CrossMark

A comparison of two downscaling methods for precipitation in China

Na Zhao[1,2] · Chuan-Fa Chen[3] · Xun Zhou[1] · Tian-Xiang Yue[1]

Abstract In most cases, climate change projections from General Circulation Models (GCM) and Regional Climate Models cannot be directly applied to climate change impact studies, and downscaling is, therefore, needed. A large number of statistical downscaling methods exist, but no clear recommendations exist of which methods are more appropriate, depending on the application. This paper compares two different statistical downscaling methods, Pre_{sim1} and Pre_{sim2}, using the Coupled Model Intercomparison Project Phase 5 (CMIP5) datasets and station observations. Both methods include two steps, but the major difference between them is how the CMIP5 dataset and the station data used. The downscaled precipitation data are validated with observations through China and Jiangxi province from 1976 to 2005. Results show that GCMs cannot be used directly in climate change impact studies. In China, the second method Pre_{sim2}, which establishes regression model based on the station data, has a tendency to overestimate or underestimate the real values. The accuracy of Pre_{sim1} is much better than Pre_{sim2} based on mean absolute error, mean relative error and root mean square error. Pre_{sim1} fuses the mode data and station data effectively. Results also show the importance of the meteorological station data in the process of residual modification.

Keywords Global climate models · Statistical downscaling method · GWR · HASM · Precipitation · China

✉ Tian-Xiang Yue
yue@lreis.ac.cn

1 State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

2 Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing 210023, Jiangsu, China

3 Geomatics College, Shandong University of Science and Technology, Qingdao 266590, China

Introduction

Precipitation as a fundamental component of the global water cycle is a key parameter of ecology, hydrology and meteorology (Goovaerts 2000; Langella et al. 2010; Li and Shao 2010; Antonellini et al. 2014; Samper et al. 2014). Understanding and quantifying the spatial variability of precipitation are of key importance in hydrological studies as precipitation drives most hydrological, environmental and agricultural processes. However, strong precipitation gradients over short distance are difficult to capture with point measurements from meteorological stations. Stations are generally located in areas which are readily accessible. It is usually low and insufficient for the use of conventional spatial interpolation techniques (Celleri et al. 2007; Ward et al. 2011). In recent years, the development of remote sensing and geographic information technology has presented us with new methods of precipitation observation (Michaelides et al. 2009). Satellite precipitation data have been widely evaluated with a better performance (Dinku et al. 2007) and used for many applications such as hydrological modeling (Li et al. 2012; Su et al. 2008; Swenson and Wahr 2009), flood prediction (Li et al. 2009), land cover (Cho et al. 2014), rainfall erosivity estimation (Vrieling et al. 2010) and climatological studies (Islam and Uyeda 2007). But studies show that different remote sensing data have different performances in China (Gao

and Liu 2013; Kan et al. 2013). Moreover, the temporal coverage of remote sensing data is limited, not long enough to resolve the decadal trends and variability in China.

