Environ Geol (2006) 50: 1183–1191
DOI 10.1007/s00254-006-0290-5

ORIGINAL ARTICLE

Ki-Dong Kim
Saro Lee
Hyun-Joo Oh
Jong-Kuk Choi
Joong-Sun Won

Assessment of ground subsidence hazard near an abandoned underground coal mine using GIS

Received: 17 October 2005
Accepted: 22 March 2006
Published online: 3 May 2006

K.-D. Kim · H.-J. Oh · J.-K. Choi
J.-S. Won
Department of Earth System Sciences,
Yonsei University, 134 Shinchon-dong,
Seodaemun-gu, Seoul 120-749, Korea

S. Lee (✉)
Geoscience Information Center, Korea
Institute of Geoscience and Mineral
Resources (KIGAM), 30, Gajung-Dong,
Yusung-Gu, Daejon 305-350, Korea
E-mail: leesaro@kigam.re.kr
Tel.: +82-42-8683057
Fax: +82-42-8619714

Abstract This study constructs a hazard map for ground subsidence around abandoned underground coal mines (AUCMs) at Samcheok City in Korea using a probability (frequency ratio) model, a statistical (logistic regression) model, and a Geographic Information System (GIS). To evaluate the factors related to ground subsidence, an image database was constructed from a topographical map, geological map, mining tunnel map, Global Positioning System (GPS) data, land use map, lineaments, digital elevation model (DEM) data, and borehole data. An attribute database was also constructed from field investigations and reports on the existing ground subsidence areas at the study site. Nine major factors causing ground subsidence were extracted from the probability analysis of the existing ground subsidence area: (1) depth of drift; (2) DEM and slope gradient; (3) groundwater level, permeability, and rock mass rating (RMR); (4) lineaments and geology; and (5) land use. The frequency ratio and logistic regression models were applied to determine each factor's rating, and the ratings were overlain for ground subsidence hazard mapping. The ground subsidence hazard map was then verified and compared with existing subsidence areas. The verification results showed that the logistic regression model (accuracy of 95.01%) is better in prediction than the frequency ratio model (accuracy of 93.29%). The verification results showed sufficient agreement between the hazard map and the existing data on ground subsidence area. Analysis of ground subsidence with the frequency ratio and logistic regression models suggests that quantitative analysis of ground subsidence near AUCMs is possible.

Keywords Ground subsidence · Abandoned underground coal mine · Frequency ratio · Logistic regression · GIS · Korea

Introduction

Since 1989, almost all underground coal mines have been abandoned and few remain near Samcheok City, Korea. The occurrence of ground subsidence around the abandoned coal mines has recently become a serious problem. However, quantitative assessment of predicted ground subsidence areas is difficult, especially in coal mining areas where the structures of the geology and mining are very complex. A method that predicts the probability of ground subsidence empirically, within surprisingly narrow limits considering the form of the input data, has been suggested (Goel and Page 1982) using: (1) the intact strength of the rock, (2) the stress field, (3) the geological structure of the rock, (4) the depth of the mining horizon, (5) the extent of the mined area, and (6) the volume extracted per unit area of mining. The National Coal Board has published a basic technique to determine the estimated area influenced by ground subsidence based on the height of the cavity, the width of the mined panel, and the angle of inclination of the coal seam (National Coal Board 1975). The method used to predict the subsidence area is very dependent on the structure of the local geology and the coal-mining method used, and the empirical methods discussed on top were developed for conditions involving horizontal coal seams and long wall mining, which are predominant in Europe. However, in Korea, due to the complicated geologic structure, there are coal seams of various widths and irregularly inclined coal seams and strata, so the slant-chute block caving method has been used. As a result, a sinkhole type of subsidence is usual, and therefore a different estimation of ground subsidence is necessary. Table 1 shows the factors that commonly affect sink-hole-type ground subsidence over time (Coal Industry Promotion Board [CIPB] 1997).

The aim of this study was to assess and predict ground subsidence for hazard mapping near an abandoned underground coal mine (AUCM) area using a Geographic Information System (GIS). To choose a study area, field investigations and reports related to ground subsidence were carefully considered. A site called Simpori was chosen, where 48 indications of ground subsidence have been identified near an AUCM at Samcheok City. The study site is between longitudes 129°00′ and 129°03′ and latitudes 37°11′ and 37°12′. The coal resource of South Korea consists almost entirely of anthracite, 85% of which was deposited during the upper Paleozoic era and the lower Mesozoic era in the Jangseong Formation of the Pyeongan Supergroup (Geological Society of Korea 1999). The study site is around the Hanyang Gallery on the Jangseong and Keumcheon Formation. The Oship Fault, Youngdong railroad, and no. 38 local road pass along the study area (CIPB 1999). The location map of this study site with ground subsidence areas is given in Fig. 1.

