



Simulation of a persistent medium-term precipitation event over the western Iberian Peninsula

S. C. Pereira¹, A. C. Carvalho², J. Ferreira², J. P. Nunes³, J. J. Keizer³, and A. Rocha¹

¹CESAM and Dept. Physics, University of Aveiro, Aveiro, Portugal

²CENSE and Dept. of Science Environmental Engineering, Faculty of Science and Technology, University NOVA of Lisboa, Almada, Portugal

³CESAM and Dept. Environment and Planning, University of Aveiro, Aveiro, Portugal

Correspondence to: S. C. Pereira (susana.cardoso@ua.pt)

Received: 20 December 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 28 January 2013

Revised: 22 July 2013 – Accepted: 6 August 2013 – Published: 2 October 2013

Abstract. This study evaluated the performance of the WRF-ARW (Weather Research and Forecasting with Advanced Research) weather prediction model in simulating the spatial and temporal patterns of an extreme rainfall period over a complex orographic region in north-central Portugal. The analysis was performed during the rainy season and, more specifically, the month of December 2009. In this period, the region of interest was under the influence of a sequential passage of low-pressure systems associated with frontal surfaces. These synoptic weather patterns were responsible for long periods of rainfall, resulting in a high monthly precipitation.

The WRF model results during the study period were furthermore evaluated with the specific objective to complement gaps in the precipitation recordings of a reference meteorological station (located in Pousadas), the data of which are fundamental for hydrological studies in nearby experimental catchments.

Three distinct WRF model runs were forced with initial fields and boundary conditions obtained from a global domain model: (1) a reference experiment with no nudging (RunRef); (2) observational nudging for a specific location, i.e. the above-mentioned Pousadas reference station (RunObsN); and (3) nudging to the analysed field (RunGridN). Model performance was evaluated, using several statistical parameters, against a dataset of 27 rainfall stations that were grouped by elevation. The three model runs had similar performances, even though RunGridN resulted in a slight improvement. Regarding the other two experiments, this improvement justifies its use for complementing

the surface measurements at the Pousadas reference station. Overall model accuracy, expressed in root mean square error (RMSE), of the three runs was comparable for the stations of the different elevations classes. Even so, it was slightly better for stations in the lowlands than the highlands. Furthermore, model predictions tended to be less accurate for stations located in rough terrain and deep valleys.

1 Introduction

Deterministic modelling of the complex interactions in nature is a valuable instrument for scientists as well as policy makers. Hydrological modelling is now widely used for addressing present and future problems such as water availability for agricultural purposes and human consumption, surface water contamination, and flooding risk.

Both short and high-intensity as well as prolonged and low-intensity rainfall events can play a key role in catchment-scale runoff generation and associated phenomena such as flooding. Flood generation processes have been described by numerous authors (e.g. Chow et al., 1988). Infiltration-excess runoff generation, when rainfall intensity exceeds the infiltration capacity of soils, can be linked with flash floods in small headwater catchments. Saturation-excess runoff generation, when large amounts of rainfall cause soils to become saturated and prevent further infiltration, can be associated with prolonged floods at larger spatial scales. The characteristic spatio-temporal scale of infiltration-excess runoff is small, ranging from minutes to hours and from 1 to 100 km²,

while the scale of saturation-excess runoff is typically related to that of storm systems and weather fronts, ranging from hours to days and from 100 km² to continental-scale river basins (Skøien and Blöschl, 2003). In the case of Mediterranean-type catchments the maximum rainfall intensity during 30 min has been indicated by several authors as critical for surface runoff generation (e.g. Castillo et al., 2003; Kirkby et al., 2005). Therefore, an analysis of the rainfall events that can provoke flooding must take these spatial and temporal scales into account.

Rainfall–runoff studies generally use measured rainfall data as input for analysis and modelling (e.g. Singh and Frevert, 2002). Although “point” rainfall measurements recorded by ground stations are considered to be reliable, they tend to be sparse and highly variable in space as well as over time (AghaKouchak et al., 2010). Furthermore, even a high-density ground network may not adequately capture the characteristic dimensions of the rainfall distribution (Hershfield, 1967). Advances in computational power are now making it possible to use numerical weather prediction (NWP) models for simulating precipitation processes with a spatial and temporal resolution that is adequate for many hydrological applications. The NWP model used in this study uses a spatial grid cell resolutions of around 1 km × 1 km and temporal resolutions of a few seconds. In fact, NWP models have been used in climatic studies in order to evaluate the uncertainties in temperature and precipitation data that are used as input for hydrological models (e.g. Kotlarski et al., 2005; Akhtar et al., 2009). He et al. (2009) used global ensemble weather predictions systems to provide a probabilistic and flood inundation forecast in an attempt to increase the forecast lead times.

The physical basis of NWP models allows to use them to test explicitly the current understanding of key meteorological processes and, thereby, to provide a more solid foundation for the explanation of meteorological measurements. In the case of precipitation simulations, the obtained results depend on key factors such as model domain horizontal resolution (e.g. Heikkilä et al., 2011; Soares et al., 2012; Luna et al., 2011), model domain size and position (Ferreira et al., 2010a), vertical resolution (Aligo et al., 2009), physical parameterisations (Fernandez et al., 2007; Awan et al., 2011), explicit cumulus and/or cumulus scheme parameterisations (Clark et al., 2007; Lenaerts et al., 2009; Luna et al., 2011), associated seasonal weather systems (Awan et al., 2011; Soares et al., 2012) and initial conditions (Lo et al., 2008; Jankov et al., 2007).

Soares et al. (2012) applied the Weather and Research Forecast (WRF) model to obtain a climatology of precipitation over Portugal, using the ERA-Interim reanalysis data as input forcing fields. Their study highlighted the importance of the fine resolution in obtaining extreme precipitation values for a grid with a 9 km resolution over north-western Portugal, i.e. the wettest region, also containing the domain used in the present study. Heikkilä et al. (2011) obtained com-

parable results with WRF for the North Atlantic and Norway (using a grid with a 10 km resolution). Luna et al. (2011) demonstrated the importance of the horizontal resolution in simulating local precipitation events over Madeira Island. In the same study, horizontal resolution was found to become a less critical factor when rainfall was averaged in space and time, indicating that the total amount of water within the domain remained approximately the same but its temporal and spatial distribution was affected. Experiments with idealised WRF runs with 3 km resolution showed that, for shallow convection, model performance tended to be worse when using the Grell cumulus parameterisation scheme and avoiding cumulus parameterisation (i.e. by explicitly resolving it) than when using the Kain–Fritsch scheme (Lenaerts et al., 2009). The above-mentioned works showed that model resolution is suitable for using explicit cumulus. However, this choice may not always correspond to better simulations.

An important motivation for the present study stemmed from the poor precipitation records that exist for the study region, hampering ongoing research work on fire-enhanced runoff and erosion (e.g. Malvar et al., 2011; Campos et al., 2012; Prats et al., 2012) and on the impacts of land use on water availability and quality (Ferreira et al., 2010b; Rial-Rivas et al., 2011). The region’s precipitation fields are not well known due to the scarcity of rain gauges and the large distances of the existing radar station. These factors led the authors to install, in 2005, the Pousadas meteorological station as a reference station for four nearby experimental catchments. However, gaps in the recordings of meteorological stations can hardly be avoided altogether, as was demonstrated well by the time period that was selected for this study (during which the station’s battery failed to recharge sufficiently due to prolonged cloudy weather). In the case of the Pousadas station, missing rainfall data cannot be estimated with sufficient accuracy from radar stations. The nearest radar station is located at approximately 250 km, and the agreement between radar-based precipitation estimates and point measurements was found to decrease with increasing distance (Sebastianelli et al., 2010). The mountainous terrain of the study area might introduce further errors in the radar-based precipitation estimates, by physically obstructing the radar’s effective coverage (Pellarin et al., 2002). NWP models are, therefore, a viable and useful alternative to estimate precipitation fields with an adequate spatio-temporal resolution.

This study assessed the performance of a NWP model – WRF – over a complex orographic region in order to provide estimates for missing data in existing rainfall time series. In particular, it addressed how different approaches to applying WRF affect the quality of the model results. To this end, three different WRF runs were carried out to evaluate if a simulation that was forced with just initial and boundary conditions would perform worse than a simulation involving data assimilation of observations over a defined location, or a simulation employing the grid-nudging technique.

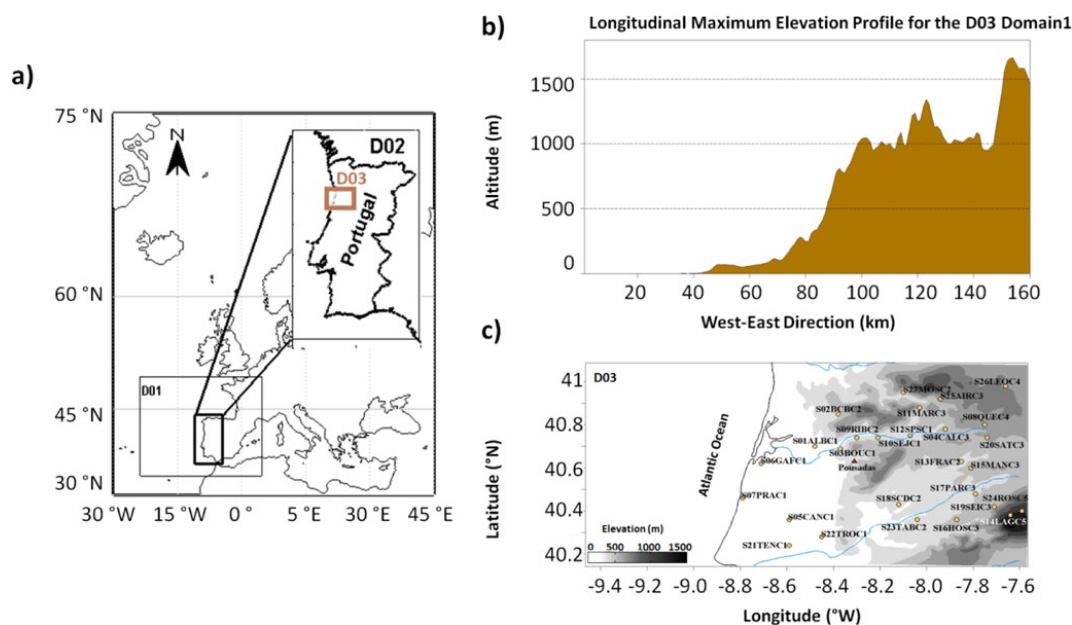


Fig. 1. Study area and stations location over the study area. (a) Nested domains for the WRF model experiments showing the outermost domain (D01) with 25-km resolution, the middle domain (D02) with 5 km of resolution and the innermost domain (D03) with 1-km resolution. The D03 frame marks the study area; (b) longitudinal maximum elevation profile for the 1-km domain; (c) location of the rain-gauge stations (yellow dots) over the 1-km domain and the location of Pousadas (red triangle).

