Regression Kriging-Based Upscaling of Soil Moisture Measurements From a Wireless Sensor Network and Multiresource Remote Sensing Information Over Heterogeneous Cropland

Jian Kang, Rui Jin, *Member, IEEE*, and Xin Li, *Senior Member, IEEE*

***Abstract*—The ground truth estimated by *in situ* measurements is important for accurately evaluating retrieved remote sensing products, particularly over heterogeneous land surfaces. This letter analyzes the role of multisource remote sensing observations on the upscaling of soil moisture observed by a wireless sensor network at the pixel scale via the regression kriging (RK) method. Three types of auxiliary remote sensing information are employed, including Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Temperature Vegetation Dryness Index (TVDI; 90 m), Polarimetric L-band Multiband Radiometer brightness temperature (700 m), and Moderate Resolution Image Spectroradiometer TVDI (1000 m). Moreover, a comparison with the ordinary kriging method is analyzed. The spatial inferences show that the RK method is more accurate and that its spatial pattern is more consistent with the auxiliary data when the trend is successfully removed, particularly when spatial continuity is destroyed by irrigation. The ASTER TVDI has a higher resolution and stronger correlation with soil moisture and yields more accurate interpolation results than the other types of remote sensing information. Although medium-resolution data do not substantially contribute to capture the spatial patterns of soil moisture, such data may still improve the prediction accuracy.**

***Index Terms*—Regression kriging (RK), remote sensing, soil moisture, spatial heterogeneity, upscaling, wireless sensor network.**

I. Introduction

Soil moisture is an important variable in the water, energy, and biogeochemical cycles within climate systems. The quantitative estimation of soil moisture is meaningful for many interrelated research and application fields, such as weather predictions, hydrology, watershed management, and flood forecasts. Soil moisture is generally obtained through ground-based observation networks, land surface models, and remote sensing technology. Model simulations and remote sensing retrievals may contain large uncertainties because of the inherent complexity of physical processes and their parameterization [1]. Therefore, grid-based approaches should be calibrated and validated against ground-based observations, traditionally via *in situ* measurements.

Recently, estimations of soil moisture at the grid scale based on point measurements, which is also known as upscaling, have received considerable attention. There are three general methods of performing these estimations: 1) empirical methods, such as the multipoint average within a pixel; 2) upscaling methods based on time information, such as time stability analyses; and 3) kriging methods, such as kriging technology that considers the spatial autocorrelations of target variables.

Soil moisture shows strong spatial variability [2] that may be caused by topography, soil texture, and vegetation coverage [3]. Understanding the spatial characteristics of soil moisture is helpful for estimating the soil moisture at heterogeneous remote sensing pixels. Kriging based on the spatial autocorrelation theory is an effective upscaling method for inferring soil moisture estimations and variance [3], [4].

Most of the previous works on kriging-based upscaling of soil moisture have been performed by directly interpolating *in situ* measurements. Such upscaling strategies are applicable when soil moisture is spatially continuous and satisfies the assumption of second-order stationary processes. Otherwise, mixed models that combine kriging and auxiliary information, such as the regression kriging (RK) method [5], are more instructive for spatial predictions. Because of the high correlations between soil moisture and topography, terrain factors such as slope and aspect are often used in combination with kriging techniques [6]. Snepvangers *et al.* [7] have conducted meaningful work in which soil moisture is predicted by employing the net water input obtained through the water balance.

In this case, the strong heterogeneity of soil moisture in the study area is primarily caused by irrigation. The spatial distribution of irrigation patterns changes with time and shows randomness. Therefore, to improve prediction accuracy, it is necessary to incorporate additional information to capture soil moisture patterns. Remote sensing information can collect the spatial characteristics of a target variable because it has spatially continuous coverage. In this letter, multiresource remote sensing information related to soil moisture is used to establish the regression models that are combined with kriging techniques.

