Smoothed functional canonical correlation analysis of humidity and temperature data

Istem Koymen Keser* and Ipek Deveci Kocakoç

Econometrics, Dokuz Eylul University, Dokuz Eylul Univ. Ikt. Id. Bil. Fak., Dokuzçeşmeler, Buca, İzmir, Turkey

(Received 28 January 2013; accepted 12 February 2015)

This paper focuses on smoothed functional canonical correlation analysis (SFCCA) to investigate the relationships and changes in large, seasonal and long-term data sets. The aim of this study is to introduce a guideline for SFCCA for functional data and to give some insights on the fine tuning of the methodology for long-term periodical data. The guidelines are applied on temperature and humidity data for 11 years between 2000 and 2010 and the results are interpreted. Seasonal changes or periodical shifts are visually studied by yearly comparisons. The effects of the ‘number of basis functions’ and the ‘selection of smoothing parameter’ on the general variability structure and on correlations between the curves are examined. It is concluded that the number of time points (knots), number of basis functions and the time span of evaluation (monthly, daily, etc.) should all be chosen harmoniously. It is found that changing the smoothing parameter does not have a significant effect on the structure of curves and correlations. The number of basis functions is found to be the main effector on both individual and correlation weight functions.

Keywords: functional data analysis; smoothing; functional canonical correlation analysis; meteorological data

1. Introduction

Most of the statistical methods include one or more observations taken from each of the individuals in a sample. As the area of interest gets broader, these observations take the form of curves or surfaces. These observed curves or surfaces are called as the functional data. It is assumed that these data which are observed on discrete points come from an underlying real function. In most cases, observations are a function of time or a closely related variable. As the capacity of tools and computers used for data collection and storage increase every day, statistical methods for analyzing functional data need to be improved. Methods for analyzing functional data are referred to as ‘functional data analysis (FDA)’ by Ramsay and Dalzell [17]. Studies in the subject continue extensively in many fields after the seminal work of Ramsay and Silverman

*Corresponding author. Email: istem.koymen@deu.edu.tr

© 2015 Taylor & Francis
The overall aim of FDA is to explain the nature of the data while focusing on the trajectories and shapes. Historical progress of functional canonical correlation analysis (FCCA) can be found in [15].

The main aim of this study is to provide a guideline for smoothed FCCA while dealing with large, seasonal and long-term data sets with an application on humidity and temperature data. There are many studies [2–8,21–23] that use seasonal and meteorological data in order to develop methodology on FDA. However, investigations were mostly carried out on the basis of the whole time span of the data. In this study, the FDA is utilized for making ‘annual’ comparisons for 35 different weather stations in Turkey. We also searched for and reported the optimal values for the number of basis functions (K) and smoothing parameters. Another aim of the study is to uncover highly correlated relationships between temperature and humidity by investigating individual functions and the correlations between the two curve sets of temperature and humidity and compare effects of land forms and climates on this relationship. We also investigated whether there are seasonal changes or periodical shifts in terms of temperature and humidity for 11 years between 2000 and 2010.

The paper is organized as follows: Section 2 presents an algorithmic guideline for applying FDA and smoothed functional canonical correlation analysis (SFCCA) and gives some suggestions for the application process. In Section 3, the SFCCA of humidity and temperature data are given with detailed results and interpretations. Comments on the effect of fine tuning of basis function number and smoothing parameter are also made. The results for individual and weight functions for 11 years are compared and summarized annually. Finally, Section 4 deals with some conclusions and suggestions.

2. Guidelines for applying SFCCA to large, seasonal and long-term data

By the insights of this extensive study on the data, on the literature and our previous experiences, the following guidelines are suggested for researchers who want to apply FDA and SFCCA to their long-term seasonal data. If the data set is small, a visual examination of the correlation surface will give some information about the variation and interaction of the curves. However, for large data sets, the surface will be very complicated to interpret. For example, for annual data, the correlation surface will have $365 \times 365$ data points and visual examination will be impossible. Therefore, SFCCA can be utilized in order to determine modes of variability in the data. There are some decisions to be made here such as the selection of basis function, number of basis functions, period of the data, smoothing approach, value of the smoothing parameters for individual and weight functions. The technical procedure is given below and the details of application are given in the next section.

