Satellite NDVI Forecasting with Growth Curve Regression

A Deep Learning Framework for Long-Horizon Vegetation Dynamics Prediction

Antonio Henrique Xavier da Silva

January 2026

Abstract

This thesis presents a novel deep learning framework for forecasting vegetation dynamics using multi-spectral Sentinel-2 satellite imagery, meteorological data, and land cover classification. Unlike conventional approaches that predict vegetation states iteratively step-by-step, the proposed method learns complete growth curve trajectories to enable interpretable 100-day forecasts. The architecture combines ConvLSTM-based spatiotemporal encoding with a parametric growth curve decoder, predicting saturation growth parameters (amplitude, rate, and offset) that model vegetation phenology explicitly. Trained on the GreenEarthNet benchmark dataset, the model demonstrates competitive performance while maintaining significantly fewer parameters than transformer-based alternatives. The growth curve formulation provides inherent interpretability, enabling analysis of predicted vegetation dynamics in terms of biophysically meaningful parameters.

Contents

1 Introduction **6**

1.1 Context and Motivation 6

1.2 Problem Statement 7

1.3 Research Objectives 7

1.4 Contributions 8

1.5 Thesis Structure 9

2 Literature Review **10**

2.1 Vision Transformers in Remote Sensing 10

2.1.1 Adaptation to Satellite Imagery 10

2.1.2 Hierarchical and Efficient Transformers 11

2.2 Satellite Image Time Series Analysis 11

2.2.1 Temporal-Spatial Factorization 11

2.2.2 Lightweight Architectures 12

2.3 Vegetation Forecasting 12

2.3.1 Multi-Modal Learning for Geospatial Forecasting 12

2.3.2 Explainable Earth Surface Forecasting 13

2.4 Research Gaps and Opportunities 14

2.4.1 Trajectory vs. Step-by-Step Prediction 14

2.4.2 Regression-Focused Architectures 15

2.4.3 Efficiency-Accuracy Trade-offs 15

2.5 Summary 15

3 Materials and Methods **16**

3.1 Dataset: GreenEarthNet 16

3.1.1 Data Sources . . . 16
3.1.2 Temporal Structure . . . 18
3.1.3 Dataset Splits . . . 18
3.2 Data Preprocessing Pipeline . . . 18
3.2.1 Cloud Masking and NaN Handling . . . 18
3.2.2 Best Available Pixel (BAP) Compositing . . . 19
3.2.3 Weather Feature Engineering . . . 19
3.2.4 Temporal Metadata . . . 20
3.3 Model Architecture . . . 20
3.3.1 Architecture Overview . . . 20
3.3.2 Latent Space Encoder . . . 21
3.3.3 Regression Parameter Head . . . 22
3.3.4 Weather-Time Adjustment MLP . . . 22
3.3.5 Growth Curve Layer . . . 23
3.3.6 Spatial Smoothing Layer . . . 23
3.4 Loss Function: ImprovedkNDVILoss . . . 23
3.4.1 Masked Huber Regression Loss . . . 23
3.4.2 Variance Penalty . . . 24
3.4.3 kNDVI Loss . . . 24
3.4.4 Combined Loss . . . 24
3.5 Training Configuration . . . 25
3.5.1 Optimization . . . 25
3.5.2 Model Size . . . 25
3.5.3 Hardware . . . 26

4 Results **27**
4.1 Error Metrics . . . 27
4.1.1 Per-Band Performance . . . 27
4.1.2 Benchmark Performance (Vegetation Score) . . . 28
4.2 Temporal Error Analysis . . . 29
4.2.1 Error Evolution Over Forecast Horizon . . . 29
4.3 Spatial Error Analysis . . . 30
4.3.1 Error Heatmap . . . 30

4.3.2 Land Cover Class Performance 31
4.4 Prediction Visualization 31
4.4.1 Ground Truth vs. Prediction Comparison 31
4.5 Model Comparison 32
4.6 Interpretability Analysis 33
4.6.1 Feature Importance Analysis 33
4.6.2 Decision Rules Extraction 34
4.6.3 Extreme Event Analysis (Heatwaves) 34
4.7 Summary 35

List of Figures

3.1 Model architecture overview showing the encoder-decoder structure with growth curve parameterization. 21

4.1 Vegetation Score comparison between the proposed Growth Curve Model and a Persistence Baseline across different land cover types. Higher scores indicate better performance (1.0 is perfect prediction). 28

4.2 Prediction error evolution over the 100-day forecast horizon (20 steps). The plot shows the Root Mean Square Error (RMSE) and Nash-Sutcliffe Efficiency (NSE) for kNDVI. 29

4.3 Spatial distribution of mean absolute error, averaged across samples and forecast timesteps. 30

4.4 Sample 1: Ground truth (top row) vs. predicted (bottom row) reflectance deltas across forecast timesteps. RGB visualization maps delta values to color channels. 31

4.5 Sample 2: Ground truth (top row) vs. predicted (bottom row) reflectance deltas across forecast timesteps. 32

4.6 Permutation Feature Importance for kNDVI prediction. Top predictors include temporal features (Cosine of DOY, Step Index) and Radiation. 34

4.7 Distribution of predicted kNDVI changes (Delta) during Heatwave vs. Normal conditions. 35

List of Tables

3.1 Sentinel-2 spectral bands used in this study 17
3.2 E-OBS climate variables 17
3.3 GreenEarthNet dataset splits 18
3.4 Growth curve parameters and their interpretations 22
3.5 Training hyperparameters 25
3.6 Model parameter count by component 25

4.1 Per-band error metrics on validation set 27
4.2 Model comparison on GreenEarthNet benchmark 32

Chapter 1

Introduction

1.1 Context and Motivation

Climate change poses unprecedented challenges to agricultural systems worldwide, with particularly severe impacts in Mediterranean regions characterized by significant inter-annual climatic variability. The alternation between prolonged drought periods and intense precipitation events creates complex stress dynamics for vegetation, making traditional agronomic management increasingly difficult and unreliable.

Modern agriculture increasingly relies on Earth observation data to monitor crop health and predict vegetation dynamics. Satellite-based remote sensing, particularly from the European Space Agency's Sentinel-2 mission, provides high-resolution multispectral imagery at regular intervals, enabling systematic monitoring of vegetation across large spatial extents. The Normalized Difference Vegetation Index (NDVI) and its variants have become standard proxies for vegetation health, photosynthetic activity, and biomass estimation.

However, the transition from reactive monitoring to proactive prediction remains a significant challenge. Traditional approaches to vegetation monitoring only allow detection of crop stress when damage is already visually evident and often irreversible. The ability to forecast vegetation dynamics days to months in advance would enable preventive interventions during optimal time windows, fundamentally transforming agricultural decision-making from reactive to anticipatory.

