International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

# Sampling design optimization of a wireless sensor network for monitoring ecohydrological processes in the Babao River basin, China

Y. Ge[a], J.H. Wang[ab], G.B.M. Heuvelink[c], R. Jin[d], X. Li[d] & J.F. Wang[a]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[b] University of Chinese Academy of Sciences, Beijing, China

[c] Soil Geography and Landscape group, Wageningen University, Wageningen, Gelderland, The Netherlands

[d] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Science, Lanzhou, Gansu, China

Published online: 10 Sep 2014.

*International Journal of Geographical Information Science*, 2014
http://dx.doi.org/10.1080/13658816.2014.948446





# Sampling design optimization of a wireless sensor network for monitoring ecohydrological processes in the Babao River basin, China

Y. Ge[a]*, J.H. Wang[a,b], G.B.M. Heuvelink[c], R. Jin[d], X. Li[d] and J.F. Wang[a]

[a]*State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing, China;* [b]*University of Chinese Academy of Sciences, Beijing, China;* [c]*Soil Geography and Landscape group, Wageningen University, Wageningen, Gelderland, The Netherlands;* [d]*Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Science, Lanzhou, Gansu, China*

(*Received 4 June 2013; final version received 21 July 2014*)

Optimal selection of observation locations is an essential task in designing an effective ecohydrological process monitoring network, which provides information on ecohydrological variables by capturing their spatial variation and distribution. This article presents a geostatistical method for multivariate sampling design optimization, using a universal cokriging (UCK) model. The approach is illustrated by the design of a wireless sensor network (WSN) for monitoring three ecohydrological variables (land surface temperature, precipitation and soil moisture) in the Babao River basin of China. After removal of spatial trends in the target variables by multiple linear regression, variograms and cross-variograms of regression residuals are fit with the linear model of coregionalization. Using weighted mean UCK variance as the objective function, the optimal sampling design is obtained using a spatially simulated annealing algorithm. The results demonstrate that the UCK model-based sampling method can consider the relationship of target variables and environmental covariates, and spatial auto- and cross-correlation of regression residuals, to obtain the optimal design in geographic space and attribute space simultaneously. Compared with a sampling design without consideration of the multivariate (cross-)correlation and spatial trend, the proposed sampling method reduces prediction error variance. The optimized WSN design is efficient in capturing spatial variation of the target variables and for monitoring ecohydrological processes in the Babao River basin.

**Keywords:** universal cokriging; linear model of coregionalization; spatial simulated annealing; optimization

## 1. Introduction

Ecological and hydrological monitoring are essential for understanding basin-scale ecohydrological processes, obtaining baseline data for assessment of impacts and conservation of regional ecosystems, and for the development and validation of ecohydrological models (Brierley *et al.* 2010). In practice, there are two ways for monitoring basin ecohydrological parameters (Li *et al.* 2013). One approach uses remote sensing data products while another uses traditional fixed-point observation stations. Remote sensing data have advantages in that they facilitate a relatively cheap and rapid means of acquiring up-to-date information over a large geographic area. However, validation studies have

*Corresponding author. Email: gey@lreis.ac.cn

shown that these products may be biased (Turner *et al.* 2006). Traditional fixed-point observation stations such as hydrological and meteorological stations and flux towers yield more accurate observations. However, their high cost, high risk and intensive labour needs limit high-density deployment in a large river basin (Zhang *et al.* 2012). Therefore, a new observation approach is required to achieve intensive data observation, to compensate for sparse traditional ground-based stations and to meet the requirements of basin-scale ecohydrological monitoring.

The emergence of wireless sensor network (WSN) technology has brought new opportunities for data-intensive observation in ecohydrological research and environmental monitoring (Hart and Martinez 2006). For watershed monitoring, an ecohydrological WSN integrates a variety of environmental observations, data processing, wireless communication and energy supply into a sensor node. This forms the WSN with a self-organizing approach, to achieve real-time sensing, monitoring, collecting and processing of information on various hydrological, ecological, and meteorological variables (Zhang *et al.* 2012). It represents an intensive low-cost observation technique, which enhances basin ecohydrological monitoring capabilities.

Prior to deploying the ecohydrological WSN, development of efficient procedures for designing information-effective monitoring networks is an essential task for accurate understanding of the spatial distribution or variation of key ecohydrological variables. Therefore, data observed by an optimized WSN should provide sufficient but not redundant information. There are currently a number of spatial sampling optimization methods for environmental or natural resource monitoring (de Gruijter *et al.* 2006). Most fall into one of the following categories: geometry-based (Royle and Nychka 1998), probability-based (Cochran 1977), and model-based sampling (Brus and de Gruijter 1997). For example, Walvoort *et al.* (2010) implemented a geometry-based spatial coverage sampling (SCS) method that takes the mean-squared shortest distance as a minimization criterion to ensure that samples cover the study area as uniformly as possible. Lesch (2005) proposed a model-based site selection algorithm that begins from the sample that is closest to the optimum response surface design calculated from experimental design theory. Among the sampling methods, the model-based ones are generally more efficient, especially when the monitoring target has significant spatial structure (Brus and de Gruijter 1997, de Gruijter *et al.* 2006). In most cases, ecohydrological variables are spatially autocorrelated; therefore, geostatistical model-based approaches are used to optimize WSN sampling locations.

Geostatistical model-based sampling commonly chooses optimal samples for which the mean kriging prediction error variance (MKV) is minimized (Brus and Heuvelink 2007). For example, van Groenigen *et al.* (1999) adopted MKV as an optimization criterion to derive an optimized sample design using spatial simulated annealing (SSA). Heuvelink *et al.* (2006) applied the simulation method to study the variation of parameters in variograms and impacts of universal kriging trend variation on optimized sampling distributions. Furthermore, Brus and Heuvelink (2007) developed a universal kriging model-based sampling method that optimizes the spatial configuration of observations, by minimizing the spatially averaged universal kriging variance with the aid of exhaustively known covariates. More recently, Melles *et al.* (2011) extended geostatistical model-based sampling optimization and applied it to produce multi-objective optimizations to monitor background values of radiation levels and conduct emergency monitoring tasks. All the above approaches focused only on one monitoring variable and its spatial distribution. In practice, however, a watershed ecohydrological WSN usually aims to monitor spatial distributions of multiple ecohydrological variables, such as soil temperature, soil moisture and precipitation. Spatial structures and relationships between these variables may vary from case to case. To address this issue, Vašát *et al.* (2010) extended univariate sampling design optimization to a multivariate sampling design and applied it to a case study from soil science. An optimal sampling design for simultaneous prediction of multiple soil properties was produced, through defining the linear model of coregionalization (LMC) (Goovaerts 1997). Besides, Yeh *et al.* (2006) developed a design approach for an optimal multivariate geostatistical groundwater quality network by proposing a network system to identify groundwater quality spatial variation using factorial kriging with a genetic algorithm. However, the above two works did not consider potential spatial trends of multiple variables within a spatially heterogeneous area.

