Rosa M. Palau● · **Marc Berenguer**● · **Marcel Hürlimann**● · **Daniel Sempere-Torres**●

Application of a fuzzy verification framework for the evaluation of a regional-scale landslide early warning system during the January 2020 Gloria storm in Catalonia (NE Spain)

Abstract The Gloria storm rainfalls affected Catalonia from 20 to 23 January 2020 and triggered multiple landslides, some of which affected buildings and infrastructures (such as roads and railways). This paper presents the rainfall and landslide datasets collected during the event, and evaluates the performance of a regional landslide early warning system (LEWS) during the Gloria storm applying a fuzzy verification method. The majority of the inventoried landslides can be classified as slides, involving a limited volume of sediment (up to 10 m³), and were triggered in cut slopes along linear infrastructures. Rainfall accumulations were significant in the whole region, especially in the Montseny area, where over 450 mm were registered in 96 h. Generally, the LEWS computed moderate and high warnings in the areas where large rainfall amounts were recorded, and showed good correspondence with the locations where landslides were reported. The fuzzy verification method has been applied using neighbouring windows of different sizes to obtain scale-dependant information on the LEWS performance. The skill of the LEWS considerably improves when enlarging the neighbouring window size from 500 m to 1 km.

Keywords Gloria storm · MORLE · Quantitative rainfall estimation · Landslide early warning system · Fuzzy verification · Uncertainty

Introduction

Multiple-occurrence regional landslide events (MORLEs) are defined as hundreds of individual landslides occurring almost simultaneously over large areas (Crozier 2005). Usually, MORLEs are constituted by shallow slides or flows that are triggered in steep slopes by intense rainstorms or earthquakes. MORLEs have been described in different regions around the globe, such as New Zealand (Crozier 2005), Taiwan (Yu et al. 2006), China (Yang et al. 2020), the USA (Campbell 1975; Whittaker and McShane 2012), Switzerland (Nicolet et al. 2013), or Italy (Crosta and Frattini 2003; Lombardo et al. 2018).

Several MORLEs also happened in the region of Catalonia (NE Spain) in the past: October 1940 (Portilla 2014), August 1963 (Portilla 2014), November 1982 (Gallart and Clotet 1988; Corominas and Alonso 1990), June 2008 (Portilla et al. 2010), or June 2013 (Shu et al. 2019). These MORLEs mainly affected the Pyrenees and Pre-Pyrenees and were associated with severe rainfall events and flooding. Most recently, from 20 to 23 January 2020, an extraordinary E-NE cyclonic storm (named Gloria) affected the region of Catalonia. The significant and widespread Gloria storm rainfalls triggered multiple landslides, especially in the Montseny (Fig. 1).

The high number of landslides and the large area affected by MORLEs normally suppose a challenge to the authorities in charge of managing the risk and the maintenance of roads and railways. In this context, regional landslide early warning systems (LEWS) may help to identify the time and location where landslides are most likely to occur and increase their preparedness (Alfieri et al. 2012; UNISDR 2015).

In the last 20 years, regional landslide early warning systems have been developed covering multiple regions, e.g. Southern California (Baum and Godt 2010), Rio de Janeiro (Calvello et al. 2015), Indonesia (Hidayat et al. 2019), Hong Kong (Lloyd et al. 2001), Japan (Osanai et al. 2010), the Zhenjiang province in China (Yin et al. 2008), the city of Busan in South Korea (Park et al. 2019), Norway (Krøgli et al. 2018), the Emilia-Romagna and Campania regions in Italy (Piciullo et al. 2017b; Segoni et al. 2018), and Catalonia in Spain (Berenguer et al. 2015; Palau et al. 2020). Usually, LEWS determine the areas that are prone to landslides employing susceptibility maps and assess whether a rainfall event might trigger a landslide using rainfall thresholds (Aleotti 2004; Guzzetti et al. 2007; Papa et al. 2013; Rossi et al. 2017; Pan et al. 2018). The majority of regional-scale LEWS use rain gauge data to assess the rainfall hazard. However, in many cases, the density of rain gauge networks is low, and landslide-triggering rainfall tends to be underestimated (Nikolopoulos et al. 2014). Other LEWS use remote sensing data such as satellite or ground-based radar rainfall products (Berenguer et al. 2015; Rossi et al. 2017; Kirschbaum and Stanley 2018).

LEWS need regular and systematic performance analysis to assure the reliability of the models. Up to date, research has mainly focused on the validation and improvement of rainfall thresholds (Gariano et al. 2015; Brunetti et al. 2018) and susceptibility maps (Kirschbaum et al. 2016). Only a few studies have put their attention in back-analysing the output warnings and its correspondence with reported landslides. Calvello and Piciullo (2016) and Piciullo et al. (2020) proposed the EDUMAP method for the evaluation of regional-scale LEWS during long periods. This methodology considers the possible occurrence of multiple landslides, and takes into account the relation between the duration of the warnings and the landslide reporting time. Kirschbaum et al. (2009) and Park et al. (2020) proposed a neighbouring window to determine the performance of global and regional-scale LEWS. In this line, fuzzy verification methods have long been employed to assess the performance of mesoscale high-resolution precipitation forecasts (Brooks et al. 1998; Atger 2001; Damrath 2004; Roberts and Lean 2008; Ebert 2008; Marsigli et al. 2008; Clark et al. 2010) and could be applied for the evaluation of LEWS performance. Fuzzy verification
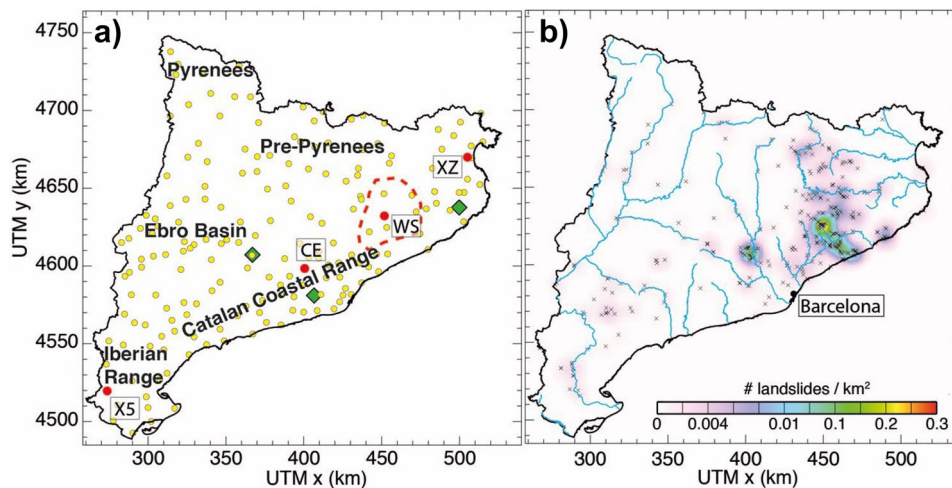
Fig. 1 a General overview map of Catalonia. The green diamonds show the location of the weather radars and yellow circles the 183 rain gauges. The four red circles show the location of the Viladrau (WS), PN dels Ports (X5), Torroella de Fluvià (XZ), and Els Hostalets de Pierola (CE) rain gauges. The red dashed polygon portrays the loca- tion of the Montseny area. **b** Density map of the landslides triggered by the Gloria storm and gathered in the inventory. The black crosses represent the landslide points of the ICGC and the #Esllavicat inven- tories. The main rivers are represented as blue lines. The location of Barcelona is indicated with a black circle

methods analyse how the evaluation results change when relaxing the condition of co-localization between simulations and observa- tions (i.e. warnings and landslide inventory points).

Having landslide inventories that are complete in space and time is crucial to establish reliable LEWS and to evaluate their performance. Historically, landslide inventories were collected focusing on small areas from the interpretation of aerial photo- graphs, remote sensing data, field surveys, and local reports (Galli et al. 2008; Guzzetti et al. 2012). Alternatively, inventory data can be obtained from data sources such as newspapers reports, and crowdsourcing (Guzzetti et al. 1994; Kirschbaum et al. 2010; Ekker et al. 2013; Juang et al. 2019). However, these inventories are often incomplete and usually biased to landslides that affected urban areas or infrastructures (Ardizzone et al. 2002).

The large number of landslides that were reported during the Gloria storm gives us a unique opportunity to analyse the perfor- mance of the existing landslide early warning system for the region of Catalonia. To do so, we propose to apply a fuzzy verification method widely employed for the verification of precipitation fore- casts using several neighbouring window sizes. The objectives of the study are (i) to analyse the Gloria storm rainfall event and the landslides that were triggered, and (ii) to assess the performance of the LEWS during the Gloria storm and deduce the extent of the LEWS skill using an inventory which has spatial and temporal uncertainties.

Settings

Geology and climate of Catalonia

Catalonia is located in the NE of the Iberian Peninsula and covers an area of around 32,000 km². Its orography (Fig. 1a) is the result of (i) the formation of the Pyrenees with peaks over 3000 m asl., the Catalan Coastal Range, and the Iberian Range during the Paleogene; (ii) the later deposition of sediments in the Ebro Basin; and (iii) the formation of a series of horst and

grabens more or less parallel to the coastline during the Miocene (Berastegui et al. 2010). The bedrock lithology is very diverse and includes igneous, sedimentary, and metamorphic materials. In many areas, the bedrock is covered by surficial formations of varied thickness. While in some areas these deposits merely consist of a few centimetres, in others, the surficial formations can be very thick, of the orders of metres. The lithology of these surficial deposits is also very variable and can be very different from one area to another.

Catalonia's climate varies, but can be classified as Mediterranean (Emberger 1952). Near the coast, the weather is mild and temperate, with a mean annual temperature of 17 °C. Inland, the climate is continental with cold winters, hot summers, and less abundant precipitation. The Pyrenees present a high-altitude climate with abundant snow and temperatures below 0 °C during winter. The rainiest seasons are generally spring and autumn, except for the Pyrenees, where the rainiest season is summer. The mean annual rainfall ranges from less than 400 mm in some parts of the Ebro Basin to over 1200 mm in the Pyrenees. In Catalonia, the 10-year return period 24-h rainfall accumulation commonly exceeds 100 mm (Clavero et al. 1996). Daily accumulations of over 200 mm can be regularly seen at least once a year in the coastal area (Martín Vide and Olcina Cantos 2001). The Gloria storm was a rather unusual event of heavy rains during the driest months of the year.

Landslides are generally triggered by either (i) convective rain- fall events with high intensities which cause widespread shallow slides and debris flows, typical from mid-summer to early autumn, or (ii) long-lasting rainfalls with moderate intensities which usually trigger or reactivate earth flows and mid-size slides, common dur- ing spring and autumn (Corominas et al. 2002, 2015; Abancó et al. 2016). The Gloria storm rainfalls happened during winter, but still triggered a significant number of landslides.

LEWS description

Herein, we briefly describe the LEWS for the region of Catalonia. More details can be found in Berenguer et al. (2015) and Palau et al. (2020). The LEWS has been designed with the aim of issuing real-time warnings to the authorities in charge of managing landslide risk in Catalonia. It is running pre-operationally at the university servers for testing purposes. The LEWS combines two input parameters: (i) a 30-m-resolution susceptibility map (Fig. 2a) and (ii) high-resolution rainfall observations. The output of the LEWS is updated every time new rainfall observations are available and consists on a map showing a qualitative warning level.

The susceptibility map (Fig. 2a) is used to depict the locations where landslides may occur. It was derived by Palau et al. (2020) applying a fuzzy logic methodology to combine slope angle and land use and land cover information.

