Geostatistical approaches to refinement of digital elevation data

Jingxiong ZHANG*a, Tao ZHUb, Yunwei TANgb and Wangle ZHANGa

aSchool of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, China; bInstitute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

(Received 22 June 2014; final version received 23 September 2014)

Data refinement refers to the processes by which a dataset’s resolution, in particular, the spatial one, is refined, and is thus synonymous to spatial downscaling. Spatial resolution indicates measurement scale and can be seen as an index for regular data support. As a type of change of scale, data refinement is useful for many scenarios where spatial scales of existing data, desired analyses, or specific applications need to be made commensurate and refined. As spatial data are related to certain data support, they can be conceived of as support-specific realizations of random fields, suggesting that multivariate geostatistics should be explored for refining datasets from their coarser-resolution versions to the finer-resolution ones. In this paper, geostatistical methods for downscaling are described, and were implemented using GTOPO30 data and sampled Shuttle Radar Topography Mission data at a site in northwest China, with the latter's majority grid cells used as surrogate reference data. It was found that proper structural modeling is important for achieving increased accuracy in data refinement; here, structural modeling can be done through proper decomposition of elevation fields into trends and residuals and thereafter. It was confirmed that effects of semantic differences on data refinement can be reduced through properly estimating and incorporating biases in local means.

Keywords: refinement; elevation data; data support; variogram deconvolution; semantic differences

1. Introduction

Multi-source geospatial data have proliferated, and their spatial scales (which are collectively defined by sampling density and data supports) are often different due to differences in data density and data supports (e.g. points, contour lines, or regular grids). Thus, it is important to understand the scales inherent to multi-source datasets and to perform change of scale (i.e. scaling) when analyzing, comparing, and transforming such data. Geospatial data can be considered to be the sums of deterministic means (or trends) and stochastic residuals, with both components being support-specific, suggesting the natural link between theory of regionalized variables and change of data support, and an important role of geostatistics for spatial data refinement, particularly in Refs. (1, 2).

Resolution refinement can be accomplished through interpolation, which is actually a kind of prediction and, in geostatistics, can be catered for by kriging. A number of authors, such as Asl and Marcotte (3), explored alternative kriging techniques for interpolation, and emphasized the relative advantages of flexible use of secondary data in interpolation through geostatistical procedures, such as kriging with external drifts vs. cokriging. A geostatistical framework for linking geographically aggregated data from different sources was proposed by Gotway and Young (4), by which upscaling (aggregation), downscaling (disaggregation), or side scaling (a term used to refer to the prediction of values on one set of spatial units from data on another set of overlapping spatial units) may be performed. Yoo and Kyriakidis (5) described the area-to-point kriging, which satisfies several critical issues in downscaling: the coherence of predictions, the explicit consideration of support differences, and the assessment of uncertainty regarding the point predictions. Yoo and Kyriakidis (6) developed a geostatistical approach, namely area-to-point kriging with external drift, which can take into account spatial dependence and spatial heteroskedasticity (if existing) for applications in hedonic price modeling where exhaustive area-averaged housing price data are available in addition to a subset of individual housing price data.

Since this paper deals with terrain elevation data, some description of elevation data is necessary. Terrain elevation data are acquired from various sources including topographic maps and remote measurements, which are often stored as digital elevation models (DEMs). Generally, there are three types of elevation data: spot heights and point DEMs (with point support), grid DEMs where cell values are convolutions which are often not flat over cells (block or area support in two-dimensional domains), and digitized isolines. As elaborated in Ref. (7), SRTM (Shuttle Radar Topography Mission, http://srtm.usgs.gov/index.php) and GTOPO30 (a global DEM data-set with a horizontal grid spacing of 30 arc seconds, i.e. about 1 km, http://eros.usgs.gov/#!/Find_Data/Products_and_Data_Available/gtopo30_info) DEMs, which were used in the experiments in this paper, have their own error margins, effective horizontal resolution, and

*Corresponding author. Email: jxzhang@whu.edu.cn

© 2014 Wuhan University
semantics, which often vary spatially. Semantics refers to meanings of elevation underlying the elevation values, and semantic differences lead to differences of elevation values at the same locations with otherwise identical value: (1) SRTM data are radar-derived elevation data, and may not refer to pure ground surface, even if the terrain surface corresponding to the grid cells of SRTM is of bare ground or non-ground (e.g. covered with man-made structures); (2) as GTOPO30 data are derived from datasets, such as Digital Terrain Elevation Data (DTED), Digital Chart of the World (DCW), and others, their semantics tend to be more complicated than those of SRTM data; for instance, DTED with a horizontal grid spacing of 3 arc-seconds (approximately 90 m) was used as the source for most of Eurasia and large parts of Africa and the Americas, and aggregation of the DTED data to GTOPO30 data was typically done by selecting a representative elevation value from 10 × 10 fine-resolution cells covering the coarse-resolution cell (for Eurasia, this representative value is the median, leading to different aggregated elevation values, as aggregation often assumes mean statistic); (3) in addition to intra-data-set semantic inhomogeneity, there are also inter-data-set semantic differences due to different meanings attached to or derivations of the values of elevation over the same locations; and (4) scale mismatch leads to differences in elevation values as elevation is scale-dependent, and should be differentiated from those due to semantic inhomogeneity.

