Estimation of Climate Zone Effects on Iranian Temperature, Humidity, and Precipitation using Functional Analysis of Covariance

S. Mohammad E. Hosseini-nasab[*] and N. Kheirollahzadeh

Shahid Beheshti University

Abstract. Functional Data Analysis (FDA) has recently made considerable progress because of easier access to the data that are essentially in the form of curves. Although functional modeling of Iranian precipitation based on temperature or humidity was done before, here we use functional analysis of variance and covariance to analyze the weather data collected randomly from Iranian weather stations in 2010. Using a functional linear model in which the covariate (climate zones) and response variable (temperature or humidity) are functions, we estimate the coefficients via functional analysis of variance. As a result, we can determine how much of temperature or humidity variation in the weather stations is affected by the geographic areas. Using a functional analysis of covariance, we can also investigate that how much of the precipitation variation, can be expressed by the temperature residual effects or humidity residual effects (temperature or humidity effects after eliminating the climate impacts) and the corresponding climate effects.

Keywords. Functional data analysis; functional linear model; functional analysis of variance; functional analysis of covariance; climate zone effects.

MSC 2010: 6207, 62J10.

---

[*] Corresponding author

1   Introduction

Functional Data Analysis (FDA) has recently made considerable progress because of easier access to the data that are essentially in the form of curves. Phenomena such as temperature, humidity and the amount of precipitation are functions of time because their amount is changing with time. While the data related to these phenomena might be collected discretely (daily), they should be treated as functional observations (Ramsay and Silverman, 2002, 2005).

Suppose that the geography of a country divides it into different climatic regions. When investigating its weather characteristics, the aim is to study the characteristic types of temperature or humidity in each zone, leading to an analysis of variance problem with some treatment groups. In particular, we are going to investigate that how much of the pattern of annual variation of temperature or of humidity in the weather stations is explainable by its geographical area. Moreover, we find out whether this effect is substantial at a specific time t. Therefore, we can use a functional analysis of variance or of covariance that takes the functional nature of the response variable into account.

There are only a few works related to statistical analyze of these phenomena in Iran in which they have generally used simple or multivariate methods for analyzing data, see, for example, Nazemosadat et al. (2006), and Kavousi and Meshkani (2009). Others, such as Moosavi (2011), have not considered statistical approaches in their works.

Tazikeh and Hosseini-nasab (2007) analyzed Iranian temperature and precipitation by using functional principal components analysis. Modeling of Iranian precipitation based on temperature and humidity with considering the essential nature of such phenomena that are continuous functions of time has been first done by Hoseeini-nasab et al. (2008) using functional regression. For each of the three phenomena, the corresponding data are generally reported as vectors by the weather stations. However, if one treats those data as multivariate observations and uses multivariate methods, he may discard the benefits of considering them as functional data (Ramsay and Silverman, 2002, 2005). We have treated a real dataset that was collected from 102 Iranian weather stations in 2010. The dataset contains measurements related to the three phenomena; temperature, humidity and precipitation. The amount of these three was reported daily by each station over the year. Because of functional nature of the data, the original discrete data should be

first converted to continuous functions, using "basis function methods" that is a way for dimension reduction (see Ramsay and Silverman, 2005). In this article, using a functional linear model in which the covariate (climate zones) and response variable (temperature or humidity) are functions, we estimate the coefficients via functional analysis of variance or of covariance. As a result, we can determine how much of temperature or of humidity variation in the stations is affected by the geographic areas. After constructing an appropriate model, we investigate that how much of the precipitation variation in the stations, can be expressed by the temperature residual effects or humidity residual effects (temperature or humidity effects after eliminating the climate impacts) and the corresponding climate effects.

There are different methods for estimating coefficients in a functional linear model. Cardot and et. al. (1999) estimated the model coefficients via the functional principal components. Cardot and et. al. (2003) used B spline to estimate the model coefficients in which the covariate variable and response variable are functions and scalar, respectively. Also James (2002) developed generalized linear models in which some covariates are functions. Zoglat (2008) generalized multivariate analysis of variance to functional analysis of variance. In this article, we will introduce the functional analysis of variance in Section 2 and in Section 3 we explain how to estimate the coefficients, using the basis functions, approximating each function with a finite number of orthonormal functions. Then, in Section 4 we express how to construct a model with more than one covariate for functional analysis of covariance. Finally, using Iranian weather stations data which were collected from across the country in 2010, we explore temperature and humidity variation based on the effect of climatic zones and precipitation based on that of temperature or humidity residual (the effect of temperature or humidity after removing the effect of climate zones) and the effect of climate zone. Then we will assess the models and interpret the results.

2 Analysis of Variance for Functional Response

Analysis of variance is used to explain the variability of the response by some factors and to estimate the effects of different levels of the factors. Here we consider three phenomena; temperature, humidity and precipitation. Data on these phenomena are generally reported daily, but are treated as functional data. We explain how to analyze a functional response variable affected by a set of scalar covariates whose values are 0's and 1's coding

the categories of factor variables. Because the dependent variable, $y(t)$, is a function of t, the methodology that is used for the classical analysis of variance is not working here. Hence, using functional analysis of variance, denoted by FANOVA, the model for the mth observation in the gth group, indicated by $y_{mg}(t)$, is

$$y_{mg}(t) = \mu(t) + \alpha_g(t) + \varepsilon_{mg}(t), \qquad m = 1, 2, .., N_g, \ g = 1, ..., q, \qquad (1)$$

where the function $\mu(\cdot)$ is the mean function, and therefore indicates the average of $y(t)$ in the population we are sampling from. The terms $\alpha_g(\cdot)$ indicate the specific effects on the response variable $y(t)$ of being in treatment g satisfying the constraint

$$\sum_g \alpha_g(t) = 0, \qquad \forall t. \qquad (2)$$

This condition is necessary for their unique identification. The residual function $\varepsilon_{mg}(\cdot)$ shows the unexplained variation specific to the mth observation within treatment g. Let $\boldsymbol{Z}$ be an $N \times (q+1)$ design matrix for this model, with one row for each individual treatment, and use the index (mg) for the row corresponding to subject m in group g; this row has a one in the first column, a one in column $g + 1$, and zeroes are in the rest. We denote the value in this row and in the jth column of $\boldsymbol{Z}$ by $z_{(mg)j}$. It should be noted that the design matrix $\boldsymbol{Z}$ has exactly the same structure as for the corresponding univariate case (for more details see Ramsay and Silverman, 2005). Therefore, the model (1) can be written as:

$$y_{mg}(t) = \sum_{j=1}^{q+1} z_{(mg)j} \beta_j(t) + \varepsilon_{mg}(t). \qquad (3)$$