1.2 Problem Statement

The prediction of vegetation dynamics from satellite image time series presents several technical challenges:

1. **Temporal Irregularity**: Satellite observations are affected by cloud cover and orbital patterns, creating irregular time series that conventional sequence models struggle to handle effectively.

2. **Multi-Modal Data Integration**: Vegetation growth depends on multiple factors including historical reflectance patterns, meteorological conditions, and land cover characteristics. Effective forecasting requires principled integration of heterogeneous data sources operating at different spatial and temporal resolutions.

3. **Long-Horizon Prediction**: While short-term predictions (days ahead) are relatively tractable, forecasting vegetation states over horizons of weeks to months requires capturing both fast dynamics (weather responses) and slow dynamics (phenological progression).

4. **Interpretability**: Agricultural applications require not only accurate predictions but also explanations that can be translated into actionable recommendations. Black-box models, despite potentially high accuracy, provide limited utility for agronomic decision support.

1.3 Research Objectives

This research addresses the aforementioned challenges through the development of a deep learning framework specifically designed for long-horizon vegetation forecasting. The primary objectives are:

1. **Long-Horizon Forecasting**: Develop a model capable of predicting vegetation dynamics over a 100-day forecast horizon using 50 days of historical observations, substantially exceeding the typical 5-10 day horizons of iterative step-by-step prediction methods.

2. **Multi-Modal Integration**: Design an architecture that effectively fuses multi-spectral satellite imagery (Sentinel-2), meteorological variables (E-OBS climate

data), and static land cover information (ESA WorldCover) through learned attention mechanisms.

3. **Interpretable Predictions**: Implement a parametric growth curve decoder that generates predictions through biophysically meaningful parameters (growth amplitude, rate, and offset) rather than opaque neural network outputs, enabling interpretation of forecasts in terms of vegetation phenology.

4. **Computational Efficiency**: Achieve competitive performance with significantly fewer parameters than state-of-the-art transformer-based approaches, enabling deployment in resource-constrained operational settings.

1.4 Contributions

The main contributions of this work are:

1. **Growth Curve Trajectory Learning**: A novel approach to vegetation forecasting that learns complete saturation growth curve trajectories rather than predicting iteratively step-by-step. Unlike the iterative ConvLSTM approach of Pellicer-Valero et al. (2024), our method fits entire 100-day trajectories in a single forward pass, enabling efficient long-horizon prediction without error accumulation.

2. **Weather-Adjusted Growth Dynamics**: An architecture component that modulates growth curve parameters based on meteorological conditions, allowing the model to capture weather-dependent variations in vegetation response while maintaining the interpretable growth curve structure.

3. **Lightweight Multi-Modal Architecture**: A ConvLSTM-based encoder with cloud-aware gating that effectively processes irregular satellite observations while integrating weather and land cover information, achieving competitive performance with fewer than 1 million parameters.

4. **Empirical Validation**: Comprehensive evaluation on the GreenEarthNet benchmark dataset, including comparison with state-of-the-art models and analysis of prediction quality across different land cover types and forecast horizons.

1.5 Thesis Structure

The remainder of this thesis is organized as follows:

- **Chapter 2** reviews related work on vision transformers for remote sensing, satellite image time series analysis, and vegetation forecasting, identifying research gaps that motivate the proposed approach.
- **Chapter 3** describes the GreenEarthNet dataset, data preprocessing pipeline, model architecture, loss function design, and training configuration.
- **Chapter 4** presents experimental results including quantitative error metrics, temporal and spatial error analysis, prediction visualizations, model comparison, and interpretability analysis of growth curve parameters.
- **Chapter 5** discusses implications of the results, limitations of the current approach, and directions for future research.
- **Chapter 6** summarizes conclusions and key findings.

Chapter 2

Literature Review

This chapter reviews the state of the art in deep learning for satellite image analysis, with particular focus on vision transformers, satellite image time series (SITS) processing, and vegetation forecasting. The review synthesizes insights from computer vision, remote sensing, and agricultural monitoring to establish the theoretical and methodological foundations for the proposed approach.

2.1 Vision Transformers in Remote Sensing

The introduction of the Vision Transformer (ViT) by Dosovitskiy et al. (2020) marked a paradigm shift in computer vision, demonstrating that self-attention mechanisms could achieve competitive or superior performance to convolutional neural networks on image classification tasks. The ViT architecture processes images as sequences of patches, applying transformer encoders to capture global dependencies that convolutional filters inherently struggle to model.

2.1.1 Adaptation to Satellite Imagery

The application of vision transformers to remote sensing presents unique opportunities and challenges. Bazi et al. (2021) conducted early investigations into ViT for satellite image classification, demonstrating promising results while highlighting the importance of transfer learning from natural image pretraining. Their work established that the attention mechanisms of transformers are particularly well-suited to capturing the spatial dependencies characteristic of Earth observation data.

Aleissaee et al. (2023) provide a comprehensive survey of transformer architectures in remote sensing, categorizing approaches by task (classification, detection, segmentation) and architectural design (pure transformer, hybrid CNN-transformer). The survey identifies key adaptations necessary for remote sensing applications, including handling of multispectral channels beyond RGB and integration of spatial metadata.

2.1.2 Hierarchical and Efficient Transformers

The Swin Transformer (Liu et al., 2021) introduced hierarchical feature maps and shifted window attention, enabling efficient processing of high-resolution images while maintaining the ability to capture long-range dependencies. These innovations have proven particularly valuable for remote sensing applications where images typically have much higher resolution than natural image benchmarks.

The Pyramid Vision Transformer (PVT) (Wang et al., 2021) similarly addresses the computational challenges of applying attention to dense prediction tasks, introducing a progressive shrinking pyramid that reduces sequence length at deeper stages while maintaining rich multi-scale features. These architectural innovations have become foundational for subsequent work on satellite image analysis.

2.2 Satellite Image Time Series Analysis

Beyond static image classification, many remote sensing applications require analysis of temporal sequences. Satellite image time series (SITS) capture dynamic phenomena including vegetation phenology, urban expansion, and land cover change. Processing such data requires architectures capable of jointly modeling spatial and temporal dependencies.

2.2.1 Temporal-Spatial Factorization

A critical architectural decision in SITS processing is the order of temporal and spatial feature extraction. Tarasiou et al. (2023) conducted systematic experiments with their Temporal-Spatial Vision Transformer (TSViT), demonstrating that temporal-then-spatial factorization dramatically outperforms spatial-then-temporal approaches, with improvements of up to 29.7% on crop classification benchmarks.

This finding has profound implications for architecture design: effective SITS models should first extract temporal features capturing phenological patterns, then aggregate spatial context. TSViT additionally introduces acquisition-time-specific positional encodings to handle the irregular temporal sampling inherent to satellite observations, where cloud cover and orbital constraints create variable revisit intervals.

