Korean Journal of Remote Sensing, Vol.24, No.5, 2008, pp.453~462

Geostatistical Integration of Different Sources of Elevation and its Effect on Landslide Hazard Mapping

No-Wook Park*† and Phaedon C. Kyriakidis**

*Department of Geoinformatic Engineering, Inha University

**Department of Geography, University of California Santa Barbara

Abstract : The objective of this paper is to compare the prediction performances of different landslide hazard maps based on topographic data stemming from different sources of elevation. The geostatistical framework of kriging, which can properly integrate spatial data with different accuracy, is applied for generating more reliable elevation estimates from both sparse elevation spot heights and exhaustive ASTER-based elevation values. A case study from Boeun, Korea illustrates that the integration of elevation and slope maps derived from different data yielded different prediction performances for landslide hazard mapping. The landslide hazard map constructed by using the elevation and the associated slope maps based on geostatistical integration of spot heights and ASTER-based elevation resulted in the best prediction performance. Landslide hazard mapping using elevation and slope maps derived from the interpolation of only sparse spot heights showed the worst prediction performance.

Key Words : Elevation, Landslide hazard, Geostatistics.

1. Introduction

Landslide hazard mapping has been regarded as an important geoscientific application field of GIS and remote sensing. Multiple spatial data should be considered simultaneously to predict future landslide hazard, since landslide occurrences are related to a large number of geomorphological and/or environmental variables (Park and Chi, 2008). Traditional GIS-based landslide hazard mapping tasks have focused on how to effectively integrate multiple spatial data, and many models based on probability, fuzzy set theory, and artificial intelligence have been proposed and tested (Chung and Fabbri, 1999; Ercanoglu and Gokceoglu, 2002; Lee *et al.*, 2006).

Since data come from various sources, however, they inevitably have varying degrees of reliability and accuracy. For example, elevation and slope maps have been regarded as the most important factors in landslide hazard mapping, but such maps can be generated from various data types and sources (e.g. spot heights, contours). These different elevation data types are usually used to construct a regular raster of elevation estimates. Alternatively, digital photogrammetric techniques can also be applied to aerial photographs

Received October 1, 2008; Revised October 10, 2008; Accepted October 18, 2008.

† Corresponding Author: No-Wook Park (nwpark@inha.ac.kr)

or satellite-based stereo images. No matter the way in which such elevation estimates are constructed, they are inherently uncertain and may affect the associated slope values and further analyses (Kyriakidis *et al.*, 1999). Thus, these effects as well as the development of effective data integration models should be considered in data integration tasks. To our knowledge, the effects of input spatial data which have been integrated for landslide hazard mapping have not been considered to date.

In this paper, the effects of elevation estimates derived from different sources are investigated for the purpose of landslide hazard mapping. Two different types of elevation estimates, elevation spot heights and a DEM from ASTER stereo images are first considered. The spot heights, which are sparsely sampled but have higher accuracy, are regarded as hard data in this work. The elevation values extracted from ASTER can be available at all pixel locations but are of lower accuracy, and thus regarded as soft data. A variant of kriging is applied to integrate the hard data with the abundant soft data. Slope values are computed from different elevation maps and the elevation and slope maps were then integrated for landslide hazard mapping. The associated performances of different landslide hazard maps are compared and discussed through a case study of Boeun area, Korea (Fig. 1).