2 Methods and data

2.1 Study area and case study

The study area spanned a mountainous region in north-central Portugal (Fig. 1). The climate is classified as wet Mediterranean, according to Köppen–Geiger climate classification, with a mean annual rainfall ranging from 800 mm at the littoral zone to 2300 mm in the inland mountains due to the marked influence of topography on spatial rainfall patterns. The Águeda river catchment is located in this region, an important watershed subject of recent studies (Figueiredo et al., 2009) and well known for its flooding risk to the old city centre of Águeda.

The present analysis focused on the month of December 2009, combining an exceptional amount of rainfall with the occurrence of various gaps in the records at the Pousadas meteorological station (a reference station for ongoing precipitation–runoff studies, as detailed earlier). The existing rainfall stations in the region recorded monthly totals, for December, that were, on average, about 88 % above their long-term median values (The Portuguese Water Institute, Instituto da Água, I.P., INAG, 2011) and, as such, corresponded to the stations' 54 to 95 percentiles for December (Table 1). The return period of monthly rainfall was approximately 3 yr, but in four stations located in the lowlands of the Mondego river valley (southern and eastern parts of the study area) the return period was higher, between 5 and 11 yr.

An analysis of maximum daily rainfall in December (Table 1) found a higher return period, c. 7 yr, with maxima on average 54 % above median values, but with a high dispersion of percentiles between 29 and 95. The stations in the south and south-east (also inside the Mondego valley) had the rainfall maxima on 6 December, with a return period between 7 and 18 yr. In this area, part of the higher-than-average monthly rainfall can be attributed to this daily event (between 15 and 25 % of total monthly rainfall). Stations in the north-west of the study area, on the sea-facing side of the coastal mountain range (Fig. 1), had the maxima on 28 December, with a return period between 2 and 7 yr.

Given the low reliability of the evaluation of return periods using daily maxima, a more detailed comparison at the (sub-)daily scale was performed for the only station with long-term intensity–duration–frequency curves (usually expressed as IDF curves) in the dataset, Santa Comba Dão (code S18SCDC2) (Table 2), which is located in the Mondego valley. The IDF curves are created with the precipitation records over a certain location in a three-axis graph and they represents a probability (e.g. return periods of certain amounts). The IDF curve for Santa Comba Dão indicate that the December 2009 values corresponded to return periods of less than 2 yr for short-term rainfall durations (< 3 h) and between 2 and 5 yr for longer durations (6–24 h), returning to under 2 yr return periods for a duration of 48 h. Therefore, the high rainfall of December 2009 could be attributed to a longer-duration event lasting between 12 and 24 h, rather than a short-term, high-intensity precipitation

Table 1. Basic precipitation statistics for each station for December 2009.

Station	Station ID	Long-term observed data					Long-term observed data						
		PP (mm)	Median PP (mm)	Dec 2009 vs. median	Percentile (of all obs. data)	Return period (yr)	Number of years	PP (mm)	Median (mm)	Dec 2009 vs. median	Percentile	Return period (yr)	Number of years
Bouçã	S03BOUC1	296.8	258.3	15 %	67 %	2.9	25	47.0	51.5	-9 %	41 %	1.7	23
São Pedro do Sul	S12SPSC1	130.7	119.1	10 %	54 %	2.2	55	53.6	32.5	65 %	89 %	7.7	45
Tentugal	S21TEMCI	285.0	126.7	125 %	89 %	8.1	56	69.1	33.4	107 %	96 %	18.3	54
Trouxenil	S22TROCI	196.5	127.6	54 %	71 %	3.3	29	45.7	27.0	69 %	93 %	10.0	29
Barragem de Castelo Burgães	S02BCBC2	301.9	214.5	41 %	69 %	3.1	71	48.6	48.6	0 %	50 %	2.0	37
Mosteiro de Cabril	S27MOSC2	204.3	159.5	28 %	63 %	2.7	60	56.2	33.0	70 %	90 %	7.3	21
Fragosela	S13FRAC2	277.4	79.5	249 %	95 %	11.5	22	49.5	24.0	106 %	90 %	7.3	21
Santa Comba Dão	S18SCDC2	262.6	110.7	137 %	83 %	5.6	77	66.4	32.1	107 %	95 %	15.6	77
Calde	S04CALC3	196.1	115.2	70 %	70 %	3.1	24	22.7	29.9	-24 %	30 %	1.5	24
Oliveira do Hospital	S16OLIC3	144.9	107.1	35 %	59 %	2.4	78	19.5	27.4	-29 %	29 %	1.4	79
Sátão	S20SATC3	340.1	111.7	204 %	91 %	9.8	48	51.9	33.3	56 %	89 %	8.2	48
Lagoa Comprida	S14LAGC5	366.7	201.5	82 %	79 %	4.5	47	142.8	63.0	127 %	> 95 %	> 6	12

Table 2. Maximum rainfall (mm) according to return period and maximum observed in December 2009 for the S18SCDC2 station (Brandão et al., 2001).

Return period (yr)	Duration (h)						
	1	2	3	6	12	24	48
2	17.5	22.4	25.9	33.3	43.1	55.4	71.2
5	25.1	31.7	36.3	45.8	57.9	73.2	92.4
10	30.0	37.7	43.1	54.1	67.2	85.3	108.3
50	41.1	51.1	58.1	72.3	88.4	111.6	141.0
100	45.7	56.8	64.4	80.0	97.3	122.7	154.7
Maximum for December 2009	9.6	14.5	18.7	34.5	49.1	67.6	68.2

event. The discrepancy in return periods for this station between Tables 1 and 2 should indicate that the actual return periods for 24 h maxima in other stations was also lower than indicated.

2.2 Model setup

The regional meteorological model used in this study is the WRF model with Advanced Research (ARW) dynamic core version 3.1.1 (Skamarock et al., 2008). WRF is a next-generation, limited-area, non-hydrostatic mesoscale modelling system, with vertical terrain-following eta coordinate designed to serve both operational and forecasting as well as atmospheric research needs. The WRF-ARW model has been widely used for simulating precipitation processes, both in forecast (Deb et al., 2010; Weisman et al., 2008) and diagnostic mode (J. Liu et al., 2012; Lou and Breed, 2011; Bukovsky and Karoly, 2009).

It has also been used in Portugal, in a sensitivity test to parameterisations for two different model operational configurations (Ferreira et al., 2010a), in climate simulations over Portugal (Soares et al., 2012) and over the Andalusia region in Spain (Argüeso et al., 2011). Previously, Fernández et al. (2007) performed regional climate simulations over the Iberian Peninsula using the predecessor of the WRF-ARW model, MM5 (short for Fifth-Generation Penn State/NCAR Mesoscale Model). These authors have performed a sensitivity test to the model parameterisations during a five-year period. Regarding precipitation, the authors pointed out that the orography representation by the model has a larger impact on the modelled precipitation in winter than in summer.

Model horizontal resolution has also importance on the simulation of local precipitation. Soares et al. (2012) highlighted the significance of the model's fine resolution in order to obtain precipitation extreme values on a 9 km-resolution grid, namely in the wettest region of Portugal, i.e. the north-western region.

In the present study the WRF-ARW model was forced with the analysis fields of the Global Forecast System (GFS),

from the United States of America's National Center for Environmental Prediction (NCEP), generated every 6 h, from 30 November 2009 until 31 December 2009. The GFS model has an approximated horizontal resolution of $0.5^\circ \times 0.5^\circ$, and the vertical domain extends from a surface pressure of 1000 to 0.27 hPa, discretised in 64 vertical unequally spaced sigma levels, from which 15 levels are below 800 hPa and 24 levels are above 100 hPa.

The WRF-ARW model was configured with three nested domains, operating in two-way nesting mode, with horizontal resolutions of 25 km (D01), 5 km (D02) and 1 km (D03), for the parent, middle and inner domains, respectively. The finer grid domain is centred over Pousadas (40.63° N, 8.31° W) (see Fig. 1).

Due to its applicability to mid-latitudes, the Lambert conformal conical projection is used with the standard parallel at 40.63° N. The three nested domains identified have the Atlantic Ocean as their western border to better capture the dominant atmospheric circulation patterns that account for the major daily precipitation observed in the region (Trigo and DaCamara, 2000). This also avoids some complications with the vertical interpolation due to differences between the GFS and WRF topography in that boundary (Lo et al., 2008). The vertical discretisation in WRF consists of 27 eta levels. The physical parameterisation schemes used in this work resulted from a previous study conducted by Ferreira et al. (2008), in which several parameterisation sets were tested against observations of temperature, water vapour, mixing ratio and wind at several stations over mainland Portugal, using the WRF-ARW model with the same configuration as the one used in the present work, only for the D01 and D02 domains. The physical parameterisation set is the following: WRF Single Moment 6 (WSM6) microphysics scheme (Hong and Lim, 2006); Dudhia shortwave radiation (Dudhia, 1989); Rapid Radiative Transfer Model (RRTM) longwave radiation model (Mlawer et al., 1997); MM5 similarity surface layer scheme (Skamarock et al., 2008), Yonsei University (YSU) planetary boundary layer scheme (Hong et al., 2006); Noah Land Surface Model (Chen and Duhia, 2001); Grell–Devenyi ensemble convective parameterisation scheme (Grell–Devenyi, 2002). A sensitivity test regarding the cumulus parameterisation in domain D03 was made for the control simulation, in which the Grell–Devenyi parameterisation was tested against an explicit precipitation computation simulation. The mean error, mean square error and the root mean square error (see Appendix for metrics definitions) of both simulations were compared for the precipitation thresholds of 0.1, 1, 2 and 3 mm h⁻¹. The results, with these metrics, are similar giving advantage to the Grell–Devenyi parameterisation scheme simulation. Hence, this parameterisation was used in all three domains.

2.3 Experimental design

Three numerical experiments, corresponding to integrations with one month of duration plus 24 h of spinup (which were discarded), were performed for December of 2009, starting at 00:00 UTC on 30 November 2009 and ending at 00:00 UTC on 1 January 2010. In order to test for improvements in the model simulations, two nudging techniques were applied (Skamarock et al., 2008) and compared with a simulation without nudging (RunRef).

