Chronicle of a forecast destined to fail:

A compilation of pitfalls for forecasters sailing between Scylla and Charybdis

bruno.gillet-chaulet@meteo.fr

Note: All times are UTC. NMS: National Meteorological Service.

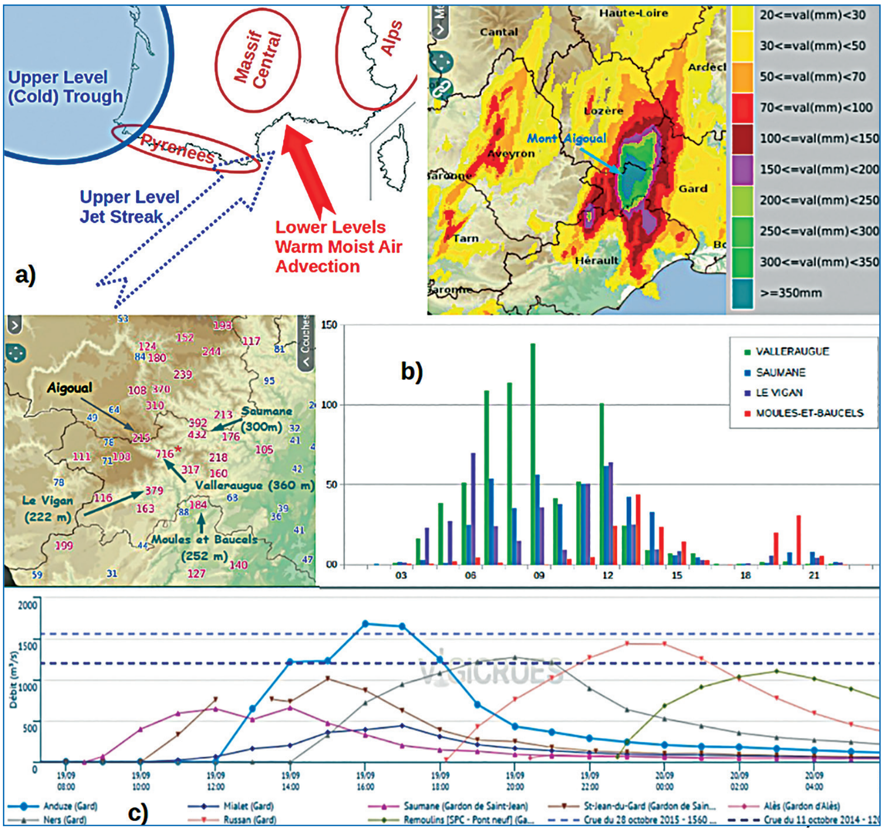
Introduction and description of the situation

This article focuses on decision making in the domain of weather warnings. It describes a situation in which all the difficulties that a forecaster may encounter were brought together in a single event.

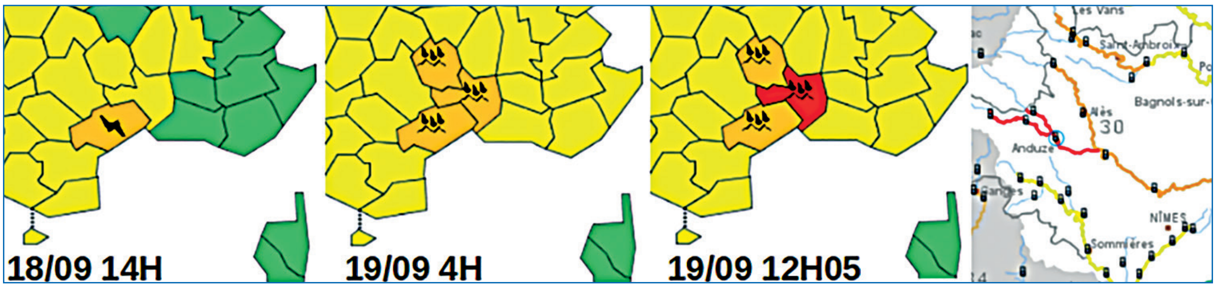
On Saturday 19 September 2020, an intense Mediterranean Episode* affected a small area of the Gard Department (a French administrative division). A quasi-stationary retrograde regenerating convective system (V-shaped thunderstorm) produced exceptional rainfall totals (100-year return period) over very short periods: more than 550 mm were recorded in twelve hours around the village of Val-d'Aigoual near Valleraugue. Most of the precipitation was focused during the period 4H/14H (**figure 1**).

