Regional ensemble forecast for early warning system over small Apennine catchments on Central Italy

Rossella Ferretti[1,2], Annalina Lombardi[1], Barbara Tomassetti[1], Lorenzo Sangelantoni[1], Valentina Colaiuda[1], Vincenzo Mazzarella[1], Ida Maiello[3,1], Marco Verdecchia[2], and Gianluca Redaelli[1,2]

[1]CETEMPS, University of L'Aquila, L'Aquila, Italy
[2]Department of Physical and Chemical Sciences, University of L'Aquila, L'Aquila, Italy
[3]Centro Funzionale Regione Abruzzo, L'Aquila, Italy

Correspondence: Rossella Ferretti (rossella.ferretti@aquila.infn.it)

Abstract. The weather forecasts for precipitation have considerably improved in recent years thanks to the increase of computational power. This allows to use both a higher spatial resolution and the parameterisation schemes specifically developed for representing sub-grid scale physical processes at high resolution. However, precipitation estimation is still affected by errors that can impact on the response of hydrological models. To the aim of considering the uncertainties in the precipitation forecast and how they propagate in the hydrological model, an ensemble approach is investigated. A meteo-hydro off-line coupled ensemble is built to forecast events in a complex orography terrain where catchments of different size are present. In this context, the meteo-hydrological ensemble forecast is implemented and tested for a severe hydrological event occurred over Central Italy on November 15, 2017, when a flood hit the Abruzzo region with precipitation reaching 200mm/24hours and producing damages with a high impact on social and economic activities.The newly developed meteo-hydro ensemble is compared with a high resolution deterministic forecast and with the observations over the same area. The Relative Operating Characteristic (ROC) statistical indicator shows how skillful the ensemble precipitation forecast is with respect to both rain gauge and radar retrieved precipitation. In addition, the meteo-hydro ensemble allows for an estimation of both the predictability of the event a few days in advance and the uncertainty of the flood. Although the modelling framework is implemented on the basins of Abruzzo region, it is portable and applicable to other areas.

1 Introduction

Floods and extreme rainfall are among the major natural hazards in Europe with over 1000 fatalities and an estimated cost of about 5.000 billion of Euros in damages, between 1998-2009 only (European Environment Agency , 2010). Italy is one of the countries more exposed to hydrogeological risk in the Mediterranean basin, with more than 90% of municipalities affected by flood and landslide risk (ISPRA , 2018). From 2013 to 2017, 67 casualties due to floods have been reported, with 26 casualties in 2018 only (IRPI-CNR , 2019). The Mediterranean basin is characterised by a highly urbanized coast and mountain ridges close to the coast. During the autumn season there is an increase of the energy available for storms (Duffourg and Ducrocq , 2011) because of large gradients of the meteorological quantities caused by the cool atmosphere and the warm sea favouring heat and moisture fluxes. That is why most of the heavy rainfall and floods are occurring in autumn in the Mediterranean area (Ferretti et

al. , 2014; Rebora et al. , 2013; Rotunno and Houze , 2007; Rotunno and Ferretti , 2003) causing natural disasters in the region. Recently, decadal observations and modelling experiments highlighted the changes of precipitation distribution, frequency and intensity (Van den Besselaar et al. , 2013; Scoccimarro et al. , 2015), and how those changes affected the hydrological cycle in terms of increasing frequency of flood events (Drobinski et al. , 2018; Marchi et al. , 2010). Specifically, a warmer atmosphere than the nowadays one with a large amount of water vapour may lead to an increment of intense to extreme precipitation events (Trenberth et al. , 2003; Willett et al. , 2008; Giorgi et al. , 2011). In a context of increasing likelihood for future weather extremes, the availability of an accurate meteorological and hydrological forecast system is essential for improving civil protection early warning systems, on which community safety and impacts reduction directly depend (Penning-Rowsell et al. , 2009). Moreover, because of the complex orography of the Italian regions with many small - to medium-sized steep and densely urbanised coastal catchments, a further reduction of the hydrological response time and an increase of flood risks is expected. Indeed, in the framework of the European Flood Directive 2007/60/CE (EU Flood Directive , 2007), these regions are mainly classified as P3 (highly dangerous) zones. Recent studies (Hally et al., 2015; Demargne et al. , 2014; Cloke and Pappenberger , 2009; Davolio et al. , 2008; Schaake et al. , 2007) are focused on the coupling between meteorological and hydrological models in order to improve the quantitative precipitation forecast (QPF) and to predict the floods with a sufficient outlook. The coupling of the meteo and hydrological models requires meteorological observed or simulated variables (mainly, but not only, precipitation and temperature) used as forcing fields in hydrological models (Cloke and Pappenberger , 2009; Alfieri et al. , 2013; Abaza et al. , 2017; Wanders and Wood , 2016; Fan et al. , 2015). Hence, the quality of hydrological forecasts is largely determined by the quality of atmospheric input (Pappenberger et al. , 2005), even if the goodness of the hydrological forecast strongly depends on the verification methodology/choice (Pappenberger et al. , 2008). Temporal and spatial scales of the atmospheric forcing have to be calibrated according to the catchment features. In case of small-sized and mountainous catchments, because of a more responsive hydrology to the precipitation events, the discharge predictions require a very accurate precipitation forecast. An accurate, in space and time, precipitation prediction represents one of the most difficult tasks in Numerical Weather Prediction (NWP), as resulting from complex processes ranging from large-scale atmospheric dynamics to clouds microphysics. The use of meteorological models with high spatial resolution improves the QPF, but the estimation of the exact location and space-time evolution is still a challenge. In addition, their high computational cost limits the length of the forecast time which is often not enough to ensure sufficient lead time for actions. A potential solution consists in Ensemble Prediction Systems (EPS) which represent one of the areas from which the largest benefits in predictive skill have been obtained in the context of NWP (Buizza et al. , 2005; Bauer et al. , 2015). Even though EPS are characterized by a lower resolution with respect to deterministic forecasting, their added value belongs mainly to two aspects: 1) providing information about forecast uncertainty; 2) severe-to-extreme events occurrence likelihood. More in detail, the analysis of ensemble members distribution allows to provide the most likely events magnitude coupled to an estimation of all potential outcomes which characterises forecast uncertainty (ensemble members standard deviation or spread). On the other hand, the portion (i.e., frequency) of ensemble members predicting values exceeding empirical thresholds corresponding to extreme events can be derived.

As already discussed in previous studies, ensemble weather prediction systems at different spatial scales and using different approaches (Marsigli et al. , 2005; Vie et al. , 2011) hold a large potential for hydrological forecasting (Demargne et al. , 2014; Cloke and Pappenberger , 2009; Schaake et al. , 2007; HEPEX , 2004). In the last decade, the scientific community paid an increasing attention to study the EPS coupled to hydrological models aiming at improving early warning systems on different spatial scale ranging from global to regional (Addor et al. , 2011; McCollor and Stull , 2007; Davolio et al. , 2008; Calvetti and Filho , 2014; Hally et al., 2015; Saleh et al., , 2016).

In this context, the possibility to quantify and estimate forecast uncertainties allows the end users of hydrological models to manage the risk and to decide the actions to be taken with the aim of reducing the possible damages (Hamill et al. , 2005; Schaake et al. , 2007). Although the main uncertainty characterising hydrological forecast results from the precipitation input, uncertainty characterising the hydrological sphere represents another point to be carefully considered. Traditionally, only uncertainty pertaining the weather forecast sphere is accounted (Cloke and Pappenberger , 2009). In fact, being forced by individual ensemble members, the same probabilistic approach previously discussed could be applied to the hydrological model as well. This would allow to characterise the range for all potential hydrological scenarios and also to assess how weather prediction uncertainty propagates into the hydrological model. This coupled probabilistic approach could further foster the level of confidence that may be associated to the forecasts. In this work the traditional approach is followed; based on Cloke and Pappenberger (2009), the total uncertainty is probably underestimated because of the lack of the hydrological model uncertainty which can be obtained by perturbing, for example, the geometry of the system, the model parameters, etc..

In this paper a preliminary evaluation of both a meteo regional ensemble and a meteo-hydro ensemble forecast chain, developed at the Center of Excellence in Telesensing of Environment and Model Prediction of Severe events (CETEMPS) is presented. The meteo-hydrological modeling chain consists on connecting the dynamically downscaled Weather Research and Forecasting model (WRF-ARW) to the CETEMPS Hydrological Model (CHyM, Tomassetti et al. (2005); Coppola et al. (2007); Verdecchia et al. (2008)). The WRF regional ensemble is built by using as initial conditions all the members (20) plus the control forecast of Global Forecast System (GFS) from the National Oceanic and Atmospheric Administration (NOAA). The CHyM ensemble is built by using the WRF-Regional 20 members ensemble plus the control. To the purpose of assessing the reliability of an operational regional ensemble, a preliminary study of a heavy precipitation event is used as test case for the above mentioned meteo-hydro ensemble chain. A statistical evaluation of the WRF-ensemble mean precipitation is performed by using Relative Operating Characteristic (ROC) curve considering rain gauge and Radar Surface Rainfall Total (SRT) data as reference products. Moreover, several experiments are performed using a few initialisation procedure: 1) CHyM is forced by WRF-ensemble mean parameters (CHyM-WRF-MEAN); 2) CHyM is forced using the 20 WRF-Regional members plus the control (CHyM-ENS).; 3) CHyM is forced by the deterministic high resolution WRF (HR); 4) CHyM is initialised using the observations (precipitation and temperature). The results of the ensemble chains will be compared to the results of both: experiments 3 and 4.

The newly developed hydrological stress index Best Discharge-based Drainage (BDD), built to detect catchment segments that are most likely to be stressed by weather extreme events, is used to analyse the results of the meteo-hydro chains in terms of maps and time series.

The novelty of this work consists in applying a coupled probabilistic approach to both the weather and the hydrological ensemble forecasts, for a small catchments in complex orography. Two different ensemble meteo-hydro configurations are proposed: 1) a pseudo-hydro ensemble forecast where the hydrological model is forced by the mean precipitation produced by the WRF 21-member ensemble; 2) a CHyM ensemble composed by 21 members initialized using the 21 WRF members. The uncertainty (i.e., ensemble members spread) and probability of extreme events (e.g., frequency of ensemble members predicting values beyond defined threshold) will be provided for both ensemble configurations by WRF regional ensemble, but for the second configuration a contribution by the hydrological component may occur. The analysis of these two configurations will help understanding how weather forecast uncertainty propagates into the hydrological modelling outputs, representing an added value in the hazardous weather-related events prediction.

Based on the results of previous studies (Cloke and Pappenberger , 2009; Marsigli et al. , 2005), in this work we assume that for the meteorological regional ensemble all the GFS members plus the control are sufficient to represent the meteorological uncertainty. Obviously, a larger ensemble would ensure a larger spread but unfortunately the 51 members of the ECMWF ensemble are not available for operational purpose. Moreover, a regional ensemble is computationally costly therefore we should reduce the ensemble in any case. This will be the topic of a next paper, following the work of Buizza and Palmer (1998), though Jaun et al., (2008) showed that using 10 members only can be sufficient for having benefits from an ensemble approach for flood forecasting.

2 Case study

During November 13, 2017 a deep upper level trough, associated with an intrusion of cold air from the Arctic region, entered the Mediterranean area and advected south-westerly flow over western central Italian regions. The surface depression was located over Central Italy, advecting easterly flow over the Adriatic regions (Fig. 1a), and the thermal front was extending from north Africa to the southern Abruzzo region. In the following 48 hours the upper level trough developed into a cut-off low (Fig. 1b) over the central Mediterranean Sea and the axis of the surface depression tilted (becoming in phase with the upper level one) advecting north-eastern flow over Abruzzo region. Hence on November 15, 2017 there was advection of warm air at low levels and cold air at the upper ones over the Adriatic sea, producing a highly unstable environment. Therefore, the event was characterized by two phases: in the first phase from November 13 to 14, 2017 the thermal front produced rainfall over southern Marche and northern Abruzzo regions (not discussed in the paper). During the second phase, on November 15, 2017 when the axis of the trough tilted , the precipitation moved southward (Fig. 2b) and the advection of cold air produced also snowfall on the east side of the mountain ridges of Abruzzo region.

