https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Discussions

Sensitivity of the Weather Research and Forecasting model (WRF) to downscaling extreme events over Northern Tunisia

Saoussen Dhib[1], Víctor Homar[2], Zoubeida Bargaoui[1], Mariadelmar Vich[2]

[1]Laboratory of Hydraulic and Environmental Modeling (LMHE), Université de Tunis El-Manar (UTM), Ecole Nationale d'ingénieurs de Tunis (ENIT), Tunis, 1002, Tunisia.
[2]Meteorology Group, Physics Department, Universitat de les Illes Balears, Palma de Mallorca, 7003, Spain.

Correspondence to: Saoussen Dhib (dhib_saoussen@hotmail.fr)

Abstract

Rainfall is one of the most important variables for water and flood management. We investigate the capacity of the Weather Research and Forecasting model (WRF) to dynamically downscale the ECMWF Re-Analysis data for Northern Tunisia. This study aims to examine the sensitivity of WRF rainfall estimates to different Planetary Boundary Layer (PBL) and Cumulus Physics (Cu) schemes. The verification scheme consists of three statistical criteria (Root Mean Square Error (RMSE), Pearson correlation, and the ratio bias coefficient). Moreover, the FSS coefficient (fraction skill score) and the quality coefficient SAL (structure amplitude latitude) are calculated. The database is composed of four heavy events covering an average of 318 rainfall stations. We mean by heavy event, each event occurred a rainfall of more than 50 mm per observed day at least in one rainfall station. The sensitivity study showed that there is not a best common combination scheme (PBL and Cu) for all the events. The average of the best 10 combinations for each event is adopted to get the ensemble map. We conclude that some schemes are sensitive and others less sensitive. The best three performing schemes for PBL and Cu parametrizations are selected for future rainfall estimation by WRF over Northern Tunisia.

Keywords: WRF-QPF, Extreme-rainfall, Sensitivity, Northern-Tunisia, Validation

1 Introduction

The occurrence of heavy rainfall makes the economy of the Tunisian country weaker. In September 2011, Zaghouan region and the lower valley of the Medjerda experienced floods. Three people dead. Huge losses occurred in the agricultural sector estimated at about 30 million Tunisian dinars and road infrastructure about (40% of the actual PIB) was subject to severe damages (Fehri, 2014). Rainfall forecasting and alert may help to surmount a part of floods impacts. The MSG MPE (Meteosat Second Generation Multi-sensor Precipitation Estimate) was used to evaluate rainfall estimation in comparison to interpolated in-situ data. Weak performance was found in detecting rainfall amounts during extreme events with daily rainfall more than 50 mm per day, in Northern Tunisia (Dhib et al., 2017). Even with two proposed corrections based on in-situ data the results were found still insufficient. Here, we seek to base on other sources of rainfall estimation. An alternative source of global rainfall information is short-range forecasts from numerical weather prediction (NWP) models. NWP models use satellite and in situ observations of atmospheric temperature and moisture as input to define the initial conditions to run models of atmospheric motion using appropriate physical parameterizations to predict rainfall (Berrisford et al.,

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

2009). WRF is one such model that has been used, among other things, to downscale ECMWF 40-year reanalysis data (Uppala et al., 2005). WRF is selected here because it is performing and widely used by the national meteorological institute (INM) (Nmiri, 2014).

However, regional climate models are sensitive to the different model physics parameterizations options. Additionally, the behavior of physics may vary depending on the location of the domain due to different climatic regimes. That is why we should study the sensitivity of WRF over our study area which is a very crucial step. Crétat et al. (2011) ran the WRF model, literally forced by ERA40 reanalysis. Twenty-seven experiments configured with three schemes of cumulus (Cu), a planetary boundary layer (PBL) and microphysics (MP) were tested at 35 km horizontal resolution to quantify the seasonal biases of rainfall. It was found that rain rates were predominantly sensitive to Cu schemes and much less to PBL and MP schemes. They found that WRF simulates accurately seasonal gradients of rainfall also the seasonal large-scale rainfall patterns. However, they noticed strong seasonal biases fluctuation from an experiment to another. We conclude from this study of (Crétat et al., 2011) that without testing numerous physical parameterizations one couldn't find satisfactory rainfall estimations.

Another sensitivity study was achieved by Evans et al. (2011), over the south of Australia, to evaluate the ability of a 36 member multi-physics WRF ensembles to reproduce four East Coast Low events. Two PBL schemes, two Cu schemes, three microphysics (Mp) schemes, and three radiation (Ra) scheme combinations of shortwave and longwave schemes respectively were used to create these 36 members. A weak sensitivity appears for weak weather systems in comparison with extreme events. In agreement with previous WRF parameterizations studies (Jankov et al., 2005; Flaounas et al, 2011), not a single preferred member is the best for all cases and all metrics.

To study WRF sensitivity over Tunisia, this paper contains four other sections organized as follows: Section 2 describes the in situ data and used WRF parametrizations, Section 3 provides the sensitivity study methodology, Section 4 represents the results, and the last section summarizes the conclusions and perspectives.

2 Data and methods

2.1 In situ data

Northern Tunisia represents the study area (Fig.1). It's hydrological division is into three parts: the Medjerda river watershed (W 5), the Meliane watershed (W 4) and the watershed composed by north coastal basins watershed (W 3). The Northern Tunisia covers an area of about 36000 km² and a population of about 6 million inhabitants. It is limited north and east by the Mediterranean Sea, south by the mountains of the Atlas and west by Algeria. The rain gauges are presented in Fig. 1 with WRF grid and the Radar Topography Mission (SRTM) map as Background.

[Figure 1]

The spatial interpolation of the in situ precipitation data was achieved using an inverse distance weighted (moving average) method (Dhib et al., 2017). The database is composed by an average of 318 rain gauges. Heavy events are defined as those daily events exceeding 50 mm/day for at least one station. A total of 77 heavy rainfall events period (Fig.2a) is result from this selection criterion during the study period which is from January 2007 to August

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

2009. 35 events were recorded during the dry period (6 months from May to October) and 42 events during the wet season (from November to April).

To undertake the present study, four days were selected among 11 important events (at least 2 stations with 50 mm/day) those that were not detected by MSGMPE (Heinemann et al.,2002).

2.2 Case studies

Figure 2a shows the spatial average of the 76 heavy events (all the colors) in comparison with the standard deviation. The important undetected events (at least 2 heavy stations in in situ data registered more than 50 mm/day) using MSG-MPE rainfall estimation are colored in Black and red. The selected events for the WRF sensitivity are represented in red color (Fig.2a). The gauges rain variability of the four case studies are presented in Fig.2b. We chose from the 76 heavy events two remarkable events 12/01/2009 with the highest spatial average rain (43.8 mm/day) and 13/09/2007 which registered the highest standard deviation (79.2). In the other hand, we chose two ordinary events. The first one is the 13/10/2007 where both the spatial average (19.3 mm/day) and the standard deviation (24.2) are near the average of all the events. The fourth event has the second highest spatial average (28.2 mm/day). The four case studies have different rainfall localization. For example we see in (Fig.2c) that the rainfall cover almost all the study area on 12/01/2009. Contrary, the three other events we could see different localization of the heaviest rain.

[Figure 2]

2.3 Interpolation: The spatial interpolation of the in situ precipitation data was done using an inverse distance weighted (moving average) method. To optimize the weight (W) of the inverse distance (IDW) interpolation method, we did a cross validation for the studied events. Fig.3(a) illustrates the correlation coefficients and the RMSE versus the Power (P) of the four events Fig.3(b).

Figure 3 highlights the importance of the cross-validation. We notice that not in all the cases the best correlation coefficient corresponds to the lowest RMSE. For example, we see the high variation of the correlation coefficient and the RMSE of the 08/03/2007 event. For a P value of 0.1 and 0.7, we find the correspondent correlation coefficient fluctuates from 0.3 to 0.48 respectively while the RMSE varies from 23 to 22 mm/day without attending the lowest value (21 mm/day). In such a case, we take into consideration the P value corresponding to the best RMSE which is 1.2.

[Figure 3]

2.3. The WRF model and the used parametrizations

WRF is a numerical weather prediction (NWP) and atmospheric simulation model. It is a mesoscale forecast and data assimilation system (Skamarock et al., 2008). WRF's boundary and initial conditions covering the study area during the studied period 2007-2009 is the latest ERA-Interim global atmospheric reanalysis product of the ECMWF (European Centre for Medium-Range Weather Forecasts) from 1 January 1989 (Berrisford et al., 2009). The variables are precipitable vapor, brightness temperatures, atmospheric motion vectors, atmospheric

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences Open Access
Discussions

refraction, scatterometer wind data, and ozone retrievals. The majority of these variables are originated from satellite-borne sensors. Some are improved by in situ measurements such as wind (u/v), upper air temperatures (T), and specific humidity (q). The ERA-Interim has a horizontal resolution about 79 km spacing on a reduced Gaussian grid. In ERA-Interim the vertical resolution is represented by 60 model layers with 0.1 hPa at the top of the atmosphere.

There are several WRF versions. This study employs WRF with the version 3.4 of the Advanced Research WRF core. Figure 4 shows coverages of WRF's domain employed in this study. The one-way nesting strategy is used. The outer domain has grid points with 30 km resolution. The inner domain has 10 km resolution and covers latitudes of 30°N–42°N and longitudes of 02°E–21°E. The ERA-Interim global atmospheric reanalysis dataset (ERA) is dynamically downscaled using WRF to obtain downscaled reanalysis at 10 km resolution. These outputs from the inner domain at 10 km resolution are employed in this study.