Step 1 – Data preparation: After the data are cleared; missing values are imputed, deleted or retained as missing, it should be arranged into a data matrix of $n \times m$ observations; $n$ being the number of argument values (knots or time points) and $m$ being the number of subjects (person, city, station, etc.). Since FDA can accommodate missing values by letting the use of different knots, imputation is not compulsory. If you want to see the seasonality in the long-term data, we recommend dividing the data set into terms (years, months, etc.) and analyzing each term separately.

Step 2 – Basis function selection: As the first step of FDA, discrete data points are converted to continuous functions by the basis function approach. Various basis functions such as B-splines, polynomials, wavelets and Fourier basis can be used. Fourier basis functions are the most suitable ones to model periodical data due to their sinusoidal structure. This is especially valid for long-term data. B-splines also give useful results for yearly data. In order to define a basis system, the number of basis functions ($K$) and period ($T$) need to be determined. According to Ramsay and Silverman [20], we should choose the number of basis functions $K$ large enough to ensure
that the regularization is controlled by the choice of smoothing parameter(s) $\lambda$ rather than that of dimensionality $K$.

**Step 3 – Smoothing parameter selection:** After deciding the structure of the basis functions, coefficients should be estimated for each individual function. Coefficient matrix will be $K \times m$. We prefer to use the roughness penalty (RP) approach which offers a greater control on smoothness instead of the least-squares approach which only considers the fitness to the curve. Determination of smoothing parameters ($\lambda_k^{RP}, k = 1, \ldots, \#curve\ sets$) can be done subjectively or by the generalized cross-validation (GCV) method. In the GCV method, the optimum lambda is defined as the lambda that minimizes the true mean square [1]. Deviations from the sinusoidal function can be obtained by examining the behavior of the harmonic acceleration operator. If strong periodicity exists, deviations would be small. In this case, the choice of the smoothing parameter does not affect the general variation structure of the curves, so a smaller and fixed lambda value ($\lambda^{RP}$) can be chosen subjectively for all curves.

**Step 4 – Estimation:** After $K$ and lambda value(s) are determined and data are smoothed, individual functions for all $m$ subjects and mean functions for each term (years, months, etc.) are estimated, visually examined and interpreted for each variable.

**Step 5 – SFCCA:** In order to determine modes of variability in the data associated with a high correlation, SFCCA is used. Since canonical correlation analysis (CCA) cannot be directly broadened to functional data, we need to apply smoothing again. FCCA produces estimates of canonic correlation that are close to 1 because FCCA is an exceedingly greedy procedure. It is in practice, essential to enforce strong smoothness on the weight functions to limit this greediness [11,13,18]. Leurgans et al. [14] refer this procedure as smoothed CCA. By examining highly correlated weight functions graphically, smoothing parameters ($\lambda_k^{CCA}, k = 1, \ldots, \#curve\ sets$) for CCA can be determined. Ramsay and Silverman [20] suggested that a single smoothing parameter may be used. In this study we found that the changes in the value of lambda have a very little effect on the overall variability structure and on individual functions and are set to the same fixed value of ($\lambda^{CCA} = \lambda^{RP}$) for all curves. The studies of He et al. [9,10,11] give alternative approaches to basis functions for CCA. At the end of SFCCA, canonical weight functions and scores are calculated for two curve sets. The number of canonical correlations is equal to the number of basis functions. The first canonical correlation gives us most of the information about the relation between the curve sets.

**Step 6 – Interpretation and additional analyses:** Weight functions and canonical correlations are visually examined and interpreted. Additional analyses are available depending on the data set. If the data are periodical, seasonality effects can also be examined. For long-term data, all terms (months, seasons, years, etc.) can be examined separately to find out if there is a shift in the structure. Phase plane plots may be utilized to show if there is a shift over the years.

Some theoretical background on the basis function approach, Fourier basis functions, RP approach and SFCCA can be read from authors’ website [12].

### 3. Application of SFCCA on humidity and temperature data

Our goal in this case study is to examine the highest variation modes between temperature and humidity and compare effects of land forms and climates on this relationship by the help of the suggested guideline. The effects of the number of basis functions and the selection of smoothing parameter on the general variability structure and correlations of temperature and humidity curves are also studied by trial and error and reported so other researchers could benefit for their studies. We also examined the seasonal changes in both individual functions and corresponding variation structures in this time span by yearly comparisons.

