ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 17, No. 1, 2022, *pp*.058-063

Statistical Downscaling and Evaluation of Summer Temperature in Jianghuai Region based on the FPCA

Shuren Cao and Chunzheng Cao[1]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
(Received February 15, 2022, accepted April 21, 2022)

Abstract: Based on the monthly temperature data of 34 stations in Jianghuai region of China from 1961 to 2005, we proposed a functional downscaling model of summer temperature, which taking into consideration the related optimal regional climate factors. The sample independence test is used to verify the robustness of the model. The ability in simulating summer temperatures in Jianghuai region is evaluated. The main results show: (1) The best selected predictors and their regions are 850hPa air temperature (27.5-37.5°N, 107.5-125°E) and sea level pressure (25-35°N, 102.5-125°E). (2) The first four principal components of the two factors can account for more than 85% of the original information after downscaling. (3) The independent sample test proves that the functional downscaling model has a good ability to simulate the spatial and temporal structure and the changes of summer temperature in Jianghuai region during the period of 1991-2005, especially for the northeast coastal part.

Keywords: Jianghuai region; Summer temperature; FPCA; Statistical downscaling model.

1. Introduction

Since the 20th century, the global climate has changed dramatically compared to the previous period, with significant warming as its main characteristic. Intergovernmental Panel on Climate Change (IPCC) has also pointed out that the global surface temperature has shown a linear increase. Climate change will have far-reaching implications for life on earth and the development of human society. Given the specificity of China's geographical location, summer temperatures in China show more complex characteristics both in terms of spatial distribution and temporal variability, which are challenging to study. In particular, the climate situation in the Jianghuai region, which lies between the subtropical and warm temperate zones, is even more complex and variable, coupled with its population concentration and economically developed characteristics, the development and improvement of temperature prediction and forecasting capabilities in this region will provide an important reference basis for enhancing the response to climate change and formulating long-term development plans in the region and in China.

A number of researchers have already made excellent contributions to the study of improving the modelling capabilities of climate models. An important practical tool for modelling future climate change scenarios is the Global Climate Models (GCMs). These models are useful to help simulate large-scale mean modes and better characterize local climate at the global scale [1,2]. However, it has difficulty to provide accurate regional climate information to meet the assessment needs of researchers. Statistical downscaling methods stand out among the many methods due to their fast simulation speed and are widely used in various fields [3, 4]. Some scholars in China have also applied downscaling models to the study of climate change in China, verifying their superiority for short-term prediction in local regions [5, 6].

Ramsay and Silverman [7] firstly propose the concept of functional data. They have made an invaluable contribution to the advancement of functional data analysis. The relationship between temperature and precipitation in Canada is discussed based on the theory of functional principal component analysis. The use of functional principal component models for forecasting was first proposed by Hyndman and Ullah [8] and applied to mortality and fertility forecasting problems. Maity [9] and Ivanescu et al.[10] have conducted research on functional principal component regression. The research on functional regression problems is mainly divided into parametric models and nonparametric methods. Wang Bo et al. [11, 12] conducted research on the Gaussian process of nonparametric functional regression for nonparametric methods. In the study of parametric models, in order to estimate the functional slope of the model, the standard functional principal component regression estimation method is proposed to regress the response of the principal component score associated with the largest eigenvalue of the function prediction

covariance operator, Since this estimation method may have great uncertainty for large samples, in order to obtain a more stable slope estimation, Hall and Horowitz [13] and Ferraty et al. [14] proposed a regularization function from different perspectives.

In view of the potential continuous generation process of climate factor data, this paper introduces a functional data analysis method and constructs a downscaling model. Based on the observed month-by-month summer temperatures and NCEP reanalysis data, downscaling simulations are carried out for 34 stations in the Jianghuai region, and the robustness of the model is tested. The rest of the article is organized as follows: Section 2 gives a brief description of the relevant data sources the relevant methods. In Section 3, we select climate influencing factors and related regions with strong correlation with air temperature. In Section 4, we test the optimal combination of climate influence factors, constructs a functional downscaling model for the month-by-month temperature in the Jianghuai region, and evaluate the model based on independent validation. Some discussions are given in Section 5.

