AN INVERSE ANALYSIS OF UNOBSERVED TRIGGER FACTORS OF THE SLOPE FAILURES BASED ON STRUCTURAL EQUATION MODELING

Hirohito KOJIMA* and Shigeyuki OBAYASHI*

* Remote Sensing Lab., Science University of Tokyo, 2641 Yamazaki Noda-City, JAPAN 278-8510
kojima_h@rs.noda.sut.ac.jp

KEY WORDS: slope stability evaluation, SEM, trigger factor, inverse analysis, satellite data, and geographical information

ABSTRACT:

This paper presents an inverse analysis of the unobserved trigger factors with respect to the slope failures based on the Structural Equation Modeling (SEM). The quantitative prediction models generally construct the relationship between the past slope failures and the causal factors (i.e. lithology, soil, slope, aspect, etc.), but does not deal with the trigger factors(i.e. rail fall, earthquake, etc.), due to the difficulties of pixel-by-pixel observation of the trigger factors itself. As a measure, in this study, an inverse analysis algorithm on the "trigger factors" is proposed, which consists of the following steps:

- **Step1:** The relationship between the slope failures (as the endogenous variables), the causal and trigger factors (as the exogenous variables) are delineated on the path diagram used in SEM.
- **Step2:** The regression weights in the path diagram are estimated in minimizing the errors between the observed and reemerged "variance-covariance matrix" by the model.
- **Step3:** As an inverse estimation, through the measurement equation in SEM between the causal and the trigger factors, a "Trigger Factor Influence map (TFI map)" is newly produced.

As an application, the TFI maps are produced with respect to the "slope failures" and the "landslides," respectively. Furthermore, as a final product, the difference of those TFI maps are delineated on a "Difference map (DIF map)." The DIF map and its interpretation are indeed useful not only for assessing the hazardous area affected by the trigger factors, but also as a "heuristic information" for locating places for setting the field measuring systems.

1. INTRODUCTION

"When, Where and What scale" of the slope failures and the landslides are the important aspects to endure the person's life as well as the social- and economical-infrastructures against the unpredictable risk. Due to the limitation of the detail field investigation, the research approaches applying the satellite remote sensing data and the various kinds of geographical information (termed "causal factor") are highly expected for identifying the hazardous area affected by the slope failures and landslides as well. However, under the present situation, the quantitative prediction models for the slope failures deals with only those causal factors (Carrara et al., 1995; Chung et al., 1995; Kasa et al., 1991; Obayashi et al., 1999), but not applying the "trigger factors", such as the local downpour, earthquake, weathering, etc., because of the difficulties of pixel-by-pixel observation of the trigger factors itself. As another viewpoint against the previous researches, the trigger factors should be treated as "unobserved factors" in terms of time and space in prediction. To estimate such trigger factors, the substantial questions are

- How can we incorporate the "trigger factors" in prediction modeling? ; and,
- Is it possible to estimate the "trigger factors," quantitatively?

With those issues as background, we have tackled the following outstanding subjects:

- To construct an inverse-analysis algorithm for the "trigger factors", based on the Structural Equation Modeling (SEM).
- To produce the Trigger Factor Influence maps (termed TFI map) with respect to the slope failures and the landslides, as well as to consider its application for the landslide hazard assessment.

2. STUDY AREA AND PREDICTION MODEL

2.1 Study Area and Spatial Input Data Set

The study area is located on Futtu in Chiba prefecture, Japan. In the rainy season between July and August in 1988, the local-downpour with continuous rainfall had caused the slope failures and landslides in this study area. Through the field investigation and the aerial photographs, those occurrences were precisely plotted on the topographical map as the training data sets for constructing the prediction model.

The quantitative prediction model constructed the relationship between those past occurrences and the following nine "causal factors": (1) Soil, (2) Surface geology, (3) Vegetation, (4) Land cover, (5) Vegetation index, (6) Slope, (7) Aspect, (8) Elevation, and (9) Drainage. Each map consists of 100×50 pixels (3.0 Km $\times$ 1.5 Km, 30m/pixels corresponding to the ground resolution of the Landsat TM data). The latter four factors were produced based on the Digital Elevation Model (DEM). The experts in each research field have made the Soil-, Surface geology- and the vegetation-map. The land cover map is made through the maximum likelihood classification for the Landsat TM data. The vegetation-index map is also produced by calculating the Normalized Vegetation Index (NVI) given by

$$NVI = (B7 - B5) / (B7 + B5) \quad (1)$$

where $B5$ and $B7$ are the digital numbers in each pixel corresponding to TM-Band 5 and TM-Band 7, respectively.

