Statistically Downscaled Summer Rainfall over the Middle-Lower Reaches of the Yangtze River

GUO Yan[1,2], LI Jian-Ping[1], and LI Yun[3]

[1] *State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*
[2] *Graduate University of Chinese Academy of Sciences, Beijing 100029, China*
[3] *Mathematics, Informatics and Statistics, Commonwealth Scientific and Industrial Research Organisation, Wembley, Western Australia 6913, Australia*

Abstract   The summer rainfall over the middle-lower reaches of the Yangtze River valley (YRSR) has been estimated with a multi-linear regression model using principal atmospheric modes derived from a 500 hPa geopotential height and a 700 hPa zonal vapor flux over the domain of East Asia and the West Pacific. The model was developed using data from 1958–92 and validated with an independent prediction from 1993–2008. The independent prediction was efficient in predicting the YRSR with a correlation coefficient of 0.72 and a relative root mean square error of 18%. The downscaling model was applied to two general circulation models (GCMs) of Flexible Global Ocean-Atmosphere-Land System Model (FGOALS) and Geophysical Fluid Dynamics Laboratory coupled climate model version 2.1 (GFDL-CM2.1) to project rainfall for present and future climate under B1 and A1B emission scenarios. The downscaled results provided a closer representation of the observation compared to the raw models in the present climate. In addition, compared to the inconsistent prediction directly from different GCMs, the downscaled results provided a consistent projection for this half-century, which indicated a clear increase in the YRSR. Under the B1 emission scenario, the rainfall could increase by an average of 11.9% until 2011−25 and 17.2% until 2036−50 from the current state; under the A1B emission scenario, rainfall could increase by an average of 15.5% until 2011−25 and 25.3% until 2036−50 from the current state. Moreover, the increased rate was faster in the following decade (2011−25) than the latter of this half-century (2036−50) under both emissions.

Keywords: statistical downscaling, summer rainfall, Yangtze River, future scenario

Citation: Guo, Y., J.-P. Li, and Y. Li, 2011: Statistically downscaled summer rainfall over the middle-lower reaches of the Yangtze River, *Atmos. Oceanic Sci. Lett.*, **4**, 191–198.

1   Introduction

In past decades, the global climate has undergone rapid changes (IPCC, 2007). The projection of future climate and its associated influence has attracted a growing interest worldwide. The general circulation models (GCMs) are good tools used to simulate large-scale, upper-layer climatology features, but they fail to reproduce surface parameters on regional and sub-grid scales; however, the later has more of a social importance. Many approaches have been developed because of the demand of providing regional climate change estimations for both impact studies and policy making; the statistical downscaling technique is one of the most useful (Fowler et al., 2007; Wilby and Wigley, 1997). It has been successfully applied to project temperature, precipitation, and other parameters on various timescales (Goodess and Palutikof, 1998; Hughes and Guttorp, 1994; Li and Smith, 2009).

Recently, the statistical downscaling technique has been investigated and applied in China. For example, Chen et al. (2003) utilized GCM-outputs to predict monthly rainfall, and Chu et al. (2008) employed the Statistical DownScaling Method (SDSM) to predict daily rainfall over the Haihe River basin. In addition, Zhu et al. (2008) developed a downscaling model for the Asia-Pacific summer monsoon rainfall with a Singular Value Decomposition (SVD) method and Fan (2009) constructed stepwise regression model to project the temperature over all of China. Huang et al. (2010) estimated future scenarios of annual rainfall in the Yangtze River basin with SDSM. However, a statistically downscaled projection of the summer rainfall over the middle-lower reaches of the Yangtze River (YRSR) has not been developed. The middle-lower reaches of the Yangtze River, located within the eastern Asian subtropical Monsoon domain, and its summer rainfall is strongly affected by the Eastern Asian Summer Monsoon (EASM), which shows pronounced inter-annual and decadal variability. Because the EASM has weakened during the past 50 years (Zhou et al., 2009), the associated monsoon rainfall has shown decadal variation (Zhai et al., 2004). Therefore, a reliable projection of the YRSR for the future is of great concern and was the main topic in this study.

The aim of this study was the following: 1) to build a statistical downscaling model for the YRSR using the multi-linear regression (MLR) method and principal components (PCs) of large-scale atmospheric parameters; and 2) to apply the downscaling model to GCM-generation as well as to project future rainfall under different emission scenarios. The framework of this study was organized as follows. Sections 2 and 3 introduce the data

Corresponding author: LI Jian-Ping, ljp@lasg.iap.ac.cn

and methodology utilized. The performance of the downscaling model is presented in Section 4. Section 5 describes the application of the downscaling model to GCM generation. Finally, Section 6 consists of the summary and discussion.

2   Data

The precipitation data were extracted from China's 160-station monthly rainfall dataset for the period 1958–2008, which were provided by the China Meteorological Administration. The total rainfall during June–August (JJA) averaged over 22 stations (Fig. 1) within the region of 28–32°N and east of 106°E is presented as the predictand.

