Hydrological modelling to predict the timing of landslide activations. Calibration, validation and sensitivity analysis of $^{GA}SAKe$

P. Iaquinta, O.G. Terranova, S.L. Gariano, R. Greco, G. Iovine

Key words: *landslide, hydrological model, genetic algorithm, calibration, validation, sensitivity analysis.*

INTRODUCTION

In Italy, a recent nationwide investigation has identified more than 1.6 landslides per km^2 (GUZZETTI *et al.*, 2008). Most of the landslides that affect the Country are usually triggered by rainfalls. Forecasting the timing of rainfall-induced landslides is then a fundamental step for Civil Protection purposes. This goal can be obtained by means of either empirical "hydrological" or physically-based, "complete" models.

The hydrological approach (adopted in the present study) tries to identify the mathematical relationship that links the rainfall series to the dates of landslide activation. With respect to shallow landslides, the dynamics of deep-seated slope movements generally shows a more complex relationship with rainfall: activations commonly require greater rainfall amounts, spanned over longer periods (from ca. 30 days to more than a single rainy season).

MODELING APPROACH

SAKe (*Self Adaptive Kernel*) is a hydrological model that can be employed to predict the occurrence of one of the most pernicious effects of rainfall on slope stability: slope movements. The model can be useful for real-time warning purposes. It is self-adaptive and based on the assumption of a linear and steady response, in terms of stability, of the slope to rainfall.

SAKe is based on an empirical approach – "black-box" type (Fig. 1) – inspired from *FLaIR* (SIRANGELO & VERSACE, 1996). By properly tuning the model parameters, a mobility function and a threshold value can be identified that allow to reconstruct the series of known landslide activations of the past by minimizing the occurrence of false alarms. The ranges of

_____

CNR-IRPI – U.O.S. di Cosenza, via Cavour 6 – 87036 Rende (CS) Italia - e-mail: g.iovine@irpi.cnr.it

parameters depend on: (i) the specific hydrological conditions that characterize the triggering events, (ii) the characteristics of the slope and, (iii) the landslide type.

According to classic hydrologic assumptions, usually made in rainfall-runoff modelling, the model output is represented by a "mobility function", *Z(t)*, defined in terms of convolution between the rains and a "filter function" (the *kernel*).
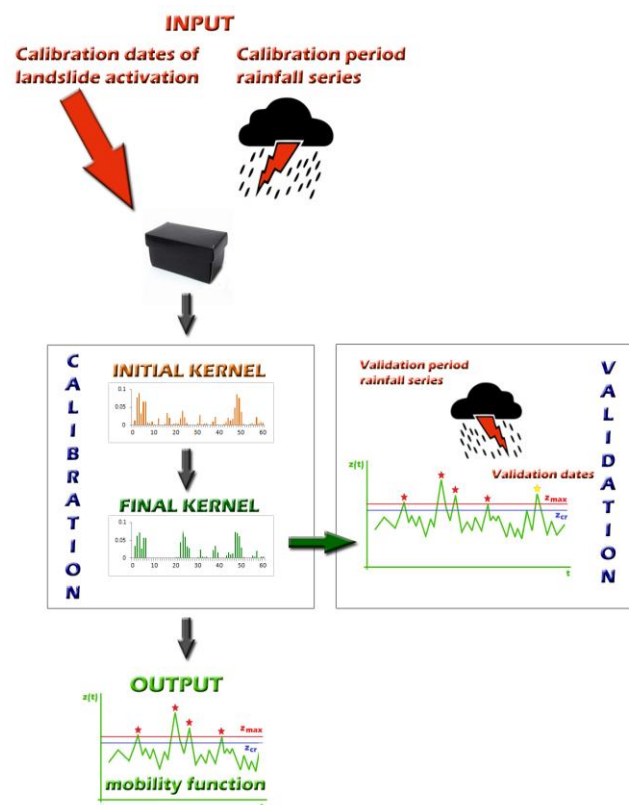

Fig. 1 – Scheme of the model *SAKe*.

The following input data are needed for a given case study:
i) the series of rains;
ii) the set of known dates of landslide activation. More specifically, either the mobilizations of a single phenomenon, or those of a homogeneous set of landslides occurred in a given area can be analyzed.

For calibration purposes, only a subset of available historical dates of activation can be employed; model validation can then be performed against the remaining subset.