Figure 2 shows the accumulated precipitation on 24 hours (from 12 UTC of 14 November) over Italy; heavy precipitation is found only along the Adriatic regions, with maximum peaks of 200mm/24h (Fig. 2a,b) recorded along the Apennine ridges. The long lasting rainfall produced effects at the ground over the Adriatic regions, particularly on Abruzzo region (Fig. 2b) as the alert called by the Civil Protection Agency (CPA) in the morning of 15 November shows (Fig. 9). This figure shows both the forecast for the alert area and the observation of flooded area as evidenced by the symbols (triangle on the figure) added by CPA

as the event develops. Figure 2c shows the 24 hours accumulated Radar Surface Rainfall Total (SRT) on Marche and Abruzzo regions. A similar areal distribution between the SRT and the rain gauges is found but a different amount of precipitation is observed. SRT data and rain gauges will be used for the statistical evaluation of both the WRF-regional ensemble and the high resolution forecasts.

3 WRF Ensemble set up and precipitation forecast

The Advanced Weather Research and Forecasting (WRF-ARW) model is used to build the regional ensemble. WRF-ARW is a non-hydrostatic model with terrain-following vertical coordinates and multiple-nesting capabilities (Skamarock et al. , 2008). The configuration for the regional ensemble is the following: one domain covering Italy (Fig. 3, yellow box) with a horizontal resolution of 9 km and 40 unequally spaced vertical levels up to 100 hPa, with higher resolution in the planetary boundary layer. The ensemble is built using all the members (20) and the control forecast (CNTR) from GFS Ensemble. The horizontal resolution of GFS ensemble system is 1^o; these analysis and forecast are used to produce a dynamically downscaled ensemble forecast at 9km. Several experiments have been performed to test the configuration and the following is the one producing the best results at this resolution:

- *Radiation:* Rapid Radiative Transfer Model (RRTM, Mlawer et al. (1997)) for long-wave, Dudhia scheme (Dudhia , 1989) for short-wave radiative processes;
- *Cumulus:* Kain-Fritsch (Kain , 2004) ;
- *Microphysics:* Hong and Lim (2006) single-moment bulk scheme, 6 class hydrometeors;
- *Boundary-layer and turbulence:* Mellor-Yamada-Janjic (Janjic , 1994) One-dimensional prognostic turbulent kinetic energy scheme with local vertical mixing;
- *Surface:* Monin-Obukhov-Janjic surface scheme with the Noah land-surface scheme (Niu et al. , 2011);

Moreover, based on the high resolution (1km) deterministic forecast operationally performed using WRF over the Abruzzo region at CETEMPS since 2016 (Pichelli et al. , 2017), a simulation initialized using the best GFS high resolution analysis and forecast at 0.25 ^{o}, update every 6 hours, is performed for this event and it is used as benchmark (HR).

All simulations start at 12UTC on Nov 13, 2017 and they end at 12UTC on Nov 16, 2017 and the boundary conditions are updated every 6 hours with the GFS members forecasts for the WRF-regional ensemble and every 6 hours with GFS high resolution deterministic forecast for the HR.

3.0.1 Ensemble members precipitation forecast

The accumulated precipitation and the associated weather characteristics produced by the 20 WRF members and the control are quite different. An example of the variability obtained forcing WRF regional ensemble using the 20 GFS members is discussed

by analyzing a few of the WRF members (01, 14, 19, 20 and CNTR). Similarly to most of the members, member 01 (Fig. 4b) clearly shows a large area of precipitation over central Tyrrhenian sea produced by the surface depression, whereas the area of heavy precipitation at the border between Marche and Abruzzo regions, reaching 200mm/24h (magenta area), is clearly driven by an orographic forcing. Member 14 shows (Fig. 4c) a similar pattern but a second maximum in the south Abruzzo is found (magenta areas). Similarly to member 14, member 19 produces two maxima of precipitation in the Abruzzo region (Fig. 4d, magenta areas) and a larger cell in the west side of the Apennine ridge with a higher maximum of precipitation (dark yellow area) than member 14. Member 20 strongly reduces the areal extent of the precipitation for both cells in the Abruzzo region with respect to member 19 and 14 showing a small area with peaks up to 200mm/24h (Fig. 4e magenta area). The CNTR is producing the largest cell in southern Abruzzo with amount of precipitation reaching values of 200mm/24h (Fig. 4a). The cell in the north side is comparable with the one produced by most of the members.

A qualitative comparison between the WRF members forecast and the observed accumulated precipitation (Fig.2b) suggests that all members catch the signal of heavy precipitation in the north side of Abruzzo region but all overestimate the areal distribution of the maximum. A good agreement with the observed precipitation is found for member 20 for what concerns the areal distribution and the maximum precipitation of the cell in the north side of Abruzzo, though a second cell in the east side is missed. In addition, both the members and the control forecasts underestimate the observed heavy rainfall along the Abruzzo coast if compared with the observations (Fig.2b).

3.1 Ensemble precipitation statistics

The information provided by the EPS relies on the analysis of three different statistics derived from the ensemble members distribution:

- ensemble mean from the 20 ensemble members and the control run;
- ensemble standard deviation;
- probability of the rainfall (or any other meteorological variables) exceeding a given threshold. This statistical analysis is derived mapping the ensemble members distribution with respect to the threshold.

To avoid the linear dependency between standard deviation value and mean ensemble precipitation, the coefficient of variation is computed allowing to assess the precipitation uncertainty independently from the amount of rainfall. For what concerns the threshold in Fig. 5c it has been arbitrarily defined and it can be adjusted in function of regional/local features, season and length of the forecast period; at each time step and for each grid point the following is computed (Eq. 1):

$$if \qquad var(i) > threshold \qquad then \qquad freq = \frac{\sum_{i=1}^{t_{mem+1}} i}{t_{mem+1}} \tag{1}$$

where var(i) is any meteorological variable of the i-member ; t_{mem} is the total number of the members.

The comparison between the 24h accumulated ensemble mean precipitation at 12 UTC on November 15, 2017 computed using all members (Fig.5a) and the control simulation clearly shows a reduction of the areal extent of the cell in the northern

Abruzzo and a reduction of the precipitation on the southern Abruzzo region with respect to the control run (Fig. 4a). This is an expected result confirming the dampening effect of the ensemble mean if compared to the deterministic simulation at the same resolution. For this event, even though the control simulation takes advantage of best GFS forecast IC, it produces a forecast poorer than the one using all members. In fact, the ensemble mean produces a forecast closer to the observations (Fig.2b) than the CNTR reducing both the areal extent of the cell in the north side of Abruzzo and the overestimation in the southern Abruzzo. However, it is noteworthy how the cell located in south Abruzzo shows higher values of the spread (Fig. 5b) even if characterized by less accumulated precipitation, suggesting a larger uncertainty in the southern Abruzzo area, for this event. Moreover, in correspondence of the most intense northern Abruzzo cell, there is a small ensemble spread. In fact, the ratio between the standard deviation and the mean is close to zero. This gives good confidence to the precipitation predicted by the ensemble mean. Beside ensemble mean value and related uncertainty, the characterization of severe-to-extreme events represents a focal point yet. At this regard, the probability of accumulated precipitation above 60 mm in 24 hours is shown in Fig.5c where almost the whole ensemble (more than 90% of the members) agrees on predicting precipitation equal or beyond such a threshold, suggesting a quite confident forecast, though on both north and south Abruzzo region.

3.2 Ensemble precipitation time series

An analysis of the EPS precipitation time series at a few stations located at the foothills of the Apennine (Tossicia, Arsita, and Villa Santa Lucia,Fig. 6a) and along the coast (Giulianova, Pescara and Chieti, Fig. 6a) is presented. A comparison is performed among the ensemble members (blue lines), the ensemble mean (cyan line), the deterministic high resolution (HR, red line) forecast and the observation (black line). The two stations along the coast (Pescara and Giulianova) show the maximum rainfall in the very early Nov 15, 2017 (Fig. 6b, c, black line) starting from Giulianova. Both the ensemble mean and the HR simulations show a good agreement with the observation in the onset of the rainfall at Giulianova and Pescara, but a large underestimation is found at Giulianova for both simulations (Fig. 6b, red and cyan lines). On the other hand a very good agreement is produced by the ensemble mean at Pescara (Fig. 6c, cyan line) for both timing and amount of precipitation for the first peak in the early morning of Nov 15, 2017, but the largest peak of precipitation at the end of the event is misrepresented by both ensemble mean and HR. At both stations the HR produces a second peak which was not observed. At Chieti station a fairly good timing is found for both ensemble mean and HR simulations, but an underestimation of the amount of precipitation is found (Fig. 6d, red and cyan lines). For what concerns the stations at the foothills (Tossicia, Arsita and Villa Santa Lucia) still a timing disagreement is found but the amount of rainfall is better reproduced by both HR and ensemble mean (Fig. 6e, f, g, black, red and cyan lines) suggesting a more accurate forecast if the orographic forcing is playing a key role. Finally, all the time series show a variability among ensemble members much smaller than the difference between ensemble mean and observations at the maximum of the rainfall, as for example at Giulianova station on Nov 15 at 00 (Fig. 6b). On the other hand, a larger variability among the members is generally found for small amount of rainfall: for example on Nov 15 at 12 UTC the spread is 15mm/3h and the difference between observation and mean is 5mm/3h, whereas at the same station on Nov 15 at 00UTC the spread is 10mm/3h and the difference between observation and mean precipitation is 23mm/3h (Fig. 6b). This would suggest a large variability of the spread depending on the maximum value of the precipitation; the same is found

for the other stations. As expected, a similar behaviour is found for the time series of the GFS ICs precipitation forecast (not shown), which does not include the maximum of observed precipitation among the members variability but it does it for other values of the rainfall. Generally, the NWP forecast at low resolution tends to underestimate the rainfall and the maximum of precipitation, as the resolution of the NWP increases the amount of precipitation tend to get closer to the observed maximum, but still a problem in the exact location and timing of the maximum can be found. Hence, we may expect that this regional ensemble (9km) underestimates the maximum of the precipitation and possibly no any member reaches the observed maximum as it is found for the GFS forecast.

3.3 Ensemble statistical evaluation

To the aim of objectively evaluate the reliability of the WRF-Regional ensemble a statistical approach is used. The Relative Operating Characteristic (ROC, Jolliffe and Stephenson (2012)) which plots the hit rate against the false alarm rate is computed to evaluate the ensemble. The 3 hourly precipitation from the ensemble mean and the 3 hourly accumulated precipitation from rain gauges (345 surface stations) and Radar Surface Rainfall Total (SRT) data on November 15 at 0900 and 1200UTC are used to built the ROC. The analysis is performed restricting the area to the one where the event occurred (Marche and Abruzzo) as shown by the inner box in Fig.3. Both Radar SRT and rain gauges data are interpolated on the model grid of the WRF-regional ensemble. To the aim of evaluating WRF-regional ensemble ability to forecast the onset of the precipitation and the convective one the following thresholds are chosen to compute the ROC: 1mm, 3mm, 5mm, 10mm, 15mm. The results of both WRF-ensemble mean precipitation and WRF-HR compared with rain gauge and SRT are shown in Fig.7. The steepness of the curve as well as the area under the curve (AUC) are an indication of the skill of the forecast (Storer et al. , 2019), therefore we concentrate on these two factors. Both WRF forecasts (ensemble and deterministic) show a high rate of increase of the Probability of Detection (POD) for thresholds of the precipitation up to 15mm on November 15 at 0900UTC suggesting a good skill for both of them if using rain gauges (Fig. 7a, black and blue lines). The comparison with Radar SRT, though showing a lower steepness than the previous ones for both WRF-ensemble mean and WRF-HR (Fig. 7a, magenta and red lines, respectively), still shows an AUC > 0.5 ensuring a skilful forecast for both. Moreover, the AUC for the HR is larger than the one for WRF-ensemble: for low thresholds it reaches higher values of POD than WRF-ensemble. On November 15 at 1200UTC, the skillfulness of the regional ensemble is confirmed: the steepest curve and the largest AUC are found for the WRF-regional ensemble compared with both rain gauges and SRT (Fig. 7b, black and magenta, respectively). The AUC is maxima if using SRT, probably because of the overestimation of the rainfall in the southern Abruzzo where the SRT is maxima(Fig.2c). The comparability of the skillfulness for WRF-regional ensemble (9km) with the WRF-HR (1km) may be due to the downgrading of the observation at the grid of the Regional Ensemble.