The universal cokriging (UCK) (Stein and Corsten 1991, Stein *et al.* 1991) model is capable of simultaneously treating multiple target variables with spatial trends and their cross-correlations and has been successfully applied in spatial data interpolation (e.g. de Carvalho *et al.* 2010, Kuhlman and Igúzquiza 2010, Biggs and Atkinson 2011, Li *et al.* 2012). Little to no research addresses spatial sampling optimization with UCK, especially when target variables are cross-correlated and have spatial trends. Therefore, the present study developed a multivariate geostatistical sampling design model to optimize WSNs using UCK with SSA, for identifying the spatial distribution of key ecohydrological variables. The proposed sampling design can optimally design a WSN monitoring network that considers not only multiple target variables but also their spatial variation. The developed model has been applied in a real ecohydrological WSN at a high-altitude alpine environment in the Babao River basin of China.

The remainder of this paper is organized as follows. Section 2 describes the study area and presents the data set used as a basis for the optimization approaches. The methodology for sampling design optimization with UCK is given in Section 3. Results for mapping multiple ecohydrological variables are also presented. The sampling optimization method is discussed in Section 4. Several conclusions are drawn and possible future research is discussed in Section 5. All analyses, simulation and evaluation were implemented using a set of functions written in the R language for statistical computing (R Core Team 2012).

## 2. Study area and data description

### *2.1. Study area*

The study area in the Babao River basin is upstream of the Heihe River. The Heihe is a typical inland river basin with arid climate in Northwest China. A vulnerable ecosystem and scarcity of water resources have significantly restricted social and economic development in this area (Li *et al.* 2009). The Babao basin has an area of 2450 km$^2$ with a latitude range 37°43’N–38°20’N and longitude 100°05’E–101°09’E. Elevation in the area is from 2339 m to 4947 m, and annual average temperature is −1°C. The local climate may be characterized as semiarid and alpine cold, with 300–500 mm annual precipitation associated with the southwest monsoon. Both temperature and precipitation vary significantly with elevation, because of the steep topography and large elevation differences (Qi and Luo 2007). Vegetation in the basin is predominately mountain forest and grassland, with extensive coverage by shrub meadow. There are various landscape belts with clear boundary lines in the vertical direction, and there is a glacier and snow-covered zone above 4500 m. At lower elevations, there are mountain meadows and bushy vegetation, intermontane basins, middle mountain forests and meadows. The vegetation has a strong influence on stream formation, water regulation and storage.

Owing to climate, topography, soil conditions and vegetation cover, ecohydrological interactions in the Babao basin have remarkable spatial heterogeneity (Chen and Qu 1992). On a regional scale, climate is the most important influence on vegetation and hydrology interactions. On the landscape scale, terrain and topography affect microclimate and soil development. These factors control the redistribution of substances (water and organic matter), especially the spatial distribution of water in soil, which directly influences growth differences in vegetation. Upper reaches of the Babao River lie in the Qilian Mountain area, where forest ecosystem and hydrology processes mainly depend on topography. Significant spatial variations, indicated by parameters such as thermal radiation, precipitation, temperature and soil water, are also caused by topographic variations. Thus, spatial heterogeneities of various topography-related environmental factors should be incorporated to properly describe the ecohydrological process and interactions in the Babao basin. Temperature, precipitation and soil moisture have significant spatial variation across the basin. There are eight automatic weather stations (AWSs) in the basin that monitor weather parameters like temperature and precipitation. These stations will be modified by adding devices for soil moisture measurement and merged with the WSN to reduce building costs.

### 2.2. *Auxiliary data description*

To assist the ecohydrological WSN design, basic geographic and environmental auxiliary data were collected in the Babao River basin. The basic geographic data describe natural geographic conditions in the basin, including basin boundary, river and road networks, existing AWSs (Figure 1) and topography (digital elevation model (DEM), slope and aspect; Figure 2). Environmental auxiliary data were mainly used to provide *a priori* information about key ecohydrological variables. As no fine-resolution ground observations were available before WSN deployment, remote sensing retrieval products and model output data were acquired as prior knowledge of annual average land surface temperature (LST), precipitation and soil moisture (Figure 3).

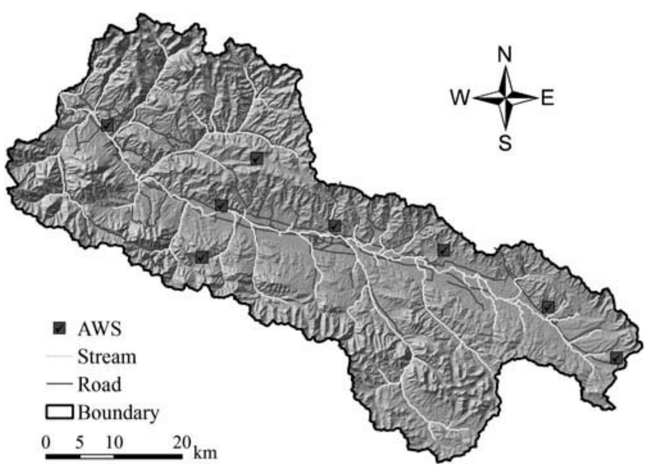


Figure 1. Basic geographic data of the Babao River basin.

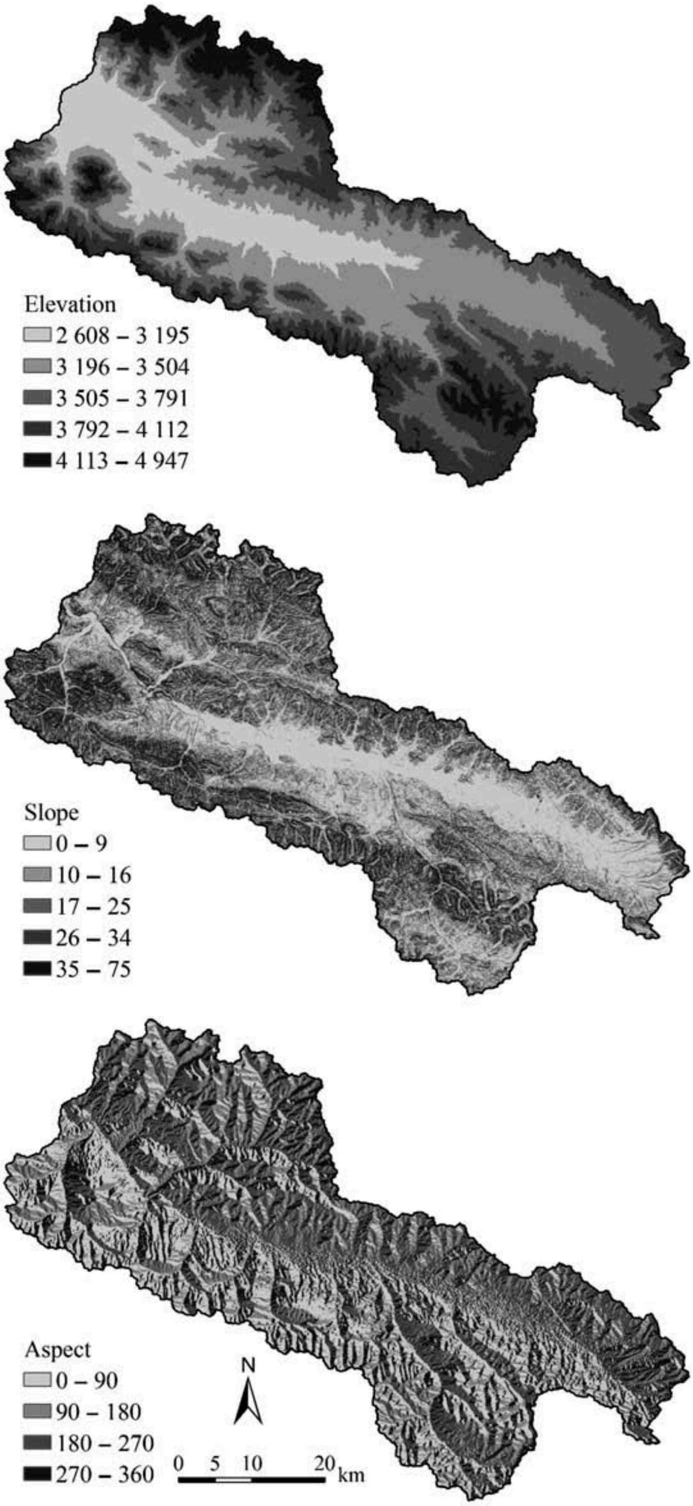


Figure 2. Topographic data of the study area. The upper panel shows elevation, the middle the slope and the lower the aspect.

