Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 174
DOI: 10.11159/icsta24.174

Exploring Universal Kriging Modelling for Functional Greenness Data Captured by Digital Webcams

Raziyeh Hosseini
Department of Statistics, Ludwig-Maximilians-University
Ludwigstr33, Munich, Germany
raziyeh.hosseini@stat.uni-muenchen.de

Abstract **-** This paper examines the application of universal kriging predictors for analysing functional greenness data obtained from digital webcam images. The goal is to predict greenness curves in areas where image data is scarce or unavailable. The study employed an unsupervised Region of Interest (uRoI) technique to extract greenness time series from 95 webcams within the AMOS database. To predict greenness curves at unsampled sites, we employed two approaches: (i) Universal Cokriging (UCoK) predictor using cross-variography and (ii) Universal Trace-Kriging (UTrK) predictor based on trace-variography means. The greenness curves were pre-processed using non-parametric fitting, with smoothing parameters selected via cross-validation.
The performance of the two universal kriging predictors was evaluated through a Monte Carlo simulation study, using the root mean square prediction error (RMSPE) as the metric. The results showed that both universal kriging predictors effectively modeled greenness curves for unsampled camera locations, with the UTrK approach producing slightly better outcomes. Additionally, the dimensionality reduction of greenness data before geostatistical analysis via UCoK method impacted the results' quality. The findings suggest that the UTrK predictor could be a preferable choice for modeling greenness curves at unsampled camera locations due to the simplicity of the model and less computational efforts in its application. Nonetheless, the UCoK predictor shows potential for improvement, particularly when incorporating additional variables such as the years of recorded images.

Keywords**:** Universal Kriging Predictor, Functional Data, Image Processing

1. Introduction

Greenness is a vital indicator in phenology studies, allowing researchers to monitor the timing and duration of plant growth. To measure vegetation greenness, remote sensing techniques such as satellite imagery and webcam images are employed. By capturing images at regular intervals from grassland sites, researchers can observe changes in plant growth and identify key phenological events. Recent advancements in computer vision and machine learning have facilitated automated approaches for detecting Regions of Interest (RoIs). One method, described in [1], utilizes a vast network of webcams from the Archive of Many Outdoor Scenes (AMOS) database to define RoIs and generate greenness values. However, challenges can arise at specific camera locations due to image quality issues, nighttime recordings, and variations in webcam positioning throughout the year. These issues can reduce the effectiveness of the unsupervised Region of Interest (uRoI) method, affecting the accurate identification of RoIs and the resulting greenness time series for the analysis year. To overcome these challenges, geostatistical techniques such as kriging modeling can be employed to help scientists, particularly phenologists, predict greenness trends at unsampled locations using measured observations.

Various prediction methods are available to derive accurate forecasts from measured observations. Danie G. Krige [2] is renowned for pioneering the spatial interpolation method known as Kriging, which aims to estimate the value of a variable at an unsampled location based on monitored observations from nearby locations. While traditional kriging deals with scalar data at specific spatial points, functional kriging extends this approach to handle functional data, where each location is associated with an entire curve or function. Functional Data Analysis (FDA) [3] offers an alternative to multivariate analysis by treating data as a single functional object with an underlying smooth dynamic that drives temporal variation. Although theoretical advancements have extended Ordinary Kriging to functional data (e.g., [4] and [5]), Universal Kriging is more frequently used in practical applications. It addresses non-stationary data by incorporating regionalized and non-regionalized terms, alongside the traditional spatial autocorrelation term employed in Ordinary Kriging [6].

This paper investigates the use of kriging-based modeling to predict greenness curves at unobserved locations based on measured greenness data from available camera locations. We analyse raw greenness values obtained from 95 AMOS webcams, primarily located in the USA. These values are derived using the (uRoI) approach applied to imagery data recorded by the selected webcams from 2013 to 2016. As part of data preprocessing, we generate a set of smooth greenness curves from the raw values in which the smoothing parameter of Lambda is selected via generalized cross-validation (GCV) measure [3]. To estimate greenness curves at unobserved locations, we implement two universal kriging predictors for the functional greenness data: Universal Trace-Kriging (UTrK), estimated through trace-variography, and Universal Co-Kriging (UCoK), estimated through cokriging of the components over a functional basis. A Monte Carlo simulation was conducted to compare the performance of these two approaches.

The paper is structured as follows: Section 2 reviews the universal kriging methodology for non-stationary, spatially correlated functional data developed by [6] and [7]. Section 3 provides an overview of data preparation, application, and prediction. The results of the Monte Carlo simulation, which evaluates the performance of the two predictors, are presented at the end of Section 3. Finally, Section 4 offers conclusions and discusses future research directions.

