







# Spatial distribution of snow depth based on geographically weighted regression kriging in the Bayanbulak Basin of the Tianshan Mountains, China

LIU Yang<sup>1,2</sup>  <http://orcid.org/0000-0001-5751-0982>; e-mail: liuyang\_dsslab@163.com

LI Lan-hai<sup>1,2</sup>  <http://orcid.org/0000-0002-1914-0180>; e-mail: lilh@ms.xjb.ac.cn

CHEN Xi<sup>1,2\*</sup>  <http://orcid.org/0000-0003-0432-0733>;  e-mail: chenxi@ms.xjb.ac.cn

YANG Jin-Ming<sup>3</sup>  <http://orcid.org/0000-0002-1843-6908>; e-mail: 2604701724@qq.com

HAO Jian-Sheng<sup>1,2</sup>  <http://orcid.org/0000-0003-3296-5838>; e-mail: haojiansheng@ms.xjb.ac.cn

\* Corresponding author

<sup>1</sup> State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

<sup>2</sup> CAS Research Center for Ecology and Environment of Central Asia, No.818 South Beijing Road, Urumqi 830011, China

<sup>3</sup> Institute of Arid Ecology and Environment, Xinjiang University, Urumqi 830046, China

**Citation:** Liu Y, Li LH, Chen X, et al. (2018) Spatial distribution of snow depth based on geographically weighted regression kriging in the Bayanbulak Basin of the Tianshan Mountains, China. *Journal of Mountain Science* 15(1). <https://doi.org/10.1007/s11629-017-4564-z>

© Science Press, Institute of Mountain Hazards and Environment, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2018

**Abstract:** Snow depth is a general input variable in many models of agriculture, hydrology, climate and ecology. This study makes use of observational data of snow depth and explanatory variables to compare the accuracy and effect of geographically weighted regression kriging (GWRK) and regression kriging (RK) in a spatial interpolation of regional snow depth. The auxiliary variables are analyzed using correlation coefficients and the variance inflation factor (VIF). Three variables, Height, topographic ruggedness index (TRI), and land surface temperature (LST), are used as explanatory variables to establish a regression model for snow depth. The estimated spatial distribution of snow depth in the Bayanbulak Basin of the Tianshan Mountains in China with a spatial resolution of 1 km is obtained. The results indicate

that 1) the result of GWRK's accuracy is slightly higher than that of RK ( $R^2 = 0.55$  vs.  $R^2 = 0.50$ ,  $RMSE$  (root mean square error) = 0.102 m vs.  $RMSE = 0.077$  m); 2) for the subareas, GWRK and RK exhibit similar estimation results of snow depth. Areas in the Bayanbulak Basin with a snow depth greater than 0.15 m are mainly distributed in an elevation range of 2632.00–3269.00 m and the snow in this area comprises 45.00–46.00% of the total amount of snow in this basin. However, the GWRK resulted in more detailed information on snow depth distribution than the RK. The final conclusion is that GWRK is better suited for estimating regional snow depth distribution.

**Keywords:** Snow depth; Spatial distribution; Regression kriging; Geographically weighted regression kriging

**Received:** 20 June 2017

**Revised:** 12 October 2017

**Accepted:** 13 November 2017

## Introduction

Snowpack is one of the most active natural elements in the cryosphere. Its special physical features (such as high emissivity, high albedo, large melting latent heat, and low thermal conductivity) and broad geographic distribution can greatly impact the surface energy budget and climate change (Tarnocai et al. 2009; Zhang 2005; Zimov et al. 2006). Among various snowpack parameters, snow depth is the main parameter that reflects the dynamic changes of snowpack. Because of this, snow depth is a significant factor researchers consider when studying hydrologic processes and assessing water resources (Bocchiola et al. 2008; Marshall et al. 1994). Therefore, snow depth monitoring has become an area of rapidly growing research when near-surface processes in cold regions are being analyzed (McCreight et al. 2014).

Distribution of snow depth is influenced to a large degree by terrain and weather conditions. Due to differences in temperature, solar radiation, and other meteorological variables, there is considerable variation in snow depth at different elevations. In addition, snow depth undergoes dynamic changes as a result of energy absorption and release (Lu et al. 2014). The Tianshan Mountains in Xinjiang, China are about 2500 km in length and the south-north width is about 250–350 km. The area consists of large mountains, inter-mountain basins, and valleys. The terrain is rather complex with many variations and the overall elevation decreases from west to east (Lu et al. 2015). Snow is one of the main water resources for the arid region in the mountain areas and affects the socio-economics and ecology of the region. Snow is rather sensitive to climate change, especially with regard to changes in temperature and solar radiation. Given the concern about global warming, the uncertainty of the temporal-spatial distribution of snow has intensified (Lu et al. 2015).

Due to the broad application of satellite remote sensing data in ice and snow study since the 1970s, researchers worldwide have utilized different remote sensing data to conduct research on snow depth inversion to determine the spatial distribution of snow depth (Che et al. 2012; Dai et al. 2015; Liang et al. 2015). There will be access to snow depth monitoring in areas with low density or areas which is lack of observation stations.

However, the relationship between snow depth and the brightness values in the optical and near infrared bands is not very clear yet (Liang et al. 2015) and microwave remote sensing is limited by the difficulty of obtaining data as well as the coarse spatial resolution (Che et al. 2012). Therefore, spatial interpolation is an effective way for obtaining timely snow depth distribution (López-Moreno and Noguéogueno 2006). Researchers worldwide have developed various methods to estimate the spatial distribution of snow depth using interpolation (Bocchiola et al. 2010; Collados-Lara et al. 2017; Erxleben et al. 2002; Huang et al. 2015; López-Moreno and Noguéogueno 2006). Interpolation methods such as ordinary kriging (OK), universal kriging (UK), ordinary co-kriging (OCK), and universal co-kriging (UCK) have been used to estimate the spatial distribution of regional snow depth. The result have shown that OK is appropriate for the estimation of snow depth when the snow depth is low and UCK is suitable for the estimation of snow covering extensive areas (Huang et al. 2015). Since snow depth is strongly influenced by the terrain conditions, potential solar radiation, and other meteorological variables, it varies greatly for different areas. Spatial interpolation based on ground-based observations of snow depth has certain limitations (Harshburger et al. 2010). With the development of spatial analysis theory, new estimation methods that combine the meteorological or terrain variables with regression tree models and generalized additive models have been put forward (Erxleben et al. 2002; López-Moreno and Nogués-Bravo 2006). These studies have indicated that snow depth estimation methods that consider explanatory variables such as terrain, potential radiation, etc. have resulted in an improvement in the estimation accuracy.