[Figure 4]

2.4. Parameters schemes characteristics

-Cumulus parameterization schemes:

There are two main types of convection: deep convection and shallow convection, which refer to convective elements development. Associated with strong ascents and precipitated quantities, deep convection warms (by the release of latent heat) and dries out (by condensation and precipitation of water vapor) the atmosphere, which is not the case for shallow convection (Dorrestijn, 2013). Convection patterns determine the vertical fluxes associated with sub-surface ancillaries and subside, compensatory motions outside the clouds, and provide vertical profiles of heat and moisture. The used cumulus (Cu) schemes in this work are briefly described in Table 1 (Skamarock et al., 2008)

-PBL parameterization schemes

PBL schemes are 1D schemes assuming a clear difference between subgrid vortices and large-scale vortices. When PBL scheme triggered, explicit vertical scattering is disabled with the assumption that the PBL scheme will handle this process. Controlling the vertical flow profiles, PBL schemes provide atmospheric tendencies of moisture, temperature, clouds, and horizontal momentum in the entire atmospheric column (Skamarock et al., 2005). Table 1 described the PBL and Cu schemes adopted here.

[Table 1]

3. Methodology

For the rest of the work we will use the four selected days out of the eleven undetected events by MSGMPE. Furthermore, based on quantile quantile comparison of the three different parameters (PBL, Cu, Mp) schemes, we will choose which parameters will be used for the sensitivity study.

3.1 Sensitivity parameters selection

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

Firstly, default parameters are used in the evaluation of the 11 chosen events (PBL (2), Cu (5)). The first run of WRF precipitation estimate was achieved using the default parameters (PBL scheme 2, Cu scheme 5, Mp scheme 6). The 11 tested events were detected rainy by WRF. Further, the sensitivity study is limited to a subsample of four events out of 11 as a first test. These four events are selected because they present different types of events where we find very high rainfall amounts covering the whole study area (12/01/2009), a high rainfall in vast areas (13/09/2008), weak rainfall in a considered area (26/03/2008), and weak rainfall in a very limited surface (23/09/2007).

We assume the three most commonly adopted parameters (PBL, Cumulus (Cu) and microphysics (Mp)) to analyze the sensitivity of WRF over the study area. Figure 5 illustrate the quantile-quantile comparison for different schemes of the three selected parameters for the extreme event of 12/01/2009.

[Figure 5]

For the PBL schemes simulation, the Cu scheme was fixed to 2 and Mp scheme to 6 (Fig.5a).

We notice that for the PBL parameter (Fig.5a), the rainfall estimation differs from one scheme to another. It is concluded that there is some WRF sensitivity for this parameter over the study area. To illustrate the sensitivity of the Cu schemes the PBL scheme was fixed to 9 and Mp scheme was fixed to 6 (Fig.5b). The quantile quantile comparison of the different Cu schemes between the WRF and the ground data shows the high difference in the estimation foremost of high rainfall (more than 70 mm/day). For the Mp schemes, the PBL parameter was fixed to 9 and Cu parameter to 2.

Based on the quantile-quantile comparison, the PBL and Cu parameters look more sensitive than MP parameter (Fig.5c) which shows a sensitivity only for high values (more than 70 mm/day). Then, in this work, MP is considered not sensitive and maybe in future work we include it in the sensitivity study.

The four evaluated events for the sensitivity study are 08/03/2007, 13/10/2007, 13/09/2008, and the 12/01/2009. The choice of these events is based on the incapability of MSGMPE to detect them. Also, we chose them because of the difference in the type of rain (scattered or very localized in space, in topographic area) and for the location difference of the extreme values in the ground.

A threshold of 0.1 mm is used in SAL and FSS verification to distinguish between rainy and no rainy pixels. In case of undetected events, they will be deleted in the SAL diagram. The number of these non-represented cases in SAL will indicate the poor forecasts. This will appear foremost for the high thresholds (30 and 50 mm/day).

3.2 Evaluation metrics of the sensitivity study

For each studied day, 99 combinations of Cu (11 schemes) and PBL (9 schemes) are simulated. The observed and forecast precipitation fields are compared. R represents the precipitation field. Observed rain and simulated precipitation are symbolized R_{obs} and R_{mod} respectively. We consider N grid in both the in situ data and WRF data. The sensitivity study verification is performed to compare the rainfall estimation by the different combinations and the in-situ data using classical scores (Zacharov et al., 2013) such as Pearson correlation coefficient , ratio bias coefficient, RMSE, SAL criterion and FSS. The ratio bias coefficient is the division of the

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

spatial Averages of WRF and on the ground. Pearson correlation coefficient is used to find how strong a relationship is between data. The formula return a value between -1 and 1, where: 1 indicates a strong positive relationship, 1 indicates a strong negative relationship, a result of zero indicates no relationship at all. RMSE is the standard deviation of the residuals which show how concentrated the data is around the line of best fit.

The use of the SAL verification method (Wernli et al., 2008) requires first the identification of individual objects. An object is contiguous rain area respecting a specific rain threshold. For here we use the simple approach introduced by Davis et al. (2006), where a threshold R^* is selected Eq.(1) to detect a coherent objects encircled by the threshold contour.

$$R^* = f\,R^{max} \quad (1)$$

R_{max} designates the maximum rainfall amount in the study area and f is a factor equal to 1/15 was selected by the fact that for most considered cases, this contour distinguishes rainfall features that correspond to easily identifiable objects.

Three components of SAL are considered going from the most complex from A to L and finally, S. The amplitude component A relates the normalized variance of the spatial average of R_{mod} and R_{obs} Eq.(2).

$$A=\frac{D(R_{mod})-D(R_{obs})}{0.5(D(R_{mod})+D(R_{obs}))} \quad (2)$$

where D(R) represents the domain average of the precipitation R.

A component varies from −2 to +2, and the impeccable forecast is indicated by A = 0 (Fig.6). A =1 designates that the model overestimates the rainfall spatial average by 3; A=-1 indicates an underestimation by a factor of 3; A = 0.4 and 0.67 means an overestimation by 1.5 and 2 respectively.

The SAL location component L is the sum of two terms L_1 Eq. (3) and L_2 Eq. (4) , L_1 and L_2 vary from 0 to 1. L1 measures the normalized distance among the mass centers of the observed and the forecast precipitation fields Eq. (3):

$$L_1=\frac{|X(R_{mod})-X(R_{obs})|}{d} \quad (3)$$

The variable d is the largest distance among two points in the specified domain.

While $X(R_{mod})$ and $X(R_{obs})$ is the mass center of the observed and modeled precipitation fields respectively.

L1 = 0 (Eq.3) designates that the mass centers of the observed $X(R_{obs})$ and the modeled precipitation $X(R_{mod})$ are the same. The component L_2 Eq. (4) indicates the mean distance between the rainy area mass center and the singular rainfall objects (Wernli et al., 2008).

$$L_2=2\left[\frac{r(R_{mod})-r(R_{obs})}{d}\right] \quad (4)$$

When the number of objects surpasses 1 in the observed or in the predicted rainfall (or both), L_1 and L_2 differs from zero.

S component allows for a comparison between the volumes of the normalized precipitation objects. It is mainly informative about the size and shape of rainy objects. For each object Rn, a Vn volume Eq. (5) is calculated based on the sum of all grid-point R(i, j):

$$V_n = \sum_{(ij)} \frac{R(i,j)}{R_n^{max}} \quad (5)$$

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

where R_n^{max} designates the maximum rainfall inside the object field. Vn designates the volume for each object in the observed and forecasted datasets. Then, for each dataset, the V value is calculated as the weighted average of the Vn over all objects. In an analogue way to A component, the S represents the normalized difference indicated in Eq. (6).

$$S=\frac{V(R_{mod})-V(R_{obs})}{0.5(V(R_{mod})+V(R_{obs}))} \quad (6)$$

The values of S are within [-2, +2]. When S is more than 0 that means the predicted rainfall objects are too outsized and/or too smooth (Fig.6), while when it is less than 0 that means that the predicted objects of rainfall are too small and/or too peaky.

[Figure 6]

The FSS (Roberts and Lean, 2008) is a neighborhood verification method. It compares the occurrence of precipitation exceeding a specified threshold in the in situ and forecasts datasets. The FSS varies from 0 to 1. For a perfect forecast, FSS gets 1. While for a total mismatch by the forecast or some surpassing values are forecasted but does not recorded the FSS gets 0. The term elementary area (EA) is used to identify a specific spatial window. Moreover, as the EA size rises, the score will progressively approach 1 and the forecast bias decreases. The FSS is defined by the Eq. (7) (Roberts and Lean, 2008):

$$FSS=1-\frac{\frac{1}{N}\sum_{j=1}^{N}(o_j-f_j)^2}{\frac{1}{N}[\sum_{j=1}^{N}{o_j}^2+\sum_{j=1}^{N}{f_j}^2]} \quad (7)$$

where o_j and f_j is the fractional area of an EA centered in the grid j by a precipitation higher than a specific threshold value respectively for observation and forecast, and N is the total of grids in the verification area. FSS score was used with a threshold of 0.1 mm.

