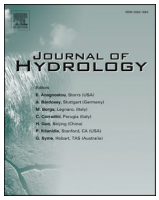
Journal of Hydrology 628 (2024) 130513

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



An empirical rainfall threshold approach for the civil protection flood warning system on the Milan urban area

Enrico Gambini [a,*], Alessandro Ceppi [a], Giovanni Ravazzani [a], Marco Mancini [a], Ismaele Quinto Valsecchi [b], Alessandro Cucchi [b], Alberto Negretti [c], Immacolata Tolone [d]

[a] Department of Civil and Environmental Engineering (D.I.C.A.), Politecnico di Milano, Piazza Leonardo Da Vinci, 32 20133 Milano, Italy
[b] Lombardy Region Administration: Functional Monitoring Centre for Natural Hazards, Piazza Città di Lombardia, 1 - 20124 Milano, Italy
[c] Terraria S.r.l.: Functional Monitoring Centre for Natural Hazards, Via Melchiorre Gioia, 132, 20125 Milano MI, Italy
[d] Lombardy Region Administration: Soil Defense Operative Unit, Piazza Città di Lombardia, 1 - 20124 Milano, Italy

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Michael Bruen, Associate Editor

Keywords:
Rainfall Thresholds
Flood Warning Systems
Civil Protection
Soil Moisture Accounting
Non-structural Risk Mitigation

ABSTRACT

Non-structural risk mitigation tools such as civil protection alerts for citizens proves highly beneficial in minimizing the impacts linked to floods. Flood forecasting represents a challenge due to complex and non-linear hydrological processes involved, especially in highly urbanized areas. In this study, a Flood Warning System (FWS) based on the development of catchment-specific empirical Rainfall Thresholds (RTs) is proposed.

Seven river catchments in the "Hydraulic node of Milan," northern Italy, were analyzed using a dataset of 25 years (1998–2022) of hourly rainfall and discharge data.

An empirical methodology, based only on historical rainfall-runoff data and applicable to any river catchment, is proposed with the aim to validate and improve the existing Rainfall Threshold (RT) defined on the same area by the Lombardy Region civil protection.

The RTs obtained using the proposed method showed improvements with respect to the existing civil protection RTs, because it allows to derive time-continuous and catchment-specific RTs. Additionally, accounting for the Antecedent Moisture Conditions (AMC) with the proposed "equivalent rainfall" approach results in more accurate RTs, suggesting its consideration for issuing civil protection alerts. The accuracy and uncertainty of the RTs were analyzed by means of binary classification measures coupled with bootstrap resampling.

The proposed procedure for constructing RTs, which is applicable to any river catchment having sufficiently long time series of rainfall and runoff data, and not necessarily on urban areas only, indicates potential for being an additional and simple FWS to mitigate flood risks for civil protection purposes.

1. Introduction

Civil protection alerts play a key role among the non-structural tools for flood risk mitigation due to their wide influence among the population. By disseminating timely and concise information about potential flood hazards, these warnings empower individuals to make informed decisions and take preventive measures to protect themselves and their possessions (del Carmen Llasat & Siccardi, 2010).

Flood forecasting systems are mostly implemented by coupling hydrological-hydraulic models with deterministic and probabilistic meteorological forecasts obtained from Numerical Weather Predictions (NWP) models (Ceppi et al., 2013; Ranzi et al., 2009; Ravazzani et al., 2016; Thielen et al., 2009): this approach is extremely useful especially when there is the necessity of long forecast horizons to supply early warnings. While these schemes have been widely applied in recent years (Das et al., 2022), the requirement of real-time runs of hydrometeorological chains could be susceptible to unforeseen system failures (wrong updating procedures, model instabilities) as well as experimenting substantial uncertainty in the description of non-linear processes inside the rainfall-runoff modelling, which is a non-negligible drawback

Abbreviations: AMC, Antecedent Moisture Conditions; CLC, Corine Land Cover; CN, Curve Number; DDF, Depth-Duration-Frequency; FWS, Flood Warning System; NWP, Numerical Weather Prediction; PCA, Principal Component Analysis; QPF, Quantitative Precipitation Forecast; RT(s), Rainfall Threshold(s); SDE, Standard Deviational Ellipse; SVD, Singular Value Decomposition; TPR, True Positive Rate; WLT, Water Level Threshold.

* Corresponding author.
E-mail address: enrico.gambini@polimi.it (E. Gambini).

https://doi.org/10.1016/j.jhydrol.2023.130513
Received 6 August 2023; Received in revised form 3 November 2023; Accepted 6 November 2023
Available online 20 November 2023

further enhanced in contexts of intense urbanization (Lombardi et al., 2018).

In the framework of small river basins, where fast evolving floods occur, the level of uncertainty of rainfall-runoff modelling could be so significant to discourage the use of hydrological models in the forecasting chain, and decision made solely on the exceedance of a predefined RT might present sufficient accuracy with enough time response (Zanchetta and Coulibaly, 2020).

Using precipitation forecast or measurements for issuing flood alerts using RTs makes the rainfall-runoff modelling not needed in real-time operations: this type of implementation is sometimes indicated in literature as an "off-line" FWS (Montesarchio et al., 2015; Rosso, 2002).

Nevertheless, it is important to highlight that RTs are always better to be used in conjunction with rainfall-runoff forecasting and real time monitoring of discharge values (Georganta et al., 2022), and thus represent a supplementary tool for early warning purposes.

RTs for flood warning can be obtained both empirically, hence using historical observations such as radar-gauges composites and satellites (Dao et al. 2020a; b) and numerically using hydrological models. Empirical RTs are very common in the context of shallow landslides and debris flow prediction (Bezak et al., 2016; Peruccacci et al., 2017; Zêzere et al., 2015) while, in the flood forecasting context, RTs were mostly derived starting from hydrological models (Amadio et al., 2003; Golian et al., 2010, 2011; Norbiato et al. 2008): this latter approach has the potential drawback of including the rainfall-runoff modelling uncertainty, that is used to derive the threshold itself, which can be very high especially when dealing with strongly urbanized areas (Salvadore et al., 2015). On the other side, the construction of RTs with empirical methods, hence using historical observations, requires the availability of an extended dataset of significant recorded events to produce a robust estimate of the threshold itself. Furthermore, the non-stationarity of the hydraulic properties in a given area (due to the change of urbanization or the construction of hydraulic infrastructures) could give the problem of a non-univocally definable value of rainfall needed to produce a certain level of risk (Mancini et al., 2002); in fact, the increasing urbanization enhance the non-stationarity of the catchment behavior. The role of urbanization is well known in the change of the hydrological response to precipitation, causing the production of higher flood peaks and volumes compared to rural basins (Brath et al., 2006; Hall et al., 2014; Zhang et al., 2018).

RTs for forecasting flood hazards have commonly been used by meteorological organizations and by Civil protection authorities to produce alerts for the citizens. In most of the Italian regions, RTs are defined for the so-called "homogeneous zones", which according to the Italian civil protection definition, are areas where the impacts of natural phenomena, like floods or landslides, tend to assume similar characteristics. In the specific case of Lombardy Region in Italy, RTs were derived through a collection of data on rainfall events and their corresponding ground effects, followed by statistical analysis aimed at finding optimal threshold values to minimize the number of false alarms and missed alerts (Regione Lombardia, 2020). These RTs are defined only each individual homogeneous zone and only for each individual time intervals (6,12 and 24 h), with values showing a decreasing rainfall intensity with respect to duration, analogously to the Depth-Duration-Frequency (DDF) curves adopted in Italy (Uboldi et al., 2014).

It is intuitive that the use of this approach lacks in temporal and spatial representativeness on the individual catchment area, because each river basin has its own characteristic in terms of time of concentration, soil type, level of urbanization, drainage density, *etc*. Hence, defining a single RT over areas which comprise multiple catchments, like homogeneous zones, could lead to incorrect estimation of the level of risk for the specific rivers inside them, as well as misinterpretations of the civil protection alert by the local municipalities and the resident population. Additionally, since most of these RTs are defined only on discrete values of duration (6, 12 and 24 h for Lombardy Region) temporal representativeness is lacking also, as this low number of durations might cause confusion in what value should be considered when a forecasted rainfall event has a duration that is between the available discrete values. Another drawback of civil protection RTs is the general absence of an antecedent soil moisture accounting scheme: the role of Antecedent Moisture Condition (AMC) in flood formation mechanisms is well known (Grillakis et al., 2016; Norbiato et al. 2008) with higher discharge values generally occurring in concomitance with more saturated soils and vice-versa (Brocca et al., 2005; De Michele & Salvadori, 2002); furthermore, the absence of a soil moisture accounting scheme potentially causes a high number of false alarms (Martina et al., 2006). Since the AMC strongly influences the runoff generation, a unique value of RT is not possible to be defined, because the same amount of rainfall falling over different initial soil saturation states produces different runoff values. A common way to proceed is to retrieve three different RTs according to the SCS approach (Soil Conservation Service, 1985) where the AMC is classified between three different categories called AMCI (dry soil), AMCII (moderately saturated soil) and AMCIII (saturated soil) by means of the 5-days antecedent rainfall (Mockus, 1964).

Nevertheless, this methodology has the disadvantage of introducing abrupt (and not realistic) changes in the values of runoff obtained; operatively, this may produce a change of alert code even with a very small difference on amounts of antecedent rain, which is physically not fully justifiable, and also not convenient in operational forecasting as well. In this study, this problem is overcome by the proposed methodology of the "equivalent rainfalls", which is derived from a modification of the SCS-CN method, and it is detailed in paragraph 3.3.

In this study, a data-driven approach was applied to derive catchment-specific RTs for seven river basins inside the "hydraulic node of Milan", northern Italy. The dataset comprised 25 years of rainfall and runoff data, publicly available from Meteonetwork and ARPA-Lombardia environmental networks. The retrieved catchment-specific RTs have the possibility to be included in the operative flood alert chain used by regional civil protection, because they provide a simple and fast tool which does not require a hydrological-hydraulic system to operate in real-time.

RTs can be used in combination with rainfall forecasts, as well as with real-time rainfall observations coming from rain gauges only, as described in more detail in Section 3.1.

As above stated, the implemented methodology to construct RTs, uses rainfall and discharge observations; furthermore, it uses the SCS-CN scheme inside the proposed "equivalent rainfall method" for the soil moisture accounting. Rainfall, discharge and CN are easily retrievable and manageable data, allowing this methodology to be easily applied on other river catchments either to compute new RTs or for an eventual updating and validation of existing flood warning RTs.

In this study we evaluated the thresholds using binary classifications scores, and we found that the retrieved RTs are able to produce enough accurate performances with good True Positive Rates (TPR) and acceptable number of False Positives (FP). When using the "equivalent rainfall" to produce new RTs, we found that the TPR slightly increased; furthermore, the number of FP systematically reduced, indicating that taking into account soil moisture using the proposed methodology gives more accurate RTs, which are able to better discriminate between an exceeding event and a non– exceeding event.

2. Study area

The study area is the so-called "hydraulic node of Milan" situated in the Lombardy Region (northern Italy) which spans approximately 1300 km^2 (Fig. 1). This region is densely urbanized with resident population exceeding 4 million inhabitants, there is the presence of lots of critical infrastructures such as highroads, railways, bridges, in a context mostly devoted to commercial and economic activities.

