applied sciences

MDPI

Article

Analysis of Ground Subsidence Vulnerability in Urban Areas Using Spatial Regression Analysis

Sungyeol Lee *, Jaemo Kang and Jinyoung Kim

Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang-si 10223, Republic of Korea; jmkang@kict.re.kr (J.K.); goldcamp@kict.re.kr (J.K.)
* Correspondence: leesy@kict.re.kr; Tel.: +82-31-910-0645

Featured Application: Authors are encouraged to provide a concise description of the specific application or a potential application of the work. This section is not mandatory.

Abstract: The main cause of ground subsidence accidents in urban areas is cavities formed by damage to underground utilities. For this reason, the attribute information of underground utilities should be used to prepare against ground subsidence accidents. In this study, attribute information (pipe age, diameter, burial depth, and density) of six types of underground utilities (water, sewer, gas, power, heating, and communication) and history information of ground subsidence were collected. A correlation analysis was conducted using the collected data, and a prediction model of vulnerability to ground subsidence was developed through the ordinary least squares (OLS) method and spatial regression analysis (spatial lag model (SLM) and spatial error model (SEM)). To do this, the target area was divided into a grid of 100 m × 100 m. Datasets were constructed using the attribute information of underground utilities included in the divided grid and the number of ground subsidence occurrences. To analyze the OLS of the constructed data, the variance inflation factor (VIF) of the attribute information of underground utilities was studied. An OLS analysis was conducted using the appropriate factors, and the results show that the spatial data were autocorrelated. Subsequently, SEM and SLM analyses, which were spatial regression analyses, were conducted. As a result, the model using SLM was selected as suitable for analyzing the vulnerability of ground subsidence, and the density of six types of underground utilities was found to be the highest influencing factor. In addition, a vulnerability map of ground subsidence in the target area was prepared using the model. The vulnerability map demonstrates that regions with frequent ground subsidence can be predicted to be highly vulnerable.

Keywords: OLS analysis; spatial regression; SEM; SLM; ground subsidence

Citation: Lee, S.; Kang, J.; Kim, J. Analysis of Ground Subsidence Vulnerability in Urban Areas Using Spatial Regression Analysis. *Appl. Sci.* **2023**, *13*, 8603. https://doi.org/10.3390/app13158603

Academic Editors: Shinya Inazumi and Jim S. Shiau

Received: 13 June 2023
Revised: 18 July 2023
Accepted: 25 July 2023
Published: 26 July 2023

1. Introduction

The causes of ground subsidence, which mainly occur in urban areas, are highly complex. However, the main reported cause of ground subsidence is damage to underground utilities (such as water and sewage). The occurrence of ground subsidence can lead to potentially catastrophic accidents, raising public anxiety [1]. For this reason, it is very important to analyze the causes of ground subsidence and be prepared. Several studies were conducted analyzing the causes of ground subsidence. Ground subsidence occurs when underground structures are damaged and develop discontinuities, which can be due to a range of causes, such as external impact, water pressure, and degradation. When this damage occurs, water paths are formed, primarily around the damaged areas, through infiltration and changes in groundwater levels. Surrounding soil particles are lost and transported into underground structures, leading to the formation of holes. As these holes progressively expand, the upper layers of the ground eventually collapse, demonstrating the mechanism of ground subsidence [2–8].

To support preparation for ground subsidence, studies on the prediction of ground subsidence risk were steadily conducted. To predict the risk level of ground subsidence, the main factors influencing ground subsidence must first be selected. According to the literature review and analytic hierarchy process analysis studies, the main factors influencing the occurrence of ground subsidence in urban areas are cavity, surcharge load, depth, thickness, soil type, shear strength, relative density, and soundness and history of underground utilities [9–11]. To predict ground subsidence caused by damage to underground utilities, finite element analysis was used to analyze the ground's behavior according to the amount of settlement and the affected area, and it was found that the location of damage, the relative density of the ground, and stratigraphic conditions were significant factors influencing the ground subsidence [12,13]. Furthermore, the attribute data of underground utilities were built using a geographic information system (GIS) program and a model, etc., to predict the risk level of ground subsidence based on machine learning, and regression analysis was developed [14–16].

The occurrence of ground subsidence is affected by spatial characteristics such as the location of underground utilities and geotechnical information. Thus, spatial correlation should be considered in the analysis to systematically predict the vulnerability to ground subsidence. For this reason, research on ground subsidence focused on indoor model experiments for mechanism determination, AHP analysis, and ground subsidence risk factor analysis, as well as risk prediction using machine learning techniques. However, there is no research using statistical techniques for spatial characteristics. Therefore, this study used GIS to develop a database by selecting the attribute information of underground utilities that are reported as the cause of ground subsidence and the history information of ground subsidence as influencing factors through a spatial regression model analysis for the purpose of developing a ground subsidence occurrence prediction model.