Also, the matrix form of the model is,

$$\boldsymbol{y}(t) = \boldsymbol{Z}\boldsymbol{\beta}(t) + \boldsymbol{\varepsilon}(t), \qquad (4)$$

where

$$\boldsymbol{y}(t) = (y_1(t), \dots, y_N(t))',$$
$$\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \dots, \varepsilon_N(t))',$$
$$\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_{q+1}(t))'.$$

2.1 Parameter Estimation

The aim is to estimate the parameter $\boldsymbol{\beta}(t)$ in model 4 so that it minimizes the least square fitting criterion:

$$\text{LMSSE}(\boldsymbol{\beta}) = \int [\boldsymbol{y}(t) - \boldsymbol{Z}\boldsymbol{\beta}(t)]^{'}[\boldsymbol{y}(t) - \boldsymbol{Z}\boldsymbol{\beta}(t)]dt.$$

We assume that all functions under consideration belong to $L_2([0,T])$, the space of all square integrable functions on $[0,T]$. Hence, we can write each function as a countable linear combination of basis functions. Now, the regression functions $\beta_j(\cdot)$ can be expressed as follow:

$$\beta_i(t) \approx \sum_{k=1}^{k_\beta} b_{ik}\theta_k(t), \qquad \text{or} \qquad \boldsymbol{\beta}(t) \approx \boldsymbol{B}\boldsymbol{\theta}(t),$$

where $\boldsymbol{B}$ is a $(q+1) \times k_\beta$ matrix of coefficients b_{ik} and the basis functions $\boldsymbol{\theta}(t)$ is a k_β-vector. Similarly, the observed function $y_i(t)$ can be written as:

$$y_i(t) \approx \sum_{k=1}^{k_y} c_{ik}\phi_k(t), \qquad \text{or} \qquad \boldsymbol{y}(t) \approx \boldsymbol{C}\boldsymbol{\phi}(t),$$

where $\boldsymbol{C}$ is an $N \times k_y$ matrix of coefficients c_{ik}. To simplify the calculations, we may choose to use the same basis functions as they were used to expand the response functions, i.e. $\boldsymbol{\theta} = \boldsymbol{\phi}$. To estimate $\boldsymbol{\beta}(\cdot)$, we can apply roughness penalties or regularization. Here, when estimating $\boldsymbol{\beta}(\cdot)$, smoothness is imposed only on the functional parameters, since smoothing $\boldsymbol{y}(t)$ as well as $\boldsymbol{\beta}(\cdot)$, may cause loosing important information in the estimate of $\boldsymbol{\beta}(\cdot)$.

Let L be a linear differential operator. Define a roughness penalty for $\boldsymbol{\beta}(\cdot)$ as

$$\text{PEN}_L(\boldsymbol{\beta}) = \int [L\boldsymbol{\beta}(s)]^{'}[L\boldsymbol{\beta}(s)]ds.$$

Since $\boldsymbol{\phi}$ is a column vector of length k_y of basis functions, $\boldsymbol{\phi}\boldsymbol{\phi}^{'}$ is a square matrix of order k_y containing all possible pairs of these functions. Now, define

$$\boldsymbol{J}_{\phi\phi} = \int \boldsymbol{\phi}(t)\boldsymbol{\phi}^{'}(t)dt, \qquad \boldsymbol{J}_{\theta\theta} = \int \boldsymbol{\theta}(t)\boldsymbol{\theta}^{'}(t)dt,$$

$$\boldsymbol{J}_{\phi\theta} = \int \boldsymbol{\phi}(t)\boldsymbol{\theta}^{'}(t)dt, \qquad \boldsymbol{R} = \int (L\boldsymbol{\theta}(t))(L\boldsymbol{\theta}(t))^{'}dt,$$

where $\boldsymbol{J}_{\phi\phi}$ is constant symmetric order k_y matrix of integrated products and similarly for the other three matrices. Therefore, the penalized least squares criterion leads to:

$$
\begin{aligned}
\text{PENSSE }(y|\boldsymbol{\beta}) &= \int \{\boldsymbol{C}\boldsymbol{\phi}(t) - \underline{\boldsymbol{Z}}\,\underline{\boldsymbol{B}}\boldsymbol{\theta}(t)\}' \{\boldsymbol{C}\boldsymbol{\phi}(t) - \underline{\boldsymbol{Z}}\,\underline{\boldsymbol{B}}\boldsymbol{\theta}(t)\} dt \\
&\quad + \lambda \int \{L\boldsymbol{B}\boldsymbol{\theta}(t))'(L\boldsymbol{B}\boldsymbol{\theta}(t)\} dt \\
&= \text{tr}(\boldsymbol{C}'\boldsymbol{C}\underline{\boldsymbol{J}}_{\phi\phi}) + \text{tr}(\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{B}\underline{\boldsymbol{J}}_{\theta\theta}\boldsymbol{B}') \\
&\quad - 2\text{tr}(\boldsymbol{B}\underline{\boldsymbol{J}}_{\phi\theta}\boldsymbol{C}'\boldsymbol{Z}) + \lambda\text{tr}(\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}'),
\end{aligned}
$$

where $tr(\cdot)$ denotes the transpose operator. Computing the first derivative of the above expression with respect to matrix $\boldsymbol{B}$ and setting the result to zero, we find that the solution $\boldsymbol{B}$ is

$$
\text{vec}(\boldsymbol{B}) = [\underline{\boldsymbol{J}}_{\theta\theta} \otimes (\boldsymbol{Z}'\boldsymbol{Z}) + \boldsymbol{R} \otimes \lambda\boldsymbol{I}]^{-1}\text{vec}(\boldsymbol{Z}'\boldsymbol{C}\underline{\boldsymbol{J}}_{\phi\theta}), \tag{5}
$$

where $\otimes$ is the Kronecker product (see Ramsay and Silverman, 2005).

