Gümüşhane University Journal of Science and Technology

GUFBD / *GUJS* (2026) 16(1): 339-349 doi: 10.17714/gumusfenbil.1825537 Research Article

Uncovering regional climate patterns in Turkey using multivariate functional data analysis

Türkiye'deki bölgesel iklim örüntülerinin çok değişkenli fonksiyonel veri analizi ile ortaya çıkarılması

Çağlar SÖZEN*[1], Yüksel ÖNER[2]

[1]*Giresun University, Görele School of Applied Sciences, Department of Finance and Banking, 28800, Giresun, Türkiye*

[2]*Ondokuz Mayıs University, Faculty of Science, Department of Statistics, 55139, Samsun, Türkiye*

• Received: 17.11.2025 • Accepted: 06.03.2026

Abstract

This study investigates regional climate patterns in Turkey through multivariate Functional Data Analysis (FDA) using daily observations recorded at 81 meteorological stations over 2012–2018 (T = 2557 days). Five coupled climate variables—mean temperature, total precipitation, temperature range, evaporation, and sunshine duration—are transformed into smooth functional trajectories using Fourier basis expansions with roughness control, where smoothing levels are selected via generalized cross-validation (GCV). To summarize joint seasonal variability and cross-variable dependence, we apply multivariate functional principal component analysis (MFPCA) and obtain low-dimensional score representations for each station. These scores are then used in a model-based clustering framework, with the number of climate regimes selected by the Bayesian Information Criterion (BIC). Finally, functional ANOVA (FANOVA) procedures are employed to evaluate whether the identified regimes differ significantly for each variable on the functional scale. The results reveal distinct data-driven climate regimes and indicate that regime separation is driven primarily by temperature range, evaporation, and sunshine duration, whereas differences in mean temperature and precipitation are comparatively weaker. Overall, the proposed FDA pipeline provides an interpretable and statistically principled framework for regional climate classification and functional inference, offering useful insights for climate monitoring and environmental modeling in Turkey.

Keywords: Climate patterns, Fourier basis, Functional clustering, Functional data analysis, Turkey

Öz

Bu çalışma, Türkiye'deki bölgesel iklim örüntülerini çok değişkenli Fonksiyonel Veri Analizi (FVA/FDA) yaklaşımıyla incelemektedir. Analizde, 2012–2018 döneminde 81 meteoroloji istasyonunda kaydedilen günlük veriler kullanılmıştır (T = 2557 gün). Beş iklim değişkeni—günlük ortalama sıcaklık, günlük toplam yağış, günlük sıcaklık farkı (temperature range), buharlaşma miktarı ve güneşlenme süresi—pürüzlülük kontrollü Fourier baz genişletmeleri ile sürekli fonksiyonel eğrilere dönüştürülmüş; düzgünleştirme düzeyi genelleştirilmiş çapraz doğrulama (GCV) ile belirlenmiştir. Değişkenler arasındaki ortak mevsimsel yapı ve bağımlılık ilişkilerini düşük boyutlu biçimde özetlemek amacıyla çok değişkenli fonksiyonel temel bileşenler analizi (MFPCA) uygulanmış ve her istasyon için skor temsilleri elde edilmiştir. Bu skorlar üzerinden model tabanlı kümeleme gerçekleştirilmiş ve iklim rejimi sayısı BIC kriteri ile seçilmiştir. Son olarak, belirlenen rejimler arasında değişken bazında fonksiyonel farklılık olup olmadığını sınamak için fonksiyonel varyans analizi (FANOVA) testleri uygulanmıştır. Bulgular, Türkiye genelinde veri güdümlü iklim rejimlerinin ortaya çıktığını ve rejim ayrışmasının özellikle günlük sıcaklık farkı, buharlaşma ve güneşlenme süresi değişkenlerinde daha güçlü olduğunu; ortalama sıcaklık ve yağış değişkenlerinde ise farklılıkların görece daha zayıf kaldığını göstermektedir. Sonuç olarak, önerilen FVA tabanlı yaklaşım Türkiye'de bölgesel iklim sınıflandırması ve fonksiyonel çıkarım için yorumlanabilir ve istatistiksel olarak tutarlı bir çerçeve sunmaktadır.

Anahtar kelimeler: *İklim örüntüleri, Fourier baz, Fonksiyonel kümeleme, Fonksiyonel veri analizi, Türkiye*

*Çağlar SÖZEN; caglar.sozen@giresun.edu.tr

1. Introduction

Climate variables exhibit pronounced seasonality, strong temporal dependence, and substantial cross-variable co-variation (e.g., among temperature, precipitation, and sunshine). They are therefore more naturally viewed as dense trajectories than as independent scalar measurements. As a result, conventional multivariate methods—typically formulated for i.i.d. finite-dimensional vectors—may struggle to capture phase-specific seasonal behavior, amplitude differences, and coupled dynamics across multiple channels. Functional Data Analysis (FDA) provides an appropriate alternative by representing time-indexed observations as smooth functions, enabling inference directly on curves and offering interpretable low-dimensional summaries of complex temporal processes (Ramsay & Silverman, 2005; Ferraty & Vieu, 2006).

