Hydrol. Earth Syst. Sci. Discuss., 10, 1423–1463, 2013
www.hydrol-earth-syst-sci-discuss.net/10/1423/2013/
doi:10.5194/hessd-10-1423-2013

Hydrology and Earth System Sciences Discussions

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

S. C. Pereira[1], **A. C. Carvalho**[2], **J. Ferreira**[2], **J. P. Nunes**[3], **J. J. Keizer**[3], **and A. Rocha**[1]

[1]CESAM and Dept. Physics, University of Aveiro, Aveiro, Portugal
[2]CENSE, Faculty of Science and Technology, University NOVA of Lisboa, Costa da Caparica, Portugal
[3]CESAM and Dept. Environment and Planning, University of Aveiro, Aveiro, Portugal

Received: 20 December 2012 – Accepted: 17 January 2013 – Published: 28 January 2013

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This study evaluates the performance of the WRF-ARW numerical weather 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 for the month of December 2009, during the rainy season in Mainland Portugal. The heavy to extreme rainfall periods were caused by several low surface pressure systems associated with frontal surfaces. Three model runs, forced with the initial fields from a global domain model, were conducted. The model experiments were conducted to compare model performance using different approaches: (1) a reference experiment with no nudging (RunRef); (2) observational nudging for a specific location (RunObsN) is included; (3) nudging is used to adjust the analysis field (RunGridN). Model performance was evaluated against an observed hourly precipitation dataset of 27 rainfall stations, grouped by altitude, using several statistical parameters. The WRF model did not show skill in reproducing the precipitation intensities but simulated reasonably the periods of precipitation occurrence. The best performance was reached for the grid-nudging experiment (RunGridN). The overall model accuracy (RMSE) was similar for all altitude classes, for the three experiments: highest for lowlands and highlands. Precipitation simulated in areas located in rough terrain and deep valleys tend to be less accurate.

1 Introduction

Deterministic modelling of the complex interactions in nature has become a valuable complementary tool for scientists and policy makers. Rainfall-runoff modelling has become an increasingly used tool due to the implications of runoff for, among others, water availability for agriculture and direct human uses, water contamination, landslides, and flash floods.

Both short, high-intensity and prolonged, low-intensity rainfall events can play a key role in catchment-scale runoff generation and associated phenomena such as

HESSD
10, 1423–1463, 2013

flooding risk. Flood generation processes have been described by numerous authors (e.g. Chow et al., 1998). 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 1–100 km^2, while the scale of saturation-excess runoff is typically related with that of storm systems and weather fronts, ranging from hours to days and at regional scales exceeding 100 km^2 (Skøien and Blöschl, 2003). In the case of Mediterranean-type catchments the maximum rainfall intensity during 30 min (I30) has been indicated by several authors (e.g. Castillo et al., 2003; Kirkby et al., 2005) as critical for surface runoff generation. An analysis of the rainfall events leading to flooding must, therefore, 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). But even for a high density rain gauge network, the characteristic dimensions of the rainfall distribution may not be adequately captured (Hershfield, 1967). Advances in computational power are making possible the use of numerical weather prediction (NWP) models for simulating precipitation processes with a spatial and temporal resolution that is adequate for many hydrological applications. The current NWP spatial grid cell resolutions of around 1 × 1 km and temporal resolutions of a few seconds, are adequate to resolve rainfall for the characteristic spatio-temporal scales for runoff generation described above. In fact, studies with NWP models have been used in climatic studies, in order to evaluate the uncertainties obtained with hydrological models when driven by temperature and precipitation results of numerical weather prediction models (e.g. Kotlarski et al., 2005; Akhtar et al., 2009). In forecasting mode, studies have been made in order to obtain

HESSD
10, 1423–1463, 2013

probability density functions of floods a few days in advance, using results from meteorological ensemble prediction systems, dynamically downscaled through regional NWP to the catchment scale (He et al., 2009).

