

An inverse and compositional analysis of unobserved trigger factors according to slope failure types

H. Kojima, Y. Taguchi, K. Nishimura & S. Obayashi

Department of Civil Engineering, Tokyo University of Science, Japan

Abstract

This paper presents an inverse- and compositional-analysis of unobserved trigger factors according to slope failure types. Due to the difficulties of pixel-based observation on the trigger factors, we had proposed the inverse analysis algorithm based on the structural equation modeling (SEM). Through the “measurement equation” defined between the causal factors (i.e., observed variables) and the trigger factor (i.e., unobserved latent variable), the trigger factor can be inversely estimated.

As the subsequent subjects for the previous studies, in this contribution, we have tried to decompose trigger factors into the “1st trigger factor” and the “2nd trigger factor” with respect to slope failure types, such as surface slope failure, deep-seated slope failure, and landslide, which had been induced by Niigata Chuetsu Earthquake (Oct. 23, 2004). The 1st and the 2nd trigger factor influence map have been also produced according to the slope failure types. As a final outcome, the differences in these TFI maps are delineated on a “difference (DIF) map” that enables us to analyze the difference of trigger factor influence with respect to slope failure types simultaneously.

Keywords: slope failure types, inverse and compositional analysis of trigger factors, geographical information, satellite remotely sensed data, structural equation modeling.

1 Introduction

A spatial data integration technology using various kinds of geographical information (termed “causal factor”) has become a current trend for identifying the hazardous area affected by slope failures and landslides. Quantitative



prediction models for slope failure occurrences generally elucidate the relationship between past slope failures and causal factors (e.g. geology, soil, slope, aspect, etc.). Due to the difficulties of pixel-by-pixel observation of trigger factors (e.g. earthquake, rainfall, weathering, etc.), The trigger factors as explanatory variables are substituted for some of causal factors in constructing models, on the assumption that there are some correlations between causal factors and trigger factors [1–3].

As a measure, we had tackled to construct a Trigger Factor Inverse analysis model (TFI model) of unobserved trigger factor, in which the relationship between past slope failures (i.e., endogenous variables), causal factors (i.e., explanatory variables), and trigger factors (i.e., unobserved variables) are delineated on the path diagram in the Structural Equation Modeling (SEM) [4, 5]. In the TFI model, through the “measurement equation” defined between the causal factors and the trigger factor (i.e., unobserved latent variable), the trigger factor can be inversely estimated [9].

In the TFI model, the pixels corresponding to past slope failures are generally used as the input data of endogenous variable (i.e., training data sets). The inverse estimated values on trigger factor are delineated on a Trigger Factor Influence map (termed “TFI map”), which depends on the distribution of the types of past slope failures used as the training data sets. Also, in our previous experiments, as for the structure of path diagram used in SEM, a “single exogenous variable” had been considered as main trigger factor of rainfall [9].

However, as either slope failures or landslides are induced by various trigger factors, the modified path model with several exogenous variables as trigger factors should be investigated to improve the identification of path model in SEM approach. This inevitable subject exactly resolves into a decompositional analysis of unobserved trigger factors according to “slope failure types”. With those issues as background, our efforts in this study are to:

- construct an inverse- and decompositional-analysis algorithm of unobserved trigger factors with respect to the “different types” of slope failures; and
- provide a pair-wise comparative strategy of the 1st and the 2nd trigger factor influence maps according to the slope failure types.

2 Background, study area and input data sets

2.1 Necessity of decompositional estimation of slope failure trigger factors

Chung and Fabbri have adopted the formulas for geologic hazard zonation as a part of “favorability function” approaches [1], and the various procedures have been applied to the landslide prediction. To improve the performance of the quantitative models as well as optimizing prediction, the practical analytical procedures had been presented as follows; i) Comparative strategy of the prediction models [6], ii) Analysis of the landslide types [7], iii) Testing on the time-robustness in prediction [8]; and iv) Sensitivity analysis of the prediction models with respect to the causal factors [2].

