Variability of spatial patterns of autocorrelation and heterogeneity embedded in precipitation

Dedi Liu, Qin Zhao, Shenglian Guo, Pan Liu, Lihua Xiong, Xixuan Yu, Hui Zou, Yujie Zeng and Zhaoli Wang

ABSTRACT

Spatial interpolation of precipitation data is an essential input for hydrological modelling. At present, the most frequently used spatial interpolation methods for precipitation are based on the assumption of stationary in spatial autocorrelation and spatial heterogeneity. As climate change is altering the precipitation, stationary in spatial autocorrelation and spatial heterogeneity should be first analysed before spatial interpolation methods are applied. This study aims to propose a framework to understand the spatial patterns of autocorrelation and heterogeneity embedded in precipitation using Moran’s I, Getis–Ord test, and semivariogram. Variations in autocorrelation and heterogeneity are analysed by the Mann–Kendall test. The indexes and test methods are applied to the 7-day precipitation series which are corresponding to the annual maximum 7-day flood volume (P-AM7FV) upstream of the Changjiang river basin. The spatial autocorrelation of the P-AM7FV showed a statistically significant increasing trend over the whole study area. Spatial interpolation schemes for precipitation may lead to better estimation and lower error for the spatial distribution of the areal precipitation. However, owing to the changing summer monsoons, random variation in the spatial heterogeneity analysis shows a significant increasing trend, which reduces the reliability of the distributed hydrological model with the input of local or microscales.

Key words | areal precipitation, spatial autocorrelation, spatial heterogeneity, time-variant spatial interpolation method, upstream area of Changjiang river basin

INTRODUCTION

All water enters the land phase of the hydrologic cycle as precipitation (Dingman 2002). Thus, to assess, predict and forecast hydrologic responses in water resource management, a hydrologist needs to understand the amount and distribution of the precipitation. As an essential tool for the hydrologist, the hydrological model is very sensitive to the precipitation input (Teegavarapu et al. 2012; Alizadeh et al. 2018; Herrnegger et al. 2018). The precipitation in the hydrological model should be spatially continuous data or converted areal precipitation for the description of the flow generation process (Syed et al. 2003; Kobold & Sušelj 2005; Gabellani et al. 2007; Cole & Moore 2008; Moulin et al. 2009; Pechlivanidis et al. 2017; Zou et al. 2017; Herrnegger et al. 2018). Although radar (NEXRAD) based precipitation estimation (Vieux 2001) can provide data with significant spatial resolution compared to point precipitation gauge measurements (Teegavarapu et al. 2012), indirect estimation of continuous surface precipitation based on the measurement of related ancillary variables must also be calibrated and validated by direct ground-based measurement through spatial interpolation methods. Spatial interpolation of precipitation data is generally classified into deterministic and geostatistical groups (Ly et al. 2011; Teegavarapu et al. 2012). The most frequently used of both
the deterministic (e.g., Thiessen polygon and inverse distance weighting) and the geostatistical (e.g., Kriging) methods are based on the spatial characteristics of precipitation. The spatial characteristics of precipitation are mainly composed of spatial autocorrelation and spatial heterogeneity. According to the assumption of high spatial autocorrelation in precipitation (Gilbert & Lowell 1997), each grid point is assigned with the value of the nearest rain gauge in a Thiessen polygons scheme (Thiessen 1911), and the influence of the measured point data is weighted according to the distance from the sampled point to the estimated point in an inverse distance weighting scheme (Shepard 1968; Wei & McGuinness 1975; Simanton & Osborn 1980; ASCE 1996; Farajzadeh & Alizadeh 2018). Spatial heterogeneity, quantified by the semivariogram, has been considered in geostatistical and variance-dependent stochastic surface interpolation methods (Krige 1951; Zimmerman et al. 1999; Grayson & Blöschl 2001; Vieux 2001; Ly et al. 2011). The variation in the precipitation spatial pattern impacts the implementation of the interpolation methods of point precipitation for areal precipitation. Although the spatiotemporal pattern of precipitation has received increasing attention in water cycle studies through point or interpolated areal precipitation data series (Larson & Peck 1974; Xu & Singh 1998; Vieux 2001; Xing et al. 2015), there are few studies on the variations in the spatial autocorrelation and spatial heterogeneity of precipitation.