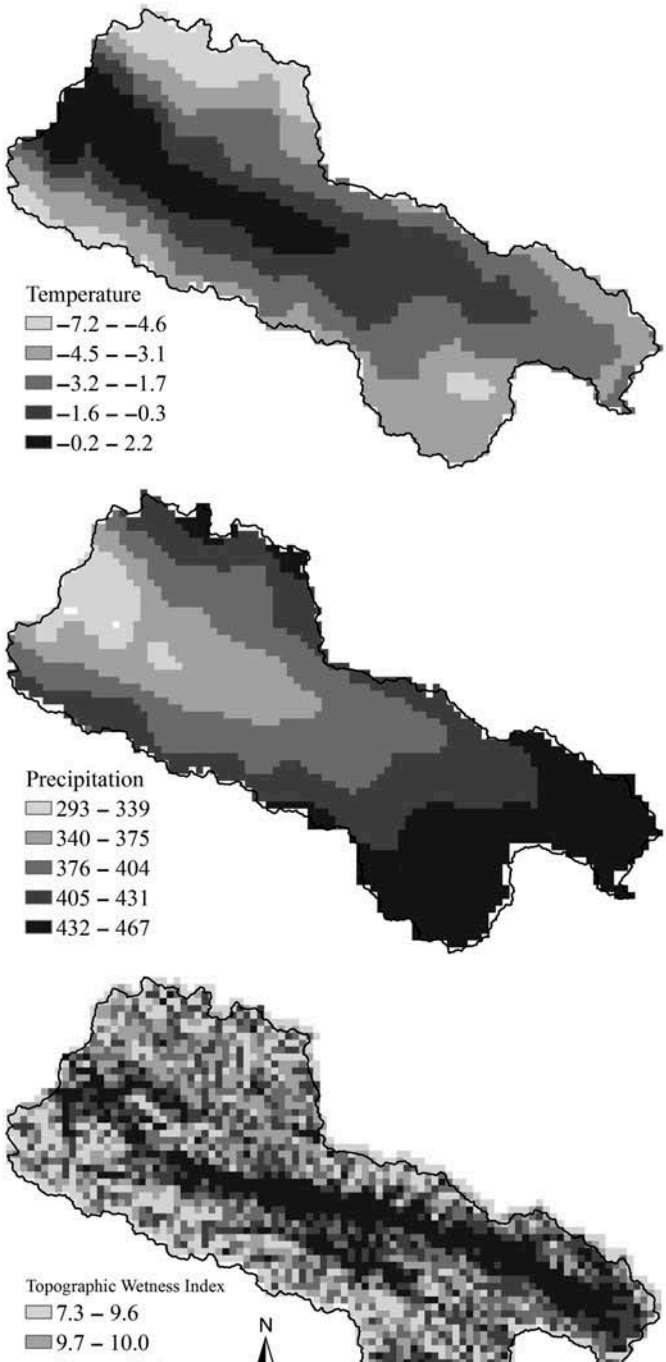


Figure 3. Environmental auxiliary variables. The upper is the annual average MODIS land surface temperature product; middle is the annual mean precipitation product and the lower is the topographic wetness index calculated from DEM data to represent soil moisture.

(1) Land surface temperature. Annual average LST was derived from moderate-resolution imaging spectroradiometer (MODIS) LST monthly average data retrieved from the NASA Level 1 and Atmosphere Archive and Distribution System (LAADS) FTP site. The data have a spatial resolution of 1 km (Wan *et al.* 2004).
(2) Precipitation. Precipitation data were obtained from *WorldClim.org*. Global precipitation data at this website were obtained by thin-plate spline interpolation, based on over 15,000 weather stations worldwide (Hijmans *et al.* 2005). Spatial resolution of these data is 1 km, which was used as prior information for precipitation.
(3) Soil moisture. The current spatial resolution of soil moisture remote sensing products is around 0.25°. Because only several pixels cover the Babao River basin, data from these remote sensing products are too coarse for the WSN sampling design. Alternative data from a DEM, called topographic wetness index (TWI) (Sørensen *et al.* 2006), were used as indirect information to describe the soil moisture spatial distribution. Spatial resolution of these data is 1 km.

In the case of no prior observations before designing a new monitoring network, the environmental auxiliary data derived from remotely sensed imagery or from model output were used as proxy data to represent the spatial distribution of ecohydrological variables. These environmental auxiliary data may be biased and have artefacts such as uncertainties from the retrieval algorithm and scaling mismatches, which may affect the quality of the sampling design. Consequently, the auxiliary data were only used to calibrate the UCK model prior to optimize the network design, while these proxies were not used for prediction after WSN deployment.

## 3. Sampling optimization design with UCK

This section presents the sampling design optimization method based on the UCK model. Steps in this method are described graphically in Figure 4. Input target variables include LST, precipitation and soil moisture, and input environmental auxiliary covariates include elevation, slope, aspect, longitude and latitude. The relationship between target variables and environmental covariates was established by linear regression models. Next, the LMC of the regression residuals was modelled and UCK was used to calculate UCK prediction

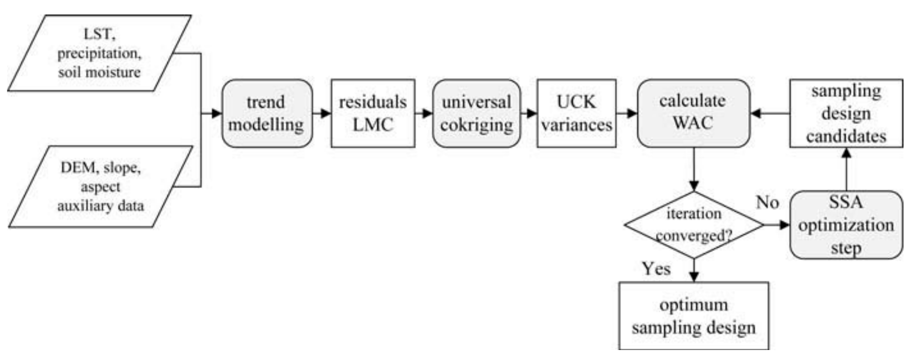


Figure 4. The procedure of the sampling optimization design with UCK.

error variances. Afterwards, the weighted-average criterion (WAC) value of the objective function and SSA optimization algorithm were used to search for the optimal sampling design. Finally, the optimum sampling configuration was obtained. All steps are described in more detail in Figure 4.