4 Hydrological model

The CHyM hydrological model was developed at CETEMPS by the hydrological group since 2002 and since then it runs operationally (Tomassetti et al. , 2005; Coppola et al. , 2007; Verdecchia et al. , 2008) producing alert mapping service to

support Abruzzo Regional Functional Centre decisions. The model is based on the kinematic wave approximation (Lighthill and Whitam , 1955) of the shallow water wave and the continuity and momentum conservation equations are used to simulate the surface routing overland and the channel flow. The CHyM model is a distributed grid-based hydrological model reaching a spatial resolution of 300m; it includes an explicit parameterization of different physical processes contributing to the hydrological cycle. CHyM is initialized using different sets of precipitation data which are assimilated and merged in a hierarchical way at each time step. The model can be used for any geographical domain up to the Digital Elevation Model (DEM) resolution and the drainage network is extracted by a sequence of native algorithm. The interpolation methods for DEM smoothing and meteorological variables spatialization follow a Cellular Automata-based algorithm (Wolfram , 2002; Coppola et al. , 2007). For the Abruzzo Region operational activity, CHyM runs at a spatial resolution of 300m and it is initialized using observed precipitation and temperature data for a spin-up time of 120 hours. The following 48 hours forecast is produced using the meteorological model forecast. For this case study, the same operational configuration is used: CHyM is forced with observed meteorological data until 23 UTC of 13 November 2017 and with WRF data for the following 48 hours. To the aim of highlighting differences resulting in the hydrological forecasts a few experiments are performed (Table1) using the WRF-CHyM chain with different initializations. An experiment is performed by using the WRF ensemble mean (CHyM-WRF-MEAN), a second one by using the deterministic high-resolution (CHyM-HR-WRF) simulation and, finally, a hydrological ensemble (CHYM-ENS) is built. Moreover, two simulations are carried out forcing CHyM with: observed data only (CHyM-OBS) which will be used as control simulation, and using the WRF-CNTR output.

Model	*Input*	*Output*	*BDD*
CHyM-WRF-MEAN	mean from all WRF members	CHyM output	BDD mean
CHyM-ENS	each WRF member	20 members CHyM output	BDD probability
CHyM-WRF-CNTR	WRF CNTR	CHyM output	BDD
CHyM-HR-WRF	HR WRF forecast	CHyM output	BDD (from HR)
CHyM-OBS	Observation	CHyM output	BDD observation

Table 1. CHyM simulations

4.1 BDD index

The analysis of the hydrological forecast is presented in terms of the hydrological stress index the Best Discharge-based Drainage (BDD, mm/h) which is able to detect catchment segments that are most likely to be stressed by severe weather. The use of a hydrological stress index is necessary because it is not straightforward to establish a threshold discharge level above which a critical event is to be expected and such threshold level should be calculated for each grid point because it depends on the size of the river bed in the selected point. In addition discharge observations in continuous time series, needed for the calibration, are often missing especially for small basins. To overcome this problem we tested different general definition of an alarm index and after simulating different case studies occurring in different basins of different size, we find that a suitable

definition could be the ratio between the maximum value of the predicted discharge, within a given time interval, and the square of hydraulic radius that is a "measure" of the river cross section for the selected point. The definition of BDD index has also a simple physical interpretation: it represents the average precipitation available for the runoff drained by each grid element from the upstream basin.

BDD index is computed at each grid point and time steps as the ratio between the flow discharge and the squared hydraulic radius that is function of the drained area, following Eq. 2:

$$BDD_{(i,j)}(t) = \frac{Q_{(i,j)}(t)}{R^2_{(i,j)}} \tag{2}$$

where Q is the discharge estimated by the model, R is the hydraulic radius and i and j are the grid points. As for many other models the hydraulic radius can be approximated as a linear function of drained area (Singh and Frevert , 2002) . In particular, $R = \beta + \gamma D^{\delta}$, where β, γ, δ are empirically constants to be optimised during the calibration phase. If the area is measured in km^2, typical values taken from literature are β=0.0015, γ=0.05 and δ close to 1 (Singh and Frevert , 2002).

In order to provide a suitable and synthetic information for flood alert mapping, it is often useful to plot the map of the maximum value of BDD index reached within a specified time interval.

Two warning thresholds are defined for the BDD index: a medium (orange) and a high warning (red), with a similar meaning to those defined by the Civil Protection authorities for the hydrometric height (Thielen et al. (2009)) Moreover, as the BDD index is based on the relationship between the computed discharge and the river geometry at each grid-point, the defined thresholds are general and applicable over the whole drainage network. These characteristics makes the index a strong user-oriented instrument.

To assess the flood event occurrences, the hydrometric level threshold exceedances or not-exceedances for the official station network belonging to the Abruzzo Functional Centre (CFA) is detected for this case study at the station level. A more complete geolocation of registered flood events, outside instrumented fluvial segments, is also inferred from local authorities reports (fire fighters, civil protection volunteers, police) as shown in Fig.8 and 9.

A preliminary evaluation of the BDD index is now presented. Figure 10 shows the comparison between two different, though related, normalised physical quantities: the time series of the hourly recorded water level and the corresponding BDD time series obtained from the CHyM control simulation (CHyM-OBS), for 7 relevant stations highlighted (yellow circle) in Fig.8. The (CHyM-OBS) is here assumed as the reference for the BDD threshold definition. In order to ease the comparison the two quantities (BDD index and water level) are normalised to their respective maximum. This preliminary validation is important because the BDD index obtained from the (CHyM-OBS) will be used in the following analysis as reference product for the validation of the whole meteo-hydro ensemble chain because of the lack of discharge observations. The comparison is qualitative and mostly focused on the threshold exceedance and maximum timing accordance between the index and the hydrometric level curves. From north to south the following rivers are accounted: Vomano, Tordino, Saline and Pescara (Fig. 9 the four red triangle-shaped, thin-bounded signs indicates the relevant hydrometers where the red hydrometric threshold has been exceeded). These sensors have been chosen because they are located close to the area where floods or critical hydrological level have actually been observed (Fig.8 and related discussion).

The BBD index correctly reproduces the timing and the hydrometric level peak (Fig.10, red and blue lines respectively) for the first 4 sensors (Fig.10 a,b,c and d). On the other hand, the Saline sensor shows (Fig.10e) an uncertainty in the prediction of the peaks of approximately one hour, which is exactly the time resolution of the series. In addition, for the sensors located in the Pescara catchments, the shape of curve is well predicted, but the hydrometric level is significantly overestimated within few hours between the 2 observed peaks (Fig.10f,g). Furthermore, it has to be considered that discharge for this river is strongly affected by the managements of hydroelectric power plants located in the upper part of the basin and, unfortunately no information are available about such managements.

Unfortunately, for the Abruzzo region we are not aware of how the hydroelectric systems are managed. For example, the hydroelectric power plants are located upstream respect to the areas involved in the event, where also precipitation maxima occurred. This suggest that the effect of the hydroelectric power plants is negligible in this case, the good agreement between BDD index and water level time series support this hypothesis.

Therefore, the good performance of the BDD index for this event for the Abruzzo rivers allows to use it for the following analysis.

5 Hydrological model results

The experiments presented in Table1 are now analyzed in terms of the BDD index and the two different ensemble meteo-hydro configurations are tested:

- a pseudo-hydro ensemble forecast where the hydrological model is forced by the mean precipitation produced by the WRF 21-member ensemble (CHyM-WRF-MEAN);
- a CHyM ensemble composed by 21 members initialized using the 21 WRF members which will be presented in the next paragraphs (CHYM-ENS).

For what concerns the hydro-ensemble, the uncertainty (i.e., ensemble members spread) is provided by WRF regional ensemble in the first case, whereas a contribution by the hydrological model is expected in the second one.

5.1 Pseudo hydro-ensemble versus hydro-deterministic

At first, the CHyM-WRF-MEAN using WRF temperature and precipitation ensemble mean is analysed. By means of quantifying reliability of ensemble mean driven hydrological forecast, 24 hours BDD stress index maps obtained considering CHyM-WRF-MEAN simulation and CHyM-OBS (Fig.11, b,c) are compared, for 15 November 2017. High values (warmer colors) of BDD highlight fluvial segments characterized by a high level of hydrological stress, where flooding are most likely to occur. The map is built assigning to each grid point of the drainage network the maximum value of the BDD index calculated according to Eq. 2. Hence, the maps in Fig.11 represent the worst expected situation from 00 UTC to 23 UTC on 15 November 2017. A qualitative analysis of the BDD index maps (Fig.11), obtained with different precipitation scenarios, show good performances of proposed alarm index by highlighting the areas where major hydrological stress have to be expected. In fact,

all the observed flood events shown in Fig.8 and 9 are correctly predicted by a critical value of the BDD index. The efficiency of proposed approach seems the same for the main channel as for the small catchments, despite a moderate overestimation of critical hydrological situation in the southern part of simulated basin. The comparison between the BDD by CHyM-OBS map (Fig.11, a) and the Fig.8 and 9, where actual flooded areas are highlighted along the central and northern Abruzzo drainage network, shows a good spatial coherence. The BDD map by CHyM-WRF-MEAN precipitation (Fig.11, b) shows a good agreement with the hydrological control run (Fig.11, a), for the main catchments over central-northern Abruzzo on the Adriatic side. However, an overestimation along the coast in the south side of Abruzzo region and an underestimation in the north side is found for the small catchments for CHyM-WRF-MEAN. This is probably due to an underestimation of the rainfall along the coast (Fig. 5,a) by the WRF-regional ensemble as well. In this condition, the main contribution to the hydrological stress (i.e. BDD index) is given by the heavy rainfall produced on the mountains which is able to charge the longest rivers, whereas the shortest streams nearby the coast do not receive enough precipitation to turn on (warm color) the BDD index. Furthermore, the hydrological stress is overestimated in the southern part of the domain (Fig.11, b). To further verify the CHyM-WRF-MEAN forecast a comparison with the deterministic high resolution forecast (CHyM-HR-WRF) is performed. The BDD index for CHyM-HR-WRF (Fig.11, c) is very similar to the index map resulting from CHyM-OBS (Fig.11, a). In this configuration, the CHyM model is able to capture a higher hydrological stress over the small catchments in the northern coastal area of Abruzzo (Fig.11, c, red color), that are missed in the CHyM-WRF-MEAN (Fig.11, b, light blue). On the contrary, the smallest flooded fluvial segment along the coast, as the Calvano stream (in the north side of Abruzzo but southern of the previous one), are not flooded by the deterministic run even if heavy precipitation was recorded. This is caused by precipitation not catch by the meteorological model because occurring only on a very small area. The stress index over the southern catchments is overestimated, as well. A further verification of the goodness of these results is obtained by comparing the BDD index maps with the alert map issued by the Department of Civil Protection of Abruzzo for these rivers (Fig. 9) where a risk of flood was issued. Hence, these results would suggest an overall good performance of both the CHyM driven by the deterministic high resolution forecast and the one driven by the WRF regional ensemble forecast.

5.2 CHyM Ensemble

Finally, a CHyM hydrological ensemble forecast is built by performing 21 simulations using the 20 members plus the control of the WRF-ensemble as initial conditions. The output of the 21 CHyM members is used to build a probability BDD_{prob} index (Fig.12) which is computed by the following equation:

$$BDD_{prob} = \frac{N_{BDD}}{N_{ens}} \tag{3}$$

where N_{BDD} is the number of members for which BDD reaches higher values than the alarm thresholds during the 24 hours and N_{ens} the total ensemble members (20, excluding the WRF control run).

