

# Are traditional hazard-based weather warnings anticipating local impact?

Jordi Morales<sup>a,b,\*</sup>, Xavier Llor<sup>b</sup>, Andreas Kaltenbrunner<sup>c</sup>, Agata Lapedriza<sup>a,d</sup>

<sup>a</sup>*AIWELL Lab, Universitat Oberta de Catalunya, Rambla del Poblenou, 154, Barcelona, 08018, Spain*

<sup>b</sup>*Hydrometeorological Innovative Solutions, Jordi Girona 1-3, ParcUPC K2M, Barcelona, 08034, Spain*

<sup>c</sup>*Department of Engineering, Universitat Pompeu Fabra, Carrer de Roc Boronat 138, Barcelona, 08018, Spain*

<sup>d</sup>*Institute for Experimental AI, Northeastern University, 360 Huntington Ave, Boston, 02115, MA, USA*

---

## Abstract

Hazard-based weather warnings alert decision makers, responders, and the population about meteorological risks, yet often overlook data on vulnerability and exposure, relying instead on fixed or climatology-based thresholds. This raises questions about their effectiveness at higher resolutions, as weather impacts are ultimately determined by local characteristics. This work addresses this gap by providing the first quantitative and high-resolution assessment of how traditional heavy rain and wind gust warnings correlate with actual impacts, using emergency calls as an impact proxy. The study works at the municipal and hourly resolution across Catalonia, in northeastern Spain, over six years (from October 2018 to February 2025), analysing warnings from two meteorological agencies: the national (AEMET, part of the EU MeteoAlert programme) and regional (SMC). We first compare the spatial distribution of issued warnings and emergency calls, assessing how they align with the rain and wind climatology of the region. Following this, overall warning performance metrics are calculated assuming a real-time setting. Further experiments analyse other critical operational variables, such as lead-time and monthly patterns. Our results reveal that while traditional

---

\*Corresponding author.

*Email address:* jordi.morales@hyds.es (Jordi Morales)

warnings can identify general impact zones, they are severely limited when considered at local scales due to a high number of false alarms, which can reduce their public use and trust. Based on these findings, we not only provide a solid evidence base on how current weather warnings relate to actual impacts, but also aim to inform and accelerate the transition from traditional, hazard-centric systems towards more actionable and precise impact-based warning systems.

*Keywords:* Impact-Based Warnings, Weather Warnings, Operational Systems, Disaster Risk Management, Impact Assessment, Severe Weather, Catalonia

---

## 1. Introduction

Weather warnings issued by official meteorological agencies stand as a critical element for reducing the impacts of increasingly frequent, prolonged, and intense climate extremes, driven by factors such as climate change or urbanisation [1, 2, 3]. Being the standard for communicating potentially harmful weather phenomena, they inform citizens, decision makers, and emergency responders about the up-to-date situation on current and future weather patterns that may cause negative impacts on both people and property. In Europe, the EMMA/MeteoAlarm project [4] was born in 2002 as an initiative from EUMETNET [5] to unify how weather warnings around Europe are disseminated, and to create a warning platform that is consistent across country borders. The service ultimately went online in 2007, and since then has provided European citizens with a Multi-Hazard Early Warning System (MHEWS) to consult weather warnings from 38 European National Meteorological and Hydrological Services (NMHSs). The initiative also served to design standards on how to issue multi-hazard warnings and foster cooperation between European NMHSs.

In 2015, the Sendai Framework for Disaster Risk Reduction [6] was adopted by the United Nations member states as the agenda to follow to reduce disaster losses and risks globally during the 2015-2030 period. The resolution signals the fragility of our climatic situation and how the increase in both frequency and intensity of natural disasters [7] will cause more and more suffering if it is not targeted adequately and decisively. In this context, it is necessary to understand disaster risks and move from a reactive to a proactive stance against natural hazards [8, 3]. DRR is evolving from a technical to

a humanitarian perspective [9], and consequently, approaches need to move from targeting hazards to impacts (i.e., the realisation of risks). Following these premises, weather warnings, intended to communicate the severity of impending disasters, should be clear and adapt to the needs of the general public in a way that actually makes sense for them: inform about the risks actually posed by the hazards and state strategies to mitigate the damage in their particular context [1]. This new paradigm, which, as exemplified in Table 1, shifts from hazard-based to impact-based warnings (IBWs), has the potential to engage more effectively with vulnerable actors by reducing misunderstandings of warning terminology and “warning fatigue” [10], and it is essential for motivating society into taking preparatory action against natural hazards.

However, there is a lack of research analysing how current warnings relate to actual impacts. This is essential to establish a baseline of the current state and to quantify the potential improvement of future IBWs. In this work, we aim to fill this gap by assessing how traditional, climatology-based warnings for heavy rain and wind gusts correlate with real impacts. The study covers Catalonia, in northeastern Spain, from October 2018 to February 2025<sup>1</sup>. As impact indicators, we use 112 emergency call data related to floods and wind gusts, a powerful proxy that captures the direct population-level consequences. As for the weather warnings, we analyse official warnings issued by the Spanish National Meteorological Agency (AEMET) [11], which operates at the national level and participates in MeteoAlarm, and the Meteorological Service of Catalonia (SMC) [12], providing an additional, more regional perspective.

## 2. Related work

### 2.1. Emergency Calls as Impact Indicators

Emergency calls have been used in the past to estimate the extent of weather damage and impact. Schuster et al. [14] studied three hailstorms in Sydney, Australia (April 1999, December 2001, February 2002), comparing hazard data from hailstone size observations and weather radar against damage data, including emergency calls data as on-the-ground impact indicators.

---

<sup>1</sup>Excluding the period between March 14th to December 31st, 2020, due to a lack of data.

Table 1: Comparison between hazard-based and impact-based warnings, including specific examples.

Warning Type	Description	Example
Hazard-based	Warnings providing information on physical meteorological magnitudes and thresholds.	"Severe thunderstorms are expected today with wind gusts exceeding 60 mph." [13]
Impact-based	Warnings communicating the potential consequences and risks to society caused by the hazard.	"Severe thunderstorms with gusts over 60 mph will result in damage to trees and power lines." [13]

Rossi et al. [15] use emergency calls in the context of convective storms. They introduce a methodology for automatically determining storm hazard levels. They employ tracking algorithms to follow and predict the storm path, while linking real-time impact data in the form of emergency calls. More recently, several case studies have demonstrated the value of calls for refining flood risk mapping and creating impact catalogues [16, 17, 18], and Sirenko et al. [19] develop a framework for assessing spatio-temporal urban vulnerability patterns through ambulance calls.

## 2.2. Impact-Based Warnings

Kaltenberger et al. [20] conducted a survey between 2017 and 2018 targeting 37 NMHSs involved in MeteoAlarm. Ultimately, 32 agencies responded, providing valuable information about the current status of how impact data is being used (or is planned to be used), and the way towards Impact-based Warnings (IBWs). In this sense, the survey highlights major gaps, which "can be identified as lack of: impact data, technical standards, cross-border exchange of impact data, impact-databases, verification methods, and resources". Steps are being taken towards IBWs, but the path towards a fully established and standardised methodology remains long. For context, the study indicates how a majority of NMHSs plan to implement some kind of impact-based criteria in the near future, but at the date of the survey, they relied mostly on fixed thresholds (31%) or climatology-based thresholds (66%).

Moving to these more informed warnings is important: research indicates that impact-based warnings, especially when paired with behavioural information, show significant potential for increasing the public's intention to take protective action [21]. One example involving IBWs can be found in the United Kingdom, where they have been issuing locally-calibrated,

impact-based warnings since 2011, moving away from a phenomenon-based system. A post-event survey following Storm Doris, analysed in Taylor et al. [22], revealed that while the personalised nature of IBWs can enhance the perception of risk and motivate response, their effectiveness depends on the public's understanding of their locally-calibrated nature. Consequently, the study underscores the importance of country-specific research to identify particularities in how warnings are interpreted. Insights like these are precisely what is needed, as implementing IBWs is complex and requires a solid foundation: a strong understanding of weather impacts, access to vulnerability data, and an effective organizational framework. Rushing their deployment before these elements are in place is a significant concern [23].

### *2.3. Weather Warnings and Impact Assessment in Spain*

A primary challenge for advancing impact-based forecasting is the scarcity of systematic impact data. AEMET's SINOBAS (Singular Atmospheric Observations Notification System) [24] was created in 2013 as an open database that allows people to report local, infrequent, and significant meteorological events with high potential societal impact. However, its focus remains primarily on rare meteorological hazards rather than the further consequences and impacts. Addressing this gap, Gaztelumendi et al. [25] describe a methodology for building a comprehensive catalogue of extreme weather events. This approach involves gathering data not only on the meteorological hazards but also on their associated societal and environmental impacts. This methodology has been made operational by the Basque Meteorology Agency (Euskalmet) to create the Basque Country Impact Weather Catalogue, serving as a pioneering model for how impact data can be systematically catalogued to better understand and anticipate the consequences of severe weather in the future.

Moving to alerts, the official warnings issued by agencies such as AEMET or SMC are fundamentally based on climatological thresholds. For AEMET in particular, as outlined in the National Meteorology Alert Plan [26], their purpose is to warn the population at risk, yet they do not explicitly incorporate data on vulnerability and exposure. Consequently, while they are effective for anticipating meteorological hazards [27], their general nature may hinder their effectiveness as warning tools and potentially lead to alert fatigue [23, 10]. However, due to a significant gap in research that quantitatively evaluates how these hazard-based warnings translate into actual

on-the-ground impacts, it is difficult to confidently draw conclusions. This work aims to address this gap.

### 3. Data

In this section, we introduce the impact and official weather warning data employed in our analysis. In particular, we introduce the different data sources in detail and explain how they are processed.

#### 3.1. Emergency calls

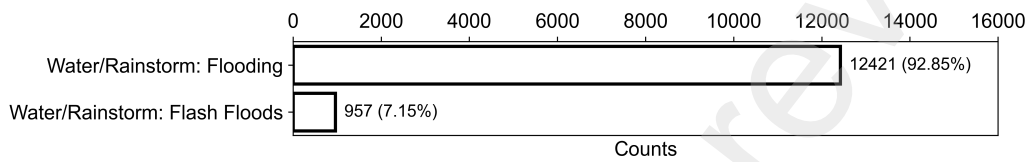
In this work, we use 112 emergency calls (equivalent to 911 in the US or Canada) as impact indicators. The data used for this study were provided by the Catalan Civil Protection authorities.

In particular, this data contains information about the position (longitude-latitude), time, and categorisation of individual calls based on the type of emergency. To ensure that precise impact indicators are considered in the study, call categories must be carefully chosen: for flood-related hazards, we consider only calls categorised as “flood” or “flash flood”. Similarly, for wind-related hazards, we choose those calls belonging to “wind storm” categories, which are related to damage of urban elements (e.g., scaffolding), and the tree and signage fall risk categories.