### *3.1. Spatial trend modeling*

To model a spatial trend in the target variables, their relationship with environmental covariates was established using multiple linear regression. As shown in Figure 3, the target variables had spatial trends. Correlation analysis showed that the environmental covariate elevation had a strong linear relationship with LST (correlation coefficient $r = -0.84$), precipitation ($r = 0.65$) and soil moisture ($r = -0.51$). Therefore, these environmental covariates can serve as explanatory variables to model a spatial trend in the target variables. Those variables were also correlated with each other. There was a strong linear relationship between LST and precipitation ($r = -0.71$). There was a weak linear correlation between LST and soil moisture ($r = 0.34$) and between precipitation and soil moisture ($r = 0.13$).

Stepwise regression coupled with the Akaike information criterion was used to select the regression model for each target variable. The results of trend modeling are shown in Table 1, in which temperature has regression relationships with elevation, slope, aspect, longitude and latitude. Precipitation has regression relationships with elevation, aspect, longitude and latitude. Soil moisture only has regression relationships with elevation, slope and aspect. These regression models were used for trend modeling in UCK.

### *3.2. Universal cokriging*

In this section, we use largely the same notation as in Pebesma (2004), in which vectors and matrices are italicized and where composite vectors and matrices as used in UCK are written in bold. Each target variable in UCK is considered the sum of a trend $X\beta$ and a residual $\varepsilon(s)$ (Christensen 1990, Pebesma 2004):

$$Z(s) = \sum_{j=0}^{m} X_j(s)\beta_j + \varepsilon(s) = X\beta + \varepsilon(s), \tag{1}$$

Table 1. Trend variables and parameters of the LMC fitted to experimental variogram.

| Variable | Trend variables | Model | Nugget | Sill | Nugget-to-sill ratio (%) | Range (m) |
|---|---|---|---|---|---|---|
| LST | $x_0, x_1, x_2, x_3, x_4, x_5$ | Spherical | 0.2495 | 0.8561 | 29.14 | 18,000 |
| Precipitation | $x_0, x_1, x_3, x_4, x_5$ | Spherical | 48.350 | 305.06 | 15.85 | 18,000 |
| Soil moisture | $x_0, x_1, x_2, x_3$ | Spherical | 0.2620 | 0.3820 | 68.59 | 18,000 |
| LST versus precipitation | — | Spherical | 3.3814 | 12.807 | 26.40 | 18,000 |
| LST versus soil moisture | — | Spherical | 0.0523 | 0.0894 | 58.50 | 18,000 |
| Precipitation versus soil moisture | — | Spherical | 0.8500 | 2.2396 | 37.95 | 18,000 |

Note: $x_0$ = intercept, $x_1$ = elevation, $x_2$ = slope, $x_3$ = aspect, $x_4$ = longtitude, $x_5$ = latitude.

where $Z(s)$ is a vector of length $n$, with observations $Z(s_1), \ldots, Z(s_n)$ taken at spatial locations $s_i$. In this study, $Z(s)$ could be LST, precipitation or soil moisture. The $X_j(s)$ represents environmental covariates (e.g. elevation, slope and aspect) at the $n$ locations. $X_0(s) = 1$ for all $s$, implying that $\beta_0$ represents the intercept. $X$ is an $n \times (m + 1)$ matrix of the covariates at all observations locations. $\beta$ is a vector with unknown regression coefficients and $\varepsilon(s)$ is a normally distributed residual with zero mean and constant variance $c(0)$. For spatial data, residuals are usually spatially autocorrelated, as quantified through a variance–covariance matrix $C$ or variogram. The best linear unbiased prediction (universal kriging) of $Z(s_0)$ at an unobserved location $s_0$ from $n$ observations $Z(s_i)$ is:

$$\hat{Z}(s_0) = x(s_0)\hat{\beta} + c'C^{-1}\left(Z(s) - X\hat{\beta}\right), \tag{2}$$

where $x(s_0)$ is the row of covariates at $s_0$, $\hat{\beta} = VX'C^{-1}Z(s)$ (let $V = (X'C^{-1}X)^{-1}$) is the generalized least-squares (GLS) estimate of the trend coefficients, $C$ is the $n \times n$ variance–covariance matrix of the $n$ residuals and $c$ is the vector of covariances between the residuals at the observation and prediction locations. $C$ and $c$ are derived from the variogram of $\varepsilon$. The corresponding prediction error variance (universal kriging variance) at $s_0$ is

$$\sigma^2(s_0) = c(0) - c'C^{-1}c + x_a{}'Vx_a, \tag{3}$$

where $x_a = x(s_0) - X'C^{-1}c$.

Now, we extend the univariate model to a multivariable one (Pebesma 2004), which involves joint modeling of spatially and cross-correlated multiple variables. For $p$ distinct variables, let $\{Z_i(s), X_i, \beta_i, \varepsilon_i, x_i(s_0), c_i, C_i\}$correspond to $\{Z(s), X, \beta, \varepsilon, x(s_0), c, C\}$ of the $i$-th variable and let $\mathbf{Z}(s) = (Z_1(s)', \ldots, Z_p(s)')'$, $\boldsymbol{\beta}(s) = (\beta^{1\prime}, \ldots, \beta^{p\prime})'$, $\boldsymbol{\varepsilon}(s) = (\varepsilon_1(s)', \ldots, \varepsilon_p(s)')'$,

$$\mathbf{X} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_p \end{bmatrix}, \quad \mathbf{x}(s_0) = \begin{bmatrix} x_1(s_0) & 0 & \ldots & 0 \\ 0 & x_2(s_0) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & x_p(s_0) \end{bmatrix},$$

$$\mathbf{c} = \begin{bmatrix} c_{1,1} & c_{1,2} & \ldots & c_{1,p} \\ c_{2,1} & c_{2,2} & \ldots & c_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p,1} & c_{p,2} & \ldots & c_{p,p} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} C_{1,1} & C_{1,2} & \ldots & C_{1,p} \\ C_{2,1} & C_{2,2} & \ldots & C_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{p,1} & C_{p,2} & \ldots & C_{p,p} \end{bmatrix},$$

where element $i$ of the $n$-vector $c_{k,l}$ is $\mathrm{Cov}(\varepsilon_k(s_i), \varepsilon_l(s_0))$, which characterizes the covariances between residuals at observation and prediction locations. Likewise, item $(i, j)$ of each $n \times n$ matrix $C_{k,l}$ refers to $\mathrm{Cov}(\varepsilon_k(s_i), \varepsilon_l(s_j))$. $\mathbf{C}$ and $\mathbf{c}$ are derived from the LMC, which is explained in Section 3.3.

The multivariate best linear unbiased prediction (UCK) at unobserved location $s_0$ is given by (Ver Hoef and Cressie 1993):

$$\hat{\mathbf{Z}}(s_0) = \mathbf{x}(s_0)\hat{\boldsymbol{\beta}} + \mathbf{c}'\mathbf{C}^{-1}\left(\mathbf{Z}(s) - \mathbf{X}\hat{\boldsymbol{\beta}}\right), \tag{4}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{V}\mathbf{X}'\mathbf{C}^{-1}\mathbf{Z}(\mathbf{s}), \mathbf{V} = \left(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\right)^{-1},$ the GLS estimator of $\boldsymbol{\beta}$. The variance of the prediction error is

$$\sigma^2(s_0) = \mathbf{c}(0) - \mathbf{c}'\mathbf{C}^{-1}\mathbf{c} + \mathbf{x}_a{}'\mathbf{V}\mathbf{x}_a, \tag{5}$$

where $\mathbf{x}_a = \mathbf{x}(s_0) - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c}$. Equation (5) is now a prediction error covariance matrix, and it can be interpreted as follows (Stein and Corsten 1991). $\mathbf{c}(0)$ is the variance–covariance matrix of the target variables under study. $\mathbf{c}'\mathbf{C}^{-1}\mathbf{c}$ is the reduction of that matrix from the best linear approximation by the observations and $\mathbf{x}_a{}'\mathbf{V}\mathbf{x}_a = \mathrm{var}\left(\mathbf{x}_a{}'\hat{\boldsymbol{\beta}}\right)$ is the estimation error variance–covariance matrix of the trend.