From a morphological standpoint, the basins of the "Hydraulic node of Milan" that converge on the city itself are elongated and narrow with a north–south trend. The riverbeds are mostly natural in the foothill part

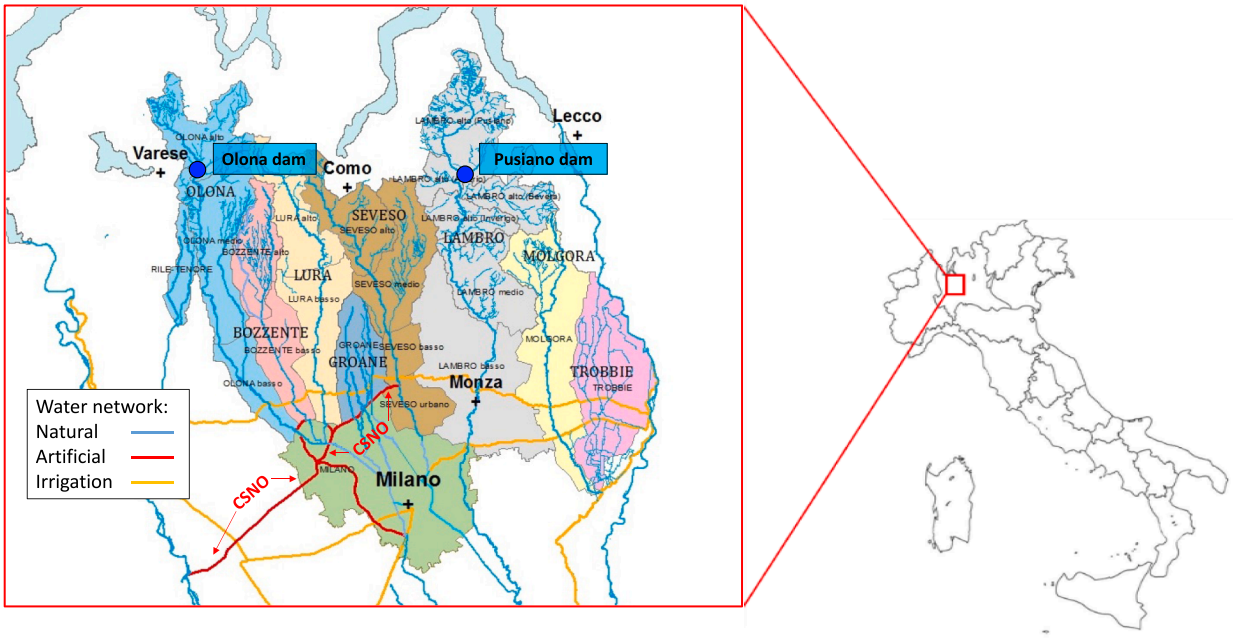


Fig. 1. The "hydraulic node of Milan".

of the area, while moving towards the plain they gradually become prismatic and, in some sections, even culverted. There is also a dense network of irrigation channels due to intensive agricultural activity. From a rainfall perspective rather, the zone experiences an average areal annual precipitation that spans from about 975 mm in the urban zone of Milan to 1570 mm in the foothill section, generally concentrated in spring and autumn periods. However, from late spring to early autumn, there are frequent storm events, with locally intense episodes.

Over the past 50 years, the territory has been subject to frequent and damaging flood events, mainly due to the rapid increase of residents and the consequent strong increase of urbanization (Becciu et al., 2018). In fact, a series of structural mitigation works were realized to reduce the discharge peaks and volumes of rivers inside the area: for instance, the diversion channel named "Canale Scolmatore di Nord-Ovest" (CSNO), is operating since 1980 and was built to collect the excess discharges of the Seveso River, preventing this excess to enter directly in the city of Milan. On the upper part of the Lambro River basin, the regulation of Lake Pusiano with the "Pusiano" dam acts mainly as a flood storage basin. Additionally, an on-stream detention basin ("Olona" dam) was constructed on the Olona River near Varese city in 2010. The reservoir has a maximum storage capacity of 1,520,000 m^3 and covers a drainage area of 3.83 km^2. To maintain downstream safety, the Olona dam is equipped with three automatic gates, limiting the released discharge to a maximum of 36 m^3/s. Since the storage capacity of the reservoir is greater than 1,000,000 m^3 this hydraulic infrastructure should be called "dam" according to Italian legislation; anyway, it works for flood mitigation purposes only. Downstream discharge regulation activates only in case of significant rainfall events; hence the storage basin remains empty most of the time.

Given the concentration of critical infrastructures and resident population, as well as the insufficiency of the capabilities of the hydraulic flood control infrastructures that is still observed nowadays, civil protection assumes a crucial role in mitigating the impact of flood events, safeguarding the well-being and resilience of people and their belongings.

3. Materials and methodology

3.1. The rainfall threshold for flood warning

The flood RT is defined as the "minimum cumulated volume of rainfall which can generate a critical water level (or discharge) at a specific river section". Once the RT for the specific cross section is computed, it can be used in real-time forecasting operations using observed rainfall and the Quantitative Precipitation Forecast (QPF) obtained from a NWP model; then, according to the level of risk on which the RT is defined, an alert could be sent if the cumulated QPF exceeds the RT for the specific event duration, as shown in Fig. 2a.

Additionally, RTs can also be used by collecting real-time rain gauge data only and by comparing the observed cumulative rainfall with the defined RT: this can potentially be useful in cases of inaccurate rainfall forecasts of NWP models both in terms of volume and location (Fig. 2b): this represents an issue that can be frequent especially in cases of convective events forecasts (Yano et al., 2018), the drawback of using the RT as reported in (Fig. 2b) is the strong reduction of the lead time of flood warning.

The critical water levels on which the RTs are computed in this study, are classified on three categories of increasing level of risk (1,2,3). The Functional Center of Hazard Monitoring of the Lombardy Region civil protection derived these three Water-Level-Thresholds (WLT) from field observations and specific hydrological-hydraulic studies (i.e., by 1D and 2D models). The exceedance of the individual WLT, according to the regional directive, corresponds to:

- WLT1: increase in water levels of main rivers and streams, generally remaining confined inside the embankments with low probability of localized flooding of adjacent areas.
- WLT2: increase in water levels of main rivers and streams, with medium probability of flooding of adjacent areas and floodplains. Possible bank erosion and sediment transport.
- WLT3: increase in water level of main rivers and streams with high probability of widespread inundations and erosions, sediment transport and channel avulsion. Phenomena of overflows, piping, seepages and levee breaches are possible.

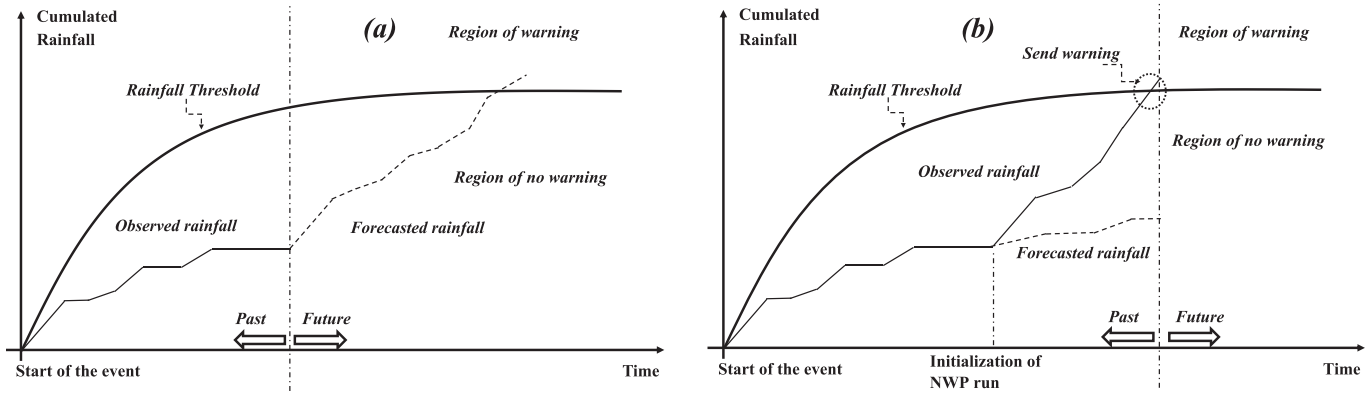


Fig. 2. A and 2b: an example of using the rt in combination with rainfall forecast (a). in figure (b) the way to make use of the rt in case of inaccurate rainfall forecast is shown. in both figures, the black continuous line indicates the observed cumulated rainfall measured from the start of the event.

The WLT2 is commonly employed as the reference for civil protection purposes as it identifies the threshold at which the events and their associated ground effects begin to have a suitable scale for calibrating the whole alerting system (Regione Lombardia, 2020). From now on the study will focus only on the definition of the RT for the WLT2, since it represents the most suitable threshold for local civil protection purposes, as above mentioned.

3.2. Dataset

To compute the empirical RTs, meteorological and hydrological data from two different sources were used.

The dataset comprised:

- Rain gauges data at hourly resolution obtained from the Regional Agency for Environmental Protection of Lombardy Region (data available from 1998 to 2022) as well as rain gauges data retrieved from the citizen-scientist network association "Meteonetwork" (Giazzi et al., 2022), available at hourly time-step from 2013 to 2022. See Fig. 3 to visualize the rain gauges locations of both meteorological networks.
- Based on historical hydrometric levels data availability and the importance of the cross-section in terms of population involved and recorded past floods, a total of seven river cross sections were selected for the study. The hydrometric levels data as well as corresponding rating curves are available at hourly resolution from the Regional Agency for Environmental Protection (ARPA-Lombardia) official website. Table 1 presents a summary of the data availability for each section, including time-range data, rating curve availability, average Curve Number (*CN*) of the catchment, and time of concentration (T_c).
- CN raster maps were derived directly from soil data type taken from the land-use maps of the Corine Land Cover (CLC) project (EUCLMS, 2023), and from the Lombardy Region geographic portal (https://www.geoportale.regione.lombardia.it/). In our study spanning from 1998 to 2022, a constant land-use type map of the year 2006, which refers to a mean *CN* condition during this period was used. This decision was based on the observation that urbanization patterns remained stable during the period of study (Ceppi et al., 2022), indicating not significant changes that would have justified the inclusion of time varying *CN* datasets.

3.3. Soil moisture accounting scheme: The "equivalent rainfall" method

The equivalent rainfall method is derived starting from the well-known event-based SCS-CN method (Soil Conservation Service, 1985) modified by the studies of (Singh, 1982) and (Mishra et al., 2004).

The SCS-CN method is based on the following balance equation:

$$P = I_a + F + Q \tag{1}$$

where:

- P is the cumulative gross precipitation, in mm.
- I_a is the so-called "initial abstraction", in mm.
- F is the cumulative infiltration, in mm.
- Q is the cumulative direct runoff, in mm.

The method defines the "maximum soil potential retention" S, in mm, as:

$$S = 254 \bullet \left(\frac{100}{CN} - 1\right) \tag{2}$$

where the *CN* (Curve Number) is a parameter ranging from 0 to 100 which parametrize the characteristics of the soil type in terms of potentiality of infiltration: soils with higher *CN* values will tend to produce more direct runoff than soils with lower *CN* values.

Combining the Eqs. (1) and (2), it is possible to write:

$$C = \frac{Q}{P - I_a} = \frac{F}{S} \tag{3}$$

where C is defined as the "runoff coefficient", and I_a can be written as:

$$I_a = \lambda S \tag{4}$$

where λ is the coefficient of initial abstraction usually taken between [0.05–0.2], most of the times is taken as 0.2.

The National Engineering Handbook of the Soil Conservation Service (SCS) suggests a classification of the *CN* values based on the Antecedent Moisture Condition (AMC) which classifies the *CN* value in three classes I, II and III for dry, average and wet conditions respectively, according to the antecedent 5 days rainfall, and the season.