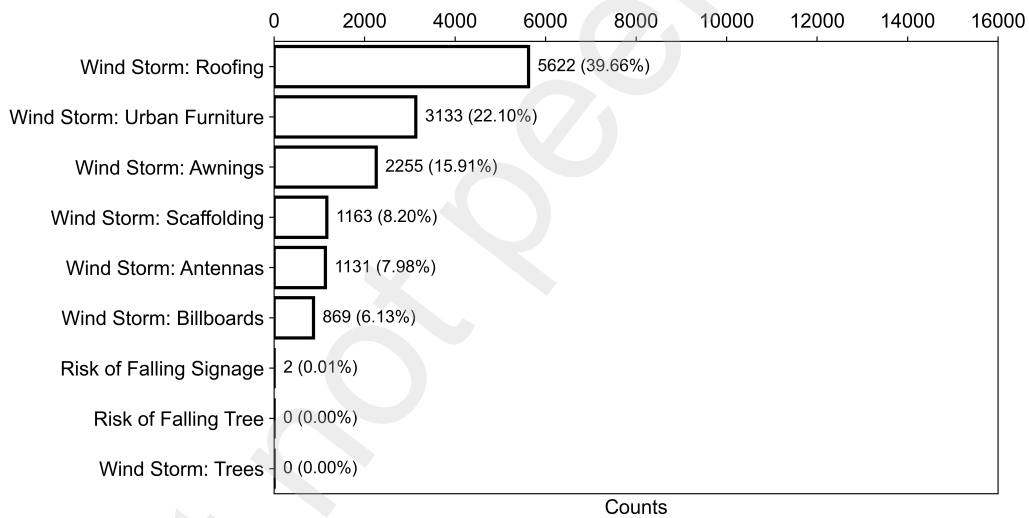
In Figures 1a and 1b, the amount and distribution of flood- and wind-related calls by type can be seen. For flooding emergencies, most of the calls (92.85%) belong to the “flooding” category, typically associated with fluvial floods but can also refer to flooding occurring within buildings. The remaining calls fall under the “flash flood” label (7.15%), which typically occur over shorter periods of time, and are more related to pluvial flooding. For wind-related calls, the distribution is less skewed, but most calls are concentrated on emergencies affecting roofing (39.66%), urban furniture (22.10%), and awnings (15.91%). Categories related to trees and falling signage, although considered, are practically non-existent for the duration of our study, probably because such events are reported directly to municipal authorities instead of the emergency service.

#### 3.2. Official warnings

Official warnings refer to those weather warnings issued by meteorological agencies to inform about meteorological hazards that could potentially pose a risk to people, activities, and infrastructure. In our research, we include



(a) Flood-related call type distribution.



(b) Wind-related call type distribution.

Figure 1: Call category distributions of (a) flood- and (b) wind-related emergency calls included in our dataset.

official warnings issued by both AEMET [11] and SMC [12]. Below, we explain the particularities of each one:

- **AEMET warnings:** AEMET warnings are issued by isoclimatic regions. These regions aggregate municipalities into broader areas sharing similar climatic characteristics. In Catalonia, after ignoring 5 regions that only cover the sea, 16 warning zones can be found. Each warning has one of four possible severity levels (from least to most severe hazard: “green”, “yellow”, “orange”, and “red”. “green” is used when there is no warning, although it is usually omitted). We convert these levels to numerical values, using 0 for “green”, up to 3 for “red”.

AEMET warnings can refer to different hazard types, but for our study, we only consider the categories of “Storm”, “1h Rain Accumulation”, and “12h Rain Accumulation” as flooding hazards, and “Wind gust intensity” and “Storm” as wind hazards. Storms are considered in both cases as, although typically related to rain, they are often accompanied by strong wind gusts (up to 70 km/h). Typical alert messages are based on physical thresholds (e.g., “1h Rainfall accumulation: 20mm”) or general hazard information (e.g., “[Storms] could be accompanied by strong or very strong wind gusts and occasional hail.”). For specific information about activation thresholds, which may vary between warning regions, please refer to the official source [28].

- **SMC warnings:** SMC warnings are issued at a regional level (by *comarques*, a sub-province administrative region scheme, similar to counties). These regions, although generally smaller than AEMET’s warning zones, are still composed of multiple municipalities. As of 2025, there are 43 regions, but this number has been changing slightly in recent years. We take this into account and, depending on the date of the warning, we use the adequate administrative boundaries.

For flood-related hazards, only warnings belonging to the categories of “Rain intensity” (20 or more millimetres of rain accumulation in 30 minutes) and “Rain accumulation” (more than 100 millimetres of rain accumulation in 24 hours) are taken into account. For wind-related hazards, “Wind gust intensity” (20 or more *km/h*, depending on the region). Warnings can be further categorized into prediction warnings or observation warnings (i.e., short-range predictions issued after a previously non-alerted hazard is observed). For this study, we only

consider prediction warnings. These are regular warnings, issued in advance based on weather forecasts, and cover 6-hour periods. For each of these warnings, a number of regions can be affected. Typical alert messages are based on physical thresholds, such as “Rainfall accumulation  $> 100$  mm/24 hours” or “Maximum wind gust  $> 108$  km/h”, representing the meteorological criteria that triggered the warning.

For SMC Warnings, the hazard level can take values between 1 and 6, from lowest to highest hazard degree. This levelling can also be understood following the classical “traffic light” categories of “green” (no warning, level 0), “yellow” (levels 1 and 2), “orange” (3 and 4), and “red” (5 and 6). In this work, we use the latter definition, taking the numerical level values.

## 4. Analysis methodology

### 4.1. Region and scope of the study

We focus on Catalonia (32,000 km<sup>2</sup>, 8.01 million inhabitants according to the Statistical Institute of Catalonia [29]), in northeastern Spain (See Fig. 2 for geographical context). The region has a typical Mediterranean climate, receiving  $\sim 700$ mm of annual rainfall with irregular interannual distribution, resulting in periods of drought and recurring flood events, including intense flash floods. Regarding wind, Catalonia, like most of Europe, is dominated by westerly winds. However, its northern and southern regions are prone to intense northerly winds.

In this study, we evaluate the relationship between weather warnings and the final consequences of predicted hazards. The spatial resolution of the analysis is at the municipal level (947 municipalities in Catalonia) and with an hourly temporal resolution ranging from November 2018 to February 2025 (both included), except for a gap in the data from March 14th to December 31st (both included) of 2020, due to the COVID-19 confinement and a later, unrelated gap of missing data in our dataset. Most of the results are provided using a lead-time of up to one hour (to reflect a typical “real-time” setting), but other lead-time values are also studied, and the results are presented in Section 6.2, as warning effectiveness is sensitive to this parameter. We also studied the performance of these systems when working at the original warning zone resolutions (isoclimatic zones for AEMET, *comarques* for SMC). Further discussions on the particular performance values will be discussed in Subsection 6.1.2.

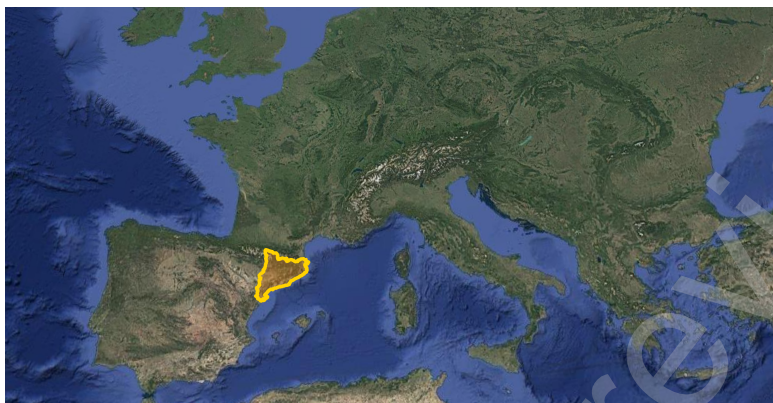


Figure 2: Map of Southern Europe with Catalonia highlighted in yellow.

#### 4.2. Definition of the evaluation framework

In this work, we assess whether the forecasted hazards contained within traditional, hazard-based weather warnings coincide with actual impact indicators, using 112 emergency call data as a proxy. We define this problem using a binary contingency table, where impacts represent the ground truth and warning activations represent the predictions. Specifically, we define an impact as the occurrence of at least one emergency call in a municipality within an hour. Similarly, a warning activation of level  $N$  is defined as the activation of a warning at a level greater than or equal to  $N$  in a municipality within an hour.

When we refer to “performance” or “effectiveness”, we specifically address the results that weather warnings obtain within this framework (i.e., the correlation with 112 impact proxies), rather than their capability to anticipate meteorological hazards.

#### 4.3. Metrics

The main metric considered in this study is the Critical Success Index (CSI; Eq. 1), a common metric used in meteorology and forecast verification [30], measuring the fraction of correctly predicted positive cases while penalizing for both false alarms and misses. While the CSI provides a general assessment of how warning activations relate to the impact proxies, a more detailed assessment and understanding requires the inclusion of the Probability of Detection (PoD; Eq. 2) and Precision (Eq. 3) must be also taken into account, as well as the total number of activations.

In particular, PoD measures the fraction of impacts detected, while the Precision serves as a complement to the False Alarm Ratio (FAR) and evaluates the proportion of forecasts that correctly detect an impact. It is important to note that the metrics presented in this section do not evaluate the operational success of the warnings themselves, but rather the relationship between these warnings and impacts (specifically 112 calls) within a binary contingency table framework.

$$\text{CSI} = \frac{\text{Hits}}{\text{Hits} + \text{Misses} + \text{False Alarms}} \quad (1)$$

$$\text{PoD} = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} \quad (2)$$

$$\text{Precision} = 1 - \text{FAR} = \frac{\text{Hits}}{\text{Hits} + \text{False Alarms}} \quad (3)$$

## 5. Geographical distribution of data

### 5.1. Population distribution

In total, 8,124,126 inhabitants live in Catalonia as of January 1st, 2025 [29], distributed across 947 municipalities. Approximately three-quarters of the total (5,959,941 inhabitants) live inside the *Barcelona Metropolitan Demarcation*, one of nine *vegueries* (high-level regional division in Catalonia). The second and third most populated ones are *Girona* with 814,218, and *Camp de Tarragona* with 564,162. Figure 3 shows the population census and density distributions for Catalonia. This highlights how Catalonia's population is heavily concentrated around major coastal cities, and particularly around the Barcelona area, with significant interior urban centres such as Lleida, Manresa, Vic, and Olot, although most of the territory is sparsely populated.

### 5.2. Warnings and impacts climatology

Regarding rainfall climatology, most of the territory is prone to short but intense episodes, particularly in coastal regions along the Prelitoral and Litoral mountain ranges [31, 32]. The Pyrenees and neighbouring mountains to the north, and the Empordà region (northeast) also register rain, but they are typically longer and less intense events. The exception is the Central

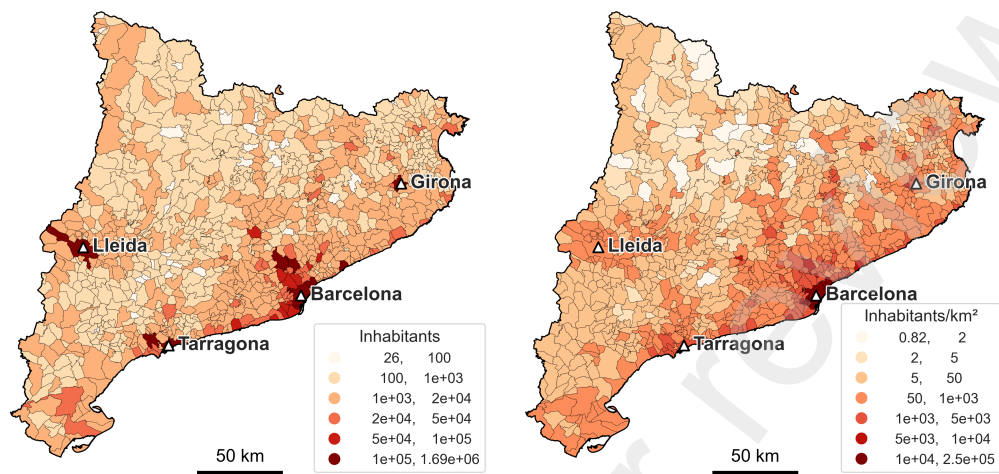
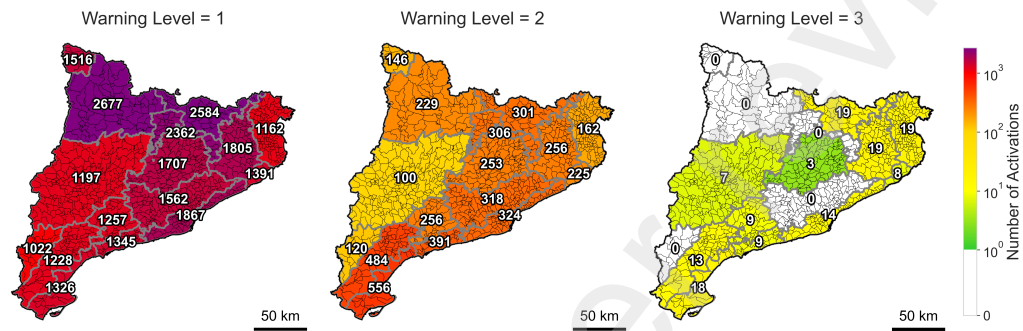


Figure 3: Catalonia's 2025 population census (left) and density (right), by municipality. Major urban centres are indicated for reference.