The consequences were dramatic: run-off, mudflows, and historic flash-floods **(figure 1c)** which led to 2 fatalities and considerable damage to homes and infrastructure amounting to tens of millions of Euros. The red vigilance, the highest level of the Météo-France warning (called the vigilance system, which is carried out at the departmental scale), issued for this event was considered to have been issued too late by the authorities **(figure 2)** and the forecasts provided were strongly criticised.

** The "Mediterranean Episodes" are intense rainstorm events affecting the regions bordering the Mediterranean Sea, occurring mainly in autumn (**figure 1a**), they are a type of Heavy Precipitation Events (**HPE**). They are known to be sudden and virulent. Their predictability is sometimes, or even often, poor because they involve small-scale convective elements in time and space. See the international research programme HyMeX (**Ducrocq and al, 2016**) for more information.*



► *Figure 1: a) Typical synoptic context favourable to Mediterranean Episode (HPE) over France. The intensity of these features (accumulated total precipitation) strongly depends on the following variables: Wet bulb potential temperature, Wind (in the lower levels) and CAPE; b) 24 Hr accumulated precipitation (radar, rain gauges in millimetres) and temporal evolution of hourly precipitation for 4 measuring stations on Saturday 19 September 2020; c) Temporal evolution of river flows for some reference stations on this day compared to the last remarkable floods.*



▼ *Figure 2: Successive vigilance watch maps issued for this event. Dates of publication on the bottom left. The* meteorological *vigilance concerns the parameters Thunderstorms and Rain-Flooding (for accumulated precipitation). The* hydrological *vigilance concerns the water levels on sections of the rivers (flooding, right map).*

The post-event feedback, from internal reports by Météo-France, clearly indicates that such accumulated precipitation could not have been predicted from the available numerical weather prediction (NWP) models. However, what lessons can be learnt from this situation? What improvements can we expect in the near future? What advice should be given to forecasters in charge of warnings?

This article, written as a chronicle, describes the framework in place beforehand and then the real time follow-up phase of this memorable event. It highlights the questions that arise at each stage of the decision-making process and proposes some answers.