The study also removed missing data and outliers from the database that was built. Using the database, an independent t-test was performed on the spatial information of underground utilities and the occurrence of ground subsidence, and a correlation analysis was performed. In addition, an ordinary least squares (OLS) analysis was performed to determine the influence relationship between ground subsidence-related influencing factors (independent variables) and the history information of ground subsidence occurrences (dependent variable), through which spatial autocorrelation was verified. Additionally, a prediction map of ground subsidence in the target area was created based on the regression equation of underground utilities and ground subsidence history information derived through the spatial regression analysis. It is expected that ground subsidence accidents can be prevented by using ground penetrating radar (GPR) and other types of equipment to explore the ground holes in the high-risk areas presented in the ground subsidence risk prediction map created in this study.

2. Flow and Data of the Study

2.1. Flow of the Study

In this study, an urban area in South Korea was selected as the target area. To develop a prediction model of ground subsidence in the target area, the target area was divided into cells with a size of 100 m × 100 m using GIS to extract the ground subsidence history information and the attribute information of underground utilities (water pipelines, sewer pipelines, communication pipe, power cables, heating pipelines, and gas pipes). After removing the missing data and outliers from the extracted data, a correlation analysis between the attribute information of underground utilities and the history information of ground subsidence was performed. Missing values in the data are values that were left blank because the survey was not fully completed. Outliers are error values, such as negative (-) or 0 values for the buried depth and age of the pipes. Then, the study singled out influencing factors that exhibited a significant correlation with ground subsidence through the correlation analysis and performed OLS analysis using the selected factors. Furthermore, after verifying multicollinearity using the variance inflation factor (VIF) and

spatial autocorrelation through the OLS analysis results, SEM and SLM analyses were performed to verify the results. Finally, the study selected an appropriate prediction model of ground subsidence vulnerability by comparing the OLS, SEM, and SLM results, and verified the attribute information of underground utilities that were significantly correlated with ground subsidence through the selected model. A map of ground subsidence vulnerability in the target area was also created. Figure 1 shows the flow chart of this study.

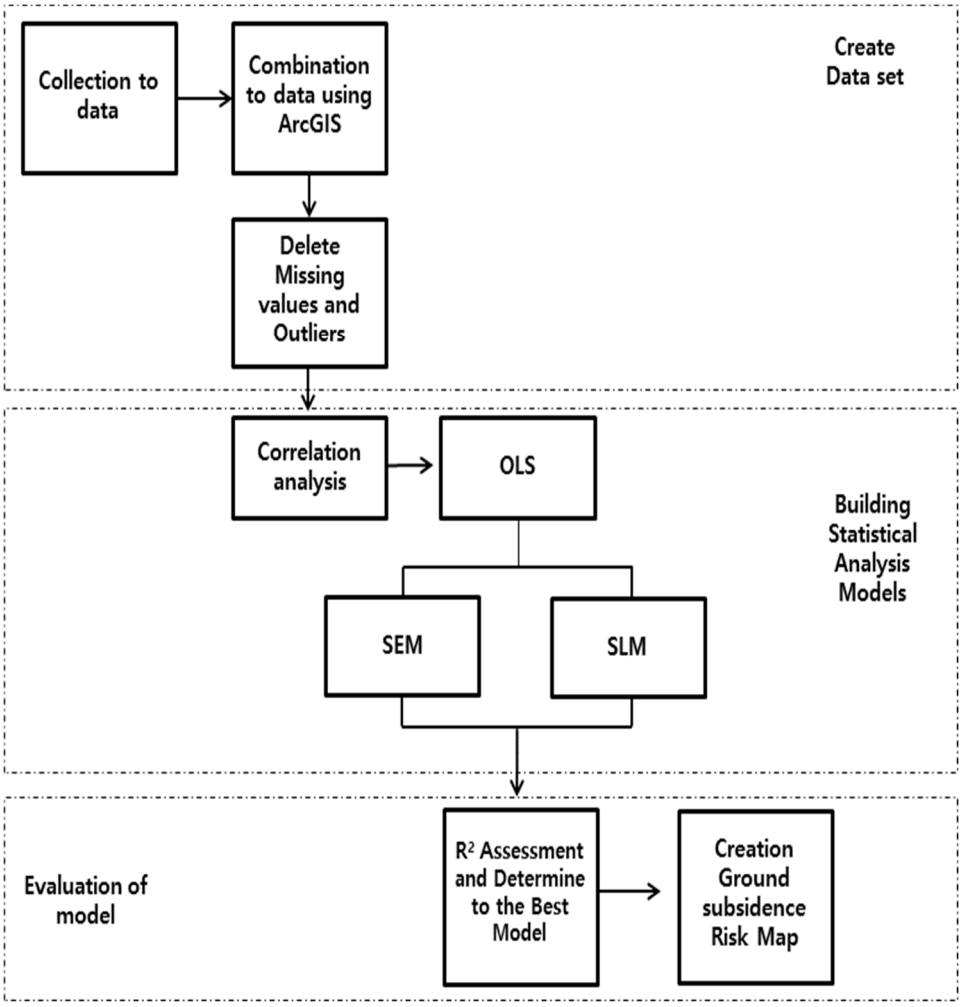


Figure 1. Flow chart.