In the World Climate Research Programme (WCRP), different global climate models (GCMs) participate in the Coupled Model Intercomparison Project Phase 5 (CMIP5). The Coupled Model Intercomparison Project Phase 5 (CMIP5) datasets have been used for the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (AR5). These simulations of GCMs have demonstrated the ability to generally replicate the precipitation trend over the second half of the twentieth century, and can offer precipitation in a longer time scale. However, GCMs have so far been too coarse to resolve this geographically well-defined region. A number of studies have been carried out to create a connection between climate change at the large scale and at the regional scale. The most straightforward approach is linear or more sophisticated methods of interpolation between large-scale grid points closest to the region to infer the regional scale. This method has attracted a lot of criticism, since it is felt that the model resolution is too coarse and the model performance is too poor to allow for interpolation of the results. To overcome the problems with direct interpolation, the approach termed downscaling can be pursued. This approach is based on the understanding that the large-scale information provided by standard coarse-grid GCMs may be postprocessed together with the regional information to specify the regional details of the present climate and its sensitivity to changes in atmospheric composition or other external anomalies. Downscaling methods are usually classified as either dynamical or statistical. Dynamical downscaling involves the use of high-resolution, limited-area climate models within the domain of interest, whereas in statistical downscaling relatively simple statistical models are used to represent the link between atmospheric circulation variables, presumably well simulated by the GCMs, and local weather variables such as precipitation and temperature (Wilby and Wigley 1997; Fowler et al. 2007; Tareghian and Rasmussen 2013; Duan and Mei 2014). Statistical downscaling method is widely undertaken because it is easy and fast to apply (Fowler et al. 2007; Haylock et al. 2006; Barfus and Bernhofer 2014). Statistical downscaling is a two-step process consisting of (1) the development of statistical relationships between local climate variables and large-scale predictors and (2) the application of such relationships to the output of GCM experiments to simulate local climate characteristic in the future. The two main challenges in statistical downscaling are the determination of the functional relationship and the identification of the predictor variables that convey the most relevant information about the predictand and the climate change signal.

Although there is a large body of literatures where an intercomparison of different downscaling methods has been made (Mehrotra et al. 2004; Diaz-Nieto and Wilby 2005; Frost et al. 2011; Liu et al. 2012), very few of these studies have compared downscaling methods from the point of data usage ways. Here, we present two statistical downscaling methods and compare them to give the optimal one for China. We use the meteorological site information over China to downscale the simulations of CMIP5 output results. Both statistical downscaling methods used here involve two steps: (1) determining a local linear model by Geographical Weighted Regression method (GWR) for every location in the prediction domain, (2) using the High Accuracy Surface Modeling method (HASM) to modify the residual produced by the first step. The major difference between them is the data used in the two steps. Then, we use the separate dataset in Jiangxi province and 10 % of the data from national scale to validate the results. At last, a conclusion is given in the final section.

Study area and data

China is located in east Asia. It is the third largest country on earth. China's topography varies enormously from high mountainous regions to inhospitable desert zones and flat, fertile plains. It is a predominantly mountainous country with a very distinct structural pattern. The extremely varied landforms of China affect the climate conditions in various ways. Precipitation over China exhibits complex space and time structures. Large interannual variability causes local precipitation to fluctuate from year to year. Several floods and droughts often occur in the same season of a year over different regions. Precipitation over China has distinct seasonal characteristics, and is largely controlled by the monsoon circulation. Traditionally, the time from mid-May to the end of August has been defined as the east Asian summer monsoon season, resulting in remarkably variable precipitation for the whole region (Wang and Li 2007).

The historical precipitation data of 752 stations across China were obtained from the national meteorological network in China for the period 1976–2005, which were further analyzed for quality control. The sampling periods of the meteorology stations are not synchronous. Only 712 stations with more than 20 complete years are selected with the exception of 30 locations with between 15 and 25 complete years, which are located in the west of China. We chose 10 % of the total sampled points to verify test results and withheld from the downscaling calculations. We also used the meteorological stations in Jiangxi province to validate the results (Fig. 1). The WCRP's Coupled Model Intercomparison Project phase 5 (CMIP5) multi-model
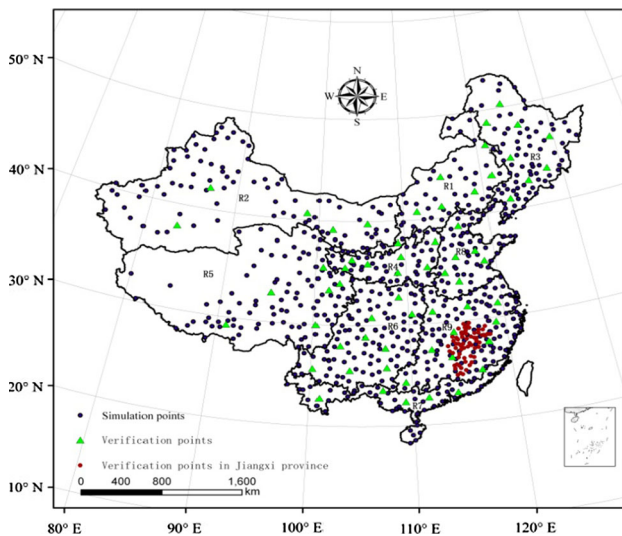
Fig. 1 Spatial distribution of the meteorological network in China