Previous studies have indicated the necessity of including humidity-related parameters to predict precipitation (Benestad, 2001; Charles et al., 1999; Von Storch et al., 1993). Therefore, sea level pressure (SLP) and 500 hPa geopotential height (H5) and humidity-related parameters, that is, zonal or meridional vapor flux at 850 hPa (ZV85, MV85) or 700 hPa (ZV7, MV7), were employed as potential predictor parameters. The atmospheric data were extracted from the reanalysis dataset on a 2.5° × 2.5° grid, which was provided by the National Center for Environment Prediction–National Center for Atmospheric Research (NCEP–NCAR).

The GCM data were derived from two models of Flexible Global Ocean-Atmosphere-Land System Model (FGOALS) and Geophysical Fluid Dynamics Laboratory coupled climate model version 2.1 (GFDL-CM2.1), which participate in the World Climate Research Programme's (WCRP's) Coupled Model Inter-comparison Project phase 3 (CMIP3) under the experiments of 20th century simulation (20c3m) and Special Report on Emission Scenarios B1 and A1B; they are available at the following website: http://www-pcmdi.llnl.gov/. Because the GCMs have different horizontal resolutions, raw GCMs outputs were interpolated to a resolution of 2.5°×2.5°,

which was the same as NCEP reanalysis data using the bi-linear interpolation method.

3   The statistical downscaling scheme

In this study, we utilized the MLR method to formulate the predictive equation. Firstly, Principal Component Analysis was performed on normalized atmospheric-parameter data to derive the PCs as predictors. It was important to select the correct number of PCs necessary to describe the rainfall time-series. To ensure the robustness of the selection, the leave-one-out cross-validation method (Stone, 1974) was introduced to hindcast the rainfall time-series. The root mean square error (RMSE) between the observed and cross-validation hindcasted rainfall was used to measure the performance of each of the PCs that was progressively added to the regression equation; a particular PC was added if it could consecutively reduce the RMSE value after being included, otherwise, it was excluded and the selection was terminated. In general, the first two or three PCs were sufficient to fit the YRSR variation.

To determine the optimum predictor parameter and domain, various downscaling models developed using distinct predictor parameters over distinct domains were compared. In this paper, an exhaustive set of possible combinations of the considered predictor parameters was examined over three domains with different sizes. Each of the predictor parameter combinations consisted of a pressure-related parameter (SLP or H5) and a humidity-related parameter (ZV85, MV85, ZV7, or MV7). Three trial domains of different spatial sizes are showed in Fig. 1 as domain 1: 20–40°N, 95–125°E; domain 2: 15–55°N, 90–130°E; and domain 3: 5–55°N, 80–140°E.

The whole data set was split into two periods, the training period (1958–92) and the test period (1993–2008). The downscaling model was fitted with an optimum predictor parameter over an optimum domain over the training period; then, it was possible to make a prediction from the 1993–2008 data to show its independent hindcast ability for YRSR. The correlation coefficient and RMSE between the downscaled and observed rainfall were used to measure the predictive skills. To quantify the degree of prediction uncertainty, a bootstrap sampling approach (Stine, 1985) was employed. The 95% confidence intervals of independent predictions were obtained from the spread of 1000 bootstrap samples with random replacement.