In this study, we investigate regional climate patterns in Turkey through multivariate FDA applied to five daily climate variables observed at $n = 81$ meteorological stations over 2012–2018 (one representative station per province, with the province used as a geographic label). Let $X_1(t)$ denote mean daily temperature, $X_2(t)$ daily total precipitation, $X_3(t)$ daily temperature range, $X_4(t)$ daily evaporation amount, and $X_5(t)$ daily sunshine duration. For each station, these variables are observed on a common daily index $t = 1, \dots, T$, where $T = 2557$ is the total number of days from 2012–2018 inclusive (including leap days). Casting these measurements as smooth functional trajectories places station-level climate records on a common functional scale and supports a characterization of regional differences in terms of seasonal dynamics rather than coarse seasonal aggregates (Ramsay & Silverman, 2002). Our analysis addresses how multivariate climate trajectories vary across Turkey and which climate variables exhibit statistically significant functional differences among data-driven regimes.

Our analysis follows a widely used FDA pipeline tailored to complex temporal data. First, daily observations are transformed into smooth functional objects via basis expansions with roughness control; smoothing levels are selected in a data-driven manner (e.g., generalized cross-validation) to balance fidelity and smoothness (Craven & Wahba, 1979; Abraham et al., 2003; Ramsay & Silverman, 2005). Second, we employ multivariate functional principal component analysis (MFPCA) jointly across the five functional variables to capture cross-channel dependence in the dominant modes of variation and to obtain low-dimensional score vectors summarizing each station's joint climate profile (Boullé, 2012; Happ & Greven, 2018). Third, we perform model-based clustering on these score vectors to discover groups of stations with similar joint dynamics, selecting the number of clusters using likelihood-based criteria such as BIC and reporting posterior membership probabilities to quantify classification uncertainty (Bouveyron & Jacques, 2011; Jacques & Preda, 2014b). Finally, to move beyond descriptive regime discovery, we incorporate functional hypothesis testing (FANOVA) to assess, variable-by-variable, whether $X_1(t), \dots, X_5(t)$ differ significantly across the inferred regimes (Shen & Faraway, 2004; Zhang, 2013; Zhang et al., 2018).

A key modeling choice is to cluster stations using MFPCA score representations rather than clustering directly on the raw functional objects. This choice is motivated by interpretability, statistical stability in high-dimensional multivariate settings, and principled uncertainty quantification via likelihood-based clustering and posterior membership probabilities (Bouveyron & Jacques, 2011; Jacques & Preda, 2014a, 2014b; Happ & Greven, 2018). More broadly, modern syntheses emphasize the practical value of the workflow "functional smoothing →(M)FPCA →model-based clustering on scores" when the applied goal is regime discovery with model selection and uncertainty reporting (Sözen, 2019; Zhang & Parnell, 2023).

FDA has become increasingly prominent in climate and meteorology because functional representations respect the continuous nature of annual cycles and avoid imposing rigid, predefined "seasons" (Ramsay & Silverman, 2005; Ferraty & Vieu, 2006). Turkey-focused evidence illustrates how Fourier-based smoothing and FPCA can reveal interpretable seasonal modes and enable meaningful regional comparisons in temperature trajectories (Sözen & Öner, 2022). Positioned within this literature, our contribution is to provide a country-scale multivariate functional characterization of Turkey's climate regimes using 81 station locations and five coupled daily climate functions, and to complement probabilistic regime discovery with variable-wise functional inference via FANOVA, thereby identifying which climate variables drive statistically significant regime differences (Shen & Faraway, 2004; Zhang, 2013; Zhang et al., 2018). By combining probabilistic regime discovery with functional inference, the analysis identifies not only climate regimes but also the specific functional drivers of regime separation.

The remainder of the paper is organized as follows. Section 2 presents the methodology, including Fourier basis smoothing with GCV-based selection, multivariate functional principal component analysis, model-based clustering with BIC-based model selection, and FANOVA procedures. Section 3 reports the empirical results and discusses the identified regional climate regimes. Section 4 concludes with a summary of the main findings, practical implications, limitations, and directions for future research.

2. Methodology

This study applies Functional Data Analysis (FDA) to daily climate records observed at N = 81 meteorological stations across Turkey over 2012–2018 (one representative station per province; province names are used only as geographic labels). For each station $i = 1, \dots, N$, we observe five daily climate variables on a common calendar index t_j $(j = 1, \dots, T)$ with $T = 2557$ total days: mean temperature ($\ell = 1$), total precipitation ($\ell = 2$), temperature range ($\ell = 3$), evaporation ($\ell = 4$), and sunshine duration ($\ell = 5$). The analysis proceeds through four steps: (i) smoothing discrete measurements into functional trajectories, (ii) summarizing multivariate dependence via multivariate FPCA (MFPCA), (iii) discovering climate regimes via model-based clustering on the score vectors, and (iv) testing regime differences for each variable using functional ANOVA (FANOVA) (Ramsay & Silverman, 2005; Ferraty & Vieu, 2006; Sözen, 2019).

