



Potentialities of ensemble strategies for flood forecasting over the Milano urban area



Giovanni Ravazzani^{a,*}, Arnau Amengual^b, Alessandro Ceppi^a, Víctor Homar^b, Romu Romero^b, Gabriele Lombardi^a, Marco Mancini^a

^a Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

^b Grup de Meteorologia, Departament de Física, Universitat de les Illes Balears, Palma, Mallorca, Spain

ARTICLE INFO

Article history:

Received 30 January 2016

Received in revised form 29 April 2016

Accepted 9 May 2016

Available online 20 May 2016

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Isabelle Braud, Associate Editor

Keywords:

Flooding

Small urban river basins

Land-use change

Convective events

Ensemble prediction systems

SUMMARY

Analysis of ensemble forecasting strategies, which can provide a tangible backing for flood early warning procedures and mitigation measures over the Mediterranean region, is one of the fundamental motivations of the international HyMeX programme. Here, we examine two severe hydrometeorological episodes that affected the Milano urban area and for which the complex flood protection system of the city did not completely succeed. Indeed, flood damage have exponentially increased during the last 60 years, due to industrial and urban developments. Thus, the improvement of the Milano flood control system needs a synergism between structural and non-structural approaches. First, we examine how land-use changes due to urban development have altered the hydrological response to intense rainfalls. Second, we test a flood forecasting system which comprises the Flash-flood Event-based Spatially distributed rainfall–runoff Transformation, including Water Balance (FEST-WB) and the Weather Research and Forecasting (WRF) models. Accurate forecasts of deep moist convection and extreme precipitation are difficult to be predicted due to uncertainties arising from the numeric weather prediction (NWP) physical parameterizations and high sensitivity to misrepresentation of the atmospheric state; however, two hydrological ensemble prediction systems (HEPS) have been designed to explicitly cope with uncertainties in the initial and lateral boundary conditions (IC/LBCs) and physical parameterizations of the NWP model. No substantial differences in skill have been found between both ensemble strategies when considering an enhanced diversity of IC/LBCs for the perturbed initial conditions ensemble. Furthermore, no additional benefits have been found by considering more frequent LBCs in a mixed physics ensemble, as ensemble spread seems to be reduced. These findings could help to design the most appropriate ensemble strategies before these hydrometeorological extremes, given the computational cost of running such advanced HEPSs for operational purposes.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The number of great natural catastrophes is increasing worldwide. Among these, flood is the worst hazard causing thousands losses of life and damage to property (Munich Re, 2014). The HyMeX (HYdrological cycle in the Mediterranean EXperiment, <http://www.hymex.org>) program is an international effort aimed at advancing in the scientific knowledge of the water cycle variability from a seamless approach. A better understanding, modelling and forecasting of hydrometeorological extremes over the flood-prone Western Mediterranean region is one of the

milestones of HyMeX (Drobinski et al., 2014). To reduce flood losses, real-time flood forecasting systems based on coupling meteorological and hydrological models are synergic to structural measures by issuing warnings in advance (Amengual et al., 2007, 2015; Rabuffetti et al., 2008; Ceppi et al., 2013). Despite the widespread use of real-time flood forecasting systems and the great steps that have been taken by hydrometeorologists, many open issues still remain when dealing with flash floods affecting urban areas and catchments of small dimensions (Silvestro et al., 2015).

One major issue concerns the land-use change due to urban developments that alters the basin response to intense precipitation. With less storage capacity and more rapid runoff, urban river basins rise more quickly during storms and have higher peak discharge rates than rural catchments. In addition, developments along stream channels and floodplains can alter the capacity of a

* Corresponding author. Tel.: +39 02 2399 6231.

E-mail address: giovanni.ravazzani@polimi.it (G. Ravazzani).

canal to convey water and can increase the height of water surface corresponding to a given discharge.

A further issue when dealing with urban flash floods is the accuracy of Quantitative Precipitation Forecasts (QPFs). Major floods at small drainage areas are mostly generated by convective systems characterized by heavy precipitation and short duration. The use of high-resolution Numerical Weather Prediction (NWP) models is effective to capture the triggering and subsequent evolution of the convectively-driven precipitation systems that are directly linked to small-scale dynamics and are highly sensitive to local topographic features (Leoncini et al., 2013; Fiori et al., 2014). Nowadays, convection-permitting models are suitable for the spatial and temporal scales of small- and medium-size flood-prone basins. QPFs can be directly used to drive rainfall–runoff models without the need of implementing additional downscaling procedures (Amengual et al., 2008; Addor et al., 2011; Vincendon et al., 2011).

Nevertheless, accurate numerical simulation of deep moist convection and extreme precipitation is difficult owing to: (i) complexity, uncertain definition and highly nonlinear character (also affecting their interrelations) of the physical parameterization schemes used in NWP models; and (ii) its sensitivity to any misrepresentation of the initial atmospheric state or boundary forcing across the relevant convective and meso-scales. Indeed, these errors can grow rapidly during the forecast horizon, strongly penalizing the quality of the nonlinear system forecast (Mullen and Baumhelfner, 1988; Toth and Kalnay, 1993; Houtekamer and Derome, 1995; Du et al., 1997). Thus, QPFs uncertainties arise from both the initial and lateral boundary conditions (IC/LBCs) and from physical parameterizations of the NWP model. When dealing with flood risk in small-size catchments, the provision of correct spatial and temporal QPF distributions is paramount, as small errors can relatively result in misleading Quantitative Discharge Forecasts (QDFs), preventing the issuance of precise and dependable early flood warnings (Le Lay and Saulnier, 2007; Amengual et al., 2007, 2009; Bartholmes et al., 2009; Cloke et al., 2013).

Ensemble prediction systems (EPSs) aim at forecasting the set of plausible outcomes and accounting for the most relevant uncertainties in the forecasting system. Uncertainties in the initial and boundary fields can be encompassed by conveniently perturbing IC/LBCs (Buizza, 2003; Gritmit and Mass, 2007; Hsiao et al., 2013). NWP model errors for QPFs arise mainly from the imperfect representation of convection, planetary boundary layer (PBL) processes, land physics and moist microphysical processes (Stensrud et al., 2000; Jankov et al., 2005; Tapiador et al., 2012). Uncertainties in model parameterizations are coped by populating the ensemble with multiple combinations of equally-skilful physical schemes.

When driving hydrological models for flood forecasting purposes, EPSs can be used to convey these external-scale uncertainties and to construct hydrological ensemble prediction systems (HEPSs). The inclusion of these external-scale uncertainties aims at improving the skill and spread of the HEPs by introducing independent information of all plausible atmospheric states. However, the most suitable methods for generating HEPs and the quantification of their added value are still under investigation (Cloke and Pappenberger, 2009; Cloke et al., 2013).

Alternative statistical methods to explore meteorological forecast uncertainty require numerous samples that satisfy the assumption of historic precipitation events sharing the same statistical characteristics with the precipitation event being forecast (Lee et al., 2013). In the current paper, we decide to explore the predictive skill of a perturbed IC/LBCs scheme (PIBL) and a multiple physical scheme (MPS) as ensemble strategies for short-range flood forecasting. To this end, we examine separately the impact of the IC/LBCs and NWP model errors into the skill of QPFs and QDFs for two of the most critical events of the last 20 years

affecting the Milano urban area, northern Italy. We assess the skill of each ensemble strategy to predict the exceeding of a given threshold through the physically-based distributed hydrological FEST-WB model (Rabuffetti et al., 2008). Performance of physically-based distributed approach is generally shown superior than a lumped one in the case of extreme flood events (Carpenter and Georgakakos, 2006; Moore et al., 2006) and can be suitable to issue public warnings with a threshold based methodology (Reed et al., 2007).

An understanding of how these distinct EPS strategies perform and how the external-scale uncertainties propagate into the HEPs is crucial for an optimal design of an operational hydro-meteorological forecasting system in the area. The rest of the paper is structured as follows: Section 2 consists of a brief description of the study area, land use change, and study cases; Sections 3 and 4 describe the hydrological and meteorological tools; Section 5 discusses the results; and Section 6 provides an assessment of the methods used, including further remarks.

2. Data and case studies

2.1. Study area

Milano is one of the most populous cities in Italy (1,316,000 inhabitants live in 182 km²), and is also one of its most important economic areas. A large region from the Italian PreAlps drains to Milano (Fig. 1). The main rivers are the Lambro (area of 500 km²), Seveso (area of 207 km²), and Olona (area of 208 km²), plus a number of minor tributaries for a total drainage surface of about 1300 km².

In the past, the Milano urban area has been subjected to a high flood hazard and, in fact, during the 1970s a series of risk mitigation works were carried out with the aim of reducing the exceeding discharges flowing through the urban areas. The main work consisted in the construction of a bypass channel (CSNO, acronym from Italian “Canale Scolmatore di Nord Ovest”) with a maximum capacity of 30 m³/s, which collects the excess of discharge from the Seveso River, preventing their entry into the city. In the upper Lambro River Basin, the regulated Pusiano Lake acts as a storage basin with respect to flood events, while more recently, in 2010 an on-stream detention basin (Ponte Gurone dam) was built on the Olona River near Varese city, just upstream the hydrometric station (Section 1 in Fig. 1). The maximum storage capacity of this reservoir is 1,520,000 m³ accounting for a total drainage area of 3.83 km². The Ponte Gurone dam is regulated through three automatic gates to keep a released rate below 36 m³/s, which is considered as the maximum allowable discharge for downstream locations. The complex existing structural engineering works alter the natural response of river basins and make calibration of hydrological models more difficult.

The complex flood protection system of the city did not completely succeed in the recent years; hence, the implementation of a hydro-meteorological chain can provide an additional support as a non-structural method for early warning systems. In fact, since lag times of these basins are a few hours (Table 1), alerts with sufficient lead time permit civil protection authorities and the public to exercise caution and take preventive measures to mitigate the impacts of flooding (Yang et al., 2015). In this study, we analyse 48 h forecasts initialized one day before the observed peak flood, as this lead time is considered sufficient and adequate by local authorities.

Available meteorological data are: precipitation and temperature, collected hourly by the telemetric monitoring system managed by the ARPA (Regional Agency for Environmental Protection) of the Lombardy region since 2003 and the

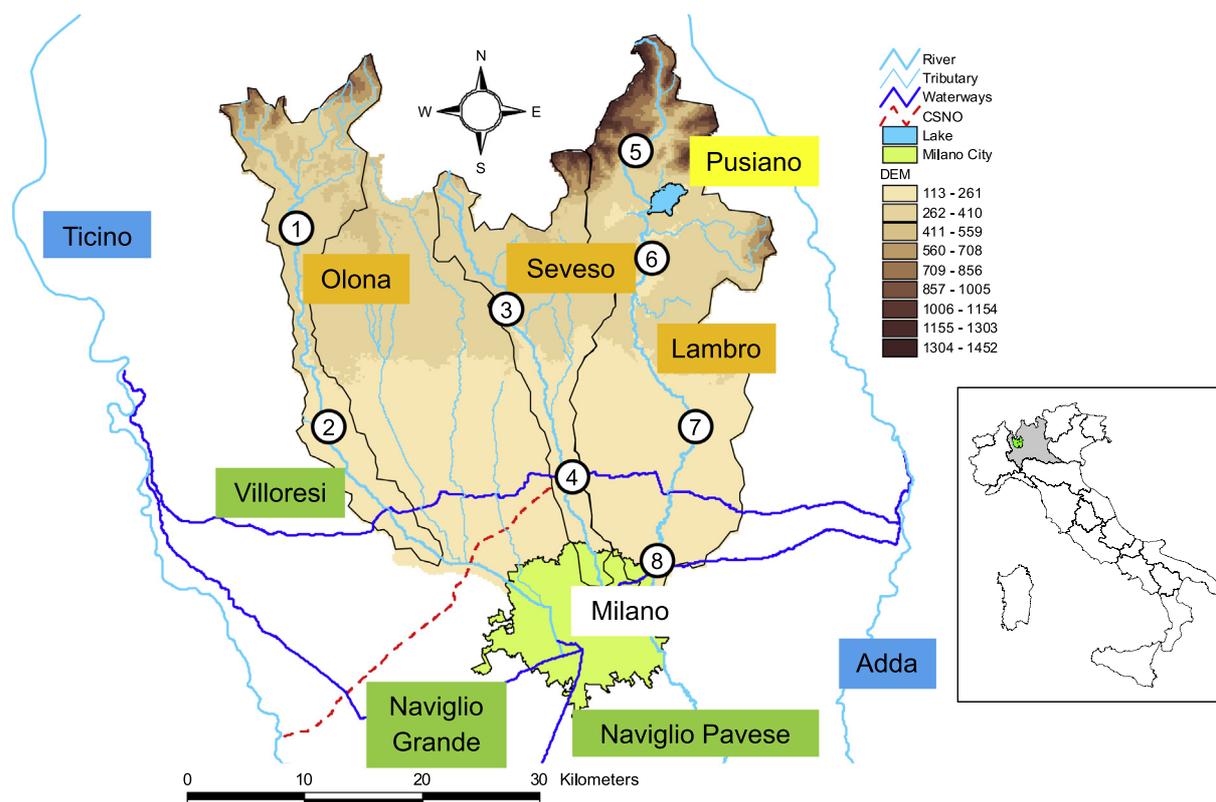


Fig. 1. River basins draining to Milano urban area. White circles denote hydrometric stations: (1) Lozza, (2) Castellanza, (3) Cantù, (4) Paderno, (5) Caslino, (6) Lambrugo, (7) Peregallo, and (8) Milano.

