https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Systematic errors analysis of heavy precipitating events prediction using a 30-year hindcast dataset

Matteo Ponzano[1], Bruno Joly[1], Laurent Descamps[1], and Philippe Arbogast[1]
[1]CRNM/GAME, Météo-France/CRNS URA 1357, Toulouse, France
Correspondence: Matteo Ponzano (matteo.ponzano@meteo.fr)

Abstract. The western Mediterranean region is prone to devastating flash-flood induced by heavy precipitation events (HPEs), which are responsible for considerable human and material damage. Quantitative precipitation forecasts have improved dramatically in recent years to produce realistic accumulated rainfall estimations. Nevertheless, challenging issues remain in reducing uncertainties in the initial conditions assimilation and the modeling of physical processes. In this study, the spatial errors resulting from a 30-year (1981-2010) ensemble hindcast which implement the same physical parametrizations as in the operational Météo-France short-range ensemble prediction system, Prévision d'Ensemble ARPEGE (PEARP), are analysed. The hindcast consists of a 10-member ensemble reforecast, run every 4-days, covering the period from September to December. 24-hour precipitation fields are classified in order to investigate the local variation of spatial properties and intensities of rainfall fields, with particular focus on the HPEs. The feature-based quality measure SAL is then performed on the model forecast and reference rainfall fields, which shows that both the amplitude and structure components are basically driven by the deep convection parametrization. Between the two main deep convection schemes used in PEARP, we qualify that the PCMT parametrization scheme performs better than the B85 scheme. A further analysis of spatial features of the rainfall objects to which the SAL metric pertains shows the predominance of large objects in the verification measure. It is for the most extreme events that the model has the best representation of the distribution of object integrated rain.

1 Introduction

Episodes of intense rainfall in the Mediterranean affect western Europe climate and can have important societal impact. During these events, daily rainfall amounts associated to a one single event can reach annual equivalent values. These rainfall events coupled with a steep orography are responsible for associated torrential floods, which may cause considerable human and material damage. In particular, Southern France is prone to devastating flash flood events such as the Aude case (Ducrocq et al., 2003), Gard (Delrieu et al., 2005), and Vaison-La-Romaine (Sénési et al., 1996), which occurred on 12–13 November 1999, 22 September 1992 and 8-9 September 2002, respectively. For instance, in the Gard case more than 600 mm were

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences Discussions

observed locally during a two-day event and 24 people were killed during the associated flash flooding. Extreme rainfall amounts generally occur in a synoptic environment favourable for such events (Nuissier et al., 2011).

A detailed list of the main basic atmospheric ingredients which contribute to the onset of HPEs (Heavy Precipitation Event) are reported by Lin et al. (2001): 1) a conditionally or potentially unstable airstream impinging on the mountains, 2) a very moist low-level jet, 3) a steep mountain, and 4) a quasi-stationary synoptic system to slow the convective system over the threat area. In the Southeastern France, the Mediterranean Sea acts as source of energy and moisture to the lower levels and a pronounced orography is present above the Massif Central, Pyrenees, and South Alps areas (Delrieu et al., 2005). Extreme rainfall amounts are enhanced especially along the Southern and Eastern foothills of mountainous chains (Frei and Schär, 1998; Nuissier et al., 2008), in particular the Southeastern part of the Massif Central (Cévennes). Nevertheless, all the other regions are also affected by rainfall events with a great variety of intensity and spatial extension. Ricard et al. (2011) studied this regional spatial distribution based on a composite analysis in order to emphasize the climatological mesoscale environment associated with heavy precipitating events. They considered four sub-domains according to the location of precipitation. They found that the synoptic and mesoscale patterns can greatly differ as a function of the location of the precipitation.

A HPE could be convective (or not) or a combination of both (Ducrocq et al., 2002). Extreme rainfall amounts are generally associated with coherent structures slowed down and enhanced by the relief, whose extension must be larger than a single thunderstorm cell. Mesoscale processes can be crucial in organizing a very large variety of precipitating systems. For some cases, a mesoscale convective system (MCS) can stay stationary for several hours, affecting a limited area. Quasi-stationary MCSs are particularly efficient in terms of rain production due to their high intensities and their spatial stationarity (Nuissier et al., 2008). This stationarity is explained by the regeneration of new convective cells at a rate compensating the advective speed of the older cells (Ducrocq et al., 2008).

If long term territory adaptations are necessary to mitigate the impact of HPEs, a more reliable and anticipating alert would be beneficial in a short term. Weather forecasting coupled with hydrological impact forecast is the main source of information for weather warning triggering. Severe weather warnings are issued for the 24-hours forecast only. However, in some cases even at 24 hours term, the forecast process could be based on a model analysis some days prior to the issuance of the severe weather warnings. A better understanding of sources of model uncertainty at such time-range may represent a major source of improvement for early diagnosis.

Uncertainties of initial conditions and lateral boundary conditions for HPEs can be investigated for deterministic models (Argence et al., 2008) or ensemble models (Vié et al., 2010). Several journal articles studied the predictability associated to intense rainfall and flash floods (Walser et al., 2004; Walser and Schär, 2004; Collier, 2007). They showed that predictability limitations increase rapidly with decreasing scale since individual convective cells are rendered unpredictable by chaotic aspects of the moist dynamics. Moreover Quantitative Precipitation Forecast (QPF) appears to be more predictable in mountainous areas, where the triggering of convection and the larger-scale uplift results in a topographic control of the precipitation. Probabilistic forecast, based on ensemble prediction systems, is a suitable tool to explore the source of uncertainty for the predictability of HPEs (Du et al., 1997; Petroliagis et al., 1997; Stensrud et al., 1999; Schumacher and Davis, 2010; World Meteorological Organization, 2012). An ensemble forecast consists of several realizations of the evolution of the state of the atmosphere, in

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences
Discussions

order to assess the uncertainty associated to the forecast. Forecast uncertainty is a mix of initial and model errors propagation. Major meteorological centres implemented different methods in order to take into account initial errors, the most common are singular vectors (Buizza and Palmer, 1995; Molteni et al., 1996) , bred vectors (Toth and Kalnay, 1993, 1997) and perturbed observation in analysis process (Houtekamer et al., 1996; Houtekamer and Mitchell, 1998). Model errors can be simulated through a multimodel approach, adding a stochastic component to the tendencies from parametrization schemes (Palmer et al., 2009), stochastically backscattering energy into the model (Berner et al., 2009) or using different parametrization schemes for each forecast member (multiphysics approach: Charron et al., 2009; Descamps et al., 2011).

The framework set-up implemented for this study is a reforecast dataset built from a simplified version of the operational Météo-France short-range ensemble prediction system, Prévision d'Ensemble ARPEGE (PEARP; Descamps et al., 2015), in which only model uncertainties are represented, by means of a multi-physics approach. A reforecast ensemble dataset can be used for the calibration of the related version of the operational model (Hamill and Whitaker, 2006; Hamill et al., 2008; Hamill, 2012; Boisserie et al., 2015). Reforecast datasets have also been used to perform characterisation of a parameter forecast extremeness relatively to a reference, by comparing ensemble distributions to reforecast distribution like in Extreme Forecast Index computations (Boisserie et al., 2015; Lalaurette, 2003). In this study, the production of a 30-year reforecast dataset provides a statistical basis for the exploration of the climatology of the model configurations implemented in the operational ensemble system. Moreover this large dataset, spreading out over a multidecadal period, may include a significant number of intense events. We adopt a 10-km grid spacing reforecast ensemble to emphasize the predictability of mesoscale events rather than scattered and isolated phenomena, which are better represented by high-resolution models. The use of a coarser model resolution ensures a longer time integration for a given computing power. Consequently, predictability can be investigated up to 4 days lead time.

Traditional rainfall verification methods can be exploited in order to assess the quality of a forecast as they are generally built on the basis of a grid-point based approach. These techniques, especially when applied to intense events, are subject to timing or position errors leading to low scores (Mass et al., 2002). This combination of both spatial and timing errors is also known as the double penalty problem (Rossa et al., 2008). Spatial verification techniques have been developed with the goal to evaluate forecast quality in a manner similar to a forecaster approach and to overcome the traditional grid-point to grid-point verification limitations. A branch of spatial techniques is represented by the object-oriented verification methods (AghaKouchak et al., 2011; Ebert and McBride, 2000; Davis et al., 2006a, 2009; Mittermaier et al., 2015; Wernli et al., 2008). In this study, the feature-based quality measure SAL (Wernli et al., 2008, 2009) is used.

The aim of this paper is to suggest a methodology suitable for evaluating the performances of an ensemble reforecast in a context of intense precipitation events, using an object-oriented approach. In particular we focus on the quality of the spatial forecasts on the basis of the region of the domain affected by the precipitations. Besides the analysis of diagnostics from the SAL-metric, a statistical analysis of the 24-hour rainfall objects identified in the forecasts and the observations is performed in order to explore the spatial properties of the rainfall fields.

The data and the methodology are presented in section 2. In detail, section 2.1 describes the reforecast ensemble dataset and section 2.2, the generation of the daily rainfall reference and the statistical stratification of this product by means of a peak-

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences Discussions

over-threshold method and a clustering analysis. Results arising from the spatial verification of the overall reforecast dataset are presented in section 3.1. Section 3.2 presents separated SAL diagnostics for each physical parameterization scheme of the ensemble reforecast, and furtherly based on individual objects spatial properties. Conclusions are given in section 4.

2 Data and methodology

2.1 PEARP hindcast

The PEARP reforecast dataset consists of a 10-member ensemble computed daily from 1800 UTC initial conditions, covering four month (from September to December), every year of a 30-year period (1981-2010). This period has been chosen since HPEs occurence in the region considered is largest during the autumn season (see Fig. 3 from Ricard et al., 2011). It uses ARPEGE (Action de Recherche Petite Echelle Grande Echelle, Courtier et al. (1991)), the global operational model of Météo-France with a spectral truncation T798, 90 levels on the vertical, and a variable horizontal resolution (mapping factor of 2.4 with a highest resolution of 10 km over France). One ensemble forecast is performed every 4 days of the four-month period up to 108-hour lead time. Our initialization strategy follows the hybrid approach described in (Boisserie et al., 2016), in which first the atmospheric initial conditions are extracted from the ERA-Interim reanalysis (Dee et al., 2011) available at the European Center for Medium-range Weather Forecasts. Second, the land-surface initialization parameters are interpolated from an offline simulation of the land-surface SURFEX model (Masson et al., 2013) driven by the 3-hourly near-surface atmospheric fields from ERA-Interim. 24-hour accumulated precipitation forecasts are extracted on a $0.1^\circ \times 0.1^\circ$ grid, that defines the domain D (see Fig. 1c), which encompasses Southeastern France (Fig. 1a). The reforecast dataset does not have any representation of initial uncertainty, but it implements the same representation of model uncertainties (multiphysics approach) as in the PEARP operational version of 2016.

Nine different physical parametrizations (see Table 1) are added to the one corresponding to the ARPEGE deterministic physical package. Two turbulent diffusion schemes are considered: the Turbulent Kinetic Energy scheme (TKE; Cuxart et al., 2000; Bazile et al., 2012) and the Louis scheme (L79; Louis, 1979). TKE_{mod} is a slightly modified version of TKE, in which horizontal advection is ignored. For shallow convection different schemes are used: a mass flux scheme introduced by Kain and Fritsch (1993) and modified by Bechtold et al. (2001), thereafter the KFB approach, the Prognostic Condensates Microphysics and Transport scheme (PCMT; Piriou et al., 2007)), the Eddy-Diffusivity/Kain-Fritsch scheme (EDKF) and the PMMC scheme (Pergaud et al., 2009). The deep convection component is parametrized by either the PCMT scheme or the Bougeault (1985) scheme (thereafter B85). Closing the equation system used in these two schemes means relating the bulk mass flux to the in-cloud vertical velocity through a quantity γ qualifying the area coverage of convection. Two closures are considered, the first one (C1) based on the convergence of humidity and the second one (C2) based on the CAPE (Convective Available Potential Energy). B85 scheme originally uses the C1 closure, while PCMT uses alternatively the closure (C1 or C2) which maximizes the γ parameter. Physics package 2 uses a modified version of the B85 scheme in which deep convection is triggered only if cloud top exceeds 3000 m (B85_{mod} in Table 1). The same trigger is used in physics package 3 in which deep convection is parametrized using the B85 scheme along with a CAPE closure (CAPE in Table 1). Finally the oceanic flux is solved by means

