Jour. of the KGS, Vol. 20, No. 7. September 2004, pp. 33~44

Damage Estimation Based on Spatial Variability of Seismic Parameters Using GIS Kriging

GIS Kriging을 이용하여 공간적으로 분포하는 지진매개변수의 분석과 손상 평가

Jeon, Sang-Soo[1] 전 상 수

요 지

지진시 측정되는 strong motion 데이터와 시설물의 손상의 상관관계를 구하기 위하여 본 연구를 수행하였다. 지진시 측정된 시간 - 지진강도 관계와 측정위치들이 포함된 대규모 지형정보시스템(GIS) 데이터에 대한 분석에 Kriging 공간 통계분석법을 사용함으로써 첨두지반속도(PGV)와 파이프라인 및 건물의 손상정도의 관계를 구할 수 있었으며, 특히, Kriging법은 측정치에 포함된 오차 등 불확실한 요소들을 고려한 90%의 신뢰도에 해당하는 지진강도-손상 관계를 구할 수 있어서 역거리 제곱법 등 다른 공간데이터 분석법에 비하여 우수한 방법이었다.

Abstract

This paper is focused on the spatial variability of measured strong motion data during earthquake and its relationship with the performance of water distribution pipelines and residential buildings. Analyses of strong motion and the correlations of peak ground velocity (PGV) and pipeline and building damage were conducted with a very large geographical information system (GIS) database including the relationship of time and earthquake intensity and the measured location, and Kriging spatial statistics. Kriging was used to develop regressions of pipeline repair rate (RR) and residential building damage ratio (DR) associated with 90% confidence peak ground velocity (PGV). Such regressions using Kriging provide an explicit means of characterizing the uncertainty embodied in the strong motion data compared with other spacial statistics such as inverse distance method.

Keywords : Building damages, Earthquake, Geotechnical engineering, GIS, Seismic parameters

1. Introduction

This paper provides an evaluation of the spatial variability of peak ground velocity (PGV) to quantify the spatial uncertainties of seismic and site conditions. It presents explanations for interpolation algorithms, such as Triangular Irregular Network (TIN) [e.g. ESRI, 1994] and kriging (e.g., Ripley, 1987), which were used to create PGV contours. ArcInfo TIN is one of the surface models in ArcInfo software (ESRI, 1994). The ArcInfo TIN model uses an exiting database including point, line, and polygon coverages and creates a tin data structure that consists of nodes and a series of edges joining these. These nodes form triangles. The nodes are the data points. The paper provides a description of regionalized variable theory and variogram models. The results of ordinary kriging and TIN procedures for PGV contours are presented and discussed (Jeon, 2002). Regression analyses are performed to develop predictive relationships between both PGV and pipeline repair

1 Member, Chief Researcher, Korea Highway & Transportation Technology, Korea Highway Corporation (ssj3@freeway.co.kr)

rates (RRs), which is pipeline repairs/km, and damage ratios (DRs), which is DR [% of existing structures with damage equal to or exceeding a particular damage factor, DF (% of building replacement cost)], for residential timber structures, respectively, for the 1994 Northridge earthquake that was collected, digitized and organized by researchers at Cornell.

Ordinary kriging is also adapted to delineate 90% confidence limits for PGV in two-dimensional space. Regressions of PGV at 90% confidence with pipeline repair rates and building damage ratios provide a means of compensating for the spatial variability of the data and for making predictions with higher confidence than those estimated on the basis of mean values.

2. Geostatistical Kriging Method

Kriging is based on statistical models representing relationships among the existing data points. Not only are these techniques capable of predicting values at unsampled locations, but they also provide a measure of the uncertainty in these predictions from which confidence values can be deduced.

Geostatistical methods for interpolation start with the recognition that the spatial variation of any continuous attribute is often too irregular to be modeled by a simple and smooth mathematical function. The variation is therefore described by a stochastic surface. Generation of the stochastic surface is accomplished through regionalized variable theory.

Burrough and McDonnell (1998) provide an explanation of regionalized variable theory. The theory starts with the assumption that the spatial variation of any variable is expressed as the sum of three major components as shown in Fig. 1. They are a structural component, which accounts for the trend of the data, a random but spatially correlated component, known as the variation of the regionalized variable, and a spatially uncorrelated random noise or residual error term, a position in one, two, or three dimensions. The value of a random variable Z at s is given in which m(s) is a deterministic function describing the structural component of Z at s, ε'(s) is the term denoting the stochastic, locally varying but spatially dependent residuals from m(s) - the regionalized variables -, and ε" (s) is a residual, spatially independent noise term.

Burrough and McDonnell (1998) indicate that the simplest case occurs when no trend or drift is either present or deduced from the data. For this condition, m(s) equals the mean value, μ, in the sampling area. The average or expected difference between Z(s) and Z(s+h) at any two points s and s+h, respectively, separated by a distance vector h, is zero:

$$E[Z(s)-Z(s+h)]=0 \quad (1)$$

Furthermore, it is assumed that the variance of differences depends only on the distance between sites, h, so that

$$E[\,Z(s)-Z(s+h)^2]=E[\,\varepsilon'(s)-\varepsilon'(s+h)^2]=2\gamma(h) \quad (2)$$

where γ(h) is known as the semivariance. The two conditions, the second-order stationarity for covariance and the intrinsic stationarity for variograms, establish the requirements for regionalized variable theory. The second-order stationarity for covariance and intrinsic stationarity for variograms assume that the covariance and the variance of the difference are the same between any two points at the same separating distance no matter which two points are chosen.