These estimation methods are based on a global regression of ordinary least squares (OLS) to establish a relationship between snow depth and the environmental variables. However, snow exhibits strong variation and spatial non-stationarity. It is difficult with OLS to determine the local characteristics of snow depth variability owing to complex terrain, vegetation cover, energy input and output and this limits the interpolation accuracy. Therefore, an approach that reflects the non-stationarity is required. Geographically

weighted regression kriging (GWRK) is a hybrid interpolation method that replaces the global regression portion of regression kriging (RK) with the local regression portion of the GWR while retaining the kriging interpolation for the residual term (Wang et al. 2012; Wasige et al. 2013). The regression coefficients of the explanatory variables, which influence the spatial distribution of the snow in the GWRK will change based on the change in the spatial position. Therefore, the interpolation result can reveal certain local changes that are masked by spatial non-stationarity, thus providing a more realistic situation of the spatial distribution of snow depth.

This study aims to determine suitable methods for the spatial interpolation of snow depth in the Bayanbulak Basin in the middle part of the Tianshan Mountains of Xinjiang. On the one hand, eight-day composite snow cover data (MOD10A2 and MYD10A2) was adopted to reflect the largest range of snow cover. These data has a high snow classification accuracy (the snow recognition rate is between 68.30%-100.00% and its average is about 87.00%). It can eliminate the influence on the accurate classification of snow cover imposed by clouds (Huang et al. 2007; Zhou et al. 2005). On the other hand, explanatory variables that are relevant to snow depth, such as the application of terrain variables. By exploring the auto-correlation of the spatial distribution of snow depth, this study compares the interpolation

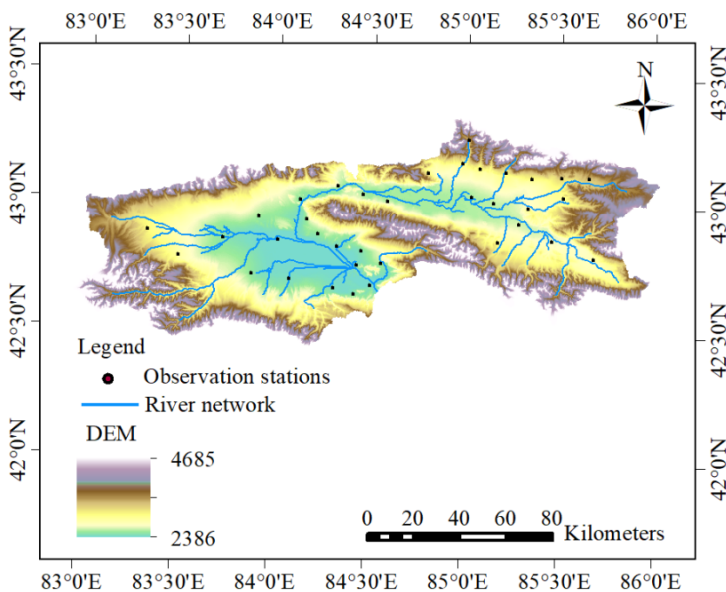
accuracy and the spatial non-stationarity of GWRK and RK by analyzing and comparing validation points. The establishment of the spatial snow depth estimation method is aimed at obtaining spatial distribution data of regional snow depth with high accuracy.

## 1 Study Site and Methods

### 1.1 Study area

The Bayanbulak Basin is located in middle of the Tianshan Mountains. It is surrounded by mountains on three sides and is a semi-closed geographic area with typical features of a high-elevation and cold basin (Figure 1). The Bayanbulak Basin includes the Big and Small Yultuz Basins with a total area of 136984 km<sup>2</sup>. The east-west length of the basin is 270 km and the north-south width is 136 km. The basin bottom is flat and the altitude ranges from 2400 m to 2600 m. The basin is surrounded by large mountains with altitudes in the range of 4000–5500 m. Areas of permanent snow and glaciers exist in the mountains. Due to its alpine conditions and distinct landforms, various alpine steppes and meadow ecosystem occur in the area. There are large areas of wet meadows and lakes with rich aquatic plants and animals, forming distinct inland wetland ecosystems that play a significant role in water yield adjustment, water storage, and maintenance of the regional water balance.

The high elevation-cold climate features are dominant in the Bayanbulak Basin. In the last 52 years, the records indicate that the winter lasts up to seven months, and the annual average temperature is  $-4.6^{\circ}\text{C}$ , with a minimum temperature of  $-48.0^{\circ}\text{C}$ . There are 139.3 snow-covered days, and the maximum depth of frozen ground is 7.50 m. Besides, an increasing trend in the annual precipitation and rainy days in the Bayanbulak Basin has occurred, with values of 9.5 mm/10 annual and 3.2 days/10 annual, respectively. However, the increase of annual precipitation in cold seasons is more than that of warm



**Figure 1** Map of the Bayanbulak region.

seasons.

The Kaidu River that originates from the Bayanbulak Basin is a river of great importance for the economy of Xinjiang. It flows into the largest inland freshwater lake in China, Bosten Lake. As the source of the Tarim River, Bosten Lake is a crucial aspect in a north-to-south water diversion project. It is responsible for delivering emergency water to the downstream ecosystem of the Tarim River. Therefore, the Bayanbulak Basin is a vital water source in the oasis of Southern Xinjiang.

## 1.2 Data Collection and Preprocessing

### 1.2.1 Field data collection

The snow depth data for this study was obtained from 36 stable observation stations (Figure 1) (set up by the Xinjiang Institute of Ecology and Geography of the Chinese Academy of Sciences) located in the Bayanbulak Basin in the central segment of Tianshan Mountains, Xinjiang. The snow depth monitoring devices are based on the principle of ultrasonic ranging to obtain automatic snow depth observations, and the data acquisition frequency is twice a day and the data acquisition times occur at 10:00 am and 20:00 pm (UTC+8).

The snow depth data on Feb 23, 2017 was used for this study. All snow depth data passed a quality control (3 $\sigma$  principle). Table 1 is the main statistics of the snow depth data. The data were evaluated with a Kolmogorov-Smirnov (*K-S*) test (Agostinelli 2003) and exhibited a normal distribution. The coefficient of variation (*CV*) was 49.40%, indicating a moderate variation.

### 1.2.2 Snow-related response variables

#### 1.2.2.1 Processing of MODIS snow cover data and virtual site arrangement

The Moderate Resolution Imaging Spectrometer (MODIS) snow cover data can be freely downloaded from the United States National Snow and Ice Data Center (NSIDC). It includes the eight-day composite snow cover data MOD10A2

and MYD10A2 of the Terra satellite (covering the area in the morning) and the Aqua satellite (covering the area in the afternoon) at 500 m resolution. Based on the longitude and latitude of the Bayanbulak Basin (Figure 1), the image with the track number h24v04 was sufficient to cover the entire Bayanbulak Basin. MOD10A2 and MYD10A2 on Feb 18, 2017 are selected for this study and the MODIS reprojection tool (MRT) is used for data extraction, reprojection, and cropping based on the extent the Bayanbulak Basin.

The establishment of virtual sites is the premise for obtaining a complete distribution of surface snow cover for the study area. In order to obtain snow cover for the Bayanbulak Basin without clouds obscuring the area, the MODIS snow cover data from multiple time periods was used to eliminate the impact of clouds. The snow cover for the corresponding pixels could be determined for a relatively short contiguous time. Figure 2a shows the snow cover results for the Bayanbulak Basin on Feb 18, 2017 after eliminating clouds.