datasets (Moss et al. 2008) were used in the period 1976–2005 with a resolution of $1^\circ \times 1^\circ$. The output databases from 21 climate models were selected for the climate change projections in China under the Representative Concentration Pathways (RCP) scenarios. The selected models include both twentieth century climate simulations and twenty-first century climate projections under the RCP2.6, RCP4.5, and RCP8.5 scenarios.

Method

The statistical downscaling method used in this study can be summarized as,

$$\text{Pre}_{sim} = \text{Pre}_{downscale} + \text{Pre}_{res} \qquad (1)$$

where Pre_{sim} is the final result, $\text{Pre}_{downscale}$ is the downscaling result and will be obtained by the GWR method. Pre_{res} is the residual produced by GWR and will be interpolated by HASM. The two downscaling methods are different according to the data used in $\text{Pre}_{downscale}$, and thus in Pre_{res}. We denote the result of the first method is Pre_{sim1} and the second is Pre_{sim2}. For Pre_{sim1}, we use CMIP5 output to form the regression function and then get $\text{Pre}_{downscale1}$, and employ station data to modify the residual to obtain Pre_{res1}. While for Pre_{sim2}, we use meteorological information to establish a statistical transfer function using latitude, longitude, elevation, and impact coefficient of aspect as independent variables to produce $\text{Pre}_{downscale2}$, and employ the results of CMIP5 to modify the residual and obtain Pre_{res2}. The second method Pre_{sim2} has been widely used in climate change research in recent years (Yue 2011; Wang et al. 2012; Fan et al. 2012).

Geographically weighted regression method

Due to the large gradients in precipitation means and variances in China, it is common practice to transform observed precipitation first:

$$\overline{\overline{\text{Pre}_i}} = \frac{\text{Pre}_i}{\max\{\text{Pre}_{i,i=1,\dots,n}\}}, \qquad (2)$$

where Pre_i is the CMIP5 simulation value in the first method or the station data in the second method, $\overline{\overline{\text{Pre}_i}}$ is the transformed data, n is the number of grids of CMIP5 results or the number of stations. This process can limit extreme values in the results.

Then, we carry Box–Cox transform of $\overline{\overline{\text{Pre}_i}}$, which can give a more normal distribution and/or improved predictions (Box and Cox 1964; Sakia 1992). The formulation of this transformation is,

$$\overline{\text{Pre}_i} = \begin{cases} \ln \overline{\overline{\text{Pre}_i}}, & \delta = 0 \\ \dfrac{\overline{\overline{\text{Pre}_i}}^\delta - 1}{\delta}, & \delta \neq 0 \end{cases} \qquad (3)$$

where $\overline{\text{Pre}_i}$ is the Box–Cox transformed data and δ is a suitable parameter, which is selected to make $\overline{\text{Pr}\,e_i}$ obey normal distribution and thus satisfy the assumption of GWR method (Fotheringham et al. 2002). In this paper, $\delta = 0.4$ in the first method and $\delta = 0.48$ in the second method. Studies have shown that this process avoids negative values in the results and is necessary for precipitation interpolation (Yue et al. 2013).