Table 1

River sections considered in this analysis, basin area (km²), warning discharge threshold (m³/s), lag time and station density both for the 2010 and 2014 event, respectively.

Id	Section	River	Area (km ²)	Warning threshold (m ³ /s)	Lag time (hours)	Weather station density (n/100 km ²)
1	Lozza	Olona	96.9	36	3.7	3.1–4.1
2	Castellanza	Olona	162.6	43	6.9	1.8–2.5
3	Cantù	Seveso	61.1	13	3.8	1.6–4.9
4	Paderno	Seveso	175.4	75	7.5	2.9–5.1
5	Caslino	Lambro	52.8	6	2.3	5.7–5.7
6	Lambrugo	Lambro	176.8	NA	5.2	3.4–3.4
7	Peregallo	Lambro	270.1	30	8.5	3.0–3.0
8	Milano	Lambro	531.5	83	13.7	2.6–4.3

Meteonetwork-Epson Meteo Centre (EMC) meteorological station network since 2012. Figs. 2a and 3a show the observed cumulative rainfall in 48 h for the two events along with geographical location of available rain gauges used in this study. Hydrometric data are provided by ARPA and are available from 2003 (Table 1, Fig. 1). The conversion to river discharge was not possible for Paderno Dugnano gauge section due to the lack of rating curve (Section 4 in Fig. 1). Analysis of available discharge time series shows that 33 flood events hit this region since 2003 (Table 2).

2.2. Land use change

The territory extending north of Milano has been subjected to significant expansion of the urban area since 1950, modifying the response of the watershed to precipitation input. Fig. 4a shows the part of the territory covered by urban areas in 1955, 1980 and 2010, and the corresponding value of Soil Conservation Service-Curve Number (SCS-CN; US Department of Agriculture,

1986). Land use data has been obtained from the Lombardy Region database (www.cartografia.regione.lombardia.it). We have chosen SCS-CN for measuring the basin runoff production, since the higher the CN is, the greater the generated runoff volume. Table 3 shows the percentage of urbanized area and the average CN in these selected years for the three major river basins draining to Milano. Results show that the overall percentage of urbanization has increased from 16.1% in 1955 to 35.1% in 1980 and up to 44.9% in 2010. Due to the change of land use the curve number has risen from 72.7 in 1955 to 75.9 in 1980 to up to 77.7 in 2010. The Seveso river basin has the highest percentage of urbanized area, up to 50.7% in 2010.

2.3. Description of the hydro-meteorological episodes

The two analysed floods unfolded under quite different synoptic conditions with different rainfall amounts and distributions as shown in Figs. 2a and 3a in the 48 analysed hours. The 17–18 September 2010 episode developed under weak mid-level dynamical conditions. During these days, a warm and moist air mass progressed north-eastwards over the eastern part of the Western Mediterranean, driven by the south-westerly cyclonic flow present at low levels (Fig. 5a and b). As warm and moist low-level maritime air impinged the central Alpine region during the late hours of 17 September, convective cells triggered and intensified within the convectively unstable environment, stimulated by the relatively cold continental air aloft (Fig. 5a and b). The first scattered convective cells initiated over the region in the afternoon of 17 and quickly evolved towards the basin of interest with high precipitation efficiency. During 18 September, as the upper level trough advanced north-eastwards towards northern Italy at mid-levels, more organized linear convective bands evolved over the region, bringing high precipitation rates exceeding 10 mm/h across the Lombardy region.

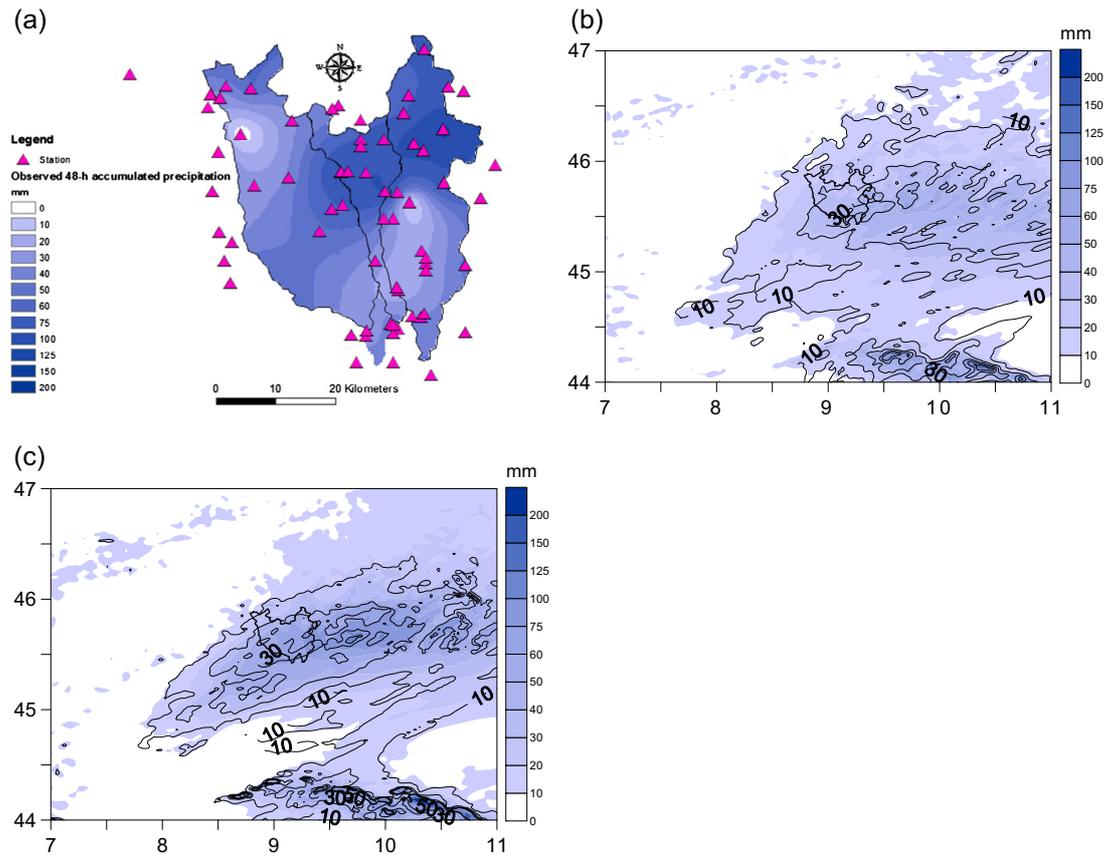


Fig. 2. 48-h rain-gauge rainfall amounts; location of rain gauges are shown in fuchsia triangles (a) and PILB (b) and MPS (c) ensemble mean 48-h accumulated precipitation fields for the 18 September 2010 event. Shaded contours according to the scale. Standard deviations also shown in (b) and (c) as black lines, in mm. The limits of the basins are highlighted in (b) and (c).

On the other side, the mid-level synoptic conditions for 7–8 July 2014 favoured upward ascents as a cold negatively tilted trough was sweeping across western Europe and extending south towards the eastern part of the Western Mediterranean. Under this system a low-level cold front evolved towards the east-southeast bringing large amounts of rainfall over the Seveso and Lambro basins from 21:30 UTC 7 July to 02:00 UTC 8 July (Fig. 5c and d). The convective systems triggered over the frontal zone during the late hours of 7 July 2014 reached the area as linear bands with embedded convective cells that exceeded 10 mm/h. The post-frontal environment, characterized by low north-westerly low-level winds, triggered isolated cells over the Rhône-Alps region that arrived at the Olona and Seveso basins with extreme precipitation intensities that exceeded 100 mm/h for certain convective systems during the evening of 8 July 2014. Rain gage data analysis shows an estimated return period of about 80–100 years over the Lambro and Seveso basins and above 200 years in the upper area of the latter.

Total economic losses caused by the two events were significant: 80 M€ and 55 M€ for the 2010 and 2014 event, respectively. During the 2010 episode the underground line 3 was seriously damaged by water and sediment that flooded it and it remained closed for recovery for about ten days.

3. Hydrological tools

We use the rainfall–runoff physically-based distributed FEST-WB model developed on top of MOSAICO library (Pianosi and Ravazzani, 2010; Ceppi et al., 2013; Ravazzani, 2013) for simulating the rainfall–runoff transformation. FEST-WB computes the

main processes of the hydrological cycle: evapotranspiration, infiltration, surface runoff, flow routing, subsurface flow, and snow melt and accumulation. The computation domain is discretized with a mesh of regular square cells (200×200 m in this study) in each of which water fluxes are calculated at hourly time step.

In particular, evolution of soil moisture, θ , for the generic cell at position i, j , is described by the water balance equation:

$$\frac{\partial \theta_{ij}}{\partial t} = \frac{1}{Z_{ij}} (P_{ij} - R_{ij} - D_{ij} - ET_{ij}) \quad (1)$$

where P is the precipitation rate, R is runoff flux, D is drainage flux, ET is evapotranspiration rate and Z is the soil depth. Runoff is computed according to a modified SCS-CN method extended for continuous simulation, where the potential maximum retention, S , is updated at the beginning of storm as a linear function of the degree of saturation, ε (Ravazzani et al., 2007).

$$S = S_1 - \varepsilon(S_1 - S_3) \quad (2)$$

where S_1 is the maximum value of S when the soil is dry (Antecedent Moisture Condition (1)), and S_3 is the maximum value of S when the soil is saturated (Antecedent Moisture Condition (3)). The actual evapotranspiration, ET , is computed as a fraction of the potential rate tuned by the beta function that, in turn, depends on soil moisture content (Montaldo et al., 2003). Potential evapotranspiration is computed according to a modified version of Hargreaves-Samani equation (Ravazzani et al., 2012) and applied to long-term simulation of hydrological balance (Ravazzani et al., 2015). Drainage flux is computed as the actual conductivity proportional to saturated conductivity according to Brooks and Corey (1964).

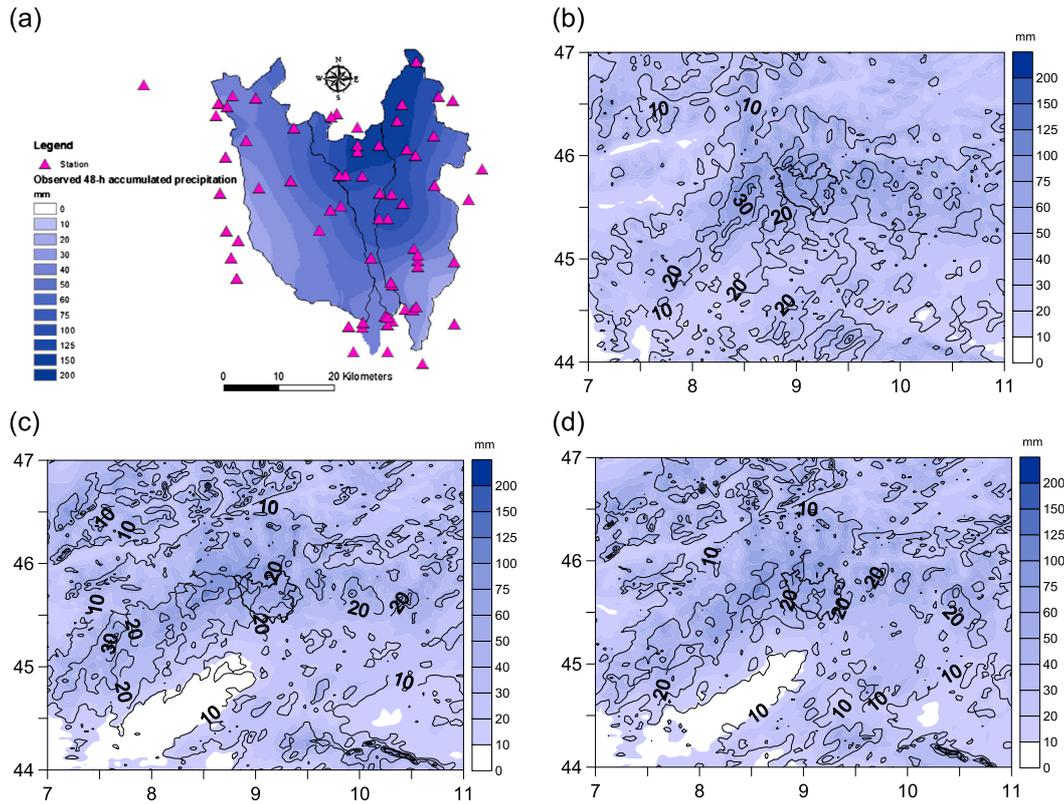


Fig. 3. As in Fig. 2 but for the 48-h rain-gauge accumulations; location of rain gauges are shown in fuchsia triangles (a), and PILB (b), MPS (c) and MPS1h (d) ensemble mean 48-h accumulated precipitation fields for the 08 July 2014 event.

Table 2
Flood events that hit Milano region since 2003.

