Replies to Anonymous Referee #2

Authors are grateful to the Referee #2 for his/her comments and interesting suggestions. Below, please find our replies to each issue raised, and the proposed changes to the text.

1) First, the adopted fitness definition deserves a more robust discussion and motivation in the manuscript. Moreover, the authors should mention if they performed an investigation on alternative metrics. In fact, according to the literature, the adopted fitness can significantly affect the results of calibration procedures, both in terms of achieved maximum and potential overfitting (which affects validation accuracy).

We agree with the Referee, and propose to extend the discussion on the adopted fitness, also providing an example of computation, by including the following sentences.

Page 1240, after line 4:

"For instance, if two dates of activation are available, the obtained fitness is $\Phi_u = 1 + ½ = 1.5$ if both are well captured by the mobility function (i.e. they correspond to the highest peaks). On the other hand, in case only one of the dates is captured and the remaining one ranks fifth, $\Phi_u = 1 + ⅕ = 1.2$."

Page 1252, after line 10:

"In this study, a 2-steps efficiency criterion was employed: the relative position of the peaks of the mobility function with respect to the dates of landslide activation was first considered, and the fitness computed. Based on the value of ΔZ_{cr}, the obtained solutions were further ranked. Average, synthetic filter functions can then be computed by selecting the best 100 kernels for successive validation purposes. Alternative metrics (cf., among the others, Krause et al., 2005) for the fitness function are being tested. However, due to uncertainties concerning input data (i.e. rainfall and dates of landslide activation), the adoption of sophisticated techniques does not sound very promising. In addition, problems of over-fitting may depend on both data uncertainties and number of parameters. Commonly, kernels characterized by a complex pattern (and then by many parameters) are needed for simulating groundwater dynamics (Pinault et al., 2001). Nevertheless, more complex kernels do not necessarily imply higher predictive uncertainties (Fienen et al., 2010; Long, 2015). Still, the adopted discrete approach allows focusing only on the timing of the peaks of the mobility function, thus somehow relieving the computational effort. Due to the cited uncertainties in input data, a "temporal window" was in fact employed to help matching dates of activation with the peaks of the mobility function. Further attempts of defining the fitness function by different metrics, and the analysis of its effects on calibration and validation, are being considered against another case study (San Benedetto Ullano, in Calabria, Southern Italy), whose mobility phases have been recently monitored by the same authors (Iovine et al., 2010; Capparelli et al., 2012)."

2) As for the family of "optimal kernels", my advice is to better explain such a concept.
In fact, a single-objective GA typically provides a single optimal (or "best") individual, thus the mentioned concept of "family of optimal kernels" provided by the procedure can confuse the reader.

The same applies to the concept of "average kernels" introduced in section 5. Currently, the reader has to struggle to understand why the authors average kernels when they have a "best kernel", or what is the origin of the 100 averaged kernels.

We agree with the Referee, but up to date - in the performed experiments related to real case studies - we never got a single "best solution" much better than the rest. Commonly, a set of optimal solutions with rather similar fitness were instead obtained. That is why we chose to average a number of individuals to synthetize the kernel to be used for validation. Accordingly, we propose to extend the discussion both on the family of optimal kernels and on the average kernel, by including the following period.

Page 1241, after line 6:

"Differently from what usually experienced in rainfall-runoff models, ^{GA}*SAKe* therefore provide multiple equivalent solutions - i.e. a number of optimal kernels with same fitness, Φ_u, despite different shapes. This may depend on the limited number of available dates of activations, and on other noises in input data (e.g. rain gauges located too far from the site of landslide activation; inaccurate information on dates of activation or on the phenomenon). The adoption of synthetic kernels – e.g. obtained by averaging a suitable set of optimal kernels – allows to synthetize the family of results for successive practical applications: in this work, the best 100 kernels obtained for each case study were in fact utilized to synthetize average kernels to be employed for validation purposes."

3) Also, in section 6, it is not clear why the progressive calibration procedure was also labelled as "self-adaptive".

The term "self-adaptive" was actually used to stress the ability of the model to react to input changes, such as new dates of landslide activations and more prolonged rainfall series. This feature represents a major advantage of the model. In particular, the self-adaptive procedure of progressive calibration was performed by considering an increasing number of dates of activation to mimic the adoption of the model in a landslide warning system. Obtained results underlined how ^{GA}SAKe easily self-adapt to external changes by optimizing its performances with increasing fitness values. To better explain the above issues, we would propose to modify the manuscript as follows.

Page 1248, line 5, replace "simulate the occurrence of known landslide activations" with "to react and self-adapt to input changes, like new dates of landslide activation,".

Page 1248, line 7, replace "In particular," with "To simulate the adoption of ^{GA}*SAKe* in a landslide warning system,".

Page 1251, line 5, at the end of the sentence, please add: "Accordingly, the results of the progressive procedure underlined how ^{GA}*SAKe* can easily self-adapt to external changes by optimizing its performances, providing increasing fitness values".

References:

Capparelli, G., Iaquinta, P., Iovine, G., Terranova, O.G., and Versace P.: Modelling the rainfall-induced mobilization of a large slope movement in northern Calabria, Nat. Hazards, 61(1), 247–256, 2012.

Fienen, M.N., Doherty, J.E., Hunt, R.J., and Reeves, H.W.: Using prediction uncertainty analysis to design hydrologic monitoring networks: example applications from the Great Lakes water availability pilot project, US Geol. Surv., Reston Virginia, Scientific Investigations Report 2010–5159, 44 pp., 2010.

Iovine, G., Lollino, P., Gariano, S.L., and Terranova, O.G.: Coupling limit equilibrium analyses and real-time monitoring to refine a landslide surveillance system in Calabria (southern Italy), Nat. Hazards Earth Syst. Sci., 10, 2341–2354, 2010.

Krause, P., Boyle D.P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Adv Geosciences, 5, 89–97, 2005

Long, A.J.: RRAWFLOW: Rainfall-Response Aquifer and Watershed Flow Model (v1.15), Geosci. Model Dev., 8, 865–880, 2015.

Pinault, J.-L., Plagnes, V., and Aquilina, L.: Inverse modeling of the hydrological and the hydrochemical behavior of hydrosystems: Characterization of karst system functioning, Water Resour Res, 37(8), 2191–2204, 2001.