Ramsay et al. [18] also examined the relationship between temperature and logarithm of precipitation ($\log(P)$) for one year and interpreted the results. Based on their study, we also examined
the relationships between 11 years of annual data for temperature and log\(P\), and humidity and log\(P\) and identified high correlations between all pairs. Since temperature and humidity have a more regular relationship, we only give here detailed results and interpretations for temperature and humidity for the year 2000 for the sake of briefness. Other results can be reached from the corresponding author.

3.1 Preliminary information about the data and its environment

In this study, daily average temperature and humidity data from 35 weather stations in Turkey, which have complete data for 11 years between 2000 and 2010, are analyzed. Data are acquired from the Head Office of Meteorology as a raw output file from their data storage. First the data are processed into a suitable data matrix for analysis by the help of MSExcel and Matlab routines. This pre-processing took a considerable time and effort. There were many stations and years with missing data with a rate over 50%. Only stations with complete data are included in this study since some of excluded stations are newly established and some of them have technical data collection problems. Resulting complete data matrix is 4015 × 35 with 365 days in 11 years for 35 stations. Some summary statistics about the raw data is given in the appendix.

Turkey lies between 36°–42° North Latitude and 26°–45° East longitude. As a result, Turkey is located in the mild temperature zone where four seasons are experienced distinctively. However, the location is not the only determinant of the climate. If it were, temperature would decrease regularly as we go from the south to the north and the same climate would reign in the western, middle and eastern regions which are on the same latitude. Whereas, in Turkey, dramatic climate changes occur in short distances. The main reasons of these changes are the rugged terrain, the orientation of the mountains near seas, being surrounded by seas on three sides and the increase in elevation from the west to the east. By the effect of all these factors, temperature, humidity and rainfall changes among regions.

Coastal regions have a milder climate due to the effect of the seas. Northern Anatolia Mountains and Taurus Mountains prevent the sea effect towards inland regions. Thus, terrestrial climate effect increases in inland regions. Inland regions are hotter in summer and colder in winter. Especially this is prominent in winter. As we go inland, temperature differences increase due to the decrease of the sea effect and humidity. Turkey’s mean elevation is approximately 1100, which is quite high. The elevation increases towards the east and causes a decrease in the temperature and an increase in terrestrial climate.

In order to make the interpretation easier, a small map of Turkey is given in Figure 1. Stations of interest are circled on the map. Gray areas are the neighboring countries and blue areas are the surrounding seas. Station names are the same for all figures without legends.

3.2 Analysis and results

Analysis and findings using the suggested guidelines are given below with some more suggestions on the analysis process.

Step 1 – Data preparation: After pre-processing of the data, our complete data matrix is 4015 × 35. Since the complete data matrix does not show the underlying seasonal variation well enough, every year is analyzed separately. Therefore, we have 11 annual data matrices which are 365 × 35. For the sake of briefness, only the results of the year 2000 are given here in detail.

Matlab code files of Ramsay [16] are modified and improved to give some required plots and outputs. Modified code files and detailed reports can be found in the corresponding author’s website [12].
Figure 1. Map of Turkey.

Figure 2. Correlation surface of humidity and temperature.

Figure 2 gives the correlation surface for the data. As can be seen from the figure, the correlation surface is nearly impossible to interpret. Thus, SFCCA is utilized to express the strength of interaction and the variation structure between the two curve sets.

**Step 2 and Step 3 – Basis function and smoothing parameter selection:** Since the data have a periodical nature, 35 individual functions, one for each station, are obtained by using the RP approach via Fourier basis functions. After some trials on the data, the number of basis functions and the period are decided to be $K = 3$ and $T = 365$, respectively. As discussed in the findings section of the application (step 6b), our aim is to see the seasonal change in daily data for one year, so we use only three basis functions, which includes only the first sine–cosine pair. As we are interested in the relation of two variables, we will have two curve sets. The RP approach is used for estimation of the coefficients. Lambda is taken as 15 consecutive values between 0.1 and 1,000,000 and all values are found to have approximately the same GCV value. Therefore, we decided subjective selection of the smoothing parameter. After some more trials, we have seen that the choice of smoothing parameter does not affect the general variation structure of the curves, so a fixed value of $\lambda^{RP} = 0.01$, which is smaller, is selected for both curves.
Step 4 – Estimation: After $K = 3$ and $\lambda^{RP} = 0.01$ are determined, individual and mean functions for all 35 subjects are estimated and then examined visually and interpreted. Individual humidity and temperature functions for 35 stations for the year 2000 are given in Figures 3 and 4, respectively.