Time-variant parameters in hydrological models have been reported to improve the accuracy of the model results (Merz et al. 2011; Peel & Blöschl 2011; Brigode et al. 2013; Jeremiah et al. 2013; Westra et al. 2014; Patil & Stieglitz 2015; Thirel et al. 2015; Deng et al. 2016). As precipitation is the largest source of uncertainty in the hydrological model for forecasting (Yatheendradas et al. 2008; Villarini et al. 2010; Hapuarachchi et al. 2011), time-variant spatial interpolation methods of areal precipitation could also be developed to improve the reliability of the results from the hydrological model. However, the stationary or nonstationary conditions for the spatial interpolation method should be tested first. Therefore, the aim of this study is to analyse the variability of the spatial autocorrelation and spatial heterogeneity of precipitation to test the suitability of the spatial interpolation method in a changing environment. If the spatial autocorrelation and spatial heterogeneity of precipitation are nonstationary, finding a suitable successor for spatial interpolation methods of areal precipitation is crucial for the adaptation to changing environment.

As the extreme precipitation in the upstream area of the Changjiang river basin is inflow to the Three Gorges Reservoir and the main source of flood for the middle and downstream area, the extreme precipitation in this area is important for the reservoir operation and flood control, and is chosen for our case study. The spatial autocorrelation indexes are specified by Moran’s I, the Getis–Ord general G, and the local Getis–Ord Gi; statistics Gi. Spatial heterogeneity is quantified by the semivariogram, and the trends of all indexes are tested by Mann–Kendall (MK) test. Therefore, the framework of the test can be as shown in Figure 1.

**STUDY AREA AND DATA**

The upstream part of the Changjiang river has a total length of 4,504 km and drainage area of 1 million km², accounting for 70% of the entire river length and 55.5% of the total river basin, from the headwater to the Yichang gauge station in Hubei province. The upper mainstream is fed by many large tributaries, such as the Yalongjiang, Minjiang, Jialingjiang and Wujiang rivers. According to the statistics of the Yangtze River Water Resources Commission (CCYRA 2006), by 2005, the number of dams in the upper reaches of the Changjiang river was 12,929. The Three Gorges Dam, the world’s largest hydropower project to date, is located 44 km upstream of the Yichang hydrological gauge station. It has a subtropical humid monsoon climate, with an annual average temperature of 15–18 °C and a mean annual precipitation of 1,232 mm (Lu & Cao 2013). Owing to the large geographic extent of the upstream part of the Changjiang river basin (24.5 °N–27.8 °N; 90.6 °E–111.5 °E), the precipitation spatial autocorrelation and heterogeneity have an irregular distribution. If the spatially uneven distribution pattern of precipitation in this upstream area changes, the instream flow to all the reservoirs located in the middle and downstream areas will be impacted. Therefore, the variability of the precipitation spatial pattern is critical for water resource management of the whole Changjiang river basin.
The flood control storage of the Three Gorges Reservoir is $221.5 \times 10^8$ m$^3$, which is based on the annual maximum 7-day flood volume (AM7FV) series. Thus, the AM7FV is the most important index for flood forecasting in the upstream area of the Changjiang river basin and for the operation of the Three Gorges reservoir. For example, the typical flood event for the Three Gorges Reservoir design occurred in 1870 and is attributed to the heavy precipitation from July 13th to 19th. The corresponding 7-day precipitation with AM7FV (P-AM7FV) is selected for the analysis of its spatial pattern variability. Although there are 169 rainfall gauge stations located in the upstream area of the Changjiang river basin, only 53 of them were not interrupted between 1960 and 2008 (as shown in Figure 2), and they

![Figure 1](https:// iwaponline.com/hr/article-pdf/50/1/215/524294/nh0500215.pdf)  
**Figure 1** | Schematic framework analyzing the variability of spatial patterns of autocorrelation and heterogeneity embedded in precipitation.  