Manuscript received January 31, 2014; revised May 5, 2014; accepted May 22, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 91125001 and Grant 41071226 and in part by the Chinese Academy of Sciences Action Plan for West Development under Grant KZCX2-XB3-15. *(Corresponding author: R. Jin.)*

The authors are with the Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China (e-mail: kangjian@lzb.ac.cn; jinrui@lzb.ac.cn; lixin@lzb.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2326775

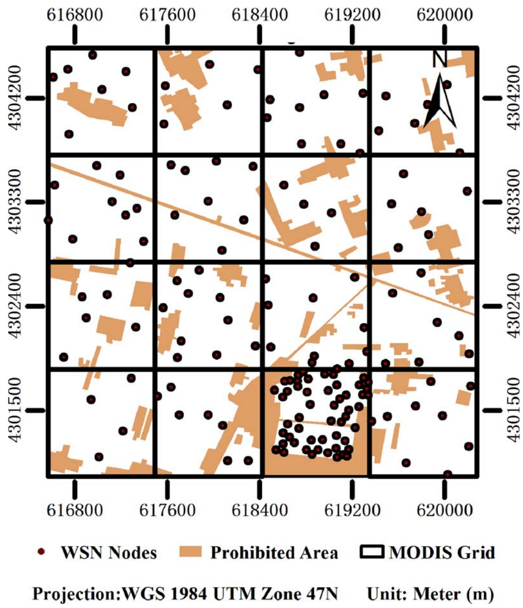

Fig. 1. Map of the study area showing the location of the EHWSN nodes.

In addition, different prediction results based on diverse remote sensing information are compared and analyzed.

II. Study Area and Data

The study area is located in the middle reach of the Heihe River Basin, which is situated in northwest China. The main crop type in the study region is seed corn, which covers approximately 75% of the total area. Other plants, such as wheat, vegetables, and fruits, are also present. There is a dense network of canals that serve as the irrigation system in this region. Soil moisture spatially varies because of the artificial allocation of irrigation water to farmland.

The Ecological and Hydrological Wireless Sensor Network (EHWSN), a multiscale observation network that is one of the fundamental experiments of the Heihe Water Allied Telemetry Experimental Research (HiWATER) project [8], is used to capture the spatial heterogeneity of soil moisture in the middle reach of the Heihe River Basin and produce accurate ground-based observations to validate the soil moisture remote sensing products (doi: 10.3972/hiwater.118.2013.db; 10.3972/hiwater.120.2013.db; 10.3972/hiwater.119.2013.db). The EHWSN distributed 163 nodes that are positioned in 4 × 4 Moderate Resolution Image Spectroradiometer (MODIS) grids (see Fig. 1) using the optimal spatial sampling method [9]. The instruments observe soil moisture and temperature at depths of 4 and 10 cm at 10-min intervals [10]. Because of the high temporal resolution of the EHWSN, the ground-based soil moisture observations can be precisely synchronized with satellite and airborne remote sensing measurements.

Remote sensing can capture the spatial heterogeneity of surface variables because of its continuous coverage of land surfaces. To enhance the prediction accuracy, multiple-source remote sensing information is employed that includes land surface temperature (LST, MOD11A1) and the Normalized Difference Vegetation Index (NDVI, MOD13A2) from MODIS. Another LST mapping with higher resolution is retrieved using the algorithm developed by Li *et al.* [11] from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER); the NDVI is also calculated from ASTER data. The microwave brightness temperature (T_B) with horizontal polarization is obtained from the airborne Polarimetric L-band Multiband Radiometer (PLMR) with a 7.5° incidence angle (doi: 10.3972/hiwater.013.2013.db; 10.3972/hiwater.017.2013.db; 10.3972/hiwater.021.2013.db).

The T_B is directly used as auxiliary information because of its strong correlation with soil moisture. The Temperature Vegetation Dryness Indexes (TVDI) are calculated using the LST and NDVI [12], which effectively indicate the spatial patterns of soil moisture and are generally used to downscale microwave soil moisture products.