When interpreting humidity curves in Figure 3 in the light of temperature data, the following conclusions can be reached for summer months where temperature increases:

- It can be seen that humidity increases along the coastline because of their mild climate in cities such as Samsun, Rize, Giresun, Adana and Mersin as expected. These cities are highlighted with thick solid lines.
- On the contrary, humidity decreases in the inland cities with high altitude such as Iğdır, Siirt and Mardin.

![Figure 3. Humidity curves for 35 stations for year 2000.](image-url)

![Figure 4. Temperature curves for 35 stations for year 2000.](image-url)
• As a result of the mild climate, coastline cities have a lower amplitude variation for humidity than inland and high altitude cities. And the period seems different between coastline cities and inland cities for humidity. This difference might be the result of mild climate of coastline and terrestrial climate of inland.

This kind of interaction between temperature and humidity is already expected. In meteorology, it is known that as the temperature increases, humidity also increases in the coastline. As we move inlands and as the altitude increases, humidity is expected to decrease. City of Muş (highlighted with bold dotted line) has a very distinctive sinusoidal structure (first increasing, decreasing and then again increasing), which leads to unusual correlation scores.

When interpreting temperature curves in Figure 4, the following conclusions can be made:

• Most of the individual functions have a regular sinusoidal structure and behave similarly.
• The coldest city, Ardahan, appears at the bottom of Figure 4 for the whole year. Ağrı follows its behavior in winter and fall. These cities are highlighted with dashed lines. Ağrı and Ardahan are in the east of Turkey where terrestrial climate is the strongest because of the high altitude and distance from sea.
• In summer days, the hottest city is Şanlıurfa, which is an inland city.
• In fall and winter, the hottest city is Mersin, which is located near the Mediterranean Sea.
• As for humidity curves, Muş has a very distinctive sinusoidal structure (this time first increasing and then decreasing) for temperature which leads to unusual correlation scores. This shows that Muş has a distinctive terrestrial climate as in humidity.

Since most of the cities are inland, it is natural to have a mean humidity curve (Figure 5(a)) that decreases with higher temperatures and increases with lower temperatures. The mean temperature curve in Figure 5(b) shows that the mean temperature starts to increase after spring and decreases after fall.

Deviations from the sinusoidal function can be obtained by examining the behavior of harmonic acceleration operator $L_x$. The behavior of $L_x$ is given in Figures 6 and 7. The curves in Figures 6 and 7 show the deviations from the sinusoidal function. Therefore, any curve that is different from a straight line means a deviation from a sinusoidal structure.

Here, it can be seen that deviations of temperature functions are nearly similar while deviations of humidity functions are more varied. Deviations of temperature functions have the same direction for the coastal and inland cities. On the contrary, deviations of humidity functions have different directions for coastal and inland cities. Especially biases of humidity for coastal cities reside near zero line because of their mild climate.
Figure 6. Deviations from sinusoidal function for the estimated humidity functions in Figure 3.

Figure 7. Deviations from the sinusoidal function for the estimated temperature functions in Figure 4.

**Step 5 – SFCCA:** In order to select the smoothing parameter for canonical correlation, lambda is taken as different values between 0.001 and 1,000,000 and the graphs of interacted weight functions are evaluated. We again found that the changes in the value of lambda have a very little effect on the overall variability structure and on individual functions and it is set to the same fixed value of \( \lambda_{CCA} = \lambda_{RP} = 0.01 \) for all curves.