![Figure 2](https://iwaponline.com/hr/article-pdf/50/1/215/524294/nh0500215.pdf)  
**Figure 2** | Locations of the rainfall and Yichang discharge gauge stations in the upstream area of the Changjiang river basin.
were selected for our study. The daily runoff data of the Yichang station from 1960 to 2008 were provided by the Bureau of Hydrology of the Changjiang (Yangtze river) Water Resources Commission of China. Daily precipitation data in the 53 rainfall gauge stations were collected from the website of the China Meteorological Bureau (http://data.cma.cn/). These stations have a relatively uniform distribution over the upstream area of the Changjiang river basin, as presented in Figure 2.

**METHODOLOGY**

The spatial patterns of precipitation include spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation was conducted by Moran’s I statistic with global clustering. The global Moran’s I statistic can provide a single index that summarises the local precipitation patterns of the study area. To identify the level of clustering at each of the 53 studied rainfall gauge stations, the Getis–Ord general G test and the local Getis–Ord general G were employed to determine the high clusters (stations clustered owing to high precipitation) or low clusters (stations clustered owing to low precipitation). As the semivariogram range quantifies the degree of similarity/dissimilarity of a continuously varying attribute between a pair of samples separated by a certain distance (Isaaks & Srivastava 1989), it has been used to analyse the spatial heterogeneity of precipitation. Then, we performed the MK test to detect the spatial autocorrelation and heterogeneity trends in precipitation at the scale of the whole study area.

**Spatial autocorrelation analysis**

**Moran's I statistic**

The Moran’s I statistic quantifies the spatial clustering and connectivity between pairs of clusters by assessing the correlation of parameters among spatially close or contiguous rainfall gauge stations (Thompson et al. 2018). As our study discusses, the association between the 53 rainfall gauge stations and their rainfall characteristics, the equation for Moran’s I is expressed as follows (Moran 1948; Anselin 1995; Blazquez & Celis 2013):

\[
I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} \cdot (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} \cdot (1/n) \sum_{i}^{n} (X_i - \bar{X})^2} \quad (i \neq j)
\]

where \(X_i\) and \(X_j\) refer to the precipitation at the \(i^{th}\) and \(j^{th}\) stations, respectively; \(W_{ij}\) is a spatial weight matrix, defining the degree of spatial closeness or contiguity of stations \(i\) and \(j\); \(\bar{X}\) denotes the spatial mean of the precipitation; and \(n\) is the total number of precipitation gauge stations used in our study. Most of the values of the Moran’s I statistic are restricted to a range from –1 to 1, but values outside of this range are also possible in extreme cases (Arbia 2014). A significant positive Moran’s I value indicates the existence of either high-value or low-value clustering for precipitation, while a negative spatial autocorrelation indicates a tendency toward the juxtaposition of low values next to high values (Zhang & Lin 2016). In other words, the more clustered the rainfall gauge stations in the study area are, the closer the value of Moran’s I is to 1, and the higher the degree of spatially negative correlation in the region, the closer the index value is to –1. The value of zero shows a randomly distributed pattern of precipitation or no spatial dependence (Soltani & Askari 2017). The statistical significance of Moran’s I is commonly transformed to a Z score test with a standard normal distribution (Cliff & Ord 1981):

\[
Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}
\]

where \(E(I)\) is the expected value of \(I\) for a random spatial pattern, and \(Var(I)\) represents the variance of \(I\). Both \(E(I)\) and \(Var(I)\) are constant (the details of the determining equations can be found in ‘How Spatial Autocorrelation (Global Moran’s I) works’ from Arcgis 10.1 help). A positive Z score with a significant level implies that the distribution of the precipitation is spatially clustered, and that the precipitation features around neighbouring stations have similar values, while a negative Z score indicates that the spatial pattern of precipitation is dispersed and the adjacent precipitation features have dissimilar values (Xie et al. 2015). A significance level of 0.05 was used in our study.
Getis–Ord test

Moran’s I is a global statistical index that summarises the values of autocorrelation over the entire study area (Luković et al. 2015). A high value of precipitation may cause serious flood events, while a low value of precipitation involves a lower possibility of disaster event occurrence. Thus, it is necessary to discriminate between those patterns that have high-precipitation or low-precipitation dominant values.