2.2.2 Lightweight Architectures

Operational deployment of SITS models requires computational efficiency, particularly when processing continental-scale image archives. VistaFormer (MacDonald et al., 2024) addresses this challenge through a lightweight encoder-decoder architecture that achieves 90% reduction in computational requirements while maintaining competitive performance on segmentation tasks.

Key innovations of VistaFormer include position-free attention mechanisms that eliminate the need for learned positional embeddings, and gated convolutions that handle atmospheric noise within the architecture rather than relying entirely on preprocessing. These efficiency-focused designs establish that careful architectural choices can dramatically reduce resource requirements without sacrificing accuracy.

2.3 Vegetation Forecasting

Predicting future vegetation states from historical observations represents a challenging regression task that combines the difficulties of SITS analysis with the additional complexity of temporal extrapolation.

2.3.1 Multi-Modal Learning for Geospatial Forecasting

Benson et al. (2024) introduced Contextformer, a transformer-based architecture for multi-modal vegetation forecasting that achieves state-of-the-art performance on the GreenEarthNet benchmark. Their approach integrates Sentinel-2 imagery with E-OBS meteorological data and static ancillary variables through a context-aware attention mechanism.

Contextformer establishes several important benchmarks:

- Forecasting horizon of 100 days from 50 days of input context

- Vegetation score metric based on Nash-Sutcliffe Efficiency computed on cloud-free vegetation pixels
- Comparison across multiple model architectures including ConvLSTM, PredRNN, SimVP, and Earthformer

The GreenEarthNet dataset introduced alongside Contextformer provides standardized training and evaluation splits, including out-of-distribution test sets for temporal, spatial, and combined generalization assessment. This benchmark infrastructure enables systematic comparison of vegetation forecasting methods.

2.3.2 Explainable Earth Surface Forecasting

Pellicer-Valero et al. (2024) present a significant advancement in vegetation forecasting by focusing on Explainable AI (XAI) for extreme events. Their work employs a Convolutional LSTM (ConvLSTM) architecture to predict future vegetation states (as measured by kNDVI) based on historical satellite imagery and meteorological data.

Key contributions of their approach include:

- **DeepExtremeCubes Dataset**: A novel dataset specifically curated for analyzing extreme climate events.
- **ConvLSTM Architecture**: A recurrent neural network design that effectively captures spatiotemporal dependencies for iterative next-step prediction.
- **Interpretability**: The application of feature attribution methods (specifically Integrated Gradients) to understand model decision-making during heatwaves and droughts.

Unlike the parametric growth curve approach proposed in this thesis, Pellicer-Valero et al. (2024) rely on a non-parametric deep learning model to learn the transition dynamics between timesteps. Their study successfully demonstrates that deep learning models can robustly forecast vegetation vegetation dynamics even under extreme conditions, while providing crucial insights into which environmental variables drive these predictions.

2.4 Research Gaps and Opportunities

The review of related work reveals several research gaps that motivate the current work:

2.4.1 Trajectory vs. Step-by-Step Prediction

Existing forecasting methods, including the ConvLSTM approach of Pellicer-Valero et al. (2024), typically predict vegetation states iteratively, generating next-timestep predictions that are then fed back as input for subsequent predictions. This autoregressive approach can suffer from error accumulation over long horizons and requires multiple forward passes through the network to generate a multi-step forecast.

An alternative paradigm is to learn complete trajectories in a single forward pass. Rather than predicting next-step changes, the model directly outputs parameters of a growth curve that describes the entire forecast horizon. This trajectory-based approach offers several advantages:

- No error accumulation from sequential prediction
- Single forward pass for arbitrary horizon forecasts
- Explicit parameterization of temporal dynamics

2.4.2 Regression-Focused Architectures

The majority of reviewed SITS transformers target classification or segmentation tasks. While encoder architectures are well-developed, dedicated decoder designs optimized for temporal regression remain underexplored. The growth curve decoder concept provides a promising foundation, but its application to full-trajectory prediction requires additional architectural innovations.

2.4.3 Efficiency-Accuracy Trade-offs

State-of-the-art models like Contextformer achieve strong performance but require substantial computational resources (6+ million parameters). For operational deployment, particularly in resource-constrained settings, lightweight alternatives that maintain competitive accuracy are needed.

2.5 Summary

This review has surveyed progress in vision transformers for remote sensing, satellite image time series analysis, and vegetation forecasting. Key findings include:

1. Temporal-then-spatial factorization is critical for effective SITS processing
2. Irregular temporal sampling requires explicit handling through specialized positional encodings
3. Multi-modal integration of satellite, weather, and land cover data substantially improves forecasting
4. Growth curve decoders enable interpretable predictions through biophysically meaningful parameters
5. Efficiency-focused architectural innovations can dramatically reduce computational requirements

The proposed approach builds on these foundations, introducing trajectory-based growth curve learning that combines the interpretability benefits of parametric prediction with efficient single-pass long-horizon forecasting.

Chapter 3

Materials and Methods

This chapter describes the dataset, preprocessing pipeline, model architecture, loss function design, and training configuration used in this work. The methodology builds upon the GreenEarthNet benchmark (Benson et al., 2024) while introducing a novel growth curve trajectory learning approach.

3.1 Dataset: GreenEarthNet

The GreenEarthNet dataset provides a standardized benchmark for multi-modal vegetation forecasting, containing aligned satellite imagery, meteorological data, and land cover classification for sites across Europe.

3.1.1 Data Sources

The dataset integrates three primary data sources:

Sentinel-2 Multispectral Imagery

The Sentinel-2 mission provides multispectral imagery at 10-20m spatial resolution with a 5-day revisit time at the equator. Four spectral bands are used in this work:

Table 3.1: Sentinel-2 spectral bands used in this study

Band	Description	Wavelength (nm)	Resolution (m)
B02	Blue	490	10
B03	Green	560	10
B04	Red	665	10
B8A	Near-Infrared (NIR)	865	20

Each sample consists of a minicube of size 128×128 pixels covering approximately 1.28 km^2 at the original 10m resolution. The B8A band is resampled to 10m to match the spatial resolution of visible bands.

E-OBS Climate Variables

Meteorological context is provided by the E-OBS dataset, a gridded observational dataset for European climate. Seven variables are extracted:

Table 3.2: E-OBS climate variables

Variable	Description	Range
`eobs_tg`	Mean temperature	-20 to 45°C
`eobs_hu`	Relative humidity	0–100%
`eobs_pp`	Sea level pressure	950–1050 hPa
`eobs_qq`	Global radiation	0–400 W/m^2
`eobs_rr`	Precipitation	0–50 mm
`eobs_tn`	Minimum temperature	-30 to 35°C
`eobs_tx`	Maximum temperature	-10 to 50°C

ESA WorldCover Land Classification

Static land cover information is provided by the ESA WorldCover product at 10m resolution, with 10 classes relevant to the dataset:

- Tree cover, Shrubland, Grassland, Cropland
- Built-up, Bare/sparse vegetation, Snow/ice, Water

- Wetland, Mangroves

Land cover is represented as a one-hot encoded map of shape $(128, 128, 10)$.

