Modelling Spatial and Non-Linear Trends in Climate Data Using Gaussian Process Regression and Generalized Additive Model

Abstract

Accurate modeling of climate variability is critical for understanding the impacts of climate change and supporting data-driven adaptation strategies. Traditional parametric models, while widely used, often struggle to capture the complex non-linear relationships and spatial dependencies that characterize climate systems, especially in regions with diverse geography such as Kenya. This study aimed to apply two non-parametric statistical approaches—Generalized Additive Models (GAM) and Gaussian Process Regression (GPR)—to model spatial and non-linear trends in climate data over Kenya. Daily climate variables, including temperature and precipitation, were obtained from the ERA5-Land dataset using Google Earth Engine, spanning the period from 2015 to 2024. GAM was used to model the smooth effects of covariates such as time, elevation, and precipitation, while GPR was implemented using a Matérn covariance kernel to capture residual spatial autocorrelation. The models were evaluated using RMSE, MAE, and ^{2}, and parameter estimation was conducted via penalized likelihood and L-BFGS optimization techniques. The results demonstrated that GAM effectively captured structured non-linear effects and provided interpretable smooth functions, while GPR performed better in modeling spatial variability and uncertainty. Both models outperformed traditional linear approaches, with GPR offering superior accuracy in areas with high spatial heterogeneity. The findings affirm that GAM and GPR are powerful and complementary tools for climate modeling in complex environmental contexts. In conclusion, this study confirms the suitability of non-parametric approaches for climate modeling in data-rich, spatially heterogeneous settings. Further research is recommended to explore integrated hybrid GAM–GPR models, extend the methodology to multivariate climate indicators, and evaluate its performance in other regions or under future climate scenarios.

Keywords: Climate Modeling; GAM; GPR

1 Introduction

Climate change is increasingly recognized as one of the most complex and urgent global challenges of our time, with far-reaching implications for ecosystems, agriculture, water resources, infrastructure, and public health (Knüsel and Baumberger (2020)). Its multifaceted nature demands a robust understanding of how climate variables behave over time and space, especially at local and regional scales where impacts are often most acutely felt. In regions like Kenya—characterized by diverse topography, variable land cover, and strong climate seasonality—the need for high-resolution, accurate climate modeling is paramount (Sagero (2019)).

Traditional climate models, including General Circulation Models (GCMs) and Regional Climate Models (RCMs), are often grounded in physical laws and rely on parametric assumptions to simplify complex processes (Rangwala et al. (2021)). While these models have contributed substantially to global climate projections, they frequently struggle to represent localized dynamics and non-linear interactions inherent in climate systems (Giorgi (2019)). Their reliance on predefined mathematical relationships can lead to oversimplified predictions, especially in heterogeneous landscapes. Moreover, their coarse spatial resolution often limits their utility for regional climate assessments and localized decision-making (Bathiany et al. (2016)).

One of the key challenges in modeling climate data lies in adequately capturing both non-linear relationships—such as the effects of elevation, precipitation, and time—and spatial dependencies, where climate variables at nearby locations are correlated due to underlying environmental and atmospheric processes (Vázquez et al. (2017)). Parametric models are often inadequate in this regard, leading to biased or incomplete interpretations of climate trends. This shortcoming becomes critical in climate-sensitive regions like Kenya, where subtle shifts in temperature or rainfall can significantly affect agriculture, water security, and livelihoods.

In recent years, non-parametric statistical methods have emerged as powerful tools for modeling complex environmental systems. These methods, unlike traditional parametric models, do not impose a fixed functional form between variables (Vansteelandt and Dukes (2022)). Instead, they allow the data to inform the structure of relationships, making them particularly suitable for capturing non-linear and context-specific interactions. Among the most prominent of these methods are the Generalized Additive Models (GAM) and Gaussian Process Regression (GPR).

GAM extends linear models by incorporating smooth, non-linear functions of the predictor variables, making it well-suited for modeling complex phenomena like climate data (Singh et al. (2023)). Its ability to model different types of covariates—temporal, topographic, or environmental—using spline functions offers high interpretability and flexibility (Lindström et al. (2014)). On the other hand, GPR is a probabilistic, kernel-based method that excels in modeling spatial processes. It assumes a Gaussian distribution over functions and utilizes covariance functions to express spatial relationships, enabling it to provide both predictions and uncertainty estimates.

Despite their strengths, GAM and GPR have often been used independently in climate studies, and their comparative application to model spatial and non-linear trends in data-rich, topographically diverse regions such as Kenya remains underexplored.

Traditional parametric models are often inadequate for representing the complex spatial and non-linear dependencies found in climate data. Their rigid assumptions limit the models' ability to capture the nuanced interactions between climate variables and geophysical features (Chen et al. (2023)). This limitation is particularly problematic in Kenya, where localized climatic variations play a critical role in shaping environmental outcomes. There is a pressing need for more flexible modeling

frameworks that can address these challenges using available high-resolution data.

The main objective of this study is to apply Generalized Additive Models (GAM) and Gaussian Process Regression (GPR) to model spatial and non-linear trends in climate data across Kenya. The goal is to evaluate how effectively each method captures the underlying structure in temperature and precipitation patterns when using environmental covariates such as elevation, time, and spatial coordinates.