It is incorrect to hold that the same linear relationship is appropriate in all places especially in the case of orographic enhancement. Unlike the ordinary linear regression model, GWR (Brunsdon et al. 1996; Loader 2004) is developed to deal with non-stationarity in the regression context, which is especially important for characterizing highly variable precipitation within China. GWR method has been successfully used in precipitation research (Brunsdon et al. 2001) and the formulation of GWR can be written as

$$\text{Pre}_{downscale} = d_{0,0}(x_i, y_j) + \sum_{i,j=1}^{N} a_{i,j} d_{i,j}(x_i, y_j) \qquad (4)$$

$\text{Pre}_{downscale}$ is the downscaling value of (i,j) grid-box in the finer scale; $d_{0,0}(x_i, y_j)$ is the intercept; $a_{i,j}$ is the explanatory variable and $d_{i,j}(x_i, y_j)$ is the corresponding coefficient which is a function of the position. x_i, y_i are the longitude and latitude, respectively. We select the independent variables from latitude, longitude, elevation, impact coefficient of aspect and sky view factor according to the value of the adjust R^2 in GWR. In this research, the most influence factors are latitude, longitude, elevation and

impact coefficient of aspect with R2 is equal to 0.92 for the first method and 0.91 for the second method.

Hasm

As an innovative surface modeling method (Yue 2011), HASM is based on the fundamental theorem of surfaces which ensures that a surface is uniquely defined by its first and second fundamental coefficients. The first fundamental coefficients reflect the local details in the surface and the second fundamental coefficients mean the macro-information of the surface. The equation of HASM is the following symmetric positive definite linear system (Zhao and Yue 2014),

$$Wx^{n+1} = v^n \tag{5}$$

Table 1 Comparison of two downscaling methods

| Validate dataset	Methods	MAE (mm)	MRE (%)	RMSE (mm)
China	Pre_{sim1}	75.24	10	119.78
	Pre_{sim2}	343.71	95	424.96
	Pre_{ms}	347.39	105	436.23
Jiangxi	Pre_{sim1}	74.43	4	97.27
	Pre_{sim2}	227.82	13	257.91
	Pre_{ms}	208.67	12	245.40

where $W = A^T A + B^T B + C^T C + \lambda^2 S^T S$, $v = A^T d + B^T q + C^T p + \lambda^2 S^T k$, and λ is a suitable parameter. The preconditioned conjugate gradient method can be used to solve Eq. (5) and the solution x is the simulated value of the residual Pre_{res} in Eq. (1).

Results and discussion

We first compare two methods in Table 1. Pre_{ms} is the CMIP5 output. Three indices, mean absolute error (MAE), mean relative error (MRE) and root mean square error (RMSE), were calculated from the station value and downscaling value at each validation sample site. The formulations of these indexes are:

$$MAE = \frac{1}{N} \sum_{k=1,\ldots,N} |Pre_{sim} - Pre_{obs}|, \ MRE$$
$$= \frac{1}{N} \sum_{k=1,\ldots,N} \left| \frac{Pre_{sim} - Pre_{obs}}{Pre_{obs}} \right|, \ RMSE$$
$$= \sqrt{\frac{1}{N} \sum_{k=1,\ldots,N} (Pre_{sim} - Pre_{obs})^2},$$