The method can be improved according to the Mishra and Singh formulation (Mishra et al., 2004), where a parameter called "antecedent" moisture, M, is defined and obtained by modifying Eqs. (3) and (4):

$$C = \frac{Q}{P - I_a} = \frac{F + M}{S + M} = S_r \tag{5}$$

$$I_a = \lambda \frac{S^2}{S + M} \tag{6}$$

$$M = \frac{(P_5 - 0.2 \bullet S_I)S_I}{P_5 + 0.8S_I} \tag{7}$$

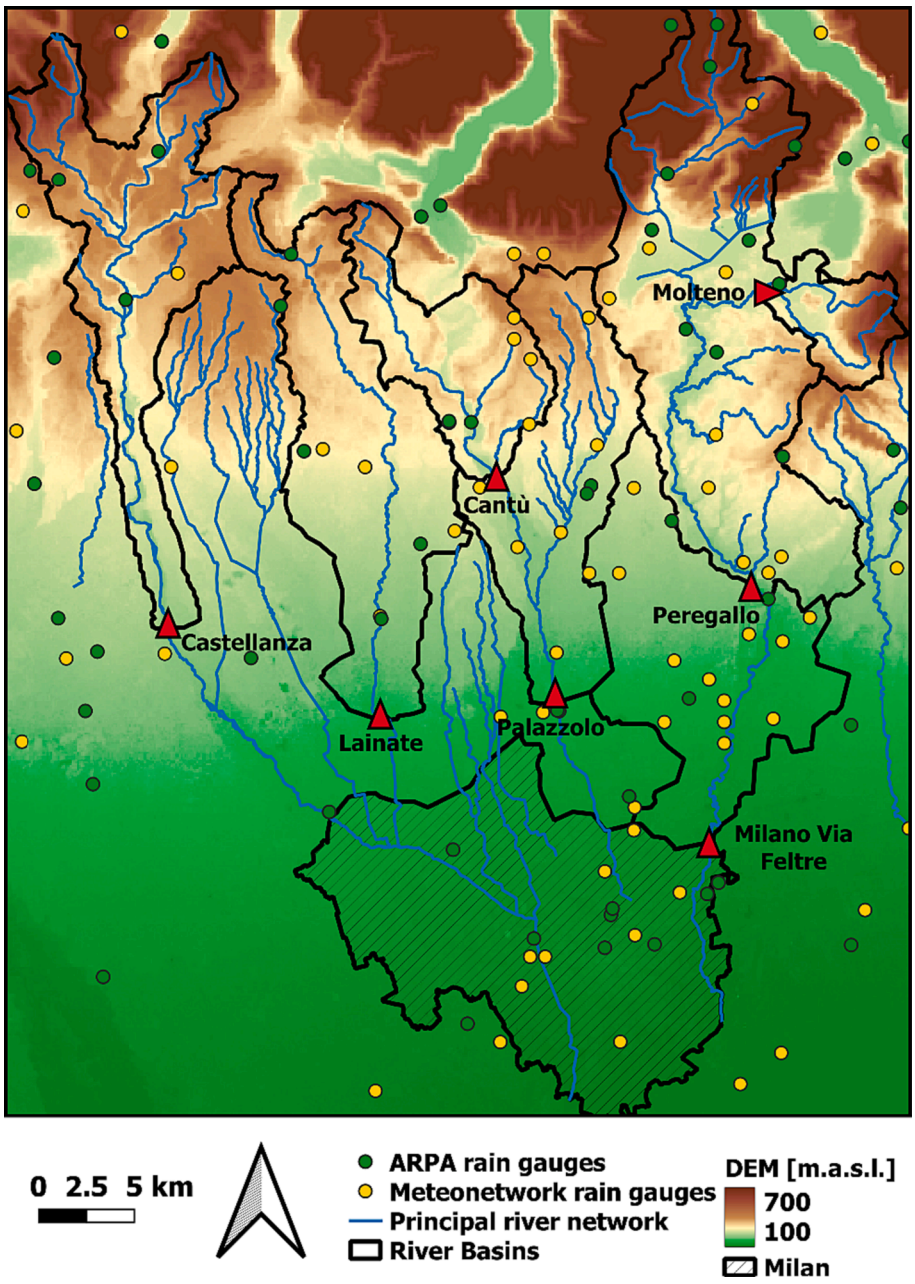


Fig. 3. River catchments contours (in black), with main rivers (in blue, rain gauges (in yellow and green dots, respectively by Meteonetwork and ARPA) and gauge sections (in red triangles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Characteristics of each river cross section and data availability.

Cross Section name	River name	Basin Area [km^2]	Available from [yyyy/mm/dd]	Rating curve	Mean CN_{II}	T_c [h]
Castellanza	Olona	143	1998/07/24	Yes	74	12
Cantù	Seveso	74	1998/07/24	Yes	77	6
Peregallo	Lambro	273	1998/07/24	Yes	79	15
Palazzolo	Seveso	187	1998/07/24	No	80	13
Milano- Via Feltre	Lambro	418	1998/07/24	Yes	80	23
Lainate	Lura	151	2014/09/25	Yes	75	10
Molteno	Bevera	29	2004/01/01	Yes	78	2

$$S_I = S + M \tag{8}$$

where S_R is the "degree of saturation" of the soil, assumed equal to C; S_I is the maximum soil potential retention considering the AMC type I class (dry), and P_5 is the antecedent 5 days precipitation amount.

Combining Eq. (5) with Eq. (1) it is possible to write:

$$Q = \frac{(P - I_a)(P - I_a + M)}{(P - I_a + S + M)} \tag{9}$$

In addition, by combining Eqs. (7) and (8), the expression of M can be generalized as :

$$M = \sqrt{S\left(P_5 + \left(\frac{1-\lambda}{2}\right)^2 S\right)} - \left(\frac{1+\lambda}{2}\right) S \; (valid\ for\ P_5 \geq \lambda S,\ M = 0\ otherwise) \tag{10}$$

Eq. (10) assumes the watershed to be dry 5 days before the onset of the rain storm (Mishra et al., 2004).

Assuming the hazard degree to be proportional to the degree of saturation S_r of the soil, it is possible to define the "equivalent rainfall" P_{eq}, associated with the predicted rainfall, P, as "the predicted rainfall on dry soil necessary to cause a S_r equal to the one caused by the rainfall P preceded by the 5-day rainfall P_5", this definition leads to the following equation:

$$S_r\left(P_{eq}, M = 0\right) = Sr(P, M(P_5, CN)) \tag{11}$$

thus, after substituting Eq. (9), Eq. (11) becomes:

$$\frac{P_{eq} - \lambda S}{P_{eq} - \lambda S + S} = \frac{P - \lambda \frac{S^2}{S+M} + M}{P - \lambda \frac{S^2}{S+M} + S + M} \tag{12}$$

finally, by solving Eq. (12) for P_{eq} and combine it with Eq. (10), we can obtain the system of equations to derive P_{eq}:

$$\begin{cases} M = \sqrt{S\left(P_5 + \left(\frac{1-\lambda}{2}\right)^2 S\right)} - \left(\frac{1+\lambda}{2}\right) S \; (valid\ for\ P_5 \geq \lambda S, M = 0\ otherwise) \\ P_{eq} = P + M\left(1 + \frac{\lambda S}{S+M}\right) = P + P_{eq0} \end{cases} \tag{13}$$

where P_{eq0} is defined as the "base equivalent rainfall", which is a function of CN and the 5-days before antecedent rainfall, P_5, and has the aim of increasing the hazard potential caused by the antecedent rainfall by summing the term to the predicted rainfall P. P_{eq0} is the amount of rainfall that causes the same degree of saturation as if the rainfall event happened on completely dry soil.

In Fig. 4 a plot highlighting the dependence of the CN and the P_5 on P_{eq0} is shown: it can be observed that generally by increasing the CN value, also P_{eq0} tends to be lower, also for high values of P_5: on mostly impermeable soils (very high CN values) the maximum soil potential retention, S, is low, and S_r (and thus C) tends to quickly approach the unit, weakly depending on the antecedent rainfall, P_5. Since soils with very high CN have low capabilities to store water, even very high values of antecedent rainfall are not able to produce a significant increase on their degree of saturation.

Conversely, low CN soils are able to store great quantities of moisture from antecedent rainfall which may be not sufficient to completely saturate the soil, giving a P_{eq0} behavior that is similar to what antecedent rainfall produces on very high CNs: this explains the higher values of P_{eq0} that arise for intermediate CN values when P_5 is kept constant.

Furthermore, it is possible to see that keeping CN constant, P_{eq0} always monotonically decreases, whatever the CN value, having the physical meaning that the more the antecedent rainfall, the more the soil will be saturated.

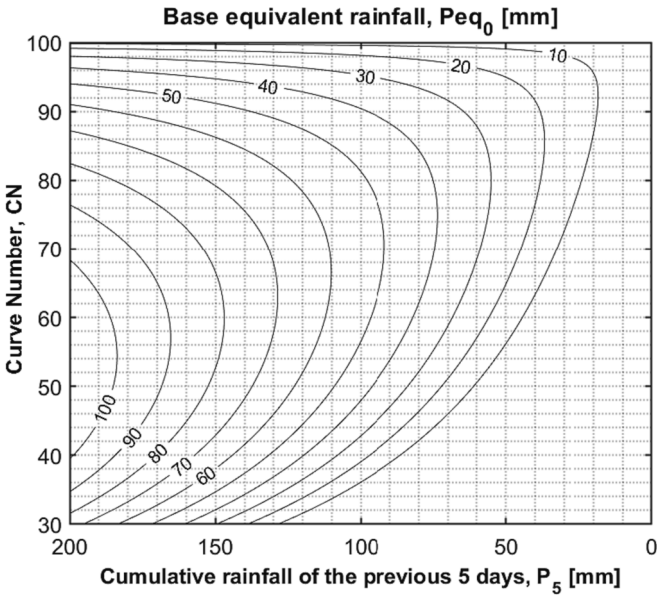


Fig. 4. Contour plot of P_{eq0} relating P_5 and CN, here $\lambda = 0.2$ is assumed.

For example, it is possible to state that an amount of equivalent rainfall of 100 mm can be obtained by forecasting 100 mm over dry soil (which means $M = 0$ at the start of the event, thus $P_{eq0} = 0$) or by forecasting a precipitation event, P, of 80 mm having a base equivalent rainfall, P_{eq0}, of 20 mm at the start of the event.

3.4. Construction of the empirical rainfall threshold

For each river catchment, the hourly time series of area-averaged precipitation is retrieved using all the available rain gauges data by means of the Thiessen Polygon's method (Taesombat & Sriwongsitanon, 2009), which was selected since it is a simple scheme known for being well suitable for retrieving area-averaged precipitation in mostly flat basins, like the ones under analysis.