**Step 6 – Interpretation and additional analyses:**

(a) **Weight functions for Year 2000:** Figure 8 shows the two interacted canonical weight functions associated with the first canonical correlation. Both functions are sinusoidal, but nearly symmetrical with respect to the \( x \)-axis. The weight function of humidity increases as the temperature increases. It reaches its maximum in mid-summer days. It then starts to decrease as the temperature decreases. The weight function of the temperature can be interpreted similar to this but just the opposite. It can be concluded that humidity and temperature are highly linearly correlated for this time span.
This result can be confirmed by examining the canonical correlation scores in Figure 9. For coastal cities of Samsun, Giresun, Ordu, Mersin, Adana and Rize where humidity increases as the temperature increases in summer, it is natural to have high canonical weight scores. Individual functions of these cities have the same structure as the weight function of humidity. Osmaniye has higher humidity functions than the other inland cities and thus have higher and positive scores. Individual functions of Siirt, Iğdır and Mardin have just the opposite structure of the canonical weight function of humidity. Their correlation scores are negative as expected. Muş, which was found to have a very distinctive individual function with a sinusoidal structure, has indeed high and negative correlation scores on the lower left corner of Figure 9.

When examining all cities, it can be seen that cities which have close individual functions are located closely in Figure 9. Since amplitude variation is smaller for temperature, canonical correlation scores have a smaller range. Since the scatter of the scores shows a near linear relationship, we can conclude that the correlation between temperature and humidity is high. Indeed, the first canonical correlation is 0.92. The second and the third are 0.54 and 0.16, respectively.

When we examine second canonical correlation scores (Figure 11), especially for years with higher canonical correlations, we can see that the main effect comes from the geographical proximity of the stations, not from their proximity to the coast.

Therefore, it can be concluded that being in coastal or inland areas has more effect on the first canonical correlations while being geographically close to each other has more effect on the second canonical correlations.

(b) Annual comparison of individual and weight functions for 11 years: All 11 years between 2000 and 2010 are examined in order to see if humidity, temperature and weight functions change
Figure 9. Canonical scores for the first pair of canonical variables for year 2000.

Figure 10. Second canonical weight functions for year 2000.

Figure 11. Canonical scores for the second pair of canonical variables for year 2000.
Table 1. Canonical correlations for 11 years.

| Year | First | Second | Third |
|------|-------|--------|-------|
| 2000 | 0.92  | 0.54   | 0.16  |
| 2001 | 0.88  | 0.65   | 0.20  |
| 2002 | 0.86  | 0.72   | 0.51  |
| 2003 | 0.84  | 0.78   | 0.52  |
| 2004 | 0.84  | 0.53   | 0.19  |
| 2005 | 0.83  | 0.77   | 0.52  |
| 2006 | 0.85  | 0.66   | 0.57  |
| 2007 | 0.87  | 0.81   | 0.48  |
| 2008 | 0.78  | 0.63   | 0.42  |
| 2009 | 0.85  | 0.46   | 0.23  |
| 2010 | 0.92  | 0.76   | 0.31  |

annually. It is found that all functions are similar between those years and there are no significant changes.

For the temperature, graphs are given in Figures S1 and S2 of Supplementary Materials and a summary of findings are given below.

- Ardahan is the coldest city and it is always at the bottom. Ağrı and Bayburt follows Ardahan.
- For all 11 years, Şanlıurfa, which is an inland city, is the hottest city in summer. Mersin, which is near Mediterranean Sea, is the hottest city in fall and winter. Coastal cities are expected to have a mild climate in winter.
- Temperature curves have a similar sinusoidal structure and the temperature changes between years and stations are not more significant than that of humidity.
- For all 11 years, mean temperature functions have no significant shifts. Temperature starts to rise in spring and to fall in fall as an expected sinusoidal structure.

For humidity, graphs are given in Figures S3 and S4 of Supplementary Materials and a summary of findings is given below.

- As temperature rises, humidity also rises in coastal cities such as Samsun, Ordu, Giresun, Rize and Mersin (highlighted by thick solid lines). However, they have a smaller amplitude variation due to their mild climate.
- As temperature rises, humidity falls in the inland city of Adıyaman. Malatya, Siirt, and Mardin follow Adıyaman with a few order changes.
- Since most of the stations are inland, mean humidity functions decrease as the temperature increases and increase as the temperature decreases. They also do not have any shifts for 11 years. It may indicate that a significant seasonal shift does not occur.

The first, second and third canonical correlations for 11 years are given in Table 1. Three basis functions are used and lambda is 0.01 for all years. It can be seen that correlations are quite high and have a mean of near 0.85.

