Increasing the accuracy of digital elevation models by means of geostatistical conflation

Cutberto U. Paredes-Hernández, Nicholas J. Tate
Kevin J. Tansey, Peter F. Fisher
Department of Geography
University of Leicester
Leicester, UK
{cup1, njt9, kjt7, pff1}@le.ac.uk

Wilver E. Salinas-Castillo
UAM Agronomía y Ciencias
Universidad Autónoma de Tamaulipas
Cd. Victoria, Tamaulipas, México
wsalinas@uat.edu.mx

Abstract— **In this paper we compare the use of two geostatistical conflation techniques, Ordinary Cokriging (OCK) and Kriging with an External Drift (KED), to increase the accuracy of a Digital Elevation Model using a set of sparsely distributed accurate Ground Control Points. Our results show that both conflation techniques produce more accurate DEMs than any of the data sources used individually, with KED producing the most accurate results.**

Keywords: ***DEM Accuracy, Geostatistical Conflation, Cokriging, Kriging with an external drift.***

I. Introduction

Digital Elevation Models (DEMs) are one of the most used inputs in environmental applications and their accuracy impacts the accuracy of the final results. Therefore, it can be expected that an increase in the accuracy of existing DEMs will result in more accurate outputs from environmental models. In this paper we present and compare the use of two geostatistical techniques for increasing the accuracy of existing DEMs by means of geostatistical conflation.

Geostatistical conflation is the term used to describe the application of geostatistical techniques to integrate datasets with different scales and different accuracies, selecting and/or combining the properties of each dataset (Kyriakidis et al., 1999) to produce more accurate and representative products than any of the input datasets (Zhang and Goodchild, 2002). Geostatistical conflation is considered of great benefit for increasing the accuracy of estimations when more than one correlated dataset exists for the area of interest and the primary information is sparsely distributed (showing poor spatial autocorrelation) and the secondary information is densely sampled (Goovaerts, 1997).

Geostatistical conflation has been previously applied to elevation by Kyriakidis et al. (1999) and Zhang and Goodchild (2002) to assess the accuracy and conflate digital elevations models and by Hengl et al. (2008) to produce topographic mapping using auxiliary variables. It has also been increasingly being applied to other variables to increase the prediction accuracy using secondary information (see for example Gohin and Langlois, 1993; Goovaerts, 2000; Zhang et al., 2009).

In the case of elevation, the geostatistical conflation of sparse ground control points (GCPs) and dense DEMs derived from remote sensors can potentially produce more accurate elevation surfaces in areas where both datasets represent bare terrain, but may deviate significantly in non-terrain areas (Zhang and Goodchild, 2002). However, the geostatistical conflation of both datasets is expected to provide more accurate and representative DEMs than any of the input datasets by themselves and reconcile the mismatches between both data sources (Zhang and Goodchild, 2002). Furthermore, it can better reproduce the spatial complexity of the terrain by including additional information in the Kriging estimators (Kyriakidis *et al.*, 1999).

The geostatistical techniques that we use in this paper to increase the accuracy of existing DEMs are introduced in the next section. Then, the datasets used in this paper and the data processing using the introduced techniques are presented. Finally, the results are discussed and some future work is suggested.

II. Methodology

The geostatistical techniques that we use to conflate and increase the accuracy of existing DEMs with sparse highly-accurate GCPs are Ordinary Cokriging (OCK) and Kriging with and External Drift (KED). OCK and KED are both geostatistical techniques that allow the incorporation of secondary variables in the Kriging estimator and are capable of reproducing the spatial complexity of the input information into the output dataset (DEM). The main difference between these two techniques is that OCK assigns the weights for the input information (GCPs and existing DEM) using measures of spatial (cross)correlation, while KED first performs regression between the primary (GCPs) and secondary information (existing DEM) and then interpolates the residuals from the regression model using Kriging (Goovaerts, 2000). The formulas of these techniques are shown below.

A. *Ordinary Cokriging*

The different versions of Kriging are based on the basic linear regression estimator, written as (Goovaerts, 1997)

$$Z^{*}(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}[Z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})] \qquad (1)$$

where $Z^{*}(\mathbf{u})$ is the estimated value, λ_{α} are the weights assigned to the sampled $z(\mathbf{u}_{\alpha})$ values, $m(\mathbf{u})$ and $m(\mathbf{u}_{\alpha})$ are the expected values (i.e. means) of the interpolated surface and the available samples, respectively.