The construction of the threshold is purely derived from observed rainfall and discharge data: for each river cross section being studied, the following steps are taken:

1. Preprocess discharge data excluding those periods not representative of the current hydraulic situation of the cross section: for instance, the hydraulic response of the Castellanza cross section was greatly influenced by the "Olona" dam construction in 2010; thus, in order to conduct the analysis to represent the actual hydraulic conditions of the basin, data after 2010 only were used in the computations, even though the time-series was available from 1998.
2. Inside the river discharge time series, the event of exceedances of the WLT2 are identified, then we define as t_{exc} the variable representing the temporal instant in which the discharge threshold corresponding to the WLT2 has been exceeded.
3. Afterwards, the cumulative precipitation value which led to the discharge threshold exceedance, p_{exc}, is found by cumulating the observed area-averaged precipitation back in time starting from t_{exc}, the cumulation is then "stopped" at the instant, t_{start}, where the "start" of the exceedance event is defined. Since we are dealing with mostly urbanized catchments, the initial abstraction term is not significant, and the start of the event can be identified as the instant when the area-averaged rainfall rate is greater than 1 mm/h (Mancini et al., 2002.; Rammal & Berthier, 2020). Thus, it is possible to

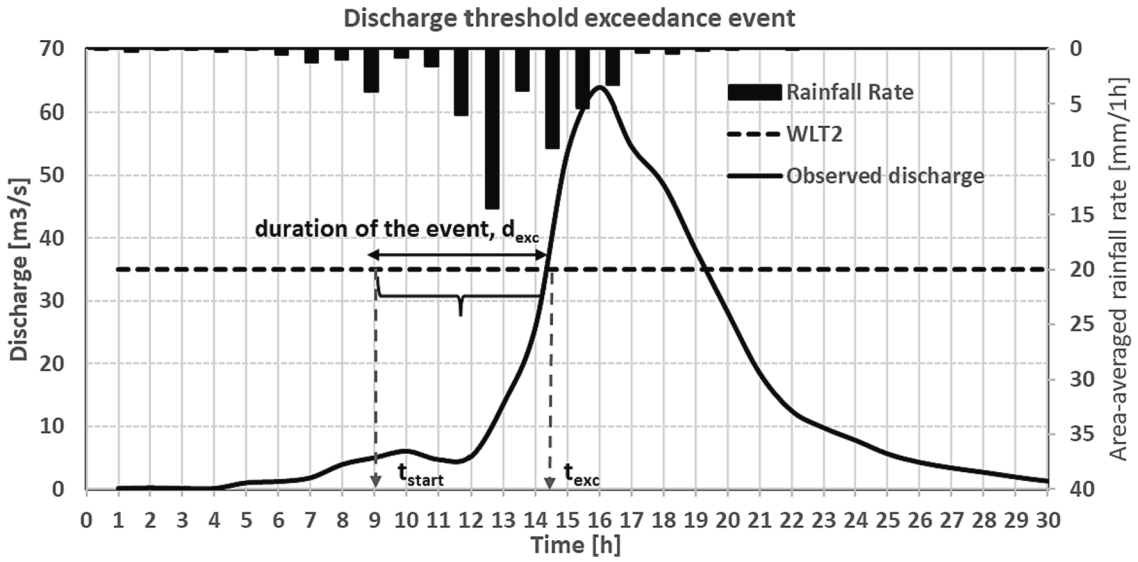


Fig. 5. Identification of an exceedance event of the WLT2.

obtain the rainfall duration which led to exceeding of the discharge threshold as $d_{exc} = t_{exc} - t_{start}$ as well as the corresponding cumulated precipitation value p_{exc}. In Fig. 5, a visual example of the procedure is reported.

4. Lastly, the construction of a dataset containing all precipitation events that did not lead to exceeding the discharge threshold is implemented. The start of each "non-exceedance" event is again taken as the temporal instant when the area-averaged rainfall rate begins to be greater than 1 mm/h, and the minimum no-rain time to separate rainfall events is set to be qual the time of concentration of the basin under investigation, T_c, because T_c is by definition the time required by water to travel from the most hydraulically remote point of a catchment to its outlet (see Fig. 6 for a schematic example). Thus, for each non-exceedance event, the couple rainfall-duration couple (p_{no_exc}; d_{no_exc}) can be defined.

Two datasets are retrieved for each analyzed cross section: one composed by the i couples ($p_{\_exc,i}$ $d_{\_exc,i}$) of the WLT2 exceedance rainfall events and the other composed by the other j couples ($p_{no\_exc,j}$, $d_{no\_exc\_j}$) of the non-exceedance of the WLT2 rainfall events.

The procedure above described is always valid when equivalent rainfall is not considered: to also obtain the datasets considering equivalent rainfall, it is possible to sum at all the above-defined times of start of the events (both for exceedance rainfall events and non-exceedance rainfall events), the area-averaged base equivalent rainfall P_{eq0} to each p_{exc} and p_{no_exc} according to Eq. (13), which says that P_{eq0} can be computed by combining the antecedent 5-days precipitation and the average CN of the river basin. This procedure allows to take into account the Antecedent Moisture Conditions (AMC) of the whole sequence of recorded rainfall events.

From a theoretical standpoint, for each duration, the cumulative rainfall, both considering and not considering equivalent rainfalls, that

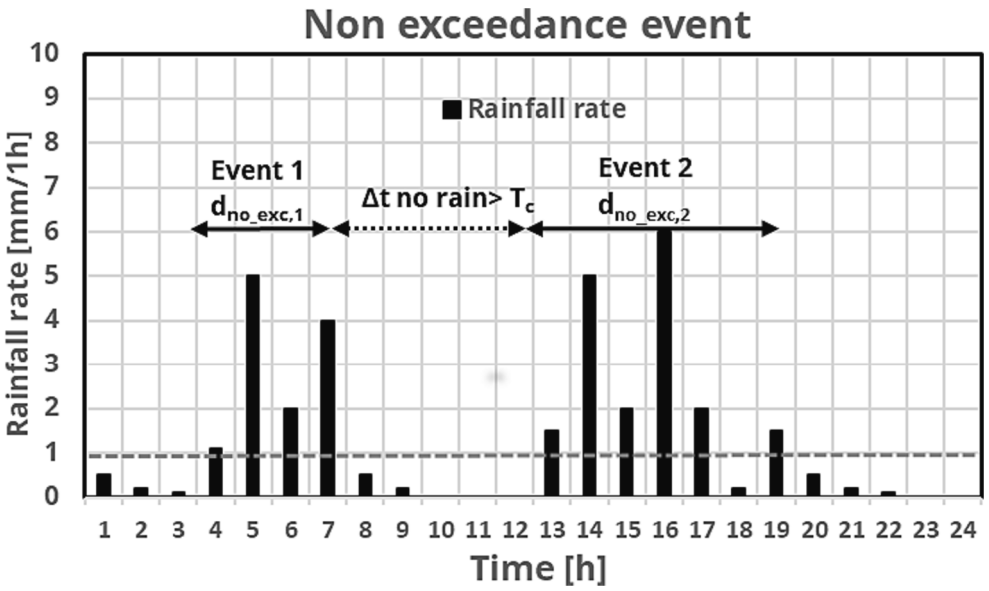


Fig. 6. Non-exceedance rainfall events separation. in this case the two events are considered as independent since the no-rain time is greater than the time of concentration, t_c, of the catchment. No-rain time is composed by all the times where area-averaged precipitation is lower than 1 mm/h. In this case T_c is supposed to be equal to 3 h. In grey dashed line, the rainfall rate threshold of 1 mm/h is highlighted.

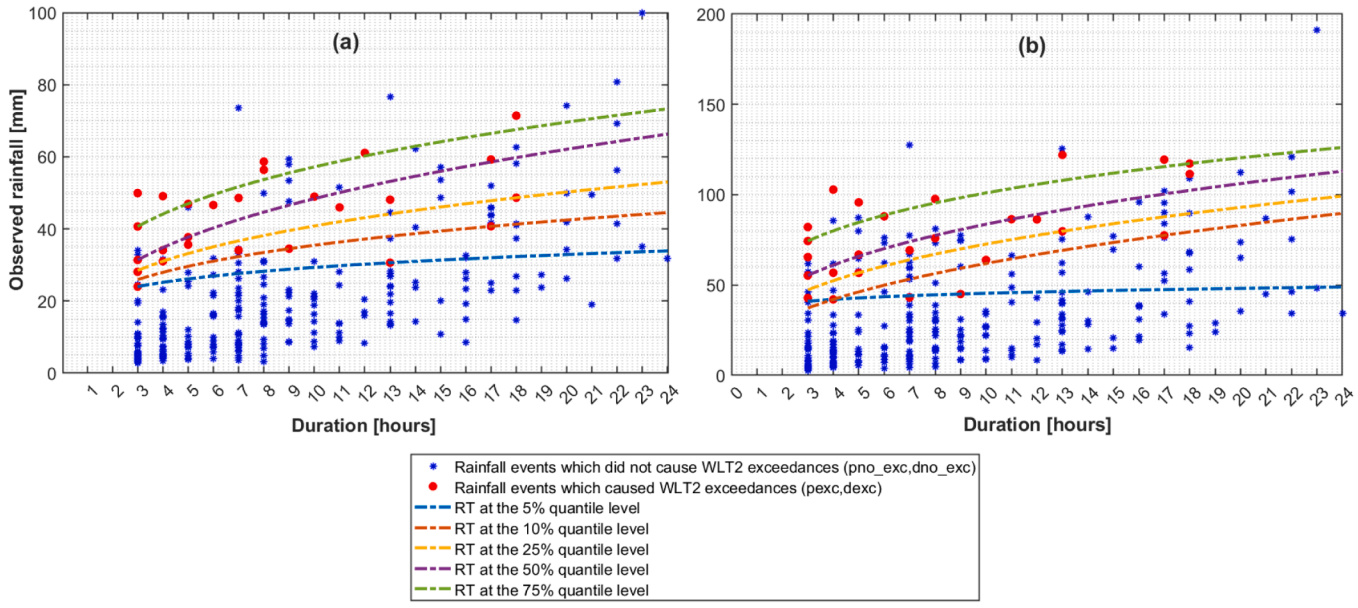


Fig. 7. An example of the result of the quantile regressions of the RTs at the Lainate river cross section, based on the WLT2: RTs are reported without considering equivalent rainfall (a), and with equivalent rainfall (b).

leads to exceedances of the WLT2, p_{exc}, should always be greater than the observed rainfall that does not cause the exceedance, p_{no_exc}; in this ideal case, the continuous RT would be obtained as the lower envelope of all the ($p_{\_exc,i}$,$d_{\_exc,i}$) couples. However, since these hypotheses are far from being achieved, a quantile regression (Koenker, 2014) applied to the ($p_{\_exc,i}$,$d_{\_exc,i}$) couples, following the approach of (Deng et al., 2022) and (Galanti et al., 2017), is used. For each quantile level selected, q, a RT is estimated as follows:

$$p_{threshold,q} = a \bullet d^n \tag{14}$$

$p_{threshold,q}$ (in mm) represent the continuous RT estimated for the *q-th* quantile level, d is the duration (in hours) of the rainfall; while a and n are the two parameters that have to be estimated, for instance, via linear regression in the $\log(p_{exc})$ vs $\log(d_{exc})$ plane, where Eq. (14) can be linearized as follows:

$$\log(p_{threshold,q}) = \log(a) + n\log(d) \tag{15}$$

The result comprises a number of thresholds equal to the number of quantile levels selected for the quantile regression, an example regression result for the Lainate river cross section is shown in Fig. 7.

The power-law functional relation of Eq. (14) between precipitation and duration was chosen because it is a simple continuous function that is in agreement with the shape of civil protection RT, as well as being widely used in the context of RTs for landslides and DDF curves (Burlando & Rosso, 1996; Roccati et al., 2020); furthermore, the choice of performing this type of regression prevents the possibility of overfitting the data with non-physically reasonable relationships.

The approach of retrieving thresholds using quantile regression extractions was chosen because it potentially mitigates the impact of variability and uncertainty of observed exceedance rainfalls for each duration; it enables to identify potential outliers and facilitates the comparison between more conservative (obtained from the regressions on low quantile levels) and less conservative (obtained from the regressions on high quantile levels) thresholds.

3.5. Rainfall threshold selection criteria

Since RTs act like binary classifiers between an alert and no-alert, it is possible to evaluate their performance by means of binary classification metrics; hence, once the RTs are fitted for each quantile level, it is possible to compute:

- True Positives *(TP)*: sum of the number of events for which $p_{exc} >= p_{threshold}$, for each observed duration, d_{exc}. They are also called "correct alarms".
- False Positives *(FP)*: sum of the number of events for which $p_{no\_exc} >= p_{threshold}$, for each observed duration, d_{no_exc}. They are also called "false alarms".
- False Negatives *(FN)*: sum of the number of events for which $p_{threshold} > p_{exc}$, for each observed duration, d_{exc}. They are also called "missed alarms".
- True Negatives *(TN)*: sum of the number of events for which $p_{threshold} > p_{no\_exc}$, for each observed duration, d_{no_exc}. They are also called "correct negatives".