2 Materials and Methods

This study focused on the geographical context of Kenya, located in East Africa and straddling the equator between latitudes 4.5°N and 4.5°S and longitudes 34°E and 42°E. Kenya's diverse topography and climatic zones makes it an ideal setting for modeling spatial and non-linear trends in climate data.

The data used in this study for Kenya were obtained through the Google Earth Engine (GEE) platform, a cloud-based geospatial processing environment that facilitates access to a wide range of satellite and environmental datasets.

A Gaussian process is a collection of random variables where any finite subset follows a multivariate normal distribution. Given input data X and the corresponding target values y, we assumed:

$$y = f_{GPR}(X) + \epsilon \tag{2.1}$$

where $f_{GPR}(X)$ was a latent function modeled as a Gaussian process:

$$f_{GPR}(X) \sim GP(\mu(X), K(X, X^{'})) \tag{2.2}$$

where $\mu(X)$ was the mean of the function at a point X and the covariance between the function values at two distinct positions x and $x^{'}$ was defined by the kernel $K(x, x^{'})$.

To model a GPR with spatial covariates, we modelled the relationship between temperature (f_{GPR}) and the spatial coordinates (latitude(x_1) and longitude (x_2)). The model assumed that:

$$f_{GPR}(x_1, x_2) \sim GP(\mu(x_1, x_2), K((x_1, x_2), (x_1^{'}, x_2^{'}))) \tag{2.3}$$

where $\mu(x_1, x_2)$ was a linear function representing a general trend in temperature across space and $K((x_1, x_2), (x_1^{'}, x_2^{'}))$ was the Matern kernel. Thus, the observed data y followed a multivariate Gaussian distribution:

$$y \sim N(0, K + \sigma_f^2) \tag{2.4}$$

where K was the covariance matrix constructed using the kernel function.

2.1 Estimation of GPR Hyperparameters

The parameters that needed to be estimated in GPR include:

- Kernel hyperparameters θ that included ℓ (controls smoothness of function) and signal variance σ_j^2 (how much the predicted values are allowed to vary up and down as the input changes),
- Noise variance, σ_n^2 which accounted for observational noise, that is, it tells the model how much random error or "noise" is expected in the data.

The marginal likelihood was obtained by integrating over the Gaussian process prior:

$$p(y/X,\theta) = N(0, K + \sigma_n^2) \tag{2.5}$$

The log marginal likelihood (LML) function was given by:

$$\log p(y/X,\theta) = -\frac{1}{2} y^T (K + \sigma_n^2)^{-1} y - \frac{1}{2}\log|K + \sigma_n^2| - \frac{n}{2}\log 2\pi \tag{2.6}$$

where

- The first term measured data fit by computing the quadratic form $y^T(K+\sigma_n^2)^{-1}y$ and ensured that predictions align closely with observed data.
- The second term penalized complexity through the determinant of the covariance matrix $\log|K+\sigma_n^2|$. A larger determinant indicated higher model flexibility, but it also increased the risk of overfitting.
- Third term represented a normalization constant, ensuring a proper probabilistic interpretation of the likelihood function.

To maximize the LML function and obtain the optimal hyperparameters, we took its derivative with respect to each parameter:

$$\frac{\partial}{\partial\theta}\log p(y|X,\theta) = \frac{1}{2} y^T K^{-1}\frac{\partial K}{\partial\theta}K^{-1}y - \frac{1}{2}tr(K^{-1}\frac{\partial K}{\partial\theta}) \tag{2.7}$$

The optimal values of θ, were obtained by solving:

$$\frac{\partial}{\partial\theta}\log p(y|X,\theta) = 0 \tag{2.8}$$

To estimate the optimal hyperparameters θ, the L-BFGS (limited-memory Broyden-Fletcher-Goldfarb-Shanno) algorithm was preferred because:

- It easily handles large data sets with minimal memory requirements.
- It approximates the Hessian matrix, accelerating convergence.
- It has been widely used in climate modeling because of its robustness.

L-BFGS is a quasi-Newton optimization method that approximates the inverse Hessian matrix to update parameters efficiently (Sadeghi-Lotfabadi and Ghiasi-Shirazi (2025)). Instead of storing the full Hessian matrix, L-BFGS maintains a history of past updates to approximate second-order information (Sadeghi-Lotfabadi and Ghiasi-Shirazi (2025)). The L-BFGS Update Formula is given by:

$$H_{k+1}^{-1} = (I - \rho_k s_k y_k^T)H_k^{-1}(I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \tag{2.9}$$

where;

- $s_k = x_{k+1} - x_k$ is the step difference,
- $y_k = \nabla f_{k+1} - \nabla f_k$ is the gradient difference,
- $\rho_k = \frac{1}{y_k^T s_k}$ ensures numerical stability.

The Matern kernel was chosen for this study due to its flexibility in modelling spatial processes with varying degrees of smoothness, which is essential for accurately representing real-world climate variability.Compared to simpler kernels like the exponential kernel (a special case of Matern with $\nu = 0.5$), the Matern kernel with $\nu = 1.5$ was used in this study to provide a better balance between model flexibility and computational stability, allowing for once-differentiable functions, which is a realistic assumption for temperature data. According to Muyskens et al. (2024), the Kernel is defined as;

$$K(X, X^{'}) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\frac{\sqrt{2\nu}||X - X^{'}||}{\ell})^{\nu} K_{\nu}(\frac{\sqrt{2\nu}||X - X^{'}||}{\ell}) \tag{2.10}$$

where:

- ν controls smoothness.
- ℓ is the length scale that controls correlation decay over distance.
- K_ν is the modified Bessel function, which governs how quickly the spatial correlation drops off as distance increases. It ensures that near locations have high correlation while distant locations have exponentially diminishing correlation.

