

Empirical atmospheric thresholds for debris flows and flash floods in the Southern French Alps

T. Turkington¹, J. Ettema¹, C.J. van Westen¹, and K. Breinl²

¹University of Twente, Enschede, The Netherlands

²Paris-Lodron University of Salzburg, Salzburg, Austria

Correspondence to: T. Turkington
(t.a.r.turkington@utwente.nl)

Abstract

Debris flows and flash floods are often preceded by intense, convective rainfall. The establishment of reliable rainfall thresholds is an important component for quantitative hazard and risk assessment, and for the development of an early warning system. Traditional empirical thresholds based on peak intensity, duration and antecedent rainfall can be difficult to verify due to the localized character of the rainfall and the absence of weather radar or sufficiently dense rain gauge networks in mountainous regions. However, convective rainfall can be strongly linked to regional atmospheric patterns and profiles. There is potential to employ this in empirical threshold analysis.

This work develops a methodology to determine robust thresholds for flash floods and debris flows utilizing regional atmospheric conditions derived from ECMWF ERA-Interim reanalysis data, comparing the results with rain gauge derived thresholds. The method includes selecting the appropriate atmospheric indicators, categorizing the potential thresholds, determining and testing the thresholds. The method is tested in the Ubaye Valley in the southern French Alps (548km²), which is known to have localized convection triggered debris flows and flash floods. This paper shows that instability of the atmosphere and specific humidity at 850hPa are the most important atmospheric indicators for debris flows and flash floods in the study area. Furthermore, this paper demonstrates that atmospheric reanalysis data is an important asset, and could replace rainfall measurements in empirical exceedence thresholds for debris flows and flash floods.

1 Introduction

A key component in risk assessments for natural hazards is quantifying the probability of occurrence in relation to specific intensities of the hazardous events. Intense short duration precipitation, long-lasting rainfall, and snowmelt are all potential triggers for hydro-meteorological hazards in mountainous areas in Europe (Brunetti et al., 2013;

Sene, 2013). However, while rainfall is often an important element in triggering hydro-meteorological hazards, the actual atmospheric conditions are often complex, with very localized rainfall.

In the European Alps and Mediterranean region, debris flows are generally caused by heavy rainfall from either intense convection, or sustained heavy frontal rainfall (Tarolli et al., 2012). Antecedent conditions, such as previous rainfall, snowmelt and evaporation, are also important, however they are often not collected or incorporated into the threshold (Guzzetti et al., 2008). Debris flows can be generated by a number of different causes, such as liquefaction of the toe part of landslides, blocking of channels, and accelerated erosion along gullies. Heavy rainfall may trigger debris flows and flash floods in the same channels filled with sediments (van Asch et al., 2013), and both events can be approached similarly in the threshold analysis. Within this paper we refer to rapid instantaneous events such as debris flows or flash floods as flash events.

The role of rainfall in triggering debris flows and flash floods can be examined using physically based models (e.g. Quan Luna et al., 2011; van Asch et al., 2013). Through the use of hydrologic and stability models, these physical models take into account not only rainfall, but other factors such as pore pressure and slope stability (Aleotti, 2004). However, the models can be computationally costly and require extensive parameterization and calibration. Therefore, the application of such models is often only feasible for relatively small areas, such as a single torrent or a few square kilometres (Brunetti et al., 2013).

For larger areas (tens of square kilometres upwards), empirical rainfall thresholds are more frequently used (e.g. Aleotti, 2004; Giannecchini, 2006; Frattini et al., 2009; Brunetti et al., 2010; Berti et al., 2012). Thresholds define minimum or maximum conditions of one or more triggering factors for a particular hazardous event (Frattini et al., 2009). The research focus in this field recently has been towards the development of objective and reproducible thresholds (Guzzetti et al., 2008). Methods include Bayesian inference, where the parameters of the threshold are fit using a probability

approach (Guzzetti et al., 2007), and a Frequentist approach, based on the frequency of conditions that have resulted in landslides (Brunetti et al., 2010). A detailed review of empirical thresholds for debris flows and landslides can be found in Guzzetti et al. (2008).

5 For debris flows a typical approach is to define a threshold based on the intensity, duration or antecedent rainfall amounts (Guzzetti et al., 2008). The general form of the rainfall threshold is as below (Eq. 1), with three examples from Caine (1980) (Eq. 2), Guzzetti et al. (2008) (Eq. 3), and Cepeda et al. (2010) (Eq. 4):

$$I = \alpha D^{\beta} \quad (1)$$

$$10 \quad I = 14.82D^{-0.39} \quad (2)$$

$$I = 2.20D^{-0.44} \quad (3)$$

$$15 \quad I = 29.14D^{-1.34} \quad (4)$$

where intensity (I) is given in mm/hr, duration (D) in hours, and α and β are curve parameters

Empirical rainfall thresholds rely on accurate rainfall measurements, often requiring sub-daily data (e.g. Aleotti, 2004; Giannecchini, 2006; Cepeda et al., 2010). However, as many hydrological and meteorological stations still collect only daily rainfall, fine resolution data is not always available. In mountainous areas, precipitation can vary greatly with altitude. Without extensive meteorological networks, the effect of orographic processes on the spatial variation of rainfall can be difficult to determine (Tobin et al., 2011). Therefore, in many threshold studies, many hazardous events are excluded from analysis. Brunetti et al. (2013) automatically excluded events where the closest rain gauge was more than 5km away or there was not sufficient rainfall data, and in Meyer et al. (2012), 20% were excluded due to insufficient information.

Other challenges for empirical rainfall thresholds include having a detailed and sufficiently complete inventory of events, and deciding and defining the indicators to use

in the thresholds. It also is often not clear how to define a rainfall event (when it starts and finishes), although recent papers have tried to address this (Brunetti et al., 2010; Berti et al., 2012). Finally, many of the empirical methods establish a threshold above which debris flows may occur, without considering non-event observation also above the threshold, as there are many more non-event days. Meyer et al. (2012) used only debris flow events to determine the threshold, then analysed the annual frequency of days above the threshold. As rainfall is not the only factor governing debris flows, there will likely always be uncertainty in the definition of rainfall thresholds (Berti et al., 2012).

One way to approach the significance of a threshold is using Bayesian probability (e.g. Berti et al., 2012). Bayesian probability takes into account the likelihood of an event given certain conditions. However, while Bayes' theorem is useful in determining the probability of an event above a certain threshold, it does not take into account the probability that an event would be below this threshold. So even if the probability of an event occurring above a particular threshold is high, many events may occur below this threshold.

The thresholds above all use rainfall directly, however, it is also possible to analyse the cause of heavy precipitation. Ingredients that can lead to precipitation include mechanisms for uplift of an air mass (such as heating at the surface or orographic lift), increased saturation of the atmosphere, or a mixing of two or more air masses (such as fronts and low pressure systems). Maddox et al. (1979) found for the US that 43% of flash floods were caused by local convection, while the rest were synoptically driven. Studies in the Mediterranean basin show heavy precipitation events are often caused by quasi-stationary local convection (e.g. Nuissier et al., 2008). Atmospheric indicators can summarize the principle atmospheric conditions leading to heavy rainfall for a particular area, depending on the different causal mechanisms.

While atmospheric indicators have not had widespread usage in threshold analysis for flash events, they have been used as indicators for heavy rainfall and downscaling climate projections. Trapp et al. (2009) used the product of convective available potential energy (CAPE) and deep-layer wind shear (DLS) as an indicator for severe

thunderstorms. Nuissier et al. (2011) used synoptic (large-scale) weather types based on the Hess- Brezowsky Grosswetterlagen classification, as well as low-level moisture flux and low-level wind direction to detect heavy precipitation events in southern France. Other examples of using atmospheric indicators for heavy precipitation include: Schmidli et al. (2007), Chen et al. (2010), and Jeong et al. (2012). Identification of synoptic atmospheric conditions that lead to flooding has also been undertaken in a number of studies (e.g. Petrow et al., 2009; Parajka et al., 2010).

Atmospheric indicators can be obtained using reanalysis data from physically-based models. Using a forecast model combined with observations, reanalysis data is both consistent with atmospheric observations and the laws of physics (Dee et al., 2011). The weighting given to the observations differs depending on the quality of the observations. Less reliable fields, such as precipitation, are less dependent on observations than more reliable fields such as mean sea level pressure (Tapiador et al., 2012). However, the quality of the output is dependent on the skill of the underlying forecasting model. Overall though, reanalysis data provides a wide range of atmospheric variables that are both spatially complete and coherent (Dee et al., 2011).

Rather than rainfall thresholds from local weather stations, this research develops empirical atmospheric thresholds for debris flows and flash floods using atmospheric indicators to identify the potential heavy rainfall events, using 63 flash events in the Southern French Alps. The main advantages are that a dense observational rain gauge network is no longer required, and that there is no need to define explicitly what a rainfall event is. Furthermore, atmospheric thresholds can lead to a better understanding of the meteorological conditions that are related to the occurrence of debris flows and flash floods. Empirical atmospheric thresholds therefore may can be an alternative to the conventional empirical rainfall thresholds where dense observational networks are not available, or where further investigation is required to the cause of the rainfall.

