

A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain

Wenbin Liu · Guobin Fu · Changming Liu ·
Stephen P. Charles

Received: 26 March 2012 / Accepted: 4 June 2012 / Published online: 21 June 2012
© Springer-Verlag 2012

Abstract Three statistical downscaling methods (conditional resampling statistical downscaling model: CR-SDSM, the generalised linear model for daily climate time series: GLIMCLIM, and the non-homogeneous hidden Markov model: NHMM) for multi-site daily rainfall were evaluated and compared in the North China Plain (NCP). The comparison focused on a range of statistics important for hydrological studies including rainfall amount, extreme rainfall, intra-annual variability, and spatial coherency. The results showed that no single model performed well over all statistics/timescales, suggesting that the user should choose appropriate methods after assessing their advantages and limitations when applying downscaling methods for particular purposes. Specifically, the CR-SDSM provided relatively robust results for annual/monthly statistics and extreme characteristics, but exhibited weakness for some daily statistics, such as daily rainfall amount, dry-spell length, and annual wet/dry days. GLIMCLIM performed well for annual dry/wet days, dry/wet spell length, and spatial coherency, but slightly overestimated the daily rainfall. Additionally,

NHMM performed better for daily rainfall and annual wet/dry days, but slightly underestimated dry/wet spell length and overestimated the daily extremes. The results of this study could be applied when investigating climate change impact on hydrology and water availability for the NCP, which suffers from intense water shortages due to climate change and human activities in recent years.

1 Introduction

Changing climate and its impacts on water resources have gained significant attention in hydrological studies (Fu et al. 2007; Chen et al. 2010). General circulation models (GCMs) are a common used tool for the assessment of climate change, but they currently remain relatively coarse in resolution and so unable to resolve sub-grid-scale features such as topography, clouds, and land use (Grotch and MacCracken 1991; Fowler et al. 2007). In particular, GCM outputs are inadequate for capturing rainfall spatial-temporal variability, which is required for hydrological modeling (Frost et al. 2011). Downscaling is thus often used to bridge the scale mismatch gap between the GCM and regional hydrological impacts studies (Maraun et al. 2010).

Downscaling methods are commonly classified as statistical and dynamic downscaling, with statistical downscaling more widely adopted in hydrological studies due to the higher computation resource requirements of dynamic downscaling, as well as inadequate spatial resolution for convective rainfall events and the effects of terrain (Fowler et al. 2007; Chen et al. 2010). In the past two decades, various statistical downscaling models/software for rainfall (or multivariable downscaling) have

W. Liu · C. Liu
Key Laboratory of Water Cycle and Related Land Surface
Processes, Institute of Geographic Sciences and Natural Resources
Research, Chinese Academy of Sciences,
Beijing 100101, China

G. Fu (✉) · S. P. Charles
CSIRO Land and Water,
Private Bag 5,
Wembley, WA 6913, Australia
e-mail: Guobin.Fu@csiro.au

W. Liu
Graduate University of Chinese Academy of Sciences,
Beijing 10049, China

been developed (Xu 1999; Wilby and Wigley 2000; Chandler 2002; Charles et al. 2004; Fowler et al. 2005; Wetterhall et al. 2006; Mehrotra and Sharma 2007; Chen et al. 2010; Chiew et al. 2010), but no single model has been found to perform well over all statistics/timescales/applications. Consequently, comparisons of different methods are very important to understand under which conditions these methods can be applied. Comparisons of different statistical downscaling methods for precipitation have been conducted in many countries and regions (Dibike and Coulibaly 2005; Khan et al. 2006; Timbal et al. 2008a; Tryhorn and DeGaetano 2010; Liu et al. 2011; Raje and Mujumdar 2011; Frost et al. 2011). However, few studies have considered the intermittency structure and daily to monthly spatial correlation of rainfall. Moreover, it is also unclear whether the total monthly, seasonal, and site-to-site variations, as required for hydrologic modeling, can be adequately reproduced by these models (Frost et al. 2011).

In China, there have been several statistical downscaling exercises for precipitation (Liao et al. 2004; Wetterhall et al. 2006; Chu et al. 2010; Chen et al. 2010; Liu et al. 2011), but there has not been a comparison of the relative performance of multi-site approaches of relevance to hydrological performance. Wetterhall et al. (2006) compared four statistical downscaling methods (two analogue methods, SDSM and a fuzzy-rule-base weather-pattern classification method: MOFRBC) on three catchments located in southern, eastern, and central China, and demonstrated that all methods capture the annual precipitation cycle, with SDSM and MOFRBC performing overall better than the analogue methods. Liu et al. (2011) compared the performance of SDSM and NHMM over an arid basin in northwest China, and determined both models showed stability with little model performance difference. However, these comparisons only used the single-site SDSM, and so did not comprehensively consider site to site correlation. Additionally, there has not been a comparison of different statistical downscaling methods for the North China Plain, where the precipitation is strongly governed by the East Asian Monsoon and that now suffers from intense water shortage (Fu et al. 2009).

The intention of this paper is to focus on an evaluation of three multi-site statistical downscaling methods (CR-SDSM, NHMM, and GLIMCLIM) in the NCP. The paper is organized as follows. The rainfall, reanalysis data, and three statistical downscaling methods used in this study are first described in Section 2. Section 3 presents the model results, followed by a discussion of each model's performance. The conclusion and proposed future research are presented in Section 4.

2 Datasets and methodology

2.1 Datasets and predictor selection

2.1.1 Observed rainfall

The study area is the North China Plain (Fig. 1), which is also known as the Huang-Huai-Hai Plain after the three major rivers that traverse it. As China's most important social, economic, and agricultural region, the NCP produces about one fourth of the country's total grain yield, and currently experiences intense water shortages and related environmental problems (Fu et al. 2009).

Observed daily rainfall from 40 weather stations (Fig. 1 and Table 1) chosen for this study was acquired from the China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn>). All station records used have complete series for the entire period (1961–2010) and have passed NMO data quality control. Observed daily precipitation less than 1.0 mm was set to zero to eliminate the impact of inconsistencies in the observation due to trace rainfall amounts (Frost et al. 2011). This threshold was also used to determine whether a day is classified as dry or wet in calculating indices.

2.1.2 Reanalysis data for atmospheric predictors

The predictor variables used in this study were from the large-scale reanalysis datasets obtained from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR, <http://www.cdc.noaa.gov/cdc/reanalysis/>). Thirty daily predictors (1961–2010) such as sea level pressure, temperature, geopotential height, wind speed and direction, and specific humidity at pressure (500, 700, and 850 hPa) and surface levels were selected as candidate predictors. Relevant predictors were then extracted for a seven by six array of grid cells ($2.5^\circ \times 2.5^\circ$) covering the chosen rainfall sites (Fig. 1). Furthermore, all candidate predictors were standardized before statistical downscaling by subtracting the long-term mean and dividing by the long-term standard deviation as:

$$\hat{u}_t = \frac{u_t - \bar{u}}{\delta_u} \quad (1)$$

Where \hat{u}_t is the normalized predictor at time t , \bar{u} is the multiyear average during study period, and δ_u is the standard deviation of u for the study period.

2.1.3 Predictor selection

The choice of predictor variable(s) is one of the most critical steps in the development of a statistical downscaling scheme

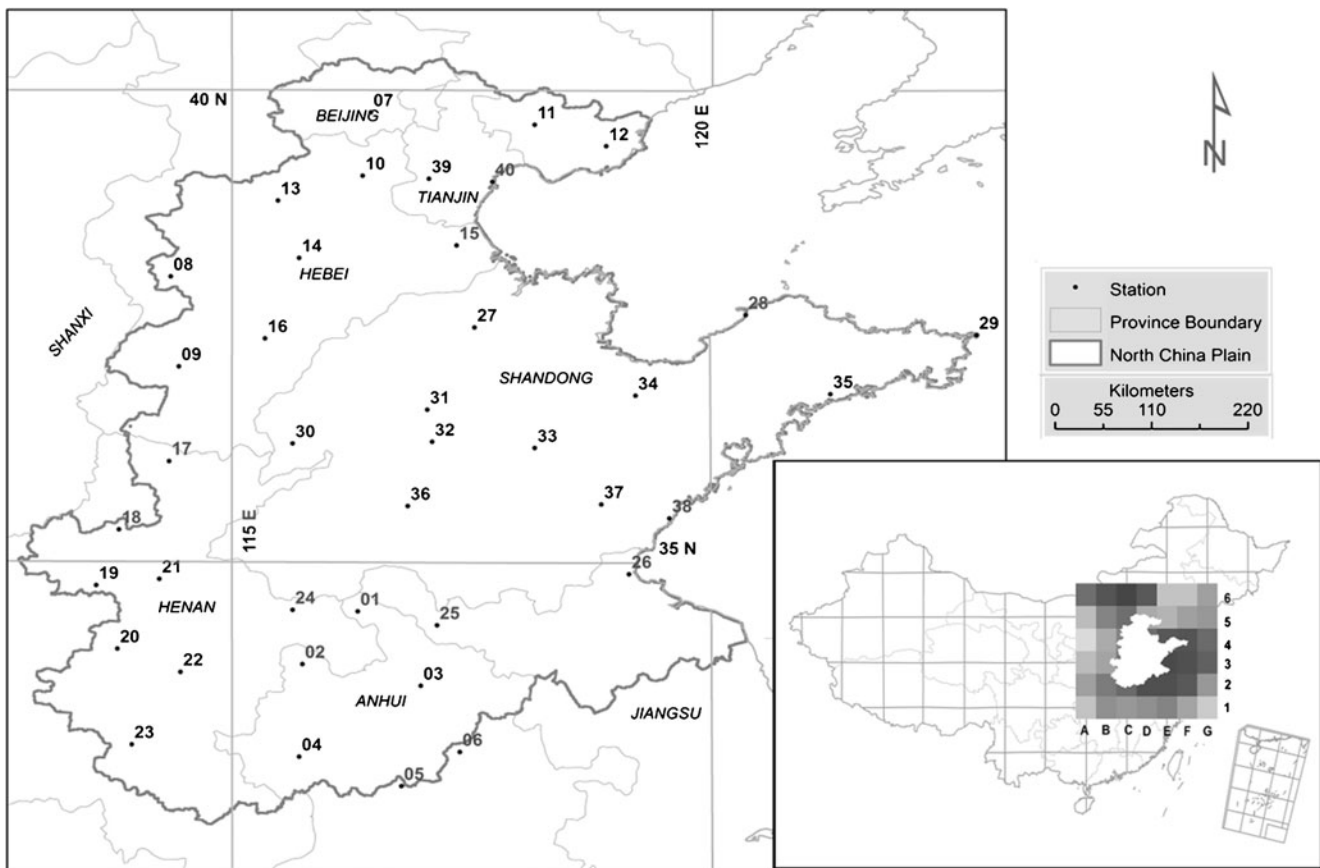


Fig. 1 Map of the NCP showing the location of the climate stations and NCEP grids used in this study (created by ArcGIS software)

because the decision largely determines the characteristics of the downscaled scenario. The selection process is complicated by the fact that the explanatory power of individual predictor variable may be low, or the power varies both spatially and temporally (Wilby et al. 2004). The basic requirements are that the selected predictors must be strongly correlated with the predictand, physically sensible, realistically represented by GCM, and multiyear variability captured (Wilby and Wigley 2000; Wilby et al. 2004; Gachon and Dibike 2007; Liu et al. 2011). Additionally, the impacts of different regions and seasons on predictor selection should also be considered (Timbal et al. 2008b). Simple procedure such as partial correlation analysis, step-wise regression, or information criteria may be used to screen most promising predictor variables from a candidate suite (Wilby et al. 2004), and the commonly used predictors in daily precipitation statistical downscaling are circulation variables, temperature and relative humidity (e.g., dew point temperature depression).