2. Overview of Universal Kriging for Functional Data

In this section, we overview the methodology of kriging-based predictors for analysing functional data based on the approach presented in [6]. Let $\{X_{s_i}(t),\ i = 1, \dots, N\ , \mathrm{t} \in T\}$ be a set of observed curves over a set of spatial location s_i from a generating process $\boldsymbol{X}_s$ which is non-stationary on $(\Omega, \mathcal{F}, \mathbb{P})$. Here the index s indicates a location in $D_s \subset \mathbb{R}^2$ and the process $\boldsymbol{X}_s$ is a random element of the Hibert space $L^2(T)$ on the time interval T, equipped with the usual inner product and the induced norm.

In Ordinary Kriging, when the process is stationary, it is assumed that the mean vector is constant for all $s \in D_s$, and the variance, covariance, and cross-covariance functions depend only on the separation vector $h = s_i - s_j$ for each pair of locations $s_i,\ s_j \in D_s$ rather than on the specific locations themselves However, in real-world applications, such as in this study, the mean value of some spatial data can vary because it also depends on the absolute location of the sample. As a result, the random process $\boldsymbol{X}_s$ can be expressed as $\boldsymbol{X}_s = m_s + \delta_s$, where the *drift* term m_s is not constant and is modeled using a functional regression model: $m_s(t) = \sum_{l=0}^{L} f_l(s) a_l(t)$, that $\{a_l(t),\ l = 0, \dots, L\}$ are functional coeffcients in $L^2(T)$, independent of the spatial locations, and $f_l(s)$ are known functions of the spatial variables on $s \in D_s$ which are constant with respect to the variable $t \in T$.

In the framework of the Universal Trace Kriging (UTrK) approach, a real-valued function called trace-covariogram by [6] denoting as $C_{tr}(\| h_{ij} \|) = \mathbb{E}[\langle \delta_{s_i}, \delta_{s_i} \rangle]$ which acts as a global measure of spatial dependence of the residual field $\{\delta_s, s \in D_s\}$ or any fixed $h_{ij} = (s_i - s_j) \in \mathbb{R}^2$.

Given the observations $\mathbf{X}_{s_1}, \dots, \mathbf{X}_{s_n}$ at the corresponding locations $s_1, \dots, s_n$ the prediction of an unobserved element $\mathbf{X}_{\mathbf{s_0}}$ at the location $\mathbf{s}_0$, is achievable by the best linear unbiased predictor $\mathbf{X}^*_{\mathbf{s}_0} = \sum_{i=1}^{n} \lambda_i^* \mathbf{X}_{s_i}$ and the weights are chosen by minimization of $\mathbb{E}\left[\left\|\mathbf{X}_{\mathbf{s_0}} - \sum_{i=1}^{n} \lambda_i \mathbf{X}_{s_i}\right\|^2\right]$ such that $\mathbb{E}\left[\left\|\mathbf{X}_{\mathbf{s_0}} - \sum_{i=1}^{n} \lambda_i \mathbf{X}_{s_i}\right\|\right] = 0$ for all weights $\lambda_1, \lambda_2, \dots, \lambda_n$.

By developing the unbiasedness constraints, using the functional form of *drift* term, and introducing $L + 1$ Lagrange multipliers, the optimization problem equivalent to solving the kriging system of:

$$\begin{pmatrix} \boldsymbol{C}_{tr} & \boldsymbol{F} \\ \boldsymbol{F}^T & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} \\ \boldsymbol{\zeta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}_{tr,O} \\ \boldsymbol{F}_0 \end{pmatrix} \tag{1}$$

where $\boldsymbol{C}_{tr} = \begin{bmatrix} C_{tr}(0) & \cdots & C_{tr}(h_{1n}) \\ \vdots & \ddots & \vdots \\ C_{tr}(h_{n1}) & \cdots & C_{tr}(0) \end{bmatrix}$, $\boldsymbol{C}_{tr,O} = \begin{bmatrix} C_{tr}(h_{10}) \\ \vdots \\ C_{tr}(h_{n0}) \end{bmatrix}$, $\boldsymbol{F} = \begin{bmatrix} f_0(s_1) & \cdots & f_L(s_n) \\ \vdots & \ddots & \vdots \\ f_0(s_n) & \cdots & f_L(s_n) \end{bmatrix}$ and $\boldsymbol{F}_0 = \begin{bmatrix} f_0(s_0) \\ \vdots \\ f_L(s_0) \end{bmatrix}$.