If the conditions specified by the above assumptions are fulfilled, the semivariance, $\hat{\gamma}(h)$, is estimated from

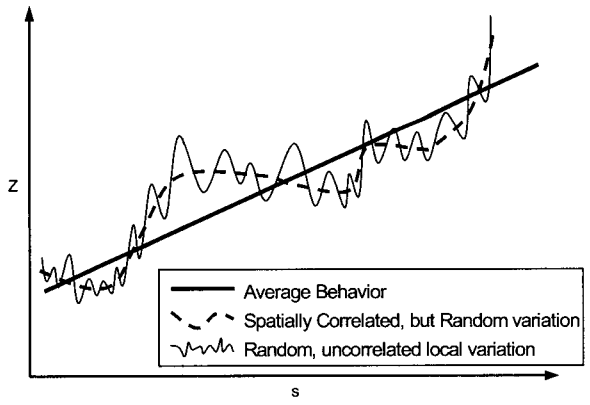

Fig. 1. Spatial variations of regionalized variable theory (after Burrough and McDonnell, 1998)

sample data as follows

$$\hat{\gamma}(\mathrm{h}) = \frac{1}{2n} \sum_{i=1}^{n} \{Z(s_i) - Z(s_i + h)\}^2 \qquad (3)$$

where n is the number of pairs of sample points of observations of the values of attribute Z separated by distance h. A plot of $\hat{\gamma}(h)$ against h is known as the experimental variogram. The experimental variogram helps to provide a quantitative description of regionalized variation. Before functions, such as interpolation and characterization of spatial patterns, can be performed a theoretical model needs to be fitted to the experimental variogram.

3. Ordinary Kriging

Ordinary kriging for PGV was performed, and the resulting PGV values were correlated with damage to cast iron pipelines, expressed as RR (repair rate), and damage to residential timber structures, expressed as DR (damage ratio). To utilize ordinary kriging it is necessary to 1) evaluate the type of distribution that best represents the seismic parameter, 2) justify the expression for the random variable that is assumed in the kriging application, and 3) select the most appropriate variogram model.

3.1 Distribution Model

It is frequently assumed that seismic parameters are natural log (ln) normally distributed. For example, Boore, et al. (1993) use the geometric mean to estimate randomly oriented components of horizontal ground motion. To determine the appropriate model for the distribution of PGV, values from 164 records of Northridge earthquake strong motion were used to develop a cumulative frequency distribution for goodness-of-fit testing using the Kolmorgorov-Smirnov (K-S) test. Cumulative frequency distributions were developed for both the arithmetic and ln values of PGV.

Figs. 2 and 3 show plots of the cumulative frequency and theoretical distribution functions in which the maximum differences, $D_{n_{max}}$, between the observed data and K-S theoretical model are indicated. There is a significantly better agreement between the theoretical distribution [F(x)] and observed cumulative frequency [$F_n(x)$] plot for ln (PGV). Comparison of $D_{n_{max}}$ with K-S values at various significance levels, D_n^a, (where α = significance level and n = 164 for PGV) shows that a normal distribution of arithmetic PGV is not significant at even the 1% level. In contrast, the normal distribution of ln (PGV) is significant to approximately the 4% level.

Similar goodness-of-fit tests were performed with thechi-square test. This type of testing shows that a normal distribution of arithmetic values fail to pass at a 1% significance, whereas a ln (PGV) distribution is significant to approximately 2.5%.

The K-S plots in Figs. 2 and 3 and the substantially higher significance levels for ln (PGV) indicate that a

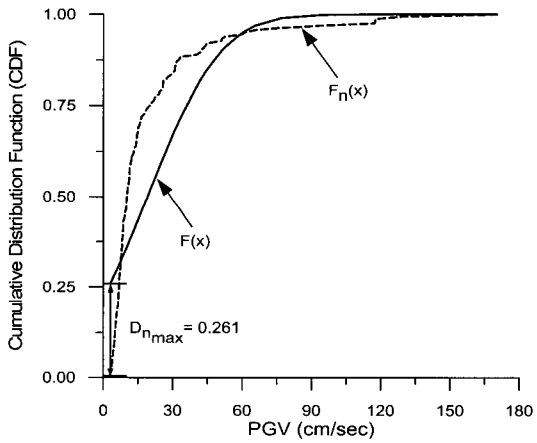

Fig. 2. Cumulative distribution of PGV

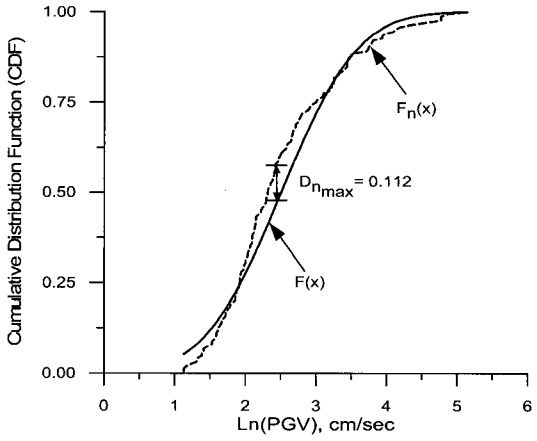

Fig. 3. Cumulative distribution of ln (PGV)

log-normal model for PGV is better than an arithmetic one. Accordingly, kriging analyses were performed for ln (PGV).

3.2 Characteristics of the Random Variables

It is assumed that the distribution of strong ground motion is a stochastic surface, governed entirely by a random process. This neglects any deterministic trend in the distribution. If an actual trend exists, such an assumption will result in a higher estimate of uncertainty relative to the case for a specified trend. Hence, the application of ordinary kriging in this instance will be biased towards a conservative assessment of uncertainty.

