Heliyon 9 (2023) e18334

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Predicting rainfall and irrigation requirements of corn in Ecuador

Miguel Flores [a], Ángel Llambo [b], Danilo Loza [b], Salvador Naya [c], Javier Tarrío-Saavedra [c,*]

[a] *Departamento de Matemática, Grupo MODES, Facultad de Ciencias, Escuela Politécnica Nacional, Ladrón de Guevara E11–253, Quito, 17–01–2759, Pichincha, Ecuador*

[b] *Departamento de Matemática, Facultad de Ciencias, Escuela Politécnica Nacional, Ladrón de Guevara E11–253, Quito, 17–01–2759, Pichincha, Ecuador*

[c] *Grupo MODES, CITIC, Departamento de Matemáticas, Escola Politécnica de Enxeñaría de Ferrol, Universidade da Coruña, Mendizábal s/n, Ferrol, 15403, A Coruña, Spain*

ARTICLE INFO

Keywords:
Functional data analysis
Regression
Evapotranspiration
Corn
Effective rainfall
Crop irrigation

ABSTRACT

This work is a case study whose objective is prediction of irrigation needs of corn crops in different regions of Ecuador; being this a fundamental basic food for the country's economy, as in the remaining countries of the Andean area. The proposed methodology seeks to help improving the quality of corn crop. Specifically, we propose the application of regression models, within the framework of Functional Data Analysis (FDA), to predict the amount of rainfall (scalar response variable) in the places with the highest production of corn in Ecuador, as a function of functional covariates such as temperature and wind speed. From the estimation of the amount of rainfall, effective precipitation is calculated. This is the fraction of water used by the crops, from which the value of real evapotranspiration or ETc is obtained and, more importantly, the irrigation requirements at each stage of the corn crop, for its adequate physiological development. Application of regression models based on functional basis, Functional Principal Components (FPC) or Functional Partial Least Squares (FPLS) for scalar response variable, allows us to use the information of variables such as wind speed and temperature (of functional nature) in a better way than using multivariate models, for predicting the amount of rainfall, obtaining, as a result, very explicative models, defined by a high goodness of fit ($R^2 = 0.97$, with 6 significant parameters and an error of 0.14) and practical utility. The model has been also applied to North Peru regions, obtaining rainfall prediction errors between 9% and 22%. Thus, the geographical limitations of the model could be the Andean regions with similar climate. In addition, this study proposes the application of FDA exploratory analysis and FDA outlier detection techniques as a common and useful practice in the specific domain of rainfall prediction studies, prior to applying the regression models.

1. Introduction

The current technological developments related to sensorics and IoT have made it possible for researchers in many areas to have large volumes of information; so, there is a great interest in learning and developing statistical tools for their analysis [1,2]. These

* Corresponding author.
E-mail address: javier.tarrio@udc.es (J. Tarrío-Saavedra).

https://doi.org/10.1016/j.heliyon.2023.e18334
Received 18 December 2022; Received in revised form 13 July 2023; Accepted 13 July 2023

Available online 20 July 2023
2405-8440/

databases, composed of features or variables continuously monitored with respect to time, and often dependent on meteorological weather, are usually complex, both in terms of variety and dimension. These characteristics in data collection mean that many variables are of a functional nature, i.e., whose realizations can be considered curves of infinite dimension [3]. Thus, developing of specific statistical techniques, such as those corresponding to Functional Data Analysis (FDA), to deal with this increasingly common paradigm of data, are needed.

FDA is a powerful and relatively new statistical tool that provides a series of techniques related to descriptive statistics, analysis of variance, regression and classification models specially designed and adapted to be applied to infinite dimensional data [4,5] such as curves or time or frequency dependent surfaces. Generally, data monitored continuously versus time have a discrete nature, so one of the first tasks in FDA is the application of smoothing methods in order to obtain smooth functions as a basic unit of information [6]. FDA has experienced exponential growth over the last two decades [3]. This development has been driven mainly by its great utility within the new paradigm of data provided by the advancement of sensor technology and the Internet of Things (IoT), within the framework of digitalization of science, and Industry and Services 4.0 [1]. Its use was popularized by the seminal monographs of Ramsay and Silverman [4] and, from a non-parametric approach, Ferraty and Vieu [5]. Now, a great deal of FDA approaches for univariate and multivariate statistical techniques are developed, ranging from descriptive analysis [7], outlier detection [8], statistical quality control [9], analysis of variance [10], classification [11], to regression [12] and time series [13], among others. All these techniques have been successfully applied to solve problems in many scientific domains such as engineering [14], material science [15], geosciences [16], medicine and genetics [17], environmental sciences (dealing many times with the spatio–temporal domain) [18], and of course agriculture [19], among many others. The current popularity of FDA techniques is also due, in part, to the large number of freely available computational tools that allow their application, available in the `R` statistical software, such as the `fda` and `fda.usc` packages [20,21], among others.

In the agriculture domain, Ogden et al. [22] proposed a FDA regression model to predict a lodging score for a rice field from digital images, whereas Yang et al. [23] developed a FDA methodology for variable selection and prediction in ecological applications. On the other hand, Shi and Cao [19] performed a robust functional principal component analysis method for dimension reduction in such frameworks with presence of outliers, as is the case of agricultural and biological datasets. The works of Kantanantha et al. [24] and Lewis-Beck et al. [25] are focused on crop studies. In fact, Kantanantha et al. [24] have proposed a functional regression model from functional principal component analysis to forecast the yield and price in crop decision planning, whereas Lewis-Beck et al. [25] proposed a FDA spatio–temporal model for predicting the time when crops reach their peak each season, from remote sensing data. Among all the crops, corn is crucial for the diet of the Ecuadorian population [26], and, by extension, for the people of whole Andean region and Latin America. In Ecuador, the primary sector is of great importance (in 2010–2014 it contributed 9.60% of the country's wealth, in terms of GDP) [27], with a great variety of crop production. In this context, corn is far the first cereal by production (1742000 ton in 2021, twice the rice production and also ahead of sorghum) [28]. Its productivity or production capacity is largely affected by climatic factors. In relation to climatic factors and their relationship with corn production, the amount of rainfall received during the planting cycle, until the harvest of the crop is of significant importance. In fact, there are certain conditions, regarding the amount of precipitation, that must be met for the correct physiological development of the plant, and in order to benefit crop productivity.

In the context of climate change studies, the estimation of the quantity of precipitation is a task currently performed by applying multivariate regression techniques [29,30]. Alternatively, the FDA techniques have been used to study the rainfall from descriptive and classification point of view [31,32]. We would like also to highlight the use of techniques that take into account the temporal and spatial dependence of the observations, such as kriging, both from the traditional approach [29,31] and also using the new functional approach [33].

In the present work, FDA regression techniques are applied to solve the specific case study of the prediction of mean daily rainfall in Ecuadorian maize regions, in a specific year, from functional data such as temperature and wind speed annual curves. Considering that regression could be defined a statistical procedure whose objective is to model the relation that exists between one or more explanatory covariates with a response variable, from a FDA approach, this relationship can be addressed by fitting Functional Linear Regression (FLR) models with scalar response. Indeed, FLR models with functional bases, FLR models with Functional Principal Components (FPC) basis, FLR model with Functional Partial Least Squares (FPLS) basis, and adaptation of the FLR model using two functional explanatory covariates [4], among others, can be applied to estimate a scalar response, such as the present case. Subsequent application of usual hydrological models to the rainfall predictions will allow us to obtain estimates of the corn irrigation requirements. This is the objective of the present work.

Currently, lack of economic and technological resources of small Ecuadorian farmers (the focus of this study) makes it impossible for them to have access to meteorological information that is crucial for the development of their crops, such as corn. In fact, there are currently no weather stations in each of the country's provinces or cantons, which makes it difficult to find accurate data on meteorological variables such as the amount of precipitation, which is the variable for which predictions are made in this study. In short, many productive areas of Ecuador, defined by the activity of small producers, suffer from a lack of information that this work aims to help amend by implementing new techniques for predicting the amount of rainfall, which will be very useful for decision making. The aim is, therefore, to provide small farmers with tools to improve their production levels.

2. Case study of corn regions in Ecuador

For development of this work, information of the most influential meteorological factors in the agricultural productivity of corn in Ecuador, such as rainfall, temperature, and wind speed has been collected. For this purpose, 25 places strategically located in the

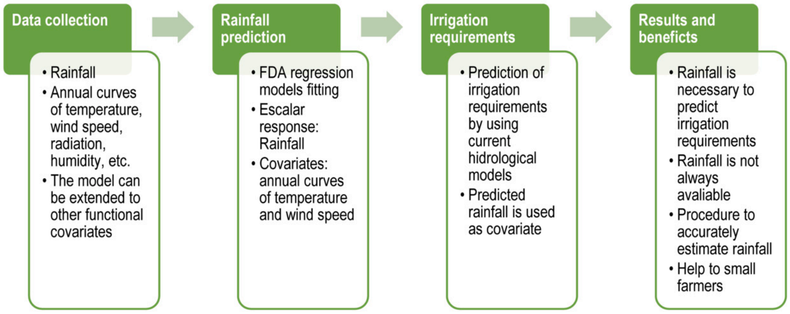


Fig. 1. The procedure to estimate the rainfall and then the irrigation requirements of corn in Ecuadorian regions.

country because of their great production of corn, have been considered. In this context, we have taken daily data between 2010 and 2020 from the NASA web page. After the collection, in the case of rainfall, the means of each place are obtained given a vector of dimension equal to $m = 25$. Regarding the temperature, daily mean is calculated for each place, obtaining a matrix structure of $n \times m = (365 \times 25)$ dimension. In a similar way, daily means for wind speed are obtained. In addition, other variables such as solar radiation and humidity are also included in the dataset. Finally, the FRL models are applied, validated and compared, with the aim to select the best option to predict precipitation, as a scalar response variable, as a function of functional explanatory variables such as temperature and wind speed. In this way, the model can be used to predict precipitation in any place of the country with corn production. The results are useful in a practical way, in order to understand the behavior of corn productivity in Ecuador and its irrigation requirements. All data generated or analyzed during this study are included in this published article (and its supplementary information files).

3. Methodology

This section shows the main parts of the methodology (Fig. 1) proposed to estimate the irrigation requirements of the corn crop in Ecuador.

Firstly, in order to estimate the amount of rainfall during a specific interval of time, in a specific region of Ecuador, the application of functional linear regression models with one or two functional covariates and scalar response is proposed. This estimation is crucial to be able to compare the real rainfall (predicted by functional regression models) with corn plant irrigation requirements. The difference will be provided by irrigation. Thus, a brief introduction to these models is also included.

Once the mean daily rainfall in a specific year is predicted, the second step is to apply traditional hydrological balances to estimate actual evapotranspiration or real water requirements of the plant, ET_c.