The structure of the paper is as follows: first an overview of the study area and the dataset is given, followed by a description of the methodology to develop atmospheric thresholds. The methodology includes dividing the flash events into those caused by

local convection, and those that are from more synoptically driven, widespread rainfall. Thresholds using weather station data are also generated for comparison. The results are then presented and discussed, with a conclusion on the main results and limitations of developing and using empirical atmospheric threshold for debris flows and flash floods.

2 Study area and data description

The Ubaye Valley is an east-west oriented valley in the Alps de Haute-Provence, France with a catchment size of around 548km² and elevation between 1100m and 3000m a.s.l (Fig. 1). The Ubaye Valley has a mountainous Mediterranean climate with snow cover at high altitudes for approximately half of the year (Malet et al., 2007). Previous investigation has found that hydro-meteorological events are generally associated with snowmelt and high intensity summer storms, although the precise triggering conditions have been difficult to determine (Flageollet et al., 1999).

Four of the five weather stations are located close to the main river channel (Fig. 1). Station 5 (Table 1) is only operational during the summer and hence only used for qualitative comparison with the other locations. Information on elevation, length of measurement series and variables for all the weather stations can be found in Table 1. All stations measure daily precipitation, and station 1 also records temperature. Stations 1 to 4 are homogeneous based on the criteria from Wijngaard et al. (2003) and three homogeneity tests (Pettitt, 1979; Alexandersson, 1986; Wang et al., 2010). Total annual precipitation amounts for stations 1 to 4 vary between 730mm and 985mm, with the mean annual daily maximum precipitation amount between 46mm (station 1) and 53mm (station 4). The correlation between station 5 and the other four stations in summer is low: between 0.02 and 0.08, based on the Kendall's tau correlation coefficient (Kendall, 1970). The correlation between stations 1 to 4 is higher: between 0.69 and 0.74.

The Ubaye Valley has an extensive landslide, debris flow, and flash flood inventory compiled from historical data in municipal archives, newspapers and technical reports (Flageollet et al., 1999). Historical records provide valuable information on temporal occurrence of larger events, although the events recorded depend on the exposure and awareness of the observers to the hazard (Ibsen and Brunnsden, 1996; Carrara et al., 2003).

The historical inventory contains 29 flash floods and 39 debris flows events observed between 1979-2010, which occurred between March and November (Fig. 2). Tarolli et al. (2012) found a similar seasonal distribution of flash floods, with events generally occurring between August to November in the western Mediterranean. On average, discharge levels between September and November closely follows the mean precipitation intensity, while the discharge increases from March to July mainly due to snowmelt (Fig. 2). As the valley is orientated west-east, north facing slopes are likely to retain snow longer than south facing slopes.

Cepeda et al. (2010) developed Eq. 4 for debris flows based on hourly precipitation from Station 1. Only 7 debris flows were used, as the others occurred before sub-daily precipitation measurements were available (1998), or the precipitation or inventory record was deemed to be not sufficient (Cepeda et al., 2010). For the threshold, 86% of the debris flow events used were correctly predicted, and 5.5% of rainfall events above the threshold resulted in a debris flows. However, no threshold was obtained using only the longer daily rainfall dataset. To obtain a threshold for a longer time period, other methods or datasets are therefore required.

ECMWF ERA-Interim reanalysis data is used for analysing the regional atmospheric variables. The data has a spatial resolution of 80km (T255) covering the period 1979-2012 (Dee et al., 2011). More information about observation and data assimilation and model characteristics for ERA-Interim can be found in Dee et al. (2011). The study area is approximately half of one grid box, so only the grid box containing the study area and those directly beside it are used (nine in total). The variables chosen (Table 2) contain commonly used predictors for statistical downscaling precipitation from Global Climate

Models at multiple atmospheric pressure levels (Chen et al., 2010; Jeong et al., 2012). In addition, convective available potential energy (CAPE), deep layer shear (*DLS*), and soil moisture fields are also included. The first two are added as they might be indicative for convective (Marsh et al., 2009) and soil moisture as part of antecedent conditions. CAPE in particular is an estimate of the energy that a parcel of air would have at the surface if it was lifted. High positive CAPE values indicate that the air may be unstable and favourable for convection. A brief description of each of the variables is also given in Table 2. Atmospheric indicators at 850hPa and 700hPa represent lower tropospheric conditions, while indicators at 500hPa and 250hPa represent the upper troposphere. The surface variables are available at 3 hourly time steps, with the others at 6 hourly time steps (Dee et al., 2011). *DLS* is estimated using the following equation and the surface wind fields (u_{10m}, v_{10m}) and 500hPa wind fields (u_{500hPa}, v_{500hPa}) (Seltzer et al., 1985):

$$DLS = \sqrt{(u_{500} - u_{10})^2 + (v_{500} - v_{10})^2} \quad (5)$$

3 Methodology

This section explains a method to establish empirical thresholds for debris flows and flash flood events (flash events) based on regional atmospheric conditions or indicators from the reanalysis dataset. Two different thresholds are considered: (1) a probabilistic threshold based on Berti et al. (2012), determining the likelihood of a flash event using a variety of indicators, and (2) a static threshold that takes into account the number of flash events below the threshold as well as the probability of occurrence. Besides defining the threshold, the methodology also examines (a) if the local weather station network was adequately capturing the rainfall causing the event, (b) whether intense convection was the main rainfall source triggering the events, and (c) if other meteoro-

logical triggers, such as snowmelt, are relevant to triggering events in the study area. The three steps of the proposed methodology are:

Section 3.1: Categorize events based on potential meteorological triggers

Section 3.2: Select appropriate atmospheric indicators for each category

5 Section 3.3: Compute the probabilistic and static thresholds and then apply these over a validation period

Based on the availability of the weather station data and reanalysis data, the period 1979-2010 was chosen as the focus study period. The years from 1989 to 2004 are used for calibration and two validation periods are selected, namely 1979-1988 and
10 2005-2010. By splitting the validation period into two segments, changes in data quality, such as measurement techniques or observational coverage, are expected to be reduced while maintaining as long as possible data period. The probabilistic and static thresholds are also established using local weather station data for direct comparison with the empirical regional atmospheric thresholds.

15 **3.1 Categorization of events**

The proposed categories are based on the governing rainfall generation processes, with a secondary subdivision based on potential antecedent conditions. The four categories are: Ls - locally generated rainfall, spring, Lr - local rainfall, summer, Ss-synoptic
(~~large scale atmosphere~~) rainfall, spring, and Sr synoptic rainfall summer. The classification is based on Merz and Blöschl (2008), who identify five categories for river floods
20 based on the type of rainfall and antecedent conditions such as snowmelt and rainfall over several weeks. The categories Ls and Ss assume snowmelt is an antecedent condition, while Lr and Sr assume no snowmelt. For this study, seasonal antecedent conditions (snowmelt or/and rainfall) are based on the average annual discharge pattern in Section 2. From Fig. 2, the discharge generally returns to near baseflow levels
25 in July. Added to this, the east-west orientation of the Ubaye Valley means that the south facing slopes will be snow-free earlier than the north facing slopes. Therefore,

the spring events were defined as flash events between March and June for south facing slopes, and between March to mid-July for north facing slopes.

The rainfall generation processes are split into types where local conditions are driving the generation, or whether it is governed by the synoptic atmospheric processes. In Done et al. (2006), the authors estimate the rate at which CAPE is being removed by convective heating as:

$$t_{CAPE} \sim \frac{CAPE}{dCAPE/dt} \quad (6)$$

where t_{CAPE} is the convective timescale and $\frac{dCAPE}{dt}$ is the rate of change of CAPE removed by convective heating.

Done et al. (2006) suggest that with convective timescales shorter than 6 hours the synoptic conditions are governing the instability of the atmosphere, while locally driven intense convection occurs when t_{CAPE} values are high. Non-convective precipitation would also have a low t_{CAPE} value, as $CAPE$ values are generally low (Molini et al., 2011). Applying the criteria by Molini et al. (2011), flash events with $t_{CAPE} > 6hr$ are classified as locally convective (L), and with $t_{CAPE} < 6hr$ corresponding to more equilibrium conditions (S).

Molini et al. (2011) and Done et al. (2006) further modified equation 6 by estimating the latent heat release using the precipitation rate. However, as hourly rainfall rates are not available for any weather station before 1998, and Done et al. (2006) explain this is just a rough indication of the convective timescale, the version in equation 6 is used.

The accuracy of the classification of rainfall generation type is dependent on the accuracy of CAPE from ERA-Interim. Molini et al. (2011) found, when comparing CAPE values from ERA-Interim with those from a near-by radiosonde, there was only modest correlation, with a coefficient of determination of approximately 60%. Differences would be expected however, when comparing the grid box average with a point location.

3.2 Indicator selection

Each day in the calibration period 1989-2004 is assigned a label as an event day (a day where one or more flash events were recorded), and non-event days (where no flash event was recorded). The atmospheric indicators that show a distinction between event days and non-event days can then be used in the development of atmospheric thresholds (Section 3.3). The silhouette index (SI) is used to identify atmospheric indicators that best differentiate between the clusters of flash events and non-flash events.

This index takes into account both the separation between the two clusters as well as the cohesion within the cluster (Rousseeuw, 1987). The index was developed as part of a tool to visualise the distinction between multiple clusters, and as a guide to the validity of the clustering and selection of number of clusters (Rousseeuw, 1987). It has since been used as a validation tool in classifying atmospheric conditions (e.g. Huth et al., 2008; Kannan and Ghosh, 2011; Kenawy et al., 2013).