To verify this BDD_{prob} index, a comparison with the alert map issued by the Civil protection agency during the event (Fig.8 and 9) is presented. Maximum flood probability is found over the northern Adriatic-side catchments (Fig. 12a), where almost all

members simulates precipitation peaks, according to CHyM-OBS and CHyM-HR-WRF (Fig.11, a,c). Small catchments stress is not simulated over this area, because of a general underestimation of the precipitation amount in the costal area, probably caused by to the lower horizontal resolution used in the meteorological ensemble. An improvement by the CHyM-ENS is clearly obtained as it is shown by the BDD_{prob} index over the southern part of the region (Fig. 12a) , where the overestimation of the stress condition detected by CHyM-WRF-MEAN (Fig.11, b) is not found, coherently with the real hydrological effect caused by the event in that area. The CHyM ensemble spread is also computed (Fig. 12b), in order to give a complementary information to the BDD_{prob} map. The large value for the spread of the precipitation in the southern Apennine's ridge produced by WRF regional ensemble (Fig. 5c) has implications on the largest catchment of southern Abruzzo: the maximum BDD spread is here obtained for the main catchment, corresponding to the Sangro river, whereas in the northern part of the domain the spread is smaller confirming the reliability of the flood forecast. These results suggest a coherent variability between WRF and CHyM.

5.3 CHyM time-series

To the aim of further evaluating the ability of the hydrological ensemble to correctly reproduce the stress distribution, an analysis of the BDD index time series at a few stations is also presented. The CHyM-OBS is compared with: the one produced by the pseudo-hydro ensemble forecast (CHyM-WRF-MEAN) where the hydrological model is forced by the mean precipitation produced by the WRF 21-member ensemble; the CHyM ensemble (CHyM-ENS) composed by 21 members initialized using the 20 WRF members plus the control; CHyM-HR-WRF simulation forced using the HR deterministic forecast. In what follows the BDD index time series are presented for all the CHyM simulations except for the mean of the 21 CHyM ensemble members because it is similar to the one produced by the CHyM-WRF-MEAN. The BDD index time series for the stations along the rivers Vomano, Tordino, Saline and Pescara (from north to south) shows (Fig. 13) a BDD spread between 10 and 20 mm/h around the peak. The CHyM-WRF-MEAN (Fig. 13, green line) is overestimated at the Pescara river stations, if compared with the BDD time series obtained by CHyM-OBS (Fig. 13, black line). Moreover, it results in a red BDD threshold exceedance at the Pescara Alanno station, that was actually affected by an orange threshold exceedance, only. As for the timing, there is a different behaviour between the northern basins and the central ones. For the northward catchments (Vomano and Tordino), the ensembles-modeled peak timing is progressively simulated up to 6 hours in advance (Fig. 13, a , b, c and d respectively green and black lines), with respect to the control hydrological simulation. At the Saline-Villa Carmine station, in the central area, the maximum of the BDD index is reproduced with high timing accuracy (Fig. 13, e, black and green lines), whereas an approximately 12 hours of delay at the Pescara river station (south area) is found (Fig. 13, f and g respectively; black and violet lines). The CHyM-HR-WRF input seems to be not affected by the aforementioned time shift. These results would suggest a contribution from the WRF-regional ensemble error, caused by the low resolution, in the timing of the maximum peak as it is found at several stations (Fig. 6) propagating in the CHyM forecasts.

Finally, all the time series show a variability among ensemble members much smaller than the difference between ensemble mean and observations at the maximum of the rainfall, because of the anticipation of the CHyM ensemble mean peak (Fig. 13, a , b, c and d). On the other hand, a larger variability among the members is found for Saline time series (10mm/h) where

the timing of the maximum is the same for both CHyM-OBS and CHyM-WRF-MEAN and the difference between the two is very small (3 mm/h, Fig. 13e). If the time lag would be set to zero by hypothetically shifting the CHyM-WRF-MEAN maximum at the right time for all time series, we would have the same variability of Saline at all the other stations, except for Vomano-Basciano. Though, the analysed time series (WRF and CHyM) are for different physical quantities a remark is necessary. The uncertainties for CHyM time series would suggest a good ensemble variability obtained by forcing CHyM with the WRF-ensemble members, albeit the reduced variability found for the WRF-regional ensemble time series (see par 3.2). A possible explanation may be found in the point-wise comparison between the models and the observations. Generally, the point-wise comparison of the precipitation between forecast and observation is penalising for the NWP forecast. In this case the WRF-regional ensemble is at 9km making the point-wise comparison even more penalising. This is not the case for the time series of a catchment which accounts for all the upstream flow making the comparison not point-wise.

6 Conclusions

On November 15, 2017 a severe hydrological event hit the Abruzzo region, causing damages with a high social and economic impact on human activities. This event is used to investigate the reliability of a meteo-hydrological ensemble chain. An operational and portable meteo-hydrological forecast system is implemented and tested at CETEMPS over the Abruzzo Region basins. The results of both the meteorological regional ensemble and the coupled meteo-hydrological ensemble chain are discussed and compared with the observations, Radar SRT and the results of a high resolution deterministic simulation.The meteorological ensemble correctly reproduce the signal of the event by catching the area of the maximum precipitation, a few days before the event. An overestimation of the maximum of the precipitation is found for the souther side of Abruzzo region. The statistical evaluation using ROC based on rain gauges and SRT supports this conclusion by showing a large AUC for the WRF-regional ensemble and a steep increase of the POD. Moreover, the comparison with the HR using ROC further confirm this conclusion. The meteo-hydrological ensemble chain results are discussed in terms of the hydrological BDD stress index able to identify catchment segments that are most likely to be stressed by weather extremes. The evaluation of the BDD index, as a user-oriented instrument to assess the flood risk over the Abruzzo Region drainage network, is carried out by comparing the occurrence of the index thresholds exceedance in the CHyM-OBS and the corresponding water level thresholds exceedance, at station level. The results show a good agreement. At this regard, it should be taken into account that the BDD index obtained from the CHyM-OBS is used as reference product for the BDD index itself and the meteo-hydro ensemble chain, because of the lack of discharge estimations and of updated rating curves. Moreover, the BDD thresholds are extendible to each grid-point of the drainage network and are used to produce a hydrological stress map over the whole spatial domain. The BDD maps produced by several CHyM simulation initialized using different WRF output are also compared with the hydrogeological criticality bulletin released during the event, in order to emphasize and confirm the spatial coherence between hydrological control simulation and detection of actual flooded areas. A very good performance of the BDD index (for both maps and time series) is found using CHyM-WRF-MEAN and CHyM-HR-WRF. Besides the BDD index, a BDD probability index and the associated spread are built using the 21 members of CHyM-ENS. The index allows for estimating the probability of a flooding

events, which is not possible to estimate by both the deterministic forecast and the CHyM forecast forced using the ensemble mean. The comparison of the BDD_{prob} map with the BDD map produced by CHyM-OBS points out a good reliability of this index for this event by both correctly identify flooded river segments and producing a small spread in these areas.

Hence we can summarize the major findings as follows:

- the pseudo-hydro ensemble forecast, that is the hydrological model forced by the mean precipitation produced by the WRF-regional ensemble, well reproduces the alert map issued by the Civil protection agency during the event, conferring reliability to this tool;

- the CHyM ensemble composed by 21 members initialized using the 21 members of the WRF-regional ensemble produces the BBD probability maps which well agrees with the alert map issued by the Civil protection agency during the event slightly reducing the overestimation produced by the pseudo-ensemble especially in the southern side of the Abruzzo region;

- the CHyM simulation forced using the HR deterministic forecast further reduces the over/underestimations produced by the ensemble mean forecast, especially over the smallest basins.

An advantage in using an ensemble prediction chain, in terms of Decision Support System (DSS) efficiency, is the low computational cost if compared with a HR deterministic modeling chain. Even though high resolution meteorological model are useful to detect hydrological stress over the smallest basins, the WRF-regional ensemble is found to be useful for an early identification of flood risk at warning area level. It has to be pointed out, that the alert map was issued in the morning of Nov 15 (i.e. the same day of the event) by the Civil Protection Agency (CPA) by using initially CHyM-HR-WRF and then it was updated using the observations. Hence, the availability of a forecast well in advance (i.e. at least the day before of the event) would allow the CPA to issue an alert map the day before. This can be achieved by using an ensemble forecast which, though at lower resolution, produces both a forecast for a longer lead time than the deterministic, and information on the probability of the forecasted event. In a DSS perspective, it is important to assign the correct alarm state at warning area level in advance, rather than focusing on the single catchment. The CHyM-HR-WRF would be used to better define the single river segment or the single catchment area of the event and to update in the following day the map. Hence, HR-deterministic and probabilistic-ensemble-based approaches should be considered both because they provide complementary information. Indeed, the computationally cheaper ensemble approach allows for longer forecast periods, identifying days ahead potential damaging events, but HR-deterministic still provides the best information in the very short term, thanks to its improved representation of regional to local scale dynamics.

Finally, an attempt is made to estimate the uncertainties in the precipitation forecast and how their errors propagate in the hydrological model if an ensemble approach is adopted. To this purpose a comparison is made between the BBD index time series extracted at station level, computed using CHyM-OBS, CHyM-WRF-ENS precipitation and temperature, and using the discharge field average by CHyM-ENS. The results do not show differences but the CHyM ensemble spread reproduces a distribution different by the WRF-regional ensemble one. Indeed, a larger spread than for WRF-regional ensemble is found

for most of the stations either for the maximum and the minimum of the precipitation. Hence we suppose that the weather prediction uncertainty propagates into the hydrological model, but still producing some problems (e.g. timing of the maximum). Therefore, to further investigate the propagation of uncertainty into the hydrological model, the same probabilistic approach used for the meteorological model should be applied to the hydrological one. In a forthcoming work a sensitivity study to the hydrological model uncertainty that can be obtained by perturbing, for example, the geometry of the system, the model factors etc. will be performed.

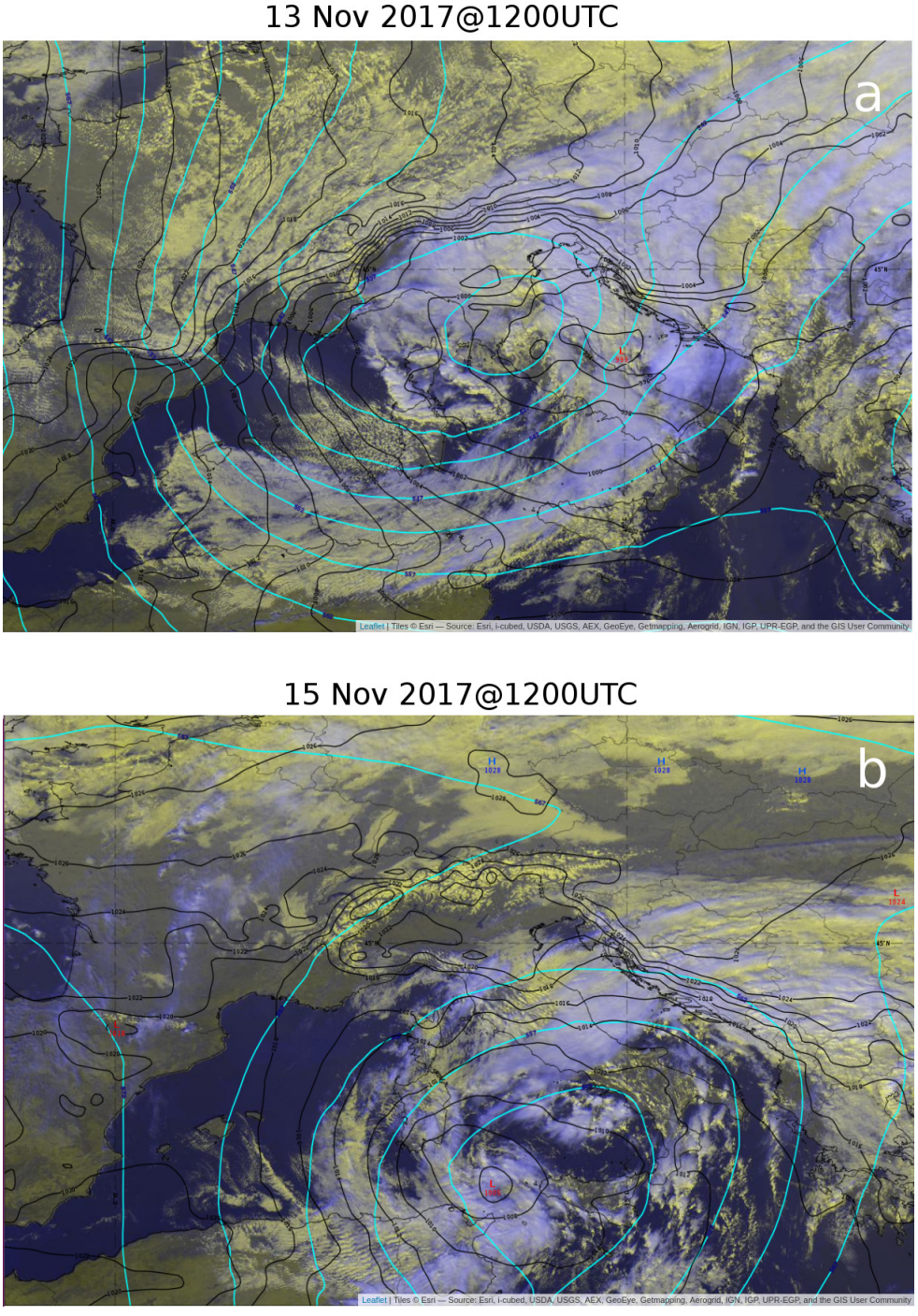

Figure 1. ECMWF analysis, Geopotential height at 500 hPa (cyan lines, contour lines=5dam) and mean sea level pressure (black lines labelled in hPa, contours lines=20hPa) and satellite WV (a) 13 Nov. 2017, 12UTC ; (b) 15 Nov. 2017, 12UTC . The maps have been retrieved from EUMeTrain/ePort archive (http://eumetrain.org/eport.html).