In this type of approach, the analyzed datasets are extremely likely to be imbalanced because the numerosity of the exceedance events is much lower than the non-exceeding ones (Krawczyk, 2016), thus it is necessary to select the optimal RT using classification metrics that do not consider the large number of True Negatives (Torgo & Ribeiro, 2009).

The metrics chosen for this study are the following:

- The True Positive Rate, *TPR*, which is also called *sensitivity*, *recall* or *hit rate*, defined as:

$$TPR = \frac{TP}{TP + FN} \tag{16}$$

which represent the ratio of correctly predicted alarms out of all the observed alarms.

- *Precision*, also named as *positive predicted value* (*PPV*):

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

which represent the ratio of correctly predicted alarms out of all the predicted alarms.

Using definitions in Eqs. (16) and (17), the *F*-measure (Van Rijsbergen, 1979) can be computed. This metric is one of the most used

when evaluating the performance of a binary classifier in case of imbalanced data, and it is defined as follows:

$$F_\beta = (1+\beta^2) \bullet \frac{Precision \bullet TPR}{\beta^2 \bullet Precision + TPR} \tag{18}$$

where β is a positive real parameter which weights the "importance" of *TPR* with respect to *Precision*.

When $\beta = 1$, the *F*-score becomes the harmonic mean of *Precision* and *TPR*, commonly referred to as the F_1-score. The F_1-score balances the *Precision* and *TPR* equally; while values of β *greater than 1* give more weight to *TPR*, making the F_β to favor models that have higher *TPR*; conversely, values of $\beta < 1$ give more weight to precision, favoring models with higher *Precision*.

The choice of the correct value of β depends on the specific needs of the problem: in this case study, as the main focus is on civil protection alerts for citizens, we chose to weight *TPR* more than *Precision*, since the consequences of missing a critical event or failing to detect a potential threat can be severe. Furthermore, high number of false negatives (missed alarms) can cause mistrusts of civil protection authority by citizens and other public entities working on the territory.

Hence, a $\beta = 2$ value was chosen, weighting *TPR* two times more than *Precision*, resulting in the so-called F_2 measure:

$$F_2 = (1+2^2)\frac{Precision \bullet TPR}{2^2 \bullet Precision + TPR} = 5\frac{Precision \bullet TPR}{4 \bullet Precision + TPR} \tag{19}$$

since the numerosity of the observed alarms can be low, the uncertainty associated to the regression parameters of RTs can be very high; thus, to quantify the overall uncertainty of the thresholds estimates, a bootstrap resampling procedure was applied to the exceeding events dataset. The bootstrap (Efron, 1983; Efron & Tibshirani, 1997) is a resample technique which generates new samples of the same size of the original data, with replacement. Replacement means that each observation in the original sample has an equal probability ($1/m$ for a sample of numerosity m) of being selected again in each bootstrap sample, allowing for duplicates and variations in sample composition.

In this study, a number of bootstrap samples of $N_{boot} = 1000$ was chosen, enabling the estimation of the parameters a and n on N_{boot} bootstrapped samples. Since the N_{boot} estimates of the parameters represent an estimate of their sample distributions, is also possible to estimate the sample distribution of *TPR*, *Precision* and F_2-*score* for each RT defined on each quantile probability level.

On a *TPR-Precision* plot (more often indicated as *Precision-Recall* plot) it is possible to represent cloud of points of each selected quantile levels according to the bootstrap distribution of the samples of *TPR* and *Precision*: the more the point cloud spreads inside the *TPR-Precision* plane, the more the estimated threshold performances will be uncertain. To better visualize the dispersion of the point cloud on the plot, Standard Deviational Ellipses (SDE) on each point cloud are plotted exploiting a Principal Component Analysis (PCA) of the *TPR-Precision* bootstrapped couples following an approach presented by (Brunton & Kutz, 2019). This procedure allows to visually quantify the uncertainty of the thresholds within the plots, as well as depicting possible *Pareto fronts*.

Concerning for instance the Palazzolo river cross section in Fig. 8, it is possible to highlight a reduction in the scores variability as the quantile level decreases both for Fig. 8 (a), in which equivalent rainfalls are not considered, and for Fig. 8 (b), where equivalent rainfalls are instead considered. In addition, it is evident that using "equivalent rainfalls" lead to systematically better scores, except for the 5 % quantile level (only for this specific river cross section).

The definitive RT for each cross section will then be selected as the RT extracted from the quantile level that presents the highest median bootstrapped F_2-score, as shown in Fig. 9. The two definitive parameters a and n of the selected threshold will be found as the median of the resampled a_i and n_i for i= *(1, 2,…,*N_{boot}*)* bootstrapped resampled parameters for the selected quantile level RT. As an example, by looking at the results of the box-whisker plots shown in Fig. 9 (a) and (b) for Milano Via Feltre cross section, the selected thresholds will be the ones having the parameters estimated considering the 5 % quantile level for the RT without considering equivalent rainfalls (Fig. 9 (a)) and the 10 % quantile level for the RTs with the consideration of equivalent rainfalls (Fig. 9 (b)).

4. Results and discussion

Inspecting the *TPR-Precision* (or *Recall-Precision*) plots with relative *SDEs* computed using the bootstrapped samples (Fig. 10) for all

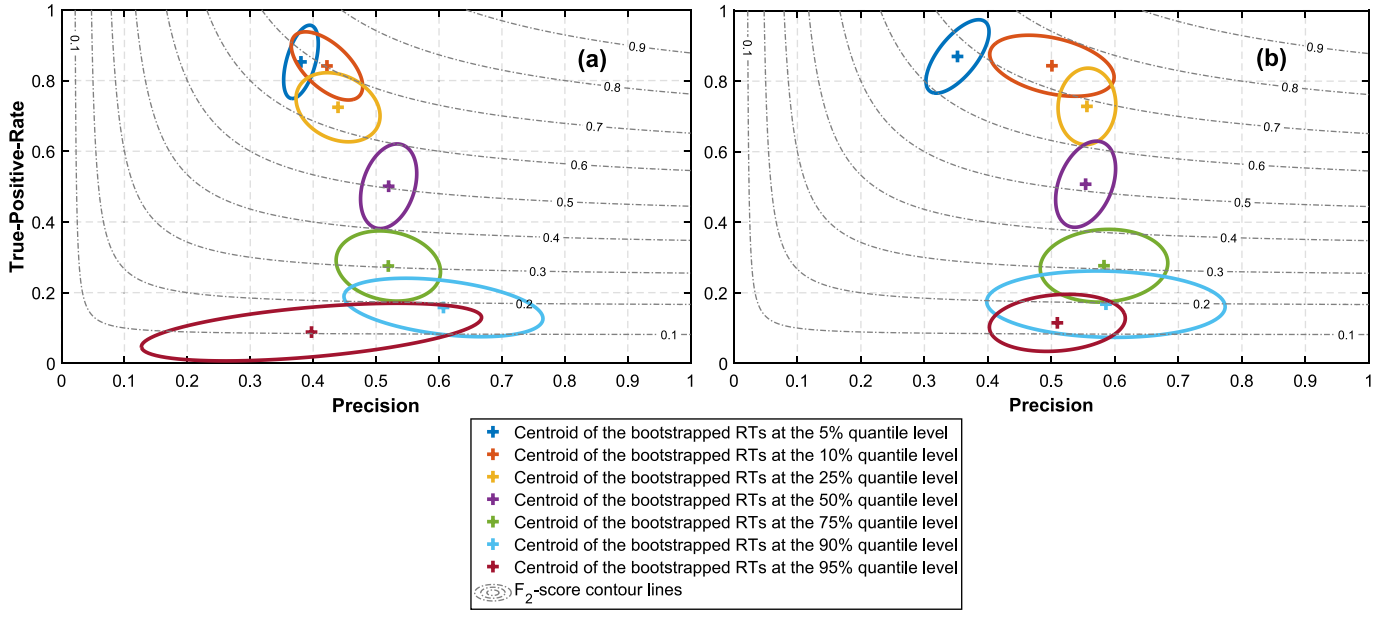


Fig. 8. TPR-Precision plots representing centroids and SDEs of the bootstrapped RTs for each quantile level for the Palazzolo river cross section. Figure (a) is obtained without considering equivalent rainfall, Figure (b) is obtained by considering equivalent rainfall. Contour lines of F_2-score are also reported in dashed gray. A perfect classifier should place himself in the upper-right corner of the plots. Here the number of standard deviations for scaling the SDEs axis is 1.

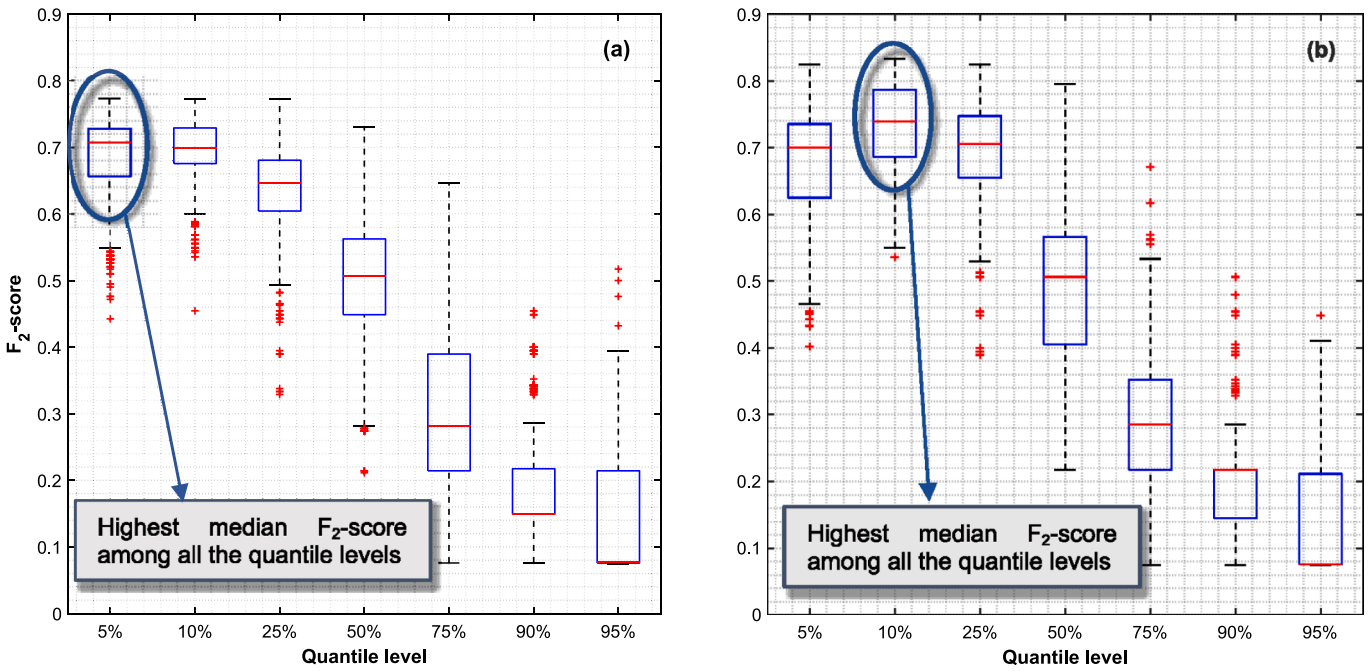


Fig. 9. Boxplots of the bootstrapped samples of the F_2-scores for each quantile level, Milano Via Feltre cross-section. Fig. 9 (a) is for RTs without considering equivalent rainfalls, Fig. 9 (b) is for RTs with the consideration of equivalent rainfalls. In this case the parameters corresponding to the 5% quantile level threshold will be selected for the construction of the RT without considering equivalent rainfalls. The 10% quantile levels parameters estimations will be selected for the construction of the RT with equivalent rainfalls.

catchments it is observable how considering equivalent rainfalls systematically produces better results in terms of F_2-score. The spread of the ellipses, which represents the uncertainty of the RTs parameters estimations and skill score performances, is as expected, dependent on the number of events available for the estimation of RTs, with scores showing less uncertainty (less spread) when the number of events available is higher. Furthermore, the spread always tends to decrease moving towards lower quantile levels, which are also the quantile levels that produce the better F_2 scores.