3.3. The methodology of the sensitivity study

Some treatment of the metrics was necessary prior to rank the ensemble members:

(i) the R (Pearson), the ratio bias and FSS scores were inverted so that smaller values (closer to zero) represent better simulations,

(ii) centered RMSE is standardized by its maxima.

Thus, all metrics are within a scale of 0–1 and are averaged. The ensemble member with the smallest metric sum corresponds to the best performing simulation.
After ranking the 99 combinations, the 20 best combinations are selected.

Then, we perform a new ranking of these 20 combinations based on the analysis of FSS, SAL, and the metrics sum to identify the finest 10 combinations. Finally, we calculate an ensemble map which is the average of the finest 10 combinations. Figure 7 depicts all the processing and sensitivity steps.

[Figure 7]

4 Results and discussion

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Discussions

a/ The evaluation of the 08/03/2007:

Figure 8 (a) shows the SAL diagram which highlights the skills of the different combination schemes in different thresholds. S component is the abscissa and A component is the ordinates. The color of the dots represents the L component (see the scale on the right).

Excellent forecasts (the three components are near zero) are found in red color in the center of the diagram. S and A components were good enough for the thresholds 0.1, 5, 10, 20 mm except for 7 overestimated combinations. L component tends to be a bit higher for thresholds 5, 10, and 20 mm in comparison with 0.1 mm threshold. The WRF model aims to estimate for some combinations larger objects for the rain exceeding 50 mm (S near 2) and sometimes peaked objects (S near -1).

Fig.8 (b) represents the FSS components of the different combinations for different thresholds (0.1, 20, 30, and 50 mm) best 20 combinations obtained by the metrics sum.

[Figure 8]

The FSS coefficient in Fig.8 (b) helped us to identify the best 10 combinations (Table 2).

[Table 2]

b/ The evaluation of the 13/10/2007 event:

Figure 9 illustrates the verification of all the assumed schemes for the 13/10/2007 event for different thresholds. The crossed lines represent the medians of S and A (Fig.9a).
The colored box symbolizes the percentiles 25th and 75th of the components S and A. The box's color indicates the median of L. The first quadrant illustrates the forecasts which overestimate both the amplitude and the structural components of SAL. The third quadrant represent the underestimation of both components.

We notice that for the threshold 0.1 mm the L component is more or less similar which is due to the presence of only one object. The threshold 5 increases the L component which is explained by the apparition of other objects. A and S components become larger showing respectively higher overestimation and larger estimated objects. For the thresholds 10 and 20 mm, SAL components are more or less similar to only larger estimated objects by the 20 mm threshold. For the threshold of 30 mm, the underestimation accentuated. Peaked objects appear clearly at the threshold of 50 mm with an important underestimation.

[Figure 9]

After achieving the ranging of the schemes based on the sum metrics methodology, we select the best 20 schemes to evaluate them using the FSS and the SAL verification method (Fig.9).

FSS helps us to select the best 10 combinations (Fig.9b) that are mentioned in Table 3. The schemes combinations are ranked from the best to the worst based on the Metrics sum coefficient.

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

[Table 3]

c/ The evaluation of the 12/01/2009 event:

From the threshold of 10 mm, S component becomes larger (median 0.7) showing large estimated objects (Fig.10).

[Figure 10]

The various FSS thresholds clarify the skills of combinations (Fig.10b). After calculating the sum of metrics, we selected the 10 best combinations (Table 4).

[Table 4]

d/ The evaluation of the 13/09/2008 event:

For all the thresholds (Fig.11) L component varies from 0 to 0.6 which indicates the presence of many objects. From the threshold 20 mm, S components become larger showing high (S near 2) and picked (S near -2) estimated objects. For the thresholds 50 mm, we notice that the number of combinations which detect this threshold decrease notably. These SAL thresholds help us to eliminate some weak combinations.

[Figure 11]

To find the best 10 combinations we represented the 20 best combinations selected previously by the metrics sum. Fig.9b helped us to identify only 9 best combinations. We select the 10th combination based on the metrics sum (Cu5Pb8) which was not so representative of FSS (Table 5).

[Table 5]

Figure 12 shows the ensembles maps of the four studied events. We notice that the rainfall gradient is similar between the ensembles and the interpolated in-situ maps. The correlation coefficient is also satisfying: 0.72, 0.58, 0.48 and 0.57 for respectively 08/03/2007, 13/10/2007, 13/09/2008 and 12/01/2009.

[Figure 12]

Figure 13 shows the sensitivity of the four events in term of PBL and Cu. We notice that there are some schemes which are sensitive and others which are less sensitive. The best performing schemes (less sensitive) are PBL 5, 7, and 99.

[Figure 13]

The best performing schemes (less sensitive) for Cumulus parametrization are Cu 1, 4, and 99.

3. Conclusion:

WRF is sensitive to the different model physics parameterizations options. Additionally, the behavior of physics may vary depending on the location of the domain due to different climatic regimes. The current study of the extreme events using climate model WRF underlines the importance of the evaluation of such estimation rainfall data before using it as a truth data mainly for daily scale for many reasons. One of the main reasons is the good performance of WRF model in the estimation of the monthly and yearly rainfall. For example in a previous evaluation of WRF over Tunisia (Fathalli et al., 2018), noticed a satisfying estimation of rainfall using this model for the monthly and yearly scale. For daily scale, we need always to improve the rainfall estimation for WRF.

We used for the four selected extreme events 99 combinations between the different Cumulus parametrization schemes and Planetary Boundary layer schemes. The metrics sum is adopted to rank the 99 combinations and to select the 20 best combinations for each event. Then, based on the analysis of FSS, SAL we performed a new ranking of these 20 combinations to identify the finest 10 combinations. Finally, we calculate the average of these finest 10 combinations to obtain an ensemble map for each event.

The results showed a good detection of all the studied events using the WRF model default parameters. Also, we notice that the use of a single verification technique could lead to a shortcoming of information about the forecast. The use of several verification techniques (statistical coefficients, SAL and FSS) is extremely helpful to choose the best combination for each event. The sensitivity study helped us to identify the sensitive parameters of our study area which will facilitate the work with WRF in the future. The ensemble map method gave a very satisfying results. Then, we suggest for Tunisian WRF users as a first result to use this schemes Cu 1, 4, and 99 and PBL 5, 7, and 99 as best performing schemes over Northern Tunisia. The operational service can use these findings in their estimation by WRF. At least this work highlighted the big difference in the estimation of rainfall by the different WRF parameters. This work will encourage them to use ensemble method to get better results. For floods estimation users, this work gave an idea about the reliability of WRF model.

Author contributions: All authors contributed to the study conception and design. Material preparation, in situ data collection and analysis were performed by Saoussen Dhib and Prof. Bargaoui Zoubeida. Prof. Victor Homar and Doctor Maria Del Mar Vich supervised Dhib Saoussen in UIB for three months financed by the Laboratory of Hydarulic and environment modeling (LMHE-ENIT). WRF data were collected, processed and archived in the UIB server. The first draft of the manuscript was written by Saoussen Dhib and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. The article charge financed by COASTEPS project.

Acknowledgements: The first author gratefully acknowledges internship support in UIB financed by LMHE-ENIT, University of Tunis-El Manar. Authors thank the Tunisian General Direction of Water Resources for the in situ rainfall data and the staff of ENIT and of UIB especially Prof. Clement Ramis, Prof. Romualdo Romero March and Mr. Jordi Vallespir for their help. We acknowledge the Agencia Estatal de Investigación of Spain (AEI); and the European Regional Development Funds (ERDF) for its support to the project CGL2017-82868-R (Severe weather phenomena in coastal regions: Predictability challenges and climatic analysis, COASTEPS).

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Bibliography

Angevine, W.M., Jiang, H., and Mauritsen, T.: Performance of an Eddy Diffusivity–Mass Flux Scheme for Shallow Cumulus Boundary Layers, Mon. Wea. Rev., 138, 2895-2912, https://doi.org/10.1175/2010MWR3142.1, 2010.

Berrisford, P., Dee, D.P., Fielding Uppala, K., Fuentes, M., Kållberg, P.W., Kobayashi, S., Uppala, S.: The ERA-Interim archive, ERA Report Series, ECMWF, 1, 16, https://www.ecmwf.int/node/8173, 2009.

Bougeault, P., and Lacarrère, P.: Parameterization of Orography-Induced Turbulence in a Mesobeta-Scale Model. Mon. Wea. Rev., 117, 1872-1890, https://doi.org/10.1175/15200493(1989)117<1872:POOITI>2.0.CO;2**,** 1989.

Bretherton, C.S., and Park, S.: A New Moist Turbulence Parameterization in the Community Atmosphere Model. Journal of Climate, 22, 3422-3448, https://doi.org/10.1175/2008JCLI2556.1, 2009.

Crétat J., Pohl B., Richard Y., Drobinski P.: Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF, Clim. Dynam., https://doi:10.1007/s00382-011-1055-8, 2011.

Davis, C.A., Brown, B., and Bullock, R.,: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. Mon. Wea. Rev., 134, P.1772–1784, https://doi.org/10.1175/MWR3145.1**,** 2006.

Dhib, S., Mannaerts, C., Bargaoui, Z., Retsios, V., and Maathuis, B.: Evaluating the MSG satellite Multi-Sensor Precipitation Estimate for extreme rainfall monitoring over northern Tunisia, Journal of Weather and Climate Extremes, 16, 14-22, https://doi: 10.1016/j.wace.2017.03.002, 2017.

Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, and Jonker H. J. J.: Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data, *Theor. Comput. Fluid Dyn.*, **27**, 133– 148, https://doi : 10.1007/s00162-012-0281-y, 2013.

Ebert, E.E.: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework, Meteorol. Appl., 15, 51–64, https://doi : 10.1002/met.25, 2008

Evans, J., Ekstrom, M., and Ji, F.: Evaluating the performance of a WRF physics ensemble over south-east Australia. Clim. Dynam., **39**, 1241–1258, doi :10.1007/s00382-011-1244-5, 2011.

Fathalli, B., Pohl, B., Castel, T., and Safi, M. J.: Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia: A multi-model and multi-member approach, Clim Dynam, 52, 335-361, https://doi: 10.1007/s00382-018-4150-2, 2018.

Fehri, N.: L'aggravation du risque d'inondation en Tunisie : éléments de réflexion , *Physio-Géo*, 8 , 149-175, https://doi : 10.4000/physio-geo.3953, 2014.

Flaounas, E., Bastin, S., and Janicot, S.: Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using WRF. Clim. Dynam, 36, 1083–1105, https://doi:10.1007/s00382-010-0785-3, 2011.

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Discussions

Grell, G.A.: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations. Monthly weather review, 121, 764-787, https://doi.org/10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2, 1993.

Han, J., and Pan, H.L.: Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System, *Wea. Forecasting*, **26**, 520–533, https://**doi.org/10.1175/WAF-D-10-05038.1**, 2011.

Heinemann, T., Lattanzio, A. and Roveda, F.: The EUMETSAT multi-sensor precipitation estimate (mpe), EUMETSAT, 8, 2002

Hong, S., and Pan, H.: Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model. *Mon. Wea. Rev.*, **124**, 2322–2339, https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2, 1996.

Hong, S.Y., and Pan, H.L.: Convective Trigger Function for a Mass-Flux Cumulus Parameterization Scheme. Mon. Wea. Rev., 126, 10, 2599–2620, https://doi.org/10.1175/1520-0493(1998)126<2599:CTFFAM>2.0.CO;2, 1998.

Hong, S.Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Wea. Rev., **134**, 2318–2341, https://doi.org/10.1175/MWR3199., 2006.

Janjic, Z. I.: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-493(1994)122<0927:TSMECM>2.0.CO;2, 1994.

Jankov, I., Gallus, W. A., Segal, M., Shaw, B., and Koch, S. E.: The Impact of Different WRF Model Physical Parameterizations and Their Interactions on Warm Season MCS Rainfall. *Wea. Forecasting*, **20**, 1048–060, https://doi.org/10.1175/WAF888.1, 2005.

Kain, J. S., and Fritsch, J. M.: A one-dimensional entraining/ detraining plume model and its application in convective parameterization. J. Atmos. Sci., 47, 2784–2802, https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2, 1990.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update. *J. Appl. Meteor.*, **43**, 170–181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU> 2.0.CO;2, 2004.

Nakanishi, M., and Nino, H.: An improved Mellor-Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog, Bound.-Layer Meteor., 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

Nmiri, A.: Regional Downscaling Case Study – (1) (Evaluation des changements climatiques sur la Tunisie). Workshop ClimaSouth, LECCE, 24, 2014.

Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395, https://doi.org/10.1175/JAM2539.1, 2007.

Roberts, N. M., and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Wea. Rev., 136, 78–97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 2. NCAR Technical Note NCAR/TND468+STR, 2005.

Skamarock, W.C., Klemp, J.B., Dudhia J., Gill, D.O., Barker, D.M., Duda M.G., Huang, X.Y., Jordan W.W., and Powers, G.: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, https://doi:10.5065/D68S4MVH. P.125, 2008.

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Sukoriansky, S., Galperin, B., Perov, V., 2006: A quasi-normal scale elimination model of turbulence and its application to stably stratified flows. Nonlinear Processes in Geophysics, European Geosciences Union (EGU), 13 (1), 9-22, hal-00302693, 2006.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Uppala, S.M., Kallberg, P.W., Simmons, AJ, Andrae, U., Da Costa Bechtold, V., Fiorino M, Gibson JK, Haseler J., Hernandez, A, Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda M.A., Beljaars, A.C.M., Van De Berg L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J-F., Morcrette, J-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J.: The ERA-40 re-analysis, Quart. J. Royal Meteorol. Soc., 131(612), 2961-3012, https://doi:10.1256/qj.04.176, 2005.

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: 10.1175/2008MWR2415.1, 2008.

Zacharov, P., Rezacova, D, and Brozkova, R.: Evaluation of the QPF of convective flash flood rainfalls over the Czech territory in 2009. Atmospheric Research 131. P. 95–107, https://doi.org/10.1016/j.atmosres.2013.03.007, 2013.

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

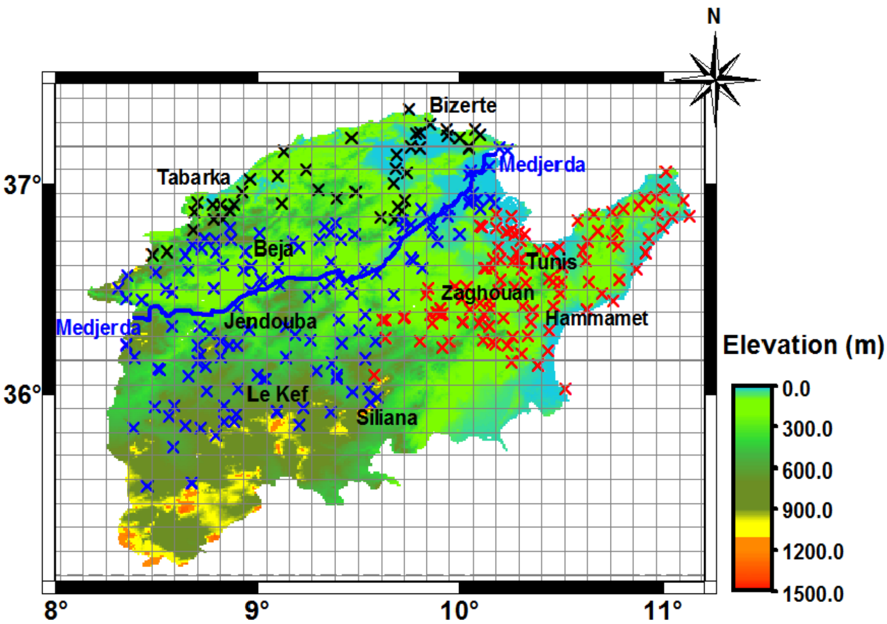


Figure 1: The rainfall network of the Northern Tunisia. The x symbols represent the rainfall stations (in Black W-3, in Yellow W-4, in Red W-5). Medjerda river is represented by the blue stream.