There are two ways to estimate trend coefficients and the variogram of the residuals. One makes use of an iterative estimation method, the other makes estimates based on a maximum likelihood-based method such as restricted maximum likelihood (Schabenberger and Gotway 2005, Lark *et al.* 2006). We adopted the first approach, because it is easy to implement and has been widely used in kriging estimation, e.g. Hengl *et al.* (2004, 2007). In this method, the trend model is first estimated using ordinary least squares, and the variance–covariance function of the regression residuals is used to calculate the GLS coefficients. Next, these are used to re-compute the regression residuals, from which an updated covariance function is obtained. This process iterates several times to obtain the final regression coefficients and covariance function of the residuals. The estimation of regression coefficients was done separately for each dependent variable, each time adjusting the residuals and recalculating the GLS estimates. Upon convergence, this yields the GLS regression coefficient estimates for all variables.

### 3.3. *Linear model of coregionalization*

To obtain the residual variance–covariance matrix in Equations (4) and (5), the LMC (Goovaerts 1997) was adopted to model residual spatial dependence of all environmental covariates. With LMC, all (cross-)variograms were then fitted with an appropriate variogram model as shown in Figure 5, following the procedure of Pebesma (2004). The criteria of the sum of squared residuals return a spherical model with range 18 km with the best fit. The other variogram parameters, i.e. the nugget, sill and range, vary among the target variables. Fitted, LMC parameters are listed in Table 1. Figure 5 shows that the residuals of LST and precipitation were strongly autocorrelated in space. Soil moisture had a large nugget variance, resulting in a high nugget-to-sill ratio of 68.6%, indicating weak spatial autocorrelation for soil moisture.

### 3.4. *Optimization criteria*

To optimize a monitoring network, one must first select an appropriate criterion with which to quantify the suitability of a given design. As stated in Section 1, the aim of this research was to design a WSN for mapping spatially distributed ecohydrological variables in the Babao basin. Towards this purpose, the optimization criterion has been commonly defined as MKV minimization (van Groenigen *et al.* 1999, Brus and Heuvelink 2007). Equation (5) shows that the prediction error variance $\sigma^2(s_0)$ depends only on the arrangement of sampling locations $s_1, s_1, \ldots, s_n$ via variance–covariance matrices $\mathbf{C}$ and $\mathbf{c}$, covariance vector $\mathbf{c}(0)$ and the spatially exhaustively known environmental covariates. Target variable observations are not used in the prediction error variance calculation. This

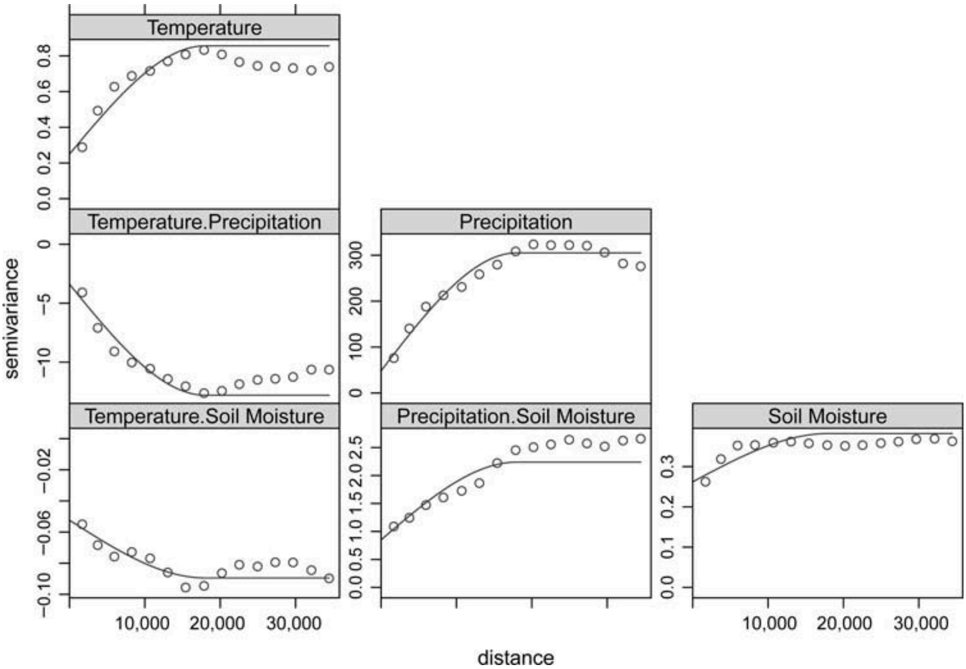


Figure 5. Direct variograms (diagonal) and cross-variograms (off-diagonal) for LST, precipitation and soil moisture variables fitted with LMC (solid line).

desirable property allows estimation of the error variance prior to collecting observations, and it was used in the WSN sampling optimization described below.

For ecohydrological WSN, one needs to optimize the variables LST, precipitation and soil moisture simultaneously. The optimization criterion can be defined as the weighted average of all three MKV values (Vašát *et al.* 2010). Because these three target variables have different measurement units, their MKVs were standardized through division by the variogram sill value. The weight denoted by $w_i$ ($i = 1,2,\ldots,n$) in Equation (6) is that given to each variable. Hence, the WAC value of the objective function is

$$\mathrm{WAC} = \frac{C_1 w_1}{S_1} + \frac{C_2 w_2}{S_2} + \frac{C_3 w_3}{S_3}, \tag{6}$$

where $C_1$, $C_2$ and $C_3$ are the MKV values of the three variables and $S_1$, $S_2$ and $S_3$ are the corresponding variogram sill values. We did not use a local search neighbourhood and calculated the MKV values using global cokriging.

### 3.5. Spatial simulated annealing

Given the UCK model, the objective is to optimize WSN node locations such that the smallest WAC value is obtained. To accelerate the optimization, the SSA optimization technique has been frequently used (van Groenigen and Stein 1998, van Groenigen *et al.* 1999, Zhu and Stein 2006, Brus and Heuvelink 2007, Baume et *al.* 2011, Melles *et al.* 2011). SSA is the spatial extension of simulated annealing (Kirkpatrick *et al.* 1983) for solving spatial optimization problems. This iterative, combinatorial algorithm has five main steps (Heuvelink et *al.* 2010, Baume *et al.* 2011): (1) Begin with an arbitrary initial sampling design and compute the associated criterion value; (2) generate a candidate new design from the current design, by random perturbation of the location of one observation location in a random direction (maximum shift is gradually decreased as the iteration continues); (3) evaluate the new candidate design criterion; (4) accept the new design if the criterion has improved or accept it with some probability (this ensures that the algorithm can escape from local optima) if the criterion has deteriorated; (5) stop after a given (large) number of iterations, or when new candidate designs have not been accepted for a given number of times.