Here, $\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}$ is the vector of the kriging weights, $\boldsymbol{O}$ is the zero matrix, the vector of and $\boldsymbol{\zeta} = \begin{bmatrix} \varsigma_0 \\ \vdots \\ \varsigma_L \end{bmatrix}$ are the Lagrange multipliers corresponding to the $(L + 1)$ unbiasedness constraints.

It should also be noted that an application of Kriging predictors requires the estimation of the trace variogram through solving the system (1). An empirical estimate of the trace variogram defining as:

$$g(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \| \delta_{s_i} - \delta_{s_j} \|^2 , \tag{2}$$

where, h represents the lag, $N(h)$ refers to the set of pairs at lag h, and $|N(h)|$ is the number of elements in the set $N(h)$. Next, a valid variogram model is fitted to the empirical estimate. Commonly used one-dimensional parametric families, such as Gaussian, Exponential, Spherical, or Matern, can be applied. (See [6] for more details.)

The projection-based approach for functional data, naming as UCoK predictor, can be considered as a generalization of the stationary approach. In UCoK method, the functional data are represented through a dimensionality reduction by utilizing functional principal component analysis. Subsequently, this approach proceeds to forecast functional data at unobserved locations using their predicted scores. Let $\varphi_k(t)$ be the basis system consisting of k basis functions, then the process X_s can be presented as $\boldsymbol{X}_s = \sum_{k=1}^{K} \xi_k(s)\varphi_k$, where $\xi_k(s)$ is the projection of $\boldsymbol{X}_s$ on the k-th element of the basis expansion. So, the process can be represented by its distributional properties of the k-dimensional random coefficients $\{\boldsymbol{\xi}(s), s \in D_s\}$, in which $\boldsymbol{\xi}(s) = (\xi_1(s), \dots \xi_k(s))^T$.

The UCoK predictor of $\mathbf{X}_{\boldsymbol{s_0}}$ is the Best Linear Unbiased Predictor $\mathbf{X}^*_{\boldsymbol{s_0}} = \sum_{i=1}^{n} \Lambda^*{}_i \mathbf{X}_{s_i}$, where $\Lambda^*{}_i$ are the operators to solve the constrained minimization problem of $\mathbb{E}\left[\left\|\mathbf{X}_{s_0} - \sum_{i=1}^{n} \Lambda^*{}_i \mathbf{X}_{s_i}\right\|^2\right]$ such that $\mathbb{E}\left[\mathbf{X}_{s_0} - \sum_{i=1}^{n} \Lambda^*{}_i \mathbf{X}_{s_i}\right] = 0$. Solving this kriging problem is equivalent to determine an optimal estimate of $\boldsymbol{\xi}(\boldsymbol{s_0})$ through the solution of $\mathbb{E}[\|\boldsymbol{\xi}(\boldsymbol{s_0}) - \sum_{i=1}^{n} \mathbb{L}_i \boldsymbol{\xi}(s_i)\|^2]$ such that $\mathbb{E}[\boldsymbol{\xi}(\boldsymbol{s_0}) - \sum_{i=1}^{n} \mathbb{L}_i \boldsymbol{\xi}(s_i)] = 0$. Therefore, the kriging weights $\mathbb{L}_i$ are found by solving the cokriging system:

$$\begin{pmatrix} \mathbb{C} & \mathbb{F} \\ \mathbb{F}^T & \mathbb{O} \end{pmatrix} \begin{pmatrix} \mathbb{L} \\ \mathbb{Z} \end{pmatrix} = \begin{pmatrix} \mathbb{C}_{\mathbf{0}} \\ \mathbb{F}_0 \end{pmatrix} \tag{3}$$

where $\mathbb{C} = \begin{bmatrix} \mathbb{C}_{11} & \cdots & \mathbb{C}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbb{C}_{K1} & \cdots & \mathbb{C}_{KK} \end{bmatrix}$, $\mathbb{C}_{\mathbf{0}} = \begin{bmatrix} \mathbb{C}_{01} \\ \vdots \\ \mathbb{C}_{0L} \end{bmatrix}$ and $\mathbb{F} = \begin{bmatrix} \mathbb{F}_{10} & \cdots & \mathbb{F}_{1L} \\ \vdots & \ddots & \vdots \\ \mathbb{F}_{K0} & \cdots & \mathbb{F}_{KL} \end{bmatrix}$. Here $\mathbb{C}_{ij}$ denotes the cross-covariance matrix between coefficient vectors $\boldsymbol{\xi}(s_i)$ and $\boldsymbol{\xi}(s_i)$, for $i, j = 1, \dots, n$ and $\mathbb{C}_{0i}$ is the cross-covariance matrix between $\boldsymbol{\xi}(s_0)$ and $\boldsymbol{\xi}(s_i)$ respectively. Moreover, for $i, j = 1, \dots, n$ and $\mathbb{F}_{iL} = diag(f_l(s_i), \dots, f_l(s_i))$ and $\mathbb{Z}_L$ is the vector of Lagrange multipliers for $l = 1, \dots, L$. It is important to note that the dimension K directly affects the number of estimated variogram and cross-variogram structures involved in solving the kriging system in equation (3). (See [7] for more details.)