The variance of the spatially dependent residuals, s^2_{diff} may be expressed as

$$s^2_{diff} = \frac{1}{n}\sum_{i=1}^{n}\{Z(s_i) - Z(s_i+h) - \mu_{diff}\}^2 \quad (4)$$

in which μ_{diff} = population average difference between any two locations. The semivariance is likewise given as

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{Z(s_i) - Z(s_i+h) - \mu_{diff}\}^2 = \frac{1}{2n}\sum_{i=1}^{n}\{Z(s_i) - Z(s_i+h)\}^2 - \mu^2_{diff} \quad (5)$$

Earthquake strong motions tend to be radially distributed around the seismic source. It can be shown that a radially symmetrical distribution of seismic parameters will result in μ_{diff} = 0. Eq. 5 reduces to the general expression for semivariance in ordinary kriging when μ_{diff} = 0. This results in a well-defined sill region of the variogram. In contrast, $\hat{\gamma}(h)$ tends to increase parabolically beyond the range when $\mu diff \neq 0$.

As will be shown in the next section, there is a well-defined sill region for the variogram developed for PGV with ordinary kriging and no evidence for increasing $\hat{\gamma}(h)$ beyond the range. Using the 164 measurements of PGV, the mean difference in measurements, $\bar{d}_{diff}$, was calculated as 1.86 cm/sec. This is a very low value that does not suggest an appreciable μ_{diff}. The characteristics of the data, therefore, support the development of a well-defined variogram.

3.3 Cressie Goodness-of-Fit Statistic of Variogram Model

The Cressie goodness of fit statistic, C (Cressie, 1993), is calculated for variogram models as follows

$$C = n\left[\frac{\hat{\gamma}(h) - \gamma(h)}{\gamma(h)}\right]^2 \quad (6)$$

where n is the number of pairs of PGV, $\hat{\gamma}(h)$ is the empirical semivariance, and $\gamma(h)$ is the theoretical semivariance. The magnitude of Cressie statistic depends on the total number of pairs of samples. Because not all samples are necessarily paired the same number of times in different directions, Clark and Harper (2000) propose a modification to remove the scaling by total number of pairs as follows

$$C = \frac{1}{\sum n}\sum n\left[\frac{\hat{\gamma}(h) - \gamma(h)}{\gamma(h)}\right]^2 \quad (7)$$

Table 1 shows the Cressie goodness-of-fit statistics, as modified by Clark and Harper (2000), of the four different variogram models. As shown by the table, there is little difference in the Cressie statistic for the different models, and very little difference between the statistic for the spherical and Gaussian models. To achieve a consistent approach to ordinary kriging, the spherical model was likewise adopted for PGV. As will be explained in a forthcoming section, contours of PGV developed by ordinary kriging with the spherical model compare very favorably with similar contours developed from the same data with TIN. This favorable comparison lends additional weight to the selection of the spherical model.

Table 1. Cressie goodness-of-fit statistic of variogrom model for PGV (cm/sec)

Variogram Model	Cressie Statistic
Gaussian	1.93
Spherical	2.18
Linear	2.55
Exponential	2.49

4. Results of Kriging

4.1 Variogram for Ln (PGV)

Fig. 4 shows the variogram for the ln (PGV) data set. The plotted line presents the actual variance. It can be seen that γ_h reaches a plateau, or sill, at a distance of approximately 237,000 ft (72 km) and remains relatively constant to about 300,000 ft (91 km), after which if decreases in an erratic fashion.

Fig. 5 shows the predicted variance using a spherical model with a range of 237,000 ft (72 km). The spherical model follows the trend of the actual variogram. The sill value is about 0.88 $(cm/sec)^2$, with a nugget of approximately zero.

4.2 Location of PGV Records

Fig. 6 shows the locations of the 164 strong motion records used in this paper and for developing PGV contours with TIN in paper. Fig. 7 shows the PGV contours developed with TIN. Superimposed on both figures is a circle with radius equal to a range of 237,000 ft (72 km) determined from the variogram. The circle is centered on the zone of highest PGV. The range helps to define a radius of influence, within which PGV values are spatially correlated. After the range, the semivariance has no variance. It means that the data points out of range do not influence on semivariance. Therefore, the data points outside of the range were not necessary to be used for the analysis.

4.3 PGV Contours

Fig. 8 shows the PGV contours generated by kriging superimposed on the PGV contours generated by TIN. Contours were developed by superimposing a grid of 300 $\times$300 cells on the area in Fig. 6. Values of ln (PGV)