These analytical procedures are crucial components of the model for optimizing the predictions. The previous quantitative prediction models,



however, generally construct the relationship between the “past slope failures” and the “causal factors”, and do not deal with the “trigger factors”, due to the difficulties of “pixel-by-pixel” observation on trigger factors. To overcome this issue, an inverse analysis algorithm on the “trigger factors” had been constructed [9]. As the structure of path diagram used in the SEM approach, a “single exogenous variable” had been considered as a main trigger factor.

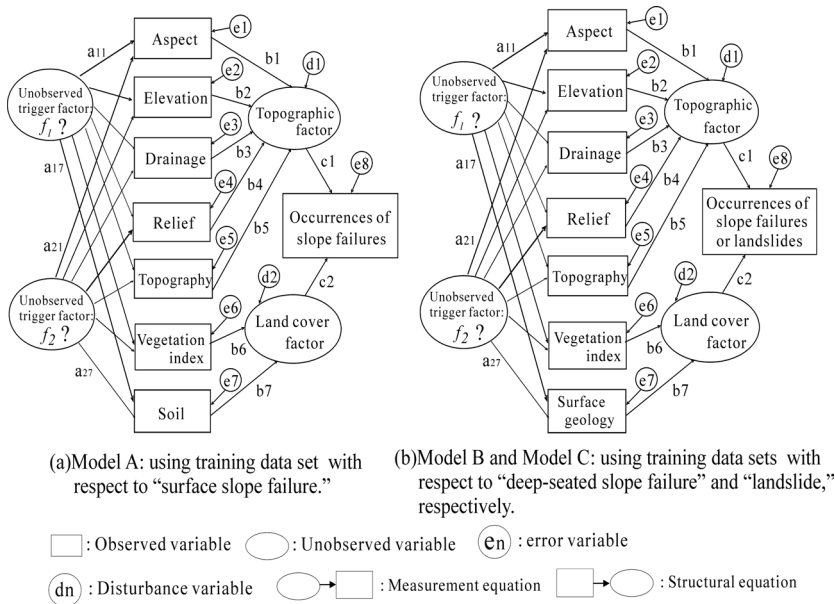


Figure 1: Inverse- and decompositional-analysis of unobserved trigger factors according to slope failure types based on the structural equation modeling.

Note that, the study area had also local-downpour for several days before the huge earthquake of Niigata Chuetsu Earthquake. Therefore the inverse- and decompositional-analysis of such “chain trigger factors” is a requisite function of quantitative models for the better assessments of slope failure hazard. As a measure, in this study, we have tackled the inverse- and decompositional-analysis of unobserved trigger factors (see Figure 1) according to “different types” of slope failures. Through this procedure, the Trigger Factor Influence (TFI) maps with respect to the 1st and the 2nd trigger factor are produced. Furthermore, a pair-wise comparative analysis for these TFI maps has been carried out based on the Hayashi’s quantification method of the fourth type.

2.2 Study area and causal factors as input data sets

The study area is located on Yamakoshi-mura in Niigata prefecture, Japan, where various types of slope failures, such as surface failure, deep-seated slope

failure, and landslide, had been induced by Niigata Chuetsu Earthquake (Oct. 23, 2004). Through the field investigation with the aerial photographs, the different types of slope failures, such as “surface slope failure”, “deep-seated slope failure”, and “landslide”, had been plotted on the topographical map as the training data sets.

In this study, the inverse and compositional analysis model of unobserved trigger factors has been constructed the relationship between those past slope failures and the following nine “causal factors”: (1) Surface geology (2) Soil, (3) Topography, (4) Vegetation index, (5) Aspect, (6) Elevation, (7) Drainage, (8) Relief, and (9) Slope. Each map consists of 120×100 pixels (3.6 Km×3.0 Km, 30m/pixels). The latter five factors were produced based on the Digital Elevation Model (DEM). The experts in each research field had produced the maps of surface geology, soil and topography. From the IKONOS data, the vegetation-index map is also produced by calculating the Normalized Vegetation Index (NVI) (see Appendix A).

3 Model specification

3.1 Conditional probabilities as input data

To evaluate the hazardous area affected by slope failure at each pixel with respect to different types of slope failures, let us consider the following proposition:

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

The conditional probabilities in each pixel, corresponding to the i^{th} category of the j^{th} causal factor, are given by

$$Prob(F_p | C_{ij}) = \frac{T_{ij}}{N_{ij}} \quad (1)$$

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, which had occurred in the area corresponding to C_{ij} . $Prob(F_p | C_{ij})$ is used as the input data (i.e., exogenous variable) for the SEM-based analysis.