3.1. Hydrological balance, relationship with crop productivity

Turk [34] proposed an index to determine the agricultural productivity of crops per hectare with greater accuracy, in which meteorological variables are involved. Thus, in the last century, different traditional mechanisms have been developed to detect the benefits provided by water use in planted crops. In recent years, modified versions of the Turk index have been developed, specifically applied in Andean countries for crops such as corn and soybean, that provide evidences of direct importance of precipitation in agricultural productivity levels [35,36].

Using these indicators, a hydrological balance is made, determined by the loss of moisture in the soil as a result of evaporation and transpiration of water, a process known as evapotranspiration. This balance will make it possible to verify the productive yield of the crop from planting to harvest. In this way, it is possible to estimate whether or not the precipitation in the places where corn is grown is sufficient for the crop to have an adequate physiological development process such that its productivity will benefit.

This balance is developed by estimation of the reference evapotranspiration (ET_o), that can be defined by the following expression proposed by Allen et al. [37], which is more suitable (exact) for the Andean country crops:

$$ET_o = \frac{0.408\Delta(R_s - G) + \gamma \frac{900}{T+273} u(c_s - c_a)}{\Delta + \gamma(1 + 0.34u)}, \tag{1}$$

where ET_o is reference evapotranspiration (mm/d), R_s is solar radiation at the ground surface $(MJm^{-2}d^{-1})$, G is heat flux density at the ground surface $(MJm^{-2}d^{-1})$, T is average temperature $(°C)$, u is average wind speed (m/s), c_s is saturation vapor pressure (kPa), c_a pressure (kPa), Δ slope of the vapor pressure curve $(kPa/°C)$, and γ psychrometric constant $(kPa/°C)$.

ET_o is defined as the evapotranspiration that occurs under optimal conditions to satisfy the plant's water requirement. In addition, the main meteorological factors affecting ET_o are radiation, temperature, atmospheric humidity and wind speed. In addition, the actual evapotranspiration (ET_c) is defined as the evapotranspiration that actually occurs under existing conditions of each crop and is calculated using

$$ET_c = K_c \cdot ET_o, \tag{2}$$

where K_c is the constant of the stage in which the crop process is. In the Initial Stage, $K_c = 0.3$, in the Development Stage, $K_c = 0.8$, in the Middle Stage, $K_c = 1.2$, whereas in the Ripening Stage $K_c = 0.35$.

Based on these indicators, it can be determined whether the water requirements are met in the study areas [38]. Productivity levels are favored or affected by these indicators that are involved in each stage of the plant physiological development process. In Ecuador, planting cycles are usually carried out during rainy seasons, since rainfall is used precisely to obtain an adequate process of the crop, so that these levels increase.

3.2. Functional linear regression models

FLR models will be applied in order to improve the predictions of rainfall quantities obtained by applying multivariate approaches. That is, we search for better explanation of relationship between the explanatory and response variables, when high dimensionality is considered.

It is proposed to use regression models for functional data to estimate the amount of precipitation as a function of meteorological variables collected in areas strategically located in the country, specially due to their agricultural properties and high corn production level. The FLR model with representation in functional bases estimates the parameters sought from a set of basis functions [4], whereas Hastie and Mallows [39] have introduced an estimator function based on the minimization of a cubic spline criterion, and Marx and Eilers [40] use a smoothed B-splines basis, introducing a penalty for the estimator. The FLR model with Functional Principal Component (FPC) basis is also applied. Its objective is to select a subset of the orthonormal basis functions, so that define the bases that are used to estimate the parameters sought [41].

On the other hand, the FLR model based on Functional Partial Least Squares (FPLS), called projection on latent structures, is a technique that combines and generalizes features of Principal Component Analysis and multiple linear regression [42]. Prediction is achieved by extracting a set of orthonormal factors called latent variables with the best predictive ability, computing functional linear regression using partial least squares of the functional data to find the principal components, and then proceeding similarly to the previous model. This criterion maximizes the covariance between the functional variable and the scalar response through the partial least squares components.

The previously mentioned functional regression models consider only one functional explanatory covariate. Therefore, it is proposed to use at least two functional covariates (whose information is commonly used in this type of studies) that allow to obtain better fits, better predictions of rainfall, and thus better values of the irrigation requirements if compared to the previous methods.

The use of FLR models will make it possible to obtain highly accurate predictive results that will be useful for understanding the behavior of agricultural productivity in Ecuador. The obtained results can be divulged in order to implement projects and/or studies to improve the socio-economic conditions of the agricultural sector.

4. Functional data analysis

FDA is a branch of statistics that studies the information contained in curves, i.e. samples of random functions and, in general, any element that is varying over a continuous space. In the practical analysis, we start working with discrete observations of the sample curves from which the functional form is reconstructed. Thus, the properties of the continuous variables are maintained and, therefore, less information is lost. The FDA has substantial advantages, since it has less demanding assumptions compared to common techniques such as time series analysis, which impose, among other assumptions, that the process is stationary, the observations are equally spaced and the process belongs to a specific class.

A variable $\{\mathcal{X}(t)\}_{t\in[0,T]}$ defined on a probability space (Ω, A, P) is said to be a functional variable if it takes values in an infinite dimensional space (functional space), i.e. a complete normed or semi-normed space. An observation χ_i of $\mathcal{X}(t)$ is called a functional data. See the Appendix A for more information about the functional variable definition.

4.1. Smoothing method

Smoothing is the method by which each high-dimensional observation is identified with a function. The objective of this method is to construct the functional data from its discrete observations [21]. The problem is that instead of using continuous observations $\left\{\chi_i\right\}_{i=1}^n$, we will start working with discrete observations $\left\{x_1(t_{i0}), \ldots, x_n(t_{im_i})\right\}$ such that $t_{ik} \in [0,T]$ is the moment with $1 \leq i \leq n$ y $1leqk \leq m_i$. Thus, the sample is considered to be determined by the vectors $x_i = (x_i(t_{i0}), \ldots, x_i(t_{im_i}))'$ such that x_{ik} is the observed value of the sample trajectory at moment t_{ik}.

It is proposed to build the functional form of the sample trajectories $S_n = \left\{\chi_i\right\}_{i=1}^n$ that describe the discrete data behavior from a representation in terms of functional basis. These recovered functions are considered functional data.

A set of functions $\left\{\phi_1, \phi_2, \ldots\right\}$ is a functional basis in $\mathcal{L}^2[0,T]$ if all function $\chi \in \mathcal{L}^2[0,T]$ have a decomposition

$$\chi(t) = \sum_{j=1}^{\infty} c_{ij}\phi_j(t), \tag{3}$$

such that $c_{ij} \in \mathbb{R}\ \forall i = 1, \ldots, n$. In practice, a finite quantity (κ_x) of *base functions* must be chosen, such that if the observed points are taken into account $x_i = \chi(t_{ik})$ para $i = 1, \ldots, n$ y $k = 1, 2, \ldots, m_i$ then the decomposition can be expressed by:

$$x_i = \chi(t_{ik}) = \sum_{j=1}^{\kappa_x} c_{ij}\phi_j(t_{ik}). \tag{4}$$

In Ramsay and Dalzell [6], some types of bases are studied, such as Fourier basis for smoothing periodic data and B-Splines basis [43] for smoothing non-periodic data. In addition, there are other useful functional basis types such as Wavelets [44], exponential, and polynomial basis [4]. See the Appendix A for more information about the description of each type of basis

4.2. Functional depth measures

Depth measures are used to summarize a functional data set. The concept of depth measures how central or deep a piece of data is with respect to a given population. That is, it is a statistical tool that provides information about the internal order of the data, so that they are ranked from the less deep to the deepest. There are many different approaches of functional data depth, namely the functional depth based on the functional median, defined by Fraiman and Muniz [45] and those based on the proximity to the mode or calculated from random projections [46], also including the variant of Tukey random depth [47] (for more information see the Appendix A).

4.3. Functional linear regression models with scalar response

The classical FLR model seeks to estimate the relationship between scalar response variable Y and a unique functional covariate $X_t = \{\mathcal{X}(t)\}_{t\in[0,T]}$ which is a second order stochastic process. Without loss of generality, we consider that $E[Y]=0$ and $E[X_t]=0\,\forall t\in[0,T]$. Therefore, the FLR model with scalar response Y and the functional explanatory covariate X_t is determined by:

$$Y = \alpha + \int_0^T X_t\beta(t)dt + \epsilon. \tag{5}$$

In other words, let $\{y_i\}_{i=1}^n$ be a sample of a random variable from Y and the sample curves $\mathcal{S}_n = \{\chi_i(t)\}_{i=1}^n$ in $\mathcal{L}^2[0,T]$. The *FLR* model is expressed by:

$$y_i = \alpha + \int_0^T \chi_i(t)\beta(t)dt + \epsilon_i, \tag{6}$$

where $\beta(t)\in\mathcal{L}^2[0,T]$ is the functional parameter and $\epsilon_i\sim N(0,\sigma^2)$ is the residual such that $Cov(\epsilon_i,\epsilon_j)=0\,\forall i\neq j$.

In this work, different FLR approaches have been applied, such as FLR model with functional basis, with functional principal components basis, with functional partial least squares basis [48], and with two functional bases (for more information see the Appendix A).

5. Results and discussion

First, the daily means, for each day of the year, corresponding to the temperature and wind speed, in each of the 25 places previously chosen, are calculated. Therefore, for each one of the meteorological variables, we have a matrix structure composed of $n=365$ rows or observations and $m=25$ columns corresponding to the places of corn production. In the case of the rainfall feature, the daily weighted average is obtained for each chosen place, giving a 25–dimensional vector. The values of the original discrete data for each corn producing region (i.e. wind speed, temperature and rainfall amount), can be observed in the map of Fig. 2. The most productive regions in terms of corn are mainly placed in the coast, as shown in Fig. 2. The highest wind speed tend to be in the Pacific coast (center and north), whereas the regions with highest amount of rainfalls tend to be those of the north, inland, close to the Convento Mountains. In addition, the highest temperatures tend to be reached in those areas more protected from the wind, i.e. the Gulf of Guayaquil and those inland areas with relatively low altitude over the sea level.

Fig. 3 shows discrete original data and functional data (after data smoothing) corresponding to the temperature and wind speed, measured in the 25 different locations studied in this work. The smoothing process was performed for the point-wise means of temperature and wind speed, since they are the functional covariates that will be used to predict precipitation. In order to smooth the temperature, a 12 dimensional functional Fourier basis has been chosen by generalized cross-validation. In the same way, a Fourier basis of 11 elements has been fitted to smooth the wind speed. In both cases, Fourier bases are used taking into account the periodic nature of each variable to be smoothed (see Fig. 3).