2.2 Quantitative Prediction Model

Figure 1 shows the inverse analysis concept of the trigger factors expanding the previous quantitative prediction model. Chung and Fabbri (1999) have adopted the formulas for geologic hazard zonation as a part of "favorability function" approaches, and the various procedures have been applied to the landslide prediction. To promote those prediction models as well as optimizing prediction, the practical analytical procedures have been presented as follows; i) Comparative strategy of the prediction models (Kojima et al., 1998, 1999), ii) Analysis of the landslide types (Kojima et al., 2000), iii) Testing on the time-robustness in prediction (Kojima et al., 2001), and iv) Sensitivity analysis of the prediction models with respect to the causal

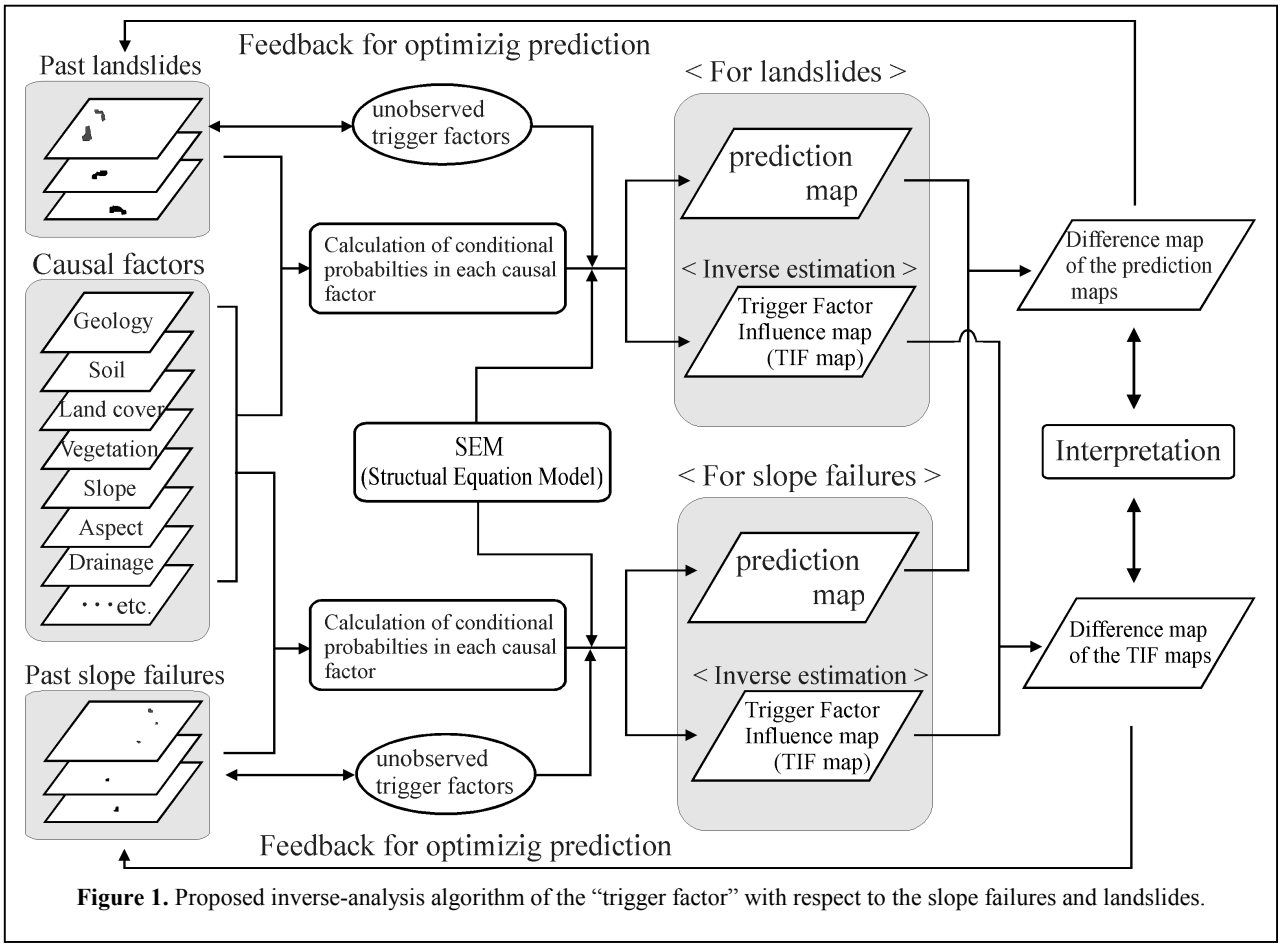


Figure 1. Proposed inverse-analysis algorithm of the "trigger factor" with respect to the slope failures and landslides.

factors (Chung et al., 2002).

Those analytical procedures are crucial components in the prediction models, and are indeed useful not only for the experts working on the landslides, but also for the end-users of the prediction models. However, the conventional prediction models generally construct the relationship between the past slope failures and the causal factors, but could not apply the "trigger factors," because of the difficulties in observing the trigger factors "pixel-by-pixel". As a measure in this study, an inverse analysis algorithm on the "trigger factors" is presented, based on the Structural Equation Modeling (termed SEM) that was originally proposed