2. Data and Methodology

2.1 Data

The surface observations of temperature used in this paper are the monthly mean temperatures for the year1961-2005, provided by the climate system of the National Climate Centre of the China Meteorological Administration. The relevant climate factor observations were selected from the National Oceanic and Atmospheric Administration's CDC derivation of the NCEP reanalysis. The climate factor fields selected include air temperature, which is directly related to surface temperature, and geopotential height, which is negatively correlated, as well as relative humidity, specific humidity, sea level pressure and longitudinal/latitudinal winds. The climate factor fields are available for a time span of 1961-2005, all at monthly intervals, and at a spatial resolution of 2.5℃×2.5℃. The specific area of Jianghuai is 27.5-35°N, 107.5-122.5°E, which contains 34 of the first-class stations.

2.2 Two-dimensional Functional Principal Components Analysis

Climate change can be considered as a random process on a space-time domain defined on the space of square-producible functions S^2. For $L \subset R^{d_1}$, $B \subset R^{d_2}$, consider the stochastic process $X : B \to S^2(L)$, where $X(\cdot, b)$ denotes the value of this stochastic process on $b \in B$.

Define the climate function as

$$X^{(t)} = X^{(t)}(l,b) = (X^{(1)}(l,b), X^{(2)}(l,b),..., X^{(n)}(l,b)) \tag{2.1}$$

$X^{(t)}$ is a matrix of $n \times m$, where $X^{(t)}(l,b) : t = 1,..., n; l = 1,..., i; b = 1,... j; i \times j = m$, t is the year of observation, l is the longitude of the observation grid and b is the latitude of the observation grid. Then the two-dimensional Karhunen-Loève expansion of the *t*-th subject can be represented as

$$X^{(t)}(b|l) = \mu(b|l) + \sum_{q=1}^{\infty} \xi_q(l)\varphi_q(b|l) = \mu(b|l) + \sum_{q=1}^{\infty}\sum_{p=1}^{\infty} \chi_{pq}\psi_{pq}(l)\varphi_p(b|l) \tag{2.2}$$

where $\{\psi_{pq}(l) : p \geq 1\}$ is the eigenfunction operator on the space $S^2(L)$, $\{\xi_q(l) : q \geq 1\}$ represents the coefficients of the centralized expansion process $X^c(\cdot, b)$, $\xi_q(l) = \sum_{q=1}^{+\infty} \chi_{pq}\phi_p(b|l)$ is the Karhunen-Loève expansion of the random function $\xi_q(l)$ on the space $S^2(B)$ which has the functional principal component χ_{pq} .

3. Selection of main influential factors

In the process of using downscaling methods for projections, the selection of appropriate climate impact factors is a matter of concern. Climate impact factors are supposed to determine the feature of future climate projection scenarios for the region under study. Therefore, when selecting factors, it is important to confirm that they have a significant physical link to the climate under study. That is to say they are supposed to be strongly correlated with the local climate and can be more accurately by the model, and weakly correlated with each other.

In order to select the climatic factors with significant correlation with the summer temperature in the Jianghuai region, we select the data from 1961 to 1990 of 13 factor fields that have strong physical significance in affecting the surface temperature, and calculates their correlation with the Jianghuai region during the same period. Correlation coefficient of regional mean summer temperature. Fig. 1 shows the time-dependent distribution of the regional average time series of the summer temperature field and each climate factor field. The blank space in the figure indicates that the time correlation coefficient in this region has not passed the significance test with a significance level of 0.1, indicating that the correlation between the influence factors in this region and the summer temperature is weak. Red represents a positive correlation, blue represents a negative correlation, and the darker the color, the stronger the correlation.

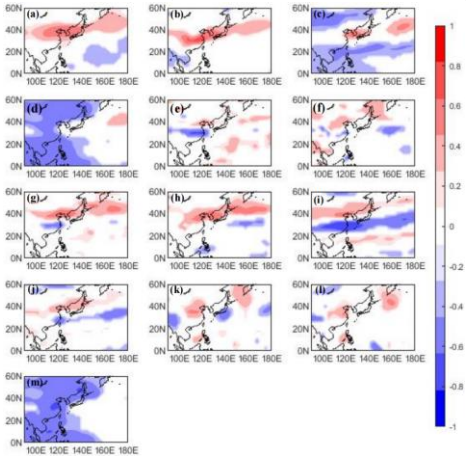

Fig. 1 Spatial distribution of regional mean time series and temporal correlation of impact factors. ((a)-(b) are air temperature, (c)-(d) are geopotential height, (e)-(f) are relative humidity, (g)-(h) are specific humidity, (i)-(j) are latitudinal wind, (k)-(l) are meridional wind , all of the above variable fields are 500 hPa for the former and 850 hPa for the latter, and finally (m) is the sea level pressure.)