The next step is to study the central position and dispersion of each variable by obtaining the functional approaches for the mean, median, and variance (Fig. 4). In order to prevent the effect of outliers on means and variance, the trimmed mean has been also calculated, discarding the 15% of the curves with less depth. In order to calculate these statistics, the concept of data depth is used. In this regard, three different approaches have been applied, that proposed by Fraiman and Muniz [45], the Mode approach, and the Random Projections alternative [46]. Depending on the functional data depth approach, the results for the statistics can be slightly different, as pointed out in Naya et al. [49]. There are not many differences in the estimates corresponding to the wind speed variable, i.e. its median, mean and variance do not almost vary regardless of the used data depth approach. Otherwise, when temperature is studied, we can observe more differences in the estimates depending on the data depth approach. The presence of

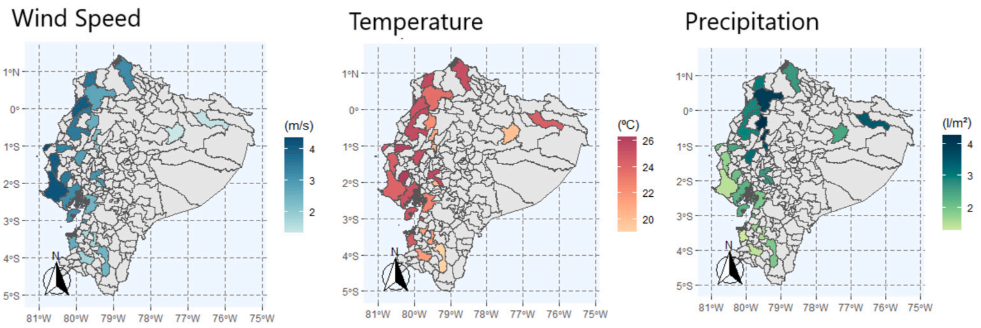
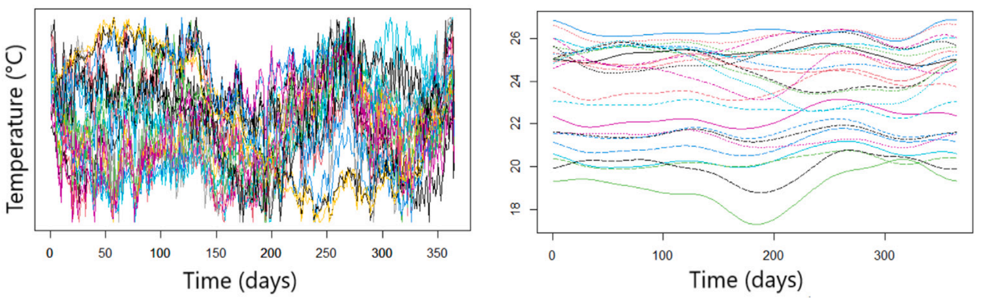
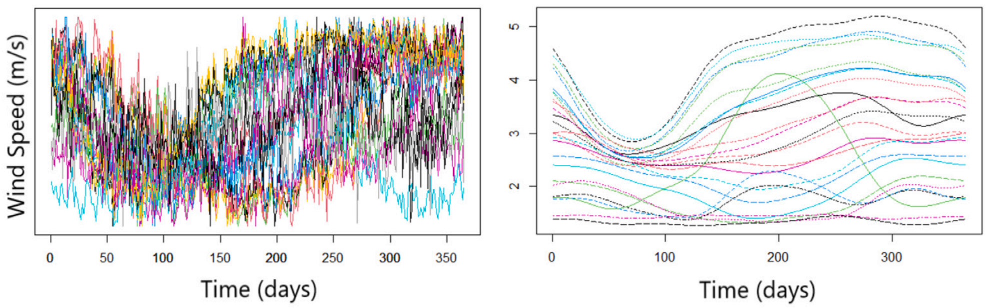


Fig. 2. Mean values of wind speed, temperature, and rainfall for each one of the most productive regions of Ecuador in terms of corn.



(a) Temperature



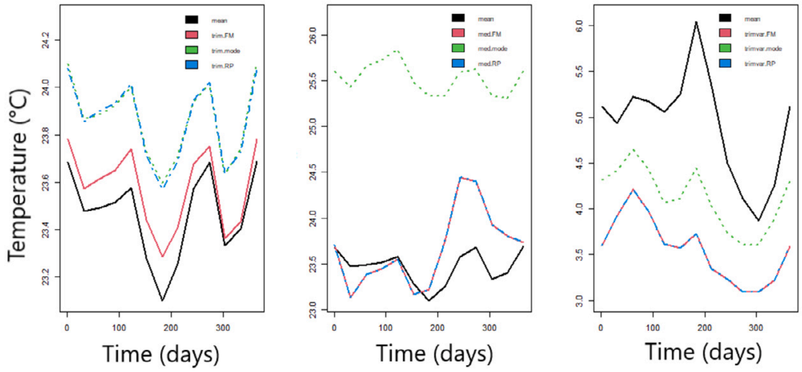
(b) Wind Speed

Fig. 3. Discrete (left) and functional data (right), after smoothing.

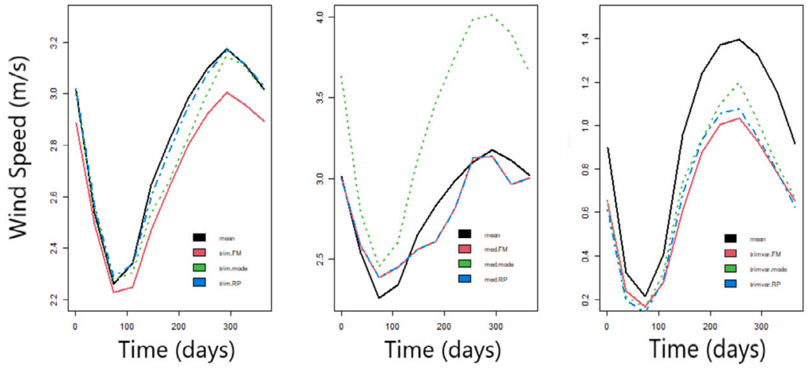
outliers could play some role in this finding. On the other hand, the functional measures of centrality and dispersion provide very interesting information about the weather in corn production areas of Ecuador. Regarding the temperatures, they do not vary much during the year, being in the equatorial region, however a tendency to decrease towards the middle of the year can be observed. The evolution of wind speed is particularly interesting, presenting a clear functional nature, with a minimum in the first months of the year (between March and April) and a clear maximum between September and October. As can be observed, maximum wind speed corresponds to local minima in temperature (see Fig. 4) and vice versa. In terms of variability, we can observe that the highest dispersion is observed in those intervals of times corresponding to the maxima and minima of the wind speed and temperature curves.

Taking into account the information of Fig. 4, the presence of outliers is expected. In order to detect these outliers, prior to the application of the FLR models, a procedure based on Fraiman and Muniz data depth and bootstrap (including also a weighted alternative) is used [49,14]. The results can be observed in Fig. 5, that shows the identification of two outlier curves of temperature and one outlier curve of wind speed. The curves corresponding to the functional temperature of Loja (21) and Loreto (25) were identified as atypical due to their low values. The wind speed curve corresponding to Loja (21) has been also detected by the two approaches (weighted or not) of the outlier detection method.

These areas are located in the extreme east (Loreto), in the Amazon, and in the extreme south (Loja), presenting, consequently, more pronounced climatic differences with respect to the other regions of the country. Taking into account that these curves of temperature and wind speed should be used to estimate the FLR models that provide the predictions of the rainfall scalar variable, the identified outliers have been previously discarded in order to obtain the best possible fit for the regression model and, consequently, more reliable predictions of the rainfall.



(a) Temperature



(b) Wind Speed

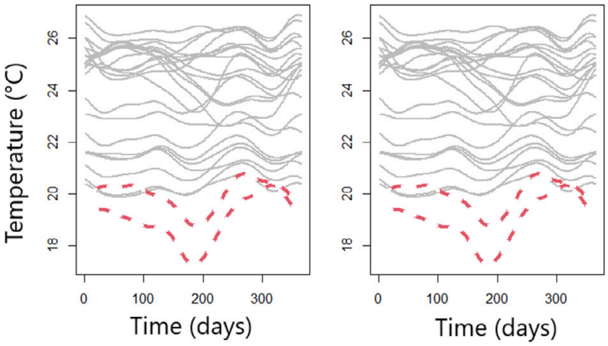
Fig. 4. Descriptive statistics of the temperature and the wind speed based on functional data depth: trimmed mean 15% (left), median (center), trimmed variance 15% (right).

Once the database contains no outliers, the FLR regression models introduced in the previous section are fitted. Before their application, it is important to stress that the overall data have been separated in two sets: a training or calibration sample from which the FLR models are fitted and the test sample, used to evaluate the prediction performance of the models previously fitted from the calibration sample. Table 1 shows which sample (training or test) correspond to each region. The data corresponding to 25 regions are associated to the calibration sample, from which the FLR regression models are fitted. Then, their prediction performance is assessed by applying them to the 6 regions of the test data.

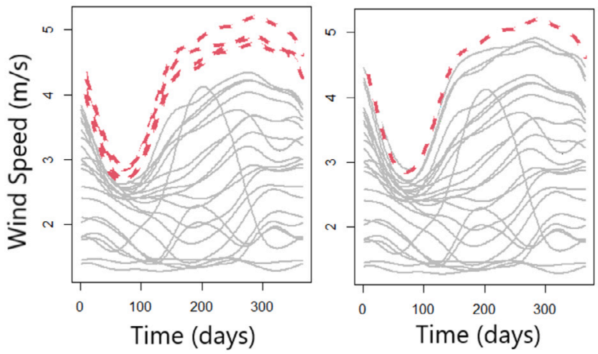
It is important to note that the FLR models are fitted only by the training sample data. Regarding the data availability, we have different climate variables provided by NASA. These are solar radiation, humidity, temperature, and wind speed. Of course, we would improve the model by studying also the role of variables such as vapor pressure or cloud cover, but we are restricted by the data availability. We will work on incorporating new variables from new source of information in order to improve the models and extend them to other climate areas. Moreover, the final model to predict rainfall (scalar variable) as a function of temperature and wind speed curves (functional variables) was chosen taking into account the criteria of increasing the determination coefficient, R^2, or proportion of explained information or variability. Table 2 shows the adjusted R^2 (see the definition in the Appendix B) corresponding to each combination of covariates.

Table 2 shows that the model with best performance, in terms of goodness of fit (adjusted R^2) is the one composed by two functional covariates, the annual curves of temperature (mean) and wind speed (adjusted $R^2 = 0.97$). In fact, its determination coefficient is the same as that corresponding to the model composed of four variables. Thus, for the sake of simplicity, the chosen model is the one that corresponds to just two covariates.