Table 1 Factors affect sink-hole type ground subsidence according to time

Occurrence of ground subsidence During time after abandoned mine	Progress	Ground collapse
• Mechanical character of rock mass	• Flow of ground water	• Mining depth
• Flow of ground water	• Structure of geology	• Height of cavity
• Structure of geology (joint, fault, dyke)	• Rate of cubical expansion	
• Caving method	• Rate of mining	
• Rate of caving		
• Back filling		

Spatial database and methodology

Many studies have identified important factors that contribute to ground subsidence around coal mines, including (CIPB 1997; Waltham 1989): depth and height of the mined cavities, excavation method, degree of inclination of the excavation, scope of mining, structure of the geology, flow of groundwater, and the mechanical characteristics of the rock mass rating (RMR). Therefore, the factors related to the occurrence of ground subsidence were collected in a vector-type spatial database. These included a 1:50,000 scale geological map, 1:5,000 scale topographic maps, 1:5,000 scale land use maps, 1:1,200 scale mined-tunnel maps, and borehole data. The data layers are shown in Table 2.

The geology data were extracted from a 1:50,000 scale geological map and the distances from lineaments were calculated using the lineament data. Contour and survey base points with elevation values read from the topographic map were extracted, and a digital elevation model (DEM) was constructed. Using the DEM, the slope gradients were calculated. There are seven classes of land use, which were extracted from the land use map of the National Geographic Institute. Most of the literature maintains that the major factor in ground subsidence is the scope of the mined cavities. Therefore, constructing a database of the depths and widths of mined cavities was very important. To achieve this object, (1) GPS measurements were used to determine the exact positions of mine heads; (2) these were used to vectorize a hard copy of the mined tunnel map; and (3) the vectorized mined tunnel map was converted to an ASCII grid file, and subtracted with the DEM raster data. There were 37 boreholes at the study site, but some boreholes did not have values, so an inverse distance weighting (IDW) interpolation method was used to contour the values for RMR, groundwater levels, and permeability factors, and these were reclassified using GIS.

This study was conducted using GIS, a frequency ratio, and logistic regression with factors that may cause ground subsidence. A flow chart describing the process of this study is shown in Fig. 2. An image database and an attribute database for ground subsidence were constructed. A key assumption using this approach is that the potential ground subsidence (occurrence possibility) will be similar to the actual frequency of ground subsidence. After the study site was selected, areas of ground subsidence were detected at the study site by field surveys. A map of existing ground subsidence was developed, and this was used to evaluate the frequency and distribution of ground subsidence at the study site.

Frequency ratio and logistic regression models were used to represent the distinction quantitatively. For this analysis, the calculated and extracted factors were mapped to a 1-m resolution grid. The raster data were