When examining variation modes and canonical correlation scores for all 11 years (graphs are given in Figures S5 and S6 of Supplementary Materials), the following can be concluded:

- Temperature and humidity weight functions are very similar like year 2000. Weight functions move symmetrically for all years. Except for small periodical shifts in some years, the general structure stays the same.
• The first canonical correlations are quite high and scores have a very similar linear appearance indicating a strong linear relationship for all 11 years. For instance, coastal cities such as Samsun, Ordu, Giresun, Rize and Mersin always have high scores for humidity weight component. Muş generally leads to unusually correlation scores. It can be concluded that the list of the most influential cities for principal variation structures does not change for years.

• For the second canonical correlations, being geographically close to each other is found to be more effective.

Since all these findings are for Turkey data, although a strong relationship between humidity and temperature is deduced, data for more years and countries may help detection of seasonal shifts and climate differences better when they exist.

(c) Findings from the fine tuning of the number of basis functions and smoothing parameter:

After many trials, we concluded that the number of time points and basis functions are closely related. For example, if our aim is to see the seasonal change in daily data for one year, using three basis functions, which includes only the first sine–cosine pair, seems to be appropriate. However, three basis functions are not enough to see the change in daily data for 11 years since data would be over-smoothed. Therefore, the number of time points and the aim of the study should be considered together. Using fewer basis functions may lead to over-smoothing which ends up missing some important changes in data. Since deciding the number of basis functions by trial and error is really cumbersome, using the RP approach after deciding a reasonable number of basis functions (for example three) seems more feasible.

In this study, we used a smoothing parameter of 0.01 for all curves. Different lambda values (changing from $10^{-1}$ to $10^6$) are used and the best value is obtained by the GCV method for each function. However, since data have a strong periodical structure, it is concluded that changing lambda has a little effect on the structure of both individual and weight functions. Neither the directions nor the magnitudes of the functions are affected by the change of lambda. Therefore, for the sake of simplicity, we decided to use a fixed lambda value of 0.01 for all curves.

The number of basis functions has a more significant effect on their structures. Since the number of canonical correlations is equal to the number of basis functions and the first canonical correlations give us most of the information about the relation between the curve sets, canonical correlation values also increase to 1 artificially as the number of basis functions increases. However, in this instance, weight functions are affected by the smallest changes in data and cannot show the seasonal effect we seek for. Therefore, the number of basis functions should be as small as possible. Fourier basis enables us to choose a small number of basis functions, whereas B-spline basis forces to use a fixed number of basis functions (#knots + order-2) and does not allow smaller numbers. In the B-spline case, #knots or the number of time points may increase the correlation artificially. Since the main affector here is the number of basis functions, not the smoothing parameter, we prefer and recommend to use Fourier functions, especially for CCA of large, seasonal and long-term data.

4. Conclusion

In this study, a detailed analysis of mean daily humidity and temperature data of 35 meteorological stations in Turkey for 2000–2010 year span is conducted by following the suggested guidelines. Since data are quite big and it is difficult and inconvenient to interpret covariance surface and individual functions for big data, SFCCA is utilized in order to obtain highly correlated variation structures between variables.

Since we use daily data, seasonal effects and locational effect of stations can explicitly be determined both in individual functions and weight functions. For humidity data, coastal and
inland cities have opposite effects on individual functions, deviations from sinusoidal function and weight functions.

Furthermore, when all years between 2000 and 2010 are examined individually, it is found that there are no significant structural changes and the cities that are the most effective on variation modes are the same. This stable relationship between humidity and temperature may enable modeling and prediction for future data. If this study is expanded to include different locations on earth for more years, seasonal shifts and variation structure would be examined better.

The effect of the number of basis functions and lambda value especially on periodical data is also examined and the number of basis functions is found to be the main effector on both individual and weight functions. It is seen that increasing the number of basis functions also increases correlations; however, weight functions are affected by the smallest changes in data and become meaningless. Setting the number of basis functions to three is found to be suitable for finding seasonal effects in annual data.

In addition, after obtaining the lambda value by GCV, trials were done for different lambda values. It is concluded that changing lambda does not have a significant effect on the structure of curves and correlations for strong sinusoidal data.