An individual silhouette value determines how similar a point is to other points in its own cluster compared to points in other clusters (Rousseeuw, 1987). The SI then the average of all the silhouette values (Huth et al., 2008), with Eq. 7 valid for two clusters:

$$SI = \frac{1}{2} \sum_{c=1}^2 \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{b_i - a_i}{\max(a_i, b_i)} \quad (7)$$

where n_c is the number of observations in cluster c , b_i is the average Euclidean distance between an observation i and all observations in the other cluster and a_i is the average Euclidean distance between i and all observations in the same cluster.

The SI varies between -1 and 1. An individual silhouette value of 1 indicates that the observation is correctly classified as a flash or non-flash event, while a near zero value indicates that the observation could belong to either cluster, and negative values indicate misclassification (Ansari et al., 2011). The highest SI indicates the best clustering (Ansari et al., 2011). An overall SI value of 1 means that the clusters are compact and well separated from each other (Kenawy et al., 2013).

5 A worked example of the SI for floods in the Ubye River is given. Days with high discharge values (flood days) are compared with no-flood days. The no-flood days chosen had similar event and antecedent rainfall amounts as the flood days. Fig. 3 shows the individual silhouette values for flood days/ and no-flood days in the Ubye Valley based on Q850 and U&V850. The left figure shows the individual silhouette values for each flood day are above 0, indicating they are more similar to the other flood days than the no-flood days. For the no-flood days, half of the days have positive silhouette values and are likely correctly classified. The other half has negative values, indicating they are more similar to the flood days. The figure on the right side plots the no-flood and flood days, and shows the separation between the two groups. It shows that generally flood days have higher specific humidity and more easterly winds compared with no-flood days.

10 The SI value is less reliable for clusters when there is a large difference between the number of flash and non-flash events. Therefore, x days are randomly selected to calculate the SI using the normalized atmospheric variables, where x is the number of flash events. This is repeated multiple times (10,000), with variables with the highest mean SI value selected for threshold analysis. Any atmospheric indicators that had more than 10% of SI values less than zero were discarded. In Sect. 4.2, only the mean SI value is given.

20 As conventional thresholds are generally defined using two variables, the analysis is performed with the two best performing indicators. Furthermore, too many indicators could create noise, or lead to over-fitting of the data (Kenawy et al., 2013). The degree of correlation between atmospheric predictors also reduces the benefit of using many predictors (Hewitson and Crane, 2006). However, where the inventory of flash events is more substantial, more three or more atmospheric variables could be used to improve the atmospheric threshold.

3.3 Probabilistic and static thresholds

Bayes' theorem expresses the conditional probability of an event A , such as a flash event, occurring given some condition or conditions, B , such as atmospheric conditions (Eq. 8). It is based on the unconditional probability of A occurring, $P(A)$, unconditional probability of the condition occurring $P(B)$, and the conditional probability of $P(B|A)$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (8)$$

Using the two indicators from Section 3.2 that had the highest SI value, the probability of a flash event occurring was calculated over the observed range of each of the indicator. This is similar to Berti et al. (2012), although extended to using atmospheric indicators.

A limitation of using probability of occurrence is that it does not take into account the percentage of flash events above the threshold. Therefore, a static threshold is also determined considering both the number of events above and below the threshold. A static threshold is taken to be a threshold where the values of the indicators remain constant. The indicators used for the static threshold are the same as for the probabilistic threshold.

A confusion matrix displays the performance of a prediction algorithm, such as a static threshold. The four classifiers in the confusion matrix (Mason and Graham, 1999) are:

- True positives (TP): the number of correctly predicted events
- False positives (FP): the number of events predicted, but where no event occurred
- False negatives (FN): the number of events that were not predicted
- True negatives (TN): the number of days that were correctly predicted as non-events

These classifiers can then be used to determine the correlation between the predicted and observed results using the Matthews Correlation Coefficient (MCC; Powers, 2011):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (9)$$

5 The MCC is similar to the Pearson's product-moment correlation coefficient applied to contingency tables (Powers, 2011). A value of 1 indicates perfect correlation, while zero indicates no relationship and negative values indicate negative correlation. Although to our knowledge the MCC has not been used in rainfall threshold assessment, it has been used in bioinformatics, as an assessment tool where there are unequal
10 events and non-events (Baldi et al., 2000; D'Este and Rahman, 2013).

The MCC is calculated for each combination of atmospheric indicators from the probabilistic threshold. The threshold with the highest MCC value is chosen as the static threshold, with the added condition in that at least 50% of the flash events are also above the threshold. These selection criteria are somewhat subjective, as the optimal
15 threshold will depend on the application.

4 Results and discussion

4.1 Categorization of events

Table 3 shows the t_{CAPE} value (Eq. 6) for all separate events in the period 1989-2004. In 66% of the events, local convection was considered to be the dominant meteorological trigger for flash events in the Ubaye Valley. The earliest local convective event
20 reported in a year occurred on the **The observed convective events occurred between the 1st** of June and the latest on the 23rd November (**numbers 5 and 13** in Table 3). The

synoptic events occurred over a wider range of months, between March and November (numbers 9 and 1 in Table 3).

It is possible that some of the flash events are in the wrong category. Four of the nine synoptic events had no rainfall recorded in at least half of the stations 1-4, which would not be expected with widespread rainfall (numbers 3,4,6,8 in Table 3). However, any misclassification would likely only reduce the efficiency of the clustering (Section 4.2), and the significance of the thresholds (Section 4.3). Therefore we used the classification as indicated in Table 3 for the subsequent analysis.

4.2 Indicator selection

The two best performing indicators for the local convective events were CAPE and specific humidity at 700hPa (Fig. 4). These two indicators showed the highest SI value, 0.32 (apart from using only Q700). CAPE especially has been used before as an indicator for intense convection (Marsh et al., 2009), as it indicates atmospheric instability. Q850 Q700 is indicative of low-level moisture, which is also necessary for locally generated precipitation. Comparatively, the U&V winds showed very low SI values, indicating that wind conditions do not separate flash event days from non-event days. This was also true for DLS and soil moisture (SWL). The vertical integral of water vapour flux was also trialled, however the SI value was also low (not shown). Temperature, vorticity and divergence showed moderate SI values, between 0.1 and 0.25 depending on what other atmospheric indicator it was paired with. The moderate SI values separate the flash events from the non-event days somewhat, but not as much as CAPE and Q700 Q850.

Figure 5 (top) shows that for all the synoptic events, only 10 indicator combinations were significant (at $p = 0.10$). The highest SI value (0.15) was also half of the value found for local convective events. As the number of significant indicator pairs was low, To improve the indicator selection, the SI was calculated again further splitting the events into the Ss and Sr categories (Fig. 5 middle and bottom). However, this was a trade-off between the limited number of events belonging to each classification and

potential differences in indicators. However this meant that there were only 4 to 5 flash events in each group. Therefore, any thresholds were unlikely to be as robust as for the local convection and weather stations, as there were fewer events to both calibrate and validate the thresholds.

Splitting the synoptic events into the Sr and Ss categories showed differences between the atmospheric indicators with the highest SI (Fig. 5, middle and bottom). For Ss events, temperature at multiple pressure levels separated days with flash events from days with no flash events. This was in combination with 8-day average mid-level divergence, temperature, CAPE or specific humidity. The highest SI value of 0.21 was for temperature (3-day) and specific humidity (8-day), both at 700hPa. These two indicators were then used as the basis of the thresholds in Section 4.3. For the Sr flash events, the significant indicators were divergence at 700hPa 850hPa (daily), low level specific humidity, SWL, and 8-day average temperature (Figure 5). The highest SI of 0.42 for the Sr flash events corresponded to specific humidity and 8-day average temperature at 700hPa. Low level moisture (Q700 and Q850) again appeared to be a key atmospheric indicator. Low level temperature was also a key indicator, although only when Ss and Sr events were separated (Fig. 5). ~~It was possible that one class was associated with colder temperatures, and the other with warmer temperature, which then cancel when combined.~~

Finally, for the local weather station data, the highest SI value of 0.29 was for the 4 day and daily total rainfall based on the data from station 3. Other stations and combination of stations were tried, but all had lower SI values. These indicators were similar to those for debris flows in Jaiswal and van Westen (2009), ~~where 1 and 5 day totals were used. A four day antecedent period was chosen over five day as it had a slightly higher SI value (0.29 compared with 0.27).~~ Intensity and duration indicators are not used, as hourly data were not available before 1998. Also, previous attempts using daily data showed all flash events were below the thresholds Eq. 2 and Eq. 3.

4.3 Probabilistic and static threshold

4.3.1 Weather station thresholds

Based on Bayes' theorem and one and four day rainfall totals, there was Using the daily values and four day values for antecedent conditions, a threshold was generated that showed increasing chance of flash events with higher rainfall totals. The highest probability of a flash event was 17% when the one-day total is above 80mm and the four-day above 96mm (Fig. 6). This is lower than the maximum probability found in the study by Berti et al. (2012) of 40-60%.