Ordinary Cokriging (OCK) incorporates a secondary variable Y into the Kriging estimator accounting for local variability in the mean of the primary and secondary variables. The OCK estimator for two variables is written as (Goovaerts, 1997, 224)

$$Z^*_{OCK}(\mathbf{u}) = \sum_{\alpha 1=1}^{n_1(\mathbf{u})} \lambda_{\alpha 1}^{OCK}(\mathbf{u}) Z(\mathbf{u}_{\alpha 1}) + \sum_{\alpha 2=1}^{n_2(\mathbf{u})} \lambda_{\alpha 2}^{OCK}(\mathbf{u}) Y(\mathbf{u}_{\alpha 2}) \tag{2}$$

with the weights of the primary variable constrained to sum to one and the weights of the secondary variable constrained to sum to zero

$$\sum_{\alpha 1=1}^{n_1(\mathbf{u})} \lambda_{\alpha 1}^{OCK}(\mathbf{u}) = 1 \qquad \sum_{\alpha 2=1}^{n_2(\mathbf{u})} \lambda_{\alpha 2}^{OCK}(\mathbf{u}) = 0 \tag{3}$$

Cokriging, as can be observed in (2), is one of the most complex geostatistical techniques since it requires the computing and modelling of three variograms (Goovaerts, 2000). Therefore, alternative techniques like Kriging with an External Drift (KED) have been developed in order to overcome this complexity (Deutsch and Journel, 1998).

B. Kriging with an External Drift

KED incorporates secondary information into the Kriging system by using auxiliary information to estimate the local mean of the primary variable and then perform Kriging on the corresponding residuals. The trend $m_{KED}(\mathbf{u})$ (i.e. local mean) is estimated within the Kriging system for each local neighbourhood as (Goovaerts, 1997, 194)

$$m_{KED}(\mathbf{u}) = a_0(\mathbf{u}) + a_1(\mathbf{u}) y(\mathbf{u}) \tag{4}$$

where $y(\mathbf{u})$ is the secondary information. The KED estimator is defined then as

$$Z^*_{KED}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KED}(\mathbf{u}) Z(\mathbf{u}_{\alpha}) \tag{5}$$

III. Datasets

In this paper we increase the accuracy of an existing DEM, extracted from remote sensing data sources, using GCPs extracted from a highly accurate Lidar DEM, which was used later to assess the accuracy increase achieved using OCK and KED.

The DEM used in this paper (Fig. 1a) is a small subset of a DEM extracted from a TerraSAR-X interferometric dataset using 4x4 multi-looks using GAMMA software. The subset used has a spatial resolution (i.e. cell size) of 5.06 m and covers a sparsely vegetated area of approximately 49 ha (900 x 550 m). The accuracy of this DEM was assessed using a Lidar DEM (Fig. 1b) as reference (Existing DEM – Lidar DEM). The error exploratory analysis of these two DEMs is presented in Table I. The accuracy of the Lidar DEM was assessed using 1531 ground control points collected on the field using Real-Time Kinematic (RTK) GPS.

The GCPs that were used to increase the accuracy of the existing DEM were extracted from the Lidar DEM, simulating a sparse field survey by extracting elevations from near the roads and important features (e.g. peaks). A total of 184 GCPs were used in this paper. The GCP distribution is shown in Fig. 1c.

IV. Data Processing

The datasets used in this paper were imported into the statistical software R (R Development Core Team, 2009) using the packages rgdal (Keitt et al., 2010) and sp (Pebesma and Bivand, 2005). After importing the datasets, Ordinary Kriging (OK), OCK and KED as implemented in the R package gstat (Pebesma, 2004) were used to produce DEMs using the imported data.

A DEM was produced using Ordinary Kriging (OK) (Fig. 1d) using only the 184 GCPs as input data. Because of the small number and distribution of available elevation points, the variogram showed different spatial autocorrelation (modelled using a Power variogram model with a small nugget) than that of the original Lidar DEM (Gaussian model). However, we continued to generate the DEM in order to evaluate the benefits of conflating the existing DEM with the GCPs using OCK and KED. The neighbourhood size was set to 48 in the `krige` command. Other neighbourhood sizes were tested but produced less accurate results, and are, therefore, not reported here. Ordinary Cokriging (OCK) was used to conflate the existing DEM with the GCPs in order to increase its accuracy.