To handle semantic inconsistency in SRTM and GTOPO30 DEMs for latter’s refinement, it would be helpful to undertake two sets of experiments in parallel: one with a data-set that suffers from semantic inconsistency (e.g. the SRTM and GTOPO30 datasets), the other with a simulated data-set (e.g. the SRTM and GTOPO30 datasets), the latter being generated by coarsening the former through simple averaging thus avoiding semantic inconsistency). The data analyses with the two datasets can be used to examine the effects of semantic differences on data refinement, and the effectiveness of methods devised to offset such effects, as will be elaborated later. Contributions of this paper are twofold: (1) assessment of effects of semantic differences between elevation datasets on data refinement, and their treatment in kriging-based data refinement, and (2) enhanced performance of geostatistical data refinement by decomposition-based structural modeling. Geostatistical methods for data refinement will be described in Section 2, while empirical results using SRTM and GTOPO30 datasets will be reported and discussed in Section 3, followed by some concluding remarks.

2. Models and methods

2.1. Variograms over point and area supports

Consider a random field (RF) \( \{Z(x), x \in \mathcal{A}\} \) where \( Z \) may represent elevation, with \( x \) referring to a location in a spatial domain \( \mathcal{A} \); at a particular location \( x \), \( Z(x) \) is a random variable. Without causing ambiguity or confusion, we may denote the RF \( \{Z(x), x \in \mathcal{A}\} \) (i.e. an ensemble of random variables \( Z(x_1), \ldots, Z(x_n) \), if \( n \) locations are used to discretize the domain \( \mathcal{A} \)) by \( Z \). According to theory of regionalized variables, we can write:

\[
Z(x) = m_x + R(x)
\]

where \( m_x \) and \( R(x) \) stand for the deterministic component and residual of \( Z \) at \( x \), respectively. Although a stationary mean \( m_x \) may be assumed over the problem domain \( \mathcal{A} \), defining and estimating local means \( m_\mathcal{A}(x) \) \((x \in \mathcal{A})\) properly is central to geostatistical analysis of RF \( Z \) and prediction of \( Z(x) \) at unsampled locations, especially in the presence of trends (i.e. spatially varying means), as will be elaborated later. We can rewrite Equation (1a):

\[
Z(x) = m_x(x) + R(x)
\]

where \( m_\mathcal{A}(x) \) and \( R(x) \) stand for the local mean and residual of \( Z \) at \( x \), respectively. This is particularly important for such RFs as elevation, which is known to exhibit strong trends.

Assuming the second-order stationarity \((1)\), spatial covariance function, \( \text{cov} \mathcal{A}(h) \), for locations separated by a lag \( h \) is defined as:

\[
\text{cov}_\mathcal{A}(h) = E[(Z(x_1) - \mu(Z(x_1)))(Z(x_1')) - \mu(Z(x_1')))] - E[R(x_1)R(x_1')]
\]

for \( x_1 - x_1' = h \)

where \( Z \), \( m \), and \( R \) are defined as in Equation (1a). The cross-covariance \( \text{cov}_{Z_1, Z_2}(h) \) for two variables \( Z_1 \) and \( Z_2 \):

\[
\text{cov}_{Z_1, Z_2}(h) = E[(Z_1(x_1) - \mu(Z_1(x_1)))(Z_2(x_1') - \mu(Z_2(x_1')))] - E[R_1(x_1)R_2(x_1')]
\]

for \( x_1 - x_1' = h \)

where \( m_{Z_1}(x_1) \) and \( m_{Z_2}(x_1') \) stand for the local means of \( Z_1 \) and \( Z_2 \) at locations \( x_1 \) and \( x_1' \), respectively, and \( R_{Z_1}(x_1) \) and \( R_{Z_2}(x_1') \) represent the residuals of variables \( Z_1 \) at \( x_1 \) and \( Z_2 \) at \( x_1' \), respectively \((1, 2)\).

The aforementioned locations \( x \) refer implicitly to point or quasi-point supports. Many spatial data including elevation data are actually defined on supports of finite size (i.e. areas in 2-dimensional cases), and are assumed to be generated from finite-support random variables defined as convolution of point-support variables. Denote such a block-support variable by \( Z(v_\alpha) \), with \( v_\alpha \) standing for a finite support centered at \( x \). The data on a block support, i.e. \( z(v_\alpha) \), can be considered as integrals of point support values within their respective supports:

\[
z(v_\alpha) = \frac{1}{|v_\alpha|} \int_{v_\alpha} z(x) dx \approx \sum_{j=1}^{n_{v_\alpha}} k_\alpha(x_j)z(x_j)
\]

where \( k_\alpha(x_j) \) is \( v_\alpha \)'s sampling kernel with \( n_{v_\alpha} \) points discretizing the extent of the finite support of \( v_\alpha \) (of area \(|v_\alpha|\)). We can define mean \( m_{\mathcal{A}}(v_\alpha) \) and average covariance \( \sigma_\mathcal{A}(v_\alpha, v_\beta) \) of \( Z \) on blocks:
\[ m_Z(v_s) = E[Z(v_s)] = \sum_{s=1}^{n_s} k_s(x_s) m_Z(x_s) \quad (4a) \]