2.1. Functional data representation and model formulation

Let $y_{i\ell}(t_j)$ denote the observed value of variable $\ell \in \{1, \dots, 5\}$ at station $i$on day t_j. We adopt the standard measurement-error model (Ramsay & Silverman, 2005):

$$y_{i\ell}(t_j) = x_{i\ell}(t_j) + \varepsilon_{i\ell j}, \mathbb{E}[\varepsilon_{i\ell j}] = 0, \mathrm{Var}(\varepsilon_{i\ell j}) = \sigma_\ell^2 \tag{1}$$

Here $x_{i\ell}(t)$ is an underlying smooth function defined on the daily time domain. FDA treats $\{x_{i\ell}(t)\}$ as functional objects, enabling inference on seasonal shape, phase shifts, and amplitude differences that are not well captured by scalar summaries (Ferraty & Vieu, 2006).

2.2. Fourier basis function approximation

Each trajectory $x_{i\ell}(t)$ is represented by a basis expansion (Ulbricht, 2004):

$$x_{i\ell}(t) = \sum_{k=1}^{K} c_{i\ell k}\ \phi_k(t) \tag{2}$$

We use a Fourier basis $\{\phi_k\}_{k=1}^{K}$, which is well-suited for periodic/seasonal behavior. A standard Fourier system over an interval of length L can be written as:

$$\phi_1(t) = 1, \phi_{2r}(t) = \sin\left(\frac{2\pi r}{L}t\right), \phi_{2r+1}(t) = \cos\left(\frac{2\pi r}{L}t\right), K = 2R + 1 \tag{3}$$

Coefficients $c_{i\ell} = (c_{i\ell 1}, \dots, c_{i\ell K})^{\mathrm{T}}$ are estimated by penalized least squares with a roughness penalty to prevent overfitting (Ramsay & Silverman, 2005):

$$\hat{\mathrm{c}}_{i\ell}(\lambda) = argmin_{c \in R^K} \left|\left|y_{i\ell} - \Phi c\right|\right|^2 + \lambda c^{\mathrm{T}} R c \tag{4}$$

where $\mathrm{y}_{i\ell} = (y_{i\ell}(t_1), \dots, y_{i\ell}(t_T))^{\mathrm{T}}$, $\Phi_{jk} = \phi_k(t_j)$, and R is a penalty matrix (commonly based on integrated squared second derivatives). The smoothed functional estimate is $\hat{x}_{i\ell}(t) = \sum_{k=1}^{K} \hat{c}_{i\ell k}\, \phi_k(t)$.

2.3. Determining the smoothing parameter

Let $\hat{\mathrm{y}}_{i\ell}(\lambda) = \mathrm{S}(\lambda)\mathrm{y}_{i\ell}$ be the fitted values, where the smoother matrix has the standard form (Craven & Wahba, 1979; Ramsay & Silverman, 2005):

$$S(\lambda) = \Phi\,(\Phi^{\top}\Phi + \lambda R)^{-1}\Phi^{\top} \tag{5}$$

We choose λ by generalized cross-validation (GCV):

$$GCV(\lambda) = \frac{T \;\| \mathrm{y}_{i\ell} - \hat{\mathrm{y}}_{i\ell}(\lambda) \|^2}{(T - \mathrm{tr}(S(\lambda)))^2} \tag{6}$$

selecting $\hat{\lambda}$ that minimizes $GCV(\lambda)$ over a candidate grid (Craven & Wahba, 1979). In the implementation, we evaluate a log-scale grid (e.g., $\log_{10}\lambda \in \{-5, -4, \ldots, 5\}$), and report the corresponding GCV outcomes by variable.

2.4. Multivariate FPCA and model-based clustering on score vectors

2.4.1. Multivariate functional principal component analysis (MFPCA)

For each station i, define the smoothed multivariate function $\hat{\mathrm{x}}_i(t) = (\hat{x}_{i1}(t), \ldots, \hat{x}_{i5}(t))$. MFPCA captures cross-variable dependence through multivariate eigenfunctions and low-dimensional score vectors (Happ & Greven, 2018). We use the truncated representation:

$$\hat{x}_{i\ell}(t) \approx \mu_{\ell}(t) + \sum_{r=1}^{d} \xi_{ir}\; \psi_{r\ell}(t), \ell = 1, \ldots, 5 \tag{7}$$

where $\mu_{\ell}(t)$ is the mean function, $\psi_{r\ell}(t)$ are the multivariate principal component functions, and $\xi_i = (\xi_{i1}, \ldots, \xi_{id})^{\top} \in \mathbb{R}^d$ are the MFPCA scores summarizing each station's joint climate profile. The dimension d is chosen to explain a high proportion of the joint variability (Happ & Greven, 2018).

2.4.2. Model-based clustering on MFPCA scores

We cluster stations using $\{\xi_i\}_{i=1}^{N}$ via a Gaussian mixture model (Tarpey & Kinateder, 2003; Bouveyron & Jacques, 2011; Jacques & Preda, 2014b):

$$p(\xi) = \sum_{g=1}^{G} \pi_g\; \varphi_d(\xi; \mu_g, \Sigma_g) \tag{8}$$

where G is the number of clusters (climate regimes), $\pi_g > 0$ with $\sum_{g=1}^{G} \pi_g = 1$, and $\varphi_d(\xi_i; \mu_g, \Sigma_g)$ denotes the d-variate Gaussian density. Parameters are estimated by the EM algorithm, and G is selected using BIC (Bouveyron & Jacques, 2011; Jacques & Preda, 2014b). Classification uncertainty is quantified via posterior membership probabilities:

$$\tau_{ig} = P(Z_i = g \mid \xi_i) = \frac{\pi_g\; \varphi_d(\xi_i; \mu_g, \Sigma_g)}{\sum_{h=1}^{G} \pi_h\; \varphi_d(\xi_i; \mu_h, \Sigma_h)} \tag{9}$$

This "smoothing →(MF)PCA →model-based clustering on scores" pipeline is widely used for interpretable regime discovery with probabilistic assignments (Happ & Greven, 2018; Jacques & Preda, 2014b).