2.2. Data

In this study, historical information on ground subsidence and attribute information on underground utilities were extracted using GIS. The target area was divided into a total of 61,707 cells with a size of 100 m × 100 m, and the number of ground subsidence incidents contained in each cell and the attribute information of six types of underground facilities were used as data. The six types of underground facilities were grouped into one type to extract data.

The history information of ground subsidence was data generated from 2008 to 2020, which consist of 9478 cases. The six types of underground utilities in the target area were water pipelines, sewer pipelines, power cables, communication pipes, heating pipelines, and gas pipes, and their attribute information was collected. The attribute information, the diameter, average depth, age, and density of the pipes were selected. The collected data of diameter and average depth were used without any change, and the age data were calculated based on the year of burial. The density was calculated through linear density

analysis using GIS. The density was calculated based on the length of pipes per unit area using GIS. It was calculated through a linear density analysis.

Table 1 presents the data characteristics (minimum, maximum, average, and standard deviation) where missing values and outliers were removed from the attribute information data of the collected underground utilities. The outliers were found mainly in the age data. If the burial year of the pipeline was difficult to identify, data marked as 103 years or older were deleted.

Table 1. Characteristics of the data.

Category		MIN	MAX	M	SD
Water pipeline	Density	0.00	0.08	0.03	0.02
	Diameter (cm)	15.00	2400.00	318.77	417.83
	Average depth (m)	0.10	25.00	1.44	0.76
	Age (years)	1.00	68.00	28.18	11.05
Sewer pipeline	Density	0.00	0.05	0.02	0.01
	Diameter (cm)	150.00	3000.00	609.45	231.36
	Average depth (m)	0.02	21.55	1.27	1.03
	Age (years)	2.00	77.00	40.92	27.30
Power cable	Density	0.00	0.05	0.01	0.01
	Diameter (cm)	50.00	300.00	164.72	27.12
	Average depth (m)	0.10	11.00	1.14	0.50
	Age (years)	1.00	42.00	11.62	7.63
Communication pipe	Density	0.00	0.09	0.02	0.01
	Diameter (cm)	1.00	1000.00	92.16	18.51
	Average depth (m)	0.20	1500.00	2.42	39.30
	Age (years)	1.00	63.00	29.51	11.48
Heating pipeline	Density	0.00	0.07	0.00	0.01
	Diameter (cm)	20.00	1100.00	322.17	230.68
	Average depth (m)	0.50	6.70	1.48	0.28
	Age (years)	2.00	38.00	20.81	10.10
Gas pipe	Density	0.00	0.06	0.02	0.01
	Diameter (cm)	20.00	750.00	162.62	93.78
	Average depth (m)	0.10	6.50	1.10	0.50
	Age (years)	1.00	49.00	25.28	8.99
All six types of underground facilities	The density of all six types of underground facilities	0.00	0.25	0.09	0.05

3. Data Correlation Analysis

To derive the attribute information of underground utilities that affected ground subsidence involving six types of underground utilities, Pearson correlation analysis was conducted between the presence, density, diameter, average depth, and age of the six types of underground utilities, the density of all six types of underground utilities, and the occurrence of ground subsidence.

Correlation analysis is an analysis that determines the degree of the linear relationship between independent and dependent variables, and the correlation coefficient has a value between -1 and 1. The correlation coefficient between the variables can be calculated through Equations (1) and (2) [17,18].

$$\mathrm{Corr}(X, Y) = \rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y} \quad (1)$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (2)$$

s_{xy}: Covariance of X and Y variables, s_x: standard deviation of variable X, s_y: and standard deviation of variable Y.

After obtaining the correlation coefficient, hypotheses of correlation in the population through Equation (3) were conducted to determine whether the correlation coefficient between two variables is statistically significant. If the test statistic is less than or equal to 0.05, there is a significant correlation between the two variables [19,20].

$$t = r\sqrt{\frac{n-2}{1-r^2}} \quad (3)$$

Table 2 presents the results of the Pearson correlation analysis. All factors, except the average depth of sewer pipelines, the average depth and age of communication pipes, the average depth of heating pipelines, and the average depth of gas pipes, were found to be significantly correlated with ground subsidence among the attribute information of underground utilities. However, most influencing factors (attribute information) did not show a high correlation with ground subsidence. This was because ground subsidence occurred due to complex causes, such as damage to underground utilities, disturbance of the surrounding ground by excavation, and poor compaction [21]. As a result, it is desirable to predict the vulnerability of ground subsidence in urban areas using influencing factors of ground subsidence as much as possible. However, as it was difficult to collect such diverse data, only the attribute information of underground utilities was used as data in this study.

Table 2. Results of Pearson correlation analysis.