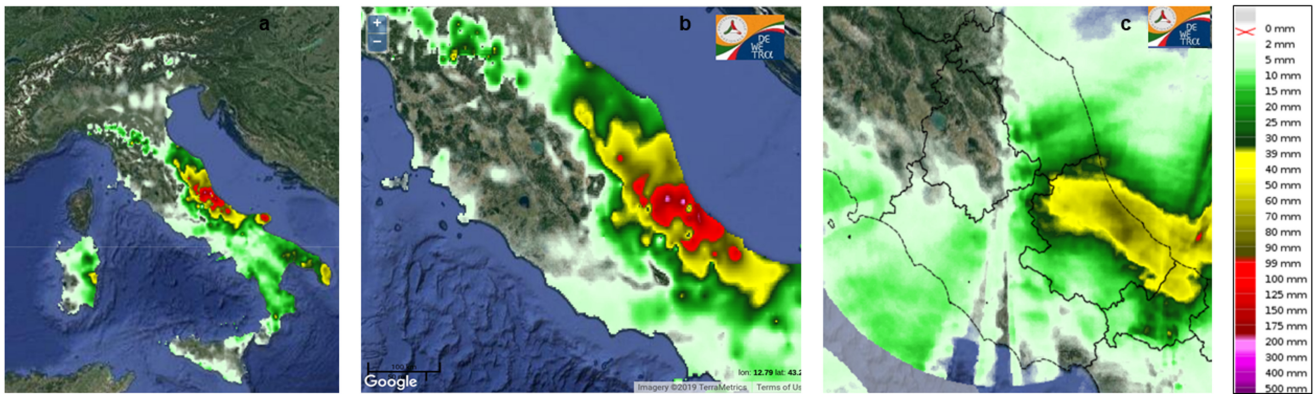

Figure 2. Observed accumulated precipitation over 24 hours starting on November 14, 2017 at 1200UTC : a) over Italy; b) over Central Italy; c) Radar Surface Rainfall Total (SRT). The daily rainfall maps and the radar data are provided by the DEWETRA Platform (Italian Civil Protection Department).

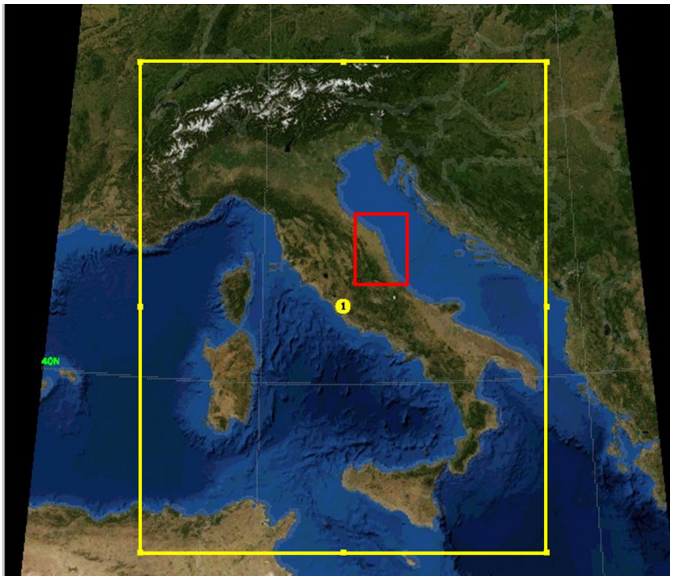

Figure 3. Domain for the WRF regional ensemble at 9km (yellow box). The red box indicates the area considered for the statical evaluation covering Marche and Abruzzo regions

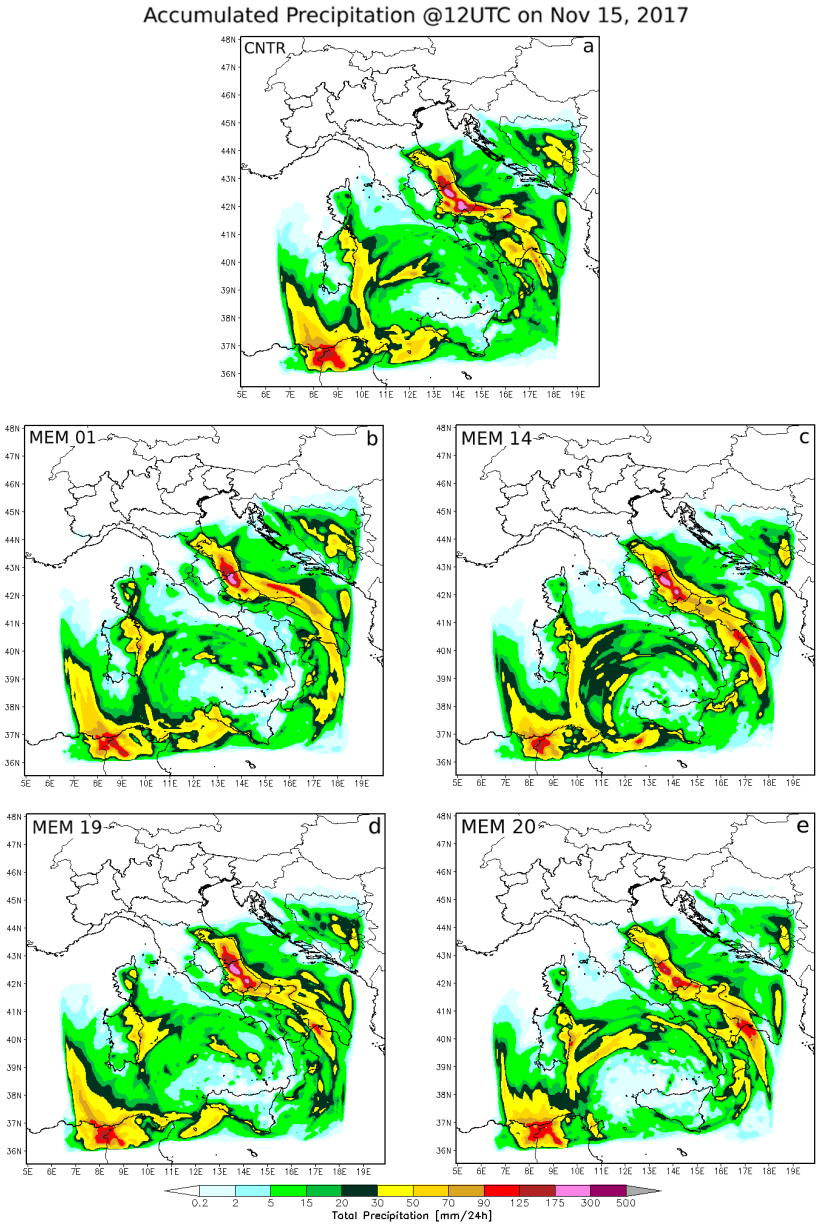

Figure 4. WRF accumulated precipitation over 24 hours produced by GFS initial conditions for : a) control (CNTR), and members b) 01; c)14; d) 19; e) 20.

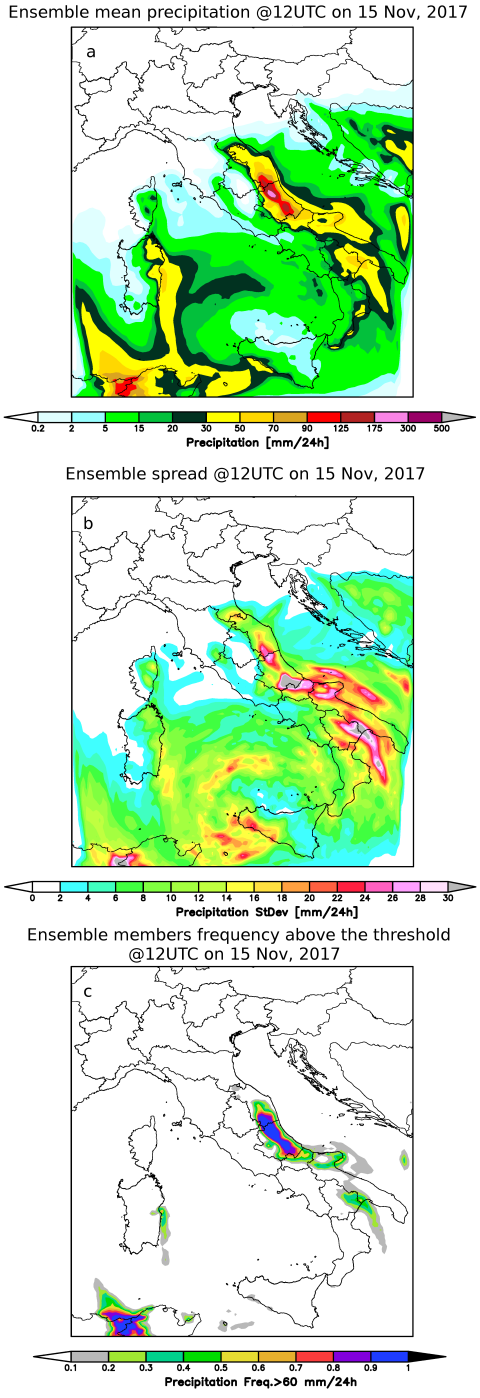

Figure 5. 24 hours accumulated precipitation at 12UTC on Nov, 15 2017: a) Ensemble Mean precipitation, b) Ensemble spread, c) Ensemble probability of precipitation above 60mm/24h produced by the members