Nudging is a method that keeps simulations close to the analysis and/or observations (input fields) over the course of integrations. In the WRF-ARW model, there are two types of nudging that can be used separately or combined. One is the observational or single location nudging that forces the simulation towards observational data. The other is the grid nudging which forces the model simulation towards a series of analysis grid point by grid point. As from the WRF version 3.1.1, the option of spectral nudging was activated allowing the nudging towards waves, under selected wave numbers. The advantage of the spectral nudging method is that it maintains the regional model in phase with large-scale circulation while permitting the small-scale flow to be calculated accordingly, without the forcing field's information. Nowadays, this is the most common method of nudging (Rummukainen, 2010). The grid-nudging technique will be applied to all scales into the WRF flow computation. A study comparing the former two techniques is presented by P. Liu et al. (2012), where it is stated that, with the appropriate choice of wave numbers, the spectral nudging outperforms grid nudging, in a WRF application with 36 km grid cell resolution. Examples of applications of grid nudging are the studies by Soares et al. (2012) and Fernández et al. (2007), who used WRF two-way nesting simulation using grid nudging in the coarser domain. Argüeso et al. (2011) also applied WRF in two-way nesting using the spectral nudging technique, over a domain covering southern Spain, with a grid cell resolution of 10 km.

In the present application, nudging was carried out to individual observations over the location of Pousadas (RunObsN), in order to evaluate the impact of local information in the computation of model precipitation at that site specifically.

In the RunObsN simulation, a single site observation with measures of wind, temperature and humidity was available at the area of interest to nudge the model integration, at Pousadas (see Fig. 1). It is expected that, when considering all stations and in the absence of nonlinear effects, this simulation will have similar results to the RunRef simulation. However, at stations near the Pousadas site, this may have some influence that may be worth noting. A third experiment, consisting in applying the grid-nudging technique (RunGridN), was conducted, with the purpose of investigating the impact of 3-D analysis nudging to constrain the circulation within the mesoscale model. The grid nudging was

applied to the entire atmospheric column except the planetary boundary layer, to wind, temperature and humidity meteorological variables, as performed by Lo et al. (2008) for all of the computational domains.

The RunObsN simulation may be regarded as a weak constraint simulation, so the RunGridN experiment was performed in order to test the results of the model to a stronger constrain, where grid nudging was performed across the three simulation domains, including the D03 domain, although only six points of the GFS analysis lie inside this domain, and two of them are over the Atlantic Ocean. This approach is different from those of Soares et al. (2012), Argüeso et al. (2011) and Fernández et al. (2007), since these authors used nesting and grid nudging only in the coarser domain. However, the present application is in line with the work of Lo et al. (2008), who applied the grid-nudging technique to the domain of analysis.

2.4 Rainfall measurements and observations from gridded data

To assess the model performance, a set of 27 existing rainfall stations from the Portuguese National Information Service of Water Resources (SNIRH; The Portuguese Water Institute, Instituto da Água, I.P., INAG, 2011) were selected for this study (Fig. 1). The SNIRH dataset consists of a series of rain gauge stations, recording with a time resolution of one minute, but only totals at the hourly resolution and above are available online (www.snirh.pt). The time period is not the same for all the stations, but the majority has a common period of 22 to 56 yr. The rain gauge locations are unevenly distributed over the study area with variable density. The data were checked for gross errors, like mistyped rainfall amounts, and then compared with nearby stations, when possible, to ensure that the rainfall amounts were consistent between stations with similar characteristics.

In addition to the rainfall data described in the previous section two additional datasets were used to assess the WRF model results. For temperature, relative humidity, sea level pressure and winds the ERA-Interim reanalysis dataset (ERA) was used, from the European Centre for Medium-Range Weather Forecasts, (ECMWF, <http://www.ecmwf.int>). For precipitation the EOBS gridded dataset from the European Climate Assessment and Dataset project (ECAD, <http://eca.knmi.nl/>) was used.

The ERA dataset is a reanalysis project of the global atmosphere covering the period starting at 1979 until present day. The dataset consists in a variety of meteorological variables with different resolutions and time steps for the several vertical pressure levels and surface. A full description of the forecast model, data assimilation method, and input datasets used to produce the ERA data, as well as the performance of the system, can be found in Dee et al. (2011). A detailed description of the ERA product archive can be found at Berrisford et al. (2009). For this study the ERA data were chosen with a

horizontal resolution of $0.25^\circ \times 0.25^\circ$ and 6-hourly time step and with 3-hourly time step for the sea level pressure.

The EOBS dataset consists in a set of gridded daily observations for precipitation. The dataset covers the period starting at 1 January 1950 and ending at 30 June 2012 covering the spatial region of Europe. A full description of the dataset can be found at Haylock et al. (2008). For this study the EOBS version 7 with a regular horizontal resolution of $0.25^\circ \times 0.25^\circ$ was used.

The chosen spatial coverage for both datasets extends from latitude 34 to 49° N and for longitudes starting at 20° W to 0.5° E. Thereby the observational grids matched the WRF coarser grid (D01 domain).

2.5 Assessment of model performance

The observations were compared with the model simulations for identical locations and times. Model data were recorded every 15 min for the 1-km-resolution domain, and hourly accumulations were calculated from these values, to match the temporal scale of the observations. Concerning the spatial scale, the observations and the model precipitation are represented on a non-matching grid. Two common methods are used for comparison, namely spatial interpolation of the modelled series of precipitation to the station location, or selection of the grid point nearest to the station location. To minimise the error, two types of series were obtained from the model: one interpolated at station location and another one from the nearest grid point to the station location.

The two-model series, the spatially interpolated and the one from the nearest point, were compared by calculating the respective deviations from the observations. The average value of the absolute deviations (MD) was calculated to investigate which had the lowest deviation. The deviation between both interpolated series and the observation series were calculated. No difference was found between the averaged MD value calculated using the interpolated series ($MD=0.57 \text{ mm h}^{-1}$) and the nearest grid point to observations ($MD=0.58 \text{ mm h}^{-1}$). Thus, the series from the model grid point nearest to the station location were chosen ignoring the correspondent error on location. Although there is not a consensual strategy concerning direct verification, i.e. “truth” observations and model precipitation, these results are consistent with those presented by Rossa et al. (2008). These authors showed that verification using the nearest grid point gives very similar overall results.

The result was a set of 27 point precipitation series of the paired observations and simulations, each one with a length of 745 elements corresponding to hourly accumulations of precipitations from 00:00 UTC on 1 December 2009 to 23:00 UTC on 31 December 2009.

Some basic statistics were calculated: mean, median, mode, standard deviation and three-hour correlation for lags between -24 and $+24$ h (every three hours).

Table 3. Contingency table of counts for a binary type of event.

		Yes Obs rain $\geq t$	No Obs rain $< t$
Yes	WRF rain $\geq t$	a (hits)	b (false alarms)
No	WRF rain $< t$	c (misses)	d (correct negatives)

The strategy of evaluation comprises a set of statistical measures following Murphy and Winkler (1987), Jolliffe and Stephenson (2003) and Wilks (2006). Two approaches were followed: one using the continuous verification measures for rain amounts, and another for the occurrence of precipitation making use of the measures derived from a contingency table. All the mathematical formulation is described in the Appendix.

The selected continuous indices were the mean error (ME) and the root mean square error (RMSE). To evaluate the model performance a skill score (SS) was derived by comparing the mean square error (MSE) with a low-skill forecast, in this case the climatological MSE (MSE_{Clim}).

When using the MSE measure as the base for calculating the skill score, the last is called reduction of variance because the skill score formulation represents the ratio between the squared deviations and the observed variance (Eq. A5 in Appendix). The skill score is expected to be maximum at a value of 1 (perfect score) and minimum at a value of 0, which indicates that the model is equivalent to climatology. For MSE, a negative value indicates model performances worse than climatology, although it does not necessary imply that the model has no skill at all (Jolliffe and Stephenson, 2003).

To validate the capability of the model in reproducing the synoptic patterns and the precipitation, the continuous measures MD and the RMSE were used. The pattern correlation coefficient (PTC) was used to measure the overall agreement between the simulations and the observations grid patterns.

The occurrence of precipitation is considered as a categorical (yes/no) event type that can be defined as the precipitation meeting or exceeding a specific threshold, t (a yes event); otherwise it is a non-event. The verification measures for these events are derived from a 2×2 contingency table of counts, as shown in Table 3, of the four possible combination pairs of y_i and o_i that meet the event criteria.

The categorical measures include the frequency bias (B), the percentage of corrected events (PC), the probability of detection (POD), the false alarm rate (F) and the equitable threat score (ETS).

Better model performances are expressed by high values of POD (varying between 0 and 1) and ETS (varying between $-1/3$ and 1) and low values of F (varying between 0 and 1), and ME, combined with B near the unity. The PC result is a measure of the model’s accuracy, giving an overall percentage of how well the model simulated the precipitation.

Table 4. Daily weather regime classification for December 2009.

Number of days (d)	Weather types	EOBS PP (mm)	WRF PP (mm)
6	NW	16	38
3	AA	1	1
12	C	109	151
8	E	16	22
2	R	9	13

The verification with continuous measures was done at each station location (to yield a score for each station individually) and also for the pooled sample of all point precipitations with observed hourly values above 0.1 mm h^{-1} . This procedure resulted in a series of simulated and observed matched pairs with different lengths for each station.

The verification measures were tested for different precipitation thresholds t (0.1, 1, 2, 3, 4, 5, 10, 15 and 20 mm h^{-1}). It is perhaps worth stressing that the model results were neither rescaled nor transformed; they were used as they were.

To test the model ability in reproducing the orographic precipitation, the stations were grouped by altitude. The division by altitude classes was made considering a 200 m interval. Thus, the first classes (C1) comprise the first 200 m, the second class (C2) ranges from 200 to 400 m and so on until reaching the last class (C5), which corresponds to altitudes above 800 m.

3 Results and discussion

The WRF-ARW model was used to simulate hourly precipitation over a high spatial resolution (1-km) study domain of complex terrain for the month of December 2009. Three experiments were performed corresponding to three different configurations of the model: without nudging (Run-Ref), with local nudging (RunObsN) and with grid nudging (RunGridN).

The synoptic features were assessed comparing the domain D01 (25-km) with observations from the EOBS and ERA datasets (Sect. 3.1).

The validation of domain D03 (1-km) consisted of the direct comparison between the observed series of precipitations from a network of rain gauges with the series extracted from the model (Sects. 3.2 and 3.3). Direct validation between in situ observations and series extracted from the model can be a source of uncertainty due to the non-matching grids.