While Fig. 6 seems reasonable (more precipitation, more likely for a flash event to occur), there are a few limitations. There are nine days with precipitation totals above 82mm where no flash event was recorded and hence zero probability of flash occurrence. The lack of recorded events may have been because of low precipitation intensity, or the amount recorded by the rain gauge was much higher than for the rest of the study area. Spatial heterogeneity of rainfall may also be the reason why during the calibration period no precipitation was recorded for one flash event, and less than 10mm for a further six flash events. From Fig. 1, the affect torrents were in some instances more than 10km away from a rain gauge, which is especially problematic for localized convection where the precipitation is confined an area less than 10km².

For the static threshold, the maximum MCC value during the calibration period, with at least 50% of events above the threshold, corresponded to the following Weather Station threshold ($Thres_{WS}$):

- $Thres_{WS}$: one-day precipitation > 20mm and four-day antecedent precipitation > 22mm

The values for the static threshold are given in Table 4. Only 8.5% of the total number of days were above the $Thres_{WS}$ ((TP+FP)/(FN+TN)), while 55% of the flash events were above the $Thres_{WS}$ (TP/(TP+FN)). Somewhat surprisingly, 45% of the event days

had less than 20mm of rainfall. The percentage of the total number of days above
5 *Thres_{WS}* was slightly lower for the two validation periods (7.5% and 6.1% respectively),
and the percentage of flash events drops even more (35.7% and 33.3% respectively).
While the likelihood of a flash event still remains higher for days above the static thresh-
old in the validation period, the drop in percentage of flash events above the threshold
10 indicates differences in the triggering conditions between the calibration and validation
periods. ~~As both the validation periods are different, this suggests that the changes
are not completely from changes in the landscape or mitigation works.~~ The torrents in
which flash events occurred are generally closer to station 3 in the earlier validation
period than the calibration period.

The results for the static threshold are comparable to those from other studies.
15 Cepeda et al. (2010) found for the same study area that their threshold is exceeded
on average 8.6 times per year, while 60% of debris flows are above the threshold (if
including all debris flows between 1998 and 2010). While the percentage of correctly
predicted events is slightly lower, the percentage of false positives is only a third of the
amount using Eq. 4. The better performance of the rainfall threshold using hourly data
20 from Station 1 indicates that rainfall intensity is important rather than daily amount. The
daily total of 20mm was in the range of Meyer et al. (2012), between 15–107mm/day.
The probability of static threshold exceedence was also similar to Meyer et al. (2012),
whose threshold was exceeded between 0 and 77 days in a year (8.5% corresponds
to 31 days a year).

25 **4.3.2 Atmospheric thresholds: Local convection**

Flash events during the summer and autumn period are more likely under high insta-
bility (CAPE) and high 700hPa specific humidity (Fig. 7). As both the instability of the
atmosphere and low level moisture increase in Fig. 7, the probability of a flash event
also increases. ~~High instability but low moisture (so less probable raincloud develop-
ment) show low probability of a flash event, as is expected.~~ The highest probability
(100%) is higher than found in Section 4.3.1 st and from Berti et al. (2012)st. This cor-

5 responds to CAPE values above 1100 J/kg and normalized Q700 greater than 1.45, although this has only been observed once between 1989 and 2003.

For the static threshold, the maximum MCC value during the calibration period, corresponded to the following threshold ($Thres_L$):

- $Thres_L$: CAPE > 250 J/kg and normalized specific humidity at 700hPa > 0.40

10 The confusion matrix results and MCC values are shown in Table 4. From this table it can be seen that 6.8% of the days are above $Thres_L$, compared with 75% of local convective flash events. In the validation periods, the percentage of days above $Thres_L$ rises to 7.8% (Validation Period 1) and 7.3% (VP2) and 71% and 80% for the local convection flash events.

15 Compared with the results in Section 4.3.1, both the probability threshold and static thresholds perform better for the local convection than for the weather station threshold. Figure 7 shows higher probabilities of flash event occurrence than Figure 6. Similarly, the MCC value for all three periods was higher for the local convection atmospheric threshold. And in In both validation periods, more flash events were above the $Thres_L$ than $Thres_{WS}$, with an even smaller number of FPs in the first validation period. Lower number of FP is important for early warning systems where the number of false alarms should be minimised.

20 While the CAPE value in $Thres_L$ was low for intense convection, similar limits have been found in other studies (e.g. for hail Niall and Walsh , 2005; Pistotnik et al., 2011, for heavy rainfall). Trapp et al. (2009) also found that availability of low level water vapour was a key component of changes in severe convection at mid-latitudes.

4.3.3 Atmospheric thresholds: Synoptic, spring

Based on Section 4.2, it appears when the Ss and Sr events are grouped together that the resulting SI values are low, performing better than when the SI values are calculated individually. However this meant that there were only 4 to 5 flash events in each group, only two or three more than the number of indicators. Therefore, the thresholds

Discussion Paper | Discussion Paper | Discussion Paper

5 were unlikely to be as robust as for the local convection and weather stations, as there were fewer events to both calibrate and validate the thresholds.

Fig. 8 shows for Ss indicators, that with warmer 700hPa temperatures and higher specific humidity the probability of flash event occurrence increases. Warm low to mid-level temperatures could be associated with melting of snow and high moisture levels could indicate rain. Fig. 8 had similar probabilities of occurrence compared to $Thres_{WS}$, with the highest probability of occurrence of 12.5%. Similar to Figure 6, the most extreme days (days with the highest 8-day moisture and warmest 3-day temperature), were not associated with flash events.

Using the criteria in Sect. 3.3 resulted in the following threshold ($Thres_{Ss}$):

- 15 – $Thres_{Ss}$: three-day mean temperature at 700hPa > 271 K and eight-day mean normalized specific humidity at 700hPa > 0.70

20 The values for the confusion matrix and MCC are in Table 4. Only 4.3% of days are above $Thres_{Ss}$, and 50% of the flash events. In the validation periods, the percentage of days above the threshold increased to 7.4% (Validation period 1) and 7.2% (VP 2), while only one of the three days in the first validation period was above the threshold. In the second validation period, there were no events in this category.

25 As was in $Thres_{Ss}$, if the three-day average temperature at 700hPa (lower troposphere) is to be above 271K, then the majority of the study area would be at above freezing temperatures. While snow could still fall at the highest elevations, it is likely that it would rain in lower regions, and that any snow on the ground may melt. The second requirement of $Thres_{Ss}$, specific humidity at 700hPa being higher than normal, also indicated possible rainfall. Therefore, $Thres_{Ss}$ indicated possible snowmelt and rainfall as triggers for flash events.

While both Fig. 8 and $Thres_{Ss}$ made physical sense, the atmospheric threshold for synoptic-spring events did not perform well in the validation periods, as well as for local convection events in Table 4. Both the percentage of flash events correctly predicted and the MCC values were lower, as well as the probability of occurrence. This may

have been ~~was partly~~ due to the small number of events, and possibly that synoptic events require more than two to correctly ~~the number of indicators trying to~~ capture the atmospheric triggering conditions.

4.3.4 Atmospheric thresholds: Synoptic, summer

Synoptic flash events in summer generally occurred with eight days of lower than normal temperature at 700hPa, and increased specific humidity at 850hPa (Fig. 9). As Sr flash events are associated with colder temperatures, compared to warmer temperatures for Ss flash events, this explains why T700 does not have a significant SI value when Sr and Ss are grouped together (Sect. 4.2). The probability of occurrence for this category was lower than any of the previous groups, with a maximum of 2%.

The Sr static threshold using the above atmospheric indicators corresponded to the following threshold ($Thres_{Sr}$):

- $Thres_{Sr}$: normalized specific humidity at 850hPa > 0.15 and eight-day mean normalized temperature at 700hPa < -0.40

The final group of values in Table 4 shows the performance of the above threshold. During the summer (July - November), 9.3% of days were above $Thres_{Sr}$, and 60% of synoptic summer flash events. However, the percentage of days above the threshold dropped in the two validation periods (8.1% and 5.9%), and no flash events were above the threshold. ~~This was true even though there were more flash events in the first validation period than during calibration. For four flash events, the Q850 value was too low, and for the other three flash events, the T700 was too warm. Therefore, while the threshold was reasonable during the calibration period, it did not hold for the validation periods.~~

Colder temperatures during a summer synoptic flash event are not unreasonable. Lower temperatures in summer may be associated with a front passing or cooler temperatures from prolonged cloud cover (and potentially rainfall). Similar to the other three atmospheric categories, high specific humidity indicated higher atmospheric moisture

and more likely rain. However, $Thres_{Sr}$ was unsuccessful in the validation period. It could be that different synoptic conditions lead to flash events in the two validation periods, or that the events were misclassified.

4.3.5 General discussion

As with any empirical threshold, accuracy and completeness of the inventory and weather data are important. During the classification and subsequent threshold analysis, it is possible that flash events were misclassified. The **spatial and temporal** resolution of ERA-Interim was not fine enough to explicitly resolve convection. Therefore, parameterization schemes are used, although Dee et al. (2011) show improvements in the convection parameterization from earlier reanalysis products. Furthermore, as the CAPE values take into account instability over the depth of the troposphere, CAPE values may be underestimated when convection is confined to a shallow layer (Niall and Walsh, 2005). As found in Section 4.1, it is likely that some events may have been misclassified as local convection or as synoptic.

