limitation lies in the restricted temporal coverage of the training data (2001–2006), which may hinder the model's ability to generalize to more recent climate conditions or capture evolving ecological responses. This is particularly relevant for interpreting outlier years like 2010, where large NDVI overpredictions were observed in western Russia. While this pattern aligns with a documented heatwave and drought, more robust evidence is needed to attribute the anomaly to extreme temperature. To strengthen this link, a follow-up analysis could involve: (1) sensitivity testing - retraining the model on varied subsets of the training years to verify the persistence of 2010 anomalies; and (2) distributional analysis - comparing temperature distributions in the training set with 2010 test conditions. If temperature values in 2010 lie outside the training range, this would suggest domain shift as a likely cause of model failure and support the interpretation of the residuals as climate-driven anomalies.

Finally, residuals are an indirect signal: they capture mismatches between modeled and observed behavior, but interpreting them as meaningful anomalies assumes that the model accurately represents expected conditions. In reality, residuals can also arise from sensor noise, data quality issues, or resolution mismatches. While this study attempts to isolate meaningful anomalies by focusing on the summer period and applying strict thresholds, some of the detected deviations may still reflect non-ecological artifacts. More rigorous anomaly detection methods could improve reliability, especially those that better account for seasonal variability or validate anomalies against independent sources of information.

Despite these limitations, the study provides a foundation for residual-based ecological anomaly detection using predictive models. Future work can refine the approach by incorporating richer datasets, using more flexible spatio-temporal models, and performing targeted validation. The combination of predictive modeling and residual interpretation remains a promising path for understanding not only how climate shapes vegetation, but also when and where that relationship breaks down.

References

[aD11] Luterbacher J Trigo RM García-Herrera R. arriopedro D, Fischer EM. The hot summer of 2010: Redrawing the temperature record map of europe. *Science*, 332(6026):220–224, 2011.

[Bre01] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[(C320] Copernicus Climate Change Service (C3S). Agrometeorological indicators from 1979 to present derived from reanalysis, 2020.

[CBK09] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[CW11] Noel Cressie and Christopher K Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2011.

[Did21] Kamel Didan. Modis/terra vegetation indices monthly l3 global 1km sin grid v061, 2021.

[EUA+25] Christina Eisfelder, Soner Uereyen, Sarah Asam, Andreas Hirner, Philipp Reiners, Juliane Huth, Felix Bachofer, Martin Bachmann, Stefanie Holzwarth, and Claudia Kuenzer. Thirty-year analyses of seasonal ndvi and climatic drivers across different land cover types and biogeographical regions in europe. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[FWM+24] Loghman Fathollahi, Falin Wu, Reza Melaki, Parvaneh Jamshidi, and Saddam Sarwar. Global normalized difference vegetation index forecasting from air temperature, soil moisture and precipitation using a deep neural network. *Applied Computing and Geosciences*, 23:100174, 2024.

[KEUR+18] Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554, 2018.

[LL17] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[MRW+24] Kaleem Mehmood, Sallahuddin Rahman, Bakhtiar Wagan, Ghazala Arain, Imran Ahmed, Sajid Iqbal, Rizwan Jamali, and Muhammad Fayaz Memon. Analyzing vegetation health dynamics across seasons and regions through ndvi and climatic variables. *Scientific Reports*, 14(1):11817, 2024.

[MS19] J. Muñoz Sabater. Era5-land monthly averaged data from 1950 to present, 2019.

[NKH+03] Ramakrishna R Nemani, Charles D Keeling, Hirofumi Hashimoto, William M Jolly, Stephen C Piper, Compton J Tucker, Ranga B Myneni, and Steven W Running. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *science*, 300(5625):1560–1563, 2003.

[PFY+19] Zhifang Pei, Shibo Fang, Wunian Yang, Lei Wang, Mingyan Wu, Qifei Zhang, Wei Han, and Dao Nguyen Khoi. The relationship between ndvi and climate factors at different monthly time scales: a case study of grasslands in inner mongolia, china (1982–2015). *Sustainability*, 11(24):7243, 2019.

[PLC+19] Shilong Piao, Qiang Liu, Anping Chen, Ivan A Janssens, Yongshuo Fu, Junhu Dai, Lingli Liu, XU Lian, Miaogen Shen, and Xiaolin Zhu. Plant phenology and global climate change: Current progresses and challenges. *Global change biology*, 25(6):1922–1940, 2019.

[PVM+05] Nathalie Pettorelli, Jon Olav Vik, Atle Mysterud, Jean-Michel Gaillard, Compton J Tucker, and Nils Chr Stenseth. Using the satellite-derived ndvi to assess ecological responses to environmental change. *Trends in ecology & evolution*, 20(9):503–510, 2005.

[RBV+94] Bradley C Reed, Jesslyn F Brown, Darrel VanderZee, Thomas R Loveland, James W Merchant, and Donald O Ohlen. Measuring phenological variability from satellite imagery. *Journal of vegetation science*, 5(5):703–714, 1994.

[RCVS+19] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

[RNH+04] Steven W Running, Ramakrishna R Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, 54(6):547–560, 2004.

[SLC+22] Qianqian Sun, Chao Liu, Tianyang Chen, Anbing Zhang, Chunyang Liu, and Yuan Tao. Adaptive decomposition and multitimescale analysis of long time series of climatic factors and vegetation index based on iceemdan-svm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6203–6219, 2022.

[VSGC+13] Sergio M Vicente-Serrano, Célia Gouveia, Jesús Julio Camarero, Santiago Beguería, Ricardo Trigo, Juan I López-Moreno, César Azorín-Molina, Edmond Pasho, Jorge Lorenzo-Lacruz, Jesús Revuelto, et al. Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences*, 110(1):52–57, 2013.

[YDO+15] Genesis T Yengoh, David Dent, Lennart Olsson, Anna E Tengberg, and Compton J Tucker III. *Use of the Normalized Difference Vegetation Index (NDVI) to assess Land degradation at multiple scales: current status, future trends, and practical considerations*. Springer, 2015.

[ZTK+01] Liming Zhou, Compton J Tucker, Robert K Kaufmann, Daniel Slayback, Nikolay V Shabanov, and Ranga B Myneni. Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *Journal of Geophysical Research: Atmospheres*, 106(D17):20069–20083, 2001.