Results show that the first method is much better than the second from these three error indexes for both datasets. The accuracy of the downscaling method Pre_{sim2} is worse than the result of CMIP5 based on the validation dataset in
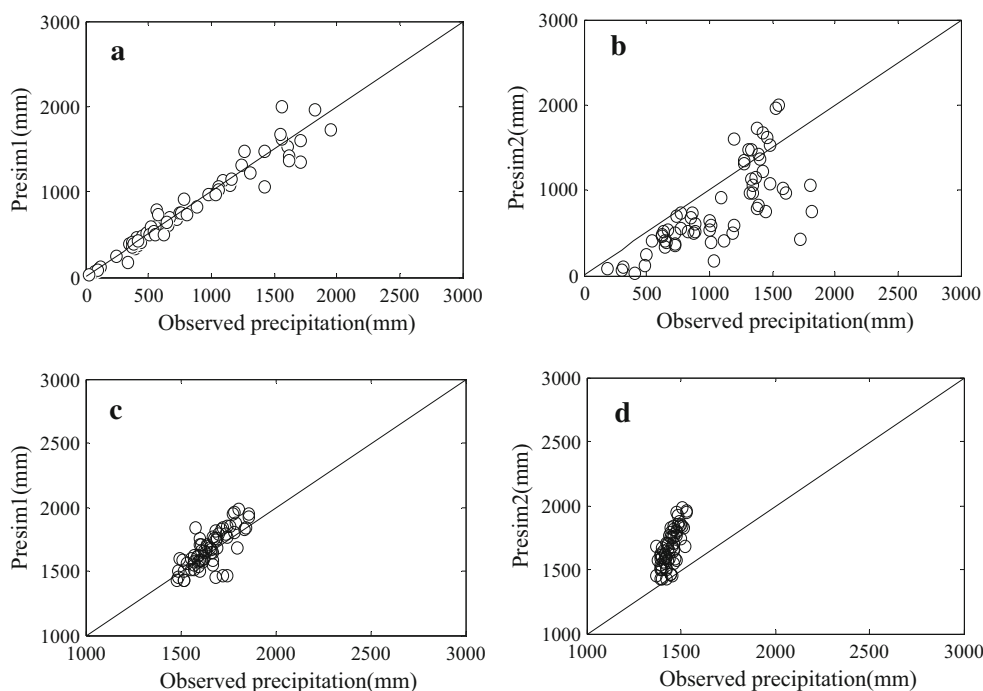
Fig. 2 Observed and estimated precipitation using different methods, **a** Pre_{sim1} for 10 % points from China, **b** Pre_{sim2} for 10 % points from China, **c** Pre_{sim1} for points from Jiangxi province, **d** Pre_{sim2} for points from Jiangxi province
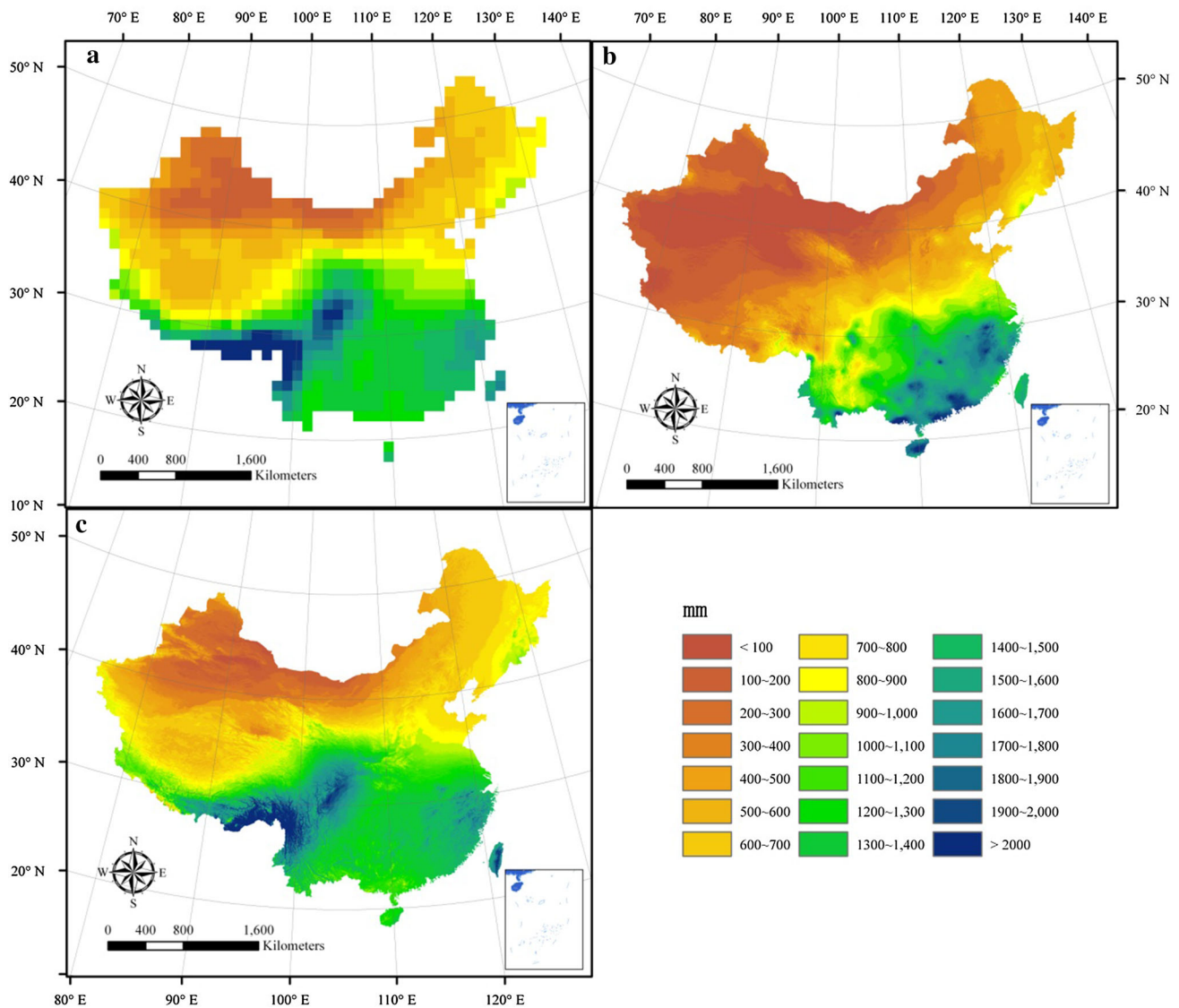
Fig. 3 The comparison of two downscaling methods, **a** original CMIP5 output, **b** the first method Pre_{sim1}, **c** the second method Pre_{sim2}