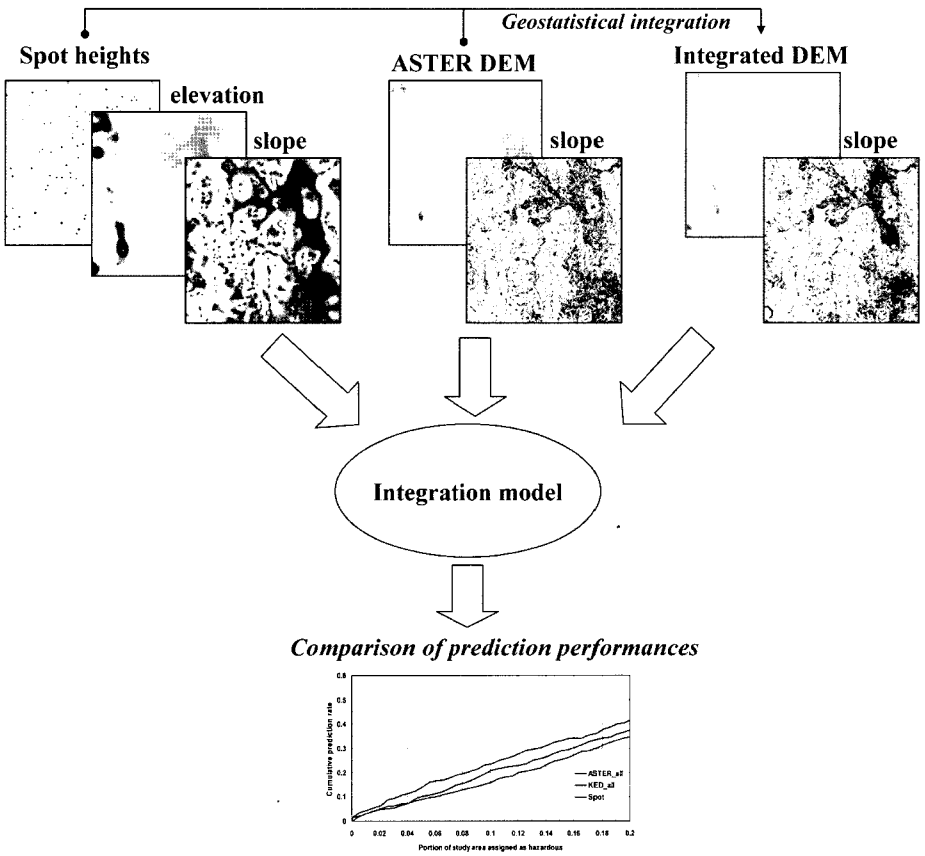

Fig. 1. Data processing flow applied in this study.

2. Study area and data set

A case study was conducted in Boeun area, Korea, which suffered heavy landslide damage following intense rainfall events in 1998 (Leet *et al.*, 2008). Among several GIS layers constructed by Lee *et al.* (2008), a landslide location map including 459 past landslides and an elevation map generated from ASTER stereo images were only considered to investigate the effects of terrain-related variables on landslide hazard mapping. The DEM was generated from ASTER 3N and 3B bands using the Y parallax, which comes from correlation analysis between the 3N band image of left side and the 3B image of right side (Lee *et al.*, 2008). The average elevation error reported in Lee *et al.*(2008) was about 6.88 m by comparison of a digital topographic map at 1:25,000 scale. In addition, 1292 elevation spot height points were extracted from the digital topographic map and were regarded as hard elevation data. The study area is discretized by a regular raster of 533 by 571 pixels, with a pixel size of 15m by 15m. The data sets used in this case study are shown in Fig. 2.