2.2   Assessing the Fit

Unlike the standard multivariate case, in the FANOVA we can not only characterize the typical response variable patterns for the weather stations in each climate zone, but also investigate whether these effects are substantial at a specific time t. In other words, it is the dependence of these quantities on t that makes the procedure different from the standard multivariate case. Thus, in FDA the model fitting is assessed by using the squared multiple correlation function (RSQ) which is defined as follows:

$$
\text{RSQ}(t) = \frac{[\text{SSY}(t) - \text{SSE}(t)]}{\text{SSY}(t)},
$$

where

$$
\text{SSY}(t) = \sum_{mg} [\boldsymbol{y}_{mg}(t) - \hat{\mu}(t)]^2, \quad \text{SSE}(t) = \sum_{mg} [\boldsymbol{y}_{mg}(t) - \boldsymbol{Z}_{mg}\hat{\boldsymbol{\beta}}(t)]^2,
$$

are the total sum of squares and error sum of squares functions, respectively. Alternatively, we can compute the F-ratio function,

$$
\text{F} - \text{ratio}(t) = \frac{\text{MSR}(t)}{\text{MSE}(t)},
$$

where

$$\text{MSE}(t) = \frac{\text{SSE}(t)}{df(error)}, \qquad \text{MSR}(t) = \frac{(\text{SSY}(t) - \text{SSE}(t))}{df(reg)},$$

in which df(error) and df(reg) are the degrees of freedom for error and model respectively (Ramsay and Silverman, 2005). At each specific value of t, if the amount of $F - \text{ratio}$ is above the 5% significance level, there are differences between the effects of the factor.

3 Functional Analysis of Covariance

The FANOVA model (1) can be extended to the following functional analysis of covariance model:

$$y_{mg}(t) = \mu(t) + \alpha_g(t) + \text{X}_1\text{Res}_{mg}(t)\beta_1(t) + \cdots + \text{X}_p\,\text{Res}_{mg}(t)\beta_p(t) + \varepsilon(t). \quad (6)$$

where $y_{mg}(t)$ is the value of the response variable at time t for the mth observation within treatment g and the variables $X_1\,Res_{mg}(t),..., X_p\,Res_{mg}(t)$ are the residual response variable after removing the response variable effect of α_g by using the estimated model (1). The reason for removing $\alpha_g(\cdot)$ effects from the response variable profiles before using them in this model is that we have already allowed for these effects in the model. In this way, $\alpha_g(\cdot)$ in the equation will not consider twice.

3.1 Parameter Estimation

Consider the multiple linear model:

$$y_i(t) = \sum_{j=1}^{q} z_{ij}(t)\beta_j(t) + \varepsilon_i(t) \qquad i = 1, 2, ..., N. \quad (7)$$

The matrix form of the above model is:

$$\boldsymbol{y}(t) = \boldsymbol{Z}(t)\boldsymbol{\beta}(t) + \boldsymbol{\varepsilon}(t), \quad (8)$$

where the $N \times q$ design matrix $\boldsymbol{Z}$ contains the z_{ij}'s, the vector coefficient function $\boldsymbol{\beta}$ of length q contains each of the regression functions, and $\boldsymbol{y}$ is a functional vector of length N of the response functions. Using a basis function expansion for each regression function $\beta_j(\cdot)$, $j = 1, \ldots, q$, with roughness

penalties to control the smoothness of the estimates for the $\beta_j(\cdot)$'s we can estimate the coefficients. To do that, for each regression coefficient function, we take a roughness penalty as follows:

$$\text{PEN}_j(\beta_j) = \lambda_j \int [L_j \beta_j(t)]^2 dt,$$

where Lj is a linear differential operator appropriate for that functional parameter. Now, the weighted regularized fitting criterion is

$$\text{LMSSE}(\boldsymbol{\beta}) = \int \boldsymbol{r}(t)' \boldsymbol{r}(t) dt + \sum_j^q \lambda_j \int [L_j \beta_j(t)]^2 dt,$$

where

$$\boldsymbol{r}(t) = \boldsymbol{y}(t) - \boldsymbol{Z}(t) \boldsymbol{\beta}(t).$$

Let regression function β_j have the expansion

$$\beta_j(t) = \sum_{k=1}^{k_j} b_{kj} \theta_{kj}(t) = \boldsymbol{\theta}_j(t)' \boldsymbol{b}_j.$$

in terms of k_j basis functions θ_{kj} and $k_\beta = \sum_{j=1}^q k_j$. Then, model (8) is expressed as:

$$\boldsymbol{y}(t) = \boldsymbol{Z}(t) \boldsymbol{\Theta}(t) \boldsymbol{b} + \boldsymbol{\varepsilon}(t), \tag{9}$$

where

$$\boldsymbol{b} = (\boldsymbol{b}_1', \boldsymbol{b}_2', ..., \boldsymbol{b}_q')', \qquad \boldsymbol{\Theta}(t) = \text{diag}(\boldsymbol{\theta}_1'(t), \cdots, \boldsymbol{\theta}_q'(t)).$$

Hence, we have

$$\boldsymbol{y}(t) = \boldsymbol{Z}^*(t) \boldsymbol{b} + \boldsymbol{\varepsilon}(t),$$

where

$$\boldsymbol{Z}^*(t) = \boldsymbol{Z}(t) \boldsymbol{\Theta}(t).$$

Define,

$$\boldsymbol{R} = \text{diag}(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_q),$$

where

$$\boldsymbol{R}_j = \lambda_j \int \{L_j \boldsymbol{\theta}_j(t)\}\{L_j \boldsymbol{\theta}_j(t)\}' dt.$$

Now, the normal equations weighted least squares solution for the composite coefficient vector $\boldsymbol{b}$ can be written as:

$$\left[\int \boldsymbol{\Theta}'(t)\boldsymbol{Z}'(t)\boldsymbol{Z}(t)\boldsymbol{\Theta}(t)dt + \boldsymbol{R}\right]\boldsymbol{b} = \int [\boldsymbol{\Theta}'(t)\boldsymbol{Z}'(t)\boldsymbol{y}(t)]dt. \tag{10}$$

For more details see Ramsay and Silverman (2005).

4 Real Data Analysis

Temperature, humidity and the amount of precipitation are essentially continuous functions of time, because their amount is changing with time. Therefore, they should be treated as functional observations.