2.5. FANOVA: testing functional differences across inferred regimes

After clustering, we test whether regimes differ significantly in each climate variable's mean function. For each variable ℓ, let $\mathcal{I}_g = \{i : Z_i = g\}$ denote the set of stations assigned to regime g, with $n_g = |\,\mathcal{I}_g\,|$ and $\sum_{g=1}^{G} n_g = N$. Define regime-specific and overall mean functions:

$$\bar{x}_{g\ell}(t) = \frac{1}{n_g}\sum_{i\in\mathcal{I}_g} x_{i\ell}(t), \bar{x}_{\cdot\ell}(t) = \frac{1}{N}\sum_{i=1}^{N} x_{i\ell}(t) \tag{10}$$

We consider the one-way functional ANOVA null hypothesis (Shen & Faraway, 2004; Zhang, 2013):

$$H_{0,\ell}: \bar{x}_{1\ell}(t) \equiv \bar{x}_{2\ell}(t) \equiv \cdots \equiv \bar{x}_{G\ell}(t) \text{ for all } t \in T \tag{11}$$

2.5.1. Pointwise F-test (PF)

At each t, define between- and within-regime sums of squares:

$$SSB_\ell(t) = \Sigma_{g=1}^{G} n_g \left(\bar{x}_{g\ell}(t) - \bar{x}_{\cdot\ell}(t)\right)^2, SSE_\ell(t) = \Sigma_{g=1}^{G}\Sigma_{i\in I_g}\left(x_{i\ell}(t) - \bar{x}_{g\ell}(t)\right)^2 \tag{12}$$

The pointwise F statistic is:

$$F_\ell(t) = \frac{SSB_\ell(t)/(G-1)}{SSE_\ell(t)/(N-G)} \tag{13}$$

Pointwise testing highlights time-localized differences, while global tests below assess overall regime separation (Shen & Faraway, 2004; Zhang, 2013). In reporting, a common summary is $PF_\ell = \sup_{t\in T} F_\ell(t)$ with calibration as described below.

2.5.2. L_2-norm (global) test

A global measure integrates between-regime variation over the domain (Zhang, 2013):

$$T_\ell = \int_T SS\,B_\ell(t)\,dt \tag{14}$$

In practice, we use the L_2-norm based approximations (naive and bias-reduced) and bootstrap calibration following the recommendations in the functional ANOVA literature (Shen & Faraway, 2004; Zhang, 2013; Zhang et al., 2018).

2.5.3. F-type global test

An F-type functional statistic compares integrated between- and within-regime variation (Zhang et al., 2018):

$$F_\ell^{glob} = \frac{\int_T SS\,B_\ell(t)\,dt/(G-1)}{\int_T SS\,E_\ell(t)\,dt/(N-G)} \tag{15}$$

2.5.4. Bootstrap calibration

Because functional data may deviate from Gaussian assumptions, we use bootstrap calibration to obtain p-values (Zhang, 2013; Zhang et al., 2018). Briefly, residual functions are constructed under H_0 by centering within regimes, resampled to generate bootstrap replicates, and the test statistics $\left(T_\ell, F_\ell^{glob}, PF_\ell\right)$ are recomputed; the p-value is the proportion of bootstrap replicates exceeding the observed statistic.

3. Results and discussion

We analyze multivariate functional climate trajectories derived from daily observations recorded by the Turkish State Meteorological Service (TSMS) over 2012–2018 at $N = 81$ meteorological stations (one representative station per province; province names are used only as geographic labels) (TSMS, 2018). The five variables are mean daily temperature $X_1(t)$, daily total precipitation $X_2(t)$, daily temperature range $X_3(t)$, daily evaporation amount $X_4(t)$, and daily sunshine duration $X_5(t)$, observed on a common daily index $t =$

$t_1, \ldots, t_T$ with $T = 2557$. Following the FDA workflow described in Section 2, we (i) smooth each station-level series using Fourier basis expansions with roughness control, (ii) summarize joint multivariate variability via MFPCA, (iii) cluster stations using model-based clustering on MFPCA score vectors with BIC-based selection, and (iv) assess regime differences using FANOVA procedures.

For each station $i = 1, \ldots, N$ and each variable $\ell = 1, \ldots, 5$, the discrete observations $y_{i\ell}(t_j)$ are converted into smooth functional trajectories $\hat{x}_{i\ell}(t)$ via Fourier basis smoothing with a roughness penalty, where the smoothing parameter λ is selected by GCV. Table 1 reports the GCV criterion evaluated over the candidate grid $\log_{10}(\lambda) \in \{-5, -4, \ldots, 5\}$. The selected $\hat{\lambda}$ values (grid minimizers) yield stable functional representations across stations and variables, supporting the subsequent MFPCA and clustering steps.