To assess if a rainfall event has the potential of triggering a landslide, the intensity–duration–frequency (IDF) curves of the Fabra meteorological observatory in Barcelona (Casas et al. 2004) are used to define four rainfall hazard levels (Fig. 2b). To separate different rainfall events, an inter-event period of 6 h without rainfall is used. Neither the antecedent rainfall nor soil moisture conditions are not taken into account in the current version of the LEWS.

Finally, the rainfall hazard and the susceptibility are combined through a warning matrix. The result is a 30-m gridded warning level map. Each warning level ('very low', 'low', 'moderate', and 'high') indicates the possibility that a landslide is triggered at a specific location. Additionally, a summary showing the maximum warning level computed within the first second- and third-order hydrological sub-basins as defined by Strahler (1957) is provided.

Additional analysis of recent rainfall events that triggered landslides in Catalonia showed that the rainfall intensity–duration (I-D) thresholds initially applied to determine the 'Moderate' and 'High' warning levels were too low. Therefore, here we have adapted the I-D thresholds employed in Palau et al. (2020). The 5-year and 20-year return period I-D curves have been used to define the 'Moderate' and 'High' rainfall hazards respectively.

Description of the Gloria storm

From 20 to 23 January 2020, the Gloria storm affected the region of Catalonia, causing several different hazards such as storm surges, erosion of beaches in coastal areas, floods, and landslides. According to the Catalan Office of the Climate Change (Canals and Miranda 2020; OCCC 2020), the economic losses due to these impacts exceeded 500 million euros. The Gloria storm was exceptional, because it took place during winter, an unusual season for torrential rainfalls in this area, and also because of its long duration.

This section first presents the meteorological situation and analyses the rainfall accumulations. Then, the landslides triggered by the Gloria storm landslide inventories are described. Finally, the LEWS Gloria storm outputs are studied.

Meteorological situation

The Gloria storm was preceded by an anticyclone that lasted over a month, during which it did not rain in Catalonia. On 18 January 2020, a cold front coming from the North Atlantic entered through the north west of the Iberian Peninsula and moved south towards the Mediterranean Sea. On the British Isles, an unusual anticyclonic situation recorded pressures up to 1050 hPa, the highest pressure since 1957 (Servei Meteorològic de Catalunya 2020a). This high pressure had an elongated shape from east to west and covered a large part of central Europe.

The Gloria storm was the result of the combination of the unusual high pressures on the British Isles and the low pressures located on the south of the Iberian Peninsula. The gradient of pressures between these two centres caused strong east-northeast winds and, provided a high humidity and abundant and widespread precipitation (Servei Meteorològic de Catalunya 2020b). The duration of the Gloria storm was long because the North Atlantic cold-air mass was stationary over Catalonia for several days.

Precipitation analysis

The rainfall datasets used in this study consist of the measurements of 187 tipping bucket rain gauges from the Meteorological Service

Fig. 2 Susceptibility map (**a**) and rainfall intensity-duration thresholds (**b**) employed by the LEWS
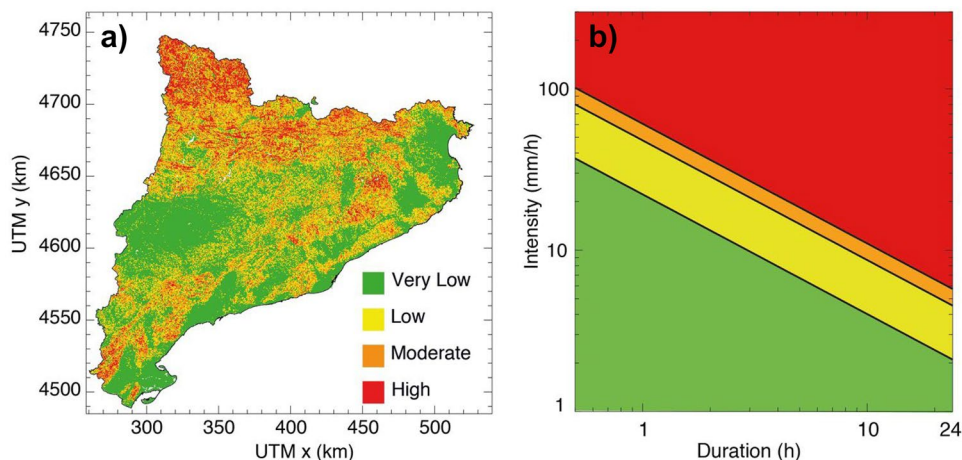
of Catalonia (SMC), and the quantitative precipitation estimates (QPEs) from the composite of the observations of the SMC radar network (XRAD). The location of the rain gauges and the radars is portrayed in Fig. 1a.

Radar QPEs have been produced from the volume scans of Creu del Vent, La Panadella, and Puig d'Arques C-band single-polarisation Doppler radars of the SMC with the integrated tool for hydrometeorological forecasting (EHIMI, Corral et al. 2009). The EHIMI tool includes a chain of quality control, correction, mosaicking, and accumulation algorithms to generate QPE products from raw radar observations. The product used here is the 30-min precipitation accumulation field with a spatial resolution of 1 km.

Rain gauge measurements and radar observations have been combined to obtain an improved QPE applying the method proposed by Velasco-Forero et al. (2009) and Cassiraga et al. (2020). This method employs a geostatistics technique known as kriging with an external drift (KED) to interpolate the rain gauge observations using radar rainfall as a secondary variable that provides the drift to the rainfall field between rain gauges. As shown by Velasco-Forero et al. (2009), this method benefits from the direct surface rainfall observations of the rain gauges located within the study area, and the radar description of the spatiotemporal variability of the rainfall field.

Figure 3 presents the daily precipitation accumulations from 20 to 23 January 2020. The evolution of the Gloria storm and the spatiotemporal variability of the rainfall field can be observed in these plots. It also shows the locations of landslide reports in relation to the rainfall.

The storm began on 20 January 2020 when snow and rain were observed in the northeast and the south (Fig. 3a). On 21 January 2020, precipitations fell over the entire region. Still, they were more abundant parallel to the coastline, where the 24-h rainfall accumulations exceeded 200 mm at the Montseny area and the Iberian range (Fig. 3b). During 22 January 2020, rainfall fell intermittently over most of Catalonia. More than 140 mm were accumulated in the Montseny area (Fig. 3c). Additionally, important rainfall accumulations were recorded at the southwest of Catalonia, the Pyrenees, and the Pre-Pyrenees. The main precipitation system moved towards the north during the morning of 23 January 2020. Rainfall fell intermittently with moderate and low intensities. Although rainfall accumulations were not as relevant as the previous days (Fig. 3d), they were still significant in the Montseny area, where over 100 mm were recorded in some areas, and in the Pre-Pyrenees.

The total accumulated rainfall during the 4 days was significant over most of Catalonia (Fig. 4). The largest rainfall amounts fell over the north-east, with around 480 mm in the Montseny area.

During the first day of the Gloria storm, no landslides were reported. In the following days, the areas that recorded the highest rainfall accumulations coincide rather well with the places where landslides were reported (black circles in Figs. 3 and 4).

Analysis of the quality of the precipitation estimates

This section presents an analysis of the quality of the precipitation estimates obtained applying the KED method. The performance has been evaluated by leave-one-out cross-validation using the observations at the rain gauges as the reference. To do so, we have applied the KED method removing one of the rain gauges from the calculation to estimate the rainfall at the location of the removed rain gauge. Then, we have compared the estimated value with the observed rainfall. This process has been repeated for every 30 min and each of the 187 considered rain gauges.

Figure 5 shows the comparison between the total precipitation accumulations during the event obtained from cross-validation and the total accumulations observed at each of the rain gauges. Additionally, four statistics have been added to the scatter plot: the bias, the standard deviation of the error (SD error), the root mean squared error (RMSE), and the root mean squared relative error (RMSR). The event KED estimates generally show a good agreement with the event accumulations recorded at rain gauges. The SD of the error and the RMSE are similar, around 31 mm; therefore, the bias is rather low. And the event RMSR is 22%.

The results from the comparison of the hyetographs obtained by cross-validation and the hyetographs from rain gauge observations for the four selected rain gauges distributed over the Catalan territory (see red circles in Fig. 1) are presented in Fig. 6. The evolution of 30-min accumulations reproduces the observations satisfactorily at the majority of the rain gauges. However, in some locations (e.g. Viladrau and PN dels Ports), the KED underestimates the measured intensities. In other sites, such as Torroella de Fluvià, the KED slightly overestimates the observed rainfall. The results for the 30-min accumulations obtained at all the available rain gauges show that the RMSE ranges between 0.15 and 1.72 mm. In the calculation of the RMSR, we have imposed a threshold of 1 mm/30 min, and the results for RMSR range between 21.7 and 107.4%, with a median value of 42.8%. The errors in small accumulations have a significant effect in the calculation of RMSR, and the larger values are obtained in areas with event accumulations between 100 and 150 mm.

The results presented in this section quantitatively describe the uncertainty in the QPEs obtained by KED during the Gloria storm. These QPEs are the precipitation inputs to the Catalonia region LEWS and, therefore, their uncertainty will affect the performance of the LEWS and the quality of the warnings during this event (see the "Analysis of the LEWS outputs" section).

Landslide inventory and impacts

The significant rainfall accumulations and high intensities registered during the Gloria storm triggered a large number of landslides over different areas in Catalonia. One of the main challenges for the evaluation of the performance of LEWS is the availability of a complete landslide inventory. In Catalonia, no systematic and official landslide inventory exists. Therefore, in this study we have used information contained in two different landslide inventories: the inventory of the Cartographic and Geological Institute of Catalonia (ICGC inventory; González et al. 2020), and the #Esllavicat inventory.