TABLE I. DEM DATA SOURCES ERROR EXPLORATORY ANALYSIS (UNITS: METRES)

DEM	Error Statistics								
	RMSE	*Mean*	*Std. Dev.*	*Min*	*Q 0.025*	*Q 0.25*	*Q 0.75*	*Q 0.975*	*Max*
Lidar DEM	0.1108	-0.0378	0.1042	-0.4526	-0.2372	-0.0777	0.0172	0.0999	0.3789
Existing DEM[a]	9.4350	2.0427	9.2115	-26.4923	-16.1796	-3.7524	8.9082	17.3024	28.3202

a. Please note that this is accuracy does not reflect the maximum accuracy that can be achieved using TerraSAR-X. See the poster paper "**An Accuracy Assessment of Spaceborne X-band (TerraSAR-X) Spotlight Mode InSAR DEMs**" in these proceedings for an exploratory analysis of multiple configurations and accuracies.

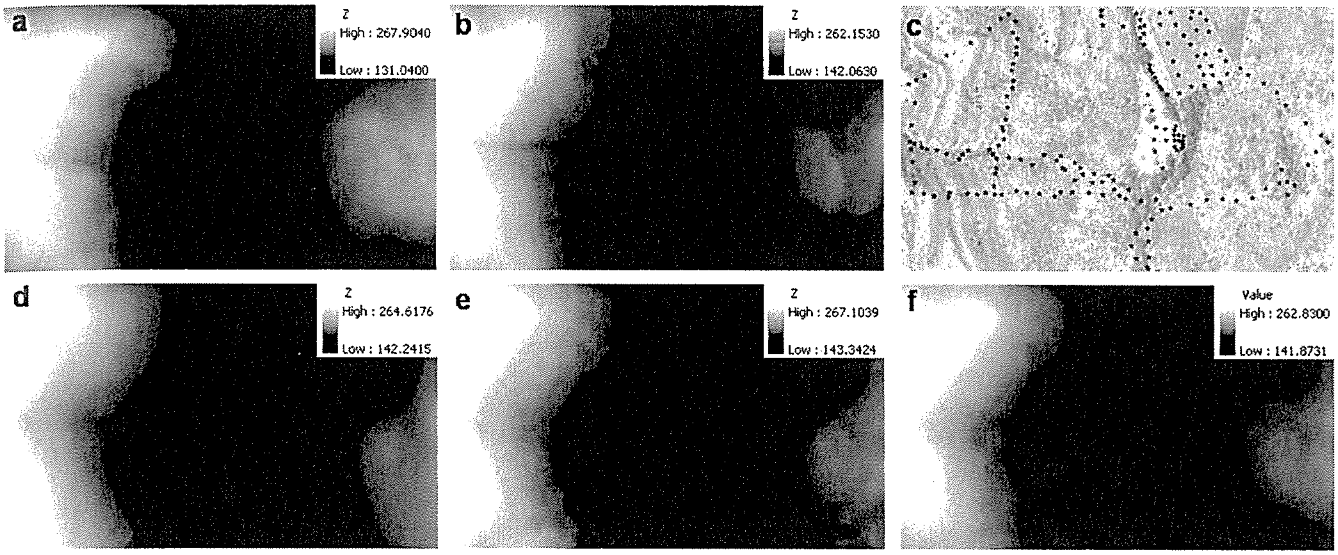

Figure 1. a. Existing DEM. b. Lidar DEM. c. Ground Control Points (GCPs) with aerial photography as background. d. Ordinary Kriging DEM using only GCPs. e. Conflated DEM using Ordinary Cokriging (OCK). f. Conflated DEM using Kriging with an External Drift (KED).

The GCPs and the existing DEM were used as primary and secondary information, respectively, in the Kriging system. The variograms (direct and cross) were modelled using the same Power model used to produce the OK DEM, since OCK requires the structure used to model the three variograms to be the same in order to guarantee the validity of the Linear Model of Coregionalisation (LCM). The neighbourhood size was set to 48 for both primary and secondary information. Other neighbourhoods were also tested but are not reported here. The output (OCK DEM) of this conflating technique is shown in Fig. 1e.The existing DEM and the GCPs were also conflated using Kriging with an External Drift (KED), using the GCPs as primary information and the DEM as the predictor variable. The elevation reported by the DEM at the GCPs location was obtained using the `overlay` command from the sp package so that it could be passed as the predictor to the `variogram` command, which was modelled using a Spherical structure. The conflated DEM (Fig. 1f) was produced using the `krige` command with the existing DEM being used as the predictor variable.

The accuracy of the OK DEM and the conflated (OCK and KED) DEMs was then assessed by subtracting the Lidar DEM from the produced DEMs, so that overestimation and underestimation resulted in positive and negative values, respectively. The results of the accuracy assessment are reported in Table II.