Developed in the mid-1990s, the Getis–Ord general G provides researchers with a straightforward way to measure how concentrated the high or low values are for a given entire study area. However, the local Getis–Ord statistics can help gain more insight into hot and cold spots of precipitation. $G_i$ is high in a cluster of high values, which is identified as a hot spot (high values next to high), and $G_i$ will be low in a cluster of low values, which is identified as a cold spot (low values next to low) (Zhang & Tripathi 2018). A high value surrounded by low values, or vice versa, is identified as an outlier. The Getis–Ord general G statistic is defined as (Getis & Ord 1992):

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} x_i x_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j}$$ (3)

The local Getis–Ord $G_i^*$ equation is presented as (Getis & Ord 1992):

$$G_i^* = \frac{\sum_j W_{ij} x_i}{\sum_i x_j}$$ (4)

Both the Getis–Ord general G statistic and the Getis–Ord $G_i^*$ statistic are interpreted using standardised Z-scores ($Z(G_i^*)$). A high positive $Z(G_i^*)$ indicates a high value of precipitation (hotspots), while a low negative $Z(G_i^*)$ indicates low precipitation (cold spots) (Getis & Ord 1992; Jana & Sar 2016).

Semivariogram for spatial heterogeneity analysis

The spatial heterogeneity of rainfall is often quantified by the semivariogram to describe its spatial structure and randomness (Zimmerman & Zimmerman 1991; Tesfamichael et al. 2009). The semi-variance measures spatial variability through squared differences between pairs of precipitation samples separated by a certain distance, which allows the building of the experimental semivariogram (Isaaks & Srivastava 1989; Subyani 2004; Dash et al. 2010). $\gamma(h)$ is defined by:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [X(x_i) - X(x_i + h)]^2$$ (5)

where $x_i$ and $x_i + h$ are two sampling locations of gauge $i$ and the gauge separated by a distance $h$ from the $i$th gauge. To assign the distance $h$, the Euclidean distance between gauges is often adopted (Skøien et al. 2003); $N(h)$ is the number of sample gauges using $h$; and $X(x_i)$ and $X(x_i + h)$ represent values of the precipitation $X$ measured at both locations.

As there are limited gauge stations, the empirical variogram cannot offer values for each distance $h$, but several empirical variogram models are able to estimate these variances. The Gaussian, spherical and exponential models for the semivariogram are broadly used and are adopted for spatial heterogeneity analysis in our study as these models adequately fit the variogram values with only a few parameters (Berne et al. 2004; Diggle & Ribeiro 2007).

The Gaussian model for the semivariogram is given by Equation (6):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C(1 - e^{-3h^2/a^2}) & h > 0 \end{cases}$$ (6)

The spherical model for the semivariogram is given by Equation (7):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C \left( \frac{3h}{2a} - \frac{1h^3}{2a^3} \right) & 0 < h \leq a \\ C_0 + C & h > a \end{cases}$$ (7)

The exponential model for the semivariogram is given by Equation (8):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C(1 - e^{-h/a}) & h > 0 \end{cases}$$ (8)
where $C$ is the structural variance, $C_0$ is the nugget variance, and $a$ is the range parameter. The performances of the above three semivariogram models were also tested using cross-validation. The results showed an acceptable fit obtained by a root mean square error (RMSE) between the measured and estimated values of approximately zero.

There are three characteristic parameters in the fitted semivariogram, namely, range, sill and nugget. The ‘nugget variance’ ($C_0$) is the random variation that usually results from either measurement errors or variation in the studied variable, which is impossible to detect in the sample range (Trangmar et al. 1986; Cressie 1993). $C_0 + C$ is called the ‘sill’ and represents the total sample variability at the semivariogram level for the patterned data. The structure variance ($C$) is the difference between the sill and nugget. The spatial dependence is often defined as $(C/(C + C_0))$, and it relates the small-scale variability to the large-scale variability (He et al. 2015). The ‘range’ referring to $a$ is the maximum distance over which the measured precipitation properties exhibit significant spatial autocorrelation. The ranges of the exponential, Gaussian and spherical models are often defined by the semivariogram, which is 95% of the sill (Bardossy 1997). Thus, their ranges are $3a$ for the exponential, $\sqrt{a}$ for the Gaussian, and $a$ for the spherical models.