D-7 to D-1, the framework ahead of the event

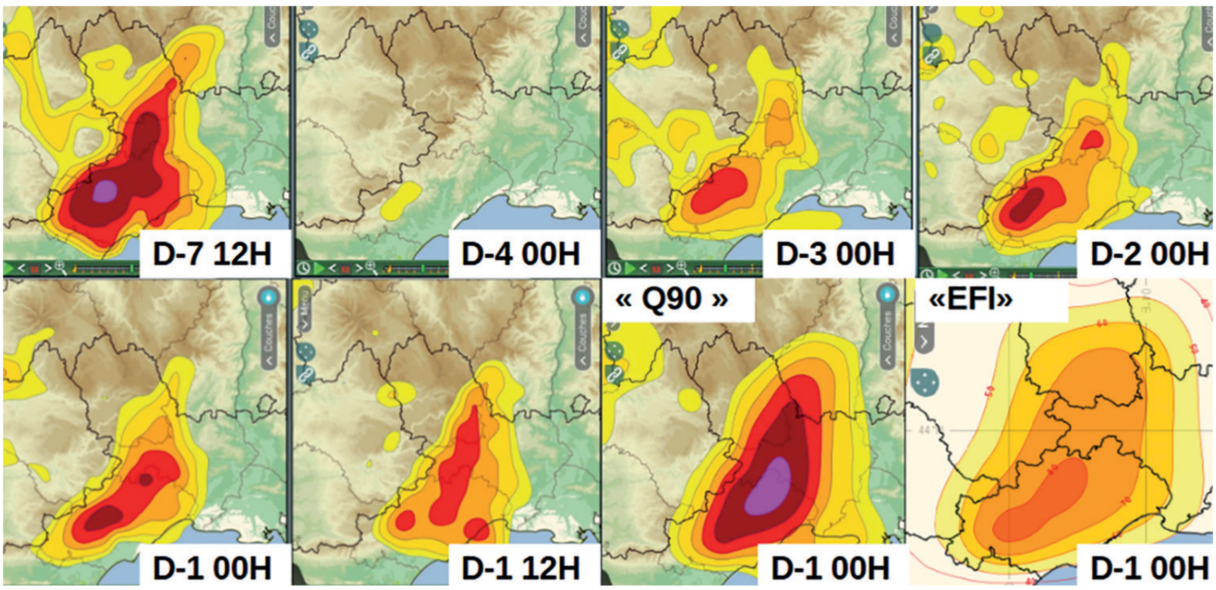
Several Days ahead

On a synoptic scale, conditions favourable for the occurrence of a Mediterranean HPE were identified several days, at least 7, before the event **(figure 1a)** with a significant potential. The study of deterministic models, ensemble forecast EPS products (probabilities, quantiles, Extreme Forecast Index -EFI-), and specific HPE diagnostics showed a weak signal but it was enough to put forecasters on alert **(figure 3)**.

However, the location was rather focused over the department of Hérault **(figure 2)** and the expected accumulated precipitation was still below the orange threshold for the concerned regions, namely:

- Over the plains: 80 mm in less than 3 hours or 120 mm in less than 24 hours
- Over relief: 120 mm in less than 3 hours or 200 mm in less than 24 hours

▼ *Figure 3: 24 H accumulated precipitation (mm, same range as in* ***figure 1****) on Saturday 19 September 2020. Successive forecasts from the ECMWF deterministic model, EPS Quantile 90 ("Q90"), EPS "EFI" for the day of the event. "Base Time" indicated at bottom right.*



Moreover, this was the first Mediterranean Episode of the autumn season, occurring in a context of severe drought with river levels at their lowest. Therefore significant consequences were not expected.

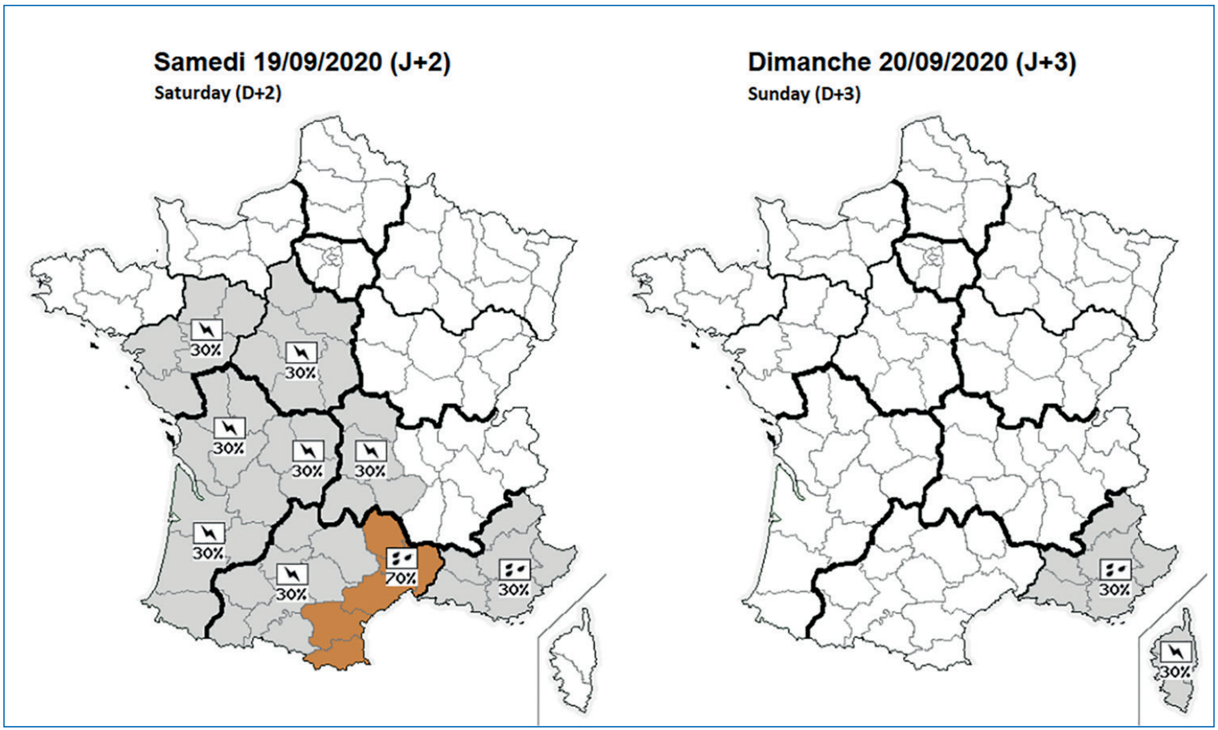
"Drawing attention, without alarming!"

