

## Standards for the performance assessment of territorial landslide early warning systems

**Abstract** Landslide early warning systems (LEWS) can be categorized into two groups: territorial and local systems. Territorial landslide early warning systems (Te-LEWS) deal with the occurrence of several landslides in wide areas: at municipal/regional/national scale. The aim of such systems is to forecast the increased probability of landslide occurrence in a given warning zone. The performance evaluation of such systems is often overlooked, and a standardized procedure is still missing. This paper describes a new Excel user-friendly tool for the application of the EDuMaP method, originally proposed by (Calvello and Piciullo 2016). A description of indicators used for the performance evaluation of different Te-LEWS is provided, and the most useful ones have been selected and implemented into the tool. The EDuMaP tool has been used for the performance evaluation of the “SMART” warning model operating in Piemonte region, Italy. The analysis highlights the warning zones with the highest performance and the ones that need threshold refinement. A comparison of the performance of the SMART model with other models operating in different Te-LEWS has also been carried out, highlighting critical issues and positive aspects. Lastly, the SMART performance has been evaluated with both the EDuMaP and a standard  $2 \times 2$  contingency table for comparison purposes. The result highlights that the latter approach can lead to an imprecise and not detailed assessment of the warning model, because it cannot differentiate among the levels of warning and the variable number of landslides that may occur in a time interval.

**Keywords** Landslide early warning systems · Rainfall thresholds · Performance · Duration matrix · Statistical indicators · Landslides · Territorial

### Introduction

Operational landslide early warning systems (LEWS) aim at reducing the loss-of-life probability by inviting stakeholders (e.g., civil protection agents, administrators, lay people) to act properly in populated areas characterized, at specific times, by an intolerable level of landslide hazard (Calvello 2017). LEWS widely differ depending on the type of landslide they address and the scale of operation, which is related to the size of the area covered by the system. Two categories of LEWS can be defined on the basis of the scale of operation (e.g., Bazin 2012): (i) local LEWS (Lo-LEWS), dealing with a single landslide system at slope scale; and (ii) territorial LEWS (Te-LEWS), dealing with multiple landslides at regional scale. The adjective “territorial” is herein preferred over the most commonly used adjective “regional” to provide a more general name for all the LEWS employed over a wide area, e.g., a nation, a region, a municipal territory, a river catchment (Piciullo et al. 2018).

In the literature, there are several proposals schematizing the structure of LEWS and highlighting the importance of the relations among different system components, as well as the role played by

the actors involved in designing and managing these systems. Di Biagio and Kjekstad (2007) employ a flow chart to outline four main sequential activities for such systems: monitoring, analysis and forecasting, warning, and response. Intrieri et al. (2013), elaborating on the well-known four-elements scheme of people-centered early warning systems proposed by the UNISDR (2006), describe LEWS as the balanced combination of four different components: design, monitoring, forecasting, and education. Calvello et al. (2015) state that the objectives of LEWS should be defined by considering the scale of analysis and the type of landslides, and they represent the process of designing and managing LEWS by a wheel with four concentric rings identifying the following: the necessary skills, the activities to be performed, the means to be used, and the basic elements of the system. Calvello (2017) illustrates the components of early warning systems for weather-induced landslides within a scheme based on a clear distinction among landslide models, warning models, and warning systems, wherein a landslide model is one of the components of a warning model and the latter is one of the components of a warning system. All these schematizations highlight the fact that all the identified system components are essential for LEWS to be effective, as the failure of any component means the failure of the whole system. Indeed, early warning systems are only as good as their weakest link as they can, and frequently do fail for a number of reasons (Maskrey 1997).

The Hyogo Framework for Action “priority for action 2” (i.e., identify, assess and monitor disaster risks, and enhance early warning) identifies as key activity the establishment of institutional capacities to ensure that early warning systems are subject to regular system testing and performance assessments (HFA 2005). The scientific literature reports many studies on LEWS, either addressing a single landslide at slope scale (Lo-LEWS, e.g., Pecoraro et al. 2019 and references therein) or concurrent phenomena over wide areas at municipal/regional/national scale (Te-LEWS, e.g., Piciullo et al. 2018 and references therein), yet the performance evaluation of the warning models employed within LEWS is often overlooked by system managers and researchers. Particularly for Te-LEWS, model performance is often assessed neglecting some important aspects peculiar to these systems, among which the occurrence of concurrent multiple landslides in the warning zone; the issued warning level in relation to the landslide spatial density in the warning zone; and the relative importance attributed, by system managers, to different types of errors (Calvello and Piciullo 2016). Indeed, in the literature, only few systems are described whose performance has been thoroughly assessed (Cheung et al. 2006; Restrepo et al. 2008; Martelloni et al. 2012; Lagomarsino et al. 2013; Calvello and Piciullo 2016; Piciullo et al. 2017a; Piciullo et al. 2017b).

A selection of indicators, available in the literature, to quantify the performance of both rainfall thresholds and EWS is presented in the following section. The paper aims at identifying the most

useful ones for the performance evaluation of Te-LEWS. Moreover, the paper describes the results for the performance evaluation of the warning model adopted by the Te-LEWS operating in Piemonte, Italy (Tiranti and Rabuffetti 2010). The evaluation is based on the application of the EDuMaP method (Calvello and Piciullo 2016), considering landslides and warnings recorded in the different warning zones of the system from 2008 to 2016. The results of the performance assessment, carried out with the EDuMaP method, have been compared with the ones obtained using a  $2 \times 2$  contingency table.

### Performance assessment of territorial landslide early warning systems

#### Rainfall threshold validation and performance of Te-LEWS

In the last decades, rainfall thresholds for landslide occurrence were thoroughly investigated, producing several different test cases and relevant technical and scientific advances. A recent literature review on rainfall thresholds (Segoni et al. 2018a), on the scientific articles published in journals indexed in SCOPUS or ISI Web of knowledge in the period 2008–2016, highlights significant advances as well as critical issues about this topic. The main concern is the validation process, which is seldom carried out. Regrettably, only 38 papers out of 115 (33%) presented a correct validation analysis performed with an independent dataset, while 31 thresholds (27.0%) were validated using the same dataset used for calibration and 46 thresholds (40.0%) were published without any evaluation of their predictive capability. About 34% and 17% of the investigated rainfall thresholds are employed for early warning purposes in LEWS, respectively in prototype and operational systems; for 58% of such thresholds, a performance analysis has been carried out. The most adopted validation criterion is the compilation of a contingency matrix and the evaluation of performance indicators derived from that matrix. The contingency matrix is almost always computed as a  $2 \times 2$  matrix, considering landslide and warning as dichotomous variables, neglecting both the different warning levels that can be issued by a LEWS and the multiple landslides that can occur simultaneously. Piciullo et al. (2018) and Pecoraro et al. (2019) show, respectively, that the majority of Te- and Lo-LEWS employ more than two warning levels (usually 4). In this circumstance, a performance analysis considering a  $2 \times 2$  contingency table can lead to incomplete or wrong performance evaluations. To solve this issue, Calvello and Piciullo (2016) proposed a method, called EDuMaP, for the performance analysis of a warning model, based on the computation of a duration matrix, to be used in place of a contingency matrix. Performance criteria and different performance indicators are applied to the computed duration matrix to evaluate the performance of the warning model.

#### Indicators used for rainfall threshold validation and performance evaluation of Te-LEWS

Piciullo et al. (2018) and Segoni et al. (2018a) show that the contingency matrix is the most used method for both rainfall threshold validation and performance evaluation of Te-LEWS. The performance indicators that can be derived from a contingency matrix are many. Table 1 summarizes the indicators employed by at least two authors for either rainfall threshold validation or performance evaluation of Te-LEWS. Indicators employing the same formula are grouped together, providing the different names

used in literature, and the related references. The formulas have been homogenized, for comparative purposes, adopting the following terms to define the four elements of the contingency table: correct alert, CA; true negative, TN; false alert, FA; missed alert, MA.

The efficiency index<sub>(1)</sub>, also called critical success index or threat score, differs from the efficiency index<sub>(2)</sub> because it does not consider TN. Therefore, the values of these two indicators can be considerably different. The same is true for the odds ratio, which evaluates the ratio between positive and negative predictions and can be computed with or without considering TN (respectively identified as odds ratio<sub>(2)</sub> and odds ratio<sub>(1)</sub> in Table 1). In the validation process of rainfall thresholds, as well as in the performance assessment of LEWS, the number of TN, which represent the absence of both warnings or landslides, is typically orders of magnitude higher than other terms of the contingency table. Thus, considering TN in the performance indicator can lead to an overestimation of the (computed) efficiency of the system. For this reason, the efficiency index and the odds ratio computed without TN are to be preferred in validation and performance analyses.