To construct a quantitative prediction model, the relationship between the past slope failures of training data sets (i.e., endogenous variables), the causal and trigger factors (i.e., exogenous variables) should be delineated on the path diagram used in the SEM approach. Figure 1 shows the path diagram used in this study, which is called a recursive model. $Prob(F_p | C_{ij})$ of equation (1) is the input data set as exogenous variables, while the pixels corresponding to occurrences and non-occurrences of slope failures are assigned to the value of “1” or “0”, respectively, that are used as the training data sets.

To exclude a multi-collinearity between causal factors, among a pair of causal factors with high correlation (e.g., above 0.7), one of a pair with high partial correlation was selected in analysis. Figures 1(a) and 1(b) illustrate the path diagrams composed of selected causal factors, and the training data sets of these models are as follows:

- Model A: using training data set of “surface slope failure”;

Table 1: Evaluation of model fit.

Measure of fit	Model A	Model B	Model C
GFI	0.999	0.998	0.999
AGFI	0.994	0.991	0.995
AIC	126.9	156.8	112.1
RMSEA	0.024	0.031	0.021

Notes: Model A: using training data set of “surface slope failure”
 Model B: using training data set of “deep-seated slope failure”
 Model C: using training data set of “landslide”.

- Model B: using training data set of “deep-seated slope failure”; and
- Model C: using training data set of “landslide”.

3.2 Evaluating model fit

Not knowing the trigger factors, the program is how to estimate the path weights of $\{a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n\}$ in Figure 1. Through the SEM-based estimation procedure, those are estimated by minimizing the errors between the observed and reemerged “variance-covariance matrix”. Among various estimation procedures (e.g., maximum likelihood estimation, asymptotically distribution-free estimation, generalized least squares estimation, etc.), the maximum likelihood estimation procedure was selected in this study, which is generally reported as a better estimator for the large population.

For evaluating model fit, 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 generally employed statistical measures of fit. Details on these fit measurements are available in Hoyle [5]. Table 1 shows the results of calculating these fit measures. By rule of thumb, GIF and AGFI need to be more than 0.9, conversely, RMSEA should be less than 0.08 for selecting reasonable model. AIC are also used to “comparison” of the model fit. The model with lower value of AIC is considered to have better fit from the other. Based on these criterions, Table 1 gives us an indication of which all models can be accepted as a model. These results imply the significance of adding the plural unobserved trigger factors (i.e., latent variables) to the path diagram.

3.3 Inverse- and compositional-analysis of unobserved trigger factors

Note that the path components connecting “unobserved variables to each other” and “observed variables to unobserved variables” are generally termed the “structural equation” and “measurement equation”, respectively (see Figure 1). In this study, through the measurement equation, the influences of the trigger

factors are inversely estimated at each pixel, and they are delineated on a “Trigger Factor Influence map (termed TFI map)”.

In Model A, the “measurement equation” between the trigger factors (i.e., unobserved variables) and the causal factors (i.e., observed variables) is given by

$$z_{ji} = a_{1j}f_{1i} + a_{2j}f_{2i} + e_{ji} \quad (2)$$

where z_{ji} is the input value of the i^{th} pixel in the j^{th} causal factor as shown in equation (2); f_{1i} and f_{2i} are the unobserved trigger factor corresponding to the 1st and the 2nd trigger factor, respectively (see Figure 1); a_{1j} and a_{2j} are the path parameters that are linked the j^{th} causal factor with the 1st and the 2nd trigger factor; and e_{ji} is the error term of the i^{th} pixel in the j^{th} causal factor.