Category	Ground Subsidence
Water pipeline presence	0.252 ***
Water pipeline density	0.221 ***
Water pipeline diameter	0.071 ***
Water pipeline average depth	0.023 ***
Water pipeline age	0.088 ***
Sewer pipeline presence	0.248 ***
Sewer pipeline density	0.237 ***
Sewer pipeline diameter	0.056 ***
Sewer pipeline average depth	−0.058 ***
Sewer pipeline age	−0.006
Power cable presence	0.254 ***
Power cable density	0.201 ***
Power cable diameter	0.073 ***
Power cable average depth	−0.044 ***
Power cable age	0.084 ***
Communication pipe presence	0.272 ***
Communication pipe density	0.238 ***
Communication pipe diameter	0.079 ***
Communication pipe average depth	0.006
Communication pipe age	0.002
Heating pipeline presence	0.039 ***
Heating pipeline density	−0.002
Heating pipeline diameter	0.075 ***
Heating pipeline average depth	−0.006
Heating pipeline age	−0.038 **

Table 2. *Cont.*

Category	Ground Subsidence
Gas pipe presence	0.252 ***
Gas pipe density	0.232 ***
Gas pipe diameter	0.133 ***
Gas pipe average depth	0.018
Gas pipe age	0.096 ***
All six types density	0.240 ***

** $p < 0.01$, *** $p < 0.001$.

4. OLS Analysis

To develop a model for predicting ground subsidence vulnerability in urban areas, a multiple linear regression analysis was conducted to determine the influence of specific ground subsidence influencing factors (underground utility attribute information) on whether ground subsidence occurs. Multiple linear regression analysis represents the relationship between multiple explanatory variables and the dependent variable (whether or not ground subsidence occurred) as a regression equation, as presented in Equation (4), and the coefficient of the regression equation is estimated through the OLS that minimizes the sum of the squares of the residuals [22,23].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (4)$$

y: Dependent variable, x_i: independent variable (explanatory variable), β_i: estimated regression coefficient, and ε: error term.

To build a multiple regression model, it is necessary to check for multicollinearity, which is the correlation between the explanatory variables. This is because the presence of multicollinearity will distort the estimated regression coefficients of the multiple regression model. Thus, multicollinearity is measured by the VIF. The VIF is generally based on a value of 10; if it is less than 10, it can be considered that multicollinearity is not present. Equation (5) presents the VIF calculation for the j-th regression coefficient estimate β_j [24,25].

$$VIF_j = \frac{1}{1 - R_j^2} \quad (5)$$

The coefficient of determination calculated from a regression model with $R_j^2 : X_j$ as the dependent variable.

Table 3 presents VIFs of the attribute information of underground utilities used in this study. The diameter, average depth, age of six types of underground utilities, and density of all six types of underground utilities exhibited a VIF of 10 or less. Accordingly, all attribute information was used to perform the OLS analysis. The Jarque–Bera and Breusch–Pagan statistics were used to test for non-normality and heteroscedasticity in the OLS model.

Table 4 presents the OLS analysis results. When performing an OLS analysis, it is necessary to consider assumptions about the presence of spatial effects, such as spatial autocorrelation. To do this, the distribution of residuals is identified using Moran's I value. Moran's I is a measure to verify the spillover effects that occur in spatial proximity. It is an indicator to check the correlation between events in a specific region and events in neighboring regions based on a spatially weighted matrix, such as Equation (6) [26].

$$I = \frac{N \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(Y_i - \bar{Y}\right) \left(Y_j - \bar{Y}\right)}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\right) \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} \quad (6)$$

N: No. of regions, Y_i: attribute of region i, Y_j: attribute of region, $\bar{Y}$: mean value, and w_{ij}: weigh value.

Values of Moran's I usually range from −1 to +1. A value close to 1 indicates a positive spatial autocorrelation, while a value close to −1 indicates negative spatial autocorrelation [27]. The value of Moran's I in the model proposed in this study is 0.127. This means that there is a positive spatial autocorrelation that increases the occurrence of ground subsidence as the number of six types of underground facilities increases. In addition, the z-score was 44.359 (p-value < 0.001). The z-score is a criterion to determine the statistical significance of Moran's I. If the z-score is statistically significant (p-value < 0.05), it confirms the existence of spatial autocorrelation in the model. Furthermore, both the Jarque–Bera and Breusch–Pagan statistics are statistically significant, indicating that there is non-normality and heteroscedasticity in the error term. Thus, since the OLS analysis model exhibits non-normality and heteroscedasticity in the error term, and the spatial autocorrelation of the residuals is confirmed, it is appropriate to conduct the analysis with a spatial regression model using a spatially weighted matrix.

Table 3. VIF of the dependent variables.

Underground Facility Information	VIF
Water pipeline diameter (cm)	1.555709
Water pipeline average depth (m)	2.338912
Water pipeline age (years)	2.712778
Sewer pipeline diameter (cm)	1.541930
Sewer pipeline average depth (m)	1.469795
Sewer pipeline age (years)	1.518273
Power cable diameter (cm)	5.369876
Power cable average depth (m)	4.826353
Power cable age (years)	1.356797
Communication pipe diameter (cm)	4.055690
Communication pipe average depth (m)	1.107161
Communication pipe age (years)	3.990815
Heating pipeline diameter (cm)	2.802712
Heating pipeline average depth (m)	3.911007
Heating pipeline age (years)	2.918603
Gas pipe diameter (cm)	2.085979
Gas pipe average depth (m)	1.150920
Gas pipe age (years)	2.768914
Density of all six types	1.758216

Table 4. OLS analysis results.