III. Methodology

A. Prediction Method

Most environmental variables have spatial autocorrelation characteristics. For such variables, the kriging estimation method, which depends on variograms, can be used to predict values at unobserved points or blocks. A spatial autocorrelation variable could be modeled as the sum of a deterministic trend $m(u)$ and a random autocorrelated residual component $\varepsilon(u)$ [13], i.e.,

$$Z(u) = m(u) + \varepsilon(u). \tag{1}$$

Remote sensing data can reflect the spatial heterogeneity of variables at its observation scale. To improve the prediction accuracy, remote sensing data are used to estimate a linear trend model. The RK method combined with a linear trend model is used to estimate target variables, and the advantages of this approach have been verified in several studies [5], [14]. The target variable, such as the soil moisture, at a new location u_0 is estimated by adding the inferred residuals to the fitted trend model. The residuals are interpolated using ordinary block kriging (OBK), which converts the residuals to a grid of the same size as the pixel of remote sensing data, i.e.,

$$\begin{aligned} z^*(u_0) &= \sum_{k=0}^{p} \beta_k^* \cdot q_k(u_0) + \sum_{i=1}^{n} w_i(u_0) \cdot \varepsilon(u_i) \\ q_0(u_0) &= 1 \end{aligned} \tag{2}$$

where p is the number of auxiliary variables, $q_k(u_0)$ is the value of the kth auxiliary variable at location u_0, n is the number of samples, and $w_i(u_0)$ is the weighted coefficient used to interpolate $\varepsilon(u)$ using OBK. Moreover, β_k^* is the coefficient of the estimated trend model and is fitted using the ordinary least squares (OLS) method, i.e.,

$$\beta_{\text{OLS}}^* = (\mathrm{q}^T \cdot \mathrm{q})^{-1} \cdot \mathrm{q}^T \cdot \mathrm{z} \tag{3}$$

where q is the $[n, p+1]$ matrix of the auxiliary variables at all EHWSN node locations, and z is the soil moisture observation vector from the EHWSN.

TABLE I
SIGNIFICANCE TESTS OF THE REGRESSION COEFFICIENTS

Predictor	Pr(>|t|)	Level	Pr(>|t|)	Level	Pr(>|t|)	Level
	June 15, 2012		June 24, 2012		Aug 11, 2012	
Intercept	5.01E-01		9.57E-02	.	2.39E-01	
ASTER TVDI	1.09E-04	****	2.00E-06	****	4.14E-02	**
X	9.80E-01		8.64E-01		6.27E-02	*
Y	4.84E-01		8.97E-02	*	1.15E-01	
	June 30, 2012		July 07, 2012		Aug 02, 2012	
Intercept	1.47E-01		7.38E-02	*	6.75E-02	*
PLMR TB	7.63E-04	****	4.73E-02	**	1.17E-05	****
X	9.23E-01		2.15E-01		9.68E-01	
Y	1.16E-01		5.49E-02	*	8.99E-02	*
	Aug 28, 2012		Sept 03, 2012		Sept 08, 2012	
Intercept	7.21E-05	****	9.21E-13	****	9.25E-10	****
MODIS TVDI	9.87E-02	*	8.93E-02	*	5.04E-02	*
X	2.40E-01		8.61E-01		4.43E-01	
Y	7.15E-05	****	2.43E-13	****	5.30E-10	****

Significance test codes: ****(0.001), ***(0.01), **(0.05) and *(0.1)

TABLE II
OLS REGRESSION MODELS AND TESTS

	Date	Regression Model	R^2	*p* value
ASTER 90 m	June 15, 2012	1.61-3.90×TVDI	0.131	1.03E-04
	June 24, 2012	1.02-4.24×TVDI	0.198	1.06E-08
	Aug 11, 2012	0.876-0.93×TVDI	0.052	4.50E-02
PLMR 700 m	June 30, 2012	14.75-0.056×TB	0.074	5.43E-04
	July 07, 2012	18.47-0.07×TB	0.065	3.08E-03
	Aug 02, 2012	49.12-0.18×TB	0.198	1.06E-08
MODIS 1000 m	Aug 28, 2012	2533.0+1.37×TVDI-5.89E-04×Y	0.242	5.50E-10
	Sept 03, 2012	2457.0+2.02×TVDI-5.71E-04×Y	0.270	2.03E-09
	Sept 08, 2012	1516.0+6.96×TVDI-3.52E-04×Y	0.118	3.13E-04