**MK test**

The non-parametric MK test (Mann 1945; Kendall 1975) has been widely used to assess the significance of the trends in hydro-meteorological time series (Liu et al. 2010; Rajah et al. 2014; Rashid et al. 2015; Wang et al. 2017; Yang et al. 2017; Foulon et al. 2018; Sun & Fang 2018), and it has also been recommended by the World Meteorological Organization (WMO) (Yue et al. 2002). Therefore, we applied the MK to detect trends in the spatial autocorrelation and heterogeneity of precipitation in this study.

The MK test is a rank correlation test of the null hypothesis in which a time series $\{x_t; t = 1, 2, \ldots, n\}$ is independent and identically distributed (Mann 1945; Kendall 1975). The test statistic $S$ is given by Equation (9) as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(x_j - x_i)$$

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(x_j - x_i)$$  \hspace{1cm} (9)

with $\text{sgn}(x_j - x_i) = \begin{cases} 1 & x_i < x_j \\ 0 & x_i = x_j \\ -1 & x_i > x_j \end{cases}$. The mean and variance of the $S$ statistic in Equation (9) given by Kendall (1975) are:

$$E(S) = 0 \hspace{1cm} (10)$$

$$\text{Var}(S) = \left[ n(n-1)(2n+5) - \frac{q}{n-1}(2t_p-1)(2t_p+5) \right]/18 \hspace{1cm} (11)$$

where $t_p$ is the extent of all the ties (equal value for the $p$th value) and $q$ is the number of tied values. The MK test statistic, $U$, which approximates a standard normal distribution (with a sufficiently large sample size, i.e., $n > 10$), can be estimated as follows:

$$U = \begin{cases} (S-1)/\sqrt{\text{Var}(S)} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ (S+1)/\sqrt{\text{Var}(S)} & \text{if } S < 0 \end{cases} \hspace{1cm} (12)$$

$U > U_{1-a/2}$ indicates a positive trend and $U < -U_{1-a/2}$ represents a negative trend in the time series at a significant level of $\alpha$. Otherwise, there is no trend in the time series at a significant level of $\alpha$.

**RESULTS AND DISCUSSION**

**Spatial autocorrelation analysis in P-AM7FV**

**Trends in the spatial autocorrelation**

The spatial autocorrelation of precipitation was assessed using Moran’s I in the upstream part of the Changjiang river when the AM7FV happened (P-AM7FV). The results of the Moran’s I analysis showed that the values were greater than 0 for every year from 1960 to 2008 (shown in Figure 3(a)). There was a positive spatial autocorrelation in the study area. P-AM7FV data were clustered but not dispersed. The degree of spatial autocorrelation was also assessed by the $Z$ values of the normal distribution. It was found that the $Z$ scores in 1961, 1962, 1963 and 1992 were less than 1.96 (shown in Figure 3(b)), and thus, the clustered patterns in these four years could have been the
result of random chance at a 5% significance level. The other 45 years exhibited tendencies towards positive spatial autocorrelations and represented clustered patterns. To show the difference between the clustered and random patterns, the Moran’s index, Z score, and p value from the Moran’s I test are shown in Figure 4, which presents the lowest Z(I) in 1992 and the highest Z(I) in 1993. Figure 4(a) shows that the Moran’s I reported for the P-AM7FV in 1992 was 0.02, which indicates a random pattern. Similar situations were found in 1961, 1962 and 1963. Figure 4(b) shows that Moran’s I is 0.60 in 1993, which indicates a high positive spatial autocorrelation or a highly clustered pattern. Given the Z(I) score of 8.55 in 1993, there is less than a 5% probability that such a clustered pattern could have resulted from chance (p < 0.0001).

The Getis–Ord general G discriminates between spatial patterns that have dominant high values or low values. Positive G values, which are found in Figure 5, indicate that high values within other high values dominate the spatial autocorrelation pattern in our study area. The Z(G) scores in Figure 5(b) from the test of statistical significance at a level of 0.05 show that the P-AM7FV features with high values tended to cluster in the upstream area, except in 1961, 1963, 1975, 1992 and 1996.