As the conclusion of the study, SFCCA can be utilized efficiently for discovering relationships especially for periodical data. As presented earlier in the study, FDA is a flexible and useful method for detecting changes in meteorological and other data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental data and research materials

Supplemental data for this article can be accessed at 10.1080/02664763.2015.1019842.

References

[1] P. Craven and G. Wahba, *Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation*, Numer. Math. 31 (1979), pp. 377–403.
[2] A. Delaigle and P. Hall, *Defining probability density for a distribution of random functions*, Ann. Statist. 38(2) (2010), pp. 1171–1193.
[3] A. Delaigle, P. Hall, and N. Bathia, *Componentwise classification and clustering of functional data*, Biometrika. 99(2) (2012), pp. 299–313.
[4] D. Gervini, *Free-knot spline smoothing for functional data*, J. R. Stat. Soc. Ser. B Stat. Methodol. 68(4) (2006), pp. 671–687.
[5] M. Guo, L. Zhou, J.Z. Huang, and Wolfgang Karl Härdle, *Functional data analysis of generalized quantile regressions*, SFB 649 Discussion Paper 2013–001, Humboldt-Universität – Berlin, 2013.
[6] P. Hall and I.V. Keilegom, *Two-sample tests in functional data analysis starting from discrete data*, Statist. Sinica. 17 (2007), pp. 1511–1531.
[7] P. Hall and M.H. Nasab, *On properties of functional principal components analysis*, J. R. Stat. Soc. Ser. B 68(1) (2006), pp. 109–126.
[8] P. Hall and C. Vial, *Assessing the finite dimensionality of functional data*, J. R. Stat. Soc. Ser. B 68 (2006), pp. 689–705.
[9] G. He, H.G. Müller, and J.L. Wang, *Extending correlation and regression from multivariate to functional data*, in *Asymptotics in Statistics and Probability*, M.L. Puri, ed., VSP, Zeist, 2000, pp. 197–210.
[10] G. He, H.G. Müller, and J.L. Wang, *Functional canonical analysis for square integrable stochastic processes*, J. Multivariate Anal. 85 (2003), pp. 54–77.
[11] G. He, H.G. Müller, and J.L. Wang, *Methods of canonical analysis of functional data*, J. Statist. Plann. Inference. 122 (2004), pp. 141–159.
Appendix. Summary statistics for the raw data