3.1 Observed and simulated synoptic features

In this section the synoptic patterns during the month of December 2009 over the region of analysis are described. The analysis follows closely the one presented by Koo and

Hong (2010). The circulation patterns were obtained from the ERA data and the precipitation from the EOBS data (described in Sect. 2.4).

Instead of analysing the mean state of the atmospheric circulation, during the time period of analysis, the concept of weather type (WT), which represents typical patterns of atmospheric synoptic circulation in a region, was used. Here, the WT calculation described by Santos et al. (2005) was applied. For Portugal, five weather types were identified plus a sixth one derived from one of the regimes. The cyclone regime (C) associated with a high density of cyclonic features, the westerly (W) associated with westerly and north-westerly winds (NW), the R regime linked with the negative phase of the North Atlantic Oscillation (NAO), the AA regime linked with positive phase of NAO and the easterly (E) regime associated with a high-pressure system over the western European basin. In the present study, each day of December 2009 was associated with a specific WT. Next, the number of days of each WT was calculated and the respective accumulated precipitation simulated by the model and observed averaged for the region of study (Table 4).

The most frequent WT was the cyclonic (C) and the north-westerly (NW). In these two regimes the precipitation is linked with travelling frontal systems that extend to south covering Portugal. The mean sea level pressure patterns of these WT are shown in Fig. 2. Although the C-regime is not the most frequent it can occasionally be the dominant feature. This was the case in December 2009 when precipitation associated with this WT represents most of the monthly rainfall (Table 4). The NW is the second-largest contributor to precipitation for the period of study. Table 4 also points to a good agreement between simulated and observed precipitation associated with each WT. This analysis shows that December 2009 was indeed characterised by synoptic regimes associated with high precipitation. To further reinforce this finding, Fig. 3a, presents the December 2009 observed anomalies relative to the mean December precipitation (PP) averaged for the time period of 1950 to 2012. This demonstrates that the period of interest had above-normal precipitation in the region coincident with the finer domain (Fig. 3a).

The ERA mean 500 hPa geopotential field shows a trough located over of the ocean west of Portugal, indicating typical conditions for heavy induced precipitation (Fig. 3b). The mean sea level pressure pattern (Fig. 3c) is consistent with the geopotential height: showing a low-pressure region north-west of Portugal with a north–south gradient. These conditions are favourable for the occurrence of precipitation. The south-westerly winds from the North Atlantic (Fig. 3d) provide the advection of moisture along the south border of the cyclonic system, which has higher moisture content.

The simulated synoptic features reproduce well mean observed conditions. The mean differences between model precipitation and observations/EOBS are shown in Fig. 4a. The model precipitation was overestimated mainly over the ocean and underestimated over land, with some exceptions near the

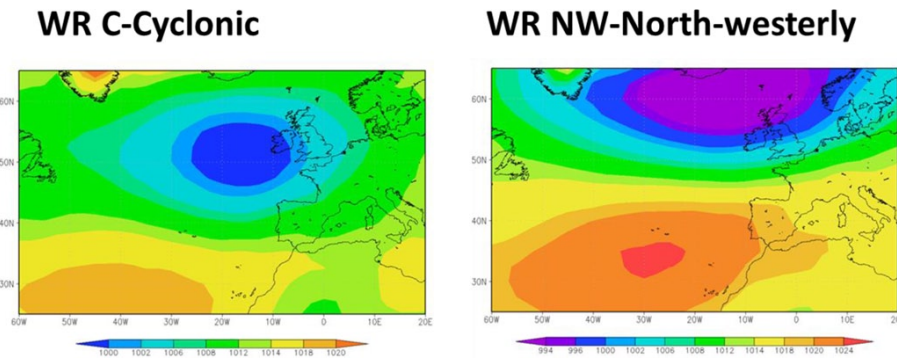


Fig. 2. Sea level pressure (SLP; hPa) depicting the weather regimes (WR) for the regime cyclonic (left panel) and north-westerly (right panel) averaged for the time period 1948–2011. Adapted from Santos et al. (2005).

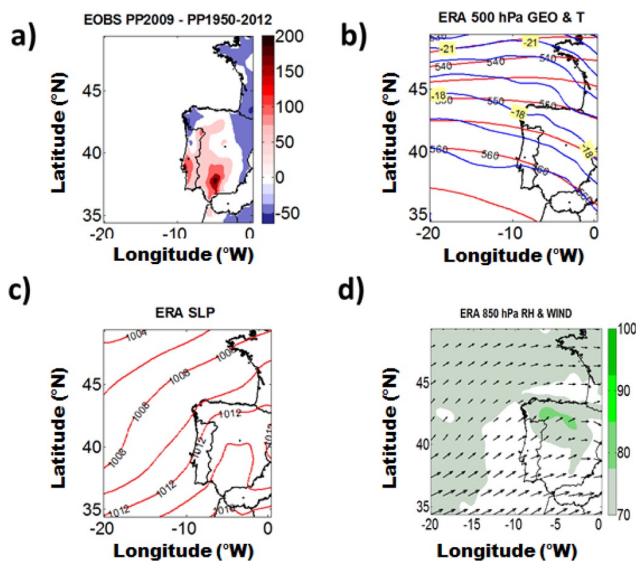


Fig. 3. (a) Precipitation difference between December 2009 and the December climatology period (1950–2012). December 2009 monthly averaged (b) 500 hPa geopotential height (dam, red contour) and temperature (K, blue contour); (c) sea level pressure (hPa) and (d) 850 hPa relative humidity (%), colour shaded) and wind vector (m s^{-1}). The reanalyses are from the ERA dataset except for precipitation that are from EOBS dataset.

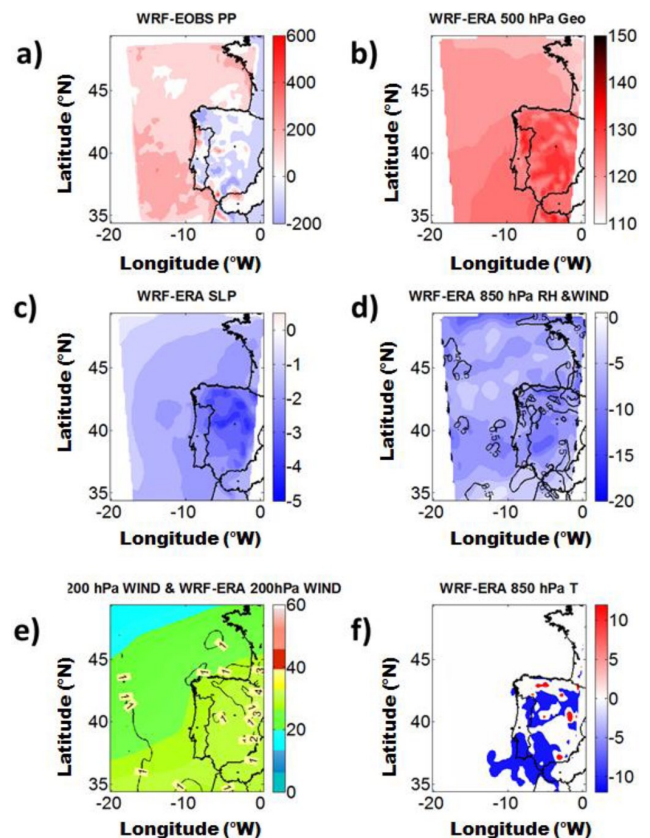


Fig. 4. Averaged differences between WRF and ERA/observations, during December 2009, for (a) precipitation (mm), (b) 500 hPa geopotential height (dam), (c) sea level pressure (hPa), (d) 850 hPa relative humidity (%), colour shaded) and wind differences (m s^{-1} , contour), (e) 200 hPa observed wind speed (m s^{-1}) averaged for December 2009 (shaded) and the difference between WRF and observations (contour) and (f) 850 hPa temperature (K). The reanalyses are from the ERA dataset except for precipitation that are from EOBS dataset.

north-western and south coast of Portugal (Fig. 4a). This positive difference is located in the northern region and covers the area defined for the finer grid domain (Fig. 1).

Mean differences between simulations and observations/ERA are shown in Fig. 4 for December 2009. The respective error measures are shown in Table 5. The WRF model overestimates the 500 hPa geopotential height (Fig. 4b; Table 5) but showed a negative bias in simulating the sea level pressure and the humidity content (Fig. 4c and d), throughout the entire domain. The highest bias values are located over land but for the central and eastern part of Iberia rather than for Portugal. The excess of model

Table 5. Statistical measurements. Mean error (ME), root mean square error (RMSE) and pattern correlation coefficient (PTC) for the WRF simulations relative to observations.

	ME	RMSE	PTC
PP ^a (mm/3 h)	0.16	7.39	0.26
500 hPa GEO ^b (dam)	229.33	90.40	0.48
SLP ^c (hPa)	3.42	7.44	0.14
850 hPa RH ^d (%)	-0.79	25.81	0.04
850 hPa U^e (ms ⁻¹)	1.89	7.84	0.02
850 hPa V^f (ms ⁻¹)	0.05	7.18	-0.03
200 hPa U^e (ms ⁻¹)	-7.97	15.91	-0.19
200 hPa V^f (ms ⁻¹)	-2.27	13.21	-0.08
850 hPa T (K)	2.27	2.74	0.57

^a PP – precipitation; ^b GEO – geopotential height; ^c SLP – sea level pressure; ^d RH – relative humidity; ^e U – zonal component of wind; ^f V – meridional component of wind.

precipitation can be caused by the enhanced 500 hPa geopotential height that tends to increase the trough located west of Portugal which are the typical conditions.

The 200 hPa wind pattern simulated by the WRF (Fig. 4e) model shows a positive bias close to the borders of the domain possibly caused by the interpolation to the observations grid. Overall, the upper troposphere winds are weaker than the observed ones. In contrast the near surface winds are zonally stronger than the observed ones (Table 5). For temperature (Fig. 4f), the WRF model underestimates the temperature for the regions located in the north and south of Iberia. There was no bias found in northern Portugal, where our area of interest is located. Overall, the WRF's higher deviations from the observations (Table 5) are related to the 500 hPa geopotential height and with the 200 hPa winds with a low pattern association. However, the differences mentioned above and shown in Fig. 4 are not statistically significant (at 5 % level).

3.2 Observed and modelled precipitation characteristics

The time lag correlations between observed and simulated precipitation were calculated (Fig. 5). The correlation values diminish when increasing the lag, with maximum values attained at lag 0 and in some cases at lag +3 h. The correlation was strong for the RunRef experiment and weak for RunGridN. For 0 and +3 lag times, the altitude class that showed strongest correspondence was C5 (> 800 m), followed by C3 (400–600 m).