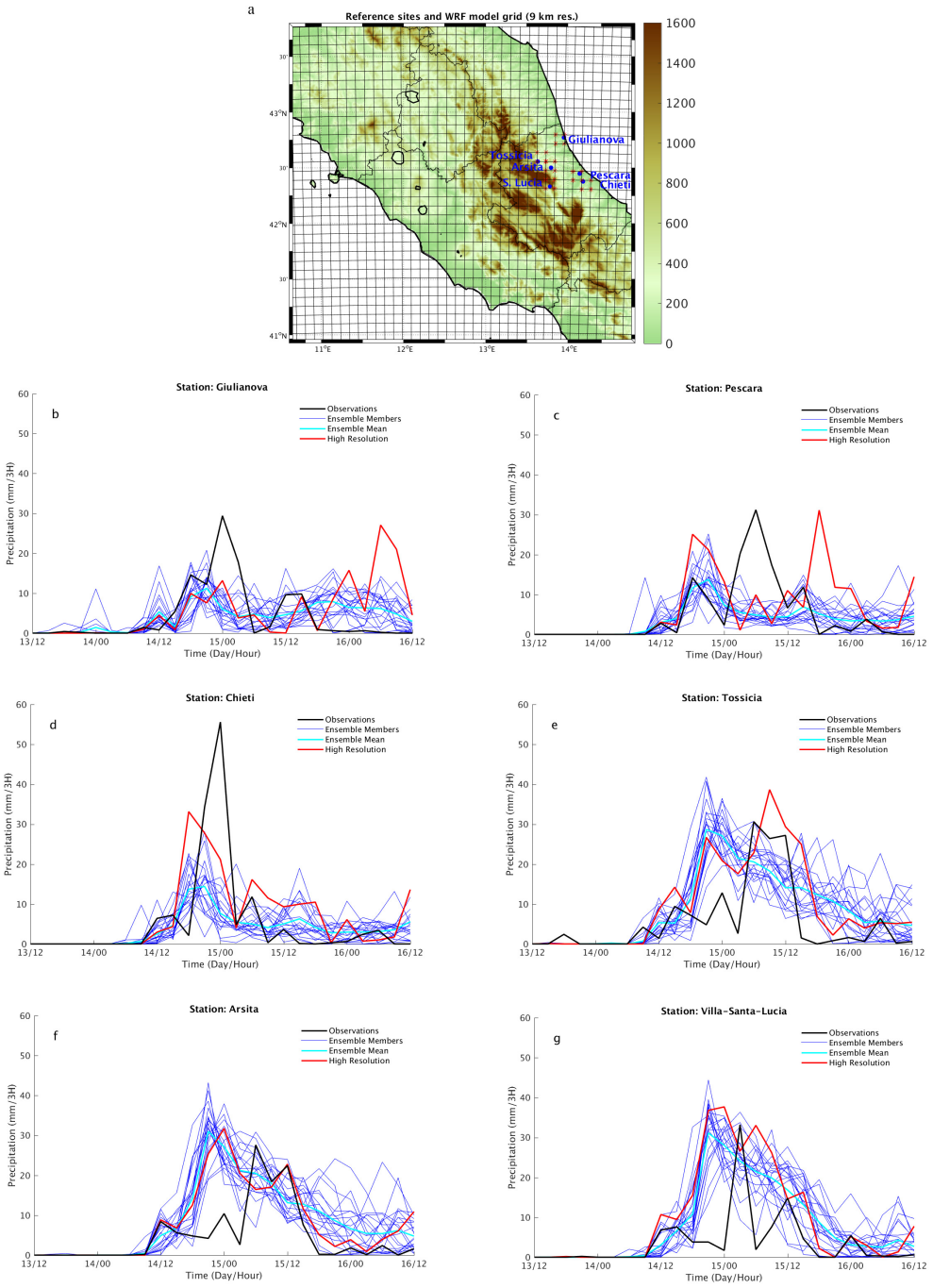

Figure 6. a) Location of the reference stations. Time series of precipitation at a few stations: b) Giulianova, c) Pescara , d) Chieti, e) Tossicia, f) Arsita, g) Villa-Santa Lucia for the ensemble mean (cyan line), ensemble members (blue lines), HR (red line) and observation (black line).

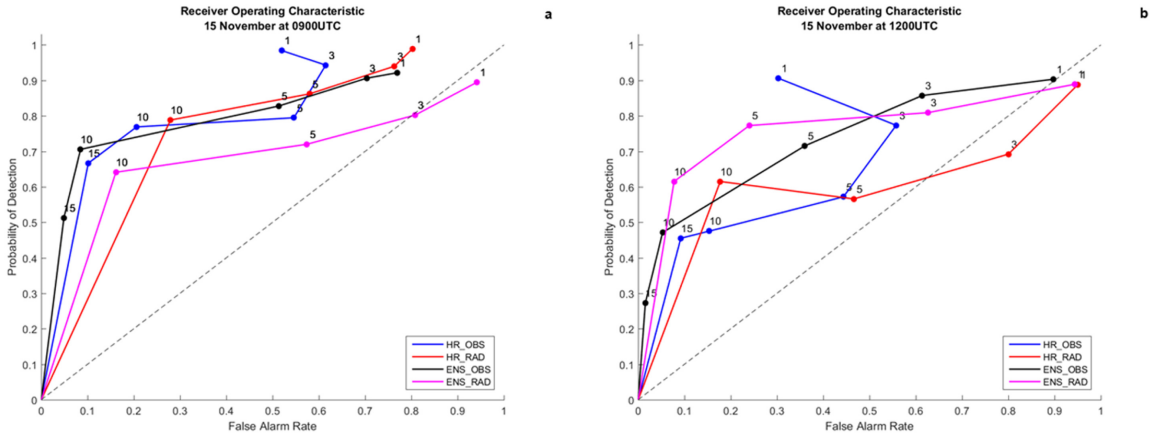

Figure 7. a) ROC plot using both rain gauge and Radar SRT for WRF-regional ensemble (black and magenta respectively) and WRF-HR (blue and red respectively) on November 15 at :a) 0900UTC; b) 1200UTC

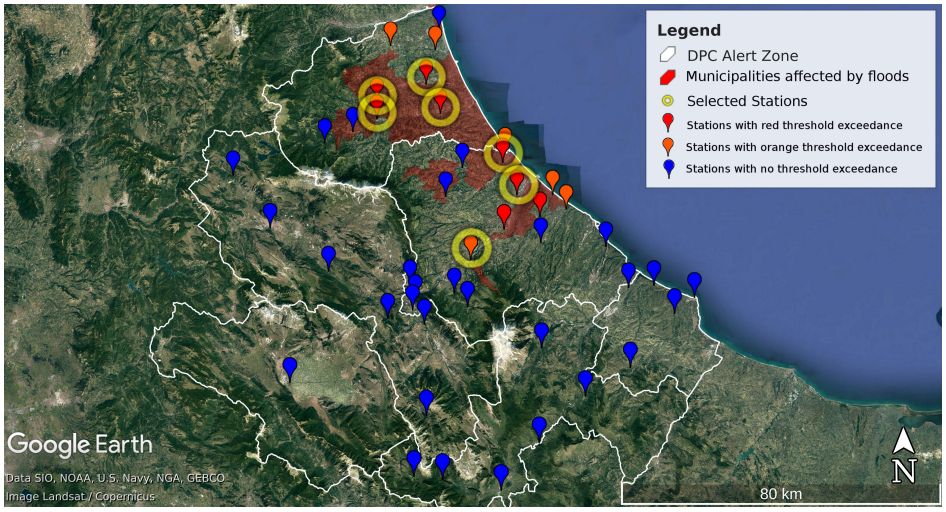

Figure 8. A map on Google Earth showing the geolocation of the hydrometric sensors over the Abruzzo Region (pinpoints) color code based on the warning threshold exceeded during the event. Yellow-circled sensors represents the subgroup where the analysis at station level is presented. The red areas correspond to the municipalities affected by flood.

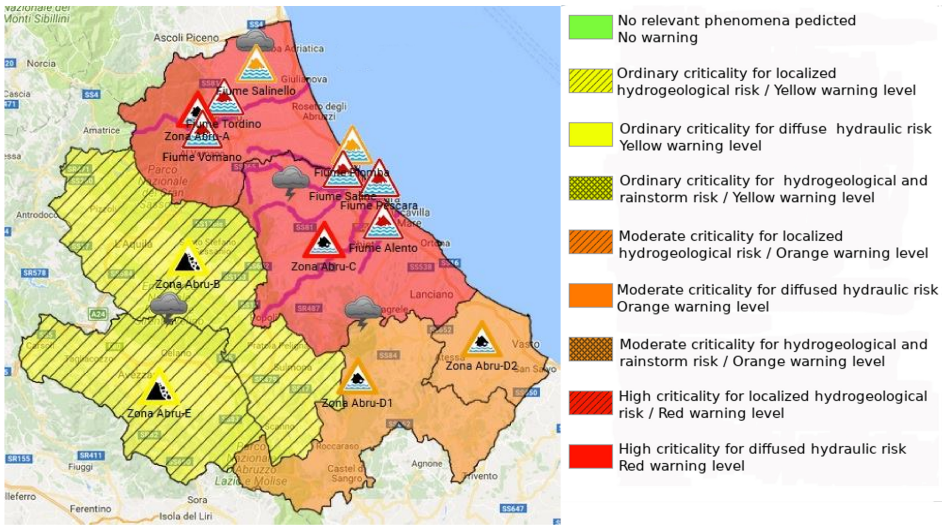

Figure 9. Hydrogeological Criticality Bullettin issued by the Civil Protection in the morning of 15 November 2017, where the Abruzzo region territory is divided into six warning areas (indicated with the prefix 'Zona Abru'), coloured according to the legend included. Triangle-shaped, thin-bounded signs are geolocated over relevant hydrometers and coloured according to the colour-code explained in figure 7, resulting from observed data. Triangle-shaped, thick-bounded signs indicates the forecasted warning level in the reference warning area. Yellow triangle indicates landslide risk at warning area level, while, thunderstorm risk is assigned through figurative icon. Purple lines on the coastal warning areas highlight river segments where the red threshold has been overpassed.

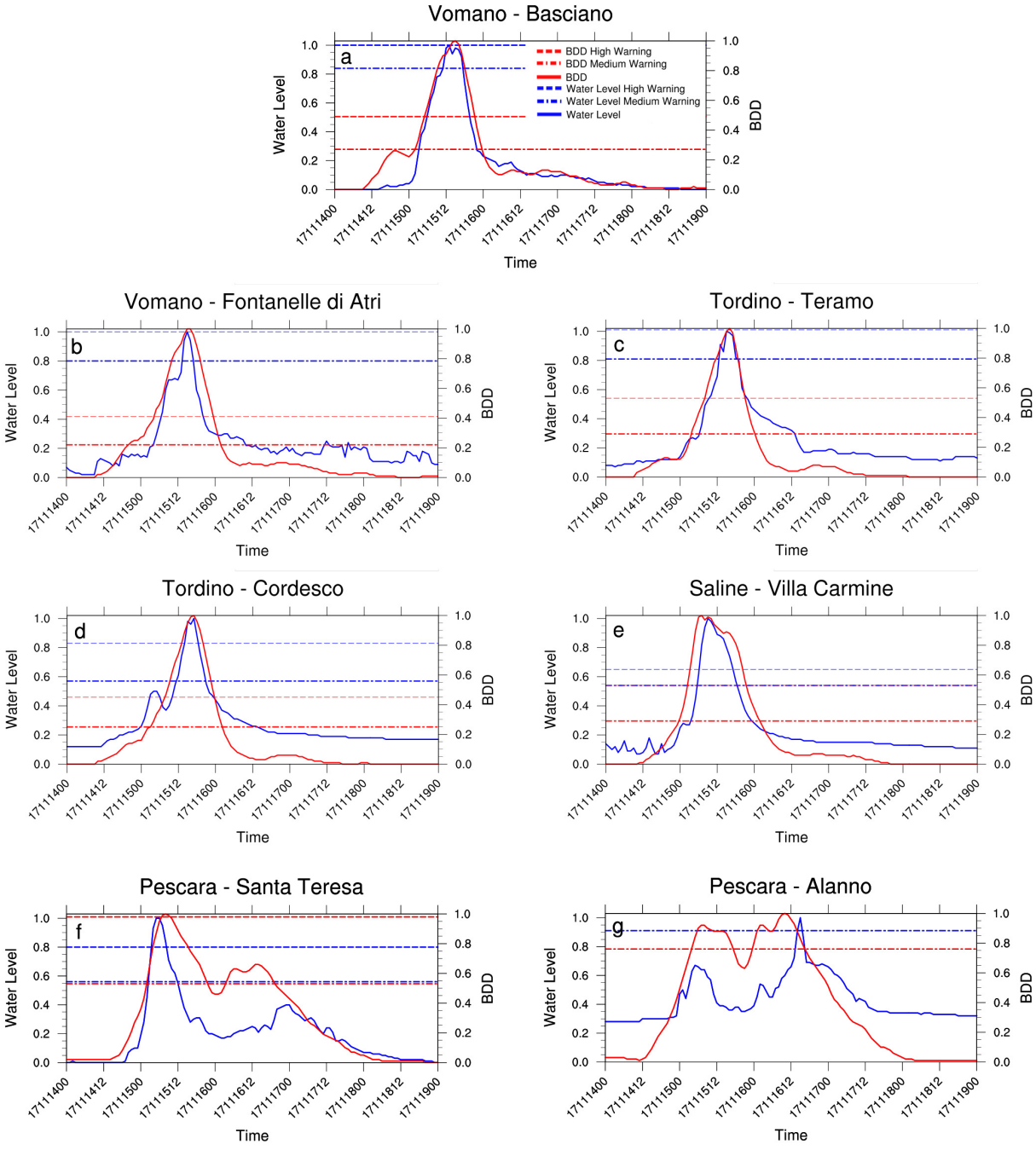

Figure 10. Normalised BDD time series and water level for catchments Vomano, Tordino, Saline and Pescara at stations: a) Basciano(Vomano); b) Fontanelle di Atri (Vomano); c) Teramo (Tordino); d) Cordesco (Tordino); e) Villa Carmine (Saline); f) Santa Teresa (Pescara) ; g) Alanno (Pescara); BDD index red lines, observed water level blue lines