The ICGC inventory gathers landslide information from several sources such as reports from different administrations (municipalities, county councils, civil protection, mountain rangers, and other institutions), interpretation of aerial photographs taken after the Gloria storm along some river banks, and media reports. It includes a total of 348 entries. However, information of these 348 landslides is not complete and many times lacks details. For example, the ICGC inventory does not include volume information. The majority
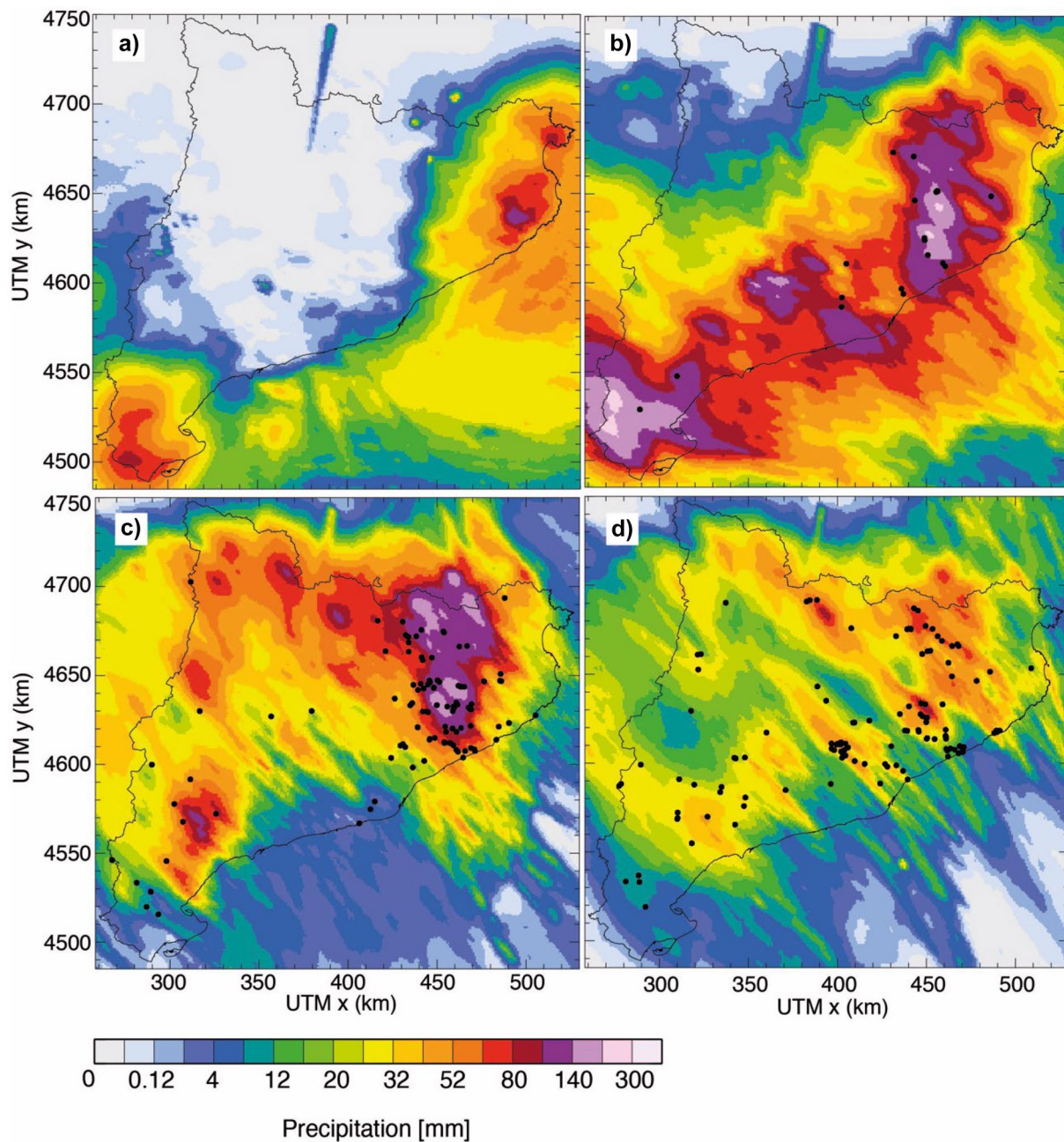
Fig. 3 Daily rainfall accumulations during the Gloria storm: **a** 20 January 2020, **b** 21 January 2020, **c** 22 January 2020, **d** 23 January 2020. Black circles represent the landslides included in the inventory each day. In the following sections, more details about the inventory and the landslides are given

of landslides are classified according to the Varnes (1978) classifications. Yet, some reports may be due to accumulation of sediment on roads associated with other processes such as water erosion. All the entries of the ICGC inventory include information on the date. However, some entries have no clear date and the day of occurrence during the Gloria storm is therefore uncertain. Additionally, the location of around 25% of the reports is uncertain, and 23 landslides are located in urban areas in flatlands, where no slope or talus could be observed in their vicinity. Therefore, these points have not been used for our analysis since we considered their spatial uncertainty was too large.

The #Esllavicat inventory collects data from social network posts of local observers. A total of 108 geolocated landslides were reported through social networks and have been included in the #Esllavicat inventory. The majority of #Esllavicat landslide reports included a photograph or video of the initiation or deposit area (see examples in Fig. 7). Most of the landslide locations have been checked by pre-storm Google Street View. Using this information,

Fig. 4 Accumulated rainfall from 20 January 2020 00:00 to 23 January 2020 24:00. The black circles represent the landslides included in the ICGC and the #Esllavicat inventories
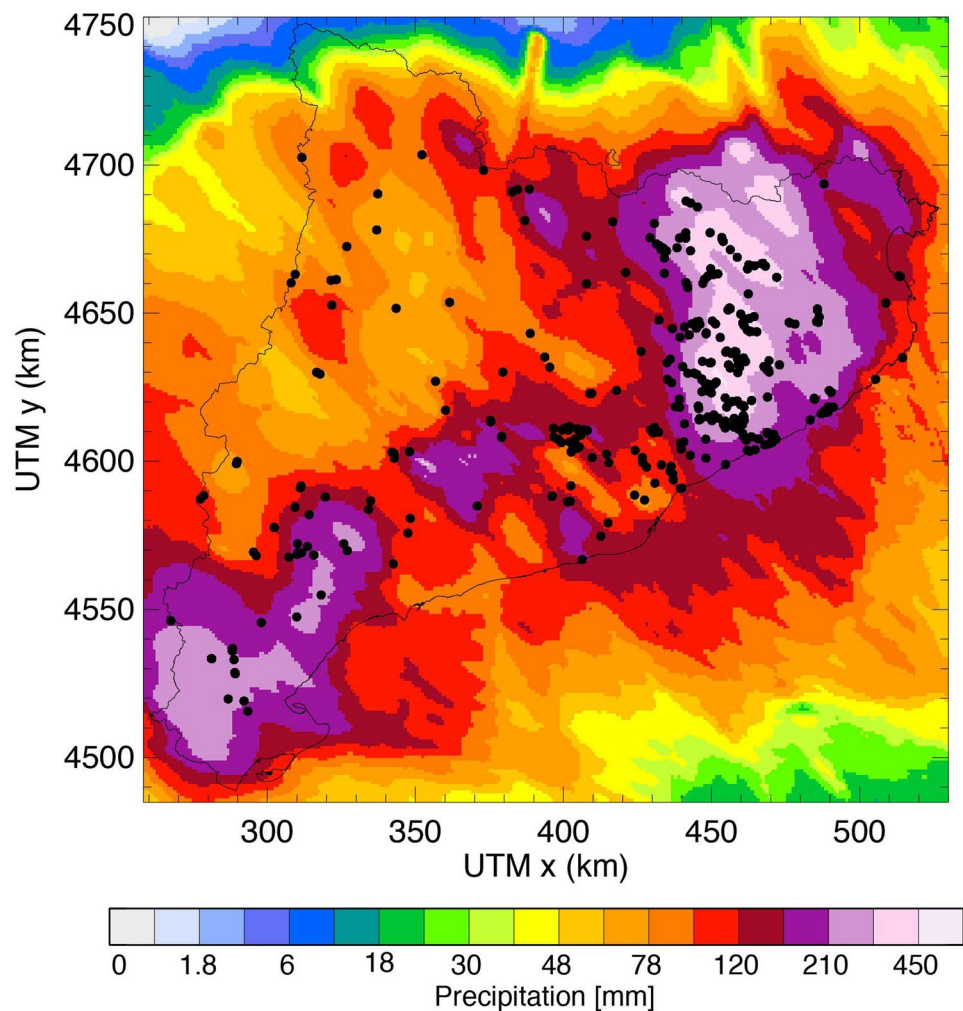
together with the descriptions provided in some posts, the landslides have been classified into different types according to the classifications proposed by Varnes (1978) and Hungr et al. (2014). Additionally, a measure of the event size has been assigned to each inventory entry to differentiate between three volume ranges: less than 1 m³, between 1 and 10 m³, and more than 10 m³. Since #Esllavicat reports were made by population, most landslides happening overnight were informed during the morning. Some other reports were made once the storm had ceased, and therefore, the precise triggering date is uncertain.

For this study, the ICGC and the #Esllavicat inventories have been merged, and duplicated points have been removed. The final landslide data set contains 108 points from the #Esllavicat inventory and 275 from the ICGC inventory, resulting in a total of 383 landslide points.

The Montseny is the area where the largest density of landslides was observed, 0.28 landslide/km² (Fig. 1b). This density is rather low, compared with the density of landslides observed for historical MORLEs in Catalonia (e.g. 1.5 landslides/km² in the Pyrenees 1982 (Corominas and Alonso 1990), and 1.16 landslides/km² in Val d'Aran 2008 (Shu et al. 2019). The differences may be partly due to the completeness of historical inventories, which fully covered smaller regions inside Catalonia with field surveys and the interpretation

of aerial photographs. This was not possible for the Gloria storm inventory due to the much larger extension and because no post-event flight surveillances were made over the most affected areas.

The characteristics of the landslides triggered by the Gloria storm and contained in the final inventory are described hereunder. The accumulated rainfall at the location of the reported landslides has been checked (Fig. 4). From the 383 landslides used for this study, more than 100 were reported in places that registered event rainfall accumulations over 300 mm in 96 h (Fig. 8a).

According to its type, 69% of the inventoried landslides were slides, 20% falls, and 3% flows (Fig. 8b). The type of the remaining 8% of events triggered by the Gloria storm is unclear. The inventories do not have enough information to determine what type of material or sliding mechanism was more predominant during the Gloria storm. Regarding the landslide volume, only information from the #Esllavicat reports is available. Fourteen percent of the landslides contained in the #Esllavicat inventory were small, with a volume of less than 1 m³. The volume of 56% of the #Esllavicat landslide reports ranged between 1 and 10 m³. Only 12% of the #Esllavicat reports correspond to landslides with volumes larger than 10 m³. The volume of 18% of the entries could not be established because not enough information was given in the report.
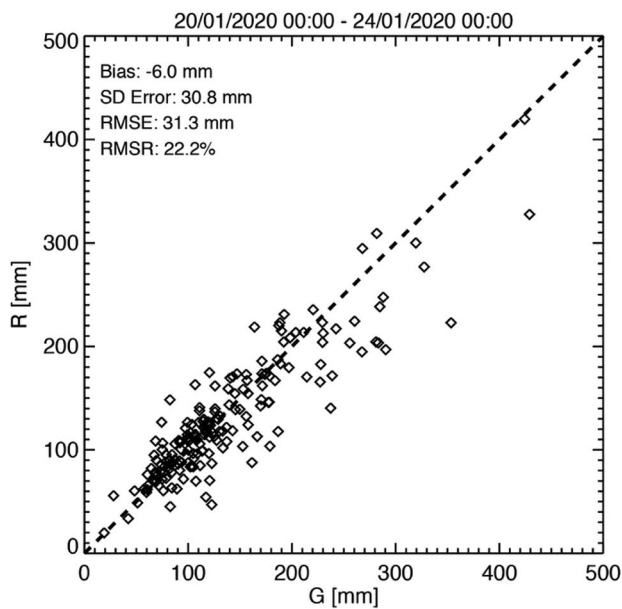
Fig. 5 Cross-validation scatter plot comparing the observed total accumulated rainfall at each of the 187 rain gauges (R) and the KED estimated value from the radar observations (G)

The 5-m-resolution DEM has been used to estimate the slope angles, and the 30-m-resolution DEM has been employed to obtain the orientation (ICGC 2013). Similarly, land use and land cover with a resolution of 30 m (MCSC-4, (CREAF 2009)) and the graph of the Catalonia infrastructures network (DGMT 2019) have been applied to analyse the most common land use and land cover classes at the landslide locations and the proximity to roads and railway lines.

The majority of landslides were located at steep slopes of over 20° (Fig. 8c). Around 27% of the events were reported in slopes with angles between 10 and 20°, and about 16% in gentle slopes with slope angles less than 10°. Such low slope angles are rather difficult to justify from a geotechnical point of view and may be related to spatial uncertainty. No clear trend can be observed in the orientation of the slopes where landslides were reported. However, the total number of events triggered in east, north-east, and south-east facing slopes is slightly larger than the sum of the events at south, south-west, and west facing slopes (Fig. 8d). The main wind direction of the Gloria storm was towards west-north west; thus, east and south-east facing slopes would be the most exposed.

Landslides most frequently occurred in forest areas (Fig. 8e) and 55 events were reported in areas with infrastructures or buildings. Two of the landslides contained in the inventory were located in water bodies, which might be related to the scouring in river banks. Most of the reported landslides were spotted close to linear infra-structures (Fig. 8f). Around 64% were triggered between 0 and 10 m away from the road or railway axis. The number of landslide reports diminishes with the distance from linear infrastructures. Only 38% of the reported landslides were located further than 200 m. These results provide two conclusions: (i) more than half of the reports were related to slope failures of road cuts and embankments in linear infrastructures, and (ii) landslides happening in remote inhabited areas may generally be unreported.