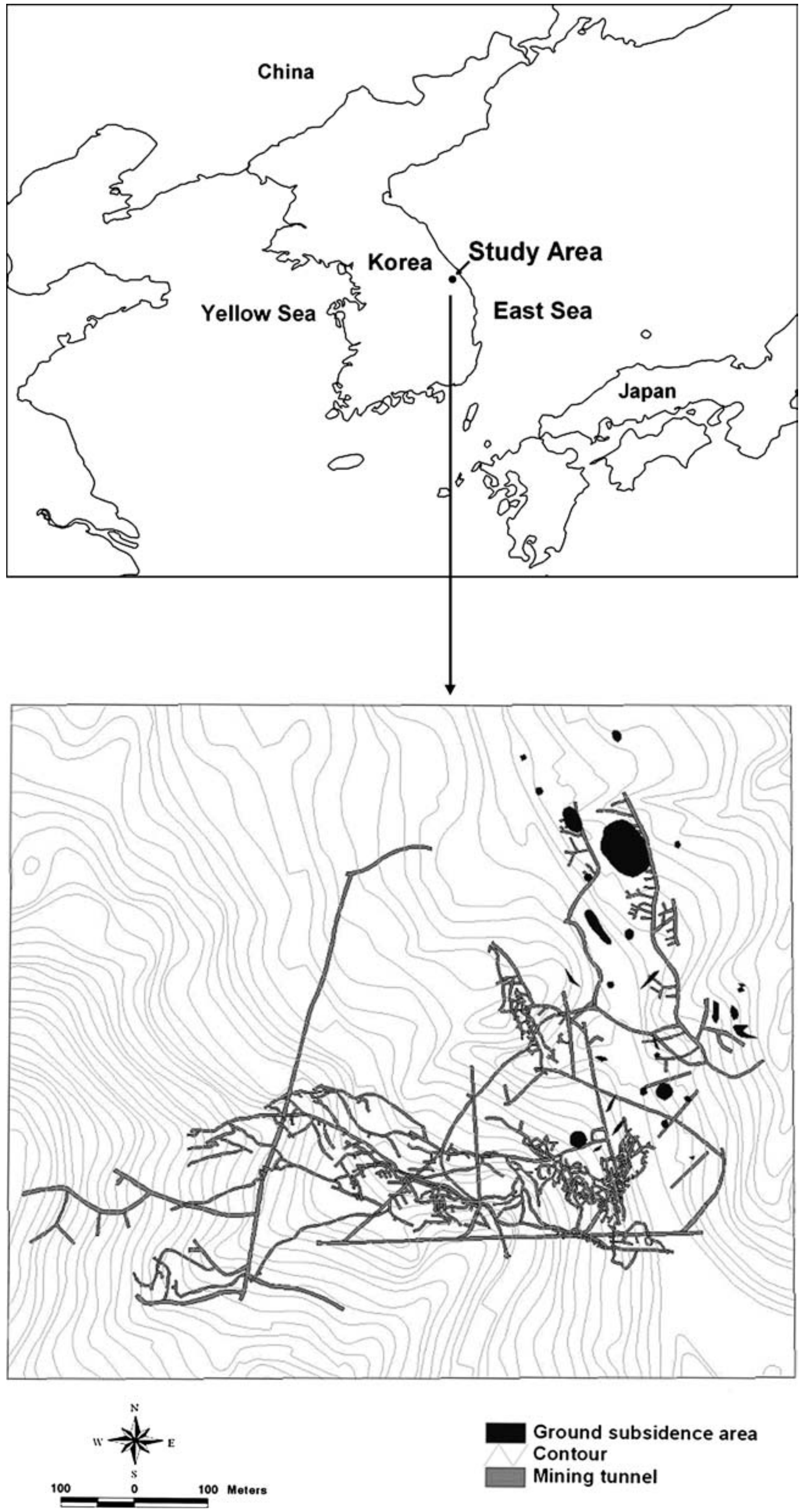

Fig. 1 Study area near Samcheok city, Korea

Table 2 Constructed GIS database including factors conneted with ground subsidence of study area

Category	Factors	Remark
Geology	Geology	Type of strata
	Distance from lineaments	Bufferig of lineament
Topography	DEM	TIN process to get elevation data
	Slope	Analyze slope by degree
Land use	Land use	Classification of landuse types
Mining tunnel map	Depth of drift	DEM minus sea level of drift
	Height of drift	Disregard this factor. Almost 0.8 –1 m along study area
Borehole[a]	RMR	IDW interpolated from 22 bore holes
	Ground water level	IDW interpolated from 12 bore holes
	Permeability	IDW interpolated from 15 bore holes

[a]Thirty-seven boreholes from investigation in Coal Industry Promotion Board (1999), some boreholes do not have value of relating factors

Abbreviations: GIS Geographic Information System; *DEM* digital elevation model; *RMR* rock mass rating; *IDW* inverse distance weighting; *TIN* triangulated irregular network

converted for the statistical program used. Then, using the frequency ratio and logistic regression models, the spatial relationships between the ground subsidence area and each ground subsidence-related factor, such as topography, depth of the mined tunnel, borehole data, geology, and land use, were analyzed with the statistical program, and a formula for the possibility of ground subsidence was extracted using the relationships. The formula was used to calculate the subsidence hazard index (SHI), which was mapped to each grid cell. Finally, the hazard map was verified using known ground subsidence areas. Success rates were calculated for quantitative verification. In this study, the GIS software ArcView 3.3 and ARC/INFO version 9.0 and the statistical software SPSS 12.0 were used as the basic analysis tools for spatial management and data manipulation.

Frequency ratio approaches are based on the observed relationships between distribution of ground subsidence areas and each subsidence-related factor, to reveal the correlation between subsidence locations and the factors. Using the frequency ratio model, the spatial relationships between subsidence occurrence location and each factor's contributing ground subsidence occurrence were derived. The frequency ratios of each factor's type or range were calculated from their relationship with ground subsidence events. In the relation analysis, the ratio is that of the area where ground subsidence occurred to the total area, so that a value of 1 is an average value. If the value is greater than 1, it means a higher correlation, and a value lower than 1 means a lower correlation.