Since there is spatial variation in the snow depth, we analyze the snow-free pixels in the rectified MODIS snow cover data and set up virtual observation stations, using the following rules 1) there is no snow in the eight neighbouring pixels of the virtual station; 2) the distance between the observation stations is no less than the smallest distance between the existing stations after adding the virtual stations to achieve an even distribution and density of observation stations.

#### 1.2.2.2 NDVI processing of MODIS data

The normalized difference vegetation index (*NDVI*) reflects the condition of vegetation cover, which can also be used to infer surface climatic information in vegetated areas without weather observation stations (Shen et al. 2015). This study uses MOD13Q1, the Terra *NDVI* data, with a spatial resolution of 250 m, a data cycle of 16 days, and a time series ranging from Jan 2001 to Dec 2016. These data are pre-processed with regard to coordinate transformation, extraction of study area,

**Table 1** Descriptive statistics of snow depth

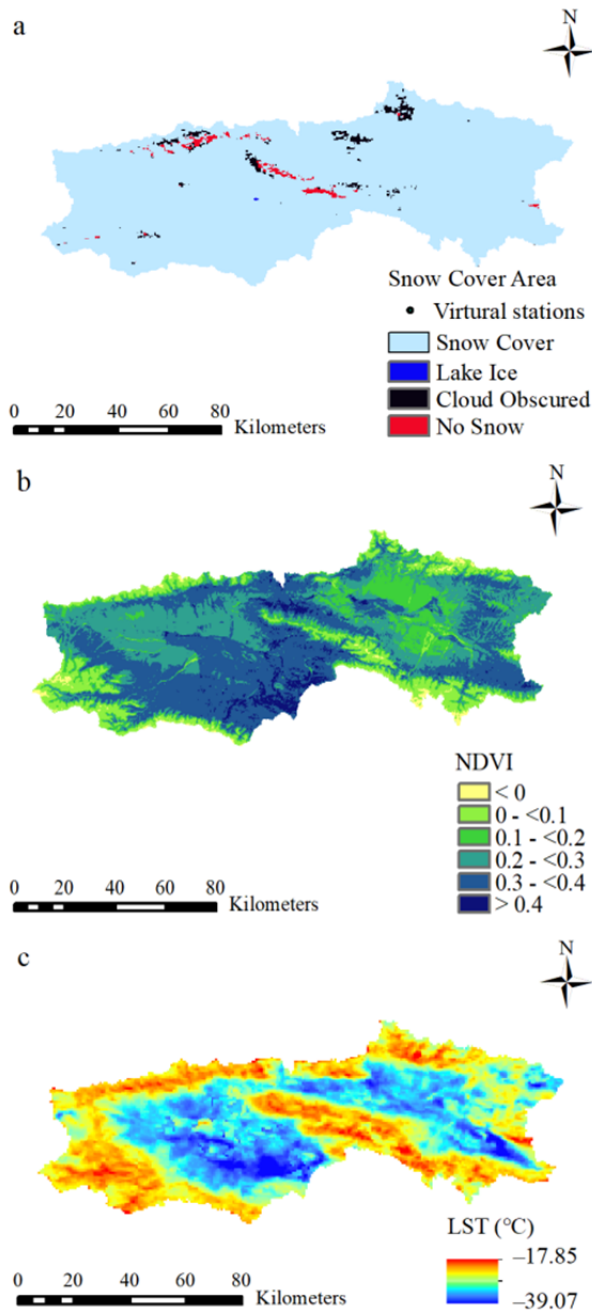
Observation number	Minimum (m)	Maximum (m)	Mean (m)	Std (m)	Kurtosis	Skewness	<i>K-S</i> Value	<i>CV</i> (%)
36	0.00	0.29	0.13	0.08	2.10	0.53	0.07	49.40

**Notes:** \* *Std* refers to Standard Deviation; *K-S*, Kolmogorov-Smirnov; *CV*, coefficient of variation.

etc. In this study, we use the 16-year average of the data to reflect the degree of vegetation cover in the study area (Figure 2b).

### 1.2.2.3 MODIS surface temperature data and processing

Land surface temperature (*LST*) is an important parameter for the mutual interaction and energy flux exchange in global and regional



**Figure 2** Snow cover distribution in the Bayanbulak region (a); spatial distribution of annual *NDVI* (b); spatial distribution of *LST* (c).

land-air interfaces. When the heat of snow layer increases (although the overall temperature of the snow layer rises), the surface temperature is slightly lower for a thick snow layer than for a thin snow layer because the temperature gradient of a thick snow layer is lower (Ge et al. 2010). The global daily surface data MYD11A1 of Aqua is used in this study. The data is based on relatively mature split-window algorithms to calculate the surface temperature (°C). The dates of data obtained in this study are Feb 22, Feb 23, and Feb 24, 2017 respectively. The maximum values of the surface temperature in the corresponding pixels are synthesized for reduce the impact of value deficiency, thus closely reflecting the *LST* conditions on Feb 23, 2017 (Figure 2c).

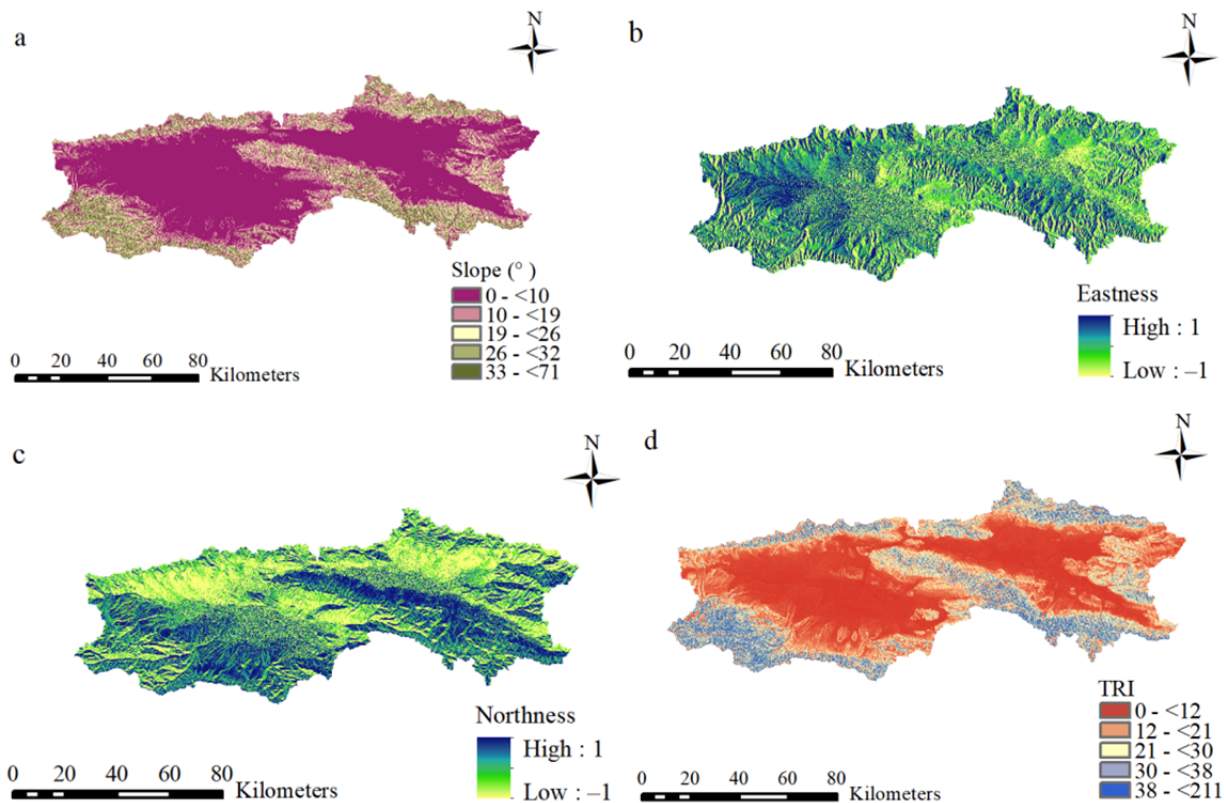
### 1.2.3 Derivation of terrain-related variables