by Joreskog and Lawley (1968). SEM is also well known as the "analysis of covariance structures," in a word, SEM could be regarded as a model integrated the regression analysis jointly with the exploratory factor analysis. The detail of SEM itself is available for the references (Joreskog et al.1968).

3. INVERSE ANALYSIS OF TRIGGER FACTOR

Figure 1 shows the proposed inverse-analysis algorithm of the "trigger factor," which consists of the following steps:

3.1 Conditional Probabilities as the Input Data

To construct a probability model for slope failure hazard, consider the following proposition:

F_p : " a pixel p will be affected by a future slope failure of a given type D."

The conditional probabilities in each causal factor given by

$$\textbf{\textit{Prob}}(F_p|C_{ij}) = T_{ij} / N_{ij} \quad (2)$$

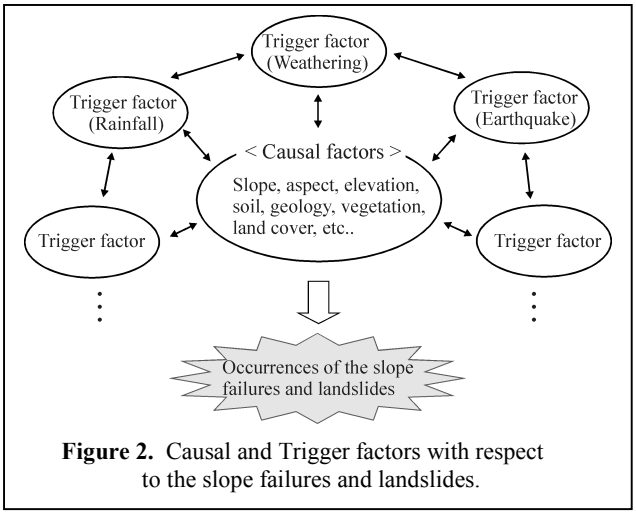


Figure 2. Causal and Trigger factors with respect to the slope failures and landslides.

where C_{ij} is the i^{th} category of the j^{th} causal factor; N_{ij} is the number of pixels of C_{ij}; and T_{ij} is number of pixels of the past slope failures or landslides that had occurred in the area corresponding to C_{ij}. $\textbf{\textit{Prob}}(F_p|C_{ij})$ are used as the input data for the SEM-based analysis.

3.2 Path Diagram

To make a prediction model, the relationship between the slope failures (as the endogenous variables), the causal and trigger factors (as the exogenous variables) should be delineated on the path diagram used in the SEM. **Figure 2** shows the intricate relation between the causal and trigger factors with respect to the