3. Case study

1) Geostatistical integration of different sources of elevation

Before integrating hard elevation spot height data

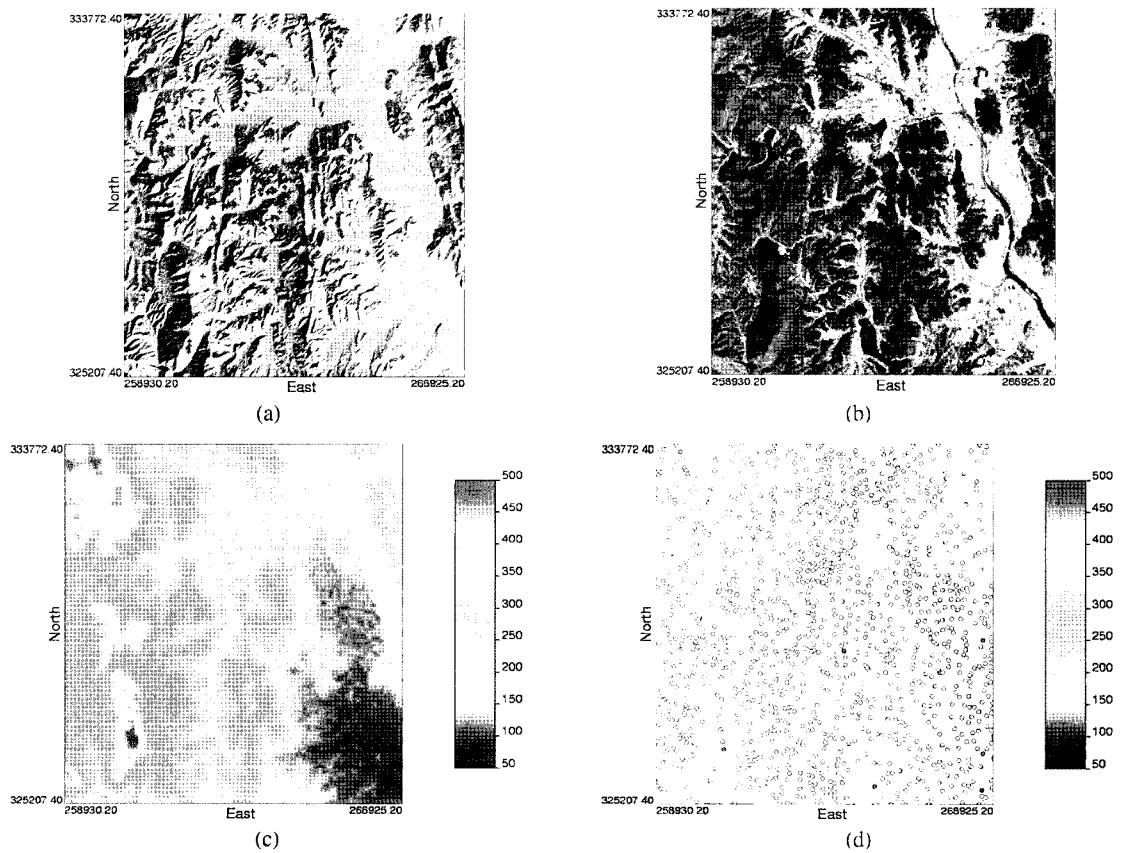

Fig. 2. Data set used in this study, (a) locations of past landslides, (b) ASTER imagery, (c) DEM from ASTER, (d) locations of elevation spot heights.

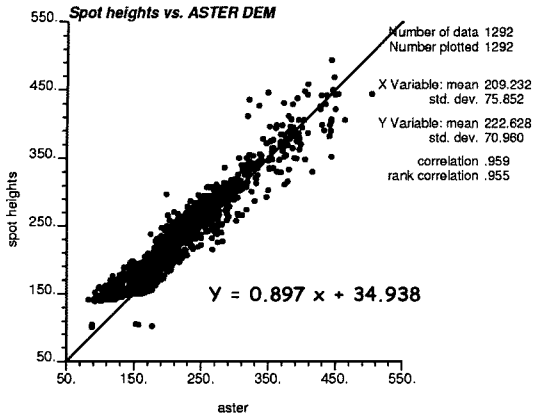

Fig. 3. Scatterplot of elevation values from spot heights and ASTER-based DEM.

with soft ASTER-based elevation data, we first examined how strong elevation values from ASTER were correlated with those at hard spot height locations. A strong linear relationship (i.e. $r \sim 0.96$) could be observed from the scatterplot shown in Fig. 3. From this linear relationship between hard and soft data, it might be expected that the integration of ASTER-based elevation would result in the improvement of accuracy of elevation values estimated at non-spot heights locations.

As for the integration of sparsely sampled hard spot height data with exhaustively sampled ASTER-based data, kriging with an external drift (KED) was chosen among various multivariate kriging algorithms. KED is an extension of kriging to accommodate a linear regression model and a smoothly varying secondary variable is used to derive the trend of the primary variable (Deutsch and Journel, 1998).

Let $y(\mathbf{u})$ be the secondary variable which corresponds to the ASTER-based elevation value at location $\mathbf{u}$. Then the trend model $m(\mathbf{u})$ is as follows:

$$m(\mathbf{u}) = a_0 + a_1\, y(\mathbf{u}) \qquad (1)$$

The KED estimator of the z variable $Z^*_{KED}(\mathbf{u})$ and the corresponding kriging system are defined as (Deutsch and Journel, 1998; Goovaerts, 1997):

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

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta^{KED}(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_0^{KED}(\mathbf{u}) + \mu_1^{KED}(\mathbf{u}) y(\mathbf{u}_\alpha) \\ = C_R(\mathbf{u}_\alpha - \mathbf{u}), \alpha = 1, ..., n \\ \sum_{\beta=1}^{n} \lambda_\beta^{KED}(\mathbf{u}) = 1 \\ \sum_{\beta=1}^{n} \lambda_\beta^{KED}(\mathbf{u}) y(\mathbf{u}_\beta) = y(\mathbf{u}) \end{cases} \qquad (2)$$

where n, $\lambda_\alpha^{KED}(\mathbf{u})$, and μ are the number of samples, the kriging weights, and Lagrange parameters, respectively. $C_R(\mathbf{u})$ is the covariance of the residual component.