Digital elevation model (DEM) data with a spatial resolution of 90 m are sourced from the United States EROS Data Center. After data alignment and re-projection, the topographic variables such as *slope* (Figure 3a) are derived from the 90-m DEM. Additionally, auxiliary variables such as *eastness* (Figure 3b), *northness* (Figure 3c), and the topographic ruggedness index (*TRI*) (Figure 3d) are also extracted from the DEM data. *Northness* is defined as the cosine of the aspect and has values ranging from -1 to 1; a value of -1 represents a slope facing directly south and a value of 1 indicates a slope facing directly north. *Eastness* (Figure 3b) is defined as the sine of the aspect with values ranging from -1 to 1; -1 represents an east-facing slope and 1 indicates a west-facing slope. The *TRI* is defined as the standard deviation of the elevation in a 270 m × 270 m (3×3 pixels) window, high *TRI* values are indicative of areas with a large fluctuation in elevation (Harshburger et al. 2010).

## 1.3 Methodology

### 1.3.1 Regression kriging

If the snow depth is related to the explanatory variables, multiple (or simple) regression relationships between snow depth and explanatory variables can be established, thus obtaining  $\hat{m}(x_0)$ , representing the deterministic tendency and the residual  $e(x_0)$ , which represents



**Figure 3** Physiographic variables used in the regression analysis: (a) *slope* (°), (b) *Eastness* , (c) *Northness*, (d) topographic ruggedness index (*TRI*).

the random portion. Interpolation is performed on the residual by OK. Finally, the two parts are summed to obtain the interpolation result of RK. This process can be expressed as:

$$Z(x_0) = \hat{m}(x_0) + \hat{e}(x_0) \quad (1)$$

where  $Z(x)$  is the estimation value when the snow depth is at the point of  $x_0$  and  $\hat{e}(x_0)$  is the residual obtained by interpolation of OK.

### 1.3.2 Geographically weighted regression kriging

GWRK is an interpolation method in which the global regression in RK is replaced by the local regression of GWR. In the geographically weighted model, the regression coefficient of the snow depth at  $x_0$  for the explanatory variable is no longer a constant obtained through global information of the explanatory variable and a multiple linear regression (MLR) but instead is an estimated coefficient of a locally weighted regression using the neighboring information of the explanatory variable. The GWR model at position  $x_0$  is

presented as:

$$y(x_0) = \beta_0(x_0) + \sum_{k=1}^p \beta_k(x_0) q_k(x_0) \quad (2)$$

where  $y(x_0)$  is the snow depth at position  $x_0$ ;  $q_k(x_0)$  is the  $k^{\text{th}}$  explanatory variable at  $x_0$ ;  $\beta_0(x_0)$  is the intercept;  $\beta_k(x_0)$  is the coefficient of the  $k^{\text{th}}$  explanatory variable;  $p$  is the number of explanatory variables.

A Gaussian function is used as the spatial weighting function, as shown in equation (3):

$$W_i(x_0) = \exp\left(-\frac{1}{2} \left(\frac{d_i}{r}\right)^2\right) \quad (3)$$

where  $d_i$  is the distance from the  $i^{\text{th}}$  observation station to the point  $x_0$ ;  $r$  is a bandwidth parameter;  $W_i(x_0)$  is the weight function. For different station densities, an adaptive bandwidth selection strategy and the corrected Akaike information criterion (AIC) are used to determine the optimal bandwidth  $r$ . After the optimal bandwidth and weighting function are determined for each point to be estimated, the weighting

matrix can be obtained and the local regression coefficient can be estimated.

The GWRK is developed from the GWR. An OK interpolation method is conducted on the residual (obtained from the GWR fitting) and using the GWR fitting trend, it is represented as:

$$y(x_0) = \beta_0(x_0) + \sum_{k=1}^p \beta_k(x_0)q_k(x_0) + \varepsilon(x_0) \quad (4)$$

where  $\varepsilon(x_0)$  is the residual of the GWR fitting at position  $x_0$ . The interpolation is conducted with the OK method.

### 1.4 Method evaluation

A 3-fold cross-validation is used to compare the accuracy of the snow depth based on the RK and GWRK methods. For the cross-validation, the measured snow depth data were equally divided into three groups and the actual snow depth data of two groups was used as the training set. The snow depth is then estimated by the RK and GWRK methods. The last group serves as a validation set to evaluate the accuracy, including mean estimation error (*MEE*), mean absolute estimation error (*MAEE*), and root mean square error (*RMSE*).

As mentioned above, the snow depth estimation was repeated three times and an average of the results was used for the three accuracy tests. Finally, the results of the estimated snow depth were obtained.

$$MEE = \frac{\sum_{i=1}^n [\hat{s}(s_i) - s(s_i)]}{n} \quad (5)$$

$$MAEE = \frac{\sum_{i=1}^n |\hat{s}(s_i) - s(s_i)|}{n} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\hat{s}(s_i) - s(s_i)]^2}{n}} \quad (7)$$

where  $\hat{s}(s_i)$  is the estimated snow depth at

position  $s_i$ ,  $s(s_i)$  is the actual snow depth at position  $s_i$ , and  $n$  is the number of snow depth observation stations.

## 2 Results

### 2.1 Descriptive statistics

#### 2.1.1 Correlation analysis for snow depth and auxiliary variables

The target variable is snow depth and the auxiliary variables include seven variables: height, slope, eastness, northness, *TRI*, *NDVI*, and *LST*. [Table 2](#) indicate that *TRI*, height, and *LST* have a significant impact on the spatial distribution of snow depth. The snow depth in the Bayanbulak Basin is significantly positively correlated with *TRI* ( $r=0.31$ ;  $p<0.01$ ) and height ( $r=0.11$ ;  $p<0.05$ ), this indicates that snow depth increases with increasing height and *TRI*. In addition, snow depth is significantly negatively correlated with *LST* ( $r=-0.09$ ;  $p<0.05$ ), which shows that snow depth is also influenced by the angle and intensity of solar radiation. Moreover, snow depth is also non-significantly positively correlated with eastness and non-significantly negatively correlated with the northness. This indicates that a southeast slope is more favorable for the accumulation of snow and may be connected with other explanatory variables such as wind speed and wind direction.