\[ \hat{c}_Z(v_s, v_y) = \sum_{s=1}^{n_s} \sum_{y=1}^{n_y} k_s(x_s) k_y(x_y) \text{cov}_Z(x_s, x_y) \quad (4b) \]

where \( n_s \) and \( n_y \) points are used to discretize the extents of the finite support \( v_s \) and \( v_y \), respectively. Similarly, block average \( \hat{c}_{Z;Z}(v_s, v_y) \), as an extension to Equation (3) according to Ref. (8), can be calculated as:

\[ \hat{c}_{Z;Z}(v_s, v_y) = \sum_{s=1}^{n_s} \sum_{y=1}^{n_y} k_s(x_s) k_y(x_y) \text{cov}_{Z;Z}(x_s, x_y) \quad (4c) \]

Point-support variogram models can be regularized over areas to derive area-support variograms, as described previously. If the experimental variograms are on areas, deconvolution needs to be performed to derive point-support variograms. Goovaerts (9) proposed an iterative approach to deconvolution, which works for both regular and irregular area units, and can be consulted for technical detail. With deconvoluted variograms, kriging can be implemented for prediction of location-specific and support-explicit Z values and associated uncertainty by combining multi-scale data (10, 11), as described in the next sub-section.

2.2. Kriging: from sparse data to dense grids or from coarser to finer grids

In general, kriging provides the best linear estimate \( \hat{z}(v_i) \) for the unknown true value \( z \) over \( v_i \), which is determined by ensuring unbiasedness and minimum dispersion of the estimator. For this paper, kriging is performed to predict values over a grid of smaller area support using: (1) sparse data (of smaller area support), and/or (2) dense data (of larger area support).

For example, simple kriging with local means can be written as:

\[ \hat{z}(v_i) = m_Z(v_i) + \sum_{s=1}^{n_s} \lambda_s(v_i) (z(v_i) - m_Z(v_i)) \quad (5) \]

where \( \hat{z}(v_i) \) and \( z(v_i) \) present the predicted value and data value at \( v_i \), respectively; \( m_Z(v_i) \) and \( m_Z(v_i) \) represent local means for \( v_i \) (to predict) and data location \( v_i \), respectively; and \( \lambda_s(v_i) \) represents the weight assigned to data location \( v_i \). Clearly, simple kriging with a stationary mean over \( A \) is easier to implement without the need to specify local means, while ordinary kriging needs to estimate the stationary but unknown mean with the consequence of increased uncertainty in the predicted values.

The linear kriging system described in Equation (5) can be adapted for point or area (the sampled locations) to point or area (locations to be predicted) predictions, depending on the relative scales (data supports) of sampled data and the locations to be predicted. Further, the relativity of points and areas can be generalized to that of smaller areas \( (v_i) \) and larger areas \( (v_a, \alpha = 1, \ldots, n) \), as is the case for SRTM and GTOPO30 DEMs used in this paper.

When local means in Equation (5) are related to an auxiliary or secondary RF \( Z_2 \), we are trying to combine information from \( Z_1 \) with that of \( Z_2 \) to improve prediction of \( Z_1 \). There are other methods of incorporating \( Z_2 \) data and information for refinement of \( Z_1 \). The so-called universal kriging often refers to the situations where by local means are functions of spatial coordinates, while kriging with an external drift (KED) uses functions of an auxiliary RF (or secondary \( Z_2 \) data) as locally defined means. There are other ways of defining local means. In regression kriging, for example, local means are determined through regression analysis, with the residuals interpolated through kriging, and the RF \( Z_1 \) being the sum of local means and kriged residuals (12, 13).

Although \( Z_2 \) data can be used as above for producing refined data with better accuracy or resolution than \( Z_1 \) data alone, \( Z_1-Z_2 \) cross-covariance and \( Z_2 \) auto-covariance \( Z_2 \) are not exploited in universal kriging and KED as in cokriging, which should be utilized for enhancing effectiveness of data refinement.

Suppose primary variable \( Z_1 \) and secondary variable \( Z_2 \) are sampled over data supports \( V_1 \) and \( V_2 \), which are denoted as \( V_1 = \{v_{ia}: Z_1(v_{ia}) \text{ known}\} \) and \( V_2 = \{v_{ib}: Z_2(v_{ib}) \text{ known}\} \), respectively. The sets of data sampled from \( V_1 \) and \( V_2 \) are denoted \( z_1(V_1) = \{z_1(v_{ia}), \alpha = 1, \ldots, n_1\} \) and \( z_2(V_2) = \{z_2(v_{ib}), \beta = 1, \ldots, n_2\} \) (vectors of lengths \( n_1 \) and \( n_2 \)), which are considered as realizations of RFs \( Z_1 \) and \( Z_2 \) over data supports \( V_1 \) and \( V_2 \), respectively.