Underground Facility Information		Linear Regression Model (OLS)
Constant		−0.0314207 ***
Underground facility Information	Water pipeline diameter (cm)	−0.0314207 ***
	Water pipeline average depth (m)	0.000063 ***
	Water pipeline age (years)	−0.010811 ***
	Sewer pipeline diameter (cm)	0.001171 ***
	Sewer pipeline average depth (m)	−0.000011 **
	Sewer pipeline age (years)	−0.0143429 ***
	Power cable diameter (cm)	0.000140 **
	Power cable average depth (m)	0.000807 ***
	Power cable age (years)	−0.033201 ***
	Communication pipe diameter (cm)	−0.000404
	Communication pipe average depth (m)	0.000415 ***
	Communication pipe age (years)	−0.000119

Table 4. *Cont.*

Underground Facility Information		Linear Regression Model (OLS)
	Heating pipeline diameter (cm)	0.000183
	Heating pipeline average depth (m)	0.000090 ***
	Heating pipeline age (years)	−0.021801 ***
	Heating pipeline diameter (cm)	−0.001326 ***
	Gas pipe average depth (m)	0.000210 ***
	Gas pipe age (years)	0.033765 ***
	Density of all six types	0.000888 ***
Spatial autocorrelation of standardized residuals	Global Moran's I	0.326620
	z-score	0.127047
Explanatory power of the model	R^2	44.358796 ***
Fit of the model	log−likelihood	0.130332
	AIC	−6183.22
	SC	12406.4
Non-normality	Jarque−Bera	12587.0
Heteroscedasticity	Breusch−Pagan	104,760.6103 ***
Multicollinearity	Multicollinearity conditional number	31,757.1448 ***
Spatial autocorrelation in spatial regression models	LM−lag	11.309430
	Robust LM−lag	2869.9710 ***
	LM−error	84.5001 **
	Robust LM−error	2807.7994 ***

** $p < 0.01$, *** $p < 0.001$.

5. Spatial Regression Analysis

The aim of this study was to develop a model to predict the vulnerability of ground subsidence in the target area using the attribute information of six types of underground utilities and the history information of ground subsidence. The OLS analysis results demonstrate that there was spatial autocorrelation. Thus, a spatial regression model was analyzed. To estimate a spatial regression model, spatial data were built first. Then, a spatial distance was defined and a spatial weight matrix that set the relationship between regions was constructed using the GeoDa 1.20 software [28]. For spatial distance, this study chose the Queen method, which determines that in gridded data, neighbors are adjacent if they share a side or corner of a given grid cell. The Queen method is used in geospatial data consisting of administrative districts, and is suitable for considering the adjacency of regions that share boundaries with each other [29].

The spatial regression analysis in this study employed SEM and SLM. SEM is a model that considers the covariance of errors in the regression model when spatial autocorrelation exists in the errors. SLM is a model that analyzes spatial dependence (spatial lagged variables) by including them in the regression model as one explanatory variable [30]. In addition, SEM and SLM were estimated through maximum likelihood. The SEM and SLM are produced using Equations (7) and (8).

$$y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \cdots + u\ u = \lambda W_u + e \tag{7}$$

y: Dependent variable, X: independent variable, U: error term with spatial autocorrelation, W: spatial weight matrix, Δ, λ : estimation coefficient, and e:error term without spatial autocorrelation.

$$y = \rho W_y + \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \ldots + \varepsilon \tag{8}$$

y: Dependent variable, X: independent variable, W: spatial weight, ρ, δ: estimation coefficient, and ε : error term.

R^2 was used as a measure of explanatory power to compare spatial regression models. To determine the fit of the model, log-likelihood, Akaike information criterion (AIC), and Schwarz criterion (SC) were used. To determine the significance of the spatial autoregressive coefficient (ρ, λ), the likelihood ratio test was used [31,32]. Table 5 presents the results of SEM and SLM.

Table 5. Spatial regression analysis results.