#### 2.1.2 Stepwise regression for snow depth and explanatory variables

A stepwise regression is used to establish the regression relationship between snow depth and auxiliary variables. The variance inflation factor (*VIF*) is used to test the collinearity among the auxiliary variables; if  $VIF>10$ , the corresponding auxiliary variables should be removed to avoid the collinearity of the variables ([Oliver et al. 2014](#)).

For the variables of height, *LST*, and *TRI*, the *VIF* is smaller than 3.1, indicating that no

**Table 2** Correlation coefficient between topographical factors and snow depth

	Height	TRI	NDVI	Eastness	Northness	Slope	LST
Snow Depth	0.11*	0.31**	-0.04	0.01	-0.08	0.03	-0.09*

**Notes:** \* $p<0.05$ ; \*\* $p<0.01$ ; topographic ruggedness index (*TRI*); normalized difference vegetation index (*NDVI*); land surface temperature (*LST*).

collinearity exists among the variables. Therefore, these three variables can be used in the regression model.

## 2.2 Estimation of snow depth

### 2.2.1 Variance function fitting of regression residuals

Snow depth (SD) is used as the target variable and height, *TRI*, and *LST* are used as explanatory variables for the MLR and GWR. The MLR model is:

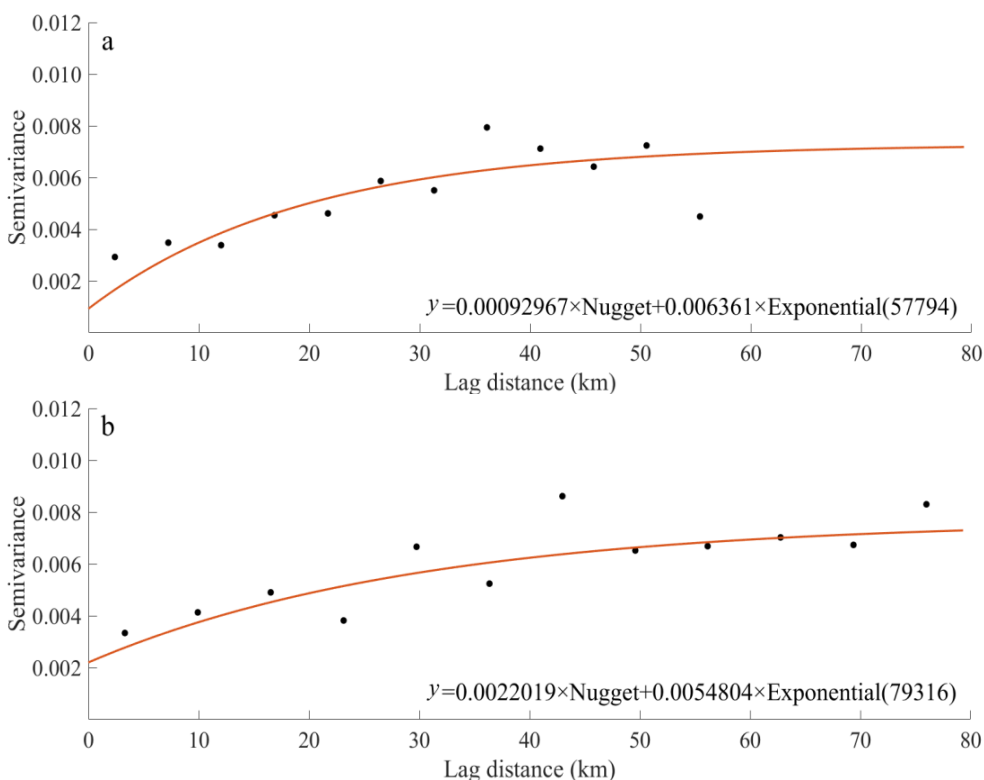
$$SD = -1.16 \times 10^{-4} \times LST + 2.21 \times 10^{-3} \times TRI - 9.75 \times 10^{-5} \times Height + 0.377 \quad (8)$$

The MLR model shows an adjusted *R*-square of 0.44 by 3-fold cross-validation. Meanwhile, the

GWR model to reveal local features could take advantage of spatial heterogeneity of regression coefficients estimated by *TRI*, *LST* and *height* as well as snow depth. The GWR model with the least AIC shows an adjusted *R*-square of 0.49 by 3-fold cross-validation.

After MLR fitting, the range of the MLR residuals is  $-0.100-0.161$ , the residual kurtosis is 2.124, and the residual skewness is 0.598. After GWR fitting, the residual range is  $-0.285-0.131$ , the residual kurtosis is 5.411, and the residual skewness is  $-0.91$ . Both residuals have passed the *K-S* test (significance level:  $\alpha=0.05$ ;  $p_{mlr}=0.629$ ;  $p_{gwr}=0.608$ ) and are therefore suitable for statistical interpolation.

An exponential variance function is used to conduct the fitting for the residuals (Massey 2012).



**Figure 4** Experimental variogram and fitted models of residuals from (a) multiple linear regression (MLR) and from (b) geographically weighted regression (GWR).

**Table 3** Variogram models and parameters fitted by using residuals from MLR and GWR models

Model	Type	Nugget ( $C_0$ )	Partial sill (C)	Sill ( $C_0+C$ )	Proportion [ $C_0/(C_0+C)$ , %]	Range (m)	Coefficient of determination ( $R^2$ )
MLR residuals	Exponential	0.0009	0.0064	0.0073	12.33	4816.16	0.71
GWR residuals	Exponential	0.0023	0.0062	0.0085	28.06	7496.08	0.77



Figure 4 and Table 3 show the residual fitting results.

The data in Table 3 show that the variance function proportion of the MLR residuals and the GWR residuals  $C_0/(C_0+C)$  are both less than 30%, indicating that the regression residuals have strong spatial auto-correlation and that OK can be used for interpolation (Yang et al. 2015). It also reflects that the snow depth variance is less influenced by random factors and is mainly controlled by structure factors. In addition, the data confirm the feasibility of using the explanatory variables for interpolation. The use of the exponential variance function for fitting the MLR residuals and the GWR residuals provides a deterministic coefficient of more than 70% and the fitting residuals all are small with a high coefficient of determination.