| Station Number | Station Name | Location | Humidity (%) | Temperature (°C) |
|----------------|--------------|----------|--------------|-----------------|
|                |              |          | Mean | Min | Max | Std. dev. | Mean | Min | Max | Std. dev. |
| 1              | Samsun       | Coastal  | 72.9 | 11.0 | 95.7 | 14.1 | 15.1 | −3.4 | 35.0 | 7.1 |
| 2              | Ordu         | Coastal  | 70.7 | 22.7 | 95.3 | 11.5 | 15.0 | −1.6 | 28.9 | 6.9 |
| 3              | Giresun      | Coastal  | 68.7 | 18.0 | 96.3 | 13.0 | 15.1 | −1.5 | 28.8 | 6.7 |
| 4              | Rize         | Coastal  | 76.8 | 31.3 | 94.0 | 8.6  | 15.0 | −0.6 | 24.9 | 7.0 |
| 5              | Ardahan      | Inland   | 72.4 | 37.3 | 95.7 | 9.2  | 4.4  | −29.2 | 24.8 | 10.7 |
| 6              | Amasya       | Inland   | 55.2 | 9.3  | 92.3 | 12.0 | 14.1 | −13.4 | 36.6 | 8.8 |
| 7              | Tokat        | Inland   | 62.9 | 20.3 | 97.0 | 11.7 | 13.1 | −16.3 | 33.1 | 8.6 |
| 8              | Gumushane    | Inland   | 68.7 | 29.0 | 96.0 | 10.2 | 9.9  | −16.2 | 29.3 | 8.7 |
| 9              | Bayburt      | Inland   | 54.7 | 13.7 | 97.7 | 14.0 | 7.6  | −20.6 | 28.0 | 9.9 |
| 10             | Sivas        | Inland   | 65.9 | 27.0 | 96.0 | 11.8 | 9.8  | −22.6 | 30.4 | 9.5 |
| 11             | Erzincan     | Inland   | 63.5 | 26.0 | 97.7 | 12.4 | 11.7 | −20.2 | 31.3 | 10.3 |
| 12             | Agri         | Inland   | 61.9 | 22.7 | 94.7 | 12.7 | 6.8  | −34.0 | 30.0 | 12.2 |
| 13             | Igdir        | Inland   | 51.0 | 8.7  | 95.3 | 15.3 | 13.0 | −22.4 | 32.4 | 10.9 |
| 14             | Eskisehir    | Inland   | 59.0 | 20.7 | 96.0 | 15.0 | 11.1 | −21.1 | 29.6 | 8.9 |
| 15             | Yozgat       | Inland   | 63.9 | 13.3 | 96.7 | 15.5 | 9.8  | −15.2 | 30.2 | 8.9 |
| 16             | Tunceli      | Inland   | 60.1 | 16.0 | 96.3 | 16.4 | 13.4 | −17.8 | 33.8 | 10.7 |
| 17             | Van          | Inland   | 59.6 | 12.7 | 97.7 | 14.9 | 10.0 | −12.2 | 28.6 | 9.5 |
| 18             | Aksaray      | Inland   | 55.1 | 16.0 | 95.3 | 15.7 | 12.8 | −15.1 | 31.0 | 9.4 |
| 19             | Nevsehir     | Inland   | 57.9 | 9.3  | 95.3 | 15.4 | 11.5 | −16.0 | 32.6 | 9.2 |
| 20             | Kayseri      | Inland   | 60.4 | 13.7 | 98.0 | 14.9 | 11.4 | −18.4 | 31.0 | 9.6 |
| 21             | Malatya      | Inland   | 52.3 | 11.7 | 97.0 | 20.0 | 14.5 | −14.6 | 34.4 | 10.4 |
| 22             | Mus          | Inland   | 65.4 | 14.7 | 98.3 | 18.0 | 10.4 | −27.6 | 31.7 | 12.3 |
| 23             | Siirt        | Inland   | 52.0 | 11.0 | 95.0 | 20.0 | 16.8 | −9.9  | 36.7 | 10.3 |
| 24             | Karaman      | Inland   | 57.9 | 16.7 | 96.3 | 15.3 | 12.5 | −19.0 | 31.2 | 9.5 |
| 25             | Nigde        | Inland   | 37.7 | 19.3 | 96.7 | 15.6 | 11.9 | −15.8 | 30.0 | 9.3 |
| 26             | Kahramanmarsas | Inland | 61.3 | 11.7 | 97.3 | 15.1 | 17.5 | −4.0  | 37.1 | 9.1 |
| 27             | Gaziantep    | Inland   | 63.6 | 14.3 | 100.0| 18.0 | 15.9 | −4.4  | 34.5 | 9.3 |
| Station | Name   | Location | Humidity (%) | Temperature (°C) |
|---------|--------|----------|--------------|------------------|
|         |        |          | Mean | Min | Max | Std. dev. | Mean | Min | Max | Std. dev. |
| 28      | Kilis  | Inland   | 51.8  | 6.7 | 97.3 | 18.6      | 17.7  | −2.9 | 37.9 | 8.7       |
| 29      | Adiyaman | Inland  | 47.8  | 3.7 | 97.0 | 20.0      | 17.8  | −4.0 | 36.6 | 9.9       |
| 30      | Sanliurfa | Inland | 51.7  | 12.0| 97.7 | 18.9      | 19.0  | −2.6 | 39.0 | 10.0      |
| 31      | Mardin | Inland   | 44.8  | 2.7 | 98.0 | 21.2      | 16.8  | −6.2 | 37.3 | 10.2      |
| 32      | Hakkari | Inland   | 54.6  | 10.7| 97.7 | 17.9      | 11.1  | −18.8| 30.8 | 11.0      |
| 33      | Mersin | Coastal  | 66.2  | 18.3| 94.7 | 12.6      | 20.3  | 4.0  | 32.3 | 6.9       |
| 34      | Adana  | Coastal  | 69.2  | 18.7| 97.3 | 12.9      | 19.5  | 2.0  | 34.0 | 7.6       |
| 35      | Osmaniye | Inland | 62.4  | 18.3| 96.7 | 12.3      | 18.4  | 0.6  | 32.6 | 7.6       |