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

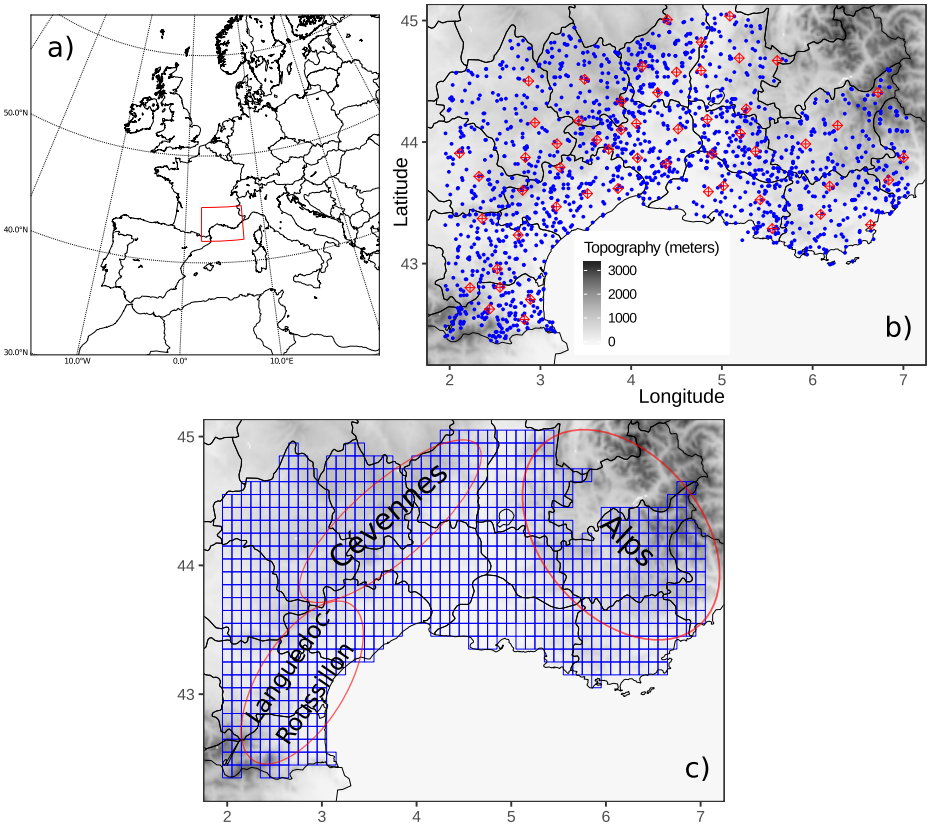


Figure 1. Panel **a** shows a situation map of the investigated area (rectangle with red edges) with respect to Western Europe and the Mediterranean Sea. Panel **b** shows the rain-gauges network used for the study. Red diamonds represent the rain-gauges selected for cross-validation testing, blue dots represent the rain-gauges selected for cross-validation training. Panel **c** shows the $0.1^\circ \times 0.1^\circ$ model grid (in blue), corresponding to the domain D, along with the location of three key areas.

of the ECUME scheme (Belamari, 2005). In $\text{ECUME}_{\text{mod}}$ oceanic fluxes are maximized. Control member and member 9 are characterized by the same parametrization set-up, but member 9 differs for the modelization of orographic waves.

2.2 Daily Rainfall Reference

24-hour accumulated precipitation is derived from the in-situ Météo-France rain-gauge network, covering the same period as
130 the reforecast dataset. 24-hour rainfall amounts collected from fourteen French departments within the reforecast domain D are used (Fig. 1b). In order to maximize the rain-gauge network density within the region, all daily available validated data covering the period have been used.

Rain-gauge observations are used to build gridded precipitation references by a statistical spatial interpolation of the observations. The aim of this procedure is to ensure a spatial and temporal homogeneity of the reference, as well as the same

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 1. Physical parametrizations used in the ensemble reforecast.

	Turbulence	**Shallow convection**	**Deep convection**	**Oceanic flux**
Ref	TKE	KFB	B85	ECUME
1	TKE	KFB	B85	$ECUME_{mod}$
2	L79	KFB	$B85_{mod}$	ECUME
3	L79	KFB	CAPE	ECUME
4	TKE_{mod}	KFB	B85	ECUME
5	TKE	EDKF	B85	ECUME
6	TKE	PMMC	PCMT	ECUME
7	TKE	KFB	PCMT	ECUME
8	TKE	PCMT	PCMT	ECUME
9	TKE	KFB	B85	ECUME

spatial resolution as the reforecast dataset. Ly et al. (2013) realized a review of the different methods for spatial interpolation of rainfall data. They showed that kriging methods outperform deterministic methods for the computation of daily precipitation. However, both types of methods were found to be comparable in terms of hydrological modeling results. For the interpolation, we use a mixed geo-statistical and deterministic algorithm, which implements Ordinary Kriging (OK; Goovaerts et al., 1997) and Inverse Distance Weighting methods (IDW, Shepard (1968)). For the kriging method three semi-variogram models (Exponential, Gaussian and Spherical) are fitted to daily sample semi-variogram drawn from raw and square root transformed data (G. Gregoire et al., 2008; Erdin et al., 2012). This configuration implies the use of six different geo-statistical interpolation models. In addition, four different IDW versions are used, by varying the geometric form parameter D used for the estimation of the weights (see eq. (2) in Ly et al. (2011)) and the maximum number n of neighbour stations involved in the IDW computation. Three versions are defined by fixing $d = 2$ and alternatively assigning n values equal to 5, 10 and N (with N being the total number of stations available for that specific day). In the fourth version we set $n = N$ and $d = 3$. For each day, a different interpolation method is used and its selection is based on the application of a cross validation approach. We select 55 rain-gauges as a training dataset (see the red diamonds in Fig. 1c) in order to have a sufficient coverage over the domain, especially on the mountainous area. Root Mean Square Error (RMSE) is used as a criterion of evaluation. For each day, the method which minimizes the RMSE computed within the rain-gauges of the training dataset is selected and the spatial interpolation is then performed on a regular high resolution grid of 0.05°. The highest resolution estimated points are then up-scaled to the 0.1°grid resolution of domain D, by means of a spatial average. This up-scaling procedure aims at reproducing the filtering effect produced by the parametrizations of the model on the physical processes that occur below the grid resolution.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

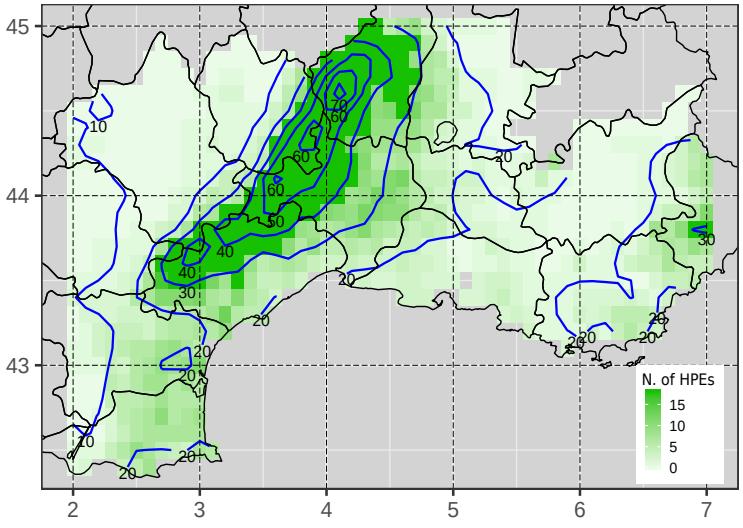


Figure 2. Annual average of HPEs occurrence per grid point (in green). The composite of daily rainfall amounts (mm/day) of the HPE dataset is represented by the blue isohyets.

2.2.1 HPE database

We implement a methodology in order to select the HPEs from the daily rainfall reference. Anagnostopoulou and Tolika (2012) have examined parametric and non-parametric approaches for the selection of rare events sampled from a dataset. Here we adopt a non-parametric peak-over-threshold approach, on the basis of the WMO guidelines (World Meteorological Organization, 2016). The aim is to generate a set of events representative of the tail of the rainfall distribution for a given region and season. Following the recommendation of Schär et al. (2016), an all-day percentile ($P_{0\leq n\leq 1}$) formulation is applied.

We proceed as follow: first the domain is split into two sub-regions based on the occurrence of climatological intense precipitations during the 30 years period. The sub-region A includes all the points whose climatological 99.5 percentile is lower or equal to a threshold T, subregion B includes all the other points. Threshold T, after several tests, has been set to 85 mm. This choice was made in order to separate the domain into two regions characterized by different frequency and intensity of HPEs. Then, a day is classified as a HPE if, for that day, there exists one point of sub-region A whose accumulated rainfall is greater than 100 mm or if there exists one point of sub-region B whose rainfall is greater than its 99.5 percentile. The selection led to a classification of 192 HPEs, corresponding to a climatological frequency of 5% over the 30-year period. The 24-hour rainfall amount maxima within the HPE dataset ranges from 100 mm to 504 mm. Figure 2 shows for each point of the domain the number of HPE, as well as the composite analysis of HPEs. The composite analysis involves computing the grid point average from a collection of cases. The signal is enhanced along the Cévennes chain and on the Alpine region. It is worth mentioning that some points are never taken into account for the HPE selection (grey points of Fig. 2), because the required

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 2. Classification of days computed from 24-hour rainfall amounts in southern France (1981-2010), percentage of HPEs and fraction of HPEs.

Cluster	Total (%)	HPEs (%)	Fraction of HPEs (%)
1	14.5	11.4	4.3
2	5.3	24.0	24.6
3	1.8	30.7	92.2
4	75.8	2.6	0.2
5	2.6	31.3	65.2
Total number of days	3660	192	

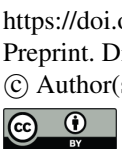
conditions are never satisfied. The analysis of the rainfall fields across the HPE database exhibits the presence of patterns of different shape and size, revealing potential differences in terms of the associated synoptic and mesoscale phenomena (not shown).

2.2.2 Clustering analysis

Clustering analysis methods can be applied to daily rainfall amounts in order to identify emergent regional rainfall patterns. This classification is largely used for assessing the between-day spatial classification of heavy rainfall (Romero et al., 1999; Peñarrocha et al., 2002; Little et al., 2008; Kai et al., 2011). We applied a cluster analysis, as an exploratory data analysis tool, in order to assess geographical properties of the precipitation reference dataset. The size of the dataset is first reduced and the signal is filtered out by means of a principal component analysis (Morin et al., 1979; Mills, 1995; Teo et al., 2011). The first 13 Principal Components (PCs), whose projection explains 90% of the variance, are retained. Then the *K*-means clustering method is applied. It is a non-hierarchical method based on the minimization of the intraclass variance and the maximization of the variance between each cluster. A characteristic of *k*-means method is that the number of clusters (K) into which the data will be grouped has to be *a priori* prescribed. Consequently, we have first to implement a methodology to find the number of clusters which leads to most classifiable subsets.

The analysis is applied to the full reference dataset, including rainy and dry days. We run 2000 tests for a range of *a priori* cluster numbers K that lie between 3 and 13, by varying a random initial guess each time. Then, for a given K, an evaluation of the stability of the assignment into each cluster is performed. The number of clusters is considered stable if each cluster size is almost constant from one test to another. $K = 5$ is retained as the most stable number of clusters and because it suggests a coherent regional stratification of the daily rainfall data. The final classification within the 2000 tests is selected by minimizing the sum of the distance between the cluster centroids from each test and the geometric medians of cluster centroids computed from all the tests. The test which minimizes this quantity has been selected as the reference classification. The results from the cluster classification are summarized in Table 2. The clusterization shows large differences in term of cluster size, more than 3/4 of the dataset is grouped in cluster 4, which collects mostly the days characterized by weak precipitation amounts or

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019



Natural Hazards and Earth System Sciences
Discussions

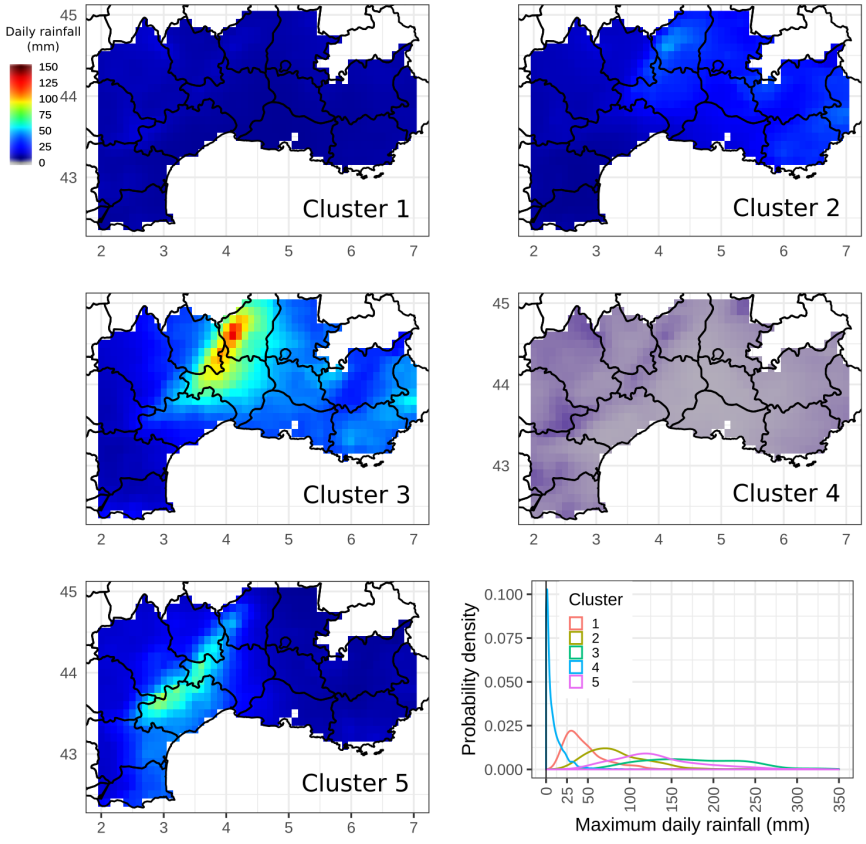


Figure 3. Rainfall composites (mm/day) for the 5 clusters selected by the *K*-means algorithm. The bottom-right panel shows the probability density distribution of the maximum daily rainfall (mm) for each cluster class.

dry days. The percentage of HPEs within the clusters shows that the most intense events are represented in clusters 2, 3 and 5, among which cluster 5 is the one with the largest proportion of HPE (86% of HPEs days within this cluster).