On Thursday afternoons, Météo-France routinely presents the weather conditions for the next few days, particularly for the upcoming weekend, to the services in charge of civil protection. This enables these stakeholders to stand up an appropriate organisation if needed. On Thursday 17 September, the event was announced with accumulated precipitation around 150 mm in 24 hours. This value was estimated as the most likely, or closest, to what was expected, given the NWP models available at that time. Unfortunately, it was interpreted as not very important (compared to the values measured during more remarkable Mediterranean Episodes) and therefore did not draw the right level of attention from the authorities.

This raises the question of how to communicate these situations well, i.e. make the audience receptive without alarming then too much when the forecast is out of the comfort zone for the forecaster dealing with uncertainty. For example, don't you think that users are too often demanding precise deterministic forecasts? That they tend to over-interpret these data and put too much trust in them? Is it preferential to better express uncertainty, even if it is unpleasant? Indeed, recent studies in human neuroscience show that natural cognitive mechanisms are in place to systematically reduce uncertainty (**Bohler, 2021**). For medium range forecasts (D-7 to D-2), Météo-France offers a specific product that gives the probability of occurrence of a severe weather event (i.e. which would require an orange or red level of vigilance, **figure 4**). The probability was high in this case (70%), but it did not indicate the precipitation intensity of the phenomenon. Is it necessary to remind people that during any autumn Mediterranean Event where the potential is high, it is advisable to keep regularly informed of the evolution of the situation, as elements of aggravation can occur very quickly?

The Day before

At Météo-France, the day before is the deadline for the meteorological vigilance procedure: the forecaster must establish the appropriate colour level for the situation. This coincides with the time that the fields are available for resolved convection as well, thanks to small-scale non-hydrostatic convection-permitting models (AROME at Météo-France, **Seity and al, 2010**). **Figure 5** shows the successive deterministic forecasts of this model.



▲ *Figure 4: Probabilistic forecast of the risk of occurrence of an orange or red vigilance level event. Production on Thursday 17 September 2020 (2 days before the event). A significant risk concerned the Languedoc-Roussillon region at D+2. The risk shifted to the east the day after (D+3).*