Analysis of the LEWS outputs

The LEWS has been run from 20 to 23 January 2020 using the KED 30-min rainfall accumulation estimates as inputs. Since the landslide inventory only has information on the reporting date, the correspondence between the 30-min warning outputs and the landslide reports could not be studied. Here the daily maximum warning level has been used to analyse the quality of the warn-ings computed each day of the Gloria storm.

Figure 9 shows the sub-basin maximum warning level sum-mary of each of the days of the Gloria storm and the positions of inventory reports. From the comparison of the warning maps of Fig. 9 and the 24-h rainfall accumulations of Fig. 3, it can be observed that generally, 'Moderate' and 'High' warnings were obtained in the areas that recorded the most significant rainfall accumulations during the corresponding day.

Generally, landslides (displayed as black circles in Fig. 9) were reported in places where the sub-basin daily warning summary was 'Moderate' or 'High'. At the eastern half of Catalonia, 'High' warnings were displayed over the area where the inventory has the highest density of landslides (Fig. 1b). 'Moderate' and 'High' warnings were obtained over the south-west of Catalonia on 21 January and over the north-west of Catalonia on 22 January 2020, but few landslides were reported in these areas (Fig. 9b, c). The Pyrenees, Pre-Pyrenees, Iberian Range, and the western Catalan Coastal Ranges have a low population density. Therefore, it may be the case that some landslides might have been unreported.

Evaluation of the performance of the LEWS during the Gloria storm

Evaluating the performance of a high-resolution LEWS over the entire Catalonia is challenging because of the spatial and tempo-ral uncertainties of the landslide inventory as well as its incom-pleteness (see the "Landslide inventory and impacts" section). Traditional verification methods match the location and time of the warnings to the precise location and time of the reported landslides to analyse the performance of a LEWS. Consequently, the uncertainties and incompleteness of landslide inventories have an effect on the results of traditional verification methods.

Fuzzy verification methods are an alternative that does not require the exact coincidence between warnings and observa-tions. Instead, such methods assume that the location and time of warnings can be close to the location and time of landslide observations but still be useful (Ebert 2008, 2009). To do so, fuzzy verification methods look in space–time neighbouring windows around each observed event to evaluate the performance of the model, granting some flexibility in the matching between the prediction and the observation.

Fuzzy verification techniques have been widely applied to measure the performance of high-resolution mesoscale precipita-tion forecasts (Damrath 2004; Theis et al. 2005; Ebert 2008; Clark et al. 2010). In many cases, fuzzy verification methods analyse

Fig. 6 Observed (black line) and estimated (blue line) hyetographs from cross-validation for four rain gauges from 20 January 2020 00:00 to 23 January 2020 24:00. The location of the rain gauges can be observed in Fig. 1. The time step is 30 min. We have imposed a 1 mm/30 min threshold for the computation of the mean relative errors. Therefore, the relative error has been computed for time-steps when the observations measured at least 1 mm/30 min. G states for the total event rainfall accumulation (Accum) recorded at each rain gauge. R refers to the estimated total event rainfall accumulation at each rain gauge location
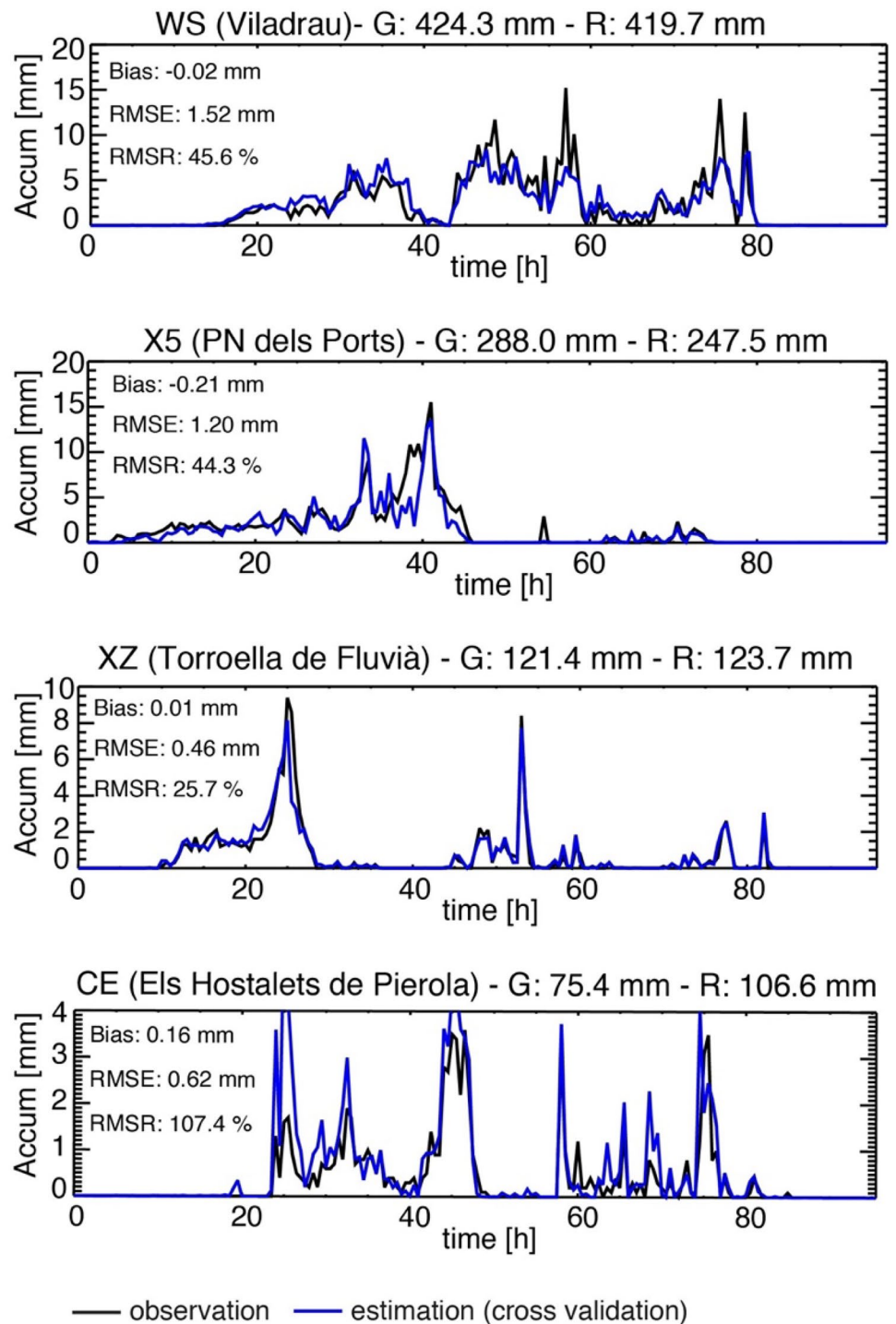
the effect of varying the size of the neighbouring windows. The results can be exploited to determine at which scales the forecasts should be used to meet certain accuracy requirements (e.g. Ebert 2008, 2009). In this section, we have applied a fuzzy verification method that is used in meteorology to evaluate the performance of the Catalonia region LEWS for the Gloria storm period, and deduce the scales for which the warnings are reliable.

Description of the verification method

The fuzzy verification method that has been applied for the evaluation of the 30-m-resolution warnings during the Gloria storm is known as the 'minimum coverage criterion' (Damrath 2004; Ebert 2008). This method supposes that the location and time of the observations are correct, and considers a neighbouring window

Fig. 7 Examples of landslides triggered by the Gloria storm in the Montseny area. **a** Rotational slide in a colluvium slope (photo courtesy of Clàudia Abancó). **b** Rotational slide that affected a road embankment and parts of natural slopes (photo courtesy of Roger Ruiz)

around each observation to search for warnings. The minimum coverage criterion method assumes events are equally likely to occur anywhere within the neighbouring window. Then, categorical scores based on the contingency table are employed for the verification of the warnings (Fawcett 2006).

For the verification purposes, we have considered that a warning has been given when the warning level was either 'Moderate' or 'High', and no warning has been given when the warning level was either 'Low' or 'Very Low'. The landslides contained in the inventory have been used as reference. Following the minimum coverage criterion, the pixels that fall inside the neighbouring windows are used to assess the true positives and the false negatives (Fig. 10): A true positive is an outcome where the LEWS correctly displays at least one warning within the neighbouring window of a landslide observation. In contrast, a false negative is an outcome where the LEWS incorrectly displays no warning within the neighbouring
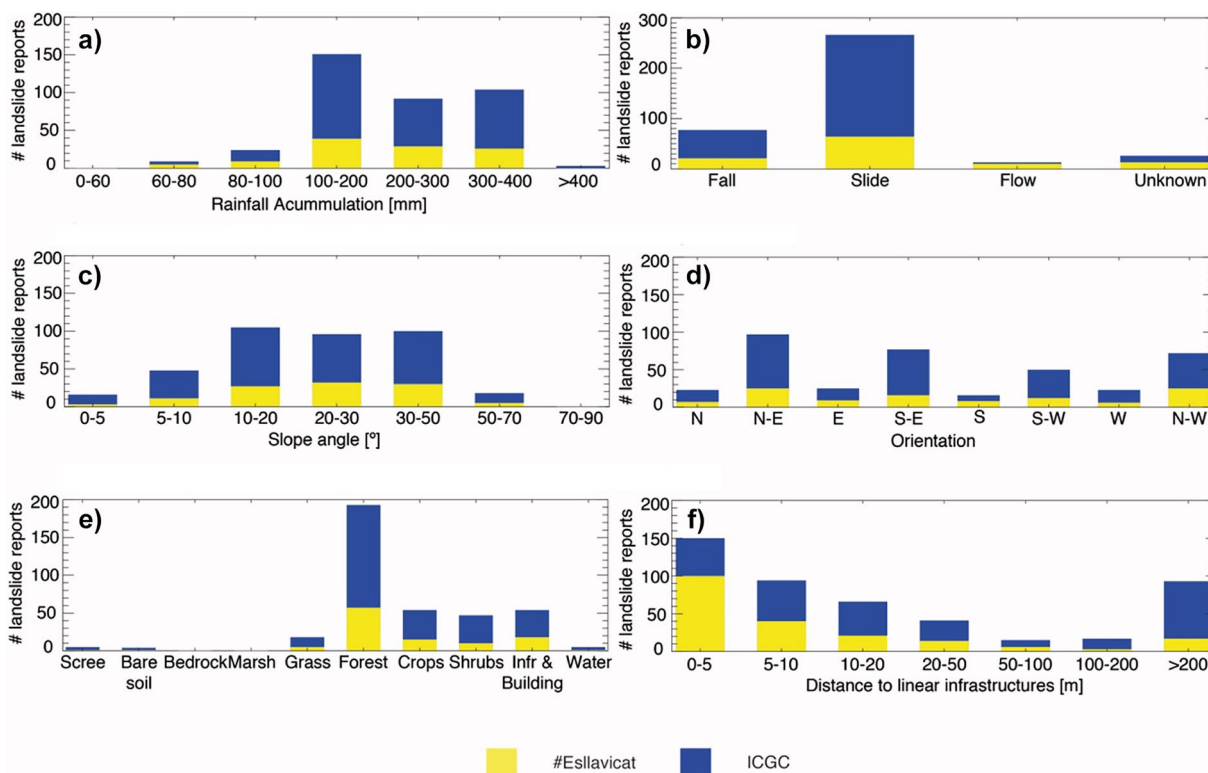
Fig. 8 Histograms showing the distribution of Gloria storm landslide reports contained in the ICGC and the #Esllavicat inventories according to **a** rainfall accumulation, **b** landslide type, **c** slope angle, **d** orientation, **e** land use and land cover, and **f** distance to the closest road or railway line axis