Several SSA parameters must be defined before optimization. The initial probability of accepting worsening designs was set to 0.2 and decreased exponentially as the iteration progressed. The SSA procedure was halted when there was no improvement in the criterion over 200 iterations, or if their total number reached 10,000. After defining the parameters, SSA was used to perform a global search for WAC optimization defined by Equation (6). At each iteration, new criterion values were calculated under a new WSN distribution. Then, an optimal WSN distribution was determined by comparing values obtained from the new WSN distribution and the old one. Given a fixed sample size, the outcome of the objective function converges to a general optimization under certain conditions, as shown in Figure 6.

### 3.6. *Final WSN sampling design*

The following two problems should be considered during WSN sampling design optimization: (1) sample size of WSN nodes and (2) available locations for WSN node installation. To determine an appropriate sample size, the relationship between sample size and corresponding minimum criterion value was investigated. Figure 7 shows minimum WAC values for various sample sizes. It was found that the WAC value decreased with sample size. When the sample size reached 50, the increase of WSN nodes did not substantially reduce the error variance of prediction. Therefore, based on a trade-off between WSN node costs, accuracy requirements and budget constraints, the total sample size was set to 50, for which 42 nodes were new and eight were existing AWSs.

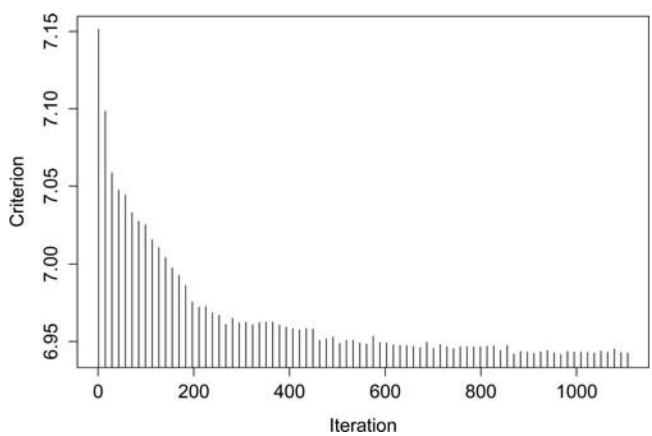


Figure 6. The trend of objective function with increasing SSA iterations.

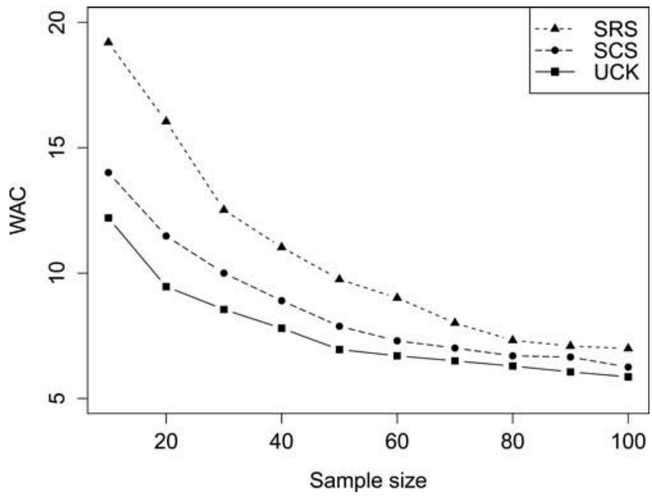


Figure 7. Trade-off between objective function WAC and sample size for SRS, SCS and UCK model-based sampling.

Regarding site availability, there are very few roads in the Babao River basin. In addition, alpine weather and mountains make it difficult to find appropriate sites for WSN nodes. Thus, new nodes were restricted to deployment within a 10-km buffer around roads, to facilitate construction and commissioning. After considering these restrictions, 42 new nodes were optimized with SSA. As expected, the criterion value decreased with the number of SSA iterations (Figure 6). After 1100 searches, the WAC produced a value of 6.95. The optimized WSN location was then obtained (Figure 8). In this figure, black dots represent the 42 new node locations, which cover the entire study area except for the grey exclusion area. Furthermore, to ensure wireless communication, the field WSN must

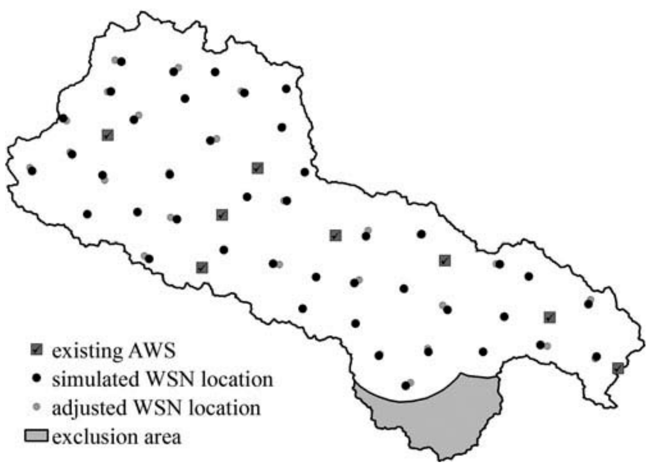


Figure 8. Final optimized WSN sampling design for multiple ecohydrological variables. Black squares represent eight existing stations, black dots represent the added WSN nodes location, grey dots represent the alternative location because of field limitation and grey area is the exclusion area by considering traffic accessibility.

be installed in open and relatively flat locations, with stable wireless signal coverage. For these reasons, final WSN installation positions were determined after field investigation. Twenty WSN points were moved less than 2 km away from planned positions; these are shown as grey dots in Figure 8. The WAC value changed from 6.95 to 6.97 after WSN adjustment. This indicates that the fine adjustment of WSN positions had little effect on the WSN sampling design optimization.

### *3.7. Sampling design comparison*

For the purpose of comparison, two commonly used spatial sampling designs, i.e. simple random sampling (SRS) (Cochran 1977) and SCS (Royle and Nychka 1998), were adopted for WSN design in the Babao River basin. In the SRS design, WSN nodes are selected randomly. This is easy to implement in practice, but may lead to uneven coverage in the study area. The SCS sampling design achieves even coverage of the study area and was implemented in the package *spcosa* (Walvoort *et al.* 2010), available at the Comprehensive R Archive Network. Subsequently, WAC values in three sampling designs were calculated with the same LMC variograms as inputs. As shown in Figure 7, the results indicate that the proposed method performed better than SRS and SCS.

In real-world applications, it is necessary for the practitioner to consider implementation costs and required accuracy before choosing an appropriate sampling design. SRS and SCS sampling have advantages, in that they do not need any auxiliary information before sampling design (Wang *et al.* 2012). They are easily applied in practice, although their sampling designs are potentially inefficient. The UCK model-based sampling potentially reduces target estimation variance, because it considers the covariance structure of the residuals in sampling design optimization. When prior data are available, such design is recommended as an efficient approach to optimize the multivariate monitoring network.

## 4. Discussion

We addressed a sampling optimization method based on a UCK model for multiple variables with non-constant trends and applied the method to WSN sampling design optimization for monitoring ecohydrological processes in the Babao basin. Through the case study above, results of this approach can be outlined as follows.