**2.2.2 Analysis of spatial distribution of snow depth**

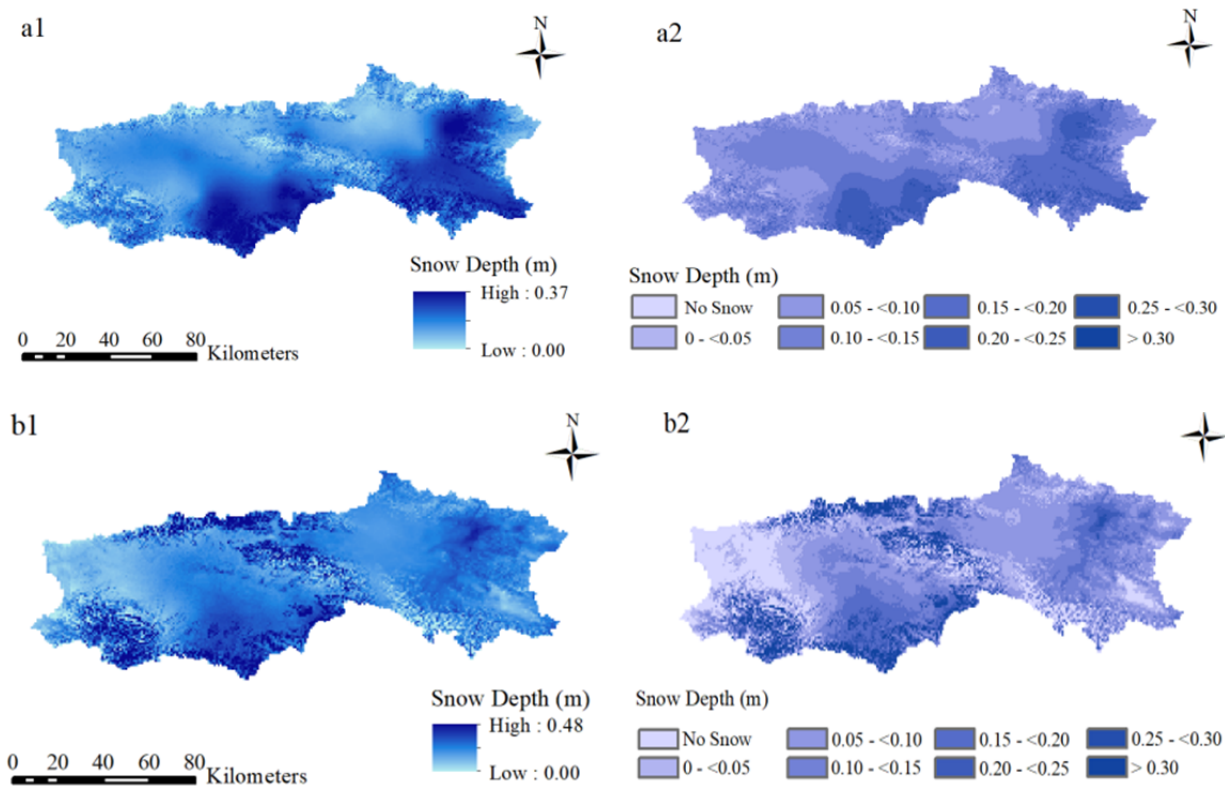
Based on the above results, the spatial distribution of snow depth for the RK and the GWRK can be obtained (Figure 5).

GWRK and RK were used in this study, and

both belong to the hybrid interpolation method. Their trend terms are fitted with GWR and MLR methods respectively, and the fitted residual is interpolated using the OK method. To determine the final snow depth distribution, the sum of the trend term and the residual term was calculated.

Figure 5 shows that the snow depth range of RK is 0.00–0.37 m and the snow depth range of GWRK is 0.00–0.48 m. The maps show that the snow depth distribution obtained by RK presents a clear transition from high to medium and to low values while the snow depth distribution obtained by GWRK is more detailed. The high values (>0.30 m) of the GWRK are broader distributed than the corresponding values of the RK and the medium values (0.10–0.30 m) are narrower than in the RK; the low values (0.00–0.10 m) are wider in the GWRK than in the RK. In addition, there are significantly more small regions with different snow depth in the GWRK than in the RK.

The study area is divided into five subareas (A to E) based on elevation (the elevations are 2386.00–2632.00 m, 2632.00–2929.00 m, 2929.00–3269.00 m, 3269.00–3644.00 m, and



**Figure 5** Spatial distribution of estimated snow depth (m) interpolated by using regression kriging (RK) (a1) and geographically weighted regression kriging (GWRK) (b1); snow cover distribution for different ranges of snow depth from RK (a2) and GWRK (b2).

3644.00–4518.00 m respectively). The snow proportions for the subareas are determined based on the RK and GWRK. According to Figure 6 and Table 4, the snow with 45.00%–46.00% of the study area in the elevation range of 2632.00–3269.00 m. In addition, the estimated maximum value of snow depth was obtained in this region, and the snow depth in the study area peaked in the intermediate elevation while low in the other elevation zones, which is in agreement with the actual observation.

### 2.3 Evaluation of accuracy

Table 5 shows the statistical results of the snow depth estimated from RK and GWRK. It can be seen that the GWRK estimation is better than the RK estimation. The accuracy analysis using the validation points shows that the average errors of snow depth obtained by RK and GWRK are both negative and are -0.056 m and -0.081 m respectively. This indicates that the overall estimation of snow depth in Bayanbulak Basin is slightly higher than the validation data. Therefore,

GWRK is better than RK for obtaining snow depth interpolation result when environmental variables are considered.

## 3 Discussion

### 3.1 Factors influencing snow depth interpolation

It is difficult to estimate the spatial distribution of snow depth for a limited number of snow depth stations, especially in areas with complex terrain. An effective and practical method for the estimation of regional snow depth distribution, especially for areas with complex terrain is the use of interpolation of snow depth data obtained by observation stations and remote sensing data because of the wide distribution of snow (Anderton et al. 2004; Harshburger et al. 2010; Collados-Lara et al. 2017). We comprehensively analyzed the correlation between terrain variables (e.g., height, slope, northness, eastness) and environmental variables (NDVI, LST)

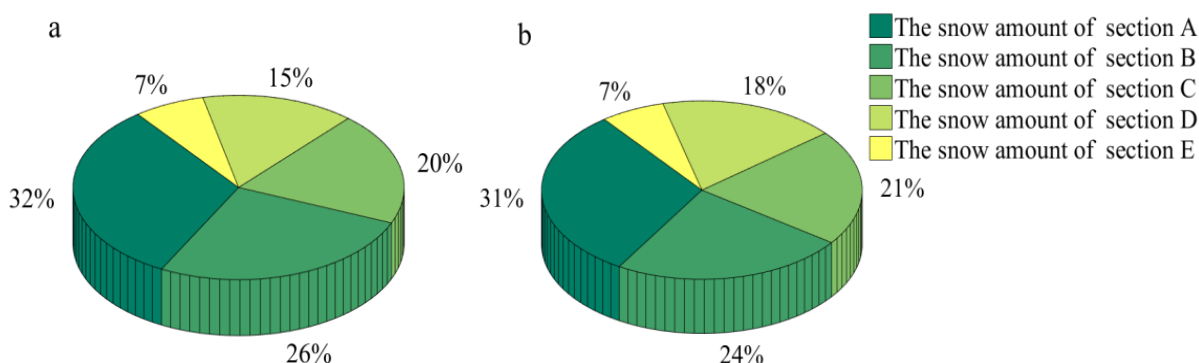


Figure 6 Snow proportions in the subareas of the region from regression kriging (RK) (a) and geographically weighted regression kriging (GWRK) (b).