Fig.1(b), the relationship between summer temperature and 850hPa air temperature in the Jianghuai region is positive, which reflects that the region is greatly affected by the subtropical high pressure belt, making the temperature in this region significantly higher. The summer temperature in the eastern sea area of the Jianghuai region has a strong positive correlation with the geopotential height of 500 hPa (Fig. 1(c)), which shows that the summer temperature in the Jianghuai region is greatly affected by the western Pacific subtropical high, which produces under its control. more warm areas. The information reflected in Fig. 1(i) and 1(b) is basically consistent, both of which reflect the strong negative correlation between summer temperature and the subtropical high belt in the Jianghuai region. The reason for this phenomenon is that the north of the summer subtropical high ridge line In addition, east wind prevails at 500hPa above the ground. By analyzing the spatial distribution of the correlation between the summer temperature and the sea level pressure field in the Jianghuai region (Fig. 1(m)), it can be clearly seen that the two have an obvious negative correlation, which shows that in summer, the Jianghuai region is mainly characterized by low pressure. Most come from the thermal factor here.

Four regional fields with significant and strong correlations with summer temperatures in the Jianghuai region are selected, namely 850hPa air temperature, 500hPa geopotential height, 500hPa latitudinal winds and sea level pressure.The ranges of strong correlations are 27.5-37.5°N,107.5-125°E, 10-25°N,95-170°E, 25-32.5°N,95-140°E and 25-35°N,102.5-125°E respectively.

4. Combination of chosen factors

We then perform a factor combination test on the selected factor fields to obtain the combination of impact factors with the best simulation results. When performing functional downscaling simulation of surface temperature in local areas, it has been proved to be more important to use large-scale temperature variables. Therefore, when selecting the combination of influencing factors in this paper, the temperature factor will be included. The observed field data from 1961-1990 were selected as the training set to build a functional downscaling regression model with precipitation observations from 34 stations during the same period, while the data from 1991-2005 were used as the test set.

Table 1 presents the fitting results under various factor combinations. We can see that when only one factor is selected for functional downscaling simulation prediction, the simulation effect of the model is obviously inferior to that of the multi-factor combined field. In addition, whether in the training set or in the

test set, the correlation coefficient obtained by the factor field a is significantly higher than that of the other three factor fields, so it is very necessary to include 850hPa air temperature in the multi-factor combination. From the training set, as the number of factor classes increases, the correlation coefficient of the simulation results increases significantly, and gets closer to 1, and the mean square error and bias also decrease significantly. When all four factor fields are included, the training set correlation coefficient is as high as 0.94 RMSE and the deviation is less than 0.25. However, this advantage cannot be maintained in the test set. From the perspective of the correlation coefficient, when the number of factor combinations is 2, the correlation coefficient is significantly higher than that of the factor combination fields with other numbers, all higher than 0.64, and the root mean square error are also lower, both below 0.63, and the deviation remains below 0.4. On the whole, selecting the combination of 850hPa air temperature and sea level pressure with the best test set effect and better training set effect can obtain more ideal prediction results, that is, selecting the air temperature field and the sea level pressure field as the predictors of the functional downscaling model.

TABLE 1. Effect of functional downscaled factors combination models.

Factor combinations	Training set			Test set		
	Cor	RMSE	Bias	Cor	RMSE	Bias
850hPa air temperature	0.82	0.42	0.60	0.68	0.59	0.42
500hPa geopotential height	0.60	0.60	0.73	0.28	0.77	0.43
500hPa latitudinal winds	0.52	0.64	0.60	0.34	0.93	0.48
sea level pressure	0.60	0.59	0.52	0.23	0.78	0.48
850hPa air temperature+500hPa geopotential height	0.89	0.34	0.36	0.73	0.55	0.38
850hPa air temperature+500hPa latitudinal winds	0.90	0.31	0.21	0.64	0.63	0.36
850hPa air temperature+sea level pressure	0.89	0.34	0.18	0.75	0.52	0.38
850hPa air temperature+500hPa geopotential height +500hPa latitudinal winds	0.92	0.28	0.25	0.63	0.64	0.34
850hPa air temperature+500hPa geopotential height +sea level pressure	0.92	0.29	0.26	0.70	0.56	0.37
850hPa air temperature+500hPa latitudinal winds +sea level pressure	0.93	0.27	0.25	0.62	0.66	0.40
850hPa air temperature+500hPa geopotential height +500hPa latitudinal winds+sea level pressure	0.94	0.24	0.23	0.60	0.67	0.43

Among them, the first four principal components of the 850hPa air temperature field explain 86.65% of the total variance, that is, they can represent more than 85% of the original data information. Therefore, the selection of the first four principal components is enough to explain the absolute variance of the original 850hPa air temperature field. most information. The first principal component of the sea level pressure field explained 89.01% of the total variance, and the first four functional principal components included 95.97% of the original data, which was enough to replace the original 50 indicators. Fig. 2 shows the spatial patterns of the first four functional principal components of the 850hPa air temperature field and the sea level pressure field under the functional downscaling model.