Jiangxi province. Scatter correlation plots for the observed and predicted precipitation (Fig. 2) suggest that the first downscaling method estimates the annual mean precipitation quite reliably, as shown in Fig. 2a and c. Many simulation points are relatively far from the straight line of $y = x$ using the second method. Underestimation of precipitation is evident for the points from national scale and overestimation of precipitation is obvious for Pre_{sim2} in Jiangxi province (see Fig. 2b, d). The correlation coefficients between predicted and observed values are 0.97 for Pre_{sim1} and 0.73 for Pre_{sim2} for the 10 % of the total sampled points in China. The correlation coefficients are 0.75 and 0.71 for Pre_{sim1} and Pre_{sim2}, respectively, in Jiangxi province.

Figure 3 illustrates the downscaling results. We can see that due to the large errors in the original CMIP5 output

(Fig. 3a), especially in southeastern of the Tibetan Plateau, the second method which used the CMIP5 output to modify the residual is worse than the first one. The distribution trends in Fig. 3a, c are similar, which show that the second downscaling method did not modify the errors produced by CMIP5. However, Fig. 3b, which is produced by the first downscaling method, agrees well with the real situation. The reason of this is the function of the meteorological station information. The accuracy of the results mainly depends on the first step in the downscaling process. For Pre_{sim1}, there are about 969 points of CMIP5 output that distribute evenly across China. While for Pre_{sim2}, 641 meteorological observations are used for downscaling which distribute extremely uneven in China. The site density is higher in eastern China than in western China, which did not well reflect the characteristics of

precipitation in China. The number and the distribution of the stations limit the accuracy of the downscaling results. The evenly distributed points of CMIP5 results and the local regress method, GWR which considers the non-stationarity of the precipitation, ensure the accuracy of the downscaling results in the first step. And further, we can see that original CMIP5 output is not good enough for use, which means that the introduction of station data is necessary to modify the local details that implemented by HASM. The comparison of the two downscaling methods also reveals that the second step in the downscaling process, that is, the residual correction, is critical important for accuracy improvement.