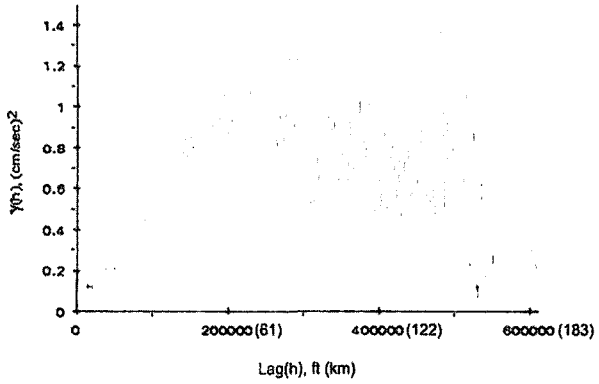

Fig. 4. Variogram model of ln (PGV) data

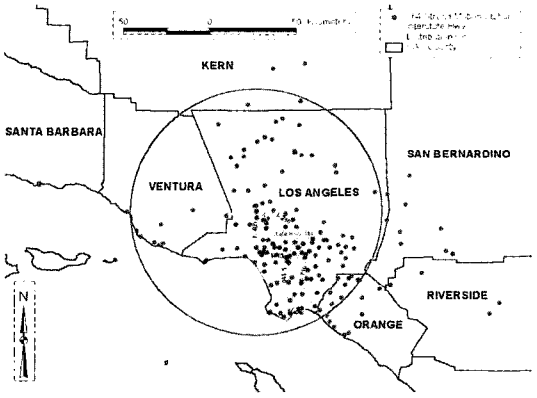

Fig. 6. Spatially distributed PGV records from 164 Strong motion stations

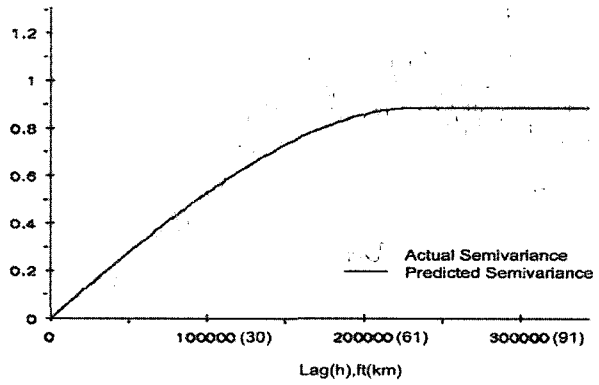

Fig. 5. Spherical variogram model of ln(PGV) data with a range = 237,000 ft (72 km)

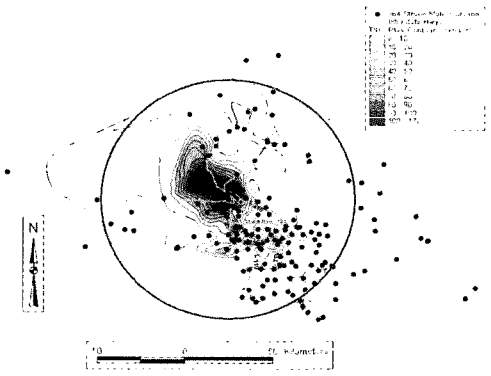

Fig. 7. PGV contours with respect to 164 strong motion stations

were estimated by ordinary kriging at nodal points centered in each cell. The ln (PGV) values were converted to PGV from which the contour lines for ordinary kriging are plotted. As shown in Fig. 8, the PGV contours developed by kriging are similar to those developed by TIN, especially for the contour lines with low PGV values.

Fig. 9 shows the ln (PGV) variance contours obtained from ordinary kriging superimposed on the locations of strong motion stations. The stations located in Los Angeles Basin and central part of San Fernando Valley areas are separated by the distance of 0.7 to 2.8 km. Because the stations are closely located each other, the variance is relatively small. In contrast, the stations located in western part of San Fernando Valley and Santa Monica Bay areas are separated by the distance of 6.7 to 8.8 km. The greater amount of variance is associated with fewer stations separated by greater distances.

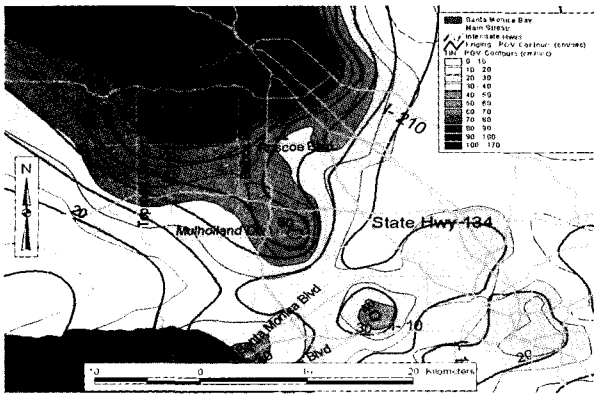

Fig. 8. PGV contours from kriging superimposed on PGV contours from TIN

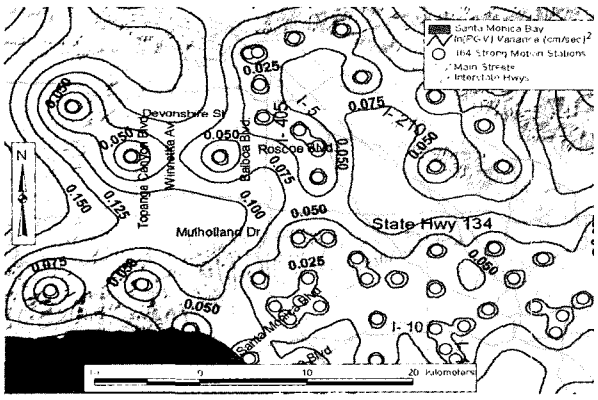

Fig. 9. Ln (PGV) variance contours superimposed on strong motion station

4.4 Correlation of Cast Iron Pipeline Damage with PGV

Fig. 10 shows that cast iron (CI) pipeline repair rate (RR) contours superimposed on PGV zones were developed by ordinary kriging. Using the GIS database, a pipeline repair rate was calculated for each PGV zone, and correlations were made between the repair rate and average PGV for each zone.

Fig. 11 shows the linear regression lines for CI RR associated with PGV obtained from kriging and TIN. There is no significant difference between the regression lines.