The first mode of the 850hPa air temperature field selection area explains the vast majority (59.95%) of the information in this field, showing an overall consistency of temperature conditions at the site and showing a gradual decrease in influence from north to south, reflecting the thermal consistency of summer temperatures in this region of the river. The second mode explains 13.92% of the information from this field, showing a latitudinal north-south gradient pattern, with a central band at 33°N spreading in both directions, in contrast to the southernmost and northern ends of the observation area, possibly related to the meridional winds. The third mode explains 8.48% of the 850hPa air temperature field selection area and represents a left-negative-right-positive dipole pattern in the northern part of the selection area, explaining some of the differences in the northern sea-land position. The fourth mode explains 3.70% of the 850hPa air temperature field selection area, representing a left-positive-right-negative dipole pattern in the southern part of the selection area, explaining some of the variance in the southern sea-land position.

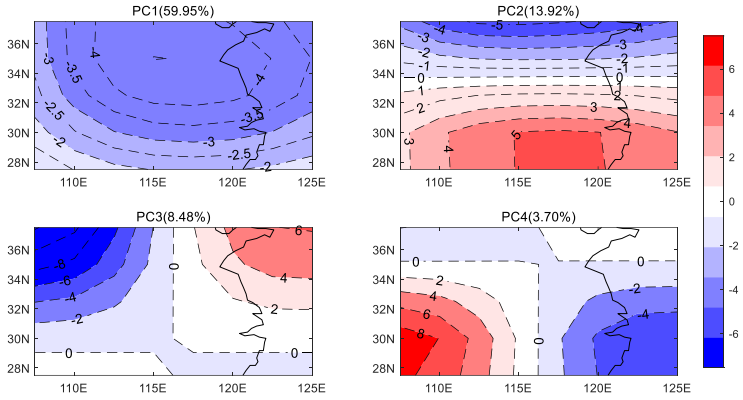

Fig.2 The first four modes of the 850hPa air temperature.

The first mode for the sea level pressure field selection region explains nearly all (89.01%) of the information from this field, demonstrating the overall positive consistency of the sea level pressure field in the region, emphasizing the consistent correlation of positive sea level pressure centred on the northwest corner. Mode 2 explains 5.36% of the total information, representing a left-positive and right-negative dipole pattern for the selected region, explaining some of the differences in sea-land position. The third mode shows a pattern centred on four locations in the northwest, northeast, southwest and southeast corners and expanding in all directions, with negative anomalies in the first and fourth locations and positive anomalies in the second and third locations. The entire mode shows signs of central symmetry about 29°S and 110°E, explaining a smaller part of the total information, only 1.78%. The fourth pattern also explains only a small part (1.38%) of the total information, showing a near-latitudinal north-south gradient pattern, with a central band at 29°N spreading in both directions, in contrast to the extreme south-western and north-western ends of the observation area, showing positive anomalies in the north and negative anomalies in the south.

5. Evaluation of the downscaling capability

Based on the test set data from 1991 to 2005, we construct a functional downscaling regression model, analyzes the simulated surface air temperature of 34 stations in the Jianghuai region, and evaluates the effect of the model.

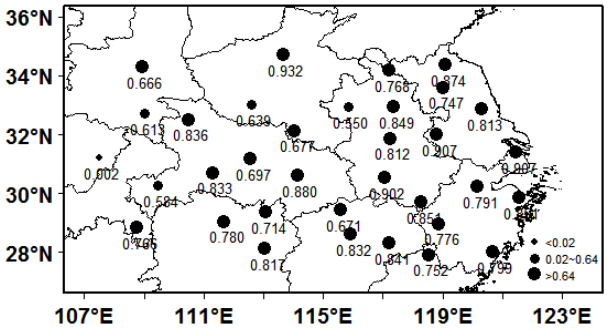

Fig. 3 Correlation coefficient between the downscaling results and the observed values.