The number of flash events limits the inferences that can be drawn from the results from this paper. The difficulty of developing atmospheric thresholds with few calibration events was borne out with the synoptic thresholds failing to capture the synoptic flash events in the validation period. However, for the convective flash events, the atmospheric threshold still captured 75% events in the validation periods. Furthermore, grouping all 63 flash events together, the atmospheric threshold still performs better than the weather station, although by a smaller margin.

As synoptic flash events generally performed the worst, further investigation on these nine events was undertaken. Based on the Hess-Brezowsky-Grosswetterlagen synoptic weather type (James, 2007), all synoptic flash event days except the 19/08/1996 event show evidence of a low pressure system near the study area. With the small number of events and variety of different locations of the fronts or low pressures, it may not have been possible to use the traditional threshold-type approach for these

5 events. A potential solution could be to identify key groups of indicators instead. Therefore, further investigation would be needed to determine if a) the events were correctly dated, b) caused by synoptic meteorological triggering conditions or c) triggered by non-meteorological factors.

10 Atmospheric thresholds, like most empirical thresholds, are reliant on near-complete inventories, and only speculations can be made about what may happen under unobserved conditions. Therefore, these methods cannot completely replace physically based models and other threshold analysis techniques. However, for the Ubaye Valley where local convection appears to be the main meteorological trigger of flash events, the atmospheric threshold improves on the local rainfall threshold. This methodology therefore has a potential to work in other areas where rainfall observations are not available, or not complete enough for the traditional empirical rainfall threshold.

15 5 Summary and Conclusions

20 The objective of this research was to develop empirical thresholds for rainfall triggered debris flows and flash floods using atmospheric indicators for the Ubaye Valley, France. Similar to rainfall thresholds, these thresholds could be used in risk assessment, early warning systems, or climate change projections. Empirical atmospheric thresholds were obtained for the Ubaye Valley, France, as well as weather station derived rainfall thresholds for comparison. Using both atmospheric indicators and weather station data, in each case two types of thresholds were obtained: a probability threshold and a static threshold, based on classification statistics and specifically the MCC value. The main conclusions are as follows:

- 25 – In general the atmospheric indicators performed better than the weather station threshold (average MCC value of 0.16 compare with 0.10, and higher probability of occurrence). They also performed better than rainfall thresholds using hourly data.

- The most important atmospheric indicators were CAPE and specific humidity at 850hPa. Both fit with convective precipitation being the main driver.
- Intense locally driven convection appears to be the main meteorological trigger for flash events in the study area (over 66% of events). Under these conditions, precipitation can be confined to a small area, and may explain why high precipitation values were not always recorded by the local weather stations.
- ~~Different atmospheric indicators in spring and summer supports snowmelt being an important antecedent condition for flash events in the study area in the spring.~~
- Even though the atmospheric thresholds performed better, there was still the high level of uncertainty in both the probabilistic thresholds and the static thresholds. This was especially true for the synoptic rainfall events.
- The number of observed events limits any statistical inference in the thresholds obtained, although this is partly mitigated using a validation dataset.
- The methodology also needs to be trialled in other locations. It may be that in areas where there is a stronger relationship between the local weather stations and rainfall at the location of the flash events that intensity - duration thresholds are more suitable.

Acknowledgements. This work was undertaken as part of the EU FP7 Marie Curie ITN project 'CHANGES' under Grant Agreement No. 263953. The authors would like to thank J.P Malet and A. Remaître for their input and help with collection of data. Data were provided by Météo France, and the ECMWF ERA-Interim data used in this study was obtained from the ECMWF data server.

References

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

- Alexandersson, H.: A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661–675, doi:10.1002/joc.3370060607, 1986.
- 5 Ansari, Z., Ahmed, W., Azeem, M.F. and Babu, A.V.: Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigation Sessions, *World of Computer Science and Information Technology Journal*, 1, 217–226, 2011.
- Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F. and Nielsen, H.: Assessing the accuracy of prediction algorithms for classification: an overview, *Bioinformatics*, 16, 412–424, 2000.
- 10 Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A. and Pizzio, M.: Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach, *J. Geophys. Res.-Earth*, 117, F4, doi:10.1029/2012jf002367, 2012.
- Breinl, K., Turkington, T. and Stowasser, M.: Stochastic generation of multi-site daily precipitation for applications in risk management, *J. Hydrol.*, 498, 23–35, 2013.
- 15 Brunetti, M.T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D. and Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, *Nat. Hazard Earth Sys.*, 10, 447–458, 2013.
- Brunetti, M.T., Luino, F., Vennari, C., Peruccacci, S., Biddoccu, M., Valigi, D., Luciani, S., Cirio, C.G., Rossi, M., Nigrelli, G., Ardizzone, F., Palma, M. and Guzzetti, F.: Rainfall thresholds for possible occurrence of shallow landslides and debris flows in Italy, in: *Dating torrential processes on fans and cones*, Schnewly-Bollschweiler, M., Stoffel, M. and Rudolf-Miklau, F. (Eds), Springer, Netherlands, 327–339, 2013.
- 20 Caine, N. : The rainfall intensity: duration control of shallow landslides and debris flows, *Geogr. Ann. A*, 62, 23–27, 1980.
- Carrara, A., Crosta, G. and Frattini, P.: Geomorphological and historical data in assessing landslide hazard, *Earth Surf. Proc. Land.*, 28, 1125–1142, doi:10.1002/esp.545, 2003.
- Cepeda, J., Colonnelli, S., Kronholm, K. et al.: Statistical and empirical models for prediction of precipitation-induced landslides. Report D1.5, SafeLand:Living with landslide risk in Europe, 137 pp, available at http://www.safeland-fp7.eu/results/Documents/D1.5_revised.pdf, 2010.
- 30 Chen, S.-T., Yu, P.-S. and Tang, Y.-H.: Statistical downscaling of daily precipitation using support vector machines and multivariate analysis, *J. Hydrology*, 385, 13–22, 2010.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally,

- 5 A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C.,
Thépaut, J.N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of
the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- D'Este, C. and Rahman, A.: Similarity weighted ensembles for relocating models of rare events,
in: *Multiple Classifier Systems*, Zhou, Z.H., Roli, F., and Kittler, J. (Eds.), Springer Berlin
10 Heidelberg, 25–36, 2013.
- Done, J., Craig, G., Gray, S., Clark, P. and Gray, M.: Mesoscale simulations of organized
convection: Importance of convective equilibrium, *Q. J. Roy. Meteor. Soc.*, 132, 737–756,
doi:10.1256/qj.04.84, 2006.
- Flageollet, J.-C., Maquaire, O., Martin, B. and Weber, D.: Landslides and climatic conditions
15 in the Barcelonnette and Vars basins (Southern French Alps, France), *Geomorphology*, 30,
65–78, doi:10.1016/s0169-555x(99)00045-8, 1999.
- Frattini, P., Crosta, G. and Sosio, R.: Approaches for defining thresholds and re-
turn periods for rainfall-triggered shallow landslides, *Hydrol. Process.*, 23, 1444–1460,
doi:10.1002/hyp.7269, 2009.
- 20 Giannecchini, R.: Relationship between rainfall and shallow landslides in the southern Apuan
Alps (Italy), *Nat. Hazards Earth Syst. Sci.*, 6, 357–364, doi:10.5194/nhess-6-357-2006, 2006.
- Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C.: Rainfall thresholds for the initia-
tion of landslides in central and southern Europe, *Meteorol. Atmos. Phys.*, 98, 239–267,
doi:10.1007/s00703-007-0262-7, 2007.
- 25 Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C.: The rainfall intensity/duration control of
shallow landslides and debris flows: an update, *Landslides*, 5, 3–17, 2008.
- Hewitson, B.C. and Crane, R.G.: Consensus between GCM climate change projections with
empirical downscaling: precipitation downscaling over South Africa, *Int. J. Climatol.*, 26,
1315–1337, 2006.
- 30 Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J. and Tveito,
O.E.: Classifications of atmospheric circulation patterns, *Ann. N.Y. Acad. Sci.*, 1146, 105–
152, 2008.
- Ibsen, M. L. and Brunsden, D.: The nature, use and problems of historical archives for the tem-
poral occurrence of landslides, with specific reference to the south coast of Britain, Ventnor,
Isle of Wight, *Geomorphology*, 15, 241–258, 1996.
- Jaiswal, P. and van Westen, C.J.: Estimating temporal probability for landslide initiation along
5 transportation routes based on rainfall thresholds, *Geomorphology*, 112, 96–105, 2009.