The data treated here were collected by the Iranian weather stations in 2010. Iran is divided into four different climate zones by Coupon method (Kaviani and Alijani, 1999). The zones are arid and semi-arid zone (SA), semi-desert zone (SD), tropical desert zone (TD) and high altitude zone (HA). The amount of temperature, humidity and precipitation were collected daily from 102 weather stations, located in four different zones, that were chosen randomly from all weather stations across the country. The original data were reported discretely (daily). These discrete data were converted to continuous functions, and denoted by the functions Temp(t), Humid(t) and Prec(t), representing the amount of temperature, humidity and precipitation at time t for each station, respectively (see Figure 1). We are going to investigate:

1. How much of the annual variation of temperature or humidity in the stations are affected by geographical zones?

2. How much of the annual variation of precipitation in the stations is expressed by climate zone effects and residual temperature effects(after removing the effects of climate zones from temperature) or residual humidity effects(after removing the effects of different climate zones from humidity)?
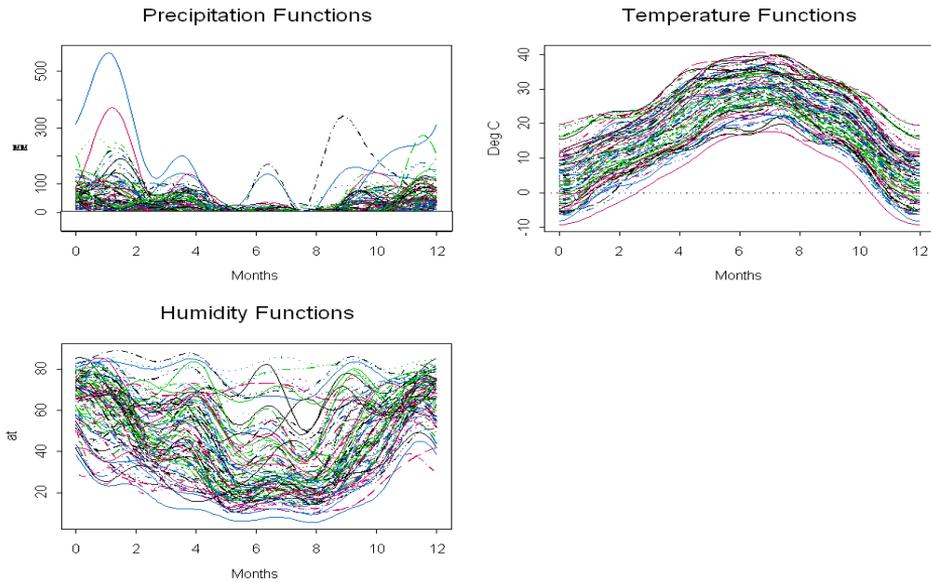
Figure 1. Daily precipitation, temperatures, and humidity for the Iranian weather stations.

4.1   Temperature Prediction Based on Climate Zone Effects

We can use model (1) to predict temperature based on climate zone effects. In this model, Temp($\cdot$) is the response variable. Therefore, the regression coefficients are estimated by using formula (5). Figures (2) and (3) show the estimated zone effects $\alpha_g(\cdot)$ and estimated effects $\mu(\cdot) + \alpha_g(\cdot)$ for four different regions, respectively. These figures show that:

- The SA weather stations appear to have a temperature average less than the Iranian average. In particular, the amount of temperature in this zone is lower than (about 3 degrees C) the Iranian average from the middle of spring until the end of summer. In the other time of the year, it is the same as Iranian temperature average.

- The SD temperature stations from the middle of June until the middle of August is a degree warmer than the Iranian temperature average and in other time of the year, there is no significant difference between them.

- Temperature of the TD stations is 10 degrees warmer than the tem-

perature average of Iran in all the time of the year.

- Temperature of the HA zone stations is 5 degrees colder than the Iranian average in all the time of the year.

We have summarized the results in Table 1.

Table 1. Summary of the results for temperature prediction based on climate zone effects

Climate zone	Temperature prediction
SA	3 degrees C colder than Iranian average in months $5-9$.
SD	1 degree C warmer than Iranina average in months $6-9$.
TD	10 degrees C warmer than Iranian average.
HA	5 degrees C colder than Iranian average.

Now for model evaluation, we use RSQ and F $-$ ratio assessment methods. Model degrees of freedom and error degrees of freedom are 3 and 97, respectively. Figure 4a presents RSQ for the fitted model. The differences between the climate zone effects in February, the middle of April until the middle of May, the middle of June until the middle of July and the middle of September until the middle of October is stronger than the effects in the middle of March until the middle of April, the middle of May until the middle of June and the middle of July until the middle of August. Figure 4b shows F-ratio for the estimated model. The F-ratio is everywhere substantially above of the 5% significance level of F(3,97). So, the effects of four different climate zones on temperature are significant at the level of 95 percent.