The cokriging estimate for the value of \( Z_1 \) over \( v_i \) is expressed as a linear combination of all the available data values of \( Z_1 \) and \( Z_2 \) in the neighborhood of \( v_i \):

\[ \hat{z}_1(v_i) = \sum_{a=1}^{n_1} \lambda_1(v_i) z_1(v_{ia}) + \sum_{b=1}^{n_2} \lambda_2(v_i) z_2(v_{ib}) \]

\[ = \sum_{k=1}^{n} \lambda^T_k Z(V_k) \quad (6) \]

where \( \lambda_1 \) and \( \lambda_2 \) stand for the weight vectors assigned to datasets \( z_1(V_1) \) and \( z_2(V_2) \), respectively. These kriging weights are derived on the criteria of minimization of the variance \( E[(Z_1(v_i) - \hat{z}_1(v_i))^2] \) and non-bias in \( \hat{z}_1(v_i) \), as described by Chilès and Delfiner (2).

2.3. De-trending and bias corrections to local means

We discuss the issues of local mean specification and bias correction. For simple kriging with local means, KED, and regression kriging, scale mismatch between \( Z_1 \) and \( Z_2 \) data is not explicitly handled in trend specification by conventional procedures. If \( Z_2 \) data (usually of coarser resolution or larger support) are used for specifying \( Z_1 \)’s local means through certain linear functions, they need to be refined to match \( Z_1 \)’s resolution, as implied, for example, by \( v_i \) in Equation (5). Refinement of \( Z_2 \) data may be carried out through interpolation,
which is itself an uncertain process, regardless of the sophistication of functions used for trend specification. Furthermore, as local means are scale dependent, the uncertainty in refining local means will be complicated by that of not knowing the intrinsic scales of $Z_1$ over space a priori due to the often sparsity of sampled $Z_1$’s data. Given these uncertainties, it is sensible to use an initially refined $Z_2$ mean surface (say by the bi-linear image resampling algorithm) directly as the local mean surface for $Z_1$. This implies that there may be local trends in de-trended $Z_2$ and hence $Z_1$ data, because the aforementioned de-trending is only approximation. Thus, residuals of a de-trended RF $Z$ may be further decomposed to local means and residuals:

$$R(x) = m_p(x) + R_p(x)$$

especially when $R(x)$ exhibits local trends due to imperfectly specified $m(x)$ (and hence $R(x)$ itself) in the first place. This strategy is particularly useful for kriging and cokriging with local means, which need to be approximately at desired finer resolution, as will be shown in Section 3.

Primary and secondary variables (i.e. $Z_1$ and $Z_2$) are believed to be generated from the same underlying processes but with different supports, when there is no semantic inconsistency between $Z_1$ and $Z_2$ true values. Otherwise, we should estimate the bias of $Z_2$’s local means (and hence $Z_2$ values) relative to $Z_1$’s assumed true local means (and hence $Z_1$ values), and make correction to this bias in $Z_2$ values; bias here is synonymous to systematic error (i.e. mean error or error in local means that are known to suffer from error due to semantic inconsistency). This means that Equation (1b) for $Z_2$ should be revised as: $Z_2(x) = m_2(x) + \epsilon(x) + R_2(x)$, where $\epsilon(x)$ stands for the bias in local mean due to semantic inhomogeneity. The deviate of mean of $Z_2$ from that of $Z_1$ can be computed as:

$$\epsilon = m_1 - m_2$$

(8a)

if the stationary means are known or can be estimated with reasonable accuracy. More often, the local difference between means should be estimated through a more laborious process. This may be done by cokriging for the difference between primary and secondary means based on the sample data:

$$\tilde{\epsilon}(v_i) = \tilde{m}_1(v_i) - \tilde{m}_2(v_i)$$

$$= \sum_{j=1}^{m_1} w_1(v_j)z_1(v_j) - \sum_{j=1}^{m_2} w_2(v_j)z_2(v_j)$$

$$= w_1^T z_1(V_1) - w_2^T z_2(V_2)$$

(8b)

where the kriging weights $w_1$ and $w_2$ assigned to datasets $z_1(V_1)$ and $z_2(V_2)$ in the neighborhood of $v_i$, respectively. Further details can be found in Refs. (14, 15).

With the difference surface between primary and secondary RF means predicted, mean surface of secondary RF can be brought in alignment with that of primary RF through:

$$m_2(v_i)_{\text{corrected}} = m_2(v_i) + \epsilon(v_i)$$

(9)

This allows for the standardized cokriging (14, 16, 17, 18) and, more importantly, the univariate kriging with bias-corrected $Z_2$ data.

3. Data and results

3.1. Data

To compare geostatistical methods for elevation data refinement and to highlight complications of scale mismatch and semantic differences for kriging-based refinement, two sets of digital terrain data (which were used also for Ref. (18)) were obtained over an area of approximately 45 km by 45 km near Zhangye in Gansu Province, China. One was from the SRTM, the other from GTOPO30, with the former used as the primary data-set and the latter as the secondary data-set, as shown in Figure 1(a) and (b), respectively.

A set of sampled data for $Z_1$ were obtained from the $Z_1$ DEM by selective sampling, in particular, the VIP (very important points) algorithm in ArcGIS (see also Ref. (18)). The sampled data-set for $Z_1$ consists of 250 SRTM grid cells, shown as dots in Figure 1(a). Data refinement was carried out based on sampled $Z_1$ data (over quasi-point support) and gridded $Z_2$ data (over area support), with results’ accuracy tested against the $Z_1$ DEM data (except for the sampled locations).