3. Universal Kriging Predictors for Greenness Curves

A total of ninety-five camera locations from the Archive of Many Outdoor Scenes (AMOS) were selected for this research. To extract the greenness percentage from the imagery dataset, we employed an automated image processing technique outlined in [1]. The data retrieval process involved the following steps: 1) Setting parameters: desired webcam (camera ID), desired year, and month. 2) Downloading images from the specified webcams. 3) Treating each pixel's colour channel as the observation unit and each time point as the variable. 4) Performing Singular Value Decomposition (SVD) on the data image matrix. 5) Conducting cluster analyses. 6)Computing the Optimal Criterion (OC) for each cluster image. The results involved the segmentation of the image into pixel clusters to identify potential regions of interest (RoI). The cluster that best met the optimal criterion (OC2) was selected as the candidate Region of Interest (RoI), and its corresponding greenness percentage time series was used as the database for further analysis. This automated image processing was performed on images from 95 webcams taken between 2013 and 2016. The resulting successful greenness time series are shown in Table 1.

Table 1: Results of URoI approach in 95 selected webcams over different years of analysis.

Years of analysis	Greenness Data Availability	Additional information
2013	54	Limited greenness data availability.
2014	85	The approach yielded satisfactory results.
2015	95	Considered as a candidate year for further analysis.
2016	24	Limited greenness data availability.

Following successful outcomes using the uRoI approach for 2014 and 2015, we concentrated our further analysis on the greenness time series data from 2015 within the functional analysis and spatial modeling. Importantly, some days featured missing values. To manage this, we omitted locations with more than 200 days of missing data, ultimately retaining 90 locations for subsequent analysis. Figure 1, in the right panel, displays the coordinates of the 90 selected webcams that provided the functional greenness data recorded in 2015.

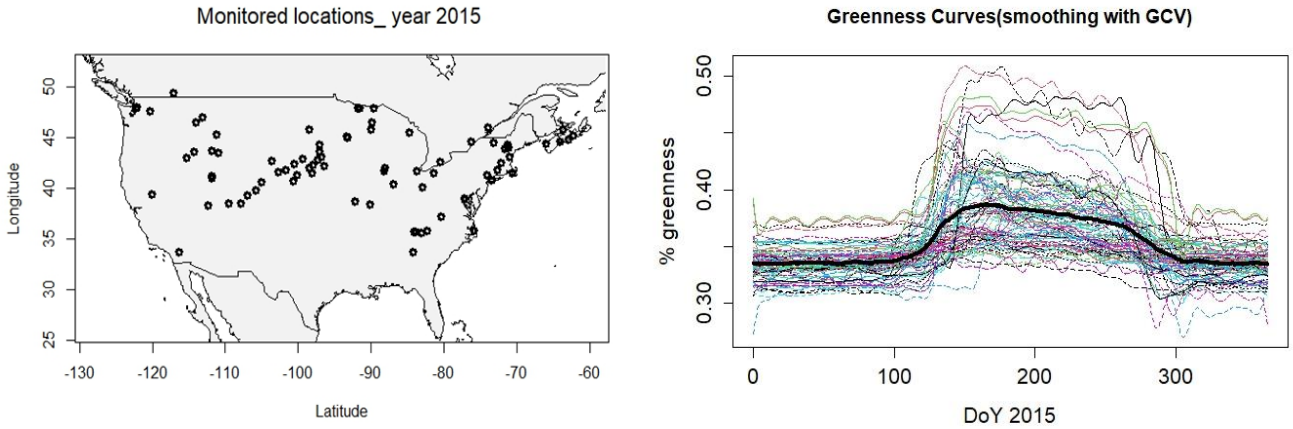

Fig. 1: Patterns of 90 camera-locations with measured greenness values (left panel); and smoothing curves of the corresponded 90 webcams (right panel)

To create greenness curves, we began by defining a set of basis functions and determining the coefficients needed to express these functions as a linear combination of the selected basis functions. For managing missing data, we used the Last Observation Carried Forward (LOCF) method. As part of the standard preprocessing in Functional Data Analysis (FDA), we employed a spline basis system, specifically B-spline basis functions. To smooth the greenness acceleration, we introduced a functional parameter object that penalizes the fourth derivative, using the Generalized Cross-Validation (GCV) method for this purpose (see [3]). Figure 1 shows the 90 greenness curves for the selected AMOS webcams in the right panel, with the mean curve indicated by the black solid line.