Summarizing, the scalar response is the rainfall, whereas the functional explanatory covariates are the wind speed and the temperature. It is also interesting to know how informative is each variable to explain the rainfall. Table 2 shows different goodness of fit measures (see definitions in the Appendix B) corresponding to each of the possible models composed of only one variable and two variables. The first fitted model is that including just one explanatory variable, the temperature. The second model is also a FLR model with one covariate; in this case the wind speed is used. Subsequently, the model that considers both functional variables, temperature and wind speed, is also fitted. This extension of traditional functional regression models, consisting in the introduction of more than one covariate (specifically the temperature and wind speed), is another contribution of the present work.



(a) Temperature



(b) Wind Speed

Fig. 5. The curves identified as outliers by a bootstrap method based on Fraiman and Muniz data depth are shown. The upper panels correspond to the temperature study, whereas the lower panels show the results for the wind speed. Moreover, the panels on the left show the results of the application of the outlier detection method, whereas, on the right, the results of the application of a weighed bootstrap approach.

Table 3 shows the results of fitting of each one of the FLR models. When just one covariate is introduced, either the temperature or the wind speed, the FLR models using functional basis, FPC basis, and FPLS basis have been fitted, obtaining their measures of goodness of fit (R^2; residual standard error, RSE; mean absolute error, MAE; mean absolute percentage error, MAPE; corrected Akaike information criterion, AIC_c; Bayesian information criterion, BIC) and including the number of significant parameters. In the case of including the two covariates, an extension of the FLR model with functional basis (B-splines) is fitted. In the case of FLR models with a single explanatory covariate, the best one in terms of goodness of fit is the FPLS approach that includes the wind speed as covariate. In fact, it explains 85% of all the variability of the rainfall, with also the lower RSE (0.6). The corresponding FPLS regression model as a function of temperature has also a similar goodness of fit. In any case, in terms of goodness of fit, the most explanatory model is the one estimated as a function of the two functional variables, i.e., temperature and wind speed. In fact, this model, which uses B-spline bases, explains 97% of the variability of the amount of precipitation or rainfall, also presenting the lowest RSE value (0.14). This model meet the assumptions of normality (p-value of Shapiro–Wilk test $= 0.28$), independence (p-value of Durbin–Watson test $= 0.83$) and homoskedasticity of the residuals (p-value of Breusch–Pagan test $= 0.3$), in addition to complying with the good specification or linearity of the model (reset test p-value $= 0.83$).

Considering the comparison of models in terms of goodness of fit, the FLR model with B-spline functional basis is chosen, as a function of the two covariates, temperature and wind speed. Using the estimated model, rainfall predictions are made in 6 cantons (apart from those used to estimate the model) producing dry hard corn. The results of the predictions are the following: the reainfall prediction for Eloy Alfaro canton is 2.32 ml/m^2, for Palenque 2.82, for Valencia 4.07, for Salitre 2.09, for San Vicente 2.15, and for Santo Domingo 4.14. Specifically, temperature and wind speed data are collected these new 6 different places located in the provinces with significant corn productivity in the framework of Ecuador and located besides the sea or close to the coast. From the new collection of the meteorological variables in the same period of time (10 years), these data are smoothed in order we can subsequently fit the FLR model in a proper and reliable way. It can be seen that there are high values of daily precipitation in the cantons of Santo Domingo and Valencia, whereas in Eloy Alfaro, Palenque, Salitre, and San Vicente the levels of daily rainfall are lower.

Table 1
The data corresponding to different corn production regions are associated to the training sample or to the test sample in order to subsequently evaluate the rainfall prediction performance of the models.

Province	Canton	Coordinates		Database sample
		Latitude	Longitude	
Esmeraldas	San Lorenzo	1,29	-78,84	Training
	Quinindé	0,33	-79,48	Training
	Esmeraldas	0,96	-79,65	Training
Manabí	Pedernales	0,08	-80,05	Training
	Chone	-0,70	-80,09	Training
	El Carmen	-0,27	-79,43	Training
	Manta	-0,96	-80,71	Training
	Portoviejo	-1,05	-80,45	Training
	Jipijapa	-1,35	-80,58	Training
	San Vicente	-0,58	-80,40	Test
Los Ríos	Buena Fe	-0,89	-79,49	Training
	Quevedo	-1,03	-79,46	Training
	Babahoyo	-1,80	-79,53	Training
	Valencia	-0.95	-79,35	Test
	Palenque	-1,43	-79,74	Test
Guayas	Balzar	-1,37	-79,88	Training
	Guayaquil	-2,20	-79,89	Training
	Simón Bolívar	-2,02	-79,48	Training
	Pedro Carbo	-1,82	-80,23	Training
	Naranjal	-2,67	-79,62	Training
	Salitre	-1.82	-79,81	Test
Santa Elena	Santa Elena	-2,23	-80,85	Training
El Oro	Pasaje	-3,33	-79,81	Training
	Zaruma	-3,68	-79,62	Training
	Arenillas	-3,55	-80,07	Training
Loja	Loja	-3,99	-79,00	Training
	Paltas	-4,07	-79,63	Training
Sucumbíos	Shushufindi	-0,19	-76,65	Training
Orellana	Loreto	-0,69	-77,31	Training
Santo Domingo	Santo Domingo	-0,25	-79,17	Test

Table 2
Goodness of fit of measurement, in terms of adjusted R^2, corresponding to the different possible FDA regression models (using functional basis) to estimate the rainfall as a function of two or three functional covariates.

Variables	Adjusted R^2
Temperature + Wind speed	0.97
Temperature + Wind speed + Radiation + Humidity	0.97
Radiation + Humidity	0.86
Temperature + Humidity	0.85
Temperature + Radiation	0.87
Wind speed + Radiation	0.92
Wind speed + Humidity	0.92
Temperature + Wind speed + Radiation	0.94
Temperature + Wind speed + Humidity	0.95
Wind speed + Radiation + Humidity	0.96
Radiation + Temperature + Humidity	0.93

Summarizing, in order to tackle the problem of rainfall estimation, from the original data we obtain the following structure:

- A scalar response variable: mean daily rainfall in one year and in one region.
- Two functional covariates: the annual mean temperature and wind speed curves (each consisting of 365 observations).

It is important to note that data have been collected from 2010 to 2021. On the one hand, functional regression models with scalar response have been fitted from data obtained between 2010 and 2020. These models then allow predicting the value of precipitation

Table 3
Goodness of fit measures for the FLR models with scalar response estimates.

Functional Regression Models with Scalar Response									
Functional Covariable	Type	R^2 *adjusted*	RSE	Significant Parameters	MAE	$MAPE$	AIC_C	BIC	$PRESS$
Temperature	Functional Basis	0.67	0.52	4	0.66	22.28	54.69	65.77	15.89
	FPC Basis	0.76	0.44	3	0.60	20.34	54.40	65.48	13.50
	FPLS Basis	0.85	0.21	3	0.60	21.54	53.98	65.05	12.01
Wind Speed	Functional Basis	0.80	0.38	5	1.11	27.50	54.36	63.84	45.57
	FPC Basis	0.65	0.52	3	1.10	27.66	54.19	63.68	44.98
	FPLS Basis	0.85	0.16	4	1.10	28.01	54.12	63.60	44.69
Temperature and Wind Speed	Functional Basis	0.97	0.14	6	0.23	8.31	66.13	79.19	2.08

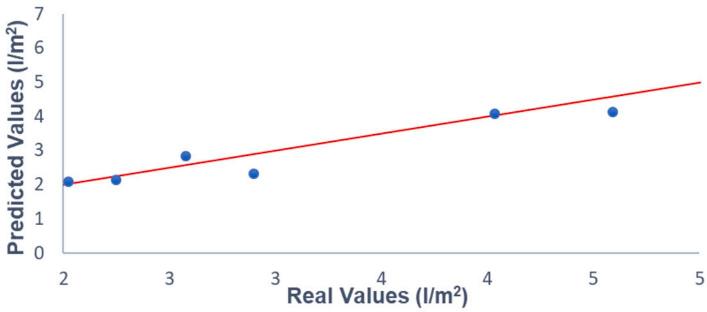


Fig. 6. Plot of observed values of precipitation versus predicted values estimated by the functional regression model, corresponding to the six studied regions.

in the year 2021, for each one of the Ecuadorian regions, defined by their corresponding temperature and wind speed curves for the years between 2010 and 2020. Then, this make it possible to anticipate irrigation needs by using the usual hydrological models.

In order to assess the prediction ability of the model, Fig. 6 shows the plot of observed values versus predicted corresponding to the six studied regions. As shown in Fig. 6, all the points are close to bisector, thus, the model is able to make good estimates (close to the real values) of the rainfall. In addition, some of the most used measurements to assess the prediction capability of regression models have been calculated, such as $R^2 = 0.884$, $MAE = 0.246$, $PRESS = 0.611$, and $MAPE = 8\%$. These results support the fact that predictions are very close to the real values.

Rainfall prediction is a task for which multivariate regression methods are usually used, as is the case of the work of Chen et al. [30] and Martínez-Cob [29]. Some of the difficulties to be faced when using the multivariate perspective are the selection of variables and the verification of starting hypotheses: distribution of the response variable, independence of the observations, non-multicolinearity, homoskedasticity, among others. Martínez-Cob [29] approached the problem of precipitation estimation from a multivariate perspective, applying three geostatistical interpolation methods of ordinary kriging (KG), cokriging (CK) and modified residual kriging (MRK). These methods are used to estimate the long-term average total Annual PREcipitation (APRE) in a mountainous region from meteorological covariates. In these works, the hypothesis of stationarity is assumed, although it is probably not fulfilled for the whole region, but it is met locally. If the present functional proposal is used, the assumption of such a starting hypothesis is not required. The results of adjusted R^2 obtained by KG, CK, modified MRK for precipitation estimation are 0.89, 0.91 and 0.88, respectively, which are slightly lower than that obtained by FLR (0.97). In addition, the MAE values corresponding to the application of the multivariate KG, CK, MRK methods are higher than that corresponding to the FRM model. On the other hand, Chen et al. [30] proposed another alternative estimation of daily precipitation using SVM classification and regression models, on the one hand, and in classification analysis and multivariate linear regression, on the other hand, as a function of predictors related to humidity and vorticity. The first step is the application of classification models to determine whether the day is dry or wet, while the second is the estimation of regression models to estimate the amount of precipitation conditional on the occurrence of a wet day. The use of SVMs tends to provide predictions closer to reality (although no goodness-of-fit indices are provided), performing well for precipitation levels less than 100 mm. The use of FLR models prevent the starting assumptions such as the requirement of no multicoliniarity of the predictors. In addition, the resulting models are simpler, in terms of the number of predictors, as they take advantage in a more complete way of the information of variables that are actually functional in nature, as is the case of temperature and wind speed.