4.2   Humidity Prediction Based on Climate Zone Effects

We use model (1) to predict humidity based on climate zone. In this model, the humidity is response variable. Therefore, estimate of regression coefficients can be obtained from phrase (5). Figures (5) and (6) show the estimated zone effects $\alpha_g(\cdot)$ for four different climate zones and estimated effects $\mu(\cdot) + \alpha_g(\cdot)$, respectively. These figures show that (see also Table 2).
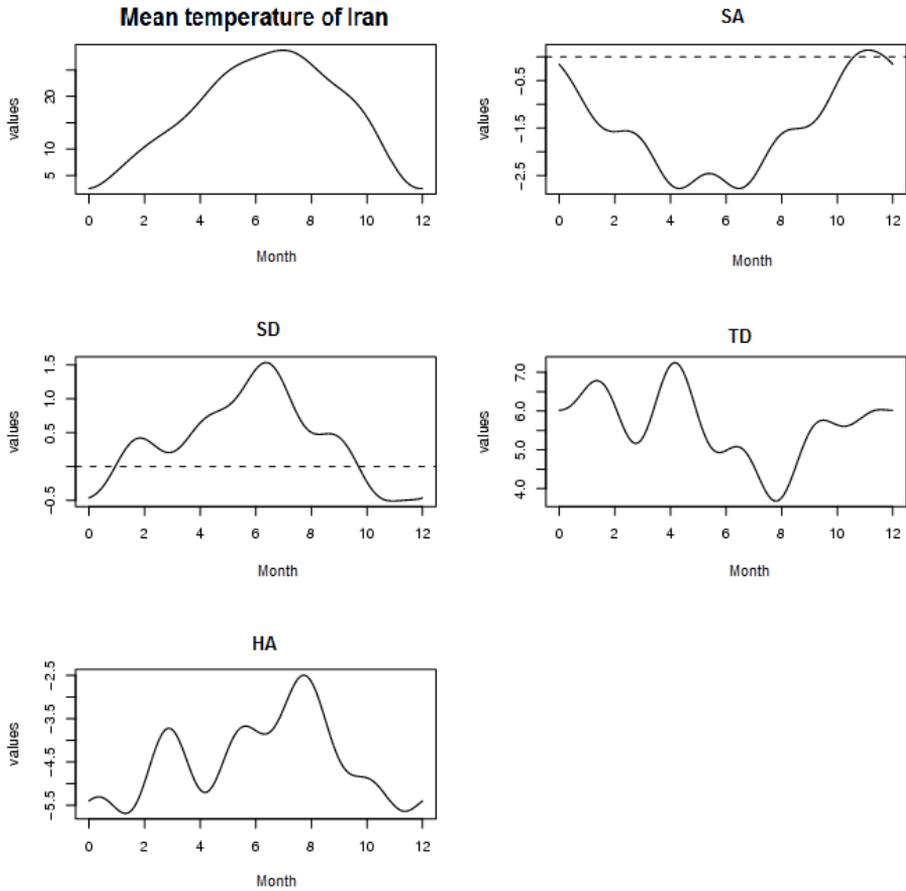
Figure 2. The estimated effects of different climatic zones of temperature in the functional analysis of variance model
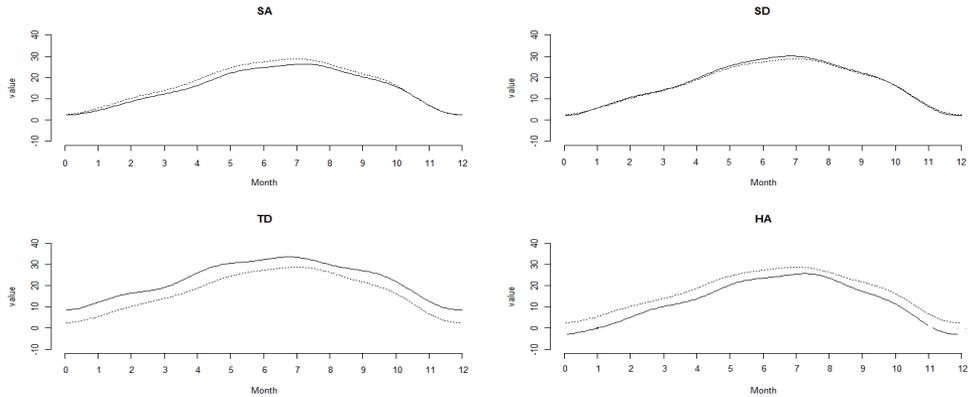
Figure 3. The estimated climate zone temperature profiles $\mu(\cdot) + \alpha_{\mathrm{g}}(\cdot)$ for the temperature function in the functional analysis of variance model (solid curves). The dashed curve is the Iranian mean function μ.
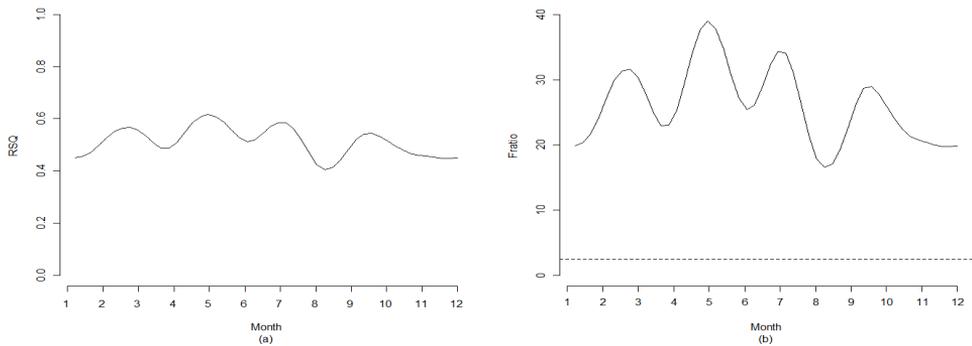

Figure 4. The left panel shows the squared multiple correlation function RSQ and the right panel represents the corresponding $\mathrm{F-ratio}$ function FRATIO in temperature. The horizontal dotted line indicates the 5% significance level for the $\mathrm{F-distribution}$ with 3 and 97 degrees of freedom.

Table 2. Summary of the results for Humidity prediction based on climate zone effects

Climate zone	Humidity prediction
SA	More than Iranian average; 20% in months $4 - 10$, 13% in month 10, and 5% in months $11 - 4$.
SD	Less than Iranian average; 13% in months $6 - 8$, 5% in months $9 - 10$ and $2 - 5$, coincides in other months.
TD	Less than Iranian average; 16% in months $4 - 5$, 2% in months $7 - 8$, 6% in other months.
HA	More than Iranian average except for months $7 - 8$; 7% in months 4 and $9 - 10$, 4% in months $1 - 2$.

- Humidity of the stations in the SA zone is more than the humidity average in all the months of the year. In particular, the amount of humidity in this zone is in early spring until the mid of October about 20%, the mid of October until the mid of November about 13% and the mid of November until the mid of March about 5% more than the humidity average of Iran.

- In general, humidity of the stations in the SD zone is less than the humidity average of Iran. The amount of humidity in this zone is from early June until the mid of August about 13% , from the mid of August until the mid of October and from the mid of February until early June about 5% less than the humidity average of Iran. In the other months of the year, the average of humidity approximately coincides with the humidity effects of the SD stations.

- The amount of humidity in the TD zone is similar to the SD zone, and is less than the average of humidity in Iran. This amount is from early April until mid of May about 16%, from the mid of July until early September about 2% and in other months of the year about 6% less than the Iran humidity average.

- Humidity in the HA zone is more than the Iran humidity average except for the months starting from early July until early August. In particular, the balance of humidity is about 7% in early April until May and the mid of September until the mid of October. It is about 4% more than Iran humidity average from early January until the mid

of February. The amount of humidity in the other months of the year approximately coincides with the humidity average.