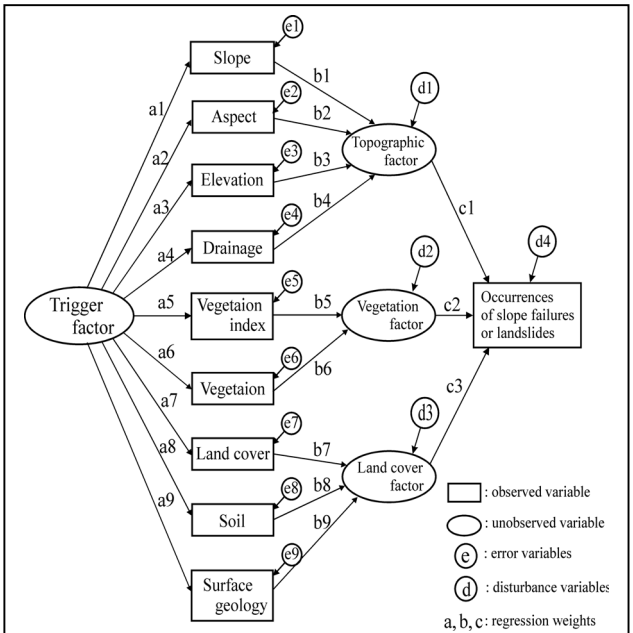


Figure 3. Path diagram for the SEM-based analysis.

slope failures and landslides. As a counterpart of **Figure 2**, let us consider the path diagram as shown in **Figure 3** that is called a recursive model. ***Prob***$(F_p \mid C_{ij})$ of **Equation 2** are the input data as the exogenous variables, while the pixels corresponding to occurrences and non-occurrences of the slope failures as well as landslides are assigned to the value "1" or "0", respectively, that are used as the endogenous variables.

3.3 Hypothesis Testing

Not knowing the trigger factors, the program is how to estimate the path weights of $\{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n\}$ in **Figure 3**. Through the estimation procedure in SEM, those are estimated by minimizing the errors between the observed variance-covariance matrix and the reemerged one. Among various estimation procedures, i.e., maximum likelihood estimation, asymptotically distribution-free estimation, generalized least squares estimation, 'scale free' least squares estimation, unweighted least squares estimation, etc., Maximum likelihood estimation procedure is selected in this study, which is generally reported as a better estimator for the large population than the others.

To make clear the significance introducing the "unobserved trigger factors" in **Figure 3**, let us consider the following path models for the slope failures and landslides, respectively:

- Model A: using both causal and trigger factors for the "slope failures," as shown in **Figure 3**.
- Model B: only using causal factors for the "slope failures."
- Model C: using both causal and trigger factors for the "landslides," as shown in **Figure 3**.
- Model D: only using causal factors for the "landslides."

As the hypothesis testing for those models, the Chi-square value, the Goodness of Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI), the Akaike Information Criterion (AIC), and the Root Mean Square Error Approximation (RMSEA) are applied as the statistical measures of fit. **Table 1** shows the results of the hypothesis test and the fit measures.

The Chi-square value for Model A is higher than that for Model B, also the probability level in Model A indicates more than 0.05, but not in Model B, which means that Model A could not be rejected under the significance level "0.05", while there is no reliability on the identification of Model B. As the other measures of fit, the GIF and the AGFI need to be more than 0.9, conversely, the RMSE should be less than 0.08 for the model selection. Furthermore, among those models, the only model with the lowest AIC should be selected.

Table 1. Hypothesis test results and the fit measures.

Measures of fit		For the slope failures		For the landslides	
		Model A	Model B	Model C	Model D
Chi-square test	Chi-square	18.3	358.5	58.8	519.1
	Degree of freedom	28	37	28	37
	Probability level	0.316	0.000	0.227	0.000
GIF		0.993	0.786	0.975	0.801
AGIF		0.986	0.778	0.951	0.705
AIC		24.5	400.5	65.0	523.6
RMSEA		0.025	0.147	0.04	0.167

Notes)
Model A: using both causal and trigger factors for "slope failures."
Model B: using only causal factors for "slope failures."
Model C: using both causal and trigger factors for "landslides."
Model D: using only causal factors for the "landslides."

Based on the above experiments, for the slope failures in the study area, we would say at least that Model A is better than Model B on the statistical grounds. Furthermore, for the landslides, **Table 1** indicates that Model C would be better fit against Model D. Those results imply that the path diagram introducing the "trigger factor" shown in **Figure 3** might be meaningful for constructing the prediction model.

3.4 Path Parameter Estimation

Table 2 shows the standardized regression weights of estimated for the selected Model A and Model C side by side. The difference of those parameter weights between Model A and Model C is obvious, which corroborates that the trigger factors with respect to the slope failures and the landslides may be different.