Once we have estimated the model that allows us to predict the amount of rainfall accurately and precisely, the next step of the methodology is to include this prediction in the standard calculations of the ideal irrigation requirement of the corn crops in Ecuador.

The predicted value of the amount of rainfall in a region will now be used to perform the water balance in the study areas, which, in turn, will provide an estimate of the irrigation requirements during the planting cycle, until the maize harvest. This balance will be carried out as indicated in section 3.1. In this way, it will be possible to estimate the production capacity of each region, taking into account their irrigation needs, and also taking into account the water resources available to them.

Table 4
Water balance corresponding to Santo Domingo and Valencia regions, obtained from the rainfall predictions.

Months	Santo Domingo			Valencia		
	ET_o mm/d	P_t mm	P_e mm	ET_o mm/d	P_t mm	P_e mm
January	3.58	128.3	78.6	4.22	126.2	100.7
February	3.48	115.9	68.7	3.78	114	93.2
March	3.4	128.3	78.6	3.73	126.2	100.7
April	3.33	124.2	75.4	3.53	122.1	98.2
May	3.19	128.3	78.6	3.69	126.2	100.7
June	3.2	124.2	75.4	3.43	122.1	98.2
July	3.6	37.1	12.3	4.12	26.4	25.3
August	4.45	19.4	1.6	4.63	15.2	14.8
September	5	29.1	7.5	5.27	18.9	18.3
October	4.89	59.2	25.5	4.91	40.3	37.7
November	4.67	64	28.4	4.84	51	46.8
December	4.31	128.3	78.7	4.67	126.2	100.7
Mean	3.93	90.5	50.8	4.24	84.6	69.6

The results of Tables 4 and 5 correspond to the year 2021. They are obtained from the prediction of the mean daily rainfall, which has been obtained from the application of the functional regression model to the annual curves of temperature and wind speed between 2010 and 2020.

To illustrate the methodology for estimating irrigation needs for corn crops in Ecuador, the specific cases of the regions of Santo Domingo and Valencia will be considered (Table 4).

Taking into account the data corresponding to these two regions, the first step calculation of the reference evapotranspiration, ET_o, as indicated in section 3.1. In parallel, the rainfall amount predicted by the present methodology will be taken to find the effective precipitation, defined as the fraction of precipitation that actually reaches the plant. The effective precipitation can be calculated using several criteria, among which we have selected the reliable precipitation criterion, one of those available in the software `CropWat` [50]:

$$P_e = 0.6P_t - 10 \text{ for } P_t < 70\ mm, \text{ and } P_e = 0.8P_t - 24 \text{ for } P_t > 70\ mm,$$

where P_e is the effective precipitation and P_t is the monthly total precipitation.

Only the months in which corn is sown in Ecuador are considered. It is important to note that corn is a transitory crop. The calculations of evotranspiration are performed by using the `Cropwat` and are shown, jointly with the effective and total precipitation, in Table 5.

The next step to estimate corn irrigation requirement is calculation of real evapotranspiration ET_c. It is the real precipitation value that the crop needs and depends on the crop phase.

Therefore, the irrigation requirements should be determined in each stage of the crop, that is initial, development, middle and final, taking into account that each stage depends on a constant K_c that is strictly defined by the type of product, in this case corn. In general, corn begins its cycle in mid-December and ends in mid-April.

Table 5 shows the results grouped by periods of 10 days for each month corresponding to corn cultivation. Specifically, it provides the ET_c values calculated according to the section 3.1, per day and per period. As previously mentioned, these values are representative of the actual amount of precipitation required by the crop at each stage. The irrigation requirements (IR) will be $IR = ET_c - P_e$, when $ET_c - P_e > 0$, and 0 when $ET_c - P_e \le 0$. This expression is used to calculate the water requirements per day (Table 5).

Table 5 shows that, in Santo Domingo, from the 3rd period of January, in the development stage, until the 2nd period corresponding to March that corresponds to the middle stage, the predicted precipitation of $4.14\ mm$ does not satisfy the real rainfall requirement necessary for the plant to carry out its physiological processes properly, i.e., productivity levels are negatively affected during all that time. Therefore, additional water requirements are necessary in these stages. For instance, in February, between the days $11-20$, that correspond to the second period of the middle stage, an irrigation system should supply $1.8\ mm$ of water per day so that the plant has good physiological development, maintaining or even increasing the levels of productivity given in this area. In Valencia region, the irrigation with $0.6\ mm$ of water per day is necessary during the third period of January, in the development stage. Thus, at least $1\ mm$ of water is needed daily in the middle and final stages.

The Santo Domingo corn-producing canton had an annual production of $1974\ tons$ during 2020, the lowest of all the cantons of Ecuador. On the other hand, the province of Los Ríos, where Valencia is located, has a corn production of $642761\ tons$, being the province with the highest corn production [51]. The implementation of procedures in order to estimate in a proper way the amount of precipitation in each period and in each region, and subsequently to calculate the corresponding irrigation requirements could help to better administrate the hydric resources, improving the health of plants, and increasing the production in regions such as Santo Domingo, Valencia and other places of Ecuador and Andean area, as is the aim of the present study. In this way, reliable precipitation values can be obtained that allow to find the water needs that the corn crop or, in general, of any other crop needs for its correct development and subsequent commercialization.

Table 5
Irrigation requirement for corn productivity, obtained from the rainfall predictions.

Irrigation Requirement							
Month	Period $p = 10$ *days*	Stage	Kc	ET_c mm/d	ET_c mm/p	Effec.Prec. Efec mm/p	Irr.Req mm/d
Santo Domingo							
Dec	2	Inic	0.3	1.4	8.4	21.6	0
Dec	3	Inic	0.3	1.36	14.9	35.2	0
Jan	1	Des	0.37	1.61	16.1	33.8	0
Jan	2	Des	0.6	2.55	25.5	33.9	0
Jan	3	Des	0.86	3.5	38.5	32.9	0.8
Feb	1	Med	1.1	4.31	43.1	31.4	1.5
Feb	2	Med	1.15	4.34	43.4	30.4	1.8
Feb	3	Med	1.15	4.32	34.6	31.4	0.8
Mar	1	Med	1.15	4.31	43.1	33	1.4
Mar	2	Fin	1.15	4.28	42.8	34	1.2
Mar	3	Fin	0.96	3.53	38.8	33.6	0.9
Apr	1	Fin	0.68	2.46	24.6	32.9	0
Apr	2	Fin	0.44	1.56	12.5	26	0
Valencia							
Dec	2	Inic	0.3	1.4	8.4	21.6	0
Dec	3	Inic	0.3	1.36	14.9	35.2	0
Jan	1	Des	0.37	1.61	16.1	33.8	0
Jan	2	Des	0.6	2.55	25.5	33.9	0
Jan	3	Des	0.86	3.5	38.5	32.9	0.6
Feb	1	Med	1.1	4.31	43.1	31.4	1.2
Feb	2	Med	1.15	4.34	43.4	30.4	1.3
Feb	3	Med	1.15	4.32	34.6	31.4	0.3
Mar	1	Med	1.15	4.31	43.1	33	1.0
Mar	2	Fin	1.15	4.28	42.8	34	0.9
Mar	3	Fin	0.96	3.53	38.8	33.6	0.5
Apr	1	Fin	0.68	2.46	24.6	32.9	0
Apr	2	Fin	0.44	1.56	12.5	26	0

Once tested the accuracy of this methodology to estimate the rainfall (and then, by applying hydrology models, the irrigation requirements) in the Ecuadorian maize regions, the next step is to estimate the scope or limitations of the present approach. This is done taking into account two perspectives, one regarding the data availability and the other taking into account the geography. Thus, on the one hand, the proposed regression model with scalar response and two functional covariates is applicable if yearly curves of temperature and wind speed are available. Without these variables of a functional nature, it is not possible to obtain estimates of the amount of rainfall and, therefore, of the subsequent irrigation needs. On the other hand, the geography (and the corresponding differences in weather) could play a significant role in the results of the proposed model. In order to check the geographical scope of the proposal, the rainfall of several new regions placed in Peru (also in de Andean region) are now estimated by using the model trained with the Ecuadorian counterparts. Thus, the performance of the model has been tested in the following two sets of regions:

- Set 1: Areas defined by similar geographical features to those corresponding to the studied Ecuadorian regions. The set is composed by the regions of Pucallpa, Tarapoto, Picota, and Lamas, between the Andes and the Amazonian area.
- Set 2: Areas defined by different geographical and weather features. Specifically, two regions of the Pacific coast specially defined by a desert and semi-desert climate, with very scarce precipitation (more arid weather than the studied Ecuadorian regions). The studied areas are Piura and Ferreñafe, on the Peru northwest Pacific coast.

Table 6 shows that the rainfall predictions for the set 1 are relatively close to the real values (the relative error is between 9% and 22%, depending on the region). Otherwise, the regression model trained with data of the Ecuadorian regions is not able to predict accurately the rainfall of the more arid regions, namely Piura and Ferreñafe. Therefore, we can infer that the results of the model will be better the closer geographically and climatically the regions to be predicted are to those from which the model is constructed. In the present case, the high dependence of the amount of rainfall on the wind speed and temperature curves has been observed in the Ecuadorian maize regions. We also observed that the estimated model can provide useful estimates of the rainfall in regions with relatively similar geography and climate. It is assumed that the greater the climatic and geographic difference, the worse the estimates will be. It remains for further studies to analyze whether in other regions of the Americas the relationship between rainfall with respect to annual temperature and wind speed curves exists in similar magnitude as in Ecuador and, therefore, to evaluate the possibility to obtain a new expression of the model corresponding to those areas. In any case, as a general conclusion of the work, the potential usefulness of the functional regression models fitting in the hydrology domain is observed.

Table 6
Rainfall real values and model predictions corresponding to six Peruvian regions. Set 1 includes those areas with similar geographical and climate features than those corresponding to the Ecuadorian regions, whereas set 2 shows two regions defined by a more arid weather.

Set	Area	Coordinates		Rainfall prediction	Real rainfall
		Latitude	Longitude		
Set 1	Pucallpa	-8.38	-74.55	2.61	2.87
	Tarapoto	-6.48	-76.36	2.42	3.11
	Picota	-6.92	-76.33	2.24	2.56
	Lamas	-6.42	-76.53	2.47	3.12
Set 2	Ferreñafe	-6.64	-79.78	3.68	0.27
	Piura	-5.20	-80.63	3.57	0.26

6. Conclusions

This work has proposed a new alternative based on the analysis of functional data for the estimation of rainfall and, as a result, of the water balance and irrigation needs in corn growing areas in Ecuador, a basic food for feeding the population of the entire Andean mountain range. Specifically, several regression models have been proposed for scalar response, precipitation, and functional predictors, wind speed and temperature (proposed by specialists in agronomy and forestry engineering), which have been applied and validated through a real case study of corn cultivation in different regions of Ecuador.