Logistic regression allows formation of a multivariate regression relation between a dependent variable and several independent variables. Logistic regression, which

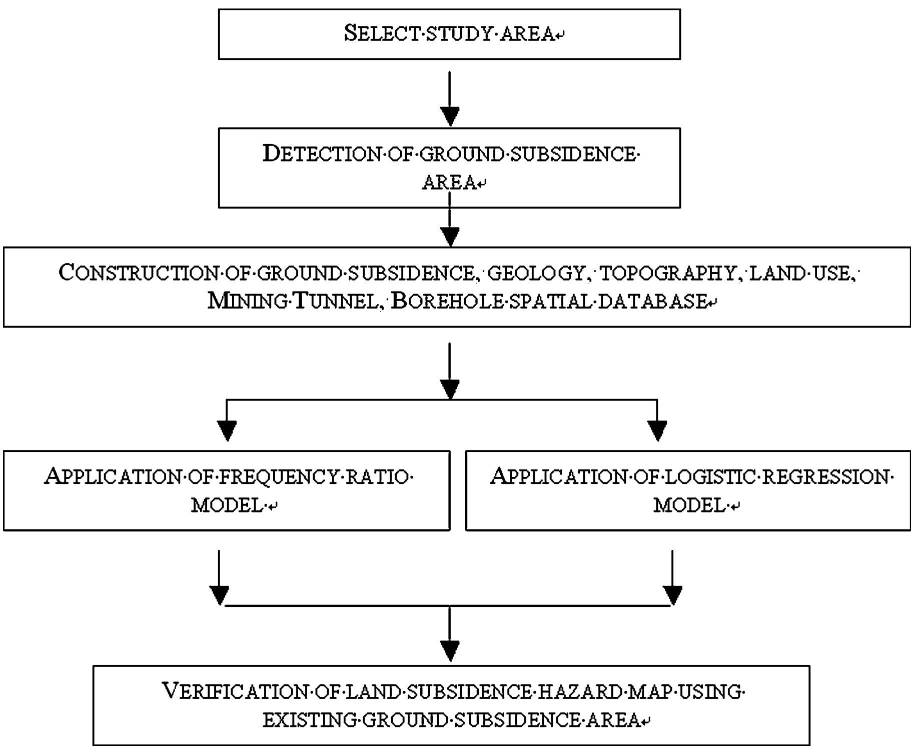

Fig. 2 Flow chart of study methodology

Table 3 Frequency ratio and coefficient value of each factor

Factor	Class	No. of pixels in domain[a]	Percentage of domain	No. of ground subsidence[b]	Percentage of ground subsidence	Ratio[c]	Logistic regression coefficients
Slope	0–9	39,415	10.31	62	2.12	0.21	−0.034
(Unit: degree)	10–13	48,432	12.67	1,113	38.08	3.01	
	14–16	40,298	10.54	328	11.22	1.06	
	17–19	37,738	9.87	330	11.29	1.14	
	20–22	40,135	10.50	394	13.48	1.28	
	23–25	39,177	10.25	200	6.84	0.67	
	26–28	39,788	10.41	123	4.21	0.40	
	29–32	39,975	10.46	280	9.58	0.92	
	33–37	33,314	8.72	29	0.99	0.11	
	38–65	23,962	6.27	64	2.19	0.35	
Distance from Drift	0–2	48,303	12.64	808	27.64	2.19	−0.095
(Unit: m)	3–7	44,382	11.61	965	33.01	2.84	
	8–14	37,867	9.91	845	28.91	2.92	
	15–25	38,488	10.07	190	6.50	0.65	
	26–39	36,894	9.65	50	1.71	0.18	
	40–54	36,401	9.52	51	1.74	0.18	
	55–73	36,494	9.55	14	0.48	0.05	
	74–100	35,338	9.25	0	0.00	0.00	
	101–154	34,283	8.97	0	0.00	0.00	
	155–318	33,784	8.84	0	0.00	0.00	
Depth of drift	0	334,059	87.40	2,048	70.07	0.80	0.001
(Unit: m)	3–57	9,925	2.60	831	28.43	10.95	
	58–122	4,925	1.29	0	0.00	0.00	
	123–139	5,200	1.36	10	0.34	0.25	
	140–152	5,075	1.33	0	0.00	0.00	
	153–159	4,750	1.24	8	0.27	0.22	
	160–169	4,775	1.25	23	0.79	0.63	
	170–182	4,900	1.28	3	0.10	0.08	
	183–196	4,500	1.18	0	0.00	0.00	
	197–237	4,125	1.08	0	0.00	0.00	
Depth to ground water level	15–27	39,009	10.21	1,527	52.24	5.12	−0.079
(Unit: m)	28–32	40,303	10.54	878	30.04	2.85	
	33–37	42,184	11.04	518	17.72	1.61	
	38–47	39,508	10.34	0	0.00	0.00	
	48–56	37,364	9.78	0	0.00	0.00	
	57–69	39,005	10.20	0	0.00	0.00	
	70–79	38,340	10.03	0	0.00	0.00	
	80–87	36,498	9.55	0	0.00	0.00	
	88–98	35,777	9.36	0	0.00	0.00	
	99–160	34,246	8.96	0	0.00	0.00	
Geology	Hongjeom series	67,026	17.54	0	0.00	0.00	−12.583
	Sadong series	315,208	82.46	2,923	100.00	1.21	0
Landuse	Field	64,212	16.80	1,063	36.37	2.16	17.296
	River	2,395	0.63	28	0.96	1.53	16.147
	Road	24,795	6.49	42	1.44	0.22	15.544
	Hybrid land	12,841	3.36	51	1.74	0.52	14.897
	Right-of-way	38,709	10.13	114	3.90	0.39	15.865
	Wood land	236,795	61.95	1,625	55.59	0.90	0
	Plot	2,487	0.65	0	0.00	0.00	14.777
Dem	371–415	39,372	10.30	441	15.09	1.46	−0.036
(Unit: m)	416–441	38,381	10.04	2,029	69.41	6.91	
	442–455	39,297	10.28	206	7.05	0.69	
	456–474	40,029	10.47	172	5.88	0.56	
	475–490	39,317	10.29	75	2.57	0.25	
	491–510	37,941	9.93	0	0.00	0.00	
	511–528	37,817	9.89	0	0.00	0.00	
	529–552	37,047	9.69	0	0.00	0.00	
	553–584	36,690	9.60	0	0.00	0.00	
	585–653	36,343	9.51	0	0.00	0.00	

