To the problem of uncertainty in interpolation of annual runoff

Ziqi Yan, Lars Gottschalk, Irina Krasovskaia and Jun Xia

ABSTRACT

The long-term mean value of runoff is the basic descriptor of available water resources. This paper focuses on the accuracy that can be achieved when mapping this variable across space and along main rivers for a given stream gauging network. Three stochastic interpolation schemes for estimating average annual runoff across space are evaluated and compared. Two of the schemes firstly interpolate runoff to a regular grid net and then integrate the grid values along rivers. One of these schemes includes a constraint to account for the lateral water balance along the rivers. The third scheme interpolates runoff directly to points along rivers. A drainage basin in China with 20 gauging sites is used as a test area. In general, all three approaches reproduce the sample discharges along rivers with postdiction errors along main river branches around 10%. Using more objective cross-validation results, it was found that the two schemes based on basin integration, and especially the one with a constraint, performed significantly better than the one with direct interpolation to points along rivers. The analysis did not allow identification of possible influence of surface water use.

Key words | annual runoff, river network, stochastic interpolation, water balance constraint

Ziqi Yan (corresponding author)
State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin,
China Institute of Water Resources and Hydropower Research,
Beijing 100038,
China
E-mail: *yanziqi@yahoo.com.cn*

Jun Xia
Institute of Geographical Sciences and Natural Resources Research,
Chinese Academy of Sciences,
Beijing 100101,
China

Lars Gottschalk
Irina Krasovskaia
Department of Geosciences,
University of Oslo, N-0315 Oslo,
Norway

INTRODUCTION

The long-term mean value of runoff is the basic descriptor of available water resources. It is an essential task to evaluate the accuracy with which these resources can be determined for a region or a larger drainage basin given the existing gauging network. Environmental change is part of the reality and many efforts have been made to estimate the magnitude and character of changes in water balance components and runoff in particular. A key issue is to confront possible changes with the accuracy in water resources estimates. The focus for the present study is to elucidate this accuracy by applying stochastic interpolation for mapping runoff across a large Chinese drainage basin. The main aim is thus to provide a general impression of the accuracy in absolute and relative values across the basin and along the river network. Differences in methodological approaches influence accuracy estimates, and to be able to evaluate this aspect three different approaches for interpolation of runoff are tested. Postdictive and predictive accuracy of the approaches are evaluated to allow comparison of performance and estimation at ungauged sites (Gauch 2003).

The long-term annual mean value is the statistical characteristic that can be expected to have the highest accuracy. It is the first order moment and as such the basic characteristic of the distribution function of runoff. It is the only characteristic that does not depend on the dynamics of the runoff process and thus can be interpolated as a purely space process. To be able to interpolate higher order moments across space, time-space interpolators are needed (Gottschalk *et al.* 2006, 2011a, b). Various methods have been developed for mapping runoff features, ranging from manual contouring to automated procedures for interpolation, and an overview is found in Gottschalk & Krasovskaia (1998).

A more recent development is to apply lumped rainfall-runoff models (Jutman 1995; Bishop *et al.* 1998; Beldring *et al.* 2002) with regionalized parameters. The spatial interpolation problem is then twofold – mapping the

doi: 10.2166/nh.2012.073

climatological input variables *viz.* precipitation and evapotranspiration and regionalization of parameters of the hydrological model. The uncertainty aspects of the problem become very complicated in this case. The chosen spatial resolution of the maps is often very high, e.g. 1×1 km, which does not really match the actual accuracy in this approach.

Here, we choose to use a stochastic interpolation approach for mapping runoff, which will allow us to explicitly grasp basic uncertainties. A later step is to look for approaches for closing the water balance with full consideration of the uncertainty in its different components (Yan *et al.* 2012). The theoretical background to the stochastic interpolation methods was introduced in the works of Kolmogorov (1941) and Wiener (1949). These methods have been developed further rather independently in several disciplines. In meteorology, Gandin's method for optimal interpolation (Gandin 1963) is a landmark in the development of objective methods in this discipline. Geostatistics (so-called kriging) is mainly directed towards the interpolation problem (Krige 1951; Matheron 1963). Cressie (1990) gives a good overview of the origin and background of stochastic interpolation methods.

Stochastic interpolation of runoff is not a standard task and special attention is needed with respect to three key issues: (i) consideration of the fact that runoff data refer to the area of drainage basins (the support); (ii) definition of distances between basins; (iii) upstream–downstream relations along the river net. The latter point has two aspects, namely, the induced dependence between nested drainage basins and the fact that lateral water balance should be satisfied along rivers. Gottschalk (1993a, b) introduced a so-called hydrostochastic approach that takes into consideration both of these aspects. This approach has been further elaborated by Sauquet *et al.* (2000). Skøien *et al.* (2006) applied a similar method (topological kriging or top kriging) for interpolation of floods at ungauged locations. This method assumes that local runoff generation or rainfall excess can be defined at each point in space and can be integrated over a catchment to form catchment runoff, and exploits the topology of nested catchments in addition to the spatial correlation of runoff. A further development concerns the estimation of time series of runoff at ungauged locations by a space-time estimation procedure, taking into account both spatial and temporal correlations of runoff (Skøien & Blöschl 2007). The study presented herein contributes to the improvement of methods for interpolation of runoff by evaluating and comparing the performance of three stochastic interpolation schemes: IS1, stochastic interpolation of runoff to a regular grid, which in a second step is integrated along the river network; IS2, stochastic interpolation of runoff directly to points along rivers; IS3, similar to IS1 but with an added water balance constraint along rivers. The methods are compared with respect to biases and accuracy.

The study area is the Huaihe River basin upstream of the site Bengbu in north-eastern China. There are a huge number of dams and enormous pressure on water resources from the intensive human activity. Thus, attempts to adequately describe runoff in the region represent a challenge and face complex problems. Manually derived contour maps are still often used in this region to describe distribution of runoff in space. That is why objective quantitative estimation of runoff characteristics has become the priority issue for water resources management in this river basin. There is substantial water use in this area but no quantitative information is available concerning water exploitation.

The paper is organized as follows: in the next section, the theoretical background for the three interpolation schemes is presented, and the study region and data are described in the following section. The results are given in the final section in terms of denested runoff, estimated spatial covariance from these denested values, interpolated runoff maps, and cross-validation.

THEORETICAL BACKGROUND FOR THE USED INTERPOLATION SCHEMES

Interpolation of runoff is more complex than interpolation of other components of the water balance. Runoff observations might be nested, i.e. the drainage basin of one gauging station is part of a larger basin contributing to another gauging station. It is therefore worthwhile to make a 'denesting' of observed runoff within a larger basin first.

It is common that a larger drainage basin with a gauging station at its outlet contains one or several other gauging stations upstream. Runoff production for the part of the basin area between the outlet station and an upstream one should be possible to obtain, in principle, by a simple subtraction of the discharge values at the upstream gauging station from those at the downstream one. For a simple case with only one upstream station, the denesting thus involves two values only $q(A_1)$ and $q(A_2)$, where A_2 is the upstream area and A_1 is the intermediate area calculated from $A_1 = A - A_2$, where A is the area at the outlet. $q(A_1)$ is estimated from:

$$q(A_1) = (Aq(A) - A_2 q(A_2))/A_1 \quad (1)$$

where q is expressed in mm/year. There is a resemblance between this approach and the Thiessen method, as the variable takes constant values over areas, which, of course, are defined from quite different principles. There is also a drawback in the denesting when Equation (1) represents a difference between too large numbers, which compromises the accuracy. On the other hand, the procedure of denesting might be a first good control of the quality of data. Errors are directly detectable. Later, when applying more sophisticated procedures for interpolation, these errors will still be present although they are more difficult to detect.

Interpolation of data with support to grid cells/sub-basins

We can now turn to the interpolation equation for runoff working with denested runoff data. The formula for estimation of runoff $\hat{q}(a_i)$ for a target area a_i by stochastic interpolation from gauged runoff at M_i neighbouring drainage basins (denested areas) $q(A_j)$, $j = 1,\ldots,M_i$ is (Gottschalk 1993a):

$$\hat{q}(a_i) = \sum_{j=1}^{M_i} \lambda_j^i q(A_j) = \Lambda_j^T \mathbf{Q} \quad (2)$$

where Λ_i^T is the transposed column vector of lambdas and $\mathbf{Q}$ is the column vector of the observed runoff. The weights λ_j^{i}; $j = 1,\ldots,M_i$, related to each of the grid cells a_i, are evaluated from $\Lambda_i = \mathbf{C}^{-1}\mathbf{C}_{0i}$ where:

$$\Lambda_i = \begin{pmatrix} \lambda_1^i \\ \lambda_2^i \\ \vdots \\ \lambda_{M_i}^i \\ \mu^i \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} \mathrm{cov}(A_1, A_1) & \mathrm{cov}(A_1, A_2) & \cdots & \mathrm{cov}(A_1, A_{M_i}) & 1 \\ \mathrm{cov}(A_2, A_1) & \mathrm{cov}(A_2, A_2) & \cdots & \mathrm{cov}(A_2, A_{M_i}) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{cov}(A_{M_i}, A_1) & \mathrm{cov}(A_{M_i}, A_2) & \cdots & \mathrm{cov}(A_{M_i}, A_{M_i}) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$$\mathbf{C}_{0i} = \begin{pmatrix} \mathrm{cov}(A_1, a_i) \\ \mathrm{cov}(A_2, a_i) \\ \vdots \\ \mathrm{cov}(A_{M_i}, a_i) \\ 1 \end{pmatrix}$$

The elements in the matrix $\mathbf{C}$ represent covariances between mean annual runoff at the M_i^q observation stations in the search neighbourhood of grid cell a_i – $\mathrm{cov}(A_i, A_j) = \mathrm{cov}(\bar{X}(A_i), \bar{X}(A_j))$ – while the elements of the column vector $\mathbf{C}_{0i}$ are the covariances between runoff at these observation stations and runoff at grid cell a_i – $\mathrm{cov}(A_j, a_i) = \mathrm{cov}(\bar{X}(A_j), \bar{X}(a_i))$, where M_i^q is the number of neighbouring observed runoff q used for interpolate runoff for a target area a_i. $\bar{X}(A)$ is the denested value of basin A. These covariances are evaluated from a river covariance model (Gottschalk *et al.* 2006, 2011a).