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Discussions

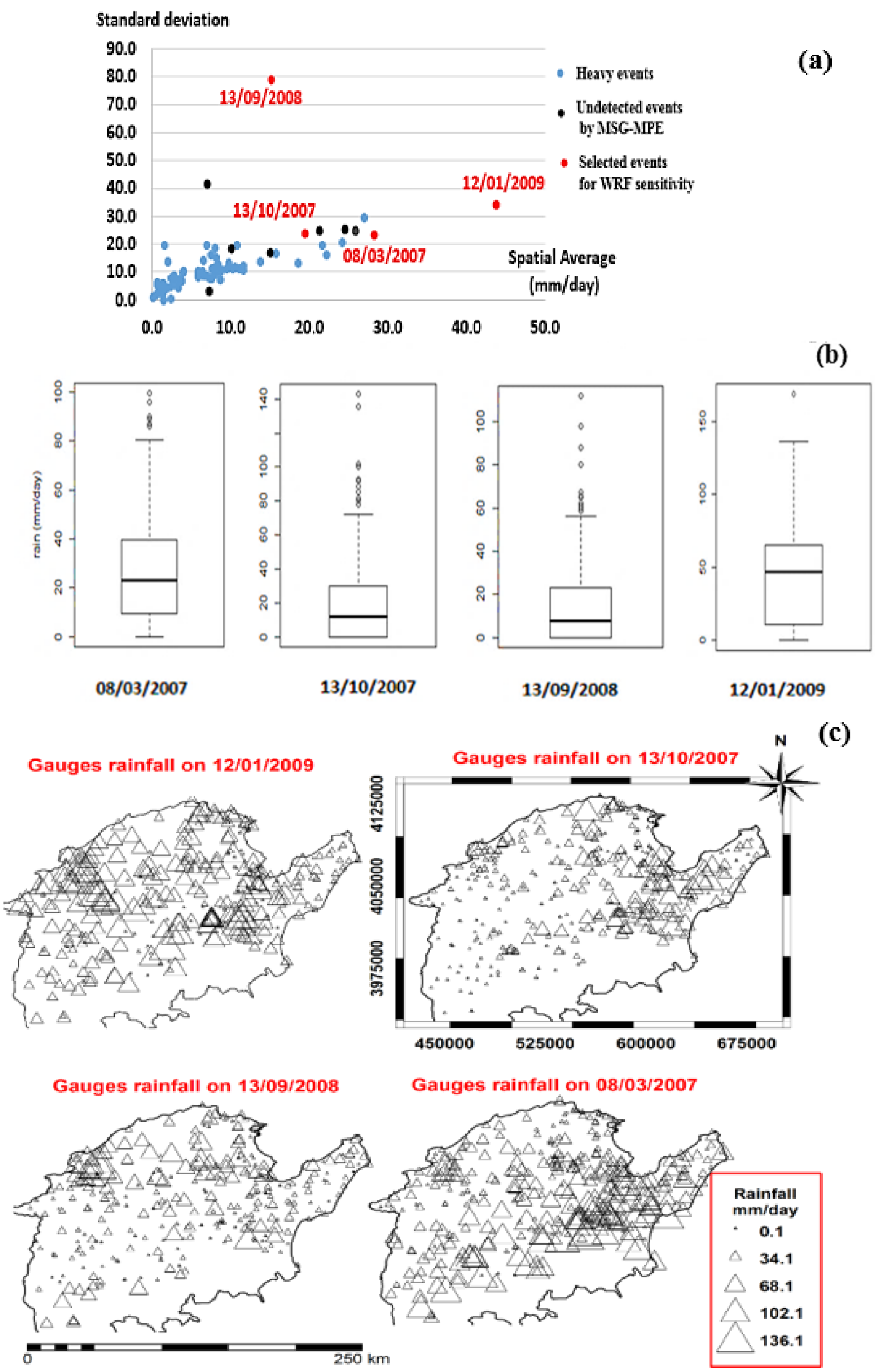


Figure 2: a) Spatial average of In-situ heavy events against their standard deviation b)The rainfall boxplot distribution for the studied cases, c) Gauges rainfall maps for the four case studies

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

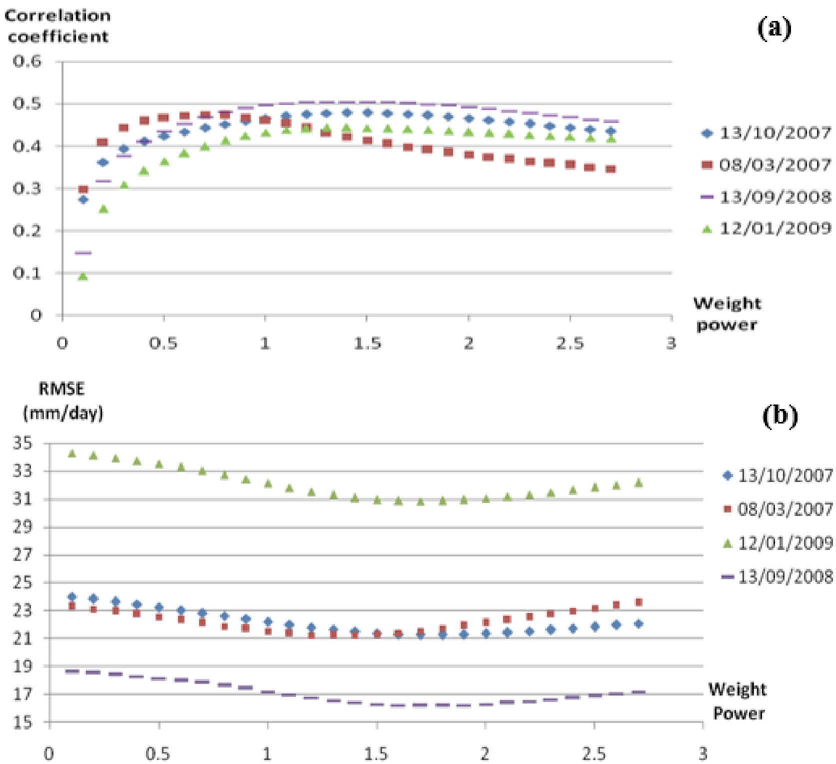


Figure 3: a) Correlation coefficients and (b) RMSE versus the Power of the IDW weight (exponent)

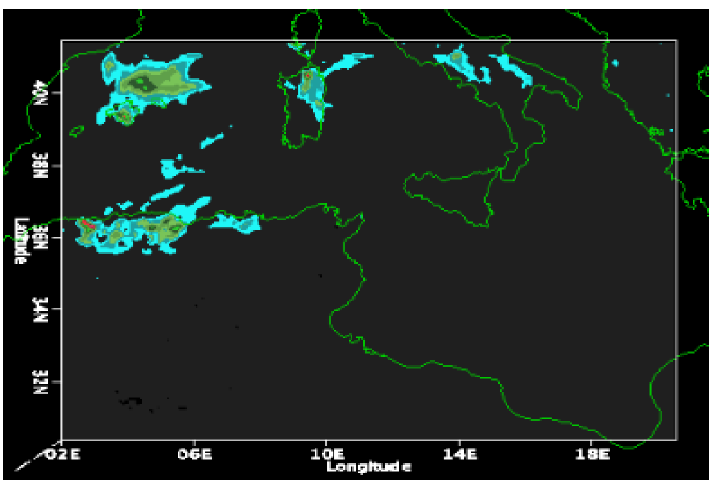


Figure 4: WRF domain of the study area

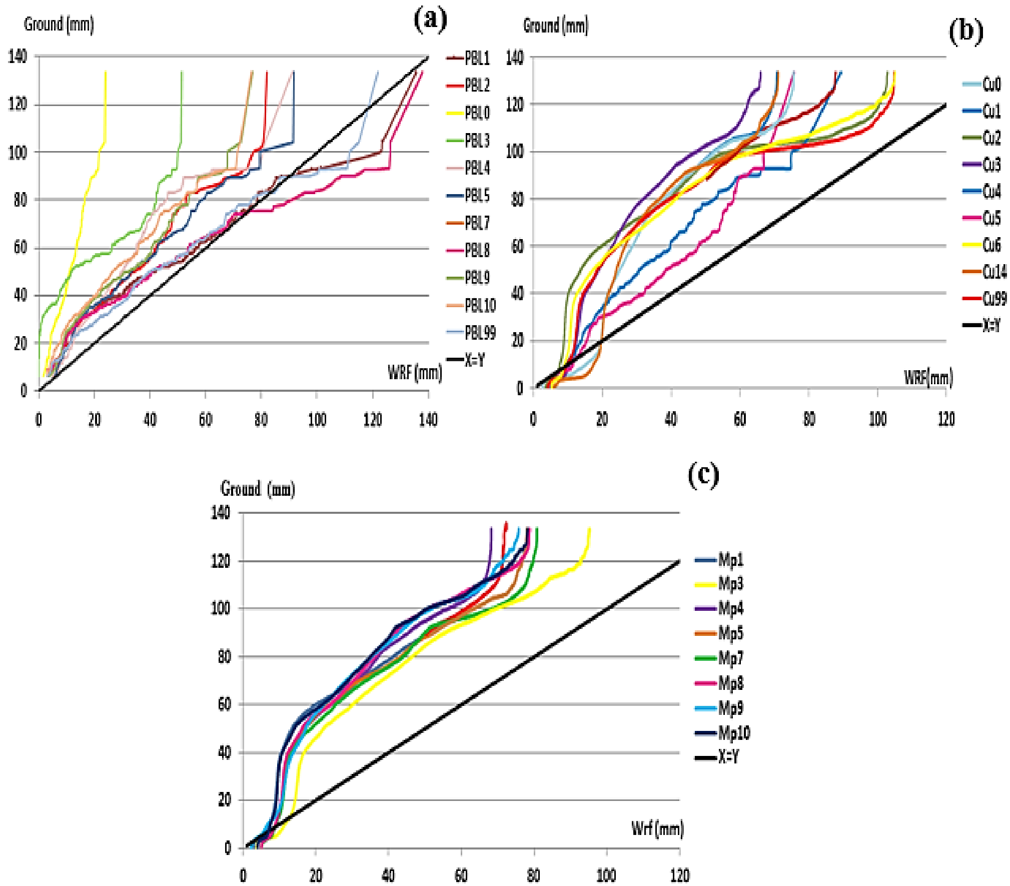


Figure 5: PBL (a), Cumulus (b), and Microphysics (c) quantile presentation of different schemes rainfall estimation by WRF in comparison with ground data for the 12/01/2009.

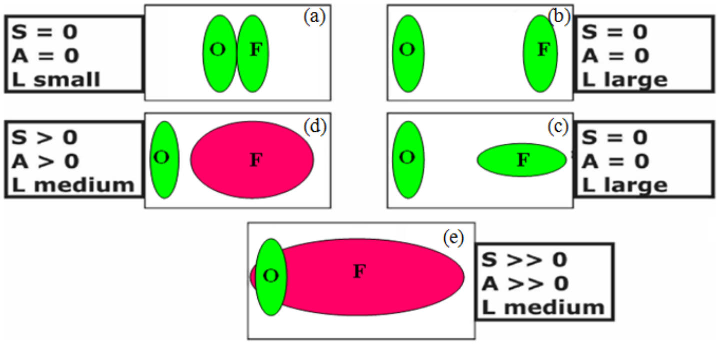


Figure 6: An example of the qualitative application of SAL for various forecast (F) and observation (O) cases).