The results obtained from the selection of the best thresholds following the procedure explicated in section 3.4 are reported in Table 2; it can be seen that considering equivalent rainfall, which allows to directly take into account antecedent moisture conditions, produce systematically better results in terms of median *F_2-score*. The achieved values of *TPR* and *Precision* are acceptable for civil protection purposes: for instance, the average *TPR* for all catchments considering equivalent rainfalls produces a value equal to 0.89. Higher values of *TPR* at the cost of lower values of *Precision* (which means higher *False Positives*) tends to be better in the framework of civil protection, this in fact lead to the selection of the *F_2-score* metric as deciding factor.

In Table 3, the median of the bootstrapped coefficients of the RTs of each cross section extracted from the best threshold are reported: in the column "range of validity" the interval of duration of the computed parameters is shown. This range is obtained by looking at the durations of observed exceeding events: the maximum observed duration is not exactly equal to 24 h for all the cross section, but it was chosen to set the maximum duration up to 24 h essentially to compare the results with the existing civil protection RTs; on the contrary the lower value of the range corresponds exactly to the minimum of observed rainfall exceedance duration.

In Fig. 11 the computed RTs are reported, both without and with considering equivalent rainfalls; the threshold representative of the whole homogeneous zone (i.e., the "hydraulic node of Milan") is also reported; this latter was obtained with a weighted average of the values of the RTs with respect to the catchment area of the single RT value for each duration: let *Ai* be the area of the catchment related to the cross section *i*, the rainfall for the homogeneous zone for the duration *d*, $p_{homogeneous}(d)$, is computed as:

$$P_{homogeneous}(d) = \frac{\sum_i A_i p_i(d)}{\sum_i A_i} \tag{20}$$

The RT without considering equivalent rainfalls for the homogeneous zone "hydraulic node of Milan" is also reported inside the plot for a comparison with the existing thresholds (Regione Lombardia, 2020): this comparison shows a very good agreement between the RT of the whole homogeneous zone, $p_{homogeneous}(d)$, retrieved from the proposed methodology, and the existing civil protection RT. The main improvement is that now $p_{homogeneous}(d)$ is a time-continuous function, hence well suitable for all possible events durations, while the existing RT values were defined only for 6,12 and 24 h durations; furthermore, here a continuous RT is estimated singularly for all the most important catchment inside the homogeneous zone, this in operative civil protection practices, can lead to more focused flood warning alerts.

Looking again at Fig. 11 we may observe how the RT obtained for Castellanza cross section is significantly higher than the others: this is an expected outcome, since the cross section is greatly influenced by the "Olona" dam, which as mentioned above is an on-line detention basin that was built specifically for flood control purposes, and thus increases the amount of rainfall necessary to exceed the WLT2 threshold. The other thresholds exhibit similar behaviors, and the lowest ones are the Lainate and Palazzolo RTs, which are catchments characterized by a strong urbanization, thus having higher runoff coefficients with respect to the others, this produces lower values of rainfall required to exceed the WLT2 threshold. Also, Lainate is a section that has a high number of exceedance events (26) and the shortest available time series length, meaning that probably its WLT2 requires lower rainfall amounts in comparison to the others.

These results remain similar when looking at the equivalent RT plots in Fig. 12, with the exception of the Cantù river cross section that exhibits a sensible increase that can either be caused by the lower

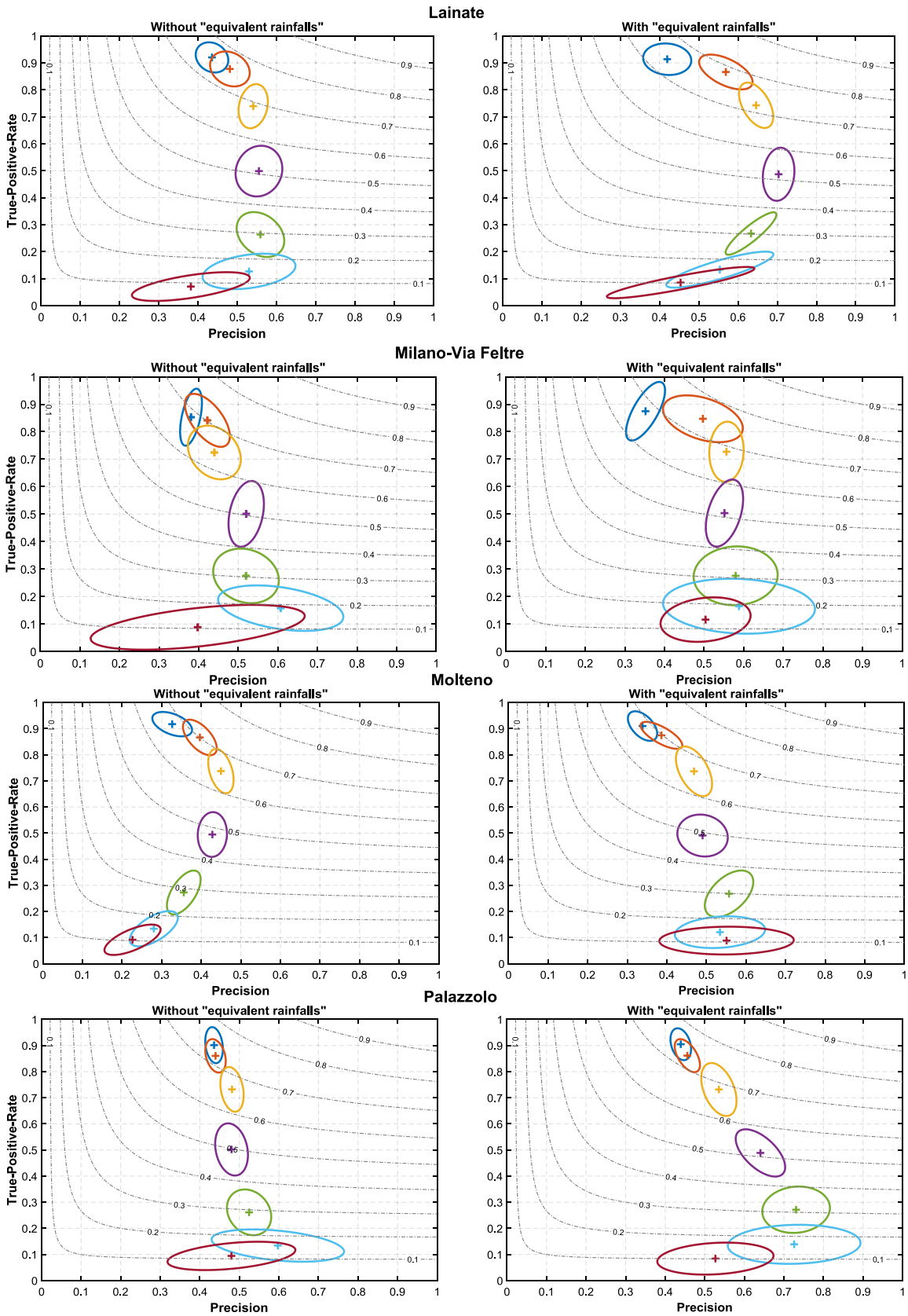


Fig. 10. TPR-Precision plots and SDEs for each cross section.

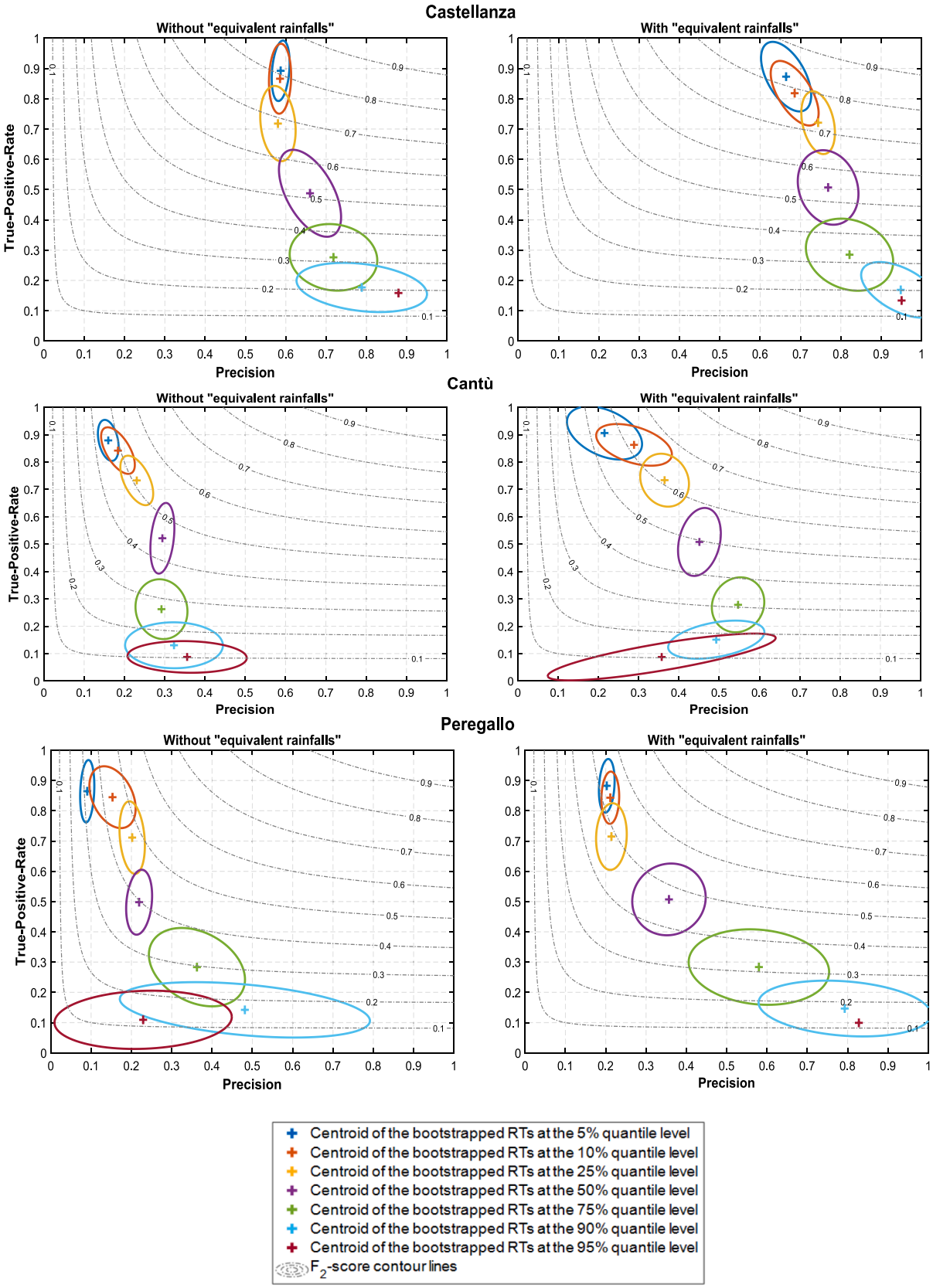


Fig. 10. (*continued*).