Fig.3 shows the spatial distribution of the temporal correlation coefficients between the functional statistical downscaling results of the 34 stations in Jianghuai and the corresponding air temperature observations in this region. The correlation coefficient of most of the stations exceeds 0.64, which also shows that the correlation coefficient of these stations has passed the significance test of α=0.01, which is enough to show that the model has a good simulation effect on the summer temperature. From the perspective of spatial distribution, the farther east the region is, the higher the station correlation coefficient, and the better the simulation effect in coastal areas, with the correlation coefficient exceeding 0.95. It shows that the model's ability to simulate the interannual variation of summer temperature for the stations in the eastern part of Jianghuai, especially the stations in the coastal areas, is higher than that of the stations in the west.

6. Conclusion

Based on the monthly summer temperature data from 34 stations in the Jianghuai region of China in1961-2005 and the NCEP reanalysis data, the optimal climate factors and their regions were selected by correlation analysis, and a functional statistical downscaling model is developed and tested for robustness.

Based on the results of the correlation analysis between the regional summer air temperature averages and the reanalysis data in the the Jianghuai region, four factors are selected, and the most effective combination of influencing factors and their regions are obtained by comparing the 850hPa air temperature (27.5-37.5°N, 107.5-125°E) with the sea level pressure (25-35°N, 102.5- 125°E). After functional statistical downscaling, the first four functional principal components that can adequately represent the original information were selected, and they explained 86.65% and 96.97% of the original climate field information of the two factors, respectively, to achieve further downscaling of the reanalysis data.

The functional downscaling model is tested based on the observation samples from 1991-2005. The model is able to simulate the spatial and temporal structure of summer temperatures in the Jianghuai region, especially in the northeastern coastal region. The spatial correlation coefficients between the annual results and the observed values after statistical downscaling exceeded 0.7. The RMSE between the simulated results and the actual observed values is less than 0.6 for most of the stations.

References:

1. X. X. Wang, D. P. Jiang, X. M. Lang, Temperature and Precipitation Changes over China under a 1.5℃ Global Warming Scenario Based on CMIP5 Models, Chinese Journal of Atmospheric Science 43(05), (2019) 1158-1170.
2. Y. X. Zhao, D. P. Xiao, H. Z. Bai, Projection and Application for Future Climate in China by CMIP5 Climate Model, Meteorological Science and Technology 47(04), (2019) 608-621.
3. J. Q. Sun, J. H. Ma, H. P. Chen, et al.,Application of Downscaling Methods in the East Asian Climate Prediction, Chinese Journal of Atmospheric Science 42(04), (2018) 806-822.
4. J. C. Shu, X. W. Jiang, X. M. Huang, et al., Statistical Downscaling Modeling Analysis of Summer Precipitation in Southwest China, Plateau Meteorology 38(02), (2019) 125-134.
5. C. Q. Ruan, J. P. Li, An Improvement in a Time-Scale Decomposition Statistical Downscaling Prediction Model for Summer Rainfall over North China, Chinese Journal of Atmospheric Sciences 40(1), (2016) 215-226.
6. J. J. Li, A. H. Wang, D. L. Guo, et al., Evaluation of extreme temperature indices over China in the NEX-GDDP simulated by high-resolution statistical downscaling models, Acta Meteorologica Sinica 77(3), (2019) 579-593.
7. J. O. Ramsay, B. W. Silverman, Functional data analysis 2nd[M], New York: Springer-Press, (2005).
8. R. J. Hyndman, M. S. Ullah, Robust forecasting of mortality and fertility rates: A functional data approach, Computational Statistics & Data Analysis 51(10), (2007) 4942-4956.
9. M. Arnab, Nonparametric functional concurrent regression models, Wiley Interdisciplinary Reviews: Computational Statistics 9(02), (2017) 1394.
10. A. Ivanescu, A. Staicu A, F. Scheipl, et al., Penalized function-on-function regression, Computational Statistics 30(02), (2015) 539-568.
11. B. Wang, A. P. Xu, Gaussian process methods for nonparametric functional regression with mixed predictors, Computational Statistics and Data Analysis 131, (2018) 80-90.
12. B. Wang, C. Tao, A. P. Xu, Gaussian process regression with functional covariates and multivariate response, Chemometrics and Intelligent Laboratory Systems 163(01), (2017) 1-6.
13. H. Peter, J. L. Horowitz, Methodology and convergence rates for functional linear regression, The Annals of Statistics35(01), (2007) 70-91.
14. F. Ferraty, G. M. G. Wenceslao, M. C. Adela, et al., Presmoothing in functional linear regression, Statistica Sinica 22(02), (2012) 69-94.