The objective is to inversely calculate the estimates for f_{1i} and f_{2i} of the unobserved trigger factors. Suppose \hat{f}_{1i} and \hat{f}_{2i} are the estimates of f_{1i} and f_{2i} , respectively, then the inverse functions are given by

$$\hat{f}_{1i} = \sum_{j=1}^p h_{1j} z_{ji} \quad (3)$$

$$\hat{f}_{2i} = \sum_{j=1}^p h_{2j} z_{ji} \quad (4)$$

where h_{1j} and h_{2j} are the inverse parameters; and p is the number of the causal factors. h_{1j} and h_{2j} are determined by minimizing the following square error:

$$Q = \sum_{i=1}^n (f_{1i} - \hat{f}_{1i})^2 + \sum_{i=1}^n (f_{2i} - \hat{f}_{2i})^2 \rightarrow \text{minimizing} \quad (5)$$

$$\therefore \frac{\partial Q}{\partial h_{1j'}} = 2 \sum_{i=1}^n z_{ji'} (f_{1i} - \sum_{j=1}^p h_{1j} z_{ji}) = 0 \quad (6)$$

$$\frac{\partial Q}{\partial h_{2j'}} = 2 \sum_{i=1}^n z_{ji'} (f_{2i} - \sum_{j=1}^p h_{2j} z_{ji}) = 0 \quad (7)$$

where n is the number of pixels in the study area. To solve equations (6) and (7), note that the average and variance of z_{ji} are standardized to “0” and “1”, respectively. Also, assuming that there is no correlation between the 1st and the 2nd trigger factor, h_{1j} and h_{2j} can be simply given by

$$h_{1j} = \sum_{j'=1}^p a_{1j'} r^{jj'} \quad (8)$$

$$h_{2j} = \sum_{j'=1}^p a_{2j'} r^{jj'} \quad (9)$$

where $r^{jj'}$ is the element (j, j') of inverse matrix for the correlation matrix between causal factors. Using equations (3) and (4), \hat{f}_{1i} and \hat{f}_{2i} can be calculated

and delineated on the 1st and the 2nd trigger factor influence maps (TFI maps), respectively.

3.4 Success rate

To evaluate the performance of model itself, the “success rates” are calculated as shown in Figure 2. 10% of X axis means the pixels with the highest 10% estimated values on the trigger factor influence are classified as hazardous. Among these pixels, the rate of correctly classifying the pixels of past occurrences of slope failures (i.e., training data sets) are indicated on Y axis as “success rate”. Similarly, for the pixels with the highest {20%, 30%, ..., 90%} of estimated values, the success rates are calculated repeatedly.

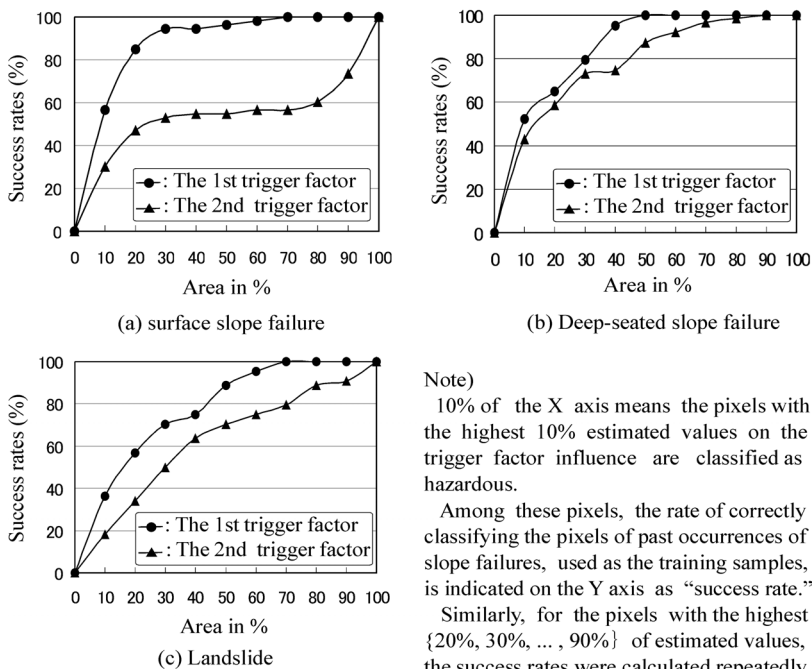


Figure 2: Success rate curves.