Underground Facility Information		Spatial Lag Model (SLM)	Spatial Error Model (SEM)
Constant		0.0235515 ***	−0.033493 ***
Underground facility Information	Water pipeline diameter (cm)	0.000063 ***	0.000071 ***
	Water pipeline average depth (m)	−0.010554 ***	−0.009999 ***
	Water pipeline age (years)	0.000899 ***	0.000957 ***
	Sewer pipeline diameter (cm)	−0.000013 ***	−0.000012 **
	Sewer pipeline average depth (m)	−0.012026 ***	−0.009926 ***
	Sewer pipeline age (years)	0.000105 *	0.000134 **
	Power cable diameter (cm)	0.000644 ***	0.000657 ***
	Power cable average depth (m)	−0.021835 ***	−0.015158 ***
	Power cable age (years)	−0.000489	−0.000689 *
	Communication pipe diameter (cm)	0.000383 ***	0.000426 ***
	Communication pipe average depth (m)	−0.000115	−0.000120
	Communication pipe age (years)	−0.000039	−0.000182
	Heating pipeline diameter (cm)	0.000093 ***	0.000100 ***
	Heating pipeline average depth (m)	−0.019548 ***	−0.017253 **
	Heating pipeline age (years)	−0.001199 ***	−0.001354 ***
	Gas pipe diameter (cm)	0.000209 ***	0.000243 ***
	Gas pipe average depth (m)	0.021559 ***	0.024353 ***
	Gas pipe age (years)	0.000644 ***	0.000735 ***
	Density of all six types	0.114037 ***	0.397172 ***
Spatial dependence	Likelihood ratio		
Spatial effect	Rho (ρ)	0.326620	
	Lambda (λ)	0.127047	
The explanatory power of the model	R^2	44.358796 ***	
Fit of the model	log-likelihood	0.130332	
	AIC	−6183.22	
	SC	12,406.4	

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Statistics were checked to determine which model out of SLM or SEM was more appropriate for predicting ground subsidence vulnerability. The model's explanatory power, R^2, showed that the SLM model was 0.173390 and the SEM model was 0.174061, indicating that the explanatory power of SEM was relatively higher. To determine the fit of the model, log-likelihood, AIC, and SC were reviewed, and the log-likelihood values of SLM and SEM were −5046.05 and −5053.52, respectively, indicating that the value of SLM was relatively higher. Both values of AIC and SC of SLM were smaller. Thus, the overall statistics exhibited that R^2 of SEM was higher, although the difference was minimal, while log-likelihood, AIC, and SC values of SLM were all more fit. Thus, the SLM model was finally selected as the most-fit model.

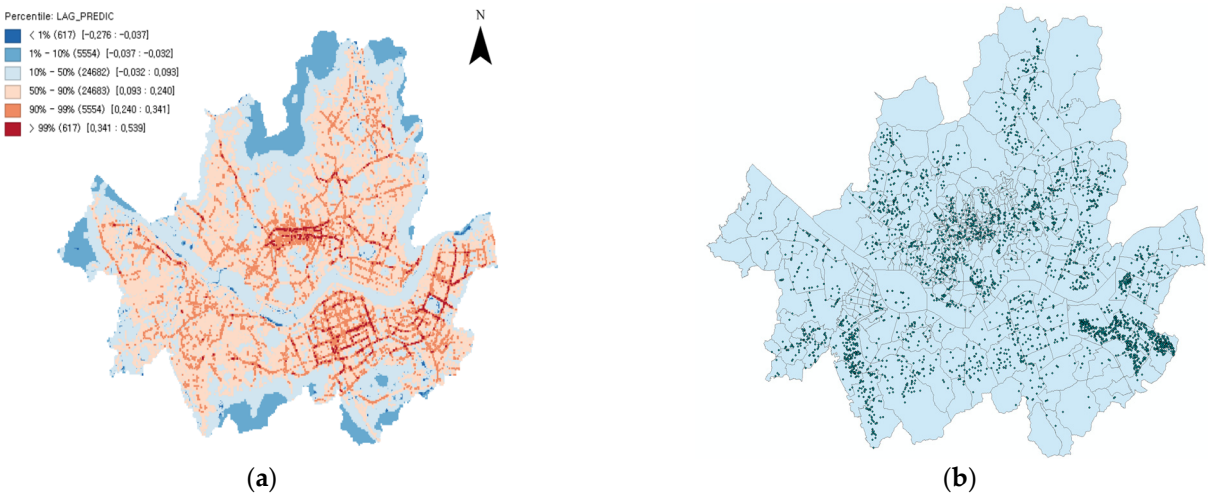
The finally selected SLM model was analyzed, and the results show that the diameter and age of water pipelines; the age of sewer pipelines; the diameter of power cables; the diameter of communication pipes; the diameter of heating pipelines; the diameter, average depth, and age of gas pipes; and the density of all six types of pipes had a positive (+) effect on the occurrence of ground subsidence. On the other hand, the average depth of water pipelines, the diameter and the average depth of sewer pipelines, the average depth of

power cables, and the average depth and age of heating pipelines were found to have a negative (-) effect on the occurrence of ground subsidence. Additionally, the age of power cables and the average depth and age of the communication pipes were not found to be significant in terms of the occurrence of ground subsidence.

Because the likelihood ratio statistic of the SLM model is significant ($p < 0.001$), it can be concluded that the fit of the model is improved by applying spatial effects. In addition, the value of ρ, the spatial effect, is 0.31534 ($p < 0.001$). This means that the occurrence of ground subsidence in a specific area that has spatial autocorrelation is affected by about 31.5% of the occurrence of ground subsidence in the surrounding ground.