Table 1. GCV criterion over the candidate smoothing-parameter grid for Fourier functional representations

	Values	$X_1(t)$	$X_2(t)$	$X_3(t)$	$X_4(t)$	$X_5(t)$
GCV	-5	1001.76	56.05391	1160.777	482.1474	894.0939
	-4	997.8755	55.83885	1156.309	480.2832	890.6692
	-3	971.9769	54.4874	1127.672	468.0453	869.3652
	-2	910.8631	51.67523	1065.507	439.8923	826.0291
	-1	857.1082	49.49375	1015.771	415.0447	792.9708
	0	826.9709	48.41473	991.0959	399.7455	778.4261
	1	816.5622	48.08578	984.7931	391.4655	**776.7679**
	2	**816.0434**	**48.03698**	**984.1055**	387.3497	781.1089
	3	820.5284	48.09322	988.8072	**386.5378**	789.639
	4	825.4053	48.19835	993.9249	387.4616	797.0927
	5	829.75	48.33542	996.1273	389.0577	776.7679

We apply MFPCA jointly to the five-variable functional vectors $\{\hat{x}_{i\ell}(t)\}_{\ell=1}^{5}$ to capture dominant coupled modes of seasonal variability and to obtain low-dimensional score vectors $\xi_i \in \mathbb{R}^d$ summarizing each station's joint climate profile. Figure 1 displays the leading multivariate component functions. The first component represents broad seasonal amplitude variation, exhibiting larger loadings for temperature range, evaporation, and sunshine duration, while the second component captures phase-shifted deviations and cross-variable compensation effects. This pattern is consistent with the annual cycle structure of Turkey's coupled temperature–radiation–moisture processes, with precipitation showing more irregular dynamics.

Using the MFPCA scores $\{\xi_i\}_{i=1}^{N}$, we fit Gaussian mixture models and select the number of clusters by BIC. Table 2 reports BIC values across candidate cluster counts G. The largest BIC is attained at $G = 3$, indicating three data-driven climate regimes among the station locations.

Table 2. Simulation of determining the number of clusters

G	BIC
2	-895676.97
3	**-852432.9**
4	-882699.97
5	-893239.81
6	-900501.54
7	-915744.96
8	-920961.45
9	-927554.87
10	-930631.85