In this study, the limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm was employed to estimate the optimal length scale parameter ℓ of the Matérn kernel used in Gaussian Process Regression (GPR). The length scale governed the spatial correlation structure in the Matérn covariance function and was critical for ensuring appropriate smoothness in the resulting spatial predictions. Accurate estimation of ℓ improved the model's ability to capture both local and global climate trends.

According to Byrd et al. (2016), the estimation procedure involves the following steps:

1. Initialization: An initial guess for the length scale parameter ℓ is provided, typically in log-transformed space to ensure positivity. The maximum memory size m, which dictates how many past gradient and parameter updates are stored, is also defined.
2. Objective Function Evaluation: At each iteration, the negative log-marginal likelihood (NLML) of the GPR model is computed. The Matérn kernel, parameterized by ℓ is used to construct the covariance matrix K, from which the NLML is calculated as:
$$\mathcal{L}(\ell) = -\log p(y|X, \ell) \tag{2.11}$$
3. Gradient Computation: The gradient of the objective function with respect to ℓ is computed analytically or numerically. This requires differentiating the kernel matrix K with respect to ℓ, which includes the Matérn kernel's dependency on both the Gamma function $\Gamma(\nu)$ and the modified Bessel function K_ν.
4. Direction Update: Using the gradient information and the stored memory of past updates, L-BFGS computes an approximate inverse Hessian matrix. The search direction p_k is then determined as:
$$p_k = -H_k \nabla \mathcal{L}_k \tag{2.12}$$
where H_k is the inverse Hessian approximation at iteration k.
5. Line Search: A line search procedure is conducted to determine a suitable step size α_k along the direction p_k that results in a sufficient decrease in the objective function, based on Wolfe or Armijo conditions. According to Jin et al. (2024), this step size must ensure a sufficient decrease in the objective function. The Armijo condition verifies that the reduction in the objective is meaningful relative to the predicted decrease. However, to avoid excessively small steps, the Wolfe conditions add a curvature requirement to ensure the gradient also decreases appropriately. These combined conditions balance efficiency and stability in the optimization process. They guide the algorithm toward the optimal solution without overshooting or stagnation.
6. Parameter Update: The length scale parameter is updated using
$$\ell_{k+1} = \ell_k + \alpha_k p_k \tag{2.13}$$
7. Memory Update: The algorithm stores the most recent parameter differences $s_k = \ell_{k+1} - \ell_k$ and gradient differences $y_k = \nabla \mathcal{L}_{k+1} - \nabla \mathcal{L}_k$, retaining only the last m pairs to maintain efficiency.
8. Convergence Check: The algorithm checks for convergence based on criteria such as the norm of the gradient, the change in the objective function, or a maximum number of iterations. If satisfied, the optimization terminates.

The final output of the L-BFGS algorithm was the optimal value of the length scale parameter ℓ^*, which maximized the marginal likelihood of the observed data under the GPR model. This optimized length scale reflects the spatial structure inherent in the climate data, enabling more accurate and reliable prediction of spatial trends.

The general form of a GAM is:

$$\eta(y) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) \tag{2.14}$$

where the link function that connects the linear predictor to the expected value of the response variable Y is $\eta(.)$, the intercept is β_0, and the smooth functions of the predictors $X_1, X_2, \cdots, X_p$ are $f_1(X_1), f_2(X_2), \cdots, f_p(X_p)$.

The smooth functions $f_j(.)$ can take various forms, including splines, loess, or kernel smoothers. Complex, non-linear interactions between each predictor and the response variable can be captured by the model thanks to these functions.

Complex, non-linear interactions between each predictor and the response variable can be captured by the model thanks to these functions. Spatial covariates, such as latitude and longitude, were incorporated into the model as predictors in order to integrate spatial effects into the GAM. The spatial patterns in the data were then captured using the smooth functions connected to these spatial variables.

The resultant form of the GAM with spatial effects became:

$$\eta(y) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) + f_{spatial}(x_1, x_2) \tag{2.15}$$

where $f_{spatial}(x_1, x_2)$ was a smooth function that models the spatial variation in the response variable based on geographic coordinates.

The average temperature was modeled in relation to geographic location (latitude and longitude), elevation, time and precipitation. The spatial effect GAM was defined as follows:

$$\eta(y) = \beta_o + f_1(time) + f_2(elevation) + f_3(precipitation) + f_{spatial}(x_1, x_2) + \epsilon. \tag{2.16}$$

2.2 Estimation of GAM Parameters

In GAM, the estimation of smooth functions was achieved by minimizing the penalized sum of squares to balance goodness-of-fit with smoothness. This process ensured that the model captures meaningful patterns without overfitting the data. The minimization problem involved optimizing a cost function that included both the residual sum of squares and a penalty term that controled function complexity.