Five rainfall periods were identified in the observed data, encompassing days 1–2, 4–6, 14–17, 19–25 and 27–31. Each one of the rainfall episodes was preceded and followed by, at least, a 12 h dry period. These five periods were reproduced well by all the three model runs, but the maximum observed intensity (30 mm h⁻¹ of precipitation at S08QUEC4 station)

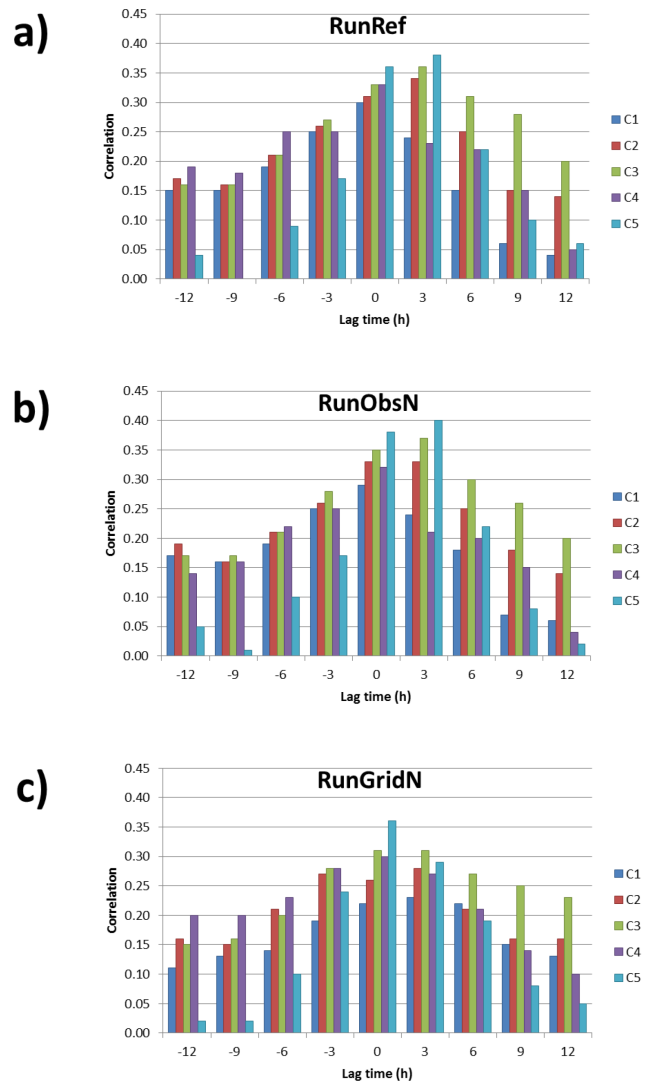
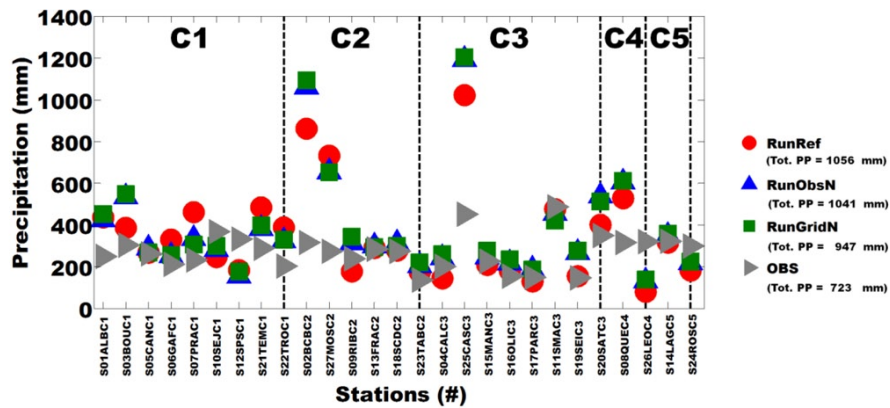


Fig. 5. Lag correlation between observations and the model series for the three experiments grouped by classes of altitude (C1 to C5).

was not. For the RunGridN experiment the majority of the series simulated a weak wet event ranging from 0 to 5 mm h⁻¹, for the 27th day, which none of the others reproduce. The total accumulated precipitation for the entire month of December at each station as well as the month totals were calculated (Fig. 6). Three locations show monthly totals much higher than the ones observed. In general, the three model runs tended to overestimate precipitation intensity.

The frequency distributions of the observed and modelled hourly rainfall amounts are shown in Fig. 7. The frequency distributions are strongly asymmetric as expected for this meteorological variable. The different model runs showed the same observed asymmetry, with median values in the range of 0.3 to 1.7 mm h⁻¹, and third-quartile values in the range of 0.3 to 5.7 mm h⁻¹. Also, the bulk of the stations revealed the existence of more extreme values (not shown)



*C1 – Altitude Classe 0-200 m; C2 – Altitude Classe 200-400 m;
 C3 – Altitude Classe 400-600 m; C4 – Altitude Classe 600-800 m;
 C5 – Altitude Classe > 800 m

Fig. 6. Monthly accumulated precipitation for the simulation period (mm) over the study domain grouped by altitude classes. The observed amounts are depicted as a grey triangle and for each of the model simulations as RunRef (red circle), RunObsN (blue triangle) and RunGridN (green square). The caption box also shows the total rainfall observed and simulated by the model in each case.

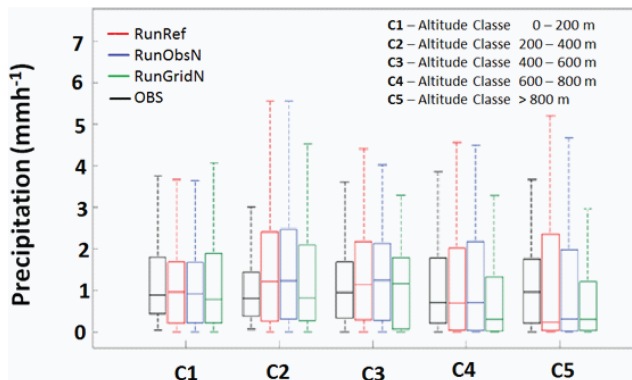


Fig. 7. Box plot for the observations (OBS; black boxes) and for the three experiments (RunRef, RunObsN and RunGridN) grouped by altitude – here represented by the C1 to C5 classes. The horizontal box line represents the median (50th percentile), the lower line the 25th percentile, and the upper line the 75th percentile. The dashed lines represent 1.5 times the IQR.

than expected, corresponding to the points lying within three times the interquartile range (IQR). The pronounced intra-variability among stations reinforced the atypical nature of the month of December 2009, as mentioned earlier. For observations only, the variability results are supported by the standard deviation values, with the majority of the individual standard deviations between the 1.0 to the 2 mm h⁻¹ interval.

3.3 Model assessment

The MD between the simulated precipitation series and observations is presented in Table 6. The MD values

range between 0.31 mm h⁻¹ (S17PARC3) and 0.92 mm h⁻¹ (S27MOSC2) for the majority of the stations, but the SO2BCBC2 (MD=1.33 mm h⁻¹) and the S25CASC3 (MD=1.50 mm h⁻¹) present an error of the same order as the respective observed mean corresponding to the three locations with monthly totals much higher than the ones observed (Fig. 6). Most stations showed a good agreement with the observations, but the stations S02BCBC2, S25CASC3 and S27MOSC2 clearly depart from the observations.

The categorical verification measures (B, PC, POD, F and ETS) together with the continuous measures (ME and RMSE) were calculated for the 27 stations as well for aggregated stations to provide a single score for the domain. The results for the pooled sample are presented in Table 7. The measures for the 0.1, 1, 2 and 3 mm-per-hour thresholds are shown. The same error measures for precipitation thresholds of 4, 5, 10, 15 and 20 mm d⁻¹ were also computed. However, for high thresholds these measures are based on very few data (very low values of a, b, c, and d, in Eqs. A7 to A12, and their robustness may be questionable and are, therefore, not shown). The experiments exhibit identical results whatever the verification measures used with a slight improvement of RunGridN.

For the categorical verification the experiments perform better for the 0.1 mm h⁻¹ thresholds and tend to deteriorate with increasing threshold value. The 0.1 mm h⁻¹ threshold yields the best pairs of verification measures with POD and ETS high and low values of F with an accuracy of 77 %. The accuracy (PC) improves with the increasing of the threshold value. However, this measure is weighted by the most frequent category and can be artificially increased by issuing more corrected negatives. Despite the F values decreasing

with threshold, POD values are low. For thresholds of 2 and 3 mm h⁻¹, 1/5 of the observed precipitation event were correctly simulated (POD = 0.2) with an ETS value of 0.1.

When analysing the continuous scores, the lowest errors were found when considering series of pairs above 0.1 mm h⁻¹. The ME values almost meet the perfect score, but the experiments overestimated the precipitation (see Fig. 8). This result was not detected in the ME due to cancellation errors. The high values obtained for RMSE suggests that the model precipitation considerably departed from the observations, as indicated by the MD values (see Table 6). The two results combined allow for the conclusion that the perfect ME achievement was supported through cancellation errors rather than agreement among the simulated and observed series. This poor scoring led to a negative skill. Usually, skill scores are designed to evaluate the model performance over some unskilled references, which in this study is the climatology. These skill results imply that the model is no better than the climatology. This could be related not to the model itself but to the horizontal resolution. J. Liu et al. (2012), while studying the best downscaling ratio for the WRF model, conclude that an increasing resolution in space may not always ensure better results in the temporal resolution, due to the higher variability of the precipitation when compared with the variability in space. Ruling out the continuous verification measures and only focusing on the categorical ones, the WRF model's best performance in capturing the occurrence of precipitation was for values above 0.1 mm h⁻¹ threshold, with 28 % of the observed precipitation events being correctly diagnosed.

Figure 8 shows the continuous measure of verification and in Fig. 9 the categorical measure, for each station grouped by altitude class. The verification measurements for pooled samples, in general, do not give the same statistics as those obtained by averaging the same verification measurement. Thus, the verification measurements presented by altitude are obtained from a pooled sample and not by averaging the individual station values.

The continuous measures of verification show that RunRef and RunObsN produce identical outcome results while RunGridN slightly departs from the others. For the three experiments the mean error is small. For all stations and altitude classes RMSE is high, leading to a negative skill.