We introduce a length coordinate l along rivers from the main drainage basin outlet and upstream. The area for this main drainage basin is denoted $A(0)$ and the drainage area for a certain upstream sub-basin – $A(l)$. Let the area $A(l)$ be approximated by M_l grid cells a_i, $i = 1,\ldots,M_l$. and $\sum_{l=1}^{M_l} A(l) = A(0)$. For each grid cell we have estimated the runoff by applying Equation (2). The runoff $q(A(l))$ at the site l can be estimated by summation over all the M_l upstream grid cells contribution to runoff at this site:

$$q'(A(l)) = \sum_{i=1}^{M_l} \frac{a_i}{A(l)} \left(\sum_{j=1}^{M_i} \lambda_j^i q(A_j) \right) \quad (3)$$

The results of the interpolated runoff grid map can then alternatively be shown as map showing only rivers with a changing colour intensity in accordance with the runoff

value estimated from Equation (3). The estimation variance for each grid a_i expressed in covariances and with the constraint on the sum of weights to be equal to one is:

$$\hat{\sigma}^2(a_i) = \sigma^2(a_i) - \sum_{k=1}^{M} \lambda_{ik}\mathrm{cov}(A_k, a_i) + \mu_i \tag{4}$$

The estimation variance of $\hat{q}(A(l))$is established as:

$$\sigma'^2(A(l)) = \mathrm{var}[q'(A(l))] = \frac{1}{A(l)^2}\sum_{j=1}^{M_l}\sum_{k=1}^{M_l} a_j\hat{\sigma}(a_j)a_k\hat{\sigma}(a_k) \times\rho(a_j, a_k) \tag{5a}$$

where $\rho(a_j, a_k)$ is the correlation in the estimation error between grids j and k. In case of equal sized grids so that $M_l = A(l)/a$ the equation can be rewritten as:

$$\sigma'^2(A(l)) = \mathrm{var}[q'(A(l))] = \frac{a}{A(l)}\sum_{j=1}^{M_l}\hat{\sigma}^2(a_j) + 2\left(\frac{a}{A(l)}\right)^2\sum_{j=1}^{M_l}\sum_{k=1}^{j-1}\hat{\sigma}(a_j)\hat{\sigma}(a_k)\rho(a_j, a_k) \tag{5b}$$

The interpolation equation to grid cells Equation (2) and the summation along rivers Equation (3) is the first interpolation scheme to be tested herein (denoted IS1 in the following).

A result with only runoff estimates along rivers in accordance with Equation (3) can also be achieved by successive interpolation of runoff along the distance coordinate l following the main branches:

$$q''(A(l)) = \sum_{k=1}^{M_l}\lambda_k^l q(A_k) \tag{6}$$

The estimation error for this case is:

$$\sigma''^2(A(l)) = \mathrm{var}[q''(A(l))] = \sigma^2(A(l)) - \sum_{k=1}^{M_l}\lambda_k^l\mathrm{cov}(A_k, A(l)) + \mu_l \tag{7}$$

Equation (6) is the second interpolation scheme to be tested (IS2).

There is no obvious argument to claim the superiority of one interpolation approach over the other, for instance, having less estimation error. This depends on the specific application. The advantage of the former Equation (3) would be that values along the river is definitely the sum of upstream runoff formation, which is not controlled in the same manner in the latter case.

Accounting for lateral water balance along rivers

A further step would be to combine the two approaches into one interpolation procedure bringing in the advantages from both. This can be done by introducing the continuity Equation (3) as a constraint in the interpolation scheme Equation (1) (Gottschalk 1993b) replacing the constraint that the sum of weights should sum up to one. The condition is that the interpolation is performed in a drainage basin with an observation station at the outlet. Let us denote the drainage area down to the outlet point by $A(0)$, as before. The water balance equation down to this point is then written:

$$q(A(0)) = \sum_{i=1}^{M_0}\frac{a_i}{A(0)}\left(\sum_{j=1}^{M}\lambda_j^i q(A_j)\right) \tag{8}$$

where M_0 is the number of grids in $A(0)$ and M the number of stations. The search neighbourhood is not allowed to be changed in this case but all M stations should be used for interpolation to each grid. From Equation (8) the following constraints for the weights can be deduced:

$$\sum_{i=1}^{M_{01}}\frac{a_i}{A(0)}\lambda_1^i = 1;\ \sum_{i=1}^{M_1}\frac{a_i}{A_j}\lambda_j^i = 0,\quad j = 2, \ldots, M \tag{9}$$

Applying the Lagrange method we find the following set of complementary equations to determine the weights:

$$\sum_{j=1}^{M}\lambda_j^i\mathrm{cov}(A_k, A_j) + (a_i/A(0))\mu_k = \mathrm{cov}(A_k, a_i);\quad k = 1, \ldots, M_0,\ i = 1, \ldots, M \tag{10}$$

The weights λ_i^j, $i = 1, \ldots, M_0, j = 1, \ldots, M$ need to be estimated simultaneously for all M_0 grid elements. Optimal

weights are thus found solving the following matrix equations:

$$\Lambda = \mathbf{C}^{-1}\mathbf{C}_i \tag{11}$$

with:

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \vdots \\ \mathbf{L}_{M_1} \\ \boldsymbol{\mu} \end{bmatrix};\ \mathbf{L}_i = \begin{bmatrix} \lambda_1^i \\ \lambda_2^i \\ \vdots \\ \lambda_M^i \end{bmatrix};\ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix}$$

$$\mathbf{C}_i = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_{M_1} \\ \mathbf{1} \end{bmatrix};\ \mathbf{G}_i = \begin{bmatrix} \mathrm{cov}(A_1, a_i) \\ \mathrm{cov}(A_2, a_i) \\ \vdots \\ \mathrm{cov}(A_M, a_i) \end{bmatrix};\ \mathbf{1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{K} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_1 \\ \mathbf{0} & \mathbf{K} & \mathbf{0} & \vdots & \mathbf{V}_2 \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{K} & \mathbf{V}_{M_i} \\ \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_{M_i} & \mathbf{0} \end{bmatrix};$$

$$\mathbf{K} = \begin{bmatrix} \mathrm{cov}(A_1,A_1) & \mathrm{cov}(A_1,A_2) & \cdots & \mathrm{cov}(A_1,A_M) \\ \mathrm{cov}(A_2,A_1) & \mathrm{cov}(A_2,A_1) & \cdots & \mathrm{cov}(A_2,A_M) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(A_M,A_1) & \mathrm{cov}(A_M,A_2) & \cdots & \mathrm{cov}(A_M,A_M) \end{bmatrix}$$

$$\mathbf{V}_i = \begin{bmatrix} a_i/A_1 & 0 & \cdots & 0 \\ 0 & a_i/A_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_i/A_1 \end{bmatrix}$$

The matrix to be inverted is a doubly-bordered block diagonal. The diagonal blocks are classical covariance matrixes and the bordering terms refer to the global hydrological constraint expressed by Equation (6). With the continuity constraint the estimation variance for a grid element is replaced by (Gottschalk 1993b):

$$\sigma^2(a_i) = \mathrm{var}(a_i) - \sum_{j=1}^{N}\left(\lambda_j^i \mathrm{cov}(A_j, a_i) + \frac{a_i}{A(0)}\lambda_j^i\mu_j\right) \tag{12}$$

The matrix Equation (11) might be very large if many stations and many grids are in use. To reduce the size of the matrix equation and to also have better control of the problem, Sauquet *et al.* (2000) suggested a hierarchical disaggregation approach in accordance with the following scheme: (i) define the total area where runoff is known and for which the water balance constraint should be satisfied; (ii) divide this area into fundamental units (grid cells/sub-basins); (iii) select gauging stations to define an observation network consistent with the scale of the fundamental units; (iv) construct a covariance function consistent with the spatial structure of the data set; (v) interpolate runoff to each element so that the runoff balance constraint is satisfied for the total area. The estimated values for the different fundamental units might now be the new starting points for resolving spatial variability within these units. The steps from (ii) to (v) are then repeated as long as new information on a smaller scale is available to define data sets for the finer resolution. This hierarchical interpolation scheme with water balance constraint is the third one to be tested (IS3).

For all three interpolation schemes, the variations in the estimation error mainly reflect the eventual nearness in a correlation space between the point of interpolation and observation sites. The correlation of runoff is very complex. High correlation is caused by nearness in space, but also by common area along nested basins, and this latter factor is obviously the strongest one. The differences in constraints between the three interpolation schemes also influence the estimation errors.

