Establishing erosion risk index, exemplified by data from Greek islands

Anna Bartkowiak,
Inst. of Computer Science, University of Wrocław, Poland,
Niki Evelpidou and *Andreas Vassilopoulos*,
Remote Sensing Laboratory, Geology Dept., University of Athens, Greece.

We seek for an erosion risk index taking values from the interval [0,1], with '1' indicating a very high and '0' a very low erosion risk. This might be done by the logistic discriminant function $L(\mathbf{x}, \boldsymbol{\alpha})$, with $\mathbf{x}$ denoting the data vector and $\boldsymbol{\alpha}$ the coefficients appearing in the definition of the logistic discriminant function (LDF). The coefficients $\boldsymbol{\alpha}$ are obtained by an iterative estimation process adjusting in subsequent steps the initially guessed values.

We have applied the method to the Sifnos erosion data (Gournelos et al., 2002). For the computations we have used the NETLAB package working in the MatLab 5 environment. The training data contained samples comprising geographical units exhibiting very high and very low erosion risk. The two samples were completely separated. It was proven by J.A. Anderson (1972) that in such case the ML estimates of the parameters of the LDF do not exist, and given the observed data $\mathbf{X}$,

$$\max_{\alpha} L(X, \boldsymbol{\alpha}) = 1,$$

which is attained when $\boldsymbol{\alpha} \to \infty$.

Despite that fact one may use the estimates obtained from some earlier steps of the iterative estimation procedure.

When analysing the Sifnos data we have stated that:

1. Finite estimates – obtained from an iterative process being at the stage when the process begins to stabilize – have a perfect ability to discriminate the data units with a very high and very low erosion risk.

2. However the LDF with parameters obtained in point 1 above is not able to indicate medium risk units: it assigns them – with a probability ≈ 1 – either to the very high or very low erosion risk class.

3. When repeating the estimation procedure with various starting values of the vector of the parameters we arrived very often to different estimates. This is in agreement with J.A. Anderson's statement (Anderson, 1972), that in the case of separate samples the estimates of the LDF are not unique.

4. Despite that the estimates of the LDF were quite different, the assigned probability of risk (the risk index) resulted in the same values.

5. Various optimization algorithms applied in the iterative estimation procedure yielded values of the parameters differing even by several ranges of magnitude. E.g. the *quasi-Newton* method yielded parameter values which were sometimes exceeding the range [-300, 300]. On the other hand, the *scg (conjugate gradients with regularization* yielded parameter values from the range [-10, 10]. Despite these great differences the assigned probabilities of erosion risk proved to be the same in the first two or 3 decimal places.

References

[1] Anderson J.A., 1972, Diagnosis by logistic discriminant function: further practical problems and results. *Biometrika* **23**, No. 3, 397–404.

[2] Gournelos Th., Evelpidou N., Vassilopoulos A., 2002, Developing an erosion risk map using soft computing. *Natural Hazards* (Kluwer), submitted.