As discussed previously, GTOPO30 and SRTM DEMs have their own meanings of elevation, which often vary spatially. Although SRTM data are more accurate than GTOPO30 data, with the former having a much finer grid spacing than the latter, the differences in data semantics suggest that it might be misleading to use the former as reference to test the accuracy of the resultant refinement. One would want two datasets with identical semantics to proceed with refinement and accuracy assessment.

In the light of the description about datasets employed in this study, we undertook two sets of experiments in parallel: one with the original datasets (i.e. the SRTM and GTOPO30 datasets downloaded), the other with the simulated datasets (i.e. the SRTM and simulated GTOPO30 datasets, with the latter being generated by coarsening the former through simple averaging over a moving window of 10 by 10 finer resolution cells falling within one coarser resolution cell, as shown in Figure 1(c)). The data analyses with the original datasets in contrast with those of simulated datasets was meant to examine the effects of semantic differences on data refinement and whether the corrections designed to offset their effects are effective. The references to errors in the results concerning the original datasets should be interpreted as being relative (i.e. pertaining to differences not inaccuracy). The next sub-section provides detail about the results.
Figure 1. Results of elevation data refinement: (a)-(c) elevation surfaces for \(Z_1\), \(Z_2\), and \(Z_2'\), respectively, with \(Z_1\) sampled locations shown as crosses in (a); (d) predicted \(Z_1\) surfaces using ordinary kriging with \(Z_1\) sample data; (e) and (f) collocated cokriging based on residuals of \(R_1-R_2\) data and residuals of \(R_1'-R_2'\) data, respectively.
3.2. Results
This sub-section describes pre-processing, variogram modeling, and (co)kriging with both the original and simulated datasets in parallel. For (co)kriging, in particular, the order of results described will be: downscaling kriging based on deconvolved variograms, simple kriging with local means (with and without bias corrections), and collocated co-kriging based on double de-trended data.

For the sample \( Z_1 \) data and gridded \( Z_2 \) data, a trend surface for the study area was derived from GTOPO30 data by averaging over a moving window of 3 by 3 grid cells, which was then interpolated to densify the trend surface at a grid spacing of 90 m. The residual surface for GTOPO30 elevation data was derived by subtracting the trend surface from the elevation surface \( (Z_2) \). De-trending for the SRTM elevation data was done similarly by subtraction of the densified trend surface from the original SRTM DEM data \( (Z_1) \). In all data analyses, the elevation surface values will be the sum of the trend and residual surfaces. The underlying residual RFs are denoted as \( R_1 \) and \( R_2 \), corresponding to original \( Z_1 \) and \( Z_2 \), respectively. Empirical variograms for both \( Z_1 \) and \( R_1-R_2 \) data were computed, with the corresponding models fitted using ordinary least squares. The resultant variogram models are listed in Table 1. As the secondary data \( Z_2 \) (and \( R_2 \)) should be treated as data defined over areas, point-based \( R_2 \) covariance and cross-covariance \( R_1 R_2 \) models were de-convolved from block data, based on the methods proposed by Goovaerts (9). The resultant model parameters are shown in Table 1 (see also Ref. 18).

With both the VIP sample from SRTM \( (Z_1) \) and the simulated GTOPO30 data \( (Z_2) \) (as shown in Figure 1(c)), we performed trend surfacing (to obtain mean surfaces \( m_1 \) and \( m_2 \)), de-trending (to obtain residuals \( R_1 \) and \( R_2 \)), and variogram modeling (with both the \( Z_1 Z_2 \) and \( R_1 R_2 \) datasets), using methods similar to those for the original datasets. The variogram models for \( Z_2 \), \( R_1 \), and \( R_2 \) are listed in Table 1. As in the case for \( Z_1 \), \( R_1 \), and \( R_2 \), point-based covariance \( (R_2) \) and cross-covariance \( (R_1 R_2) \) models were de-convolved from block data. The resultant model parameters are also shown in Table 1.

As base-line results, ordinary kriging was performed using \( Z_1 \) sample data alone to derive predictions of \( Z_1 \) over the 499 by 499 grid-node points (this gridding template will be the same for all the following procedures), as shown in Figure 1(d). As the complete fine-resolution \( Z_1 \), surface is available, mean error (ME), standard deviation (SD), root mean squared error (RMSE), and mean absolute error (MAE) were computed for the kriged \( Z_1 \) surface by testing it against the reference \( Z_1 \) surface. Results are shown in Table 2, and will provide the basis against which improvements in accuracy obtained with other approaches can be assessed.

With de-convolved variograms, it is possible to perform downscaling kriging (i.e. simple kriging with local means using a de-convolved variogram) for \( Z_1 \), surfacing based on denser (though coarser resolution) \( R_2 \) data (corrected for bias) alone. This resulted in a map of predicted elevation values at the SRTM data resolution. Results are reported in Table 2. As with \( R_2 \) data, downscaling kriging was performed for \( Z_1 \) surfacing with \( R_2 \) data alone, with error statistics reported in Table 2.