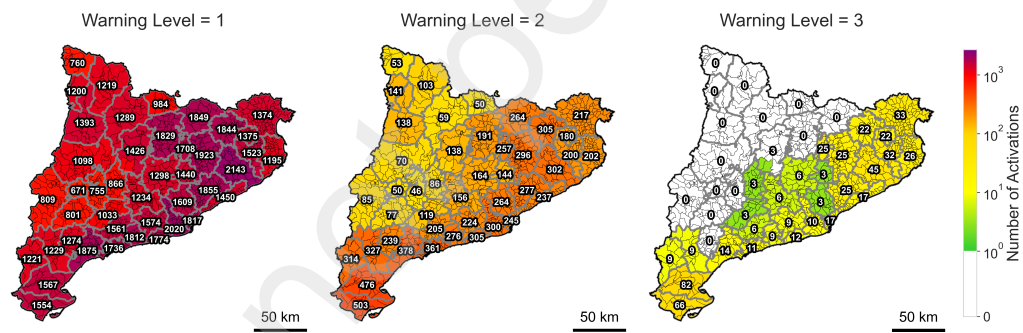
Depression (mid-western region), which generally receives less total rainfall compared to the rest of Catalonia.

Figure 4 shows the number of AEMET and SMC rainfall-related activations (total hours a warning was active) per zone for different warning levels. The distribution of warnings generally reflects the previously described climatology. As expected, the regions with the most activations are the Pyrenees and the coastal ranges, which are prone to the most frequent and intense rainfall. At Level 1, the northern mountains show the highest activity for AEMET warnings, consistent with their propensity for longer, less intense rain events, while SMC warnings are more frequent around the coast. However, as the warning level increases, the distribution shifts, and coastal regions become the most prone to activations in both cases, particularly toward the south. This aligns with the idea that short, intense downpours are more likely to trigger more severe warnings. An exception is found in some Level 3 AEMET activations in the interior, which are related to uncommon 12-hour accumulation events. The Central Depression (mid-west) appears to have a volume of Level 1 activations similar to southern regions, the Empordà (northeast) or the Aran Valley (northwest), highlighting how warnings are regionally calibrated to account for different local climates.

In Figure 5, the number of wind-related activations for both AEMET and SMC is shown. Considering the typical wind climatology in Catalonia,

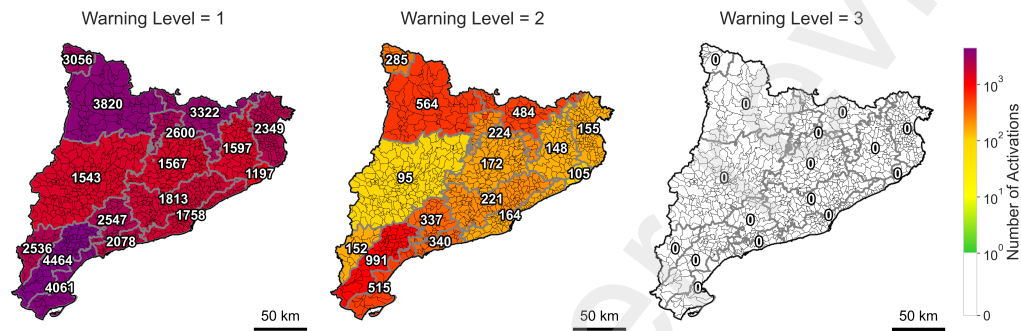


(a) AEMET rain-related activations (1-hour and 12-hour rain accumulation, and storm warnings).

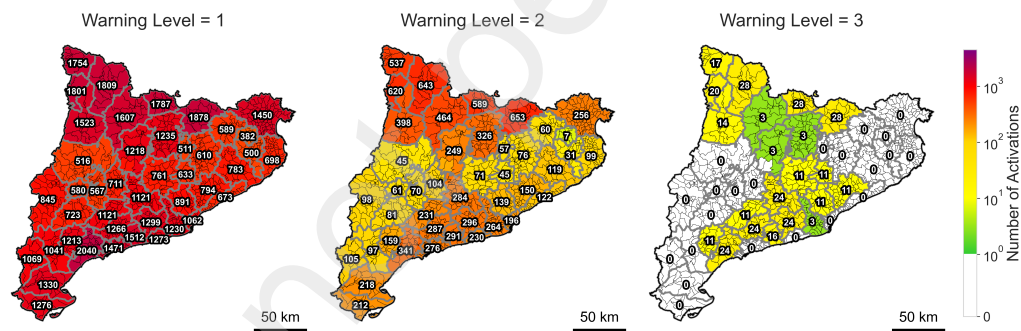


(b) SMC rain-related activations (rain intensity and rain accumulation).

Figure 4: Number of rain-related warning activations for both (a) AEMET and (b) SMC, with their respective warning zones. An activation is defined as an hour during which a warning at a specific level was active. Data covers the period from November 2018 to February 2025, with a gap between March 14 and December 31, 2020.



(a) AEMET wind-related activations (wind gust and storm warnings).



(b) SMC wind-related activations (wind gust warnings).

Figure 5: Number of wind-related warning activations for both (a) AEMET and (b) SMC, with their respective warning zones. An activation is defined as an hour during which a warning at a specific level was active. Data covers the period from November 2018 to February 2025, with a gap between March 14 and December 31, 2020.

most of the territory experiences gusts of up to 20 m/s a few times per year [31, 32, 33]. More intense winds occur primarily during the cold months and are concentrated in the Pyrenees, the northeastern Empordà region, and the Ebro Valley to the south. In these areas, northerly winds are channelled, giving rise to strong tramontane gusts. While wind warnings are issued across the entire territory (as thresholds are calibrated regionally), a clear divide is visible, with the highest frequency of Level 1 and 2 warnings occurring in the Pyrenees to the north and the southern regions along the coast, while more central areas have significantly fewer activations. This is common for both AEMET and SMC warnings. It is interesting to note that the Empordà region, despite being typically associated with strong winds, does not receive as many AEMET warnings as other parts of the Pyrenees, probably due to more severe activation thresholds. Regarding the most severe events, AEMET has not issued any Level 3 warnings for the period covered by this study. The SMC, in contrast, has issued several, affecting the central and western Pyrenees and areas along the Prelitoral and Litoral mountain ranges primarily.

Finally, flood- and wind-related impacts (Fig. 6) do not perfectly align with the described climatologies and warning distributions. Most reported impacts are concentrated in coastal regions, particularly in the central area around Barcelona, which is expected, as this area contains most of Catalonia's population. However, when we analyse the number of calls per 10,000 inhabitants per year (a measure of a municipality's susceptibility to making emergency calls), a different pattern emerges: the map of flood susceptibility shows that the Barcelona area becomes more homogeneous and actually concentrates most of the low values. In contrast, areas to the north, northeast, and south appear more susceptible, with some of the highest values observed towards the interior. Wind impact susceptibility shows even more homogeneity across the entire coast, with the highest values almost exclusively located in the interior and the northeast.

It should be noted that the susceptibility metric may not be fully reliable in municipalities with very few calls. Due to data scarcity, a single call can lead to a high susceptibility value, potentially overstating the risk in small municipalities where the underlying call distribution cannot be accurately approximated. Nonetheless, our measure of susceptibility serves to show how the contrast between where impacts are most frequent (high-population areas) and where susceptibility is highest (some low-population areas) underscores how population distribution modulates the visible consequences of

climatological hazards.

## 6. Results

### 6.1. Overall performance

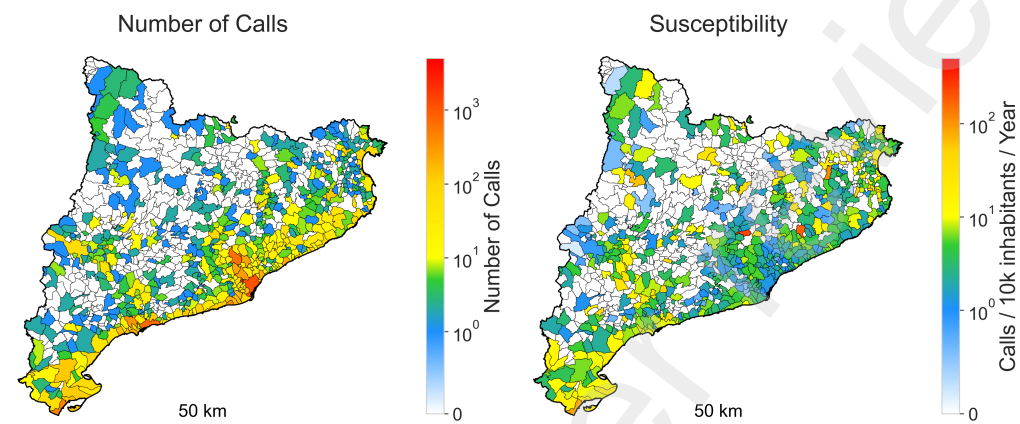
In this section, we provide an overall summary of the impact-anticipation capabilities of weather warnings, using the three evaluation metrics mentioned. The results are shown at an hourly temporal resolution with a lead-time of up to one hour (“real-time” temporal resolution). Furthermore, we include a discussion and performance comparison between using the original warning resolution versus the municipal level. While we recognize that the weather warnings analysed are not designed to work at such a high resolution, we argue that localized impact assessment is essential for effective emergency response.

#### 6.1.1. Real-Time Results

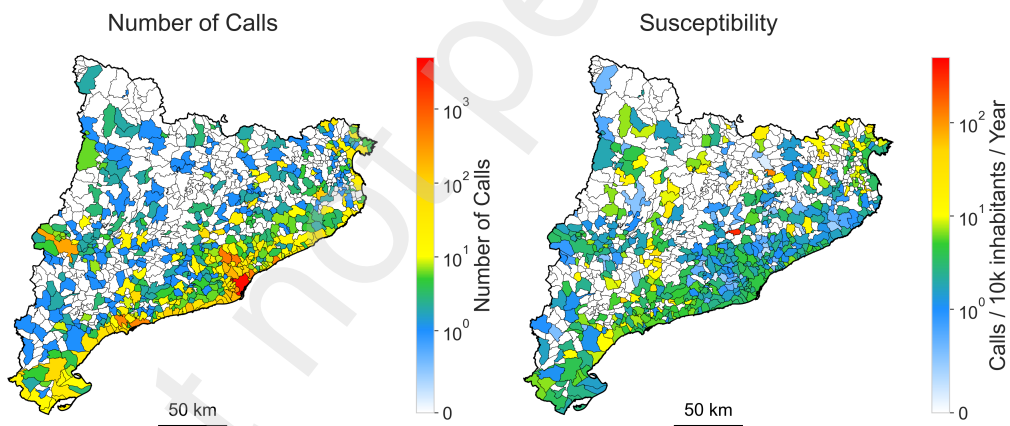
Figure 7 shows the overall performance of AEMET and SMC rain warnings in anticipating flood- and wind-related impacts, using different warning levels as activation thresholds (all warnings greater than or equal to the specified level are considered). Given the hourly resolution of the study, we treat the warnings as being in real-time (with a lead-time of up to one hour): for each hour in the study period, we consider the warnings that were active during that hour. As expected, the number of warning activations diminishes as the warning level threshold is raised.