For model assessment, we use RSQ and F-ratio values obtained from the data and shown by Figure 6. The model degrees of freedom and error degrees of freedom are 3 and 97 respectively. Figure 7a represents the RSQ for the fitted model. As it shows, difference between climate zones from the mid of November until early February and from the mid of April until the mid of May is stronger than the mid of July until early September and from the mid of May until the mid of June. Figure 7b displays F-ratio for each $t \in [0, 12]$. We can see that it is substantially everywhere above the 5% significance level of F(3,97). Therefore, the effects of four different climate zones in significant level of 95 percent for humidity are significant.
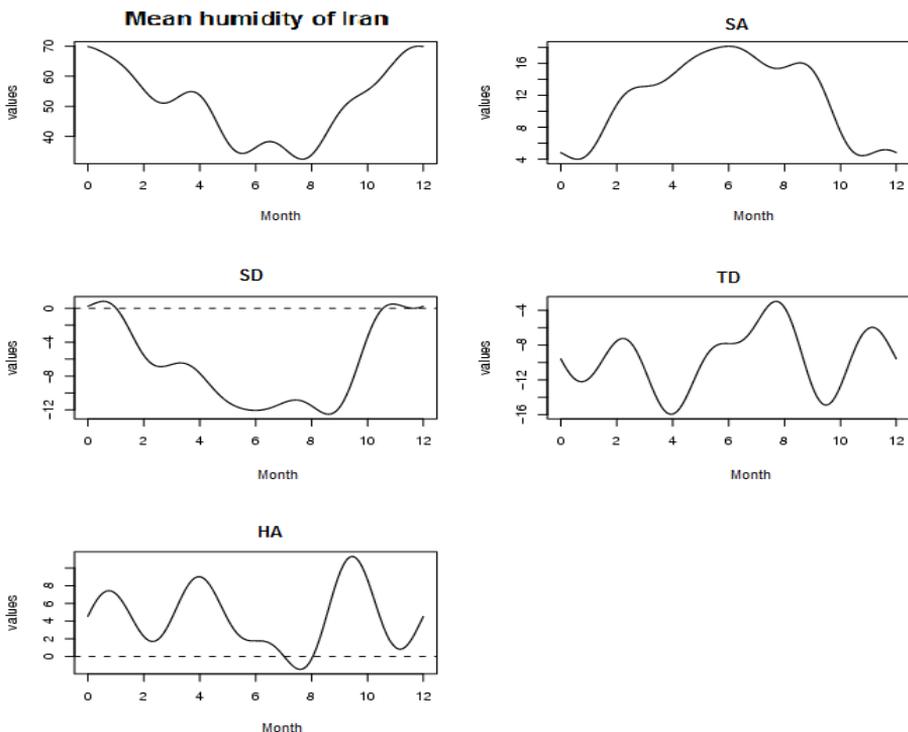

Figure 5. The estimated effects of different climatic zones of humidity in the functional analysis of variance model.
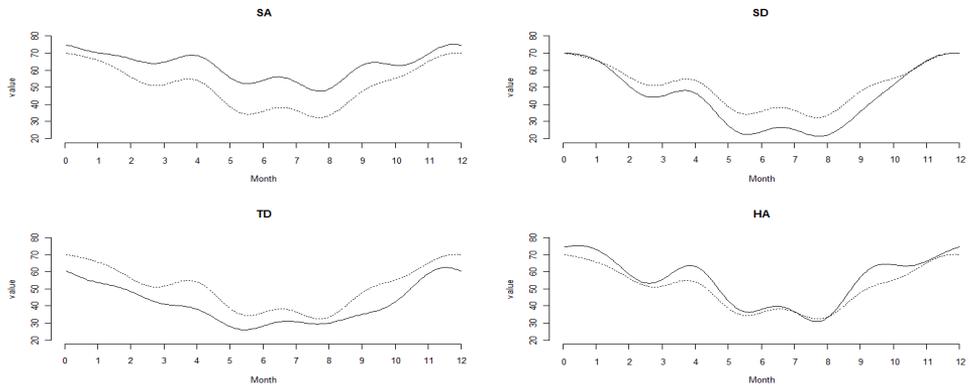
Figure 6. The estimated climate zone humidity profiles $\mu(\cdot) + \alpha_g(\cdot)$ for the humidity function in the functional analysis of variance model (solid curves). The dashed curve is the Iranian mean function μ.
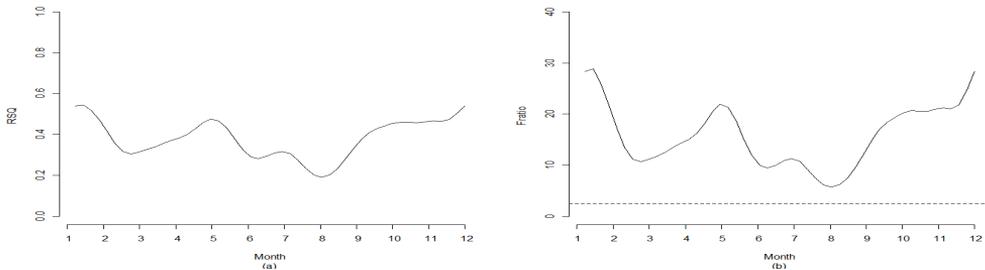

Figure 7. The left panel shows the squared multiple correlation function RSQ and the right panel represents the corresponding $F-$ratio function FRATIO in humidity. The horizontal dotted line indicates the 5% significance level for the $F-$distribution with 4 and 97 degrees of freedom.

4.3   Precipitation Prediction Based on Climate Zone and Temperature Residual

We are now going to answer the second question discussed in Section 3. Because the amount of precipitation in the climate zones is highly different, and experience indicates that logging magnitudes tends to improve the fitting power of the model. So, we decided to use the logarithm of the response. Now, we use model (6) to predict the logarithm of precipitation based on the climate zone and temperature residual. In this model, $p = 1$ and $X_1 Res_{mg}(\cdot)$ is considered to be $TempRes(\cdot)$. Therefore, the estimate of the regression coefficients is available by using formula (10). Figure 7 shows the estimate of $\mu(\cdot) + \alpha_g(\cdot) + TempRes(\cdot)$. The figure shows that (see also Table 3):

- Precipitation of the stations located in the SA zone is more than Iranian precipitation average. The amount of precipitation is about 1.4 millimeters more than the Iran precipitation average from the mid of June until early August and from the mid of September until early October. We observe that differences between the precipitation average, zones' effects and the temperature residual are low.