Given the GAM:

$$Y_i = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \cdots + f_p(x_{pi}) + \epsilon_i \tag{2.17}$$

where $f_j(x)$ are smooth functions estimated non-parametrically, the estimation of these functions is done by minimizing the penalized sum of squares:

$$RSS_P = \sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p} f_j(x_{ji}))^2 + \sum_{j=1}^{p} \lambda_j \int (f_j^{''}(x))^2 dx \tag{2.18}$$

where:

- The first term represents the residual sum of squares.
- The second term is a penalty function that discourages excessive curvature in $f_j(x)$.
- λ_j are smoothing parameters controlling the trade-off between fit and smoothness.

The penalized sum of squares was minimized by solving:

$$\min_{\beta_0, f_1, \cdots, f_p} \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{j=1}^{p} f_j(x_{ji}))^2 + \sum_{j=1}^{p} \lambda_j \int (f_j^{''}(x))^2 dx \tag{2.19}$$

To achieve this, we represent each smooth function as a linear combination of basis functions:

$$f_j(x) = \sum_{k=1}^{K} \beta_{jk} \beta_{jk}(x) \tag{2.20}$$

where $\beta_{jk}(x)$ are basis splines and β_{jk} are their coefficients.

Rewriting the penalized sum of squares in matrix form, we obtained

$$RSS_P = ||Y - X\beta||^2 + \beta^T S \beta \tag{2.21}$$

where:

- X is the design matrix of basis functions
- β is the vector of coefficients
- S is the penalty matrix, which penalizes large second derivatives

Taking the derivative and setting it to zero gave the penalized normal equations given by;

$$(X^T X + \lambda S)\beta = X^T Y \tag{2.22}$$

From which

$$\beta = (X^T X + \lambda S)^{-1} X^T Y \tag{2.23}$$

The selection of λ was critical and was typically done using generalized cross validation (GVC), which provides an objective, data-driven means of balancing model complexity against goodness of fit. In this study, tensor product splines were exclusively applied within the GAM framework to model the complex, non-linear relationships present in the climate data. This choice was necessitated by the multi-dimensional nature of the predictors (particularly spatial (longitude and latitude) and temporal variables), which operate on different scales and often exhibit non-isotropic behavior.

3 Results and Discussion

The temperature data was collected from NOAA for a period of 10 years (2015-2025). In order to achieve parsimonious results, summary tables were created, graphs were drawn, and the results were extensively discussed. The analysis was performed using R and Python software.

3.1 GPR with Spatial Covariates

This section presents the results of applying a Gaussian Process Regression (GPR) model to spatially referenced temperature data collected across Kenya. The goal of this model was to capture underlying spatial dependencies and nonlinear interactions between temperature and selected environmental covariates, while providing a principled framework for uncertainty quantification. GPR, as a fully Bayesian and non-parametric approach, enables both the prediction and probabilistic understanding of temperature dynamics in areas with variable data density and heterogeneous climate conditions.

3.1.1 Model Formulation

The primary objective for this component of the study was to develop a GPR model that can flexibly account for spatial variability in temperature, particularly where residual structure remains after accounting for structured environmental effects such as elevation and precipitation. Unlike traditional regression models, which are parametric and assume a fixed mathematical form (such as a straight line or specific curve) to describe relationships between variables, GPR is a non-parametric approach. This means it does not begin with a predefined equation; instead, it places a Gaussian process prior over functions and allows the data itself to shape the form of the relationship. This flexibility makes GPR especially useful in capturing complex and irregular temperature patterns across space, where traditional models may fall short.

This objective was achieved by selecting a model structure in which the inputs consisted of latitude, longitude, elevation, and precipitation, and the temperature observations were treated as noisy realizations of a latent temperature surface. This formulation allowed the model to learn both the spatial continuity of the temperature field and the uncertainty associated with predictions, which is particularly important in regions with sparse measurements or strong spatial gradients.

Moreover, the GPR model provided a natural extension to the additive structure developed using the Generalized Additive Model (GAM). While the GAM captured deterministic trends (that is, systematic and predictable effects of variables like elevation, precipitation, and time on temperature), the GPR framework complemented it by modeling the stochastic spatial behavior, which refers to the random, unpredictable variation that remains even after accounting for known factors. In simpler terms, deterministic components explain what we can predict based on known relationships, while stochastic components capture the unexplained or random spatial fluctuations in the data. This combination allowed the study to satisfy its research aim of integrating both predictable (deterministic) and random (stochastic) aspects of spatial climate variation within a unified modeling framework.

3.1.2 Kernel Specification and Parameter Learning

The fitted kernel was a Matérn kernel with smoothness parameter $\nu = 1.5$, which corresponds to a once-differentiable function and yields moderate smoothness (that is, a realistic assumption for temperature fields that exhibit small-scale variability without abrupt discontinuities).

After hyperparameter optimization, the final form of the kernel included:

- A signal variance of $\sigma_j^2 = 0.837^2 = 0.700$.
- Independent length scales for each input dimension $\ell = [1.42, 1.64, 1.76]$
- An additive Gaussian noise term σ_n^2, incorporated into the full covariance as $K + \sigma_n^2 I$

The varying length scales suggest anisotropy in the spatial process (meaning that temperature does not vary evenly in all directions. Instead, it changes more rapidly along some variables than others). For instance, the smaller length scale along latitude (1.42 degrees) compared to elevation (1.76 meters) indicates stronger spatial variability in the latitudinal direction. This directional dependence suggests that temperature patterns respond more sensitively to positional changes north or south—potentially due to climatic gradients such as the transition from the humid equatorial belt to the arid northern regions—than they do to changes in elevation. In simple terms, anisotropy refers to the idea that temperature patterns vary differently depending on direction or environmental dimension, and the model captures this through distinct length scales for each input variable.