When applying UCoK predictors, we conduct a functional principal component analysis (FPCA) to define a sample of 90 curves or functions defining as the expansion system then proceeds to spatially interpolate the coefficient of basis expansion. The most common criterion for choosing the number of components K is the Fraction of Variance Explained (FVE). Based on the FPC analysis of the greenness data and its visual representations, we concentrate on the first three functional principal components(fpcs), which together explain about 95% of the variability in the data (see Figure 2, the top-left panel). The fact that unrotated functional principal components are easier to interpret, highlights the necessity for a rotation of functional principal components (fpcs) to uncover more insightful variation components. To achieve this, we utilized the VARIMAX rotation algorithm.

In Figure 2, we also illustrate the rotation of fpcs of the measured greenness observations recorder by the 90 webcams in 2015. The rotated FPCs are represented with the mean curve (solid line) and "+" and "-" symbols, indicating the effects of adding or subtracting small quantities of each principal component, which describe variations around the mean curve. The first rotated FPC explains 60.5% of the total variation, reflecting a significant shift in spring, potentially

marking the start of the growing season. It shows a notable vertical shift in the mean greenness curve during summer and mid-autumn, followed by a substantial deviation in late autumn, which likely corresponds to the end of leaf colouring in autumn.

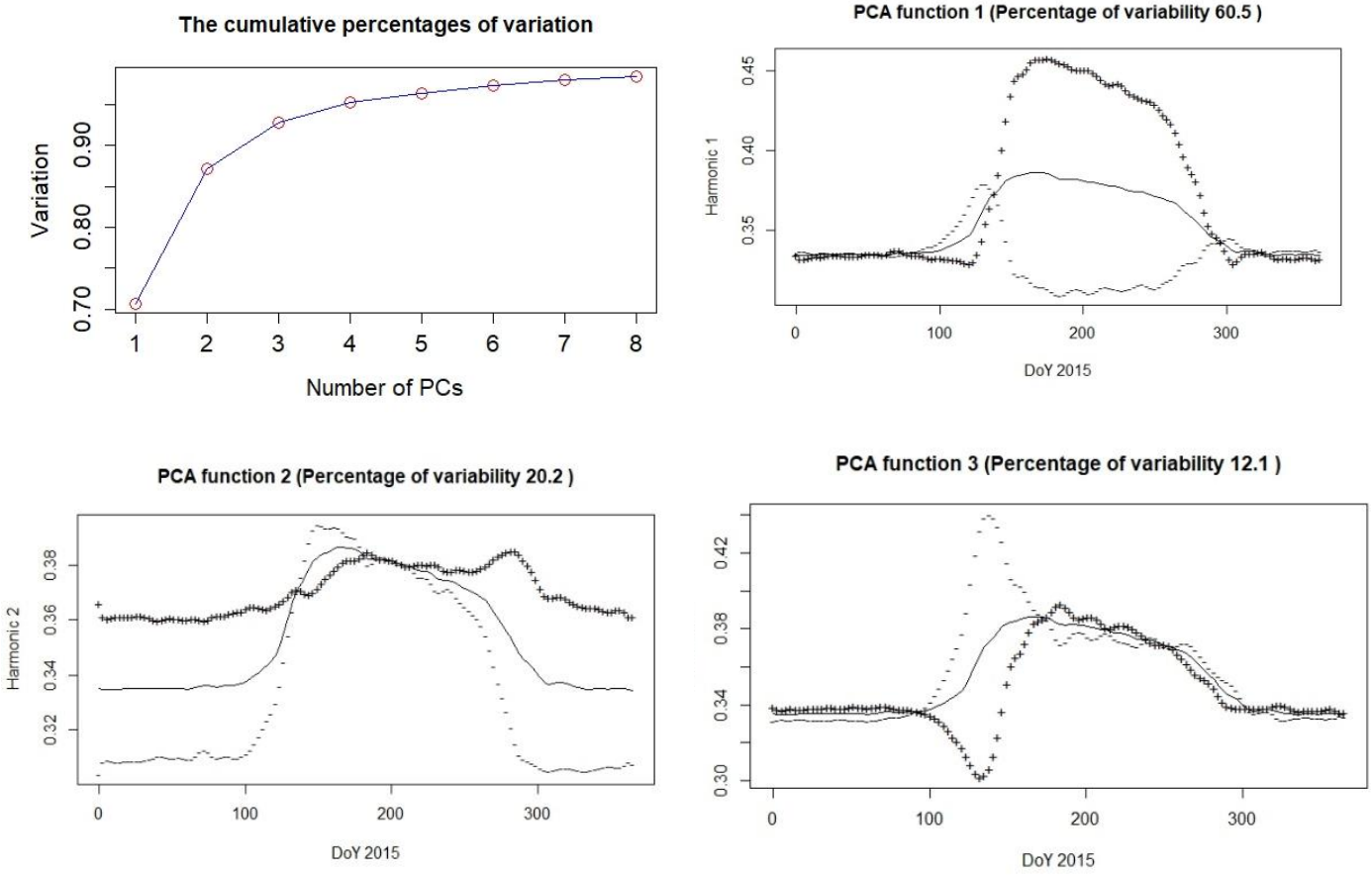

Fig. 2: FPCA plots: three rotated fpcs of 90 greenness curves

To implement UCoK predictors, we analyse the variograms and cross-variograms of the residuals obtained by fitting a Matern model to the scores of the varimax rotated functional principal components, as illustrated in Figure 3. It also displays the three omni-directional autocorrelation variograms (rfpc1, rfpc2, and rfpc3) along with three cross-variograms of the greenness curves.

For the application of UTrK, we construct a functional meta-model and estimate the variogram for the greenness curves. We fit an Exponential model to estimate the drift and trace-variogram. In Figure 3, the empirical trace-variogram is shown alongside the fitted model, with symbols representing empirical estimates and the solid line indicating the fitted model. Numbers provided indicate the number of location pairs used for each empirical estimate. The bottom panels of Figure 3 display a subsample of 30 predicted greenness curves for a randomly selected set of camera locations within the AMOS database. The left panel shows predictions generated using the UCoK predictor, while the right panel shows those generated using the UTrK predictor.