Fig. 9 Daily maximum warning level sub-basin summary. **a** 20/01/2020. **b** 21/01/2020. **c** 22/01/2020. **d** 23/01/2020. The black circles represent the landslides contained in the inventory
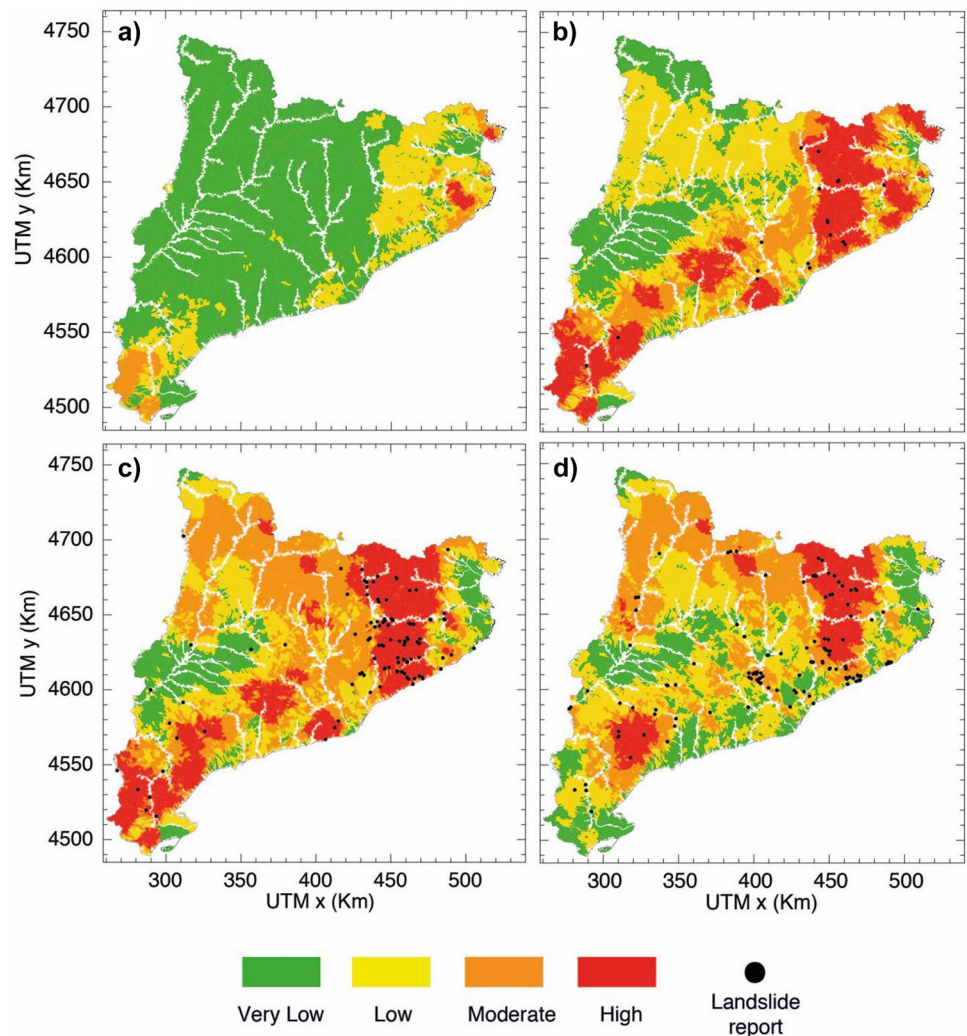
window. The pixels that fall outside the neighbouring windows are used to calculate the false positives and the true negatives (Fig. 10): A true negative is an outcome where the LEWS correctly computes no warning at a grid cell that falls outside the neighbouring windows of landslide observations. And a false positive is an outcome when the LEWS incorrectly computes a warning at a grid cell that falls outside the neighbouring windows of landslide observations.

To study the performance of the LEWS, we have selected three different metrics:

(i) The true positive rate (TPR),

$$TPR = \frac{TP}{TP + FN} \qquad (1)$$

(ii) The false positive rate (FPR)

$$FPR = \frac{FP}{FP + TN} \qquad (2)$$

(iii) And, the true skill statistic (TSS)

$$TSS = TPR - FPR = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \qquad (3)$$

where TP, FP, and TN are, respectively, the number of true positives, false positives, and true negatives. TPR and FPR values range from 0 to 1. Ideally, a LEWS should have no false negatives and no false positives; therefore, the perfect TPR and FPR should be 1 and 0 respectively. The TSS combines the TPR and the FPR. It measures how well the warning map can separate points with landslide observations from points where landslides have not been observed. Its scores range from −1 to 1. Since an ideal LEWS would have a TPR equal to 1 and a FPR equal to 0, the best TSS score is 1 and a score of 0 indicates no skill.

Next, the minimum coverage method has been applied considering different neighbouring window sizes, which provide additional information on the quality, and space and time representativeness of the warnings. Here, we have used 30 m, 500 m, 1 km, 2 km, and 10 km squared neighbouring windows around each landslide observation. Additionally, we have included two types of polygon neighbouring windows that we considered of special interest: hydrological sub-basins and municipalities. As explained in the "LEWS description" section, hydrological sub-basins are used for the regional visualisation of LEWS outputs. Its mean area and standard deviation are 2.1 km² and 1.6 km², respectively. Municipalities have been chosen since they are relevant from an emergency

Fig. 10 Fuzzy verification at an area of Catalonia on 23 January 2020 applying **a** 500 m, **b** 1 km, and **c** sub-basins neighbouring windows. The black diamonds represent the location of the landslide observations. The polygons and squares are the neighbouring windows. TP stands for true positive, and FN for false negative. All the grid cells with 'Moderate' or 'High' warnings outside the neighbouring windows are false positives. The pixels outside the neighbouring windows with 'Very Low' and 'Low' warnings are true negatives. Note that more pixels with moderate or high warning fall within the neighbouring windows as the neighbouring window size increases. Thus, the number of true positives increases and the number of false positives decreases
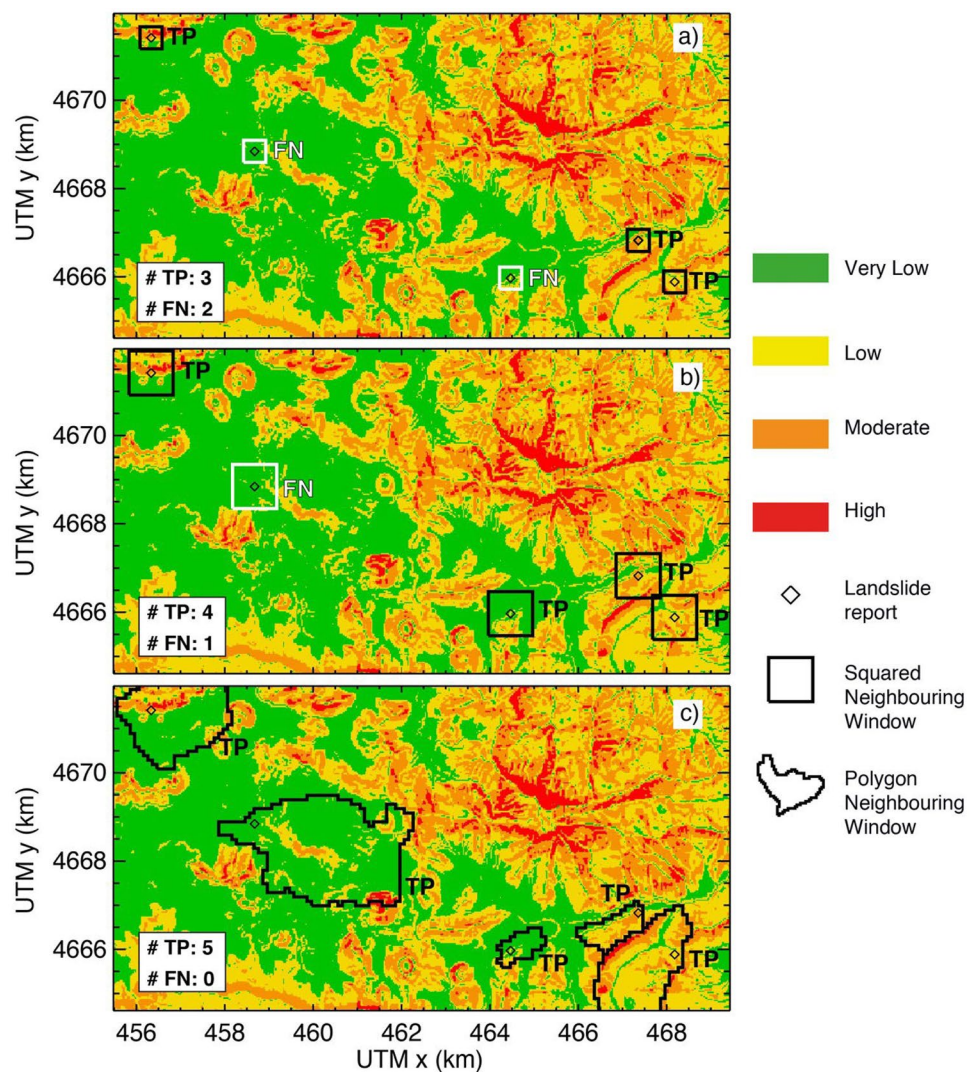
management point of view. The area of municipalities is very variable. Its mean area is 26.9 km², and the standard deviation is 30 km². Indeed, the largest municipality has an area of 303 km².

The landslide inventory only has information on the reporting date, many landslides triggered overnight were reported during the morning, and 138 landslides were reported after the end of the storm. Therefore, two different time windows have been applied for the verification of the LEWS outputs. First, a time window of 48 h comprising the day of the landslide report and the day before has been used to overlook the time uncertainties of the landslides that were triggered overnight. Second, a time window of the entire duration of the Gloria storm has been employed for the event verification of the warnings to include the landslides that were reported after the end of the storm.

The results of fuzzy verification indicate the performance of the LEWS as a function of the neighbouring window size. The application of the largest space and time neighbouring windows is subject to granting large flexibility in the matching between the warnings and the observations. Since it is easier to find grid cells with warning within a larger domain (Fig. 10), we can expect the TPR to increase when using larger neighbouring windows.

Conversely, the FPR is expected to decrease because the number of pixels with warning outside the neighbouring windows decreases when increasing the neighbouring window size (Fig. 10). Finally, the verification results for the different neighbouring windows have been compared to find the scale for which the change in the LEWS performance is the most significant. In the verification of high-resolution precipitation forecasts, this scale is interpreted as the effective resolution of the model for which the forecasts are most representative (Damrath 2004; Ebert 2008). By analogy, we can relate this scale to the effective resolution of the LEWS.

Performance evaluation using a 48-h window

First, the minimum coverage criterion has been implemented using as a reference the 245 inventory entries that had a specific reporting date. Thus, for every landslide observation, each of the different space neighbouring windows has been jointly applied with the time window of 48 h. The time window has been used to check for the warnings displayed within the space neighbourhood windows during the day of the landslide report and the day before.