For rain-related warnings (Fig. 7a), the Level 1 threshold achieves the highest Probability of Detection (PoD) (78.69% for AEMET and 75.43% for SMC), but its low Precision (0.27% and 0.29%) results in an equally low Critical Success Index (0.27% and 0.29%). As higher warning thresholds are considered, the CSI increases. This improvement is driven mostly by a reduction in False Alarms (increasing Precision), even as the PoD decreases significantly. While Level 3 warnings show the best CSI (2.20% and 1.73%), they detect less than ~5% of total impacts, and their Precision remains low (3.67% and 2.48%).

A similar pattern is observed in Figure 7b for wind-related warnings and impacts: higher warning levels lead to better Precision and CSI but at the expense of detecting fewer impacts. The results show slightly worse values across all metrics compared to flood-related impacts. For SMC warnings, for which we have Level 3 activations, the best results are observed with a Level

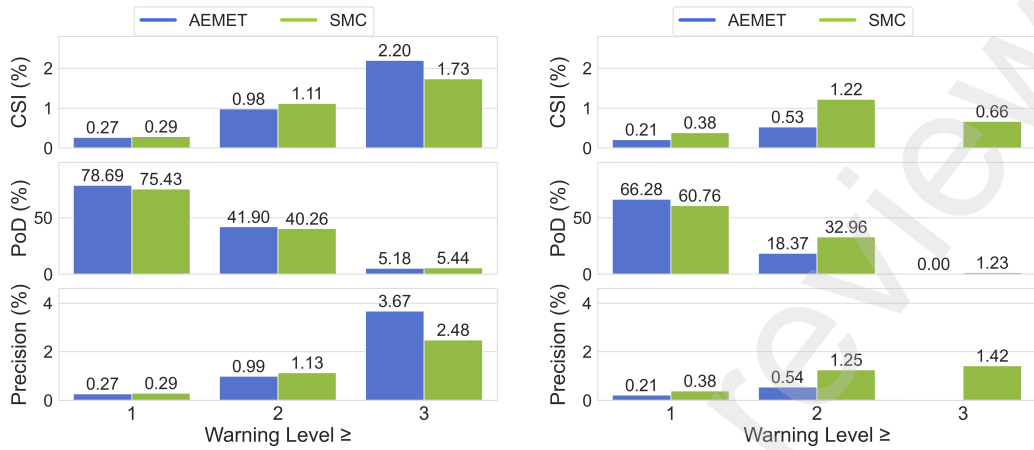


(a) Flood-related impacts.



(b) Wind-related impacts.

Figure 6: Total number of calls and susceptibility maps for flood and wind impacts at the municipality level. Susceptibility is defined as the number of calls per 10,000 inhabitants per year. The call data cover the period from November 2018 to February 2025, with a gap between March 14 and December 31, 2020.



(a) Performance of rain-related weather warnings for anticipating flood impacts.

(b) Performance of wind-related weather warnings for anticipating wind impacts.

Figure 7: Performance of rain- and wind-related weather warnings for anticipating flood and wind impacts, respectively. The “ $\geq$ ” symbol indicates the activation threshold; all warnings at or above the specified level are considered active. AEMET did not issue any Level 3 wind-related warnings during the evaluated period, and, therefore, CSI and Precision metrics are omitted for this warning level.

2 threshold. This occurs because both thresholds have very similar Precision (from 1.25% at Level 2 to 1.42% at Level 3), while the PoD falls dramatically from 32.96% to 1.23%.

### 6.1.2. Performance comparison: local vs. original warning-zone resolution

Figure 8a shows the performance difference when evaluating AEMET rain-warning activations at two different spatial resolutions: the municipality level (same results as in Subsection 6.1.1) and the original “warning zone” level (AEMET’s isoclimatic regions). The Probability of Detection shows comparable trends in both cases, which is expected since the warned areas cover the same spatial regions and, therefore, the same impacts. The slightly lower PoD observed at the warning-zone level occurs because an impact spread across multiple municipalities counts as multiple detections at the municipal level but only as one at the warning level. The main difference, however, is in the number of false alarms. Translating a large warning zone into dozens of municipalities multiplies the number of activations by a similar factor, drastically reducing Precision and, consequently, the Critical Success Index (e.g., the Level 1 threshold CSI falls from 6.35% to 0.27%). At

the original resolution, a Level 2 threshold achieves the best balance between detection and false alarms (12.58% CSI), whereas at the municipality level, Level 3 generally performs better (2.20% CSI).

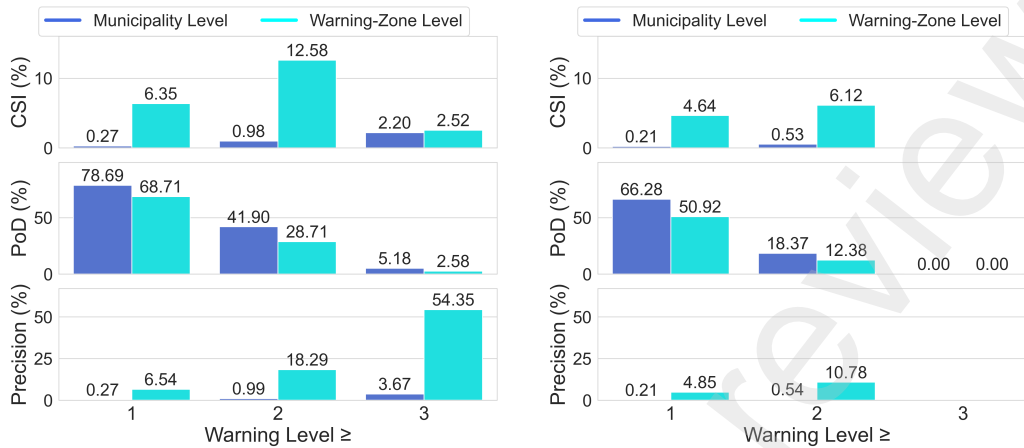
This pattern also holds for wind-related warnings and impacts (Fig. 8b). Again, the PoD is comparable, though slightly lower at the zonal level, while Precision and CSI are substantially higher, confirming the behaviour is a general consequence of the change in resolution. In Appendix A we explore this same situation with SMC warnings, arriving to comparable conclusions.

This is an expected behaviour, and we acknowledge that the increased resolution inherently represents a “disadvantage” for the weather warnings, which are designed neither to operate at the local/hourly scale nor to forecast any impacts. From a hazard-based perspective, if a weather warning correctly anticipates a hazard within the warned region, it can be considered as successful. However, within the scope of this study, we argue that emergency calls are events closely related to local characteristics. Analysing them at a larger “warning zone” level negates their localized nature and loses critical meaning. Since the focus of this work is to examine the relationships between forecasted hazards and their resulting impacts at the local level, we have chosen to work at the municipality level.

### *6.2. Lead-time analysis*

The results shown in Subsection 6.1.1 can be understood as “real-time”, as for each evaluated hour we considered the most updated warning status (lead-time of up to one hour). In the analysis in this section, we explore the effect of increasingly higher lead-times (up to 48 hours) on performance to evaluate the anticipation capabilities of warning activations.

Figure 9 shows the evolution of the different metrics for both AEMET and SMC rain-related activations as longer lead-times are used, across various warning level thresholds. For the Level 1 threshold, CSI trends slightly upward, as does Precision, while the PoD steadily decreases until the detection power is more than halved (from 78.59% and 75.43% to 31.16% and 27.40% for AEMET and SMC, respectively). For Level 2, the PoD continues to decrease with longer lead-times, while Precision dips slightly before picking up an upward trend (AEMET increased from 0.99% to 2.31% at 48h of lead-time. Meanwhile, SMC increases from 1.13% to 1.96% at 36h, to then decrease to 1.51% at 48h). This same behaviour is reflected in the CSI. For Level 3, the CSI starts with the highest value across all warning levels at the real-time configuration, with SMC warnings achieving even higher values



(a) AEMET's rain-related warnings performance (b) AEMET's wind-related warnings performance

Figure 8: Performance of AEMET's rain- and wind-related weather warnings for anticipating flood and wind impacts, respectively, at municipality (local) and warning-zone spatial resolutions. The " $\geq$ " symbol indicates the activation threshold; all warnings at or above the specified level are considered active. AEMET did not issue any Level 3 wind-related warnings during the evaluated period, and, therefore, CSI and Precision metrics are omitted for this warning level.

with some lead-times around 12 to 18 hours, after which they quickly decrease due to a lack of activations for lead-times longer than 32 hours. This trend is linked primarily to the PoD, which starts low and steadily decreases. More generally, the upward trend observed on the Precision curves might be due to a corresponding decrease in the number of activations as lead-time increases (Fig. 10).

This general behaviour of a reduction in PoD and an improvement in Precision is common to both AEMET and SMC warnings, and is generally observed for both flood and wind-related impacts. Figure 11 shows the performance for wind-related activations. The best CSI values are achieved with a Level 2 threshold, where they remain relatively stable for up to 24 hours of lead-time for SMC and 36 hours for AEMET, after which the PoD's downward trend accelerates. This behavior does not appear to have a direct correlation with the total volume of wind-related activations (Fig. 12), which steadily decreases as the lead-time increases.

The distinct behavior between Levels 1 and 2 (CSI improves or is more or less maintained with higher lead-times) and Level 3 (Better performance with smaller lead-times) could be explained by taking into account that Level

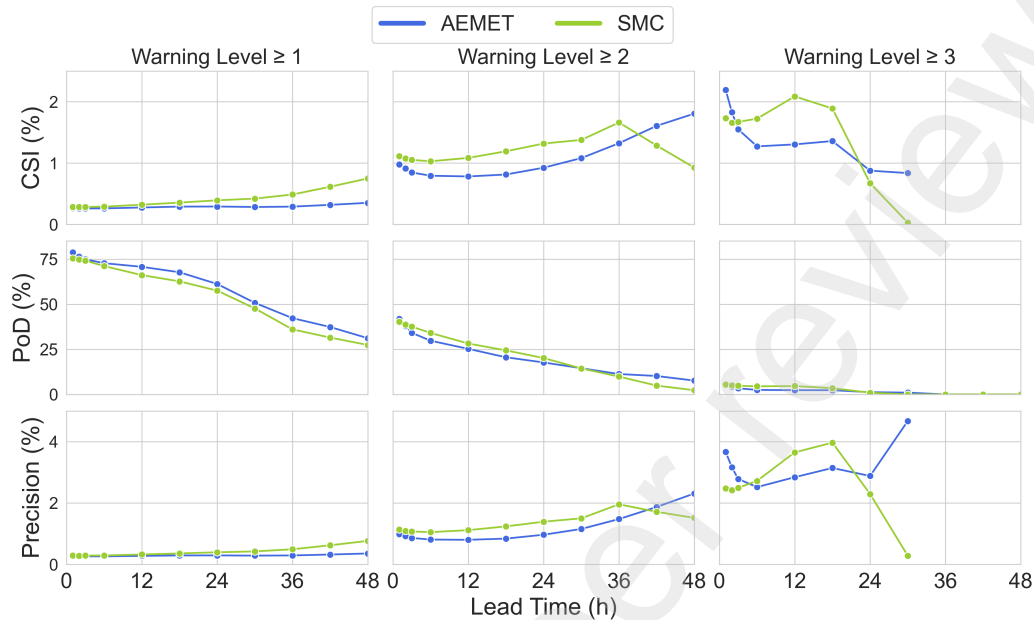


Figure 9: Performance of AEMET and SMC rain-related warnings across various warning level thresholds as lead-time increases (up to 48 hours). Note that for the Level 3 threshold, no warnings were issued with a lead-time greater than 32 hours.