4.5 Correlation of Residential Building Damage with PGV

Table 2 lists total number of existing buildings that

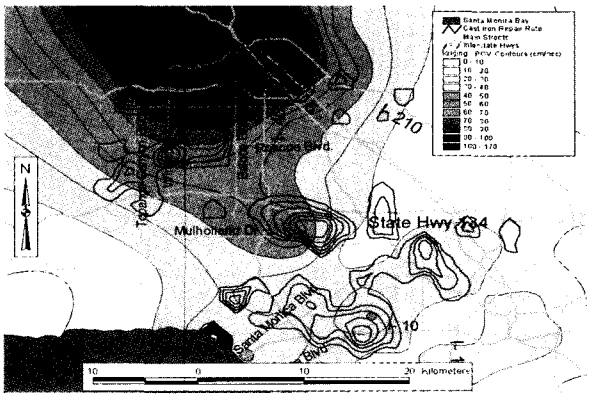

Fig. 10. CI pipeline repair rate plotted relative to PGV from kriging

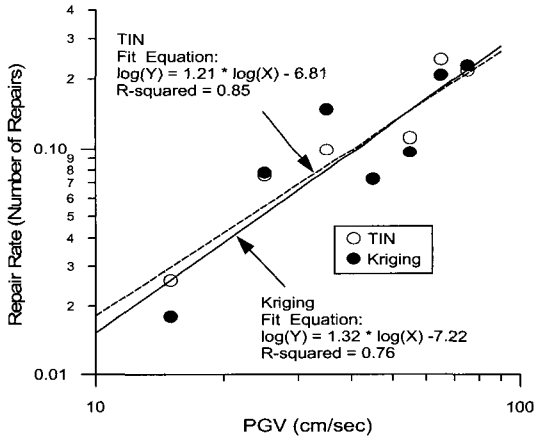

Fig. 11. CI pipeline repair rate correlation with PGV from TIN and kriging

Table 2. Total number of existing buildings associated with PGV (cm/sec) for tax assessor grid data

PGA (cm/sec)	Number of Existing Buildings	Percentage of Existing Buildings (%)
5		
15	36952	13.3
25	34804	12.5
35	34862	12.5
45	45038	16.2
55	41041	14.7
65	23158	8.3
75	12372	4.4
85	12629	4.5
95	7966	2.9
105	8356	3.0
115	7547	2.7
125	7939	2.8
135	3457	1.2
145	1415	0.5
155	1126	0.4
165	0	0.0
Total	278662	100.0

were affected by various PGVs for tax assessor grid data. The data were assembled from the GIS by summing all damaged buildings inside each PGV zone. The PGV assigned to each zone is the mid-range value.

Tables 3 and 4 summarize data associated with residential timber frame buildings in a matrix format for tax assessor grid data. Table 3 lists the number of buildings that were affected by various PGV according to damage factor (DF) expressed as a percentage. The data were assembled from the GIS by summing all damaged buildings of a given DF inside each PGV zone. Table 4 lists the DRs that were computed for various combinations of DF and PGV. Damaged buildings were not observed in areas with PGV < 10 cm/sec. The weighted average damage ratios, calculated by summing all damaged buildings for a given PGV, are shown in the bottom row.

To establish a procedure that minimizes local biases and at the same time utilizes a sample size small enough to evaluate data trends over a meaningful range of variables, a minimum sample size of approximately 2.0 - 2.5% of the population as described in Appendix A was adopted when developing regressions among damage ratio, PGV, and DF.

Fig. 12 shows the linear regression of damage ratio (DR) vs. peak ground velocity (PGV) for various damage

Table 3. Number of residential timber-frame buildings according to PGV (cm/sec) and DF (%) for tax assessor grid data

PGV\DF	≥ 5%	≥ 10%	≥ 20%	≥ 30%	≥ 40%	≥ 50%	≥ 60%	≥ 70%	≥ 80%	≥ 90%
5										
15	111	54	19	14	11	8	4	4	3	2
25	814	344	114	70	56	48	35	29	21	13
35	1226	593	242	118	86	62	41	33	25	19
45	1923	894	303	180	135	402	76	71	56	43
55	2619	1174	402	231	151	446	79	65	53	38
65	1351	627	202	96	62	49	31	22	21	17
75	1022	430	168	93	76	56	36	34	25	19
85	1248	614	221	104	67	44	23	19	9	6
95	852	411	161	76	58	43	21	15	10	9
105	822	411	175	74	44	33	15	14	10	9
115	656	292	122	70	59	42	23	20	17	7
125	637	333	140	79	59	48	40	28	22	15
135	295	157	67	31	22	17	13	11	8	5
145	140	65	30	20	11	7	3	1	1	1
155	64	32	17	11	8	7	6	5	5	
165										
Total	13780	6431	2383	1267	905	682	446	371	286	203

factors (DFs). Most regressions for PGV are quite consistent with the data trend as indicated by relatively high r^2. The r^2 values are slightly less than those resulting from TIN. Fig. 12 (b) compares the DR regressions resulting from kriging with those resulting from TIN. As shown in the figure, linear regressions of DRs from kriging and TIN are very similar. The predicted values of DRs from TIN are slightly higher than those from kriging.

The data in Fig. 12 are replotted in Fig. 13 as a linear regression that accounts for the relationship among DR, DF, and PGV. As evinced by high r^2 and generally good characteristics with respect to residuals, the regression in Fig. 13 provides a good fit to the data.

Fig. 14 shows standardized residuals (Draper and Smith, 1981) plotted with respect to scaled PGV for tax assessor

Table 4. Damage ratio (%) of residential timber-frame buildings according to PGV (cm/sec) for DF(%) for tax assessor grid data

PGV\DF	≥ 5%	≥ 10%	≥ 20%	≥ 30%	≥ 40%	≥ 50%	≥ 60%	≥ 70%	≥ 80%	≥ 90%
5										
15	0.30	0.15	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.01
25	2.34	0.99	0.33	0.20	0.16	0.14	0.10	0.08	0.06	0.04
35	3.52	1.70	0.69	0.34	0.25	0.18	0.12	0.09	0.07	0.05
45	4.27	1.98	0.67	0.40	0.30	0.23	0.17	0.16	0.12	0.10
55	6.38	2.86	0.98	0.56	0.37	0.28	0.19	0.16	0.13	0.09
65	5.83	2.71	0.87	0.41	0.27	0.21	0.13	0.09	0.09	0.07
75	8.26	3.48	1.36	0.75	0.61	0.45	0.29	0.27	0.20	0.15
85	9.88	4.86	1.75	0.82	0.53	0.35	0.18	0.15	0.07	0.05
95	10.70	5.16	2.02	0.95	0.73	0.54	0.26	0.19	0.13	0.11
105	9.84	4.92	2.09	0.89	0.53	0.39	0.18	0.17	0.12	0.11
115	8.69	3.87	1.62	0.93	0.78	056	0.30	0.27	0.23	0.09
125	8.02	4.19	1.76	1.00	0.74	0.60	0.50	0.35	0.28	0.19
135	8.53	4.54	1.94	0.90	0.64	0.49	0.38	0.32	0.23	0.14
145	9.89	4.59	2.12	1.41	0.78	0.49	0.21	0.07	0.07	0.07
155	5.68	2.84	1.51	0.98	0.71	0.62	0.53	0.44	0.44	
165										
W.Aver.	4.95	2.31	0.86	0.45	0.32	0.24	0.16	0.13	0.10	0.07