The procedure adopted for selecting suitable predictors in this study is as follows: (1) The potential variables are extended to more than 4,000 predictors by calculating the gradients between two grid cells (including north–south, west–east, northwest–southeast, and northeast–southwest), basing on the 30 candidate predictors mentioned in Section 2.1.2; (2) The

Pearson partial correlation is used to screen the most promising variables. This leads to twelve predictors (Table 2) being selected for wet season (April to September) and dry season (October to next March). This is because the atmosphere circulation features in NCP, which are strongly controlled by the East Asian monsoon, are quite different between wet season and dry season (Chu et al. 2010). The selected predictors were directly used in SDSM and NHMM. For GLIMCLIM, the predictors are further validated through the likelihood ratio statistics and residual analysis while the occurrence and amount models are being fitted. Besides these predictors, other predictors reflecting seasonality, autocorrelation, inter-site dependence, etc. are also used for GLIMCLIM (Table 3).

2.2 Model descriptions

2.2.1 CR-SDSM

The statistical downscaling model (SDSM) is a hybrid between a regression-based method and a stochastic weather generator, in which the local-scale weather generator parameters are linearly conditioned by large-scale predictors at individual sites (Wilby et al. 2003). It can be described as (Wilby et al. 2003; Chu et al. 2010):

Table 1 Rainfall stations in NCP selected in this study

Label	Station ID	Name	Latitude	Longitude	Altitude (m)
01	58015	Dangshan	34°26'	116°20'	44.2
02	58102	Haozhou	33°52'	115°46'	37.7
03	58122	Suxian	33°38'	116°59'	25.9
04	58203	Fuyang	32°52'	115°44'	32.7
05	58215	Shouxian	32°33'	116°47'	22.7
06	58221	Bengbu	32°55'	117°23'	21.9
07	54511	Beijing	39°48'	116°28'	31.3
08	53698	Shijiazhuang	38°02'	114°25'	81
09	53798	Xingtai	37°04'	114°30'	77.3
10	54518	Langfang	39°07'	116°23'	9
11	54534	Tangshan	39°40'	118°09'	27.8
12	54539	Leting	39°26'	118°53'	10.5
13	54602	Baoding	38°51'	115°31'	17.2
14	54606	Raoyang	38°14'	115°44'	19
15	54624	Huanghua	38°22'	117°21'	6.6
16	54705	Nangong	37°22'	115°23'	27.4
17	53898	Anyang	36°03'	114°24'	62.9
18	53986	Xinxiang	35°19'	113°53'	73.2
19	57083	Zhengzhou	34°43'	113°39'	110.4
20	57089	Xuchang	34°02'	113°52'	66.8
21	57091	Kaifeng	34°47'	114°18'	73.7
22	57193	Xihua	33°47'	114°31'	52.6
23	57290	Zhumadian	33°00'	114°01'	82.7
24	58005	Shangqiu	34°27'	115°40'	50.1
25	58027	Xuzhou	34°17'	117°09'	41.2
26	58040	Ganyu	34°50'	119°07'	3.3
27	54725	Huiminxian	37°29'	117°32'	11.7
28	54753	Longkou	37°37'	120°19'	4.8
29	54776	Chengshantou	37°24'	122°41'	47.7
30	54808	Chaoyang	36°14'	115°40'	37.8
31	54823	Jinan	36°36'	117°03'	170.3
32	54826	Taishai	36°15'	117°06'	1,533.7
33	54836	Yiyuan	36°11'	118°09'	305.1
34	54843	Weifang	36°45'	119°11'	22.2
35	54863	Haiyang	36°46'	121°11'	40.9
36	54916	Yanzhou	35°34'	116°51'	51.7
37	54936	Lvxian	35°35'	118°50'	107.4
38	54945	Rizhao	35°26'	119°32'	36.9
39	54527	Tianjin	39°05'	117°04'	2.5
40	54623	Tanggu	39°03'	117°43'	4.8

$$\omega_t = \alpha_0 + \sum_{j=1}^n \alpha_j \hat{u}_t^{(j)} \quad (2)$$

Where ω_t is the conditional probability of rainfall occurrence on day t , $\hat{u}_t^{(j)}$ is the normalized atmospheric predictor, and α_j is regression coefficients calculated using least

squares regression. The rainfall occurrence was determined by a uniformly distributed random number r_t ($0 \leq r_t \leq 1$), if the rainfall occurs ($\omega_t \leq r_t$), the rainfall amount can be expressed by a z score as:

$$Z_t = \beta_0 + \sum_{j=1}^n \beta_j \hat{u}_t^{(j)} + \beta_{t-1} + \varepsilon \quad (3)$$

in which Z_t is the z score, β_j is the regression parameter estimated using least squares regression, and ε is a error term represented by the normal distribution $\varepsilon \sim N(0, \delta^2)$, and:

$$Z_t = \phi^{-1}[F(y_t)] \quad (4)$$

Where ϕ is the normal cumulative distribution function and $F(y_t)$ is the empirical distribution function of y_t . More detailed information on SDSM can refer to these studies (Hay et al. 2000; Wilby et al. 2002; Dibike and Coulibaly 2005; Khan et al. 2006; Chu et al. 2010; Liu et al. 2011). Detailed technical information on SDSM can be found in Wilby et al. (2002) and the corresponding software toolkit can be downloaded from <http://co-public.lboro.ac.uk/cocwd/SDSM/main.html>.

SDSM is best described as a single-site model, but it can be extended to multi-site applications via conditional resampling (CR-SDSM, Wilby et al. 2003; Harpham and Wilby 2005). Applying SDSM to multi-site daily rainfall downscaling includes two steps: (1) the daily rainfall at a “marker” site (in this study, the area average amounts) is first downscaled by the single-site SDSM; (2) Daily rainfall amounts are then “resampled from the empirical distribution of area averages, conditional on the large-scale atmospheric forcing and the stochastic error term. The actual daily rainfall is determined by mapping the modeled normal cumulative distribution value onto the observed cumulative distribution of amounts at the marker site” (Wilby et al. 2003). Ultimately, the marker site rainfall is resampled to the constituent amount falling on the same day from each station in the multi-sites array (Harpham and Wilby 2005).

Thus, if the marker series is based on an unweighted average of all sites, the conditional resampling will preserve both the areal average of the marker series and the spatial covariance of the multi-site rainfall (Wilby et al. 2003). Additionally, using area average, instead of individual sites as the marker series, reduces the risk of employing a non-homogeneous/non-representative record and increases the signal to noise ratio of the predictand (Wilby et al. 2003).

2.2.2 NHMM

The non-homogeneous hidden Markov model (NHMM) relates the atmospheric predictors to point rainfall at multi-sites using a hidden weather state process (Bates et al. 1998;

Table 2 Candidate predictors selected from the NCEP datasets

Predictors	Description	Grid
Wet season		
Temp NW–SE	Surface temperature northwest to southeast gradient	B4–C3
DPT500	Dew point temperature depression at 500 hPa	(D1+D2+E2)/3
SHUM 500	Specific humidity at 500 hPa	(F3+G3+G4)/3
MSLP	Mean sea level pressure	(A2+A3+B2+B3)/4
UWD 850 N–S	Zonal velocity component at 850 hPa north to south gradient	((F5–F4)+(G5–G4))/2
VWD 700	Meridional velocity component at 700 hPa	(E2+F2)/2
Dry-season		
Temp 850 W–E	Temperature at 850 hPa west to east gradient	((D3–E3)+(D4–E4)+(E3–F3)+(E4–F4))/4
DPT 500	Dew point temperature depression at 500 hPa	(D1+D2+E1+E2)/4
SHUM 700	Specific humidity at 700 hPa	(D1+E1+E2+F2+F3+G3)/6
MSLP N–S	Mean sea level pressure north to south gradient	((A3–A2)+(B3–B2))/2
UWD850	Zonal velocity component at 850 hPa	(F3+F4+G4)/3
VWD NW–SE	Surface meridional velocity component northwest to southeast gradient	A2–B1

Hughes et al. 1999; Charles et al. 1999, 2003, 2004, 2007; Chiew et al. 2010; Frost et al. 2011; Fu and Charles 2011).

The NHMM models multi-site patterns of daily rainfall as a finite number of “hidden” (i.e., unobserved) weather states, and the temporal evolution of these daily states is modeled as a first-order Markov process with state-to-state transition probabilities conditioned on a small number of synoptic-scale atmospheric predictors (Fu and Charles 2011). Generally, the NHMM can be expressed by the following assumptions (Charles et al. 2004):

$$P(R_t|S_1^T, R_1^{t-1}, X_1^t) = P(R_t|S_t) \quad (5)$$

$$P(S_t|S_1^{t-1}, X_1^T) = P(S_t|S_{t-1}, X_t) \quad (6)$$

in which R_t denotes a multivariate vector giving rainfall occurrences at an n stations' network at time t , X_t is the vector of atmospheric measures at time t ($1 \leq t \leq T$), and S_t presents the weather state at time t . The notation X_1^t is used to present the sequence of atmospheric data (from time 1 to T) and similar for S_1^T and R_1^t , and specific NHMMs are defined by the parameterizations chosen for $P(R_t|S_t)$ and $P(S_t|S_{t-1}, X_t)$ (Hughes et al. 1999; Charles et al. 2004). The first assumption states that the rainfall process is conditionally independent given the current weather state and the second assumption states that the rainfall process depends only on the previous weather state and the current atmospheric data (Charles et al. 2004).

The most appropriate number of hidden states is estimated via the Bayes Information Criterion (BIC, Robertson et al. 2004). Conditioned on the state process, rainfall at a network of stations is modeled using tree averaged multivariate copulas as described in Kirshner (2007).

A detailed description of the current-generation NHMM, including its assumptions, mathematical parameterizations, and estimation algorithms can be found in Kirshner (2005), with a corresponding software toolkit available at: <http://www.stat.purdue.edu/~skirshne/MVNHMM/>.