The efficiency index<sub>(1)</sub> and the odds ratio<sub>(1)</sub> are related by the following expression:  $1/EI - 1/OR = 1$ , so it could be sufficient to select one of them in performance analyses. Furthermore, the hit rate and the missed alert rate are complementary, as well as positive predictive power and false missed alert rate. Among the indicators used to quantify errors, it is worth mentioning the missed and false alerts balance, which defines the percentage of MA among the erroneous predictions and thus it ranges between 0 and 1. From the perspective of reducing the number of MA, which may cause higher negative consequences compared with FA, missed and false alerts balance values should be as low as possible.

The considerations above have led to the selection of 2 main performance indicators, for the alert classification criterion (criterion A in the following): (i) efficiency index<sub>(1)</sub> and (ii) missed and false alerts balance.

#### A tool for the application of the EDuMaP method

The EDuMaP is a method for the performance analysis of a warning model, based on the computation of a duration matrix, to be used in place of a contingency matrix. Performance criteria and different performance indicators are applied to the computed duration matrix to evaluate the performance of the warning model. The model is fully described in Calvello and Piciullo (2016). An Excel tool for the application of the EDuMaP method has been recently programmed in Visual Basic for Applications. The Excel spreadsheet comprises an initial “home” page and some other tabs. The left side of the home page is set to define the input data for the performance analysis and to run different subroutines, following the main structure of the EDuMaP method (Fig. 1). The right side of the home page presents the chosen performance criteria, the computed duration matrix, and the final results of the analysis in terms of performance indicators. The values of the 10 input parameters (i.e., warning levels, landslide density criterion, lead time, landslide typology, minimum interval between landslide events, over time, area of analysis, spatial discretization adopted for warnings, time frame of analysis, temporal discretization of analysis), as well as the landslide and warning datasets for the period of analysis, are defined in separate tabs. Once the datasets

**Table 1** Performance indicators adopted by different authors engaged in rainfall threshold validation or LEWS performance assessment, grouped by formula

Name	Formula	Employed for rainfall threshold validation	Employed for LEWS performance
Critical success index <sup>1</sup> , efficiency index_(1) <sup>2</sup> , threat score <sup>3</sup>	$CA/(CA+MA+FA)$	Tiranti and Rabuffetti 2010 <sup>3</sup> ; Staley et al. 2013 <sup>3</sup> ;	Cheung et al. 2006 <sup>1</sup> ; Piciullo et al. 2017b <sup>2</sup> ; Calvello and Piciullo 2016 <sup>3</sup>
efficiency index_(2)	$(CA+TN)/(TN+CA+MA+FA)$	Lagomarsino et al. 2015	Martelloni et al. 2012; Lagomarsino et al. 2013; Calvello and Piciullo 2016;
Probability of detection <sup>1</sup> , sensitivity <sup>2</sup> , hit rate <sup>3</sup> , true positive rate <sup>4</sup>	$CA/(CA+MA)$	Tiranti and Rabuffetti 2010 <sup>3</sup> ; Staley et al. 2013 <sup>4</sup> ; Gariano et al. 2015 <sup>1</sup>	Cheung et al. 2006 <sup>1</sup> ; Restrepo et al. 2008 <sup>1</sup> ; Martelloni et al. 2012 <sup>2</sup> ; Lagomarsino et al. 2013 <sup>2</sup> ; Calvello and Piciullo 2016 <sup>3</sup> ; Piciullo et al. 2017b <sup>3</sup>
Positive predictive power	$CA/(CA+FA)$		Martelloni et al. 2012; Calvello and Piciullo 2016; Piciullo et al. 2017b
Odds ratio_(1)	$CA/(MA+FA)$		Piciullo et al. 2017b
Odds ratio_(2)	$(CA+TN)/(MA+FA)$		Martelloni et al. 2012; Calvello and Piciullo 2016
False positive rate <sup>1</sup> , probability of false detection <sup>2</sup>	$FA/(FA+TN)$	Staley et al. 2013 <sup>1</sup> ; Gariano et al. 2015 <sup>2</sup>	Martelloni et al. 2012 <sup>1</sup> ; Lagomarsino et al. 2013 <sup>1</sup>
False negative rate <sup>1</sup> , missed alert rate <sup>2</sup>	$MA/(MA+CA)$		Martelloni et al. 2012 <sup>1</sup> ; Lagomarsino et al. 2013 <sup>1</sup> ; Calvello and Piciullo 2016 <sup>2</sup> ; Piciullo et al. 2017b <sup>2</sup>
False alert rate <sup>1</sup> , probability of false alarms <sup>2</sup>	$FA/(FA+CA)$	Tiranti and Rabuffetti 2010 <sup>1</sup> ; Gariano et al. 2015 <sup>2</sup>	Cheung et al. 2006 <sup>1</sup> ; Restrepo et al. 2008 <sup>1</sup> ; Calvello and Piciullo 2016 <sup>1</sup> ; Piciullo et al. 2017b <sup>1</sup>
Missed and false alert balance	$MA/(FA+MA)$		Piciullo et al. 2017b

are inserted, it is possible to generate landslide and warning events, i.e., to group landslides and warnings on the basis of the values of the input parameters. After that, the element value of the duration matrix,  $d_{ij}$ , can be computed. Then, two sets of performance criteria need to be defined. Figure 1 reports the two performance criteria that will be used for the analysis presented in this paper. They are named, respectively, alert classification (criterion A) and grade of correctness (criterion B). The first criterion (A) employs an alert classification that groups together some elements of the matrix to identify: correct alerts, *CA*; false alerts, *FA*; missed alerts, *MA*; true negatives, *TN*. The second criterion (B) assigns a color code to the elements of the matrix in relation to their grade of correctness, herein classified in four classes as follows: green, *Gre*, for the elements which are assumed to be representative of the best model response; yellow, *Yel*, for elements representative of minor model errors; red, *Red*, for elements representative of significant model errors; purple, *Pur*, for elements representative of the worst model errors. Once the two performance criteria are defined, the performance indicators can be computed, and the results are shown in both tabular and graphical formats.

The performance indicators employed in the Excel tool and, adopted in this paper, are a revised and reduced version of what has been proposed by Calvello and Piciullo (2016) and Piciullo et al. (2017a). They refer both to alert classification criterion (A), grade of correctness criterion (B), and a mix of the two (A+B). The indicators adopted herein for criterion A have been discussed in the “Indicators used for rainfall threshold validation and performance evaluation of Te-LEWS” section. For the reasons described in that section, the computation of all the performance indicators does not include the element  $d_{ii}$ , which represent the amount of

time associated with the simultaneous absence of warning and landslide events. Table 2 shows the indicators used, their formulas, and the reference to the performance criterion considered.

### Case study

#### Piemonte “SMART” LEWS

Arpa Piemonte (the Regional Agency of Environmental Protection of Piemonte) developed its first shallow landslides early warning system in 2008 (Tiranti and Rabuffetti 2010). The LEWS, called SMART (Shallow landslides Movement Announced through Rainfall Thresholds), are based on an empirical intensity-duration (ID) model where the thresholds have been identified by back analysis, considering the relationship between historical widespread shallow landslide events that occurred in 1990 and 2002 and rainfall data recorded by the regional rain gauge network (more than 400 rain gauges distributed over an area of 25,873 km<sup>2</sup>).

SMART operates both in real-time and in forecasting mode, coherently with the setup of the Regional Warning System for Geohydrological and Flooding Risk in Piemonte (RWS) (Rabuffetti et al. 2003; Rabuffetti and Barbero 2005). SMART operates in two macro-zones of the Piemonte region, called “homogeneous zones”: Alps and Apennine; hilly environment including Tertiary Piedmont Basin (TPB) and Torino Hill (Fig. 2a). The two zones are characterized by two different sets of thresholds (Eqs. 1 and 2), developed considering the rain gauge locations within the two zones.