ID	Period	Event type
1	27 November 2003	Stratiform
2	5 August 2004	Convective
3	1 November 2004	Stratiform
4	29 November 2004	Stratiform
5	9 September 2005	Convective
6	9 April 2006	Stratiform
7	8 December 2006	Stratiform
8	21 August 2007	Convective
9	11 April 2008	Convective
10	16 May 2008	Convective
11	12 July 2008	Convective
12	12 September 2008	Convective
13	4 November 2008	Stratiform
14	30 November 2008	Stratiform
14	5 February 2009	Stratiform
16	26 April 2009	Stratiform
17	6 July 2009	Convective
18	14 July 2009	Convective
19	23 December 2009	Stratiform
20	2 May 2010	Stratiform
21	10 May 2010	Stratiform
22	10 August 2010	Convective
23	17 September 2010	Convective
24	30 October 2010	Stratiform
25	15 November 2010	Stratiform
26	13 July 2010	Convective
27	25 October 2011	Stratiform
28	23 April 2012	Convective
29	20 May 2012	Stratiform
30	4 November 2012	Stratiform
31	10 November 2012	Stratiform
32	26 November 2012	Stratiform
33	7 July 2014	Convective

The surface and subsurface flow routing is based on the Muskingum–Cunge method in its non-linear form with the time variable celerity, ω , (Montaldo et al., 2007).

$$\omega(t) = 5/3V_m(t) \tag{3}$$

with V_m (m/s) denoting the mean velocity in the reach at a given time step as estimated from the Manning–Gauckler–Strickler friction equation

$$V_m = c_s R^{2/3} S_0^{0.5} \tag{4}$$

where R (m) is the hydraulic radius, c_s ($m^{1/3} s^{-1}$) is the Gauckler–Strickler roughness coefficient, and S_0 is the river bed slope ($m m^{-1}$).

River basins analysed in this study have been delineated considering the contribution of urban sewage system that, in some cases, deviates runoff from the natural flow direction, modifying the contributing area. For further details upon development and application of the FEST-WB, the reader can refer to Ravazzani et al. (2014a,b,c, 2015) and Boscarello et al. (2014).

3.1. Calibration and validation

This section describes the method used to calibrate and validate the hydrological model in those sections where accurate river discharge observations are available. To this purpose, the Paderno Dugnano gauging section was not considered, because its rating curve is missing and Lambrugo has not been calibrated either, because observed discharge is strongly affected by Pusiano Lake regulation during floods.

The calibration/validation procedure was performed over two independent periods. For each section, we split the available data

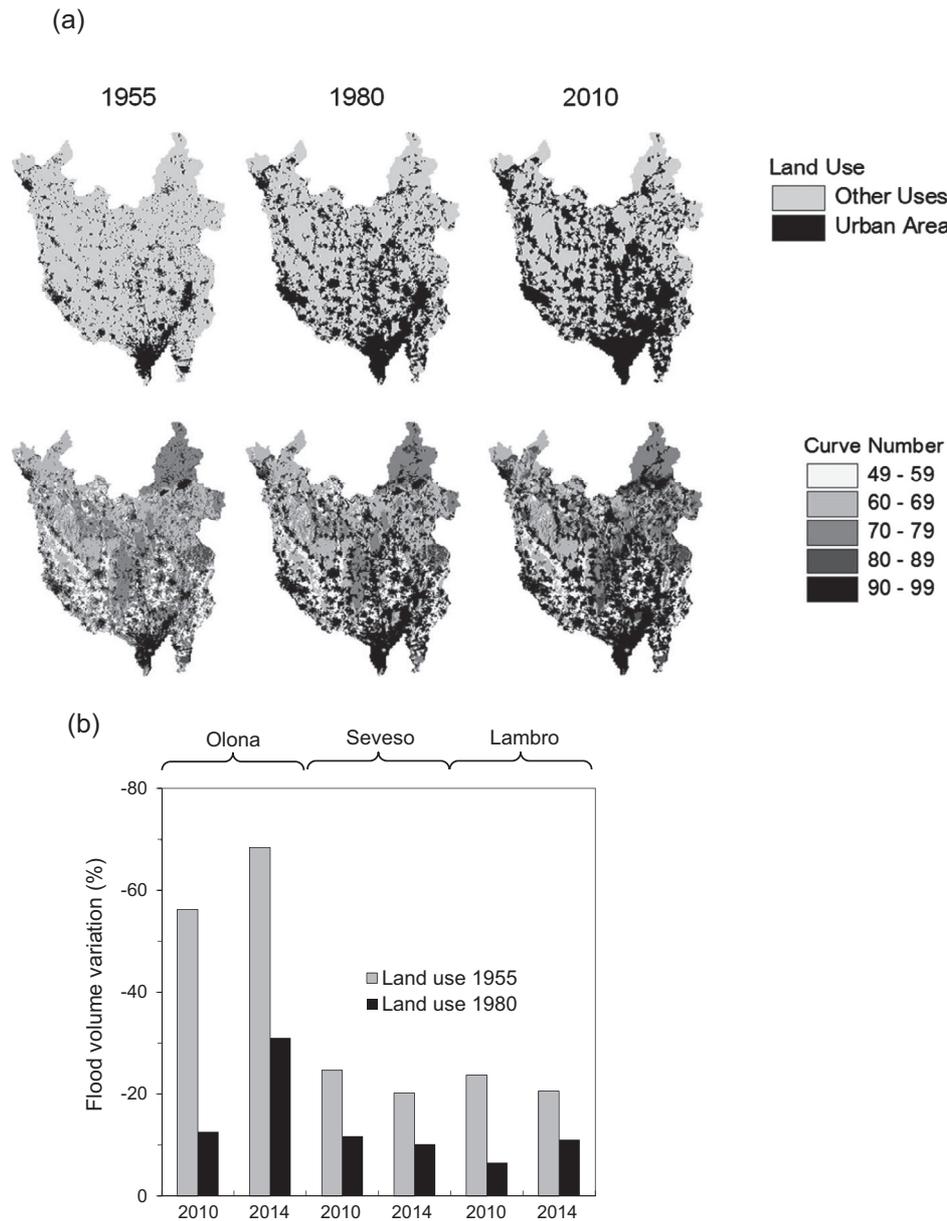


Fig. 4. (a) Urban development and curve number in 1955, 1980, and 2010 for the river basins North of Milano, and (b) percentage of flood volume variation calculated with a land use of 1955 (grey) and 1980 (black) in comparison with the one simulated during the events of 2010 and 2014 over the three basins.

Table 3

Percentage of area covered by urban development, *Urban* (%), and curve number, *CN*, in the years 1955, 1980, and 2010 for the three main river basins.

	1955		1980		2010	
	Urban (%)	CN	Urban (%)	CN	Urban (%)	CN
Olona	16.4	70.0	33.4	73.1	42.9	75.1
Seveso	17.1	73.9	38.3	77.3	50.7	80.1
Lambro	15.3	74.6	33.8	77.8	43.7	79.6
Total	16.1	72.7	35.1	75.9	44.9	77.7

into two datasets so that both stratiform and convective events were well represented in the two periods (Table 2). In particular, the September 2010 episode was used in the validation phase, while the July 2014 flood event was not included in calibration and validation, but was added to the study presented herein for its relevance.

The number of events used for calibration ranges from a minimum of 6 events for Peregallo section to 10 for Caslino and Castellanza sections. Flood events used for validation range from a minimum of 3 events for Peregallo to 4 for Caslino.

As the object of this work is the forecast of flood inducing discharge, the model was calibrated and validated considering the peak discharge relative error, ϵ_p . Model performance was considered acceptable when peak discharge error was lower than 25%, considered as the mean error associated to discharge measurement in this area (Di Baldassarre and Montanari, 2009). In order to achieve this goal, two parameters were modified: soil saturated hydraulic conductivity that affects flood volume and flood peak magnitude, and roughness coefficient used for routing runoff that affects hydrograph shape and time to peak.

For a general assessment of model's performances, other indexes well known in literature were computed, the root mean square error, *RMSE*, and the Nash and Sutcliffe (1970) efficiency,

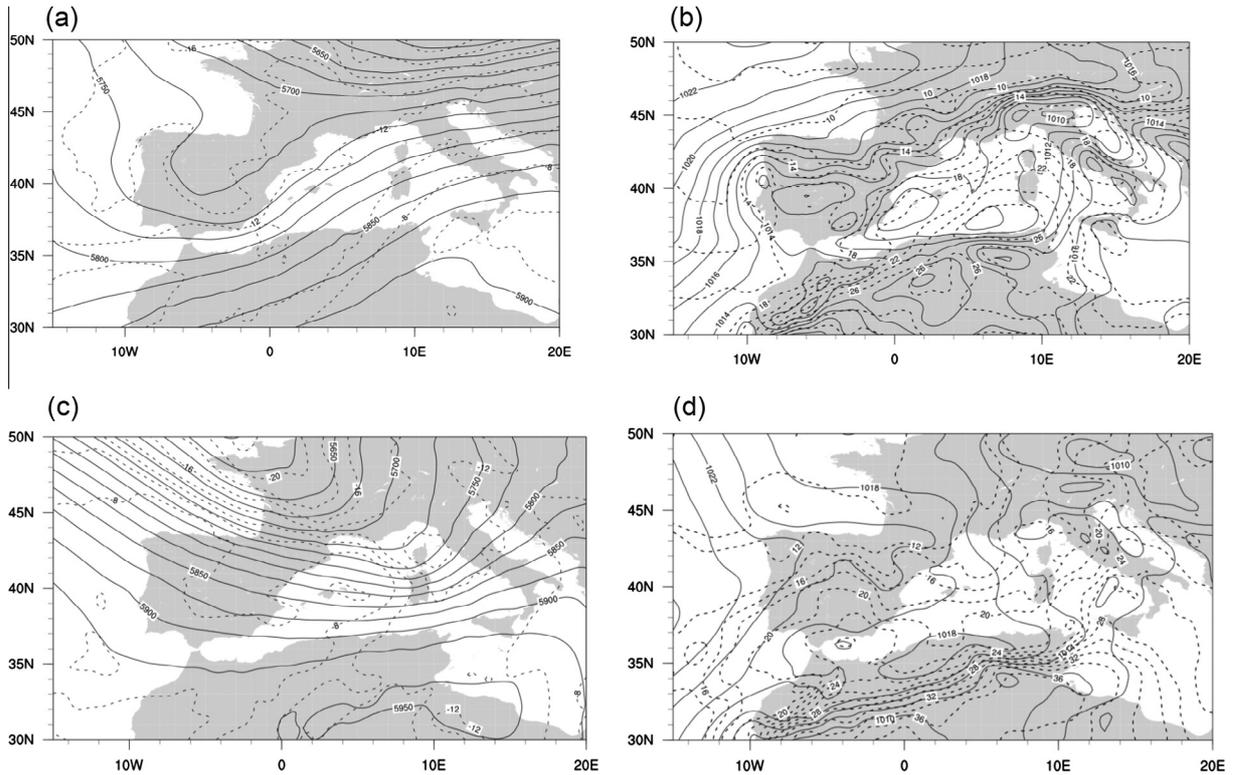


Fig. 5. ECMWF analyses for the 2 cases considered: (a and c) 500 hPa geopotential height (solid line, in gpm) and 500 hPa temperature (dashed line, in °C); and (b and d) sea-level pressure (solid line, in hPa) and 925 hPa temperature (dashed line, in °C); for: (a and b) 18 September 2010 at 00 UTC; and (c and d) 08 July 2014 at 00 UTC.

η . The model was not optimized against these as they are not suitable to characterize errors in peak flow that are crucial in flash-flood forecasting.

4. Meteorological tools

Individual ensemble members are generated using the Weather and Research Forecasting (WRF) model version 3.4 (Skamarock et al., 2008). We use a single computational domain of 750 × 550 grid-points at 2.5 km resolution, centred over north Italy and spanning the entire north-western Mediterranean region (Fig. 6); in the vertical, 35 computational levels are used. This configuration allows for deep moist convection to be explicitly resolved in our simulations. All simulations span 48 h as forecast horizon – from

17 to 19 September 2010 00 UTC and from 07 to 09 July 2014 00 UTC, respectively, thus encompassing the initiation phase and mature evolution of the most active convective systems for the 18 September 2010 and 08 July 2014 floods. These simulations are driven by the global ECMWF forecasts, using either the reference run for nesting our MPS ensemble or a collection of ECMWF-EPS members for the PILB approach. It is interesting to note that the ECMWF ensemble yielded a potential risk of extreme precipitation over the Alpine region for both case studies, according to the Extreme Forecast Index (EFI; defined in Lalaurette (2003) and Zsótér (2006)). The resulting WRF-EPS hourly QPFs are finally used to force unidirectionally the FEST-WB to ultimately generate HEPSSs one day in advance before the main peak flood: a suitable time to issue warnings and advice local authorities.