4   Downscaled results

Using the training data from 1958–92, distinct statistical downscaling models were trained with different predictor parameters over different domains and they were evaluated by cross-validation. Figure 2 shows the performances of various downscaling models measured by the RMSE and the correlation coefficient between the observation and the cross-validated hindcasts over the time period of 1958–92. It seemed that the model trained using H5+ZV7 within domain 1 performed best with a
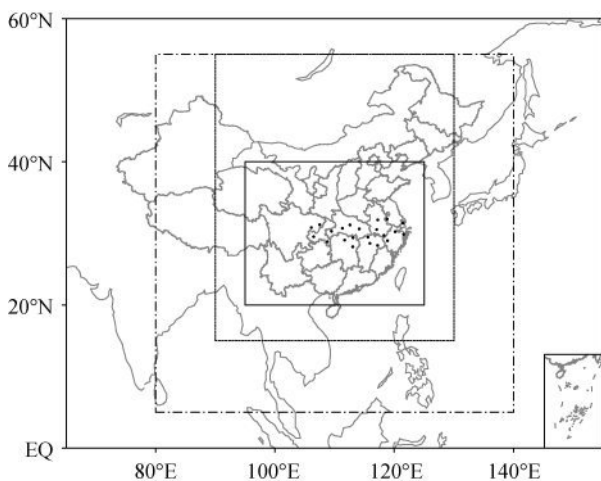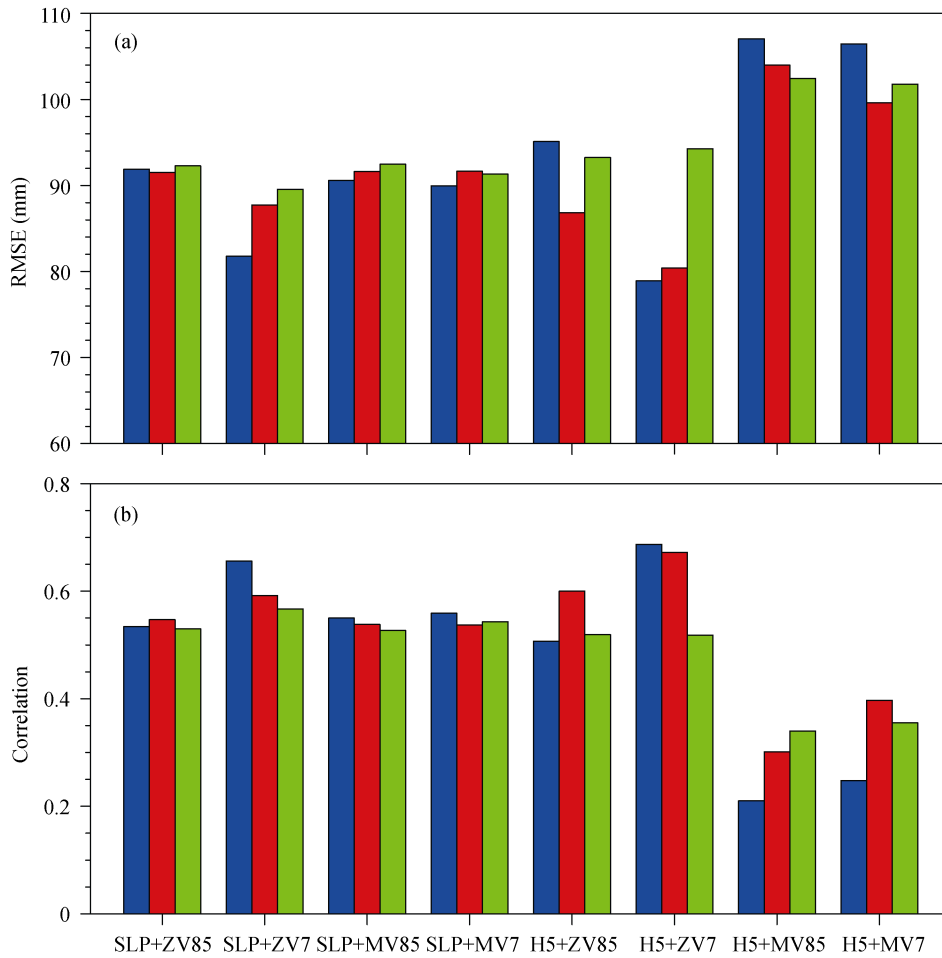

Figure 1   Twenty-two gauge stations were used to represent the region of middle-lower reaches of the Yangtze River (black dots) and three predictor domains (rectangles) were tested in this study. Please see the main text for details.

Figure 2 (a) RMSE and (b) correlation coefficient between the observed and cross-validation hindcasted rainfall from distinct downscaling models trained using different parameter combinations (*x*-coordinate labeled) over different domains of domain 1 (20–40°N, 95–125°E; blue), domain 2 (15–55°N, 90–130°E; red), and domain 3 (5–55°N, 80–140°E; green).

minimal RMSE value of 78.9 mm and the maximal correlation coefficient of 0.69; thus, H5+ZV7 was determined as the final predictor parameter and domain 1 was the final predictor domain.

The downscaling model was trained using the observed H5+ZV7 within domain 1 from 1958–92. In the regression equation, only two leading PCs were selected because the regression equation, which contained the first two PCs, achieved the minimum RMSE value of 77.04 mm in the cross-validation, indicating there was no need to include the other PCs in the equation. Figures 3a–f show the spatial modes and the associated normalized time series, which individually explained 55.3% and 15.9% of the total variance, respectively. The first PC primarily showed the decadal variation, while the second PC primarily showed the inter-annual variation; they were both significantly related to the meridional displacement of western North Pacific Subtropical High ridge with correlation coefficients of −0.46 and −0.52 calculated by detrended data (significant at the 0.05 level). In the 700 hPa zonal vapor flux field, there were anomalous eastward vapor transfer responses over the middle-lower reaches of the Yangtze River valley in the two modes,

which were responsible for the anomalous wet summer there.

Using the two leading PCs, the regression equation was formulated in the following form:

$$Y=483.42+44.88PC1+69.13PC2, \qquad (1)$$

where *Y* is the YRSR series from 1958–92 and PC1 and PC2 are the normalized series of the first two PCs. The regression coefficients were both significant at the 0.05 level.