Table 3 (Contd.)

Factor	Class	No. of pixels in domain[a]	Percentage of domain	No. of ground subsidence[b]	Percentage of ground subsidence	Ratio[c]	Logistic regression coefficients
Permeability	400–430	39,997	10.46	1,548	52.96	5.06	−1.338
	431–435	61,965	16.21	440	15.05	0.93	
	436	70,549	18.46	51	1.74	0.09	
	437–438	55,968	14.64	92	3.15	0.21	
	439	33,137	8.67	49	1.68	0.19	
	440–442	28,968	7.58	104	3.56	0.47	
	443–447	26,547	6.95	52	1.78	0.26	
	448–452	25,129	6.57	103	3.52	0.54	
	453–461	20,841	5.45	308	10.54	1.93	
	461–449	19,133	5.01	176	6.02	1.20	
RMR	300–338	38,256	10.01	0	0.00	0.00	0.368
	339–354	39,577	10.35	0	0.00	0.00	
	355–368	39,497	10.33	16	0.55	0.05	
	369–389	39,407	10.31	19	0.65	0.06	
	390–401	41,822	10.94	61	2.09	0.19	
	402–411	38,276	10.01	204	6.98	0.70	
	412–420	39,164	10.25	470	16.08	1.57	
	421–433	37,498	9.81	241	8.24	0.84	
	434–446	34,797	9.10	1,635	55.94	6.14	
	447–479	33,940	8.88	277	9.48	1.07	

[a]Number of total cells in study area: 382,234
[b]Number of ground subsidence cells: 2,923
[c]Percentage ground subsidence/percentage domain
RMR rock mass rating

is one of the multivariate analysis models, is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. The advantage of logistic regression is that through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types and they do not necessarily have normal distributions. In the case of multi-regression analysis, the factors must be numerical, and in the case of a similar statistical model, discriminant analysis, the variables must have a normal distribution. In the present situation, the dependent variable is a binary variable representing presence or absence of ground subsidence. Where the dependent variable is binary, the logistic link function is applicable (Atkinson and Massari 1998). For this study, the dependent variable must be input as either 0 or 1, so the model applies well to ground subsidence possibility analysis. Logistic regression coefficients can be used to estimate ratios for each of the independent variables in the model.

Quantitatively, the relationship between the occurrence and its dependency on several variables can be expressed as:

$$P = 1/(1 + e^{-z}) \tag{1}$$

where P is the probability of an event occurring. In the present situation, the value P is the estimated probability of subsidence occurrence. The probability varies from 0 to 1 on an S-shaped curve and z is the linear combination. It follows that logistic regression involves fitting an equation of the following form to the data:

$$Z = B_0 + B_1X_1 + B_2X_2 + \ldots + B_NX_N \tag{2}$$

where b_0 is the intercept of the model, the b_i ($i = 0, 1, 2, ..., n$) are the slope coefficients of the logistic regression model, and the x_i ($i = 0, 1, 2, ..., n$) are the independent variables. The linear model formed is then a logistic regression of the presence or absence of ground subsidence (present conditions) on the independent variables (pre-failure conditions).

Application of the frequency ratio and logistic regression models

To calculate the frequency ratio, the area ratio of ground subsidence occurrence to non-occurrence was calculated for the class or type of each factor, and an area ratio for the class or type of each factor to the total area was calculated. The frequency ratios for the class or type of each factor were calculated by dividing the ground subsidence occurrence ratio by the area ratio. The frequency ratios are shown in Table 3. The frequency ratios for the type or class of each factor were summed to calculate the SHI, as shown in Eq. 3.

$$\mathrm{SHI} = \sum \mathrm{Fr} \text{ (where Fr = frequency ratio of each factor's type or class)} \quad (3)$$

A high SHI value indicates a high hazard of ground subsidence; a lower value indicates a lower risk of ground subsidence. The SHI is shown in Fig. 3 for visual interpretation. The index is classified into four classes based on equal areas for easy visual interpretation. The minimum value is 1.27 and the maximum value is 52.21, the mean value is 8.996 and the standard deviation is 6.773.