Parameter estimation was achieved through log marginal likelihood (LML) maximization, where the optimal kernel parameters were selected to maximize the probability of observing the training data under the assumed GP prior. The use of the L-BFGS optimization algorithm ensured convergence

to a locally optimal set of hyperparameters while preserving computational efficiency, even for high-dimensional inputs.

3.1.3 Prediction Accuracy and Model Evaluation

To evaluate the generalization capability of the GPR model (that is, its ability to make accurate predictions on new, unseen data rather than just fitting the training data), it was tested on a holdout dataset that was deliberately excluded during model training. This approach ensures that the model's performance reflects its real-world predictive power rather than its ability to memorize known patterns. By applying the trained GPR model to this separate validation dataset, we could assess how well it captures underlying spatial and environmental structures in temperature data across different locations. Performance metrics were as follows:

METRIC	**VALUE**
Explained Variance (R^2)	0.872
Root Mean Square Error (RMSE)	1.6134
Mean Absolute Error	1.2957
Mean Predictive Standard Deviation (σ)	0.3475

Table 1: GPR Model Predictive Accuracy and Evaluation

The R^2 score of 0.8742 indicates that the model was able to explain nearly 87.4% of the variability in observed temperature values on the test set. This high level of explained variance, especially in geostatistical contexts, reflects the model's ability to capture both the deterministic and stochastic structure of spatial climate variation.

The RMSE value of $1.6134°C$ and the MAE of $1.2957°C$ further emphasize the model's predictive strength. The relatively low gap between RMSE and MAE implies that the prediction errors are not skewed by large outliers and that the GPR predictions maintain a high level of consistency across different regions. In particular, the model demonstrated strong predictive reliability in both well-sampled and under-sampled regions, adapting its confidence in accordance with spatial data availability and local smoothness properties.

One of the most significant advantages of the GPR framework is its ability to provide distributional predictions – i.e., mean and variance – instead of just point estimates. This facilitates risk-aware modeling, which is particularly critical in climate-related applications such as disaster preparedness, resource allocation, and agricultural planning.These values indicate that the model was generally highly confident in its predictions (with a mean uncertainty of less than 0.35℃). Predictive standard deviation was lowest in areas with dense observational coverage or strong spatial autocorrelation, and highest in topographically complex or data-sparse regions, such as northern Kenya or regions near the Lake Victoria basin.

Importantly, the spread between the minimum and maximum standard deviation demonstrates the model's ability to scale uncertainty appropriately. This behavior is expected from a well-calibrated Bayesian model and ensures that users are alerted to potential regions of uncertainty or increased climatic variability.

The GPR model produced a smooth and realistic temperature surface that was consistent with both known climatic gradients and topographic influences. For instance:

- Cooler temperatures were consistently predicted in the highland regions, such as Mt. Kenya and the Rift Valley escarpment,
- Warmer temperatures were observed in the eastern lowlands and along the Indian Ocean coast,
- Transitional zones between arid and humid regions were clearly delineated with gradual changes in prediction values and variance.

The spatial length scales learned by the model further imply that temperature is highly structured over space, but that the strength and scale of correlation vary by direction and context. For example, locations within 1.5° of one another (roughly 160 km) shared considerable mutual influence in their predicted temperatures, while farther points became increasingly independent.

This has important implications for spatial sampling strategies and climate station placement, suggesting that inter-station distances should be kept within 1.5° to retain high spatial fidelity in temperature monitoring.

3.2 GAM with Spatial Effects

3.2.1 Model Formulation

To model the nonlinear relationship between average temperature and a set of environmental covariates, a Generalized Additive Model (GAM) was employed. GAM extends the classical linear model by incorporating smooth functions for each predictor, allowing flexibility in capturing nonlinear patterns without requiring explicit specification of their functional form. The structure of the model was:

$$\eta(y) = \beta_o + f_1(time) + f_2(elevation) + f_3(precipitation) + f_{spatial}(x_1, x_2) + \epsilon. \tag{3.1}$$

where:

- $\eta(y)$ is the expected average temperature,
- β_0 is the intercept,
- f_1, f_2, f_3 are univariate smooth functions for time, elevation and precipitation
- $f_{spatial}$is a tensor-product spline function for spatial coordinates (longitude, latitude),
- $\epsilon \sim N(0, \sigma^2)$ represents normally distributed error.

The model captures both temporal trends, topographical influences, climatic variables, and geographical spatial effects in a flexible, data-driven manner.

3.2.2 Data Preparation and Model Fitting Procedure

Data used in the model consisted of 730,400 observations, with variables including date, elevation, precipitation, longitude, latitude, and the average temperature of the response variable. The date variable was transformed into a numerical format (days since the first record) to allow for spline

smoothing.

To ensure model robustness and prevent overfitting, the dataset was randomly split into training (80%) and testing (20%) sets. The model was implemented using the pyGAM library in Python, which allows the specification of different smoothing structures. In this case:

- Cubic splines were used for date, elevation, and precipitation.
- A 2D tensor-product spline was used to model the joint spatial effect of longitude and latitude.