The same composite analysis as the one previously applied to HPE class, is now computed for each cluster class (Fig. 3). It reveals significant differences between clusters. Not only the relative intensity of events is different for each of the clusters, but also the location differs. Rainfall range is weak for cluster 1 and close to zero for cluster 4. Cluster 2 includes some moderate 24-hour rainfall amounts related to generalized precipitation events and a few of HPEs. For cluster 1, composite values are slightly higher on the northwestern area of the domain, while for cluster 2, rainfall amounts values are more enhanced on the eastern side of the domain D. Clusters 3 and 5 contains together 63% of the HPEs of the whole period, but rainfall events seem to affect different areas. Cluster 3 includes most of the events impacting the Cévennes mountains and the eastern departments on the southern side of the Alps. Cluster 5 average rainfall is enhanced along the southern side of the Cévennes, especially the Languedoc-Roussillon region.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences
Discussions

The bottom-right panel of Fig. 3 shows the density distributions computed from the maximum daily rainfall for each cluster. It is worth noting how each distribution samples different ranges of maximum grid-point daily rainfall amounts. Cluster 4 includes all the dry days. As this paper focuses on the most severe precipitation events, results will only be shown for cluster 2, 3 and 5 for the remainder of the paper.

2.3 The SAL verification score

2.3.1 The SAL score definition

The SAL score is an object-based quality measure introduced by Wernli et al. (2008) for the spatial verification of numerical weather prediction (NWP). It consists in computing three different components: structure **S** is a measure of volume and shape of the precipitations patterns, amplitude **A** is the normalized difference of the domain-averaged precipitation fields, and location **L** is the spatial displacements of patterns on the forecast/observation domains.

Different criteria for the identification of the precipitation objects could be implemented: a threshold level (Wernli et al., 2008, 2009), a convolution threshold (Davis et al., 2006a, b), or a threshold level conditioned to a cohesive minimum number of contiguous connected points (Nachamkin, 2009; Lack et al., 2010). The threshold level approach needs only one estimation parameter, so it has been preferred to the other methods for its simplicity and interpretability. Since we focus on the patterns associated to the HPEs, we decided to adapt the threshold definition given by $T_f = x_{max} \times f$, where x_{max} is the maximum precipitation value of the points belonging to the domain and f is a constant factor ($= 1/15$, in the paper of Wernli et al. (2008)). Here the coefficient f has been raised up to $1/4$, because a smaller value results in excessively large objects spreading out over most of the domain D. Choosing an higher f factor enables to obtain more realistic features within the considered domain. Thresholds levels T_f are computed daily for the reforecast and the reference dataset.

If we consider the domain D, the amplitude A is computed as follows:

$$A = \frac{\langle R_{\text{for}} \rangle_D - \langle R_{\text{obs}} \rangle_D}{0.5\,(\langle R_{\text{for}} \rangle_D + \langle R_{\text{obs}} \rangle_D)} \in [-2, 2]\,, \tag{1}$$

where $\langle \rangle_D$ denotes the average over the domain D. R_{for} and R_{obs} are the 24-hour rainfall amounts over D associated to the forecast and the observation, respectively. A perfect score is achieved for $A = 0$. The domain-averaged rainfall field is overestimated by a factor 3 if $A = 1$, similarly it is underestimated by a factor 3 if $A = -1$. The amplitude is maximal ($A = 2$) if $\frac{\langle R_{\text{for}} \rangle_D}{\langle R_{\text{obs}} \rangle_D} \to +\infty$ and minimal ($A = -2$) if $\frac{\langle R_{\text{for}} \rangle_D}{\langle R_{\text{obs}} \rangle_D} \to 0$.

The two other components require the definition of precipitation objects (thereafter $\{Obj\}$), also called features, which represent contiguous grid points belonging to the domain D, characterized by rainfall values exceeding a given threshold. The location L is a combined score defined by the sum of two contributions, $L1$ and $L2$. $L1$ measures the magnitude of the shift between the center of mass of the whole precipitation field for both in the forecast ($\bar{x}_{\text{for}}$) and observation ($\bar{x}_{\text{obs}}$):

$$L1 = \frac{|\bar{x}_{\text{for}} - \bar{x}_{\text{obs}}|}{d} \in [0, 1]\,, \tag{2}$$

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences
Discussions

where d is the largest distance between two boundary points of the considered domain D. The second metric $L2$ takes into account the spatial distribution of the features inside the domain, that is the scattering of the objects:

$$r = \frac{\sum_{n=1}^{N} M_n |\bar{x} - x_n|}{\sum_{n=1}^{N} M_n}, \tag{3}$$

where M_n is the integrated mass of the object n, x_n is the center of mass of the object n, N is the number of objects and $\bar{x}$ is the center of mass of the whole field.

$$L2 = 2\frac{|r_{\text{for}} - r_{\text{obs}}|}{d} \in [0,1]\,, \tag{4}$$

$$L = L1 + L2 \in [0,2]\,. \tag{5}$$

$L2$ aims at depicting objects differences between observed and forecasted scattering of the precipitation objects. We can notice that the scattering variable (eq. (3)) is computed as the weighted distance between the center of total mass and the center of mass of each object. Therefore L is a combination of the information provided by the global spatial distribution of the fields ($L1$) and the difference in the scattering of the features over the domain ($L2$). The location score is perfect if $L1 = L2 = 0$, so if $L = 0$ all the centers of mass match each others.

The S-component is based on the computation of the integrated mass M_k of one object k, scaled by the maximum rainfall amount of the object k:

$$V_k = \frac{M_k}{\max\limits_{x \in Obj_k} R(x)}. \tag{6}$$

Then, the weighted average V of all features is computed, in order to obtain a scaled, weighted total mass:

$$V = \frac{\sum_{n=1}^{N} M_n V_n}{\sum_{n=1}^{N} M_n}, \tag{7}$$

$$S = \frac{V_{\text{for}} - V_{\text{obs}}}{0.5(V_{\text{for}} + V_{\text{obs}})} \in [-2,2]\,. \tag{8}$$

Then, S represents the difference of both forecasted and observed volumes, scaled by their half-sum. It is important to scale the volume so that the structure is less sensitive to the mass, meaning that it relates more to the shape and extension of the features rather than their intensities. In particular $S < 0$ means that the forecast objects are large and/or flat compared to the observations. Inversely, peaked and/or smaller objects in the forecast give positive values of S. We refer to Wernli et al. (2008) for the exploration of the behaviour of SAL for some idealized examples.

On the basis of the definition of the score, it can be noticed that A and $L1$ components are not affected by the object identification and depend only on the total rainfall fields.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

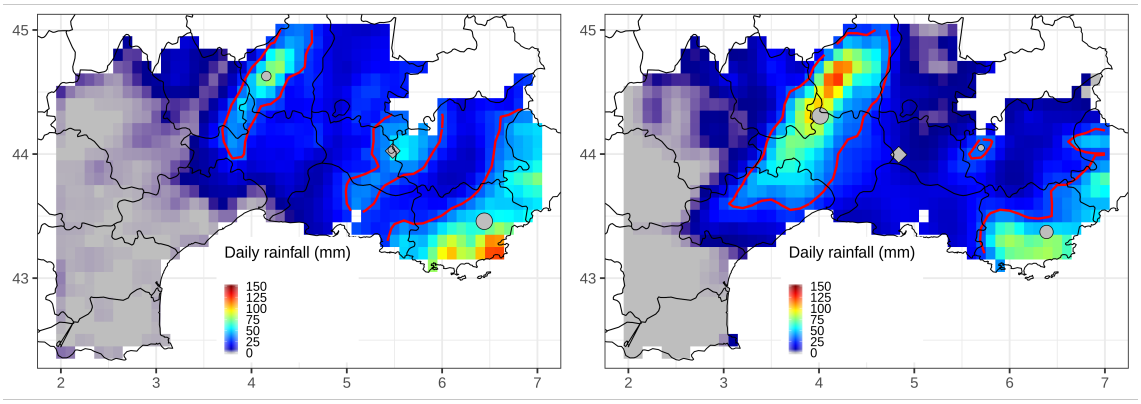


Figure 4. SAL pattern analysis for the case of 28 october 2004, applied on the observation data (left panel), and one 60-hour lead time forecast (right panel). Base contour of the identified objects are in red lines. Gray points stand for the rain barycentre of each pattern, gray diamond depicts the rain barycenter for the whole field. Size of the barycentre points is proportional to the the integrated mass of the associated object.

2.3.2 A selected example of the application of SAL

An example of the SAL score applied to an HPE, that occurred on the 28 Oct 2004, is shown in Fig. 4 (60-hour lead time forecast run using the physical package n.8). For the rainfall reference, a 24-hour rainfall maximum value (121.3 mm), was registered in the south-Eastern coastal region. Therefore the threshold level T_f is set to 30.3 mm. For the forecast, the maximum value is 123.1 mm ($T_f = 30.8$ mm) and, in contrast with the reference, it is located on the Cévennes. The number of objects, three, is equivalent in both fields. The value of A is 0.08, which means that the domain-averaged precipitation field of the forecast is nearly similar to the reference one. The structure S-components is positive (0.28), which could be explained by the larger forecast object over the Cévennes area, while the object along the south-eastern coast is smaller and less intense. The contribution of the third object is negligible for the computation of S. The location L-component L is equal to 0.23, with $L1$=0.13 and $L2$=0.10. The location error $L1$ means that the distance between the centres of total mass (see diamonds in Fig. 4) is $13/100$ of the largest distance between two boundary points of the considered domain. This error is mostly due to the fact that the most intense rainfall patterns are far apart from each other in the observations and the forecast.

3 Analysis of the reforecast HPEs representation

SAL verification score has been applied to the reforecast dataset to perform statistical analysis of QPF errors. The reforecast dataset is considered as a testbed model in order to study sources of systematic errors of the forecast. The overall reforecast performance is first examined for HPE/non-HPE cases, then according to the clusters. In a second step, the behaviour of the

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 3. Contingency table computed for rainy and dry days.

Contingency table	Obs rainy day	Obs dry day
Model rainy day	3258	84
Model dry day	226	62

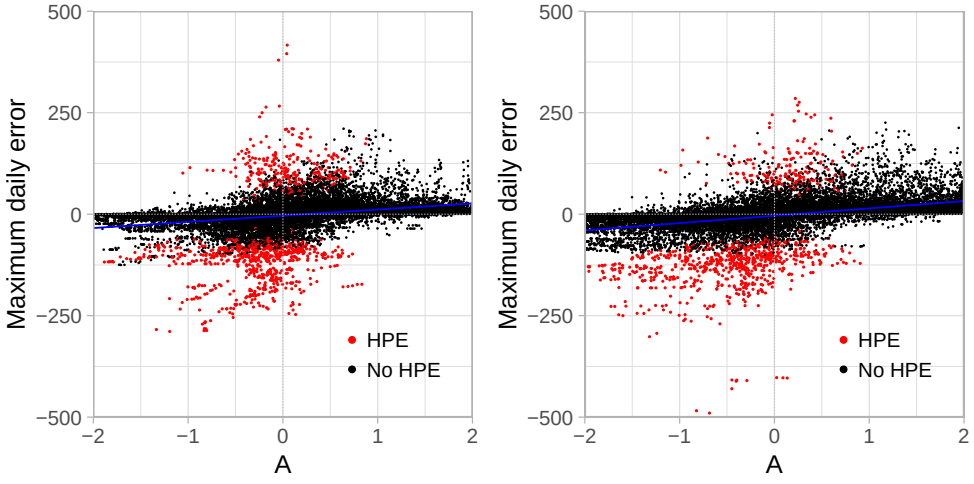


Figure 5. Relationship between the daily rainfall gridpoint maximum algebric error and the A-component of the SAL score. HPEs days are plotted in red, while other days are in black. Left panel is for LT12 lead time, right panel shows LT34 lead time. Linear regression analysis is added to the plot.

different physics schemes is analysed by considering separately the SAL results of each reforecast member. Similarly, the analysis is again allocated to HPE/non-HPEs cases and subsequently to each cluster.

For both the reforecast and the reference, we set all the days with at least one grid point beyond 0.1 mm as a rainy day. In order to facilitate the comparison between the parametrizations, SAL verification is only performed when all the members and the reference are classified as rainy day. Table 3 shows the contingency table of the rainy and dry days. Therefore 84 false alarms, 226 missed cases, and 62 correctly forecast dry days are not involved in the SAL analysis. No HPE days belong to the misses. The SAL measure is then applied to the 3258 rainy days.