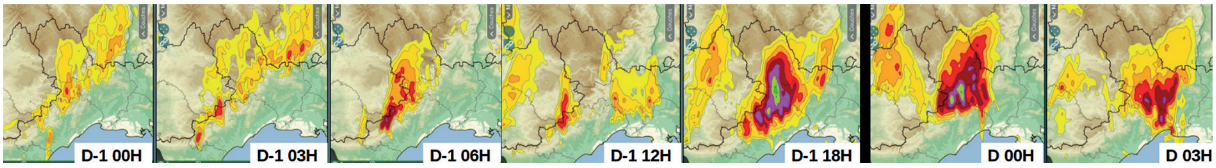
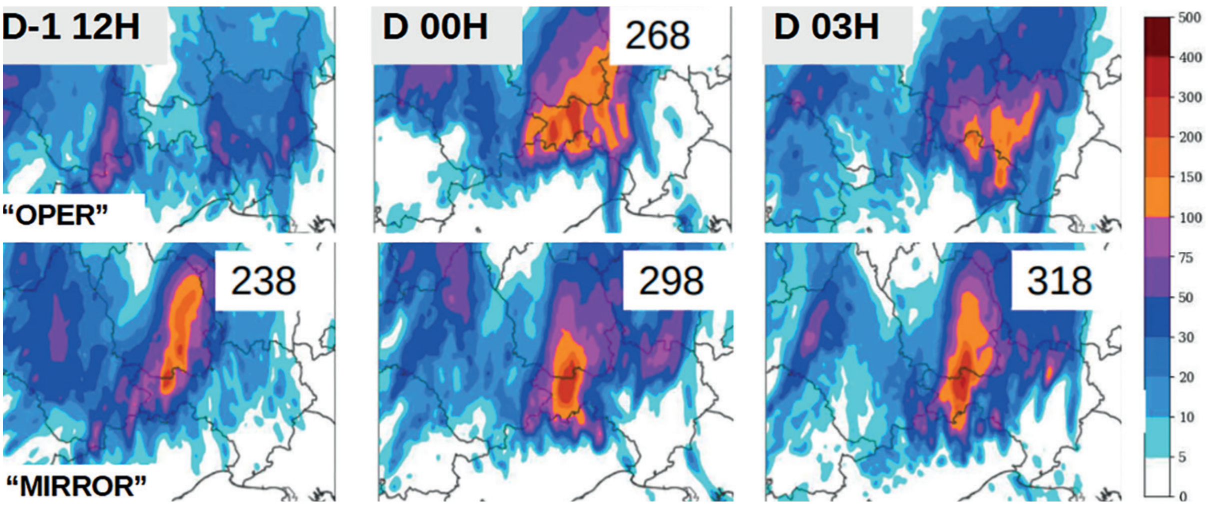


Figure 5: 24 H accumulated precipitation (mm, same range as in figure 1) on Saturday 19 September 2020. Successive forecasts from AROME model for the day of the event. "Base Time" indicated at bottom right.

These forecasts showed a strong variability. They essentially targeted the Hérault department (at least the oldest ones, in coherence with the larger scale forecasts presented previously, **figure 3**). They generally proposed surprisingly low precipitation; given the ingredients gathered in this situation, **(figure 1a)**, lower than the orange vigilance thresholds. The high quantiles (Q90, QMAX) of the ensemble forecast based on the AROME model (PEARO) available (but not shown) unfortunately did not give higher precipitation than the highest proposed by the deterministic runs. None of the ensemble members therefore allowed for a more pessimistic scenario. The Expected Utility **(Gillet-Chaulet, 2020)** could not therefore lead to a correct decision in this case. The important point to emphasize here is that the forecasters on duty nevertheless opted for an orange vigilance on the basis of the available forecasts (those up to D-1 6H, **figure 5**, for the first vigilance watch map of 18/09 14Hr, **figure 2**, issued for the event). This decision making which deviated from the NWP data, based on the experience of the forecasters, is reminiscent of many cases described by **Klein (1999)** using the Recognition-Primed Decision (RPD) Model. Even if the department is not correctly targeted, as in this case which is a problem in terms of false alarm and non-detection, it is considered that the vigilance is easier to amend after the decision to issue an orange has been taken both in terms of location and intensity of the phenomenon. This is what occurred on the following vigilance watch map (published on 19/09 04Hr, **figure 2**) based on the forecast runs D-1 12H, 18H and D 00H **(figure 5)**. It can be noted that the forecast run D-1 18H **(figure 5)** was the closest to the reality. Unfortunately, the following forecasts which were closer to the event's deadline became much less relevant! Moreover, such an underestimation was unusual by the AROME model. This behaviour was surprising.

Post-event experiments showed that this situation was exacerbated by sensitivity to calculation conditions. Indeed, at that time, Météo-France was preparing for the arrival of its new supercomputers (see: **News and updates from a selection of NMSs, The European Forecaster, Newsletter of the WGCEF N°26, September 2021**). A version identical to the operational version of the AROME model was tested in parallel on these computers; a so-called 'mirror' version, same computer code, same observations. Surprisingly, it gave very different and better results than the operational version (**communication Pierre Brousseau, CNRM/GMAP, figure 6**). This type of chaotic



▲ *Figure 6: An experiment on the sensitivity to computational conditions: the same model, with the same observations, was run in operational mode ("OPER") and in mirror mode ("MIRROR") on two different computers. The forecasts are very different! Maximum 24 H accumulated precipitation (in mm) over 200 mm, top right. "Base Time" indicated at top left; Valid time: the day of event (19 September 2020).*

behaviour, well known with the medium-range forecasts, is less well known in the short-range, at convective scale. However, the same causes are at the origin of it.