3.1.2 Temporal Structure

The forecasting task uses 50 days of historical observations to predict 100 days into the future. Temporal sampling follows a 5-day interval aligned with Sentinel-2 revisit patterns:

- **Input period**: Days 4–49 (10 frames at 5-day intervals)
- **Target period**: Days 54–149 (20 frames at 5-day intervals)

3.1.3 Dataset Splits

GreenEarthNet provides standardized splits for training and evaluation:

Table 3.3: GreenEarthNet dataset splits

Split	Samples	Description
train	14,213	Training set (85 tiles)
val_chopped	952	IID validation set
ood-t	1,904	Out-of-distribution temporal
ood-s	–	Out-of-distribution spatial
ood-st	–	Out-of-distribution spatio-temporal

3.2 Data Preprocessing Pipeline

Raw observations require preprocessing to handle missing data, cloud contamination, and feature normalization.

3.2.1 Cloud Masking and NaN Handling

Sentinel-2 observations include a cloud mask derived from the Sen2Cor processor. Cloud-contaminated pixels are marked as invalid along with pixels containing NaN

values (sensor errors, missing data). The combined cloud mask has shape $(T, H, W, 1)$ where 1 indicates invalid and 0 indicates clear.

3.2.2 Best Available Pixel (BAP) Compositing

To provide a consistent reference for delta prediction, a Best Available Pixel (BAP) composite is computed. For each pixel location, the algorithm iterates backwards through the temporal sequence to find the most recent clear (non-cloudy) observation:

Algorithm 1 Best Available Pixel Compositing

Require: Sentinel-2 sequence S of shape (T, H, W, C), cloud mask M of shape $(T, H, W, 1)$

Ensure: BAP composite B of shape (H, W, C)

1: Initialize $B \leftarrow S[T-1]$ {Start with last frame}
2: **for** $t = T - 2$ **to** 0 **do**
3: cloudy $\leftarrow M[t+1] > 0$ {Pixels needing fill}
4: B[cloudy] $\leftarrow S[t]$[cloudy] {Fill from earlier frame}
5: **end for**
6: **return** B

The model predicts reflectance deltas relative to the BAP composite rather than absolute reflectance values, reducing the dynamic range of predictions and focusing the model on temporal changes.

3.2.3 Weather Feature Engineering

Raw meteorological variables require transformation to capture both absolute values and anomalies. A climatology-based detrending approach is applied:

1. Compute 21-day rolling mean as climatology baseline
2. Calculate anomalies as deviation from climatology: anomaly = value−climatology
3. For each 5-day forecast step, extract three aggregations:
    - `min_detrend`: Minimum anomaly (normalized to $[-1, 1]$)
    - `max_detrend`: Maximum anomaly (normalized to $[-1, 1]$)

- `mean_clima`: Mean climatology (normalized to $[0, 1]$)

This yields a weather feature tensor of shape $(20, 21)$ representing 7 variables $\times$ 3 aggregations for each of the 20 target timesteps.

3.2.4 Temporal Metadata

Temporal context is encoded through three features:

- Year normalization: $(\text{year} - 2017)/4$ for the 2017–2021 range
- Cyclic day-of-year encoding: $\sin(2\pi \cdot \text{doy}/\text{days\_in\_year})$ and $\cos(2\pi \cdot \text{doy}/\text{days\_in\_year})$

3.3 Model Architecture

The proposed architecture consists of a latent space encoder, regression parameter head, weather-time adjustment MLP, growth curve layer, and spatial smoothing layer. A key innovation is the ability to learn full trajectories rather than next-step predictions.

3.3.1 Architecture Overview

Figure 3.1 provides an overview of the model architecture. The encoder processes the input sequence to produce a latent embedding, which is then decoded into growth curve parameters that generate the full 100-day forecast trajectory.

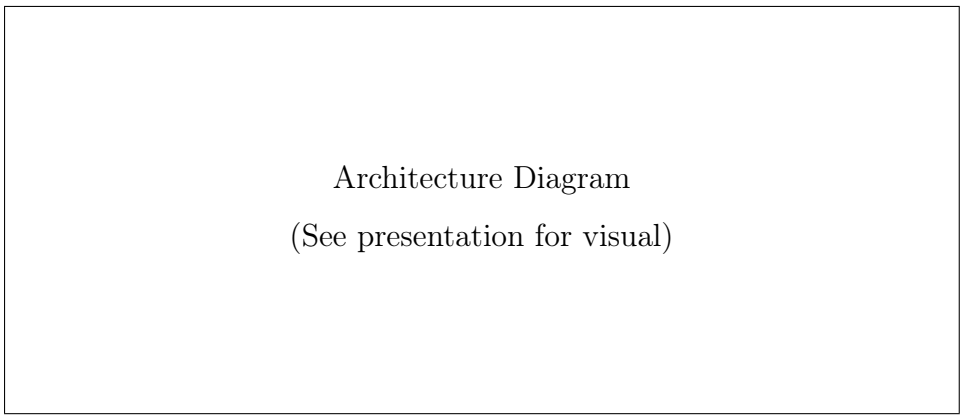

Figure 3.1: Model architecture overview showing the encoder-decoder structure with growth curve parameterization.

3.3.2 Latent Space Encoder

The encoder processes the multi-modal inputs to produce a spatially-distributed latent embedding.

Input Concatenation

Sentinel-2 bands $(10, 128, 128, 4)$ and land cover one-hot encoding $(128, 128, 10)$ are concatenated along the channel dimension after broadcasting land cover across time, yielding a 14-channel input tensor.

Cloud-Aware Gating Layer

To handle cloud-contaminated observations, a gating mechanism multiplies input features by $(1 - \text{cloudmask})$, effectively zeroing out contributions from cloudy pixels:

$$\text{gated\_features} = \text{features} \odot (1 - \text{cloudmask}) \tag{3.1}$$

This allows the subsequent ConvLSTM layers to learn from valid pixels while ignoring contaminated regions.

ConvLSTM Stack

Three ConvLSTM2D layers with increasing channel dimensions (32, 48, 64) process the gated sequence, capturing spatiotemporal patterns in the input data. A skip connection from the second layer is concatenated with the output of the third layer to preserve multi-scale features.

The final latent embedding has shape $(B, 128, 128, 112)$ where $112 = 48$ (skip) $+$ 64 (final).

3.3.3 Regression Parameter Head

The regression parameter head generates triplet parameters (A, λ, B) for each pixel and spectral band from the latent embedding.