2.2.3 GLIMCLIM

Generalized linear models (GLMs) are an extension of classical regression and are well established in the statistical literature (Chandler 2002; Yang et al. 2005; Yan et al. 2006). The GLIMCLIM model (generalised linear model for daily climate time series) provides an alternative conceptualization of the rainfall process (Chandler 2002; Frost et al. 2011), modeling rainfall occurrence using logistic regression and rainfall amounts using a gamma distribution with a common dispersion parameter.

The logistic regression can be described as follows (Chandler 2002, Yang et al. 2005):

$$\ln \frac{p_i}{1-p_i} = x_i' \beta \quad (7)$$

Where p_i is the rainfall probability for the i th case in the dataset conditional on a covariate vector X_i with coefficient

Table 3 GLIMCLIM model (occurrence and amount) specification

GLIMCLIM model fitted parameter	Parameter name	
Occurrence model	Constant	
	1. Legendre polynomial 1 for eastings	
	2. Legendre polynomial 1 for northings	
	3. Legendre polynomial 2 for eastings	
	4. Legendre polynomial 2 for northings	
	5. Legendre polynomial 3 for eastings	
	6. Legendre polynomial 3 for northings	
	7. Daily seasonal effect, cosine component	
	8. Daily seasonal effect, sine component)	
	9. Ln(1+Y[t-1])	
	10. TEMP NW-SE (wet)/TEMP 850 W-E (dry)	
	11. DPT 500	
	12. SHUM 500 (wet)/SHUM700 (dry)	
	13. MSLP	
	14. UWND 850 (wet)/UWND 850 N-S (dry)	
15. VWND NW-SE (wet)/VWND700 (dry)		
Occurrence interactions	1-2, 1-14, 2-8, 2-9, 2-10, 2-11, 2-13, 2-14, 3-10, 3-13, 4-11, 4-12, 4-14, 5-14, 7-8, 7-9, 7-11, 7-12, 7-13, 8-10, 8-11, 8-12, 9-14, 11-12, 11-13, 12-15	
	Occurrence spatial dependence	Spatial dependence parameter (powered exponential correlation function)
		Amount model
	16. Legendre polynomial 1 for eastings	
	17. Legendre polynomial 1 for northings	
18. Legendre polynomial 2 for eastings		
19. Legendre polynomial 2 for northings		
20. Legendre polynomial 3 for eastings		
21. Legendre polynomial 3 for northings		
22. Daily seasonal effect, cosine component		
23. Daily seasonal effect, sine component		
24. Ln(1+Y[t-1])		
25. TEMP NW-SE (wet)/TEMP 850 W-E (dry)		
26. DPT 500		
27. SHUM 500 (wet)/SHUM700 (dry)		
28. MSLP		
29. UWND 850 (wet)/UWND 850 N-S (dry)		
30. VWND NW-SE (wet)/VWND700 (dry)		
Amount interactions	2-10, 2-11, 2-14, 7-13, 7-11, 8-11, 10-11, 12-13	
Amount spatial dependence	Spatial dependence parameter (powered exponential correlation function)	

The explanations of terms please see Chandler (2002)

vector β . The rainfall amount for i th wet month has, conditional on a covariate vector ξ_i and coefficient vector γ , a gamma distribution with mean μ_i , where

$$\ln(\mu_i) = \xi_i' \gamma \quad (8)$$

The shape parameter of the gamma distribution (ν) is assumed constant for all observations.

To describe the climatology of the region, other covariates representing spatial dependence, seasonal variation, interactions terms and persistence should also be included in the occurrence and amount models in GLIMCLIM. Moreover, the GLIMCLIM can output rich information to check for unexplained structure, mean Pearson residuals. For an observation Y_i , the Pearson residual is described as follows (Chandler 2005):

$$r_i = \frac{(Y_i - \mu_i)}{\delta_i} \quad (9)$$

in which Y_i denotes the observed response for the i th case, μ_i is the modeled mean and δ_i is the standard deviation. If the model is correctly fitted, the Pearson residuals should all come from distribution with mean 0 and variance 1 (Ambrosino et al. 2010).

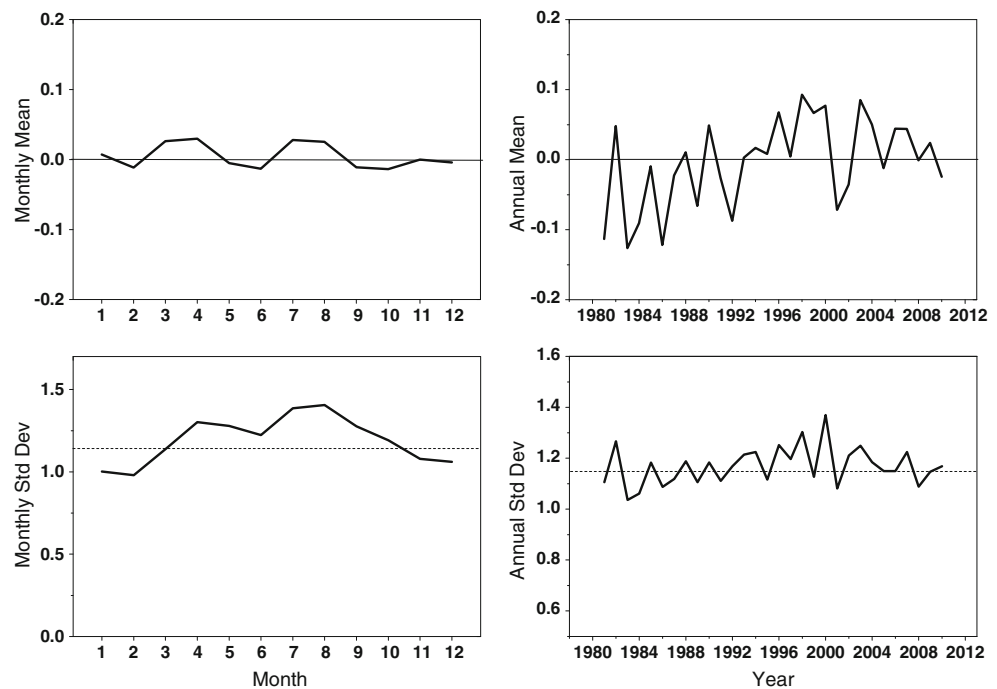
Refer to Chandler and Wheater (2002) and Yang et al. (2005) for further details, and it can be freely download from http://www.homepages.ucl.ac.uk/~ucakarc/work/rain_glm.html.

2.3 Model calibration and verification

All three models were calibrated on a half year basis (wet season and dry season). The period 1981–2010 was chosen for calibration due to the availability of high-quality rainfall data, while the period 1961–1980 was chosen for validation. The SDSM was only built at the “marker” site, and then the single-site results were extended to multi-site synthesis of daily rainfall via conditional resampling. We used the fourth root transformation to convert the original rainfall to a normal distribution, and the ordinary least square method for optimization (Wilby et al. 2002). The percentages of explained variance in the “maker” site were 40.8 % for dry season and 36.1 % for wet season. Moreover, the determination coefficients in the calibration period were 0.729 for dry season and 0.674 for wet season. For the NHMM model, the main step is to choose the appropriate number of hidden states from a fitted NHMM by using the BIC. When the BIC reaches its minimum value, the corresponding hidden states are chosen. In this study, the numbers of hidden state are five for wet season (log-likelihood of data set, $-2.099055e+05$; log-posterior of the data set, $-2.103359e+05$) and four for dry season (log-likelihood of data set, $-6.914106e+04$, log-posterior of data set: $-6.958825e+04$), respectively.

For either logistic or gamma model in GLIMCLIM, a baseline model was firstly developed using some basic

Fig. 2 The monthly and annual means and standard deviations of Pearson residuals from GLIMCLIM amount model for the calibration period (1981–2010). The *dashed lines* show the standard deviations expected under the model



factors influencing precipitation variability (i.e., seasonal and geographical factors). Different terms were progressively added following a perceived order of importance and the most significant candidates were added to the new model (Ambrosino et al. 2010). The likelihood ratio test and Pearson residual means were used to check the statistical significance and the model structure. A complex model composed of basic factors, circulation predictors, interaction terms, and spatial dependence structure (powered exponential correlation function: $\phi=0.6587$; $\kappa=0.7002$) is finally fitted. As mentioned in Section 2.2.3, the fit of either model can be assessed by the mean Pearson residuals. The mean and standard deviation of Pearson residuals in the logistic (occurrence) model are 0.0001 and 1.0038 for dry season and are -0.0004 and 1.022 for wet season, while they are 0.000 and 1.091 for dry season and 0.000 and 1.129 for wet season, respectively, for the amount model. In addition, the monthly and annual residual plots for the amounts model are shown in Fig. 2, which suggests that the seasonal structure and trends in the rainfall sequences are overall well presented by the amount model.

The verification statistics chosen focus on the measures considered important in the runoff generation process (Frost

et al. 2011). Assessment of whether a method should be used depends on the application and performance over a range of timescales (e.g., daily, monthly and annual statistics). Additionally, several evaluation metrics, such as relative error, correlation coefficient, and quantile–quantile plot were used for comparing statistical characteristics and probability distributions of observed and simulated rainfall.

3 Results and discussion

In this section, downscaling results (100 stochastic realizations) of each model in the calibration and validation periods are presented. A range of annual, monthly, and daily statistics was calculated, chosen on the basis of providing a robust test of the model for hydrological applications.