Simple kriging with local means was performed for \( R_1 - R_2 \) data-set. The local means for \( R_1 \) were assumed to be co-located \( R_2 \) trend surface values. The reason for using local means in kriging with \( R_2 \) data is that means of residuals (i.e. the \( R \) data) are assumed to be spatially varied (i.e. residuals of the \( R \) data are still non-stationary in terms of first moment). Results reported in Table 2, where it is shown that simple kriging with local means on the basis of \( R_1 - R_2 \) data generated better results than those based on \( Z_1 \) data and marginal improvement (only in terms of ME) over that obtained by downscaling

| Table 1. Parameters for different spatial covariance models (units: meters for lag \( h \)). |
|-----------------------------------------|------------------|------------------|
| Model                                   | Variable         | Estimated parameter value |
| Auto-covariance model                   | \( Z_1 \)         | 1,150,000 \( h^{0.6} \) |
| Auto-covariance model                   | \( R_1 \)         | 46,340 exp \( (-h^2/1080^2) \) |
| Auto-covariance model                   | \( R_2 \)         | 9887 exp \( (-h^2/756^2) \) |
| De-convolved point-based model          | \( R_1 - R_2 \)   | 17,585 exp \( (-h^2/645^2) \) |
| Cross-covariance model                  | \( R_1 - R_2 \)   | 16,110 exp \( (-h^2/756^2) \) |
| De-convolved point-based model          | \( R_1 - R_2 \)   | 24,429 exp \( (-h^2/690^2) \) |
| Auto-covariance model                   | \( R_1 - R_2 \)   | 6410 exp \( (-h/1080) \) |
| Auto-covariance model                   | \( R_1 - R_2 \)   | 10,498 exp \( (-h/802) \) |
| Cross-covariance model                  | \( R_1 - R_2 \)   | 15,369 exp \( (-h/1080) \) |
| Auto-covariance model                   | \( R_1 - R_2 \)   | 25,614 exp \( (-h/802) \) |
| Auto-covariance model                   | residuals of \( R_1 \) | 43,761 exp \( (-h/864) \) |
| Auto-covariance model                   | residuals of \( R_2 \) | 9101 exp \( (-h/864) \) |
| Cross-covariance model                  | residuals of \( R_1 - R_2 \) | \(-15,518 \ exp \( (-h/1080) \) |
| Auto-covariance model                   | residuals of \( R_1 \) | 18,144 exp \( (-h/864) \) |
| Auto-covariance model                   | residuals of \( R_2 \) | 1218 \ exp \( (-h/648) \) |
| Cross-covariance models                 | residuals of \( R_1 - R_2 \) | \(-2150 \ exp \( (-h/1080) \) |
kriging of \( R_2 \) data. The bias surface between \( Z_1 \) and \( Z_2 \) was derived by estimating the bias values \( \hat{\varepsilon}(v_i) \) in Equation (8b) using the \( R_1-R_2 \) data-set. The derivation of the bias surface enabled bias correction to local means of \( Z_2 \) (Equation (9)). Simple kriging with bias-corrected local means was then performed for the \( R_1-R_2 \) data-set, resulting in a predicted \( Z_1 \) map, with the error statistics shown in Table 2. Clearly, simple kriging with bias-corrected local means outperformed simple kriging with uncorrected local means. The explanation is that bias-corrected local means were drawn closer to the \( Z_1 \) trend surface than the uncorrected ones.

The method of simple kriging with local means was also applied to \( R_1 \) data. The local means for \( R_1 \) were assumed to be co-located \( R_2 \)-trend surface values. Error statistics of the kriged map of \( Z_1 \) reported in Table 2 show that simple kriging of \( R_1 \) with local means performs much better than downscaling kriging. To highlight negative consequences of imprudent use of bias corrections to datasets free of semantic differences, we performed simple kriging with bias-corrected local means to the de-trended data, after derivation of the bias surface between \( Z_1 \) and \( Z_2 \) (using \( R_1 \) and \( R_2 \) data). The error statistics for the resultant \( Z_1 \) predictions are shown in Table 2, indicating that simple kriging with bias-corrected local means led to much worse performance than with uncorrected local means. The explanation is that bias-correction to local means in the case of simulated data-set was improper, as \( Z_1 \) and \( Z_2 \) data (and \( R_1 \) and \( R_2 \) data) were made semantically consistent via simulation.