Due to their physical basis, NWP allow to explicitly test the current understanding of key meteorological processes and provide a more solid foundation for the explanation of meteorological measurements. For precipitation simulations, several aspects may play an important role on the obtained results, such as domain resolution (e.g. Heikkilä et al., 2011; Soares et al., 2012; Luna et al., 2011), and domain (Ferreira et al., 2010), 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).

Similar results were found by Heikkilä et al. (2011) with the same model operated in a climatic study over the Atlantic North and Norway (represented with a 10 km cell resolution). Luna et al. (2011) have also shown the importance of model horizontal resolution for the simulation of local precipitation over Madeira Island. However, if precipitation is integrated over time and space, resolution turns out to be less effective. For the same period, the same authors have found that the precipitation amounts were not dependent on the cumulus schemes parameterisations for grid cells resolutions of 1-km against the explicit precipitation calculation. Experiments with WRF idealised runs with 3-km resolution showed that, for shallow convection, Grell Cumulus parameterisation scheme or explicit physical options the model tend to perform worse (Lenaerts et al., 2009).

The motivation for the present work emerged from the need of having precipitation fields which would be used in runoff applications in a later stage. The study region is a mountainous region near the Central-North Portuguese coast line. For this particular region the precipitation pattern is not well known, partly due to the scarcity of the rain gauges, the lack of radar-based information as well as the possible precipitation

HESSD
10, 1423–1463, 2013

gradients induced by the orography. Hence, the hydrological-erosion response of six experimental catchments are being monitored, in the aim of ongoing scientific projects (Fernandes et al., 2010; Rial-Rivas et al., 2011; Campos et al., 2012; Machado et al., 2012), using the Pousadas meteorological station as reference station for high-quality local rainfall records. Nonetheless, data gaps can hardly be avoided altogether, as was the case for the time period selected for this study. In the present case, however, the usefulness of the two existing radar stations is somewhat doubtful. Firstly, the study area lies a considerable distance from the nearest radar station (c. 250 km), whereas the agreement between radar-based estimates and point measurements was found to decrease with increasing distance (Sebastianelli et al., 2010). Secondly, the study area is mountainous, and mountains may introduce 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 a good spatio-temporal resolution.

This study evaluates the performance of a NWP model over a complex orographic region in an attempt to evaluate the model's suitability for providing those estimates of precipitation fields and time series in order to fill the measured gaps and for further hydrological applications. In particular, it addresses how different approaches for model application impact the quality of model results. For this, three experiments were performed to test if the data assimilation of observations over a defined location, or the grid nudging technique would yield better results than a simulation forced only with initial and boundary conditions.

2 Materials and methods

2.1 Study area and case study

The study area spanned a mountainous region in north-central Portugal (Fig. 1). The climate is wet Mediterranean, with a mean annual rainfall ranging from 800 mm at the

HESSD
10, 1423–1463, 2013

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 of 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 c. 3 yr, but in four stations located in the lowlands of the Modego river valley (S and E part 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 S and SE (also inside the Mondego valley had the rainfall maxima in 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 NW of the study area, in the sea-facing side of the coastal mountain range (Fig. 1), had the maxima in 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 IDF long-term IDF curves in the dataset, Santa Comba Dão (code S18SCDC2) (Table 2), located in the Mondego valley. It indicated that the December 2009 values corresponded to return periods of less than 2 yr for short-term rainfall durations (< 3 h) and between 2 to 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

HESSD
10, 1423–1463, 2013

attributed to a longer-duration event lasting between 12 and 24 h, rather than a short-term, high intensity precipitation 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 Weather Research and Forecasting (WRF) Model with Advanced Research WRF (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 in diagnostic mode (Liu, 2012; Lou and Breed, 2011; Bukovsky and Karoly, 2009).

It has also been used in Portugal, in a sensitivity test to parameterizations for two different model operational configurations (Ferreira et al., 2010), in climate simulations over Portugal (Soares et al., 2010) and over the Andaluzia 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. These authors have performed a sensitivity test to the model parameterizations during a five year period. Regarding precipitation, the authors pointed out that the orography representation by the model has a larger impact on the winter modelled precipitation than the one computed on summer.