Covariance of long-term mean values

The estimation of the covariance between mean values is a different problem compared to the calculation of covariance from the original values of a spatial random process. The premises of spatial homogeneity are hardly satisfied for hydrologic random processes and it is common that the correlation structure demonstrates homogeneity while the covariances do not. The reason for this is the fact that the correlation is based on a standardization of the initial data and by this the non-homogeneity in the mean and the variance is eliminated. This non-homogeneity across space is commonly assumed to behave in a regular way, i.e. a deterministic trend surface. Temporal mean values behave as irregular in space as individual events do (Gottschalk 1993b). They are the result of complex interaction between

a heterogeneous landscape and hydrometeorological processes giving rise to a pattern of variability that is described as a spatial random process (random field). We only have one mean value per point, i.e. only one realization of the random process, but each of these mean values is estimated from a time sequence of data at each site. The point cov_p spatial covariance of mean values is derived as:

$$\mathrm{cov}_p(\mathbf{u}_i, \mathbf{u}_j) = \mathrm{cov}_p\left(\bar{X}(\mathbf{u}_i), \bar{X}(\mathbf{u}_j)\right) = \rho\left(X(\mathbf{u}_i), X(\mathbf{u}_j)\right) \sigma_{\bar{X}(\mathbf{u}_i)}\sigma_{\bar{X}(\mathbf{u}_j)} + \left(\bar{X}(\mathbf{u}_i) - \bar{X}(\mathbf{u}_\bullet)\right)\left(\bar{X}(\mathbf{u}_j) - \bar{X}(\mathbf{u}_\bullet)\right) \tag{13}$$

where $\bar{X}(\mathbf{u}_\bullet)$ is the point process in space $\mathbf{u}$. It is generally recommended to use semivariogram for characterizing the spatial variation when only one realization of the random process is available. For the present case, the point semivariogram is:

$$\begin{aligned}\gamma\left(\bar{X}(\mathbf{u}_i), \bar{X}(\mathbf{u}_j)\right) &= \frac{1}{2}\sigma^2_{\bar{X}(\mathbf{u}_i)} + \frac{1}{2}\sigma^2_{\bar{X}(\mathbf{u}_j)} \\ &\quad - \rho\left(X(\mathbf{u}_i), X(\mathbf{u}_j)\right)\sigma_{\bar{X}(\mathbf{u}_i)}\sigma_{\bar{X}(\mathbf{u}_j)} \\ &\quad + \frac{1}{2}\left(\bar{X}(\mathbf{u}_i) - \bar{X}(\mathbf{u}_\bullet)\right)^2 + \frac{1}{2}\left(\bar{X}(\mathbf{u}_i) - \bar{X}(\mathbf{u}_\bullet)\right)^2 \\ &\quad - \left(\bar{X}(\mathbf{u}_i) - \bar{X}(\mathbf{u}_\bullet)\right)\left(\bar{X}(\mathbf{u}_j) - \bar{X}(\mathbf{u}_\bullet)\right) \\ &= \frac{1}{2}\mathrm{var}\left(\bar{X}(\mathbf{u}_i)\right) + \frac{1}{2}\mathrm{var}\left(\bar{X}(\mathbf{u}_i)\right) \\ &\quad - \mathrm{cov}\left(\bar{X}(\mathbf{u}_i), \bar{X}(\mathbf{u}_j)\right)\end{aligned} \tag{14}$$

For the interpolation of the mean values it is completely equivalent to use covariances (Equation (13)) or semivariogram (Equation (14)) for interpolation. In the following we use the covariance function. The point covariance function needs to be averaged over the basin area to apply to runoff:

$$\mathrm{cov}(X(A_1), X(A_2)) \approx \frac{\sigma(A_1)\sigma(A_2)}{A_1 A_2}\int_{A_1}\int_{A_1} \mathrm{cov}_p(|\mathbf{u}' - \mathbf{u}''|)\,\mathrm{d}\mathbf{u}'\,\mathrm{d}\mathbf{u}'' \tag{15}$$

Distance measure

The stochastic approach is widely used for interpolation of meteorological fields (Creutin & Obled, 1982; Dingman *et al*. 1988; Goovaerts 2000) but it needs to be modified for interpolation of runoff characteristics since runoff observations are related to specific areas rather than to points. In particular, a relevant distance between pairs of basins has to be defined. Huang & Yang (1998) and Merz & Blöschl (2005) allocate the representative value of the runoff depth to the centre of gravity of the basin and thus choose the distance between centres of gravity. The alternative is to use the average of all possible distances between two basins, here named Ghosh distance (Ghosh 1951). The Ghosh distance thus replaces the classical Euclidian distance. We denote the density function of all possible distances between/within line segments or areas by $f(\lambda)$ and its expected value by m_Λ, which defines the distance measure.

The use of the Ghosh distance also allows finding an approximate numerical method for the integral expression in Equation (15). The covariance between two basins A_1 and A_2 can be expressed by (Matérn 1960):

$$\begin{aligned}\mathrm{cov}(A_1, A_2) &= \int_{A_1}\int_{A_2} \mathrm{cov}(\mathbf{u}, \mathbf{v})\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v} \\ &= \int_{\min(\lambda)}^{\max(\lambda)} \mathrm{cov}(|\lambda|)f(\lambda)\mathrm{d}\lambda = E[\mathrm{cov}(|\lambda|)]\end{aligned} \tag{16}$$

This expression invites to further simplifications by expanding into a Taylor series, i.e.:

$$E[\mathrm{cov}(|\lambda|)] = \mathrm{cov}(E[|\lambda|]) + \frac{1}{2}\frac{\partial^2 \mathrm{cov}}{\partial\lambda^2}\Big|_{E[|\lambda|]}\mathrm{Var}[|\lambda|] + \frac{1}{6}\frac{\partial^3 \mathrm{cov}}{\partial\lambda^3}\Big|_{E[|\lambda|]}\mu_3[|\lambda|] + \cdots \tag{17}$$

To be able to use this approximation the lower order moments, the mean $m_\Lambda = E[|\lambda|]$, the variance $\sigma^2_\Lambda = \mathrm{Var}[|\lambda|]$, etc. of the distribution function $f(\lambda)$, should be known. For runoff and drainage basins with complex geometry, it can highly simplify the analysis (Gottschalk 1993a; Gottschalk *et al*. 2011b). A truncation after the first term results in the approximation:

$$\mathrm{cov}(A_1, A_2) \approx \mathrm{cov}(m_\Lambda(A_1, A_2)) \tag{18}$$

which will be further exploited herein.

GEOGRAPHICAL SETTING AND DATA USED

The study area is located upstream of Bengbu Sluice in the downstream part of the Huaihe River basin, and has an area of 121,000 km^2 (Figure 1). The climate in this area is in general characterized by monsoonal weather conditions with significant seasonal changes of precipitation, which is mainly concentrated to summer (60–70% of the annual precipitation). There is a gradient in the average annual precipitation across the basin from about 1,000 mm in the southeast to less than 600 mm in the northwest, while the highest precipitation is observed in the inner mountain areas. Temperature variation in the basin is significant due to the continental climatic conditions, with high summer temperature 30° (July) and low winter temperature 0° (January). Water problems are significant in this region. Both spatial and temporal distribution of water resources is uneven and the human impact on water resources is huge. Detailed information on water use is unfortunately lacking.

The basic data used are average monthly runoff series at 20 hydrological stations located in the study area. These time series covering the period from 1958 to 1998 were used to estimate the long-term annual average runoff at each station (cf. Table 1).

The hydro-stochastic approach used requires a unique numeric description of the hierarchical structure of the river network as a background for runoff interpolation. A digital elevation model (DEM) with a resolution of 1×1 km was used to extract river network. Flow direction from each grid cell was obtained by standard methods assuming that there is only one outflow from each grid cell. Most of the study area is situated in flatland and the accuracy in extracting the river network directly from the DEM in these parts might be relatively poor. Corrections were needed using a