V. Results

By visually inspecting the conflated DEMs (Fig. 1e,f) and comparing them with the DEMs of original data sources (Fig. 1a,d) it can be noted that the noise commonly present in DEMs extracted from remote sensing data sources is reduced by both techniques, mostly because of the smoothing effect of Kriging interpolators; and that the texture of both conflated DEMs is more similar to the Lidar DEM (Fig. 1b) than the texture of the DEMs produced using the data sources individually.

Furthermore, the error exploratory analysis of the conflated DEMs (Table II) shows that OCK and KED are both appropriate techniques for conflating dense DEMs with sparse highly-accurate GCPs. Both conflation techniques produce more accurate DEMs than any of the data sources used individually, with KED producing the most accurate results. The mean error and its standard deviation were reduced using both techniques, which means that the spread of the errors in the DEM was reduced, as confirmed also by the quantiles shown in Table II. However, even when the range of the errors was reduced, the maximum (overestimation) error was increased when using OCK. Nevertheless, the Root Mean Squared Error (RMSE) of the existing DEM (9.4350 m) was reduced by approximately 25% using OCK and halved using KED, confirming that the accuracy of existing DEMs can be increased by means of geostatistical conflation.

TABLE II. CONFLATED DEMs ERROR EXPLORATORY ANALYSIS (UNITS: METRES)

DEM	Error Statistics								
	RMSE	*Mean*	*Std. Dev.*	*Min*	*Q 0.025*	*Q 0.25*	*Q 0.75*	*Q 0.975*	*Max*
OK	8.8750	2.6060	8.4840	-16.1316	-8.9840	-0.8486	3.2974	30.5404	54.3690
OCK	6.8836	1.7667	6.6532	-15.1279	-7.9120	-1.1689	2.7428	21.8827	42.4899
KED	4.3180	0.8111	4.2413	-9.7800	-5.2799	-1.4234	1.7412	12.0229	24.3982

FUTURE WORK

The results presented in this paper were produced using two of many conflation techniques and should be compared with those of other techniques such as Standardised Cokriging, Simple Kriging with varying local means (SKlm), Regression Kriging and simulated annealing among others. Furthermore, the accuracy increase achieved using these techniques can possibly be larger if non-terrain segments are filtered from the existing DEM before undertaking geostatistical conflation

ACKNOWLEDGEMENTS

The TerraSAR-X interferometric data was provided by DLR under the research proposal CAL0428. The authors would like to thank Edzer Pebesma, Paulo Justiniano Ribeiro and the r-sig-geo community for promptly answering questions about Kriging methods in the gstat package and R in general. Cutberto would like to thank the Mexican National Council for Science and Technology (CONACYT) for funding his PhD at the University of Leicester and to the Universidad Autónoma de Tamaulipas, Mexico for funding his research project.

REFERENCES

Deutsch, C.V. and Journel, A.G. (1998). *GSLIB. geostatistical software library and user's guide*, New York: Oxford University Press, 369pp.

Gohin, F. and Langlois, G. (1993). Using geostatistics to merge in situ measurements and remotely-sensed observations of sea surface temperature. *International Journal of Remote Sensing*. 14, 9–13.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press, 483pp.

Goovaerts, P. (2000) Geostatistical Approaches for Incorporating Elevation into the Spatial Interpolation of Rainfall. *Journal of Hydrology*. 228(1-2), 113-129.

Hengl, T., Bajat, B., Dragan, B. and Reuter, H.I. (2008). Geostatistical modeling of topography using auxiliary maps. *Computers & Geosciences*. 34(12), 1886-1899.

Keitt, T.H., Bivand, R., Pebesma, E. and Rowlingson, B. (2010). *Rgdal: bindings for the geospatial data abstraction library*. R package version 0.6-25. http://CRAN.R-project.org/package=rgdal

Kyriakidis, P.C., Shortridge, A.M. and Goodchild, M.F. (1999) Geostatistics for conflation and accuracy assessment of digital elevation models. *International Journal of Geographical Information Science*. 13(7), 677-707.

Pebesma, E.J., (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*. 30(7), 683-691.

Pebesma, E.J. and Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News*. 5 (2), 9-13.

R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Zhang, J. and Goodchild, M.F. (2002) *Uncertainty in geographical information*. Research Monographs in GIS Series, London: Taylor and Francis, 266pp.

Zhang, J., Kyariakidis, P. and Kelly, R. (2009). Geostatistical approaches to conflation of continental snow data. *International Journal of Remote Sensing*. 30(20), 5441-5451.