- In general, the amount of precipitation in the SD zone, from early June until mid of July and from early September until the mid of October is about 0.7 millimeters less than the Iranian precipitation average and form early January until the mid of April is about 0.5 millimeters more than the Iranian average. In the other months of the year, however, there is no significant difference between them.

- The amount of precipitation in the TD zone, from the mid of April until the mid of July and from early September until early November is about 1 millimeters and in the other months is about 0.5 millimeters less than the Iranian average. In the other months of the year, however, there is no significant difference between them.

- The amount of precipitation in HA zone, from the mid of April until mid of May and from early October until early November, is about 0.7 millimeters more than the Iranian average. In the other months of the year, there is no significant difference between them.

Table 3. Summary of the results for precipitation prediction based on climate zone and temperature residual effects

Climate zone	Precipitation prediction
SA	More than Iranian average; 1.4 millimeters in months $6 - 8$ and $9 - 10$, coincides in other months.
SD	0.7 millimeters less than average in months $9 - 10$; 0.5 millimeters more than average in months $1 - 5$, coincides in other months.
TD	Less than Iranian average; 1 millimeter in months $4 - 7$ and $9 - 11$, and 0.5 millimeters in other months.
HA	More than Iranian average; 0.7 millimeter in months 4 and $10 - 11$, coincides in other months.

We use the RSQ and F-ratio for model assessment. We have computed these quantities from the data and presented in Figure 9. The model degrees of freedom and the error degrees of freedom are 5 and 96, respectively. Figure 9a shows RSQ for the model that the logarithm of precipitation based on different climate zones and the $TempRes(\cdot)$ variable is expressed. It is seen that the differences between climate zones and temperature residual from the early May until early June and from early September until early October are stronger than the differences from mid of July until mid of August and from mid of October until mid of November. Figure 9b displays in the mid of July until August,the effects of the climate zones and $TempRes(\cdot)$ aren't significant. However, in the other months they are significant, since their corresponding F-ratio is substantially above of the 5% significance level of F(5,96), meaning that variation of the logarithm of precipitation can be expressed by these factors.

4.4   Precipitation Prediction Based on Climate Zone and Humidity Residual

The model (6) can be used to predict the logarithm of precipitation based on the climate zones and humidity residual. In this model, the response variable is the logarithm of precipitation, $p = 1$ and $X_1Res_{mg}(\cdot)$ is taken to be $HumidRes(\cdot)$. Therefore, estimate of the regression coefficients can be obtained from formula (10). Figure 10 also has similar interpretation as Figure 8.

Model degrees of freedom and error degrees of freedom are 5 and 96 respectively. Figure 11a shows the RSQ for the logarithm of precipitation

fitting based on different climate zones and the $HumidRes(\cdot)$ variable. It can be seen that differences between climate zones and humidity residual from the early September until early October and from mid of April until mid of May are stronger than differences between the two from mid of July to mid of August and mid of December to early February. Figure 11b shows that from the mid of July until mid of August, the effects of the climate zones and $HumidRes(\cdot)$ variable aren't significant but in the other months they are significant. The largest percentage of the logarithm of precipitation variation is expressed by the factors because the value of F-ratio is above of the line F(5,96,0.05).
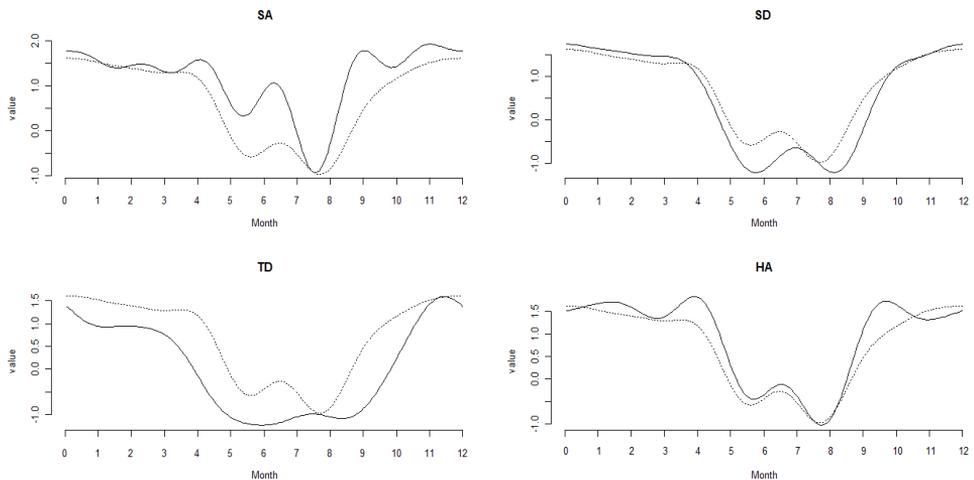

Figure 8. estimated climate zone and temperature residual profiles $\mu(\cdot) + \alpha_{\mathrm{g}}(\cdot) + TempRes(\cdot)$ for the logarithm of precipitation function in the functional analysis of covariance model (solid curves). The dashed curve is the Iranian mean function μ.
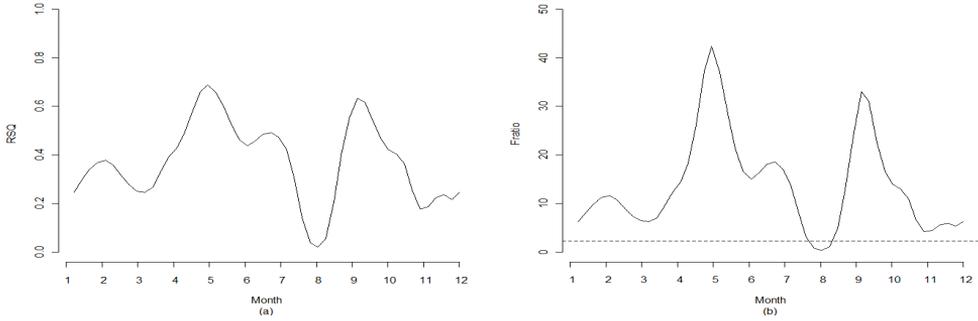
Figure 9. The left panel shows the squared multiple correlation function RSQ and the right panel reveals the corresponding $F-\mathrm{ratio}$ function FRATIO in the logarithm of precipitation based on climate zone and temperature residual. The horizontal dotted line indicates the 5% significance level for the $F-\mathrm{distribution}$ with 5 and 96 degrees of freedom.
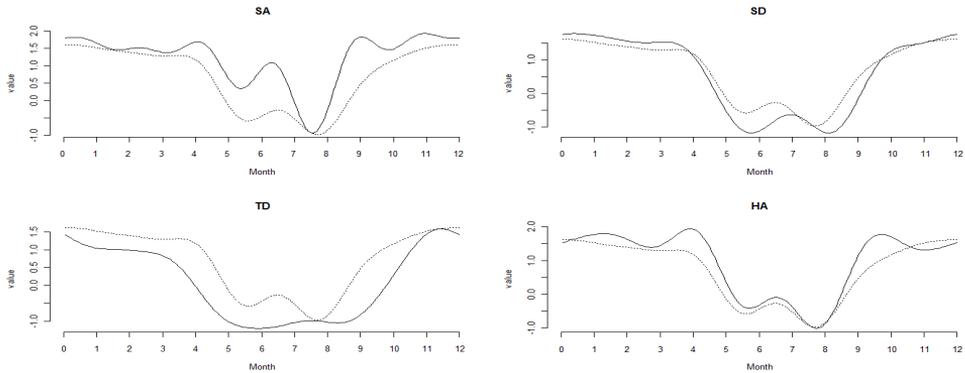