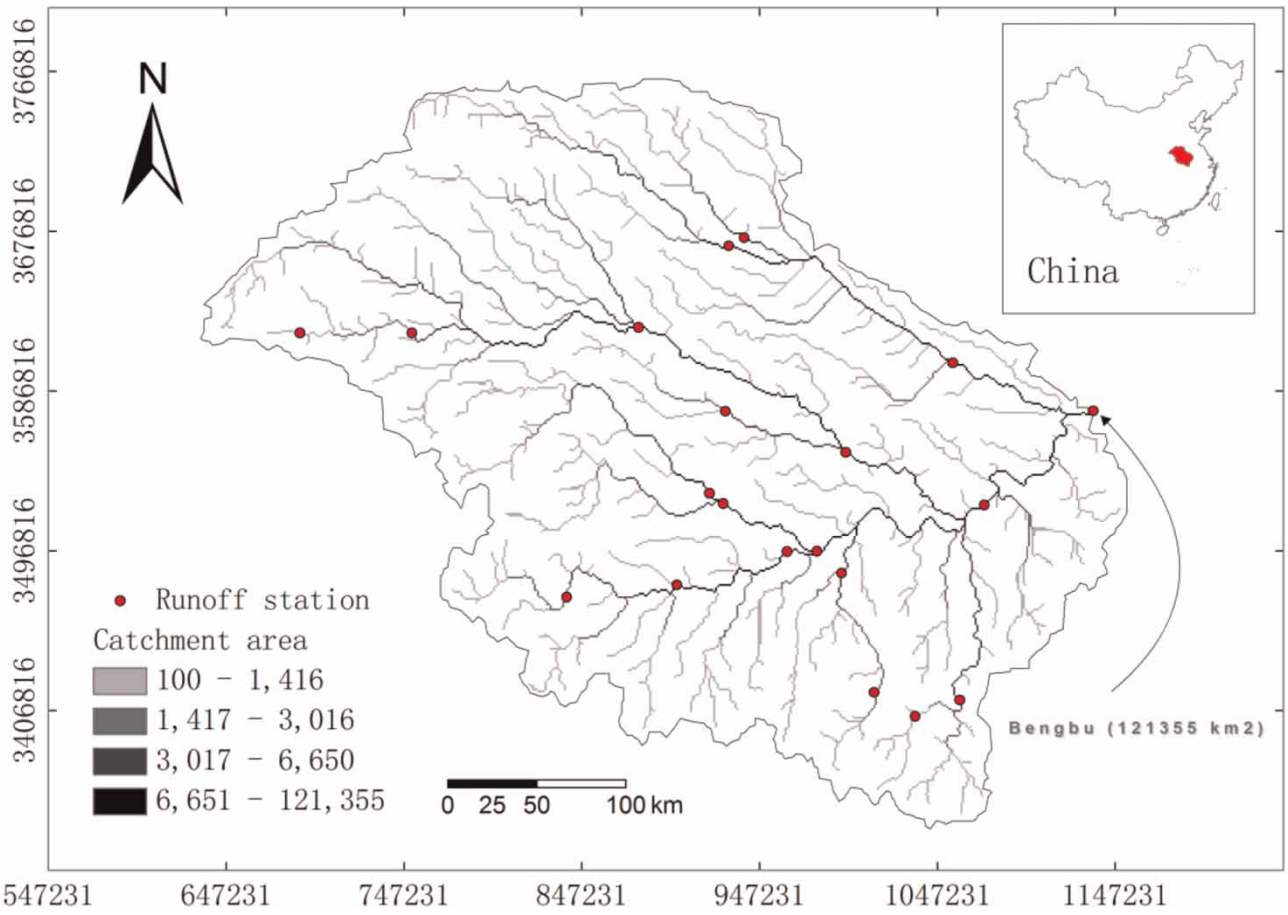

Figure 1 | The Huaihe Drainage basin in South-East China with the river network extracted from DEM and 20 gauging stations upstream Bengbu.

Table 1 | Long-term annual average runoff for the period 1958 to 1998 for 20 gauging stations in Huaihe basin, China

Station	Basin area (km^2)	m^3/s	mm/year
Baiguishan	2,730	25	293
Bantai	11,280	84	235
Bengbu	121,330	948	246
Changtaiguan	3,090	38	386
Fuyang	35,246	156	140
Hengpaitou	4,370	108	783
Huaibin	16,005	195	385
Jiangjiaji	5,930	99	528
Lutaizi	88,630	794	283
Meishan	1,970	44	700
Mengcheng	15,475	40	81
Shenxiu	3,094	13	132
Wangjiaba	30,630	316	326
Xianghongdian	1,476	35	749
Xincai	4,110	27	205
Xixian	10,190	134	415
Xuanwu	4,020	7	56
Zhaopingtai	1,416	17	371
Zhoukou	25,800	114	139
Zhuanqiao	3,410	7	61

previously digitized stream network. Since the accuracy of the extracted river net can directly affect the integration process along the river network, there are two requirements that the identified river network should meet:

1. The river network should pass through the 20 hydrometric stations in Table 1.
2. The drainage basin areas extracted for the 20 hydrometric stations should be as close as possible to the 'official' value (from the archives).

In order to meet these two requirements, it is essential to repeatedly extract and correct the digital river network until the extraction achieves the required accuracy. The extracted river network structure and the location of hydrological stations in Table 1 are shown in Figure 1. In this study, information about official basin area from the governmental database (China Ministry of Water Resources 2006) was used to check the derived area. Figure 2 uses a log-log scale to compare the final catchment areas A' extracted from DEM and the 'true' catchment areas A, which reveals that some uncertainties remain in the delineation of the 20 drainage basins in relation to the official figures. The regression line is given by $A' = 1.0494\ A – 0.198$ and the correlation coefficient is equal to 0.9552. All but one of the drainage basins seem to be correctly reproduced (deviation is less than 10%). The high deviation originates from one basin in plain area.

RESULTS

The application of the interpolation schemes contains four steps: (i) denesting runoff; (ii) estimation of a covariance structure of the denested values; (iii) interpolation to grids and integration along rivers or direct interpolation along rivers; and finally, (iv) objective evaluation of the performance of the three schemes applying cross-validation.

Denesting runoff

The principle for denesting was described above under Theoretical background for the used interpolation schemes. The background for the denesting is the structure of the identified river network illustrated in Figure 1 and the information concerning gauging stations shown in Table 1. The discharge(s) (m^3/s) measured at the neighbouring upstream gauging station(s) are subtracted from the value observed at a downstream gauging station successively from headwaters towards the outlet. From the gauged discharge in 20 basins new discharge values are determined for 20 denested basins. To eliminate scale effect within the data set due to the size of the basin, runoff $q(A_i)$ generated for denested sub-basin is expressed in mm/year. Figure 3 shows the average runoff in each denested sub-basin after the normalization of data to mm/year. The runoff $q(A_i)$ for the denested basins is the background data for the construction of a covariance function and for the interpolation exercises developed in the following.

The denested runoff map is very informative concerning the spatial signal in our data, i.e. the pattern of variability of runoff across space that the available gauging network is able to reveal. There are, for example, two small basins in the south-western part that show high runoff values that

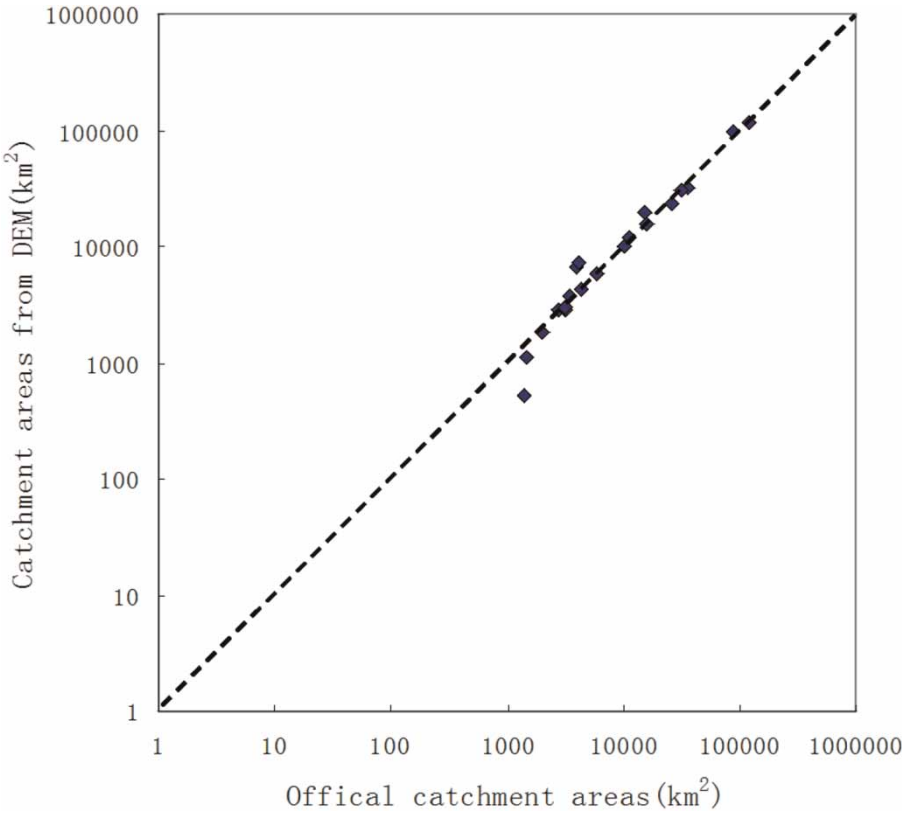

Figure 2 | Comparison of DEM estimated drainage areas with those from the official archives. A log-log scale is used.

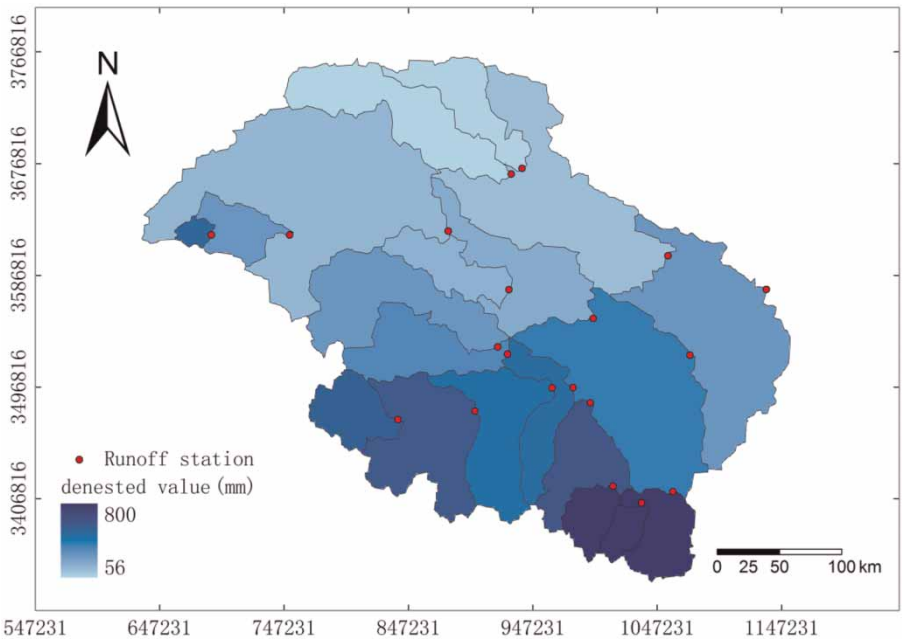

Figure 3 | Denested runoff estimated by disaggregation procedure (mm/year).

are probably representative for this part of the Huaihe basin. The downstream basins, in which these two small ones are nested, cover a large area that also borders the much drier north western part. As a result, the runoff for this larger basin averages out the dry north and the wet south and we are not able to detect any details. In the present study, we will accept the available runoff data as a base for the runoff mapping exercise. In an applied situation it would, of course, be necessary to look for supplementary information.