Cokriging with \( R_1-R_2 \) and \( R_1-R_2 \) data was found to have mixed results, and is thus omitted here. To enhance the performance of cokriging, collocated cokriging in which only collocated secondary data are used was performed on further decomposed \( R_1-R_2 \) data (i.e. data of trends and residuals of \( R_1-R_2 \) data that are themselves residuals). The reasons for collocated cokriging based on double-detrended data is that once de-trended data still exhibit significant trends and should thus be detrended further. The decomposition of \( R_1 \) and \( R_2 \) data was similar to that of \( Z_1 \) and \( Z_2 \): trend surface of \( R_2 \) was generated through averaging over 3 by 3 moving windows, with residuals of \( R_2 \) being de-trended components of the \( R_2 \) surface; trend surface of \( R_1 \) was densified version of trend surface of \( R_2 \), thus enabling the derivation of residuals of \( R_1 \) at sampled locations. The auto- and cross-covariance models for residuals of \( R_1 \) and \( R_2 \) are in Table 1. Collocated cokriging (MM2) \((19)\) was then carried out using residuals of \( R_1 \) and \( R_2 \), and generated the predicted \( R_1 \)'s residuals, which should be added to the \( R_1 \)'s trends surface and then \( Z_2 \)'s trends surface (interpolated at \( Z_1 \)'s resolution) to derive predicted \( Z_1 \) surface, as shown in Figure 1(e) (see also Ref. (18)). Error measures are shown in Table 2, indicating that collocated cokriging generated the best of all results obtained with \( Z_1 \) and \( R_1-R_2 \) datasets.

Similar to the case of \( R_1 \) \( R_2 \) data, collocated cokriging was performed using further decomposed \( R_1 \) \( R_2 \) data (trends and residuals of \( R_1 \) and \( R_2 \)). The auto- and cross-covariance models for residuals of \( R_1 \) and \( R_2 \) are in Table 1. Collocated cokriging (MM2) was then carried out using residuals of \( R_1 \) and \( R_2 \), and generated the predicted residuals of \( R_1 \), which should be added to the \( R_1 \)'s trends surface and then \( Z_2 \)'s trends surface (interpolated at 90 m resolution) to derive predicted \( Z_1 \) surface, as shown in Figure 1(f). Error measures are shown in Table 2, indicating that collocated cokriging generates the best results of all.

### 3.3. Discussion

First, we discuss the performance of downscaling kriging based on deconvolved variograms using \( R_2 \) and \( R_2 \) data, and reiterate the importance of sampling density for accuracy in data refinement. As shown in Table 2, downscaling kriging with \( R_2 \) data generated a \( Z_1 \) surface with increased prediction accuracy as opposed to simple kriging with \( R_2 \) as local means in terms of all error measures except for ME. An overall interpretation of the results with \( R_2 \) data-based downscaling kriging is that sampling density (approximately 1% for \( R_2 \) data as opposed to about 0.1% for \( R_1 \) data as mentioned previously) appears competitive against use of fine-resolution data (e.g. \( R_1 \) data alone or in combination with \( R_2 \) data as in simple kriging with local means). This confirmed the importance of sampling in data refinement, although the values of relatively more densely sampled data are offset by the semantic differences between \( R_1 \) and \( R_2 \) data, as seen by the larger ME reported with \( R_2 \) data-based downscaling kriging. As also shown in Table 2, downscaling simple kriging of \( R_2 \) (i.e. using de-trended

---

**Table 2.** Error statistics for data conflation based on the original and the simulated datasets (in each pair of error statistics (except for ordinary kriging): left for the original data-set, right for the simulated data-set; units for all measures: meters).

| Method | ME  | SD  | RMSE | MAE |
|--------|-----|-----|------|-----|
| Ordinary kriging with \( Z_1 \) | 23  | 93  | 96   | 64  |
| Downscaling simple kriging of \( R_2/R_2 \) using de-convolved variogram | 15  | –2  | 88   | 50  | 53  | 28  |
| Simple kriging with \( R_2/R_2 \) as local means without bias-correction for \( R_2/R_2 \) | 14  | –1  | 95   | 49  | 96  | 49  | 56  | 28  |
| with bias-correction for \( R_2/R_2 \) | 10  | 10  | 92   | 72  | 93  | 73  | 53  | 42  |
| Collocated cokriging with residuals of \( R_1-R_2 \) | –3  | 0   | 67   | 49  | 67  | 49  | 38  | 28  |
simulated data alone) with de-convolved variograms produced results that ranked the best in terms of MAE, only inferior to those obtained by collocated cokriging with residuals of $R_1-R_2$ and simple kriging with $R_2$ as local means by very small margins; the two more accurate approaches have actually taken advantages of densely sampled $R_2$. This reconfirms the value of data density in kriging-based downscaling in addition to proper variogram deconvolution.

Second, as shown previously, effects of semantic differences can be offset to some extent by incorporating bias corrections to local means of the secondary variable with respect to the primary variable. The effects of bias correction are, however, twofold: (1) bias-corrected $R_1$–$R_2$ data produced more accurate results than uncorrected data, as bias surfaces are known to exist due to semantic differences between $R_1$ and $R_2$. (2) bias corrections should not be considered for simulated GTOPO30 data (negative effects can be seen in Table 2 where bias correction leads to decreased accuracy as opposed to that without bias correction), which are upscaled version of SRTM DEM (implying that there is no semantic inhomogeneity between $R_1$ and $R_2$).

Third, for both the original and the simulated datasets (i.e. de-trended $R_1$–$R_2$, $R_1$–$R_2$), collocated cokriging generated the most accurate predictions of $Z_t$. In addition to the primary reason that trends still exist in $R_1$–$R_2$ and $R_1$–$R_2$ data which should thus be de-trended, there are further explanations: spatial correlation is better accommodated via collocated secondary data, as collocated data are more informative and make cokriging solution more stable; due to the large scale difference between $Z_1$ and $Z_2$ data and nature of double-detrended data (which means the spatial information has been used to a large extent), it is believed that only collocated $Z_2$ data contain useful extra information for cokriging.