- James, P.M.: An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe, *Theor. Appl. Climatol.*, 88, 17–42, doi:10.1007/s00704-006-0239-3, 2007.
- Jeong, D., St-Hilaire, A., Ouarda, T.M.J. and Gachon, P.: CGCM3 predictors used for daily temperature and precipitation downscaling in Southern Québec, Canada, *Theor. Appl. Climatol.*, 107, 389–406, doi:10.1007/s00704-011-0490-0, 2012.
- 10 Kannan, S. and Ghosh, S.: Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output, *Stoch. Env. Res. Risk. A.*, 25, 457–474, doi:10.1007/s00477-010-0415-y, 2011.
- Kenawy, A., López-Moreno, J. and Vicente-Serrano, S.: Summer temperature extremes in northeastern Spain: Spatial regionalization and links to atmospheric circulation (1996-2006), *Theor. Appl. Climatol.*, 113, 387–405, doi: 10.1007/s00704-012-0797-5, 2013.
- 15 Kendall, M.: Rank correlation methods, 4th edition, Griffin, London, U.K., 1970.
- Maddox, R.A., Chappell, C.F. and Hoxit, L.R.: Synoptic and meso- α aspects of flash flood events, *B. Am. Meteorol. Soc.*, 60, 115–123, doi:1175/1520-0477-60.2.115, 1979.
- 20 Malet, J.P., Durand, Y., Remaître, A., Maquaire, O., Etchevers, P., Guyomarch, G., Déqué, M. and van Beek, L.P.H.: Assessing the influence of climate change on the activity of landslides in the Ubaye Valley, in: Proceedings of the International Conference on Landslides and Climate Change — Challenges and Solutions, McInnes, R. and Jakeways, J. and Fairbank, H. and Mathie, E. (Eds), Wiley, United Kingdom, 195–205, 2007.
- 25 Marsh, P., Brooks, H. and Karoly, D.: Preliminary investigation into the severe thunderstorm environment of Europe simulated by the Community Climate System Model 3, *Atmos. Res.*, 93, 607–618, doi:10.1016/j.atmosres.2008.09.014, 2009.
- Mason, S. and Graham, N.: Conditional probabilities, relative operating characteristics and relative operating levels, *Weather Forecast.*, 14, 713–725, 1999.
- 30 Merz, R. and G. Blöschl: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, 2008.
- Meyer, N.K., Dyrddal, A.V., Frauenfelder, R., Etzelmaller, B. and Nadim, F.: Hydrometeorological threshold conditions for debris flow initiation in Norway, *Nat. Hazard Earth Sys.*, 12, 3059–3073, doi:10.5194/nhess-12-3059-2012, 2012.
- 5 Molini, L., Parodi, A., Rebora, N. and Craig, G. C.: Classifying severe rainfall events over Italy by hydrometeorological and dynamical criteria, *Q. J. Roy. Meteor. Soc.*, 137, 148–154, doi:10.1002/qj.741, 2011.

- Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C. and Anquetin, S.: A numerical study of three catastrophic precipitating events over southern France. I: Numerical framework and synoptic ingredients, *Q. J. Roy. Meteor. Soc.*, 134, 111-130, 2008.
- 10 Niall, S. and Walsh, K.: The impact of climate change on hailstorms in southeastern Australia, *Int. J. Climatol.*, 25, 1933–1952, 2005.
- Nuissier, O., Joly, B., Joly, A., Ducrocq, V. and Arbogast, P.: A statistical downscaling to identify the large-scale circulation patterns associated with heavy precipitation events over southern France, *Q. J. Roy. Meteor. Soc.*, 137, 1812-1827, doi:10.1002/qj.866, 2011.
- 15 Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J. and Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine - Carpathian range, *J. Hydrol.*, 394, 78–89, doi:10.1016/j.jhydrol.2010.05.015, 2010.
- Petrow, T., Zimmer, J. and Merz, B.: Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns, *Nat. Haz. Earth Sys.*, 9, 1409–1423, doi:10.5194/nhess-9-1409-2009, 2009.
- Pettitt, A. N.: A non-parametric approach to the change-point problem, *J. Roy. Stat. Soc. C-App*, 28, 126–135, 1979.
- 25 Pistornik, G., Schneider, S. and Wittmann, C.: Detection of convective initiation by objective analysis methods and its use for precipitation nowcasting, 6th European Conference on Severe Storms, Palma de Mallorca, Spain, 3–7 October 2011, 2011.
- Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation, *Journal of Machine Learning Technologies*, 2, 37–67, 2011.
- Quan Luna, B., Blahut, J., van Westen, C.J., Sterlacchini S., van Asch, T.W.J and Akbas, S.O.: The application of numerical debris flow modelling for the generation of physical vulnerability curves, *Nat. Haz. Earth Sys.*, 11, 2047–2060, doi:10.5194/nhess-11-2047-2011, 2011.
- 30 Rousseeuw, P.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53–65, 1987.
- Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundscha, Y., Ribalaygua, J. and Schmih, T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, *J. Geophys. Res.-Atmos.*, 112, D4, doi:10.1029/2005jd007026, 2007.
- 5 Seltzer, M.A., Passarelli, R.E. and Emanuel, K.A.: The possible role of symmetric instability in the formation of precipitation bands, *J. Atmos. Sci.*, 42, 2207–2219, 1985.

- Sene, K.: Debris flows, in: *Flash Floods*, Springer, Netherlands, 271–291, doi: 10.1007/978-94-007-5164-4_9, 2013.
- 10 Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L.A.T., Angelis, C.F., Salio, P., Kidd, C., Huffman, G.J., de Castro, M.: Global precipitation measurement: Methods, datasets and applications, *Atmos. Res.*, 104–105, 70–97, 2012.
- Tarolli, P. and Borga, M. and Morin, E. and Delrieu, G.: Analysis of flash flood regimes in the North-Western and South-Eastern Mediterranean regions, *Nat. Hazard Earth Sys.*, 12, 1255–1265, doi:10.5194/nhess-12-1255-2012, 2012.
- 15 Tobin, C., Nicotina, L., Parlange, M. B., Berne, A. and Rinaldo, A.: Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region, *J. Hydrol.*, 401, 77–89, doi:10.1016/j.jhydrol.2011.02.010, 2011.
- Trapp, R., Diffenbaugh, N. and Gluhovsky, A.: Transient response of severe thunderstorm forcing to elevated greenhouse gas concentrations, *Geophys. Res. Lett.*, 36, L01703, doi:10.1029/2008GL036203, 2009.
- 20 van Asch, T.W.J., Tang, C., Alkema, D. Zhu, J. and Zhou, W.: An integrated model to assess critical rainfall thresholds for run-out distances of debris flows, *Nat. Hazards*, 1-13, doi: 10.1007/s11069-013-0810-z, 2013.
- Wang, X., Chen, H., Wu, Y., Feng, Y. and Pu, Q.: New techniques for the detection and adjustment of shifts in daily precipitation data series, *J. Appl. Meteorol. Clim.*, 49, 2416–2436, doi:10.1175/2010jamc2376.1, 2010.
- 25 Wijngaard, J. B. and Klein Tank, A. M. G. and Können, G. P.: Homogeneity of 20th century European daily temperature and precipitation series, *Int. J. Climatol.*, 23, 679–692, 2003.

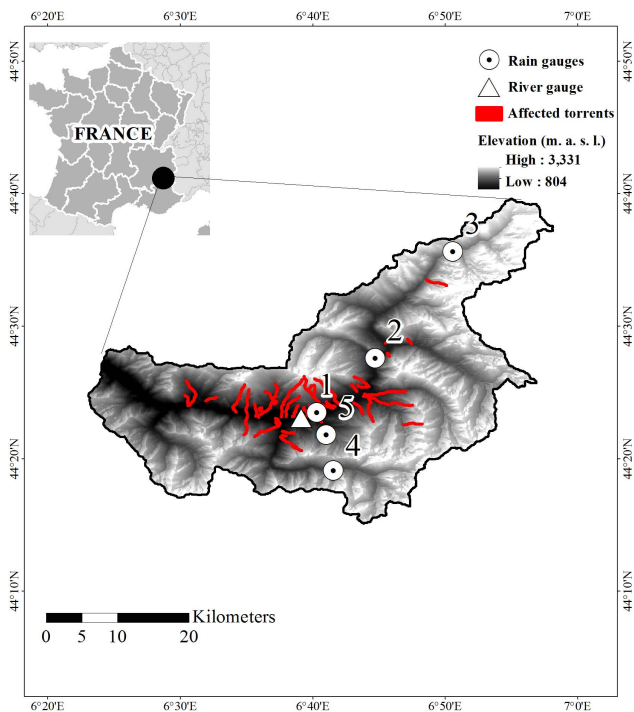


Fig. 1. The study area including the location of rain gauges and a single river gauging station. Red lines depicts the affected torrents where debris flows or flash floods occurred between 1979 and 2010 (map based on Breinl et al., 2013)

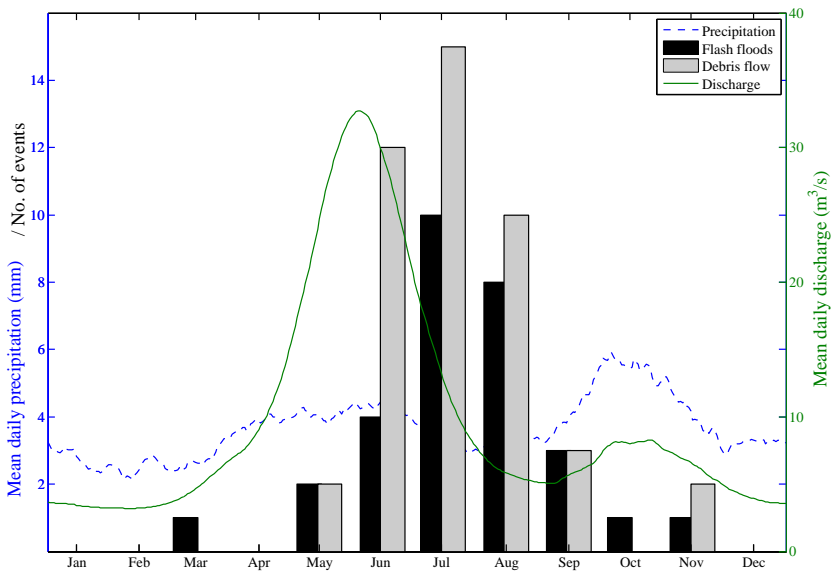


Fig. 2. Running 30 day mean daily precipitation and discharge for the period 1979-2009, for the Barcelonnette weather station and river gauge in the Ubaye River. The bar graph displays the number of flash floods and debris flows observed between 1979 and 2010.