This example therefore highlights, if that were needed, that in some cases of intense, high impact convection the deterministic approach has its limits. A high variability of successive forecasts should be a warning. Above all, the most recent forecasts are not necessarily the best, even if it is natural to give them more weight in the warning process. This type of behaviour can be found in many extreme weather situations. An ensemble approach is therefore required. However, the availability of an efficient ensemble forecast at a convective scale still largely remains a challenge.

D-day, the follow-up phase of the event

Once the level of vigilance has been set, we enter the field of *monitoring the situation*. The task consists of assessing its evolution and if necessary amending the products as soon as possible. The operational organisation of the forecasting services means that it is not necessarily the person who issued the warning who is in charge of the monitoring. This has its advantages: It can avoid certain cognitive biases such as confirmation bias **(Cadet and Chasseigne, 2009)**. It also has disadvantages: how can the memory of earlier, more pessimistic forecasts be retained when the most recent forecasts suggest that a false alarm is becoming likely, demobilising the new forecasters attention? The handover between forecasters on the arrival of a new team is therefore a 'key moment' at the start of this monitoring process.

"a succession of pitfalls"

04H00: While the most recent numerical forecasts showed decreasing accumulated precipitation (see above, **figure 5**), the event seemed to be starting, with already high rainfall intensities **(figure 1b)**. The numerical predictions did not agree with the observations and the now-casting products were therefore irrelevant. Have you ever encountered this type of situation? Doesn't the forecaster feel helpless or lost? On which foundation should the forecast now be based? Who can claim to make an accurate forecast in this context?

05H00-06H30: Hourly intensities exceeded 50 mm. The onset of the event, which had been set for 8 am, had to be brought forward. The production was amended accordingly, with total accumulated precipitation revised upwards (300 mm) in view of those already observed **(figure 1b)**. Surprisingly, the civil security authorities were requesting information on the evolution of the situation for the next day, over regions located further east, the forecast development area of the rainstorm activity. This request was disorienting as while a legitimate concern was apparent among forecasters, the current situation was not drawing the right level of attention. How to focus the required attention of the authorities to the current situation?

08H00: The flood forecasting services noted the first reactions of the rivers which had been all placed in green vigilance (they were almost dry at that time, see above, **figure 1c**). Should the vigilance level have been raised to yellow? In fact, everything was moving very fast. A flash flood occurred on the Gardon d'Anduze. The orange level was required! A new vigilance watch map was therefore issued at 08H48 (not shown).
At the same time, questions appeared regarding the values recorded by the Valleraugue's rain gauge located at the epicentre of the phenomenon. The rain gauge indicated values higher than those of the rates given by the radars. This measuring device was not operated by Météo-France: So were these data reliable? As open-source information, this was quickly picked up and quoted on the social media networks: How could it be qualified by in real time? This type of situation, which is fairly recent in the history of NMSs, may put the latter in a difficult position when everyone now has access to this type of doubtful information. An *a posteriori* verification showed a systematic, but weak, overestimation by 6% with this instrument.

09H00: On the meteorological side, the observed accumulated rainfall was approaching red level vigilance values. The national authorities were contacted to ask for information about the impacts on the ground. Indeed, this criterion is taken into account in the decision to change to this colour. At that moment there was no knowledge of any particular intervention. At the end of the morning, the first images of flooding appeared on social networks. On the hydrological side, the level of the Gardon de Saint-Jean was worrying. The flood forecasting services were also considering a red update.

10H00: Finally, a lull seemed to be taking shape. Rainfall intensities were decreasing **(figure 1b)**. Could this have been the end of the episode? It was decided to remain in orange vigilance. The situation was serious enough for the meal break to be shortened.