Covariance of runoff

The first step in the construction of a covariance function is the calculation of the distances between each pair of non-nested drainage basins following the hierarchy of the drainage network. The distances are estimated as Ghosh distances, i.e. as the average of all possible distances between the pair of basins. With these distances as a background, an experimental covariogram for non-nested basins can be drawn. This step is followed by a selection of possible theoretical models for the point process covariance function cov_p. Basin integrated covariance values were estimated for each pair of non-nested drainage basins applying Equation (15) and compiled for distance classes to be able to estimate a theoretical basin integrated covariogram (Figure 4). The choice of the best point process model is based on a graphical comparison between the experimental and theoretical covariograms. For the present data, a theoretical spherical covariance function with the following expression:

$$\mathrm{cov}_p() = \begin{cases} 0 & d > T \\ C_0\left[1 - 1.5\dfrac{d}{T} + 0.5\left(\dfrac{d}{T}\right)^3\right] & d \le T \end{cases} \quad (19)$$

gives an acceptable fit to the experimental one. The analysis was limited to pairs of sub-basins less than 100 km apart.

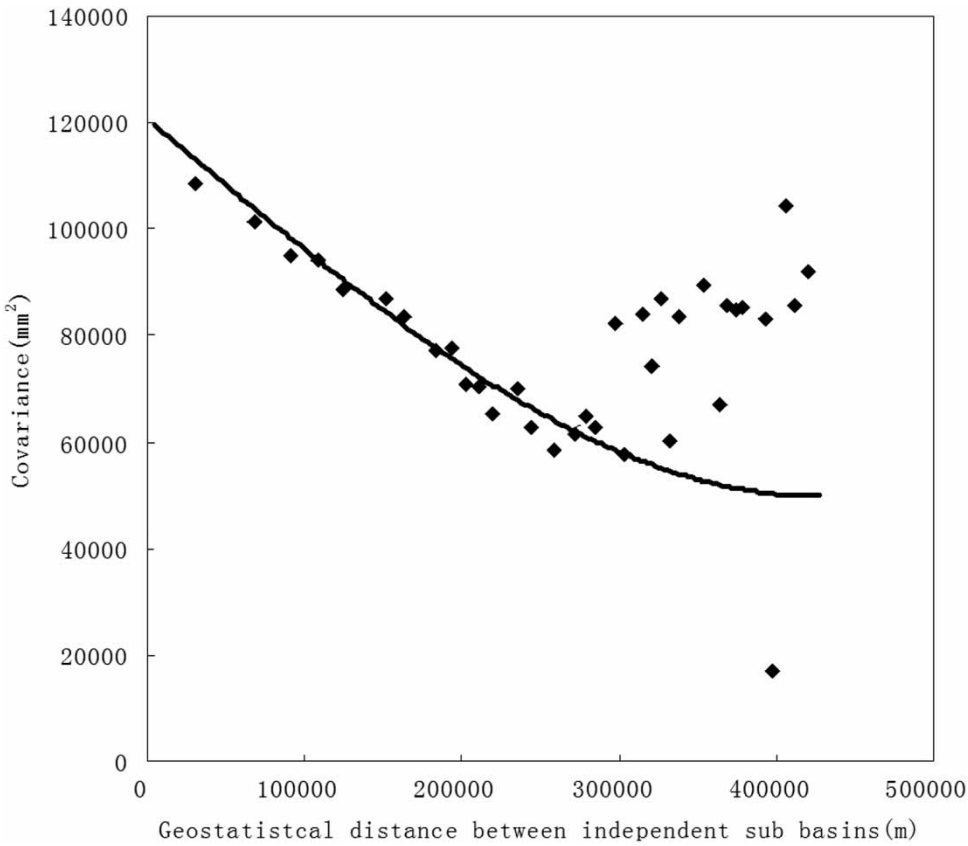

Figure 4 | Empirical covariogram estimated from denested runoff data and the fitted theoretical covariogram (Equation (19)).

The spatial scale coefficient was set to $T = 380$ km and $C_0 = 115{,}400$ (mm/year)^2.

Interpolated runoff maps

Two of the interpolation schemes, namely IS1 and IS3, in a first step interpolate to a grid cell network. A grid size of 20×20 km^2 was chosen to keep a balance between the spatial signal that is obtained from 20 gauging stations and the runoff map resolution. The resulting runoff grid map obtained with the scheme IS1 is shown in Figure 5, together with the map of the corresponding estimation variance. The value *q(a)* for a grid cell can be interpreted as the average surface runoff from this cell that may flow towards the river channels. The runoff map mirrors the main features already revealed by the denested map (Figure 3). The step function character of the denested map is, of course, filtered out to arrive at a smoother variation across space, but the footprints of the individual drainage basins are still strongly present in the map. The interpolation procedure obviously cannot compensate for the incompleteness in the spatial data coverage, as commented earlier in connection to the denested runoff map. The average runoff across all cells in the map is 238 mm/year, which is only slightly less than the runoff at the outlet 246 mm/year, i.e. even without introducing a lateral water balance constraint in the interpolation scheme along rivers this balance is well kept.

The estimation variance of interpolation scheme IS1 (Figure 5(b)) mainly reflects the location of the observation sites. As the drainage basin and grid cell support are taken into account in the interpolation scheme, the errors are not reduced to zero at these locations. Furthermore, the smallest errors are, as a rule, found upstream of the gauging station. In terms of estimation errors, the average across all cells is 38 mm/year and the minimum and maximum is 6 and 99 mm/year, respectively. On average, the relative interpolation error is about 15%, and in extreme cases it can be as high as 30%.

The interpolated runoff map by the scheme IS3 exposes, of course, the same main patterns as the map obtained with IS1 and is, for briefness, not shown here. It is rather the differences between these two maps (IS3-IS1), illustrated in Figure 6, that are interesting. On average, the differences are not significant in relation to the estimation errors. The average positive change is 11 mm/year over 114 cells of a total of 306 cells and the average negative change is –6 mm/year over 192 cells. Local maximum deviation is 124 mm/year and local minimum is –26 mm/year. The differences are not randomly distributed as they mainly follow drainage basin boundaries. A special feature of IS3 is increasing/decreasing upstream values if they deviate from the downstream value. The differences are thus explained by inconsistencies in values along rivers in the accumulated runoff that do not totally satisfy the lateral water balance. We can see at least two reasons for this. Firstly, the estimated long-term average discharge at different sites along the river can contain systematic errors, which can be positive and negative randomly among the sites. Secondly, river stream reaches can be losing water and/or there are water retrievals from the river. It is difficult to identify the real cause with the available information at present. There is, for instance, nothing in the pattern of deviations that would reveal large scale water retrieval activities. It is known that there is a substantial water use in the area but we have not been able to achieve estimates on how large they are in comparison with the deviations found here.

The estimation variance of the interpolation scheme IS3 is slightly higher than IS1, as can be expected from a theoretical point of view, since a constrained search for minimum variance might not find the absolute minimum as in a free search. On average, the estimation variance is 14% higher for IS3 compared to IS1. The pattern of estimation errors (cf. Figure 5) is exactly the same for both cases except for the slightly higher values of IS3.

In a second step the integrated values along rivers were calculated for these two maps and compared with the directly interpolated values to rivers using interpolation scheme IS2. The river interpolation procedure (IS2) gives us the estimated runoff value and estimated variances directly along the river network. We chose to illustrate the result as a map of discharges along the river network (Figure 7) and in a graph of discharge as a function of the distance from outlet (Figure 8).