3 warnings reflect very hazardous and uncommon situations, which typically undergo more rigorous validation processes before being issued. Lower-level warnings are issued more routinely. The lower false alarm ratio at longer lead-times might be attributed to forecast confidence: events are generally more predictable closer to their occurrence. If an event can be predicted with certainty multiple days ahead, it typically signifies a large and high-likelihood meteorological scenario.

### 6.3. Monthly distribution of activations and impacts

Figure 13a compares the number of rainfall-warning activations and flood-related impacts across different warning level thresholds, revealing that both follow similar monthly trends. To ensure consistency, only years without missing data (i.e., excluding 2018, 2020, and 2025) are considered in this section's analyses. The highest amounts are observed around the summer months and the start of autumn, while the lowest occur in the first months of the year and in December. For a Level 1 threshold, the total number of activations is orders of magnitude greater than the number of impacts

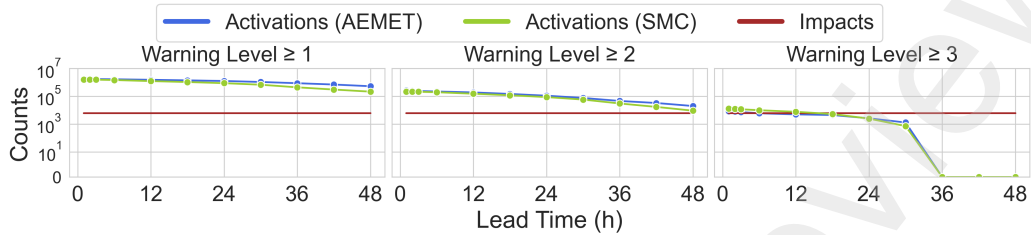


Figure 10: Number of AEMET and SMC rain-related activations, with respect to the lead-time. The number of flood-related impacts is also shown as a constant value to add context.

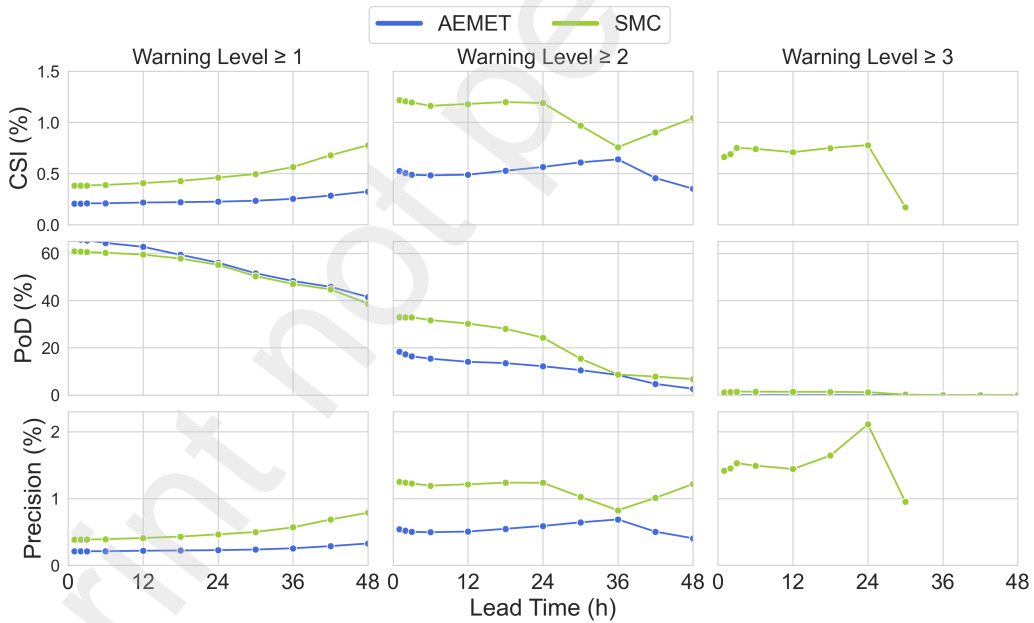


Figure 11: Performance of AEMET and SMC wind-related warnings across various warning level thresholds as lead-time increases (up to 48 hours). AEMET did not issue any Level 3 warnings of this kind, while SMC only issued them with lead-times up to 32 hours.

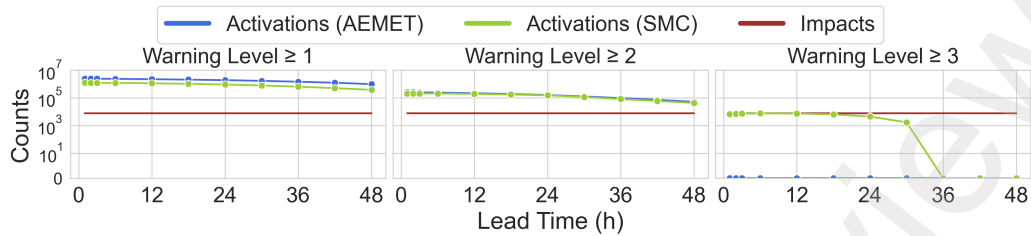


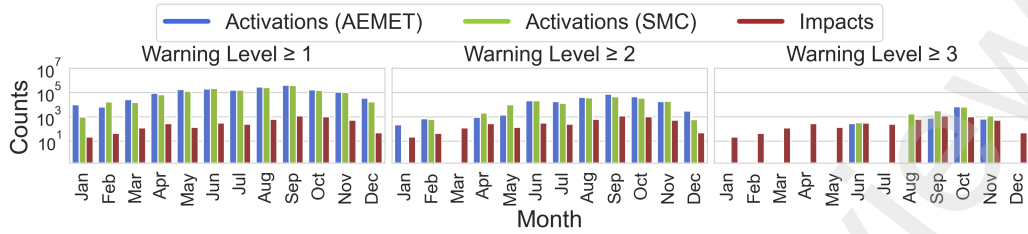
Figure 12: Number of AEMET and SMC wind-related activations, with respect to the lead-time. The number of wind-related impacts is also shown as a constant value to add context.

(activations between  $10^4$  and  $10^6$  while impacts have at most  $10^3$ ). With a Level 2 threshold, the number of activations still generally exceeds that of impacts, but we begin to see months without activations (March for AEMET; January and March for SMC). When using a Level 3 threshold, the difference between activations and impacts in active months is less significant, and in some cases, impacts even surpass activations. The primary issue, however, is that most months of the year do not register any Level 3 activations.

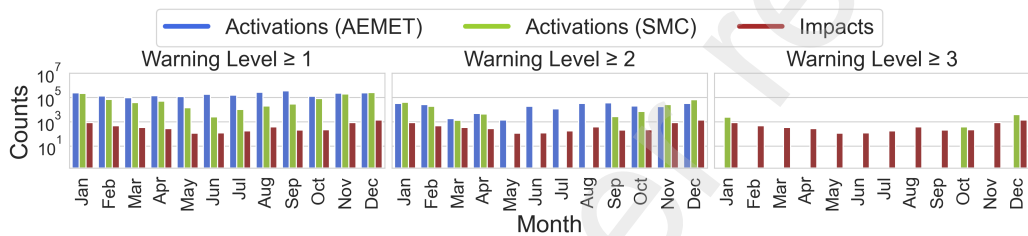
Figure 13b depicts the same analysis for wind-related activations and impacts. The trend changes, with the highest number of SMC activations and impacts now occurring at the beginning and end of the year. AEMET activations do not follow this same trend, probably because our analysis of AEMET wind warnings also includes storms, which can produce high winds but fall outside the typical patterns of “isolated” wind events. In any case, both activation and impact frequency decrease with higher warning level thresholds: at Level 2, SMC has four months without activations, and at Level 3, only SMC has activations during three months.

Regarding metrics, warning performance follows a pattern similar to activation counts, particularly for the PoD. Higher values correspond with the rainy season for flood-related emergencies, and with the beginning- and end-of-the-year months for wind-related impacts. This, however, also signals how warning effectiveness is highly dependent on the month of the year. See Appendix B for a more extensive analysis.

While performance often aligns with climatology, without a corresponding reduction in false alarms, the warnings could lose effectiveness at the municipality level. Furthermore, the lack of activations during certain months, despite the presence of impacts throughout the year, signals how warning procedures should better adapt to seasonal risk variations and their ultimate



(a) AEMET and SMC rain-related activations and flood-related impacts.



(b) AEMET and SMC wind-related activations and impacts.

Figure 13: Number of rainfall (a) and wind (b) warning activations (AEMET and SMC) and their related impacts, aggregated monthly. Note that the impact counts are independent of the warning level and are therefore consistent across all levels.

consequences.

#### 6.4. Effect of Increasing the Impact Threshold

Throughout this work, we have defined weather impacts as the occurrence of at least one emergency call in a municipality per hour, a call impact threshold of 1. However, not all emergencies have the same severity. To establish a more nuanced impact definition, we explore how performance varies when a higher number of calls is used as the impact threshold. Figure 14 shows this evolution for flood-related 112 calls and rainfall warnings across different warning levels. While this section focuses on rainfall-related data, the same notions and conclusions apply to wind-related performance (refer to Appendix C for specific results).

The Probability of Detection (PoD) improves as the impact threshold increases. This makes sense, as the total number of impacts is reduced (Fig. 15), and the analysis focuses on rarer, more severe events. Less severe events, which might be less correlated with the warnings, are ignored. It is interesting to note that the relative improvement in PoD is more significant at higher

warning levels: if we focus on AEMET warnings, with a Level 1 threshold, the PoD increases from 78.69% (call threshold of 1) to 95.73% (call threshold of 10), a 21.65% relative increase. In contrast, with a Level 3 threshold, it increases from 5.18% to 11.59%, a 123.75% relative increase. For SMC, this only holds for Level 1 and 2 warning thresholds, as the detection capabilities of Level 3 warnings remain nearly constant (from 5.44% to 5.49%).

Despite this improvement in detection, the overall CSI exhibits a downward trend and worsens quickly across all warning levels. This is due to a significant decrease in Precision, as the number of warning activations remains the same across all call thresholds, while the number of impacts becomes increasingly smaller, often by orders of magnitude, and further concentrated in densely populated areas.

We can conclude that the significance of a single emergency call is not uniform across the territory. Receiving one call in Barcelona, a city with nearly two million inhabitants, may not indicate an impactful weather event, while the same call in a small municipality in the Pyrenees could signal a major impact. Exploring this call-impact threshold at a local level would be interesting towards adapting IBWs to the local needs of the population.

## 7. Discussion and conclusions

The results of this work provide a quantitative assessment of the performance of traditional, hazard-based weather warnings in anticipating flood and wind impacts, using emergency calls as a high-resolution impact proxy. We define this performance as the co-occurrence of both warning activations and impacts within the framework presented in this study, rather than a measure of their operational success in anticipating meteorological hazards, which is their intended objective. With this in mind, our analysis reveals that, while these warnings are relatively effective at identifying a general area of risk, their ultimate effectiveness is severely limited due to the high number of false alarms when considered at the municipal scale. This disparity is a direct cause of the localized nature of impacts, and particularly of emergency calls, which are ultimately determined by specific local factors such as population distribution or exposure.