Model horizontal resolution also has an importance for the simulation of local precipitation. Soares et al. (2012) highlight the importance of the model fine resolution in order to obtain precipitation extreme values on a 9 km resolution grid, namely in the wettest region of Portugal (NW).

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

HESSD
10, 1423–1463, 2013

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° × 0.5° and the vertical domain extends from a surface pressure of 1000 to 0.27 hPa, discretized 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-nesting way mode, with resolutions of 25 km (D01), 5 km (D02) and 1 km (D01), for the parent, middle and inner domains, respectively. The finer grid domain is centered over Pousadas (40.63° N, 8.31° W) (see Fig. 1).

The Lambert conformal conical projection, due to its applicability to mid-latitudes, is used with the standard parallel at 40.63° N. The three nested domains identified have the Atlantic Ocean as 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 discretization consists of 27 eta levels. The physical parameterization schemes used in this work resulted from a previous study made by Ferreira et al. (2008), in which several parametrizations sets were tested against observations of temperature, water vapor, mixing ratio and wind at several stations over mainland Portugal, using the WRF-ARW with the same configuration as the one used in the present work, only for the D01 and D02 domains. The physical parametrization configuration is the following: WRF Single Moment 6 class scheme microphysics (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 (Noh et al., 2003); Noah Land Surface Model (Chen and Duhia, 2001); Grell Devenyi ensemble convective parameterization scheme (Grell and Devenyi, 2002). This parameterization set was used in all three domains.

HESSD
10, 1423–1463, 2013

2.3 Experimental design

Three numerical experiments, corresponding to integrations with one month of duration plus 24 h of spinup which were discarded, were made for December of 2009, starting at 00:00 UTC 30 November 2009 until 00:00 UTC 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, 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.

In their studies Soares et al. (2012), Argüeso et al. (2011) and Fernández et al. (2007) applied the model using two-way nesting, having all performed grid nudging in the coarser domain. In this study, nudging was carried out to individual observations over the location of Pousadas (RunObsN), in order to evaluate the impact of local circulation in the computation of model precipitation. A third experiment, consisting in applying the grid nudging technique (RunGridN), was made, with the purpose to investigate the impact of 3-D analysis nudging to constrain the large-scale 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.

2.4 Rainfall measurements

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, Table 8). The SNIRH dataset consists of a series of rain gauge stations, recording with

HESSD
10, 1423–1463, 2013

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.

2.5 Assessment of model performance

The observations were compared with the model simulations for identical locations and times. Model data was 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, spatially interpolation of the modelled series of precipitation to the station location, or selection of the grid point nearest 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. So, the average of the absolute deviations can be obtained by:

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

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

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 1 December 2009 to 23:00 UTC of 31 December 2009.

HESSD
10, 1423–1463, 2013

Some basic statistics were calculated: mean, median, mode, standard deviation and correlation for lags between −24 to +24 h. The correlation was tested against the null hypothesis at significant level of 1 % for 744 degrees of freedom. The critical value was 0.098 indicating that correlation values above this threshold are statistically significant and the null hypothesis can be rejected.

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.

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

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).

The equations for calculating the continuous measures are:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i) = \overline{y} - \overline{o} \quad (2)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2 \quad (3)$$

$$RMSE = \sqrt{MSE} \quad (4)$$

$$SS = 1 - \frac{MSE}{MSE_{Clim}} \quad (5)$$

where,

HESSD
10, 1423–1463, 2013

$$\mathrm{MSE}_{\mathrm{Clim}} = \frac{1}{n}\sum_{i=1}^{n}\left(\overline{o} - o_i\right)^2 \quad (6)$$

and $\overline{o}$ stands for the climatological mean of hourly precipitation. When using the MSE measure as base for calculating the skill, the last is called reduction of variance because Eq. (5) is represents the ratio between the squared deviations and the observed variance. 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 does not necessary imply that the model has no skill at all (Jolliffe and Stephenson, 2003).

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 meets the event criteria.