If the model is “reasonably good”, then one would expect that the success rate of the corresponding first class with 10% estimated value, which is defined as hazardous area, should be much higher than 10%. From the success rate curves in Figure 2, the following points can be made:

- For the pixels with the highest 30% of estimated value, all the success rates for the 1st trigger factors according to the slope failure types indicate more than 70%, which corroborates the inverse and decompositional operation in estimating the 1st trigger factors is functioning.

- Furthermore, the estimated value of the 1st trigger factor is higher than that of the 2nd trigger factor. Note that the huge earthquake with the antecedent rainfall had caused different types of slope failures in this study area. The 1st trigger factor (i.e., main trigger factor) and the 2nd trigger factor can be considered to have reflected the influence of the earthquake and rainfall, respectively.
- The lowest success rate curve with respect to the 2nd trigger factor comes from the “surface slope failure”, which suggests the surface slope failure is influenced only from the 1st trigger factor (i.e., earthquake in this study case).

These results suggest that the proposed approach on the inverse and compositional analysis is applicable for estimating the unobserved trigger factor with respect to the different types of slope failures. To evaluate the relationship between the 1st and the 2nd trigger factors according to the slope failure types, let us consider the pair-wise comparative strategy of TFI maps based on Hayashi's quantification method of the fourth type, which is well known as one of the multivariate statistical analysis.

3.5 Pair-wise comparison of trigger factor influence maps

As a final product of the inverse- and compositional-analysis of trigger factors, the estimated values of \hat{f}_{1i} and \hat{f}_{2i} are delineated on the “Trigger Factor Influence map (termed TFI map)”. Figure 3 indicates the difference maps (termed “DIF map”) with all combination cases of TFI maps, with respect to “surface slope failure”, “deep-seated slope failure”, and “landslide”, respectively. Note that the legend for these DIF maps lead to the following interpretation on the difference of trigger factor influence:

- **Shade of red:** The estimated values in each pixel of TFI map-A (see Figure 3) are larger than that of TFI map-B;
- **White:** The estimated values in each pixel of TFI map-A are almost equivalent to that of TFI map-B; and
- **Shade of blue:** The estimated values in each pixel of TFI map-A are lower than that of TFI map-B.

As mentioned above, note that, before the huge earthquake of Niigata Chuetsu Earthquake (Oct. 23, 2004), the study area had also caught in local-downpour for several days. So, it is possible to say that the 1st trigger factor (i.e., main trigger factor) and the 2nd trigger factor correspond to “earthquake” and “rainfall”, respectively.

Based on this assumption, we would say that the larger the values in each pixel of TFI map on the 1st trigger factor are, the higher the risk of the given types of slope failures affected by “earthquake (in this study case)” as a main trigger factor. On the other hand, the larger the values in each pixel of TFI map on the 2nd trigger factor are, the higher the risk of the slope failures affected by “rainfall” (in this study case) as the 2nd trigger factor.

Figure 3 indicates that the pixels with higher estimated values on the 1st trigger factor on the landslide and the deep-seated slope failure tend to be

distributed in either the central or the east part in the study area. Such “heuristic information” would be useful not only for assessing the unobserved trigger factors introducing the expert’s opinion, but also for improving the cost-effectiveness in locating the places for setting the field measuring systems (i.e., rain gage, tiltmeter, tensiometer, etc.).