We also trace the Moran’s I and Getis–Ord general G values over time (as shown in Figures 3 and 5). Statistically significant increasing trends in Moran’s I were observed in the upstream area of the Changjiang river with an MK U value of 2.13 at a significance level of 0.05, as presented in Table 1. However, the series of general G did not pass the trend test. Therefore, the trend of the spatial cluster pattern increased significantly for P-AM7FV data from 1960 to 2008, while the trend of the high-value spatial pattern for the P-AM7FV did not increase significantly. According to
the statistic’s structure and the characteristics of Moran’s I and the Getis–Ord general G (Getis & Ord 1992), P-AM7FV data values from 1960 to 2008 tended to increase by the same proportion rather than to increase by the same absolute amount.

**Spatial analysis for identification of hot spots**

Both the global statistic Moran’s I and the general G include all the precipitation gauge stations and provide an overall measure of spatial autocorrelation. The spatial autocorrelation associated with one particular spatial gauge station can be assessed through the hot spot analysis tool, which is calculated by the local Getis–Ord G* statistic (Equation (4)). The hot spot analysis leads to spatial cluster identification and spatial filtering. Z($G_i^*$) from the test of statistical significance at the 0.05 significance level was used to determine whether features with high or low values tended to cluster. Clusters of precipitation gauge stations with high P-AM7FV values were considered hot spots, whereas clusters of precipitation gauge stations with low P-AM7FV values were considered cold spots. A ring map, which is a spatial visualisation tool combing temporal and spatial data for more meaningful interpretation of trends (Huang et al. 2008; Chan et al. 2015), was used to explore the hot and cold spots of P-AM7FV (Figure 6).

The centre of Figure 6 is the digital elevation model (DEM) distribution map of the upstream part of the Changjiang river basin. The ring map is surrounded by a set of concentric, segmented rings that possess a circular shape indicating $Z(G_i^*)$ values. The four shades represent significant hot spots ($Z(G_i^*) > 1.96$), no significant hot spots ($0 < Z(G_i^*) \leq 1.96$), no significant cold spots ($-1.96 \leq Z(G_i^*) < 0$) and significant cold spots ($Z(G_i^*) < -1.96$). Each ring displays a $Z(G_i^*)$ value that represents high or low clusters of P-AM7FV at a particular location. There are 49 rings created corresponding to the number of years considered in our study. The inner ring displays the $Z(G_i^*)$ value from 1960, whereas the outer ring displays the $Z(G_i^*)$ value from 2008. The 53 spokes surrounding the base map represent the 53 precipitation gauge stations.

It is clearly noticeable that the $Z(G_i^*)$ values in the east and southeast parts of the ring map mainly show the hot spots in most of the years, while there are cold spots in the west area with higher elevations (higher DEM). According to the distribution of the precipitation gauge station's...
location shown in Figure 2, the east and southeast parts of the ring map mainly cover the Three Gorges Reservoir area. The west area covers the Jinshajiang river sub-basin. The south and south-west parts of the ring map, which represent the confluence area of the Jinshajiang and Daduhe river basins, shows Z($G_i$) values from negative to highly positive. In other words, the cold spots of P-AM7FV are becoming hot spots over time. However, the hot spots have tended to change to cold spots in the northeast part of the ring map. The distribution of the hot and cold spots is associated with the distribution of the elevations. The high (west) and low (east) elevational areas are the cold and hot spot areas, respectively. However, the transition zone between the high and low elevation areas is divided into two parts. The first part, located in the upper area of the ring map or at the north of the transition zone, is where the transition occurs with spot changes from hot to cold, whereas the other part, located to the south of the transition zone, has shown a tendency to change spots from cold to hot.

Spatial heterogeneity analysis in P-AM7FV

Variogram fitting on P-AM7FV data

To estimate the semivariogram parameters of the three semivariogram models (Gaussian, spherical and exponential
models), the observed precipitation datasets from the 53 gauge stations are used to optimise the RMSE values for the model's goodness of fit for every year. RMSE is a measure to describe the semi-variance difference between the observed and modelled values. The semi-variance can be determined through Equation (5). To ensure enough data for the improvement of the stability of the semivariogram parameter estimation, the value $h$ in Equation (5) is set to 35 km, according to the maximum separation distance between the pairs of observation gauges in our study (Journel & Huijbregts 1978; Burrough 1991) and the minimum $N(h)$ value is 30 in Equation (5). Boxplots are utilised to show graphically the RMSE from 1960 to 2008 (shown in Figure 7). Figure 7 indicates that the performances of all three semivariogram models are acceptable for describing the spatial heterogeneity of the P-AM7FV.