The test data were projected onto the two leading modes, obtaining the associated PCs time-series from 1993–2008. Using Eq. (1), the YRSR from 1993–2008 was hindcasted. Figure 3g shows the observed and downscaled rainfall from both the training period and independent validation period; the uncertainty of the 95% confidence intervals in terms of bootstrapping with predictions after 1992 are indicated by red dash lines. In general, the downscaling model provided a relatively accurate representation of observations, even for the verification period. The correlation coefficient and the ratio of RMSE to the climatology between the downscaled and observed rainfall were 0.76 and 14% in training period as well as 0.72 and 18% in validation period, respectively.
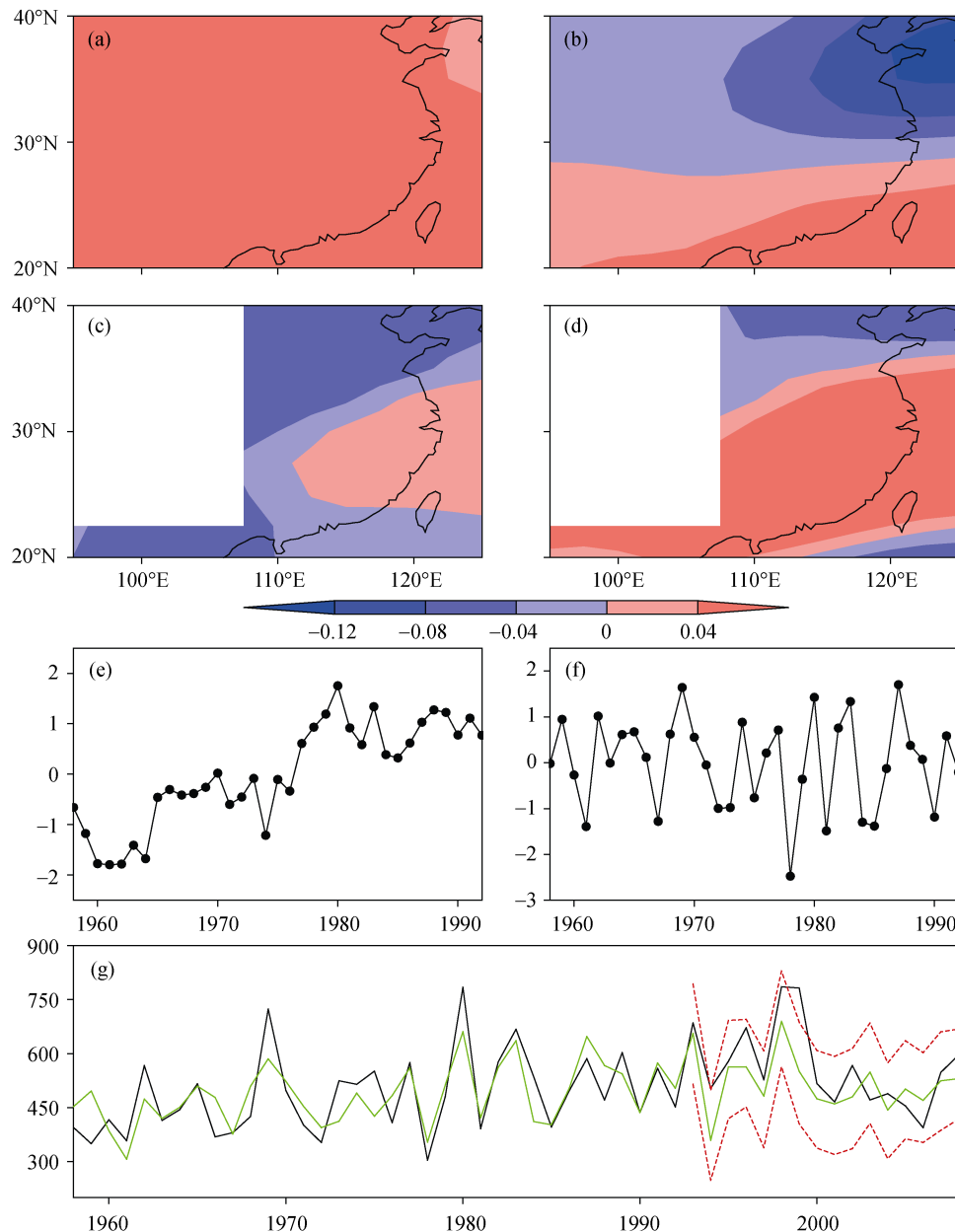
Figure 3 The first (left column) and second (right column) leading modes of the combined parameter fields of (a, b) 500 hPa geopotential height, (c, d) 700 hPa zonal vapor flux, and (e, f) their associated normalized time-series; (g) The observed (black) and downscaled (green) rainfall (mm) during the training period of 1958–92 and the independent test period from 1993–2008. The red dashed lines indicate the 95% confidence intervals for independent prediction based on 1000 bootstrap samples. The blank in (c, d) indicates the surface of the Tibet Platean.