The categorical verification measures, derived from the above table, are defined by:

$$B = \frac{a+b}{a+c} \quad (7)$$

$$\mathrm{PC} = \frac{a+d}{N} \times 100\% \quad (8)$$

$$\mathrm{POD} = \frac{a}{a+c} \quad (9)$$

$$F = \frac{b}{b+d} \quad (10)$$

HESSD
10, 1423–1463, 2013

$$\mathrm{ETS} = \frac{a - a_r}{a + b + c - a_r} \tag{11}$$

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

$$a_r = \frac{1}{N} \times (a + b)(a + c) \tag{12}$$

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

Better model performances are expressed by high values of POD (varies between 0 and 1) and ETS (varies between −1/3 and 1) and low values of F (varies between 0 and 1), ME and MSE, 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.

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 per hour). Meaning, that an event with precipitation above t is considered a yes event, otherwise it is a non-event. Worth stressing is perhaps that the model results were neither rescaled nor transformed.

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 meters interval. Thus, the first classes (C1) comprise the first 200 meters, the second class (C2) ranges from 200 to 400 m and so on until you reach the last class (C5) corresponding to altitudes above 800 m.

HESSD
10, 1423–1463, 2013

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 (RunRef), with local nudging (RunObsN) and with grid-nudging (RunGridN).

The validation consisted of the direct comparison between the observed series of precipitations from a network of rain gauges with the series extracted from the model. Direct validation between in situ observations and series extracted from the model can be a source of uncertainty due to the non-matching grids. To minimize 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 deviation between both interpolated series and the observation series were calculated.

The two difference series showed no differences. 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 to those presented by Rossa et al. (2008). These authors showed that verification using the nearest grid point gives very similar overall results.

3.1 Observed and modelled precipitation characteristics

The mean absolute error (MD) between the simulated precipitation series and observations is presented in Table 4. The results among experiments are identical. The MD values range between 0.31 to 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 = 1.49 mm h^{-1}) present an error of the same order as the respective observed mean consistent with the values found for the average hourly precipitation. Most stations showed a

HESSD
10, 1423–1463, 2013

good agreement with the observations but the stations S02BCBC2, S25CASC3 and S27MOSC2 clearly depart from the observations.

The correlation between observed and simulated of precipitation was calculated (Fig. 2, for convenience only the RunRef case is shown). The spatial pattern of correlation is quite variable, with stations showing correlations of 0.5 with the neighbouring mesh points while others showed a weak association, e.g. the S01ALBC1 station, but all are significant. Tables 5–7 present the lag correlations found for precipitation. The correlation values diminish with increasing lag, with maximum values attained at lag 0 and in some cases at lag +3, pointing for the exclusion of phase errors. The association among series was strongest for the RunRef experiment and weakest for the RunGridN. The altitude class that showed strongest correspondence were C5 (> 800 m) and C1 (0–200 m), the former with slightly worst results.

Table 8 shows the mean, median, mode and the standard deviation for the 27 rain gauges stations over the period of evaluation. The correspondent hourly mean values and the standard deviation from the model simulations are also shown. The average hourly value observed range from 1.80 to 0.96 mm h^{-1} and the observed variability varies from 0.74 to 2.56 mm h^{-1}.

The observed and modelled hourly rainfall amounts during December 2009 are shown in Fig. 3. 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 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) was not. For the RunGridN experiment the majority of the series simulated a week wet event ranging from 0 to 5 mm h^{-1}, for the 27th day, that none of the others reproduce. The total accumulated precipitation for the entire December at each station as well the month totals were calculated (Fig. 4). The experiments overestimated the precipitation. This result was not detected in the ME. Although, three locations show monthly totals much higher than the ones observed corresponding to

HESSD
10, 1423–1463, 2013

those which already showed a high error (Table 4). 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. 5. The frequency distributions are strongly asymmetric. 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 quite a lot extreme values (not shown), corresponding to the points lying out in the three times the IQR area. The pronounced intra-variability among stations reinforced the atypical nature of the 2009 December month, as mentioned earlier. For observations only, the variability results are supported by the standard deviation values presented in Table 8, with the majority of the individual standard deviations between the 1.0 to the 2 mm h^{-1} intervals.