Trends in the spatial heterogeneity of P-AM7FV

As the characteristic parameters $C_0$ (random variation), $C_0 + C$ (total sample variability), $C/(C + C_0)$ (relates small-scale variability to large-scale variability), and range (exhibits significant spatial autocorrelation) fully quantify the spatial heterogeneity, the trend of spatial heterogeneity in P-AM7FV was tested by the calibrated parameters of the semivariogram model from 1960 to 2008. As seen in Figure 8,
Figure 8 | Variation of the calibrated characteristic parameters for the three semivariogram models from 1960 to 2008: (a) $C_0$; (b) $C_0 + C_1$; (c) $C/C_0 + C_2$; (d) range.
all the parameters of all the models changed every year, and thus it was indispensable to validate the parameters of the models for the precipitation interpolation. However, the parameter values from the three models and their trends were similar to their spatial heterogeneity simulation.

The random variation $C_0$ is typically used to model the white noise effect in spatial heterogeneity analyses. Without considering the measurement errors in the spatial heterogeneity, the nugget $C_0$ only refers to microscale variations (Cressie 1993). The P-AM7FV in the upstream area of the Changjiang river was mainly influenced by the interactions between monsoons and elevation, while the effects of monsoons on precipitation varied owing to climate change (Niu et al. 2015; Wang et al. 2018). As a result, the trend in random variation was a significantly increasing trend at a significance level of 0.05 (as presented in Table 2). The maximum variability of P-AM7FV $C_0 + C$ (sill) was composed of the spatial random variations and structure variations, and it showed a non-significant increasing trend at a significance level of 0.05 (as shown in Figure 8(b)). The spatial structure variation $C$ represents the amount of the observed variation that can be explained by the distance between observations. As the distance can be interpreted as the surrogate of different spatial effects of the monsoons on the P-AM7FV, the $C$ can be physically attributed to the spatial distributions and intensities of the monsoons. If the ratio of $C/[C + C_0]$ is used to define distinct classes of spatial structure or dependence of the P-AM7FV, the P-AM7FVs can be strongly, moderately or weakly dependent spatially when the ratios are >75%, between 25% and 75%, and <25%, respectively (Cambardella et al. 1994). From Figure 8(c), it is evident that 41%, 53%, and 61% of the years are strongly dependent spatially on the Gaussian, spherical and exponential models, respectively, and fewer than 10% of the years are weakly dependent spatially on the P-AM7FV in all the models. However, their spatial dependence decreased non-significantly from 1960 to 2008 at a significance level of 0.05 (shown in Table 2). Therefore, the influences of the monsoons on the P-AM7FV were decreasing, while the randomness of the precipitation, including the effects of geomorphology and local climate facts, were increasing at the microscale.

The range is the maximum length scale of the spatial autocorrelation. The ranges fluctuate in the three models (as shown in Figure 8(d)), resulting in non-constant autocorrelation distances. As the influences of the monsoon on the P-AM7FV were decreasing, the importance of local elevations to the P-AM7FV was relatively enhanced. Range had a significant increasing trend. Therefore, the degree of global spatial autocorrelation represented by Moran’s I had a significant increasing trend, which may have been partly related to the significant increasing trend of the range in spatial autocorrelation. However, both the randomness, with a significant increasing trend, and the spatial dependence, with a non-significant decreasing trend, may be attributable to changes in the distribution of hot and cold spots at the microscale.