11H00-12H00: Everything accelerated. As the rainy activity resumed, the emergency authorities were furious with Météo-France, accused of having "underestimated" the event. The feedback from the field was dramatic. One person was reported missing. Bridges were "washed away". The switch to red vigilance at 12H05 was considered to be very late in this context **(figure 2)**.

This warning was reduced the next day, on Sunday 20 September at 1 am (local time). This episode of unusual magnitude finally came to an end. The criticisms lasted a little longer. An evil for a good, a few days later, the unforgettable storm Alex occurred in the Alpes-Maritimes department (2 October 2020). The memory of the Gard event helped the forecasters not to procrastinate: (thanks also to a better predictability) the red vigilance was triggered right on time, with congratulations!

Conclusion

The narrative of the convective episode of 19 September 2020 in Gard (France) provided a compilation of the pitfalls encountered by forecasters in their warning exercise for severe weather. The following difficulties were combined: the intensity of the phenomenon escaped the NWP, both deterministic models and ensemble prediction systems, with the simulations closest to the event being the least accurate; now-casting products were of little use, being too far from the observed reality; insufficiently mobilised civil security organisations due to difficulties in communication (with *a posteriori* reproaches for them not having been awakened sufficiently!); rainfall observations considered dubious in real time, but widely disseminated as truth on the social media; Lack of on-the-ground situational awareness (or even a misleading absence of reports); a lull in precipitation at mid-episode (giving the impression that the worst had passed); extremely rapid but completely unexpected hydrological consequences due to a pre-existing drought.

This sequence of events led to a crisis situation in which the civil security authorities expressed their dissatisfaction. Paradoxically, the forecasters on duty felt that they had played their part and done some good work given the difficulties encountered, in particular for issuing an orange warning even though NWP forecasts did not reach the orange warning thresholds.

Even if numerical weather prediction continues to improve, it is likely that we, i.e. forecasters and authorities, will still be confronted with this type of event whose frequency could increase with climate change: similar situations may again escape even the best models. In order to improve the management of such events, some recommendations need to be sought.

Above all, it is useful to remember that these events have a highly non-linear dynamic. The consequences 'follow' a geometric progression, going from benign to catastrophic in a very short time, rather like the level of a river during a flash flood **(figure 1c)**. However, numerous psychological studies have shown that we, including 'seasoned experts', have difficulty conceiving this type of explosive evolution **(Bronner, 2021)**. We need to be aware of this, increasingly so with the expected consequences of climate change!

Progress is of course to be sought in the knowledge of these events, in particular the recognition of the cold pools which intervene in the convection regime leading to persistence of these violent storms. Better real-time qualification of these convective systems is needed: such as radar products which give indications on wind shear and could be very interesting for forecasters in that domain. Pending long-term improvements in numerical weather prediction, short-term improvements are to be sought systematically from the list of difficulties above. Such improvements will surely help to provide the precious minutes that will save lives and mitigate damage during future severe weather events.

References

Bohler S., 2021: “*Où est le sens, les découvertes récentes sur notre cerveau qui changent l’avenir de notre civilisation*”, Pocket, 358 p.

Bronner G., 2021: “*Apocalypse Cognitive*”, PUF, 386 p.

Cadet B., Chasseigne G., 2009: “*Psychologie du jugement et de la décision, des modèles aux applications*”, De Boeck, Bruxelles, 480 p.

Ducrocq V., Davolio S., Ferretti R., Flamant C., Santaner V. H., Kalthoff N., Richard E. and Wernli H., 2016 : “*Advances in understanding and forecasting of heavy precipitation in the Mediterranean through the HyMeX SOP1 field campaign*” - Special issue 142, SI, Q.J.R. Meteor. Soc., 2016.

Gillet-Chaulet B., 2020: “*Expected Utility, a benefit for the forecaster*”, The European Forecaster, Newsletter of the WGCEF, 25, September 2020.

Klein G., 1999: “*Sources of Power, How People Make Decisions*”, The MIT Press, Cambridge, Massachusetts, 332 p.

Seity Y., Brousseau P., Malardel S., Hello G., Benard P., Bouttier F., Lac C. and Masson V., 2010: “*The AROME-FRANCE convective-scale operational model*”, Mon. Wea. Rev., 139 (3), 976-991.