$$\text{Zone 2 } I = 25 \cdot d^{-0.45} \quad (1)$$

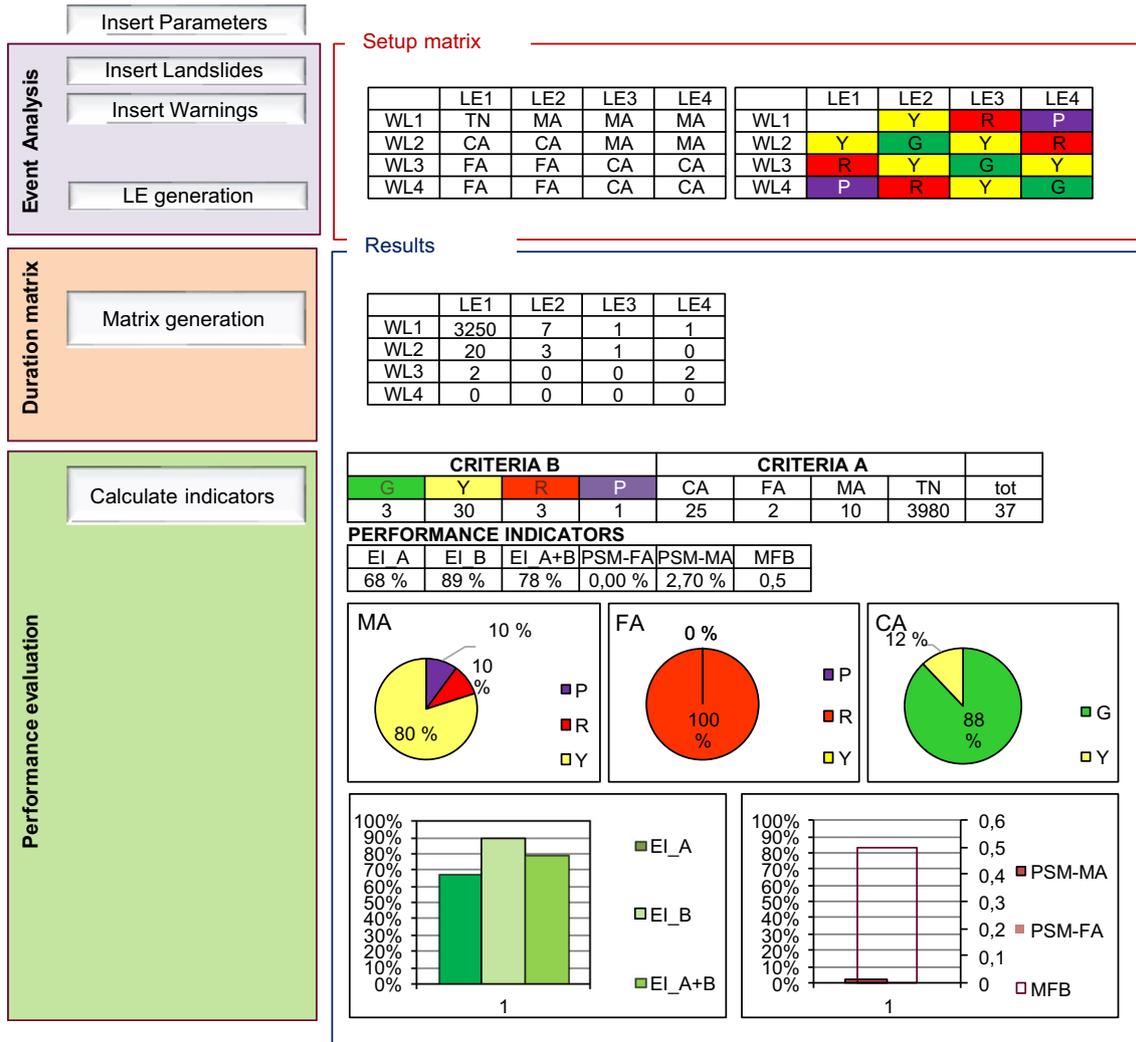


Fig. 1 Excel tool interface for the application of the EDuMaP method

$$\text{Zone 2 } I = 40 \cdot d^{-0.65} \quad (2)$$

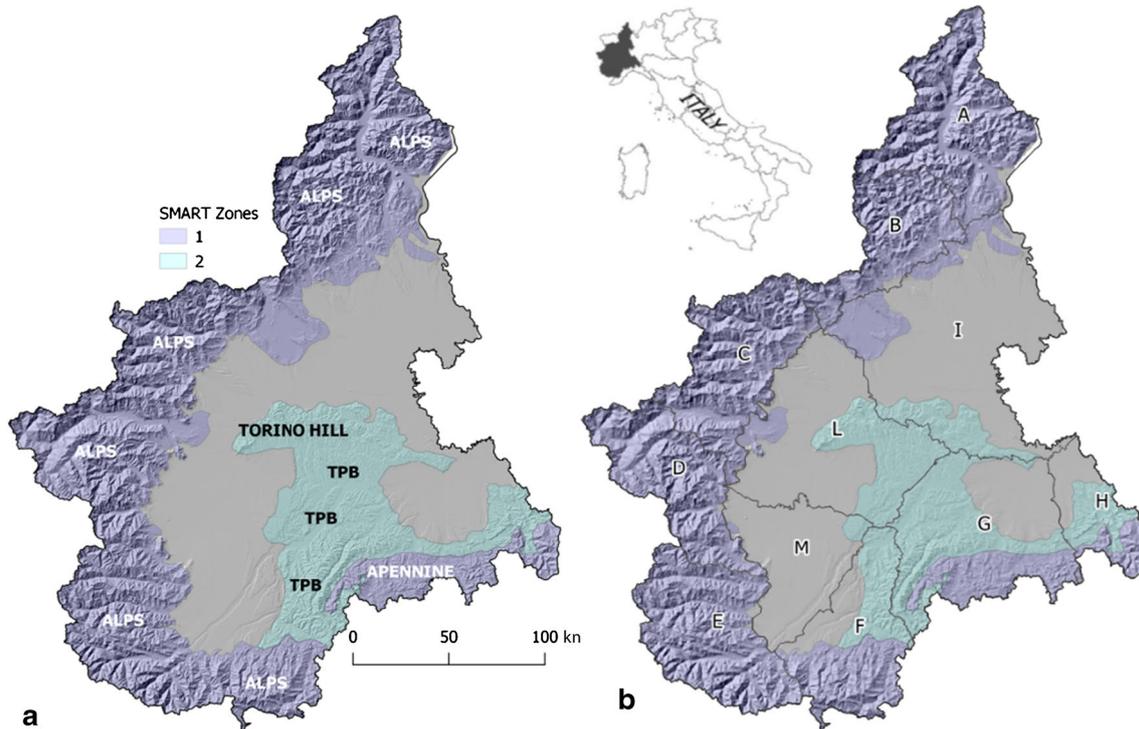
Thresholds become operative for rainfall duration exceeding 12 h. Indeed, for rainfall lasting less than 12 h, thresholds overpassing may indicate the probability of occurrence of other

phenomena commonly triggered by short and intense rainstorms, such as accelerated soil erosion due to widespread surface runoff or channelized debris flows in small Alpine catchments. Intersection between the two homogeneous zones of Fig. 2a and the 11 warning zones of the RWS produces warning zones (Fig. 2b) for the prediction of shallow landslides.

SMART does not employ a probabilistic approach, and therefore, an issued warning has the same degree of severity whether

Table 2 Performance indicators considered for the analyses

Performance indicator	Performance criterion (Fig. )	Symbol	Formula
Efficiency index_(1)	Criterion A	EI_A	$CA / \sum_{ij} d_{ij}$ (excluding $d_{11}$ )
	Criterion B	EI_B	$(Gre + Yel) / \sum_{ij} d_{ij}$ (excluding $d_{11}$ )
	Criteria A+B	EI_A+B	$(Gre + Yel + CA) / 2 * \sum_{ij} d_{ij}$ (excluding $d_{11}$ )
Probability of serious missed alerts	Criteria A+B	$P_{SM-MA}$	$MA_{P_{ur}} / MA$
Probability of serious false alerts	Criteria A+B	$P_{SM-FA}$	$FA_{P_{ur}} / FA$
Missed and false alert balance	Criteria A+B	MFB	$(MA_{P_{ur}} + MA_{Red}) / (MA_{P_{ur}} + MA_{Red} + FA_{P_{ur}} + FA_{Red})$



**Fig. 2** a Areas with homogeneous behavior in the SMART model (Homogeneous areas). b 11 warning zones deriving from the intersection between the RWS Alerting areas and the SMART homogeneous areas

the threshold value is just reached or whether it is exceeded by a considerable amount. However, three levels of warnings are defined, based on an indirect estimation of the expected landslides, estimated as a function of the number of rain gauges for which the rainfall threshold is exceeded in real-time or in forecasting mode.

In addition to the no-warning condition, corresponding to a negligible probability of shallow landslide occurrences, the other warning levels are as follows: (1) yellow (isolated triggering of shallow landslides); (2) orange (diffuse but not widespread triggering of shallow landslides, equivalent to less than 10 landslides in a warning area); (3) red (widespread triggering of shallow landslides, equivalent to more than 10 landslides in a warning area).