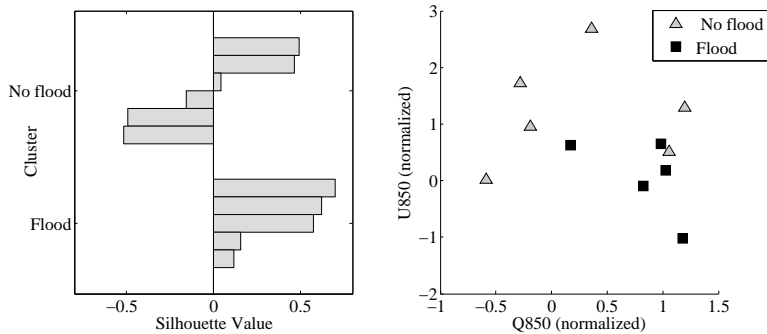


Fig. 3. A worked example of calculating SI values. The right side plots the specific humidity at 850hPa and U wind at 850hPa for five flood days and six non-flood days. These values were then used to derive the individual silhouette values in the plot on the left.

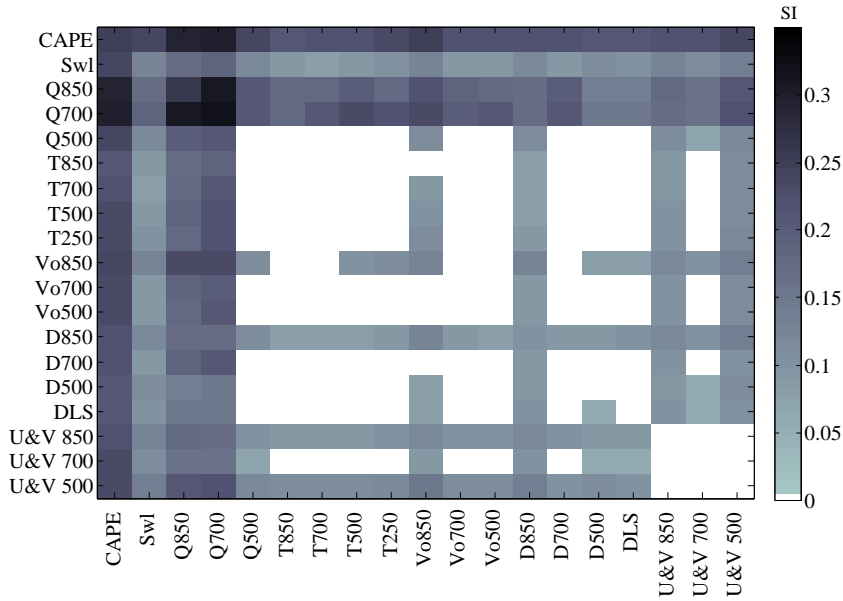


Fig. 4. The SI value for each pair of atmospheric indicators in Table 2 for local convection events using daily values. Any value that was not significant at $p = 0.05$ level was given a value of zero.

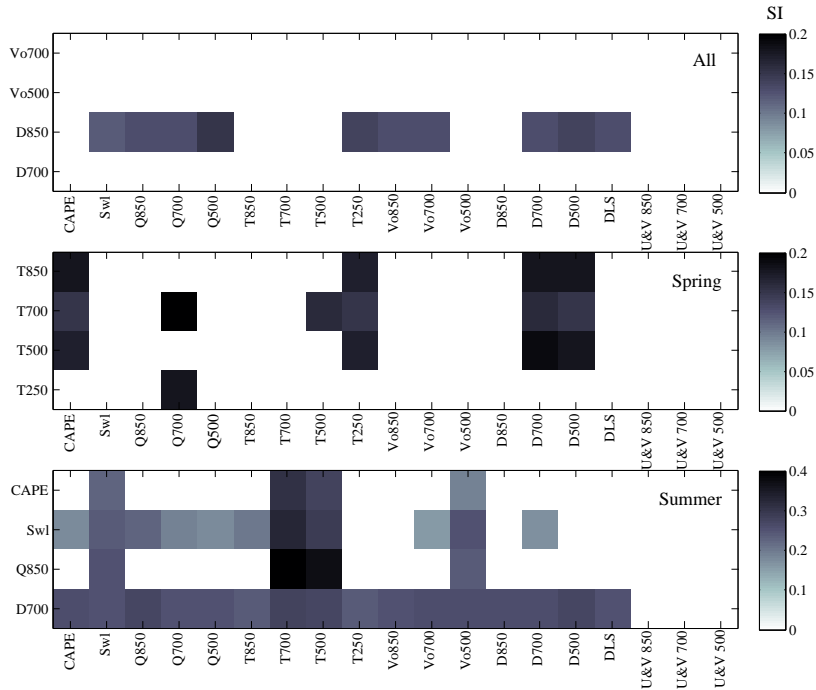


Fig. 5. Top: The SI value for each pair of atmospheric indicators for all synoptic events using the daily value and the mean value over 10 days. Middle: The SI value for each pair of atmospheric indicators for the Ss events (3-day and 8-day averages). Bottom: The SI value for each pair of atmospheric indicators for Sr using (daily value and 8-day average). Any value that was not significant at $p = 0.10$ level was given a value of zero.

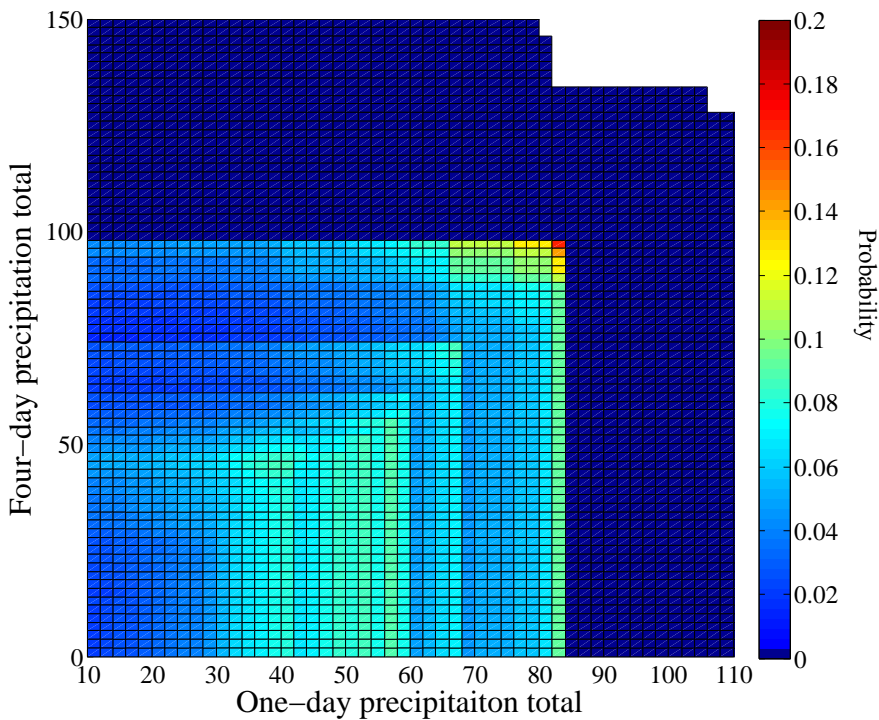


Fig. 6. Probability of a flash event based on 1 day and 4 day precipitation totals from a local rain gauge. Dark blue indicates zero probability of occurrence.

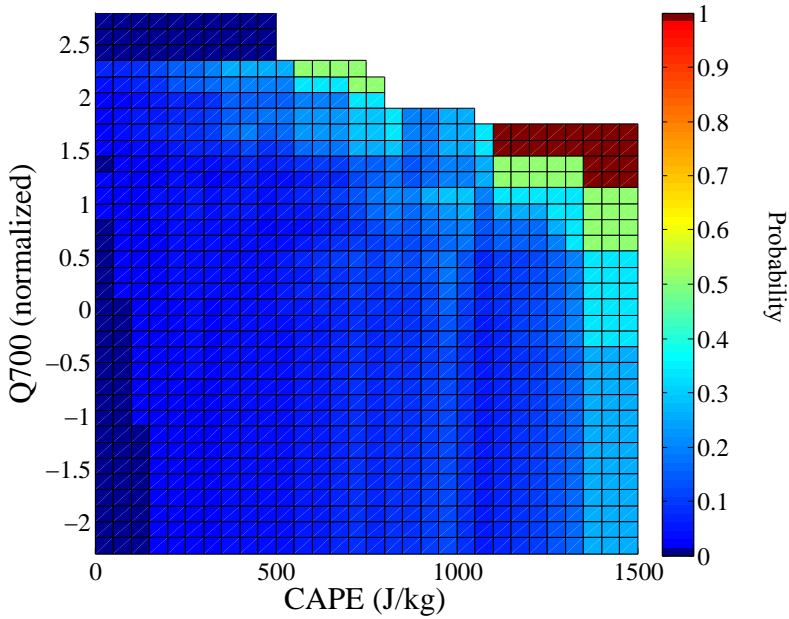


Fig. 7. Probability of a local convection flash event based on atmospheric indicators CAPE and normalized specific humidity at 700hPa (between 1989-2003).