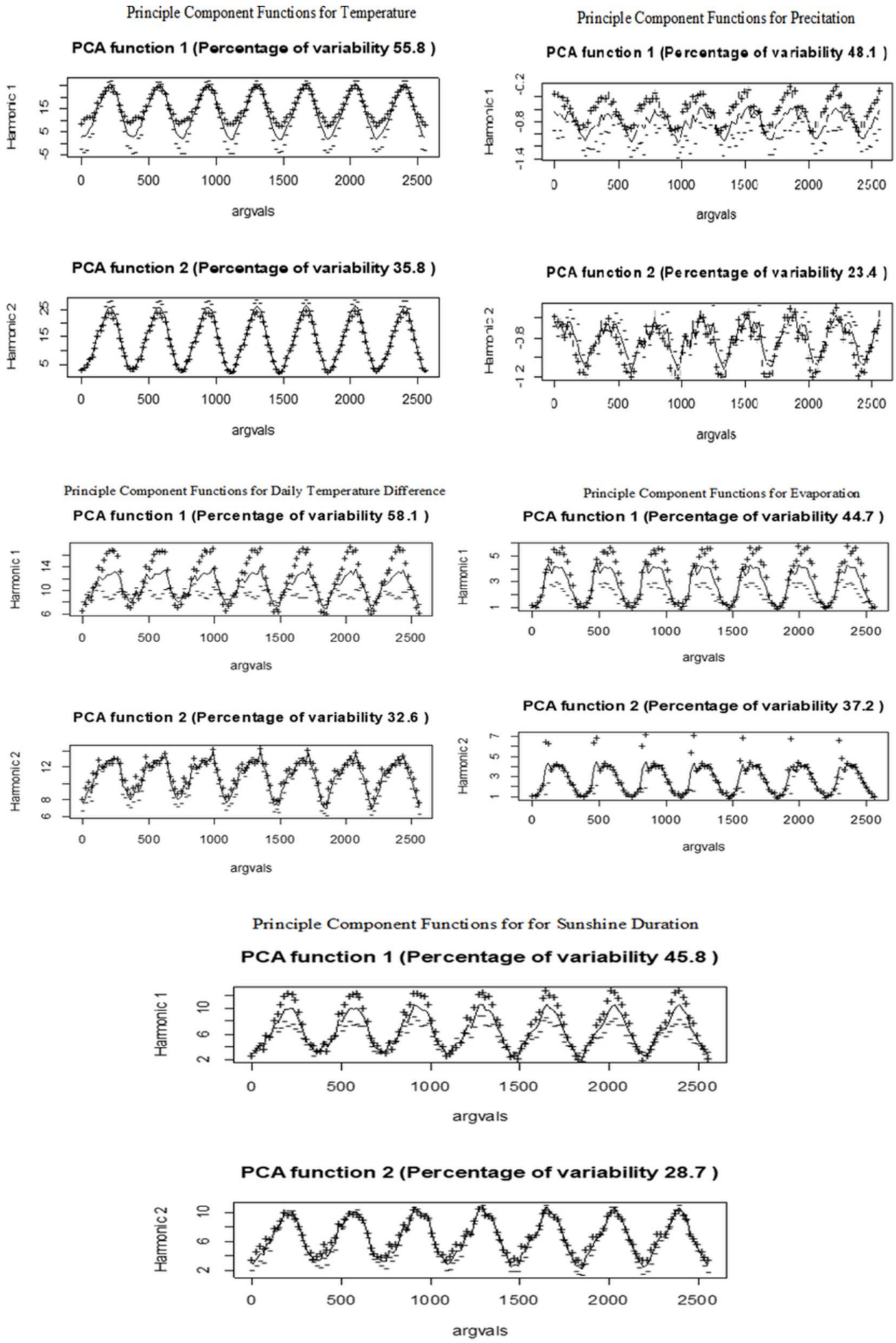

Figure 1. Principal component functions for climate variables

Regime composition is summarized in Table 3, and the corresponding mixing proportions $\hat{\pi}_g = n_g/N$ are reported in Table 4. Regime 2 contains the majority of stations ($\hat{\pi}_2 \approx 0.691$), while regime 3 accounts for $\hat{\pi}_3 \approx 0.284$. Regime 1 is small ($\hat{\pi}_1 \approx 0.025$), containing two stations (Kars and Çanakkale), suggesting an atypical joint seasonal profile rather than geographic proximity as the primary driver of this grouping.

Table 3. Model-based clustering results

Clusters	1	Kars, Çanakkale
	2	Erzincan, Bingöl, Malatya, Artvin, Giresun, Ordu, Rize, Trabzon, Amasya, Bartın, Çankırı, Çorum, Kastamonu, Samsun, Sinop, Tokat, Zonguldak, Bilecik, Bolu, Düzce, Kocaeli, Sakarya, Yalova, Edirne, Kırklareli, Tekirdağ, İstanbul, Afyonkarahisar, Denizli, İzmir, Kütahya, Manisa, Uşak, Ankara, Karaman, Konya, Aksaray, Kayseri, Kırıkkale, Nevşehir, Niğde, Sivas, Yozgat, Burdur, Isparta, Kahramanmaraş, Adıyaman, Batman, Diyarbakır, Gaziantep, Kilis, Mardin, Siirt, Şanlıurfa, Şırnak
	3	Ağrı, Ardahan, Erzurum, Bayburt, Iğdır, Bitlis, Elazığ, Hakkâri, Muş, Tunceli, Van, Gümüşhane, Karabük, Bursa, Eskişehir, Adana, Antalya, Muğla, Balıkesir, Aydın, Hatay, Mersin, Osmaniye

Table 4. Cluster probabilities of cities

Clustering	**1**	**2**	**3**
Probability	0.024691	0.691358	0.283951

To visualize regime separation on the functional scale, Figure 2 plots the regime-wise mean functions $\bar{x}_{g\ell}(t)$ for each variable ℓ and regime $g \in \{1,2,3\}$. Differences across regimes are most pronounced for temperature range $X_3(t)$, evaporation $X_4(t)$, and sunshine duration $X_5(t)$, while mean temperature $X_1(t)$ and precipitation $X_2(t)$ show more similar average profiles.

Table 5 reports the FANOVA test statistics and p-values for assessing whether the inferred climate regimes differ on the functional scale across variables. The results show that regime separation is mainly driven by temperature range, evaporation, and sunshine duration. Among these, evaporation provides the strongest evidence of regime differences, followed by sunshine duration, while temperature range shows a weaker but still consistent signal. In contrast, mean temperature and precipitation exhibit comparatively limited separation across regimes. Overall, the findings suggest that the identified regimes are differentiated more by variability and energy-related dynamics than by level-type climatic features.