Conclusion

Precipitation, as a fundamental component of the global water cycle, is a key parameter in ecology, hydrology and meteorology. Precipitation data with accurate, high spatial resolution are crucial for improving our understanding of basin-scale hydrology. In this study, we compare two statistical downscaling methods using two datasets. One dataset scatters randomly in the whole of China and another is located in Jiangxi province. As expected, the results show that GCMs cannot be used directly in climate change impact studies. In China, the second method Pre_{sim2} which establishes regression model based on the station data has a tendency to overestimate or underestimate the real values. The advantage of the first method is obvious, which fuses the mode data and station data effectively. Results also show the importance of the meteorological station data in the process of residual modification. China is such a vast area, precipitation is affected by many geographical and topographical factors, which means that more accurate results can be obtained in different regions with different explanatory variables, especially for short time scales. Except the variables considered in this study, further researches should concentrate on more explanatory variables to gain more accurate results.

References

Antonellini M, Dentinho T, Khattabi E, Mollema PN, Silva V, Silverira P (2014) An integrated methodology to assess future water resources under land use and climate change: an application to the Tahadart drainage basin (Morocco). Environ Earth Sci 71:1839–1853

Barfus K, Bernhofer C (2014) Assessment of GCM performances for the Arabian Peninsula, Brazil, and Ukraine and indications of regional climate change. Environ Earth Sci 72:4689–4703

Box GEP, Cox DR (1964) An analysis of transformation. J R Stat Soc B 26:211–252

Brunsdon C, Fotheringham S, Charlton M (1996) Geographically weighted regression-modelling spatial non-stationarity. Geogr Anal 28:281–289

Brunsdon C, McClatchey J, Unwin DJ (2001) Spatial variations in the average rainfall-altitude relationship in Great Britain: an approach using geographically weighted regression. Int J Climatol 21:455–466

Celleri R, Willems P, Buytaert W, Feyen J (2007) Space–time rainfall variability in the Paute basin, Ecuadorian Andes. Hydrol Process 21(24):3316–3327

Cho J, Lee YW, Yeh PJF, Han KS, Kanae S (2014) Satellite-based assessment of large-scale land cover change in Asian arid regions in the period of 2001-2009. Environ Earth Sci 71(9):3935–3944

Diaz-Nieto J, Wilby RL (2005) A comparison of statistical downscaling and climate change factor method: impacts on low flows in the River Thames, United Kingdom. Clim Change 69:245–268

Dinku T, Ceccato P, Grover-Kopec E, Lemma M, Connor SJ, Ropelewski CF (2007) Validation of satellite rainfall products over East Africa's complex topography. Int J Remote Sens 28:1503–1526

Duan K, Mei YD (2014) A comparison study of three statistical downscaling methods and their model-averaging ensemble for precipitation downscaling in China. Theor Appl Climatol 116:707–719

Fan ZM, Yue TX, Chen CF, Sun XF (2012) Downscaling simulation for the scenarios of precipitation in China. Geogr Res 31(12):2283–2291

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, London

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modeling to impacts studies: recent advances in downscaling techniques for hydrological modeling. Int J Climatol 27:1547–1578

Frost AJ, Charles SP, Timbal B, Chiew FHS, Mehrotra R, Nguyen KC, Chandler RE, McGregor J, Fu G, Kirono DGC, Fernandez E, Kent D (2011) A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. J Hydrol 408:1–18

Gao YC, Liu MF (2013) Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. Hydrol Earth Syst Sci 17:837–849

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228:113–129

Haylock MR, Cawley GC, Harpham C (2006) Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. Int J Climatol 26(10):1397–1415

Islam MN, Uyeda H (2007) Use of TRMM in determining the climatic characteristics of rainfall over Bangladesh. Remote Sens Environ 108:264–276