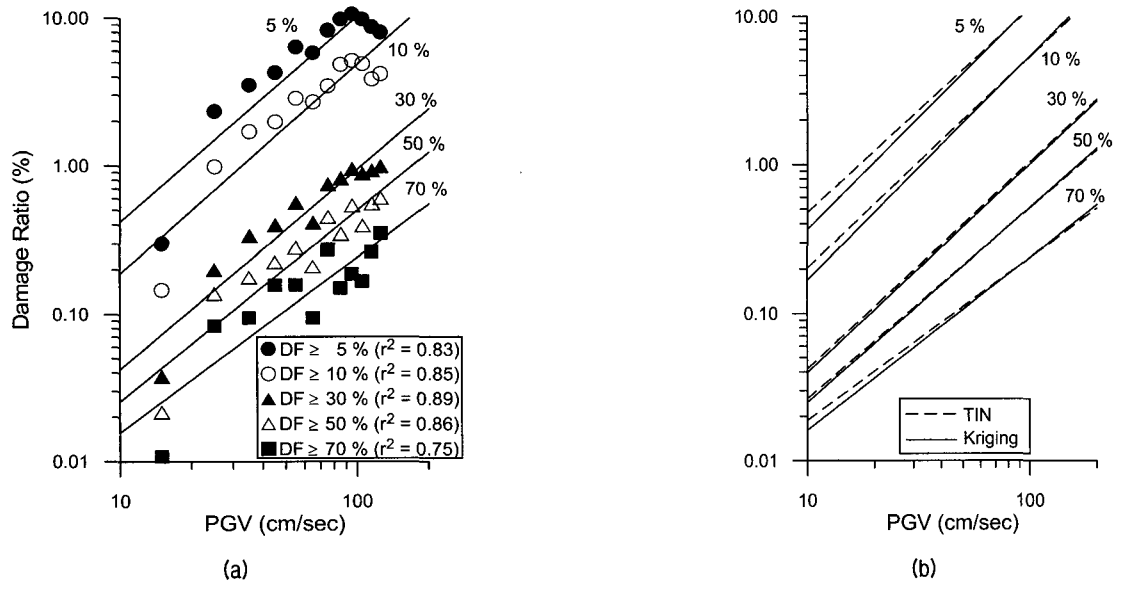

Fig. 12. Damage ratio regression for PGV

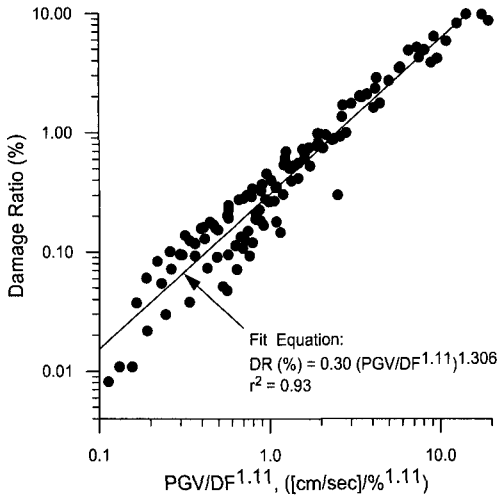

Fig. 13. Damage ratio regression for scaled PGV

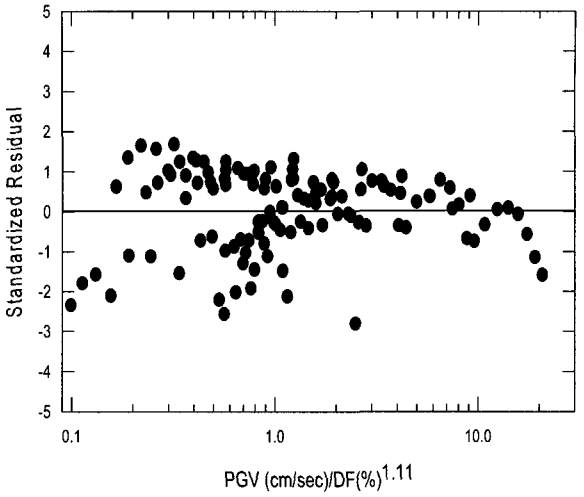

Fig. 14. Standardized residual vs. scaled PGV

grid data. The data are distributed in a reasonably uniform fashion about the population regression. There appears to be some bias towards larger variation at lower values of scaled PGV.

5. Confidence Level of PGV

One of the strengths of using a statistical approach is that it is possible to calculate a statistical measure of uncertainty for the prediction of PGV. In this case, the PGV at unsampled locations can be estimated. Kriging provides the estimated value at each location and its variance. If we assume that these values represent the population mean and variance, we can determine the PGV that corresponds to a 90% exceedance value. In other words, we can estimate a level of PGV at each location that has a 90% chance of being exceeded. This leaves only a 10% chance that a lower value may have occurred. This adjustment allows us to compensate for the spatial variability of the data and make predictions with higher confidence than those predicted on the basis of the mean value.