Focusing on the path-relation between the trigger factor and the causal factors, in Model C for the landslides, the highest weight "0.869" comes from the path between the trigger factor and the land-cover. While in Model A for the slope failures, it is interesting to note that the negative weights come from the path between the elevation, drainage and the land-cover. Comparing such differences of the path weights, we can evolve the factor analysis from various points of view on the slope failures and the landslides.

3.5 Prediction Map

Based on the path model of **Figure 3**, the prediction maps could be produced. **Plate 1(a)** and **Plate 1(b)** show the prediction maps with respect to the slope failures and the landslides, respectively. **Table 3** shows the description of those prediction maps, which are made by the mini-max discriminate method that classifies the pixels into two groups as "occurrence and non-occurrence (Kasa, Kojima, et al., 1991)." The classified results of these kinds are indeed useful for supporting the decision-making of the landslide

Table 2. Results of the standardized parameter estimates.

Path discription			Parameters in Figure 3	Model A for slope failures	Model C for landslides
Trigger factor	→	Slope	a1	0.232	0.705
	→	Aspect	a2	0.011	0.007
	→	Elevation	a3	-0.030	0.159
	→	Drainage	a4	-0.052	0.447
	→	Vegetaion Index	a5	0.001	0.010
	→	Vegetation	a6	0.247	0.459
	→	Land cover	a6	-0.203	0.869
	→	Soil	a7	0.294	0.273
	→	Surface geology	a8	0.260	-0.075
Slope	→	Topographic factor	b1	0.411	-0.004
Aspect	→		b2	0.748	0.703
Elevation	→		b3	0.451	0.163
Drainage	→		b4	0.271	0.680
Vegetaion Index	→	Vegetaion factor	b5	0.662	0.637
Vegetation	→		b6	0.749	0.768
Land cover	→	Land cover factor	b7	-0.034	0.953
Soil	→		b8	0.856	0.122
Surface geology	→		b9	0.449	-0.094
Topographic factor	→	Occurrences of slope failures or landslides	c1	0.105	0.100
Vegetaion factor	→		c2	0.065	0.093
Land cover factor	→		c3	0.051	0.061

prevention plans, against the general prediction map ranked with several hazardous levels.

In **Plate 1(a)** and **Plate 1(b)**, it is obvious that the prediction patterns are fairly different, which means the causal and the trigger factors between the slope failures and the landslides might be different. In the next section, as further investigation, let us discuss about the inverse estimation of the trigger factor for the slope failures and landslides as well.

3.6 Inverse Estimation of Trigger Factor

Note that the path components connecting "unobserved variables to each other" and "observed variables to unobserved variables" are often called the "structural equation" and "measurement equation," respectively. In this study, through the measurement equation, the influence of the trigger factor are inversely estimated pixel-by-pixel, and those are delineated on a "Trigger Factor Influence map (termed TFI map)." In the path diagram shown in **Figure 3**, the measurement equation between the trigger factors (as unobserved variables) and the causal factors (as observed variables) is given by

$$z_{ji} = a_j f_i + e_{ji} \tag{3}$$

where z_{ji} is the conditional probability of the j^{th} causal factor in the i^{th} pixel as shown in **Equation 2**; a_j is the path parameter in **Table 2** linked the j^{th} causal factor with the trigger factor; f_i is the real value of the degree on the trigger factor influence

Table 3. Description of the prediction maps shown in Plate 1(a) and Plate 1(b).

		Training data sets (Past situation)	
		Occurrence	Non-occurrence
Prediction map	Hazardous	Red	Green
	Non-hazardous	Blue	White

in the i^{th} pixel; and e_{ji} is the error for the j^{th} causal factor in the i^{th} pixel. The objective is how to calculate the estimates of $\hat{f}_i$ of the trigger factor inversely, based on **Equation 3**. The inverse function is given by

$$\hat{f}_i = \sum_{j=1}^{p} b_j z_{ji} \tag{4}$$

The parameters of $\{b_1, \ldots, b_n\}$ are determined by minimizing the following least square error:

$$Q = \sum_{i=1}^{n} (f_i - \hat{f}_i)^2 = \sum_{i=1}^{n} (f_i - \sum_{j=1}^{p} b_j z_{ji})^2 \rightarrow \min$$

$$\frac{\partial Q}{\partial b_{j'}} = -2\sum_{i=1}^{n} z_{j'i}(f_i - \sum_{j=1}^{p} b_j z_{ji}) = 0 \tag{5}$$