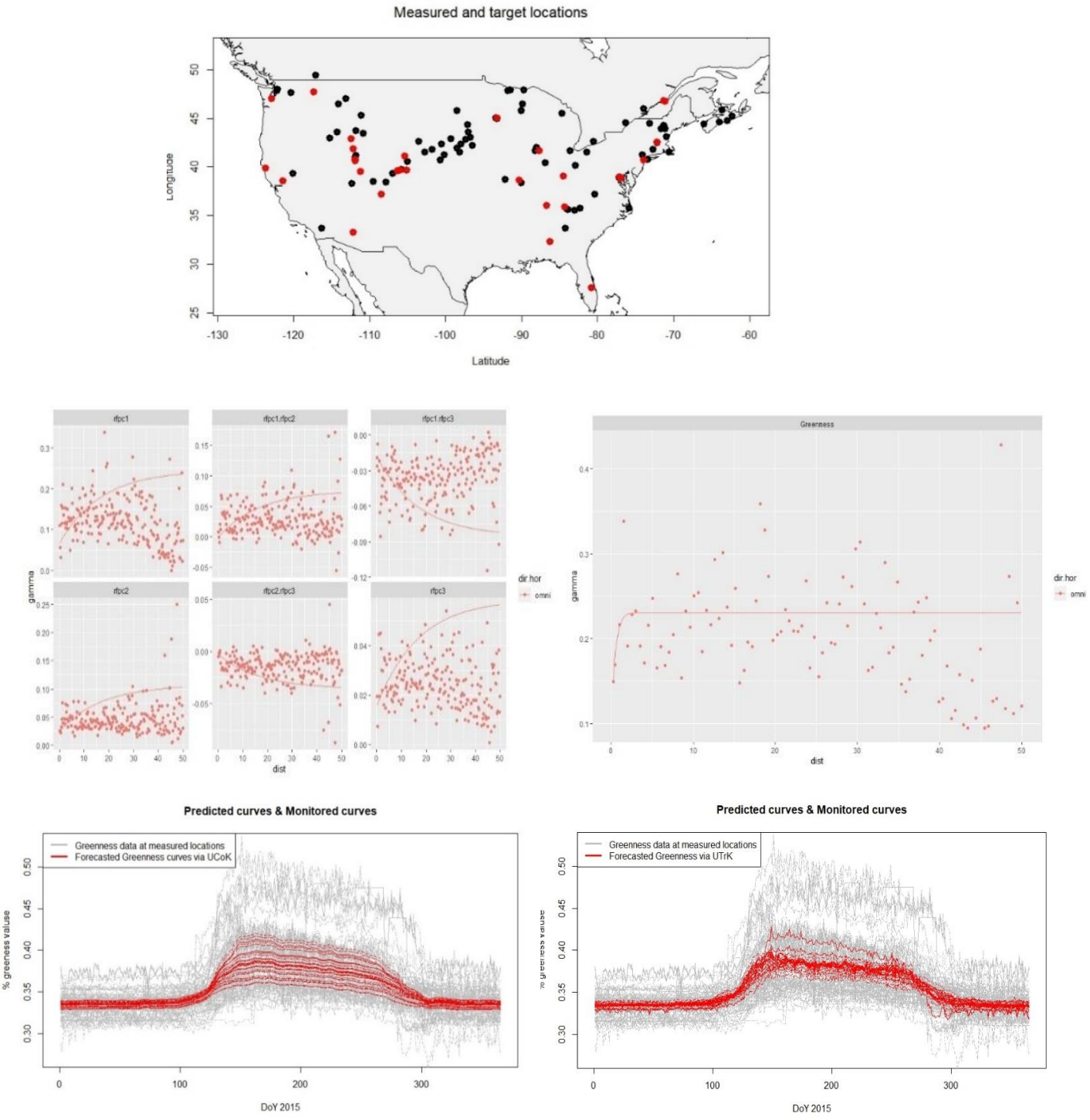

Fig. 3: Top panel: Coordinates of the sampled camera locations (90 black dots) and the predicted sites (30 red dots). From left to right: estimated variograms and cross-variograms of the scores along the first three rotated FPCs; empirical trace-variogram; forecasted greenness curves using the UCoK predictor (left) and the UTrK predictor (right)

To evaluate the effectiveness of the two universal Kriging predictors, we conducted a Monte Carlo simulation study. We estimated the Root Mean Squared Prediction Errors (RMSPEs) by running experiments with 300 randomly selected

training and test sets for each parameter value mmm. The dataset was randomly divided into training sets containing m% of the data and test sets with $(100 - m)$%. We examined training sets with m values of 80, 70, 60, and 50.

Additionally, we investigated how varying numbers of functional principal components affected the UCoK method by running simulations with 2, 3, and 4 components. Figure 4 displays box plots of the mean and median indices using 2, 3, and 4 functional principal components to aid in comparison.