### Performance analysis

The performance of the warning model employed in the SMART has been evaluated adopting the EduMaP method (Calvello and Piciullo 2016) using the Excel tool described in the “A tool for the application of the EDuMaP method” section. The analysis was performed considering the values of the 10 input parameters reported in Table 3. Landslide events (LE) are defined, according to Calvello and Piciullo (2016), as a series of landslides grouped together based on their spatial and temporal characteristics. The performance assessment was conducted considering the landslide events (LE) and the warning events (WE) registered in Piemonte between January 2008 and December 2016 (Table 4) in 10 warning zones (from A to L). One warning area (M, see Fig. 2) was not considered since no landslides occurred in that area during the period of analysis.

Figure 3a and b show the results obtained for the ten warning areas, reporting the number of elements of the 10 durations matrices for the two performance criteria reported in Fig. 1, i.e., the alert classification criterion—herein called criterion A—and the grade of correctness criterion—herein called criterion B. The time unit considered in the duration matrix is *day*, consistently with the temporal discretization available for the considered data sets ( $\Delta t = 1$  day). Therefore, considering the time frame of the analysis ( $\Delta T = 9$  years), the total number of elements for each duration matrix is 3287 days.

Figure 4a and b show the results in terms of performance indicators for the ten warning zones. Comparing the efficiency indexes (EI\_A, EI\_B, and EI\_A+B), the higher values are reached for EI\_B, due to the significant number of Yel elements observed in all the warning zones. This means that most of the MA and FA observed in the period of analysis are associated with minor errors of the model. The results provided by these three indicators generally agree in pointing out that the best-performing models are those adopted for zones A and C. It is worth mentioning that in 6 cases out of 10, the EI\_A is lower than 50%, indicating that the sum of MA and FA is higher than the number of CA (especially for zones I and L).

Among the error indicators, the probability of serious missed alerts indicator ( $P_{SM-MA}$ ) is higher than 10% in 6 cases out of 10, pointing out that the majority of severe model errors are related to missed alert of very large landslide events. This can be explained, as discussed in Stoffel et al. (2014), considering that temperature changes cause important modifications of the slopes’ hydrological cycle, as well as of the precipitation type and behavior, such as

**Table 3** Input parameters for the classification, identification, and temporal analysis of landslide events (LE) and warning events (WE)

Parameters of event analysis	Symbol	Relevant for	Value
1. Warning levels	$W_{lev}$	Classification of WE	4
2. Landslide density criterion	$L_{den(k)}$	Classification of LE	LE <sub>1</sub> : 0–1 landslide LE <sub>2</sub> : 2–5 landslides LE <sub>3</sub> : 6–9 landslides LE <sub>4</sub> : ≥ 10 landslides
3. Lead time	$t_{LEAD}$	Identification of WE	0
4. Landslide typology	$L_{typ}$	Identification of LE	Rainfall-induced
5. Minimum interval between Landslide Events	$\Delta t_{LE}$	Identification of LE	1 day
6. Over time	$t_{OVER}$	Identification of LE	0
7. Area of analysis	$A$	Temporal analyses of LE and WE	Piemonte
8. Spatial discretization adopted for warnings	$\Delta A_{(k)}$	Temporal analyses of LE and WE	10 warning zones
9. Time frame of analysis	$\Delta T$	Temporal analyses of LE and WE	3287 days
10. Temporal discretization of analysis	$\Delta t$	Temporal analyses of LE and WE	1 day

increase of winter precipitations and decrease of autumn ones; shortening of snow cover persistence during spring. The accelerated snowmelt contributes significantly to the triggering of shallow landslides, also in presence of spring rainfall of moderate intensity, because water deriving from snow melting completely infiltrates the ground. On the other hand, although the probability of serious false alerts indicator ( $P_{SM-FA}$ ) is equal to zero in 7 cases out of 10, in the remaining 3 cases, more than 30% of the FA are Pur errors. Regarding the missed and false alert balance (MFB), which represents the ratio of MA over the sum of MA and FA, Piciullo et al. (2017a) recommended values lower than 0.25 for operational Te-LEWS to be considered efficient (i.e., the duration of MA should be less than one-third of the duration of FA). This condition is respected only in 1 case out of 10, while MFB is equal to 1 for warning zones E, I, and L.

Figure 5 reports a detailed analysis on the grade of severity of MA and FA and the grade of correctness of CA, distinguishing respectively among Pur, Red, Yel and Gre, Yel.

In all the warning zones, some LE are missed in the period of analysis and in 7 cases out of 10, several Pur errors occur (i.e., a LE of class 4 missed). However, in almost all the cases (8 out of 10), most errors are Yel errors. The exceptions are represented by G and L, the two warning zones characterized by the highest numbers of missed alerts (15 and 12, respectively). The presence of a significant number of Pur and Red is probably due to adopted rainfall thresholds that are inadequately high for these warning zones. The number of FA is generally lower than the number of MA (except for zone H). Besides, in 3 cases out of 10, only MA and no FA are observed in the period of analysis. It should also be mentioned that, when FA occurs, most of them are characterized by Pur and Red errors, revealing that in many cases, warning level (WL) 3 and 4 were issued without the occurrence of large LE. Finally, the warning model was able to correctly predict the occurrence of several landslide events in all the warning zones, especially in A, B, and C. However, as already noted, a relatively slight number of correct alerts are associated with the best performance of the model (i.e., Gre elements) and in three warning zones (E, H, and L), only Yel elements were observed.

## Discussion

### Metrics of success and error for Te-LEWS

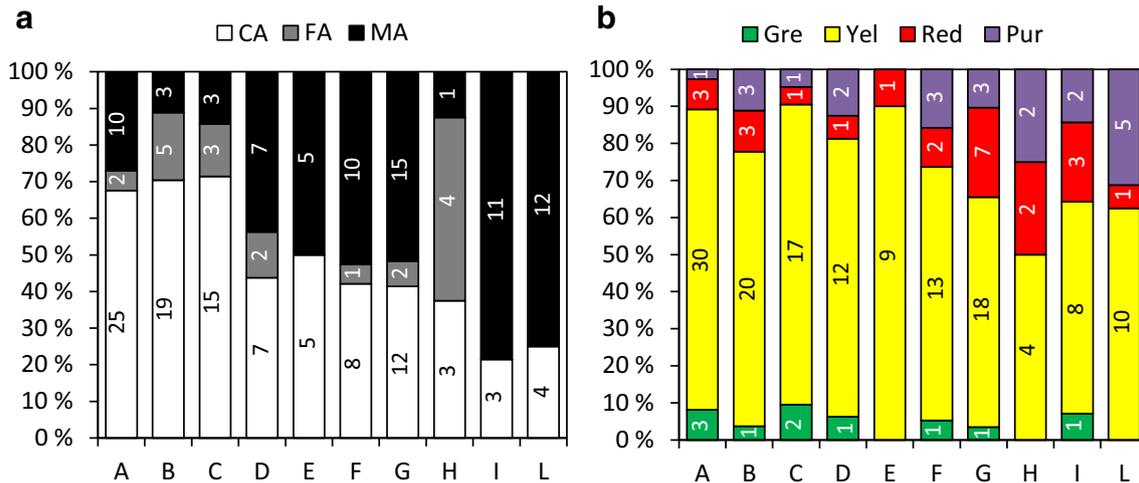
Different performance indicators are available in the scientific literature for rainfall threshold validation and performance of LEWS (see the “Performance assessment of territorial landslide early warning systems” section). The following three indicators (see Table 2) are herein used for the comparison of the performance of the SMART model with different models adopted in other Te-LEWS: efficiency index<sub>(1)</sub>, performance criterion A, (EI\_A); and missed and false alerts balance (MFB). These values are discussed in relation to the values provided in the literature by different authors.

In the LEWS operating in Hong Kong, two warning models currently coexist. Indeed, a SWIRLS Landslip Alert (SLA) model was developed and added to the system to provide some lead time (up to 3 h) to the warnings before the standard landslip warning criteria are exceeded. The SLA model considers the rolling 21 h of measured rainfall plus a 3-h rainfall forecast, whereas the standard landslip warning model is based on the measured 24-h accumulated rainfall. The performance for the period 2001–2004 of both models has been reported in Cheung et al. (2006). The EI\_A of the models resulted, respectively, equal to 61% and 78%, for the SLA and the landslip warning models. Based on these values, the author stated that both the SLA and landslip warning were found to be generally effective. To compare the performance of these models with the performance evaluation carried out in this manuscript for the SMART system, the data provided in the paper by Cheung et al. (2006) have been used to compute the missed and false alert balance. The SLA and landslip warning models in the period 2001–2004 showed MFB values equal to 33% and zero.