Table 1 Skill scores obtained using the 48-h time window for the different spatial neighbouring window scales. True positives (TP), false negatives (FN), false positive (FP) area, true negative (TN) area, true positive rate (TPR), false positive rate (FPR), and true skill statistic (TSS) for the different neighbouring window types

Neighbouring window size	Date [dd/mm/yyyy]	Number of events	TP	FN	FP-area [km²]	TN-area [km²]	TPR	FPR	TSS
30 m	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	10	10	3176	36,367	0.50	0.08	0.41
	22/01/2020	103	48	55	6119	33,424	0.46	0.15	0.31
	23/01/2020	122	46	76	6392	33,151	0.37	0.16	0.21
500 m	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	12	8	3176	36,367	0.60	0.08	0.51
	22/01/2020	103	56	47	6119	33,424	0.54	0.15	0.38
	23/01/2020	122	54	68	6392	33,151	0.44	0.16	0.28
1 km	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	17	3	3168	36,351	0.85	0.08	0.76
	22/01/2020	103	86	17	6075	33,347	0.83	0.15	0.68
	23/01/2020	122	108	14	6348	33,054	0.88	0.16	0.72
2 km	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	19	1	3151	36,304	0.94	0.07	0.87
	22/01/2020	103	91	12	5960	33,149	0.88	0.15	0.73
	23/01/2020	122	113	9	6240	32,813	0.92	0.15	0.76
10 km	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	20	0	2681	35,110	1.00	0.07	0.92
	22/01/2020	103	102	1	4149	29,198	0.99	0.12	0.86
	23/01/2020	122	122	0	4380	28,263	1.00	0.13	0.86
Sub-basins	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	18	2	3149	36,329	0.89	0.07	0.82
	22/01/2020	103	89	14	5995	33,249	0.86	0.15	0.71
	23/01/2020	122	112	10	6273	32,930	0.91	0.16	0.75
Municipalities	20/01/2020	0	0	0	200	39,344	–	0.00	–
	21/01/2020	20	20	0	2972	35,773	1.00	0.07	0.92
	22/01/2020	103	100	3	4974	31,684	0.97	0.13	0.83
	23/01/2020	122	118	4	5143	30,467	0.96	0.14	0.82

Table 1 shows the results for the fuzzy verification of the warnings issued with the LEWS for the Gloria event. As expected, the number of true positives increases with the size of the neighbouring window employed for the LEWS verification (Table 1). In contrast, the number of false negatives decreases when the neighbouring window size increases (Table 1). As a consequence, the TPR increases with the neighbouring window size (Fig. 11a). The worst TPR value is 0.37 for 30-m neighbouring windows on 23 January 2020, whereas its highest score is 1.00 for the 10-km neighbouring window during 21 and 23 January 2020, and for the municipalities neighbouring window during 21 January 2020. In fact, for the verifications applying large neighbouring windows (1 km, 2 km, and 10 km, sub-basins and municipalities), the TPR scores are generally high (> 0.83). This indicates that when using such neighbouring windows, a warning could be found at the surroundings of 83% of the landslide observations.

Fig. 11 Daily verification skill scores using the 48-h time window for the different spatial neighbouring windows. **a** True positive rate. **b** False positive rate. **c** True skill statistic
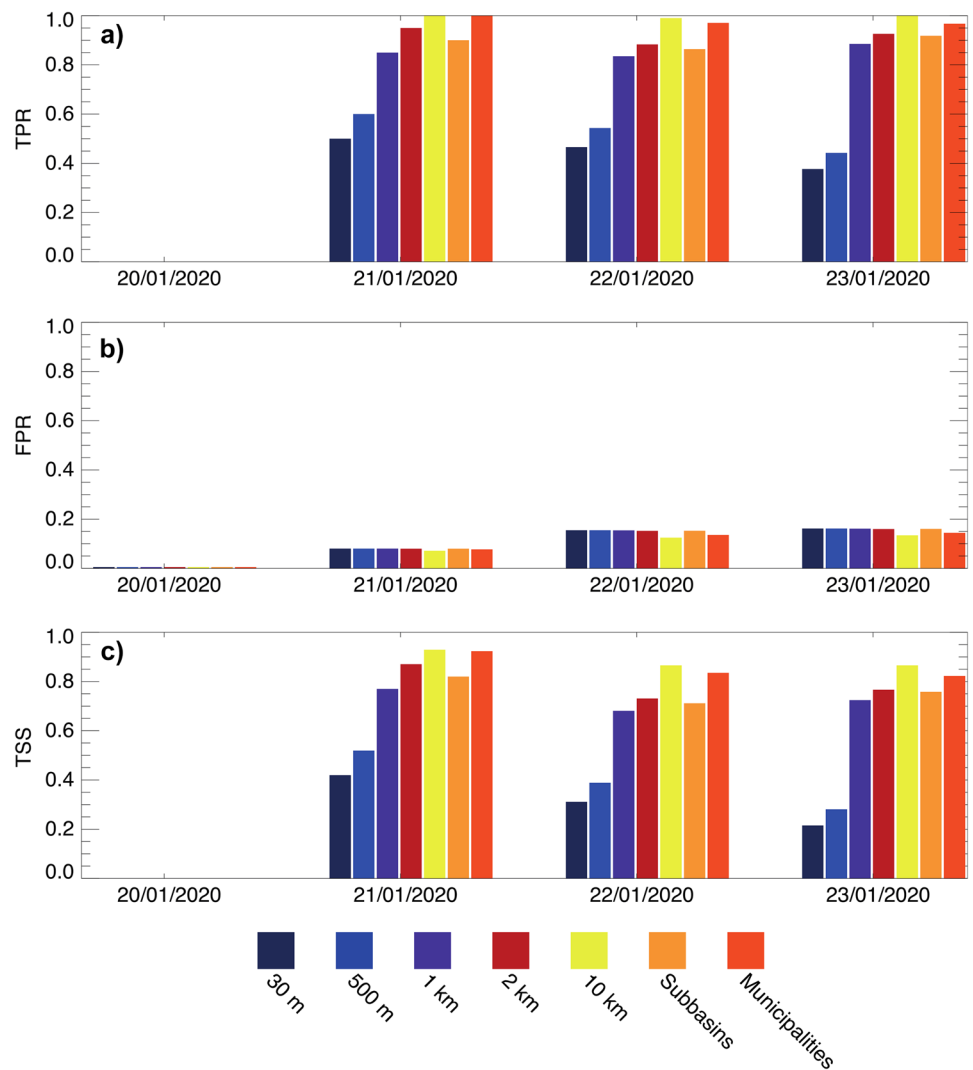
The results in Table 1 show that as foreseen the number of false positives decreases when increasing the neighbouring window size. Because of the incompleteness of the landslide inventory, especially in the less densely populated areas, the number of false positives is expected to be larger than if an ideal landslide inventory had been used. However, the false positive area is an order of magnitude smaller than the area where true negatives are displayed (Table 1). Therefore, the FPR values are rather low and range from 0.00 to 0.16 for all the neighbouring windows used (Fig. 11b).

Regarding the TSS, the highest score, 0.92, is achieved employing 10-km neighbouring windows for 21 January 2020 (Table 1 and Fig. 11c). Both the 30-m and 500-m neighbouring window verifications have relatively low TSS values, especially for 22 and 23 January 2020. In contrast, TSS values are always above 0.68 when using the larger neighbouring windows.

Additionally, Table 1 and Fig. 11a, b show that TPR and TSS scores are very similar for the verifications applying squared 1-km, 2-km, and sub-basin neighbouring windows. They also resemble the 10-km and municipality neighbouring windows.

These results are reasonable because the area of most sub-basins ranges between the area of 1- and 2-km neighbouring windows. The area of the largest municipalities is also similar to the area of 10-km neighbouring windows. A significant improvement has been observed in the skill scores when increasing the neighbouring window size from 500 to 1 km. Hence, if a warning is given, we would probably be able to find a landslide within a surrounding area of 1 km and a 48-h period after the warning is computed.
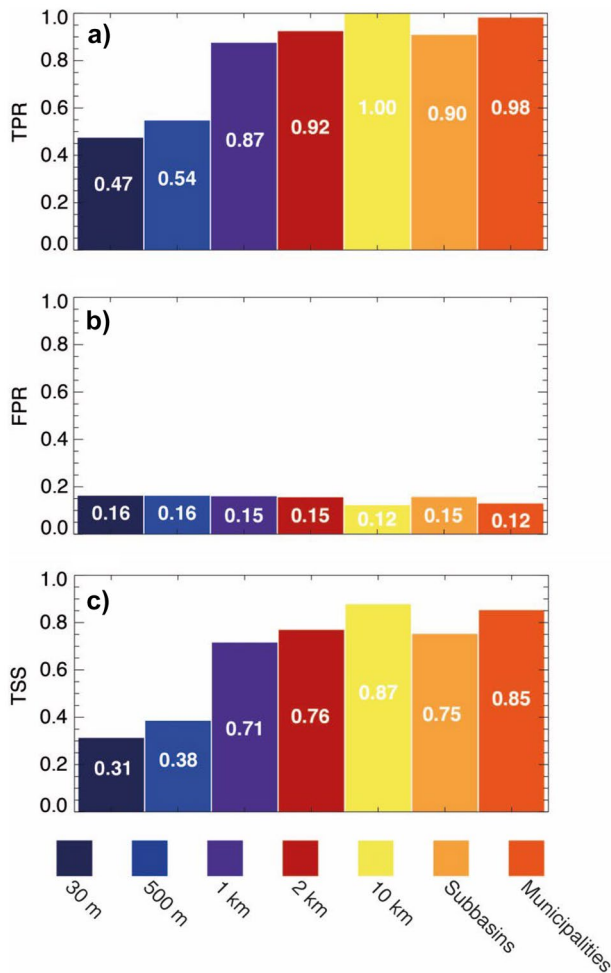
Performance evaluation using the entire storm duration window

Some landslide reports were made after the end of the storm. Thus, an event verification allows us to include the additional 138 landslides that were reported after the end of the storm. Here, all the 383 landslide events included in the landslide inventory have been used. The TPR, FPR, and TSS skill scores have been computed considering the maximum warning computed during the 96 h that lasted the Gloria event and applying the different spatial windows mentioned in the "Description of the verification method" section.

As expected, the obtained number of TP is larger than the obtained number of TP for the verification using 48-h time windows, and the number of FP is lower. As a consequence, the scores obtained when applying a 96-h neighbouring window for the LEWS verification improve. This is partly because the inventory employed for the verification using the entire storm duration includes a larger amount of landslide reports. Additionally, the uncertainties on the triggering time are more overlooked when using a longer time window.

As in the verification using 48-h time windows, the larger the spatial neighbouring windows, the higher the TPR and TSS are. Except for the 30-m and 500-m neighbouring window verifications, the TPR and TSS are relatively good with values above 0.87 and, 0.71, respectively (Fig. 12a, c). FPR values are rather low, around 0.15 for the evaluations using the different neighbouring windows (Fig. 12b).

The sub-basin neighbouring window verification achieves slightly higher TPR and TSS scores than the 1-km neighbouring window verification. Both TPR and TSS are also very similar for the verifications using 10-km and municipality neighbouring windows.

Fig. 12 Verification skill scores applying the 96-h time window and the different spatial neighbouring windows. **a** True positive rate. **b** False positive rate. **c** True skill statistic

As observed in "Performance evaluation using a 48-h window", the results of the verification using the 96-h time window show a significant improvement in the LEWS skill when increasing the size of the neighbouring window from 500 to 1 km. The verification results do not change significantly for the larger neighbouring windows (2 km, 10 km, sub-basins, and municipalities). Hence, if a warning is computed, we will probably be able to find a landslide within a surrounding area of 1 km². We could interpret this result as an effective resolution of the LEWS, at which the warnings are more reliable.