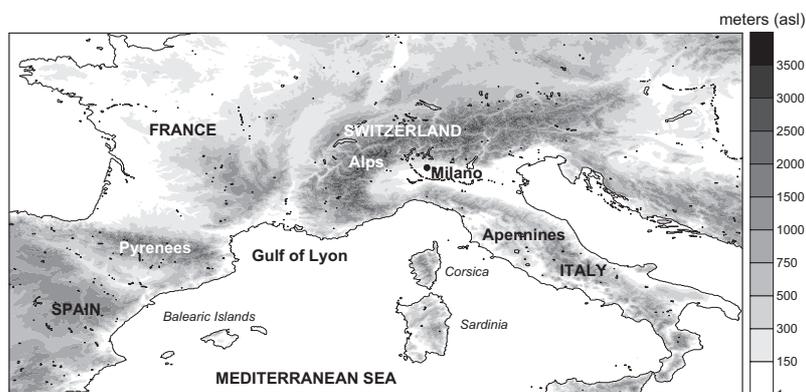


Fig. 6. Computational domain used for the WRF numerical simulations. Main geographical features mentioned in the text are indicated. Also shown the WRF model orography according to the scale.

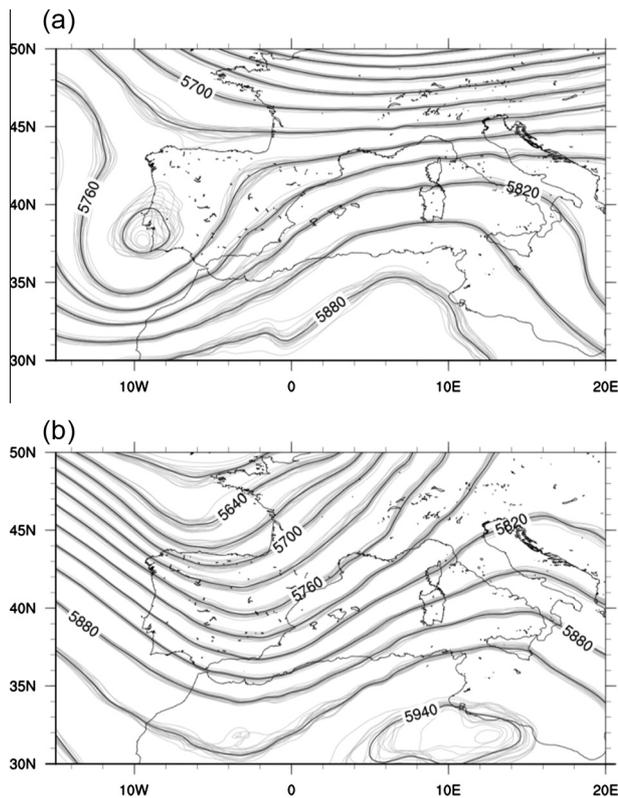


Fig. 7. Initial states of the 500 hPa geopotential heights (in gpm) for the ECMWF analyses (black lines) and the 20 selected ECMWF-EPS members (grey lines) on (a) 17 September 2010 00 UTC and (b) 07 July 2014 00 UTC. The ensemble members have been selected after applying a clustering technique (see text). Note that areas with high ensemble spread denote greater uncertainty.

4.1. Perturbed initial and lateral boundaries (PILB) ensemble

The operational European Centre for Medium Range Weather Forecasts – global Ensemble Predictions System (ECMWF-EPS) aims at sampling the distribution of plausible atmospheric states, given the bulk of observational and modelled information available. It consists of 50 members, operating at T639 spectral resolution (~ 32 km), that are generated by perturbing an initial analysis. Perturbations are derived from flow-dependent singular vectors computed daily at ECMWF in order to span the synoptic-scale uncertainties of the day (Buizza and Palmer, 1995; Molteni et al., 1996). Specifically, the PILB system is designed to dynamically downscale the global ECMWF ensemble forecasts, thus relying on the sampling of the IC/LBCs uncertainty subspace provided by the global system (Marsigli, 2009).

Under weak-to-moderate mid-levels forcing regimes, the synoptic perturbations inherited from the ECMWF-EPS into the PILB ensemble will, on average, spread out slower than in the MPS members for these episodes with active mesoscale processes. In order to alleviate this, the 20 ECMWF-EPS members exhibiting

the largest diversity over our numerical domain are identified and used as initial and boundary conditions for the entire PILB ensemble. To this end, we applied to the 50 ECMWF-EPS members a *k*-means clustering algorithm using the Principal Components of the 500 hPa geopotential and 850 hPa temperature fields over the area spanned by the WRF domain (Fig. 6). These dynamically relevant fields are often used in weather pattern classification studies (e.g. Romero et al., 1999a, 1999b; Trigo and DaCamara, 2000; Seibert et al., 2007; Huth et al., 2008). A total of 20 clusters are produced and, the 20 members closest to the centroid of each cluster are selected in the PILB (Fig. 7). The aim of this approach is to identify representative members of 20 “distinct circulation types of the day” in which the ECMWF-EPS is classified.

All PILB ensemble members use the same set of physical parameterizations (Table 4): the WRF single-moment 6-class microphysical scheme incorporating graupel; the 1.5-order Mellor–Yamada–Janjic PBL scheme; the Dudhia short-wave scheme (Dudhia, 1989); the RRTM long-wave scheme (Mlawer et al., 1997); the unified Noah land surface model (Tewari et al., 2004); and the Eta similarity surface layer (Janjic, 1994). Lateral boundary conditions for all ensemble members are updated every 3 h.

4.2. Mixed-physics (MPS) ensemble

Sensitivity analyses of mixed-physics ensembles have shown that no single WRF model configuration performs the best, as different meteorological variables are sensitive to different physical parameterizations (Jankov et al., 2005; Evans et al., 2012). We adopt a multiphysics ensemble generation strategy, which consists in combining 5 microphysical and 4 boundary layer schemes (i.e. 20 members; Table 4). Convective processes are explicitly allowed within the spatial scales resolved by the WRF model, thus these are not considered in the sampling methodology of the physical uncertainties that defines the MPS ensemble. Microphysics schemes are used to model the processes resulting in the several forms of precipitation. All selected schemes involve the simulation of explicitly resolved liquid water, cloud and precipitation, including mixed-phase transformations (i.e. the interaction of ice and liquid water). However, each microphysical parameterization treats differently the interaction among five or six moisture species (i.e. water vapour, cloud water, rain, cloud ice, snow and graupel); the physical processes of rain production, fall and evaporation; the cloud water accretion and auto-conversion; condensation; and saturation adjustment and ice sedimentation (Skamarock et al., 2008).

The PBL schemes are used to parameterize the sub-grid turbulent vertical fluxes of heat, momentum and moisture within the boundary layer and throughout the atmosphere (Pielke and Mahrer, 1975). The PBL representation can be a determinant factor in accurately simulating mesoscale weather phenomena owing to the critical role that these fluxes exert in the onset of severe phenomena. Thus, the choice of different PBL schemes can substantially affect temperature and moisture profiles in the lower troposphere and the effects of turbulence in daytime convective conditions (Hu et al., 2010; Coniglio et al., 2013).

Table 4
The different physical schemes combined for generating the MPS ensemble. Note that the physical parameterizations used in the PILB ensemble are highlighted in bold italic.

Microphysics scheme	Planetary boundary scheme
Purdue Lin (Lin et al., 1983)	Yonsei University (YSU; Hong et al., 2006)
Ferrier (1994)	Mellor–Yamada–Janjic (MYJ; Janjic, 1994)
WRF single-moment 6-class (WSM6; Hong and Lim, 2006)	Mellor–Yamada Nakanishi Niino level 2.5 (MYNN; Nakanishi and Niino, 2006)
Goddard scheme (Tao et al., 1989)	Asymmetric convection model 2 scheme (ACM2; Pleim, 2007)
New Thompson (Thompson et al., 2008)	

Table 5
River sections considered in calibration and validation analysis, and peak discharge relative error.

Section	ε_p (%)			RMSE (m^3/s)			η		
	Before calibration	After calibration	Validation	Before calibration	After calibration	Validation	Before calibration	After calibration	Validation
Lozza	-50.7	0.27	26	13.76	10.36	6.10	-0.71	0.07	0.57
Castellanza	-51.8	0.12	16.7	15.93	10.14	7.65	-0.29	0.36	0.63
Cantù	-65.5	-10.9	-8.9	6.85	5.31	3.51	-0.56	0.30	0.46
Caslino	78.4	0.57	14.5	11.62	7.67	5.46	0.27	0.41	0.45
Peregallo	-72.1	1.5	1.2	20.65	15.11	6.00	-0.83	0.43	0.63
Milano	-74.8	-3.54	-25.5	30.96	19.26	12.49	-0.79	0.03	0.35

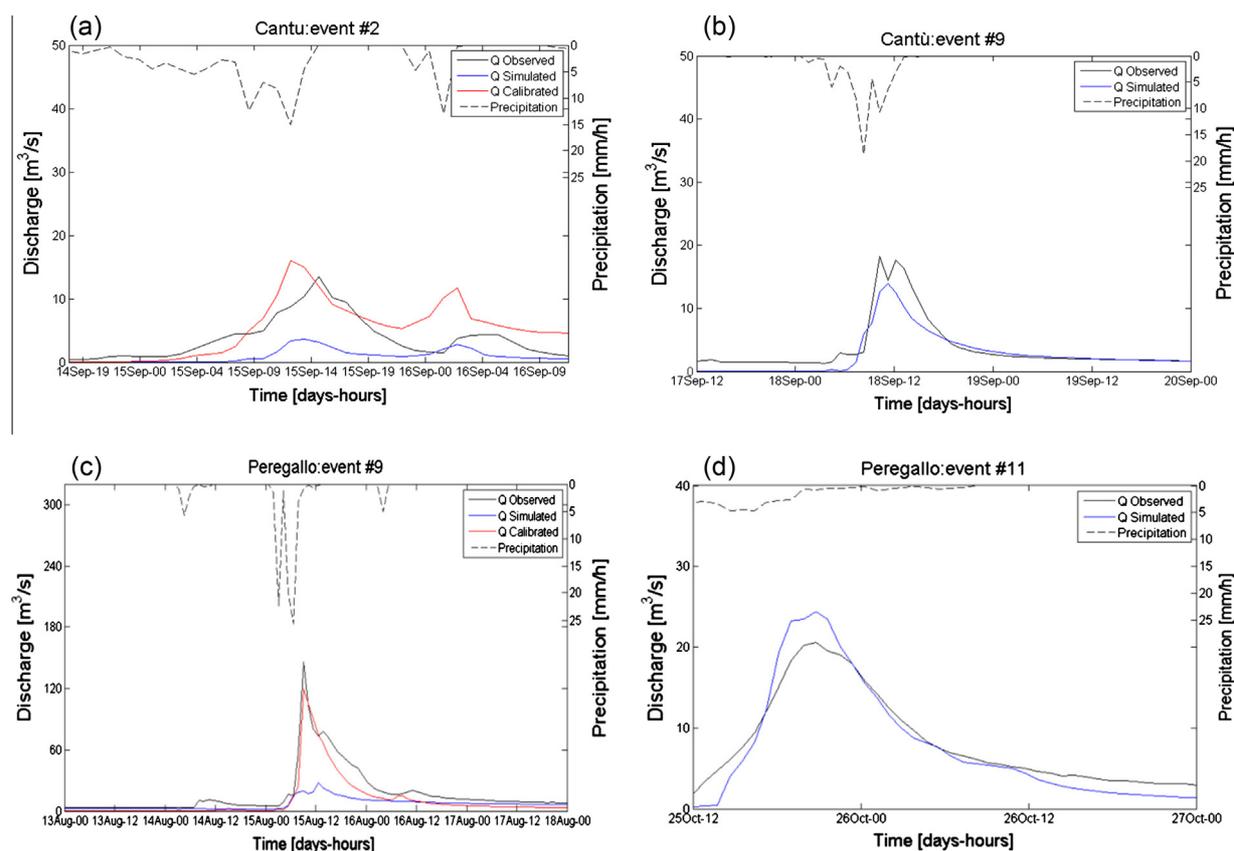


Fig. 8. Simulated hydrograph before and after calibration compared to observed discharge: (a) stratiform calibration event at Cantù, (b) convective validation event at Cantù, (c) convective calibration event at Peregallo, and (d) stratiform validation event at Peregallo.

Initial and lateral boundary conditions are the same for all mixed-physics experiments and each case study and come from the ECMWF-EPS reference (unperturbed) forecast. Lateral boundary conditions are updated every 3 h. Since early 2014, ECMWF archives hourly forecasts for the reference run. This opens the window to investigate the sensitivity of HEPS skill to the boundary conditions update frequency. To this end, an additional mixed-physics ensemble with hourly update –labelled as *MPS1h*– is tested on the 08 July 2014 episode.

5. Results and discussion

5.1. Performance of the hydrological model

Calibration of FEST-WB model allowed the removal of systematic errors in simulated peak discharges, as shown in Table 5 where peak discharge relative error is compared before and after calibration.

Some underestimation is still present in the Cantù and Milano gauge sections after calibration, but error lies inside the acceptable bounds. Increase of peak discharge errors in validation for Lozza and Castellanza is due to the effect of Ponte Gurone dam which is not simulated into the hydrological model. Peak error lies within or slightly exceeds acceptable bound even in validation. In Fig. 8, simulated hydrograph before and after calibration is compared to observed hydrograph for one stratiform calibration event at Cantù, one convective validation event at Cantù, one convective calibration event at Peregallo, and one stratiform validation event at Peregallo.