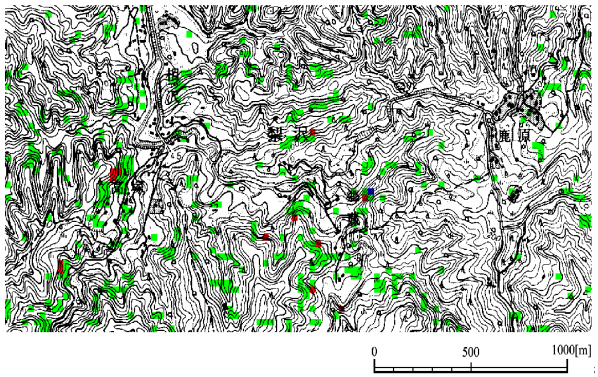
Note that the average and the variance for z_{ji} and f_i are standardized to "0" and "1", respectively. So, the error parameter could be ignored in **Equation 5**. Hence, $\{b_1, \ldots, b_n\}$ are simply given by

$$b_j = \sum_{j'=1}^{p} a_{j'} r^{jj'} \tag{6}$$

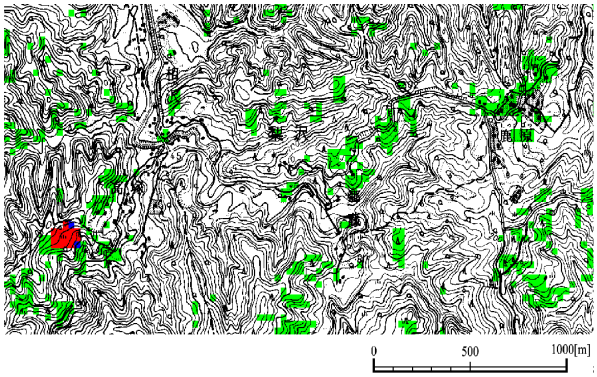
where $r^{jj'}$ is the element (j , j') of the inverse matrix for the correlation matrix. Through **Equation 4**, $\hat{f}_i$ is calculated pixel-by-pixel, and is delineated on the Trigger Factor Influence map (TIF map). **Plate 1(c)** and **Plate 1(d)** show the TIF maps with respect to the slope failures and the landslides, respectively. From those TIF maps jointly with the prediction maps, the following points could be indicated:

- At a glance, it is found that the estimated patterns between the TIF maps with respect to the slope failures and the landslides are clearly different.
- In **Plate 1(c)** on the slope failures, the steeper the slope is, the higher the trigger factor influence is for the most part. On the other hand, in **Plate 1(d)** on the landslides, the gentler the slope is, the higher the trigger factor influence is. Those results corroborate that the trigger factors could be estimated "pixel-by-pixel" through the path model shown in **Figure 3**.
- However, in **Plate 1(c)** on the slope failures, we can also see the areas where the trigger factor influence is high in spite of relatively gentle slope. Especially, for those areas, the detail field investigation should be carried out.

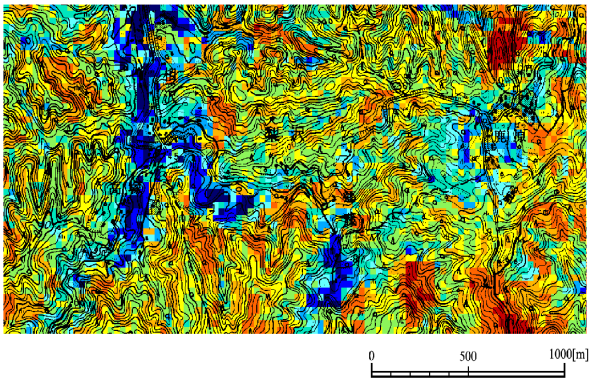
For the practical utilization of the TIF maps, let us consider the

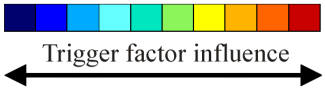


(a) Prediction map with respect to the "slope failures." (cf. Table 3 for the legend)

