Nat. Hazards Earth Syst. Sci. Discuss.,
doi:10.5194/nhess-2017-24-AC2, 2017

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "Performance evaluation of the national Norwegian early warning system for weather induced landslides" *by* Luca Piciullo et al.

Luca Piciullo et al.

lucapiciullo@gmail.com

Received and published: 10 March 2017

The Authors thank Reviewer #2 for his/her interest in our manuscript. We carefully revised the manuscript according to the many valuable comments and recommendations provided.

1) R: We thank Reviewer #2 for this comment that gives us the possibility to better explain the reasons of choosing this case study and dataset. The "Vestlandet" region was chosen as it is one of the areas most prone to landslides in Norway. Moreover, for this area the landslide database is more reliable and complete than in the rest of Norway. As the second most populated area of the Nation, more information on landslides are available. The Norwegian national landslide early warning system (LEWS)

is a realtively new system that became operational in 2013. The analyses presented in this manuscript started in 2015 and only data for 2013-2014 were available at that time. A large work of collection and checking of landslide information from different sources (NVE, rails and roads Authority, other databases, media) was carried out, with the aim of avoiding repetitions and providing a reliable dataset. However, to answer this comment, we checked the number of warning issued in 2015-2016 in Vestlandet. There were only few days with Orange warnings and no one with Red warnings. The table below shows the number of warnings and the warning levels issued in Vestlandet in the period 2013-2016.

Warning levels yellow orange red 2013 21 0 0 2014 34 5 0 2015 20 2 0 2016 21 0 0

As shown, the red level would still be missing even if we considered the period 2015-2016. According to the meaning of warning levels presented at http://www.varsom.no/en, the red level defines "an extreme situation that occurs very rarely, it requires immediate attention and may cause severe damages within a large extent of the warning area". Concluding, incorporating these data would not change the results of the performance analysis and would not add anything significant towards the main aim of the paper, i.e. proposing an extension of the EDuMaP method for the performance evaluation of LEWSs issuing warnings over zones characterised by a variable size. Finally the title of the paper has been modified in "Adapting the EDuMaP method to test the performance of the Norwegian early warning system for weather-induced landslides", for better clarifying the aim of the paper and avoiding confusion in the reader.

2)R: The EDuMaP method comprises three successive steps: identification and analysis of landslide and warning Events (E), from available databases; definition and computation of a Duration Matrix (DuMa), and evaluation of the early warning model Performance (P) by means of performance criteria and indicators. The parameters needed to carry on the events analysis (E) are ten. Among them, there is the spatial discretization adopted for warnings, $\Delta A(k)$, which describes if the warning zone is fixed or variable.

For instance, the LEWS employed in Rio de Janeiro considers fixed warning zones, on the contrary the system adopted in Norway uses variable warning zones. In earlier studies, the EDuMaP method has been applied to analyse the performance of regional landslide EWSs adopting a fixed spatial discretization for warnings. When the landslide EWS employs variable warning zones, this characteristic significantly influences the first two steps of the EDuMaP method. Section 3.3 was rewritten for increasing the comprehensibility of the methodology. It explains how to define landslide events (LEs) and warning events (WEs) and how to compute the duration matrix in case of variable warning zones. The landslides are grouped in LEs as a function of the warning zone in which they occur. A warning zone can be seen as an area alerted with the same level of warning (i.e., green, yellow, orange, red). The EDuMaP method evaluates the duration of each level of warning (i.e., green, yellow, orange, red) and the class of landslide event (i.e: the number of landslides) occurred over the time in a warning zone. In the EDuMaP method, a warning can be considered successful as a function of both the level of warning issued and the number of landslide occurred in the zone alerted. The number of landslides expected for each warning level often is defined by the LEWS managers, otherwise can be evaluated considering a landslide density criterion, Lden(k).