Using the logistic regression model, the spatial relationship between ground subsidence occurrence and factors influencing ground subsidence was assessed. The spatial databases for each factor were converted to ASCII format files for use in the statistical package, and the correlations between ground subsidence and each factor were calculated (Table 3). Logistic regression formulae were also created as shown in Eq. 4. Finally, the probability of subsidence occurrence was calculated using the spatial database, data from Table 2, and Eqs. 1 and 4:

$$\begin{aligned} z = {} & (-0.034 \times \mathrm{SLOPE}) + (-0.095 \times \mathrm{DISTDRIFT}) \\ & + (-0.079 \times \mathrm{WATERLEVEL}) + (-0.036 \times \mathrm{DEM}) \\ & + (-1.338 \times \mathrm{PERMEABILITY}) + (0.368 \times \mathrm{RMR}) \\ & + (0.001 \times \mathrm{DEPTH\ DRIFT}) \\ & + \mathrm{GEOLOGY}_b + \mathrm{LAND\ USE}_b + 3.952 \end{aligned} \quad (4)$$

where SLOPE is the slope; DISTDRIFT is the distance from the drift; WATERLEVEL is the depth to the groundwater level; DEM is the elevation; DEPTHDRIFT is depth of drift; PERMEABILITY is permeability value; RMR is the RMR value; $\mathrm{GEOLOGY}_b$ is the lithology type; and $\mathrm{LANDUSE}_b$ are the logistic regression coefficient values listed in Table 2 and z is a prediction parameter.