For categorical measures (Fig. 9), estimated quantities of B above unity indicate a tendency to overestimate precipitation occurrences and the opposite to underestimate. Stations within the same altitude class were pooled to yield a single score for that particular class. The majority of the stations showed a tendency to overestimate precipitation occurrences with a few exceptions mainly in the first altitude class. The accuracy (PC) among stations is high and, therefore, within altitude classes. The PC values ranged from about 75 % for the first class to about 73 %, with RunGridN more accurate than the others. The individual measurements for POD and F are similar among stations and experiments. On aver-

age, the model was able to diagnose 67 % (average POD) of the observed 'yes' precipitation with a correspondence between hits of 29 % (average ETS). The model incorrectly diagnosed 20 % (average F) of the simulated precipitation as a rain event when it was not. The analysis by altitude class slightly exceeded the individual scores specially when considering the RunGridN experiment. For this case the C4 class scored better than the others, with 28 % (ETS) of correspondence between the observed and diagnosed hits and with 88 % (POD) of the observed events being correctly diagnosed by the model against 28 % (F) of hits on non-rain days.

Limited-area NWP models used for this type of studies show a high dependence on initial conditions. Lo et al. (2008) compared three WRF model setups with one single initialisation, with and without nudging of the meteorological fields every 6 h, and weekly initialisation, all with an update of boundary conditions every 6 h, for a medium- to long-term run. They came to the conclusion that one single initialisation with fields nudging gives the best scores. Jankov et al. (2007) have found that the simulated rainfall amounts and rates were dependent on the initial conditions used in the WRF-ARW model, as well as on the physical schemes applied. Specifically, they have come to the conclusion that, for hydrological purposes where higher accuracy on the amount of rain is the most important variable, the WRF model was sensitive to the initialisation datasets and physical parameterisations, whereas the rain rate was more sensitive to the cumulus parameterisations applied. Limited-area models also show seasonal dependence of model skill. The HIRLAM (High Resolution Limited Area Model) regional climate model shows best precipitation skills during winter time and worst results during summer when evaluated over Denmark (Larsen et al., 2012). On long-term simulations of one year over the European Alpine region, WRF and MM5 do not show similar skills of precipitation simulation in winter and summer seasons as demonstrated by Awan et al. (2011). These authors have shown that precipitation results have more spread during summer (due to local-scale phenomena imposed to the large-scale circulation) than during winter. The WRF-ARW model parameterisations also introduce more variability on the results over this region. Both Awan et al. (2011) and Jankov et al. (2007) point out, in their sensitivity studies to physical parameterisations, the importance of spending time and effort on the analyses of the influence of the planetary boundary layer, radiation and cumulus schemes on the precipitation results obtained over the domain of interest.

4 Conclusions

The purpose of this study was to conduct an evaluation of the WRF model for simulating wet-season precipitation over a complex orographic domain in a Mediterranean climate region, with the ultimate objective of using predicted rainfall

Table 6. Mean absolute error (MD) between the simulations and the observations for each location grouped by altitude class.

Altitude classes	WRF experiments	Stations									
		S01ALBC1	S03BOUC1	S05CANC1	S06GAFC1	S07PRAC1	S10SEJC1	S12SPSC1	S21ITEMC1	S22TROCI	
C1	RunRef	0.65	0.72	0.45	0.45	0.48	0.55	0.42	0.64	0.50	
	RunObsN	0.63	0.71	0.47	0.44	0.53	0.53	0.41	0.61	0.50	
	RunGridN	0.65	0.61	0.51	0.56	0.77	0.54	0.47	0.74	0.58	
C2	RunRef	1.33	0.92	0.50	0.42	0.49	0.35				
	RunObsN	1.28	0.91	0.46	0.44	0.48	0.33				
	RunGridN	1.09	0.95	0.38	0.48	0.50	0.31				
C3	RunRef	0.41	1.50	0.39	0.36	0.31	0.72	0.39	0.66		
	RunObsN	0.39	1.50	0.35	0.33	0.30	0.74	0.37	0.68		
	RunGridN	0.33	1.25	0.33	0.29	0.26	0.72	0.28	0.53		
C4	RunRef	0.78	0.42								
	RunObsN	0.79	0.41								
	RunGridN	0.72	0.40								
C5	RunRef	0.58	0.45								
	RunObsN	0.57	0.44								
	RunGridN	0.54	0.41								

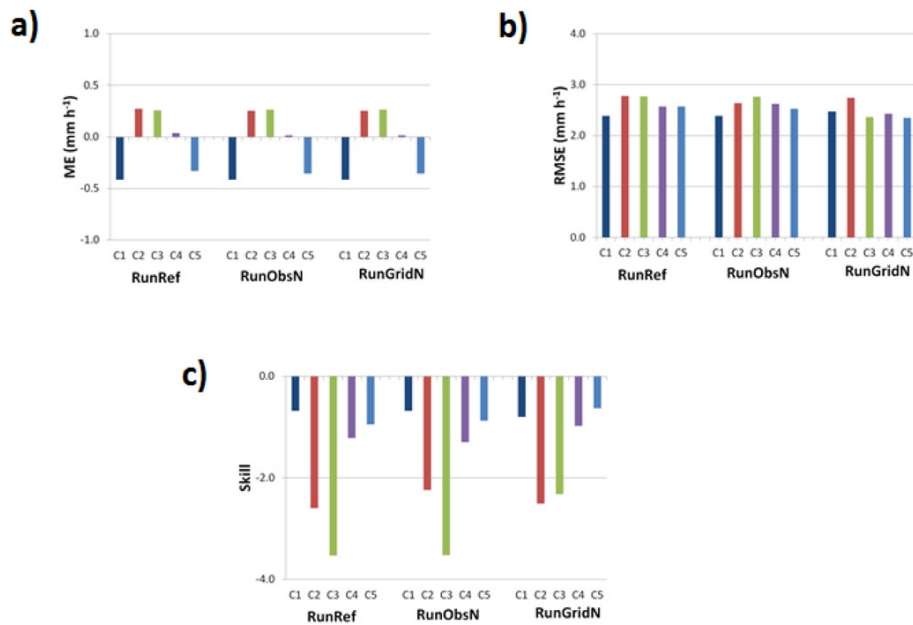
Table 7. Verification measures of hourly precipitation for the December 2009 evaluation period (745 h of simulation) for the three WRF model experiments: RunRef, RunObsN and RunGridN.

Threshold (mm h ⁻¹)	WRF experiments	B ^a	PC ^b (%)	POD ^c	F ^d	ETS ^e	ME ^f (mm h ⁻¹)	MSE (mm h ⁻¹)	RMSE ^g (mm h ⁻¹)	SS ^h
$t = 0.1$	RunRef	1.2	77	0.7	0.2	0.3	0	6.9	2.6	-1.7
	RunObsN	1.2	77	0.7	0.2	0.3	0	6.7	2.6	-1.6
	RunGridN	1.3	77	0.7	0.2	0.3	-0.2	6.2	2.5	-1.4
$t = 1$	RunRef	0.8	85	0.4	0.1	0.2	-0.8	10.6	3.3	-0.7
	RunObsN	0.8	85	0.4	0.1	0.2	-0.8	10.5	3.2	-0.7
	RunGridN	0.9	84	0.4	0.1	0.2	-1.1	9.7	3.1	-0.5
$t = 2$	RunRef	0.7	90	0.2	0.0	0.1	-1.7	14.8	3.8	-0.2
	RunObsN	0.7	90	0.2	0.0	0.1	-1.7	14.5	3.8	-0.2
	RunGridN	0.8	90	0.2	0.0	0.1	-2.1	14.0	3.7	-0.2
$t = 3$	RunRef	0.6	93	0.2	0.0	0.1	-2.6	21.0	4.6	0.0
	RunObsN	0.6	93	0.2	0.0	0.1	-2.6	20.8	4.6	0.0
	RunGridN	0.7	93	0.1	0.0	0.1	-3.0	20.8	4.6	0.0

^a B – frequency bias; ^b PC – percentage of correct events; ^c POD – probability of detection; ^d F – false alarm rate; ^e ETS – equitable threat score; ^f ME – mean error; ^g RMSE – root mean square error; ^h SS – skill.

amounts to fill gaps in precipitation measurements. The region of interest is located in north-central Portugal and includes the Águeda river catchment, which is well known for its elevated flood risks during the wet season in particular. Three different model configurations exploring three different nudging options were considered (RunRef, RunObsN and RunGridN). Hourly values of precipitation with 1-km

horizontal resolution were extracted from the simulations and the observed records. Model performance was evaluated by calculating statistical verification measures of model-simulated precipitation, namely the contingency table (B – frequency bias; PC – percentage of corrected events; POD – probability of detection; F – false alarm rate; ETS equitable



*ME – mean error; *RMSE – root mean square error; *Skill

Fig. 8. Continuous verification measurements of the hourly precipitation (above 0.1 mm h^{-1}) for the stations grouped by altitude class.

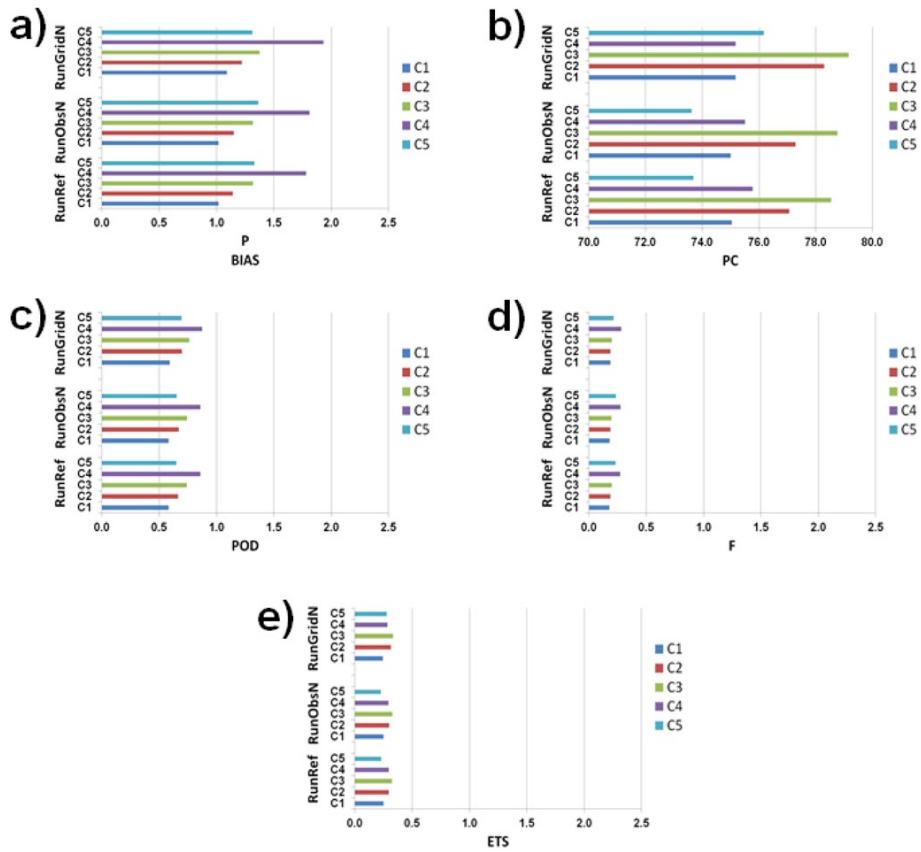
threat score), and the continuous measures ME – mean error; RMSE – root mean square error; and SS – skill score.