Across all experiments, the analysed weather warnings are capable of detecting most impacts. The Probability of Detection (PoD) can be further improved by filtering for emergencies that received a significant amount of

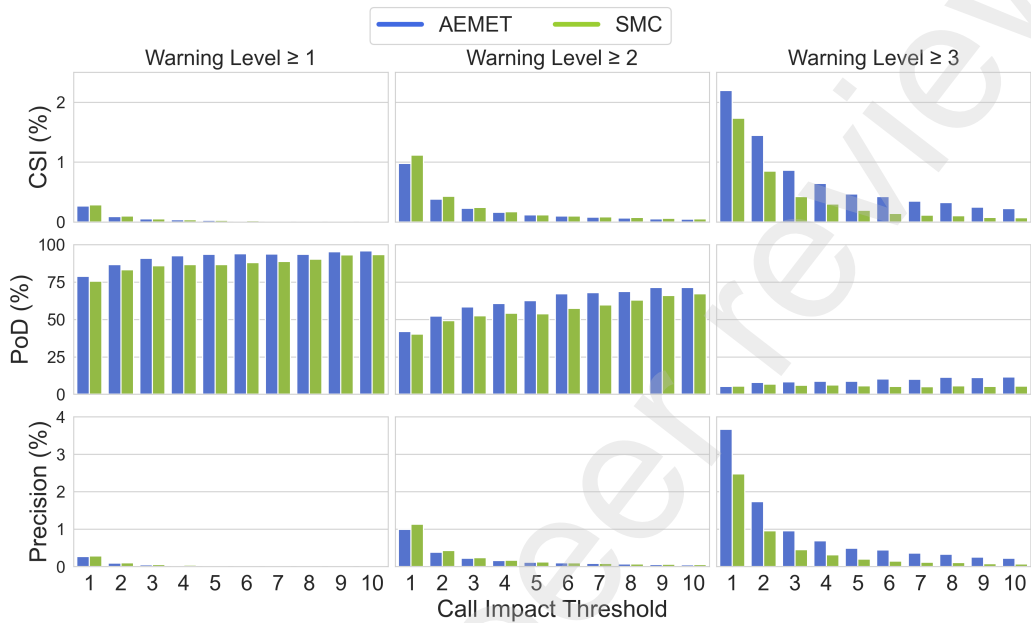


Figure 14: Performance of AEMET and SMC rainfall-related warnings with increasingly higher call impact thresholds, across various warning levels.

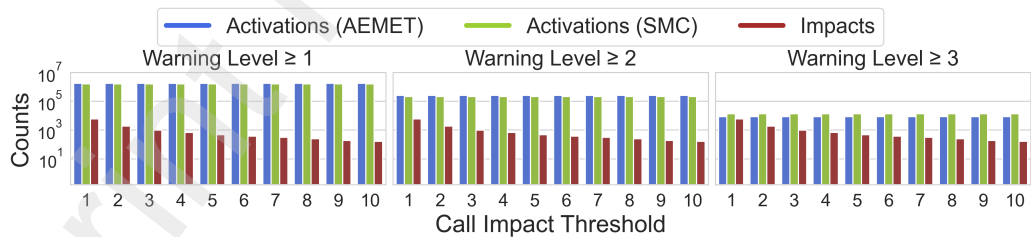


Figure 15: Number of flood-related impacts with respect to the call impact threshold, and the corresponding number of rainfall warning activations from AEMET and SMC, which remain constant for each warning level threshold.

calls (a threshold that should vary depending on the municipality), corresponding with more severe events. However, the Precision at the municipal level remains very low, negating the good PoD results and driving the Critical Success Index, our main metric for overall performance, to equally low values. This is especially significant when considering the results in Section 6.2, where longer lead times appear to yield higher CSI values; however, this is a result of a lower total number of activations (and a corresponding increase in Precision) rather than an actual improvement in PoD.

This issue is common to the warnings from the two meteorological agencies examined. Based on these findings, we argue that the core problem is the spatial scale: a warning may be correct for a warning zone, but not for every municipality within it. Consequently, current warnings can serve as a good first filter for determining high-risk areas, but a further processing step, driven by a nuanced catalogue of impact data and local characteristics, is necessary to downscale zone-level alerts into actionable, local-level information. In Section 6.4, we explore a simple case of this approach by adjusting the global impact threshold. Results are promising, showing a substantial increase in PoD, but these thresholds should be defined per municipality to ensure that impact definitions are adapted to local contexts. While we acknowledge that the practical implementation of such systems may be complex and costly [34], it is vital to start working towards this goal now rather than waiting until everything is perfectly defined [25]. This transition could start by first prioritizing the collection of impact data, followed by an iterative development and improvement of impact modelling techniques.

We propose that the design of this downscaling step should not rely on a single threshold for a particular feature, and instead integrate multiple data streams in a dynamic fashion [3]. Such sources should include historical impact data (e.g., archived emergency calls, damage reports, outages); vulnerability and exposure information (e.g., population census and density distribution, land usage maps, presence of critical infrastructure); other information layers relevant to specific hazards (e.g., flood maps for heavy rain warnings); and temporal context (time of year and day). This could be operationalized as a post-processing step following current weather warnings or forecasts. The definition of the thresholds could be achieved through the use of machine learning algorithms [35, 36], with an emphasis on simple and explainable approaches, as this downscaling layer should be completely interpretable by forecasters and emergency responders.

Moreover, regarding the monthly patterns observed in Section 6.3, it is

understandable that current weather warnings might not activate for entire months if the underlying phenomena behind them are not as intense as those observed during other periods. However, the reality is that impacts occur year-round, and therefore, impact-based warnings should also adapt accordingly and be calibrated based on these monthly patterns.

The implementation of such localized, impact-based systems faces significant practical challenges. A limitation of our study is that emergency calls inherently depend on human reporting, which becomes particularly scarce and less representative as we consider less populated municipalities, directly affecting our capability to assess performance in these areas. Moreover, even in major urban centres, emergency calls may not capture small-scale impacts, representing instead emergencies that pose an immediate risk to the individuals making the calls. This underscores a critical need for systematic and multi-source impact data. For instance, Giner Pérez De Lucia et al. [37] successfully reconstructed the events of the 2024 Valencia flood [38] by integrating 156,000 geolocated social media messages with other data sources, demonstrating how automated data collection pipelines can be invaluable. Building effective frameworks that allow for the cataloguing of impact information at all levels, from emergency calls to insurance claims or transport incidents, is essential to form a complete picture of the consequences that weather events have on the population. Still, relying solely on forecast and warning systems is unrealistic, as perfect predictions are practically unattainable. Therefore, other non-structural measures should be explored and adopted in parallel, as they can address significant gaps in the warning value chain. These measures may include improving risk communication [34], educating communities on risk awareness and safe behaviours [9], and clarifying the respective roles and merits of both hazard- and impact-based warnings [22, 34].

This work aims to provide not only a methodology for quantifying impact assessment at a high resolution, but also a solid evidence base to inform and accelerate the transition from traditional, hazard-centric systems towards more actionable and precise impact-based warning systems. The goal should not be to replace current systems [39, 36], but to augment hazard forecasting with intelligent impact anticipation layers. By demonstrating both the potential and limitations of the systems currently in place, we highlight the necessity of a closer collaboration between meteorological agencies, civil protection, and data holders. This collaboration is essential to overcome the scarcity of open and varied impact data through standardised processes for

its collection, sharing, and integration into operational warning pipelines [34, 40].

### *7.1. Future Work*

Future work should aim to extend this analysis to other geographical regions with different climatologies and population distributions, to test the robustness of the methodology and compare the observed conclusions and limitations. It would also be valuable to explore the use of other impact indicators, ranging from publicly available, such as social media posts, to private, like insurance claims. This would improve our understanding of the currently available data, identify information gaps, and develop collaborative frameworks for data acquisition. Furthermore, research should investigate how multiple impact data sources can be integrated simultaneously. Lastly, experimental studies on operationalizing this impact data for calibration and the issuance of weather warnings would be a necessary step towards real-time implementation.

## **8. Acknowledgments**

Jordi Morales' research is partially funded by project 2024DI00041 of the Catalan Industrial Doctorates Plan.

The writing of the article was co-funded by the European Union's Horizon Europe GoBeyond project under grant agreement No. 101121135. The opinions expressed in this paper solely reflect the views of the authors; the EU is not responsible for any use that may be made of the information it contains.

The authors would like to acknowledge the Catalan Civil Protection authorities, and particularly the CECAT, for providing the 112 emergency call data and for the fruitful discussions regarding this work.

## **9. Author contributions**

Conceptualization: JM, AK, AL, XL; data curation: JM; formal analysis: JM; investigation: JM, AK, AL, XL; methodology: JM; funding acquisition: AK, AL, XL; supervision: AK, AL, XL; Writing (original draft preparation): JM; Writing (review and editing): JM, AK, XL

## References

- [1] Intergovernmental Panel On Climate Change (IPCC), Christopher B. Field, Vicente Barros, Thomas F. Stocker, and Qin Dahe. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change. Technical report, UN Intergovernmental Panel On Climate Change, 2012.
- [2] Intergovernmental Panel On Climate Change (IPCC). Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, UN Intergovernmental Panel On Climate Change, 2023.
- [3] Serena Schroeter, Harald Richter, Craig Arthur, David Wilke, Mark Dunford, Martin Wehner, and Elizabeth Ebert. Forecasting the Impacts of Severe Weather. *Aust. J. Emerg. Manag.*, pages 76–83, 2021. doi: 10.47389/36.1.76.
- [4] EUMETNET. MeteoAlarm project, 2025. URL <https://www.meteoalarm.org>. Accessed 7 July 2025.
- [5] EUMETNET. Home Page, 2025. URL <https://www.eumetnet.eu/>. Accessed 7 July 2025.
- [6] United Nations Office for Disaster Risk Reduction. Sendai Framework for Disaster Risk Reduction 2015 - 2030. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, Switzerland, 2015. URL <https://www.undrr.org/quick/11409>.
- [7] Bruno Merz, Günter Blöschl, Sergiy Vorogushyn, Francesco Dottori, Jeroen C. J. H. Aerts, Paul Bates, Miriam Bertola, Matthias Kemter, Heidi Kreibich, Upmanu Lall, and Elena Macdonald. Causes, Impacts and Patterns of Disastrous River Floods. *Nat. Rev. Earth Environ.*, 2(9):592–609, 2021. doi: 10.1038/s43017-021-00195-3.
- [8] Florian Pappenberger, Hannah L. Cloke, Dennis J. Parker, Fredrik Wetterhall, David S. Richardson, and Jutta Thielen. The Monetary Benefit of Early Flood Warnings in Europe. *Environ. Sci. Policy*, 51:278–291, 2015. doi: 10.1016/j.envsci.2015.04.016.