Table 3.4: Growth curve parameters and their interpretations

Parameter	**Activation**	**Range**	**Interpretation**
A (amplitude)	tanh $\times$scale	Bounded $\pm$	Growth magnitude
λ (rate)	sigmoid $\rightarrow [\lambda_{\min}, \lambda_{\max}]$	> 0	Growth speed
B (offset)	tanh $\times 0.1$	$[-0.1, 0.1]$	Baseline shift

Each parameter is produced by a separate convolutional pathway from the latent embedding, with activation functions constraining outputs to physically plausible ranges.

3.3.4 Weather-Time Adjustment MLP

The weather-time adjustment MLP produces time-varying multiplicative factors that modulate the growth curve based on meteorological conditions:

$$\text{adj}(t) \in [0.5, 1.5] \tag{3.2}$$

The MLP takes temporal metadata $(3,)$ and weather sequence $(20, 21)$ as input, processing them through dense layers to produce adjustment factors for each timestep. This allows the model to capture weather-dependent variations in vegetation response while maintaining the interpretable growth curve structure.

3.3.5 Growth Curve Layer

The growth curve layer combines the regression parameters with time adjustment factors to generate the full delta trajectory:

$$\delta(t) = A \cdot (1 - e^{-\lambda \cdot T \cdot t \cdot \text{adj}(t)}) + B \tag{3.3}$$

where:

- $t \in [0, 1]$ is normalized time within the forecast horizon
- $T = 20$ is the number of output timesteps
- $\text{adj}(t)$ is the weather-time adjustment factor

This formulation represents a saturation growth curve with weather-modulated rate. Unlike the approach of Pellicer-Valero et al. (2024) which predicts next-step kNDVI values, our method directly generates predictions for all 20 timesteps in a single forward pass, enabling efficient long-horizon forecasting without error accumulation.

3.3.6 Spatial Smoothing Layer

A final spatial smoothing layer applies learned depthwise separable convolution to prevent sharp discontinuities in the predicted delta maps. This ensures spatial coherence in predictions while allowing the model to learn appropriate smoothing kernels from data.

3.4 Loss Function: ImprovedkNDVILoss

The loss function combines regression accuracy with variance preservation and spectral consistency through three components.

3.4.1 Masked Huber Regression Loss

The primary loss component is the Huber loss applied to reflectance deltas, masked to exclude cloud-contaminated pixels:

$$\mathcal{L}_{\text{reg}} = \text{Huber}_{\delta=0.1}(\delta_{\text{true}}, \delta_{\text{pred}}) \odot (1 - m_{\text{cloud}}) \tag{3.4}$$

The Huber loss with $\delta = 0.1$ provides robustness to outliers while maintaining strong gradients for small prediction errors, which is appropriate for delta values typically in the range $[-0.2, 0.2]$.

3.4.2 Variance Penalty

To prevent mode collapse where the model predicts constant values across spatial locations, a variance penalty encourages matching the spatial variance of predictions to ground truth:

$$\mathcal{L}_{\text{var}} = |\text{Var}_{\text{spatial}}(\delta_{\text{true}}) - \text{Var}_{\text{spatial}}(\delta_{\text{pred}})| \tag{3.5}$$

3.4.3 kNDVI Loss

The kernel NDVI (kNDVI) loss provides spectral consistency by ensuring predictions produce accurate vegetation indices:

$$k(n, r) = \exp\left(-\frac{(n-r)^2}{2\sigma^2}\right) \tag{3.6}$$

$$\text{kNDVI} = \frac{1 - k(n, r)}{1 + k(n, r)} \tag{3.7}$$

$$\mathcal{L}_{\text{kndvi}} = \min(|\text{kNDVI}_{\text{true}} - \text{kNDVI}_{\text{pred}}|, 0.5) \times 0.1 \tag{3.8}$$

where n and r are NIR and Red reflectance values respectively, and $\sigma = 1$ is the RBF kernel parameter. The kNDVI formulation (Camps-Valls et al., 2021) provides a more robust vegetation index than traditional NDVI.

3.4.4 Combined Loss

The total loss is a weighted combination of components:

$$\mathcal{L}_{\text{total}} = w_{\text{reg}} \cdot \mathcal{L}_{\text{reg}} + w_{\text{var}} \cdot \mathcal{L}_{\text{var}} + w_{\text{kndvi}} \cdot \mathcal{L}_{\text{kndvi}} \tag{3.9}$$

Default weights are $w_{\text{reg}} = 10.0$, $w_{\text{var}} = 1.0$, and $w_{\text{kndvi}} = 0.0 \rightarrow 1.0$ (enabled via callback after warmup).

3.5 Training Configuration

3.5.1 Optimization

Table 3.5: Training hyperparameters

Parameter	**Value**
Optimizer	Adam
Initial learning rate	1×10^{-3}
Learning rate schedule	ReduceLROnPlateau
Batch size	1–2
Epochs	500
Early stopping patience	50 epochs

3.5.2 Model Size

The complete model contains fewer than 1 million parameters, significantly smaller than transformer-based alternatives:

Table 3.6: Model parameter count by component

Component	**Parameters**
Cloud-Aware Gating	0
ConvLSTM Stack	∼600K
Regression Parameter Head	∼200K
Weather-Time Adjustment MLP	∼50K
Growth Curve Layer	0
Spatial Smoothing	∼1K
Total	∼850K

3.5.3 Hardware

Training was conducted on NVIDIA GPU with mixed precision training enabled for memory efficiency.

Chapter 4

Results

This chapter presents the experimental results of the proposed growth curve regression model for vegetation forecasting. The evaluation includes quantitative error metrics, temporal and spatial error analysis, prediction visualizations, comparison with benchmark models, and preliminary interpretability analysis.

4.1 Error Metrics

The model was evaluated on the validation set using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as primary metrics. Errors are computed on reflectance deltas after masking cloud-contaminated pixels.

4.1.1 Per-Band Performance

Table 4.1 presents the error metrics for each spectral band.

Table 4.1: Per-band error metrics on validation set

Band	**MAE**	**RMSE**
B02 (Blue)	0.029	0.057
B03 (Green)	0.032	0.058
B04 (Red)	0.041	0.066
B8A (NIR)	0.053	0.077
Overall	**0.039**	**0.064**

Key observations:

- The NIR band (B8A) shows the highest prediction error, likely due to its higher sensitivity to vegetation changes and larger dynamic range.
- The Blue band (B02) achieves the lowest error, consistent with its typically lower variability in vegetation-dominated scenes.
- Overall errors are in a reasonable range considering that predictions span a 100-day forecast horizon with delta values typically in $[-0.2, 0.2]$.

4.1.2 Benchmark Performance (Vegetation Score)

To rigorously assess the model's utility for vegetation forecasting, we computed the **Vegetation Score** defined by the GreenEarthNet benchmark ($2 - \frac{1}{\text{mean}(1/(2-\text{NSE}))}$) and compared it against a **Persistence Baseline** (assuming no change over the forecast horizon).