Table 4 Minimum, maximum, mean, and standard deviation (Std) of snow depth in the subareas of the study area

Section	Minimum (m)		Maximum (m)		Mean (m)		Std (m)	
	RK	GWRK	RK	GWRK	RK	GWRK	RK	GWRK
A	0.047	0	0.272	0.407	0.131	0.125	0.047	0.061
B	0.045	0	0.300	0.489	0.142	0.136	0.053	0.095
C	0	0	0.331	0.525	0.129	0.125	0.051	0.113
D	0	0	0.374	0.508	0.109	0.111	0.051	0.116
E	0	0	0.324	0.592	0.097	0.083	0.054	0.103

Table 5 Comparison of the precision indices for the two approaches (RK and GWRK)

Method	R <sup>2</sup>	MEE	RMSE	MAEE
		(m)		
RK	0.50	-0.056	0.077	0.285
GWRK	0.55	-0.081	0.102	0.188

Notes: MEE, mean estimation error; RMSE, root mean square error; MAEE, absolute estimation error.

and find that the spatial variance of snow depth is simultaneously influenced by structural factors (terrain variables) and stochastic factors (environmental variables). The terrain factors influence the distribution and changes in snow depth as a result of *solar radiation, temperature*, and other meteorological conditions, while the environmental variables are used as a characteristic variable significantly correlated with snow depth and reflecting the snow depth distribution.

Based on correlation and *VIF*, we used height, *TRI*, and *LST* as the explanatory variables for snow depth spatial interpolation. This shows that the thick snow layer and lower *LST* would occur with the increase of height and *TRI*. This conclusion conforms to actual conditions. Therefore, terrain variables and environmental variables are used as explanatory variables and by referring to limited snow depth information obtained from observation stations, a reliable spatial distribution of snow depth in the Bayanbulak Basin can be obtained using GWRK.

### 3.2 Limitations of snow depth interpolation assessments

The kriging methods used in this study are representative of present spatial interpolation research methods (Huang et al. 2015). Although kriging methods have the advantages of making full use of structural features of samples and showing the estimation errors of the interpolation results, they fail to consider environmental variables which are closely related to snow depth. In addition, the interpolation accuracy is also influenced by sampling density and sampling mode. However, RK cannot only be used to develop a global regression model with a correlation between snow depth and environmental variables but an OK interpolation can also be conducted using sample information on spatial auto-correlation of the residuals (Somarathna et al. 2016). A GWR model relies on flexible changes of the weight function, thus overcoming the imbalanced importance of explanatory variables in different spatial positions in the regression model (Jie et al. 2011). It can capture local spatial structural factors and stochastic factors and, combined with OK can make full use of the residual information to improve the

estimation accuracy. This study results also prove that the snow depth results are more detailed for GWRK than for RK.

### 3.3 Limitations of remote sensing and meteorological observation data

Eight-day composite snow cover data (MOD10A2 and MYD10A2) was adopted to reflect the largest range of snow cover within eight days to compensate for a low sampling density. The image pixels in these products were determined by snow cover on certain days, or, over the entire eight days (Huang et al. 2007). Therefore, After eliminating the impact of the clouds by integrating MOD10A2 and MYD10A2 snow cover products on Feb 18, 2017 into a SCA data, the virtual sites that exist in bare lands were established and stabilized. When combing the slope measurements, it was found that the bare lands were mainly distributed between the slope of 22.5° and 67.8°. Because these field in mountain receive more solar radiation than flatland, therefore, non-conductive to snow accumulation, the introduced virtual stations can reduce the smoothness of the interpolation results from GWRK and RK. This allows the spatial variation of snow depth and the characteristics of the snow depth distribution to be accurately reflected. However, the SCA data used in the study was gathered from the cloud-eliminating method for adjacent pixels, and came from only one source, which used a simple method. Therefore, future studies should use multiple sources when collecting remote sensing data for the collaborative elimination of clouds (Molotch et al. 2004; Pardo-Igúzquiza et al. 2017). This would ensure the improvement of cloud-elimination efficiency and snow recognition rate.

There are various environmental and terrain factors that influence the accumulation and variation of snow depth. This study uses GWRK and RK with *height, TRI*, and *LST* as explanatory variables. However, there are other terrain and environmental variables (e.g. *wind speed, wind direction*, etc.) that are closely related to snow depth and are not included in this study. In addition, certain other factors are ignored in this study (e.g. *land cover*) (Vos et al. 2005) and may also influence the interpolation results. Therefore, the interpolation accuracy may be further

optimized if the mutual reaction mechanism between snow depth and environmental/terrain factors could be analyzed in more detail; therefore, additional suitable auxiliary variables should be selected for future spatial distribution analyzes of snow depth.

#### 4 Conclusions

This study tries to address certain limitations of previous studies on the correlation between environmental variables and snow depth. GWRK, a local approach that emphasizes the spatial non-stationarity and variability combined with spatial auto-correlation, is used to estimate the snow depth in the Bayanbulak Basin.

Two geostatistical hybrid approaches (GWRK and RK) are compared in this study for snow depth estimation and height, *TRI*, and *LST* are used as

explanatory variables. An estimation of snow depth with a spatial resolution of 1 km is obtained. This research indicates that the snow depth results are more detailed and better for the GWRK than the RK because GWRK is a spatial interpolation that considers a regression relationship, spatial non-stationarity, and residual auto-correlation simultaneously, thus providing more detail than RK.

#### Acknowledgments

This study is supported by Projects of International Cooperation and Exchanges NSFC (grant: 41361140361), the Special fund project of Chinese Academy of Sciences (grant: Y371164001) and the key deployment project of Chinese Academy of Sciences (Grant No. KZZD-EW-12-2, KZZD-EW-12-3). The authors thank the reviewers for helpful comments during the peer review process.