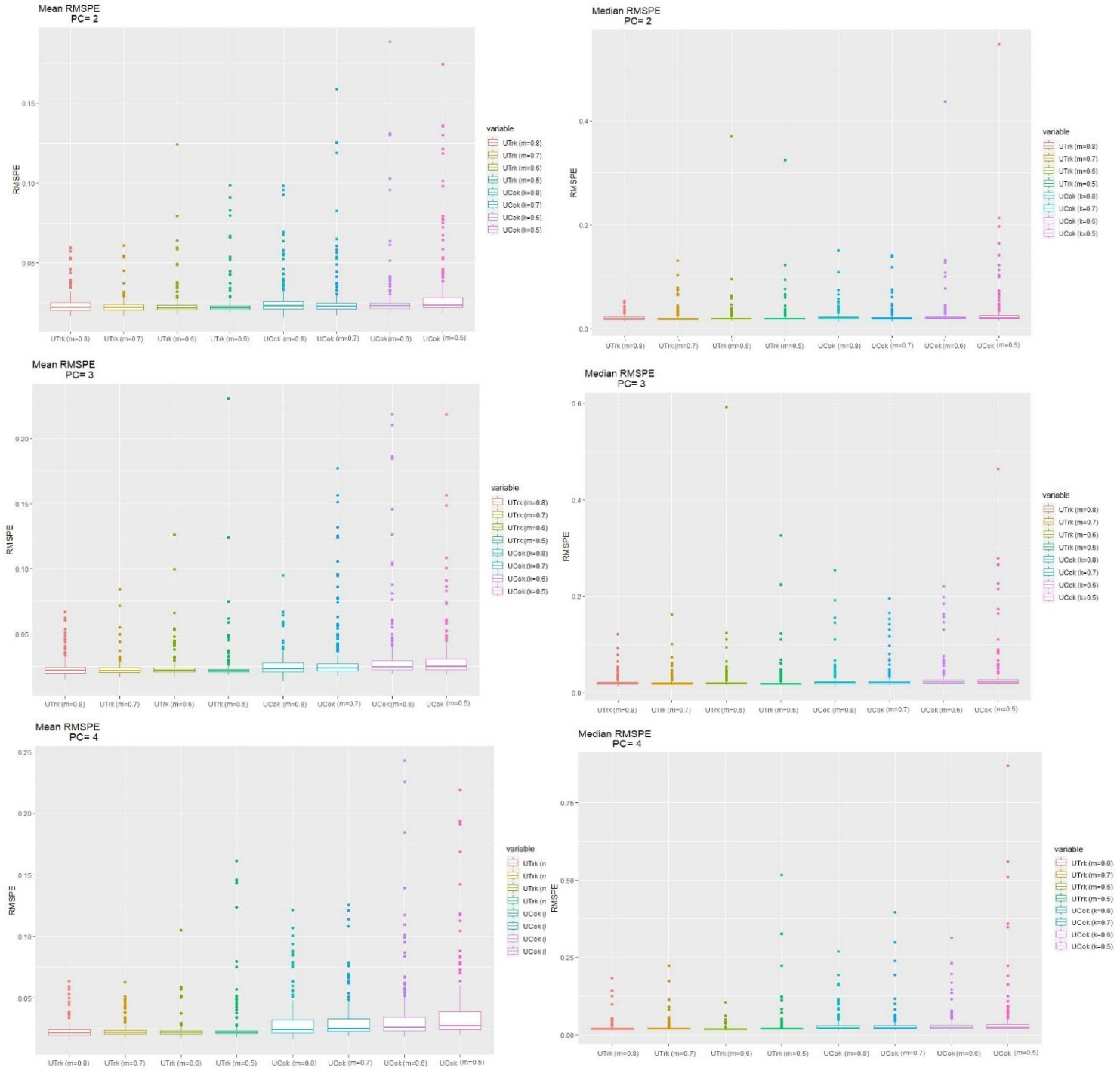

Fig. 4: Boxplots of Mean and Median RMSPE indices for UCoK and UTrK predictors

The Monte Carlo results indicate that both universal predictors provide nearly identical forecasts, with UTrK showing slightly better performance. In both models, predictive accuracy improves as the training set size increases, highlighting the impact of the parameter m on the amount of information available for predictions. Although the UCoK predictor could potentially lead to better results, the relatively straightforward nature of the UTrK predictor likely accounts for its marginally superior performance due to its simpler model and less computational works in application.

4. Conclusion

A significant challenge when working with image data from digital webcams is the variability in image quality and resolution. A recent unsupervised image processing method, uRoI, can analyse webcam images on a large scale. While this approach generally yields satisfactory results, certain limitations can arise in conditions with poor image quality, nighttime recordings, or significant changes in camera positioning. By employing geostatistical techniques, some limitations of the uRoI method for camera-specific data can be addressed.

In this study, we used two Universal Kriging predictors on greenness data obtained from digital webcam images via uRoI approach. The investigated methods can predict greenness curves in areas where image data is scarce or unavailable. The study demonstrated the effectiveness of both kriging predictors in modeling greenness curves for unobserved locations. Additionally, reducing the dimensionality of greenness data before geostatistical analysis with the UCoK method was found to affect result accuracy.