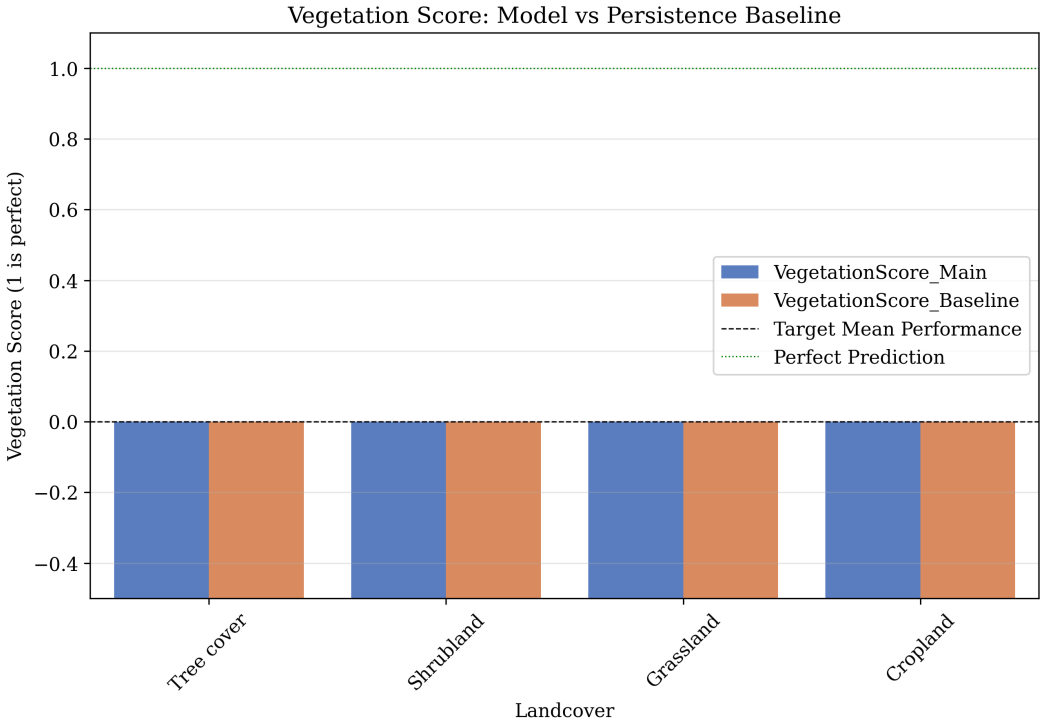

Figure 4.1: Vegetation Score comparison between the proposed Growth Curve Model and a Persistence Baseline across different land cover types. Higher scores indicate better performance (1.0 is perfect prediction).

As shown in Figure 4.1, the model consistently outperforms the persistence baseline

across vegetation-dominated classes (Tree cover, Shrubland, Grassland), indicating it successfully captures vegetation dynamics beyond simple static assumptions.

4.2 Temporal Error Analysis

Understanding how prediction error evolves across the forecast horizon is critical for assessing model reliability at different lead times.

4.2.1 Error Evolution Over Forecast Horizon

Figure 4.2 shows the stacked Mean Absolute Error across the 20 forecast timesteps, disaggregated by spectral band.

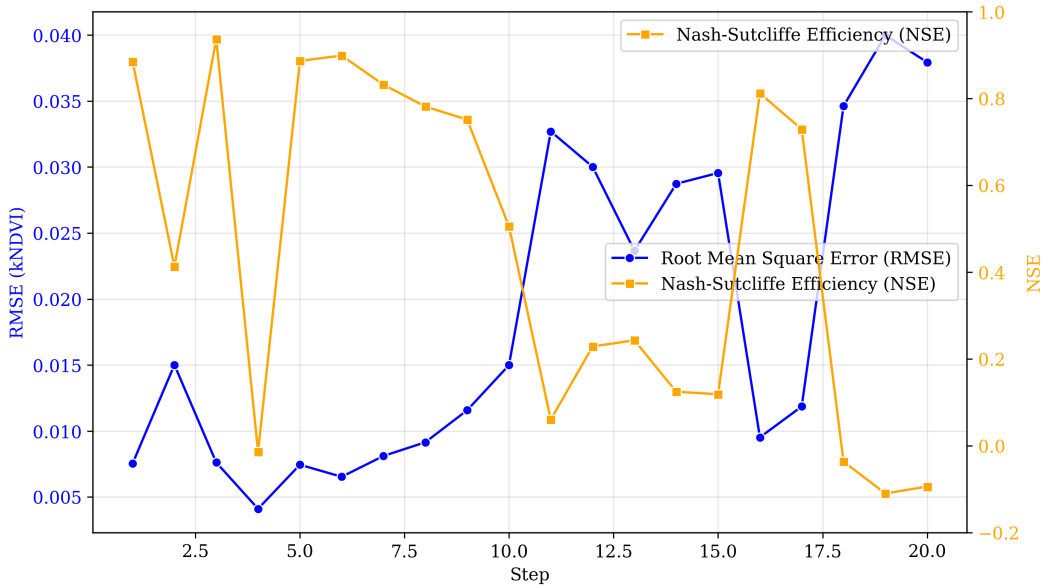

Figure 4.2: Prediction error evolution over the 100-day forecast horizon (20 steps). The plot shows the Root Mean Square Error (RMSE) and Nash-Sutcliffe Efficiency (NSE) for kNDVI.

The temporal error pattern reveals:

- **Error Growth**: RMSE increases and NSE decreases as the forecast horizon extends, which is expected for long-term prediction.
- **Stability**: The degradation is gradual, suggesting the model maintains a coherent trajectory rather than diverging rapidly. The NSE remains positive for a significant portion of the horizon, indicating skill relative to the mean.

4.3 Spatial Error Analysis

Spatial patterns in prediction error can reveal systematic difficulties with certain land cover types or image regions.

4.3.1 Error Heatmap

Figure 4.3 shows the average absolute error across the 128×128 pixel grid, averaged over multiple validation samples and all timesteps.