The final model was fitted using penalized maximum likelihood estimation, with smoothing parameters λ selected automatically through generalized cross-validation (GCV).

3.2.3 Model Diagnostics and Performance

In evaluating the performance of the Generalized Additive Model (GAM) fitted to the climate dataset, a comprehensive set of statistical diagnostics was used. These metrics offer insight into the overall goodness-of-fit, the explanatory power of the model, the appropriateness of the functional form, and the general reliability of the results. The key indicators assessed include the pseudo R-squared, log-likelihood, Akaike Information Criterion (AIC), corrected AIC (AICc), Generalized Cross Validation (GCV) score, scale parameter, and effective degrees of freedom (EDoF). Each of these metrics contributes to a holistic understanding of how well the GAM captures the underlying structure in the data. The statistical summary of the fitted model is presented below:

METRIC	**VALUE**
Number of Samples	584,320
Effective Degrees of Freedom (DoF)	107.23
Log-Likelihood	-1,124,627.66
AIC	2,249,471.78
AICc	2,249,471.82
Generalized Cross-Validation (GCV)	2.1725
Scale	2.1718
Pseudo R-squared	0.8939

Table 2: GAM Model Diagnostics and Performance

The pseudo R-squared value, also referred to as the explained deviance, provides a measure analogous to the traditional R^2 in linear regression, although it is computed differently in the context of generalized models. In this study, the pseudo R-squared value was 0.8939, which implies that the

GAM was able to explain approximately 89.4% of the variability in the average temperature data.

This is a substantial proportion, indicating a highly effective model. A pseudo R-squared value closer to 1 denotes that the model accounts for nearly all the variability in the response variable. In environmental and climatic data modeling—where complex and nonlinear relationships are common—achieving a pseudo R^2 of this magnitude is a strong indicator of model adequacy. It also validates the appropriateness of including non-linear smoothing components and spatial splines in the model structure, confirming that simple linear assumptions would have likely been insufficient to describe the patterns in the data.

Moreover, the high explained deviance highlights that the major climatic trends (e.g., temporal seasonality, elevation influence, precipitation interaction, and spatial temperature gradients) were successfully captured by the selected covariates and functional forms. The model's performance suggests minimal unexplained variance, meaning that residual errors are relatively small and likely to be composed of random noise rather than systematic omission of important predictors.

The log-likelihood value of the model was reported as -1,124,627.66. The log-likelihood represents the logarithm of the likelihood function, which measures the probability of the observed data under the specified model. In practice, models with higher (i.e., less negative) log-likelihoods are preferred, as they indicate that the model fits the data well.

However, log-likelihood values on their own are not directly comparable across models of different complexity, which is why adjusted metrics such as Akaike Information Criterion (AIC) and Corrected AIC (AICc) are also computed. In this study, the AIC was 2,249,471.78 and the AICc was 2,249,471.82. These metrics penalize model complexity, striking a balance between goodness-of-fit and parsimony.

The AIC is defined as:

$$AIC = 2k - 2\log(L) \tag{3.2}$$

where:

- k is the number of estimated parameters,
- $\log(L)$ is the log-likelihood of the model.

The corrected AIC (AICc) is particularly useful when the sample size is not vastly larger than the number of model parameters. Although the sample size here (n = 584,320) is very large, the number of estimated degrees of freedom (over 100) still warrants the consideration of AICc. The minimal difference between AIC and AICc in this model indicates that the correction is negligible due to the very large sample size.

Lower AIC values generally indicate a better model when comparing multiple candidates. While only a single model was reported here, the AIC provides a useful benchmark for future model refinement or when comparing this model against nested or more complex alternatives.

The GCV score is a critical metric used during the automatic selection of smoothing parameters in GAMs. It approximates the leave-one-out cross-validation (LOOCV) error without explicitly removing each observation in turn. The GCV for this model was 2.1725, a relatively low value, especially given the scale of the response variable (temperature, typically ranging from around 10℃ to 35℃ in the study region).

GCV is especially important in regularized regression frameworks like GAMs, where each smooth term is penalized to prevent overfitting. A well-chosen penalty ensures that the model

remains generalizable to new data. A high GCV score would indicate that the model may be overfitting to noise in the training data, while a very low score paired with a low pseudo R-squared might suggest underfitting.

In this model, the combination of a high pseudo R-squared and low GCV score demonstrates that the model fits the training data well and is also likely to perform well on unseen data. This balance is essential for modeling real-world phenomena such as temperature, where predictive stability and interpretability are equally important.

The scale parameter, reported as 2.1718, represents the estimated variance of the residuals, or how much deviation remains unexplained after the model has been fitted. It is especially relevant when assuming a Gaussian distribution of the residuals, as was done here.

In the context of temperature modeling, this residual variance suggests that the average deviation of the predicted temperatures from the observed values is around $1.47^{\circ}C$ (since the square root of the scale gives the standard deviation). This is within acceptable limits for daily or monthly temperature modeling, especially across broad spatial domains.

The effective degrees of freedom (EDoF) of the model, reported as 107.23, represent the complexity of the model. Unlike the fixed degrees of freedom in linear regression, the EDoF in a GAM quantifies how flexible each smooth term is. Higher EDoFs indicate more ”wiggly” or complex smooth functions.