### *4.1. Multivariate optimization*

The optimization criterion combined weighted-average MKV values of LST, precipitation and soil moisture residuals of cokriging error variance. As a result, the UCK model achieved comprehensive optimization of all variables, as well as comparative optimization of each target variable. Cross-variations caused by correlation between variables were added to the UCK variance–covariance matrix calculation. Consequently, methods for sampling design optimization based on the UCK model fully considered correlations between target variables, which are appropriate for joint sampling and sampling design optimization for multiple variables. It should be noted that in this particular case, where all variables are sampled at the same locations, the extra effort in using cokriging instead of kriging will not lead to a dramatic increase in interpolation accuracy. However, the optimization of the sampling designs could not be done for each variable separately because at each monitoring station we collect all variables. This given, it seems most appropriate to use a cokriging approach for multivariate optimization (Vašát et *al.* 2010), also because the extra work involved with cokriging is relatively modest.

### *4.2. Optimizing both geographic space and attribute space for target variables*

The UCK variance incorporates both the prediction error variance of the residual (first two terms on right side of Equation (5)) and estimation error variance of the trend (third term on the right side of Equation (5)). To minimize the first component, sample points should cover as much as possible the geographic space; this explains the uniform distribution of sampling points shown in Figure 8. To minimize the second component, sample points should have a large spread in attribute space so that estimates of the linear regression coefficients are more accurate, which means optimization in attribute space (Brus and Heuvelink 2007). The attribute density distribution of the variables, including LST, DEM, slope and aspect, is presented in Figure 9. Here, grey and black lines represent the population (variables in Figures 2 and 3) and sample attribute density histograms, respectively. As shown in Figure 9, the target variables (LST) and environmental covariates (DEM, slope and aspect) have similar curve shapes, revealing that substantial attribute representativeness was achieved after sampling design optimization.

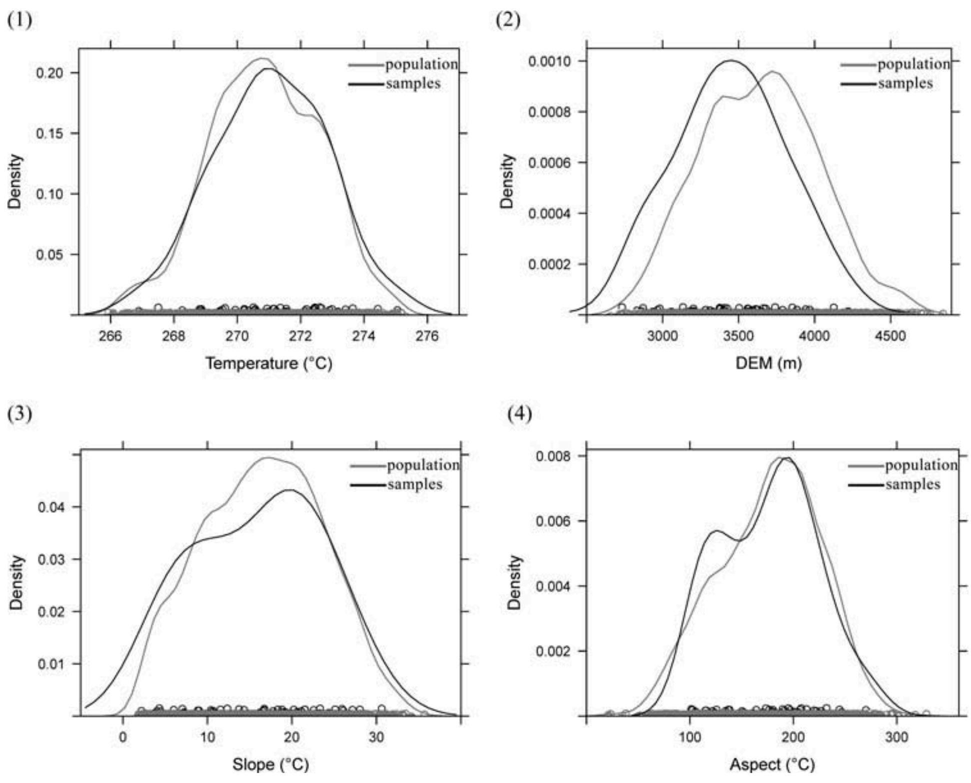


Figure 9. Comparison of optimized sample attributes density distribution and corresponding population density distribution. (1) LST, (2) DEM, (3) slope and (4) aspect.

### 4.3. Different spatial variation in multiple variables

Spatial variations of LST, precipitation and soil moisture varied across the Babao River basin. Variations of LST and precipitation were similar. These were continuous in space, and their annual averages varied with elevation. Consequently, nugget values of the corresponding variogram were small, and sampling points were uniformly distributed in space. Soil moisture was affected by a variety of factors, such as precipitation, soil texture, vegetation, evaporation, terrain and topography. As a result, it was discontinuous in space and had weak spatial autocorrelation. This resulted in a large nugget value and nugget-to-sill ratio of the corresponding variogram (Figure 5). For a weak spatial autocorrelation target variable optimization, the sampling points tend to be distributed randomly in space (Heuvelink *et al.* 2006). In this study, soil moisture contributed up to 52.2% of the WAC value in the objective function, whereas percentages contributed by LST and precipitation were 28.3% and 19.5%, respectively. Therefore, the spatial variation of soil moisture significantly affected the spatial distribution of sampling points. Furthermore, for different spatial variations of LST, precipitation and soil moisture, it is important that the final WSN sampling design is optimal for the full design and sub-optimal for the separate target variables.

### 4.4. Issues of auxiliary data

Since there was no sampling in the study area prior to designing the sampling scheme, remote sensing products and model data were used to provide primary information on the targets such as variograms for LMC. Since the auxiliary maps (Figure 3) show averages for a given time, whereas the three target variables vary in space and time, those maps cannot be used to depict the target variables and were only used to calibrate the UCK model. We assumed that these auxiliary maps could represent the spatial correlation structure and relationships with the covariates. However, two issues need to be addressed, as follows.

(a) Scale issue: spatial resolutions of auxiliary maps of LST, precipitation and soil moisture are 1 km; therefore, the maps represent spatial patterns of target variables at 1-km scale. This scale may not be the same as that of spatial patterns characterized by WSN samples on target variables.
(b) Uncertainty issue: the data may have uncertainties stemming from remote sensing data and the model. For example, TWI, which was used to represent soil moisture, is affected not only by terrain but also by other factors such as meteorological conditions and land cover. Therefore, TWI will directly affect LMC fitting and sampling optimization results. The trend elimination process and residual error of LMC fitting also introduce uncertainties in the optimization process, which affects the accuracy and reliability of WSN sampling design.

## 5. Conclusions

Using WSN to monitor ecohydrological processes at watershed scale is an efficient method. This study developed and implemented a UCK model-based sampling design approach for optimal selection of observation locations, towards monitoring multivariate ecohydrological processes such as LST, precipitation and soil moisture in the Babao River basin of China. The method begins with multivariate regression trend modeling, to remove spatial trends associated with the target variables. Then, on the assumption that the regression residuals are stationary, variograms and cross-variograms of residuals are calculated and fitted with the LMC. By defining an objective function for the multiple ecohydrological variables, the optimal sampling design for WSN can be achieved using a spatially simulated annealing algorithm. Results demonstrate that this method can optimize the distribution of multiple variables in geographic space and attribute space simultaneously, with respect to spatial autocorrelation and cross-correlation of target variables. Spatial heterogeneity characteristics of key ecohydrological variables can be effectively obtained through optimized WSN observations, which can provide high accurate data for monitoring and simulating the ecohydrological processes.