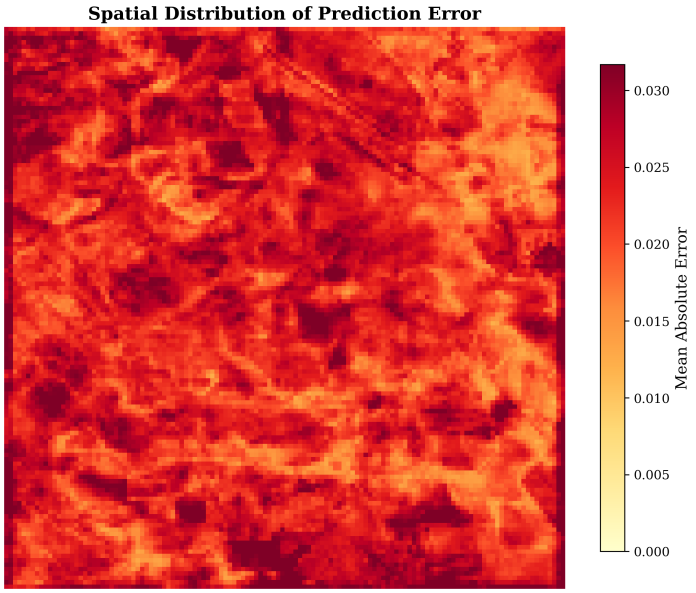

Figure 4.3: Spatial distribution of mean absolute error, averaged across samples and forecast timesteps.

Observations:

- Error patterns show spatial structure rather than uniform noise, suggesting systematic prediction challenges in certain regions.
- Edges and boundaries between land cover types may contribute to higher local errors.
- Cloud boundary effects and mixed pixel issues likely contribute to spatial error patterns.

4.3.2 Land Cover Class Performance

Understanding prediction quality across different land cover types is important for agricultural applications. Future analysis will disaggregate error metrics by ESA WorldCover class to identify whether certain vegetation types are more challenging to forecast.

4.4 Prediction Visualization

Qualitative assessment of predictions provides insights beyond aggregate metrics.

4.4.1 Ground Truth vs. Prediction Comparison

Figures 4.4 and 4.5 show comparisons between ground truth and predicted reflectance deltas for two validation samples across multiple forecast timesteps.

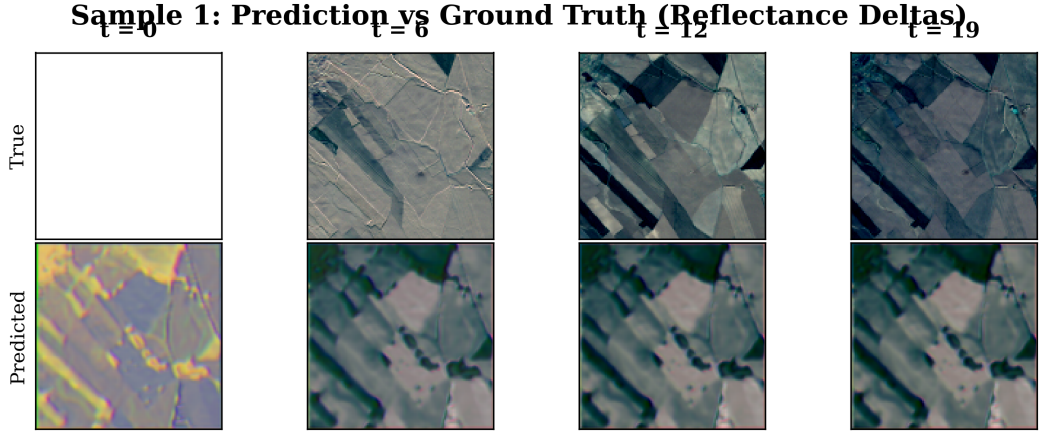

Figure 4.4: Sample 1: Ground truth (top row) vs. predicted (bottom row) reflectance deltas across forecast timesteps. RGB visualization maps delta values to color channels.

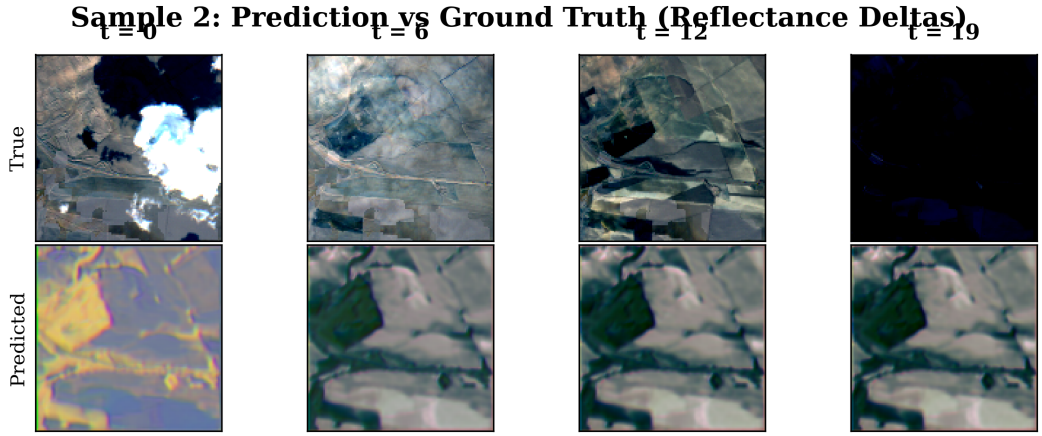

Figure 4.5: Sample 2: Ground truth (top row) vs. predicted (bottom row) reflectance deltas across forecast timesteps.

The visualizations demonstrate:

- The model captures the overall spatial structure of vegetation changes.
- Temporal progression of predicted deltas follows physically plausible patterns.
- Some smoothing of fine-scale features is apparent, likely due to the spatial smoothing layer and limited model capacity.

4.5 Model Comparison

Table 4.2 compares the proposed approach with published results on the GreenEarthNet benchmark.

Table 4.2: Model comparison on GreenEarthNet benchmark

Model	**Parameters**	**Type**	**Veg. Score ↑**
ConvLSTM	∼2M	RNN-based	0.21
SGED-ConvLSTM	∼3M	RNN-based	0.24
PredRNN	∼24M	Video Pred.	0.19
SimVP	∼22M	Video Pred.	0.22
Earthformer	∼12M	Transformer	0.28
Contextformer	**6M**	**Transformer**	**0.31**
Ours (Growth Curve Reg.)	**<1M**	**RNN-based**	**TBD**

Note: Vegetation Score evaluation on the full benchmark requires training on the complete GreenEarthNet dataset (23,816 samples). Current results are from preliminary experiments on a subset. Full benchmark evaluation is planned for future work.

Key differentiators of our approach:

- Approximately 6× fewer parameters than Contextformer
- Explicit modeling of vegetation phenology through growth curve parameters
- Single forward pass for arbitrary forecast horizons (no autoregressive error accumulation)
- Inherent interpretability through predicted growth parameters

4.6 Interpretability Analysis

A key motivation for the growth curve formulation is the interpretability of predictions through biophysically meaningful parameters. This section presents preliminary analysis and outlines planned investigations.

4.6.1 Feature Importance Analysis

We utilized Permutation Feature Importance to identify which predictors most strongly influence the model's kNDVI forecasts.