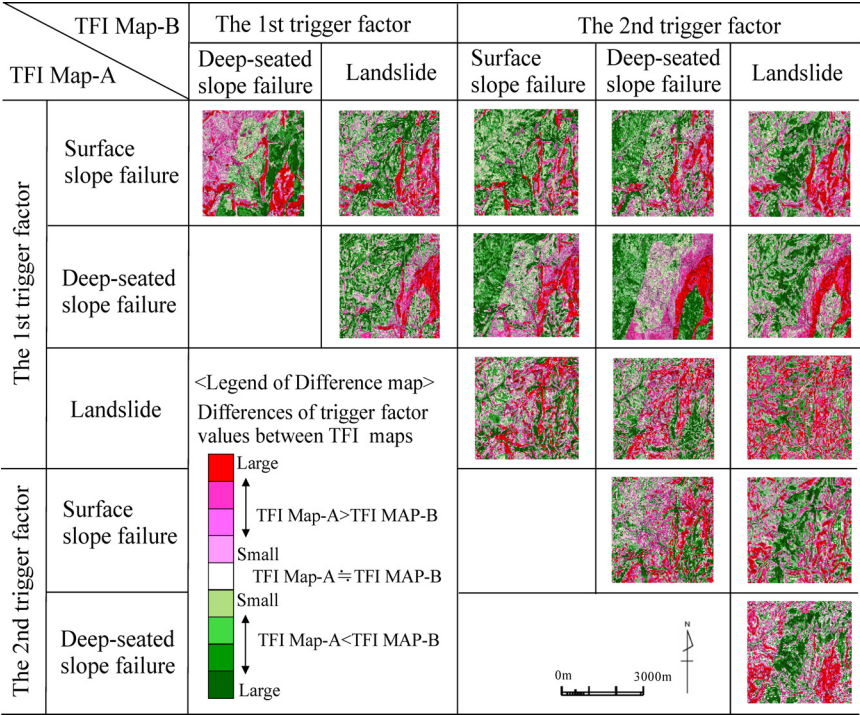


Figure 3: Difference maps (DIF maps) between trigger factor influence (TFI maps) according to “slope failure types”.

3.6 Exploratory analysis of decomposed trigger factors

The pair-wise comparative strategy as shown in Figure 3 are useful for clarifying the spatial differences between the TFI maps, however, there are limitations in analyzing the mutual relationships between decomposed trigger factors according to slope failure types. As a measure, in this study, the Hayashi’s quantification method of the fourth type is introduced, which is a set of related multivariate analysis method (e.g., multi-dimensional scaling) often used in data visualization for exploring similarities or dissimilarities in multivariate data. The dissimilarity measure (DI) used in this study is as follows:

$$DI = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (10)$$

where n is the number of pixels in the study area, x_i and y_i are the estimated values in each pixel of TFI map-A and TFI map-B (see Figure 3), respectively.

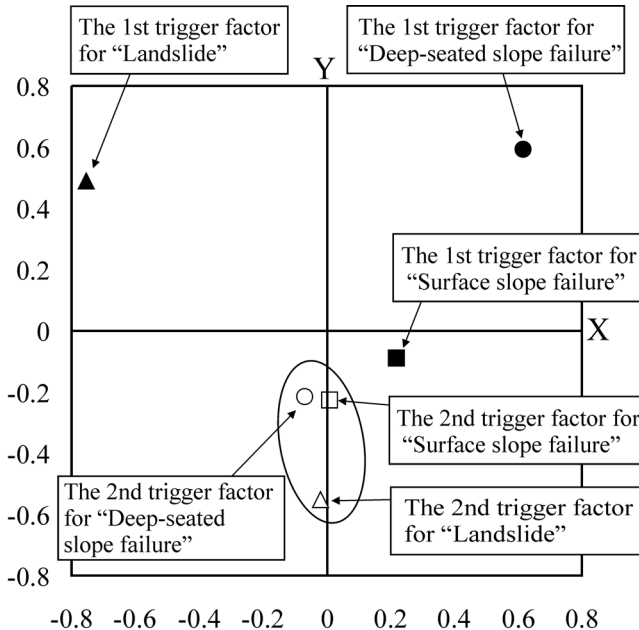


Figure 4: Scatter diagram with respect to decomposed trigger factors.

Based on a matrix consisted of the dissimilarity measures (DI) between all pairs of items, a location of each item is plotted on the N-dimensional space. The items correspond to the 1st and the 2nd trigger factors according to the slope failure types. Figure 4 illustrates a scatter diagram with respect to these items, in which the axis X and Y correspond to the 1st and 2nd eigen value, respectively.

The scattered distribution on three items of the 1st trigger factors corroborates the inverse- and compositional operation according to the slope failure types would be functioning. On the other hand, the items of 2nd trigger factors are closely distributed, which means the models shown in Figure 1 suggest the difficulty of compositional estimation of other trigger factors except the main trigger factor of slope failure.