Discussion

The fuzzy verification method that has been applied for the verification of the warnings in the "Performance evaluation using a 48-h window" and "Performance evaluation using the entire storm duration window" sections is widely used in meteorology to analyse the performance of high-resolution mesoscale precipitation forecasts (Ebert 2008, 2009), and has been tested for the first time for LEWS. The largest change in the LEWS skill can be observed when increasing the neighbouring window size from 500 to 1 km. In the evaluation of precipitation forecasts, the most significant change in the forecast performance would determine the effective resolution of the model. However, since landslide inventories are not totally complete and suffer from spatial and temporal uncertainties, the extent in which the applied fuzzy verification method provides information on the effective resolution of the LEWS or the inventory resolution is not totally known.

Because neighbouring windows are used only around observations, and because it is easier to get a warning within a larger area, the analysis that has been conducted benefits larger neighbouring windows. As an alternative, a modified version of the minimum coverage criterion in which neighbouring windows are centred in each grid cell of the LEWS domain can be applied. By doing so, the domain is discretised in a series of homogeneous neighbouring windows that are used to compute not only the TP and FN, but also the FP and TN. A similar approach was adopted by Piciullo et al. (2017a) who applied homogeneous windows of a given size over their entire analysis domain.

We have also tested this approach to analyse the performance of the LEWS during the Gloria storm, and compared it with the original fuzzy verification method. As expected, results show an increasing TPR and FPR with the neighbouring window size. This has an impact to the TSS, which reaches its maximum values when using 500-m neighbouring windows for the verification of the warnings.

One of the shortcomings of this modified version of the minimum coverage criterion is that the number of times each grid cell is used for the computation of the skill scores increases with the neighbouring window size. This fact significantly enlarges the number of FP when applying medium and large neighbouring windows. In contrast, an advantage of this modified fuzzy verification method is that the interpretation of the results can be more intuitive. The scale for which the maximum TSS values are obtained is the resolution for which warnings are most representative, and by analogy could be related to the LEWS effective resolution. Moreover, since neighbouring windows are centred in the grid cells of the analysed domain, the results may be less influenced by the resolution of the inventory.

Conclusions

The significant Gloria storm precipitations triggered multiple landslides at a regional scale in Catalonia. First in this study, the Gloria storm rainfalls and landslides that were triggered have been analysed. Then, we have taken the opportunity of this unique event to evaluate the performance of a regional-scale LEWS applying a fuzzy verification method. In the following, the Gloria storm rainfall data, the uncertainties of the landslide inventory, and the results of the fuzzy verification will be shortly discussed, and the main conclusions of this work described.

To analyse the Gloria storm rainfall, KED estimates combining radar observations and rain gauges surface measurements have been obtained. Additionally, the uncertainties of the Gloria storm estimated QPEs have been quantified by cross-validation. Results show that rainfall estimates generally are in good agreement with the rainfall observations. Because the QPEs constitute the precipitation input of the Catalonia region LEWS, their errors are a source of uncertainty of the LEWS and influence its performance. It is worth noticing that the present version of the Catalonia region LEWS uses real-time rainfall observations, and shallow slides and debris flows are rapid phenomena that happen during or just after the triggering rainfall. Hence, to issue effective early warnings, the LEWS should implement rainfall nowcasts or forecasts (Alfieri et al. 2012).

The available Gloria storm landslide inventory data has spatial and temporal uncertainties. Additionally, it is affected by reporting biases: the majority of landslides included in the inventory were triggered in forests, adjacent to railway lines and roads, and some affected buildings. Landslides that occurred in remote uninhabited areas have probably been unreported. In addition, it has been seen that many events were reported in gentle slopes. Such events can be attributed to the spatial uncertainty of the landslide inventories. In future, aerial photographs and satellite images could be used to identify landslides that have detectable sizes. Applying such techniques may improve the current and future landslide inventories in terms of number of entries and location.

Regarding the fuzzy verification method, which has been applied to assess the LEWS, we should first state that the use of neighbouring windows is meant to provide flexibility in the matching between warnings and landslide observations. Similar approaches have been used by Park et al. (2020) and Kirschbaum et al. (2009) who applied neighbouring windows to assess the performance of LEWS algorithms. The main advantage of using fuzzy verification methods varying the neighbouring window size is obtaining scale-dependent information on the LEWS skill. Our analysis shows that the LEWS has little predictability at small scales (30 m and 500 m), yet a significant improvement of the LEWS performance can be observed when increasing the neighbouring window size from 500 m to 1 km. Hence, if the inventory had been complete and the time and location of landslide reports had not been uncertain, these results would indicate that the LEWS effective resolution is around 1 km. If an exhaustive landslide inventory had been available, we could possibly find a landslide within an area of less than 1 km² from a warning. Thus, the results of the fuzzy verification for smaller neighbouring windows would have probably been better. For this reason, it could be hypothesised that the actual LEWS effective resolution might be higher than 1 km, between 500 m and 1 km.

Since the applied fuzzy verification method uses neighbouring windows only around landslide observations it benefits larger neighbouring windows. Alternatively, a modified version of the fuzzy verification method employing neighbouring windows centred in each of the grid cells of the LEWS domain has been tested. The highest LEWS performance is achieved when using 500-m neighbouring windows, indicating that if the time and location of landslide reports had not been uncertain the LEWS effective resolution would be around 500 m. The effective resolution is slightly higher than the obtained resolution when applying the original minimum coverage criterion (between 500 m and 1 km). This difference might be because the original method uses neighbouring windows centred at landslide observations. Therefore, it probably is more influenced by the resolution of the inventory.

It is interesting to notice that the obtained LEWS effective resolution is similar to the resolution of the rainfall data that has been employed to compute the warnings, which is 1 km. Although this fact seems to indicate that the rainfall resolution might affect the scale at which warnings have a useful skill, the effective resolution could also depend on other factors such as the resolution of the susceptibility map. Further research needs to be conducted in order to determine in which extent both factors influence the effective resolution of the warnings.

Additionally, it has been observed that the LEWS skill from the verification using sub-basin neighbouring windows is somewhat better than the skill obtained from the 1-km neighbouring window verification. Therefore, our results confirm that sub-basins are indeed a suitable mapping unit to visualise the LEWS outputs at a Catalan scale, as Palau et al. (2020) proposed.

Although the present study has focused on analysing the LEWS performance during a relatively short period (4 days), results seem to indicate that fuzzy verification methods could be a helpful tool to deduce the scale for which the warnings are most representative. However, their outcomes can be influenced by the resolution and uncertainties of the landslide inventory. From an emergency management point of view, knowing the effective resolution of the LEWS can be useful as it provides information on the scale at which the warnings should be trusted (Ebert 2009). In future, if landslide inventories are available covering longer periods, it would be interesting to apply fuzzy verification to assess the long-term LEWS performance and compare it with the results obtained for the Gloria storm. Furthermore, if future inventories include more accurate time information, the relation between the duration of the warnings and time of landslide occurrence could be analysed as proposed by Piciullo et al. (2020).

Declarations

Conflict of interest   The authors declare no competing interests.

References

Abancó C, Hürlimann M, Moya J, Berenguer M (2016) Critical rainfall conditions for the initiation of torrential flows. Results from the Rebaixader catchment (Central Pyrenees). J Hydrol 541:218–229. https://doi.org/10.1016/j.jhydrol.2016.01.019

Aleotti P (2004) A warning system for rainfall-induced shallow failures. Eng Geol 73:247–265

Alfieri L, Salamon P, Pappenberger F et al (2012) Operational early warning systems for water-related hazards in Europe. Environ Sci Policy 21:35–49. https://doi.org/10.1016/j.envsci.2012.01.008

Ardizzone F, Cardinali M, Carrara A et al (2002) Impact of mapping errors on the reliability of landslide hazard maps. Nat Hazard 2:3–14. https://doi.org/10.5194/nhess-2-3-2002

Atger F (2001) Verification of intense precipitation forecasts from single models and ensemble prediction systems. Nonlinear Process Geophys 8:401–417. https://doi.org/10.5194/npg-8-401-2001

Baum RL, Godt JW (2010) Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides 7:259–272. https://doi.org/10.1007/s10346-009-0177-0

Berastegui X, Casas JM, Liesa M et al (2010) Històtria geològica de Catalunya. Atles geològic de Catalunya 68–77

Berenguer M, Sempere-Torres D, Hürlimann M (2015) Debris-flow forecasting at regional scale by combining susceptibility mapping and radar rainfall. Nat Hazards Earth Syst Sci 15:587–602. https://doi.org/10.5194/nhess-15-587-2015

Brooks HE, Kay M, Hart JA (1998) Objective limits on forecasting skill of rare events. 19th Conf. Severe Local Storms. American Meteorological Society, Minneapolis, MN, USA, pp 552–555

Brunetti M, Massimo M, Peruccacci S et al (2018) How far are we from the use of satellite rainfall products in landslide forecasting? Remote Sens Environ 210:65–75. https://doi.org/10.1016/j.rse.2018.03.016

Calvello M, D'Orsi RN, Piciullo L et al (2015) The Rio de Janeiro early warning system for rainfall-induced landslides: analysis of performance for the years 2010–2013. Int J Disaster Risk Reduct 12:3–15. https://doi.org/10.1016/j.ijdrr.2014.10.005

Calvello M, Piciullo L (2016) Assessing the performance of regional landslide early warning models: the EDuMaP method. Nat Hazard 16:103–122. https://doi.org/10.5194/nhess-16-103-2016

Campbell RH (1975) Soil slips, debris flows, and rainstorms in the Santa Monica Mountains and vicinity, Southern California. US Geological Survey Professional Paper 851, p 51

Canals M, Miranda J (2020) Sobre el temporal Gloria (19–23.01.20), els seus efectes sobre el país i el que se'n deriva: Report de Resposta Rapida (R3), 1a edició. institut d'Estudis Catalans. Secció de Ciències i tecnologia, Barcelona

Casas MC, Codina B, Redano A, Lorente J (2004) A methodology to classify extreme rainfall events in the western Mediterranean area. Theoret Appl Climatol 77:139–150. https://doi.org/10.1007/s00704-003-0003-x

Cassiraga E, Gómez-Hernández JJ, Berenguer M et al (2020) Spatiotemporal precipitation estimation from rain gauges and meteorological radar using geostatistics. Math Geosci. https://doi.org/10.1007/s11004-020-09882-1

Clark AJ, Gallus WA, Weisman ML (2010) Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. Weather Forecast 25:1495–1509. https://doi.org/10.1175/2010WAF2222404.1

Clavero P, Martín-Vide J, Raso-Nadal J (1996) Atles Climàtic de Catalunya. Termopluviometria i RadiacióSolar; Barcelona, Spain, Spain

Corominas J, Alonso E (1990) Geomorphological effects of extreme floods (November 1982) in the southern Pyrenees. In: Hydrology in mountainous regions. II - Artificial Reservoirs; Water and Slopes. IAHS no. 194, Lausanne, Switzerland, pp 295–302

Corominas J, Moya J, Hürlimann M (2002) Landslide rainfall triggers in the Spanish Eastern Pyrenees. In: 4th EGS Plinius Conference "Mediterranean Storms." Editrice, Mallorca, Spain, pp 1–4

Corominas J, Mateos RM, Remondo J (2015) Review of landslide occurrence in Spain and its relation to climate. In: Ken H, Lacasse S, Picarelli L (eds) Slope safety preparedness for impact of climate change. Workshop on Slope Safety Preparedness for Effects of Climate Change. Taylor & Francis, Napoli, Italy, pp 1–28378

Corral C, Velasco D, Forcadell D et al (2009) Advances in radar-based flood warning systems. The EHIMI system and the experience in the Besòs flash-flood pilot basin. Flood Risk Management Research and Practice Extended Abstracts Volume 332 Pages Full Paper CDROM 1772, pp 1295–1303