3.1 SAL Evaluation of the HPEs forecast

3.1.1 HPE/non-HPE cases

First the relationship between the A-component of SAL and the maximum grid-point error is investigated (Fig. 5). 36-hour and 60-hour lead times (LT12 hereafter) and 84-hour and 108-hour lead times (LT34 hereafter) are grouped together. Maximum daily absolute errors ranges between -250 mm and 250 mm. Rare higher values are observed, which are likely related to strong

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

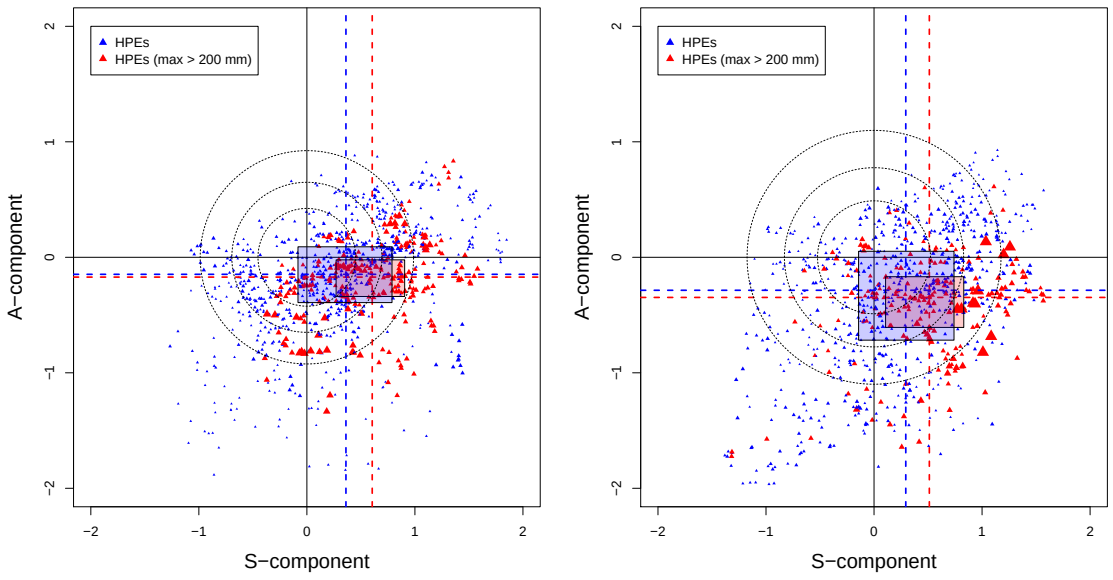


Figure 6. Relationship between the A-component and the S-component of the SAL score (SAL diagrams) for HPEs events only, for lead times LT12 (left) and LT34 (right). Blue triangles represent HPEs events with rainfall gridpoint maximum under 200 mm/day, and red triangles for rainfall amount beyond 200 mm, triangles are proportional to the rainfall value. Some main characteristics of the component distribution are plotted, the median value (dashed lines), percentile 25% and 75% delimitate the boxes. Circles represent the limits 25%, 50% and 75% percentiles to the best score (A=0, S=0).

double penalty effects that often occur in gridpoint-to-gridpoint verification. Points are mostly scattered along the amplitude axis showing that the error dependence on A-component is weak. Concerning HPEs, the scatter plot shows A-component values under 1, which means that the scaled average precipitation in the forecast never exceeds three times the observation. On the opposite, A-component negative values are predominant, in particular at LT34, in relation with strong underestimations of the domain-averaged rainfall field. Some cases of significant maximum grid-point errors in conjunction with moderate negative A-component must be related to strong location errors. In these cases, the domain-averaged field may be similar to the observed one while the maximum rainfall is spatially deviated. For the non-HPE days, we can see that, especially for LT34, the model could significantly overestimate both the A-component and the maximum grid-point error.

The relationship between the different SAL components might help to understand sources of model error. In Fig. 6 the S and A components are drawn for the HPE days only. Perfect scores are reached for the points located on the origin *O* of the diagram. A very few number of points are located on the top left-hand quadrant. This indicates that an overestimation of precipitation amplitude associated with too small rainfall objects is rarely observed. The points, especially for LT34, are globally oriented from the bottom left-hand corner to the top right-hand corner. This suggests a linear growth of the A-component as a function of the S-component, which means that the average rainfall amount is roughly related to the structure of the spatial extension. For

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

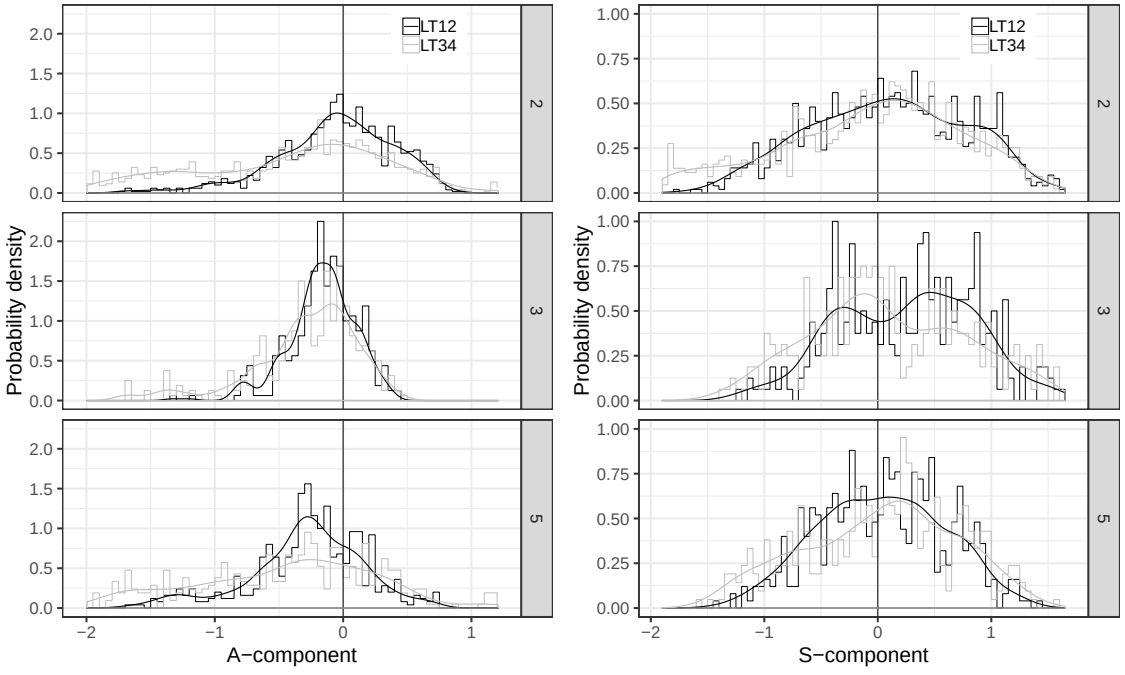


Figure 7. A-component (left column) and S-component (right column) normalized histograms and probability density functions for clusters 2, 3 and 5. Results for lead time LT12 are plotted in black lines and results for lead times LT34 are in grey.

the two diagrams, it can also be noticed that many of the points are situated in the lower-right quadrant, suggesting the presence of too large and/or flat rainfall objects compared to the reference while the corresponding A-component is negative. This is supported by the values of the medians of the two components distribution (dashed lines) and the quartile values (respective limits of the boxes). This positive bias in the S-component is even stronger for the most extreme HPEs (red triangles). This distortion of S-component error compared to A-component shows that the model has more difficulties reproducing the complex spatial structure than simulating the average volume of a heavy rainfall. An hypothesis to explain such a result might be that in order to reach rainfall amounts that occurs in HPEs, the model needs to produce rainfall processes of larger extension.

For each point of the diagram of Fig. 6 we compute its distance from the origin (perfect score (A=0; S=0)). The dotted circles respectively contain the 25%, 50% and 75% points with the smallest distance. The radius of the circles are much larger for LT34, confirming a degradation of the scores for higher lead time ranges.

3.1.2 Clusters

We use our clustering procedure (as defined in section 2.2.2) to analyze the characteristics of the forecast QPF errors along with the regional properties. SAL components are stated for each day of each cluster associated with HPEs, i.e. C2, C3 and C5. In Fig. 7, PDFs (Probability Density Functions) are drawn from the corresponding normalized histograms for the two lead times LT12 and LT34. The distributions of the A-component are negatively skewed for all the clusters. This reveals that the model

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 4. Pearson correlation between the daily mean S-component and the maximum daily rainfall for the three cluster classifications. A t-test is applied to the individual correlations. For the three clusters, the null hypothesis (true correlation coefficient is equal to zero) is rejected.

Cluster	LT12	LT34
2	**0.50**	**0.44**
3	**0.59**	**0.50**
5	**0.37**	**0.46**

tends to produce too weak domain-averaged rainfall in the case of heavy rainfall. This is even more important for clusters 3 and 5. For long lead times, the distributions are flatter, showing that the left tail of the A-component PDFs spreads far away from the perfect score.

The distributions of the S-component (right panels) are positively skewed in cluster 2 and 3, while they are more centred for cluster 5. For all the clusters, the spread of the S-component distributions is less dependent on the lead time, compared to the A-component distributions. It is interesting to examine whether a relationship between the S-component and the intensity of the rainfall is identifiable. A Pearson correlation coefficient is computed between the daily mean of S-component estimated within the ten members of the reforecast and the maximum observed daily rainfall for each cluster class (table 4). A positive correlation is found for all the three clusters which corroborates the results from Fig. 6 where HPEs correspond to the highest S-component values. Maximum correlation is found for cluster 3.

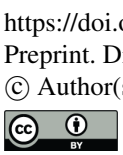
3.2 Sensitivity to physical parametrizations

The SAL measure is now analysed separately for the ten different physical packages to study corresponding systematic errors. More specifically, we raise the following questions: Do the errors based on an object-quality measure and computed for the different physics implemented in an ensemble system show different rainfall structure properties? Which physical packages are more sensitive to the intense rainfall forecast errors? As in section 3.1, we first distinguish the results for the HPEs group before the cluster ones.

3.2.1 HPEs

Probability density distributions for each SAL component are separately computed for each physics reforecast (Fig. 8), considering only the HPEs days. Colours lines correspond to four categories, depending on the parametrization of the deep convection. The figure highlights that members from each of the two main parametrization schemes, B85 and PCMT have similar behaviours. Considering the A-component, PCMT members are more centred around zero for LT12 than B85. This effect is higher for LT34, for which B85 and PCMT density distributions are more shifted. For LT34, more events with a positive A-component are associated to PCMT, whereas negative values are more recurrent in B85. The A-component never exceeds +1, but strong underestimations are observed. This range of values stems from the fact that this forecast verification is applied to a subsample of the observation limited to the most extreme events. For these specific events, a model underestimation is

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019



Natural Hazards and Earth System Sciences
Discussions

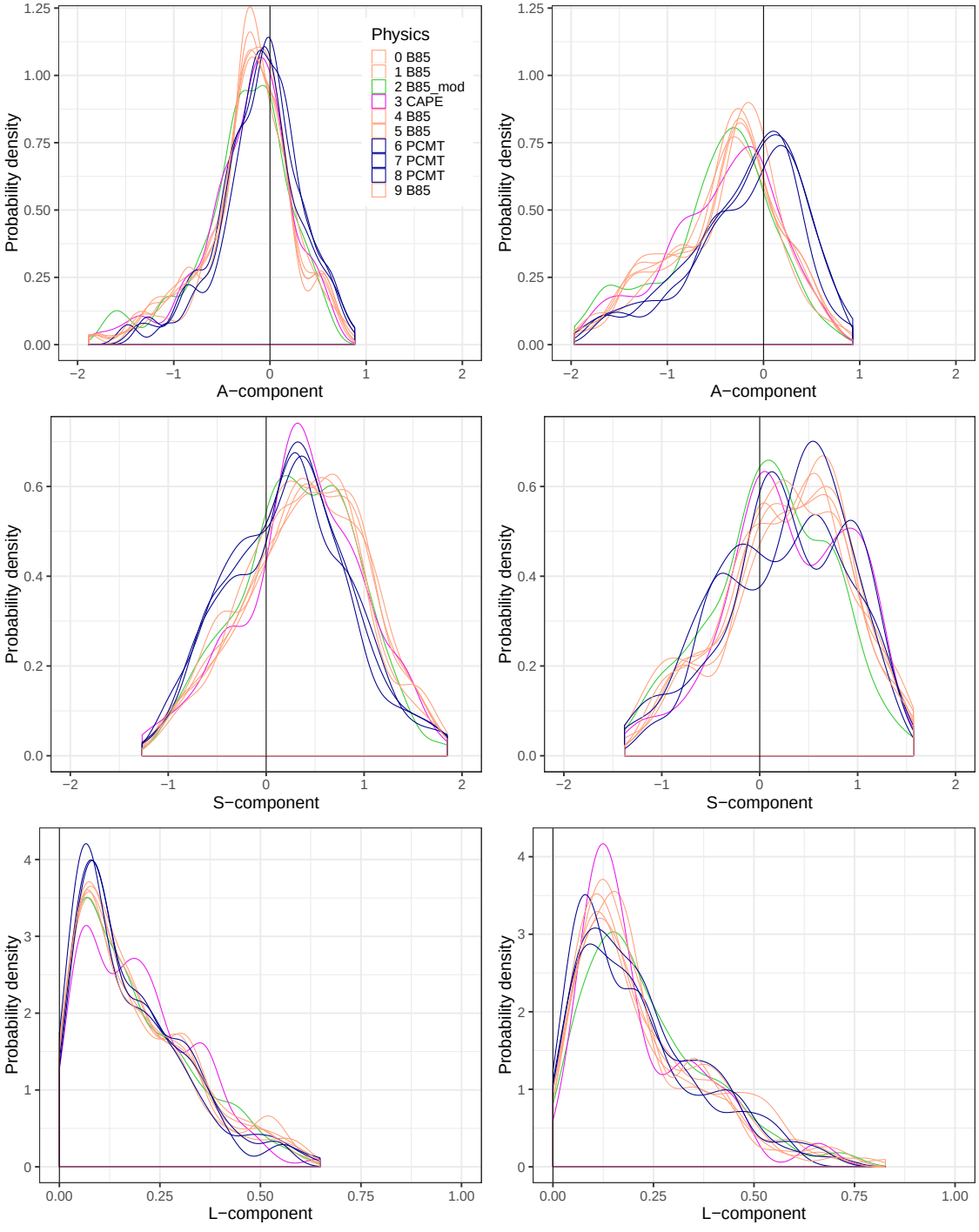


Figure 8. Probability density functions of the three SAL components for the HPEs and for each physics of the reforecast system (colored lines). Physics scheme are gathered in four categories depending on the parametrization of the deep convection (PCMT (blue), B85 (orange), $B85_{mod}$ (green), CAPE (purple). Left column corresponds to lead time LT12, and right column relates to lead time LT34.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