The scheme IS2 directly yields the estimation variance along rivers for the integrated case applying Equation (7). For the other two schemes, Equation (5) allows an estimation of the variances as an integrated value for the error estimation maps with consideration of the spatial

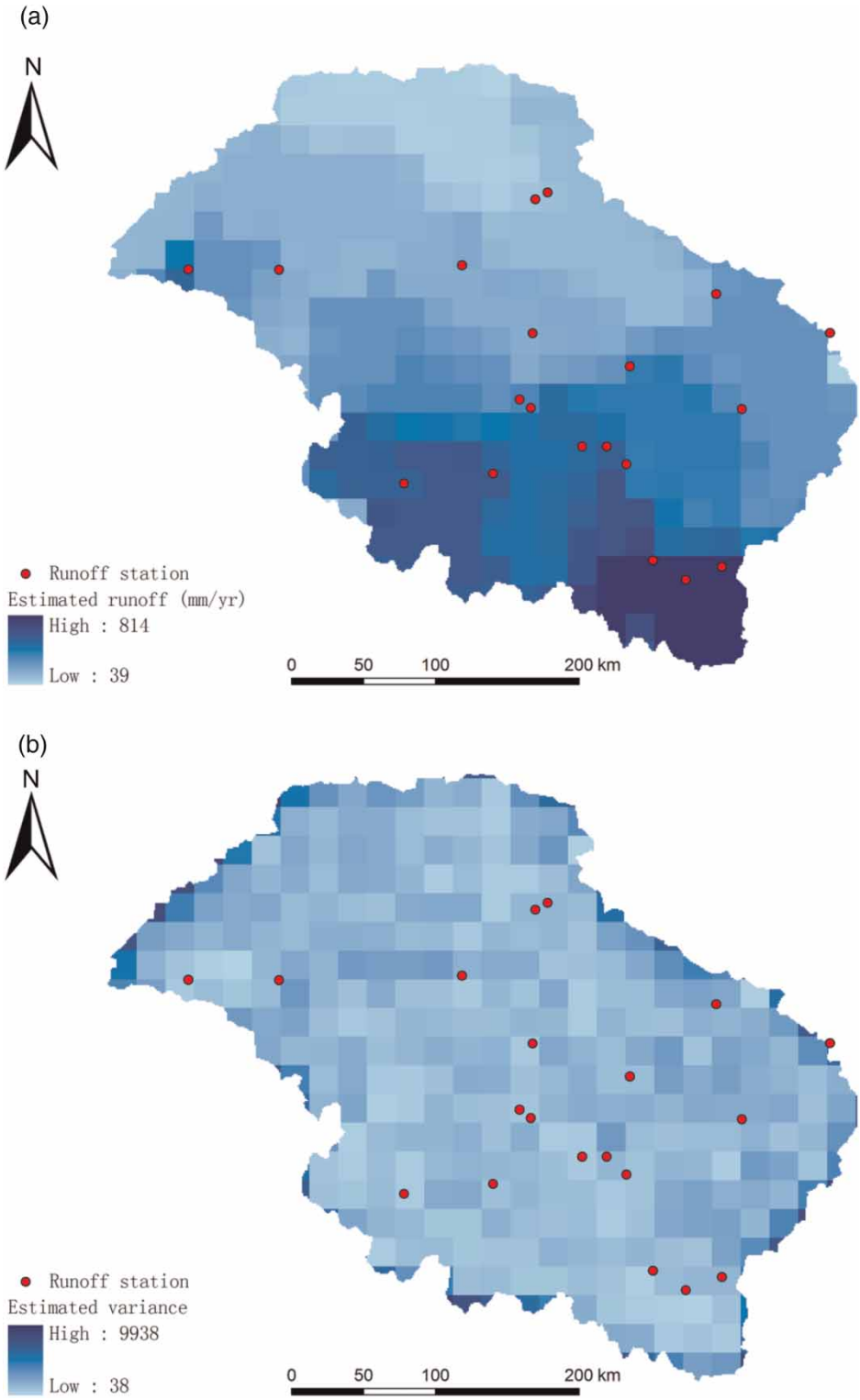

Figure 5 | Estimated runoff for grid cells calculated by interpolation scheme IS1 (a) and corresponding estimation variance (b).

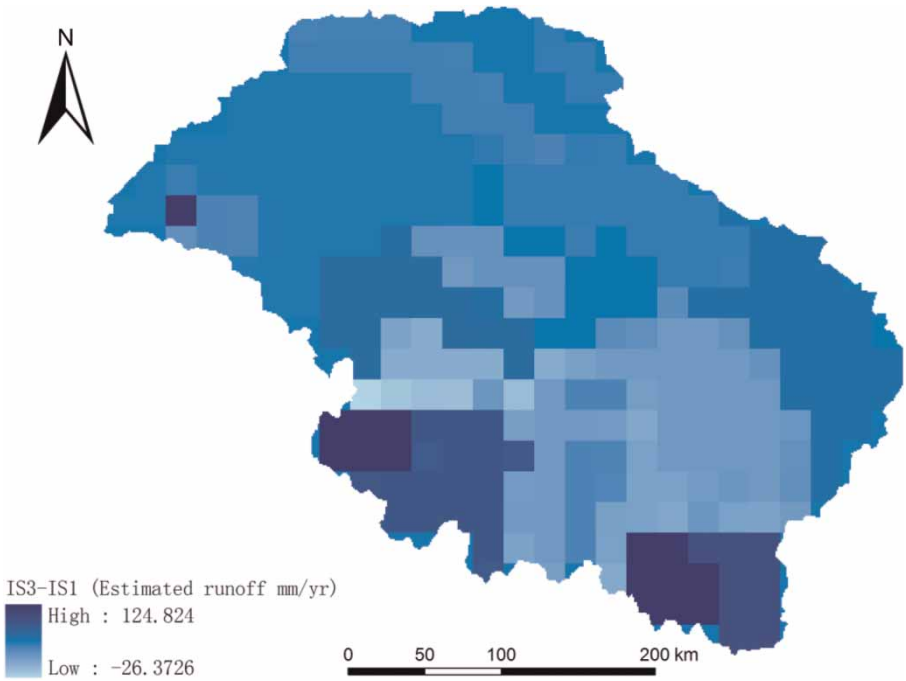

Figure 6 | Map of differences in the estimated runoff obtained by subtracting grid estimates of IS1 from IS3.

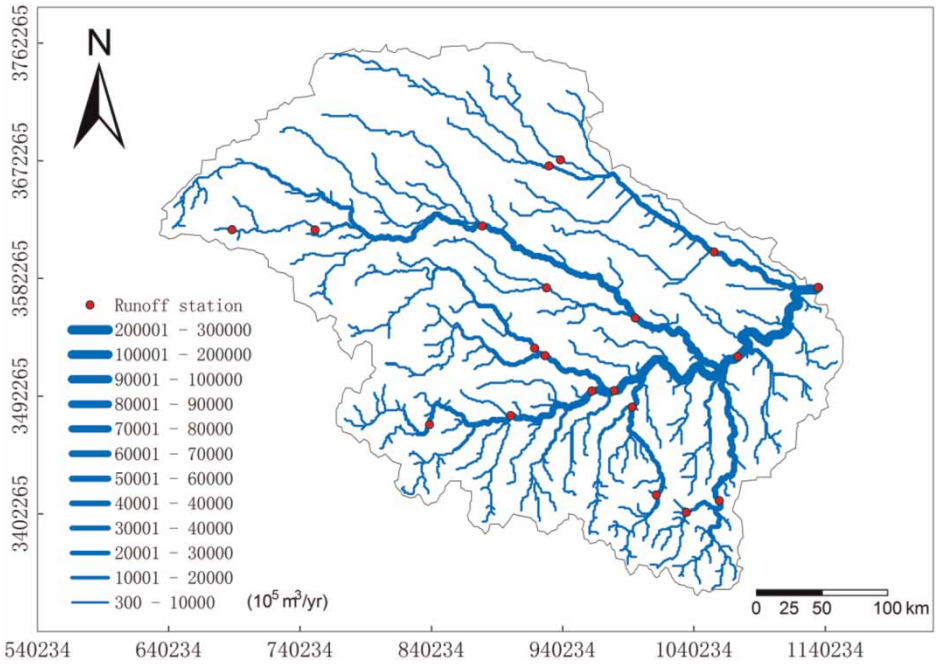

Figure 7 | Map of estimated long-term mean discharge for the Huaihe River basin along rivers for basins larger than 300 km^2 obtained by the IS3 scheme.

correlation in this error. In case of the estimation errors for the three schemes along the main branch (Figure 9), the IS3 scheme gives absolute estimation errors ranging from 25 to 35 mm/year, which is obviously higher than the other two schemes (20–30 mm/year). In terms of the relative estimation error, all three schemes yield values around 10%

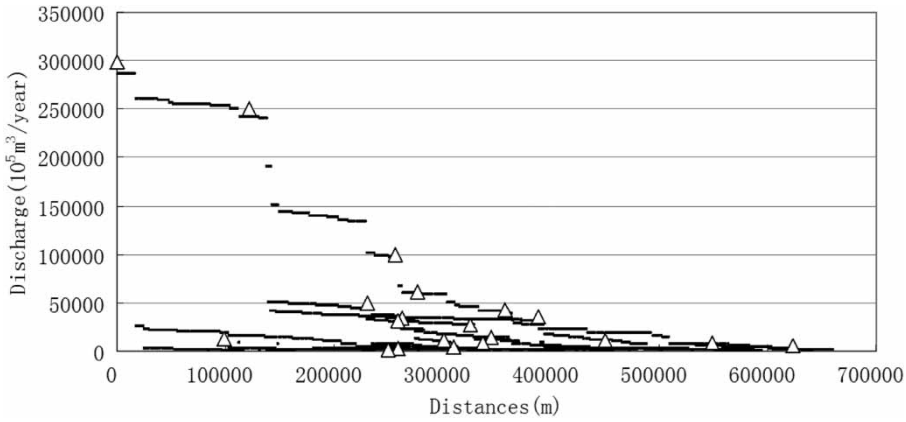

Figure 8 | Graphs of long-term mean discharge as a function of distance along the river to the outlet. Open triangles show observed long-term mean values.