Figure 10. The estimated climate zone and humidity residual profiles $\mu(\cdot) + \alpha_{\mathrm{g}}(\cdot) + HumidRes(\cdot)$ for the logarithm of precipitation function in the functional analysis of covariance model (solid curves). The dashed curve is the Iranian mean function μ.
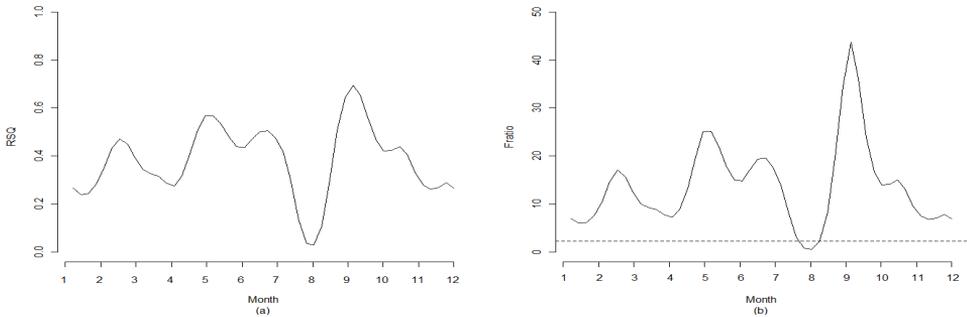
Figure 11. The left panel reveals the squared multiple correlation function RSQ and the right panel shows the corresponding F − ratio function FRATIO in the logarithm of preci− pitation based on climate zone and humidity residual. The horizontal dotted line indicates the 5% significance level for the F − distribution with 5 and 96 degrees of freedom.

5   Conclusion

When predicting temperature based on climate zone effects, it is 3 degrees C colder than Iranian average for climate zone SA in months 5-9. Temperature of climate zone SD is predicted to be 1 degree C warmer than the average in months 6-9. In the two other climate zones TD and HA, temperature is 10 degrees C warmer and 5 degrees C colder than the average, respectively. When predicting humidity based on climate zone effects, it is more than Iranian average for SA; from 5% in months 11-4 to 20% in months 4-10. Humidity in SD is less than the average; from 5% in months 9-10 to 13% in months 6-8. It is 2% less than the average for TD is months 7-8 up to 16% in months 4-5. Humidity of HA is predicted to be 4% and 7% more than the average in months 1-2 and months 4 and 9-10, respectively.

When predicting precipitation based on climate zone and temperature residual, it is more than Iranian average for SA; 1.4 millimeters in months 6-8 and 9-10, coincides in other months. Precipitation of SD is 0.7 millimeters less than average for months 9-10; 0.5 millimeters more than average for months 1-5, coincides for other months. Precipitation of TD is less than Iranian average; 1 millimeter for months 4-7 and 9-11, and 0.5 millimeters for other months. Precipitation for HA is more than the average; 0.7 millimeter in months 4 and 10-11, coincides in other months. The results of precipitation

prediction based on climate zone and humidity residual are the same.

Acknowledgement

References

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model, *Statistics & Probability Letters*, **45**, 11-22.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline Estimators for the functional linear model, *Statistica Sinica*, **13**, 571-591.

Hosseini-nasab, S.M.E., Kheirolah-zadeh, N. and Tazikeh Miyandarreh, N. (2008). Functional Modeling of Iranian Precipitation Based on Temperature and Humidity, *Journal of Statistical Research of Iran*, **5**, 95-121.

James, G.M. (2002), Generalized linear models with functional predictors, *J . Roy. Statist. Soc. Ser. B* **64**, 411-432.

Kaviani, M. and Alijani, B. (1999). *Basics of Climatology*. Payamnoor Publication.

Kavousi, A. and Meshkani, M.R. (2009). Spatial Analysis of Humidity and Temperature of Iran *Computational Science and Its Applications   Lecture Notes in Computer Science,* 2009, **5592**, 94-106.

Moosavi, M. (2011). An analysis to historic roots of climatic design in ancient architecture of central zone of Iran, *2nd International Conference on Humanities, Historical and Social Sciences IPEDR, IACSIT Press, Singapore* , **17**, 93-97.

Nazemosadat, M.J., Samani, N., Barry, D.A. and Molaii Niko, M. ENSO Forsing on Climate Change in Iran: Precipitation Analysis, *Iranian Journal of Science & Technology, Transaction B, Engineering,***30**, 555-565.

Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis*. Springer, New York.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd edition*. Springer, New York.

Tazikeh M., N. and Hosseini-nasab, M. (2007). Functional analysis of Iranian temperature and precipitation by using functional principal components analysis, *Journal of Statistical Research of Iran*, **4**, 109-128.

Zoglat, A. (2008). Functional analysis of variance, *Appl. Math. Sci.*, No. 21-24, 1115-1129.

S. M. E. Hosseini-nasab
Department of Statistics,
Faculty of Mathematical Sciences,
Shahid Beheshti University,
Tehran, Iran.
email: *m_hosseininasab@sbu.ac.ir*

N. Kheirollahzadeh
Department of Statistics,
Faculty of Mathematical Sciences,
Shahid Beheshti University,
Tehran, Iran.