Table 5 shows also the other two computed performance indexes: the root mean square error and the Nash and Sutcliffe (1970) efficiency. They confirm improvement of simulations after calibration, although they were not explicitly considered for optimization since they are mainly sensitive to simulation of the shape of hydrograph and, not the flood peak, that is the objective of this work.

5.2. The impact of land-use change on runoff

The change of land use due to urban development has the consequence of increasing the curve number that, in turn, can affect basin runoff production. In order to assess the impact of urban development on floods, the two flood events occurred in 2010 and 2014 were simulated with the CN map based on current land use (2010), and compared to simulation runs with the CN maps derived using the 1955 and 1980 land use, following Rulli and Rosso (2002). Variation of flood volume obtained with current land use respect to the value obtained with 1955 and 1980 land use is reported in Fig. 4b. The Seveso and Lambro basins show a decrease of flood volume in the range –24.7 to –20.2 with the 1955 land use, and –11.7 to –6.6 with the 1980 land use. The greatest impact is shown on the Olona river basin with a maximum volume change of about –68% registered for the 2014 flood simulated assuming the 1955 land use. However, the Olona basin was subjected to small precipitation amounts in both analysed events and, as a consequence, its flood volume is reasonable small. When precipitation amount is small, the initial abstraction of the SCS-CN method, that defines the precipitation that is lost by vegetation and interception, becomes significant over runoff generation. Initial abstraction is a function of SCS-CN itself: an increase of CN implies a decrease of initial abstraction and, therefore, an increase of runoff. This explains the remarkable impact reported on the Olona river floods.

5.3. Verification of the weather predictions

Operational pluviometric networks consisted of 34 and 41 rain-gauges over the basins during the 18 September 2010 and 08 July 2014 floods, respectively (Fig. 1). Ensemble QPFs for each case study are bilinearly interpolated onto the rain-gauge locations (Akima, 1978, 1996). Next, observed and simulated precipitations over the three catchments are accumulated over 48 h periods (i.e., 17–19 September 2010 00 UTC and 07–09 July 2014 00 UTC, Figs. 2 and 3).

First, the skill of the EPSs is assessed by means of Taylor diagrams (Fig. 9; Taylor, 2001). The PILB and MPS ensembles show weak differences in skill in terms of spatial distributions of the 48 h accumulated precipitations for both flood events, although the latter exhibits some members with slightly better statistical scores for the 18 September 2010 flash-flood. In general, both ensemble strategies feature too small spatial variability for the flood episodes (Fig. 9a and b). Regarding the MPS1h test for the 08 July 2014 episode, no significant benefits are obtained when compared with the more standard MPS strategy (Fig. 9c). Spatial correlations are rather low for many ensemble members and both case studies, highlighting the difficulties for accurately forecasting the spatial distributions of the convectively-driven rainfall fields at such small-scales.

Secondly, the ability of the different EPS strategies to predict forecast skill has been assessed by comparing the ensemble spread against forecast error. To this end, we have computed the ensemble root-mean-square error (RMS_{ENS}) and the root-mean-square error spread (SP_{RMSE}) statistical indices (Scherrer et al., 2004). The RMS_{ENS} is obtained after applying the root-mean-square error (RMSE) to each ensemble member, i :

$$RMSE_i = \sqrt{\frac{1}{M} \sum_{m=1}^M (f_{m_i} - o_m)^2} \quad (5)$$

where f_{m_i} and o_m are the forecasted and observed values of the variable of interest at the rain-gauge m , respectively; and M is the total number of rain-gauges. Then, the RMS_{ENS} is defined as follows:

$$RMS_{ENS} = \frac{1}{N} \sum_{i=1}^N RMSE_i \quad (6)$$

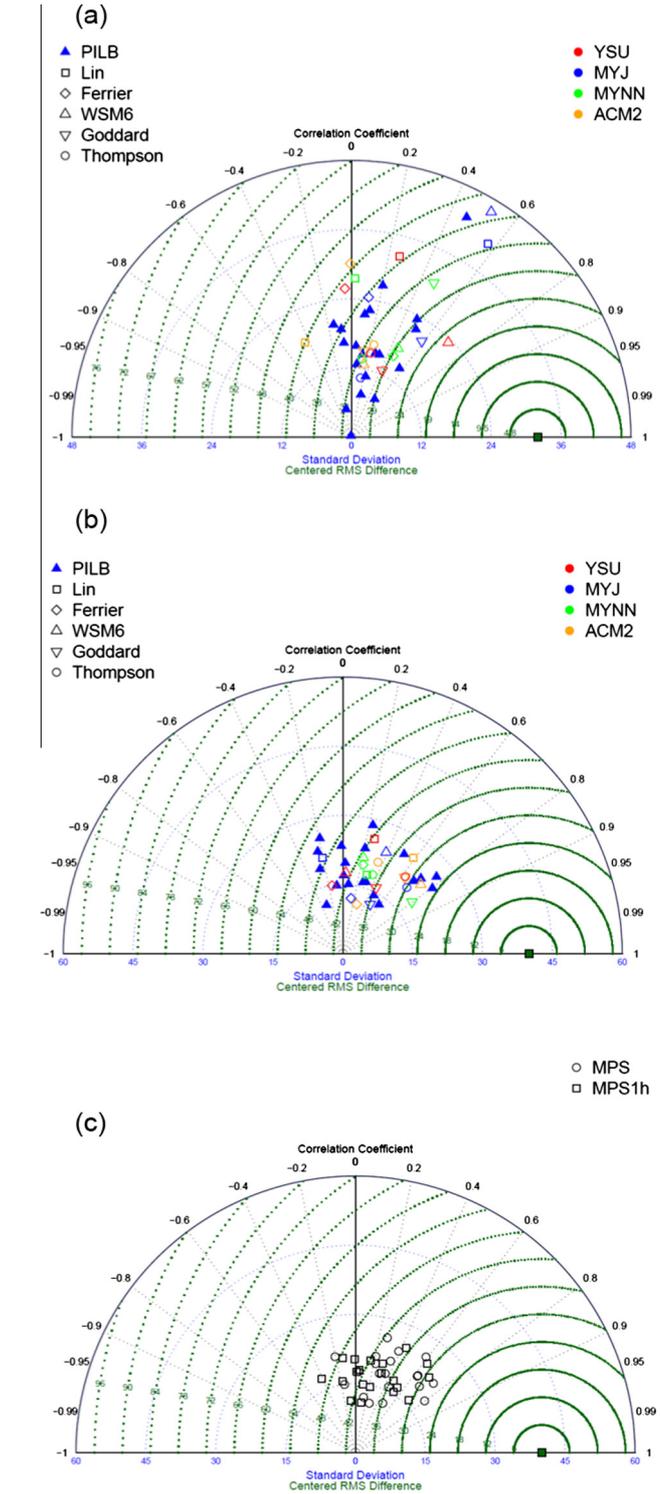


Fig. 9. Taylor diagrams of the PILB and MPS ensembles for the (a) 18 September 2010 and (b) 08 July 2014 floods. The Taylor diagram of the MPS and MPS1h ensembles for the 08 July 2014 flood is also shown (c). Standard deviations and centred root-mean-square (RMS) differences in mm.

where N is the ensemble size. The SP_{RMSE} is computed as the average of the sum of the RMSE between every ensemble member and the remaining elements as follows:

$$T_{ij} = RMSE[W_i, W_j] \quad (7)$$

$$\Delta = \sum_i \sum_j T_{ij} \quad i < j \quad (8)$$

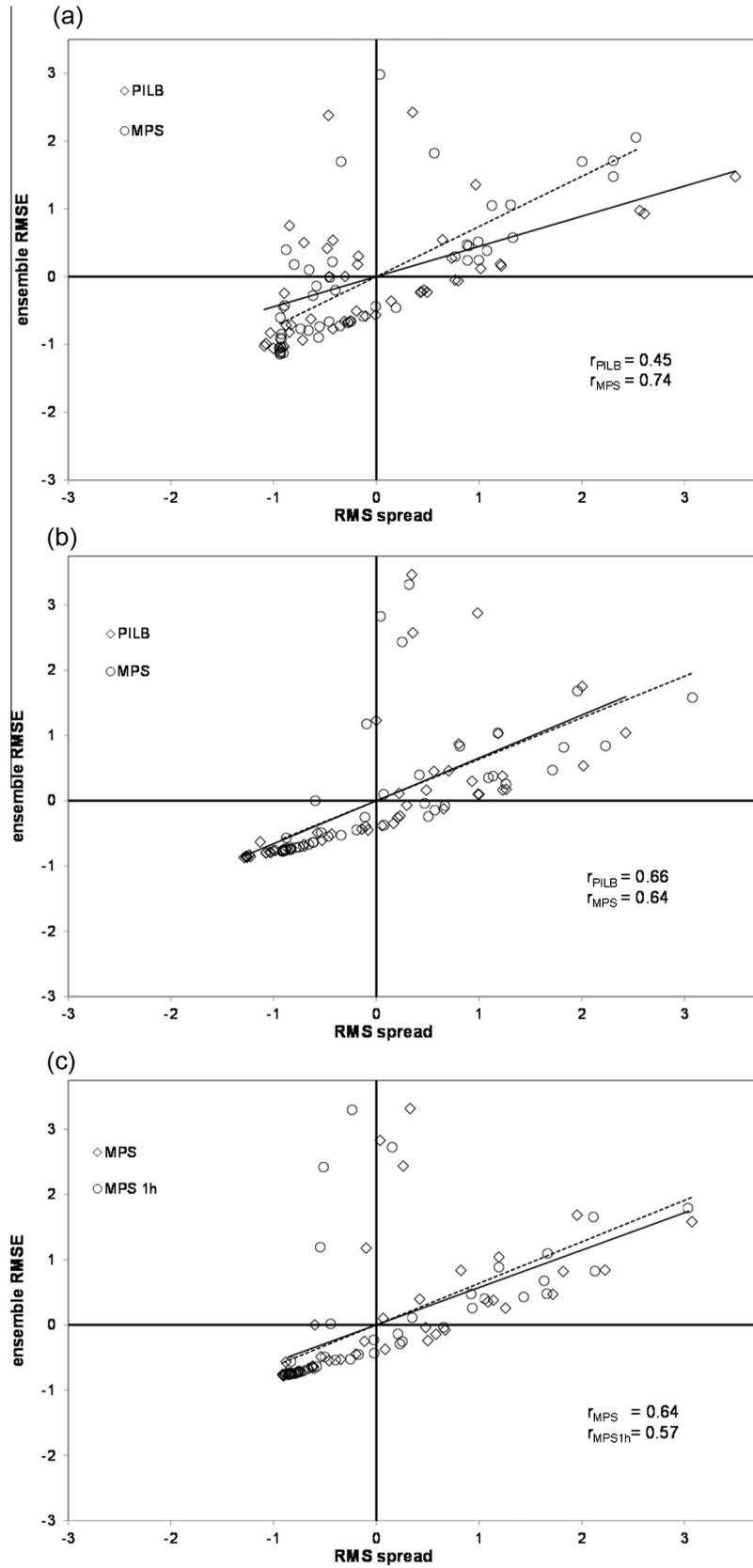


Fig. 10. Spread-skill scatterplots of (a) the PILB and MPS ensembles for the 18 September 2010 flood; and (b) the PILB and MPS, and; (c) the MPS and MPS1h ensembles for the 08 July 2014 flood. Also shown linear regressions as continuous and dashed black lines for the PILB and MPS ensembles in (a) and (b), respectively. Continuous and dashed black lines denote the MPS1h and MPS experiments in (c). Note that spread and skill scores have been normalized.

Table 6

Percentage of exceeding the warning threshold in each analysed basin for the two events with different WRF combinations.

Exceeding threshold		Seveso		Lambro		
		Cantù	Peregallo	Milano		
2010 event	PILB-HEPS	20%	35%	10%		
	MPS-HEPS	50%	85%	40%		
2014 event	PILB-HEPS	25%	50%	10%		
	MPS1h-HEPS	25%	50%	10%		
	MPS-HEPS	15%	55%	10%		

$$SP_{RMSE} = \frac{2 \cdot \Delta}{N \cdot (N - 1)} \quad (9)$$

where W_i and W_j are calculated over the $N \times M$ points for the meteorological field of interest. These statistical scores have been computed from the hourly observed and simulated precipitations at all available rain-gauges and over the entire 48-h forecasting period for each case study. Next, the quality of the distinct ensemble strategies are assessed by (Scherrer et al., 2004): (i) building contingency tables based on the spread-skill relationships, and (ii) plotting scatterplots of spread versus skill and computing linear correlation coefficients.

Contingency tables are built by counting all individual spread-skill pairs that lie within the different quadrants in the skill-spread scatterplots. Skillful probabilistic EPSs show small spreads mainly associated with high forecasting skills, whereas small spreads linked to low ensemble skills denote more unreliable EPSs. As large spreads are not always associated with low forecasting skills, the fanning in the spread-skill scatterplots is an inherent property of skillful probabilistic EPS systems (Fig. 10; Molteni et al., 1996; Scherrer et al., 2004). Note that we have normalized the ensemble spreads and skills in the scatterplots (i.e., $z = \frac{x-\bar{x}}{S_x}$ where z is the standardized variable, \bar{x} is the mean, and S_x is the standard deviation), so large skills and small spreads appear with negative numbers.