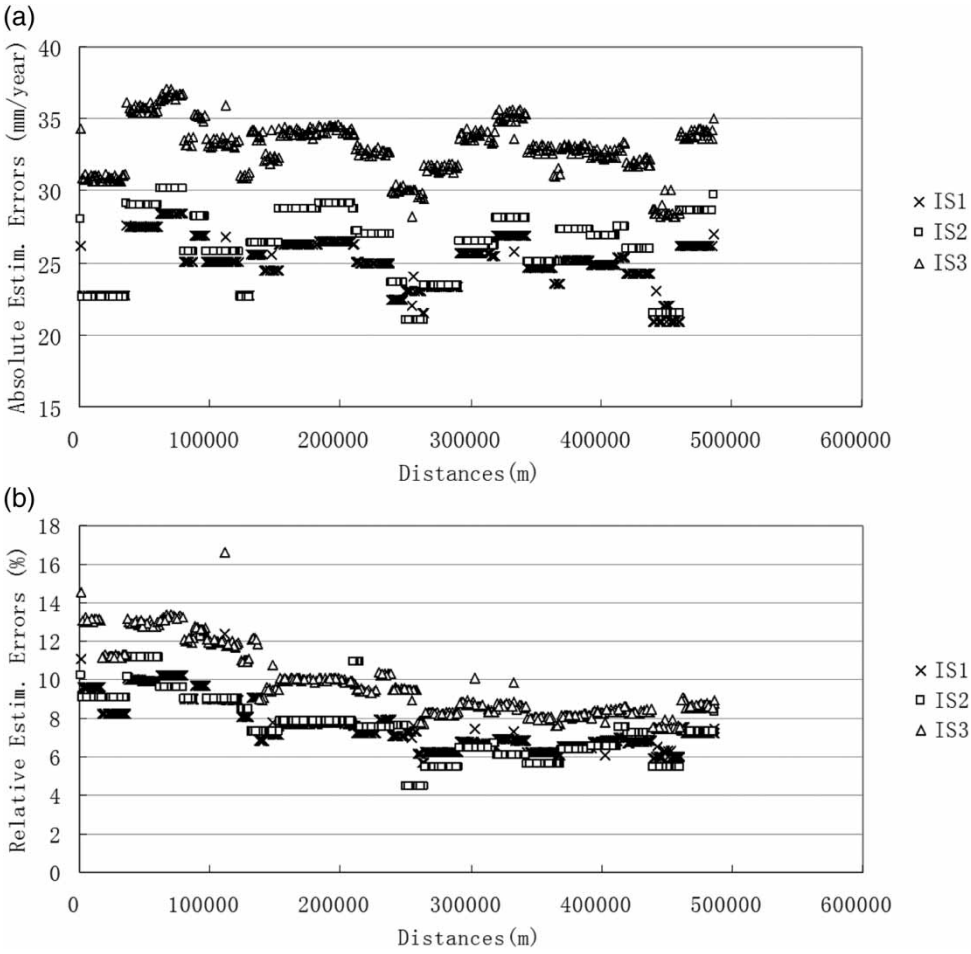

Figure 9 | Estimation error along the main branch of the Huaihe River in absolute values (mm/year) (a) and in relative values (%) (b).

on average for schemes IS1 and IS2 and a slightly higher value for scheme IS3. The ranges are 7–14% and 5–12%, respectively. In general, IS1 and IS2 show almost the same estimation errors and thus in the same way reflect the basic uncertainty related to the available observation sites and the estimated covariance structure. The methodological

differences in the derivation of the equations have very minor influence. The constraint of the IS3 scheme on the other hand slightly increases the uncertainty.

Cross-validation

In this paragraph, a jack knife procedure is applied for comparing the prediction accuracy of three interpolation schemes, i.e. the prediction error at ungauged sites. The analysis consists of excluding one gauging station in turn from the network and then estimating runoff at this site by applying the interpolation procedure on the remaining stations with the covariance model cov_p fitted to the full data set. It is assumed that the fitted covariance function is only slightly influenced by excluding one observation site. For each removed station, the observed runoff is compared to the estimated runoff produced by the watershed upstream of this point. This calculation implies withdrawing the subbasins upstream from the excluded station in the target partition. In the analysis, all stations apart from the outlet station were withdrawn in turn, i.e. the procedure is repeated 19 times so that each observation in the sample is used once as the validation data.

In order to compare the estimated value with that observed for the schemes IS1 and IS3, two steps are still required before assessing efficiency: firstly, runoff generated by the elements of the partition is calculated by interpolation to grids, and then the runoff value for the respective gauging station is obtained by aggregation along the river network to its site. For IS2, the procedure is easier as it estimates runoff at points along the river directly.

The performance of the three interpolation schemes are illustrated in Figure 10, where the observed values are plotted against the estimated ones. The correlation coefficient between observed and estimated values for the first scheme IS1 is equal to 0.93. The result is good but the regression line indicates an underestimation for some high values. The highest relative deviation is 53% and the same site also has the highest absolute deviation of 196 mm. The prediction standard error for this case is equal to 39 mm.

The second interpolation scheme, IS2, gives a correlation coefficient equal to 0.83 between observed and estimated values, which is significantly lower than the first case. The regression line also shows more deviation from the one-to-one line. Three sites have very high relative deviations of more than 100% and five sites have absolute deviations bigger than 150 mm. The prediction standard error is equal to 79 mm, which is twice as high compared to the previous case.

The results for the third interpolation scheme, IS3, are almost identical to those of IS1. The correlation coefficient for IS1 is equal to 0.94, which again is a good result, but as for IS1 the regression line indicates an underestimation for some headwater catchments. The highest relative deviation is found for the same site as for scheme IS1. The relative deviation is 33% and absolute deviation is 121 mm. The prediction standard error is equal to 37 mm, which is only slightly lower than that of IS1.

DISCUSSION AND CONCLUSIONS

Assessment of the uncertainties in the estimation of water resources is important for their sustainable management, when decisions have to be taken based on several uncertain factors. Environmental and climate change impact studies need a proper consideration of the magnitude of the uncertainties in original data and interpolation procedures to critically evaluate the magnitude of changes against them. Water balance estimation at ungauged basins is still another topic that could benefit from consideration of these uncertainties. An important issue is to be able to grasp the basic uncertainty in water balance estimates before speculating about possible changes.

The stochastic interpolation methods applied here are mainly used as a diagnostic tool for analysing the uncertainty of available hydrologic data of runoff. Interpolation of runoff is not a standard task that can be solved directly with the use of commercial geostatistical software packages. There are three key issues that should be accounted for, namely that runoff data refer to the area of drainage basins (the support) and consequently the distances to be considered are not between the points but between the irregular drainage basins. Furthermore, upstream–downstream relations along the river net need to be taken into consideration. These aspects have been accounted for by applying the so-called hydro-stochastic approach (Gottschalk 1993a, b) in the present study.