3.1 Annual statistics

The annual mean, standard deviation (SD), and the coefficient of skewness (CS) provide a summary of whether a model can reproduce long term (e.g., water availability and

Table 4 Annual statistics for each model

	Calibration			Validation		
	Mean (mm/day)	SD (mm/day)	CS	Mean (mm/day)	SD (mm/day)	CS
Observation	1.82	8.24	9.41	1.93	8.43	9.10
CR-SDSM	1.85	8.23	9.42	1.94	8.39	9.15
NHMM	1.93	8.08	8.08	2.07	8.49	7.79
GLIMCLIM	1.89	8.21	8.92	2.10	9.01	8.84

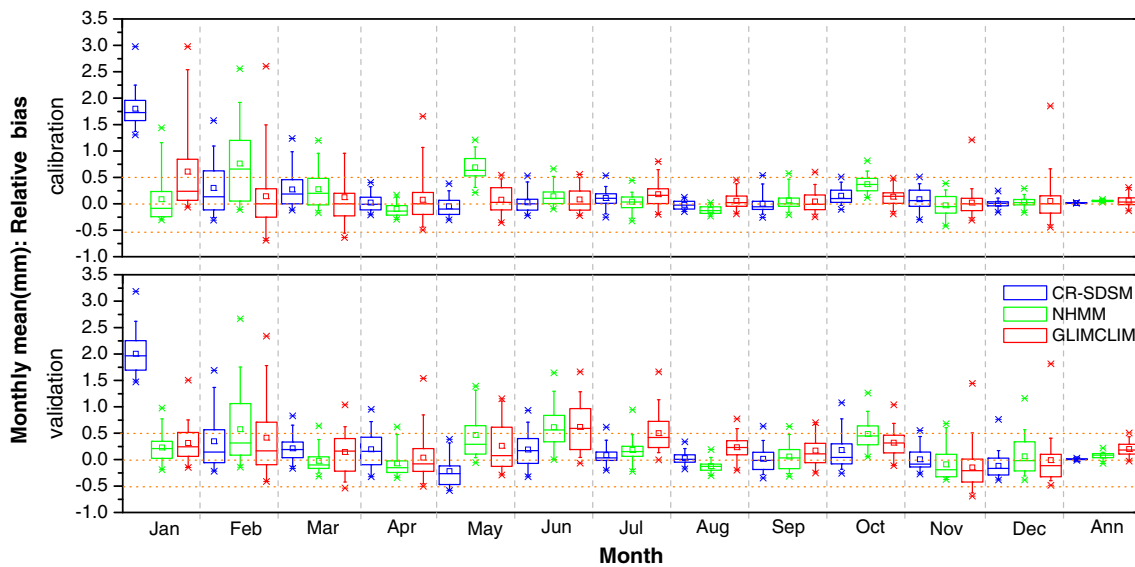


Fig. 3 Monthly and annual mean for each model. (Relative bias to observed values across all sites) (created by Origin)

drought) characteristics (Frost et al. 2011). The overall mean, SD and CS are generally reproduced well although all three models overestimated the annual mean and underestimated the SD, except for NHMM and GLIMCLIM in the validation period (Table 4). The CR-SDSM yielded better results for both annual mean and SD (Table 4). These results are also consistent with the relative errors for annual mean and SD in Figs. 3 and 4. CR-SDSM estimated the CS, which relates to the occurrence of extreme values, much more accurately than NHMM and GLIMCLIM (Table 4), indicating that the CS-SDSM outperforms the NHMM and GLIMCLIM in downscaling annual extremes. Additionally, the performances of GLIMCLIM and NHMM for annual means are consistent with the findings of Frost et al. (2011).

Figure 5 shows both GLIMCLIM and NHMM overestimated the annual rainfall in the calibration and validation periods. Additionally, both GLIMCLIM and NHMM reproduced reasonable annual dry days/wet days, but CR-SDSM grossly overestimated the number of annual wet days, with corresponding underestimation of annual dry days (Fig. 5). This could be due to either an insufficient pool of existing rainfall days for conditional resampling or the metric used to choose rainfall days for the given climate predictors inadequately reproduces observed daily variability (Frost et al. 2011). So although CR-SDSM performed relatively well for annual rainfall, it appears that the model produces rainfall on too many days, with overall underestimation of amount on those days (see Fig. 12).

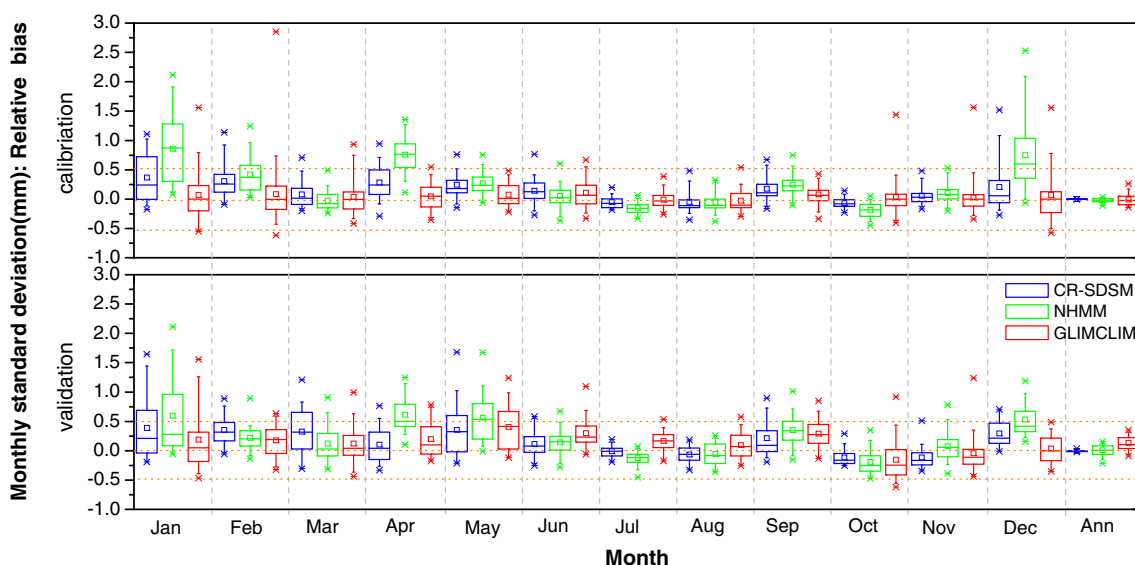


Fig. 4 Monthly and annual standard deviation for each model. (Relative bias to observed values across all sites) (created by Origin)

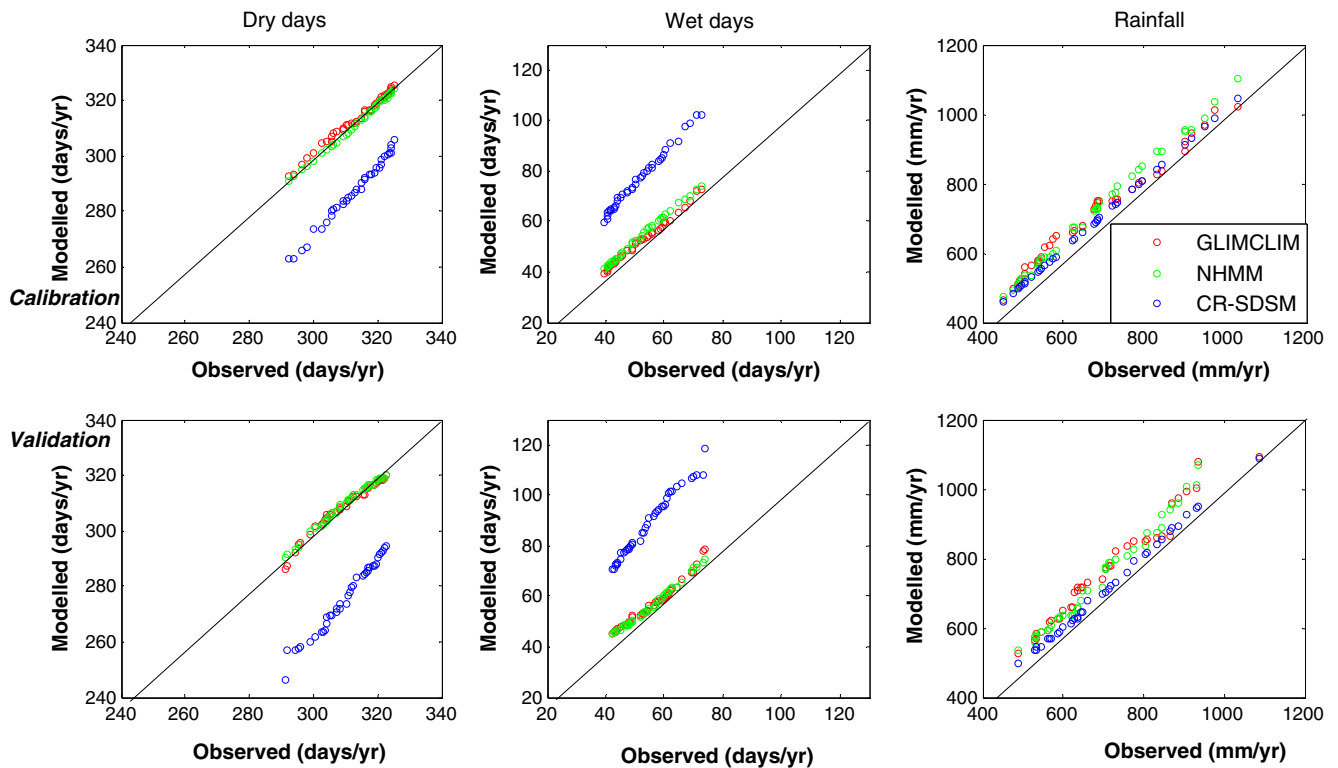


Fig. 5 Quantile–quantile plots of observed and simulated annual dry days, annual wet days, and annual rainfall (created by Matlab)

3.2 Monthly statistics

The monthly statistics (e.g., seasonality, intra-annual variation) are also important for water availability and drought studies. Relative biases of monthly rainfall are exhibited in Fig. 3. Generally, CR-SDSM performed the best among all three models in most months, except for January where it markedly overestimated the mean rainfall. Moreover, compared to

NHMM, GLIMCLIM performed better in most months (e.g., February, March, April, May, June, August, September, and October) in the calibration period but performed relatively worse in the validation period (e.g., January, March, June, July, August, September, November, and December).

GLIMCLIM better reproduced monthly standard deviations in the calibration period compared to the other two models (Fig. 4), with the median relative errors for all