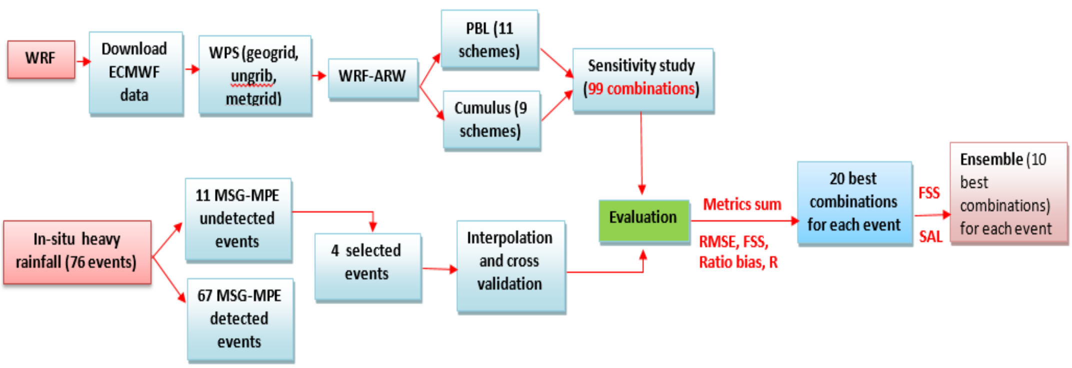
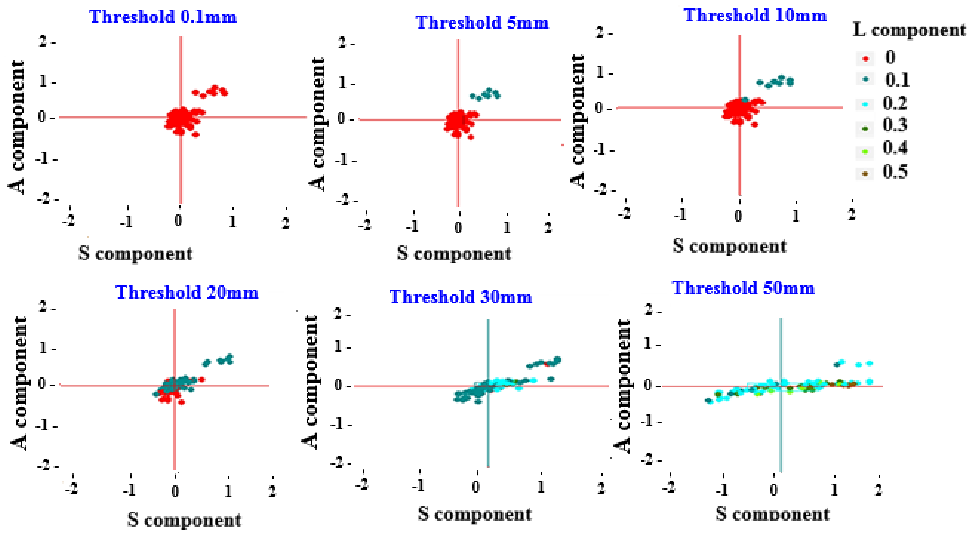


Figure 7: Steps of Processing and sensitivity study

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

(a) SAL



(b) FSS

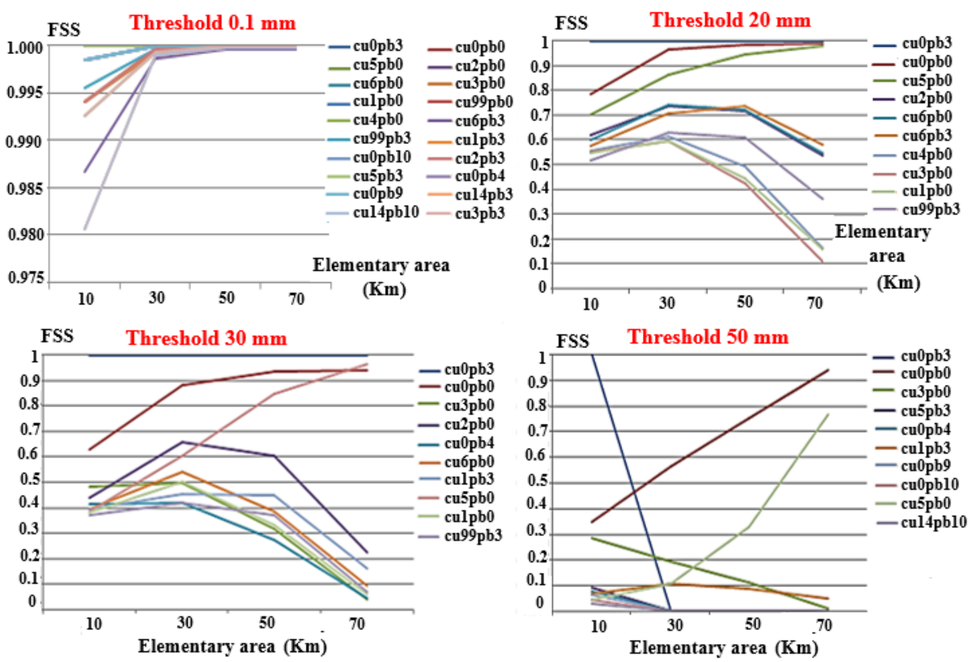


Figure 8: (a) The SAL evaluation components and (b) the FSS verification of the 08/03/2007 event

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

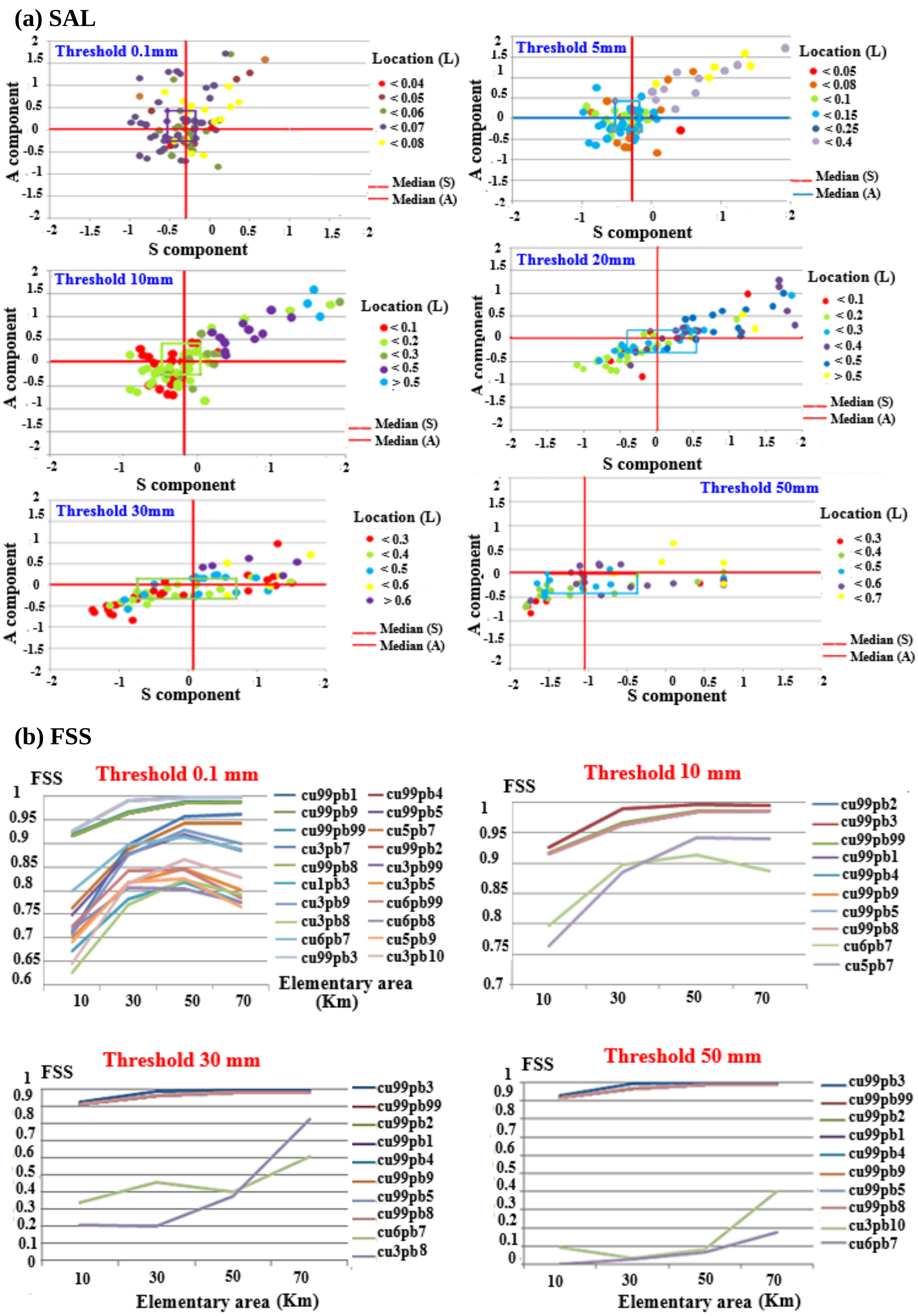


Figure 9: (a) The SAL evaluation components and (b) the FSS verification of the 13/10/2007 event

