Uno strumento per la modellazione statistica della suscettibilità da frana realizzato in R

Rossi Mauro (\*), Guzzetti Fausto (\*), Peruccacci Silvia (\*), Reichenbach Paola (\*) & Cardinali Mauro (\*)

Abstract

A statistical modeling tool in R for the evaluation of landslide susceptibility

The assessment of landslide spatial hazard, also known as landslide susceptibility, is a key issue for the assessment of landslide risk. Many studies have attempted to evaluate landslide susceptibility, using a variety of statistical classification techniques. However, only a few attempts have been made to compare the results of models obtained using different classification techniques. In this work, a statistical modeling tool in R for landslide susceptibility assessment is presented. In the tool, four statistical classification techniques are implemented, including: (i) a linear discriminant analysis by least square, (ii) quadratic discriminant analysis, (iii) logistic regression, and (iv) a self-optimizing neural network. Other classification techniques can be added. The proposed tool is scale independent, and it requires in input a configuration file and a data file containing the grouping and the explanatory variables, which can be dummy or numeric. A diagnostic procedure is implemented to detect collinearity in the explanatory variables. A preliminary procedure for the stepwise selection of explanatory variables in linear and quadratic discriminant analysis is used. The performance of each model – measured by its ability to “predict” the original data – is tested. Results are shown using contingency tables and Fourfold plots. Receiver Operating Characteristic (ROC) plots, providing a measure of the forecast’s accuracy, are prepared, with confidence intervals calculated using a bootstrapping technique. We tested the landslide susceptibility statistical modeling tool in the Staffora River basin, which extends for about 275 square kilometers in the northern Italian Apennines (Lombardy Region). For the analysis, we used the partition of the study area into 2243 litho-hydro-morphological units and an initial set of 58 dummy and numeric explanatory thematic variables, including morphological, lithological, structural and land use variables. Results indicate a good accuracy of the obtained susceptibility forecasts, with the best performance obtained using the self-optimizing neural network model.

Key words: Frana, suscettibilità, modello, software.

INTRODUzIONe

La stima della pericolosità spaziale di frana, conosciuta anche come suscettibilità da frana, è un elemento chiave per la stima del rischio da frana. Molti lavori sono stati realizzati per valutare la suscettibilità da frana, utilizzando differenti tecniche statistiche di classificazione, tuttavia solo pochi lavori sono stati realizzati per comparare i risultati ottenuti utilizzando differenti tecniche di classificazione. In questo lavoro viene presentato uno strumento software (SW) realizzato in R (un ambiente di sviluppo specifico per l’analisi statistica con licenza GNU GPL) per la valutazione della suscettibilità da frana. Nello strumento sono state implementate quattro tecniche di classificazione statistica: (i) un’analisi discriminante lineare, (ii) un’analisi discriminante quadratica, (iii) un’analisi di regressione logistica, e (iv) una rete neurale auto ottimizzata. Altre tecniche di classificazione possono essere aggiunte. Lo strumento SW proposto può essere applicato a varie scale e richiede in ingresso un file di configurazione ed un file dati contenente la variabile di raggruppamento e le variabili esplanatorie numeriche o di comodo. Lo strumento include una procedura diagnostica per evidenziare problemi di collinearità tra le variabile esplanatorie. La performance di ogni modello di classificazione statistica, misurata dalla abilità di “predire” i dati originali, è stata valutata. I risultati sono mostrati utilizzando tabelle di contingenza e grafici *fourfold*. Per fornire una misura dell’accuratezza della predizione sono state calcolate le curve ROC (*Receiver Operating Characteristic*) relative ad ogni modello statistico utilizzato, e sono stati calcolati i relativi intervalli di confidenza utilizzando la tecnica statistica di bootstrap. Lo strumento SW è stato testato nell’area studio di Collazzone, nella Regione Umbria in Italia Centrale. Ai fini dell’analisi, l’area è stata suddivisa in 894 unità di versante, ed un set di 51 variabile esplanatorie derivate da informazioni morfologiche, litologiche, strutturali e di uso del suolo. Infine è stata predisposta una procedura per la combinazione dei risultati dei quattro modelli di classificazione statistica.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(\*) Consiglio Nazionale delle Ricerche, Istituto di Ricerca per la Protezione Idrogeologica, via Madonna Alta 126, 06128 Perugia

I risultati hanno indicato una buona accuratezza delle previsioni ottenute dai differenti modelli statistici di suscettibilità. La miglior performance è ottenuta dalla rete neurale auto ottimizzata, mentre gli altri modelli hanno mostrato performance simili. I risultati ottenuti dalla procedura di combinazione delle predizioni dei quattro modelli statistici, dipendono fortemente dal modello a rete neurale auto ottimizzata, mentre escludendo quest’ultimo dalla combinazione, i risultati non dipendono da un unico modello statistico indicando che più modelli partecipano alla predizione.

References

Guzzetti F., Galli M., Reichenbach P., Ardizzone F. & Cardinali M. (2006) *Landslide hazard assessment in the Collazzone area, Umbria, central Italy*. Natural Hazards and Earth System Sciences, **6**, 115-131.

Guzzetti F., Reichenbach P., Ardizzone F., Cardinali M. & Galli M. (2006) *Estimating the quality of landslide susceptibility models*. Geomorphology, **81**, 166-184.