Increasing the Accuracy of Low Spatial Resolution Digital Elevation Models using Geostatistical Conflation

C.U. Paredes-Hernández[1], N.J. Tate[1], K.J. Tansey[1], P.F. Fisher[1], W.E. Salinas-Castillo[2]

[1]Department of Geography, University of Leicester, Leicester, United Kingdom, LE2 6UE
Email: {cup1, njt9, kjt7, pff1}@le.ac.uk

[2]UAM Agronomía y Ciencias, Universidad Autónoma de Tamaulipas, Cd. Victoria, Tamaulipas, México, C.P. 87149
Email: wsalinas@uat.edu.mx

1. Introduction

Digital Elevation Models (DEMs) are an important data source for a range of scientific and commercial applications, which accuracy depends on the accuracy of the input DEMs. Examples of such applications are hydrological studies, topographic mapping and landscape modelling, among others.

Different technologies (e.g. Lidar, radar, photogrammetry) exist for producing accurate high resolution DEMs. However, DEMs produced using these technologies are generally limited to small areas and are expensive. In contrast, low spatial resolution DEMs cover most of the planet and are freely available. Consequently, these DEMs are commonly used in applications with limited resources (Hirt et al. 2010).

However, the accuracy of low resolution DEMs that are freely available (e.g. Aster GDEM, SRTM) is undocumented for specific study areas and only an estimate of the global or regional accuracy is provided with them, adding uncertainty to their use.

In this paper we recommend the use of geostatistical conflation to reduce the uncertainty associated with the use of low resolution DEMs by increasing their vertical accuracy by means of conflating them with a set of sparsely distributed Ground Control Points (GCPs) using Kriging with an External Drift (KED).

2. Geostatistical Conflation

The application of geostatistical techniques to the integration of datasets with different accuracies and spatial resolutions is known as geostatistical conflation (Kyriakidis et al. 1999). The main objective of geostatistical conflation is to combine the properties of the input datasets to produce more accurate and representative products (Zhang and Goodchild 2002).

Geostatistical conflation has been previously applied to elevation to assess accuracy and uncertainty using Cokriging as the conflation technique (Kyriakidis et al. 1999), to increase the accuracy of photogrammetric DEMs using Cokriging and simulated annealing (Zhang and Goodchild 2002) and to produce more accurate DEMs using auxiliary variables using Regression Kriging (RK) (Hengl et al. 2008). The technique we use in this paper, KED (described below), is equivalent to RK when the same neighbourhood is used in the Kriging system (Hengl et al. 2007).

2.1 Kriging with an External Drift

Kriging with an External Drift (KED) is a geostatistical technique that can be used to conflate multiple datasets since it incorporates auxiliary information in the Kriging estimator and reproduces the spatial complexity of the input datasets into the output surface (Goovaerts 1997).

KED incorporates multiple datasets into the Kriging system by using the auxiliary information (DEM in this case) to estimate the local mean (i.e. trend) of the primary variable (GCPs) and then perform Kriging on the corresponding residuals. The trend is estimated within the Kriging system for each local neighbourhood as (Goovaerts 1997:194)

$$m_{KED}(\mathbf{u}) = a_0(\mathbf{u}) + a_1(\mathbf{u})y(\mathbf{u}) \tag{1}$$

where $y(\mathbf{u})$ is the secondary information. The KED estimator is defined then as (Goovaerts 2000:121)

$$Z^*_{KED}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KED}(\mathbf{u})[Z(\mathbf{u}_\alpha) - m_{KED}(\mathbf{u})] + m_{KED}(\mathbf{u}) \tag{2}$$

3. Study Case

3.1 Study Area

In this paper we use a 1x10 km study area located in Veracruz, Mexico (centred at 19º32’24” N and 96º30’44” W). The area is rich in topography with flat areas in the centre and mountainous terrain near the borders. The elevations range from 111 m to 358 m, with a mean slope of 9º, as reported by Lidar datasets. The area is sparsely vegetated, with the exception of the last 2 km to the south of the study area.

3.2 Datasets

Three low spatial resolution DEMs are freely available for the study area: the Aster GDEM (METI-NASA 2009), the SRTM DEM v4 (Jarvis et al. 2008) and the Inegi DEM (the Mexican national elevation dataset; INEGI 2003).