Unlike simple kriging with varying local means, the unknown regression coefficients for modeling the linear relationship between primary and secondary variables are locally estimated through the kriging system with each search neighborhood (Goovaerts, 1997). Interested readers should refer to Goovaerts (1997) and Kyriakidis *et al.* (1999) for a detailed description on KED and geostatistical integration for digital elevation models, respectively.

For comparison purposes, elevation values at unsampled locations were estimated through ordinary kriging (OK) using only the 1292 spot height elevation data. For OK, experimental variograms of spot height values were estimated and then variogram modeling was carried out. For KED, experimental variograms were estimated from residual values which can be computed by subtracting ASTER-based elevation values from spot heights values, and then modeled. The OK-based elevation map in Fig. 4 shows much smoother patterns of elevation values than those from ASTER and KED, which is a typical characteristic of kriging.

Leave-one-out cross validation was carried out to compare the prediction performances of OK and KED. In our work, the difference between true and estimated elevation values was quantified using the

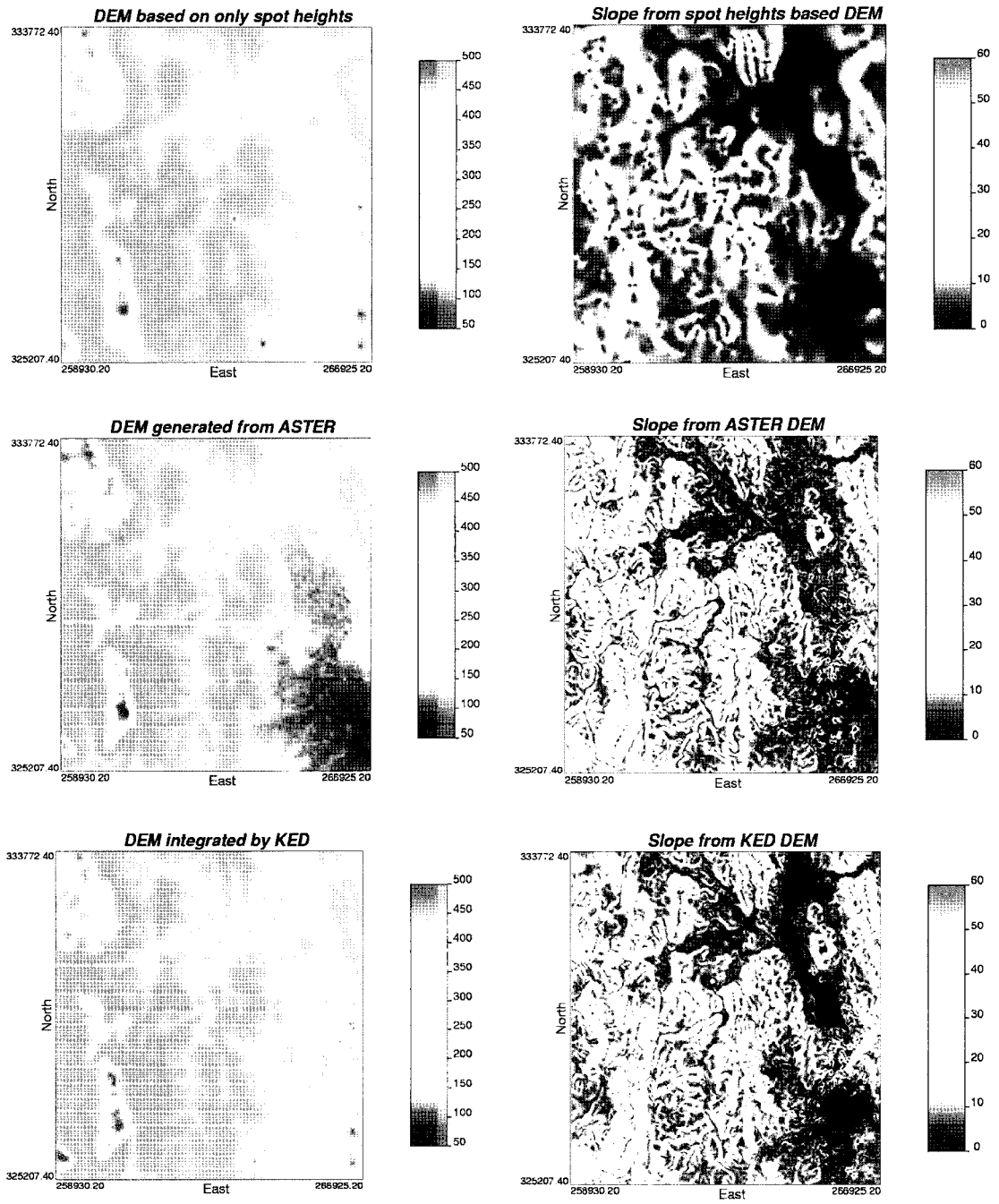

Fig. 4. Elevation and slope maps derived from three different sources of elevation.