more frequent than an overestimation. At short lead times, the separation between the two deep convection schemes is also well established for the S-component (middle left panel), but it becomes mixed up for LT34 (middle right panel). A reason for this behaviour could be that predictability decreases for LT34, so that discrepencies of spatial rainfall structure assigned to the physics families become less identifiable. The S-component is positively skewed in all cases (in particular for the B85 physics at LT12 lead time). This supports the previous analysis of the S-component (Fig. 6 and 7), showing that for intense rainfall, model mostly produces larger and more flat rainfall signal. The results for the S-component also highlight better skills for PCMT schemes for HPEs events, especially at first lead times. Focusing on high values of S, B85 exhibits a stronger distribution tail at LT12, while both schemes seem comparable for LT34.

For the L-component, the maxima of the density distributions are higher for PCMT than B85 for LT12, implying a more significant number of good estimations of pattern location. Regarding the tail of the L-component PDF, it is globally more pronounced in LT34 than LT12. This means that the location of HPEs is poorly forecasted at long lead times. Concerning the behaviour of the forecasts that use the CAPE or $B85_{mod}$ schemes, their A-component PDFs are close to the B85 PDFs. This is not observed for the other components. For the S-component, the CAPE distribution follows, at LT12, the PCMT one. For the L-component, $B85_{mod}$ PDF is close to the B85 ones, while CAPE shows a behaviour different from all the others physics.

3.2.2 Clusters

According to the results of the previous section, which shows that the predictability of intense rainfall events is sensitive to the parametrization of the deep convection, we continue analysing the four different deep convection schemes model behaviours: B85, $B85_{mod}$, CAPE, and PCMT. The link between the behaviour of the physical schemes and the belonging to a particular cluster is statistically assessed through the SAL components differences between the schemes.

Any parametric goodness-of-fit tests, which assumes normality, have been discarded, because SAL values are not normally distributed. We choose the *k*-sample Anderson–Darling (AD) test (Scholz and Stephens, 1987; Mittermaier et al., 2015), in order to evaluate whether differences between two given distributions are statistically significant. It is an extension of the two-sample test (Darling, 1957), originally developed starting from the Classic Anderson-Darling test (Anderson and Darling, 1952). The *k*-sample AD test is a non parametric test designed to compare continuous or discrete sub-samples of the same distribution. In this case the test is implemented for the evaluation of the pairs of distributions. The two sample goodness-of-fit statistic A^2_{mn} is a sum of the integrated squared differences between two distributions functions:

$$A^2_{mn} = \frac{mn}{N} \int_{-\infty}^{+\infty} \frac{\{F_m(x) - G_n(x)\}^2}{H_N(x)\{1 - H_N(x)\}} dH_N(x) \tag{9}$$

where $F_m(x)$ is the proportion of the sample $X_1, \ldots, X_m$ that is not greater than x and $G_n(x)$ is the empirical distribution function of the second independent sample $Y_1, \ldots, Y_n$ obtained from a continuous population with distribution function $G(x)$ and $H_N(x) = \{(mF_M(x) + nG_n(x))/N$, with $N = m + n\}$ is the empirical distribution function of the pooled sample. Since n can differ from m, the test does not require samples with the same size. The above integrand is appropriately defined to be zero whenever $H_N(x) + 1$ is equal to zero. Under the null hypothesis H_0, for which $F(x) = G(x)$, the expected value of A^2_{mn}

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 5. The table provides the *p*-values computed from the *k*-sample AD test at 0.05 significance level for LT12. Upper right side of the table displays the values for the A-component obtained for pairs of distributions from the four physical package families. The elements between brackets represent the *p*-values computed for each of the three clusters retained after the clusterization. The lower left side shows results for the S-component. Bold values indicates where the null hypothesis is rejected, meaning that the difference between two distributions is statistically significant.

Physics	B85	$B85_{mod}$	CAPE	PCMT
B85	/	(0.77)(0.46)(0.65)	(0.83)(0.74)(0.56)	(0.16)(**0.00**)(**0.02**)
$B85_{mod}$	(0.24)(0.39)(0.60)	/	(0.99)(0.99)(0.78)	(0.14)(**0.01**)(0.13)
CAPE	(0.25)(0.80)(0.74)	(0.33)(0.66)(0.45)	/	(0.23)(**0.02**)(0.70)
PCMT	(**0.05**)(**0.02**)(**0.01**)	(0.61)(0.72)(0.38)	(0.13)(0.39)(**0.02**)	/

Table 6. As in Tab. 5, but for lead time LT34.

Physics	B85	$B85_{mod}$	CAPE	PCMT
B85	/	(0.16)(0.25)(0.91)	(0.45)(0.28)(0.13)	(**0.01**)(**0.00**)(**0.00**)
$B85_{mod}$	(0.23)(0.73)(0.40)	/	(0.92)(1.00)(0.40)	(**0.00**)(**0.01**)(0.08)
CAPE	(0.84)(0.77)(0.35)	(0.64)(0.93)(0.28)	/	(**0.01**)(**0.02**)(0.79)
PCMT	(0.24)(0.37)(0.50)	(0.63)(0.81)(0.77)	(0.50)(0.49)(0.47)	/

is 1. The test statistic is *standardized* using the expected value and the variance of A^2_{mn}, σ_N:

$$T_N = \frac{A^2_{mn} - 1}{\sigma_N} \tag{10}$$

The null hypothesis H_0 is rejected if T_N exceeds the critical value $t(\alpha)$, where α is the significance level, here set to 0.05. If this condition is verified, distributions are significantly different from each other at the 5% level.

The tests are performed for the comparison of each pairs of PDFs combined from the four deep convection families and from the three clusters classifications (Tab. 5 (LT12) and Tab. 6 (LT34)). Statistically significant differences are found for A-components for both LT12 and LT34. For the most intense events, $B85_{mod}$ and CAPE perform as B85 (see *p*-values for cluster 3 in the upper side of Tab. 5 and 6). It is worth noting that when comparing $B85_{mod}$ to CAPE, *p*-values are close to 1, so that these parametrizations exhibits the same mean behaviour for the A-component over the full dataset. As already observed in the HPEs analysis section, PCMT physics distributions depart significantly from B85 schemes at all lead times. PCMT, $B85_{mod}$ and CAPE exhibit differences for some clusters, especially at LT34. By focusing the attention on the most extreme clusters, it is worth noting that PCMT, $B85_{mod}$ and CAPE distributions are equivalent for cluster 5, but not for cluster 3. This means that $B85_{mod}$ and CAPE are alternatively close to B85 or PCMT, depending on the cluster.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

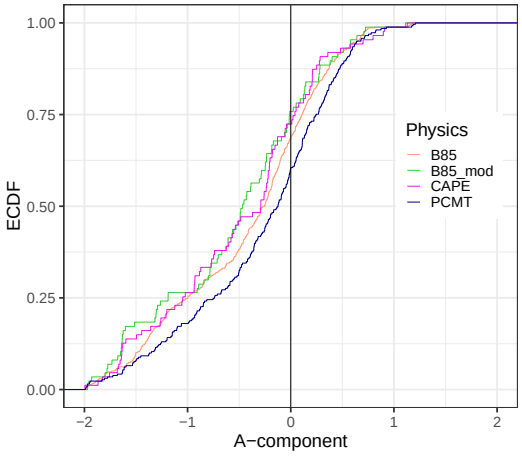


Figure 9. Empirical cumulative distribution function of of the A-component computed from the cluster 2 and lead time LT34 for the four classes of physics schemes.

In respect to the S-component distributions, k-sample AD tests show significant differences between B85 and PCMT physics for LT12. For longer lead times (LT34) structure quality measures converge towards a homogeneous distribution of the physics schemes (see p-values in the bottom side of Tab. 6), meaning that the differences between physics are negligible.

The test applied to the location component (not shown) does not reveal significant differences between the PDFs. We suppose that the limited dimensions of domain employed in this study, as well as its irregular shape, may lead to a less coherent estimation of the location, resulting in a degradation of the score significance. Since the L-component result is not informative about HPEs, it is ignored hereafter.

Once the statistical differences between the PDFs of the physics have been examined, it is interesting to compare the relative error on amplitude and structure components. S and A component errors are estimated by comparing the shapes of their distributions. Empirical Cumulative Density Functions (ECDF) of S and A components are computed separately for each cluster and lead time (LT12 and LT34). We show an example of an ECDF for cluster 2 and LT34 (Fig. 9). Forecasts are perfect when the ECDF tends towards an Heaviside step function, which means that the distribution tends towards the Dirac delta function centred on zero. The departure from the perfect score could be quantified, by estimating the area under the ECDF curve on the left side, and the area above the ECDF curve on the right side:

$$err_- = \int_{-2}^{0} F(t)dt - \int_{-2}^{0} H(x)dx = \int_{-2}^{0} F(t)dt - 0 = \int_{-2}^{0} F(t)dt, \tag{11}$$

$$err_+ = \int_{0}^{2} H(x)dx - \int_{0}^{2} F(x)dx = 2 - \int_{0}^{2} F(x)dx, \tag{12}$$

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 7. Forecast errors computed from the A-component distribution for the B85 and PCMT physics for each cluster classifications and lead time LT12. Grey lines denote clusters for which B85 and PCMT differences are not statistically significant.

	B85			PCMT		
Cluster	A_-	A_+	A_{tot}	A_-	A_+	A_{tot}
2	0.216	0.136	0.352	0.185	0.163	0.347
3	0.215	0.029	0.244	0.140	0.054	0.194
5	0.369	0.045	0.413	0.295	0.075	0.370

Table 8. As in Table 7, but for lead time LT34.

	B85			PCMT		
Cluster	A_-	A_+	A_{tot}	A_-	A_+	A_{tot}
2	0.530	0.107	0.637	0.429	0.146	0.575
3	0.312	0.023	0.336	0.294	0.061	0.355
5	0.608	0.064	0.672	0.451	0.138	0.589

405

$$err = err_- + err_+ = 2 - \int_0^2 F(x)dx + \int_{-2}^0 F(t)dt, \tag{13}$$

where $F(x)$ is the ECDF computed for A or S, $H(x)$ is the Heaviside step function and err is the forecast error for a given component. Since the previous *k*-sample AD test pointed out significant differences within the two main classes B85 and PCMT, the evaluation of the errors is restrained to these two specific classes.

410 The results for the A-component are shown in Table 7 (LT12) and Table 8 (LT34). Negative and positive errors increase with lead time. We note that the negative errors are always more important than the positive ones. This behaviour is strengthened at LT34, especially for clusters 3 and 5. This is not surprising since those two clusters collect the most extreme rainfall events. Indeed, the uncertainty of the forecast is supposed to be higher in the case of most intense rainfall events. Forecast hardly

Table 9. As in Table 7, but for the S-component.

	B85			PCMT		
Cluster	S_-	S_+	S_{tot}	S_-	S_+	S_{tot}
2	0.230	0.338	0.569	0.285	0.303	0.588
3	0.128	0.434	0.562	0.171	0.316	0.487
5	0.187	0.264	0.451	0.268	0.178	0.446

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

Table 10. As in Table 8, but for the S-component.

	B85			PCMT		
Cluster	S_-	S_+	S_{tot}	S_-	S_+	S_{tot}
2	0.326	0.305	0.631	0.381	0.279	0.660
3	0.214	0.337	0.551	0.246	0.310	0.556
5	0.278	0.289	0.568	0.269	0.265	0.535

produce as many rainfall amount as it is observed, especially for the longest lead times, since temporal error can lead to strong underestimations.

PCMT produces overall better A-component statistics, in particular the A-component negative contribution is reduced. Concerning the S-component evaluation, results are shown in Tables 9 and 10. The best forecasts are observed for clusters 3 and 5. This suggests that the forecast of the structure of the object depends on the considered phenomenon. In particular, neglecting the location error, shape and size of rainfall patterns are better forecasted for heavy rainfall events (clusters 3 and 5), rather than for the remaining classes of events. In contrast to the A-component, the S-component exhibits the highest error on the right side of the distribution for B85 scheme, whereas this trend is not systematic in PCMT physics. Restraining the analysis to LT12 PCMT globally performs better than B85, as the positive bias of the S-component is reduced. As with the amplitude A, the S-component gets worse for longer lead time, resulting in a shift to more negative values of the distribution for both B85 and PCMT physics.