The Aster (Figure 1a) and the Inegi (Figure 1b) DEMs spatial resolution is 1 arc-second or 30 m. The expected vertical accuracy of the first is between 7 m and 14 m (METI-NASA 2009). The accuracy of the Inegi DEM is undocumented. The SRTM DEM (Figure 1c) spatial resolution is 3 arc-seconds or 90 m with an expected accuracy of 10m (METI-NASA 2009).

The set of GCPs (Figure 1e) that was conflated with the DEMs was extracted from Lidar data available for the study area. A total of 312 GCPs were extracted near the roads and important topographic features with a minimal horizontal spacing of 30 m.

A Lidar DSM (Digital Surface Model; Figure 1d) was used as reference to assess the accuracy of the DEMs. The original spatial resolution of the Lidar DSM is 1 m with a vertical accuracy of 0.1469 m (assessed using 965 GCPs collected using Real-Time Kinematic GPS).

3.2 Data processing

The data processing was undertaken in R (R Development Core Team 2009). First, all the datasets were imported into R using the package rgdal (Keitt et al., 2010). Then, to assess the accuracy of the DEMs the Lidar dataset was resampled to match the spatial resolution of the different DEMs using the package raster (Hijmans and van Etten 2010). The resampled Lidar DSM was subtracted from the three DEMs to assess their vertical accuracy. The results are reported in Table 1.

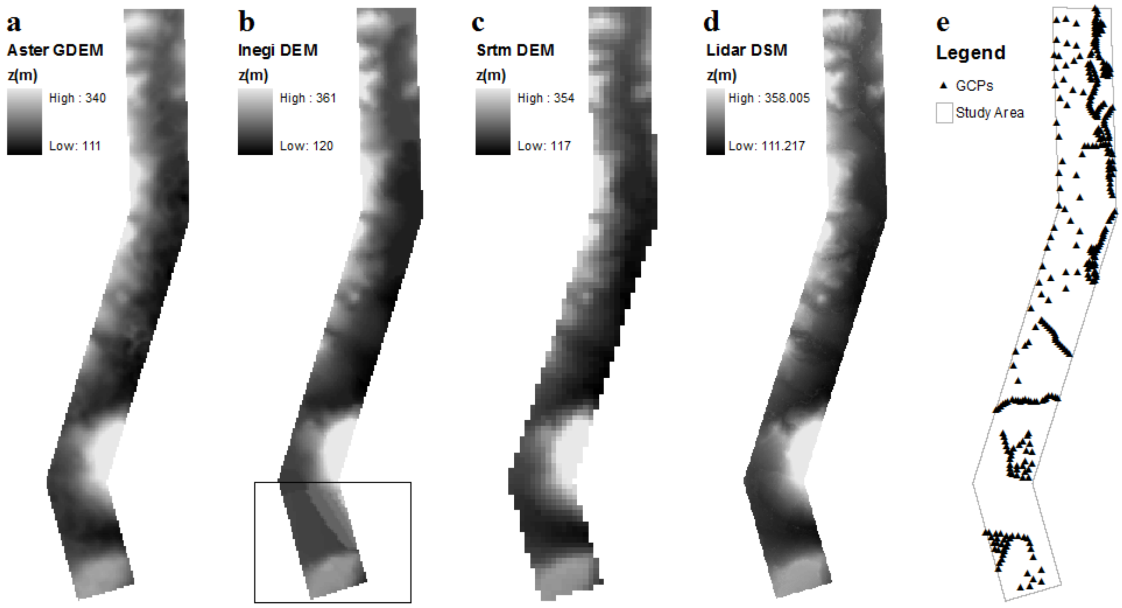

Figure 1. Datasets used in this paper. a. Aster GDEM. b. Inegi DEM. The area enclosed in the box was removed, reported in the error statistics as Inegi DEM (ER). c. SRTM DEM. d. Lidar DSM. e. Set of GCPs conflated with the DEMs.

In order to assess the accuracy of a DEM produced using only the GCPs, Ordinary Kriging (OK) as implemented in the R version of gstat (Pebesma 2004) was used to produce DEMs matching the spatial resolution of the existing DEMs (30 m and 90 m). Their accuracy was assessed using the resampled Lidar DSM and reported in Table 2.

Each of the low resolution DEMs was conflated with the GCPs using Kriging with an External Drift (KED), also using gstat. The elevation reported by the DEMs was used as the regressor and the GCPs as the primary information. The accuracy of the conflated DEMs is reported in Table 3.