#### References

- Agostinelli C (2003) Robust stepwise regression. *Quality Engineering* 48(5): 557-558. <https://doi.org/10.1080/02664760220136168>
- Anderton SP, White SD, Alvera B (2004) Evaluation of spatial variability in snow water equivalent for a high mountain catchment. *Hydrological Processes* 18(3): 435-453. <https://doi.org/10.1002/hyp.1319>
- Bocchiola D, Janetti EB, Gorni E, et al. (2008) Regional evaluation of three day snow depth for avalanche hazard mapping in Switzerland. *Natural Hazards and Earth System Sciences* 8(4): 685-705. <https://doi.org/10.5194/nhess-8-685-2008>
- Bocchiola D, Groppelli B (2010) Spatial estimation of snow water equivalent at different dates within the Adamello Park of Italy. *Cold Regions Science and Technology* 63(3): 97-109. <https://doi.org/10.1016/j.coldregions.2010.06.001>
- Che T, Dai L, Wang J, et al. (2012) Estimation of snow depth and snow water equivalent distribution using airborne microwave radiometry in the Binggou Watershed, the upper reaches of the Heihe River basin. *International Journal of Applied Earth Observation and Geoinformation* 17: 23-32. <https://doi.org/10.1016/j.jag.2011.10.014>
- Collados-Lara AJ, Pardo-Iguzquiza E, Pulido-Velazquez D, et al. (2017) Spatio-temporal estimation of snow depth using point data from snow stakes, digital terrain models and satellite data. *Hydrological Processes* 31(10): 1966-1982. <https://doi.org/10.1002/hyp.11165>
- Dai L, Che T, Ding Y (2015) Inter-Calibrating SMMR, SSM/I and SSMI/S Data to Improve the Consistency of Snow-Depth Products in China. *Remote Sensing* 7(6): 7212-7230. <https://doi.org/10.3390/rs70607212>
- Erxleben J, Elder K, Davis RE (2002) Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrological Processes* 16(18): 3627-3649. <https://doi.org/10.1002/hyp.1239>
- Ge Y, Gong G (2010) Land surface insulation response to snow depth variability. *Journal of Geophysical Research Atmospheres* 115(D8): 462-474. <https://doi.org/10.1029/2009JD012798>
- Huang XD, Zhang XT, Xia LI, Liang TG (2007) Accuracy analysis for modis snow products of mod10a1 and mod10a2 in northern Xinjiang area. *Journal of Glaciology and Geocryology* 29(5): 722-729. <https://doi.org/10.3969/j.issn.1000-0240.2007.05.008>
- Harshburger BJ, Humes KS, Walden VP, et al. (2010) Spatial interpolation of snow water equivalency using surface observations and remotely sensed images of snow-covered area. *Hydrological Processes* 24(10): 1285-1295. <https://doi.org/10.1002/hyp.7590>
- Huang CL, Wang HW, Hou JL (2015) Estimating spatial distribution of daily snow depth with kriging methods: combination of MODIS snow cover area data and ground-based observations. *The Cryosphere Discussions* 9(5): 4997-5020. <https://doi.org/10.5194/tcd-9-4997-2015>
- Jie Lin, Robert Cromley, Chuanrong Zhang (2011). Using geographically weighted regression to solve the areal interpolation problem. *Annals of GIS* 17(1): 1-14. <https://doi.org/10.1080/19475683.2010.540258>
- Lopez-Moreno JI, Nogués-Bravo D (2006) Interpolating local snow depth data: an evaluation of methods. *Hydrological Processes* 20(10): 2217-2232. <https://doi.org/10.1002/hyp.6199>
- Lu H, Wei WS, Liu MZ, et al. (2014) Observations and modeling of incoming longwave radiation to snow beneath forest canopies in the west Tianshan Mountains, China. *Journal of Mountain Science* 11(5): 1138-1153. <https://doi.org/10.1007/s11629-013-2868-1>
- Liang J, Liu X, Huang K, et al. (2015) Improved snow depth retrieval by integrating microwave brightness temperature and visible/infrared reflectance. *Remote Sensing of Environment* 156: 500-509. <https://doi.org/10.1016/j.rse.2014.10.016>

- Lu H, Wei WS, Liu MZ, et al. (2015) Variations in seasonal snow surface energy exchange during a snowmelt period: an example from the Tianshan Mountains, China. *Meteorological Applications* 23(1), 14-25. <https://doi.org/10.1002/met.1511>
- Marshall SE, Roads JO, Glatzmaier GA (1994) Snow hydrology in a general circulation model. *Journal of Climate* 7(8): 1251-1269. [https://doi.org/10.1175/1520-0442\(1994\)007<1251:SHIAGC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1251:SHIAGC>2.0.CO;2)
- Molotch NP, Fassnacht SR, Bales RC, Helfrich SR (2004) Estimating the distribution of snow water equivalent and snow extent beneath cloud cover in the salt-Verde River basin, Arizona. *Hydrological Processes*, 18(9), 1595-1611. <https://doi.org/10.1002/hyp.1408>
- Massey JF (2012) The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* 46(253): 68-78. <https://doi.org/10.1080/01621459.1951.10500769>
- McCreight JL, Small EE, Larson KM (2014) Snow depth, density, and SWE estimates derived from GPS reflection data: Validation in the western U. S.. *Water Resources Research* 50(8): 6892-6909. <https://doi.org/10.1002/2014WR015561>
- Oliver MA, Webster R (2014) A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113(2): 56-69. <https://doi.org/10.1016/j.catena.2013.09.006>
- Pardo-Igúzquiza E, Collados-Lara AJ, Pulido-Velázquez D (2017) Estimation of the spatiotemporal dynamics of snow covered area by using cellular automata models. *Journal of Hydrology*, 550, 230-238. <https://doi.org/10.1016/j.jhydrol.2017.04.058>
- Somarathna PDSN, Malone BP, Minasny B (2016). Mapping soil organic carbon content over new south wales, australia using local regression kriging. *Geoderma Regional* 7(1): 38-48. <https://doi.org/10.1016/j.geodrs.2015.12.002>
- Shen X, Liu B, Zhou D (2015) Using GIMMS NDVI time series to estimate the impacts of grassland vegetation cover on surface air temperatures in the temperate grassland region of China. *Remote Sensing Letters* 7(3):229-238. <https://doi.org/10.1080/2150704X.2015.1128131>
- Tarnocai C, Canadell JG, Schuur EA, et al. (2009) Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles* 23(2): 1-11. <https://doi.org/10.1029/2008GB003327>
- Vos BD, Meirvenne MV, Quataert P, et al. (2005) Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal* 69(2): 500-510. <https://doi.org/10.2136/sssaj2005.0500>
- Wang K, Zhang C, Li W (2012) Comparison of Geographically Weighted Regression and Regression Kriging for Estimating the Spatial Distribution of Soil Organic Matter. *Gis Science and Remote Sensing* 49(6): 915-932. <https://doi.org/10.2747/1548-1603.49.6.915>
- Wasige JE, Groen TA, Smaling EM, et al. (2013) Monitoring basin-scale land cover changes in Kagera Basin of Lake Victoria using ancillary data and remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 21(1): 32-42. <https://doi.org/10.1016/j.jag.2012.08.005>
- Yang SH, Zhang HT, Guo L, et al. (2015) Spatial interpolation of soil organic matter using regression Kriging and geographically weighted regression Kriging. *Chinese Journal of Applied Ecology* 26(6): 1649-1536. <https://doi.org/10.13287/j.1001-9332.20150331.023>
- Zhang T (2005) Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics* 43(4): 589-590. <https://doi.org/10.1029/2004RG000157>
- Zhou X, Xie H, Hendrickx JMH (2005) Statistical evaluation of modis snow cover products with constraints from streamflow and snotel measurement. *Remote Sensing of Environment*, 94(2), 214-231. <https://doi.org/10.1016/j.rse.2004.10.007>
- Zimov S, Schuur EA, Chapin FS (2006) Permafrost and the Global Carbon Budget. *Science* 312(5780): 1612-1613. <https://doi.org/10.1126/science.1128908>