In Restrepo et al. (2008), a performance analysis of the prototype debris flow warning system for recently burned areas in Southern California has been carried out for the winter of 2005/2006 (first year of operation). In this case, the probability of detection (92%) and the false alert rate (72%) (see Table 1) have been evaluated. Considering the same database, for comparative purposes, the EI\_A and missed alert balance have been computed. The values are quite low for the success indicator EI\_A: 28%. The

**Table 4** Number of days with landslide events (LE) and warning events (WE) of level 2, 3, and 4 (warning levels reported within circles) in the ten warning zones for the period of analysis

WZ	2008	2009	2010	2011	2012	2013	2014	2015	2016
A	LE 3②	3④	3②	2②3③	3②				
	WE 6②	4②2③	3②	2②	3②	1②	2②3③		3②
B	LE	1②3④		1②					
	WE 5②	2②1③1④	5②1④	1②1④	1②	1②	1②2③		1③
C	LE 1②	3②		②					2②
	WE 5②1③	1②1③	3②	4②		1②			1②1④
D	LE 2②	1②	1②	2②			1②		3④
	WE 2②2③	1②	2②	1②					1③
E	LE 2②2③			1②					
	WE 2②	1②	1②	1②				1②	
F	LE	2②3④	1②	1②			1②		3④
	WE 1②	2②	1②	3③			1②	1②	1②1③1④
G	LE 2④	2②2④	2②	2②3③					3③
	WE 3②	3②	2②	1③			3②		1②2③
H	LE			②					
	WE 1②	1③	2②	1④			1③1④		
I	LE 3②	3④	1②2③	2②		1②			
	WE 1②	1②	1②	1②					
L	LE	1②3④	2②	3④		3②			
	WE 1②	1②	2②	1②					



**Fig. 3** a, b Relative distribution of CA, MA, and FA (criterion A) and Pur, Red, Yel, and Gre (criterion B) in the ten warning zones, not considering the TN elements

missed alert balance is low due to the high number of false alerts compared with both missed alerts and correct alerts. Since the values of the success indicators are quite low, the indicators have not been included in the comparison in Fig. 6.

The performance analysis of the SIGMA model, employed in the LEWS operational in the Emilia Romagna region, Italy, and described in Martelloni et al. (2012), and Lagomarsino et al. (2013), reports very high values for the odds ratio and the efficiency index, which prompt the authors of those papers to highlight the very good predictive power of the model. However, those performance indicators have been computed including TN, and they are significantly influenced by the very high number of TN in the period of analysis. The analysis performed for the SMART system, purposefully exclude TN elements from the computation of the performance indicators, for the reasons described in the “Indicators used for rainfall threshold validation and performance evaluation of Te-LEWS” section. For comparative purposes, the data provided in the papers by Martelloni et al. (2012), and Lagomarsino et al. (2013), have been used to compute new values of the performance indicators not including TN. The results for the three selected indicators are, respectively, for the two papers: EI\_A equal to 15.9%, 14%; 0.17; missed and false alert balance equal to 6.9%, 5%. Since the values of the success indicators are quite low, also in this case the indicators have not been included in the comparison (Fig. 6). However, some updates to increase the performance of SIGMA model have been recently realized and published in Segoni et al. (2018b).

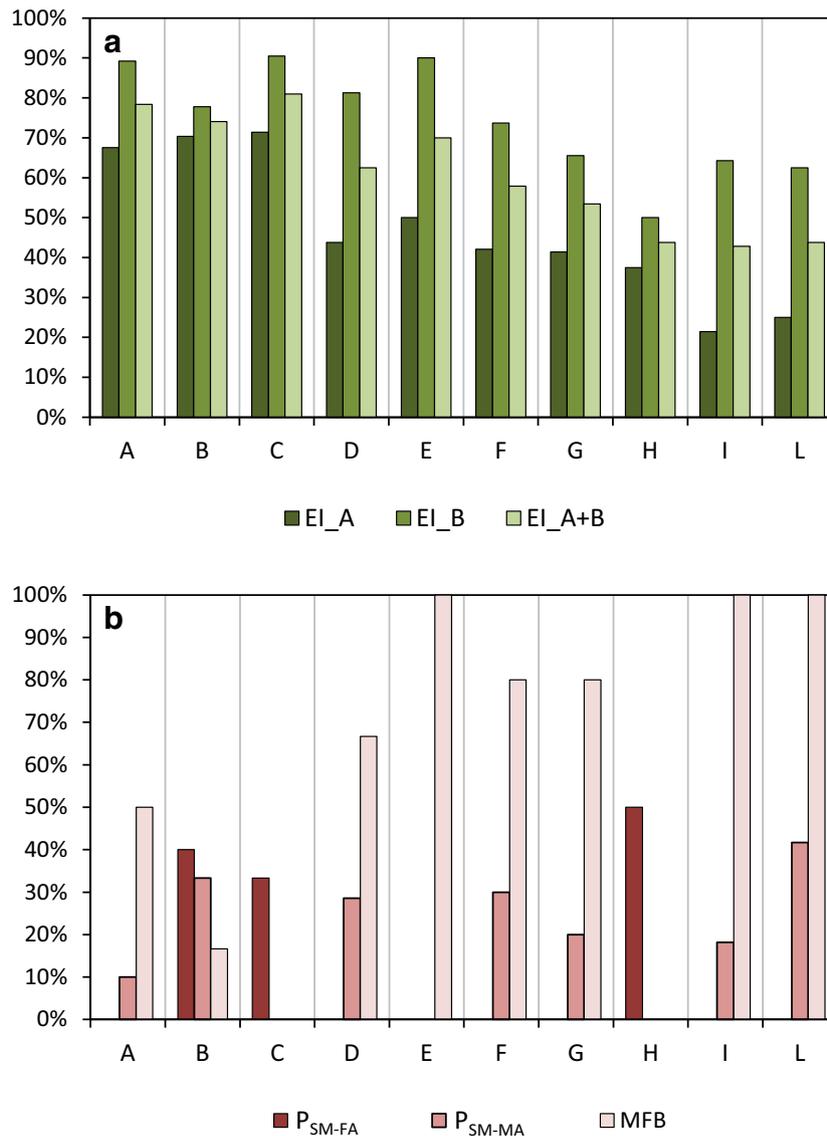
Calvello and Piciullo (2016) reported the first application of the EDuMaP method. They applied it to the municipal early warning system operating in Rio de Janeiro, Brazil, for which they carried out a parametric analysis. They also presented a list of indicators for the performance evaluation of LEWS. Among them, the efficiency index was evaluated in the same way of EI\_A of Table 1 (considering how criterion A was applied); thus, they are directly comparable with the results obtained for the SMART model. The two performance indicators were evaluated for two warning zones (out of 4) of the municipality: baia de Guanabara and zona Sul. In these two zones, the authors report values of EI\_A equal to 75%

(baia de Guanabara) and 66% (zona Sul). The MFB, herein calculated considering the data provided in the paper by Calvello and Piciullo (2016), has the following values: 14.5% for baia de Guanabara and 3.4% for zona Sul.

To compare the performance evaluation of the SMART model with the literature case studies previously mentioned, a radar chart is used (Fig. 6) where the literature experiences are reported for all the 10 warning zones. The computed values of the three versions of the efficiency index, computed without considering the TN (EI\_A, EI\_B, EI\_A+B), are compared with the following values: 61% (SLA from Cheung, 2006); 78% (landslip warning from Cheung, 2006) 75% (baia de Guanabara from Calvello and Piciullo 2016); 66% (zona Sul from Calvello and Piciullo 2016). The computed values of the missed and false alert balance (MFB) are compared with the following values: 33% (SLA from Cheung, 2006); 14.5% (baia de Guanabara from Calvello and Piciullo 2016); 3.4% (zona Sul from Calvello and Piciullo 2016), and 25% (Piciullo et al. 2017a). Figure 6 clearly shows that the application of the performance criterion A is the most conservative (see blue markers) and that criterion B is the one providing the highest values of the indicators (see red markers). Zones A, B, C, E, F, and G have high values of EI\_A compared with the references from the literature. Piciullo et al. (2017a) recommended a value of MFB lower than 25% for considering efficient a warning model within an operational Te-LEWS (i.e., only one wrong alert out of four is a MA). This condition is respected for the 3 zones out of 10 in our analyses: B, C, H. This comparison shows that the SMART model employed in zone C is giving the best performance. On the contrary, zones L and I should be considered for thresholds improvement, since their performance is quite poor.