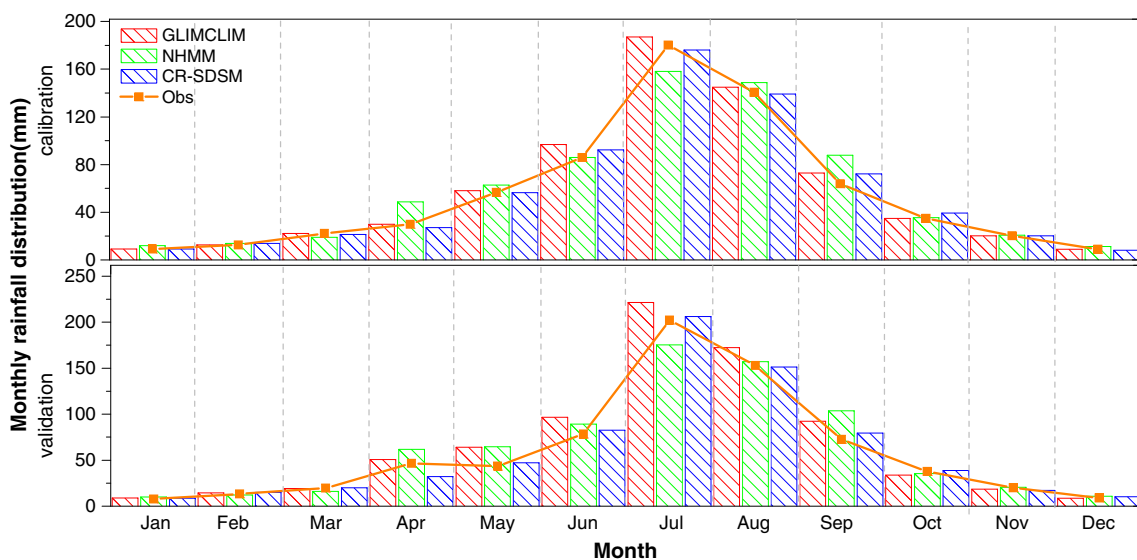
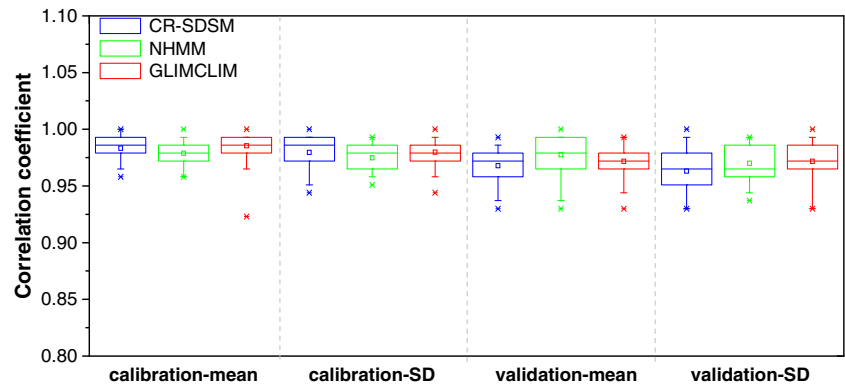


Fig. 6 Monthly rainfall distribution for each model (created by Origin)

Fig. 7 Box plots of intra-annual correlation coefficients for both mean and standard deviation for each observed station/month (created by Origin)



stations lower than 15 % in all months. Moreover, CR-SDSM performed relatively better than NHMM in all months except June and August. In the validation period, CR-SDSM performed well relative to the other two models in most months. The monthly performance of NHMM and GLIMCLIM are also consistent with the results of Frost et al. (2011) for Australia condition.

For hydrological applications, it is essential that simulations capture the monthly distribution and intra-annual

variability of rainfall. In Fig. 6, all three models reasonably reproduce the monthly distribution, with simulations in dry seasons somewhat better than in wet seasons. Specifically, GLIMCLIM performed relatively well in the calibration period, but tended to overestimate the wet season precipitation in the validation period. The CR-SDSM performed well in both calibration and validation periods (Fig. 6), but as already noted, this encompasses self-cancelling biases in wet days (too many) and wet day amounts (too little).

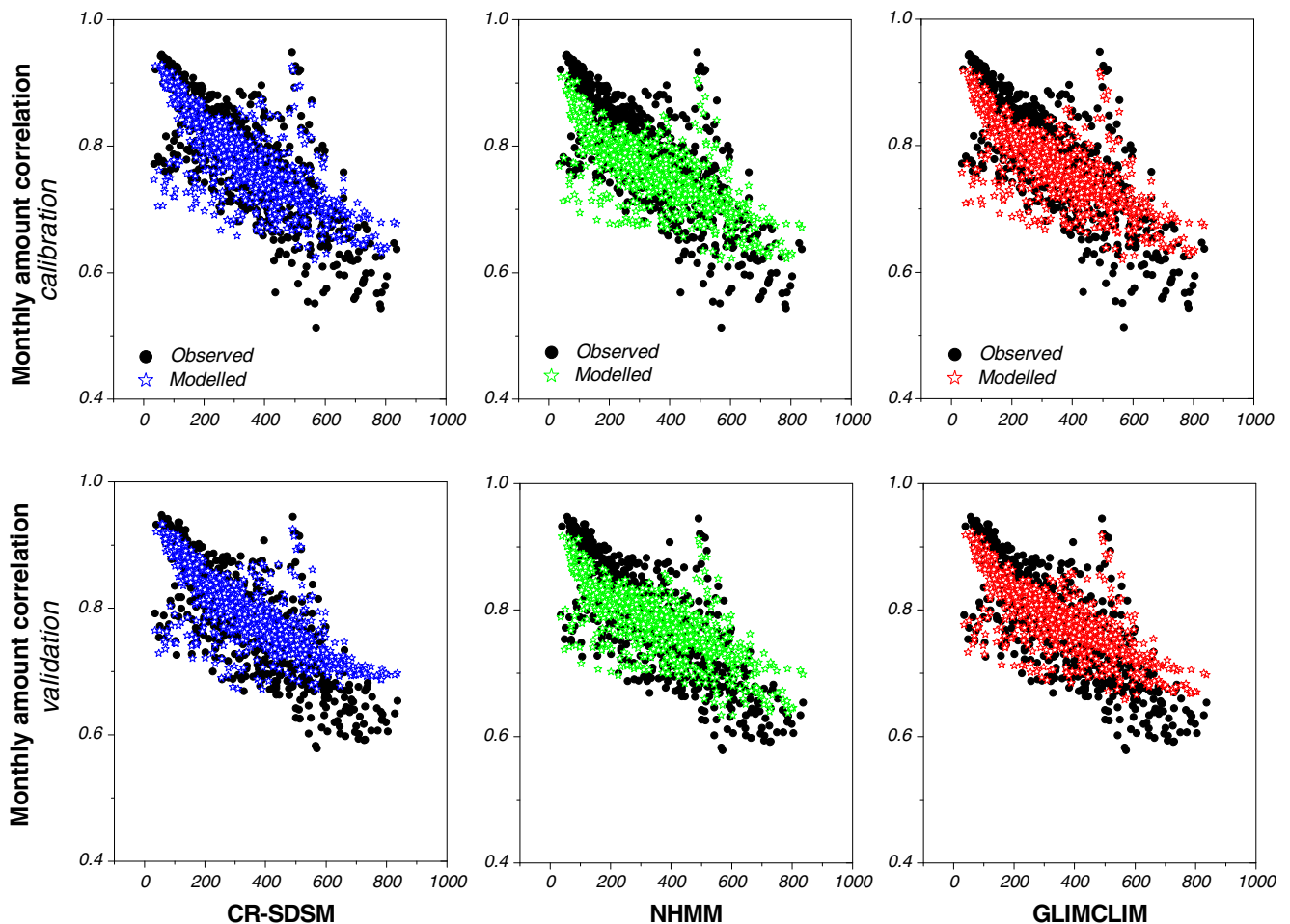


Fig. 8 Correlations between pairs of station monthly rainfall amount vs. their separation distance for all possible combinations of station pairs in NCP (created by Origin)

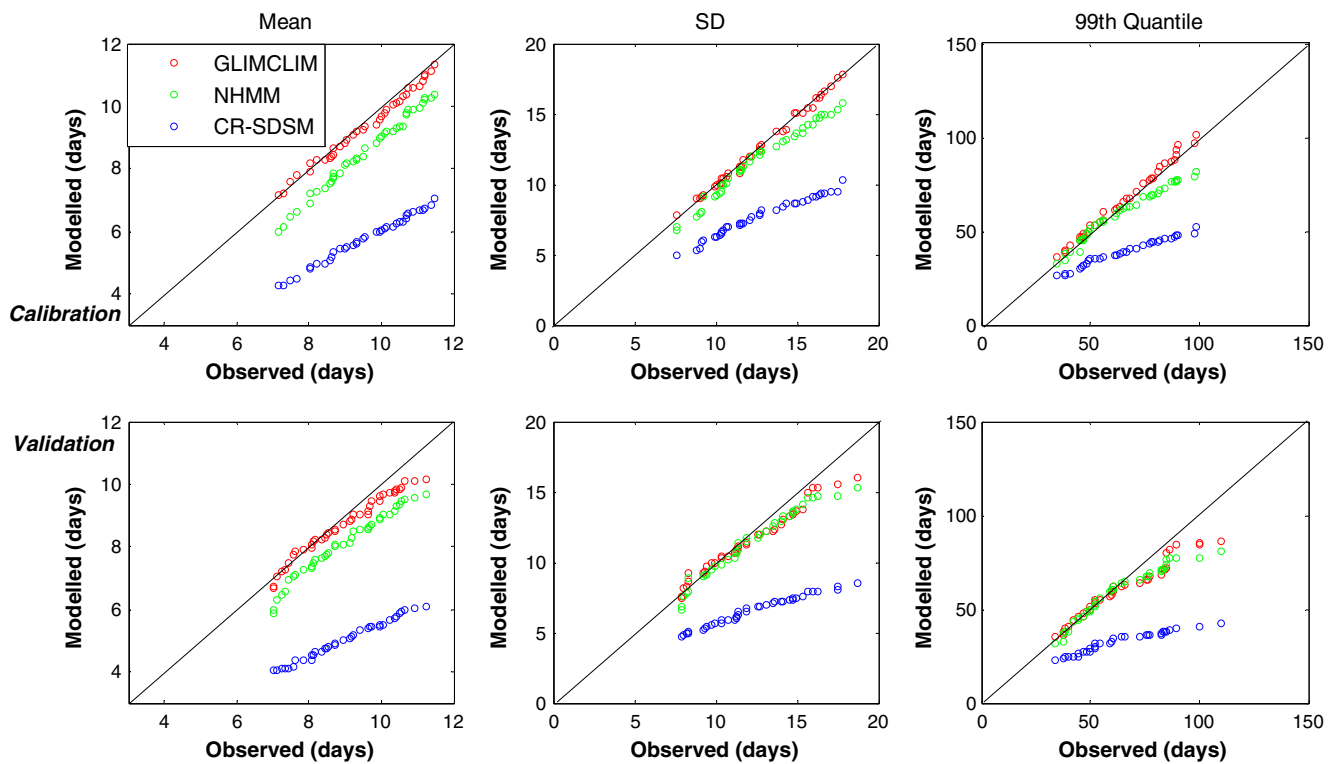


Fig. 9 Quantile–quantile plots of observed and simulated means, standard deviations, and 99th quantiles of dry-spell length (created by Matlab)

Furthermore, all models reproduce intra-annual correlations for monthly mean and standard deviation for both the calibration and validation periods, with correlation coefficients larger than 0.97 (Fig. 7). NHMM