3.2.3 Rainfall object analysis

We now analyze the physical properties of the objects, i.e. the number of objects from a rainfall field, and the object integrated volumes and surfaces, according to the different clusters. All the statistics are applied separately to the B85, PCMT physics and observations. For each day of the dataset period, the thresholds defined in subsection 2.3.1 lead to the identification of a certain number of precipitating objects. The frequency of this number of objects per day is plotted by means of normalized histograms for the three clusters (Fig. 10). Cluster 2 and 3 show maximum frequency for one and three objects range, whereas cluster 5 is dominated by one object per day. A visual inspection of the individual cases of single object rainfall patterns in cluster 5 suggests that the zone of the domain affected by objects can be crucial. Single objects extend mainly over the Languedoc-Roussillon region (cluster 5), while the other clusters display rainfall accumulated bands frequently split over the domain, typically over the Cévennes and Alpine regions. Objects identification for PCMT forecast shows the existence of an overestimation of single objects days compared to the observation and to B85 physics scheme, a behaviour emphasized in clusters 3 and 5.

More details about the magnitude of the objects can be achieved by computing the integrated mass per object, M_k (see subsection 2.3.1). First, for each day, objects are sorted from the largest to the smallest integrated mass. Integrated mass distribution of the two *heaviest* objects (noted O_1 and O_2) are then dispatched as a function of the number of objects for each

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

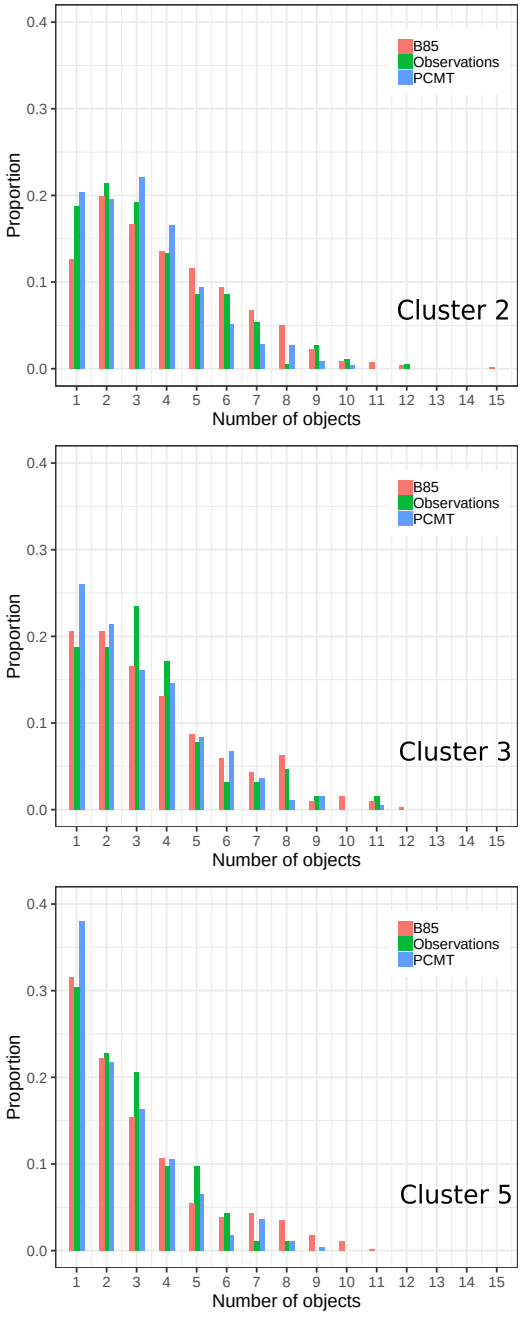


Figure 10. Normalized histograms of the daily number of SAL patterns, for B85 physics scheme (red), PCMT (blue), Observation (green). Panels correspond to the 3 clusters classification.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

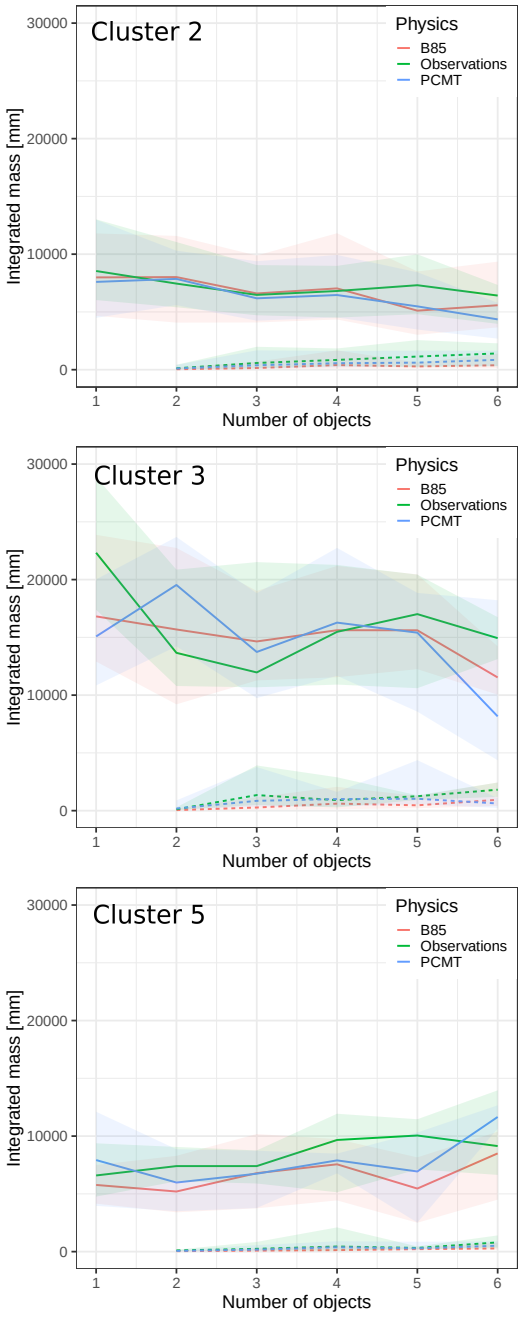


Figure 11. Distribution of SAL first pattern O_1 rain amount according to the number of patterns per day. Curves stand for the median of the distribution, shaded areas range between 25% and 75% percentiles. The dashed lines correspond to the second ranked SAL pattern O_2 rain amount.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

cluster on Fig. 11. First, the range value of M is highly variable from one cluster to another. Maximum values are observed for cluster 3, while the magnitude for clusters 2 and 5 is comparable. The decrease of O_1 M is more clear for cluster 3, meaning that a high number of objects over the domain leads to a natural decrease of the M value of the heaviest ones. We think that a part of the total integrated mass is then redistributed to the other objects. This is confirmed on O_2 curves since its mass increases with the number of objects. Conversely, for cluster 5, O_1 mass increases with the number of the objects, while O_2 is almost stable. The gap between O_1 and O_2 masses is maximum in the most extreme clusters (3 and 5). This suggests that when computing the volume V (see eq. ((7))) and L2 (see ((4))), the weighted average is dominated by the object O_1. This implies that the verification could be considered as a single to single object metric.

The integrated mass M is only partially informative about the intensity of accumulated rainfall because it depends also on the spatial extension of object, also called the object base area. We define as R^* the integrated individual object mass M, weighted by its base area. The same statistics than previously are shown for R^* in Fig. 12. Compared to M, the gap between O_1 and O_2 is significantly reduced for R^*, even if R^* is still larger for O_1 than for O_2. R^* reaches the greatest values for cluster 3, while, in contrast with the results from M, cluster 5 exhibits higher R^* compared to cluster 2. This difference is explained by considering the object base area values. Pattern spatial extension are frequently larger for cluster 2, than for cluster 5. Orography leads cluster 2 to have more spatially extended objects with a weaker scaled object mass R^* than those of cluster 3.

The clusters associated with rainfall events impacting the Cevennes and eastern area of the domain D (clusters 3 and 5) are characterized by similar values of base area (not shown). Accordingly, they collect similar phenomena, but for two distinct classes of intensities. It can also be noted for cluster 5 that R^* is slightly decreasing, meaning that base area values increase faster than integrated mass values per number of identified objects.

Comparing observed and forecast objects we can see that the scaled pattern mass criterion highlights the gap between observations and models for O_1 and O_2, especially for clusters 3 and 5. B85 physics usually underestimates R^* compared to the observations except for the highest number of objects. On the contrary, for PCMT the departures between models and observations for R^* are higher in the most extreme clusters (3 and 5), showing a relation between the error and the magnitude of the observed variable.

We now examine the ratio between the daily maximum rainfall of objects O_1 and O_2. This ratio ranges between 1.5 and 3 which means that O_1 represents the essential contribution of the daily rainfall peak, even when its scaled object mass R^* is close to O_2. Since O_1 base area tends to be significantly larger that O_2, the information related to the inner object maximum rainfall is diluted in the large base area, resulting in a flat weak mean intensity of the object. This last result appears to support the fact that SAL metric gives more weight to the object that contains the most intense rainfall.

The comparison between the model reforecast physics and the observations is addressed using the whole distribution of daily mass M from the objects O_i identified across the full reforecast dataset, where i ranges between 1 and the total number N of objects. We proceed separately for each physical package. For a given scheme and cluster, the quantile values corresponding to the selected dataset are sorted in ascending order, and then plotted versus the quantiles calculated from observations (Fig. 13). Half of the quantile distributions are not visible as they correspond to very weak pattern masses. For cluster 2 and PCMT physics most of the distribution of object mass is close to the observations, however all the other physics distributions are

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

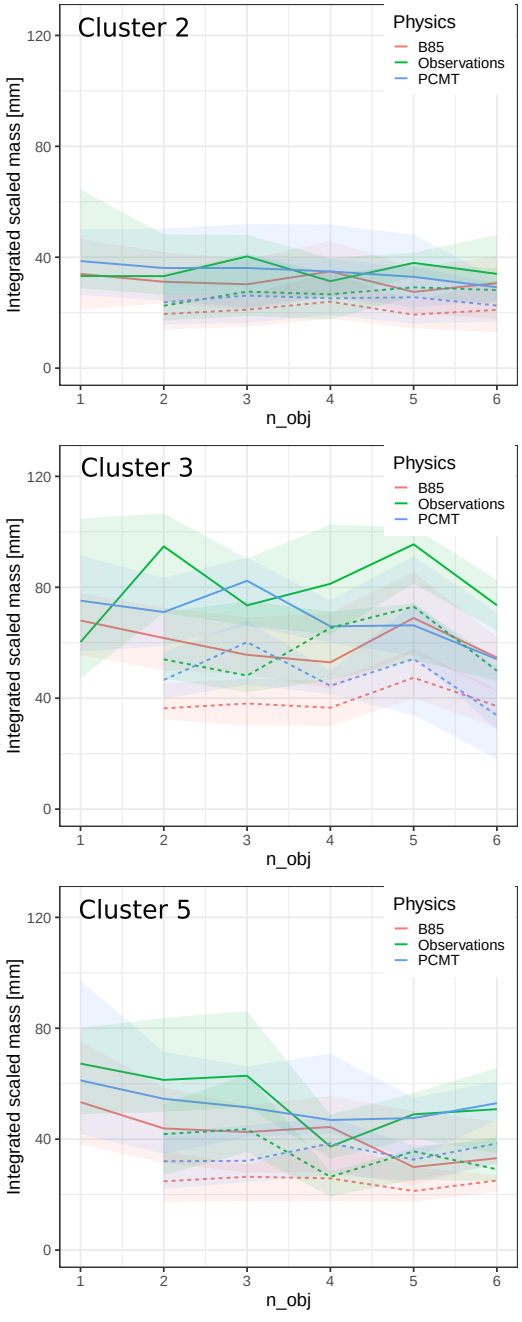


Figure 12. As in Fig. 11, but for rain amounts scaled by the pattern base area surface.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions

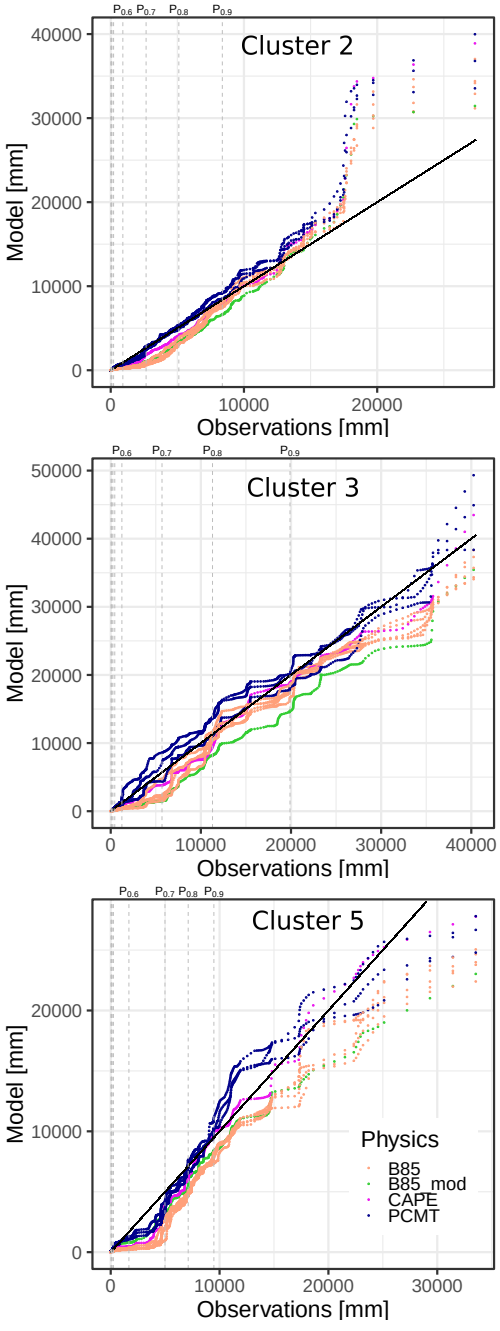


Figure 13. Quantile-quantile plot between SAL pattern rain amounts from the model (Y-axis) and from the observation (X-axis). Physics schemes are gathered into 4 classes (B85, PCMT, B85mod, CAPE). Observation deciles correspond to the vertical dashed lines.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences Discussions

skewed to the right compared to the observations for values below 10000. This behaviour is also observed for cluster 5 and it involves PCMT physics as well, for values between percentile 0.5 and percentile 0.7. Overall, in the quantile-quantile plot for cluster 5, the PCMT outperforms B85. In cluster 3, discrepancies between PCMT, B85 and the observations are of opposite sign, PCMT being slightly above the observations. CAPE physics distribution is left skewed compared to the observations and to the others physics. These results reveal some interesting properties of the models in predicting the rainfall objects. For the most extreme clusters, object mass distribution of physics is similar to the distribution drawn from the observation, especially for cluster 5. This means that the forecast is able to reproduce the same proportion of rainfall amounts inside a feature as the observations, even concerning the extreme right tail of the distributions, which corresponds to the the major events of the series.

4 Summary and conclusions

In this study we have characterized the systematic errors of 24-hour rainfall amounts from a reforecast ensemble dataset, covering a 30-year fall period. A 24-hour rainfall reference has been produced with the same model resolution as the reforecast one in order to have access to a point-to-point verification. We applied an object-based quality measure in order to evaluate the performance of the forecasts of any kind of HPEs. Then, we take advantage of a rainfall clustering to analyse the dependence of systematic errors to clusters.

The selection of the HPEs within the reference dataset was based on a peak-over-threshold approach. The spatial regional discrepancies between HPEs are studied based on the k-means clustering of the 24-hour rainfall. Finally, we analysed the rainfall objects properties repectively in the model and in the observation to underline the rainfall field object properties for which the model acts distinctly.

The peak-over-threshold criterion leads to the selection of 192 HPEs, confirming that the most impacted region are the Cévennes area and part of the Alps. Even though HPEs affects predominantly the mountainous areas, severe precipitating systems can occur in plain areas, especially on the foothills oriented towards the meridional fluxes. The composite analysis for the five clusters reveals that each cluster is associated to a specific class and location of 24-hour precipitation events. It was found that 86% of the total of HPEs are included in clusters 2, 3 and 5. Cluster 2 and 3 patterns impact predominantly the Cévennes and Alps area, while the cluster 5 the Languedoc-Roussillon region. Moreover clusters 3 and 5 are the most extreme ones, while cluster 4 contains weak rainfall events or dry days. Diagnostics for clusters 2, 3 and 5 only are considered.

Model performances analysis have lead to several distinct results that we outline in the following.

SAL object-quality measure has been applied distinctly to the ten model physics members of the reforecast dataset and compared to the rainfall reference. This shows that the model overall behaviour is characterized by negative A-components and positive S-components. The model objects are generally more flat and large objects than the observed ones, and moreover their corresponding domain-average amplitude is weaker. For all computed performance diagnostics, it has been found a degradation of SAL scores along with the lead times, comparatively with quantitative rainfall diagnostics.

When SAL diagnostics are performed according to the clusters, the A-component is negative-skewed, and it is enhanced notably for the most extreme clusters (over the Cévennes and over the Languedoc-Roussillon). Concerning the structure S-

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019


Natural Hazards and Earth System Sciences Discussions

component behaviour, diagnostic is dependent to the clusters. It is slightly positively skewed for cluster 2 and 3, while for cluster 5 the distribution of the S-component is more centred. This might indicate that heavy rainfall episodes over the relief regions (Cévennes, Alps) are represented by the model by flat and large pattern spreading out on a larger zone compared to the observations. For cluster 5, this effect is not found, and at that point it is difficult to determine whether this is characterising more a contrast in the model behaviour or whether it is due to the physical properties of the cluster 5 events.

The performances of the model are then investigated separately for each physical scheme composing the reforecast dataset, emphasising mostly the role of the deep convection physical parametrisation. In terms of SAL, the two deep convection schemes, B85 and PCMT, clearly determine the behaviour of the model. However, for lead time ranges higher than three days, no significant differences appear. It has been measured quantitatively that PCMT members performs better than B85 ones in terms of both SAL diagnostics, A and S components. S-component analysis shows to be better also for HPEs rather than for weak or moderate events which means that the predictability of pattern structure is higher for HPEs.

The second part of the study was dedicated to the characterization of rainfall objects properties in the model and in the reference, cluster by cluster. Cluster 5, which depicts essentially the precipitating objects that impacts the Languedoc-Roussillon, is the only cluster characterized by single object rainfall field. The analysis of object masses distribution of the two first sorted objects (O_1, O_2), shows that the second ranked object weight is weaker also in term of inner rainfall maximum which means that the weight of the larger object O_1 is preponderant in the SAL-analysis.

The analysis of the ranked distributions (quantile-quantile analysis) of the object masses shows that weakest precipitations are overestimated by all physics schemes. On another hand, the object mass distributions are relatively close between all the physics scheme and the observation for most extreme rainfall events, specially for the PCMT deep convection scheme.

The inter-comparison between some model physics deep convection scheme and their role in HPEs predictability shows it is of course very sensitive for designing multi-physics type of ensemble forecasting systems. Even if the sensitivity to the initial perturbations was not studied in this work, the forecast of intense rainfall seems to be mainly driven by the classes of deep convection parametrizations. Since physical parametrization set-up is built by replicated schemes, the model error representation might lack of an exhaustive sampling of the forecasted trajectories. Using more than two deep convection parametrization schemes may improve the representation of model errors, at least for heavy precipitating events.

Data availability. Research data can be accessed by contacting Matteo Ponzano at his e-mail address matteo.ponzano@meteo.fr and the other authors.

Author contributions. MP, BJ, and LD conceived and designed the study. MP carried out the formal analysis, wrote the whole paper, made the literature review, and produced the observation reference dataset. BJ built the hindcast dataset. BJ, LD, and PA reviewed and edited the original draft.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

References

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E.: Evaluation of Satellite-Retrieved Extreme Precipitation Rates across the Central United States, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2010JD014741, 2011.

Anagnostopoulou, C. and Tolika, K.: Extreme Precipitation in Europe: Statistical Threshold Selection Based on Climatological Criteria, Theoretical and Applied Climatology, 107, 479–489, https://doi.org/10.1007/s00704-011-0487-8, 2012.

Anderson, T. W. and Darling, D. A.: Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes, The Annals of Mathematical Statistics, 23, 193–212, https://doi.org/10.1214/aoms/1177729437, 1952.

Argence, S., Lambert, D., Richard, E., Chaboureau, J.-P., and Söhne, N.: Impact of Initial Condition Uncertainties on the Predictability of Heavy Rainfall in the Mediterranean: A Case Study, Quarterly Journal of the Royal Meteorological Society, 134, 1775–1788, https://doi.org/10.1002/qj.314, 2008.

Bazile, E., Marquet, P., Bouteloup, Y., and Bouyssel, F.: The Turbulent Kinetic Energy (TKE) scheme in the NWP models at Meteo France, in: Workshop on Workshop on Diurnal cycles and the stable boundary layer, 7-10 November 2011, pp. 127–135, ECMWF, ECMWF, Shinfield Park, Reading, 2012.

Bechtold, P., Bazile, E., Guichard, F., Mascart, P., and Richard, E.: A Mass-Flux Convection Scheme for Regional and Global Models, Quarterly Journal of the Royal Meteorological Society, 127, 869–886, https://doi.org/10.1002/qj.49712757309, 2001.

Belamari, S.: Report on uncertainty estimates of an optimal bulk formulation for surface turbulent fluxes, MERSEA IP Deliverable 412, pp. 1–29, 2005.

Berner, J., Shutts, G. J., Leutbecher, M., and Palmer, T. N.: A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System, Journal of the Atmospheric Sciences, 66, 603–626, https://doi.org/10.1175/2008JAS2677.1, 2009.

Boisserie, M., Descamps, L., and Arbogast, P.: Calibrated Forecasts of Extreme Windstorms Using the Extreme Forecast Index (EFI) and Shift of Tails (SOT), Weather and Forecasting, 31, 1573–1589, https://doi.org/10.1175/WAF-D-15-0027.1, 2015.

Boisserie, M., Decharme, B., Descamps, L., and Arbogast, P.: Land surface initialization strategy for a global reforecast dataset, Quarterly Journal of the Royal Meteorological Society, 142, 880–888, https://doi.org/10.1002/qj.2688, http://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.2688, 2016.

Bougeault, P.: A Simple Parameterization of the Large-Scale Effects of Cumulus Convection, Monthly Weather Review, 113, 2108–2121, https://doi.org/10.1175/1520-0493(1985)113<2108:ASPOTL>2.0.CO;2, 1985.

Buizza, R. and Palmer, T. N.: The Singular-Vector Structure of the Atmospheric Global Circulation, Journal of the Atmospheric Sciences, 52, 1434–1456, https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2, 1995.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gagnon, N., Mitchell, H. L., and Michelin, L.: Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System, Monthly Weather Review, 138, 1877–1901, https://doi.org/10.1175/2009MWR3187.1, 2009.

Collier, C. G.: Flash Flood Forecasting: What Are the Limits of Predictability?, Quarterly Journal of the Royal Meteorological Society, 133, 3–23, https://doi.org/10.1002/qj.29, 2007.

Courtier, P., Freydier, C., Geleyn, J., Rabier, F., and Rochas, M.: The ARPEGE project at Météo-France, ECMWF Seminar proceedings, vol. II. ECMWF Reading, UK, pp. 193–231, 1991.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Cuxart, J., Bougeault, P., and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations, Quarterly Journal of the Royal Meteorological Society, 126, 1–30, https://doi.org/10.1002/qj.49712656202, 2000.

Darling, D. A.: The Kolmogorov-Smirnov, Cramer-von Mises Tests, The Annals of Mathematical Statistics, 28, 823–838, https://doi.org/10.1214/aoms/1177706788, 1957.

Davis, C., Brown, B., and Bullock, R.: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas, Monthly Weather Review, 134, 1772–1784, https://doi.org/10.1175/MWR3145.1, 2006a.

Davis, C., Brown, B., and Bullock, R.: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems, Monthly Weather Review, 134, 1785–1795, https://doi.org/10.1175/MWR3146.1, 2006b.

Davis, C. A., Brown, B. G., Bullock, R., and Halley-Gotway, J.: The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program, Weather and Forecasting, 24, 1252–1267, https://doi.org/10.1175/2009WAF2222241.1, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delrieu, G., Nicol, J., Yates, E., Kirstetter, P.-E., Creutin, J.-D., Anquetin, S., Obled, C., Saulnier, G.-M., Ducrocq, V., Gaume, E., Payrastre, O., Andrieu, H., Ayral, P.-A., Bouvier, C., Neppel, L., Livet, M., Lang, M., du-Châtelet, J. P., Walpersdorf, A., and Wobrock, W.: The Catastrophic Flash-Flood Event of 8–9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes–Vivarais Mediterranean Hydrometeorological Observatory, Journal of Hydrometeorology, 6, 34–52, https://doi.org/10.1175/JHM-400.1, 2005.

Descamps, L., Labadie, C., and Bazile, E.: Representing model uncertainty using the multiparametrization method, in: Workshop on Representing Model Uncertainty and Error in Numerical Weather and Climate Prediction Models, 20-24 June 2011, pp. 175–182, ECMWF, ECMWF, Shinfield Park, Reading, https://www.ecmwf.int/node/9015, 2011.

Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., and Cébron, P.: PEARP, the Météo-France Short-Range Ensemble Prediction System, Quarterly Journal of the Royal Meteorological Society, 141, 1671–1685, https://doi.org/10.1002/qj.2469, 2015.

Du, J., Mullen, S. L., and Sanders, F.: Short-Range Ensemble Forecasting of Quantitative Precipitation, Monthly Weather Review, 125, 2427–2459, https://doi.org/10.1175/1520-0493(1997)125<2427:SREFOQ>2.0.CO;2, 1997.