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

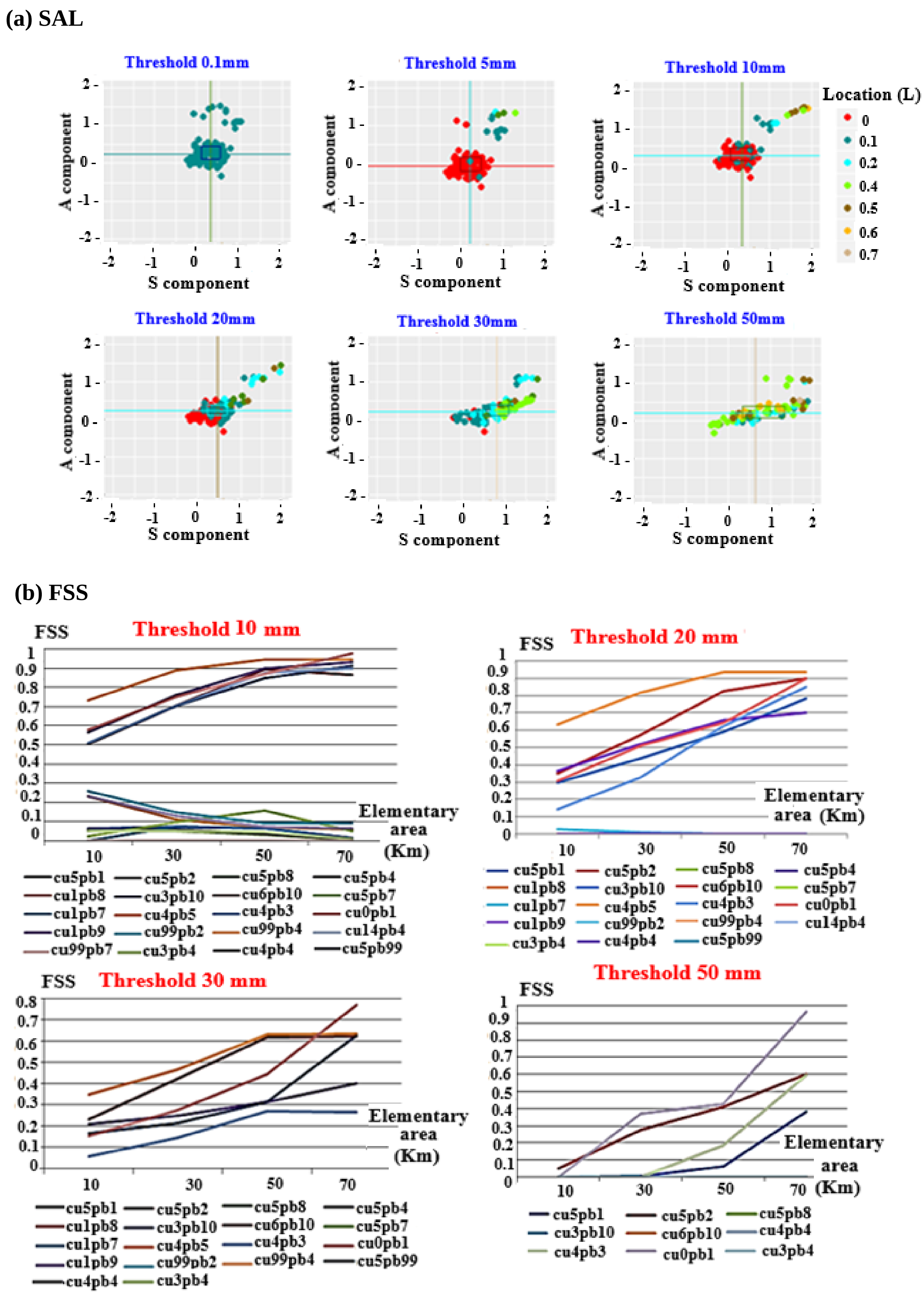
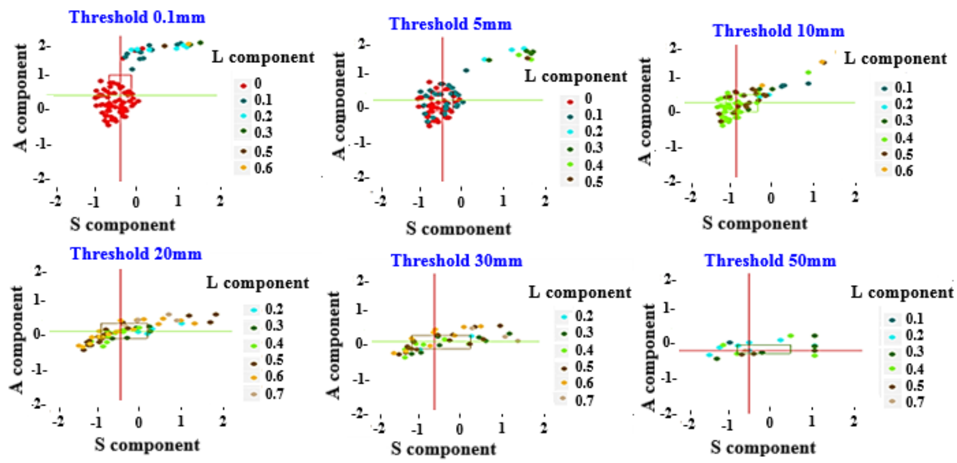


Figure 10: **The SAL components of the best 20 combinations for the event 12/01/2009.**

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

(a) SAL



(b) FSS

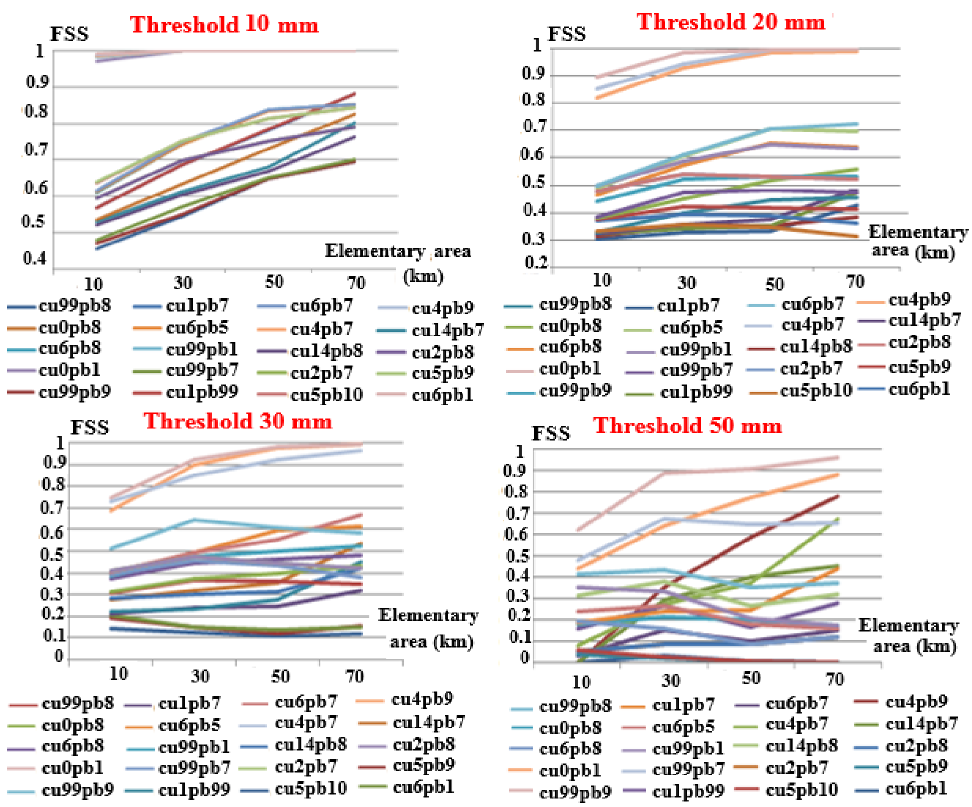


Figure 11: SAL components for different thresholds for all the combinations for the 13/09/2008