CREAF (2009) Mapa de Cobertes del Sòl de Catalunya (MSC-4), V4 edn

Crosta GB, Frattini P (2003) Distributed modelling of shallow landslides triggered by intense rainfall. Nat Hazards Earth Syst Sci 3:81–93. https://doi.org/10.5194/nhess-3-81-2003

Crozier MJ (2005) Multiple-occurrence regional landslide events in New Zealand: hazard management issues. Landslides 2:247–256. https://doi.org/10.1007/s10346-005-0019-7

Damrath U (2004) Verification against precipitation observations of a high density network – what did we learn? In: International Verification Methods Workshop. Montreal, Canada

DGMT DG d'Infraestructures de MT (2019) Graf d'Infraestructures Vectorial de la DGIMT. Barcelona

Ebert EE (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorol Appl 15:51–64. https://doi.org/10.1002/met.25

Ebert EE (2009) Neighborhood verification: a strategy for rewarding close forecasts. Weather Forecast 24:1498–1510. https://doi.org/10.1175/2009WAF2222251.1

Ekker R, Kværne K, Os A, Humstad T (2013) regObs - public database for submitting and sharing observations. In: International Snow Science Workshop. Grenoble, Chamonix Mont-Blanc, p 5pp

Emberger L (1952) Sur le quotient pluviothermique. C.R. Acadé´mie Science: Paris; 2508–2510. Comptes Rendus De L'academie Des Sciences 234:2508–2510

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27:861–874

Gallart F, Clotet N (1988) Some aspects of the geomorphic processes triggered by an extreme rainfall event: the November 1982 flood in Eastern Pyrenees. Catena Supplement 79–95

Galli M, Ardizzone F, Cardinali M et al (2008) Comparing landslide inventory maps. Geomorphology 94:268–289. https://doi.org/10.1016/j.geomorph.2006.09.023

Gariano SL, Brunetti M, Iovine G et al (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. Geomorphology 228:653–665. https://doi.org/10.1016/j.geomorph.2014.10.019

González M, Pinyol J, Micheo MJ et al (2020) El temporal Gloria (19–23/01/2020): Els efectes dels procesos geológics sobre el territori. Institut Cartogràfic i Geològic de Catalunya, Barcelona

Guzzetti F, Cardinali M, Reichenbach P (1994) The AVI project: a bibliographical and archive inventory of landslides and floods in Italy. Environ Manage 18:623–633. https://doi.org/10.1007/BF02400865

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol Atmos Phys 98:239–267. https://doi.org/10.1007/s00703-007-0262-7

Guzzetti F, Mondini AC, Cardinali M et al (2012) Landslide inventory maps: new tools for an old problem. Earth Sci Rev 112:42–66. https://doi.org/10.1016/j.earscirev.2012.02.001

Hidayat R, Sutanto SJ, Hidayah A et al (2019) Development of a landslide early warning system in Indonesia. Geosciences 9:451. https://doi.org/10.3390/geosciences9100451

Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. Landslides 11:167–194. https://doi.org/10.1007/s10346-013-0436-y

ICGC (2013) Model d'Elevacions del Terreny de Catalunya 5x5metres v1.0 (MET-5 v1.0). ICGC

Juang CS, Stanley TA, Kirschbaum DB (2019) Using citizen science to expand the global map of landslides: introducing the Cooperative Open Online Landslide Repository (COOLR). PLoS One 14:e0218657. https://doi.org/10.1371/journal.pone.0218657

Kirschbaum DB, Adler R, Hong Y, Lerner-Lam A (2009) Evaluation of a preliminary satellite-based landslide hazard algorithm using global landslide inventories. Nat Hazard 9:673–686. https://doi.org/10.5194/nhess-9-673-2009

Kirschbaum DB, Adler R, Hong Y et al (2010) A global landslide catalog for hazard applications: method, results, and limitations. Nat Hazards 52:561–575. https://doi.org/10.1007/s11069-009-9401-4

Kirschbaum DB, Stanley T, Yatheendradas S (2016) Modeling landslide susceptibility over large regions with fuzzy overlay. Landslides 13:485–496. https://doi.org/10.1007/s10346-015-0577-2

Kirschbaum DB, Stanley T (2018) Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. Earth's Future 6:505–523. https://doi.org/10.1002/2017EF000715

Krøgli IK, Devoli G, Colleuille H et al (2018) The Norwegian forecasting and warning service for rainfall- and snowmelt-induced landslides. Nat Hazard 18:1427–1450. https://doi.org/10.5194/nhess-18-1427-2018

Lloyd DM, Wilkinson PL, Othmann MA, Anderson MG (2001) Predicting landslides: assessment of an automated rainfall based landslide warning system. In: KKS H, KS L (eds) 14th South East Asia Geotechnical Conference. Balkema, Hong Kong, pp 135–139

Lombardo L, Opitz T, Huser R (2018) Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. Stoch Env Res Risk Assess 32:2179–2198. https://doi.org/10.1007/s00477-018-1518-0

Marsigli C, Montani A, Paccangnella T (2008) A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. Meteorol Appl 15:125–143. https://doi.org/10.1002/met.65

Martín Vide F, Olcina Cantos J (2001) Climas y tiempos de España. Alianza Editorial

Nicolet P, Foresti L, Caspar O, Jaboyedoff M (2013) Shallow landslide's stochastic risk modelling based on the precipitation event of August 2005 in Switzerland: results and implications. Nat Hazard 13:3169–3184. https://doi.org/10.5194/nhess-13-3169-2013

Nikolopoulos EI, Crema S, Marchi L et al (2014) Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. Geomorphology 221:286–297. https://doi.org/10.1016/j.geomorph.2014.06.015

OCCC (2020) L'impacte de la tempesta Gloria. Informe intern, Barcelona

Osanai N, Shimizu T, Kuramoto K et al (2010) Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network. Landslides 7:325–338. https://doi.org/10.1007/s10346-010-0229-5

Palau RM, Hürlimann M, Berenguer M, Sempere-Torres D (2020) Influence of the mapping unit for regional landslide early warning systems: comparison between pixels and polygons in Catalonia (NE Spain). Landslides 17:2067–2083. https://doi.org/10.1007/s10346-020-01425-3

Pan HL, Jiang YJ, Wang J, Ou GQ (2018) Rainfall threshold calculation for debris flow early warning in areas with scarcity of data. Nat Hazard 18:1395–1409. https://doi.org/10.5194/nhess-18-1395-2018

Papa MN, Medina V, Ciervo F, Bateman A (2013) Derivation of critical rainfall thresholds for shallow landslides as a tool for debris flow early warning systems. Hydrol Earth Syst Sci 17:4095–4107. https://doi.org/10.5194/hess-17-4095-2013

Park J-Y, Lee S-R, Lee D-H et al (2019) A regional-scale landslide early warning methodology applying statistical and physically based approaches in sequence. Eng Geol 260:105193. https://doi.org/10.1016/j.enggeo.2019.105193

Park J-Y, Lee S-R, Kim Y-T et al (2020) A regional-scale landslide early warning system based on the sequential evaluation method: development and performance analysis. Appl Sci 10:5788. https://doi.org/10.3390/app10175788

Piciullo L, Dahl M-P, Devoli G et al (2017a) Adapting the EDuMaP method to test the performance of the Norwegian early warning system for weather-induced landslides. Nat Hazard 17:817–831. https://doi.org/10.5194/nhess-17-817-2017

Piciullo L, Gariano SL, Melillo M et al (2017b) Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. Landslides 14:995–1008. https://doi.org/10.1007/s10346-016-0750-2

Piciullo L, Tiranti D, Pecoraro G et al (2020) Standards for the performance assessment of territorial landslide early warning systems. Landslides 17:2533–2546. https://doi.org/10.1007/s10346-020-01486-4

Portilla M, Chevalier G, Hürlimann M (2010) Description and analysis of major mass movements occurred during 2008 in the Eastern Pyrenees. Nat Hazards Earth Syst Sci 10:1635–1645. https://doi.org/10.5194/nhess-10-1635-2010

Portilla M (2014) Reconstrucción y análisis de ocurrencias regionales de múltiples eventos de movimientos en masa generados por lluvias históricas en los Pirineos. In: TDX (Tesis Doctorals en Xarxa). Universitat Politècnica de Catalunya

Roberts NM, Lean HW (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon Weather Rev 136:78–97. https://doi.org/10.1175/2007MWR2123.1

Rossi M, Luciani S, Valigi D et al (2017) Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data. Geomorphology 285:16–27. https://doi.org/10.1016/j.geomorph.2017.02.001

Segoni S, Rosi A, Fanti R et al (2018) A regional-scale landslide warning system based on 20 years of operational experience. Water 10:1297. https://doi.org/10.3390/w10101297

Servei Meteorològic de Catalunya S (2020a) Nota de Premsa. Balanç d'una llevantada històrica a Catalunya. Barcelona, Spain

Servei Meteorològic de Catalunya S (2020b) Butlletí climàtic mensual. gener del 2020. Barcelona, Spain

Shu H, Hürlimann M, Molowny-Horas R et al (2019) Relation between land cover and landslide susceptibility in Val d'Aran, Pyrenees (Spain): historical aspects, present situation and forward prediction. Sci Total Environ 693:133557. https://doi.org/10.1016/j.scitotenv.2019.07.363

Strahler AN (1957) Quantitative analysis of watershed geomorphology. EOS, Transactions - American Geophysical Union 38:913–920. https://doi.org/10.1029/tr038i006p00913

Theis SE, Hense A, Damrath U (2005) Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. Meteorol Appl 12:257. https://doi.org/10.1017/S1350482705001763

UNISDR (2015) Sendai framework for disaster risk reduction. In: Third United Nations World Conference on Disaster Risk Reduction, pp 1–25

Varnes DJ (1978) Slope movements types and processes. Landslides analysis and control transportation research board. National Academy of Sciences, Special Report 176:11–33

Velasco-Forero CA, Sempere-Torres D, Cassiraga EF, Jaime Gómez-Hernández J (2009) A non-parametric automatic blending methodology to estimate

rainfall fields from rain gauge and radar data. Adv Water Resour 32:986–1002. https://doi.org/10.1016/j.advwatres.2008.10.004

Whittaker KA, McShane D (2012) Comparison of slope instability screening tools following a large storm event and application to forest management and policy. Geomorphology 145–146:115–122. https://doi.org/10.1016/j.geomorph.2012.01.001

Yang H, Yang T, Zhang S et al (2020) Rainfall-induced landslides and debris flows in Mengdong Town, Yunnan Province, China. Landslides 17:931–941. https://doi.org/10.1007/s10346-019-01336-y

Yin K, Chen L, Zhang G (2008) Regional landslide hazard warning and risk assessment. Earth Sci Front 14:85–93. https://doi.org/10.1016/s1872-5791(08)60005-6

Yu F-C, Chen T-C, Lin M-L et al (2006) Landslides and rainfall characteristics analysis in Taipei city during the typhoon Nari Event. Nat Hazards 37:153–167. https://doi.org/10.1007/s11069-005-4661-0

Rosa M. Palau (✉) **· Marc Berenguer · Daniel Sempere-Torres**

Centre of Applied Research in Hydrometeorology, Universitat Politècnica de Catalunya, Jordi Girona 1-3 (C4), 08034 Barcelona, Spain
Email: palau@crahi.upc.edu

Rosa M. Palau · Marc Berenguer · Marcel Hürlimann · Daniel Sempere-Torres

Department of Civil and Environmental Engineering, Division of Geotechnical Engineering and Geosciences, Universitat Politècnica de Catalunya, Jordi Girona 1-3 (D2), 08034 Barcelona, Spain

Rosa M. Palau

Natural Hazards Division, Norwegian Geotechnical Institute, N-0806 Oslo, Norway