mean absolute error (MAE). As expected from Fig. 3, the integration of soft data which have a strong linear relationship with the hard data led to the improvement of prediction performance (Table 1). The relative improvement of MAE in percentage of KED over OK are 27.04 %. This result means that accounting for ASTER-based elevation data can complement the sparsely sampled spot height observations and thus

Table 1. Comparison of the prediction performances of OK and KED

	MAE	Relative improvement of MAE*
OK	12.47	-
KED	9.10	27.04

* Relative improvement of MAE (%) = $(MAE_{OK} - MAE_{KED}) / MAE_{OK} \times 100$

improve the accuracy of elevation values estimated at non-spot height locations.

For subsequent for landslide hazard mapping purposes, three slope maps were generated from three different elevation maps using 1) OK of spot heights only, 2) ASTER-based DEM, 3) KED of spot heights and ASTER-based DEM (Fig. 4).

2) Data integration for landslide hazard mapping

Landslide hazard maps in the study area were constructed by integrating different elevation and slope maps to investigate the effects of different accuracy arising from different terrain-related maps on the prediction performance of future landslide hazard.

The integration methodology consisted of a likelihood ratio model (Chung, 2006) with empirical kernel density estimation. For elevation and slop maps, the respective two empirical density functions from landslide areas and non-landslide areas were computed and then the density ratio was taken to highlight the differences between two empirical distributions. Note that the spread parameter in the Gaussian kernel function for density estimation was experimentally chosen at 2%. Empirical density functions and likelihood ratio values for the different terrain-related variables are shown in Figs. 5 and 6. In the case of elevation, the likelihood ratio values in all three elevation maps were greater than one, in the range of elevation values between 180m and 280m (Fig. 5). This means that most landslides occurred in areas between those altitudes and thus the probability

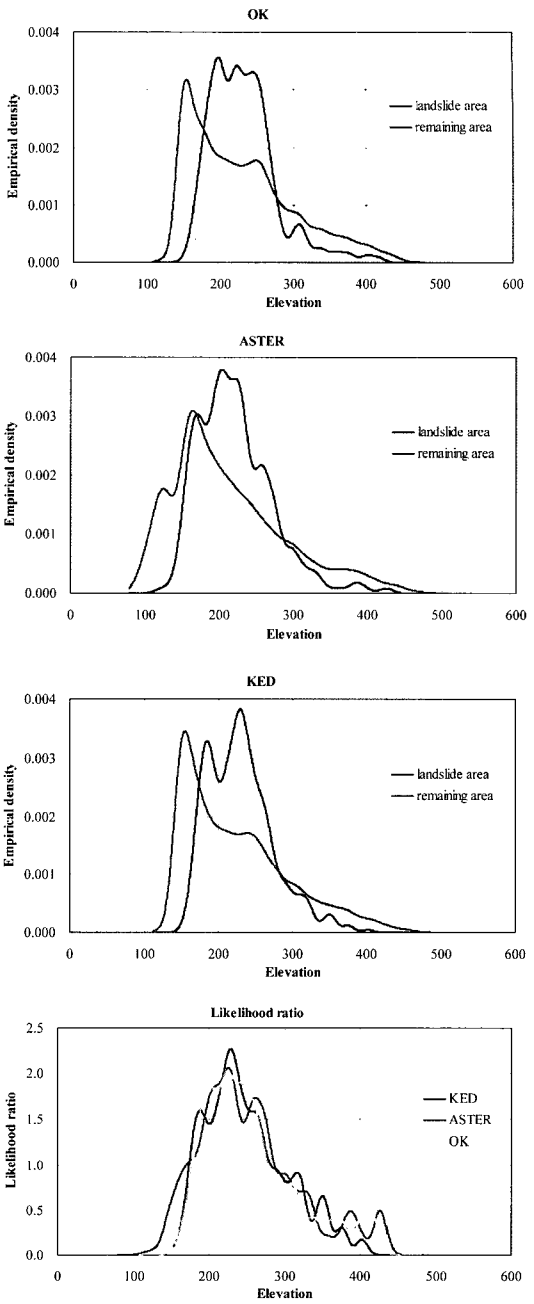

Fig. 5. Empirical density functions and likelihood ratio values of elevation based on different sources of elevation data.

of landslide occurrence is high for that elevation range. When comparing the three elevation maps, the overall patterns from OK and KED are similar, but a broader distribution of non-landslide areas was observed in ASTER-based elevation.