Table 5. FANOVA test statistic results and p values according to variables

Tests	**Variables**				
	$\boldsymbol{X_1(t)}$	$\boldsymbol{X_2(t)}$	$\boldsymbol{X_3(t)}$	$\boldsymbol{X_4(t)}$	$\boldsymbol{X_5(t)}$
$\boldsymbol{L_2}$**N**	T_ℓ =209589.1	T_ℓ = 3476.598	T_ℓ =129841.9	T_ℓ =138542.8	T_ℓ =60409.82
	$\hat{\beta}_\ell$= 22043.98 $\hat{d}_\ell$= 4.577366	$\hat{\beta}_\ell$= 59.80377 $\hat{d}_\ell$ = 52.67402	$\hat{\beta}_\ell$= 6571.005 $\hat{d}_\ell$= 10.90537	$\hat{\beta}_\ell$= 1204.793 $\hat{d}_\ell$ = 46.97927	$\hat{\beta}_\ell$= 2180.092 $\hat{d}_\ell$= 13.79395
	p_ℓ =0.0712876	p_ℓ =0.2812835	p_ℓ =0.04688431	p_ℓ = $1.248358e^{-7}$	p_ℓ = 0.01421683
$\boldsymbol{L_2}$**B**	T_ℓ =209589.1	T_ℓ =3476.598	T_ℓ =129841.9	T_ℓ =138542.8	T_ℓ =60409.82
	$\hat{\beta}_\ell$=21133.17 $\hat{d}_\ell$=4.723361	$\hat{\beta}_\ell$=39.12208 $\hat{d}_\ell$=80.46857	$\hat{\beta}_\ell$=6036.247 $\hat{d}_\ell$ =11.82021	$\hat{\beta}_\ell$=831.5826 $\hat{d}_\ell$=68.01206	$\hat{\beta}_\ell$=1962.804 $\hat{d}_\ell$ =15.26971
	p_ℓ =0.06626803	p_ℓ =0.2443702	p_ℓ =0.04034888	p_ℓ =$3.033204e^{-10}$	p_ℓ =0.01057928
$\boldsymbol{L_2}$**b**	T_ℓ^* =209589.1	T_ℓ^* =3476.598	T_ℓ^* =129841.9	T_ℓ^* =138542.8	T_ℓ^* =60409.82
	p_ℓ = 0.607	p_ℓ = 0.2165	p_ℓ =0.7386	p_ℓ =0.3374	p_ℓ =0.1265
FT	F_ℓ^{glob} =2.077127	F_ℓ^{glob} =1.103645	F_ℓ^{glob} =1.811936	F_ℓ^{glob} =2.44774	F_ℓ^{glob} =2.008834
	p_ℓ = 0.07639064	p_ℓ = 0.2844771	p_ℓ =0.05046737	p_ℓ = $2.38694e^{-7}$	p_ℓ = 0.01604624
Fb	$F_\ell^{glob,*}$ =2.077127	$F_\ell^{glob,*}$ =1.103645	$F_\ell^{glob,*}$ =1.811936	$F_\ell^{glob,*}$ =2.44774	$F_\ell^{glob,*}$ =2.008834
	p_ℓ =0.7615	p_ℓ =0.3313	p_ℓ =0.8918	p_ℓ =0.7856	p_ℓ =0.1741
PF	PF_ℓ =1.94262	PF_ℓ =1.091962	PF_ℓ =2.125067	PF_ℓ =2.723363	PF_ℓ =1.345824
	$\hat{\beta}_\ell$ = 0.2053331 $\hat{d}_\ell$ = 4.998297	$\hat{\beta}_\ell$= 0.02039162 $\hat{d}_\ell$ = 50.33027	$\hat{\beta}_\ell$ = 0.07188914 $\hat{d}_\ell$ = 14.27637	$\hat{\beta}_\ell$= 0.0201402 $\hat{d}_\ell$ = 50.95856	$\hat{\beta}_\ell$= 0.04550941 $\hat{d}_\ell$ = 22.55173
	p_ℓ =0.09195204	p_ℓ =0.3518099	p_ℓ =0.009936569	p_ℓ =$1.458928e^{-9}$	p_ℓ =0.1467945