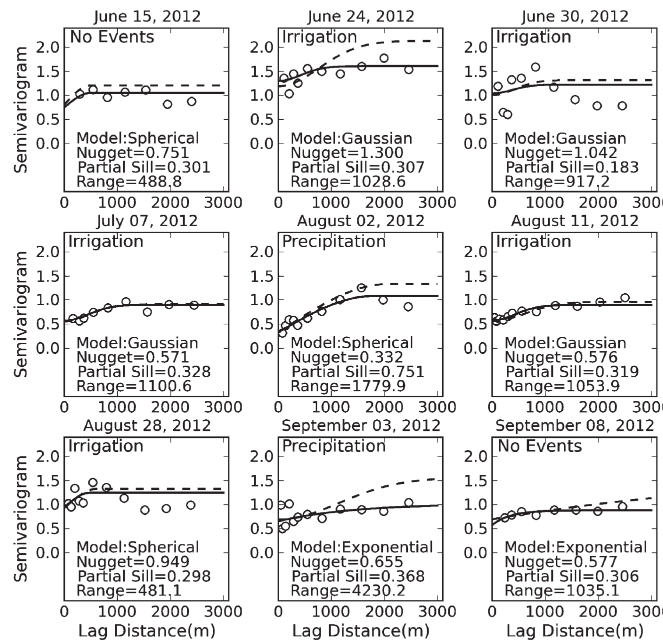

Fig. 2. Experimental semivariograms (circles) and soil moisture variograms (dotted line) predicted by the EHWSN and residual variograms (solid line) predicted by OLS under various environmental conditions (e.g., precipitation and irrigation).

B. Evaluation

K-fold cross validation is a common method used to evaluate prediction accuracy. The EHWSN samples are randomly divided into k equal size subsamples. K-fold cross validation is performed by eliminating a single subsample at a time as the validation data and subsequently predicting values of this subsample by using the remaining k-1 subsamples. We examine the difference between the observed and predicted values using the mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r), i.e.,

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}\left[z(u_i) - z^*(u_i)\right] \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[z(x_i) - z^*(x_i)\right]^2} \tag{5}$$

$$r = \frac{\sum_{i=1}^{N}\left[z(u_i) - \overline{z(u_i)}\right]\left[z^*(u_i) - \overline{z^*(u_i)}\right]}{\sqrt{\sum_{i=1}^{N}\left[z(u_i) - \overline{z(u_i)}\right]^2 \cdot \sum_{i=1}^{N}\left[z^*(u_i) - \overline{z^*(u_i)}\right]^2}} \tag{6}$$

where $z^*(u_i)$ is the predicted value, $z(u_i)$ is the observed value, and N is the number of samples. For k-fold cross validation, the evaluation index is defined as follows:

$$Q = \frac{1}{k}\sum_{i=1}^{k} I_i \tag{7}$$

where k is the number of subsamples, and I_i may be the RMSE, ME, or r of the ith subsample. A common choice for K-fold cross validation is $K = 10$.

IV. RESULTS

A. Data Analysis and Regression Modeling of Trends

Kriging methods require that data conform to a normal distribution. To enhance normality and symmetry, the soil moisture raw data are transformed using logit transformation. The details of this method can be found in [14].