#### Comparison with simpler validation techniques

The validation of the thresholds used in the SMART LEWS was conducted in 2008 using a  $2 \times 2$  contingency table (Tiranti and Rabuffetti 2010), evaluating the joint distribution of “yes”/“no” and “landslide forecast”/“occurrence.” The performance evaluation was conducted considering the whole set of widespread shallow landslide events that occurred between 1990 and 2002. The



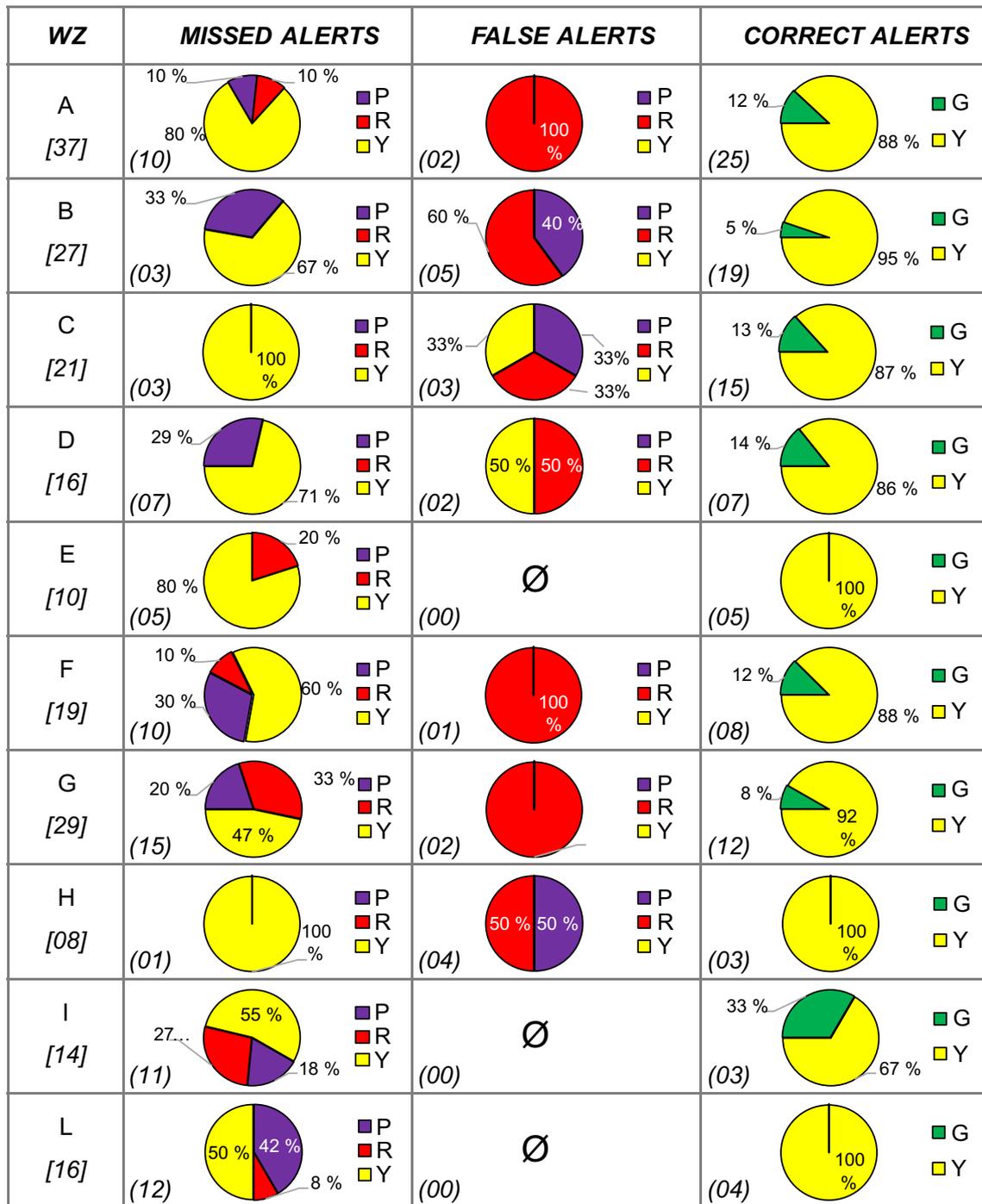
**Fig. 4** Bar charts showing the values of success (a) and error (b) indicators for each warning zone. Values shown as percentages on the primary vertical axes: efficiency indexes (EI\_A, EI\_B, EI\_A+B); probability of serious false alerts (P<sub>SM-FA</sub>); probability of serious missed alerts (P<sub>SM-MA</sub>). Absolute values on the secondary vertical axis: missed and false alert balance (MFB)

indicators considered for the analysis were hit rate (HR), false alert rate (FAR), and efficiency index<sub>(1)</sub> (EI). Formulas are reported in Table 1.

The results showed the following values for the three analyzed indicators: HR = 0.83, FAR = 0.45, and EI = 0.49 (Tiranti and Rabuffetti 2010). Among all the single landslides that occurred in the period of analysis, 83% has been correctly predicted. Yet, the high number of false alerts produced a rather low value of EI and a high value of FAR. As already mentioned, when employing a 2 × 2 contingency table for the performance evaluation of LEWS, it is not possible neither to distinguish among different warning levels nor considering the number of landslides. Consequently, it is not possible to identify the warning levels that cause false alerts. Usually, when a low warning level is issued, one or few landslides are expected. However, it is not always possible to record all the landslides that occur in a warning zone, since the area covered is

always very wide and the places that are not urbanized are significant. Consequently, one should judge with care the assessment of the non-occurrence of one or few landslides when a low warning level has been issued.

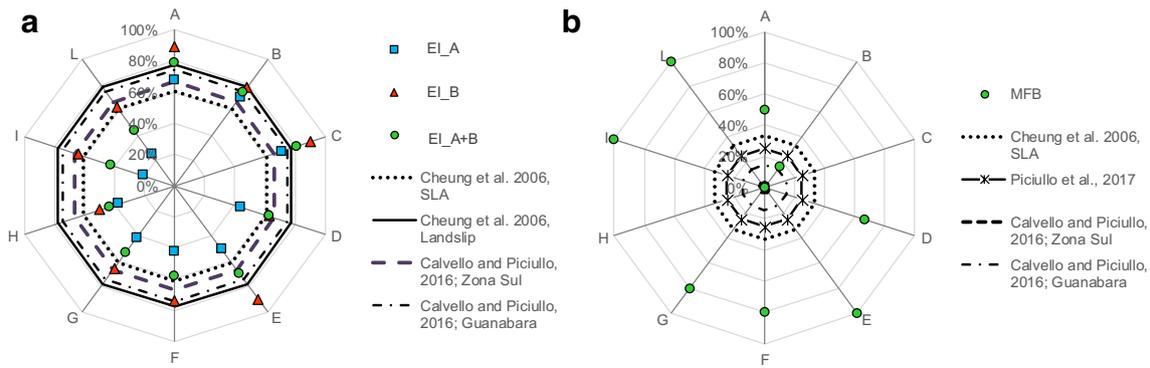
Table 5 reports the results of a performance evaluation of the SMART model conducted adopting the same validation technique (i.e., 2 × 2 contingency table) adopted in Tiranti and Rabuffetti (2010), using the database described in the “Performance analysis” section, for the period 2008–2016. Two different comparisons can be derived from the results of this new analysis: (i) comparison of the performance of the SMART model in two different time periods, using a simple validation technique, and (ii) comparison of the results obtained conducting the performance assessment in two different ways, i.e., by employing the EDuMaP method and a simpler validation technique.



**Fig. 5** Grade of correctness distribution for MA, FA, and CA. For each warning zone (WZ), the total number of days without TN is reported in square brackets, and the number of MA, FA, and CA is reported in round brackets

Concerning the first issue, i.e., comparison of the performance of the SMART model in two different time periods, the results clearly highlight a decrease of EI, whose values change from 0.49 (period 1990–2002) to 0.14 (period 2008–2016), demonstrating that the general accuracy of the SMART model significantly decreased in a relatively short period of time. This is also confirmed by the low value of HR (0.28), mainly due to the relevant

number of missed landslide events (70 out of 97). Besides, an increasing number of false alarms can be also observed, as the value of FAR is equal to 0.79. The very different performance of the SMART model in the two periods could be associated with the non-stationarity of the rainfall characteristics in the two periods. The shallow landslide events' behavior has changed between 1960 and 2016 in Piemonte, as shown by the data



**Fig. 6 a, b** Comparison between the performance indicators of the SMART model for all the 10 warning zones of the Piemonte region, with literature case studies. The computed values of the three versions of the efficiency index (EI\_A, EI\_B, EI\_A+B) are compared with the following values: 61% (SLA from Cheung, 2006); 78% (landslip warning from Cheung, 2006); and 51% (Staley et al. 2013). The computed values of the missed and false alert balance (MFB) are compared with the following values: 33% (SLA from Cheung, 2006) and 25% (Piciullo et al. 2017a)

reported in Table 6. Landslide events until around 2000 were characterized by an average return period of about 5 years, high number of phenomena (from 1000 to more than 10,000) during a single event, and a higher frequency of occurrence in the fall season (September–November). After the year 2000, the frequency of the landslide events has increased (about one event per year), the main season of occurrence became spring (March–June), and the events are typically characterized by a lower number of landslides (from 50 to about 2000), as already reported by Stoffel et al. (2014) and subsequently updated by Tiranti et al. (2019). All that considered, the performance of the SMART model has also been most likely influenced by the significant changes of the weather pattern that have been occurring in the area in a relatively short time (Cremonini and Tiranti 2018). In fact, the SMART model was calibrated considering landslide events that occurred between 1990 and 2002, thus practically using landslide data before the recorded (almost abrupt) change in behavior of the temporal and spatial distribution of widespread shallow landslide events.