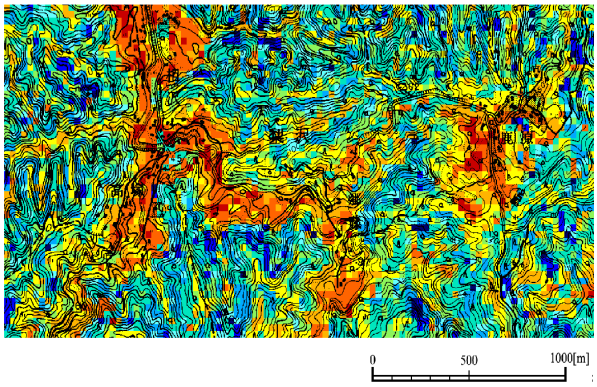


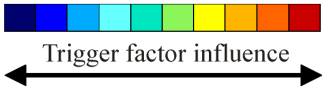
(b) Prediction map with respect to the "landslides." (cf. Table 3 for the legend)





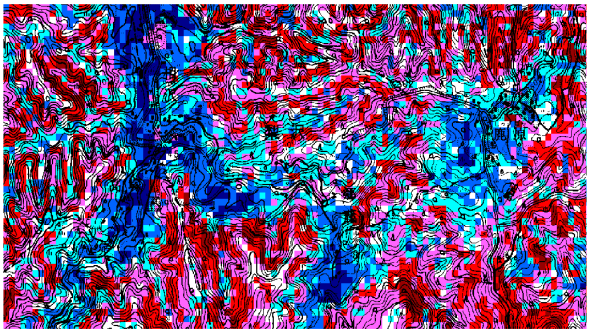
(c)Trigger Factor Influence map (TFI map) with respect to the "slope failures."





(d)Trigger Factor Influence map (TFI map) with respect to the "landslides."

Visualization of differences



(e) Difference map (DIF map) between the trigger factor influence maps with respect to the slope failures and the landslides, respectively.

< Legend of Plate 1(e) >

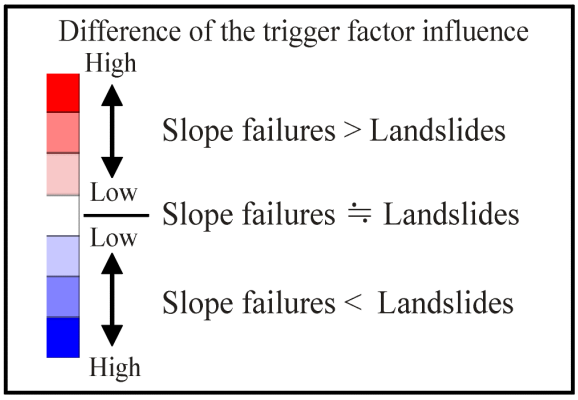


Plate 1 Prediction maps and the Trigger Factor Influence maps (termed TFI map) with respect to the slope failures and the landslides, respectively.

interpretation for differences of the TIF maps with respect to the slope failures and the landslides, respectively.

4. INTERPRETATION OF THE TRIGGER FACTOR INFLUENCE MAP (TIF MAP)

As the conventional interpretation of the prediction maps, the difference map (termed DIF map) is produced between two prediction maps with respect to the slope failures and landslides (Kojima, et al. 1998, 2000, 2001). Similarly, to make clear the difference of the trigger factor influence, the DIF map between two TIF maps for the slope failures and landslides is produced as shown in **Plate 1(e)**. Attention should be paid that we can interpret the difference of the trigger factor influence according to the legend for **Plate 1(e)** as follows:

- **Shade of red:** The degree of trigger factor influence for the slope failures is higher than that of the Landslides.
- **White:** The degree of trigger factor influence for the slope failures is almost equivalent to that of the Landslides.
- **Shade of blue:** The degree of trigger factor influence for the slope failures is lower than that of the Landslides.

Such "heuristic information" might be useful not only for assessing the hazardous area affected by the trigger factors in terms of the types of the slope failures and landslides, but also for improving the cost-effectiveness in locating the places for setting the field measuring systems, i.e. the tensiometer, the rain gage, etc.