The correlations between the logit transformation of soil moisture and predictors (i.e., T_B, TVDI, X, and Y coordinates) are analyzed using the t-test statistic of the regressions (see Table I) to determine the optimal predictor to remove the trend. Most of the X coordinate coefficients are not statistically significant at a less than 90% confidence level, which indicates that the X coordinates have almost no correlation with soil moisture. All of the predictor coefficients from the remote sensing information exhibit statistical significance at a probability of at least 0.9. Compared with MODIS, the coefficients of

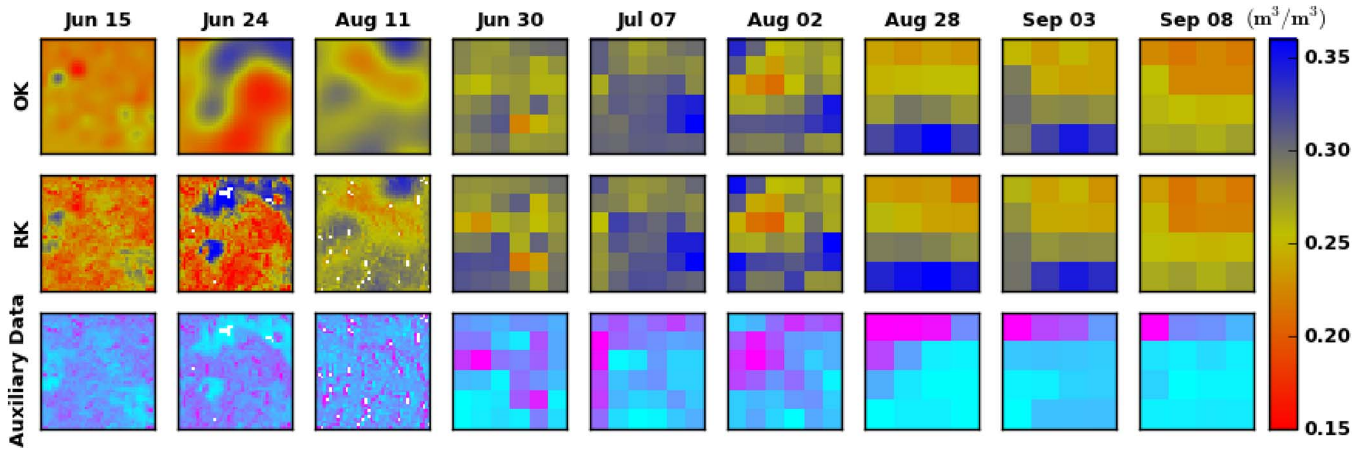

Fig. 3. Soil moisture interpolated using the RK and OK methods.

ASTER and PLMR have a higher confidence. Because artificial irrigation was performed from June 20, 2012 to September 1, 2012, the soil moisture was spatially reallocated. The Y coordinate is negatively correlated with soil moisture, particularly during the latter period of irrigation (August 28, September 3, and September 8).

The predictors with regression coefficients that exhibit low statistical significance are removed, and the linear model is then reestablished with the remaining predictors. The coefficient of determination (R^2) and test statistic (p value) are used to verify the goodness of fit and significance of the regression models, respectively (see Table II). Although the R^2 values of the regression models for July 7 and August 11 are very low, these both regression models are still statistically significant at the $p < 0.05$ level.

B. Variogram Analysis of the Random Residuals

Fig. 2 shows the variograms of both *in situ* soil moisture measurements and the residuals predicted by OLS for various environmental conditions, e.g., precipitation and irrigation. The residuals are calculated by subtracting the EHWSN soil moisture observations from the corresponding model predictions according to Table II. Three universal variogram models, including the Gaussian, spherical, and exponential models, are compared, and the theoretical model with the smallest residual sum of squares relative to the experimental semivariograms is then selected. The soil moisture variogram is used for interpolation with the OBK method, and the variogram of the residuals is used in the RK method.

The variograms of the residuals represent smaller ranges of spatial dependence and smaller sills than the variograms from soil moisture, which suggests that the trend is successfully removed. However, the soil moisture variograms for July 7 and August 11 are nearly identical to the residual variograms. This similarity is caused by the trend models with low coefficients of determination. The effect of removing the trend is more remarkable for the ASTER and MODIS TVDIs than the PLMR T_B.

Local irrigation spatially redistributes soil moisture and enhances surface heterogeneity. This phenomenon damaged the structure of the variograms for June 30 and August 28. As a result of the increase in heterogeneity, the variograms display a shorter range of spatial dependence than that under the precipitation condition, which makes the ground more homogeneous, and the variograms display a longer range of spatial dependence, particularly for September 03. Without precipitation and irrigation events, the variograms faithfully reflect the spatial variability of soil moisture.