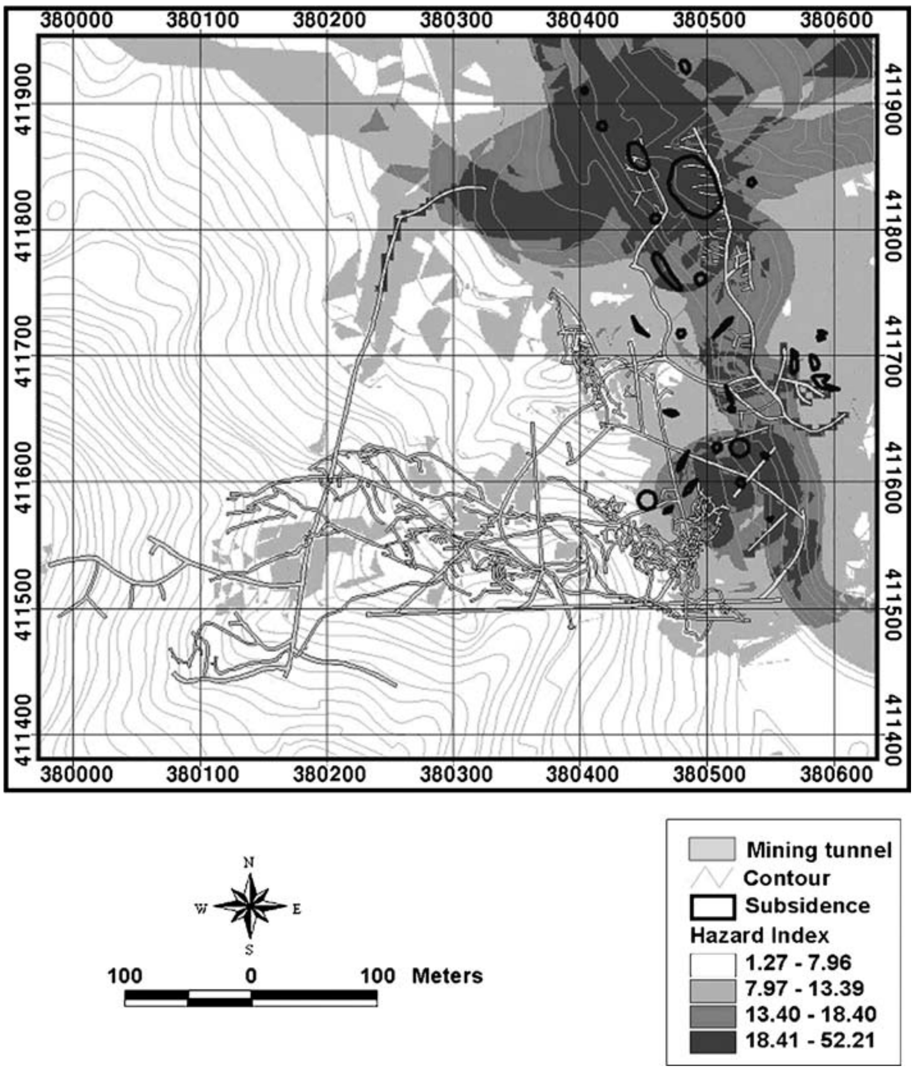

Fig. 3 Ground subsidence hazard map using frequency ratio model

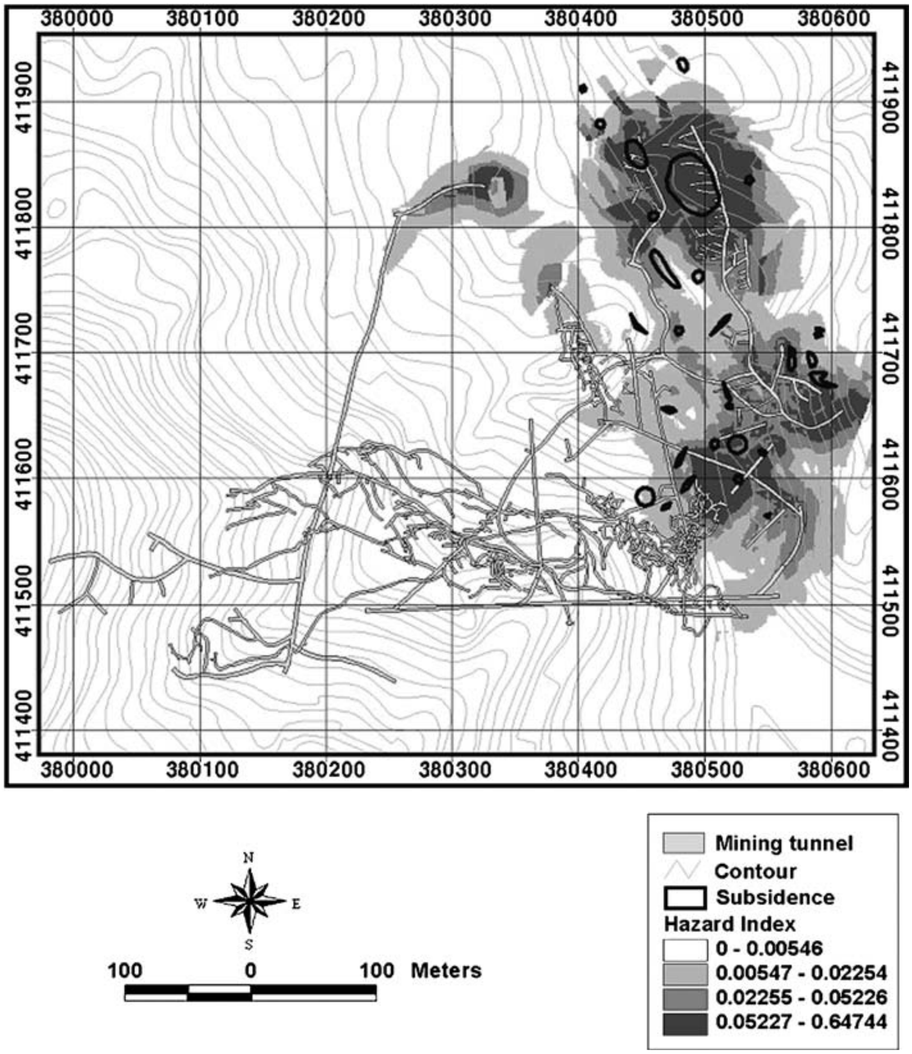

Fig. 4 Ground subsidence hazard map using logistic regression model plotted against percentage of probability

The possibility of subsidence was calculated using Eqs. 1 and 4 and a subsidence hazard map was constructed. The distribution of calculated possibilities is shown in Fig. 4. The possibilities were classified into four classes based on equal areas for easy visual interpretation. The minimum value is 0.00 and the maximum value is 0.64744, the mean value is 0.008961 and the standard deviation is 0.03482.

Verification of ground subsidence maps

The subsidence hazard analysis results were verified using known ground subsidence locations. Verification was performed by comparing the known ground subsidence location data with the subsidence hazard map. Each factor used and its frequency ratio was compared. Rate curves were created and the areas under the curves were calculated for two cases. The rate explains how well the model and the factor predict the subsidence. Thus, the area under the curve can quantitatively estimate the prediction accuracy. To obtain the relative rank for each prediction pattern, the calculated index values for all cells in the study area were sorted in descending order. The ordered cell values were then divided into 100 classes at accumulated 1% intervals. The rate verification results appear as a line in Fig. 5. For example, in the case of the frequency model used, the 90–100% (10%) class of the study area where the subsidence hazard index had a high rank could explain 73% of all subsidence. The 80–100% (20%) class of the study area where the subsidence hazard index had a high rank could explain 85% of subsidence. In the case of the logistic regression model used, the 90–100% (10%) class of the study area where the subsidence hazard index had a high rank could explain 79% of all subsidence. The 80–100% (20%) class of the study area where the subsidence hazard index had a high rank could explain 96% of subsidence. To compare the results quantitatively, the areas under the curves were