As regards model calibration, different automated tools can be utilized, based on either iterative cluster modification – $^{CM}SAKe$ – or Genetic Algorithms (HOLLAND, 1975) – $^{GA}SAKe$ – as recently done for debris flows modelling by IOVINE *et al.* (2005). As a result, a family of optimal, discretised kernels can be obtained from initial standard analytical functions, whose values represent the probability of landslide activation given that a critical threshold is overcome.

The family of optimal kernels – that maximize the adopted fitness function – is selected by means of a calibration technique based on elitist Genetic Algorithms. In this way, the values of model parameters are iteratively changed, aiming at improving the fitness of the tested solutions. A suitable fitness function must first be defined, able to minimise the number of false alarms. In the present study, fitness is defined as follows:

a) the *N* available dates of landslide activation are sorted, in chronological order, in the vector $S = \{S_1, ..., S_i, ..., S_N\}$;

b) the vector of the relative maxima of the mobility function $Z = \{z_1, ..., z_k, ..., z_M\}$, where *M* is the number and *k* is the rank of the relative maxima, is sorted in decreasing order;

c) the vector of the partial fitness values is defined as $\varphi = \{\varphi_1, ..., \varphi_i, ... \varphi_N\}$.

The following cases can occur:

i) $\varphi_i = 1$ if the date of the *i-th* activation, S_i, matches, within a pre-fixed tolerance (Δ), the date of a *k-th* relative maximum of the mobility function, z_k, with $k \leq N$;

ii) $\varphi_i = (k-N+1)^{-1}$ - i.e. it is inversely proportional to the rank of z_k in *Z* – if the date of the *i-th* activation, S_i, matches, within a pre-fixed tolerance, the date of a *k-th* relative maximum of the mobility function, z_k, with $k > N$;

iii) $\varphi_i = 0$ if the date of the *i-th* activation, S_i, does not match any date of the relative maxima in *Z*.

Accordingly, the fitness of a generic kernel is:

$$\Phi = \frac{1}{N} \sum_{i=1}^{N} \varphi_i \ .$$

$^{GA}SAKe$ is based on two levels of optimization. The first optimization process allows to search for the mobility function that gives the maximum fitness function f_n. The second optimization process looks for the maximum difference between the critical threshold, Z_{cr}, and the smallest value of the relative maxima of the mobility function, Z_s. Note that the operational levels can modify the filter function, producing an integral that differs from the unity. Thus, a normalization process is also necessary.

The genetic algorithm is an iterative process based on the following operators (Fig. 2):

- *Selection* of the elements to be copied in the "mating pool" for reproduction purposes (in this study, it is of elitist type);
- *Crossover* of pairs of elements (based on probability, p_c);
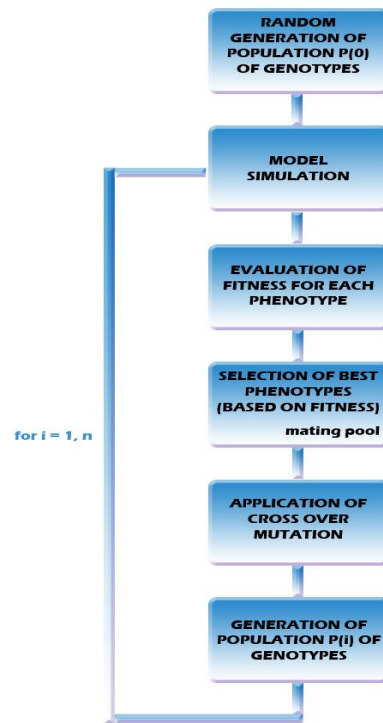- *Mutation* of elements (based on probability, p_m).

Fig. 2 – Steps of the Genetic Algorithm for model calibration.

The mating pool contains *N* elements of the population; these elements eventually originate *N* new mobility functions. Minimum, average and maximum values of the fitness functions for each iteration are evaluated and plotted. New generations are obtained by repeating the steps above.

An example of model optimization is presented, with reference to the Uncino rock slide (maximum width = 200 m, length > 650 m) (Fig. 3), developed in clay and conglomerate (Late Miocene) overlaying gneiss and biotitic schist (Palaeozoic). The landslide repeatedly threatened the northern rim of the San Fili village, in Northern Calabria. For this case study, *N*=7 historical dates of mobilization are available. On such occasions, the railroad connecting Cosenza to the Tyrrhenian coast was damaged or even interrupted. Preliminary results of model calibration, carried out using all the dates of activation, are shown in Fig. 4. Aiming at temporal validation purposes, only a subset of the dates of landslide activation has been used; sensitivity analysis results are discussed.