6. A Map of Ground Subsidence Vulnerability

A map of ground subsidence vulnerability in the target area was created using the SLM model where spatial dependence was applied. To create a map of vulnerability, a program called GeoDa 1.20, which can perform spatial data and modeling, was used. A Queen-based spatial weight matrix was used and the target area was divided into a 100 m × 100 m grid. Figure 2a shows the map of ground subsidence vulnerability predicted by the SLM model, in which the vulnerable spots are indicated with red while the relatively safer spots are indicated with blue. Figure 2b shows a map of points where ground subsidence actually occurred in the target area. The figures indicate that the vulnerability of ground subsidence was highly well predicted in the central and eastern parts, where ground subsidence occurred frequently. On the other hand, in the western part where there was no significant ground subsidence, the SLM model did not predict vulnerability well. This is a problem caused by the model's inaccuracy. In the case of the western part, it is estimated that the number of occurrences of ground subsidence caused by factors other than damage to underground utilities was included. Therefore, it is expected that a high-accuracy ground subsidence vulnerability map will be achieved by securing information on various underground spaces, such as ground information, the groundwater level, and subway, and reflecting it in the model in the future.



(**a**) (**b**)

Figure 2. Comparison of ground subsidence vulnerability maps. (**a**) A map of ground subsidence vulnerability; and (**b**) location of actual ground subsidence occurrences.

7. Conclusions

This study employed the attribute information of six types of underground facilities and the history information of ground subsidence in the target area to select factors influencing ground subsidence through correlation analysis and present a model of ground subsidence vulnerability through OLS and spatial regression analysis. In addition, a map of ground subsidence vulnerability in the target area was created through the selected optimal model.

The global Moran's I was checked through OLS analysis and the results show that the spatial data exhibited autocorrelation. Thus, the model of ground subsidence vulnerability in the target area employed spatial regression models (SEM and SLM). The results of analysis using the spatial regression models show that the model estimated by SLM was more fit. The results of an analysis of correlation coefficients regarding the attribute information of six types of underground utilities in the SLM model show that the correlation of the density of all six types of underground utilities was the highest. This is a similar finding to those of previous studies, which showed that the density of underground utilities is strongly correlated with ground subsidence [10,33]. Furthermore, the ρ value of the SLM model verified that the ground subsidence is influenced by underground utilities and spatial ripple effects.

Using the attribute information of underground utilities in the same target area, the map of ground subsidence risk prediction based on machine learning and the map of ground subsidence vulnerability in this study were compared [15]. The comparison results show that the central and eastern parts of the target area where ground subsidence frequently occurred had the same high vulnerability to ground subsidence. Thus, the central and eastern parts of the area and regions with a high density of underground utilities should be checked as a priority to prepare for a ground subsidence accident in the target area.

In this study, a model to predict ground subsidence vulnerability was developed using the attribute information of only six types of underground utilities. However, ground subsidence can have more complex causes. Thus, a model that more reliably predicts ground subsidence vulnerability can be developed in the future by securing high-quality ground subsidence, influencing factor data (subway, underpass, high-rise building information, etc.).

Author Contributions: Conceptualization, J.K. (Jaemo Kang) and J.K. (Jinyoung Kim); Methodology, S.L.; Software, S.L.; Writing—original draft, S.L.; Writing—review & editing, S.L., J.K. (Jaemo Kang) and J.K. (Jinyoung Kim). All authors have read and agreed to the published version of the manuscript.

Funding: Research for this paper was carried out under the KICT Research Program (project no. 20230116-001, Underground Utilities Diagnosis and Assessment Technology (4/4).) funded by the Ministry of Science and ICT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, S.Y.; Kang, J.M.; Kim, J.Y. Development of Machine Learning Model to predict the ground subsidence risk grade according to the Characteristics of underground facility. *J. Korean Geo-Environ. Soc.* **2022**, *23*, 5–10.
2. Kuwano, R.; Horii, T.; Kohashi, H.; Yamauchi, K. Defects of sewer pipes causing cave-in's in the road. In Proceedings of the 5th International Symposium on New Technologies for Urban Safety of Mega Cities in Asia, Phuket, Thailand, 16–17 November 2006; pp. 347–353.
3. Kuwano, R.; Kohata, Y.; Sato, M. Case study of ground cave-in due to subsurface erosion in old landfill. In Proceedings of the 6th International Conference on Scour and Erosion, France, Paris, 27–31 August 2012; pp. 56–62.
4. Mukunoki, T.; Otani, J.; Nonaka, S.; Horii, T.; Kuwano, R. Evaluation of cavity generation in soils subjected to sewerage defects using X-ray CT. In Proceedings of the 2nd International Workshop on X-ray CT for Geomaterials, GeoX 2006, Grenoble, France, 4–7 October 2006; pp. 365–371.
5. Mukunoki, T.; Kuwano, N.; Otani, J.; Kuwano, R. Visualization of three-dimensional failure in sand due to water inflow and soil drainage from defected underground pipe using X-ray CT. *Soils Found.* **2009**, *49*, 959–968. [CrossRef]
6. Renuka, S.; Kuwano, R. Formation and evaluation of loosened ground above a cavity by laboratory model test with uniform sand. In Proceedings of the 13th International Summer Symposium, Uji, Japan, 26 August 2011; pp. 211–214.
7. Sato, M.; Kuwano, R. Effects of Buried Structures on the Formation of Underground Cavity. In Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris, France, 2–6 September 2013; pp. 1769–1772.