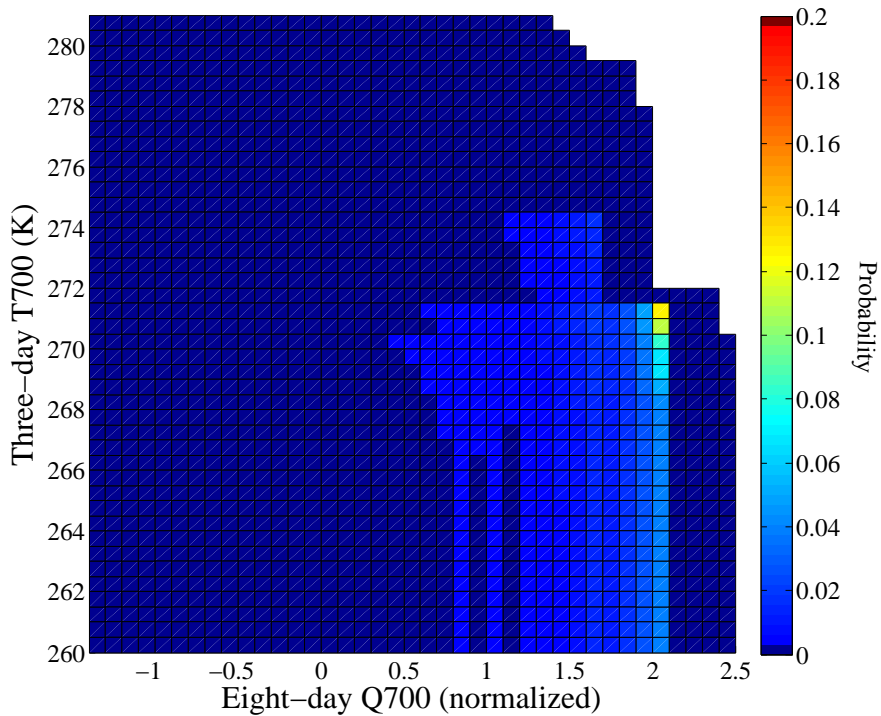


Fig. 8. Probability of a flash event from spring synoptic rainfall based on eight-day mean specific humidity at 700hPa and three-day mean temperature at 700hPa between 1989-2003.

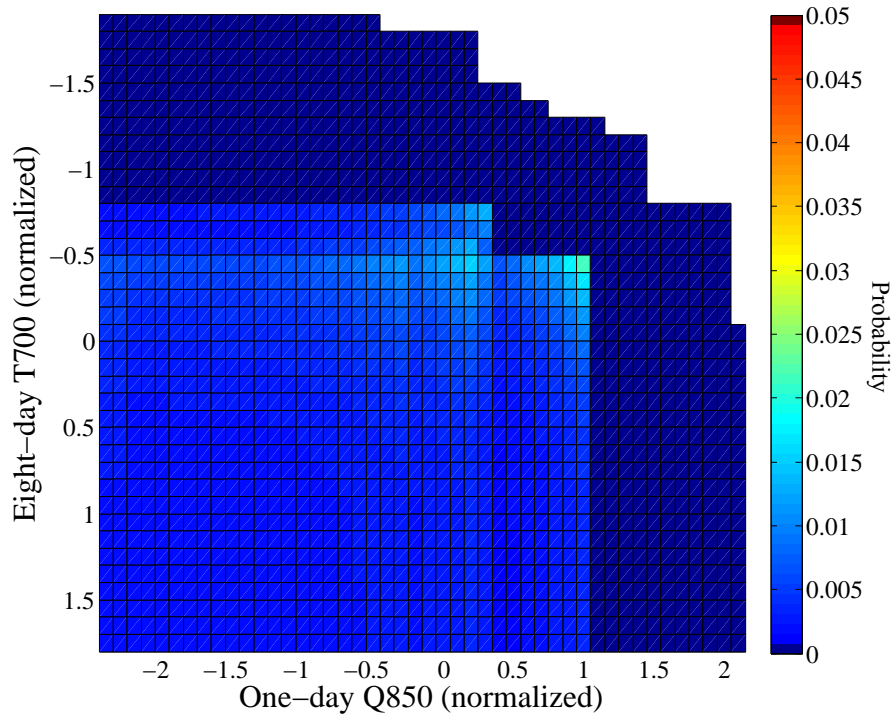


Fig. 9. Probability of a flash event from summer synoptic rainfall based on eight-day mean temperature at 700hPa and three-day mean specific humidity at 850hPa between 1989-2003. The y-axis is inverted to highlight that the figure represents the probability of a flash event given that T700 is less than a particular value and Q850 is greater than a particular value.

Table 1. Weather station information for the Ubaye Valley. The numbers refer to those in Fig. 1. The useable years show the percentage of years that are homogeneous and have at least 99% of days where the gauge was working. T = temperature, P = precipitation. The final column is the mean annual total precipitation, where applicable.

Site	Elevation (m a.s.l.)	Variable	Time Period	Useable Years	Summary (mm)
1. Barcelonnette	1,152	T	1961-2010	92%	
		P	1928-2010	80%	740
2. Condamine	1,325	P	1955-2004	98%	670
3. Saint-Paul	1,903	P	1971-2010	90%	930
4. Uvernet	1,660	P	1955-2010	95%	980
5. Super - Sauze	1,950	P	1996-2004	Summer only	

Table 2. ERA-interim variables used in this study, along with abbreviations used. A brief description of each variable is also given.

Variable	Pressure level	Description
Precipitation (RR)	Surface	Rain and snow
CAPE	Surface	Estimate of instability of the atmosphere
Soil moisture(SWL)	Surface	Soil moisture for top layer (0-7cm).
Specific humidity, (Q)	850hPa, 700hPa, 500hPa	Atmospheric moisture
U&V wind	10m, 850hPa, 700hPa, 500hPa	Meridional (V) and zonal (U) wind speed
Temperature (T)	850hPa, 700hPa, 500hPa, 250hPa	Temperature
Vorticity (Vo)	850hPa, 700hPa, 500hPa, 250hPa	Local spinning motion of the air
Divergence(D)	850hPa, 700hPa, 500hPa	Expansion or spreading out of a vector field

Table 3. Classification of the flash events in the calibration period 1989-2003. The list contains the date of event, the t_{CAPE} value, and its category : Ls - local rainfall, spring Lr - local rainfall, summer, Ss synoptic rainfall, spring , Sr synoptic rainfall, summer.

Local convection				Synoptic			
Date	t_{CAPE}	Group		Date	t_{CAPE}	Group	
1.	18/06/1989	7	Ls	1.	03/11/1991	-1	Sr
2.	14/08/1990	13	Lr	2.	02/06/1992	-1	Ss
3.	29/09/1991	46	Lr	3.	12/07/1993	0	Sr
4.	06/10/1991	24	Lr	4.	11/05/1994	0	Ss
5.	01/06/1992	19	Ls	5.	13/05/1994	0	Ss
6.	18/06/1992	93	Ls	6.	06/07/1996	0	Sr
7.	21/07/1992	8	Lr	7.	19/08/1996	1	Sr
8.	27/09/1992	9	Ls	8.	25/07/1997	0	Sr
9.	10/07/1993	12	Ls	9.	22/03/2001	0	Ss
10.	05/11/1994	77	Lr				
11.	28/08/1997	9	Lr				
12.	12/08/2000	9	Lr				
13.	13/08/2000	20	Lr				
14.	23/11/2000	8	Lr				
15.	26/07/2001	55	Lr				
16.	05/06/2002	27	Ls				
17.	23/06/2002	15	Ls				
18.	27/07/2003	10	Lr				
19.	05/08/2003	110	Lr				
20.	08/08/2003	20	Lr				

Table 4. Results for the static threshold for the calibration period (1989-2003) and validation period 1 (1979-1988) and validation period 2 (2004-2010). The total number of days (TP+FN+FP+TN) is the same for the weather station and local convection. The number is lower for Ss and Sr, as they are only applied over spring and summer respectively.

	TP	FN	FP	TN	MCC
Weather station					
Calibration	16	13	412	5037	0.13
Validation 1	10	18	244	3381	0.10
Validation 2	2	4	145	2406	0.06
Local convection					
Calibration	15	5	332	5126	0.17
Validation 1	14	5	255	3397	0.18
Validation 2	4	1	171	2381	0.13
Synoptic -snowmelt					
Calibration	2	2	93	2198	0.10
Validation 1	1	2	105	1422	0.05
Validation 2	0	0	70	967	-
Synoptic -rainfall					
Calibration	3	2	193	2097	0.09
Validation 1	0	6	115	1409	0
Validation 2	0	1	60	1010	0