As the summer monsoons are the intrinsic variation factors in the spatial pattern of the P-AM7FV, it is necessary to understand further the trends of the summer monsoons in the upstream area of the Changjiang river basin. The P-AM7FV in the study area was mainly influenced by the South Asia summer monsoon (SASM), Plateau summer monsoon, and elevation. The SASMs are often quantified by the dynamically normalised seasonality index (Li & Zeng 2002) while the Plateau summer monsoon index (PSMI) is based on the height departure values of the low-pressure systems in the surface layer over the plateau (Tang et al. 1984). The SASM sector (5°N–22.5°N, 35°E–97.5°E) is composed of two independent components, the SWASM (2.5°N–20°N, 35°E–70°E) and the SEASM

### Table 2 | MK test statistics ($U$ values) for the calibrated characteristic parameters of the three semi-variogram models

<table>
<thead>
<tr>
<th>Semi-variogram models</th>
<th>Characteristic parameters</th>
<th>Statistic $U$ value</th>
<th>Tendency$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$C_0$</td>
<td>2.40</td>
<td>↑$^*$</td>
</tr>
<tr>
<td></td>
<td>$C_0 + C$</td>
<td>0.78</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td>$C/(C_0 + C)$</td>
<td>−1.90</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>2.54</td>
<td>↑$^*$</td>
</tr>
<tr>
<td>Spherical</td>
<td>$C_0$</td>
<td>2.86</td>
<td>↑$^*$</td>
</tr>
<tr>
<td></td>
<td>$C_0 + C$</td>
<td>0.51</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td>$C/(C_0 + C)$</td>
<td>−2.19</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>2.77</td>
<td>↑$^*$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$C_0$</td>
<td>2.83</td>
<td>↑$^*$</td>
</tr>
<tr>
<td></td>
<td>$C_0 + C$</td>
<td>1.01</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td>$C/(C_0 + C)$</td>
<td>−2.47</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>3.00</td>
<td>↑$^*$</td>
</tr>
</tbody>
</table>

$^*$Passed the test at the significant 0.05 level.
(2.5°N–20°N, 70°E–110°E), with different relations than the monsoon rainfall in South Asia (Li & Zeng 2002). PSM refers to the centre position (32.5°N, 90°E) and the four positions (32.5°N, 80°E; 25°N, 90°E; 32.5°N, 100°E; 40°N, 90°E) around the centre (Tang et al. 1984). Higher index values for the SWASMI, SEASMI and PSM mean that higher precipitation happened in the upstream area of the Changjiang river basin. Lower index values mean lower precipitation. The trend of the SEASM was significantly decreasing, while the SEASMI and PSM had non-significant increasing trends. If the SWASMI, SEASMI and PSM are normalised and the trend of their average values (AVSWEP) is tested, they show a non-significant decreasing trend, which consists of the AM7FV trend and the range and spatial structure variation trends (as presented in Table 3). As the sources of the summer monsoons are different, their trends are different, especially under climate change. Their impacts on the extents and intensities of precipitation are also different every year. Therefore, there are more uncertainties in the variation of the spatial pattern of P-AM7FV at the microscale than that at the global scale.

### CONCLUSION

This study evaluated the variation in spatial autocorrelation and heterogeneity of the precipitation in the upstream area of the Changjiang river basin, China. The spatial autocorrelation of P-AM7FV represented by Moran’s I had a statistically significant increasing trend over the whole study area. The range of spatial heterogeneity also had a significant increasing trend. Spatial interpolation schemes for precipitation depending on the spatial autocorrelation and variance can lead to better estimation and lower error in the spatial distribution of the areal precipitation for the lumped hydrologic model. However, owing to the changing effects of summer monsoons on precipitation, the random variation in the spatial heterogeneity analysis had a significant increasing trend. It may have impacted the accuracy of the spatial interpolation schemes at the microscale and further reduced the reliability of the distributed hydrological model with the input of local or microscale P-AM7FV in the upstream area of the Changjiang river basin. Also, the effects of duration of precipitation (e.g., 3 days or 15 days) on their spatial patterns of autocorrelation and heterogeneity will be studied in the future.

### ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (No. 51579183, 91647106, 51879194 and 51525902), the Science and Technology Program of Guangzhou City (No. 201707010072) and the 111 Project (No. B18037).

### REFERENCES


Arbia, G. 2014 *A Primer for Spatial Econometrics with Applications in R.* Palgrave MacMillan, Basingstoke, UK.


First received 28 April 2018; accepted in revised form 30 August 2018. Available online 26 September 2018