Kan BY, Su FG, Tong K (2013) Analysis of the applicability of four precipitation datasets in the upper reaches of the yarkant river, the Karakorum. J Glaciol Geocryol (China) 35(3):710–722

Langella G, Basile A, Bonfante A, Terribile F (2010) High-resolution space–time rainfall analysis using integrated ANN inference systems. J Hydrol 387:328–342

Li M, Shao QX (2010) An improved statistical approach to merge satellite rainfall estimates and raingauge data. J Hydrol 385:51–64

Li L, Hong Y, Wang JH, Adler RF, Policelli FS, Habib S (2009) Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational flood prediction system in Nzoia basin, Lake Victoria, Africa. Nat Hazards 50:109–123

Li XH, Zhang Q, Xu CY (2012) Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. J Hydrol 426–427:28–38

Liu W, Fu G, Liu C, Charles SP (2012) A comparison of three multi-site statistical downscaling models for daily rainfall in the North Chain Plain. Theor Appl Climatol 111:585–600

Loader C (2004) Smoothing: local regression techniques. In: Gentle JE, Hardle WK, Mori Y (eds) Handbook of computational techniques. Springer, Berlin

Mehrotra R, Sharma A, Cordery I (2004) Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrence. J Geophys Res 109:D14107

Michaelides S, Levizzani V, Anagnostou E, Bauer P, Kasparis T, Lane JE (2009) Precipitation: measurement, remote sensing, climatology and modeling. Atmos Res 94:512–533

Moss R, Babiker M, Brinkman S, Calvo E, Carter T, Edmonds J, Elgizouli I, Emori S, Erda L, Hibbard K, Jones R, Kainuma M, Kelleher J, Lamarque JF, Manning M, Matthews B, Meehl J, Meyer L, Mitchell J, Nakicenovic N, O'Neill B, Pichs R, Riahi K, Rose S, Runci P, Stouffer R, van Vuuren D, Weyant J, Wilbanks T, van Ypersele JP, Zurek M (2008) Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies. Technical Summary. Intergovernmental Panel on Climate Change, Geneva

Sakia RM (1992) The Box-Cox transformation technique: a review. J R Stat Soc D 41:169–178

Samper J, Li YM, Pisani B (2014) An evaluation of climate change impacts on groundwater flow in the Plana de La Galera and Tortosa alluvial aquifers (Spain). Environ Earth Sci. doi:10.1007/s12665-014-3734-3

Su F, Hong Y, Lettenmaier DP (2008) Evaluation of TRMM multisatellite precipitation analysis (TMPA) and its utility in hydrologic prediction in the La Plata basin. J Hydrometeorol 9:622–640

Swenson S, Wahr J (2009) Monitoring the water balance of Lake Victoria, East Africa, from space. J Hydrol 370:163–176

Tareghian R, Rasmussen PF (2013) Statistical downscaling of precipitation using quantile regression. J Hydrol 487:122–135

Vrieling A, Sterk G, de Jong SM (2010) Satellite-based estimation of rainfall erosivity for Africa. J Hydrol 395:235–241

Wang SW, Li WJ (2007) Climate of China. China Meteorol Press, Beijing

Wang CL, Yue TX, Fan ZM, Zhao N (2012) HASM-based climatic downscaling model over China. J GEO-Inf Sci (China) 14(5):599–610

Ward E, Buytaert W, Peaver L, Wheater H (2011) Evaluation of precipitation products over complex mountainous terrain: a water resources perspective. Adv Water Resour 34(10):1222–1231

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21(4):530–548

Yue TX (2011) Surface modeling: high accuracy and high speed methods. CRC Press, New York

Yue TX, Zhao N, Ramsey RD, Wang CL, Fan ZM, Chen CF, Lu YM, Li BL (2013) Climate change trend in China, with improved accuracy. Clim Change 120:137–151

Zhao N, Yue TX (2014) A modification of HASM for interpolating precipitation in China. Theor Appl Climatol 116:273–285