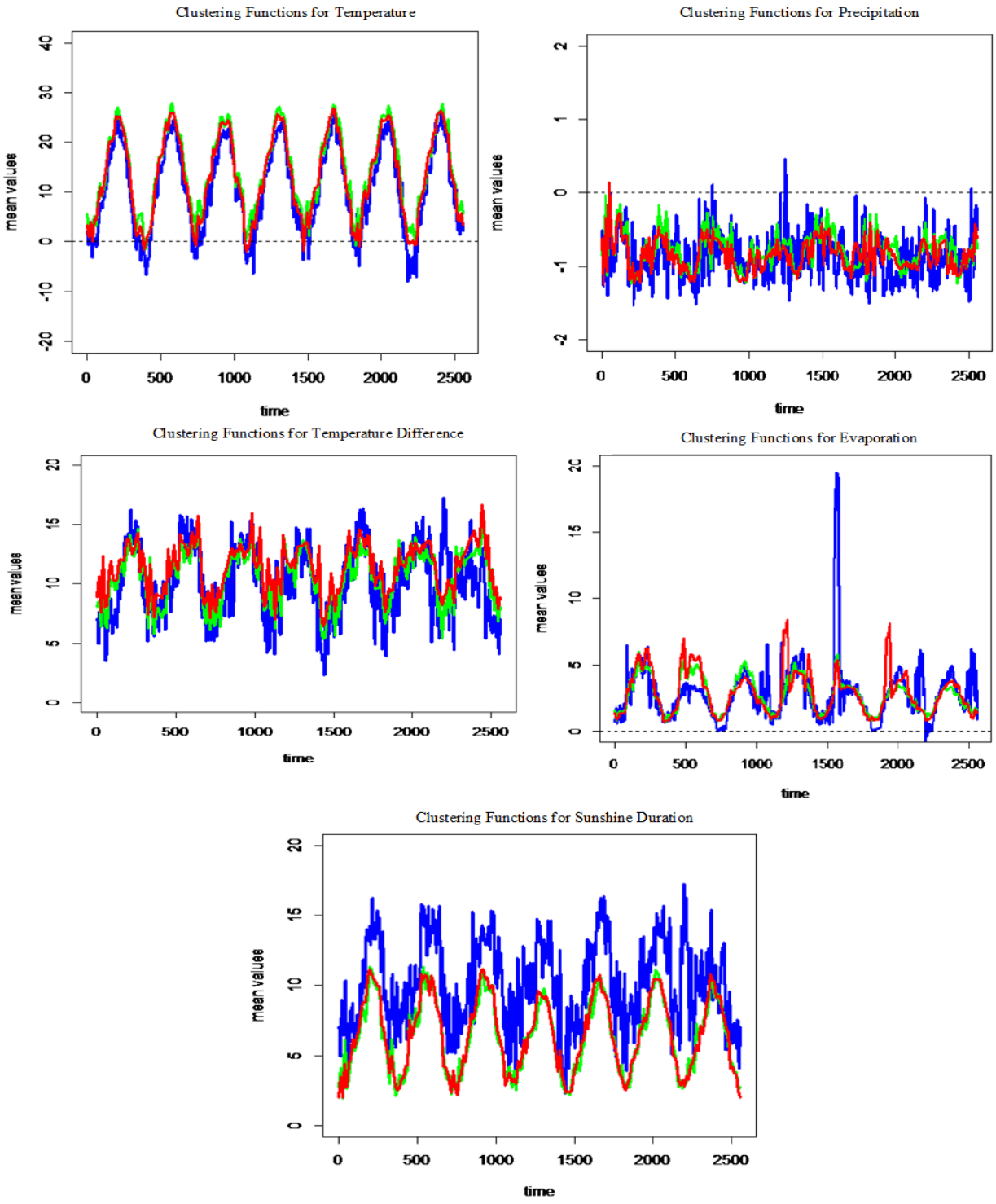

Figure 2. Clustering functions for climate variables

4. Conclusion

This study investigated regional climate patterns in Turkey using a multivariate Functional Data Analysis framework applied to daily observations from 81 meteorological stations over 2012–2018. Five coupled climate variables—mean temperature, precipitation, temperature range, evaporation, and sunshine duration—were transformed into smooth functional trajectories through Fourier basis expansions with roughness control, where smoothing levels were selected by generalized cross-validation. Multivariate functional principal component analysis was then used to summarize joint seasonal variability and cross-variable dependence in a

low-dimensional score representation, enabling statistically stable regime discovery via model-based clustering with BIC-based selection.

The results support the presence of three data-driven climate regimes across Turkey with unequal cluster sizes, indicating that functional regime separation is driven more by systematic differences in seasonal dynamics than by simple geographic proximity. FANOVA-based inference suggests that regime differences are strongest for evaporation and sunshine duration, and are also present—though more moderately—for temperature range. In contrast, mean temperature and precipitation exhibit weaker evidence of functional separation across the inferred regimes. Bootstrap-calibrated global tests provide a more conservative assessment of regime differences, underscoring the importance of calibration choice in functional inference under heterogeneous variability.

Overall, the proposed multivariate FDA pipeline offers an interpretable and robust framework for regional climate classification and functional hypothesis testing. By leveraging smooth functional representations and multivariate dependence structure, the approach yields insights into regime-specific seasonal dynamics that are not easily captured by traditional scalar summaries. Future work may extend the framework by incorporating spatial dependence, covariate-adjusted functional clustering, and time-varying regime dynamics to further support climate monitoring and environmental decision-making in Turkey.

Acknowledgement

This article was derived from the first author Çağlar SÖZEN's doctoral dissertation completed at Ondokuz Mayıs University, Department of Statistics.

Declaration of ethical code

The authors of this article declare that the materials and methods used in this study do not require ethical committee approval and/or legal-specific permission.

Conflicts of interest

The authors declare that they have no conflict of interest.

References

Abraham, C., Cornillon, P. A., Matzner-Løber, E., & Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics, 30*(3), 581–595.

Boullé, M. (2012). Functional data clustering via piecewise constant nonparametric density estimation. *Pattern Recognition, 45*(12), 4389–4401.

Bouveyron, C., & Jacques, J. (2011). Model-based clustering of time series in group specific functional subspaces. *Advances in Data Analysis and Classification, 5*(4), 281–300.

Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik, 31*, 377–403.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer.

Happ, C., & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, *113*(522), 649-659.

Jacques, J., & Preda, C. (2014a). Functional data clustering: A survey. *Advances in Data Analysis and Classification, 8*(3), 231–255.

Jacques, J., & Preda, C. (2014b). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis, 71*, 92–106.

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. Springer.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer.

Shen, Q., & Faraway, J. (2004). An F-test for linear models with functional responses. *Statistica Sinica, 14*(4), 1239–1257.

Sözen, Ç. (2019). *Investigation of multivariate functional data with clustering methods: The case of Turkey climate* (Doctoral dissertation, Ondokuz Mayıs University, Department of Statistics).

Sözen, Ç., & Öner, Y. (2022). The investigation of temperature data in Turkey's Black Sea Region using functional data analysis. *Journal of Applied Statistics, 49*(9), 2403–2415.

Tarpey, T., & Kinateder, K. (2003). Clustering functional data. *Journal of Classification, 20*(1), 93–114.

Turkish State Meteorological Service. (2018). *Daily mean temperature, daily total precipitation, daily temperature range, daily sunshine duration, and daily evaporation*. Ankara, Turkey.

Ulbricht, J. (2004). *Representing functional data as smooth functions* (Master's thesis, Humboldt University of Berlin, Institute of Statistics and Econometrics).

Zhang, J. T. (2013). *Analysis of variance for functional data* (Monographs on Statistics and Applied Probability 127).

Zhang, J. T., Cheng, M. Y., Wu, H. T., & Zhou, B. (2018). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis*.

Zhang, M., & Parnell, A. (2023). Review of clustering methods for functional data. *ACM Transactions on Knowledge Discovery from Data*, *17*(7), 1-34.