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

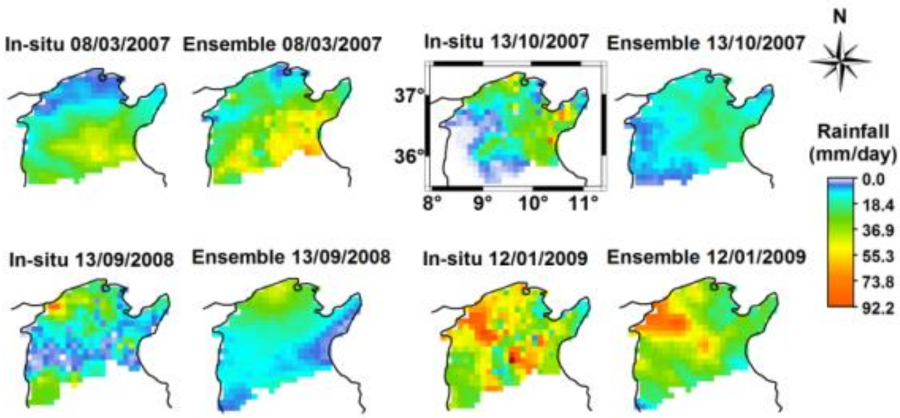


Figure 12: Studied events and the WRF ensembles

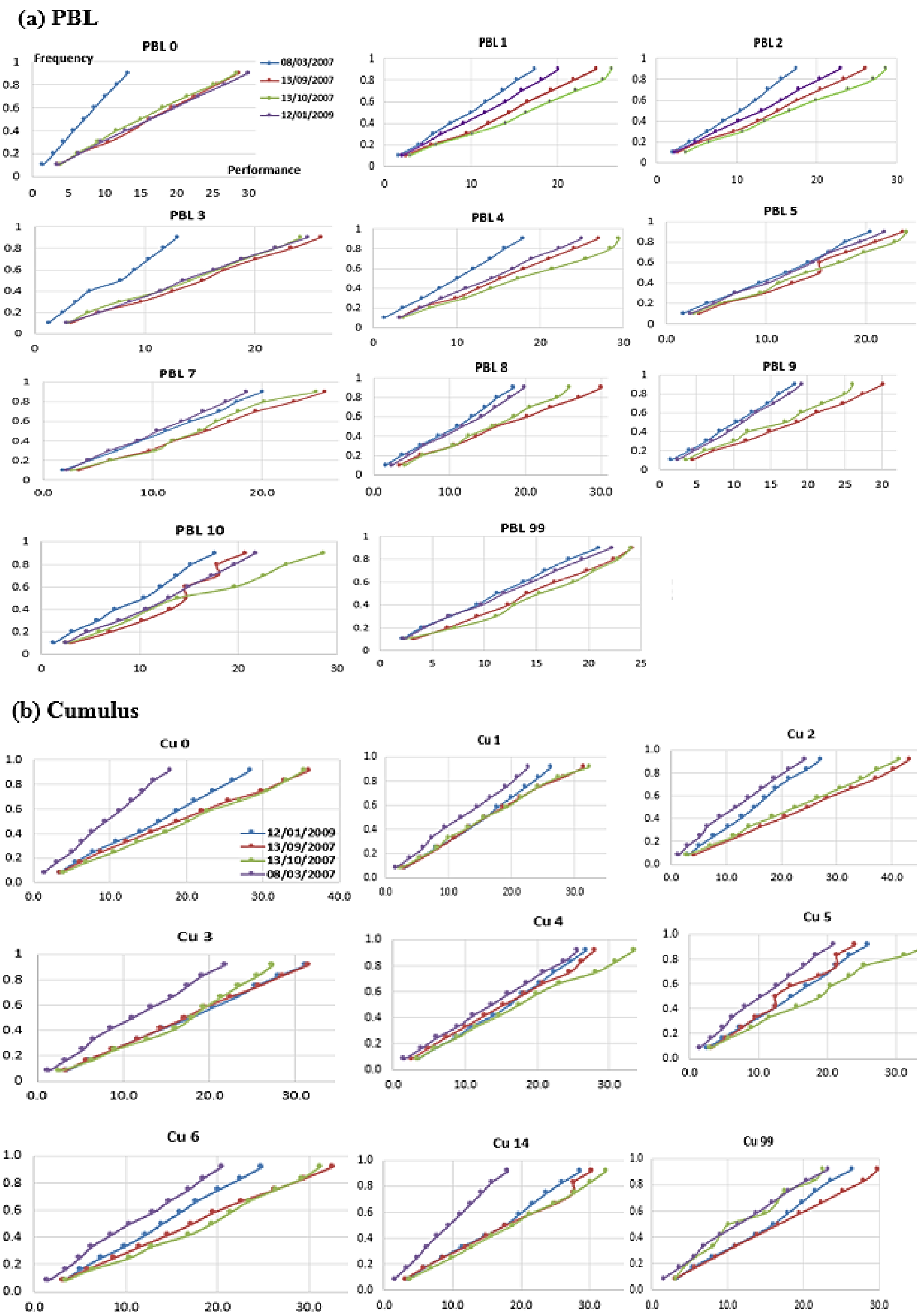


Figure 13 : The sensitivity study of the four events for (a) the different PBL schemes (The legend mentioned at PBL 0) and (b) the different Cu schemes (The legend mentioned at Cu 0)

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

605 **Table 1: Cu_physics and PBL parameterization schemes used in our study**

Cu_physics nomenclature number	Scheme	Description	PBL_physics nomenclature number	Scheme	Description and reference
0		no cumulus	0		no PBL
1	Kain-Fritsch (KF)	Convection deep and shallow, mass flux with downdrafts and CAPE exclusion time scale (Kain 2004)	1	YSU (Yonsei University)	Parabolic profile in the mixed layer, Non-local-K, entrainment layer explicit (Hong et al. (2006),
2	Betts-Miller-Janjic (BMJ)	Well-mixed profile, Operational Eta scheme. (Janjic (1994)	2	MYJ (Mellor-Yamada-Janjic)	One-dimensional prognostic turbulent kinetic energy. (Janjic (1994)
3	Grell-Devenyi (GD) ensemble	Ensemble using 144 sub-grid members, Multi parameter, multi-closure,	3	GFS	Predicts TKE and other second-moment terms (Hong and Pan (1996).
4	Old SAS (OSAS)	Scheme of Simple mass-flux with quasi-equilibrium ending with shallow mixing. (Hong and Pan 1998)	4	QNSE (Quasi-Normal Scale Elimination)	Option of TKE-prediction using a new theory of stably stratified regions. (Sukoriansky, et al. 2006)
5	Grell-3 -D (G3)	Improved version of the GD scheme (option cugd_avedx) is turned on Grell (1993).	5	MYNN2	Nakanishi and Niino with Level 2.5 (Nakanishi and Niino (2006), Mellor-Yamada
6	Tiedtke	Mass-flux with the CAPE-removal, shallow component and momentum transport. Tiedtke (1989)	7	Asymmetric Convective Model (ACM2)	Downward mixing, and upward mixing for local and nonlocal (Pleim (2007).
14	New SAS (NSAS)	New scheme of mass-flux using deep and shallow mechanisms and momentum transport (Han and Pan (2011))	8	BouLac (Bougeault-Lacarrère)	Option of TKE-prediction useful with urban model (BEP) (Bougeault and Lacarrere (1989).
99	Old Kain-Fritsch (old KF)	Scheme deep convection based on mass flux theory with downdrafts and CAPE without time scale (Kain and Fritsch (1990))	9	UW (Bretherton and Park)	CESM climate model with option TKE scheme (Bretherton and Park (2009).
			10	TEMF (Total Energy - Mass Flux)	Total energy Prognostic variable with mass-flux. Angevine, et al. (2010)
			99	MRF	KF older version using an implicit approach of entrainment layer mixed layer (Hong and Pan (1996)

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

Table 2: The metrics of the best 10 combinations

Combinations	**RMSE**	**Ratio bias**	**R (Pearson)**	**FSS**	**Metrics sum**
cu99pb3	18.42	0.50	0.76	1.00	1.25
cu5pb3	18.14	0.52	0.74	1.00	1.25
cu0pb3	18.78	0.50	0.74	1.00	1.27
cu1pb3	17.72	0.55	0.73	0.99	1.27
cu6pb3	18.45	0.51	0.71	0.99	1.33
cu3pb0	13.76	0.90	0.59	0.99	1.35
cu2pb0	12.77	0.86	0.62	0.99	1.39
cu0pb4	16.28	1.15	0.56	1.00	1.40
cu5pb0	14.48	0.90	0.55	1.00	1.41
cu0pb0	14.67	0.86	0.56	1.00	1.43

Table 3: Comparison between the different combination schemes skills of the 13/10/2007 event.

Combinations	RMSE	Ratio bias	R (Pearson)	FSS	Metrics sum
cu99pb1	13.09	0.90	0.53	0.92	1.0
cu99pb4	13.09	0.90	0.53	0.92	1.0
cu99pb9	13.22	0.83	0.54	0.92	1.1
cu99pb5	13.42	0.86	0.50	0.92	1.1
cu99pb99	14.14	0.65	0.56	0.92	1.3
cu5pb7	16.09	0.98	0.29	0.92	1.5
cu99pb2	14.66	0.86	0.33	0.92	1.6
cu99pb8	14.76	0.74	0.37	0.92	1.6
cu6pb7	20.58	1.41	0.37	0.92	2.1
cu99pb3	18.30	0.28	0.33	0.93	2.2

Table 4: The best 10 combinations metrics

	RMSE	Ratio bias	R	FSS	Metrics sum
cu99pb9	20.61	0.74	0.54	0.98	1.70
cu4pb7	22.46	0.67	0.55	1.00	1.81
cu6pb8	22.36	1.00	0.35	0.97	1.81
cu6pb5	21.37	0.90	0.39	0.97	1.81
cu99pb8	24.59	0.88	0.42	0.98	1.87
cu4pb9	23.70	0.62	0.49	1.00	1.91
cu14pb8	30.31	1.05	0.43	0.98	1.94
cu2pb7	24.64	0.81	0.50	0.93	1.96
cu2pb8	24.64	0.81	0.50	0.93	1.96
cu99pb1	21.68	0.75	0.39	1.00	2.02

https://doi.org/10.5194/nhess-2020-376
Preprint. Discussion started: 5 January 2021

Natural Hazards and Earth System Sciences
Open Access
Discussions

Table 5: The best 10 combination metrics

Combinations	RMSE	Ratio bias	R (Pearson)	FSS	Metrics sum
cu5pb1	13.78	0.96	0.52	0.95	1.63
cu5pb2	16.81	1.12	0.39	0.92	1.80
cu5pb8	14.39	0.96	0.42	0.67	1.97
cu1pb8	14.48	0.96	0.13	0.83	2.31
cu4pb4	16.19	0.81	0.05	0.96	2.39
cu1pb7	14.59	0.88	0.10	0.84	2.45
cu4pb5	22.74	1.05	-0.16	0.96	2.48
cu4pb3	22.59	1.06	-0.15	0.96	2.49
cu0pb1	23.31	1.24	-0.11	0.98	2.52
cu1pb9	15.35	0.89	0.01	0.84	2.55