Table 7 and Fig. 10 show that small spreads are predominantly linked with high skills for all the ensemble strategies. Therefore, the hourly observed precipitations lie within the forecast ensembles. Spread-skill correlations range from 0.45 to 0.74, well above the minimum value of 0.4 which would indicate unreliable EPSs (Scherrer et al., 2004). Moreover, SP_{RMSE} is created more rapidly than the RMS_{ENS} for the 18 September 2010 flash-flood ($m_{PILB} = 0.59$; $m_{MPS} = 0.82$ for the unnormalized slopes of the linear regressions), whereas errors are generated faster than spread for the 08 July 2014 episode ($m_{PILB} = 1.12$; $m_{MPS} = 1.03$). In addition, the 65–88% of the small (large) spread is linked to the corresponding high (low) skill for all the ensemble strategies, pointing out a remarkable ability of the different EPS approaches to estimate skill from spread. Note that not remarkable differences are found between the MPS and MPS1h experiments for the 08 July 2014 flood.

Thus, accounting for the largest diversity in the IC/LBCs is an effective approach in order to avoid reduced PILB ensemble

spreads when compared to mixed-physics ensembles. However, the enhanced PILB ensemble spread does not result in higher skill for these flood events when compared with the MPS ensemble, provided that the unperturbed IC/LBCs yield skillful forecasts. The crucial role of the local orography for triggering and organizing the simulated convective activity masks the benefits of accounting for an enhanced diversity in the IC/LBCs. That is, the influential role of local topography diminishes the sensitivity to this kind of uncertainties. Furthermore, the use of more frequent LBCs results in a decrease of the ensemble spread, ($m_{MPS} = 1.03$; $m_{MPS1h} = 0.94$), negatively impacting the ensemble skill when forecasting the 48 h rainfall amounts (Clark et al., 2008).

5.4. Performance of the hydro-meteorological prediction system

The performance of the hydro-meteorological coupling system is here discussed for the 2010 and 2014 events in the three analysed gauge sections of the two catchments (the Seveso at Cantù, and the Lambro at Peregallo and Milano). Analyses are carried out considering the WRF forecasts and a control simulation for precipitation and discharge predictions in comparison with observed data. Cumulated precipitation forecasts shown in Figs. 11 and 12 (top) refer to initializations, respectively, on 17 September 2010 00 UTC, 36–48 h before the measured peak flow of the following day occurred between 16 and 22 UTC depending on the basin section, and on 07 July 2014 00UTC, 24–36 h before the observed peak flow (between 04 and 11 UTC on the day after).

Fig. 11 shows the performance for the 2010 event over the three close sections in terms of precipitation (top) and discharge (bottom) forecasts. It clearly appears how the PILB forecast underestimates the cumulated precipitation over the 48 h in all the three sections when comparing the ensemble median and percentiles with observed data. On the contrary, the MPS performs much better, achieving a good agreement (−8% at Cantù and Peregallo, −3% at Milano). From a hydrological point of view, a good result is also obtained by the FEST-WB model (green rhombi) initialized with measured values coming from the hydro-meteorological regional network: in particular simulation errors compared with observed data (red circle) are equal to +5% at Cantù, +12% at Peregallo and −4% at Milano. The underestimation of the PILB precipitation forecasts is consequently propagated to the PILB-HEPS over the three basins; on the contrary, MPS-HEPS limits this error prediction (−28% at Cantù, −23% at Peregallo and −38% at Milano), and most of the ensembles go beyond the respective alert threshold (10 in Cantù, 17 in Peregallo and 8 in Milan out of the total 20), providing an useful advice to local authorities and civil protection agencies.

Because of spatial variability of precipitation, and associated return period (Section 2.3), quantitative precipitation forecast becomes very challenging even with an ensemble strategy approach: this is mainly due to the east–west movement of convective cells over the two elongated shape basins placed north–south; therefore, even a small shift in the target forecast can have a huge impact on hydrological prediction. In fact, QPFs turn out to be far from the observed values in all the three sections (Cantù, Peregallo, and Milano) with the probabilistic WRF simulations

Table 7

Contingency table for the SP_{RMSE} spread and RMS_{ENS} skill derived from the observed and simulated hourly precipitations and the different EPS approaches. “Small spread-high skill” and “large spread-low skill” values are in bold. Note that high skill denotes low RMS_{ENS} .

Percent (%)	2010 flood				2014 flood					
	PILB		MPS		PILB		MPS		MPS1h	
	Small spread	Large spread								
High skill	45.8	14.6	54.2	2.1	47.9	12.5	56.3	10.4	56.3	8.3
Low Skill	18.8	20.8	10.4	33.3	0.0	39.6	2.1	31.3	8.3	27.1

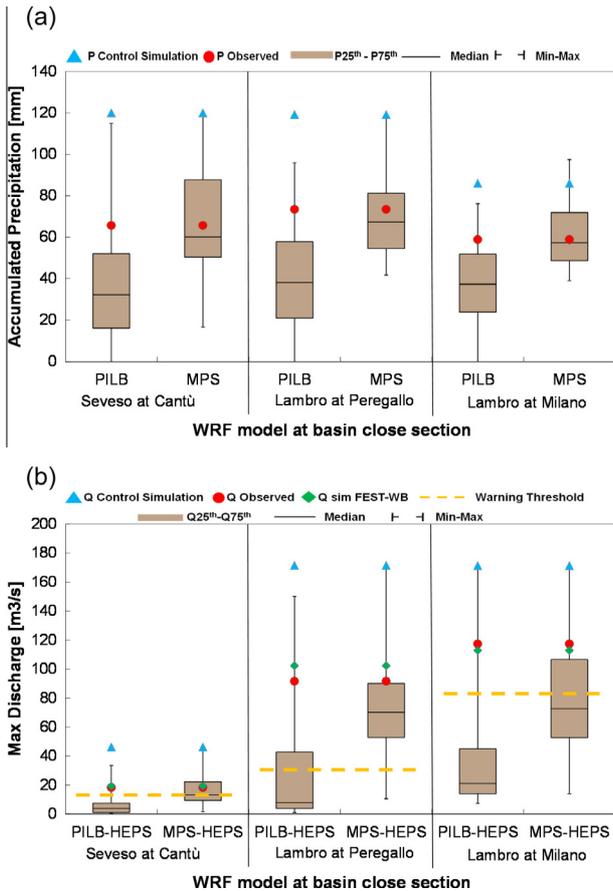


Fig. 11. QPFs (a) and QDFs (b) for the 18-09-2010 event over three close sections of the two hydrological basins with two different WRF simulations. Observed values are shown in red circles, the FEST-WB peak runoff is shown with green rhombi, the control forecast with blue triangles and WRF simulations with the brown boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

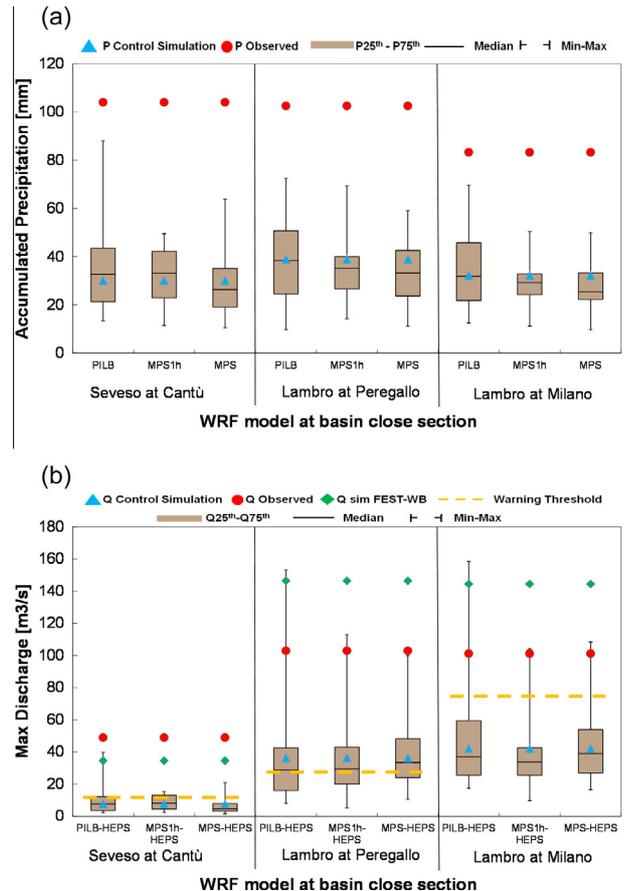


Fig. 12. QPFs and QDFs for the 08-07-2014 event over three close sections of the two hydrological basins with three different WRF simulations. Observed values are shown in red circles, the FEST-WB peak runoff is shown with green rhombi, the control forecast with blue triangles and WRF simulations with the brown boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 12) in the 2014 event; consequently, this underestimation is transferred in the peak discharge prediction as it is highlighted in Fig. 11 (bottom), despite the FEST-WB model forced with observed data shows an overestimation of the peak discharge at Peregallo (+42%) and Milano (+43%) gauging sections.

5.5. System reliability for civil protection

Notwithstanding the underestimation shown by the HEPS for the 2010 event and, above all, for the 2014 case, if we investigate deeply the ensemble spread, some of the WRF simulations exceed the warning threshold of the respective basins. From a civil protection point of view, the exceeding percentage given by the WRF ensemble members can provide more information than a single forecast can do. In Table 6 we summarize the exceeding probability with all possible WRF combinations in both events, in terms of how many ensemble members go beyond the alert threshold.

It clearly appears how the MPS-HEPS performance is better for the 2010 event in all three basins, while the PILB-HEPS gives, on average, a high performance in the 2014 event. Best results among the three basins are achieved at Peregallo, where at least a 10% of exceeding probability is always present. To better appreciate the value of the HEPS, we join in a single plot all the ensemble members of the WRF forecasts and calculate the deviation from the observed peak discharge and time over the three catchments both

for the 2010 and 2014 events, enhancing the innovative exploration of the all WRF combinations.

Following the approach of the “Peak Box” proposed by Zappa et al. (2013), Figs. 13 and 14 summarize all the simulations and forecasts carried out for the three basins in this study, considering on the y-axis the deviation from the observed peak discharge and on x-axis from the observed peak time, which is another important factor to be considered. The red circle is the true observation used as normalized reference; the green rhombus is the FEST-WB model simulation forced with observed data; the purple rhombi and brown rectangles are respectively the PILB-HEPS and MPS-HEPS. The warning threshold (orange dashed line) is shown in terms of cubic meters per second under or over the observed value, in order to display whether an ensemble member goes beyond or not this given alert line although its forecasted value can be far from the actual one. In fact, since hydrological alarms are given according to the threshold exceeding, this probability becomes a key factor for civil protection: i.e., we are far from what are going to observe, but we are able to predict rather well the exceeding threshold, and issue right alerts.

Figs. 13 and 14 also highlight the deviation forecast in terms of time from the observed peak which is another noteworthy feature for early warning systems and flood risk. In particular, most of the ensemble members of the two HEPS predicted a delay of some hours in comparison with the observed peak time in the 2010 event (Fig. 13) over the three watersheds; however, many