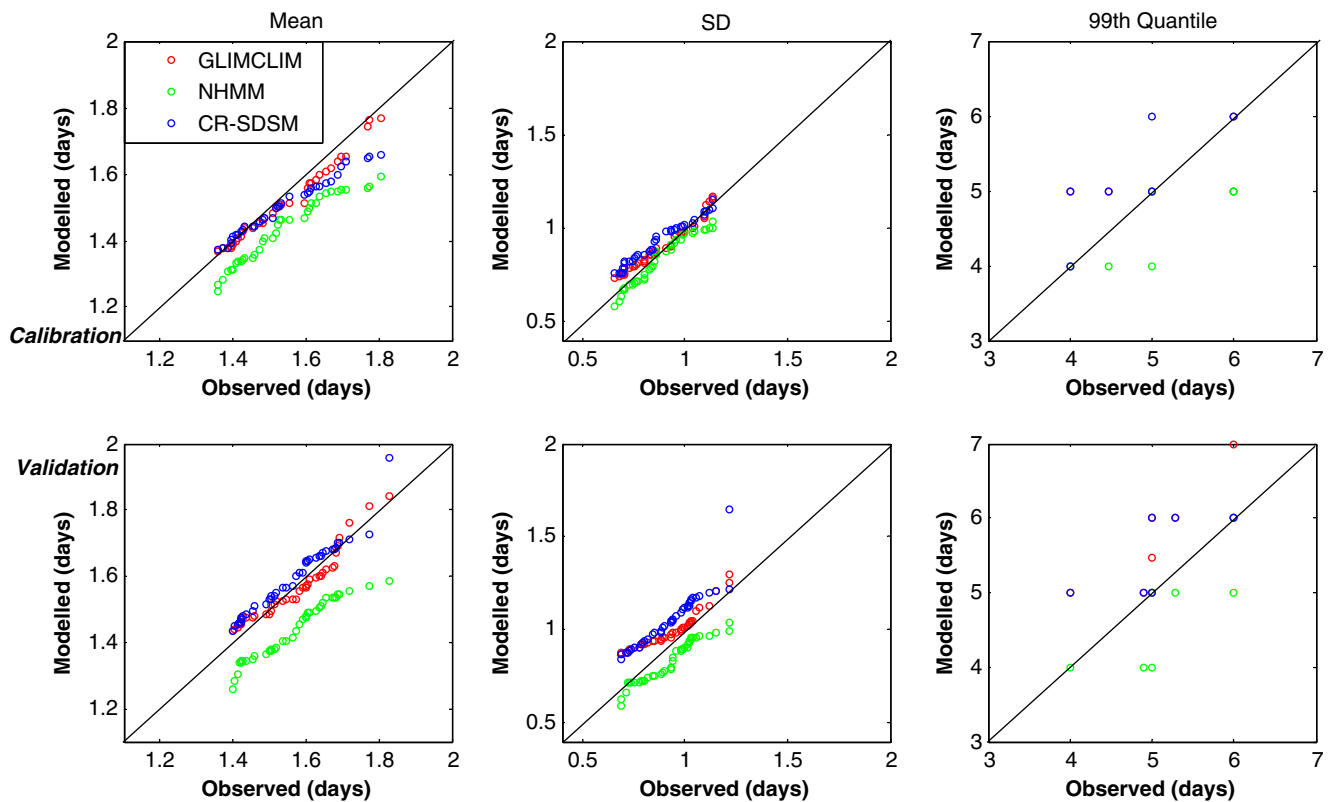


Fig. 10 Quantile–quantile plots of observed and simulated means, standard deviations, and 99th quantiles of wet-spell length (created by Matlab)

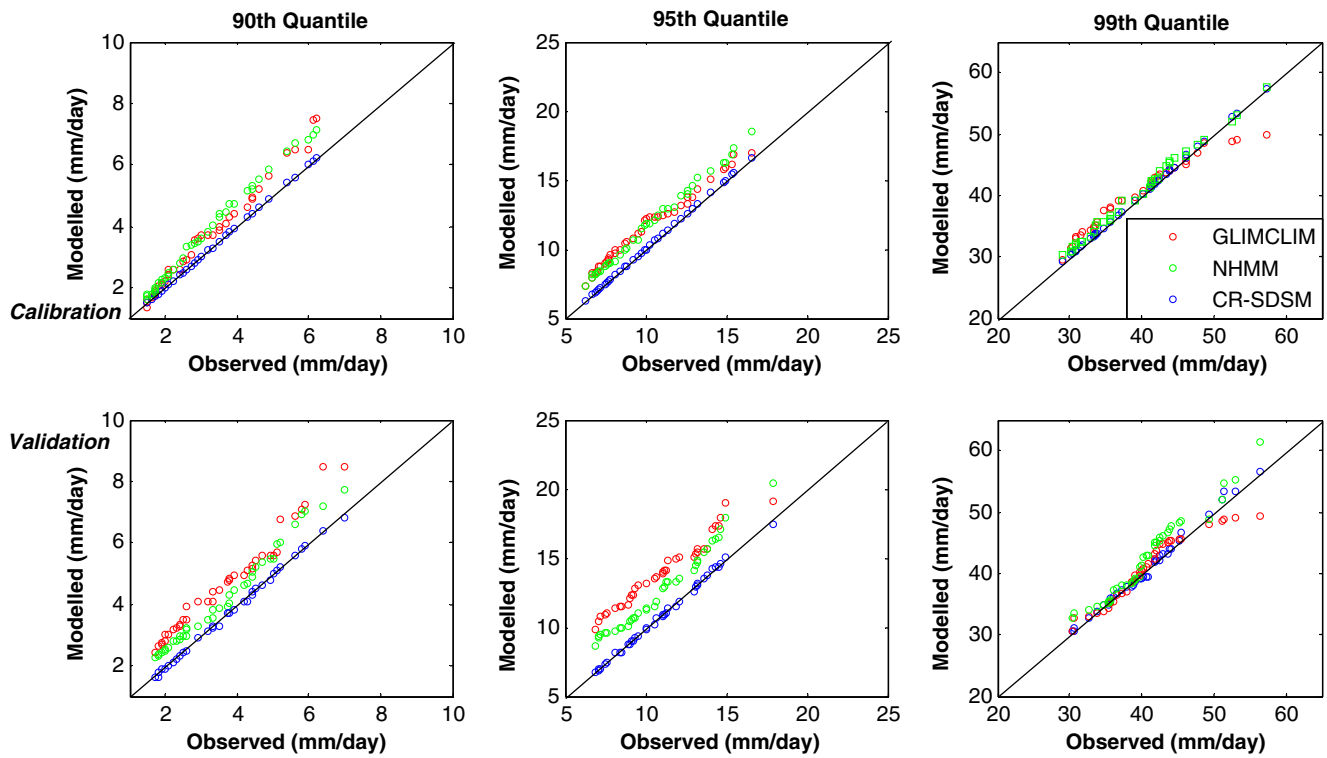


Fig. 11 Quantile–quantile plots of observed and simulated 90th, 95th, and 99th quantiles of daily rainfall (created by Matlab)

performed a little worse than GLIMCLIM and CR-SDSM in the calibration period, whereas it performed the best in the validation period. Additionally, for the

standard deviation (Fig. 7), CR-SDSM reproduced the best intra-annual correlations in the calibration period, whereas it performed worse than GLIMCLIM and

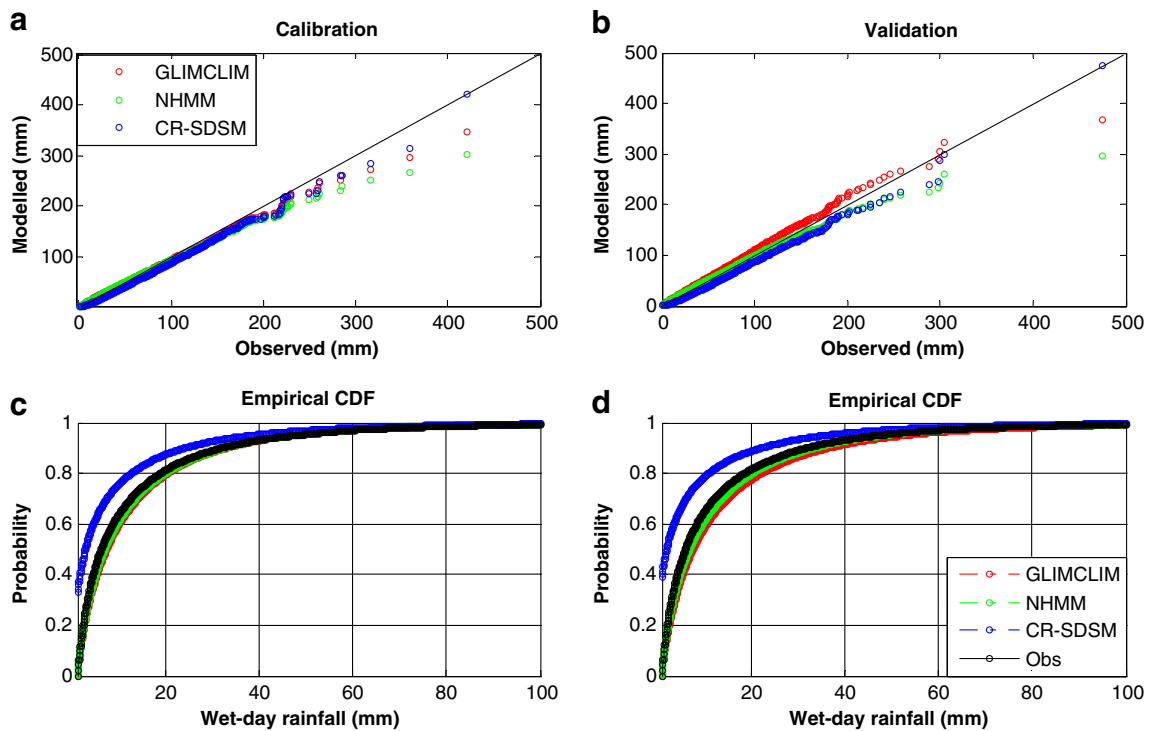


Fig. 12 Quantile–quantile and empirical distribution of wet-day rainfall (millimeters) (calibration: a, c validation: b, d) (created by Matlab)

NHMM in the validation period. These results are inconsistent with other studies (Harpham and Wilby 2005; Liu et al. 2011), which have reported SDSM did not reproduce reasonably monthly rainfall variability. The modeled intra-annual and inter-annual variations of rainfall are improved when humidity was included in the predictor set of SDSM (Wetterhall et al. 2006).

Spatial correlation statistics of monthly rainfall amount are shown in Fig. 8. Overall, the monthly spatial variability was relatively well reproduced by CR-SDSM. Wetterhall et al. (2006) indicated that the SDSM had a disadvantage modeling the spatial coherency of multi-sites when applied to the multiple sites individually, but the conditional resampling method used here obviously improves on this weakness. For all models, the monthly spatial correlations at short distances were slightly underestimated while they were slightly overestimated at long distances. CR-SDSM and GLIMCLIM markedly overestimated the long distance correlations while NHMM markedly underestimated the spatial correlations at short distances. GLIMCLIM uses correlation-

based structures for spatial dependence of rainfall occurrence and amounts between sites, with these results indicating further development is also required to adequately capture monthly spatial scale dependence at short and long distances.