8. Sato, M.; Kuwano, R. Model tests for the evaluation of formation and expansion of a cavity in the ground. In Proceedings of the 7th International Conference on Physical Modelling in Geotechnics, Zurich, Switzerland, 28 June–1 July 2010; pp. 581–586.
9. Park, J.Y.; Jang, Y.J.; Kim, H.J.; Im, M.H. Classification of Ground Subsidence Factors for Prediction of Ground Subsidence Risk. *J. Eng. Geol.* **2017**, *27*, 153–164.
10. Kim, J.Y.; Kang, J.M.; Choi, C.H.; Park, D.H. Correlation Analysis of Sewer Integrity and Ground Subsidence. *J. Korean Geo-Environ. Soc.* **2017**, *18*, 31–37.
11. Jin, Y.S. The Analysis on Correlation of Precipitation and Risk Factors to the Soil Subsidence. Ph.D. Thesis, Chonnam National University, Gwangju, Republic of Korea, 2018; pp. 104–105.
12. Oh, D.W.; Kong, S.M.; Lee, D.Y.; Yoo, Y.S.; Lee, Y.J. Effects of Reinforced Pseudo-Plastic Backfill on the Behavior of Ground around Cavity Developed due to Sewer Leakage. *J. Korean Geo-Environ. Soc.* **2015**, *16*, 13–22. [CrossRef]
13. Takeuchi, D.; Fukatani, W.; Miyamoto, T. Using decision tree analysis to extract factors affecting road subsidence. *J. Jpn. Sew. Work. Assoc.* **2007**, *54*, 124–133.
14. Lee, S.Y.; Kim, J.Y.; Kang, J.M.; Baek, W.J. Comparison of Machine Learning Models to Predict the Occurrence of Ground Subsidence According to the Characteristics of Sewer. *J. Korean Geo-Environ. Soc.* **2022**, *23*, 5–10.
15. Lee, S.Y.; Kang, J.M.; Kim, J.Y. Prediction Modeling of Ground Subsidence Risk Based on Machine Learning Using the Attribute Information of Underground Utilities in Urban Areas in Korea. *Appl. Sci.* **2023**, *13*, 5566. [CrossRef]
16. Seo, J.W.; Ryu, D.W.; Yum, B.W. Logistic Regression and GIS-based Urban Ground Sink Susceptibility Assessment Considering Soil Particle Loss. *Korean Soc. Rock Mech. Rock Eng.* **2020**, *30*, 149–163.
17. Pearson, K. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242.
18. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1964**, *26*, 211–252. [CrossRef]
19. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [CrossRef]
20. Fisher, R.A. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1921**, *1*, 3–32.
21. Seoul Seokchon-dong Cavity Cause Investigation Committee. *Cause Analysis of Cavity at Seokchon Underground Roadway and Road Cavity*; Seoul Seokchon-dong Cavity Cause Investigation Committee: Seoul, Republic of Korea, 2014.
22. Hayes, A.F. Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behav. Res. Methods* **2007**, *39*, 709–722. [CrossRef]
23. Hayes, A.F.; Mattels, J. Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behav. Res. Methods* **2009**, *41*, 924–936. [CrossRef]
24. Chan, J.Y.-L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.-W.; Chen, Y.-L. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics* **2022**, *10*, 1283. [CrossRef]
25. Kimon, N.; Alex, K.; Andreas, A. Interdependency Pattern Recognition in Econometrics: A Penalized Regularization Antidote. *Econometrics* **2021**, *9*, 44. [CrossRef]
26. Wu, D.; Liu, J. Spatial and Temporal Evaluation of Ecological Footprint Intensity of Jiangsu Province at the County-Level Scale. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7833. [CrossRef]
27. Gerad, H.; Rob, B.; KHoogh Danielle, V.; John, G.; Paul, F.; David, B. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578.
28. Arthur, G.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, *24*, 189–206.
29. Anselin, L. The Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]
30. Yang, L.; Chau, K.W.; Chu, X. Accessibility-based premiums and proximity-induced discounts stemming from bus rapid transit in China: Empirical evidence and policy implications. *Sustain. Cities Soc.* **2019**, *48*, 101561. [CrossRef]
31. Park, S.; Ko, D. ASpatial Regression Modeling Approach for Assessing the Spatial Variation of Air Pollutants. *Atmosphere* **2021**, *12*, 785. [CrossRef]
32. Wei, G.; Zhang, Z.; Ouyang, X.; Shen, Y.; Jiang, S.; Liu, B.; He, B.J. Delineating the spatial-temporal variation of air pollution with urbanization in the Belt and Road Initiative area. *Environ. Impact Assess. Rev.* **2021**, *91*, 106646. [CrossRef]
33. Zhao, C.; Wang, B. How does new-type urbanization affect air pollution? Empirical evidence based on spatial spillover effect and spatial Durbin model. *Environ. Int.* **2022**, *165*, 107304. [CrossRef] [PubMed]