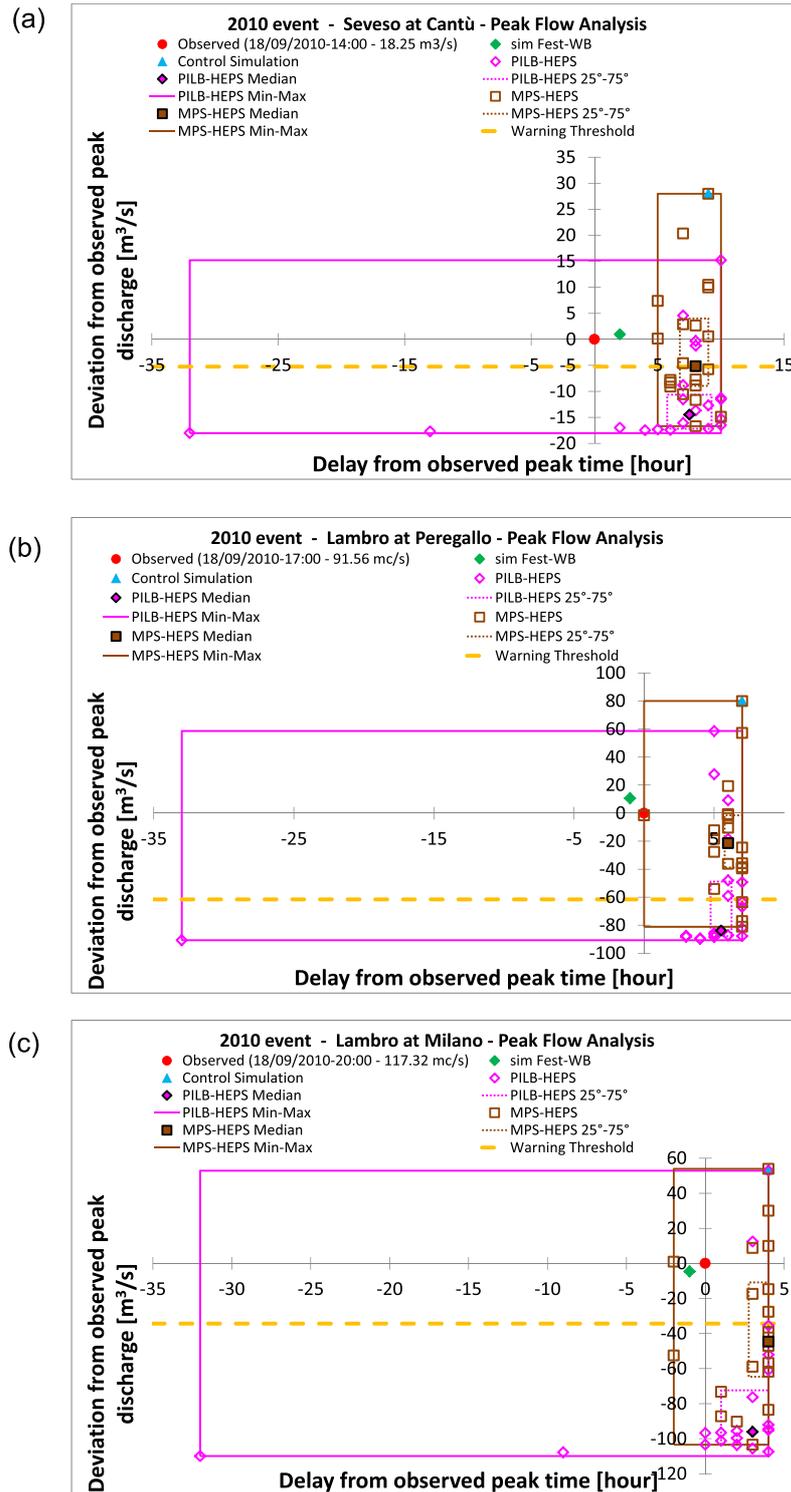


Fig. 13. Peak flow analysis for the Seveso at Cantù (a), the Lambro at Peregallo (b) and at Milano (c) for the 2010 event. Y-axis shows the deviation from the observed peak discharge and x-axis the deviation from the observed peak time. The green rhombi show the deviation of the FEST-WB simulations initialized with observed data, while the blue triangles are the control forecast errors. Ensemble forecast errors in terms of deviation from the observed peak discharge and time are shown with purple rhombi and brown triangles for the PILB-HEPS and MPS-HEPS respectively; all deviations are normalized from the observed value (red circle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

members have exceeded the warning threshold. Regarding the 2014 event (Fig. 14), the forecasting delay from observed peak time is also observed, except at Peregallo section where it is not so marked as in the other two basins. Nevertheless, if we take into account all the 60 WRF simulations, we count 13 ensembles at

Cantù, 31 at Peregallo and 6 at Milano that exceed the warning thresholds, providing a probabilistic information, which is more meaningful if compared with single forecasts, that can be addressed to local Italian civil protection agencies for preventive actions and support decision systems.

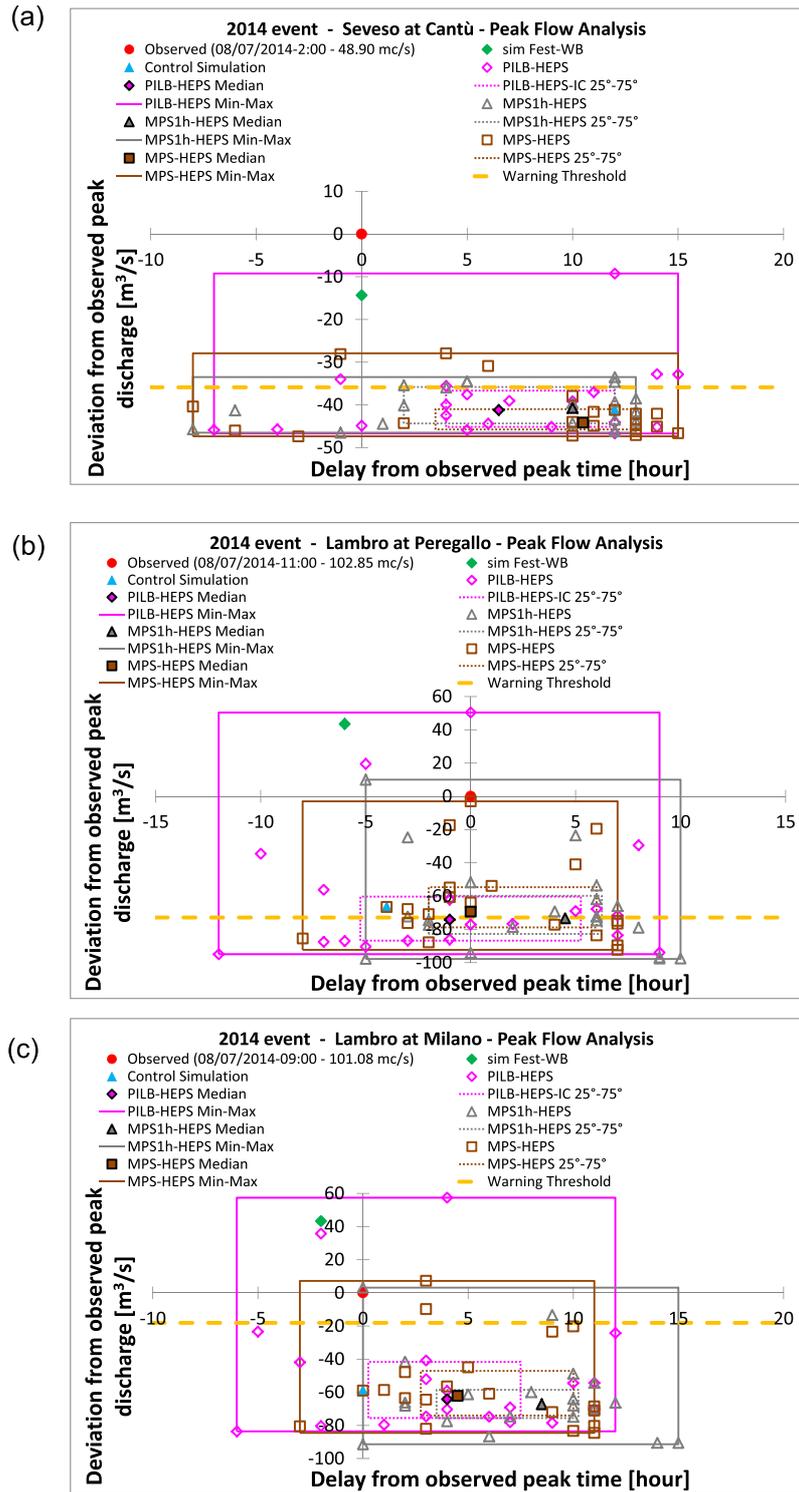


Fig. 14. Peak flow analysis for the Seveso at Cantù (top), the Lambro at Peregallo (centre) and at Milano (bottom) for the 2014 event. Y-axis shows the deviation from the observed peak discharge and x-axis the deviation from the observed peak time. The green rhombi show the deviation of the FEST-WB simulations initialized with observed data, while the blue triangles are the control forecast errors. Ensemble forecast errors in terms of deviation from the observed peak discharge and time are shown with purple rhombi, brown and grey triangles for the PILB-HEPS, MPS-HEPS, and MPS1h-HEPS respectively; all deviations are normalized from the observed value (red circle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

In this work we reanalysed the two most significant flood episodes that recently hit the Milano urban area, in order to explore different ensemble strategies for flood forecasting purposes. The

two flood events were among the most difficult to predict because of the high intensity and spatial variability of precipitation. Far from characterizing the statistical properties of the proposed ensemble hydrometeorological forecasting systems, our results show the potential of PILB and MPS to produce reliable extreme

flood forecasts over very small and rapid catchments as the Milano river basin. The Main findings of this work are:

- The FEST-WB model, forced with observed data, is reasonably satisfactory in simulating the peak discharges triggered by both stratiform and convective events in the calibration and validation phase. Future developments can be based on merging radar data to rain gauge data to provide more accurate rainfall observations on the ground, as shown in Löwe et al. (2014), to improve the hydrological model skill.
- The land-use change due to urban development that affected the Milano area starting from 1955 had the consequence of increasing the curve number and, therefore, a significant increase of runoff volume.
- The use of hydrological ensemble prediction systems is an effective and promising non-structural measure to help mitigate flood risk in Milano urban area, synergic with the existing structural engineering works. A forecast horizon of two days is required for an operational chain over the three basins, and accurate quantitative forecasts are necessary at least one day in advance, as already mentioned in Meneguzzo et al. (2004).
- No substantial differences have been found between the PILB and MPS ensemble strategies. It appears that for convectively-driven floods generated under weak-to-moderate synoptic dynamic forcing during the warm season, to account for an enhanced disparity in the IC/LBCs is an effective strategy to avoid underdispersion when compared with a mixed physics ensemble. However, note that an enhancement of the PILB ensemble spread would not always be beneficial for flood forecasting purposes over small- and medium-size basins, since more ensemble members would forecast rainfall amounts outside the basin boundaries, resulting in underproductions by the HEPSSs.
- No benefits are obtained by using more frequent unperturbed LBCs in a mixed physics ensemble as this results in a decrease of the ensemble spread.

Furthermore, the indistinguishable performance of PILB and MPS, at least for the two case studies analysed, suggests that both sources of uncertainty contribute similarly to produce adequate levels of spread in the forecasts. This opens the opportunity to explore mixed-type ensembles which merge both strategies, although the optimal combination of PILB and MFS members is an issue that deserves further research.

The EPSs have shown to be valuable when exceeding warning thresholds and issue the correct alerts, even if discharges can be far from observations. These limitations have arisen when dealing with convective-scale forecasting despite of using ensemble approaches. It remains as future work to test whether the application of different model output statistics techniques to the WRF outputs can help to overcome these difficulties.

Acknowledgements

Prof. Georgakakos, scientific editor of Journal of Hydrology, and two anonymous reviewers are deeply acknowledged for their efforts to improve the quality and contents of this manuscript. We thank ARPA Lombardia (<http://www.arpalombardia.it>) for providing meteorological and discharge observations from automatic stations and the Meteonetwork Association (<http://www.meteonetwork.it>) for the provision in weather data. This work has been sponsored by CGL2011-24458 (PREDIMED) and CGL2014-52199-R (EXTREMO) Spanish projects, which are partially supported with FEDER funds.

References

- Addor, N., Jaun, S., Fundel, F., Zappa, M., 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.* 15, 2327–2347.
- Akima, H., 1978. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Trans. Math. Softw.* 4, 148–164.
- Akima, H., 1996. Algorithm 761: scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans. Math. Softw.* 22, 362–371.
- Amengual, A., Homar, V., Jaume, O., 2015. Potential of a probabilistic hydrometeorological forecasting approach for the 28 September 2012 extreme flash flood in Murcia, Spain. *Atmos. Res.* 166, 10–23.
- Amengual, A., Romero, R., Alonso, S., 2008. Hydrometeorological ensemble simulations of flood events over a small basin of Majorca Island, Spain. *Q. J. R. Meteorol. Soc.* 134, 1221–1242.
- Amengual, A., Romero, R., Gómez, M., Martín, A., Alonso, S., 2007. A hydrometeorological modeling study of a flash flood event over Catalonia, Spain. *J. Hydrometeor.* 8, 282–303.
- Amengual, A., Romero, R., Vich, M., Alonso, S., 2009. Inclusion of potential vorticity uncertainties into a hydrometeorological forecasting chain: application to a medium size basin of Mediterranean Spain. *Hydrol. Earth Syst. Sci.* 13, 793–811.
- Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The European flood alert system EFAS Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.* 13, 141–153.
- Boscarello, L., Ravazzani, G., Rabuffetti, D., Mancini, M., 2014. Integrating glaciers raster-based modelling in large catchments hydrological balance: the Rhone case study. *Hydrol. Process.* 28 (3), 496–508. <http://dx.doi.org/10.1002/hyp.9588>.
- Brooks, R.H., Corey, A.J., 1964. Hydraulic properties of porous media. *Hydrology*, Pap. 3, Colo. State Univ., Fort Collins.
- Buizza, R., Palmer, T.N., 1995. The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.* 52, 1434–1456.
- Buizza, R., 2003. *Weather Prediction: Ensemble Prediction*. Encyclopaedia of Atmospheric Sciences. Academic Press, London.
- Carpenter, T.M., Georgakakos, K.P., 2006. Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *J. Hydrol.* 329 (1–2), 174–185.
- Ceppi, A., Ravazzani, G., Salandini, A., Rabuffetti, D., Montani, A., Borgonovo, E., Mancini, M., 2013. Effects of temperature on flood forecasting: analysis of an operative case study in Alpine basins. *Nat. Hazard Earth Sys.* 13 (4), 1051–1062. <http://dx.doi.org/10.5194/nhess-13-1051-2013>.
- Clark, A.J., Gallus Jr., W.A., Chen, T.C., 2008. Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Weather Rev.* 136 (6), 2140–2156.
- Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. *J. Hydrol.* 375, 613–626.
- Cloke, H.L., Pappenberger, F., van Andel, S.J., Schaake, J., Thielen, J., Ramos, M.-H., (Eds.), 2013. Special Issue on Hydrological Ensemble Prediction Systems (HEPS), *Hydrol. Processes*, vol. 27, no. 1, pp. 1–163.
- Congiglio, M.C., Correia Jr., J., Marsh, P.T., Kong, F., 2013. Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Weather Forecast.* 28 (3), 842–862.
- Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrol. Earth Syst. Sci.* 13, 913–921. <http://dx.doi.org/10.5194/hess-13-913-2009>.
- Drobinski, P. et al., 2014. HyMeX: a 10-year multidisciplinary program on the Mediterranean water cycle. *Bull. Am. Meteor. Soc.* 95, 1063–1082.
- Du, J., Mullen, S.L., Sanders, F., 1997. Short-range ensemble forecasting of quantitative precipitation. *Mon. Weather Rev.* 125, 2427–2459.
- Dudhia, J., 1989. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.* 46, 3077–3107.
- Evans, J.P., Ekstrom, M., Ji, F., 2012. Evaluating the performance of a WRF physics ensemble over South-East Australia. *Clim. Dyn.* 39, 1241–1258. <http://dx.doi.org/10.1007/s00382-011-1244-5>.
- Ferrier, B.S., 1994. A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.* 51 (2), 249–280.
- Fiori, E., Comellas, A., Molini, L., Reborá, N., Siccardi, F., Gochis, D.J., Tanelli, S., Parodi, A., 2014. Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: the Genoa 2011 case. *Atmos. Res.* 138, 13–29.
- Grimit, E.P., Mass, C.F., 2007. Measuring the ensemble spread-error relationship with a probabilistic approach: stochastic ensemble results. *Mon. Weather Rev.* 135, 203–221.
- Hong, S.-Y., Lim, J.-O.J., 2006. The WRF single-moment 6-class microphysics scheme (WSM6). *J. Kor. Meteor. Soc.* 42, 129–151.
- Hong, S.-Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.* 134, 2318–2341.
- Houtekamer, P.L., Derome, J., 1995. Methods for ensemble prediction. *Mon. Weather Rev.* 123, 2181–2196.
- Hsiao, Ling-Feng, Yang, Ming-Jen, Lee, Cheng-Shang, Kuo, Hung-Chi, Shih, Dong-Sin, Tsai, Chin-Cheng, Wang, Chieh-Ju, Chang, Lung-Yao, Chen, Delia Yen-Chu, Feng, Lei, Hong, Jing-Shan, Fong, Chin-Tzu, Chen, Der-Song, Yeh, Tien-Chiang, Huang, Ching-Yuang, Guo, Wen-Dar, Lin, Gwo-Fong, 2013. Ensemble forecasting of