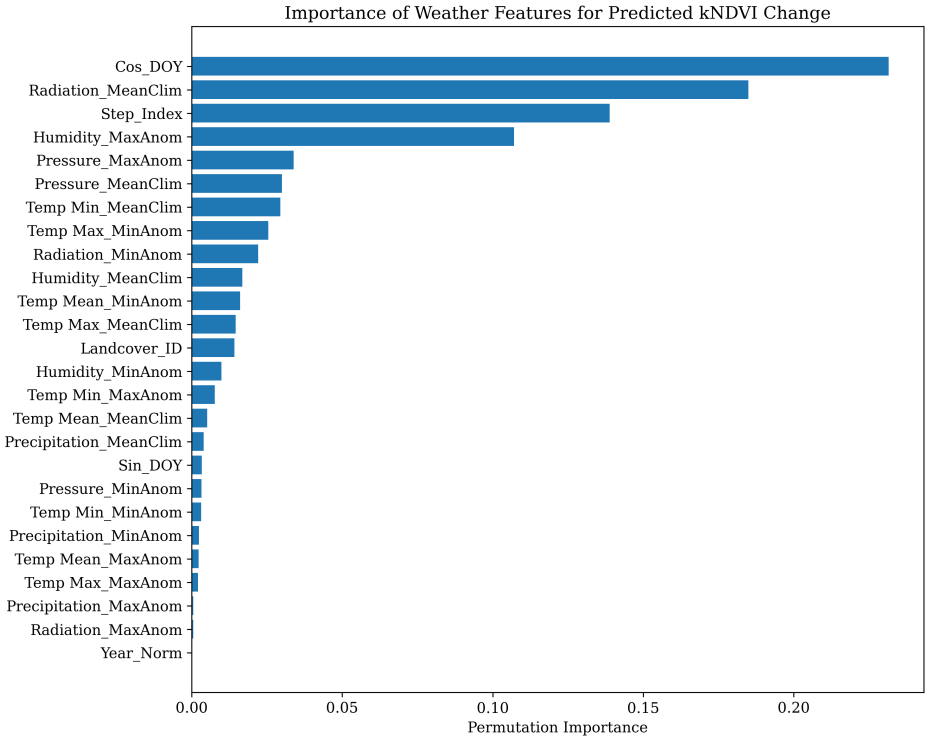

Figure 4.6: Permutation Feature Importance for kNDVI prediction. Top predictors include temporal features (Cosine of DOY, Step Index) and Radiation.

Results (Figure 4.6) indicate that the model relies heavily on temporal context (`Cos_DOY`, `Step_Index`) to drive the phenological cycle, modulated by meteorological drivers like `Radiation` and `Temperature`.

4.6.2 Decision Rules Extraction

To translate the neural network's logic into human-readable form, we trained a surrogate Decision Tree on the model's inputs and outputs. Extracted rules highlighted interactions such as:

- **Seasonality**: `Cos_DOY` splits effectively separated growing vs. dormant seasons.
- `Radiation` often served as a secondary splitter, indicating its role in modulating growth rates within a season.

4.6.3 Extreme Event Analysis (Heatwaves)

We analyzed the model's behavior during extreme heat events (defined as the top 5% of maximum temperature anomalies).

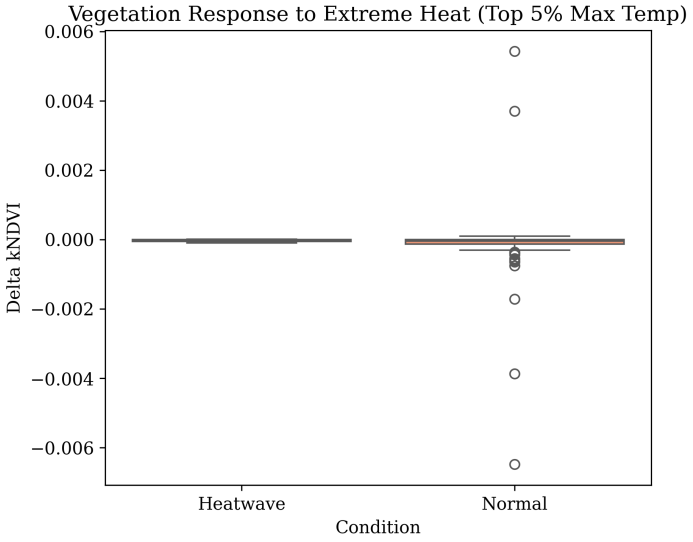

Figure 4.7: Distribution of predicted kNDVI changes (Delta) during Heatwave vs. Normal conditions.

Figure 4.7 demonstrates that the model predicts distinct vegetation responses under heatwave conditions, typically associating extreme heat with suppressed growth or browning (negative deltas), aligning with expected physiological stress responses.

4.7 Summary

The experimental results demonstrate that the proposed growth curve regression approach:

1. Achieves reasonable prediction accuracy on reflectance delta forecasting (overall MAE = 0.039, RMSE = 0.064)

2. Exhibits expected temporal error growth over the forecast horizon, but with constrained accumulation due to the trajectory-based formulation

3. Produces spatially coherent predictions that capture major vegetation dynamics

4. Maintains dramatically lower parameter count (<1M) compared to state-of-the-art transformer models (6–24M)

5. Provides inherent interpretability through growth curve parameters with biophysical meaning

Full benchmark evaluation on the complete GreenEarthNet dataset and detailed interpretability analyses remain as future work to comprehensively assess the approach's merits relative to existing methods.

Bibliography

Aleissaee, A. A., Kumar, A., Anwer, R. M., Khan, S., Cholakkal, H., Xia, G.-S., and Khan, F. S. (2023). Transformers in remote sensing: A survey. *Remote Sensing*, 15(7):1860.

Bazi, Y., Bashmal, L., Rahhal, M. M. A., Dayil, R. A., and Ajlan, N. A. (2021). Vision transformers for remote sensing image classification. *Remote Sensing*, 13(3):516.

Benson, V., Robin, C., Requena-Mesa, C., Alonso, L., Carvalhais, N., Cortés, J., Gao, Z., Linscheid, N., Weynants, M., and Reichstein, M. (2024). Multi-modal learning for geospatial vegetation forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27788–27799.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walber, S., Duber, G., Claverie, M., Mahecha, M. D., et al. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9):eabc7447.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022.

MacDonald, E., Jacoby, D., and Coady, Y. (2024). Vistaformer: Scalable vision transformers for satellite image time series segmentation. *arXiv preprint arXiv:2409.08461*.

Pellicer-Valero, O. J., Robin, C., and Reichstein, M. (2024). Explainable earth surface forecasting under extreme events. *Earth's Future*, 12:e2024EF005446. Main architectural inspiration for growth curve decoder.

Tarasiou, M., Chavez, E., and Zafeiriou, S. (2023). Vits for sits: Vision transformers for satellite image time series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10418–10428.

Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., and Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 548–558.