3.2 Model assessment

The categorical verification measures (B, PC, POD, F and ETS) together with the continuous measures (ME and MSE) 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 9. For convenience only measurements for the 0.1, 1, 2 and 3 mm per hour thresholds are shown. The RunRef and the RunObsN experiment exhibit identical results whatever the verification measures used.

For the categorical verification measures, the RunGridN experiment slight outperforms the others. This characteristic is more pronounced for the 0.1 mm h^{-1} threshold and tends to deteriorate with increasing threshold value. The 0.1 mm/h 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 measurement is weighted by the most frequent category and can be artificially increased by issuing more corrected negatives. Despite, the F values decreased with threshold, POD values are low. For thresholds of 2 and 3 mm h^{-1}, 1/5 of

HESSD
10, 1423–1463, 2013

the observed precipitation event were correctly simulated (POD = 0.2) with an ETS value of 0.1.

Analysing the continuous scores the lowest errors were found when considering series of pairs above 0.1 mm h^{-1} and for the RunGridN experiment. The ME values almost meet the perfect score. The high values obtained for MSE and RMSE suggests that the model precipitation considerably departed from the observations, as indicated by the MD values (see Table 4). 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 can be related not to the model itself but to the horizontal resolution. Liu et al. (2011) 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 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 6 shows the continuous measure of verification and in Fig. 7 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

HESSD
10, 1423–1463, 2013

experiments the mean error is small. For all stations and altitude classes MSE and RMSE are high leading to a negative skill.

For categorical measures (Fig. 7), estimated quantities of B above the 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 the RunGridN more accurate than the others. The individual measurements for POD and F are similar among stations and experiments. On average, 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 slighted exceeded the individual scores specially when considering the RunGridN experiment. For this case the C4 class scored better than the others with a 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 set ups 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 initialization

HESSD
10, 1423–1463, 2013

datasets and physical parameterisations, whereas the rain rate was more sensitive to the cumulus parameterisations applied. Limited area models also show seasonal dependency of model skill. The HIRLAM 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 on 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 on 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 in simulating precipitation over a complex orographic domain, which could be used subsequently to fill measured gaps and in hydrological applications. The region of interest is located in north-central Portugal and includes the Águeda river catchment, which is an important watershed in the region. Three different model configurations corresponding to three different nudging options were considered. Hourly values of precipitation with 1-km horizontal resolution were extracted from simulations and observed records. The evaluation was performed by calculating statistical verification measures of model simulated precipitation namely, the contingency table (B, PC, POD, F, ETS), the continuous measures ME, MSE and a skill score.

The WRF model did not show skill in reproducing the precipitation intensities but simulated reasonably the periods of precipitation occurrence. No phase or spatial errors

HESSD
10, 1423–1463, 2013

were identified. The best performance was reached for the grid-nudging experiment (RunGridN) with the model being able to diagnose 30 % of the observed precipitation events (ETS 0.3) for threshold of 0.1 mm h^{-1}, and 70 % of the observed precipitation were simulated as a hit. Also, for this simulation and threshold value, 20 % of the cases were incorrectly diagnosed as precipitation occurrence. For the majority of the indices the RunGridN experiment scored better than the local nudging (RunObsN) and the no nudging (RunRef) runs. The overall model accuracy (RMSE) was similar for all altitude classes, for the three experiments: highest for lowlands and highlands (C1, C4 and C5 altitude classes). Precipitation simulated in areas located in rough terrain and deep valleys (C2 to C4 altitude classes) tend to be less accurate.

The lack of skill (SS) showed by the model can be related with the grid horizontal resolution. Liu et al. (2011) 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 suggestion of improvement would be to investigate the downscaling ratios and/or to test other coarser grid resolutions because the continuous verification measures may be negatively affected by excessive information of the small scales.