About the second issue, i.e., comparison of results obtained conducting the performance assessment in two different ways, the performance computed with the simpler validation technique is generally poor for all the warning zones (especially for E, H, and L), as highlighted by the very low values of EI (Table 5). Looking at HR, it is worth mentioning that in almost all the cases (apart from B and C) more than half of the occurred landslides were not forecasted by the model. Besides, the high values of FAR suggest that for all the warning zones, most of the warnings issued are false alarms. On the other hand, the performance evaluation carried out with the EDuMaP method highlighted a relatively good model performance in several warning zones (especially in A, B, and C). This can be explained considering that the EDuMaP method allows for a more detailed analysis on the severity of the errors and the correctness of the predictions. The performance analyses, carried out with the two methods, also indicate different warning zones as the best-performing ones: A and C using the EDuMaP method; D using the simpler validation technique. This difference can be related again to the possibility of a more detailed

**Table 5** Performance analysis of the SMART model from 2008 to 2016, conducted considering the validation technique adopted by Tiranti and Rabuffetti (2010): performance indicators (HR, FAR, and EI) for each warning zone

Warning zone	Contingency table 2 × 2			HR	FAR	EI [%]	EDuMaP			EI A+B [%]
	CA	MA	FA				CA	MA	FA	
All	27	70	101	0.28	0.79	14				
A	6	9	22	0.40	0.79	16	25	10	2	78
B	4	3	20	0.57	0.83	15	19	3	5	74
C	3	3	15	0.50	0.83	14	15	3	3	81
D	4	6	6	0.40	0.60	25	7	7	2	62
E	1	5	5	0.17	0.83	9	5	5	0	70
F	3	9	7	0.25	0.70	16	8	10	1	58
G	3	13	13	0.19	0.81	10	12	15	2	53
H	0	1	7	0.00	1.00	0	3	1	4	44
I	2	10	2	0.17	0.50	14	3	11	0	43
L	1	11	4	0.08	0.80	6	4	12	0	44

**Table 6** Shallow landslide events observed from 1962 to 2016 in Piemonte (fall events in italic—modified from Tiranti et al. 2019). *Tertiary Piedmont Basin (TPB)*

Event date	Involved area	Number of landslides
<i>November 1962</i>	<i>Alps and TPB</i>	> 2000
<i>November 1968</i>	<i>Alps</i>	≈ 2000
February–May 1972	Torino Hill and TPB	> 200
May 1974	Torino Hill and TPB	≈ 500
May 1977	Alps	≈ 100
<i>October 1977</i>	<i>Alps and TPB</i>	≈ 1000
August 1978	Alps	≈ 700
August 1987	Alps	≈ 100
<i>September and October 1993</i>	<i>Alps and TPB</i>	≈ 200
<i>November 1994</i>	<i>TPB and Alps</i>	≈ 10,000
<i>October 1996</i>	<i>Alps</i>	≈ 50
<i>September 1998</i>	<i>Alps</i>	≈ 50
May 1999	Alps and TPB	≈ 50
<i>October 1999</i>	<i>TPB and Apennine</i>	≈ 50
June 2000	Alps and TPB	≈ 100
<i>October 2000</i>	<i>Alps</i>	> 2000
<i>November 2000</i>	<i>Alps</i>	≈ 100
March 2002	Alps	≈ 50
May 2002	Alps	≈ 100
June 2002	Alps	≈ 100
<i>November 2002</i>	<i>TPB and Alps</i>	≈ 500
May 2004	Alps	≈ 10
June 2007	Torino Hill	≈ 20
May 2008	Alps	≈ 300
December 2008	TPB	≈ 100
April 2009	TPB and Alps	≈ 1000
May 2010	TPB	≈ 10
June 2010	Alps	≈ 100
March 2011	TPB	≈ 100
<i>November 2011</i>	<i>Alps</i>	≈ 300
April–May 2013	Hills	≈ 200
May 2013	Torino Hill and Alps	≈ 200
<i>November 2014</i>	<i>Alps and TPB</i>	≈ 200
<i>November 2016</i>	<i>Alps and TPB</i>	> 2000

assessment of the model performance when the EDuMaP method is used. In this case, this highlights that the large majority of MA and FA, in some warning zones, are not severe errors of the warning model.

### Conclusions

The performance evaluation of LEWS is often overlooked; however, different indicators are available in the literature and can be

employed for this task. These indicators have been homogenized and proposed in the “Indicators used for rainfall threshold validation and performance evaluation of Te-LEWS” section. Few of them have been judged by the authors to be essential for describing the performance of a LEWS (see Table 2). The most important indicators that can give a general overview of the system performance are the efficiency index (EI) and the missed and false alert balance (MFB). The first can be

considered to evaluate the general success rate of a Te-LEWS; the latter can be used to evaluate the percentage of missed alerts among the wrong predictions (sum of false and missed alerts). Then, to have a more detailed understanding on the severity of the missed and false alerts (i.e., wrong predictions that belong to the purple cells), it is relevant to evaluate and analyze the probability of serious missed alerts and probability of serious false alerts ( $P_{SM-FA}$ ,  $P_{SM-MA}$ ). They quantify, respectively, the percentage of the serious no-warning mistakes (i.e., missed alerts of a high LE class) and of serious no landslides mistakes (i.e., false alerts with high levels of warning issued).

According to the results of these four indicators, it is possible to fully evaluate the system performance and to identify the warning levels and, consequently, the thresholds that need to be refined. Concerning the use of the efficiency index to evaluate different criteria (A, B, A+B), it is possible to state that the values of the EI for the criteria A and B correspond, respectively, to a lower and an upper bound (Fig. 6a). The use of EI for the combined criterion (A+B) is, however, to be preferred for the performance analysis. By comparing the performance of different Te-LEWS, it is possible to state that a system should have a EI higher than 60%. However, EI (A+B) in an efficient system should exceed 80%, as it is the case for zones A and C in our application. Concerning MFB, it should not be greater than 20%, better if its value is lower than 10%. For evaluating the indicators, the element  $d_{11}$  (i.e., the TN values) of the duration matrix has been purposefully neglected, to avoid an overestimation of the performance (see the “Metrics of success and error for Te-LEWS” section).

The definition of the landslide events deserves some remarks. It is influenced by a series of choices the analyst needs to make in selecting and grouping landslides (Calvello and Piciullo 2016). The definition of limit values to differentiate among  $k$  classes of landslide events (see Table 3) has been discussed at length with the SMART system managers. Standard or commonly used procedures do not exist in literature since the classification of landslide events varies, as it should, as a function of the LEWS under investigation. Indeed, this classification depends on how the warning levels, and their thresholds, have been defined, as well as on the expected number of landslides associated with each warning level. For these reasons, it is of great importance that this parameter, as well as the definition of the performance criteria (see the “A tool for the application of the EDuMaP method” section), is defined by the analyst in accordance with the system managers. A parametric analysis carried out in Piciullo et al. (2017b) shows how the performance evaluation can differ as a function of the landslide criterion and how its definition is a crucial point to obtain a correct performance evaluation of the warning model.

Performance assessment of Te-LEWS is a fundamental issue to run an efficient warning model. Often the performance analysis of Te-LEWS is carried out considering a  $2 \times 2$  contingency matrix. Yet, with this method, it is not possible to differentiate among different warning levels and number of occurred landslides in a given time interval. For instance, a missed alert of just 1 landslide is judged in the same way of a missed alert of many landslides. Moreover, the error associated with the highest level of warning issued when no landslides occurred is judged in the same way as the one associated with any lower level of warning issued with no

landslides. To overcome these issues, an advanced method for the performance evaluation of LEWS should be used.