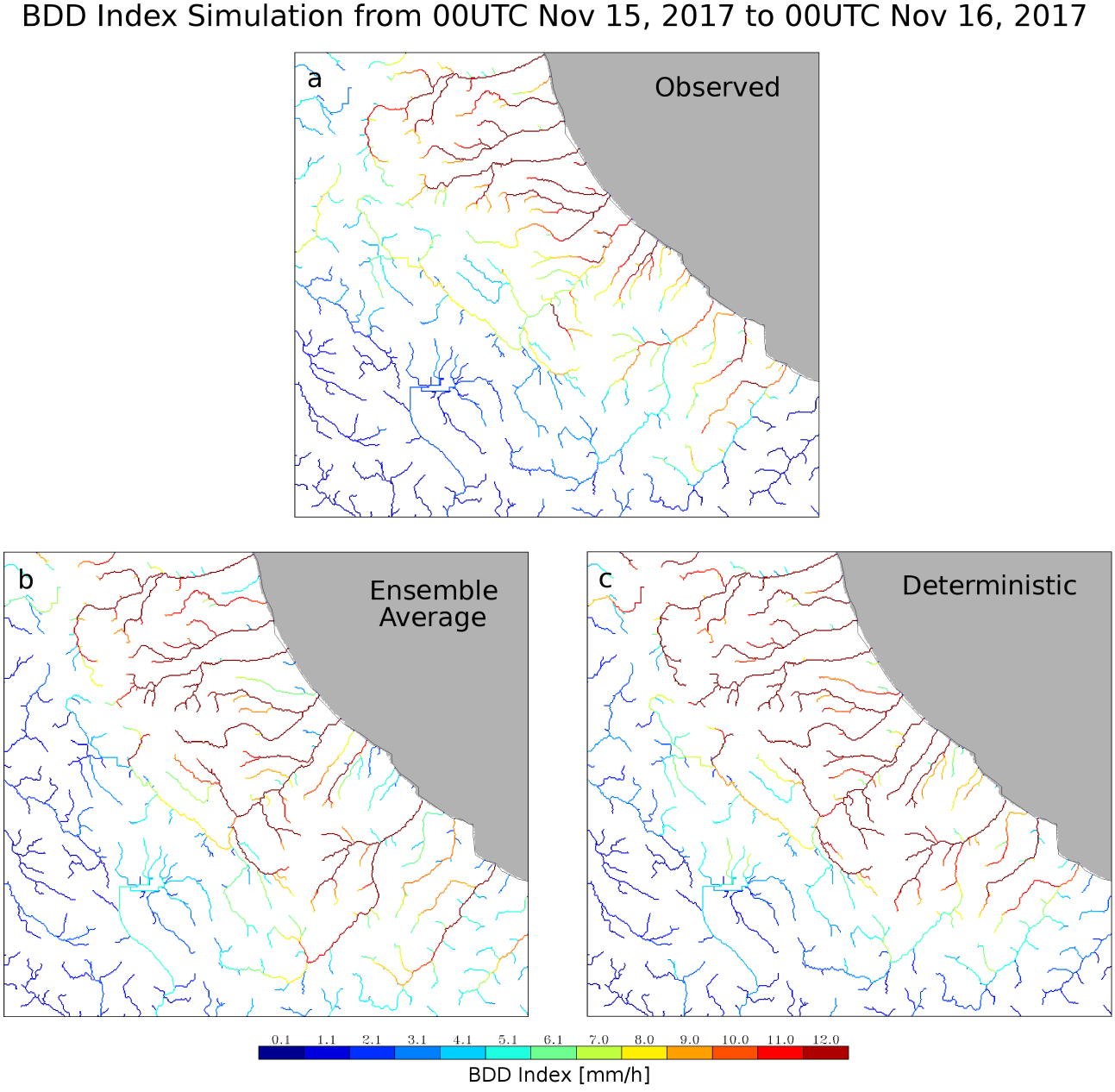

Figure 11. 24 hours BDD index computed by CHyM using: a) the observed accumulated precipitation; b) the mean precipitation produced by the WRF ensemble; c) the precipitation produced by the deterministic wrf @ 1km (HR)

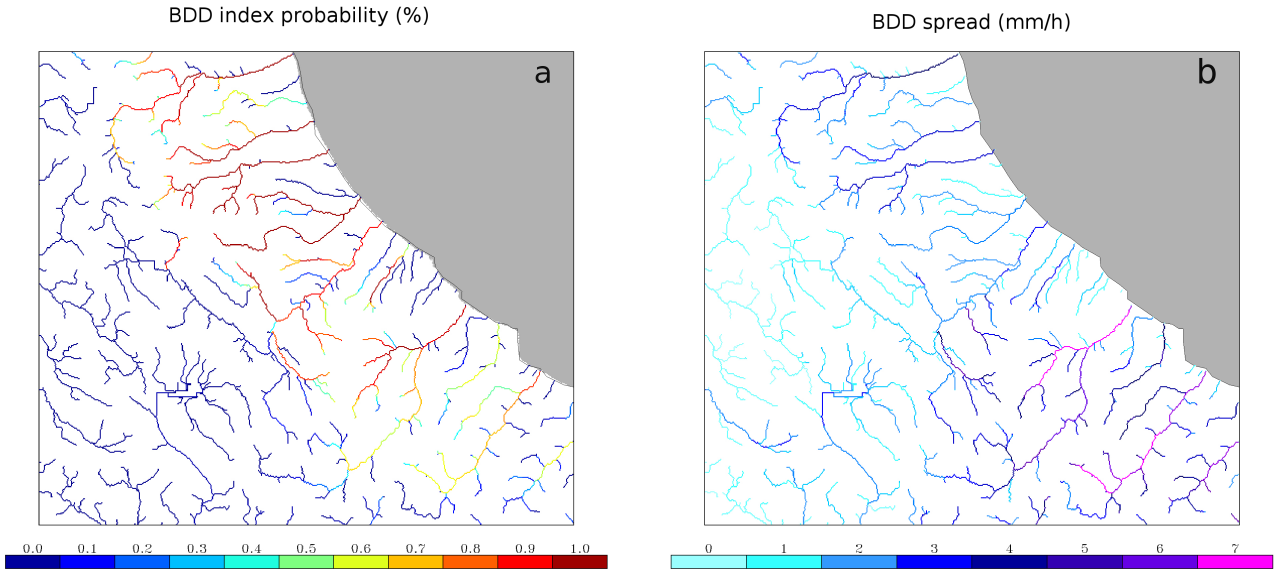

Figure 12. a) 24 hours BDD probability index computed forcing CHyM with the 21 WRF members; b) CHyM ensemble spread.

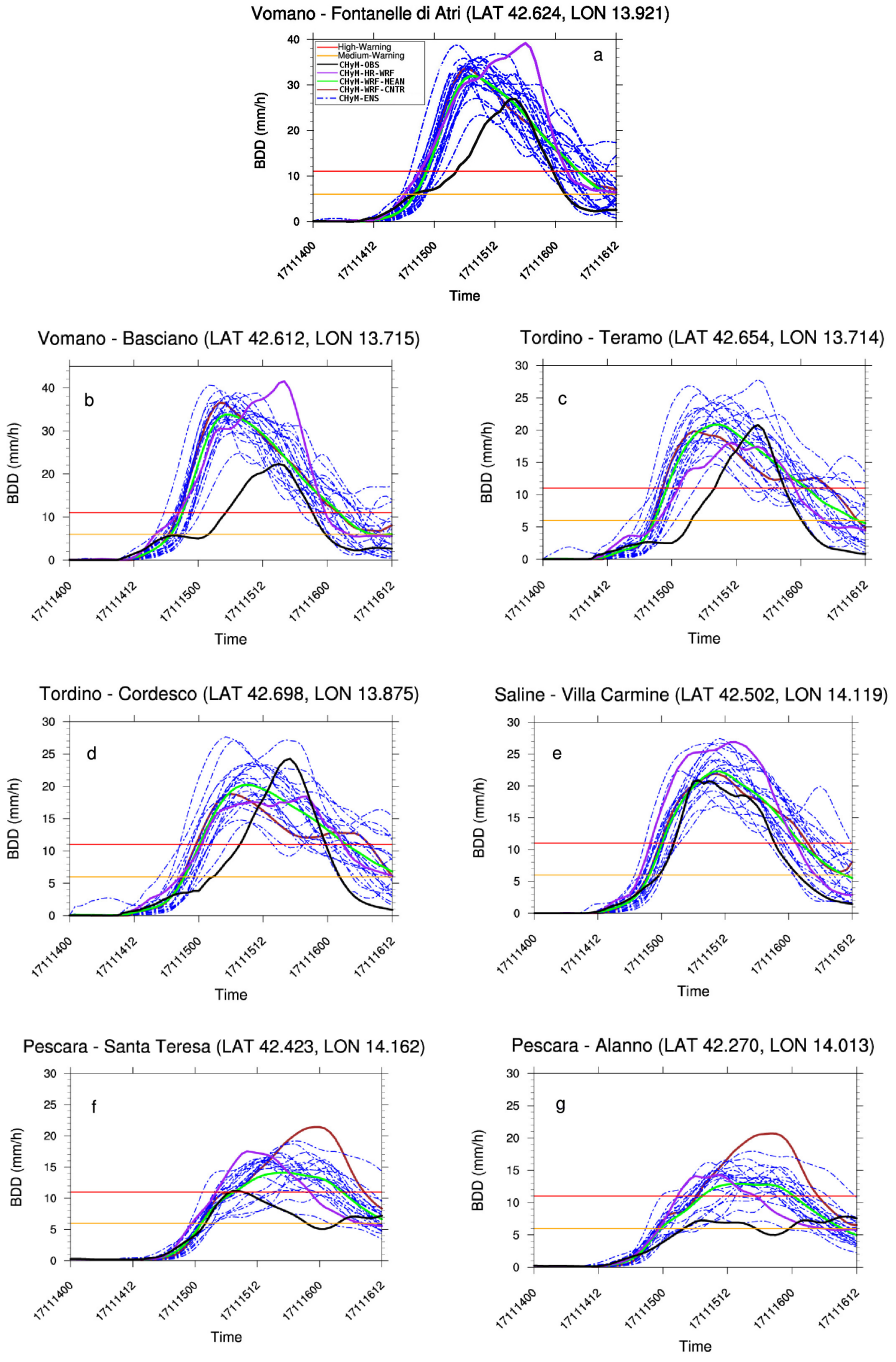

Figure 13. BDD time series for catchments Vomano, Tordino, Saline and Pescara at stations: a) Fontanelli di Atri (Vomano); b) Basciano (Vomano); c) Teramo (Tordino); d) Cordesco (Tordino); e) Villa Carmine (Saline); f) Santa Teresa (Pescara); g) Alanno (Pescara)

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

Acknowledgements. The authors acknowledge: the National Center for Atmospheric Research (NCAR) and the Center of Excellence in Telesensing of Environment and Model Prediction of Severe events (CETEMPS) for financial and computing resources; NCAR for WRF-ARW source code; the Italian Civil Protection department for rain-gauges data.

References

Abaza, M., Anctil, F., Fortin, V. and Luc Perreault: On the incidence of meteorological and hydrological processors: Effect of resolution, sharpness and reliability of hydrological ensemble forecasts. *Journal of Hydrology*, 555, 371-384, 2017. ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2017.10.038.