3.3 Daily statistics

Daily rainfall characteristics, such as dry/wet spell length, extremes, daily rainfall distribution and spatial correlations are critical for hydrological modeling. The mean, SD and 99th quantile value of dry-spell length and wet-spell length were all well reproduced by GLIMCLIM in both the calibration and validation period (Figs. 9 and 10). Compared to GLIMCLIM, NHMM performed relatively poorer for the means of both dry-spell length and wet-spell length. Furthermore, CR-SDSM showed a constant underestimation for mean, SD and 99th quantile of dry-spell length (recall bias in number of dry days), but performed better than NHMM for the mean and 99th quantile of wet-spell length. Both

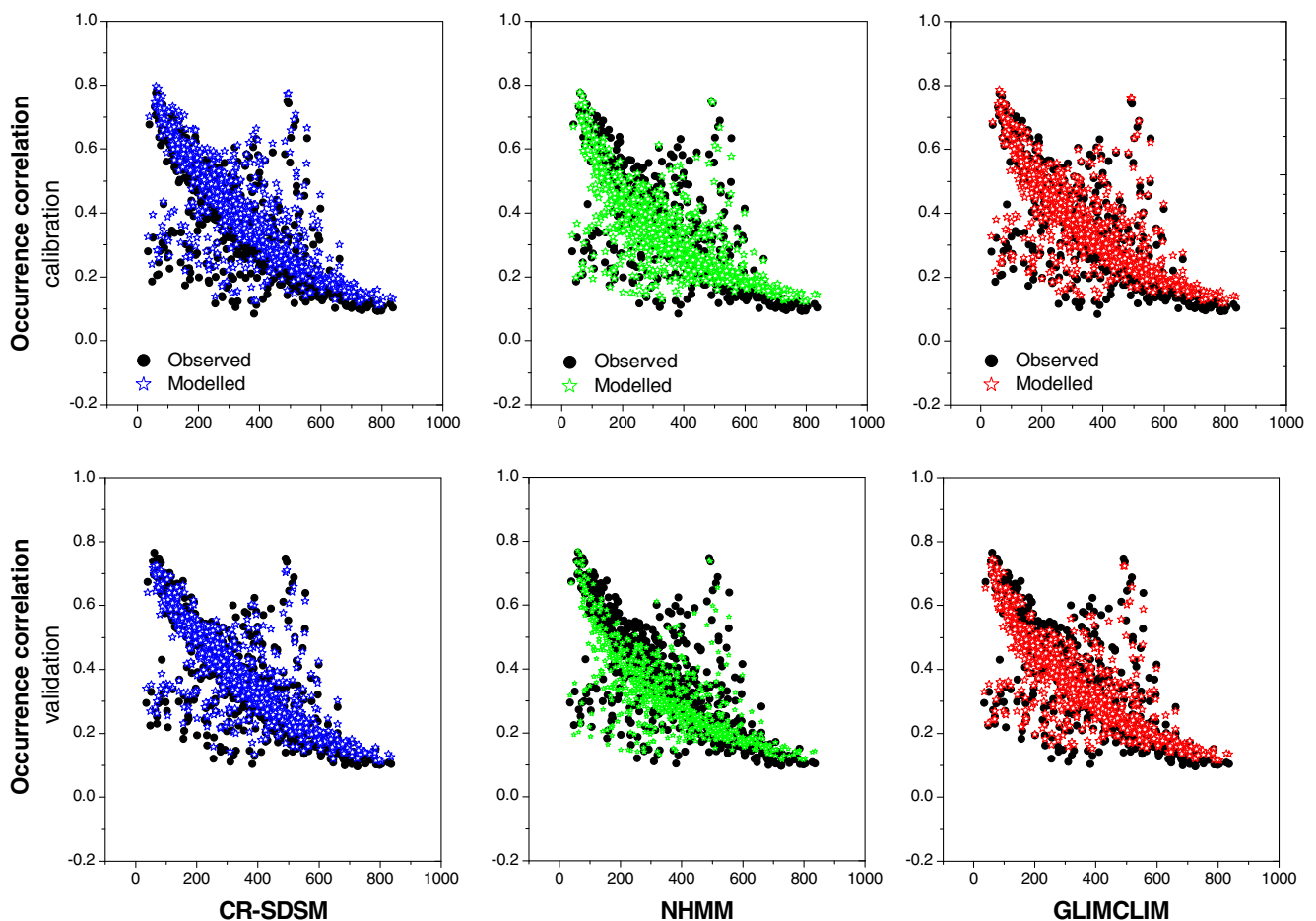


Fig. 13 Correlations between pairs of station daily rainfall occurrence vs. their separation distance for all possible combinations of station pairs in NCP (created by Origin)

GLIMCLIM and NHMM reproduced reasonable dry-spell length extremes and slightly overestimated the wet-spell length extremes, while CR-SDSM slightly underestimated the wet-spell length extreme (with several points overlying each other in Fig. 10).

In general, all three models were able to capture the daily rainfall extremes (Fig. 11). CR-SDSM produced almost perfect results (Fig. 11). This result is consistent with Harpham and Wilby (2005), but not with Liu et al. (2011) in an arid basin of China. Both GLIMCLIM and NHMM overestimated the 90th and 95th quantiles and produced reasonable 99th quantiles values of daily rainfall in the validation and calibration period.

A comparison of daily rainfall amount distributions using quantile–quantile plots and cumulative distribution functions (CDFs) is shown in Fig. 12. GLIMCLIM and NHMM gave the best fits to the CDF for the validation period, whereas the CR-SDSM model poorly reproduced it. The CR-SDSM tended to underestimate daily rainfall quantiles and hence overestimate the probabilities of rainfall amounts. Conversely, the GLIMCLIM overestimated the daily rainfall

quantiles (and consequently monthly/annual rainfall) and so underestimated the rainfall probabilities. In comparison, the NHMM better reproduced the daily rainfall distributions (Fig. 12).

Daily spatial correlation statistics (correlations of rainfall occurrence and amount between sites) is important in determining whether localised flooding occurs. In Figs. 13 and 14, almost all models performed well for daily rainfall occurrence and amount inter-site correlations. CR-SDSM tended to slightly overestimate, while NHMM tended to slightly underestimate the spatial correlations. GLIMCLIM produced better spatial dependence of daily rainfall than CR-SDSM and NHMM in both calibration and validation periods. Frost et al. (2011) found the spatial correlations of rainfall occurrence model were underestimated at short distances, and for amount model they were sometimes underestimated with distance under Australia conditions. The results showed that the improved version of GLIMCLIM (which considered the decay of correlation with inter-site separation in large region/basin simulation) could improve the simulation of spatial dependence for daily rainfall in large regions.

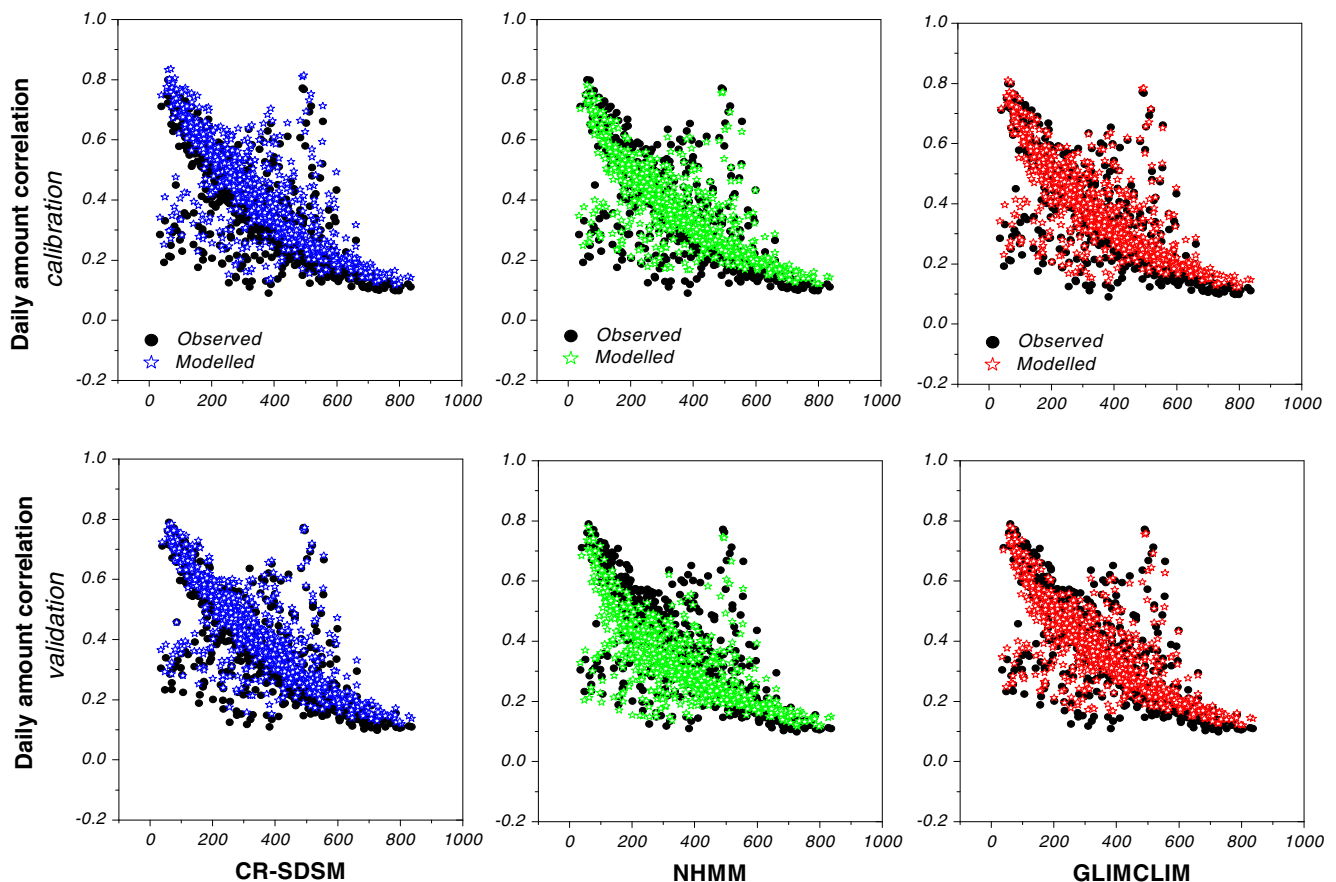


Fig. 14 Correlations between pairs of station daily rainfall amount vs. their separation distance for all possible combinations of station pairs in NCP (created by Origin)

4 Conclusions

A comparison of three multi-site daily rainfall statistical downscaling methods conditional on reanalysis predictors, applied to 40 sites in the NCP, were presented in this study. There were several advantages and drawbacks associated with each downscaling method, when applied for the first time across the North China Plain. More specifically, the following conclusions are made:

CR-SDSM provided relatively robust results for a range of statistics such as the extreme of daily rainfall (90th, 95th, and 99th quantiles), monthly rainfall (mean, SD, and distribution), and annual rainfall (mean, SD, and CS). However, it also exhibited obvious weaknesses: the daily rainfall amount was underestimated whilst its distribution was overestimated, the annual wet days were markedly overestimated (and consequently annual dry days were underestimated), and the dry-spell length was also underestimated. Consequently, CR-SDSM should be used with caution for typical yield/flood risk studies (which rely on the accurate prediction at annual timescales and also extremes at the daily timescale) and other hydrological applications on monthly/annual timescale.