4 Conclusions

In this contribution, we have discussed inverse- and compositional- analysis of unobserved trigger factors according to the slope failure types, based on the SEM approach. The results of this study are summarized as follows:



- Due to the difficulties and limitations of “pixel-based observations” of slope failure trigger-factors, we strongly point out the necessity for the inverse- and compositional-analysis of “unobserved trigger factors”. As a measure, through the measurement equation (defined in SEM) between the causal factors (i.e., observed variables) and the trigger factors (i.e., unobserved variables), a “Trigger Factor Influence map (termed TFI map)” is produced;
- As a compositional analysis of the trigger factors, the trigger factors are decomposed into the “1st trigger factor” and the “2nd trigger factor”. The Trigger Factor Influence maps (TFI map) with respect to these trigger factors are also produced according to the different types of slope failures; and
- As a final outcome, the differences in these TFI maps are delineated on a “difference (DIF) map” that enables us to analyze the difference of trigger factor influence with respect to slope failure types simultaneously. Furthermore, a pair-wise comparative analysis for these TFI maps has been carried out based on the Hayashi’s quantification method of the fourth type.

In order to further proceed in the practical application of the inverse- and compositional-analysis algorithm, the following observations can be made:

- As a compositional analysis of the trigger factors, we only considered the dual “exogenous variables” as shown in Figure 3. However, either slope failures or landslides can be affected by various kinds of trigger factors. So, the modified path models with three or more exogenous variables should be investigated to improve the identification of models in the SEM approach; and
- As occasion demand of investigators and specialists working on the slope failures, we can readily add the training data sets of other types of slope failures or landslides in the analysis. The analytical procedure proposed in this study can contribute to the slope stability evaluation as one of the systematic approaches.

We should admit that “precise estimation” of unobserved trigger factors with respect to slope failures and landslides is impossible. However, to overcome some of the limitations of conventional approaches for quantitative prediction modeling, the inverse and compositional analysis procedure of unobserved trigger factor presented in this study can contribute to the evaluation of hazardous area affected by various types of slope failures.

Acknowledgements

Funding was partially supported by the Grant-in-Aid for Exploratory Research program (No.19651079) from The Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.



Appendix A

The Normalized Vegetation Index (NVI) used in this study as one of the causal factors is given by

$$NVI = \frac{NIR - RED}{NIR + RED} \quad (11)$$

where *NIR* and *RED* are the Digital Number (DN) values of pixels corresponding to Near-infrared- and Red-band of satellite IKONOS data, respectively.

References

- [1] Chung, C.F. & Fabbri, A.G., Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing*, Vol.65, No.12, pp.1389-1399, 1999.
- [2] Chung C.F., Kojima, H. & Fabbri, A.G., Stability analysis of prediction models for landslide hazard mapping, In: Allison, R.J. (Eds.), *Applied Geomorphology: Theory and Practice*, John Wiley & Sons Ltd, pp.3-19, 2002.
- [3] Crozier, M.J. & Glade, T., Landslide hazard and risk: Issues, Concepts and Approach, In: Glade T., Anderson M. & Crozier, M.J. (Eds.), *Landslide Hazard and Risk*, pp.22-23, 2004.
- [4] Joreskog, K.G. & Lawley, D.N., New methods in maximum likelihood factor analysis, *British J. Math. Statist. Psychol.*, 21, pp.85-96, 1968.
- [5] Hoyle, R.H., Structural equation modeling approach: Basic concepts, and fundamental issues. In: Hoyle, R.H. (Eds.), *Structural Equation Modeling: Concepts, Issues, and Applications*, Sage publications, pp.1-15. 1995.
- [6] Kojima, H., Chung, C.F., Obayashi, S. & Fabbri, A.G., Comparison of strategies in prediction modeling of landslide hazard zonation, *Proceedings of IAMG 1998, International Association for Mathematical Geology Annual Meeting, Ischia, Italy*, pp.218-223, 1998.
- [7] Kojima, H., Chung, C.F. & van Westin, C.J., 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, 2000.
- [8] Kojima, H. & Chung, C.F., Testing on the time-robustness of a landslide prediction model, *Proceedings of IAMG 2001, International Association for Mathematical Geology Annual Meeting, Cancun, Mexico*, pp.41-46, 2001.
- [9] Kojima, H. & Obayashi, S., An inverse analysis of unobserved trigger factor for slope stability evaluation, *Computers & Geosciences*, Vol.32, Issue 8, pp.1069-1078, 2006.