In the case of slope, slope values from OK of spot heights are located in a narrower range, compared with the range of ASTER- and KED-based slopes (Fig. 6). This trend is caused by the smoother values of OK-based elevations. That is, similar elevation values in local neighborhood resulted in small slope values. In the OK-based slope map, most landslides occurred in areas where the slope angle falls in the range between 8 and 28 degrees. On the contrary, the likelihood ratio values from ASTER- and KED-based slopes increase according to the slope angle. The only difference between those two slopes is observed in areas where the slope value is greater than 48 degrees. From Figs. 5 and 6, we can observe differences due to the three different sources of elevation in landslide and non-landslide areas. Such differences may affect the final integration results and lead to different prediction performances when it comes to landslide hazard mapping.

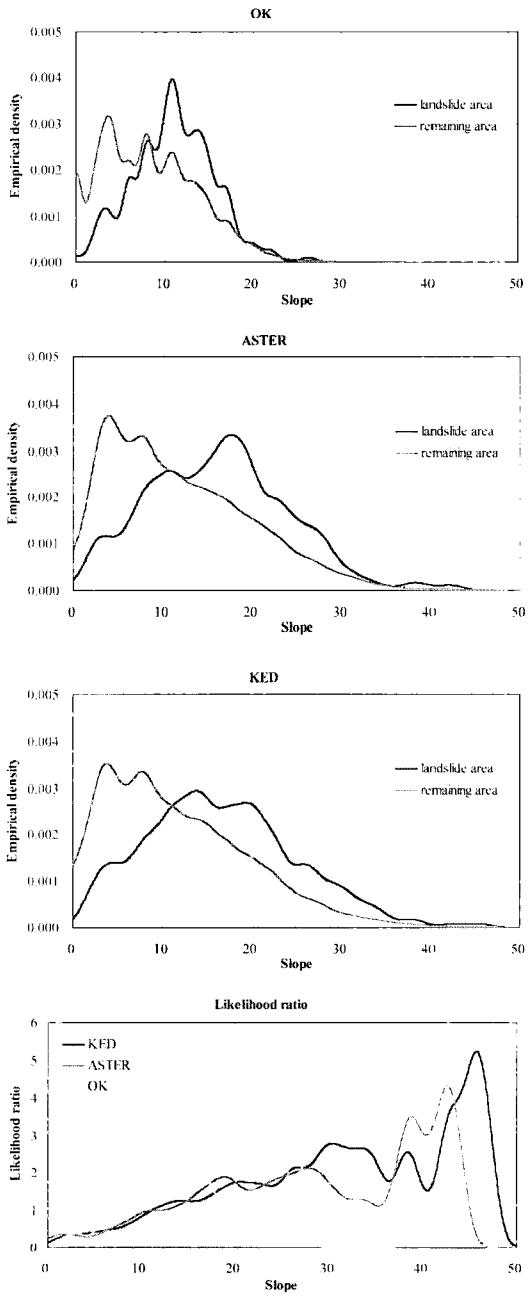

Fig. 6. Empirical density functions and likelihood ratio values of slope based on different sources of elevation data.

The likelihood ratio values from elevation and slope maps were finally multiplied to obtain the joint likelihood ratio values under the assumption of conditional independence. These joint likelihood ratio values were transformed into rank values to visualize relative hazard levels in the study area (Fig. 7). The overall patterns of hazard levels in the three hazard maps are very similar to those of slope values shown in Fig. 4. Since the slope patterns were computed from different sources of elevation, it is anticipated that the prediction performance associated with the resulting three hazard maps would be different.