GLIMCLIM performed well for the statistics of dry-spell/wet-spell length, annual wet/dry days, and spatial correlations of daily rainfall (and consequently monthly rainfall), but overestimated daily rainfall (and consequently annual rainfall). It could be recommended for some extreme events studies (e.g., drought/flooding) due to daily extreme statistic reproduction and for hydrologic simulations that rely on methods that capture a fuller range of rainfall characteristics (Frost et al. 2011). Furthermore, NHMM provided relatively robust results for daily, monthly and annual rainfall and annual wet/dry days, but slightly underestimated dry-spell length and wet-spell length and slightly overestimated the daily extremes (90th and 95th quantile). Therefore, it could be used for water availability and planning studies due to its relatively comprehensive performance across daily to annual timescales.

NCP climate is strongly controlled by the East Asian monsoon, with quite a difference between the atmosphere circulation in winter and summer (Chu et al. 2010). Consequently, it is a huge challenge to choose predictors adequate for this wide tempo-spatial space (Samel et al. 1999; Chu et al. 2010). In this study, a common predictor set was chosen a priori (via correlation analysis) before applying them to downscaling models. The influences of predictor selection (e.g., same predictors for all model vs different predictors for each model) should be further investigated. Given that further applications of statistical downscaling methods are to

predict daily rainfall in the future and use the data to hydrological studies, additional work should verify the performances of the different methods with GCM data as inputs. Furthermore, the effect of uncertainty introduced due to GCM scenarios, GCM choice, and GCM bias should also be considered in future (Vidal and Wade 2008; Leith and Chandler 2010; Frost et al. 2011).

Acknowledgments This work was supported by the National Basic Research Program of China (2010CB428406) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDA05090309). We wish to thank Dr. Richard E. Chandler from University College London (UK) and Dr. Chi Yang from Beijing Normal University (China) for the help of running GLIMCLIM model, and Dr. David Post, Ms. Jin Teng and two anonymous reviewers for their invaluable comments and constructive suggestions used to improve the quality of the manuscript.

References

- Ambrosino C, Chandler RE, Todd MC (2010) Southern African monthly rainfall variability: an analysis based on generalized linear models. *J Climate* 24:4600–4617. doi:10.1175/2010JCLI3924.1
- Bates BC, Charles SP, Hughes JP (1998) Stochastic downscaling of numerical climate model simulations. *Environ Model Softw* 13 (3–4):325–331
- Chandler RE (2002) GLIMCLIM: Generalised linear modelling for daily climate Time Series (software and user guide). Research Report No.227, Department of Statistical Science, University College London. <http://www.ucl.ac.uk/Stats/research/abs02.html#227>
- Chandler RE (2005) On the use of generalized linear models for interpreting climate variability. *Environmetrics* 16:699–715
- Chandler RE, Wheater HS (2002) Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resour Res* 38(10):1192. doi:10.1029/2001WR000906
- Charles SP, Bates BC, Hughes JP (1999) A spatial-temporal model for downscaling precipitation occurrence and amounts. *J Geophys Res* 104(D24):31657–31669. doi:10.1029/1999JD900119
- Charles SP, Bates BC, Viney NR (2003) Linking atmospheric circulation to daily rainfall patterns across the Murrumbidgee River Basin. *Water Sci Technol* 48:233–240
- Charles SP, Bates BC, Smith IN, Hughes JP (2004) Statistical downscaling of daily precipitation from observed and modeled atmospheric fields. *Hydrol Process* 18:1373–1394. doi:10.1002/hyp.1418
- Charles SP, Bari MA, Kitsuos A, Bates BC (2007) Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia. *Int J Climatol* 27: 1673–1690. doi:10.1002/joc.1508
- Chen ST, Yu PS, Tang YH (2010) Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J Hydrol* 385:13–22. doi:10.1016/j.jhydrol.2010.01.021
- Chiew FHS, Kirono DGC, Kent DM, Frost AJ, Charles SP, Timbal B, Nguyen KC, Fu GB (2010) Comparison of runoff modeled using rainfall from different downscaling methods for historical and future climates. *J Hydrol* 387:10–23. doi:10.1016/j.jhydrol.2010.03.025
- Chu JT, Xia J, Xu CY, Singh VP (2010) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for

- climate change scenarios in Haihe River, China. *Theor Appl Climatol* 99:149–161. doi:10.1007/s00704-009-0129-6
- Dibike YB, Coulibaly P (2005) Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *J Hydrol* 307:145–163. doi:10.1016/j.jhydrol.2004.10.012
- Fowler HJ, Kilsby CG, O'Connell PE, Burton A (2005) A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change. *J Hydrol* 308(1–4):50–66. doi:10.1016/j.jhydrol.2004.10.021
- 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. doi:10.1002/joc.1556
- Frost AJ, Charles SP, Timbal B, Chiew FHS, Mehrotra R, Nguyen KC, Chandler RE, McGregor JL, Fu GB, Kirono DGC (2011) A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. *J Hydrol* 408:1–18. doi:10.1016/j.jhydrol.2011.06.021
- Fu GB, Charles SP (2011) Statistical downscaling of daily rainfall for southeastern Australia. In: *Hydro-climatology: variability and change (Proceedings of symposium JH02 held during IUGG2011 in Melbourne, Australia, July 2011)*. IAHS Publ 344:69–74
- Fu GB, Charles SP, Chiew FHS (2007) A two-parameter climate elasticity of streamflow index to assess climate change effects on annual streamflow. *Water Resour Res* 43:W11419. doi:10.1029/2007WR005890
- Fu GB, Charles SP, Yu JJ, Liu CM (2009) Decadal climatic variability, trends and future scenarios for the North China Plain. *J Climate* 22:2111–2123. doi:10.1175/2008JCLI2605.1
- Gachon P, Dibike Y (2007) Temperature change signals in northern Canada: convergence of statistical downscaling results using two driving GCMs. *Int J Climatol* 27:1623–1641. doi:10.1002/joc.1582
- Grotch SL, MacCracken MC (1991) The use of general circulation models to predict regional climatic change. *J Climate* 4:286–303
- Harpham C, Wilby RL (2005) Multi-site downscaling of heavy daily precipitation occurrence and amounts. *J Hydrol* 312:235–255. doi:10.1016/j.jhydrol.2005.02.020
- Hay LE, Wilby RL, Leavesley GH (2000) A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *J Am Water Resour As* 36:387–397. doi:10.1111/j.1752-1688.2000.tb04276.x
- Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov model for precipitation occurrence. *J R Stat Soc C-Appl* 48(1):15–30
- Khan MS, Coulibaly P, Dibike Y (2006) Uncertainty analysis of statistical downscaling methods. *J Hydrol* 319:357–382. doi:10.1016/j.jhydrol.2005.06.035
- Kirshner S (2005) Modeling of multivariate time series using hidden Markov models. University of California, Irvine, Dissertation
- Kirshner S (2007) Learning with tree-average densities and distributions. *Adv Neural Inform Process Syst*, NIPS 2007, December 2007
- Leith NA, Chandler RE (2010) A framework for interpreting climate model outputs. *J Roy Stat Soc C* 59:279–296. doi:10.1111/j.1467-9876.2009.00694.x
- Liao Y, Zhang W, Chen D (2004) Stochastic modeling of daily precipitation in China. *J Geogr Sci* 114(4):417–426
- Liu ZF, Xu ZX, Charles SP, Fu GB, Liu L (2011) Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. *Int J Climatol* 31:2006–2020. doi:10.1002/joc.2211
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Suater T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Ehiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48:RG3003. doi:10.1029/2009RG000314
- Mehrotra R, Sharma A (2007) Preserving low-frequency variability in generated daily rainfall sequences. *J Hydrol* 345(1–2):102–120. doi:10.1016/j.jhydrol.2007.08.003
- Raje D, Mujumdar PP (2011) A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrol Process* 25(23):3575–3589. doi:10.1002/hyp.8083
- Robertson AW, Kershner S, Smyth P (2004) Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. *J Climate* 17:4407–4424
- Samel AN, Wang WC, Liang XZ (1999) The monsoon rainband over China and relationships with the Eurasian circulation. *J Climate* 12:115–131
- Timbal B, Hope P, Charles S (2008a) Evaluating the consistency between statistically downscaled and global dynamical model climate change projections. *J Climate* 21:6052–6059. doi:10.1175/2008JCLI2379.1
- Timbal B, Fernandez E, Li Z (2008b) Generalization of a statistical downscaling model to provide local climate change projections for Australia. *Environ Model Softw* 24:342–358. doi:10.1016/j.envsoft.2008.07.007
- Tryhorn L, DeGaetano A (2010) A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. *Int J Climatol* 31(13):1975–1989. doi:10.1002/joc.2208
- Vidal JP, Wade SD (2008) Multimodel projections of catchment-scale precipitation regime. *J Hydrol* 353:143–158. doi:10.1016/j.jhydrol.2008.02.003
- Wetterhall F, Bárdossy A, Chen DL, Halldin S, Xu CY (2006) Daily precipitation-downscaling techniques in three Chinese regions. *Water Resour Res* 42:W11423. doi:10.1029/2005WR004573
- Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. *Int J Climatol* 20:641–661
- Wilby RL, Dawson CW, Barrow EM (2002) SDSM—A decision support tool for the assessment of regional climate change impacts. *Environ Model Softw* 17:147–159
- Wilby RL, Tomlinson OJ, Dawson CW (2003) Multi-site simulation of precipitation by conditional resampling. *Climate Res* 23:183–194
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. IPCC Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA), http://ipcc-ddc.cru.uea.ac.uk/gu-idelines/StatDown_Guide.pdf
- Xu CY (1999) From GCM2 to river flow: a review of downscaling methods and hydrologic modeling approaches. *Progr Phys Geogr* 23(2):229–249
- Yan Z, Bate S, Chandler RE, Isham V, Wheeler H (2006) Changes in extreme wind speeds in NW Europe simulated by generalized linear models. *Theor Appl Climatol* 83:121–137. doi:10.1007/s00704-005-0156-x
- Yang C, Chandler RE, Isham VS, Wheeler HS (2005) Spatial-temporal rainfall simulation using generalized linear models. *Water Resour Res* 41:W11415. doi:10.1029/2004WR003739