The fitted scalar response functional regression models have shown good performance in terms of goodness of fit, providing reliable, accurate and precise rainfall predictions. Given the functional nature of the explanatory variables, the use of these techniques in agriculture should be considered as an alternative approach to the usual methods of multivariate statistics, such as multivariate regression models or outlier detection methods.

Regarding the choice of the model, it is concluded that the best FLR model is the one that considers two functional explanatory covariates to obtain the prediction of precipitation, also fulfilling all the starting assumptions, i.e., normality, independence and homoscedasticity of the residuals, in addition to the assumption of specification or linearity of the model.

On the other hand, in terms of goodness of fit, the best FLR model as a function of functional covariates explains 97% of the variability of the data ($R^2 = 0.97$), being also defined by 6 significant parameters and an error of 0.14.

Concerning the limitations of the functional regression model, it has been also applied to North Peru regions close to Ecuadorian areas where the model has been estimated, obtaining rainfall prediction errors between 9% and 22%. The magnitude of the error is related to the climate differences with respect to the Ecuadorian regions. Thus, the geographical limitations of the model could be those Andean regions with similar climate.

Regarding the practical information obtained from the case study in the agricultural field, it was possible to calculate the water balance and irrigation requirements during the physiological development cycle of hard corn in the corn producing cantons of Ecuador based on the estimation of precipitation through FDA techniques, including descriptive statistics, outlier detection methods, and the fitting of FLR models. This methodology intends to help to assign properly the water resources in corn crops in Ecuador.

CRediT authorship contribution statement

Conceptualization, Miguel Flores, Danilo Loza, Ángel Llambo; **methodology**, Miguel Flores and Danilo Loza, Ángel Llambo; **software**, Miguel Flores, Danilo Loza, Ángel Llambo; **validation**, Miguel Flores, Salvador Naya, Javier Tarrío-Saavedra; **formal analysis**, Danilo Loza, Ángel Llambo, Miguel Flores; **investigation**, Danilo Loza, Ángel Llambo, Miguel Flores, Salvador Naya, Javier Tarrío-Saavedra; **resources**, Miguel Flores, Salvador Naya; **data curation**, Miguel Flores, Danilo Loza, Ángel Llambo; **writing- original draft preparation**, Danilo Loza, Ángel Llambo, Miguel Flores; **writing-review and editing**, Javier Tarrío-Saavedra, Salvador Naya; **visualization**, Miguel Flores, Danilo Loza, Ángel Llambo, Javier Tarrío-Saavedra; **supervision**, Miguel Flores, Salvador Naya, Javier Tarrío-Saavedra; **project administration**, Miguel Flores, Salvador Naya; **funding acquisition**, Salvador Naya, Javier Tarrío-Saavedra. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data included in the supplementary material.

Additional information

Supplementary content is published online at https://power.larc.nasa.gov/data-access-viewer/.

Acknowledgements

This research has been supported by Ministerio de Ciencia e Innovación grant PID 2020-113578RB-100 and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020-14 and Centro de Investigación del Sistema universitario de Galicia ED431G2019/01), all of them through the ERDF. The authors would like to thank its support to the Escuela Politécnica Nacional, specifically to "Entidad Operativa Desconcentrada Unidad de Gestión de la Investigación y Proyección Social-EOD-UGIPS" belonging to "VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y VINCULACIÓN DIRECCIÓN DE INVESTIGACIÓN". Moreover, we would like to strongly thank all the guidance and comments of the two reviewers and the editor. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Functional data analysis definitions

Functional Random Variable. A variable $\{\mathcal{X}(t)\}_{t\in[0,T]}$ defined on a probability space (Ω, A, P) is said to be a functional variable if it takes values in an infinite dimensional space (functional space), i.e. a complete normed or semi-normed space. An observation χ_i of $\mathcal{X}(t)$ is called a functional data.

Let Ω be a measurable compact subset of $\mathbb{R}$ with positive Lebesgue measure. We define the Hilbert space $\mathcal{H} = \{\mathcal{L}^2[T] : T = [0,T] \in \Omega\}$ which is the space of square integrable functions over the real interval T and is determined by $\mathcal{L}^2[T] = \{f : T \to \mathbb{R} : \int_T f^2(t)dt < \infty\}$, whose usual scalar product is

$$\langle f, g\rangle = \int_T f(t)g(t)dt \quad \forall f, g \in \mathcal{L}^2(T). \tag{A.1}$$

This product is the equivalent of the inner product of vectors in $\mathbb{R}^n$. Consequently, the FDA explores data that are functions belonging to the space $\mathcal{L}^2[T]$ with the aim of extending multivariate data analysis methodologies to functional data. Thus, it is possible to extract concepts such as:

- Sample mean: $\bar{\mathcal{X}}(t) = \frac{1}{n}\sum_{i=1}^n \chi_i(t)$.
- Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^n [\chi_i(t) - \bar{\mathcal{X}}(t)]^2$.

Fourier basis For this case study, Fourier bases were used in order to smooth the discrete data, so that the building of the functional data $\chi_i(t)$ is performed from discrete data [4]. This approximation is achieved according to the following expression:

$$\hat{\chi}(t) \approx \frac{a_0}{2} + \sum_{i=1}^{N}(a_i cos\frac{2\pi i t}{N} + b_i sen\frac{2\pi i t}{N}),$$

where a_0, a_i y b_i $(i = 1, 2, \dots, N)$ are constants.

This type of basis is generally used when the trajectories to be estimated are regular and have a periodic nature.

B-splines basis A spline is a function that is constructed by sections, i.e. a spline is formed by pieces of polynomials that are connected to each other by a node [43]. The number of sections is usually denoted as L. In order to determine a spline function one must have a number of values greater than $L + g + 1$ where g is the degree of the polynomial used in each interval. The number $g + 1$ is called the order of the spline and the number $L + g + 1$ is called the degrees of freedom of the spline family.

A function $\phi : T \to \mathbb{R}$ is a spline function if a partition $\{\tau_l\}_{l=1,\dots,L}$ of T exists, in addition to polynomial functions $\{\beta_k\}_{k=1,\dots,L}$ with each $\beta_k : \tau_k \to \mathbb{R}$ of degree m such that

1. $\phi_{\tau_k} = \beta_k$ and
2. for every pair of adjacent polynomials β_j, β_i, its derivatives coincide at the point $t \in T$ such that $\phi(t) = \beta_j(t) = \beta_i(t)$, for $q = 1, \dots, m-2$.

The set of splines functions make up a vector space whose dimension is given by the degrees of freedom. Therefore, various bases can be used to generate the mentioned space. Thus, any spline function is represented $\beta(t)$ as follows:

$$\beta(t) = \sum_{i=0}^{N+k} c_i B_{i,k}(t), \quad t \in [t_0, t_{N+1}],$$

where c_i are called control points or Boor points. For a B-spline of degree k with N interior nodes, there are $M = N + k + 1$ control points.

Fraiman-Muniz Depth (FMD). Let $F_{n,t}$ be the empirical distribution of the sample i.i.d $\mathcal{S}_n = \{\chi_i\}_{i=1}^n$ of a functional random variable X_t. For each $t \in [0,T]$ the univariate depth $Z_i(t)$ of the data $\{\chi_i(t)\}$ is given by:

$$Z_i(t) = 1 - \left|\frac{1}{2} - F_{n,t}(\chi_i(t))\right|.$$

Then, the *FM* depth is defined as the average of the univariate depth along the domain of the functional data:

$$FMD(\chi_i) = \int_0^1 Z_i(t)dt \ \ \forall i = 1, \dots, n.$$

The FMD will match the deepest data, i.e. the function is ordered $FMD(\chi_i(t)$ for which $FMD(\chi_i)$ is maximal [45].

Modal Depth (MD). In Cuevas et al. [46], a depth measurement is based on how surrounded the curves are with respect to a distance (metric or semi-metric), selecting the path most densely surrounded by other paths of the process. Let the sample i.i.d $\mathcal{S}_n = \{\chi_i\}_{i=1}^n$ of a functional random variable. The kernel is defined by:

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}.$$

The h-depth over a neighborhood z is determined by:

$$MD_h(z) = n^{-1} \sum_{i=1}^{n} K_h\left(\|z - x_i\|\right), \tag{A.2}$$

where $K_h(t) = h^{-1}K(\frac{t}{h})$ is the rescaled kernel and h is a fitting parameter.

Random Projections Depth (RPD). In Cuevas et al. [46], the depth measurement calculated through Random Projections (RP) are also considered. Let the sample i.i.d $\mathcal{S}_n = \{\chi_i\}_{i=1}^n$ of a functional random variable. Let $h \in \mathcal{H}$ a direction of the independent direction process $\mathcal{H}$. The projection $\{\chi_i\}$ along the direction h is given by:

$$P_i^h = \langle h, \chi_i \rangle = \int_0^T h(t)\chi_i(t)dt.$$

The RPD is defined as:

$$RPD(\chi_i, h) = \mathcal{D}(P_i^h),$$

where $\mathcal{D}$ is the univariate depth measure. If we consider a collection of random projections $\{h_j\}_{j=1}^m$, the depth is obtained using all the projections

$$RPD(\chi_i, \{h_j\}_{j=1}^m) = \frac{1}{m} \sum_{j=1}^{m} \mathcal{D}(P_i^{h_j}).$$

Tukey Random Depth (RTD). On the other hand, in Cuesta-Albertos and Nieto-Reyes [47], a variant of the depth of random projections is defined as:

$$RTD(\chi_i, \{h_j\}_{j=1}^m) = \min \mathcal{D}(P_i^{h_j}). \tag{A.3}$$

FLR Model with Functional Basis. It is proposed to estimate the function $\beta(t)$ using functional bases. This procedure considers a functional basis large enough to express the estimated function $\hat{\beta}$ of the regression (6). To do this, β is approximated using a finite number of functional bases of the B-splines type. That is, the functional basis $\{\theta_j\}_{j=1}^{\kappa_B}$ is considered such that:

$$\hat{\beta}(t) = \sum_{j=1}^{\kappa_B} b_j \theta_j(t) = \mathbf{b}^T \boldsymbol{\theta}. \tag{A.4}$$

Now, we consider the smoothing of the functional data $\chi_i(t)$ through the representation of bases given by:

$$\chi_i(t) = \sum_{j=1}^{\kappa_x} c_{ij} \phi_j(t) = \mathbf{c}_i^T \boldsymbol{\phi}. \tag{A.5}$$

Replacing these matrix representations in the model (6) you get that

$$\begin{aligned} y_i &= \alpha + \int_0^T [\mathbf{c}_i^T \boldsymbol{\phi}][\theta^T(t)\mathbf{b}]dt + \epsilon_i \\ &= \alpha + \mathbf{c}_i^T \left[\int_0^T \boldsymbol{\phi}(t)\boldsymbol{\theta}^T(t)dt \right] \mathbf{b} + \epsilon_i \\ &= \alpha + \mathbf{c}_i^T [\boldsymbol{J}_{\boldsymbol{\phi}\boldsymbol{\theta}}]\mathbf{b} + \epsilon_i, \end{aligned} \tag{A.6}$$

such that $J_{\phi\theta}$ is the matrix of internal products given by $\boldsymbol{\phi}(t)$ and $\theta(t)$ of size $(\kappa_x \times \kappa_B)$. Thus, we consider the matrix $Z_i = [\mathbf{c}_i^T J_{\phi\theta}]$ and the vector $\mathbf{b} = (b_1, \ldots, b_{\kappa_B})$ in such a way that the problem is reduced in the following way:

$$y_i = \alpha + Z_i\mathbf{b} + \epsilon_i. \tag{A.7}$$

This regression model allows estimating α and the coefficients $\mathbf{b}$ of $\beta(t)$ by least squares:

$$\hat{\mathbf{b}} = (Z^T Z)^{-1} Z^T Y.$$

Therefore, the following problem arises:

$$y_i = c_0 + \int_0^T [\chi_i(t)][\beta(t)]dt + \epsilon_i. \tag{A.8}$$

The functional parameters $\beta(t)$ are estimated using a B-spline basis.