3) Validation results

To quantitatively evaluate the prediction performance of the three different landslide hazard maps shown in Fig. 7, a cross-validation approach based on random spatial partitioning of past landslides was carried out. 459 past landslides were first randomly divided into 2 groups including 230 and 229 landslides. Then, the landslide hazard map was generated using group 1 as the training group and rank-based relative hazard values at landslide locations in group 2 that were not used for

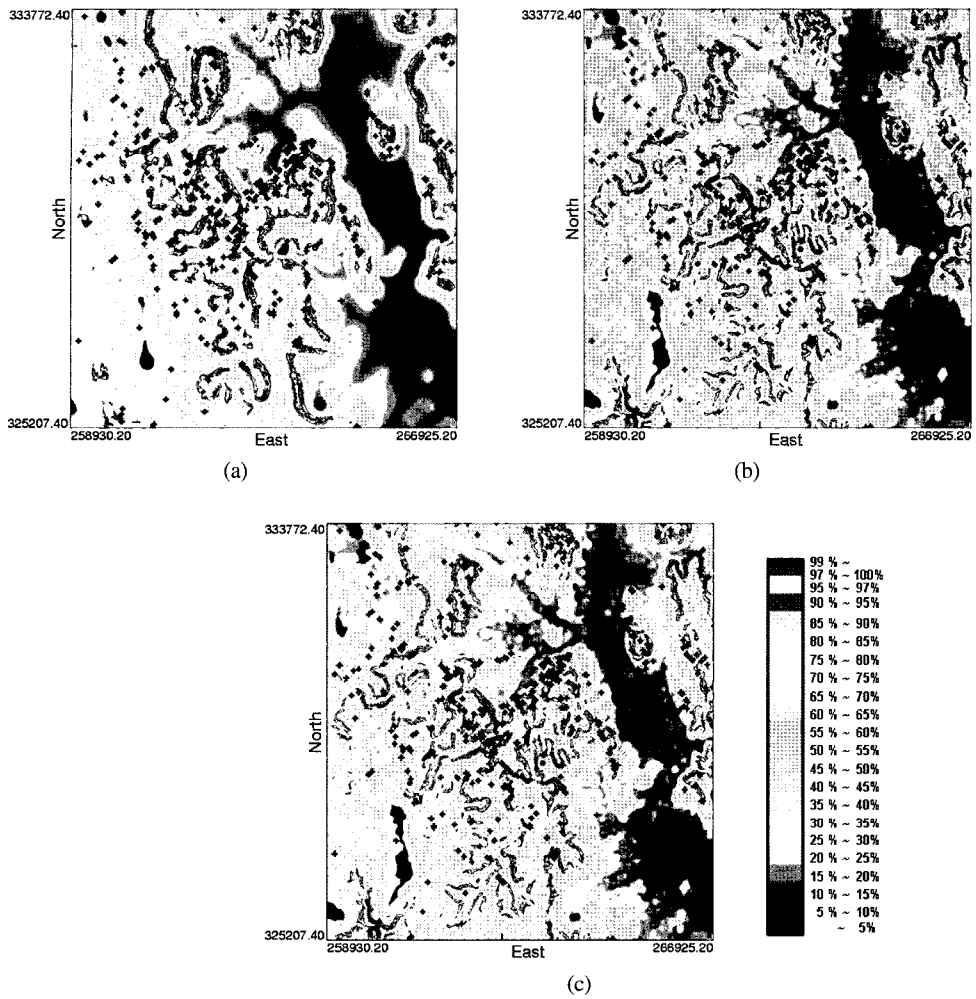

Fig. 7. Landslide hazard maps based on elevation and slope, from: (a) OK of spot heights, (b) ASTER, (c) KED of spot heights and ASTER elevation. Black dots denote past landslide locations.

constructing the landslide hazard map were computed and stored. By changing group 1 into group 2, the same experiment was repeated. Finally, the prediction rate curve (Chung and Fabbri, 1999) was computed from relative hazard values at all landslide locations as a quantitative prediction of future possible landslides.