The WRF model did not reproduce well the precipitation intensities but simulated reasonably well the periods of precipitation occurrence. No phase or spatial errors were identified for time lags of zero hours, but the RunRef and RunObsN experiments revealed a phase or spatial error up to time lags of +3 h (for the C2, C3 and C5 altitude classes). The grid-nudging experiment (RunGridN) was able to diagnose, using a threshold of 0.1 mm h^{-1} , 30 % of the observed precipitation events ($\text{ETS}=0.3$), and simulated 70 % of the observed precipitation as a hit. At the same time, RunGridN incorrectly diagnosed precipitation occurrence in 20 % of the cases. The RunGridN experiment performed better than the local nudging (RunObsN) and the no-nudging (RunRef) experiment for the majority of indices. The three experiments, however, revealed similar overall model accuracies (RMSE) for the different altitude classes. RMSE's were highest for the lowlands as well as the highlands (altitude classes C1 – 0–200 m; C4 – 200–400; and C5 – above 800 m). Precipitation simulated in areas located in rough terrain and deep valleys (C2 to C4 altitude classes) tended to be less accurate.

The lack of skill (SS) shown by the WRF model could be related to the grid horizontal resolution. Even so, J. Liu et al. (2012) suggested that increasing horizontal resolution may not lead to better results, due to the higher spatial variability of precipitation associated with more grid points. One possibility for improvement would be to investigate

the downscaling ratios and/or testing other model configurations in terms of nesting (1-way instead of 2-way), nudging (spectral nudging in the outermost domain) and then comparing the 1-km resolution with the 5-km horizontal resolution. The WRF model performance could be improved using the ERA-Interim as boundary forcing instead of GFS, mainly due to the ERA-Interim's better horizontal resolution (0.25°) than GFS (0.50°) and because it is known to be of good quality for Europe. Cardoso et al. (2012) used the ERA-Interim as boundary data, in a downscaling simulation, for two nested grids of 27 and 9 km horizontal resolution, over the Iberian Peninsula, and the model results compared well with observations.

The simulated precipitation appeared to be of insufficient quality for event-based hydrological modelling, as this typically requires precipitation amounts at (sub-)hourly intervals. However, it seemed adequate for continuous hydrological modelling at a daily scale, as can be deduced by the low ME, the reasonable agreement between simulated and observed daily maxima, and the correct simulation of the temporal rainfall patterns during the study month of December 2009. Overall, the RunObsN and RunGridN experiments provided the best match with the observations, but the performance of the two experiments varied between meteorological stations, although the RunGridN experiment had a slight improvement over RunObsN. The good performance of WRF in simulating the spatial patterns in precipitation constituted an important advantage for hydrological modelling, especially



*B – frequency bias; *PC – percentage of correct events; *POD – probability of detection;
 *F – false alarm rate; *ETS – equitable threat score

Fig. 9. Categorical verification measurements for hourly precipitation (above 0.1 mm h⁻¹) for stations grouped by altitude.

in mountainous regions with high precipitation amounts and few ground stations as is the case of the present study area.

Appendix A

Statistical measures used for model verification

This section describes the mathematical formulation of the basic statistical measurements used throughout this study.

The average of the absolute deviations can be obtained by

$$MD = \frac{1}{n} \sum_{i=1}^n |y_i - o_i|, \quad (A1)$$

where o_i and y_i designate the observations and simulated precipitation series, respectively, for the i -th station location, and n is the number of records.

The equations for calculating the continuous measures are as follows:

$$\text{mean error: } ME = \frac{1}{n} \sum_{i=1}^n (y_i - o_i) = \bar{y} - \bar{o} \quad (A2)$$

$$\text{mean square error: } MSE = \frac{1}{n} \sum_{i=1}^n (y_i - o_i)^2 \quad (A3)$$

$$\text{root mean square error: } RMSE = \sqrt{MSE} \quad (A4)$$

$$\text{skill score: } SS = 1 - \frac{MSE}{MSE_{Clim}} \quad (A5)$$

where

$$MSE_{Clim} = \frac{1}{n} \sum_{i=1}^n (\bar{o} - o_i)^2 \quad (A6)$$

and \bar{o} stands for the climatological mean of hourly precipitation.

The categorical verification measures, derived from Table 3, are defined by

$$\text{frequency bias : } B = \frac{a + b}{a + c} \quad (\text{A7})$$

$$\text{percentage of corrected events : } PC = \frac{a + d}{N} \times 100\% \quad (\text{A8})$$

$$\text{probability of detection : } POD = \frac{a}{a + c} \quad (\text{A9})$$

$$\text{false alarm rate : } F = \frac{b}{b + d} \quad (\text{A10})$$

$$\text{equitable threat score : } ETS = \frac{a - a_r}{a + b + c - a_r}, \quad (\text{A11})$$

where a_r represents the expected hits by chance and is determined by

$$a_r = \frac{1}{N} \times (a + b)(a + c) \quad (\text{A12})$$

and N represents the total number of elements of the table.

The pattern correlation coefficient (PTC) is expressed as

$$PTC = \frac{\sum_{i=1}^N (y_i - \bar{y}_i)(o_i - \bar{o}_i)}{\left[\sum_{i=1}^N (y_i - \bar{y}_i)^2 \sum_{i=1}^N (o_i - \bar{o}_i)^2 \right]^{1/2}} \quad (\text{A13})$$

In this case the statistics were calculated for each grid point and the summations were performed for the entire grid; N stands for the total of grid points and the bar denotes spatial averaging over the grid.

Acknowledgements. This work has been supported by the Fundação para a Ciência e Tecnologia (FCT) via PhD grant no. SFRH/BD/65982/2009 (S. Pereira), SFRH/BD/31465/2006 (J. Ferreira) and postdoctoral researcher SFRH/BPD/39721/2007 (J. P. Nunes).

This study was supported by FEDER funds through the Programa Operacional Factores de Competitividade – COMPETE and by Portuguese national funds through FCT – Fundacao para a Ciencia e a Tecnologia, within the framework of the following project(s):

- HIDRIA Project Reference PTDC/CTE-GEX/71651/2006;
- RESORT Project Reference PTDC/CTE-ATM/111508/2009;
- CLIPE Project Reference PTDC/ AAC-CLI/111733/2009.

Edited by: E. Morin

References

AghaKouchaka, A., Bárdossy, A., and Habiba, E.: Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula, *Adv. Water Resour.*, 33, 624–634, doi:10.1016/j.advwatres.2010.02.010, 2010.

Akhtar, M., Ahmad, N., and Booij, M. J.: Use of regional climate model simulations as input for hydrological models for the Hindukush-Karakorum-Himalaya region, *Hydrol. Earth Syst. Sci.*, 13, 1075–1089, doi:10.5194/hess-13-1075-2009, 2009.

Aligo, E. A., Gallus, W. A., and Segal, M.: On the Impact of WRF Model Vertical Grid Resolution on Midwest Summer Rainfall Forecasts, *Weather Forecast.*, 24, 575–594, doi:10.1175/2008WAF2007101.1, 2009.

Argüeso, D., Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Esteban-Parra, M. J., Dudhia, J., and Castro-Díez, Y.: Evaluation of WRF Parameterizations for Climate Studies over Southern Spain Using a Multistep Regionalization, *J. Climate*, 24, 5633–5651, doi:10.1175/JCLI-D-11-00073.1, 2011.

Awan, N. K., Truhetz, H., and Gobiet, A.: Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble Based Analysis, *J. Climate*, 24, 3107–3123, doi:10.1175/2011JCLI3674.1, 2011.

Berrisford, P., Dee, D. P., Fielding, K., Fuentes, M., Kållberg, P., Kobayashi, S., and Uppala, S. M.: The ERA-Interim Archive, ERA Report Series No. 1, ECMWF, Reading, UK, 2009.

Brandão, C., Rodrigues, R., and Pinto, J. C.: Análise de fenómenos extremos, precipitações intensas em Portugal Continental, Lisboa, DSRH, available at: http://www.isa.utl.pt/der/Hidrologia/relatorio_prec_intensa.pdf (last access: 19 September 2013), 2001.

Bukovsky, M. S. and Karoly, D. J.: Precipitation Simulations Using WRF as a Nested Regional Climate Model, *J. Appl. Meteorol. Clim.*, 48, 2152–2159, doi:10.1175/2009JAMC2186.1, 2009.

Campos, I., Abrantes, N., Vidal, T., Bastos, A. C., Gonçalves, F., and Keizer, J. J.: Assessment of the toxicity of ash-loaded runoff from a recently burnt eucalypt plantation, *Eur. J. Forest Res.*, 131, 1889–1903, doi:10.1007/s10342-012-0640-7, 2012.

Cardoso, R. M., Soares, P. M. M., Miranda, P. M. A., and Belo-Pereira, M.: WRF high resolution simulation of Iberian mean and extreme precipitation climate, *Int. J. Climatol.*, 33, 2591–2608, doi:10.1002/joc.3616, 2012.

Castillo, V. M., Gómez-Plaza, A., and Martínez-Mena, A.: The role of antecedent soil water content in the runoff response of semi-arid catchments: a simulation approach, *J. Hydrol.*, 284, 114–130, doi:10.1016/S0022-1694(03)00264-6, 2003.

Chen, F., and Dudhia, J.: Coupling an advanced land-surface/hydrology model with the Penn State/ NCAR MM5 modeling system, Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, 129, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Chow, V. T., Maidment, D. R., and Mays, L. W.: *Applied Hydrology*, McGraw-Hill, New York, 1988.

Clark, A. J., Gallus, J. R., William, A., and Chen, T.-C.: Comparison of the Diurnal Precipitation Cycle in Convection-Resolving and Non-Convection-Resolving Mesoscale Models, *Mon. Weather Rev.*, 135, 3456–3473, doi:10.1175/MWR3467.1, 2007.