Table 2
Summary of the results of the skill scores for each river cross section. The statistics of *TPR*, *Precision* and F_2-score refers as the bootstrapped median value of the selected quantile level RT. The"*" superscript indicates the scores of the selected best threshold with the bootstrapped median F_2-score rule.

Cross section	N° of events considered	*TPR**	*Prec.**	F_2*	*TPR** *(Eq. rainfall)*	*Prec. ** *(Eq. rainfall)*	F_2* *(Eq. rainfall)*
Lainate	26	0.88	0.35	0.67	0.88	0.39	0.69
Milano Via-Feltre	16	0.88	0.38	0.71	0.88	0.55	0.74
Molteno	29	0.86	0.44	0.70	0.90	0.38	0.70
Palazzolo	24	0.86	0.41	0.74	0.92	0.44	0.76
Castellanza	12	0.92	0.59	0.82	0.92	0.63	0.83
Cantù	17	0.76	0.27	0.53	0.88	0.34	0.63
Peregallo	14	0.71	0.2	0.47	0.86	0.2	0.53

Table 3
Median bootstrapped coefficients of the rts both computed considering equivalent rainfall and without considering equivalent rainfalls.

Cross section	*a**	*n**	*a** *(Equivalent rainfall)*	*n** *(Equivalent rainfall)*	*Range of Validity [hours]*
Lainate	22.22	0.21	23.38	0.42	$3 \leq d \leq 24$
Milano Via-Feltre	8.53	0.64	39.19	0.29	$4 \leq d \leq 24$
Molteno	13.15	0.50	16.49	0.59	$2 \leq d \leq 24$
Palazzolo	8.47	0.59	13.96	0.62	$3 \leq d \leq 24$
Castellanza	35.58	0.22	34.87	0.39	$6 \leq d \leq 24$
Cantù	16.45	0.45	50.38	0.33	$3 \leq d \leq 24$
Peregallo	9.53	0.62	10.62	0.75	$3 \leq d \leq 24$

urbanization of the catchment itself, which lead to the equivalent rainfall to weight more, or possibly by WLT2 that is much higher in comparison to the others; in addition, it should be emphasized that even with using the bootstrap to account for uncertainty in the parameters estimation, the intrinsic uncertainty in the method cannot ever be totally excluded.

4.1. Considerations about the limitations of the study

The use of a purely empirical approach to construct RTs for flood warning purposes can produce limitations in the application and reliability of the procedure, the most important are here reported:

- the major constraint of the method is related to the potential low number of exceedance events of the hydrometric threshold. The number of events varies across sections and depends on the length of the available historical series as well as the WLT; furthermore, the possible non-stationarity during the available time-series of the hydraulic response of the catchment can strongly reduce the number of events suitable for the RT parameters estimation, even in case of long time-series. One way to assess if the available number of events is sufficient, is to use a resampling procedure to estimate confidence intervals and assess whether the uncertainty is acceptable.
- Another source of uncertainty in this method lies in the fact that the area-averaged rainfall is assumed as well representative for the entire basin area. This assumption holds well in cases of small basins, like the ones in this study. For larger basins, different discharge levels can be obtained even if the same value of area-averaged rainfall is observed: this can happen for instance when the highest rainfall intensities occur inside highly urbanized areas, in which the runoff production will be higher than if the same amount was concentrated inside a rural area. This drawback could be enhanced in case of rapidly varying precipitation events, like for instance, convective systems of precipitation.
- Area-averaged rainfall values strongly depend on the distribution of ground-based rain gauges also. It is possible that the precipitation

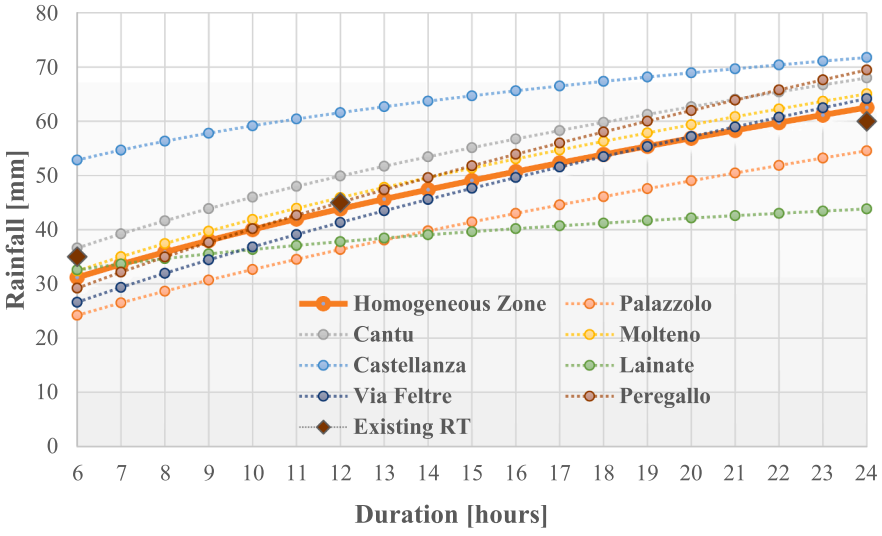


Fig. 11. Empirical RTs for all the cross sections in the interval 6–24 h of rainfall duration. The values of the existing RTs for the whole homogeneous zone area are also reported.

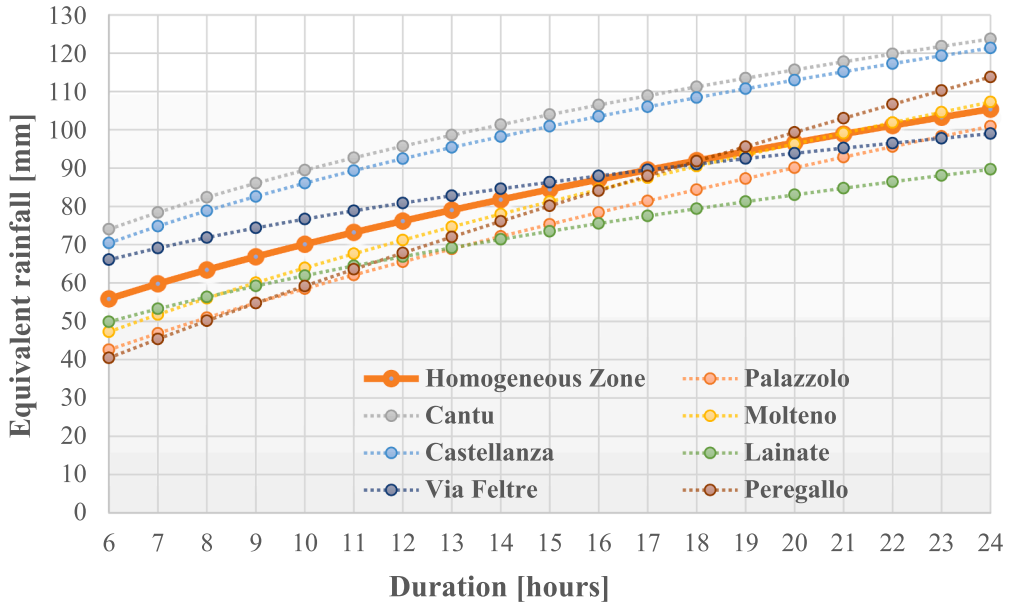


Fig. 12. Empirical RTs using equivalent rainfalls for each cross section in the interval 6–24 h of rainfall durations.

field may not be accurately captured by the rain gauges, particularly again during convective events, resulting in a significant underestimation of areal average precipitation. In these cases, radar or satellite information would be crucial for more accurate estimation of the precipitation field. Further studies should be focused on the sensibility of the area-averaged rainfall value, with respect to the rain gauge network density.

5. Conclusions

In this study, an empirical methodology for the derivation of catchment-specific empirical RTs was employed. The analysis focused on seven river catchments within the "Hydraulic node of Milan" in northern Italy, utilizing a 25-year dataset (1998–2022) of hourly rainfall and discharge data.

Empirical RTs for flood are simple and direct tools for issuing civil protection alerts in the context of hydraulic risk: the coupling RTs with the available real-time flood forecasting systems provides an additional instrument to the civil protection warning system, especially in cases of high uncertainty in the NWP forecasts or in river catchments where the response to the meteoric solicitation is strongly uncertain and difficult to be predicted by hydrological models.

The civil protection alerts that use RTs do not require a hydrological model to run; also, the proposed methodology to compute the RTs is applicable to any river catchment, hence inside any other civil protection homogeneous zone, as far as enough rainfall-runoff data and exceedance discharge events are available. Estimating the accuracy of the RTs using binary classification measures and their uncertainty using bootstrap resampling procedure can represent a way to assess whether the historical data length is sufficient to define appropriate RTs, in accordance with the uncertainty and accuracy requirements of the civil protection forecasting operations.

The RTs obtained in this study represent an improvement with respect to the existing civil protection RTs: in fact, the proposed methodology allows to estimate catchment-specific RTs on a continuous rainfall duration interval, while the existing civil protection RTs were defined only for the homogeneous zones, that are areas that usually comprise multiple catchments, and only for three discrete values of duration (6, 12 and 24 h).

Furthermore, also AMC can be accounted continuously using the proposed "equivalent rainfall" method, which gave better results in terms of *TPR* and *Precision*: on average the *TPR* increased from 0.84 to 0.89 and the *Precision* from 0.38 to 0.42. This proposed method overcomes the operative limitations of using different numbers of RTs for different levels of AMC; nevertheless, since no equivalent RT is now available for a comparison, a period of further validation (2–3 years) is suggested.

Further studies on the methodology can focus on the use of weather radar/satellite data in case of insufficient rain gauge spatial–temporal availability, also possibly taking into account AMC continuously by means of satellite data-assimilation.

In addition, an effort should be done in also considering the sensibility of hydrological response on the spatial–temporal variability of rainfalls inside the area of the catchment, especially when dealing with larger river basins, validating the applicability of the procedure also using properly calibrated hydrological models.

CRediT authorship contribution statement

Enrico Gambini: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Alessandro Ceppi:** Resources, Data curation, Writing – review & editing. **Giovanni Ravazzani:** Resources, Data curation, Writing – review & editing. **Marco Mancini:** Supervision, Project administration. **Ismaele Quinto Valsecchi:** Methodology, Writing – review & editing, Visualization. **Alessandro Cucchi:** Methodology, Writing – review & editing, Visualization. **Alberto Negretti:** Methodology, Writing – review & editing, Visualization. **Immacolata Tolone:** Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank ARPA-Lombardia (https://www.arpalombardia.it/) for providing precipitation and discharge observations, and to Meteonetwork (https://www.meteonetwork.it/) for providing rainfall observations.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Amadio, P., Mancini, M., Rabuffetti, D., Ravazzani, G., 2003. A Real Time Flood Forecasting System based on Rainfall Thresholds working on the Arno Watershed: definition and reliablity analysis. Mediterranean Storms.

Becciu, G., Ghia, M., Mambretti, S., 2018. A Century of works on River Seveso: from unregulated development to basin reclamation. International Journal of Environmental Impacts 1 (4), 461–472.

Bezak, N., Šraj, M., Mikoš, M., 2016. Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. J. Hydrol. 541, 272–284. https://doi.org/10.1016/j.jhydrol.2016.02.058.

Brath, A., Montanari, A., Moretti, G., 2006. Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). J. Hydrol. 324 (1–4), 141–153. https://doi.org/10.1016/j.jhydrol.2005.10.001.

Brocca, L., Melone, F., Moramarco, T., 2005. Progress in surface and subsurface water studies at plot and small basin scale. Empirical and Conceptual Approaches for Soil Moisture Estimation of Event-Based Rainfall-Runoff Modelling 1–8.

Brunton, S.L., Kutz, J.N., 2019. Data-Driven Science and Engineering. Cambridge University Press. https://doi.org/10.1017/9781108380690.