The cross-validation results based on different sources of elevation are shown in Fig. 8. The best prediction performance was obtained from KED-based elevation and slope. The map of elevation constructed using OK of spot heights only showed the worst prediction rate values. If the most hazardous 10% of the area is considered, then about 24% of the landslides are located in the KED-based landslide hazard map. In the case of ASTER- and OK-based hazard maps, about 21% and 16% of landslides are located in that area, respectively. When interpreting the prediction rate curve, the prediction rate values in the upper most categories (e.g. top 20% area) are of greatest interest. The superiority of the KED-based hazard map can be easily found in the pattern of the prediction rate curve within the top 30% shown in Fig. 8(b). By using elevation estimates which can

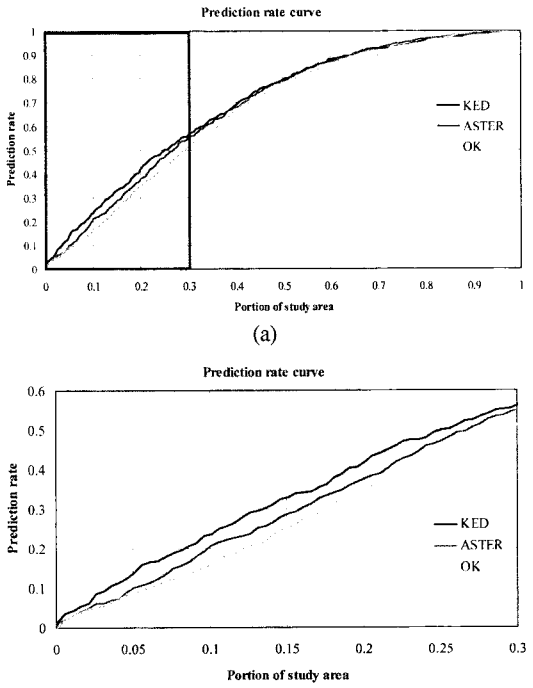

Fig. 8. (a) Prediction rate curve, (b) enlarged prediction rate curve within the black rectangle in (a).

account for both sparsely sampled hard spot heights and satellite-based soft elevation data, more realistic topographic data with less uncertainty could be generated and thus the best prediction performance could be obtained. When considering that ground-based field surveys are limited by the cost of sampling and accessibility, soft information from remote sensing data, which provides exhaustive information over the area of interest, would be a useful information source for thematic mapping.

4. Conclusions

GIS-based spatial data integration tasks have used exhaustive thematic maps generated from sparsely sampled data or satellite-based data. Due to a simplification of reality and error in mapping procedures, such spatial data are usually imperfect and of different accuracy. As a result, uncertainty from input spatial data may propagate through spatial data integration procedures and affect the final interpretation of integration results. Thus, analysis of this kind of effect in connection with input data should be properly done to obtain the best reasonable interpretation.

In this paper, the effects of different sources of elevation estimates on spatial prediction of landslide hazard were investigated. Three different sources of elevation were considered; 1) ground-based sparsely sampled spot height data, 2) exhaustive elevation data derived from ASTER stereo images and 3) an integrated elevation data set which accounted for above two different sources of elevation and was generated via geostatistical kriging. In this case study, more reliable and less uncertain elevation and slope values derived from geostatistical integration led to the best prediction performance for landslide hazard mapping purposes, compared with landslide hazard maps based on only spot heights or only ASTER-based elevation.

To thoroughly investigate the effects of input data in GIS-based spatial data integration tasks, as well as the degree in which soft data from space observation can complement hard data as a function of the sampling density of the former, a stochastic simulation framework will be developed in the near future.

Acknowledgments

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-611-C00003). The work of N.-W. Park was in part supported by a grant (08KLSGC03) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded

by Ministry of Land, Transport and Maritime Affairs. The authors would like to thank Drs. Saro Lee and Sung-Soon Lee at the Korea Institute of Geoscience and Mineral Resources for providing spatial data sets used in this work.

References

Chung, C. F., 2006. Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment, *Computers & Geosciences*, 32(8): 1052-1068.

Chung, C. F., and A. G. Fabbri, 1999. Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing*, 65(12): 1389-1399.

Deutsch, C. V. and A. G. Journel, 1998. *GSLIB: Geostatistical Software Library and User's Guide*, 2nd Edition, Oxford University Press, New York, NY, USA.

Ercanoglu, M. and C. Gokceoglu, 2002. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach, *Environmental Geology*, 41: 720-730.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, NY, USA.

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

Lee, S., J. H. Ryu, M. J. Lee, and J. S. Won, 2006. The application of artificial neural networks to landslide susceptibility mapping at Janghung, Korea, *Mathematical Geology*, 38(2): 199-220.

Lee, S., H. J. Oh, N. W. Park, and S. S. Lee, 2008. Extraction of landslide-related factors from ASTER imagery and its application to landslide susceptibility mapping using GIS, *Geomorphology*, under revision.

Park, N. W. and K. H. Chi, 2008. Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model, *International Journal of Remote Sensing*, 29(1): 247-264.