An Excel-based tool (freely available on request), programmed in VBA, has been recently released to increase and speed the applicability. In this paper, EDuMaP has been employed, using the Excel tool, to the SMART warning model operational in the LEWS of Piemonte region, Italy. The results highlight that the SMART model has a good performance in some warning zones: A, B, and C (see Fig. 4). Detailed insights emerge by analyzing the results of the performance evaluation carried out with this method (see Figs. 4 and 5 and Table 6). The same considerations and analyses could not be carried out with simpler methods (see the “Comparison with simpler validation techniques” section). For instance, the EDuMaP method allows for a more detailed assessment of the seriousness of the errors and of the correctness of the predictions. In the specific case of LEWS operating in Piemonte, the EDuMaP method highlighted that the large majority of MA and FA, in some warning zones, were not severe errors. As expected, the warning zones showing the highest performance differ when different performance evaluations are carried out: A and C using the EDuMaP method; D using the simpler validation technique (see the “Comparison with simpler validation techniques” section). Finally, it is worth mentioning that, after being operational for almost 20 years, the SMART model will be soon replaced by a new model, named SLOPS (Tiranti et al. 2019), that upgrades some weakness aspects of the previous model. In the near future, the performance of the SLOPS model in the early prediction of landslides will be evaluated with an advanced performance model and compared with the SMART model.

#### Acknowledgments

This paper has been conceived in the context of the Klima 2050 project—*Risk reduction through climate adaptation of buildings and infrastructure*, and it is included in the *Work Package 3. Landslides triggered by hydro-meteorological processes*. The authors thank the Research Council of Norway for financing the Open Access publication of this paper.

#### Funding Information

Open Access funding provided by Norwegian Geotechnical Institute.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bazin S (Ed.) (2012) SafeLand deliverable 4.8. Guidelines for landslide monitoring and early warning systems in Europe – design and required technology. European Project SafeLand, Grant Agreement No. 226479, Deliverable 4.8, 153 pp., available at: <http://www.safeland-fp7.eu>, 2012 <https://www.ngi.no/eng/Projects/SafeLand> (Last access: Apr 2017)
- Calvello M (2017) Early warning strategies to cope with landslide risk. *Rivista italiana di geotecnica* 2/2017, 63–91. Patron editore. doi: <https://doi.org/10.19199/2017.2.0557-1405.063>
- Calvello M, Piciullo L (2016) Assessing the performance of regional landslide early warning models: the EDuMaP method. *Nat Hazards Earth Syst* 16:103–122. <https://doi.org/10.5194/nhess-16-103-2016>
- Calvello M, d'Orsi R, Piciullo L, Paes NM, Magalhaes MA, Coelho R, Lacerda WA (2015) The community-based alert and alarm system for rainfall induced landslides in Rio de Janeiro, Brazil. In: Engineering geology for society and territory “landslide processes”, Proc. XII Int. IAEG Congress, Torino, Italy, 2, 653–657, doi:[https://doi.org/10.1007/978-3-319-09057-3\\_\(1\)09](https://doi.org/10.1007/978-3-319-09057-3_(1)09)
- Cheung PY, Wong MC, Yeung HY (2006) Application of rainstorm Nowcast to realtime warning of landslide hazards in Hong Kong. In: WMO PWS Workshop on Warnings of Real-Time Hazards by Using Nowcasting Technology, Sydney, Australia, 9–13 October 2006
- Cremonini R, Tiranti D (2018) The weather radar observations applied to shallow landslides prediction: a case study from north-western Italy. *Front Earth Sci* 6:134. <https://doi.org/10.3389/feart.2018.00134>
- Di Biagio E, Kjekstad O (2007) Early warning, instrumentation and monitoring landslides. 2nd Regional Training Course, RECLAIM II, 29th January - 3rd February 2007
- Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova OG, Vennari C, Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* 228:653–665. <https://doi.org/10.1007/s11069-014-1129-0>
- HFA (2005) Hyogo Framework for Action 2005–2015: building the resilience of nations and communities to disasters. World Conference on Disaster Reduction 18–22 January 2005, Kobe, Hyogo, Japan, Final report, p 22. (<http://www.unisdr.org/we/inform/publications/1037>, Accessed 14 Jan 2020)
- Intrieri E, Gigli G, Casagli N, Nadim F (2013) Brief communication “landslide early warning system: toolbox and general concepts”. *Nat Hazards Earth Syst* 13:85–90. <https://doi.org/10.5194/nhess-13-85-2013>
- Lagomarsino D, Segoni S, Fanti R, Catani F (2013) Updating and tuning a regional scale landslide early warning system. *Landslides* 10:91–97. <https://doi.org/10.1007/s10346-012-0376-y>
- Lagomarsino D, Segoni S, Rosi A, Rossi G, Battistini A, Catani F, Casagli N (2015) Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting. *Nat Hazards Earth Syst* 15(2413–2423):2015–2423. <https://doi.org/10.5194/nhess-15-2413-2015>
- Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* 9(4):485–495. <https://doi.org/10.1007/s10346-011-0308-2>
- Maskrey A (1997) Report on national and local capabilities for early warning. IDNDR secretariat, Geneva, p 33 (<http://lib.riskreductionafrica.org/handle/123456789/1255>, last accessed: 14/1/2020)
- Pecoraro G, Calvello M, Piciullo L (2019) Monitoring strategies for local landslide early warning systems. *Landslides* 16(2):213–231. <https://doi.org/10.1007/s10346-018-1068-z>
- Piciullo L, Dahl M-P, Devoli G, Colleuille H, Calvello M (2017a) Adaptation of the EDuMaP method for the performance evaluation of the alerts issued on variable warning zones. *Nat Hazards Earth Sys* 17(6):817–831. <https://doi.org/10.5194/nhess-17-817-2017>
- Piciullo L, Gariano SL, Melillo M, Brunetti MT, Peruccacci S, Guzzetti F, Calvello M (2017b) Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides* 14(3):995–1008. <https://doi.org/10.1007/s10346-016-0750-2>
- Piciullo L, Calvello M, Cepeda JM (2018) Territorial early warning systems for rainfall-induced landslides. *Earth Sci Rev* 179:228–247. <https://doi.org/10.1016/j.earscirev.2018.02.013>
- Rabuffetti D, Barbero S (2005) Operational hydro-meteorological warning and real-time flood forecasting: the Piemonte region case study. *Hydrol Earth Syst Sci* 9(4):457–466
- Rabuffetti D, Mensio L, Barbero S, Pelosini R, Bovo S (2003) An operating alert procedure for meteo-hydrological risk forecasting and management, Proceedings of the Sixth European Conference on Applications of Meteorology, Rome, 15–19 September 2003
- Restrepo P, Jorgensen DdP, Cannon SH, Costa J, Laber J, Major J, Martner B, Purpura J, Werner K (2008) Joint NOAA/NWS/USGS prototype debris flow warning system for recently burned areas in Southern California. *Am Meteorol Soc* 2008:1845–1851
- Segoni S, Piciullo L, Gariano SL (2018a) Preface: landslide early warning systems: monitoring systems, rainfall thresholds, warning models, performance evaluation and risk perception. *Nat Hazards Earth Syst Sci* 18:3179–3186. <https://doi.org/10.5194/nhess-18-3179-2018>
- Segoni S, Rosi A, Fanti R, Gallucci A, Monni A, Casagli N (2018b) A regional-scale landslide warning system based on 20 years of operational experience. *Water* 10(10):1297
- Staley DM, Kean JW, Cannon SH, Schmidt KM, Laber JL (2013) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides* 10:547–562. <https://doi.org/10.1007/s10346-012-0341-9>
- Stoffel M, Tiranti D, Huggel C (2014) Climate change impacts on mass movements – case studies from the European Alps. *Sci Total Environ* 493:1255–1266. <https://doi.org/10.1016/j.scitotenv.2014.02.102>
- Tiranti D, Rabuffetti D (2010) Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation. *Landslides* 7:471–481. <https://doi.org/10.1007/s10346-010-0198-8>
- Tiranti D, Nicolò G, Gaeta AR (2019) Shallow landslides predisposing and triggering factors in developing a regional early warning system. *Landslides* 16(2):235–251. <https://doi.org/10.1007/s10346-018-1096-8>
- United Nations Inter-Agency Secretariat of the International Strategy for Disaster Reduction (UN ISDR) (2006) Global survey of early warning systems: an assessment of capacities, gaps and opportunities towards building a comprehensive Global Early Warning System for All Natural Hazards. Available at: <http://www.unisdr.org/2006/ppew/info-resources/ewc3/Global-Survey-of-Early-Warning-Systems.pdf>, Accessed July 2020)

**L. Piciullo** (✉) · **J. M. Cepeda**

Norwegian Geotechnical Institute,  
Oslo, Norway  
Email: luca.piciullo@ngi.no

**D. Tiranti**

Department of Natural and Environmental Risks,  
Regional Agency for Environmental Protection of Piemonte,  
Turin, Italy

**G. Pecoraro** · **M. Calvello**

Department of Civil Engineering,  
University of Salerno,  
Fisciano, Italy