3)R: Among LEWSs at a regional scale, the performance of the system is evaluated principally by computing the joint frequency distribution of landslides and alerts. Empirical evaluations are often carried out by simply analyzing the time frames during which significant high-consequence landslides occurred in the test area (Keefer et al., 1987; Aleotti, 2004; Cheung et al., 2006; Baum and Godt, 2010; Capparelli and Tiranti, 2010). Alternatively, the performance evaluation is based on 2 by 2 contingency tables computed for the joint frequency distribution of landslides and alerts, both considered as dichotomous variables (Yu et al., 2003; Cheung et al., 2006; Godt et al., 2006; Restrepo et al., 2008; Tiranti and Rabuffetti, 2010; Kirschbaum et al., 2012; Martelloni et al., 2012; Peres and Cancelliere, 2012; Staley et al., 2013; Lagomarsino et al., 2013, 2015; Greco et al., 2013; Segoni et al., 2014; Gariano et al., 2015; Stähli et al., 2015).

The performance of the systems operational in Norway and Rio de Janeiro was analysed applying the EDuMaP method considering: the possible occurrence of multiple landslides in the warning zone; the duration of the warnings in relation to the time of occurrence of the landslides; the level of the issued warning in relation to the landslide spatial density in the warning zone; the relative importance system managers attribute to different types of errors. In general it's difficult to compare the performance of LEWSs, especially if it has been evaluated with different methods. The values to evaluate the statistical indicators derive from different reasoning, for example, on what is considered as false, missed or correct alerts. Substantial differences may be observed among a 2x2 contingency table and a nxm duration matrix. The latter compares the n levels of warning in relation to the m classes of landslide events. The EDuMaP method evaluates the performance of a LEWS considering the number of warning levels and the classes of landslide events, thus, warnings and landslides are not considered as dichotomous variables as it is for contingency tables. A benchmark could be defined, but it would require a separate analysis and a comparison of a relatively high number of different LEWSs evaluated with the EDuMaP method. Because system managers of LEWSs may attribute a relative importance to different aspects (i.e.: missed alerts, false alerts, purple errors, correct alerts, greens, the level of warning issued, classes of landslide, etc..). As a consequence, different performance criteria are needed to be chosen in order to consider the system managers choices and to carry on the performance analysis. Currently the authors are still working on a comparison among the performance evaluation of different LEWSs in order to provide "functioning standards".

4) R: We thank the Reviewer for giving us the possibility to clarify some important concepts of the duration matrix, that erroneously we have neglected to mention in the manuscript. The component d11 ("no warning issued – no event observed") of the matrix expresses the number of hours when no warnings are issued and no landslides occur. Both criteria (1 and 2) purposefully neglect element d11, whose value is typically orders of magnitude higher than the values of the other elements of the matrix because it also includes all hours without rainfall, for which a LEWS is not designed to deal with,

specifically. Thus, d11 component is neglected in our analysis in order to avoid an overestimation of the performance and to allow a more useful relative assessment of the information located in the remaining part of the duration matrix. So, in figure 9 a, b (currently figure 6 a,b) the d11 component of the duration matrix is neglected. According to the suggestion provided we have modified the description for figure 9. Here are the new sentences: "In terms of criterion 2, Case B shows slightly higher values of Green (14%) than Case A (7%). This means that considering the reduced set of landslides (Set b), there is a slightly better correspondence between the LE classes and corresponding warning levels issued". However, it doesn't mean a better performance for Case B, because figure 9 (currently figure 6) shows only preliminary results. With the EDuMaP method the performance is evaluated through the evaluation of statistical indicators (fig. 12 and tab. 11- currently fig. 9 and tab. 9)

5) R: The dataset B is composed by a catalogue of landslides with a known typology. On the contrary the dataset A includes also landslides in soil of unknown typology that can be, anyway, classified as rainfall-induced landslides. For this reason we decided to keep both the datasets. Finally, the results coming from the two datasets were compared to evaluate the differences in terms of performance indicators arising from uncertainties in the landslide database.

6) R: According to the suggestion all the references in Norwegian have been cancelled because considered not useful to improve the comprehension of the manuscript. A comparison between the EDuMaP method and other methodologies for the evaluation of the performance lies outside the scope of the paper, which is focused on the definition of an original approach, to be implemented in the EDuMaP method, for the computation of the elements of the duration matrix in the case of early warning models issuing alerts on variable warning zones. Many references to different approaches for the performance evaluation were presented in Calvello and Piciullo 2016, and Piciullo et al., 2016. Anyway, following the comment, the revised manuscript includes references to other authors on performance evaluation of warning systems.

Finally, according to minor comments: the abstract has been shortened, section 3.3 has been rewritten to better explain the method, some figures and tables cancelled.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2017-24, 2017.