FLR Model with Functional Principal Components.

This method allows associating a multivariate regression model with a functional regression model by using the scores of the FPC components. From the functional observations $S_n = \left\{\chi_i\right\}_{i=1}^n$ and the covariance function $Cov(s,t)$ the first functional principal components PC are used, $\left\{\xi_j(t)\right\}_{j=1}^p$, so that we have the decomposition given by:

$$\chi_i(t) = \sum_{j=1}^{p} \xi_j(t) f_{ij}. \tag{A.9}$$

The model associated with the functional regression (6) is expressed by

$$y_i = b_0 + \sum_{j=1}^{p} b_j(t) f_{ij} + \epsilon_i. \tag{A.10}$$

This model is a standard multivariate regression model, that is, solved by least squares. Thus, the estimate of $\mathbf{b} = \left(b_1, \ldots b_p\right)$ is obtained by:

$$\hat{\mathbf{b}} = (F^T F)^{-1} F^T Y, \tag{A.11}$$

where $F = \left\{f_{ij}\right\}_{nxp}$. Since the components *FPC* form an orthogonal system, it is taken

$$\beta_{FPC}(t) = \sum_{j=1}^{p} b_j \xi_j(t),$$

so that you can get to the model (6) through

$$y_i = b_0 + \int_0^T [\chi_i(t)][\sum_{j=1}^{p} b_j \xi_j(t)]dt + \epsilon_i. \tag{A.12}$$

After performing the functional regression using FPC components, the estimated functional regression function is recovered, determined by

$$\hat{\beta}_{FPC}(t) = \sum_{j=1}^{p} \hat{b}_j \xi_j(t).$$

Therefore, the following problem arises:

$$y_i = c_0 + \int_0^T [\chi_i(t)][\beta_{FPC}(t)]dt + \epsilon_i. \tag{A.13}$$

In this case, the $\beta_{FPC}(t)$ parameters are estimated as functional principal components.

FLR Model with Functional Partial Least Squares Basis.

This process is based on decomposing the response variable and the functional covariate in terms of random variables, so that the prediction is maximized. The use of the least squares criterion to estimate the model (6) results in a poorly posed problem because the covariance operator, in general, is not invertible [48]. In fact, the estimation of the regression coefficient function $\beta(t)$ under the least squares criterion results in the Wiener Hopf integral equation:

$$E(Y \cdot X_t) = \int_0^T E[X_t \cdot X_s]\beta(s)ds. \tag{A.14}$$

The PLS approach is an efficient solution to the inverse problem presented by the above equation [42]. From the functional observations $\mathcal{S}_n = \{\chi_i\}_{i=1}^{n}$, the FPLS component scores, $\{n_h\}_{h=1}^{p}$, are used as shown below:

$$\begin{aligned} \chi_i &= \sum_{h=1}^{p} p_h(t)\eta_h \\ y_i &= \sum_{h=1}^{p} c_h(t)\eta_h \end{aligned} \tag{A.15}$$

the way to get the FPLS components is recursive, then

$$\begin{aligned} p_h(t) &= \frac{Cov(\chi_{h-1}, \eta_h)}{Var(\eta_h)} = \frac{E[\chi_{h-1}\eta_h]}{E[\eta_h^2]} \\ c_h(t) &= \frac{Cov(y_{h-1}(t)\eta_h)}{Var(\eta_h)} = \frac{E[y_{h-1}\eta_h]}{E[\eta_h^2]} \end{aligned} \tag{A.16}$$

The model associated with the functional regression (6) is expressed by:

$$y_i = c_o + \sum_{h=1}^{p} c_h\langle \chi_i(t), \kappa_h(t)\rangle + \epsilon_i, \tag{A.17}$$

where

$$\kappa_h(t) = \frac{Cov(\chi_{h-1}(t), y_{h-1})}{\left\|Cov(\chi_{h-1}(t), y_{h-1})\right\|}. \tag{A.18}$$

From here, η_h can be rewritten as

$$\eta_h = \langle \chi_{h-1}, \kappa_h\rangle = \langle \chi_i, \phi_h\rangle,$$

such that

$$\phi_h = \kappa_h - \langle p_1, \kappa_h\rangle\phi_h - \ldots - \langle p_{h-1}, \kappa_h\rangle\phi_{h-1},$$

with $\phi_1 = \kappa_1$. So that

$$y_i = c_0 + \langle \chi_i(t), \sum_{h=1}^{p} c_h\phi_h(t)\rangle + \epsilon_i. \tag{A.19}$$

This model is a standard multivariate regression model that is solved for least squares. In this way, the estimate of $\mathbf{b} = \left(c_1, \ldots c_p\right)$ is given by:

$$\hat{\mathbf{b}} = (S^T S)^{-1} S^T Y, \tag{A.20}$$

where $S = \left\{\phi_h\right\}_{nxp}$. Therefore, it is considered that

$$\beta_{FPLS}(t) = \sum_{h=1}^{p} c_h\phi_h(t),$$

so that the model (6) can be reached by:

$$y_i = c_0 + \int_0^T [\chi_i(t)][\sum_{h=1}^{p} c_h\phi_h(t)]dt + \epsilon_i. \tag{A.21}$$

After performing the functional regression using FPLS components, the estimated function can be retrieved, given by:

$$\hat{\beta}_{FPLS}(t) = \sum_{h=1}^{p} c_h\phi_h(t). \tag{A.22}$$

Therefore, the following problem is presented:

$$y_i = c_0 + \int_0^T [\chi_i(t)][\beta_{FPLS}(t)]dt + \epsilon_i. \tag{A.23}$$

In this case the model estimates the parameters $\beta_{FPLS}(t)$ as partial least squares components. For the development of the models (A.8), (A.13) and (A.23) is considered as the sample of the scalar variable $\left\{y_i\right\}_{i=1}^{n}$ as the sample of the Y scalar variable, represented by the precipitation and the sample $S_n = \{\chi_i\}_{i=1}^n$ corresponds to the temperature as a functional explanatory covariate. Similarly, the same process will be performed using the wind speed as the functional explanatory covariate.

FLR Model with two Functional Bases. The FLR model is generalized to estimate the relationship between a scalar response variable Y and a finite number q of functional covariates $X_t^j = \{\mathcal{X}^j(t)\}_{j=1}^q$. Without loss of generality, we consider that $E[Y] = 0$ and $E[X_t^j] = 0\ \forall t \in [0, T]$. In the case of a non-centered process should be used $\left\{X_t^j - \mu(t)\right\}$. Thus, the FLR model with scalar response Y and a finite number q of functional explanatory covariates X_t^j is determined by:

$$Y = \alpha + \sum_{j=1}^{q} \int_0^T X_t^j \beta^j(t)dt + \epsilon. \tag{A.24}$$

We consider $\{y_i\}_{i=1}^n$ as sample of the scalar response variable Y and a set of sample curves $\{S_n^1, \dots, S_n^q\}$ such that $\left\{S_n^j\right\}_{j=1}^q = \left\{\chi_i^j(t)\right\}_{i=1}^n$ in $\mathcal{L}^2[0, T]$. The regression function can be expressed by:

$$y_i = \alpha + \sum_{j=1}^{q} \int_0^T \chi_i^j(t)\beta^j(t)dt + \epsilon_i, \tag{A.25}$$

where $\beta(t) \in \mathcal{L}^2[0, T]$ is the functional parameter and $\epsilon \sim N(0, \sigma^2)$ is the residual such that $Cov(\epsilon_i, \epsilon_j) = 0\ \forall i \neq j$.

In this work, for the case study on the estimation of precipitation, an adaptation of this model is proposed using temperature and wind speed as functional explanatory covariates. Thus, the following problem arises:

$$y_i = \alpha + \int_0^T \chi_i^1(t)\beta^1(t)dt + \int_0^T \chi_i^2(t)\beta^2(t)dt + \epsilon_i. \tag{A.26}$$

In this case, $\{y_i\}_{i=1}^n$ is the sample of the scalar variable Y, the precipitation, whereas the samples $S_n^1 = \{\chi_i^1\}_{i=1}^n$ and $S_n^2 = \{\chi_i^2\}_{i=1}^n$ correspond to the temperature and wind speed functional explanatory variables, respectively. Thus, the estimation of $\beta^1(t)$ and $\beta^2(t)$ is obtained by the representation in functional bases of the B-spline type.

Appendix B. Definitions of goodness of fit and error measurements

Mean Absolute Error: It is a measure of the absolute prediction error defined by

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{B.1}$$

where y_i are the real observations, $\hat{y}_i$ are the predicted values and n is the sample size.

Mean Absolute Percent Error: It is an error measurement expressed in percent and defined by

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|}. \tag{B.2}$$

This measure expresses the errors as a percentage, which allows the study of the errors. As it is measured in percentage terms, it allows the performance of a model to be evaluated in absolute terms, without having to compare it with others.

Predicted Residual Error Sum of Squares: It is a measure of the deviation between the fitted values and the observed values of the forecast error. It is defined by

$$PRESS = \sum_{i=1}^{n} |y_i - \hat{y}_i|^2. \tag{B.3}$$

Adjusted R^2: It measures the proportion of explained variability of the response variable by fitting a regression model and it is defined by

$$R^2_{Adjusted} = 1 - \frac{SSR/(n-(k+1))}{SSY/(n-1)}, \tag{B.4}$$

whereby, SSR is the residual sum of squares, $SSR/(n-(k+1))$ is the residual variance, SSY is the response variable sum of squares, and $SSY/(n-1)$ is the response variable variance.