In this model, the spatial smooth component had the highest EDoF (55.8), indicating that it was modeling a substantial amount of complexity in the geographic variation of temperature. The date, elevation, and precipitation smooths also had moderate EDoFs (ranging from $\approx$14 to 20), supporting the assumption that each of these covariates contributes nonlinearly to the temperature variability.

The total EDoF also informs model selection and overfitting risk. The balance observed in this case – moderate EDoFs with a low GCV and high pseudo R squared – suggests that the model captures sufficient complexity without overfitting.

3.2.4 Interpretation of Model Terms

The Generalized Additive Model (GAM) fitted in this study comprises four key components: three univariate smooth functions and one bivariate tensor product spline. Each term in the model was chosen based on its theoretical relevance and empirical role in influencing surface temperature across the Kenyan landscape. This section offers a detailed interpretation of each model term, highlighting its contribution to the overall model, its statistical significance, and the implications for environmental and climatological analysis. The summary is shown in the table below:
Table 3 above summarizes the estimated effects of core covariates on surface temperature using an Ordinary Least Squares (OLS) regression model with spline basis expansions for nonlinear relationships and a tensor product spline to model spatial interactions. The model was fitted to over half a million observations of climate data across Kenya and achieves a high explanatory power with an R^2 of 0.8939, indicating that approximately 89% of the variability in temperature is captured by the selected predictors.

The model intercept is estimated at 28.20°C with a remarkably small standard error of 0.028 and a corresponding t-statistic of 1007.98. While the intercept is technically the expected temperature when all spline basis functions are at their reference (typically zero) values, it lacks a straightforward physical interpretation due to the transformed nature of the predictors. Nonetheless, it reflects the

Variable	Coefficient	Std. Error	t-value	p-value
Intercept	28.2025	0.028	1007.98	0.000 ***
Date	-0.8484	0.018	-47.12	0.000 ***
Elevation	-13.5498	0.037	-369.95	0.000 ***
Precipitation	-7.2726	0.060	-121.29	0.000 ***
Longitude $\times$ Latitude	10.0634	0.070	143.49	0.000 ***

Table 3: Summary of Key Variable Effects

baseline around which deviations from predictors such as elevation, precipitation, and time vary. The significance of the intercept confirms the model is properly scaled and centered.

The selected coefficient representing the temporal trend is -0.8484 with a standard error of 0.018, statistically significant at the 0.001 level. This coefficient corresponds to one of the spline basis components derived from date_num, a continuous representation of calendar dates. The negative sign suggests a declining trend in temperature for that portion of the time domain captured by the spline. This may represent a transition toward a cooler period (e.g., seasonal movement from a warm to a cold phase or dry to rainy season), consistent with Kenya's bimodal rainfall pattern and associated temperature oscillations. The temporal smooth captures intra-annual climatic cycles, and this significant term indicates the model is effectively detecting and utilizing seasonal signals.

One of the most influential predictors in the model is elevation, with a coefficient of -13.55°C and a very small standard error of 0.037, again highly significant. This large negative coefficient underscores the well-established environmental lapse rate—the rate at which temperature decreases with increasing altitude. Kenya's varied terrain, from the coastal lowlands to highlands like Mount Kenya and the Aberdares, provides a natural laboratory for observing such variation. The magnitude of this coefficient is particularly notable: it suggests that even modest changes in elevation can have a profound impact on surface temperature. This finding has major implications for modeling agriculture, human comfort indices, and vector-borne disease dynamics, all of which are sensitive to elevation-driven microclimates.

Precipitation's selected spline component coefficient is -7.27°C ($SE = 0.060, p < 0.001$), suggesting a strong inverse relationship between rainfall intensity and temperature during the modeled period. This likely reflects the short-term cooling associated with convective rainfall events, which are common in equatorial regions. Rainfall in Kenya is often associated with cumulonimbus cloud formation, which reduces solar insolation and leads to temporary surface cooling. Additionally, the latent heat of vaporization during rain events and the associated increase in surface moisture can suppress temperature. These findings are in line with prior studies in equatorial East Africa that report rapid cooling during and after precipitation events, especially in high-rainfall seasons.

The model includes a tensor product spline to capture interactions between longitude and latitude, producing a spatial smooth surface. The reported coefficient of 10.06°C (SE = 0.070,

t = 143.49) is for a specific tensor basis function within this smooth. This significant positive interaction term highlights how certain spatial regions—likely those closer to the equator or with lower elevations—experience systematically higher temperatures. This spatial component is critical for capturing latent geographical variation in temperature not explained by date, elevation, or precipitation alone. For instance, areas in northwestern Kenya (e.g., Turkana and West Pokot), which lie in lower latitudes and arid zones, consistently record higher temperatures than central highlands or lake basin areas. The spatial interaction term effectively models such regional heterogeneity and is crucial for spatial interpolation or kriging approaches in climate mapping.