Simulated precipitation does not appear to have sufficient quality for event-based hydrological modelling, where the correct precipitation amount and timing are usually required. They seem, however, to be able to provide data for daily hydrological models which require additional sub-daily information such as maximum rainfall intensity, as can be deduced by the low ME, the relatively reasonable agreement between simulated and observed daily maxima, and the correct simulation of rainfall evolution patterns within the month of December (Fig. 3). In the latter case, however, it should be noted that observations are better matched with RunObsN or RUNGridN, depending on the meteorological station. Furthermore, the good performance of WRF in simulating the spatial patterns of precipitation provides an additional advantage for hydrological modelling, especially in mountainous regions with high precipitation such as the selected study area. In these areas, modelling precipitation can be used to overcome the

HESSD
10, 1423–1463, 2013

scarcity of rain-gauges; especially their distribution is inadequate to capture the spatial distribution of rainfall, either by design or by occasional data collection failures.

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), and also framed in the research projects: HIDRIA Project Reference PTDC/CTE-GEX/71651/2006; RESORT Project Reference PTDC/CTE-ATM/111508/2009.

References

AghaKouchaka, A., Bárdossyb, 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.

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. M. A. N., Abrantes, N. J. C., 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.

HESSD

10, 1423–1463, 2013

Castillo, V. M., Gómez-Plaza, A., and Martínez-Mena, A.: The role of antecedent soil water content in the runoff response of semiarid 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, 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.

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

Fernandes, I. A. C., Nunes, J. P., Ferreira, R. S. V., Pereira, L. M. G., and Keizer, J. J.: Escorrência e erosão a múltiplas escalas espaciais numa área florestal mista recentemente ardida no centro de Portugal, Territorium, 17, 135–144, 2010.

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, 5–7 May 2008, http://www.ame-web.org/images/stories/Congresos/30-Zaragoza-2008/3A_Ferreira_2.pdf, last access: 25 January 2013, Zaragoza, Spain, 2008.

Ferreira, J., Carvalho, A., Carvalheiro, L., Rocha, A., and Castanheira, J.: Sensitivity of a simulated extreme precipitation event to spatial resolution, parametrisations and assimilation, 10th EMS Annual Meeting, 8th European Conference on applied Climatology, 13–17 September 2010, EMS2010-647, Zurich, Switzerland, 2010.

HESSD

10, 1423–1463, 2013

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.

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, Proc. Gen. Assembly of Bern (September–October 1967), Int. Ass. Sci. Hydrol, Pub No. 78, 177–188, http://iahs.info/redbooks/a078/iahs_078_0177.pdf, last access: 25 January 2013, Bern, Switzerland, 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.

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.

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. D., 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., doi:10.1007/s00382-012-1513-y, in press, 2012.

HESSD
10, 1423–1463, 2013

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. Western 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, 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, 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.

Machado, A. I., Nunes, M. I., Cerqueira, M. A., Pinto, R., Martins, M., Patoilo, D., and Keizer, J. J.: Perdas de nutrientes por escorrência superficial em três encostas florestais recentemente ardidas, Recursos Hídricos, 33, 47–59, doi:10.5894/rh33n1-4, 2012.

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, 1987.

Noh, Y., Cheon, W. G., Hong, S.-Y., and Raasch, S.: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data, Bound.-Lay. Meteorol., 107, 401–427, 2003.

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, 2002.

Rial-Rivas, M. E., Nunes, J. P., Boulet, A. K., Ferreira, J. D. A., Coelho, C. O. A., and Keizer, J. J.: Adressing input data uncertainties in the nydrological simulation of a small forested catchment in north-central Portugal, Bodenkultur, 62, 105–110, 2011.

HESSD

10, 1423–1463, 2013

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., Springer-Verlag Berlin, Heidelberg, 417–450, 2008.

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, doi:10.1029/2002WR001736, 2003.

Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., de Medeiros, J., Belo-Pereira, M., and Espirito-Santo, F.: WRF high resolution dynamical downscaling of ERA-Interim for Portugal, Clim. Dynam., 39, 2497–2522, doi:10.1007/s00382-012-1315-2, 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, 2008.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: An Introduction, 2nd Edn., International Geophysics Series Vol. 59, Elsevier, Oxford, 627 pp., 2006.