Let $\phi(s)$ be the distribution function of the standard normal variate, s. The probability of a seismic parameter, SP, being less than a certain value $[\ln(SP)]_{0.10}$ at the 10% level is:

$$\phi\left[\frac{\{\ln(SP)\}_{0.10}-\mu}{\sigma}\right]=0.10 \tag{8}$$

where ϕ is the standard deviation or square root of the variance.

Eq. 8 reduces to

$$\frac{\{\ln(SP)\}_{0.10}-\mu}{\sigma}=\phi^{-1}(0.10)=-\phi^{-1}(0.90)=-1.28 \tag{9}$$

which means

$$[\ln(SP)]_{0.10}=-1.28 \tag{10}$$

In this case, μ and σ are the mean and standard deviation of either the ln (PGV), both of which are obtained directly from ordinary kriging. It is recognized that $[\ln(SP)]_{0.10}$ refers to a probability P = 10% that a lower value exists. Correspondingly, $[\ln(SP)]_{0.10}$ refers to a probability 1 - P = 90% that a larger value exists. Thus, we can develop regressions using such values with 90% confidence that they will not represent overestimated parameters.

It should be noted that the statistics associated with characterizing the spatial variability of PGV are not based on random sampling. The purpose of the statistical interpretation is to predict seismic parameters at un-sampled locations for a specific earthquake and not to predict seismic parameters during future earthquakes. The confidence reflects a stochastic process that governs the

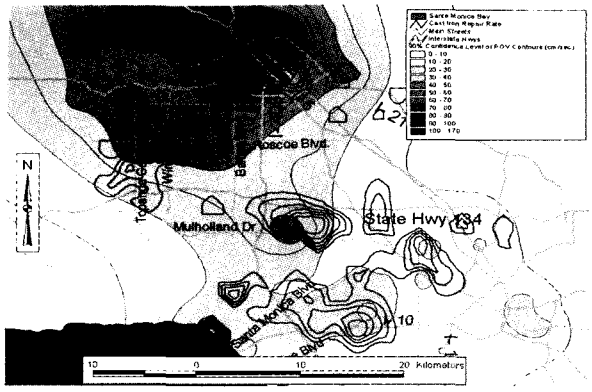

Fig. 15. 90% confidence level of PGV contours superimposed on mean PGV contours

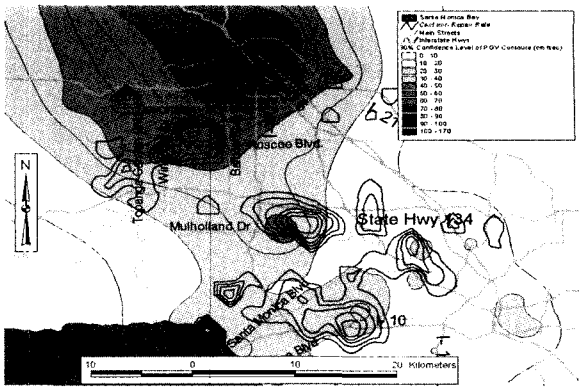

Fig. 16. CI pipeline repair rate plotted to 90% confidence level of PGV contours

distribution of strong motion. Relationships are deduced from the regression analyses of the Northridge data. Assuming similarity is preserved in future earthquake events for the characteristics of strong motion, pipelines, and residential buildings, the relationships obtained in this study can be applied for earthquake loss estimation, irrespective of the time and location of future seismic activity.

Fig. 15 shows the 90% confidence level of PGV contours superimposed on the mean PGV contours, which are shown in Fig. 10, generated by ordinary kriging. Ln (PGV) values at the locations of strong motion instruments were used to predict ln (PGV) values in a grid of 300×300 cells using the spherical variogram model with a range of 237,000 ft (72 km). The 90% confidence level of ln (PGV) in each cell was calculated by Eq. 10. Ln (PGV) values were converted to arithmetic PGV values and then PGV contours were generated using TIN. As shown in the figure, the 90% confidence level of PGV contours is very similar to the mean PGV contours and the PGV zones of 90% confidence level are positioned at relatively higher PGV zones of mean.

5.1 Correlation of Cast Iron Pipeline Damage with PGV

Fig. 16 shows cast iron (CI) pipeline repair rate contours superimposed on 90%-confidence-level PGV zones. The 90% confidence level PGV zones were developed by Eq. 10 with the PGV records obtained from

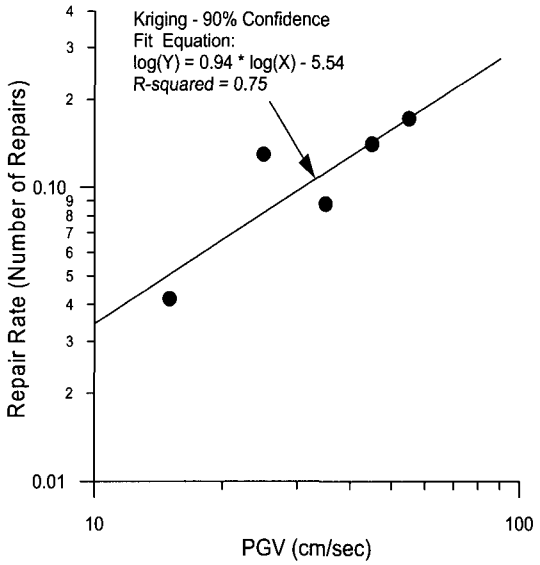

Fig. 17. CI pipeline repair rate correlation with 90% confidence values of PGV

164 strong motion instruments. Using the GIS database, a pipeline repair rate was calculated for each PGV zone, and correlations were made between the repair rate and average PGV for each zone.