REFERENCES

GUZZETTI F., PERUCCACCI S., ROSSI M. & STARK C.P. (2008) – *The rainfall intensity-duration control of shallow landslides and debris flow: an update*. Landslides, **5**, 3-17.

HOLLAND J.H. (1975) – *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

IOVINE G., D'AMBROSIO D. & DI GREGORIO S. (2005) – *Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects.* Geomorphology, **66**(1-4), 287-303.

SIRANGELO B. & VERSACE P. (1996) – *A real time forecasting for landslides triggered by rainfall.* Meccanica, **31**, 1–13.

Fig. 3 – Localization and photo of the Uncino rock slide at San Fili (CS), Calabria, Southern Italy (photo courtesy of L. Antronico).
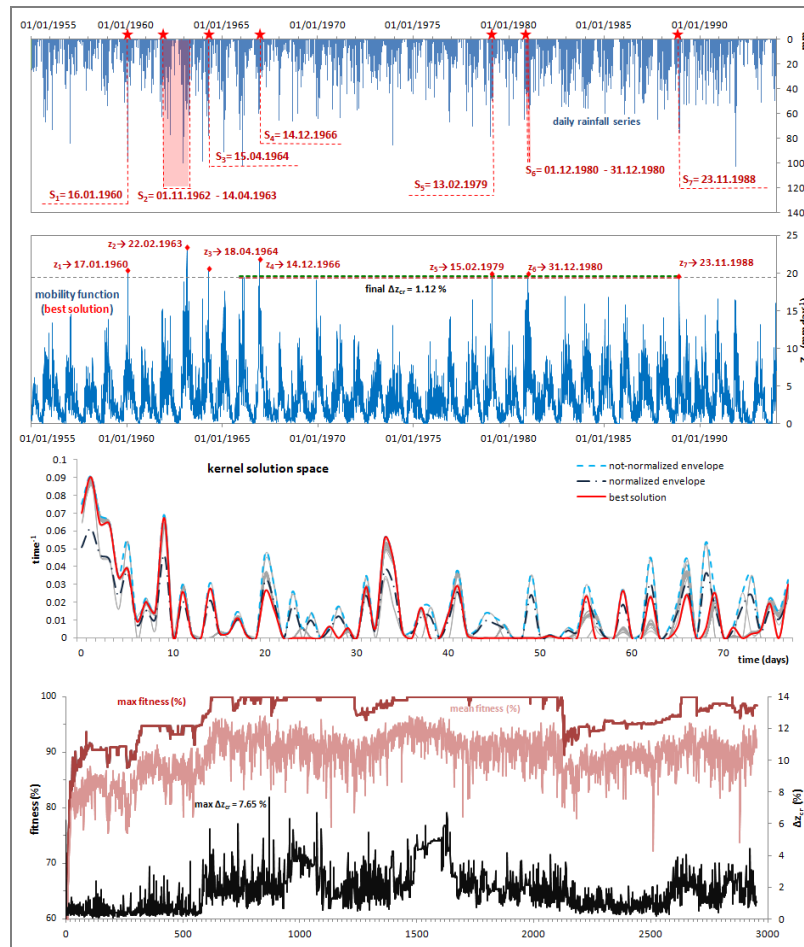

Fig. 4 – Example of output of $^{GA}SAKe$ (elitist approach) for the Uncino case study at San Fili. On top, the red stars mark the known dates of landslide activations ($S_1...S_5$). The series of daily rainfalls (from 01.01.1955 to 31.12.1993), the best mobility function (fitness=100%) and the final value of Δz_{cr} are also shown. In the middle the solution space of the kernel with the best solution (in red, $t_b = 77$), the not-normalized envelope including all the solutions (in dashed, light blue) built considering the maximum value for each element of the kernel function, and the normalized-envelope of the kernel function (in dashed, blue) are shown. At bottom, the evolution of the fitness function, in terms of maximum (red) and average (pink) values per iteration are shown. The evolution of the maximum values of Δz_{cr}, (black line) per iteration, is also shown.