Ducrocq, V., Ricard, D., Lafore, J.-P., and Orain, F.: Storm-Scale Numerical Rainfall Prediction for Five Precipitating Events over France: On the Importance of the Initial Humidity Field, Weather and Forecasting, 17, 1236–1256, https://doi.org/10.1175/1520-0434(2002)017<1236:SSNRPF>2.0.CO;2, 2002.

Ducrocq, V., Aullo, G., and Santurette, P.: The extreme flash flood case of November 1999 over Southern France, La Météorologie, 42, 18–27, 2003.

Ducrocq, V., Nuissier, O., Ricard, D., Lebeaupin, C., and Thouvenin, T.: A Numerical Study of Three Catastrophic Precipitating Events over Southern France. II: Mesoscale Triggering and Stationarity Factors, Quarterly Journal of the Royal Meteorological Society, 134, 131–145, https://doi.org/10.1002/qj.199, 2008.

Ebert, E. E. and McBride, J. L.: Verification of Precipitation in Weather Systems: Determination of Systematic Errors, Journal of Hydrology, 239, 179–202, https://doi.org/10.1016/S0022-1694(00)00343-7, 2000.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Natural Hazards and Earth System Sciences Discussions

Erdin, R., Frei, C., and Künsch, H. R.: Data Transformation and Uncertainty in Geostatistical Combination of Radar and Rain Gauges, Journal of Hydrometeorology, 13, 1332–1346, https://doi.org/10.1175/JHM-D-11-096.1, 2012.

Frei, C. and Schär, C.: A Precipitation Climatology of the Alps from High-Resolution Rain-Gauge Observations, International Journal of Climatology, 18, 873–900, https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9, 1998.

G. Gregoire, T., Lin, Q. F., Boudreau, J., and Nelson, R.: Regression Estimation Following the Square-Root Transformation of the Response, Forest Science, 54, 597–606, 2008.

Goovaerts, P. et al.: Geostatistics for natural resources evaluation, Oxford University Press on Demand, 1997.

Hamill, T. M.: Verification of TIGGE Multimodel and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States, Monthly Weather Review, 140, 2232–2252, https://doi.org/10.1175/MWR-D-11-00220.1, 2012.

Hamill, T. M. and Whitaker, J. S.: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application, Monthly Weather Review, 134, 3209–3229, https://doi.org/10.1175/MWR3237.1, 2006.

Hamill, T. M., Hagedorn, R., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation, Monthly Weather Review, 136, 2620–2632, https://doi.org/10.1175/2007MWR2411.1, 2008.

Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, Monthly Weather Review, 126, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., and Mitchell, H. L.: A System Simulation Approach to Ensemble Prediction, Monthly Weather Review, 124, 1225–1242, https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2, 1996.

Kai, T., Zhong-Wei, Y., and Yi, W.: A Spatial Cluster Analysis of Heavy Rains in China, Atmospheric and Oceanic Science Letters, 4, 36–40, https://doi.org/10.1080/16742834.2011.11446897, 2011.

Kain, J. S. and Fritsch, J. M.: Convective Parameterization for Mesoscale Models: The Kain-Fritsch Scheme, in: The Representation of Cumulus Convection in Numerical Models, edited by Emanuel, K. A. and Raymond, D. J., Meteorological Monographs, pp. 165–170, American Meteorological Society, Boston, MA, https://doi.org/10.1007/978-1-935704-13-3_16, 1993.

Lack, S. A., Limpert, G. L., and Fox, N. I.: An Object-Oriented Multiscale Verification Scheme, Weather and Forecasting, 25, 79–92, https://doi.org/10.1175/2009WAF2222245.1, 2010.

Lalaurette, F.: Early detection of abnormal weather conditions using a probabilistic extreme forecast index, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 129, 3037–3057, 2003.

Lin, Y.-L., Chiao, S., Wang, T.-A., Kaplan, M. L., and Weglarz, R. P.: Some Common Ingredients for Heavy Orographic Rainfall, Weather and Forecasting, 16, 633–660, https://doi.org/10.1175/1520-0434(2001)016<0633:SCIFHO>2.0.CO;2, 2001.

Little, M. A., Rodda, H. J. E., and McSharry, P. E.: Bayesian Objective Classification of Extreme UK Daily Rainfall for Flood Risk Applications, Hydrology and Earth System Sciences Discussions, 5, 3033–3060, https://doi.org/https://doi.org/10.5194/hessd-5-3033-2008, 2008.

Louis, J.-F.: A Parametric Model of Vertical Eddy Fluxes in the Atmosphere, Boundary-Layer Meteorology, 17, 187–202, https://doi.org/10.1007/BF00117978, 1979.

Ly, S., Charles, C., and Degré, A.: Geostatistical Interpolation of Daily Rainfall at Catchment Scale: The Use of Several Variogram Models in the Ourthe and Ambleve Catchments, Belgium, Hydrol. Earth Syst. Sci., 15, 2259–2274, https://doi.org/10.5194/hess-15-2259-2011, 2011.

Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: a review, BASE, 2013.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does Increasing Horizontal Resolution Produce More Skillful Forecasts?, Bulletin of the American Meteorological Society, 83, 407–430, https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2, 2002.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, https://hal.archives-ouvertes.fr/hal-00968042, 2013.

Mills, G. F.: Principal Component Analysis of Precipitation and Rainfall Regionalization in Spain, Theoretical and Applied Climatology, 50, 169–183, https://doi.org/10.1007/BF00866115, 1995.

Mittermaier, M., North, R., Semple, A., and Bullock, R.: Feature-Based Diagnostic Evaluation of Global NWP Forecasts, Monthly Weather Review, 144, 3871–3893, https://doi.org/10.1175/MWR-D-15-0167.1, 2015.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF Ensemble Prediction System: Methodology and Validation, Quarterly Journal of the Royal Meteorological Society, 122, 73–119, https://doi.org/10.1002/qj.49712252905, 1996.

Morin, G., Fortin, J.-P., Sochanska, W., Lardeau, J.-P., and Charbonneu, R.: Use of Principal Component Analysis to Identify Homogeneous Precipitation Stations for Optimal Interpolation, Water Resources Research, 15, 1841–1850, https://doi.org/10.1029/WR015i006p01841, 1979.

Nachamkin, J. E.: Application of the Composite Method to the Spatial Forecast Verification Methods Intercomparison Dataset, Weather and Forecasting, 24, 1390–1400, https://doi.org/10.1175/2009WAF2222225.1, 2009.

Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C., and Anquetin, S.: A Numerical Study of Three Catastrophic Precipitating Events over Southern France. I: Numerical Framework and Synoptic Ingredients, Quarterly Journal of the Royal Meteorological Society, 134, 111–130, https://doi.org/10.1002/qj.200, 2008.

Nuissier, O., Joly, B., Joly, A., Ducrocq, V., and Arbogast, P.: A Statistical Downscaling to Identify the Large-Scale Circulation Patterns Associated with Heavy Precipitation Events over Southern France, Quarterly Journal of the Royal Meteorological Society, 137, 1812–1827, https://doi.org/10.1002/qj.866, 2011.

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., and Weisheimer, A.: Stochastic parametrization and model uncertainty, ECMWF Technical Memorandum, p. 42, https://doi.org/10.21957/ps8gbwbdv, https://www.ecmwf.int/node/11577, 2009.

Peñarrocha, D., Estrela, M. J., and Millán, M.: Classification of Daily Rainfall Patterns in a Mediterranean Area with Extreme Intensity Levels: The Valencia Region, International Journal of Climatology, 22, 677–695, https://doi.org/10.1002/joc.747, 2002.

Pergaud, J., Masson, V., Malardel, S., and Couvreux, F.: A Parameterization of Dry Thermals and Shallow Cumuli for Mesoscale Numerical Weather Prediction, Boundary-Layer Meteorology, 132, 83, https://doi.org/10.1007/s10546-009-9388-0, 2009.

Petroliagis, T., Buizza, R., Lanzinger, A., and Palmer, T. N.: Potential Use of the ECMWF Ensemble Prediction System in Cases of Extreme Weather Events, Meteorological Applications, 4, 69–84, https://doi.org/10.1017/S1350482797000297, 1997.

Piriou, J.-M., Redelsperger, J.-L., Geleyn, J.-F., Lafore, J.-P., and Guichard, F.: An Approach for Convective Parameterization with Memory: Separating Microphysics and Transport in Grid-Scale Equations, Journal of the Atmospheric Sciences, 64, 4127–4139, https://doi.org/10.1175/2007JAS2144.1, 2007.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Ricard, D., Ducrocq, V., and Auger, L.: A Climatology of the Mesoscale Environment Associated with Heavily Precipitating Events over a Northwestern Mediterranean Area, Journal of Applied Meteorology and Climatology, 51, 468–488, https://doi.org/10.1175/JAMC-D-11-017.1, 2011.

Romero, R., Ramis, C., and Guijarro, J. A.: Daily Rainfall Patterns in the Spanish Mediterranean Area: An Objective Classification, International Journal of Climatology, 19, 95–112, https://doi.org/10.1002/(SICI)1097-0088(199901)19:1<95::AID-JOC344>3.0.CO;2-S, 1999.

Rossa, A., Nurmi, P., and Ebert, E.: Overview of Methods for the Verification of Quantitative Precipitation Forecasts, in: Precipitation: Advances in Measurement, Estimation and Prediction, edited by Michaelides, S., pp. 419–452, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-77655-0_16, 2008.

Schär, C., Ban, N., Fischer, E. M., Rajczak, J., Schmidli, J., Frei, C., Giorgi, F., Karl, T. R., Kendon, E. J., Tank, A. M. G. K., O'Gorman, P. A., Sillmann, J., Zhang, X., and Zwiers, F. W.: Percentile Indices for Assessing Changes in Heavy Precipitation Events, Climatic Change, 137, 201–216, https://doi.org/10.1007/s10584-016-1669-2, 2016.

Scholz, F. W. and Stephens, M. A.: K-Sample Anderson-Darling Tests, Journal of the American Statistical Association, 82, 918–924, https://doi.org/10.2307/2288805, 1987.

Schumacher, R. S. and Davis, C. A.: Ensemble-Based Forecast Uncertainty Analysis of Diverse Heavy Rainfall Events, Weather and Forecasting, 25, 1103–1122, https://doi.org/10.1175/2010WAF2222378.1, 2010.

Sénési, S., Bougeault, P., Chèze, J.-L., Cosentino, P., and Thepenier, R.-M.: The Vaison-La-Romaine Flash Flood: Mesoscale Analysis and Predictability Issues, Weather and Forecasting, 11, 417–442, https://doi.org/10.1175/1520-0434(1996)011<0417:TVLRFF>2.0.CO;2, 1996.

Shepard, D.: A Two-Dimensional Interpolation Function for Irregularly-Spaced Data, in: Proceedings of the 1968 23rd ACM National Conference, ACM '68, pp. 517–524, ACM, New York, NY, USA, https://doi.org/10.1145/800186.810616, 1968.

Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S., and Rogers, E.: Using Ensembles for Short-Range Forecasting, Monthly Weather Review, 127, 433–446, https://doi.org/10.1175/1520-0493(1999)127<0433:UEFSRF>2.0.CO;2, 1999.

Teo, C.-K., Koh, T.-Y., Chun-Fung Lo, J., and Chandra Bhatt, B.: Principal Component Analysis of Observed and Modeled Diurnal Rainfall in the Maritime Continent, Journal of Climate, 24, 4662–4675, https://doi.org/10.1175/2011JCLI4047.1, 2011.

Toth, Z. and Kalnay, E.: Ensemble Forecasting at NMC: The Generation of Perturbations, Bulletin of the American Meteorological Society, 74, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2, 1993.

Toth, Z. and Kalnay, E.: Ensemble Forecasting at NCEP and the Breeding Method, Monthly Weather Review, 125, 3297–3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2, 1997.

Vié, B., Nuissier, O., and Ducrocq, V.: Cloud-Resolving Ensemble Simulations of Mediterranean Heavy Precipitating Events: Uncertainty on Initial Conditions and Lateral Boundary Conditions, Monthly Weather Review, 139, 403–423, https://doi.org/10.1175/2010MWR3487.1, 2010.

Walser, A. and Schär, C.: Convection-Resolving Precipitation Forecasting and Its Predictability in Alpine River Catchments, Journal of Hydrology, 288, 57–73, https://doi.org/10.1016/j.jhydrol.2003.11.035, 2004.

Walser, A., Lüthi, D., and Schär, C.: Predictability of Precipitation in a Cloud-Resolving Model, Monthly Weather Review, 132, 560–577, https://doi.org/10.1175/1520-0493(2004)132<0560:POPIAC>2.0.CO;2, 2004.

Wernli, H., Paulat, M., Hagen, M., and Frei, C.: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, Monthly Weather Review, 136, 4470–4487, https://doi.org/10.1175/2008MWR2415.1, 2008.

https://doi.org/10.5194/nhess-2019-302
Preprint. Discussion started: 25 September 2019

Wernli, H., Hofmann, C., and Zimmer, M.: Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique, Weather and Forecasting, 24, 1472–1484, https://doi.org/10.1175/2009WAF2222271.1, 2009.

World Meteorological Organization, ed.: Guidelines on Ensemble Prediction Systems and Forecasting, 1091, WMO, 2012.

World Meteorological Organization, ed.: Guidelines on the definition and monitoring of extreme weather and climate events, Task Team on definitions of Extreme Weather and Climate Events (TT-DEWCE), 2016.