This comparative analysis provides a valuable reference for researchers, particularly phenologists interested in conducting comprehensive spatial sampling analyses to estimate greenness curves in areas with limited or no image data. Effectively implementing a universal kriging model in unsampled locations requires collecting the most informative greenness curves from a set of sample locations. Therefore, optimizing the spatial pattern of these sample locations for greenness mapping, considering the variances of the two examined universal kriging predictors, shows significant promise.

Moving forward, it is planned to conduct comprehensive sampling analyses to identify the optimal sampling pattern that minimizes the variance of the Universal Kriging predictors.

Acknowledgements

This work is part of the Ph.D. research of Raziyeh Hosseini, conducted under the supervision of Professor Dr. Göran Kauermann. The author sincerely thanks the AMOS team for granting direct access to the recorded imagery database, which greatly enhanced the data preparation process.

References

[1] L. Bothmann, A. Menzel, B. H. Menze, C. Schunk, and G. Kauermann, "Automated processing of webcam images for phenological classification," PloS One, vol. 12, no. 2, p. e0171918, 2017.

[2] D. G. Krige, "A statistical approach to some basic mine valuation problems on the Witwatersrand," Journal of the Southern African Institute of Mining and Metallurgy, vol. 52, no. 6, pp. 119-139, Jun. 1951.

[3] J. Ramsay and B. Silverman, "Functional Data Analysis," Springer Series in Statistics, Springer, New York, 2006.

[4] S. Ullah and C. F. Finch, "Applications of functional data analysis: A systematic review," BMC Medical Research Methodology, vol. 13, pp. 1-12, 2013.

[5] D. Nerini, P. Monestiez and C. Manté, "Cokriging for spatial functional data," in Journal of Multivariate Analysis, vol. 101, no. 2, pp. 409-418, 2010.

[6] A. Menafoglio, P. Secchi, and M. Dalla Rosa, "A universal kriging predictor for spatially dependent functional data of a Hilbert space," Electronic Journal of Statistics, vol. 7, pp. 2163-2189, 2013.

[7] A. Menafoglio, O. Grujic, and J. Caers, "Universal kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales," Spatial Statistics, vol. 15, pp. 39-55, 2016.