Since it was observed that the Inegi DEM reported erroneous elevations in the vegetated area, this area was removed and KED was undertaken using the rest of the area and 277 GCPs. The error statistics from this subset are reported as ER (Error Removed) in the tables below. The area that was removed is shown in Figure 1b.

Table 1. Original DEMs Error Statistics.

DEM	RMSE	Mean	Std. Dev.	Min	Max
Aster GDEM	5.4329	1.4096	5.2471	-28.3202	20.8639
Inegi DEM	11.5371	5.2352	10.2814	-31.6234	75.4420
Inegi DEM (ER)	7.4441	3.7828	6.4117	-26.2211	33.3725
Srtm DEM	8.2137	2.4419	7.8462	-35.2507	43.9907

Table 2. Ordinary Kriging (OK) DEM Error Statistics.

DEM	RMSE	Mean	Std. Dev.	Min	Max
OK 30m	17.0868	-0.2210	17.0862	-50.4549	201.2829
OK 90m	15.9790	-0.5821	15.9749	-46.1691	130.0117

4. Results

Table 3 shows the results of the accuracy assessment of the conflated DEMs using KED. The Root Mean Squared Error (RMSE) of the three conflated DEMs was reduced, the mean error was brought close to zero and the error standard deviation was also reduced.

Furthermore, in all cases the accuracy of the conflated DEMs is higher than that of the DEMs produced using individual datasets (Tables 1 and 2). However, in the case of the Aster GDEM the accuracy increase is marginal, possibly due to the presence of artefacts in its current version (Hirt et al. 2010) which cause instability in KED (Goovaerts 1997:195).

To sum up, the results presented here suggest that the geostatistical conflation of low spatial resolution DEMs with a set of sparsely distributed GCPs increases the vertical accuracy of the DEMs, which should also increase the accuracy of the applications where these DEMs are used and reduce the uncertainty associated with their use.

Table 3. Conflated DEMs Error Statistics.

DEM	RMSE	Mean	Std. Dev.	Min	Max
KED Aster	5.3136	-0.1923	5.3104	-23.0801	24.2297
KED Inegi	7.2816	0.6868	7.2494	-30.4990	61.4068
KED Inegi (ER)	5.6448	0.2903	5.6376	-23.7957	27.2045
KED Srtm	6.5961	1.4416	6.4393	-22.1884	30.3891

Acknowledgements

The authors would like to thank Edzer Pebesma 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

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, Heuvelink GBM and Rossiter DG, 2007, About Regression-Kriging: From Equations to Case Studies. *Computers & Geosciences*, 33(10):1301-1315.

Hengl T, Bajat B, Dragan B and Reuter HI, 2008, Geostatistical Modeling of Topography Using Auxiliary Maps. *Computers & Geosciences*, 34(12):1886-1899.

Hijmans RJ and van Etten J, 2010, *raster: Geographic analysis and modeling with raster data*. R package version 1.0.0-1. http://CRAN.R-project.org/package=raster

Hirt C, Filmer MS and Featherstone WE, 2010, Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9S ver3 digital elevation models over Australia, *Australian Journal of Earth Sciences*, 57(3):337-347.

INEGI (Instituto Nacional de Estadística y Geografía, México), 2003, Sistema de Descarga del Continuo de Elevaciones Mexicano (CEM), Available from: http://mapserver.inegi.org.mx/DescargaMDEWeb/. Accessed: 20 April 2010.

Jarvis A, Reuter HI, Nelson A and Guevara E, 2008, *Hole-filled SRTM for the globe Version 4*, available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org. Accessed: 20 April 2010.

Keitt TH, Bivand R, Pebesma E and Rowlingson B, 2010, *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 0.6-26. http://CRAN.R-project.org/package=rgdal

Kyriakidis PC, Shortridge AM and Goodchild MF, 1999, Geostatistics for Conflation and Accuracy Assessment of Digital Elevation Models. *International Journal of Geographical Information Science*, 13(7):677-707.

METI-NASA (Ministry of Economy, Trade and Industry of Japan - National Aeronautics and Space Administration, USA), 2009, *ASTER Global Digital Elevation Model (GDEM)*. Available from: http://www.gdem.aster.ersdac.or.jp/. Accessed: 20 April 2010.

Pebesma EJ, 2004, Multivariable Geostatistics in S: the gstat Package. *Computers & Geosciences*, 30(7):683-691.

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 MF, 2002, *Uncertainty in Geographical Information*. Research Monographs in GIS Series, London: Taylor and Francis, 266pp.