HESSD

10, 1423–1463, 2013

Table 1. Month precipitation totals, daily maxima and correspondent percentile observed at some stations used in this study, for the December 2009 time period.

		Monthly rainfall December 2009	Long-term observed data					Maximum daily rainfall December 2009	Long-term observed data				
Station	Station ID	PP (mm)	Median PP (mm)	December 2009 vs. Median	Percentile (of all obs. data)	Return period (yr)	Number of years	PP (mm)	Median	December 2009 vs. Median	Percentile	Return period (yrs)	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
Tentúgal	S21TEMC1	285.0	126.7	125 %	89 %	8.1	56	69.1	33.4	107 %	96 %	18.3	54
Trouxemil	S22TROC1	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

HESSD
10, 1423–1463, 2013

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

HESSD
10, 1423–1463, 2013

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)

HESSD
10, 1423–1463, 2013

Table 4. Mean absolute error (MD) between the simulations and the observations for each location divided by altitude classes.

Altitude classes	WRF experiments	Stations								
		S01ALBC1	S03BOUC1	S05CANC1	S06GAFC1	S07PRAC1	S10SEJC1	S12SPSC1	S21TEMC1	S22TROC1
C1	RunRef	0.65	0.72	0.45	0.45	0.48	0.55	0.42	0.64	0.51
	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
		S02BCBC2	S27MOSC2	S09RIBC2	S13FRAC2	S18SCDC2	S23TABC2			
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			
		S04CALC3	S25CASC3	S15MANC3	S16OLIC3	S17PARC3	S11SMAC3	S19SEIC3	S20SATC3	
C3	RunRef	0.41	1.50	0.39	0.36	0.31	0.72	0.39	0.66	
	RunObsN	0.39	1.48	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	
		S08QUEC4	S26LEOC4							
C4	RunRef	0.78	0.42							
	RunObsN	0.79	0.41							
	RunGridN	0.72	0.40							
		S14LAGC5	S24ROSC5							
C5	RunRef	0.58	0.45							
	RunObsN	0.57	0.44							
	RunGridN	0.54	0.41							

HESSD
10, 1423–1463, 2013

Table 5. Lag correlation between observations and the model series for the RunRef experiment, divided by classes of altitude.

	RunRef					
lag	C1	C2	C3	C4	C5	Mean
−12	0.15	0.17	0.16	0.19	0.04	0.14
−9	0.15	0.16	0.16	0.18	0.00	0.13
−6	0.19	0.21	0.21	0.25	0.09	0.19
−3	0.25	0.26	0.27	0.25	0.17	0.24
0	0.30	0.31	0.33	0.33	0.36	0.33
3	0.24	0.34	0.36	0.23	0.38	0.31
6	0.15	0.25	0.31	0.22	0.22	0.23
9	0.06	0.15	0.28	0.15	0.10	0.15
12	0.04	0.14	0.20	0.05	0.06	0.10

HESSD
10, 1423–1463, 2013

Table 6. Same as Table 4 but for the RunObsN experiment.

	RunObsN					
lag	C1	C2	C3	C4	C5	Mean
−12	0.17	0.19	0.17	0.14	0.05	0.14
−9	0.16	0.16	0.17	0.16	0.01	0.13
−6	0.19	0.21	0.21	0.22	0.10	0.19
−3	0.25	0.26	0.28	0.25	0.17	0.24
0	0.29	0.33	0.35	0.32	0.38	0.33
3	0.24	0.33	0.37	0.21	0.40	0.31
6	0.18	0.25	0.30	0.20	0.22	0.23
9	0.07	0.18	0.26	0.15	0.08	0.15
12	0.06	0.14	0.20	0.04	0.02	0.09

HESSD
10, 1423–1463, 2013

Table 7. Sames as Table 4 but for the RunGridN experiemnts.