- [9] Claudio Rossi, Antonella Frisiello, Giacomo Falcone, and Fabrizio Dominici. Best practices on Public Warning Systems for Climate-Induced Natural Hazards, 2018. URL [https://drmkc.jrc.ec.europa.eu/Portals/0/innovation/SupportSystem/11\\_Lithuania/JRC\\_PublicWarningonNaturalHazard\\_05\\_PWS\\_final.pdf](https://drmkc.jrc.ec.europa.eu/Portals/0/innovation/SupportSystem/11_Lithuania/JRC_PublicWarningonNaturalHazard_05_PWS_final.pdf).
- [10] Brenda Mackie. *Warning Fatigue: Insights from the Australian Bushfire Context*. PhD thesis, University of Canterbury. Media and Communications, 2014. URL <https://ir.canterbury.ac.nz/items/46203cf9-3397-470a-b63b-122e0f1c3b44>.
- [11] Spanish National Meteorological Agency (AEMET). Home Page, 2025. URL <https://www.aemet.es/es/>. Accessed 27 May 2025.
- [12] Meteorological Service of Catalonia (SMC). Home Page, 2025. URL <https://www.meteo.cat/>. Accessed 27 May 2025.
- [13] World Meteorological Organization. WMO guidelines on multi-hazard impact-based forecast and warning services. Technical report, World Meteorological Organization (WMO), Geneva, Switzerland, 2015. URL <https://www.preventionweb.net/quick/84473>.
- [14] S. S. Schuster, R. J. Blong, R. J. Leigh, and K. J. McAneney. Characteristics of the 14 April 1999 Sydney hailstorm based on ground observations, weather radar, insurance data and emergency calls. *Nat. Hazards Earth Syst. Sci.*, 5(5):613–620, 2005. doi: 10.5194/nhess-5-613-2005.
- [15] Pekka J. Rossi, Vesa Hasu, Kalle Halmevaara, Antti Mäkelä, Jarmo Koistinen, and Heikki Pohjola. Real-Time Hazard Approximation of Long-Lasting Convective Storms Using Emergency Data. *J. Atmos. Ocean. Technol.*, 30(3):538–555, 2013. doi: 10.1175/JTECH-D-11-00106.1.
- [16] Antonio Oliva and Jorge Olcina. Floods and Emergency Management: Elaboration of Integral Flood Maps Based on Emergency Calls (112)—Episode of September 2019 (Vega Baja del Segura, Alicante, Spain). *Water*, 15(1):2, 2022. doi: 10.3390/w15010002.
- [17] Ana M Camarasa-Belmonte and M Paz Caballero López. Luvias in situ en la comunidad valenciana. relación entre indicadores pluviométricos,

- llamadas al centro de coordinación de emergencias (112) y relación de daños, durante el episodio de 26-30 de noviembre de 2016. In *Publicaciones de la Asociación Española de Climatología. Serie A; 11*, XI Congreso de la Asociación Española de Climatología, Cartagena, España, 17–19 October 2018, 2018. URL <http://hdl.handle.net/20.500.11765/9904>.
- [18] Guadalupe Ortiz, Pablo Aznar-Crespo, Antonio Oliva, Jorge Olcina-Cantos, and Antonio Aledo. Uses and opportunities of emergency calls as a resource for flood risk management. *IJDRR*, 100:104160, 2024. doi: 10.1016/j.ijdr.2023.104160.
- [19] Mikhail Sirenko, Tina Comes, and Alexander Verbraeck. The rhythm of risk: Exploring spatio-temporal patterns of urban vulnerability with ambulance calls data. *Environ. Plan. B: Urban Anal. City Sci.*, 52(4): 863–881, 2025. doi: 10.1177/23998083241272095.
- [20] Rainer Kaltenberger, Andreas Schaffhauser, and Michael Staudinger. “What the weather will do” – results of a survey on impact-oriented and impact-based warnings in European NMHSs. *Adv. Sci. Technol. Res. J.*, 17:29–38, 2020. doi: 10.5194/asr-17-29-2020.
- [21] Philippe Weyrich, Anna Scolobig, David N. Bresch, and Anthony Patt. Effects of Impact-Based Warnings and Behavioral Recommendations for Extreme Weather Events. *Weather Clim. Soc.*, 10(4):781–796, 2018. doi: 10.1175/WCAS-D-18-0038.1.
- [22] Andrea L. Taylor, Astrid Kause, Barbara Summers, and Melanie Harrowsmith. Preparing for Doris: Exploring Public Responses to Impact-Based Weather Warnings in the United Kingdom. *Weather Clim. Soc.*, 11(4):713–729, 2019. doi: 10.1175/WCAS-D-18-0132.1.
- [23] Qinghong Zhang, Liye Li, Beth Ebert, Brian Golding, David Johnston, Brian Mills, Shannon Panchuk, Sally Potter, Michael Riemer, Juanzhen Sun, Andrea Taylor, Sarah Jones, Paolo Ruti, and Julia Keller. Increasing the value of weather-related warnings. *Sci. Bull.*, 64(10):647–649, 2019. doi: 10.1016/j.scib.2019.04.003.
- [24] Spanish National Meteorological Agency (AEMET). SINOBAS, 2026. URL <https://sinobas.aemet.es/>. Accessed 17 Feb 2026.

- [25] Santiago Gaztelumendi, Joseba Egaña, Miriam Ruiz, and Eguzkiñe Iturrioz. The Basque Impact Weather Catalogue. *Adv. Sci. Res.*, 21:41–48, 2024. doi: 10.5194/asr-21-41-2024.
- [26] Spanish National Meteorological Agency (AEMET). Plan Nacional de Predicción y Vigilancia de Fenómenos Meteorológicos Adversos MeteoAlerta, 2025. URL [https://www.aemet.es/documentos/es/eltiempo/prediccion/avisos/plan\\_meteoalerta/plan\\_meteoalerta.pdf](https://www.aemet.es/documentos/es/eltiempo/prediccion/avisos/plan_meteoalerta/plan_meteoalerta.pdf). Accessed 6 February 2026.
- [27] Juan de Dios del Pino Corredera. Análisis del uso operativo de los indicadores de verificación basados en tablas de contingencia en el pronóstico de fenómenos meteorológicos adversos. In *XXXIV Jornadas Científicas de la AME y el 17º Encuentro Hispano-Luso de Meteorología*, volume 34, 2016. URL <https://pub.ame-web.org/index.php/JRD/article/view/1943>.
- [28] Spanish National Meteorological Agency (AEMET). Anexo 1 - umbrales y niveles de aviso, 2022. URL [https://www.aemet.es/documentos/es/eltiempo/prediccion/avisos/plan\\_meteoalerta/METEOALERTA\\_ANX1\\_Umbrales\\_y\\_niveles\\_de\\_aviso.pdf](https://www.aemet.es/documentos/es/eltiempo/prediccion/avisos/plan_meteoalerta/METEOALERTA_ANX1_Umbrales_y_niveles_de_aviso.pdf). Accessed 30 June 2025.
- [29] Statistical Institute of Catalonia. Population as of 1 January, 2025. counties and aran, areas and provinces, 2025. URL <https://www.idescat.cat/indicadors/?id=aec&n=15224&lang=en&t=202500>. Accessed 16 December 2025.
- [30] Henry R. Stanski, Laurence Wilson, and William Burrows. Survey of Common Verification Methods in Meteorology. Technical report, World Meteorological Organization, Ontario, Canada, 1990.
- [31] Meteorological Service of Catalonia (SMC). Home page, 2025. URL [https://www.meteo.cat/climatologia/atles\\_climatic/](https://www.meteo.cat/climatologia/atles_climatic/). Accessed 6 October 2025.
- [32] Meteorological Service of Catalonia (SMC). Catalonia’s IDF Maps, 2025. URL <https://www.meteo.cat/wpweb/climatologia/dades-i-productes-climatices/corbes-idf/idf-mapes/>. Accessed 6 October 2025.

- [33] Protecció Civil de Catalunya. VENTCAT - Pla especial d'emergències per risc de vent a Catalunya, 2017. URL [https://dsp.interior.gencat.cat/bitstream/handle/20.500.14007/2849/Document\\_Pla\\_VENTCAT.pdf?sequence=2&isAllowed=y](https://dsp.interior.gencat.cat/bitstream/handle/20.500.14007/2849/Document_Pla_VENTCAT.pdf?sequence=2&isAllowed=y). Accessed 6 February 2026.
- [34] Sally Potter, Thomas Kox, Brian Mills, Andrea Taylor, Joanne Robbins, Carolina Cerrudo, Faye Wyatt, Sara Harrison, Brian Golding, Will Lang, Adam J.L. Harris, Rainer Kaltenberger, Stefan Kienberger, Harold Brooks, and Andrew Tupper. Research gaps and challenges for impact-based forecasts and warnings: Results of international workshops for High Impact Weather in 2022. *IJDRR*, 118:105234, 2025. doi: 10.1016/j.ijdr.2025.105234.
- [35] Scientific Advice Mechanism to the European Commission. Artificial Intelligence in Emergency and Crisis Management: Rapid Evidence Review Report. Technical report, SAPEA, 2025.
- [36] Thomas Kox, Sara Harrison, Ferdinand Ziegler, and Lars Gerhold. Perceptions, hopes, and concerns regarding the possibilities of artificial intelligence in weather warning contexts. *IJDRR*, 130:105817, 2025. doi: 10.1016/j.ijdr.2025.105817.
- [37] José Giner Pérez De Lucia, Adrián López-Ballesteros, Julio Fernández-Pedauy, Javier Senent-Aparicio, and José M. Cecilia. Harnessing social sensing for real-time flood event reconstruction: A digital autopsy of the 2024 Valencia DANA. *International Journal of Disaster Risk Reduction*, 132:105966, 2026. doi: 10.1016/j.ijdr.2025.105966.
- [38] Spanish National Meteorological Agency (AEMET). Informe sobre el episodio meteorológico de precipitaciones torrenciales y persistentes ocasionadas por una DANA el día 29 de octubre de 2024. Technical report, Ministry for Ecological Transition and the Demographic Challenge, Madrid, Spain, 2024. URL <http://hdl.handle.net/20.500.11765/16360>.
- [39] Tobias Geiger, Thomas Rössli, David N. Bresch, Bodo Erhardt, Andreas M. Fischer, Dominik Imgrüth, Stefan Kienberger, Laura Mainetti, Gudrun Mühlbacher, and Raphael Spiekermann. How to provide actionable information on weather and climate impacts?—A summary of

strategic, methodological, and technical perspectives. *Front. Clim.*, 6: 1343993, 2024. doi: 10.3389/fclim.2024.1343993.

- [40] Monique M. Kuglitsch, Ivanka Pelivan, Serena Ceola, Mythili Menon, and Elena Xoplaki. Facilitating adoption of AI in natural disaster management through collaboration. *Nat. Commun.*, 13(1):1579, 2022. doi: 10.1038/s41467-022-29285-6.

## Appendix A. SMC performance comparison: local vs. original warning-zone resolution

Figure Appendix A.1 shows the performance difference when evaluating SMC rainfall- and wind-related warning activations at the municipal level versus the original “warning zone” (counties for SMC) level. A similar behaviour to that observed for AEMET warnings is displayed, where the lower resolution results in overall higher Precision (and consequently CSI). While the PoD appears lower at the warning-zone resolution, both configurations have the same detection capabilities, as the total warned area is the same. The difference in PoD stems from how multiple impact instances at the municipal level are aggregated into a single instance at the county level.

SMC warnings at the original resolution exhibit lower Precision (and CSI) values than their AEMET counterparts. This is primarily because SMC’s warning zones have a higher spatial resolution (counties instead of AEMET’s larger isoclimatic zones), which are closer in scale to the municipal resolution.