Burlando, P., Rosso, R., 1996. Scaling and muitiscaling models of depth-duration-frequency curves for storm precipitation. J. Hydrol. 187 (1–2), 45–64. https://doi.org/10.1016/S0022-1694(96)03086-7.

Ceppi, A., Ravazzani, G., Salandin, A., Rabuffetti, D., Montani, A., Borgonovo, E., Mancini, M., 2013. Effects of temperature on flood forecasting: analysis of an operative case study in Alpine basins. Nat. Hazards Earth Syst. Sci. 13 (4), 1051–1062. https://doi.org/10.5194/nhess-13-1051-2013.

Ceppi, A., Gambini, E., Lombardi, G., Ravazzani, G., Mancini, M., 2022. SOL40: Forty Years of Simulations under Climate and Land Use Change. Water 14 (6), 837. https://doi.org/10.3390/w14060837.

Dao, D.A., Kim, D., Kim, S., Park, J., 2020a. Determination of flood-inducing rainfall and runoff for highly urbanized area based on high-resolution radar-gauge composite rainfall data and flooded area GIS data. J. Hydrol. 584, 124704 https://doi.org/10.1016/j.jhydrol.2020.124704.

Dao, D.A., Kim, D., Park, J., Kim, T., 2020b. Precipitation threshold for urban flood warning - an analysis using the satellite-based flooded area and radar-gauge composite rainfall data. J. Hydro Environ. Res. 32, 48–61. https://doi.org/10.1016/j.jher.2020.08.001.

Das, J., Manikanta, V., Nikhil Teja, K., Umamahesh, N.V., 2022. Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. Hydrol. Sci. J. 67 (3), 477–493. https://doi.org/10.1080/02626667.2021.2023157.

De Michele, C., Salvadori, G., 2002. On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition. J. Hydrol. 262 (1–4), 245–258. https://doi.org/10.1016/S0022-1694(02)00025-2.

del Carmen Llasat, M., Siccardi, F., 2010. A reflection about the social and technological aspects in flood risk management – the case of the Italian Civil Protection. Nat. Hazards Earth Syst. Sci. 10 (1), 109–119. https://doi.org/10.5194/nhess-10-109-2010.

Deng, R., Liu, H., Zheng, X., Zhang, Q., Liu, W., Chen, L., 2022. Towards establishing empirical rainfall thresholds for shallow landslides in guangzhou, guangdong province. China. *Water* 14 (23), 3914. https://doi.org/10.3390/w14233914.

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. J. Am. Stat. Assoc. 78 (382), 316–331. https://doi.org/10.1080/01621459.1983.10477973.

Efron, B., Tibshirani, R., 1997. Improvements on cross-validation: the 632+ bootstrap method. J. Am. Stat. Assoc. 92 (438), 548–560. https://doi.org/10.1080/01621459.1997.10474007.

EUCLMS. (2023). *https://land.copernicus.eu/pan-european/corine-land-cover/*.

Galanti, Y., Barsanti, M., Giannecchini, R., D'Amato Avanzi, G., Benvenuto, G., 2017. Statistical Methods for the Assessment of Rainfall Thresholds for Triggering Shallow Landslides: A Case Study. In: Mikoš, M., Casagli, N., Yin, Y., Sassa, K. (Eds.), Advancing Culture of Living With Landslides. Springer International Publishing, Cham, pp. 429–436.

Georganta, C., Feloni, E., Nastos, P., Baltas, E., 2022. Critical rainfall Thresholds as a tool for urban flood identification in attica region, greece. Atmos. 13 (5) https://doi.org/10.3390/atmos13050698.

Giazzi, M., Peressutti, G., Cerri, L., Fumi, M., Riva, I.F., Chini, A., Ferrari, G., Cioni, G., Franch, G., Tartari, G., Galbiati, F., Condemi, V., Ceppi, A., 2022. Meteonetwork: an open crowdsourced weather data system. Atmos. 13 (6), 928. https://doi.org/10.3390/atmos13060928.

Golian, S., Saghafian, B., Maknoon, R., 2010. Derivation of probabilistic thresholds of spatially distributed rainfall for flood forecasting. Water Resour. Manag. 24 (13), 3547–3559. https://doi.org/10.1007/s11269-010-9619-7.

Golian, S., Saghafian, B., Elmi, M., Maknoon, R., 2011. Probabilistic rainfall thresholds for flood forecasting: evaluating different methodologies for modelling rainfall spatial correlation (or dependence). Hydrol. Process. 25 (13), 2046–2055. https://doi.org/10.1002/hyp.7956.

Grillakis, M.G., Koutroulis, A.G., Komma, J., Tsanis, I.K., Wagner, W., Blöschl, G., 2016. Initial soil moisture effects on flash flood generation – A comparison between basins of contrasting hydro-climatic conditions. J. Hydrol. 541, 206–217. https://doi.org/10.1016/j.jhydrol.2016.03.007.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T.R., Kriaučiūnienė, J., Kundzewicz, Z.W., Lang, M., Llasat, M.C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R.A.P., Plavcová, L., Rogger, M., Salinas, J.L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., Blöschl, G., 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrol. Earth Syst. Sci. 18 (7), 2735–2772.

R. Koenker Quantile Regression N. Balakrishnan T. Colton B. Everitt W. Piegorsch F. Ruggeri J.L. Teugels Wiley StatsRef: Statistics Reference Online 1 Wiley.

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence 5 (4), 221–232. https://doi.org/10.1007/s13748-016-0094-0.

Lombardi, Gabriele, Alessandro, Ceppi, Ravazzani, Giovanni, Silvio Davolio, Marco, Mancini, 2018. From deterministic to probabilistic forecasts: The 'shift-target'approach in the Milan urban area (northern italy). Geosciences 8 (5), 181.

Regione Lombardia. (2020, December 21). *D.g.r. 21 dicembre 2020 - n. XI/4114, https://www.regione.lombardia.it/wps/wcm/connect/e148f424-8b59-4ed8-849a-c607dced194f/dgr-4114-2020-aggiornamento-direttiva-gestione-sistema-allertamento-protezione-civile.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-e148f424-8b59-4ed8-849a-c607dced194f-nBHegVi*.

M. Mancini P. Mazzetti S. Nativi D. Rabuffetti G. Ravazzani P. Amadio R. Rosso DEFINIZIONE DI SOGLIE PLUVIOMETRIHE DI PIENA PER LA REALIZZAZIONE DI UN SISTEMA DI ALLERTAMENTO IN TEMPO REALE PER IL BACINO DELL'ARNO A MONTE DI FIRENZE (n.d.).

Mancini, M., Mazzetti, P., Rabuffetti, D., Ravazzani, G., Amadio, P., & Rosso, R. (2002, September). Definizione di Soglie Pluviometriche di Piena per la realizzazione di un sistema di allertamento in tempo reale per il bacino dell'Arno a monte di Firenze. *28° Convegno Di Idraulica e Costruzioni Idrauliche*.

Martina, M.L.V., Todini, E., Libralon, A., 2006. A Bayesian decision approach to rainfall thresholds based flood warning. Hydrol. Earth Syst. Sci. 10 (3), 413–426.

Mishra, S.K., Jain, M.K., Singh, V.P., 2004. Evaluation of the SCS-CN-Based Model Incorporating Antecedent Moisture. Water Resour Manage 18 (6), 567–589.

Mockus, V., 1964. National engineering handbook Vol. 4.

Montesarchio, V., Napolitano, F., Rianna, M., Ridolfi, E., Russo, F., Sebastianelli, S., 2015. Comparison of methodologies for flood rainfall thresholds estimation. Nat. Hazards 75 (1), 909–934. https://doi.org/10.1007/s11069-014-1357-3.

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., Anquetin, S., 2008. Flash flood warning based on rainfall thresholds and soil moisture conditions: an assessment for gauged and ungauged basins. J. Hydrol. 362 (3–4), 274–290. https://doi.org/10.1016/j.jhydrol.2008.08.023.

Peruccacci, S., Brunetti, M.T., Gariano, S.L., Melillo, M., Rossi, M., Guzzetti, F., 2017. Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology 290, 39–57. https://doi.org/10.1016/j.geomorph.2017.03.031.

Rammal, M., Berthier, E., 2020. Runoff losses on urban surfaces during frequent rainfall events: A review of observations and modeling attempts. Water 12 (10), 2777.

Ranzi, R., Bacchi, B., Ceppi, A., Cislaghi, M., Ehret, U., Jaun, S., Marx, A., Hegg, C., Zappa, M., 2009. Real-time demonstration of hydrological ensemble forecasts in map d-phase. La Houille Blanche 95 (5), 95–104. https://doi.org/10.1051/lhb/2009061.

Ravazzani, G., Amengual, A., Ceppi, A., Homar, V., Romero, R., Lombardi, G., Mancini, M., 2016. Potentialities of ensemble strategies for flood forecasting over the Milano urban area. J. Hydrol. 539, 237–253. https://doi.org/10.1016/j.jhydrol.2016.05.023.

Roccati, A., Paliaga, G., Luino, F., Faccini, F., Turconi, L., 2020. Rainfall threshold for shallow landslides initiation and analysis of long-term rainfall trends in a mediterranean area. Atmos. 11 (12), 1367. https://doi.org/10.3390/atmos11121367.

R. Rosso Manuale di Protezione Idraulica del Territorio 2002 Cusl.

Salvadore, E., Bronders, J., Batelaan, O., 2015. Hydrological modelling of urbanized catchments: a review and future directions. J. Hydrol. 529, 62–81. https://doi.org/10.1016/j.jhydrol.2015.06.028.

Singh, K. P. (1982). *Runoff conditions for converting storm rainfall to runoff with SCS Curve Numbers*.

Soil Conservation Service, 1985. Hydrology. In *National Engineering Handbook*.

Taesombat, W., Sriwongsitanon, N., 2009. Areal rainfall estimation using spatial interpolation techniques. ScienceAsia 35 (3), 268. https://doi.org/10.2306/scienceasia1513-1874.2009.35.268.

Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., 2009. The European flood alert system – Part 1: concept and development. Hydrol. Earth Syst. Sci. 13 (2), 125–140. https://doi.org/10.5194/hess-13-125-2009.

Torgo, L., & Ribeiro, R. (2009). *Precision and Recall for Regression* (pp. 332–346). https://doi.org/10.1007/978-3-642-04747-3_26.

Uboldi, F., Sulis, A.N., Lussana, C., Cislaghi, M., Russo, M., 2014. A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution. Hydrol. Earth Syst. Sci. 18 (3), 981–995. https://doi.org/10.5194/hess-18-981-2014.

Van Rijsbergen, C.J., 1979. *Information Retrieval* (Butterworth-Heinemann, Ed, 2nd ed.).

Yano, J.-I., Ziemiański, M.Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S.L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V.T.J., Soares, P.M.M., Wyszogrodzki, A.A., 2018. Scientific challenges of convective-scale numerical weather prediction. Bull. Am. Meteorol. Soc. 99 (4), 699–710. https://doi.org/10.1175/BAMS-D-17-0125.1.

Zanchetta, A., Coulibaly, P., 2020. Recent advances in real-time pluvial flash flood forecasting. Water 12 (2), 570.

Zêzere, J.L., Vaz, T., Pereira, S., Oliveira, S.C., Marques, R., Garcia, R.A.C., 2015. Rainfall thresholds for landslide activity in Portugal: a state of the art. Environ. Earth Sci. 73 (6), 2917–2936. https://doi.org/10.1007/s12665-014-3672-0.

Zhang, W., Villarini, G., Vecchi, G.A., Smith, J.A., 2018. Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. Nature 563 (7731), 384–388. https://doi.org/10.1038/s41586-018-0676-z.