- typhoon rainfall and floods over a mountainous watershed in Taiwan. *J. Hydrol.* 506, 55.
- Hu, X.-M., Nielsen-Gammon, J.W., Zhang, F., 2010. Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.* 49, 1831–1843.
- Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Tveit, O.E., 2008. Classifications of atmospheric circulation patterns. *Ann. N. Y. Acad. Sci.* 1146 (1), 105–152.
- Janjic, Z.I., 1994. The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* 122, 927–945.
- Jankov, I., Gallus Jr., W.A., Segal, M., Shaw, B., Koch, S.E., 2005. The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. *Weather Forecast.* 20 (6), 1048–1060.
- Lalurette, F., 2003. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Q. J. R. Meteorol. Soc.* 129, 3037–3057.
- Le Lay, M., Saulnier, G.M., 2007. Exploring the signature of climate and landscape spatial variabilities in flash flood events: case of the 8–9 September 2002 Cévennes-Vivarais catastrophic event. *Geophys. Res. Lett.* 34 (13).
- Lee, C.S., Ho, H.Y., Lee, K.T., Wang, Y.C., Guo, W.D., Chen, Y.C., Hsiao, L.F., Chen, C.H., Chiang, C.C., Yang, M.J., Kuo, H.C., 2013. Assessment of sewer flooding model based on ensemble quantitative precipitation forecast. *J. Hydrol.* 506, 101–113. <http://dx.doi.org/10.1016/j.jhydrol.2012.09.053>.
- Leoncini, G., Plant, R.S., Gray, S.L., Clark, P.A., 2013. Ensemble forecasts of a flood-producing storm: comparison of the influence of model-state perturbations and parameter modifications. *Quart. J. Roy. Meteorol. Soc.* 139 (670), 198–211.
- Lin, Yuh-Lang, Farley, R.D., Orville, H.D., 1983. Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Met.* 22, 1065–1092.
- Löwe, R., Thorndahl, S., Mikkelsen, P.S., Rasmussen, M.R., Madsen, H., 2014. Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar. *J. Hydrol.* 512, 397–407, 11.
- Marsigli, C., 2009. COSMO Priority Project Short Range Ensemble Prediction System (SREPS): Final Report. COSMO Technical Report No. 13, p. 33.
- Meneguzzo, F., Pasqui, M., Menduni, G., Messeri, G., Gozzini, B., Griffoni, D., Rossi, M., Maracchi, G., 2004. Sensitivity of meteorological high-resolution numerical simulations of the biggest floods occurred over the Arno river basin, Italy, in the 20th century. *J. Hydrol.* 288, 37–56.
- Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated- k model for the longwave. *J. Geophys. Res.* 102, 16663–16682.
- Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.* 122, 73–119.
- Montaldo, N., Ravazzani, G., Mancini, M., 2007. On the prediction of the Toce alpine basin floods with distributed hydrologic models. *Hydrol. Process.* 21, 608–621.
- Montaldo, N., Toninelli, V., Albertson, J.D., Mancini, M., Troch, P.A., 2003. The effect of background hydrometeorological conditions on the sensitivity of evapotranspiration to model parameters: analysis with measurements from an Italian alpine catchment. *Hydrol. Earth Syst. Sci.* 7 (6), 848–861.
- Moore, R.J., Cole, S.J., Bell, V.A., Jones, D.A., 2006. Issues in flood forecasting: ungauged basins, extreme floods and uncertainty. *Front. Flood Res.* 305, 103–122.
- Mullen, S.L., Baumhefner, D.P., 1988. Sensitivity to numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations. *Mon. Weather Rev.* 116, 2289–2329.
- Munich Reinsurance Company (Munich Re), 2014. Topics Geo Natural Catastrophes 2013: Analyses, Assessments, Positions, Munich Re, München, Germany, 65 p.
- Nakanishi, M., Niino, H., 2006. An improved Mellor–Yamada level 3 model: its numerical stability and application to a regional prediction of advecting fog. *Bound. Lay. Meteor.* 119, 397–407.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through the conceptual models. Part 1: A discussion of principles. *J. Hydrol.* 10 (3), 282–290.
- Pianosi, F., Ravazzani, G., 2010. Assessing rainfall–runoff models for the management of Lake Verbano. *Hydrol. Process.* 24 (22), 3195–3205. <http://dx.doi.org/10.1002/hyp.7745>.
- Pielke, R., Mahrer, Y., 1975. Representation of the heated planetary boundary layer in mesoscale models with coarse vertical resolution. *J. Atmos. Sci.* 32.
- Pleim, J.E., 2007. 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.
- Rabuffetti, D., Ravazzani, G., Corbari, C., Mancini, M., 2008. Verification of operational Quantitative Discharge Forecast (QDF) for a regional warning system – the AMPHORE case studies in the upper Po River. *Nat. Hazards Earth Syst. Sci.* 8, 161–173.
- Ravazzani, G., 2013. MOSAICO, a library for raster based hydrological applications. *Comput. Geosci.* 51, 1–6. <http://dx.doi.org/10.1016/j.cageo.2012.08.007>.
- Ravazzani, G., Barbero, S., Salandini, A., Senatore, A., Mancini, M., 2015. An integrated hydrological model for assessing climate change impacts on water resources of the Upper Po river basin. *Water Resour. Manage.* 29 (4), 1193–1215. <http://dx.doi.org/10.1007/s11269-014-0868-8>.
- Ravazzani, G., Corbari, C., Morella, S., Gianoli, P., Mancini, M., 2012. Modified Hargreaves–Samani equation for the assessment of reference evapotranspiration in Alpine river basins. *J. Irrig. Drain. Eng.* 138, 592–599.
- Ravazzani, G., Ghilardi, M., Mendlik, T., Gobiet, A., Corbari, C., Mancini, M., 2014a. Investigation of climate change impact on water resources for an Alpine basin in northern Italy: implications for evapotranspiration modeling complexity. *PLoS ONE* 9 (10), e109053. <http://dx.doi.org/10.1371/journal.pone.0109053>.
- Ravazzani, G., Gianoli, P., Meucci, S., Mancini, M., 2014b. Assessing downstream impacts of detention basins in urbanized river basins using a distributed hydrological model. *Water Resour. Manage.* 28 (4), 1033–1044. <http://dx.doi.org/10.1007/s11269-014-0532-3>.
- Ravazzani, G., Gianoli, P., Meucci, S., Mancini, M., 2014c. Indirect estimation of design flood in urbanized river basins using a distributed hydrological model. *J. Hydrol. Eng.* 19 (1), 235–242. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000764](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000764).
- Ravazzani, G., Mancini, M., Giudici, I., Amadio, P., 2007. Effects of soil moisture parameterization on a real-time flood forecasting system based on rainfall thresholds. In: Boegh, E., Kunstmann, H., Wagener, T., Hall, A., Bastidas, L., Franks, S., Gupta, H., Rosbjerg, D., Schaake, J. (Eds.), *Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management*, vol. 313. IAHS Publication, pp. 407–416.
- Reed, S., Schaake, J., Zhang, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.* 337 (3–4), 402–420.
- Romero, R., Ramis, C., Guijarro, J.A., 1999a. Daily rainfall patterns in the Spanish Mediterranean area: an objective classification. *Int. J. Climat.* 19 (1), 95–112.
- Romero, R., Sumner, G., Ramis, C., Genovés, A., 1999b. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *Int. J. Climat.* 19 (7), 765–785.
- Rulli, M.C., Rosso, R., 2002. An integrated simulation method for flash-flood risk assessment: 1. Frequency predictions in the Bisagno River by combining stochastic and deterministic methods. *Hydrol. Earth Syst. Sci.* 6, 267–284. <http://dx.doi.org/10.5194/hess-6-267-2002>.
- Scherrer, S.C., Appenzeller, C., Eckert, P., Cattani, D., 2004. Analysis of the spread-skill relations using the ECMWF ensemble prediction system over Europe. *Weather Forecast.* 19 (3), 552–565.
- Seibert, P., Frank, A., Formayer, H., 2007. Synoptic and regional patterns of heavy precipitation in Austria. *Theor. Appl. Climat.* 87 (1–4), 139–153.
- Silvestro, F., Rebora, N., Giannoni, F., Cavallo, A., Ferraris, L., 2015. The flash flood of the Bisagno Creek on 9th October 2014: an “unfortunate” combination of spatial and temporal scales. *J. Hydrol.* <http://dx.doi.org/10.1016/j.jhydrol.2015.08.004>.
- Skamarock, W.C., et al., 2008. A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-4751STR, 125 p.
- Stensrud, D.J., Bao, J.-W., Warner, T.T., 2000. Using initial and model physics perturbations in short-range ensemble simulations of mesoscale convective events. *Mon. Weather Rev.* 128, 2077–2107.
- Tao, W.K., Simpson, J., McCumber, M., 1989. An ice-water saturation adjustment. *Mon. Weather Rev.* 117, 231–235.
- Tapiador, F.J., Tao, W.K., Shi, J.J., Angelis, C.F., Martinez, M.A., Marcos, C., Rodríguez, A., Hou, A., 2012. A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain. *J. Appl. Meteorol. Climatol.* 51 (3), 489–504.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106, 7183–7192.
- Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M.A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., Cuenca, R.H., 2004. Implementation and verification of the unified Noah land surface model in the WRF model. In: 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, pp. 11–15.
- Thompson, G., Field, P.R., Rasmussen, R.M., Hall, W.D., 2008. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Weather Rev.* 136, 5095–5115.
- Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteor. Soc.* 74, 2317–2330.
- Trigo, R.M., DaCamara, C.C., 2000. Circulation weather types and their influence on the precipitation regime in Portugal. *Int. J. Climat.* 20 (13), 1559–1581.
- Vincendon, B., Ducrocq, V., Nuissier, O., Vié, B., 2011. Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting. *Nat. Haz. Earth Syst. Sci.* 11, 1529–1544.
- Yang, Tsun-Hua, Yang, Sheng-Chi, Ho, Jui-Yi, Lin, Gwo-Fong, Hwang, Gong-Do, Lee, Cheng-Shang, 2015. Flash flood warnings using the ensemble precipitation forecasting technique: a case study on forecasting floods in Taiwan caused by typhoons. *J. Hydrol.* 520, 367–378.
- Zappa, M., Fundel, F., Jaun, S., 2013. A ‘Peak-Box’ approach for supporting interpretation and verification of operational ensemble peak-flow forecasts. *Hydrol. Process.* 27 (1), 117–131. <http://dx.doi.org/10.1002/hyp.9521>.
- Zsótér, E., 2006. Recent developments in extreme weather forecasting. *ECMWF Newsletter*, vol. 107. Spring, pp. 8–17.