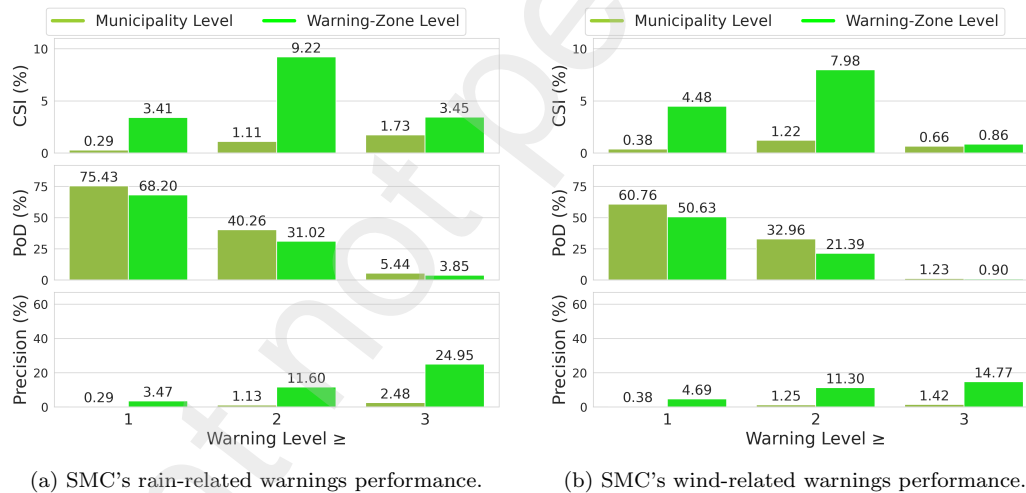


Figure Appendix A.1: Performance of SMC’s rain- and wind-related weather warnings for anticipating flood and wind impacts, respectively, at municipality (local) and warning-zone spatial resolutions. The “ $\geq$ ” symbol indicates the activation threshold; all warnings at or above the specified level are considered active.

## Appendix B. Monthly performance metrics

Regarding monthly performance metrics, Figure Appendix B.1 shows the performance of rain warnings for anticipating flood-related impacts by month, using various warning levels as thresholds. With a Level 1 threshold, the PoD ranges from approximately 20% in the early months of the year to between 60% and 90% later, coinciding with the increase in activations. However, this does not translate into a similar pattern for the CSI, as all months maintain a very low Precision that, at best, does not reach 0.5%. This suggests that the higher detection values are driven more by extensive spatial coverage of warnings than by more accurate forecasts.

Choosing a higher warning threshold can improve CSI values in certain months, with notable peaks in AEMET warnings, such as April for Level 2 (5.35%) and November for Level 3 (4.91%). These peaks also correspond to the lowest false alarm ratios (Precision of 6.59% and 8.42%, respectively). SMC obtains similar values, but the peaks are less pronounced. Nonetheless, the lack of activations in many months, especially for Level 3, renders these warnings ineffective for anticipating impacts for entire periods of the year.

Shifting to wind, Figure Appendix B.2 displays the performance of wind-related activations. Here, the PoD for SMC activations varies significantly from month to month, particularly for Level 1, while for AEMET, the variations are smoother. Similar to the activation distribution, the best values are observed at the beginning and end of the year. Interestingly, these months also see the best Precision values, signalling that warnings are more accurate during these periods despite the higher volume of activations. With a Level 2 threshold, the PoD falls significantly. In several months towards the middle of the year, PoD reaches 0% despite having activations. For Level 3, AEMET did not issue any wind warnings, while SMC issued them during three months, for which only December obtained a noticeable PoD, and the highest Precision.

## Appendix C. Effect of wind-related impact thresholds

In Section 6.4 we analysed how increasing the impact threshold (i.e. the number of calls in an hour required to consider a sample an impact) affected the performance of rainfall-related warnings in anticipating flood-related impacts. In Figure Appendix C.1, we present the metrics corresponding to wind-related warning performance. Increasing the impact threshold results

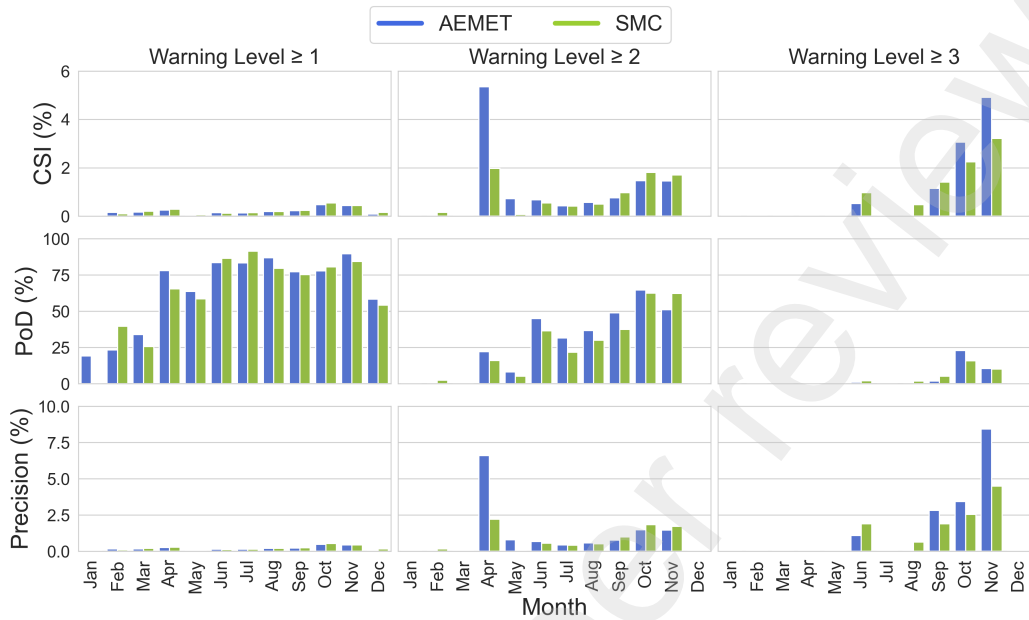


Figure Appendix B.1: Performance of AEMET and SMC rainfall-related warnings, aggregated by month, across various warning level thresholds.

in a higher PoD. This corresponds with a reduction in the total number of impacts (Fig. Appendix C.2), suggesting that a higher impact threshold helps filter out less severe emergencies, as more severe events are more related to the climatology. It is interesting to note the difference between SMC and AEMET in the PoD for the warning level 2 threshold, with SMC nearly doubling the detection capabilities despite the total number of activations being similar in both cases. This translates into higher CSI values for SMC compared to AEMET in this case. We believe this difference could stem from the distinct threshold definitions used by each agency.

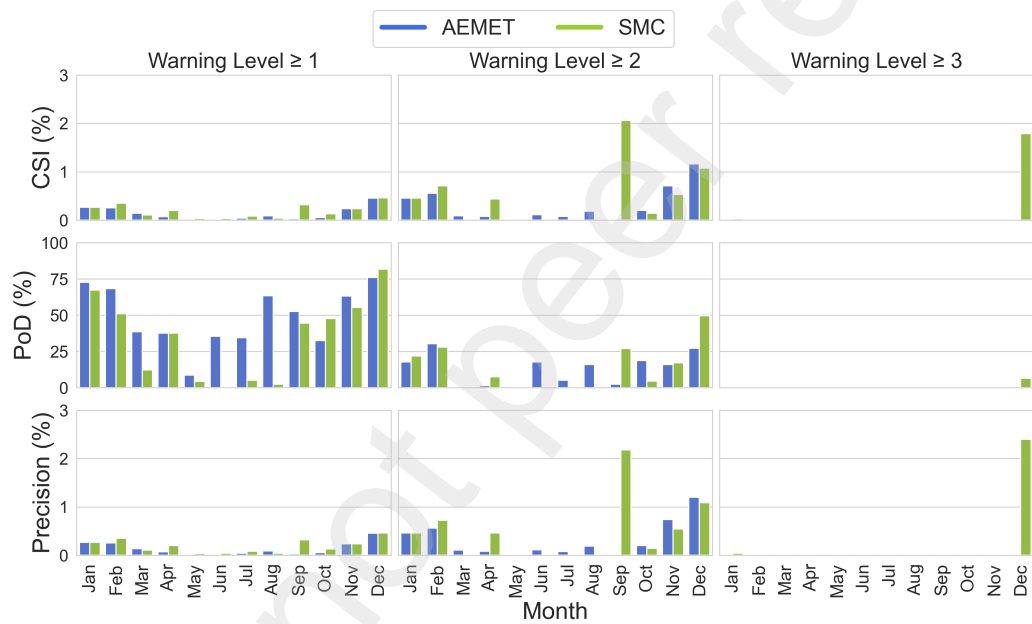


Figure Appendix B.2: Performance of AEMET and SMC wind-related warnings, aggregated by month, across various warning level thresholds.

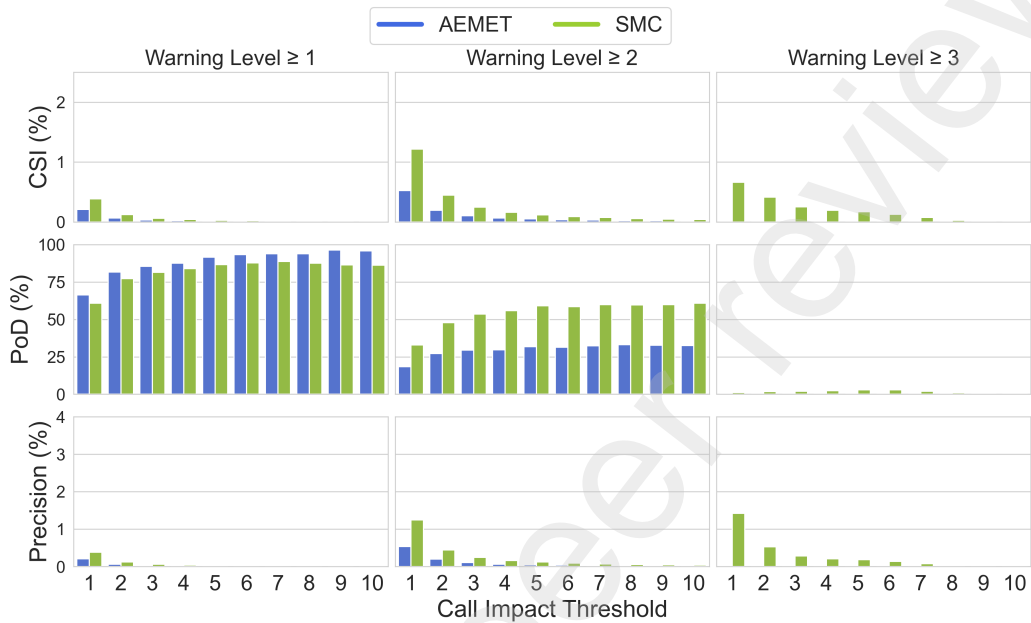


Figure Appendix C.1: Performance of AEMET and SMC wind-related warnings with increasing higher call impact thresholds, across various warning levels.

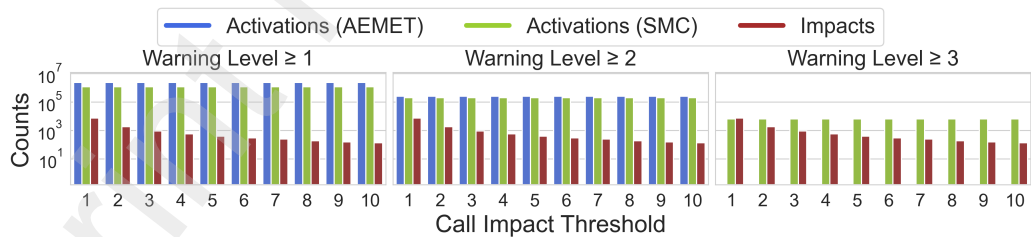


Figure Appendix C.2: Number of wind-related impacts with respect to the call impact threshold, and the corresponding number of rainfall warning activations from AEMET and SMC, which remain constant for each warning level threshold.