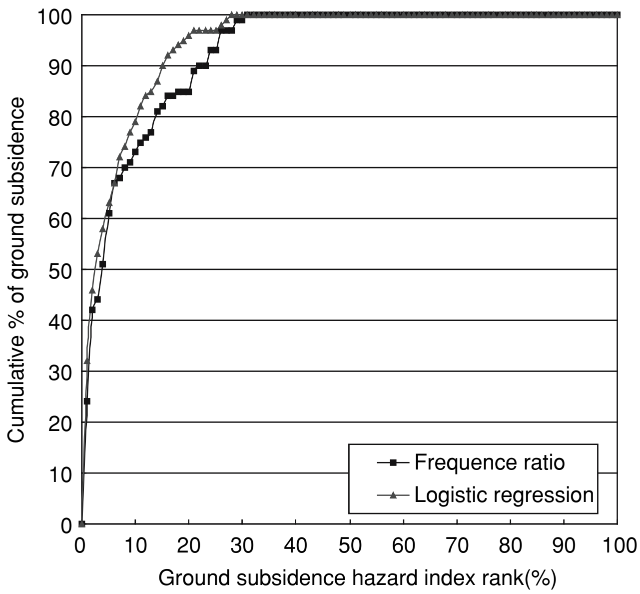

Fig. 5 Cumulative frequency diagram showing ground subsidence hazard rank occurring in cumulative percent of ground subsidence occurrence

recalculated as a total area of 1, which means perfect prediction accuracy.

Therefore, the area under a curve can be used to assess the prediction accuracy quantitatively. In the case of the logistic regression model used, the area ratio was 0.9501, so it can be said that the prediction accuracy is 95.01%. In the case of the frequency ratio model used, the area ratio was 0.9329, so it can be said that the prediction accuracy is 93.29%. Overall, the logistic regression model used showed a higher accuracy than the frequency ratio model used.

Results and discussion

Ground subsidence is among the most hazardous of artificial disasters. Government and research institutions worldwide have attempted for years to assess subsidence hazards and risks, and to show their spatial distribution. In this study, a statistical approach to identifying hazardous areas of subsidence using GIS shows considerable promise.

Ground subsidence maps were constructed using frequency ratio and logistic regression models. These showed very high prediction accuracy: 93.29 and 95.01% with the frequency ratio model and logistic regression model, respectively. Thus, the logistic regression model showed a better result than the frequency ratio model.

The subsidence areas of this study are around railroad, road, and other facilities above shallow mine workings. Therefore, the low elevation and depth of the mined tunnel are important factors, as well as the groundwater level. The data on groundwater levels were obtained during field surveys, without considering the amount of rainfall at the time. However, precipitation history is a meaningful value and should be considered in calculating the safety of a base rock. In further studies, exact maps of mined tunnels, and many borehole and geophysical data will be required to design an underground model of this study area to analyze ground subsidence more quantitatively.

The frequency ratio model is simple and the process of input, calculation, and output is easily understood. Moreover, because the frequency ratio value can be used as a rating, there is no need to convert the database to another format. A large amount of data can be processed in a GIS environment quickly and easily. The logistic regression model requires data to be converted to ASCII format for use in the statistical package and later reconverted to incorporate it into the GIS database. Moreover, large amounts of data cannot be processed by the statistical package quickly and easily. However, the degree of ground subsidence hazard can be analyzed quantitatively.

Acknowledgment The Coal Industry Promotion Board provided whole investigation reports and basic GIS database.

References

Atkinson PM, Massari R (1998) Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Compute Geosci 24:373–385

Coal Industry Promotion Board, CIPB (1997) A study on the Mechanism of subsidence over abandoned mine area and the construction method of subsidence prevention. Coal Industry Promotion Board, Seoul, 97-06:1–67

Coal Industry Promotion Board, CIPB (1999) Fundamental investigation report of the stability test for Simpori. Coal Industry Promotion Board, Seoul, 97-08:5–38

Goel SC, Page CH (1982) An empirical method for predicting the probability of chimney cave occurrence over a mining area. Int J Rock Mech Min, Sci Geomech Abstr 19:325–337

National Coal Board (1975) Subsidence engineer's handbook. National Coal Board Mining Department, London, p 111

The Geological Society of Korea (1999) Geology of Korea. Sigma Press, Seoul, 550–556

Waltham AC (1989) Ground subsidence. Blackie & Son Ltd, New York, pp 49–97