5 Application to GCM-generation

The downscaling model was applied to the GCM-generated predictors for both the current and future climate. Before application, we updated the downscaling model with the entire data set from 1958–2008. The two leading modes, derived using the entire data set, were almost the same as the modes derived using the training data with spatial correlation coefficients of 0.99 for H5 and 0.96 for the ZV7 in Mode 1 and 0.97 for H5 and 0.96 for ZV7 in Mode 2, respectively. The regression equation was refitted over the whole period with the associated two leading PCs, which was applied to the GCM-generation to make projections.

Among all the GCMs participating in the WCRP's CMIP3, an evaluation was performed for the simulated two leading modes of H5 and ZV7 over domain 1 (20–40°N, 95–125°E) from 1958–99. The FGOALS and GFDL-CM2.1 exhibited better simulation skills than the others. Figure 4 shows the simulated two leading modes from 1958–99 using the FGOALS and GFDL-CM2.1. The spatial correlation coefficients between the observed leading modes and FGOALS (GFDL-CM2.1) simulated modes were 0.71 (0.77) for H5 and 0.39 (0.25) for ZV7 in Mode 1 and 0.94 (0.97) for H5 and 0.8 (0.68) for ZV7 in Mode 2, respectively. Thus, the outputs from the FGOALS and GFDL-CM2.1 were employed.
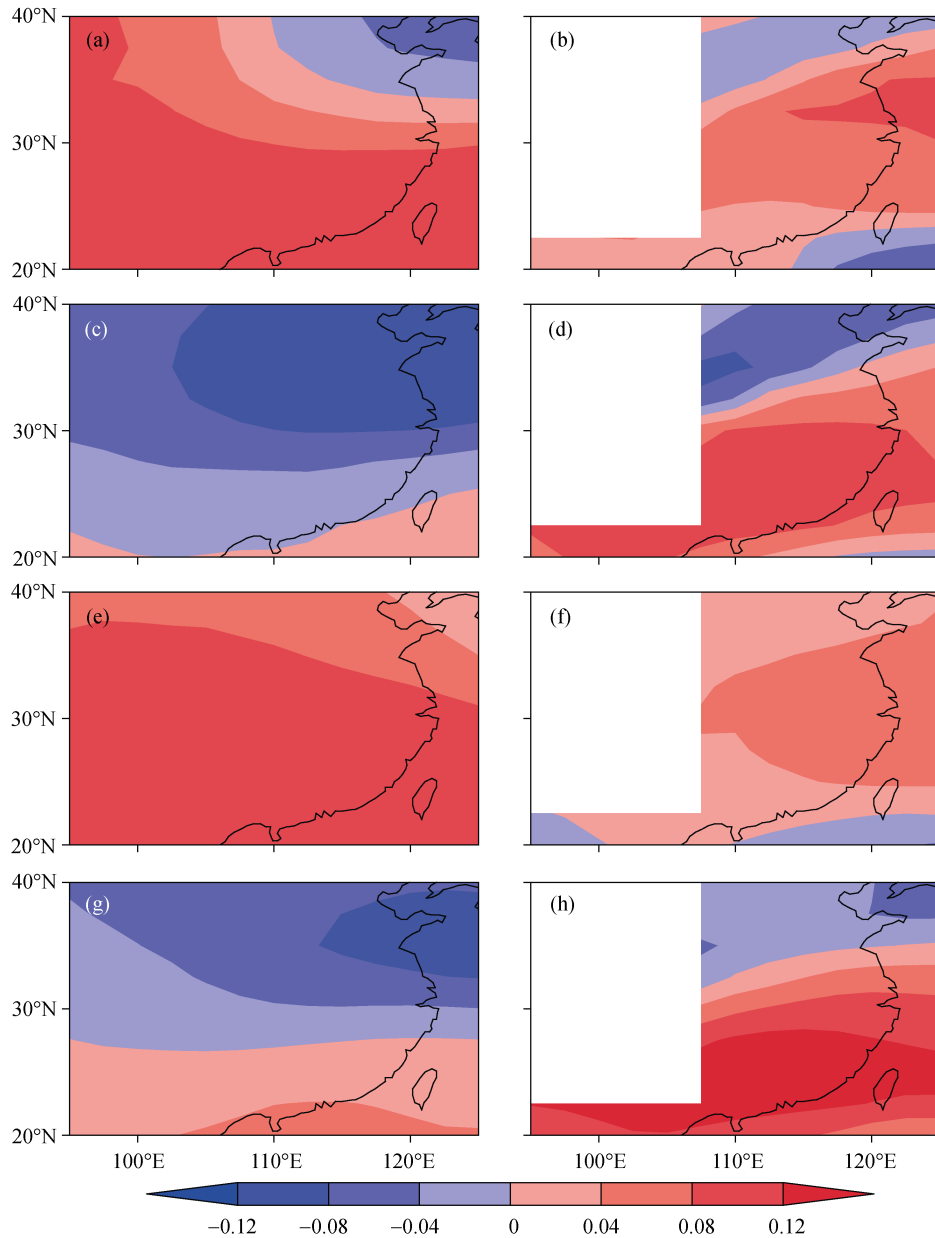
To ensure that the downscaled values from the GCM-