The resultant GAM Equation is of the form;

$$\begin{aligned}\eta(y) = 28.2025 - 0.8484(Time) - 13.5498(Elevation) - 7.2726(Precipitation) \\ + 10.0634(Spatial) + \epsilon\end{aligned} \tag{3.3}$$

4 Conclusions

The Gaussian Process Regression (GPR) model exhibited excellent predictive performance in modeling spatial temperature variability across Kenya. By employing a Matérn kernel and optimizing hyperparameters through the L-BFGS algorithm, the model achieved a high coefficient of determination ($R^2 \approx 0.87$) and a low root mean square error (RMSE $\approx 1.61^\circ$C). GPR effectively captured spatial autocorrelation and adjusted well to heterogeneous geographic features. Furthermore, its ability to produce predictive uncertainty made it a valuable tool for probabilistic forecasting and risk-sensitive climate planning. However, the model's computational complexity, especially with large datasets, remains a notable limitation.

The Generalized Additive Model (GAM) provided a flexible and interpretable framework for modeling nonlinear relationships between temperature and explanatory variables such as time, elevation, and precipitation. By integrating spatial smooth functions, the model effectively accounted for spatial trends in the data. While GAM does not offer predictive uncertainty in the Bayesian sense, it allowed for clear identification and interpretation of the influence of each covariate. Its computational efficiency and ease of communication make it well-suited for policy applications and environmental analysis.

The GPR and GAM models offer complementary strengths. GPR is advantageous for uncertainty quantification and spatial interpolation in data-sparse regions, whereas GAM excels in identifying and explaining nonlinear covariate effects. Together, they provide a comprehensive approach to climate modeling. Integrating these models into a hybrid framework enhances robustness, enabling precise, interpretable, and spatially aware climate predictions tailored to Kenya's diverse geography.

Building on the promising results of both the GPR and GAM models, future work will explore the development of scalable hybrid frameworks that combine the interpretability of GAM with the uncertainty quantification of GPR. One direction involves extending the current models to handle multi-output prediction tasks, such as simultaneous forecasting of temperature and precipitation. Moreover, studies could be conducted in a different geographic location and conduct a comparative analysis for the models. Additionally, incorporating remote sensing indices like NDVI or land surface temperature could enhance the environmental context and improve model performance. To address the computational burden of GPR on large datasets, sparse approximations or variational inference techniques will be investigated. Finally, adapting the hybrid model into a near-real-time predictive tool for climate early warning systems in Kenya would significantly enhance its practical utility for decision-makers.

References

B. Knüsel and C. Baumberger, "Understanding climate phenomena with data-driven models," *Studies in History and Philosophy of Science Part A*, vol. 84, pp. 46–56, 2020.

P. O. Sagero, "Assessment of past and future climate change as projected by regional climate models and likely impacts over kenya," Ph.D. dissertation, PhD Thesis]. School of Pure and Applied Science, Kenyatta University, 2019.

I. Rangwala, W. Moss, J. Wolken, R. Rondeau, K. Newlon, J. Guinotte, and W. R. Travis, "Uncertainty, complexity and constraints: how do we robustly assess biological responses under a rapidly changing climate?" *Climate*, vol. 9, no. 12, p. 177, 2021.

F. Giorgi, "Thirty years of regional climate modeling: where are we and where are we going next?" *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 11, pp. 5696–5723, 2019.

S. Bathiany, H. Dijkstra, M. Crucifix, V. Dakos, V. Brovkin, M. S. Williamson, T. M. Lenton, and M. Scheffer, "Beyond bifurcation: using complex models to understand and predict abrupt climate change," *Dynamics and Statistics of the Climate System*, vol. 1, no. 1, p. dzw004, 2016.

D. P. Vázquez, E. Gianoli, W. F. Morris, and F. Bozinovic, "Ecological and evolutionary impacts of changing climatic variability," *Biological Reviews*, vol. 92, no. 1, pp. 22–42, 2017.

S. Vansteelandt and O. Dukes, "Assumption-lean inference for generalised linear model parameters," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 84, no. 3, pp. 657–685, 2022.

A. Singh, J. Nagar, J. Amutha, and S. Sharma, "P2ca-gam-id: Coupling of probabilistic principal components analysis with generalised additive model to predict the k- barriers for intrusion detection," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107137, 2023.

J. Lindström, A. A. Szpiro, P. D. Sampson, A. P. Oron, M. Richards, T. V. Larson, and L. Sheppard, "A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates," *Environmental and ecological statistics*, vol. 21, pp. 411–433, 2014.

S. Chen, G. Long, J. Jiang, D. Liu, and C. Zhang, "Foundation models for weather and climate data understanding: A comprehensive survey," *arXiv preprint arXiv:2312.03014*, 2023.

A. Sadeghi-Lotfabadi and K. Ghiasi-Shirazi, "Speeding up l-bfgs by direct approximation of the inverse hessian matrix," *Computational Optimization and Applications*, pp. 1–28, 2025.

A. Muyskens, B. W. Priest, I. R. Goumiri, and M. D. Schneider, "Identifiability and sensitivity analysis of kriging weights for the matern kernel," *arXiv preprint arXiv:2410.08310*, 2024.

R. H. Byrd, S. L. Hansen, J. Nocedal, and Y. Singer, "A stochastic quasi-newton method for large-scale optimization," *SIAM Journal on Optimization*, vol. 26, no. 2, pp. 1008–1031, 2016.

Q. Jin, R. Jiang, and A. Mokhtari, "Non-asymptotic global convergence analysis of bfgs with the armijo-wolfe line search," *arXiv preprint arXiv:2404.16731*, 2024.