## Acknowledgments

The authors are grateful to the anonymous referees for their constructive comments, which helped to improve the quality of the article.

## Funding

This work was supported by the 863 High Technology of China [2012AA12A305].

## References

Baume, O.P., *et al.*, 2011. Network optimization algorithms and scenarios in the context of automatic mapping. *Computers & Geosciences*, 37, 289–294. doi:10.1016/j.cageo.2010.04.014.

Biggs, E.M. and Atkinson, P.M., 2011. A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK. *Hydrological Processes*, 25, 795–810. doi:10.1002/hyp.7869.

Brierley, G., *et al.*, 2010. What are we monitoring and why? Using geomorphic principles to frame eco-hydrological assessments of river condition. *Science of the Total Environment*, 408, 2025–2033. doi:10.1016/j.scitotenv.2010.01.038.

Brus, D.J. and de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80, 1–44. doi:10.1016/S0016-7061(97)00072-4.

Brus, D.J. and Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138, 86–95. doi:10.1016/j.geoderma.2006.10.016.

Chen, L.H. and Qu, Y.G., 1992. *Rational development and utilization on water and soil resources in Hexi region*. Beijing: Science Press.

Christensen, R., 1990. *Linear models for multivariate, time and spatial data*. New York: Springer.

Cochran, W.G., 1977. *Sampling techniques*. 3rd ed. New York, NY: Wiley.

de Carvalho, L.G., *et al.*, 2010. Multivariate geostatistical application for climate characterization of Minas Gerais State, Brazil. *Theoretical and Applied Climatology*, 102, 417–428. doi:10.1007/s00704-010-0273-z.

de Gruijter, J.J., *et al.*, 2006. *Sampling for natural resource monitoring*. Berlin: Springer.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Hart, J.K. and Martinez, K., 2006. Environmental sensor networks: a revolution in the earth system science? *Earth-Science Reviews*, 78, 177–191. doi:10.1016/j.earscirev.2006.05.001.

Hengl, T., Heuvelink, G.B.M., and Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. *Computers & Geosciences*, 33, 1301–1315. doi:10.1016/j.cageo.2007.05.001.

Hengl, T., Heuvelink, G.B.M., and Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120, 75–93. doi:10.1016/j.geoderma.2003.08.018.

Heuvelink, G.B.M., Brus, D.J., and de Gruijter, J.J., 2006. Chapter 11 optimization of sample configurations for digital mapping of soil properties with universal kriging. *In*: A.B.M.P.

Lagacherie and M. Voltz, eds. *Digital soil mapping – an introductory perspective*. Amsterdam: Elsevier, 137–151.
Heuvelink, G.B.M., *et al*., 2010. Optimization of mobile radioactivity monitoring networks. *International Journal of Geographical Information Science*, 24, 365–382. doi:10.1080/13658810802646687.
Hijmans, R.J., *et al*., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. doi:10.1002/joc.1276.
Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., 1983. Optimization by simulated annealing. *Science*, 220, 671–680. doi:10.1126/science.220.4598.671.
Kuhlman, K.L. and Igúzquiza, E.P., 2010. Universal cokriging of hydraulic heads accounting for boundary conditions. *Journal of Hydrology*, 384, 14–25. doi:10.1016/j.jhydrol.2010.01.002.
Lark, R.M., Cullis, B.R., and Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, 57, 787–799. doi:10.1111/j.1365-2389.2005.00768.x.
Lesch, S.M., 2005. Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. *Computers and Electronics in Agriculture*, 46, 153–179. doi:10.1016/j.compag.2004.11.004.
Li, L.F., *et al*., 2012. Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. *Atmospheric Environment*, 55, 220–228. doi:10.1016/j.atmosenv.2012.03.035.
Li, X., *et al*., 2013. Heihe watershed allied telemetry experimental research (HiWATER): scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, 94, 1145–1160. doi:10.1175/BAMS-D-12-00154.1.
Li, X., *et al*., 2009. Watershed allied telemetry experimental research. *Journal of Geophysical Research-Atmospheres*, 114, D22103. doi:10.1029/2008JD011590.
Melles, S.J., *et al*., 2011. Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computers & Geosciences*, 37, 280–288. doi:10.1016/j.cageo.2010.04.007.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30, 683–691. doi:10.1016/j.cageo.2004.03.012.
Qi, S.Z. and Luo, F., 2007. Environmental degradation problems in the Heihe River Basin, northwest China. *Water and Environment Journal*, 21, 142–148. doi:10.1111/j.1747-6593.2006.00059.x.
R Core Team, 2012. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org/
Royle, J.A. and Nychka, D., 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers & Geosciences*, 24, 479–488. doi:10.1016/S0098-3004(98)00020-X.
Schabenberger, O. and Gotway, C.A., 2005. *Statistical methods for spatial data analysis*. Boca Raton, FL: Chapman & Hall/CRC.
Sørensen, R., Zinko, U., and Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, 10, 101–112. doi:10.5194/hess-10-101-2006.
Stein, A. and Corsten, L.C.A., 1991. Universal kriging and cokriging as a regression procedure. *Biometrics*, 47, 575–587. doi:10.2307/2532147.
Stein, A., Eijnsbergen, A.C., and Barendregt, L.G., 1991. Cokriging nonstationary data. *Mathematical Geology*, 23, 703–719. doi:10.1007/BF02082532.
Turner, D.P., *et al*., 2006. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment*, 102, 282–292. doi:10.1016/j.rse.2006.02.017.
van Groenigen, J.W., Siderius, W., and Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87, 239–259. doi:10.1016/S0016-7061(98)00056-1.
van Groenigen, J.W. and Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environment Quality*, 27, 1078–1086. doi:10.2134/jeq1998.00472425002700050013x.
Vašát, R., Heuvelink, G.B.M., and Borůvka, L., 2010. Sampling design optimization for multivariate soil mapping. *Geoderma*, 155, 147–153. doi:10.1016/j.geoderma.2009.07.005.
Ver Hoef, J.M. and Cressie, N., 1993. Multivariable spatial prediction. *Mathematical Geology*, 25, 219–240. doi:10.1007/BF00893273.

Walvoort, D.J.J., Brus, D.J., and de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers & Geosciences*, 36, 1261–1267. doi:10.1016/j.cageo.2010.04.005.

Wan, Z., *et al*., 2004. Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing*, 25, 261–274. doi:10.1080/0143116031000116417.

Wang, J.H., *et al*., 2012. Effect of the sampling design of ground control points on the geometric correction of remotely sensed imagery. *International Journal of Applied Earth Observation and Geoinformation*, 18, 91–100. doi:10.1016/j.jag.2012.01.001.

Yeh, M.S., Lin, Y.P., and Chang, L.C., 2006. Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms. *Environmental Geology*, 50, 101–121. doi:10.1007/s00254-006-0190-8.

Zhang, Y., *et al*., 2012. An eco-hydrology wireless sensor demonstration network in high-altitude and alpine environment in the Heihe River Basin of China. *Wireless Sensor Network*, 4, 138–146. doi:10.4236/wsn.2012.45020.

Zhu, Z.Y. and Stein, M.L., 2006. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural Biological and Environmental Statistics*, 11, 24–44. doi:10.1198/108571106X99751.