Figure 4 The (a, b) first and (c, d) second modes of combined parameter fields of 500 hPa geopotential height (left column) and 700 hPa zonal vapor flux (right column) calculated using FGOALS simulation over 1958−99. (e−h) are the same as (a−d), but for GFDL-CM2.1. The blank indicates the surface of the Tibet Plateau.

outputs were free from the GCM's systematical bias, the GCM-simulated data were normalized by its mean and standard deviation over the base period of 1958−99. Then the normalized data were projected onto the observed two leading modes to obtain the GCM-generated PCs. Figure 5 compares the observed and GCMs directly predicted as well as the downscaled long-term-mean rainfall for the current (1970−99) and future climate (2011−25 and 2036−50) under B1 and A1B emission scenarios. The downscaled value for the future was accompanied by 50% and 95% confidence intervals (horizontal lines in Fig. 4), which indicated the uncertainty associated with the downscaling model as estimated by the bootstrap procedure. For present day predictions, the downscaled values in

both cases showed smaller errors compared to the raw GCM simulations. For future projections, under the B1 emission, FGOALS model directly projected a slow increase in rainfall, opposite to GFDL-CM2.1 simulation; under the A1B emission, the two models projected a consistent scenario of a slight increase from 2011−25, followed by a slow decrease from 2036−50. In contrast, the downscaled results provided a consistent projection under both emissions, which indicated increasing rainfall this half century. Under the B1 emission, rainfall increased on average 11.9% until 2011−25 and 17.2% until 2036−50 from the current state; under A1B emission, rainfall increased 15.5% on average until 2011−25 and 25.3% until 2036−50 from the current state. In addition, the A1B
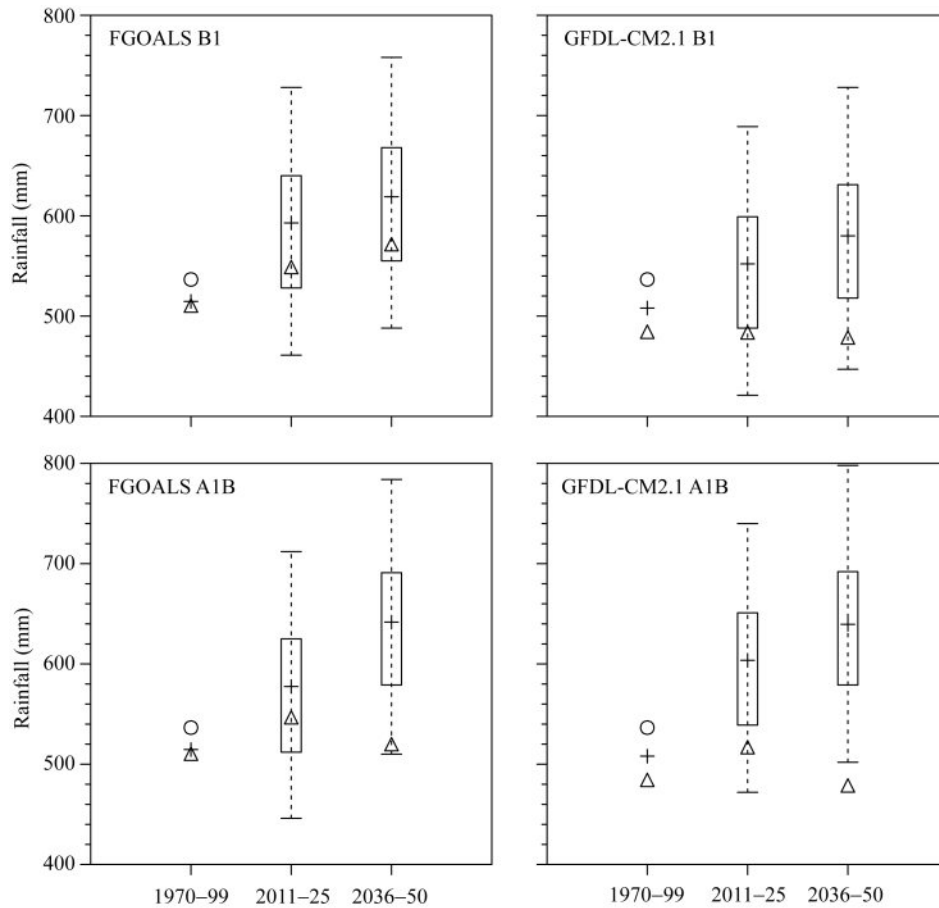
Figure 5 Present (1970−99) and near future (2011−25 and 2036−50) rainfall under B1 and A1B emission scenarios from FGOALS and GFDL-CM2.1. Circles represent observation, triangles represent raw GCM-output, and crosses represent downscaled values. Boxes and error bars represent 50% and 95% confidence intervals, respectively.

emission seemed wetter than the B1 emission for the next 40 years. Moreover, in both cases, the increasing rate was faster in the following decade (2011−25) than the latter of this half-century (2036−50) under both emission scenarios.