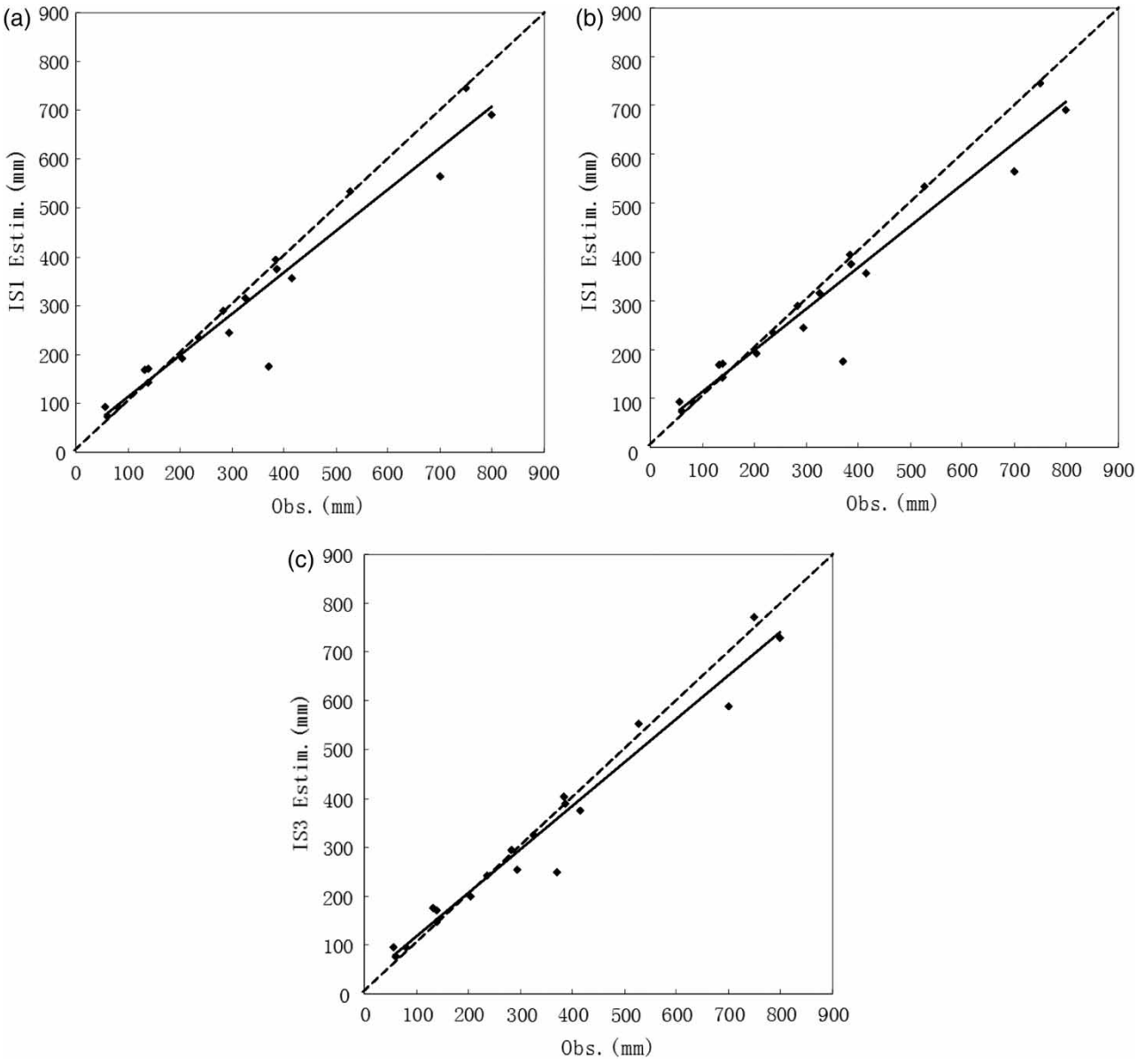

Figure 10 | Cross-validation of interpolated estimates of mean annual runoff with observations for IS1 (a), IS2 (b) and IS3 (c) (the regression line (dashed) and the one-to-one line (black) are both represented).

The results of the interpolation of long-term mean runoff investigated herein can be displayed in two different ways, either as estimated runoff values for cells in a regular grid across space expressed in mm/year or as estimated values along main river branches in a drainage basin. It is only in the latter case that it is possible to explicitly compare the estimated values with the observed ones. The grid runoff map is as such an implicit result and to be able to verify it by observations it must be assumed that the lateral surface water balance along rivers is satisfied. With this in mind the only objective method, in the sense that it is free from any basic assumptions, is interpolating runoff to points in the rivers, called interpolation scheme IS2 in the paper. In the other two interpolation schemes, IS1 and IS3, the values are first interpolated to a grid cell and in a second step the values are integrated along rivers assuming that surface water accumulates downstream without any losses.

In general, all three approaches perform acceptably with estimates of discharge along rivers with a postdiction accuracy along main river branches of around 10–12%. The scheme IS3 that contains a water balance constraint shows the smallest deviations which, of course, is due to this constraint. If we could be convinced that the lateral water balance was preserved to 100%, this may be the preferred scheme. The IS3 scheme adjusts the interpolated values so that they agree with the downstream values. To confirm this statement supplementary information is needed about water extractions from the river, possible influent river reaches, as well as a control of rating curves, etc., at the gauged sites in order to detect possible systematic errors in the estimates of long-term mean values. The detected deviations between the different schemes are in the order of 5%, which can be explained well by these factors.

While all three interpolation schemes are able fit to the sample equally well, there are significant differences in the results when turning to the cross-validation by the jackknifing procedure. The two schemes that include an interpolation to grids and then integration to points along main rivers, IS1 and IS3, give results with an accuracy twice as high as that of the scheme IS2, the prediction relative standard error compared to the global mean runoff is 15 and 33%, respectively. The fact that the water balance equation is included in the procedure for IS1 and IS3 schemes is a likely reason for these better results. The difference in accuracy between the schemes is minor, but the scheme IS3 benefits from the control at the outlet yielding slightly better results. A general conclusion would be that the inclusion of basic physical laws into statistical methods can drastically improve the estimation at ungauged sites.

It is obvious that estimated runoff/discharge along main rivers appears to perform better than estimates for grid cells across space, as the upstream–downstream control is present in the former case. On the other hand, the grid cell map gives the realistic picture of the existing spatial signal that the existing observation network is able to grasp. As there are no more gauging stations in the basin, based on only discharge data this is the best possible map that can be established. The estimation error was on average 15%, which is a theoretical value derived from the covariance structure and the geometry of the observation points. To this we should add a model uncertainty. We can get an idea of the order of magnitude of this error by comparing the grid maps produced by schemes IS1 and IS3. For further improvements supplementary data are needed, for instance long-term mean values of precipitation and actual evapotranspiration to close also the vertical water balance. The joint task to minimize the estimation errors in all three water balance components with preservation of water balance is developed further by Yan *et al.* (2012).

The resolution of grids here was subjectively chosen to be 20×20 km, and not 1×1 km as the original DEM, to avoid a false impression of precision. On average, the estimation error in the grid map is 15% and in extreme cases 30%. The choice of the grid resolution can be determined in a more objective way by optimizing the signal to noise ratio, as the larger the cells are the less of the spatial signal is captured but at the same time the estimation error is reduced. This is a topic for further investigations.

ACKNOWLEDGEMENTS

This research was supported by the National Basic Research Program of China (2010CB428406), National Key Water Project (No. 2009ZX07210-006), the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 51021006) and the Key Project of the Natural Science Foundation of China (No. 40730632). The constructive comments by the anonymous reviewer are highly appreciated and helped improve the manuscript.

REFERENCES

Beldring, S., Roald, L. A. & Voksø, A. 2002 Avrenningskart for Norge (Runoff map of Norway, in Norwegian). Norwegian Water and Energy Directorate Report 2: Oslo, Norway.

Bishop, G. D., Church, M. R. & Aber, J. D. 1998 A comparison of mapped estimates of longterm runoff in the northeast United States. *J. Hydrol.* **206**, 176–190.

China Ministry of Water Resources 2006 *Water Year Book in Huaihe River Basin.* Volumes 1–6, China Water Power Press, Beijing.

Cressie, N. 1990 The origins of kriging. *Math. Geol.* **22** (3), 239–254.

Creutin, J. D. & Obled, C. 1982 Objective analysis and mapping techniques for rainfall fields an objective comparison. *Water Resour. Res.* **18**, 413–431.

Dingman, S. L., Seely-Reynolds, D. M. & Reynolds, R. C. 1988 Application of kriging to estimating mean annual precipitation in a region of orographic influence. *Water Resour. Bull.* **24**, 329–339.

Gandin, L. S. 1963 *Objectivnij analis meteorologiskikh polej. (Objective analysis of meteorological fields)*. Gidrometerologicheskoe Izdatelstvo, Leningrad.

Gauch, H. G. 2003 *Scientific Method in Practise*. Cambridge University Press, Cambridge.

Ghosh, B. 1951 Random distances within a rectangle and between two rectangles. *Bull. Calcutta Math. Soc.* **43**, 17–24.

Goovaerts, P. 2000 Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **228** (1–2), 113–129.

Gottschalk, L. 1993a Correlation and covariance of runoff. *Stoch. Hydrol. Hydraul.* **7**, 85–101.

Gottschalk, L. 1993b Interpolation of runoff applying objective methods. *Stoch. Hydrol. Hydraul.* **7**, 269–281.

Gottschalk, L. & Krasovskaia, I. 1998 Development of grid-related estimates of hydrological variables. Report of the WCP-Water Project B.3, Geneva, WCP/WCA.

Gottschalk, L., Krasovskaia, I., Leblois, E. & Sauquet, E. 2006 Mapping mean and variance of runoff in a river basin. *Hydrol. Earth Syst. Sci.* **10**, 1–16.

Gottschalk, L., Leblois, E. & Skøien, J. 2011a Correlation and covariance of runoff revisited. *J. Hydrol.* **398**, 76–90.

Gottschalk, L., Leblois, E. & Skøien, J. 2011b Distance measures for hydrological data having a support. *J. Hydrol.* **402**, 415–421.

Huang, W. C. & Yang, F. T. 1998 Streamflow estimation using Kriging. *Water Resour. Res.* **34**, 1599–1608.

Jutman, T. 1995 *Runoff In: Climate, Lakes and Rivers: National Atlas of Sweden*. SNA Publishing, Stockholm, pp. 106–111.

Kolmogorov, A. N. 1941 Interpolation and extrapolation of stationary sequences. *Izvestia Akademii Nauk SSSR, Seria Matematich*. **5**, 3–14.

Krige, D. G. 1951 A statistical approach to some basic mine valuation problems of the Witwatersrand. *J. Chem. Metal. Min. Assoc. SA* **52**, 119–139.

Matérn, B. 1960 Spatial Variation. *Meddelanden från Statens Skogsforskningsinstitut* **49** (5), 144.

Matheron, G. 1963 Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266.

Merz, R. & Blöschl, G. 2005 Flood frequency regionalisation – spatial proximity vs. catchment attributes. *J. Hydrol.* **302**, 283–306.

Sauquet, E., Gottschalk, L. & Leblois, E. 2000 Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme. *Hydrol. Sci. J.* **45** (6), 799–815.

Skøien, J. O. & Blöschl, G. 2007 Spatiotemporal topological kriging of runoff time series. *Water Resour. Res.* **43**, 1–21.

Skøien, J. O., Merz, R. & Blöschl, G. 2006 Top-kriging – Geostatistics on stream networks. *Hydrol. Earth Syst. Sci.* **10**, 277–287.

Wiener, H. 1949 *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. MIT Press, Cambrideg, Mass.

Yan, Z., Gottschalk, L., Leblois, E. & Xia, J. 2012 Joint mapping of water balance components in a large Chinese basin. *J. Hydrol.* **450–451**, 59–69.

First received 29 July 2010; accepted in revised form 13 September 2011. Available online 6 June 2012