5. CONCLUDING REMARKS

In this contribution, we have discussed about an inverse analysis of the unobserved trigger factors with respect to the slope failures as well as the landslides, based on the Structural Equation Modeling (SEM). The results of this study are summarized as follows:

- Due to the difficulties in observing the trigger factors "pixel-by-pixel," we strongly point out the necessity for the inverse estimation of the "unobserved trigger factors" itself. As a measure, through the measurement equation between the causal factors (as observed variables) and trigger factors (as unobserved variables), a "Trigger Factor Influence map (termed TFI map)" is newly produced.
- As an application of the proposed inverse-analysis algorithm, the TFI maps are produced with respect to the slope failures and the landslides, respectively. Furthermore, as a final product, the differences of those TFI maps are delineated on a "Difference map (termed DIF map)."
- Through the DIF map, we can evolve the analysis on the "trigger factor influence" with respect to the slope failures and the landslides, jointly with the prediction maps and the expert's opinions as well.

As for the subsequent subjects, in order to corroborate the practicality of the inverse-analysis algorithm (**Figure 1)**, the additional investigation for other study areas should be carried out. As occasion demands of the investigators, we can readily add the training data sets of other types of slope failures and landslides in analysis. In this point of view, the analytical procedure shown in **Figure 1** is expected to contribute to the landslide hazard assessment as one of the standards.

Furthermore, as for the structure of the path diagram in **Figure 3**, a single "exogenous variable" is considered as the "trigger-factor." However, in practice, either slope failures or landslides are caused by various trigger factors as shown in **Figure 2**. So, the modified path models adding the several exogenous variables as the trigger factors should be investigated to improve the identification of the models itself.

It is no exaggeration to say that the precise estimation of the trigger factor itself is impossible. As one of the measures, the inverse-analysis algorithm presented in this study, as well as "heuristic information" on the DIF map of the trigger factor influence maps, might be effective for identifying the hazardous area affected by the different types of slope failures and the landslides. Such a systematic analysis procedure, as against the limitation of the conventional research approaches for prediction modeling, might be essential to optimize prediction as well as to promote the former quantitative prediction models.

REFERENCES

Carrara, A., M. Cardinali, F. Guzzetti, and P. Reichenbach, 1995. GIS Technology in mapping landslide hazard, *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Academic Publishers, The Netherlands, pp.135-175.

Carrara, A., 1998. Current limitations in modeling landslide hazard, *Proceedings of The International Association for Mathematical Geology*, No.4, pp.195-203.

Chung, C.F., A.G. Fabbri, and C.J. van Westen, 1995. Multivariate regression analysis for landslide hazard zonation, *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Academic Publishers, The Netherlands, pp.107-133.

Chung, C.F., and A.G. Fabbri, 1999. Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing*, Vol.65, No.12, pp.1389-1399.

Chung, C.F., H. Kojima, and A.G. Fabbri, 2002. Applied Geomorphology: theory and practice, Stability analysis of prediction models applied to landslide hazard mapping, *John Wiley & Sons Publication (in press).*

Joreskog, K.G., and D.N. Lawley, 1968. New methods in maximum likelihood factor analysis, *British J. Math. Statist. Psycol*., 21, 85-96.

Kasa, H., M. Kurodai, H. Kojima, and S. Obayashi, 1991. Study on the landslide prediction model using satellite data and geographical information, Landslides, *Bell(ed.), Balkerma, Rotterdam, International landslide symposium*, ISBN90 5410 032 X, pp.983-988.

Kojima, H., C.F. Chung, S. Obayashi, and A.G. Fabbri, 1998. Comparison of strategies in prediction modeling of landslide hazard zonation, *Proceedings of The International Association for Mathematical Geology*, No.4, pp.195-203.

Kojima, H., C.F. Chung, and C.J. van Westin, 2000. Strategy on the landslide type analysis based on the expert knowledge and the quantitative prediction model, *International Archives of Photogrammetry & Remote Sensing*, Vol.33, Part-B7, pp.701-708.

Kojima, H., and C.F. Chung, 2001. Testing on the time-robustness of a landslide prediction model, *Proceedings of The International Association for Mathematical Geology*, Vol.7, pp.41-46.

Obayashi, S., H. Kojima, and C.F. Chung, 1999. Comparison of the accuracy of the slope stability evaluation models and its practical application, *Journal of the Construction Management and Engineering*, *Japan Society of Civil Engineers (in Japanese)*, No.630, 6-44, pp.77-89.