Akaike Information Criteria: It is a measure of the relative quality of a statistical model, and is defined by

$$AIC = 2k - 2\ln(L[\hat{\beta}(k)]), \tag{B.5}$$

whereby $(L[\hat{\beta}(k)]$ is the maximum likelihood function of the observations and k is the number of independent parameters estimated within the model. The first term indicates a penalty that increases as the number of parameters increases; while the second term can be interpreted as a measure of goodness of fit. As lower is the AIC as better is the model.

Corrected Akaike Information Criteria: It is a measure of second-order correction of the value of the AIC, defined by

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \tag{B.6}$$

where n is the sample size.

Bayesian/Schwarz Information Criterion: It is an alternative measure to AIC with a Bayesian approach, defined by

$$BIC = k\ln(n) - 2\ln(L[\hat{\beta}(k)]). \tag{B.7}$$

This coefficient penalizes the number of parameters by means of $\ln(n)$ becoming much more reliable as the sample size increases. It estimates the loss of information when approximating a real model with an estimated model. The best model will be the one with the smallest value of BIC.

Appendix C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18334.

References

[1] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, P. Nillaor, IoT and agriculture data analysis for smart farm, Comput. Electron. Agric. 156 (2019) 467–474.
[2] D.D. Mühl, L. de Oliveira, A bibliometric and thematic approach to Agriculture 4.0, Heliyon 8 (2022) e09369.
[3] S. Ullah, C.F. Finch, Applications of functional data analysis: a systematic review, BMC Med. Res. Methodol. 13 (2013) 1–12.
[4] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, vol. 40, Springer, New York, NY, 2005.
[5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, vol. 76, Springer, 2006.
[6] J.O. Ramsay, C.J. Dalzell, Some tools for functional data analysis, J. R. Stat. Soc., Ser. B, Stat. Methodol. 53 (1991) 539–572.
[7] G. Hébrail, B. Hugueney, Y. Lechevallier, F. Rossi, Exploratory analysis of functional data via clustering and optimal segmentation, Neurocomputing 73 (2010) 1125–1141.
[8] M. Flores, J. Tarrio-Saavedra, R. Fernandez-Casal, S. Naya, Functional extensions of Mandel's h and k statistics for outlier detection in interlaboratory studies, Chemom. Intell. Lab. Syst. 176 (2018) 134–148.
[9] C. Capezza, A. Lepore, A. Menafoglio, B. Palumbo, S. Vantini, Control charts for monitoring ship operating conditions and CO2 emissions based on scalar-on-function regression, Appl. Stoch. Models Bus. Ind. 36 (2020) 477–500.
[10] J. Tarrío-Saavedra, S. Naya, M. Francisco-Fernández, R. Artiaga, J. Lopez-Beceiro, Application of functional ANOVA to the study of thermal stability of micro–nano silica epoxy composites, Chemom. Intell. Lab. Syst. 105 (2011) 114–124.
[11] A. Baíllo, A. Cuevas, R. Fraiman, Classification Methods for Functional Data, Oxford Handbooks Online, 2018.
[12] M. Febrero-Bande, P. Galeano, W. González-Manteiga, Estimation, imputation and prediction for the functional linear model with scalar response with responses missing at random, Comput. Stat. Data Anal. 131 (2019) 91–103.
[13] R.J. Hyndman, M.S. Ullah, Robust forecasting of mortality and fertility rates: a functional data approach, Comput. Stat. Data Anal. 51 (2007) 4942–4956.
[14] M. Flores, S. Naya, R. Fernández-Casal, S. Zaragoza, P. Raña, J. Tarrío-Saavedra, Constructing a control chart using functional data, Mathematics 8 (2020) 58.
[15] J. Tarrío-Saavedra, M. Francisco-Fernández, S. Naya, J. López-Beceiro, C. Gracia-Fernández, R. Artiaga, Wood identification using pressure DSC data, J. Chemom. 27 (2013) 475–487.
[16] J. Tarrío-Saavedra, N. Sánchez-Carnero, A. Prieto, Comparative study of FDA and time series approaches for seabed classification from acoustic curves, Math. Geosci. 52 (2020) 669–692.
[17] X. Leng, H.G. Müller, Classification using functional data analysis for temporal gene expression data, Bioinformatics 22 (2006) 68–76.
[18] R. Giraldo, P. Delicado, J. Mateu, Continuous time-varying kriging for spatial prediction of functional data: an environmental application, J. Agric. Biol. Environ. Stat. 15 (2010) 66–82.
[19] H. Shi, J. Cao, Robust functional principal component analysis based on a new regression framework, J. Agric. Biol. Environ. Stat. (2022) 1–21.
[20] J.O. Ramsay, S. Graves, G. Hooker, fda: Functional Data Analysis, 2020, R package version 5.1.9.
[21] M. Febrero Bande, M. Oviedo de la Fuente, Statistical computing in functional data analysis: the R package fda.usc, J. Stat. Softw. 51 (2012) 3–20.
[22] R.T. Ogden, C.E. Miller, K. Takezawa, S. Ninomiya, Functional regression in crop lodging assessment with digital images, J. Agric. Biol. Environ. Stat. 7 (2002) 389–402.
[23] W.H. Yang, C.K. Wikle, S.H. Holan, M.L. Wildhaber, Ecological prediction with nonlinear multivariate time-frequency functional data models, J. Agric. Biol. Environ. Stat. 18 (2013) 450–474.
[24] N. Kantanantha, N. Serban, P. Griffin, Yield and price forecasting for stochastic crop decision planning, J. Agric. Biol. Environ. Stat. 15 (2010) 362–380.
[25] C. Lewis-Beck, Z. Zhu, V. Walker, B. Hornbuckle, Modeling crop phenology in the us corn belt using spatially referenced SMOS satellite data, J. Agric. Biol. Environ. Stat. 25 (2020) 657–675.
[26] F. Sichiqui, J.G. Huilca, A. García-Cedeño, J.C. Guillermo, D. Rivas, R. Clotet, M. Huerta, Agricultural information management: a case study in corn crops in Ecuador, in: The International Conference on Advances in Emerging Trends and Technologies, Springer, 2019, pp. 113–124.
[27] P.H.d.A. Paiva, C.J.C. Bacha, The gross domestic product (gdp) shares of the agriculture sector and the hydrocarbon and mining sector in the countries of south america between 1960 and 2014, CEPAL Rev. (2019).
[28] Food and Agriculture Organization of the United Nations (FAO), GIEWS Country Brief Ecuador, Global Information and Early Warning System on Food and Agriculture (GIEWS) 15-June (2022) 1–2.
[29] A. Martinez-Cob, Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain, J. Hydrol. 174 (1996) 19–35.
[30] S.T. Chen, P.S. Yu, Y.H. Tang, Statistical downscaling of daily precipitation using support vector machines and multivariate analysis, J. Hydrol. 385 (2010) 13–22.
[31] D. Ballari, R. Giraldo, L. Campozano, E. Samaniego, Spatial functional data analysis for regionalizing precipitation seasonality and intensity in a sparsely monitored region: unveiling the spatio-temporal dependencies of precipitation in Ecuador, Int. J. Climatol. 38 (2018) 3337–3354.

[32] M.A. Hael, Y. Yongsheng, B.I. Saleh, Visualization of rainfall data using functional data analysis, SN Appl. Sci. 2 (2020) 1–12.
[33] W. Caballero, R. Giraldo, J. Mateu, A universal kriging approach for spatial functional data, Stoch. Environ. Res. Risk Assess. 27 (2013) 1553–1563.
[34] L. Turk, Évolution des besoins en eau d'irrigation. évapotranspiration potentielle. Formule climatique simplifiée et mise a jour, Ann. Agron. 12 (1961) 13–49.
[35] C. Perez, C. Nicklin, O. Dangles, S. Vanek, S.G. Sherwood, S. Halloy, K.A. Garrett, G.A. Forbes, Climate change in the high andes: implications and adaptation strategies for small-scale farmers, Int. J. Environ. Cultur. Econ. Soc. Sustain. 6 (2010) 71–88.
[36] N. Borja, J. Cho, K.S. Choi, The influence of climate change on irrigation water requirements for corn in the coastal region of Ecuador, Paddy Water Environ. 15 (2017) 71–78.
[37] R.G. Allen, L.S. Pereira, D. Raes, M. Smith, et al., Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56, FAO, Rome 300 (1998) D05109.
[38] Y. Zhang, Y. Wang, H. Niu, Effects of temperature, precipitation and carbon dioxide concentrations on the requirements for crop irrigation water in China under future climate scenarios, Sci. Total Environ. 656 (2019) 373–387.
[39] T. Hastie, C. Mallows, A statistical view of some chemometrics regression tools: discussion, Technometrics 35 (1993) 140–143.
[40] B.D. Marx, P.H. Eilers, Generalized linear regression on sampled signals and curves: a P-spline approach, Technometrics 41 (1999) 1–13.
[41] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Stat. Probab. Lett. 45 (1999) 11–22.
[42] C. Preda, G. Saporta, PLS regression on a stochastic process, Comput. Stat. Data Anal. 48 (2005) 149–158.
[43] C. De Boor, Package for calculating with B-splines, SIAM J. Math. Anal. 14 (1977) 441–472.
[44] A. Haar, Zur Theorie der orthogonalen Funktionensysteme, Math. Ann. 69 (1910) 331–371.
[45] R. Fraiman, G. Muniz, Trimmed means for functional data, Test 10 (2001) 419–440.
[46] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Stat. 22 (2007) 481–496.
[47] J.A. Cuesta-Albertos, A. Nieto-Reyes, The random tukey depth, Comput. Stat. Data Anal. 52 (2008) 4979–4988.
[48] A.M. Aguilera, M.C. Aguilera Morillo, C. Preda, Penalized versions of functional PLS regression, Chemom. Intell. Lab. Syst. 154 (2016) 80–92.
[49] S. Naya, J. Tarrío-Saavedra, J. López-Beceiro, M. Francisco-Fernández, M. Flores, R. Artiaga, Statistical functional approach for interlaboratory studies with thermal data, J. Therm. Anal. Calorim. 118 (2014) 1229–1243.
[50] M. Smith, CROPWAT: A Computer Program for Irrigation Planning and Management, vol. 46, Food & Agriculture Org, Rome, Italy, 1992.
[51] J. Márquez, Boletín Técnico y presentación de la ESPAC 2020, Technical Report, Instituto Nacional de Estadísticas y Censos, 2021, https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2020/BoletinTecnicoESPAC2020.pdf.