C. Bias and Prediction Accuracy

The estimated OLS residuals at the EHWSN nodes are interpolated using OBK and added to the trend. The predicted block size is consistent with the resolution of the remote sensing data. The final predicted results using the RK method are shown in Fig. 3, whereas the ordinary kriging (OK) results using only the EHWSN observations are shown for comparison.

The OK maps exhibit smooth transitions relative to the RK method. The RK maps exhibit more detailed spatial patterns than the OK maps when the high-resolution ASTER TVDI is used. Because of the higher correlation with soil moisture, the RK results on June 24 are more similar to the auxiliary data relative to the results on June 15 and August 11. With the increasing size of the interpolation grid, the differences between the RK and OK methods are small using the PLMR and MODIS data. Therefore, the coarse-resolution auxiliary data provide a smaller contribution when capturing soil moisture spatial characteristics. However, low-resolution data remain effective for improving the prediction accuracy.

Table III shows several assessment indexes used to evaluate prediction accuracy. We expect the predictions to be unbiased; if prediction values are unbiased, the ME should be approximately zero. There are six cases where the RK biases are smaller than the OK biases, and three of these cases use the ASTER data. The degree of bias is described by the RMSE. The RK RMSE is only worse than the OK RMSE for July 07, and the RMSE is much closed for August 11. To compare the consistency between the measurements and predictions, correlation coefficients are calculated. The RK method exhibits a distinct advantage over the OK method, and the RK method using the ASTER TVDI data yields the best results; the second best case uses the MODIS TVDI data. Although the MODIS data resolution is coarser than that of the PLMR T_B, the Y

TABLE III
VALIDATION OF BOTH THE RK AND OK METHODS

Date	ME		RMSE		r	
	OK	RK	OK	RK	OK	RK
June 15, 2012	-0.0015	-0.0008	0.0396	0.0387	0.2738	0.3864
June 24, 2012	-0.0022	-0.0015	0.0654	0.0609	0.4279	0.5501
June 30, 2012	-0.0039	-0.0029	0.0540	0.0537	0.3244	0.3528
July 07, 2012	0.0001	0.0006	0.0389	0.0394	0.2969	0.3328
Aug 02, 2012	-0.0005	0.0002	0.0483	0.0453	0.4576	0.5252
Aug 11, 2012	0.0042	0.0046	0.0486	0.0484	0.2219	0.2104
Aug 28, 2012	0.0011	0.0025	0.0614	0.0593	0.6758	0.6997
Sept 03, 2012	0.0006	0.0005	0.0493	0.0474	0.5854	0.6235
Sept 08, 2012	-0.0003	-0.0002	0.0435	0.0413	0.2414	0.3613
Average performance time in R language (seconds): OK 0.11 and RK 0.13						

coordinate, which exhibits a strong correlation, is also used to establish the trend model. Therefore, the accuracy of the RK method using the MODIS TVDI data is higher than that when using the PLMR T_B.

V. DISCUSSION AND CONCLUSIONS

The OK method is based on strict second-order stationary assumptions, but irrigation events may make it difficult to satisfy this hypothesis. However, the RK method avoids this problem, which is one of reasons why the accuracy of RK is generally higher than that of OK. The differences between the RK and OK methods are examined in this study. When high-resolution data are used for the interpolations, the RK-interpolated results exhibit more detailed spatial patterns and higher prediction accuracy relative to the OK method. The contribution of coarse-resolution data in capturing the structural characteristics of the target variable is weak; however, such data may improve the prediction accuracy. In addition, employing auxiliary data does not substantially reduce the performance efficiency (see Table III). The computational expense of the RK method is low; thus, the assistant data should be properly considered.

Correlation analyses between the predictors and target variables are required. As the correlations between the target variables and predictors increase, the effect of removing the trend becomes more pronounced, which would improve the interpolation accuracy. If inappropriate predictors are selected, the RK method may produce estimates that are similar to or even worse than the OK method.