Addor, N., S. Jaun, F. Fundel, and M. Zappa: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.*, 15, 2327-2347, 2011.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger: GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, 17, 1161-1175, 2013. doi:10.5194/hess-17-1161-2013.

Bauer, P., Thorpe, A. and Brunet, G.,:The quiet revolution of numerical weather prediction. *Nature*, 525(7567), 47-55, 2015. doi: 10.1038/nature14956.

Buizza, R. and Palmer, T.N. (1998) Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126(9), 2503–2518.

Buizza R, Houtekamer P. L., Toth Z., Pellerin G., Wei M., Zhu Y.: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction system. *Mon. Wea. Rev.*, 133, 1076-1097, 2005.

Calvetti, L. and Pereira Filho, A.J.,: Ensemble hydrometeorological forecasts using WRF hourly QPF and top model for a middle watershed. *Advances in Meteorology*, ID 484120, 12 pag., 2014. http://dx.doi.org/10.1155/2014/484120.

Cloke, H.L. and Pappenberger, F.,: Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3-4), 613-626, 2009.

Coppola E., B. Tomassetti, L. Mariotti, M. Verdecchia and G. Visconti, 2007. Cellular automata algorithms for drainage network extraction and rainfall data assimilation, *Hydrol. Sci. J.*, 52(3), pp. 579-592 .

Davolio, S., Miglietta, M.M., Diomede, T., Marsigli, C., Morgillo, A. and Moscatello, A.,:A meteo-hydrological prediction system based on a multi-model approach for precipitation forecasting. *Natural Hazards and Earth System Science*, 8(1), pp.143-159, 2008.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D-J., Hartman, R., Herr, H.D., Fresch, M., Schaake, J. and Zhu, Y., (2014). The science of NOAA's operational hydrologic ensemble forecast service. *Bull Am Meteorol Soc* 95, 79-98, 2014. doi: 10.1175/BAMS-D-12-00081.1

Drobinski, P., Silva, N.D., Panthou, G. et al.,: Scaling precipitation extremes with temperature in the Mediterranean: past climate assessment and projection in anthropogenic scenarios. *Clim. Dyn.*, 51, 1237-1257, 2018. https://doi.org/10.1007/s00382-016-3083-x

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, 46, 3077-3107,1989.

Duffourg, F. and Ducrocq, V., : Origin of the moisture feeding the Heavy Precipitating Systems over Southeastern France, *Nat. Hazards Earth Syst. Sci.*, 11, 1163-1178, 2011. doi:10.5194/nhess-11-1163-2011.

European Environment Agency, 2010. Mapping the impacts of natural hazards and technological accidents in Europe: an overview of the last decade. *EEA Technical Report* No 13/2010, ISBN: 978-92-9213-168-5.

European Flood Directive 2007: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32007L0060

Fan, F.M., Collischonn, W., Quiroz, K.J., Sorribas, M.V., Buarque, D.C. and V.A., Siqueira,: Flood forecasting on the Tocatins River using ensemble rainfall forecasts and real-time satellite data estimates., *Journal of Flood Risk Management,* 9, 278-288, 2015. doi: 10.1111/jfr3.12177

Ferretti R., E. Pichelli, S. Gentile, I. Maiello, D. Cimini, S. Davolio, M. M. Miglietta, G. Panegrossi, L. Baldini, F. Pasi, F. S. Marzano, A. Zinzi, S. Mariani, M. Casaioli, G. Bartolini, N. Loglisci, A. Montani, C. Marsigli, A. Manzato, A. Pucillo, M. E. Ferrario, V. Colaiuda,

and R. Rotunno: Overview of the first HyMeX Special Observation Period over Italy: observations and model results. *Hydr. Earth Syst. Sci.*, 18, 1953-1977, 2014, doi:10.5194/hess-18-1953-2014.

Giorgi, F., E.-S. Im, E. Coppola, N. S. Diffenbaugh, X. J. Gao, L. Mariotti, and Y. Shi,: Higher Hydroclimatic Intensity with Global Warming. *J. Clim.*, 24, 5309-5324, 2011. doi:10.1175/2011JCLI3979.1. http://journals.ametsoc.org/doi/abs/10.1175/2011JCLI3979.1

Hally A., Caumont O., Garrote L., Richard E., Weerts A., Delogu F., Fiori E., Rebora N., Parodi A., Mihalovic A., Ivkovic M., Dekic L., van Verseveld W., Nuissier O., Ducrocp V., D'Agostino D., Galizia A., Danovaro E, Clematis A.: . Hydrometeorological multi-model ensemble simulations of the 4 November 2011 flash flood event in Genoa, Italy, in the framework of the DRIHM project. *Nat. Hazard Earth Syst. Sci.*, 15, 537-555, 2015.

Hamill, T., Clark, M., Schaake, J., and Buizza, R.,: Second HEPEX Workshop Summary Report, Boulder, Colorado, 2005. :http://hydis8.eng.uci.edu/hepex/scndwksp/HEPEX05-Summary.pdf.

HEPEX: https://hepex.irstea.fr

Hong S.-Y. and J.-O.J. Lim,: The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *J. Korean Meteor. Soc.*, 42, 129-151, 2006.

IRPI-Istituto di Ricerca per la Protezione Idrogeologica Consiglio Nazionale delle Ricerche: Rapporto periodico sul Rischio posto alla Popolazione italiana da Frane e Inondazioni. *Annuale 2018*. IRPI, Jan, 2019.

ISPRA-Istituto Superiore per la Protezione e la Ricerca Ambientale: Dissesto idrogeologico in Italia: pericolositá e indicatori di rischio. *ISPRA, Report* no. 287/2018, ISBN: 978-88-448-0901-0

Jaun, S., Ahrens, B., Walser, A., Ewen, T., Schar, C.: A probabilistic view on the August 2005 floods in the upper Rhine catchment. *Natural Hazards Earth System Sciences* 8, 281-291, 2008.

Janjic, Z. I.: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure scheme. *Mon. Wea. Rev.*, 122, 927-945,1994.

Jolliffe, I.T. and Stephenson, D.B: Forecast Verification: A Practitioner's Guide in Atmospheric Science. *Chichester: John Wiley & Sons*, 2012.

Kain, J.S.: The Kain-Fritsch convective parameterization: an update. *J. Appl. Meteorol.*, 43, 170-181, 2004.

Lighthill M. J.and F.R.S., G.B. Whitam,: On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. R.Met. Soc. A.*, 317-345, 1955. DOI: 10.1098/rspa.1955.0089

Marchi, L.,Borga, M., Preciso, E. and Gaume, E.: Characterisation of Selected Extreme Flash Floods in Europe and Implications for Flood Risk Management. *Journal of Hydrology.* 394. 118-133, 2010. doi:10.1016/j.jhydrol.2010.07.017.

Marsigli C., F. Boccanera, A. Montani, T. Paccagnella. The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics*, European Geosciences Union (EGU), 12 (4), 527-536, 2005.

McCollor, D. and Stull, R.: Hydrometeorological short-range ensemble forecasts in complex terrain. Part I: Meteorological evaluation. *Weather and Forecasting*, 23(4), 533-556, 2008.

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-model for the longwave. *J. Geophys. Res.*, 102, 16663-16682, 1997.

Niu G-Y, Yang Z-L, Mitchell KE, Chen F, Ek MB, Barlage M, Longuevergne L, Kumar A, Manning K, Niyogi D, Rosero E, Tewari M, Xia Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local scale measurements. *J. Geophys. Res.*, 116, 2011. D12109. http://dx.doi.org/10.1029/2010JD015139.

Pappenberger, F., Beven, K. J., Hunter, P. D., Bates, B. T., Gouweleeuw, J., Thielen, J., and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), *Hydrol. Earth Syst. Sci.*, 9, 381-393, 2005. http://www.hydrol-earth-syst-sci.net/9/381/2005/.

Pappenberger, F., K. Scipal R. Buizza: Hydrological aspects of meteorological verification.*Atmos. Sci. Let.* 9, 43?52, 2008.

Penning-Rowsell, E., Tunstall, S., Tapsell, S., Parker, D.: The benefits of flood warnings: real but elusive, and politically significant. *Journal of the Chartered Institution of Water and Environmental Management* 14, 7-14, 2009.

Pichelli E.. R. Rotunno, Ferretti R.: Effects of the Alps and Apennines on forecasts for Po Valley convection in two HyMeX cases. *Q. J. R. Meteorol. Soc.*,143, 2420-2435, 2017. DOI: 10.1002/qj.3096.2017

Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., Pignone, F., Siccardi, F., Silvestro, F., Tanelli, S., and Parodi, A.: Extreme rainfall in the Mediterranean: what can we learn from observations?, *J. Hydrometeorol.*, 14, 906-922, 2013.

Rotunno R. and R. Ferretti: Orographyc analysis of rainfall in MAP cases IOP2B and IOP8. *Q. J. R. Meteorol. Soc.*, 129B, 373-390, 2003, IF 2.977

Rotunno, R. and Houze, R. A.: Lessons on orographic precipitation from the Mesoscale Alpine Programme, *Q. J. R. Meteorol. Soc.*, 133, 811-830, 2007.

Saleh, F., Ramaswamy, V., Georgas, N., Blumberg, A.F. and Pullen, J.: A retrospective streamflow ensemble forecast for an extreme hydrologic event: a case study of Hurricane Irene and on the Hudson River basin. *Hydrology and Earth System Sciences*, 20(7), 2649-2667, 2016.

Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark : HEPEX: The Hydrological Ensemble Prediction Experiment. *Bull. Amer. Meteor. Soc.*, 88, 1541-1547, 2007. doi:https://doi.org/10.1175/BAMS-88-10-1541.

Scoccimarro, E., Villarini, G., Vichi, M., Zampieri, M., Fogli, P.G., Bellucci, A. and Gualdi, S.: Projected changes in intense precipitation over Europe at the daily and subdaily time scales. *J. of climate*, 28(15), 6193-6203, 2015.

Singh V.P., Frevert D.K. : Mathematical models of small watershed hydrology and application. *Water ResourcePublications*, LLC, Highlands Ranch, Colorado, USA, 2002.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, Powers JG.: *'A Description of the Advanced Research WRF Version 3'*, NCAR Technical Note NCAR/TN-475+STR, 125 pp, 2008. NCAR: Boulder, CO, USA.

Storer, N.L., Gill, P.G. and P.D., Williams: Multi-Model ensemble predictions of aviation turbulence. *Meteorol Appl.*, 26, 416-428,2019. DOI: 10.1002/met.1772

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, 13, 125-140, 2009. https://doi.org/10.5194/hess-13-125-2009,

Trenberth, K.E., Dai, A., Rasmussen, R.M. and Parsons, D.B.: The changing character of precipitation. *BAMS*, 84(9), 1205-1218, 2003.

Tomassetti B., E. Coppola, M. Verdecchia, and G. Visconti: Coupling a distributed grid based hydrological model and MM5 meteorological model for flooding alert mapping, *Adv. Geosci.*, 2, 59-63, 2005.

Van den Besselaar, E.J.M., Klein Tank, A.M.G. and Buishand, T.A.: Trends in European precipitation extremes over 1951-2010. *Int. J. of Climatology*,, 33(12), 2682-2689, 2013.

Verdecchia M., E. Coppola, C. Faccani, R. Ferretti, A. Memmo, M. Montopoli, G. Rivolta, T. Paolucci, E. Picciotti, A. Santacasa, B. Tomassetti, G. Visconti and F. S. Marzano, Flood forecast in complex orography coupling distributed hydrometeorological models and in-situ and remote sensing data. *Meteorol. Atmos. Phys.*, 101, 267-285, 2008.

Vie B., O. Nuissier, V. Ducrocq: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, 139, 403-423, 2011.

Wanders, N. and Wood, E.F.: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environ. Res. Lett.*, 11, 2016. 094007

Willett, K.M., Jones, P.D., Gillett, N.P. and Thorne, P.W.: Recent changes in surface humidity: Development of the HadCRUH dataset. *J. of Climate*, 21(20), 5364-5383, 2008.

Wolfram, S.: A new kind of science. *Wolfram Media*, 1197 pag., 2002.