6 Summary

In this paper, a MLR downscaling model was developed, which linked YRSR with two principal atmospheric modes derived using 500 hPa geopotential height and 700 hPa zonal vapor flux over the domain of East Asia and West Pacific. The independent validation from 1993−2008 indicated that the MLR downscaling model had a relatively high hindcast capability for YRSR with a correlation coefficient of 0.72.

The downscaling model was then applied to the outputs of the FGOALS and the GFDL-CM2.1 models and projected the rainfall for both the current (1970−99) and near future (2011−25 and 2036−50) climate under B1 and A1B emission scenarios. For the current climate, the downscaled values in both cases showed smaller errors than the raw models. This superiority indicated that the downscaled predictions were more reliable for representing the present day climate, which implied a better representation of the future climate. For the future climate,

which was in contrast to the raw GCM simulation, the downscaled results provided consistent projections between different GCMs. This indicated a clear increase in the rainfall for this half-century; on average, the increase was larger under the A1B emission scenario vs. the B1 emission scenario. Moreover, the increasing rate seemed faster in the following decade (2011−25) compared to the latter of this half-century (2036−50) under both emissions.

Like other statistical downscaling models, the underlying stationary hypothesis may be questionable. Previous studies have emphasized the importance of assessing the robustness of the relationship in the future (Paul et al., 2008). We have performed the principal component analysis using model simulations from 2001−50 under B1 and A1B emission scenarios. Figure 6 shows the simulated two leading modes under the B1 emission scenario. Compared to the simulated leading modes under the current state (1958−99), the dominant modes in the future were maintained, and the situation using the A1B emission scenario was the same (figure not shown). To some extent, it provided reliability to the downscaled projection for the near future.

The reliability of the downscaled future projection was strongly dependent on the GCM's simulation of predict-
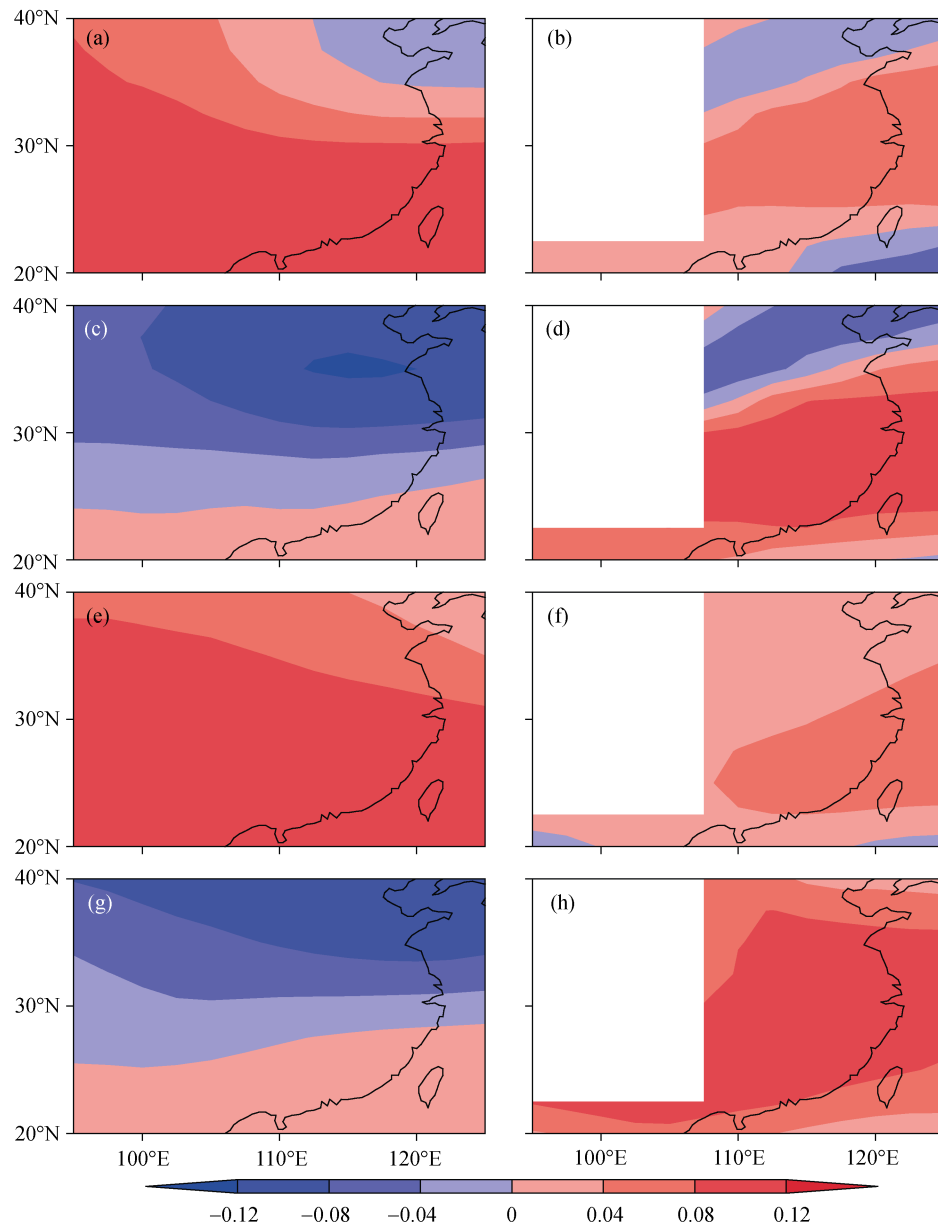
Figure 6   The (a, b) first and (c, d) second modes of combined parameter fields of 500 hPa geopotential height (left column) and 700 hPa zonal vapor flux (right column) calculated using FGOALS generated from 2001−50 under the B1 emission scenario. (e−h) are the same as (a−d), but for GFDL-CM2.1. The blank indicates the surface of the Tibet Platean.