RK relies on the data quality of the target variable. If the data are from a biased sampling design, the prediction may be even worse than with a simple method, such as the inverse distance interpolation method. An incorrect data point may lead to a bad regression model, which would affect the global predictions. However, undersampling is another factor that might affect the accuracy of RK. For variogram modeling, the number of paired points must be adequate for various spacings, and at least 50 points are recommended to estimate a variogram. For regression modeling, more than 50 observations per predictor are strongly advised to prevent overfitting [15]. The results are not affected by the aforementioned factors because the data quality has been controlled and the sampling design considers both variogram modeling and spatial prediction accuracy [9].

Irrigation damages the structure of variograms, which increases their uncertainty. Further research should introduce irrigation data for upscaling. It is difficult to apply a single upscaling method to all scenarios. Therefore, based on the spatial data, the autoselection of a suitable method should be an important component of future work.

REFERENCES

[1] X. Li, "Characterization, controlling, reduction of uncertainties in the modeling and observation of land-surface systems," *Sci. China Earth Sci.*, vol. 57, no. 1, pp. 80–87, Jan. 2014.
[2] D. Bosch, V. Lakshmi, T. Jackson, M. Choi, and J. Jacobs, "Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003," *J. Hydrol.*, vol. 323, no. 1–4, pp. 120–137, May 2006.
[3] W. T. Crow *et al.*, "Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products," *Rev. Geophys.*, vol. 50, no. 2, Jun. 2012.
[4] V. Pandey and P. K. Pandey, "Spatial and temporal variability of soil moisture," *Int. J. Geosci.*, vol. 1, no. 2, pp. 87–98, Aug. 2010.
[5] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. New York, NY, USA: Oxford Univ. Press, 1997.
[6] Y. Qiu, B. Fu, J. Wang, and L. Chen, "Spatial variability of soil moisture content and its relation to environmental indices in a semi-arid gully catchment of the Loess Plateau, China," *J. Arid Environ.*, vol. 49, no. 4, pp. 723–750, Dec. 2001.
[7] J. Snepvangers, G. Heuvelink, and J. Huisman, "Soil water content interpolation using spatio-temporal kriging with external drift," *Geoderma*, vol. 112, no. 3/4, pp. 253–271, Mar./Apr. 2003.
[8] X. Li *et al.*, "Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific objectives and experimental design," *Bull. Amer. Meteorol. Soc.*, vol. 94, no. 8, pp. 1145–1160, Aug. 2013.
[9] J. Kang *et al.*, "Hybrid optimal design of the eco-hydrological wireless sensor network in the middle reach of the Heihe River Basin, China," *Int. J. Appl. Earth Observ. Geoinf.*, to be published.
[10] R. Jin *et al.*, "A nested eco-hydrological wireless sensor network for capturing the surface heterogeneity in the midstream areas of the Heihe River Basin, China," *IEEE Geosci. Remote S.*, vol. 11, no. 11, pp. 2015–2019, Nov. 2014.
[11] H. Li *et al.*, "Evaluation of the VIIRS and MODIS LST products in an arid area of northwest China," *Remote Sens. Environ.*, vol. 142, pp. 111–121, Feb. 2014.
[12] I. Sandholt, K. Rasmussen, and J. Andersen, "A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status," *Remote Sens. Environ.*, vol. 79, no. 2/3, pp. 213–224, Feb. 2002.
[13] I. Odeha, A. McBratney, and D. Chittleborough, "Spatial prediction of soil properties from landform attributes derived from a digital elevation model," *Geoderma*, vol. 63, no. 3/4, pp. 197–214, Nov. 1994.
[14] T. Hengl, "A practical guide to geostatistical mapping of environmental variables," Joint Research Centre, Luxemburg, IA, USA, Tech. Rep., 2007, EUR., Sci.
[15] T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter, "About regression-kriging: From equations to case studies," *Comput. Geosci.*, vol. 33, no. 10, pp. 1301–1315, Oct. 2007.