4. Conclusions
This paper has explored geostatistical methods for the refinement of elevation data that are known to be complicated with semantic inhomogeneity. As demonstrated by experimental results, bias corrections to local means are effective for significantly reducing such effects and increasing the accuracy of refined data, but improper use of bias corrections in source data can lead to undesired results. Moreover, geostatistical approaches, when implemented properly, are effective for increasing accuracy in data refinement. In particular, double de-trending of the underlying RFs, proper specification of local means, variogram deconvolution, and proper use of kriging are key contributors to accuracy in data refinement. Further research is needed on data refinement methods that can accomplish fusion and refinement of gridded DEM data efficiently while handling semantic difference and scale mismatch effectively.

Acknowledgments
Research reported in this paper is supported by the National Natural Science Foundation of China (grant numbers 41171346, 41471375).

Notes on contributors
Jingxiong Zhang, having a PhD in geography from the University of Edinburgh in 1997, is now a professor and maintains his research interests in remote sensing and GIScience.

Tao Zhu, a master research student, has research interests in geomatics.

Yunwei Tang, having obtained PhD in GIScience from Wuhan University in 2013, is now an assistant research scientist and has research interests in remote sensing and geostatistics.

Wangle Zhang, a PhD student, has research interests in remote sensing and geostatistics.

References
(1) Journel, A.G.; Huijbregts, Ch.J. Mining Geostatistics; Academic Press: London, 1978.
(2) Chilès, J.P.; Delfiner, P. Geostatistics: Modeling Spatial Uncertainty; John Wiley: New York, NY, 1999.
(3) Aslì, M.; Marcotte, D. Comparison of Approaches to Spatial Estimation in a Bivariate Context. Math. Geol. 1995, 27 (5), 641–658.
(4) Gotway, C.A.; Young, L.J. A Geostatistical Approach to Linking Geographically-Aggregated Data from Different Sources. J. Comput. Gr. Stat. 2007, 16 (1), 1–21.
(5) Yoo, E.-H.; Kyriakidis, P.C. Area-to-Point Prediction Under Boundary Conditions. Geog. Anal. 2008, 40 (4), 355–379.
(6) Yoo, E.-H.; Kyriakidis, P.C. Area-to-Point Kriging in Spatial Hedonic Pricing Models. J. Geog. Sci. Syst. 2009, 11 (4), 381–406.
(7) Zhang, J.; Atkinson, P.M.; Goodchild, M.F. Scale in Spatial Information and Analysis; CRC Press: Boca Raton, FL, 2014.
(8) Kyriakidis, P.C. A Geostatistical Framework for Area-toPoint Spatial Interpolation. Geog. Anal. 2004, 36 (3), 259–289.
(9) Goovaerts, P. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. Math. Geol. 2008, 40 (1), 101–128.
(10) Young, L.J.; Gotway, C.A.; Yang, J.; Kearney, G.; DuClos, C. Linking Health and Environmental Data in Geographical Analysis: It’s So Much More Than Centroids. Spatial Spatio-temporal Epidemiol. 2009, 1 (1), 73–84.
(11) Zhang, J.; Kyriakidis, P.; Kelly, R. Geostatistical Approaches to Conflation of Continental Snow Data. Int. J. Remote Sens. 2009, 30 (20), 5441–5451.
(12) Lark, R.M.; Cullis, B.R.; Welham, S.J. On Spatial Prediction of Soil Properties in the Presence of a Spatial Trend: The Empirical Best Linear Unbiased Predictor (E-BLUP) with REML. Eur. J. Soil Sci. 2006, 57 (6), 787–799.
(13) Minasny, B.; McBratney, A.B. Spatial Prediction of Soil Properties Using EBLUP with the Matérn Covariance Function. Geoderma 2007, 140 (4), 324–336.
(14) Goovaerts, P. Ordinary Cokriging Revisited. Math. Geol. 1998, 30 (1), 21–42.
(15) Papritz, A.; Flühler, H. Temporal Change of Spatially Autocorrelated Soil Properties: Optimal Estimation by Cokriging. Geoderma 1994, 62 (1–3), 29–43.
(16) Bishop, T.F.A.; Lark, R.M. The Geostatistical Analysis of Experiments at the Landscape-Scale. *Geoderma* 2006, 133 (1–2), 87–106.

(17) Papritz, A. Standardized vs. Customary Ordinary Cokriging: Some Comments on the Article “The Geostatistical Analysis of Experiments at the Landscape-Scale” by T.F.A. Bishop and R.M. Lark. *Geoderma* 2008, 146 (1–2), 391–396.

(18) Zhang, J.X.; Tang, Y.W.; Goodchild, M.F. Conflation of Sparsely Sampled and Gridded Elevation Data: Scale Mismatch and Semantic Differences. *Int. J. Image Data Fusion*, in press.

(19) Ma, X.; Journel, A.G. An Expanded GSLIB Cokriging Program Allowing for Two Markov Models. *Comput. Geosci.* 1999, 25 (6), 627–639.