ors. Although FGOALS and GFDL-CM2.1 were well simulation for the 500 hPa geopotential height and 700 hPa zonal vapor flux fields over the East Asia and West Pacific domain in present climate, it could not ensure the reliability of simulation under future climate change conditions. Therefore, caution should be used when interpreting the downscaled results for the future. A consistent projection from different kinds of downscaling models could provide increased reliability and confidence. More downscaled projections based on additional GCMs or the results from regional climate models are needed to support this projection and these will be performed in future work.

References

Benestad, R. E., 2001: A comparison between two empirical downscaling strategies, *Int. J. Climatol.*, **21**(13), 1645−1668.

Charles, S. P., B. C. Bates, P. H. Whetton, et al., 1999: Validation of downscaling models for changed climate conditions: Case study of southwestern Australia, *Climate Res.*, **12**, 1−14.

Chen, L. J., W. J. Li, P. Q. Zhang, et al., 2003: Application of a new

downscaling model to monthly precipitation forecast, *J. Appl. Meteor. Sci.* (in Chinese), **14**(6), 648−655.

Chu, J. T., J. Xia, and C. Y. Xu, 2008: Statistical downscaling the daily precipitation for climate change scenarios in Haihe river basin of China, *J. Nat. Res.* (in Chinese), **23**, 1068−1077.

Fan, L. J., 2009: Statistically downscaled temperature scenarios over China, *Atmos. Oceanic Sci. Lett.*, **2**, 208−213.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modeling, *Int. J. Climatol.*, **27**(12), 1547−1578.

Goodess, C. M., and J. P. Palutikof, 1998: Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling, *Int. J. Climatol.*, **18**(10), 1051−1083.

Huang, J., J. Zhang, Z. Zhang, et al., 2010: Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method, *Stoch. Environ. Res. Risk Assess*, **25**(6), 781−792.

Hughes, J. P., and P. Guttorp, 1994: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res.*, **30**(5), 1535−1546.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. H. Qing, M. Manning, et al., (Eds.), Cambridge University Press, Cambridge and New York, 996pp.

Li, Y., and I. Smith, 2009: A statistical downscaling model for southern Australia winter rainfall, *J. Climate*, **22**(5), 1142−1158.

Paul, S., C. M. Liu, J. M. Chen, et al., 2008: Development of a statistical downscaling model for projecting monthly rainfall over East Asia from a general circulation model output, *J. Geophys. Res.*, **113**, D15117, doi:10.1029/2007JD009472.

Stine, R. A., 1985: Bootstrap prediction intervals for regression, *J. Amer. Stat. Assoc.*, **80**(392), 1026−1031.

Stone, M., 1974: Cross-validatory choice and assessment of statistical predictions, *J. Roy. Stat. Soc. Ser. B*, **36**(2), 111−147.

Von Storch, H., E. Zorita, and U. Cubasch, 1993: Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime, *J. Climate*, **6**(6), 1161−1171.

Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations, *Prog. Phys. Geog.*, **21**(4), 530−548.

Zhai, P. M., Q. C. Chao, and X. K. Zou, 2004: Progress in China's climate change study in the 20th century, *J. Geogra. Sci.*, **14**(1), 3−11.

Zhou, T. J., D. Y. Gong, J. Li, et al., 2009: Detecting and understanding the multi-decadal variability of the East Asian summer monsoon recent progress and state of affairs, *Meteor. Z.*, **18**(4), 455−467.

Zhu, C. W., C. K. Park, W. S. Lee, et al., 2008: Statistical downscaling for multi-model ensemble prediction of summer monsoon rainfall in the Asia-Pacific region using geopotential height field, *Adv. Atmos. Sci.*, **25**(5), 867−884.