Fig. 17 shows the linear regression of RR versus 90% confidence values of PGV. There is a close statistical fit of the regression line to the data, indicating a good correlation of repair rate with PGV. Fig. 18 shows CI repair rates obtained from the regression line corresponding to 90% confidence values as well as the mean PGVs determined by ordinary kriging. As expected, the regression for the 90% confidence values plots above the mean PGV regression.

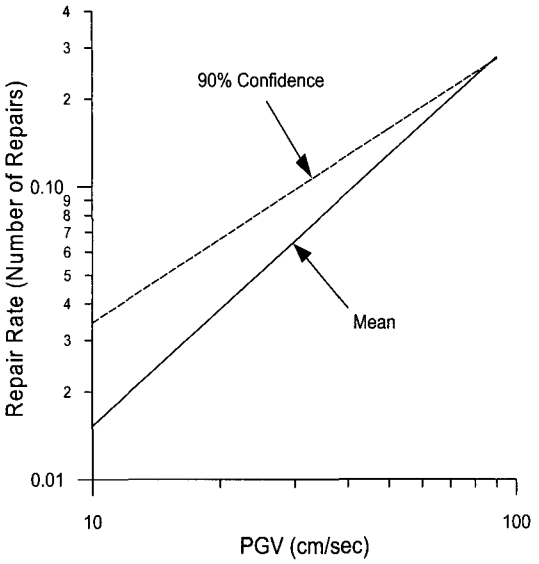

Fig. 18. CI pipeline repair rate correlation with 90% confidence values of PGV and mean PGV

6. Conclusions

To investigate the spatial variability of peak ground velocity (PGV), ordinary kriging was performed. Ordinary kriging is a geostatistical process for making unbiased estimates at unsampled locations with the smallest estimation variance (e.g., Burrough and McDonnell, 1998; Clark and Harper, 2000). It does so by quantifying the spatially correlated variance of the differences in seismic parameters at all sampled locations in the form of a variogram. The variogram is then used to make weighted average estimates at unsampled locations from neighboring recorded values.

Kriging is an interpolation procedure, like inverse squared distance or Triangulated Irregular Network (TIN) techniques, by which strong motion parameters are projected from geocoded measurements to locations where no measurement is available. One of the great strengths of ordinary kriging is that it provides a statistical measure of uncertainty for estimates of seismic parameters.

In contrast, interpolation procedures, such as TIN, provide no statistical measure of uncertainty. Regression analyses for predicting pipeline and residential building damage without statistical procedures are based on the assumption that seismic parameters are known without error and that any statistical measure of uncertainty will apply only for pipeline repair rates and residential building damage ratios.

To utilize ordinary kriging it is necessary to evaluate the type of statistical distribution model that best represents the seismic parameter of interest. Goodness-of-fit assessments, based on Kolmogorov-Smirnov and chi-square tests of the strong motion measurements show that a log-normal model for PGV is much better than an arithmetic one. Moreover, an evaluation of the Cressie goodness-of-fit statistics (Clark and Harper, 2000) for different variogram models shows that a spherical model is best suited for the application of kriging with the Northridge earthquake data set.

Contours of PGV developed with ordinary kriging compare very favorably with those developed with TIN procedures. Moreover, regressions of pipeline repair rate (RR) and residential building damage ratio (DR) vs. PGV developed from ordinary kriging are remarkably similar to and consistent with those generated by TIN interpolations.

By assuming that kriging estimates represent the population mean and variance, PGV values were readily calculated for 90% exceedance values. In other words, at each location estimates were made of PGV with 90% chance of being exceeded. This leaves only a 10% chance that a lower value may have occurred.

Regressions of pipeline RR and residential building DR were made for 90% exceedance values of PGV and compared with the mean values from kriging as well as PGV estimates from TIN procedures. Regressions based on 90% confidence address concerns about the uncertainty of spatially variable data. They provide a means of compensating for this uncertainty and engaging in loss estimation with higher confidence than possible with regressions developed on the basis of mean or apparent average values.

Acknowledgements

The research described in this paper was supported through the Multidisciplinary Center for Earthquake

Engineering Research (MCEER), Buffalo, NY and the National Science Foundation (NSF). Support for GIS visualization development was also provided by the Institute for Civil Infrastructure Systems (ICIS), New York, NY. Special thanks are expanded to Ron Eguchi and Charles Huyck of ImageCat, Inc., Long Beach, CA, for their assistance in acquiring and interpreting databases related to Northridge earthquake building damage.

References

1. Boore, D. M., W. B. Joyner, and T. E. Fumal (1993), "Estimation of Response Spectra and Peak Acceleration from Western North American Earthquakes : An Interim Report", *Open-File Report 93-509*, Menro Park, CA, U.S. Geological Survey.
2. Burrough P. A. and R. A. McDonnell (1998), *Principles of Geographical Information Systems*, Oxford University Press, Oxford, England, p.333.
3. Clark I. and W. V. Harper (2000), Practical Geostatistics 2000, Ecosse North America Llc, Columbus, OH.
4. Cressie, N. A. C. (1993), *Statistics for Spatial Data*, John Wiley & Sons, Inc., Revised Edition, NY.
5. Draper, N. R. and H. Smith (1981), *Applied Regression Analysis, Second Edition*, John Wiley and Sons, New York, NY, p.709.
6. ESRI (1994), The Arc/Info User's Guide, *Environmental Systems Research Institute, Inc.*, Redlands, CA.
7. Jeon, S.-S. (2002), "Earthquake Performance of Pipelines and Residential Buildings and Rehabilitation with Cast-In-Place Pipe Lining Systems", Ph.D. Thesis, Cornell University, Ithaca, NY.
8. Ripley, B. D. (1987), *Spatial Statistics*, John Wiley & Sons, New York, NY.

(received on Mar. 2, 2004, accepted on Jul. 26, 2004)