Deb, S. K., Kishtawal, C. M., Bongirwar, V. S., and Pal, P. K.: The simulation of heavy rainfall episode over Mumbai: impact of horizontal resolutions and cumulus parameterization schemes, *Nat. Hazards* 52, 117–142, doi:10.1007/s11069-009-9361-8, 2010.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer,

- A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, 46, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2, 1989.
- Fernández, J., Montávez, J. P., Sáenz, J., González-Rouco, J. F., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, *J. Geophys. Res.*, 112, D04101, doi:10.1029/2005JD006649, 2007.
- Ferreira P., Castanheira, J. M., Rocha, A., and Ferreira J.: Estudo de sensibilidade das previsões de superfície em Portugal pelo WRF face à variação das parametrizações físicas, XXX Jornadas Científicas da AME – AGUA Y CAMBIO CLIMÁTICO, IX Encuentro Hispano Luso de Meteorología y XII Congreso Latinoamericano e Ibérico de Meteorología, 5 a 7 de mayo de 2008, Zaragoza, Espanha, http://www.ame-web.org/images/stories/Congresos/30-Zaragoza-2008/3A_Ferreira2.pdf (last access: 17 September 2013), 2008.
- Ferreira, J., Carvalho, A., Carvalheiro, L., Rocha, A., and Castanheira, J.: Sensitivity of a simulated extreme precipitation event to spatial resolution, parametrizations and assimilation, EMS2010-647, 10th EMS Annual Meeting, 8th European Conference on applied Climatology, 13–17 September 2010, Zurich, Switzerland, 2010a.
- Ferreira, R. S. V., Cerqueira, M. A., Condesso de Melo, M. T., de Figueiredo, D. R., and Keizer, J. J.: Spatial patterns of water quality in the Cértima River basin, central Portugal, *J. Environ. Monitor.*, 12, 189–199, 2010b.
- Figueiredo, E., Valente, S., Coelho, C., and Pinho, L.: Coping with risk: analysis on the importance of integrating social perceptions on flood risk into management mechanisms – the case of the municipality of Águeda, Portugal, *J. Risk Res.*, 12, 581–602, 2009.
- Grell, G. A. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophys. Res. Lett.*, 29, 1693–1697, doi:10.1029/2002GL015311, 2002.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded dataset of surface temperature and precipitation, *J. Geophys. Res.-Atmos.*, 113, D20119, doi:10.1029/2008JD010201, 2008.
- He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., and McGregor, M.: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, *Meteorol. Appl.*, 16, 91–101, doi:10.1002/met.132, 2009.
- Heikkilä U., Sandvik, A., and Sorteberg, A.: Dynamical downscaling of ERA-40 in complex terrain using the WRF regional climate model, *Clim. Dynam.*, 37, 1551–1564, doi:10.1007/s00382-010-0928-6, 2011.
- Hershfield, D.: Rainfall input for Hydrological models. Symp on Geochem, Precipitation, Evaporation, Soil-moisture, Hydrom, Int. Ass. Sci. Hydrol., Pub. No. 78, http://iahs.info/redbooks/a078/iahs078_0177.pdf, Proc. Gen. Assembly of Bern, September–October 1967, Bern, 177–188, 1967.
- Hong, S.-Y. and Lim, J.-O. J.: The WRF Single-Moment 6-Class Microphysics Scheme (WSM6), *J. Korean Meteorol. Soc.*, 42, 129–151, 2006.
- Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, *Mon. Weather Rev.*, 134, 2318–2341, doi:10.1175/MWR3199.1, 2006.
- Jankov, I., Gallus, W. A., Segal, M., and Koch, S. E.: Influence of Initial Conditions on the WRF–ARW Model QPF Response to Physical Parameterization Changes, *Weather Forecast.*, 22, 501–519, doi:10.1175/WAF998.1, 2007.
- Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practitioner’s Guide in Atmospheric Science, John Wiley and Sons, Chichester, 2003.
- Kirkby, M. J., Bracken, L. J., and Shannon, J.: The influence of rainfall distribution and morphological factors on runoff delivery from dry land catchments in SE Spain, *Catena*, 62, 136–156, 2005.
- Koo, M.-S., and Hong, S.-Y.: Diurnal variations of simulated precipitation over East Asia in two regional climate models, *J. Geophys. Res.-Atmos.*, 1984–2012, 2010.
- Kotlarski, S., Block, A., Böhm, U., Jacob, D., Keuler, K., Knoche, R., Rechid, D., and Walter, A.: Regional climate model simulations as input for hydrological applications: evaluation of uncertainties, *Adv. Geosci.*, 5, 119–125, doi:10.5194/adgeo-5-119-2005, 2005.
- Larsen, M. A., Thejll, P., Christensen, J. H., Refsgaard, J. C., and Jensen, K. H.: On the role of domain size and resolution in the simulations with the HIRHAM region climate model, *Clim. Dynam.*, 40, 2903–2918, doi:10.1007/s00382-012-1513-y, 2012.
- Lenaerts, J. T. M., van Heerwaarden, C. C., and de Arellano, J. V.-G.: Shallow convection over land: a mesoscale modelling study based on idealized WRF experiments. *Tethys, J. Weather Clim. West. Mediterran.*, 6, 51–66, doi:10.3369/tethys.2009.6.04, 2009.
- Liu, J., Bray, M., and Han, D.: Sensitivity of the Weather Research and Forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation, *Hydrol. Process.*, 26, 3012–3031, doi:10.1002/hyp.8247, 2012.
- Liu, P., Tsimpidi, A. P., Hu, Y., Stone, B., Russell, A. G., and Nenes, A.: Differences between downscaling with spectral and grid nudging using WRF, *Atmos. Chem. Phys.*, 12, 3601–3610, doi:10.5194/acp-12-3601-2012, 2012.
- Lo, J., Yang, Z., and Pielke Sr., R. A.: Assessment of dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) Model, *J. Geophys. Res.*, 113, D01303, doi:10.1029/2007JD009216, 2008.
- Lou, X. F. and Breed, D.: Model evaluations for winter orographic clouds with observations, *Chin. Sci. Bull.*, 56, 76–83, doi:10.1007/s11434-010-4249-2, 2011.
- Luna, T., Rocha, A., Carvalho, A. C., Ferreira, J. A., and Sousa, J.: Modelling the extreme precipitation event over Madeira Island on 20 February 2010, *Nat. Hazards Earth Syst. Sci.*, 11, 2437–2452, doi:10.5194/nhess-11-2437-2011, 2011.
- Malvar, M. C., Prats, S. A., Nunes, J. P., and Keizer, J. J.: Post-fire overland flow generation and inter-rill erosion under simulated rainfall in two eucalypt stands in north-central Portugal, *Environ. Res.*, 111, 222–236, doi:10.1016/j.envres.2010.09.003, 2011.

- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave, *J. Geophys. Res.*, 102, 16663–16682, doi:10.1029/97JD00237, 1997.
- Murphy, A. H. and Winkler, R. L.: A general framework for forecast verification, *Mon. Weather Rev.*, 115, 1330–1338, doi:10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2, 1987.
- Pellarin, T., Delrieu, G., Saulnier, G.-M., Andrieu, H., Vignal, B., and Creutin, J.-D.: Hydrologic Visibility of Weather Radar Systems Operating in Mountainous Regions: Case Study for the Ardèche Catchment (France), *J. Hydrometeorol.*, 3, 539–555, doi:10.1175/1525-7541(2002)003<0539:HVOVRS>2.0.CO;2, 2002.
- Prats, S. A., MacDonald, L. H., Monteiro, M., Ferreira, A. J. D., Coelho, C. O. A., and Keizer, J. J.: Effectiveness of forest residue mulching in reducing post-fire runoff and erosion in a pine and a eucalypt plantation in north-central Portugal, *Geoderma*, 191, 115–125, doi:10.1016/j.geoderma.2012.02.009, 2012.
- Rial-Rivas, M. E., Nunes, J. P., Boulet, A. K., Ferreira, J. D. A., Coelho, C. O. A., and Keizer, J. J.: Addressing input data uncertainties in the hydrological simulation of a small forested catchment in north-central Portugal, *Bodenkultur*, 62, 105–110, 2011.
- Rossa, A., Nurmi, P., and Ebert, E.: Overview of methods for the verification of quantitative precipitation forecasts, in: *Precipitation: Advances in Measurement, Estimation and Prediction*, edited by: Michaelides, S. C., http://link.springer.com/content/pdf/10.1007/978-3-540-77655-0_16.pdf, Springer-Verlag, Berlin, Heidelberg, 417–450, 2008.
- Rummukainen, M.: State-of-the-art with regional climate models, *WIREs Clim. Change*, 1, 82–96, doi:10.1002/wcc.8, 2010.
- Santos, J. A., Corte-Real, J., and Leite, S. M.: Weather regimes and their connection to the winter rainfall in Portugal, *Inter. J. Climatol.*, 25, 33–50, doi:10.1002/joc.1101, 2005.
- Sebastianelli, S., Russo, F., Napolitano, F., and Baldini, L.: Comparison between radar and rain gauges data at different distances from radar and correlation existing between the rainfall values in the adjacent pixels, *Hydrol. Earth Syst. Sci. Discuss.*, 7, 5171–5212, doi:10.5194/hessd-7-5171-2010, 2010.
- Singh, V. and Frevert, D. (Eds.): *Mathematical Models of Small Watershed Hydrology and Applications*, Water Resources Publications, LLC, Chelsea, Michigan, 2002.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Yu, Wang, W., and Powers, J. G.: A description of the Advanced Research WRF version 3, NCAR/TN475+STR, NCAR, Boulder, Colorado, 125 pp., 2008.
- Skøien, J. O. and Blöschl, G.: Characteristic space scales and timescales in hydrology, *Water Resour. Res.*, 39, 1304–1323, 2003.
- Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., Medeiros, B.-J., Pereira, M., and Santo, E.-F.: WRF high resolution dynamical downscaling of ERA-Interim for Portugal, *Clim. Dynam.*, 39, 2497–2522, 2012.
- Trigo, R. M. and DaCamara, C. C.: Circulation Weather Types and Their Influence on the Precipitation Regime in Portugal, *Int. J. Climatol.*, 20, 1559–1581, 2000.
- Weisman, M. L., Davis, C., Wang, W., Manning, K. W., and Klemp, J. B.: Experiences with 0–36-h Explicit Convective Forecasts with the WRF-ARW Model, *Weather Forecast.*, 23, 407–437, doi:10.1175/2007WAF2007005.1, 2008.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences: An Introduction*, 2nd Edn., International Geophysics Series Vol. 59, Elsevier, Oxford, 627 pp., 2006.