	RunObsN					
lag	C1	C2	C3	C4	C5	Mean
−12	0.11	0.16	0.15	0.20	0.02	0.13
−9	0.13	0.15	0.16	0.20	0.02	0.13
−6	0.14	0.21	0.20	0.23	0.10	0.17
−3	0.19	0.27	0.28	0.28	0.24	0.25
0	0.22	0.26	0.31	0.30	0.36	0.29
3	0.23	0.28	0.31	0.27	0.29	0.28
6	0.22	0.21	0.27	0.21	0.19	0.22
9	0.15	0.16	0.25	0.14	0.08	0.16
12	0.13	0.16	0.23	0.10	0.05	0.13

HESSD
10, 1423–1463, 2013

Table 8. List of rain-gauge locations used for assessing model performance ranked by class altitude (from lowest to highest). Sample statistic for the December 2009 hourly precipitation above 0.1 mm h^{-1}.

Station name	Station ID	Lat (° N)	Lon (° W)	Alt (m)	Mean (mm h^{-1})	Median (mm h^{-1})	Mode (mm h^{-1})	Std (mm h^{-1})
Albergaria-a-Velha	S01ALBC1	40.70	−8.48	131	1.5	0.9	0.2	1.7
Bouça	S03BOUC1	40.69	−8.37	152	1.6	1.0	0.2	1.7
Cantanhede	S05CANC1	40.36	−8.59	58	1.6	0.9	0.2	1.9
Gafanha da Nazaré	S06GAFC1	40.62	−8.71	17	1.4	0.8	0.2	1.6
Praia de Mira	S07PRAC1	40.46	−8.79	17	1.8	0.9	0.2	2.6
Sejães	S10SEJC1	40.74	−8.21	157	1.7	1.1	0.2	1.9
S. Pedro do Sul	S12SPSC1	40.75	−8.07	182	1.3	0.7	0.2	1.6
Tentúgal	S21TEMC1	40.24	−8.59	72	1.8	0.8	0.2	2.2
Trouxemil	S22TROC1	40.28	−8.45	79	1.5	0.9	0.3	1.6
Barragem Castelo Burgães	S02BCBC2	40.85	−8.38	306	1.4	1.0	0.2	1.4
Mosteiro Cabril	S27MOSC2	40.95	−8.10	389	1.5	0.9	0.3	1.7
Ribeiradio	S09RIBC2	40.74	−8.30	228	1.3	0.9	0.6	1.2
Fragosela	S13FRAC2	40.63	−7.85	376	1.4	0.8	0.2	1.5
Santa Comba Dão	S18SCDC2	40.43	−8.12	289	1.5	0.9	0.3	1.7
Tábua	S23TABC2	40.36	−8.04	220	1.0	0.6	0.2	1.1
Calde	S04CALC3	40.78	−7.92	505	1.0	0.7	0.2	1.0
Castro d'Aire	S25CASC3	40.92	−7.94	584	1.6	1.2	0.2	1.5
Mangualde	S15MANC3	40.60	−7.81	512	1.1	0.7	0.2	1.2
Oliveira do Hospital	S16OLIC3	40.36	−7.87	468	1.0	0.7	0.2	0.8
Paranhos da Beira	S17PARC3	40.48	−7.79	402	1.0	0.6	0.2	0.9
S. Martinho das Moitas	S11SMAC3	40.88	−8.03	408	1.8	1.3	0.2	1.6
Seia	S19SEIC3	40.42	−7.71	447	1.0	0.7	0.2	0.7
Sátão	S20SATC3	40.74	−7.74	570	1.5	1.0	0.2	1.5
Queriga	S08QUEC4	40.80	−7.75	685	1.1	0.5	0.2	1.9
Leomil	S26LEOC4	40.98	−7.66	704	1.5	1.0	0.2	1.4
Lagoa Comprida	S14LAGC5	40.38	−7.64	1560	1.6	0.9	0.2	2.3
Vale Rossim	S24ROSC5	40.40	−7.59	1427	1.4	1.0	0.2	1.3

HESSD
10, 1423–1463, 2013

Table 9. Verification measurements of hourly precipitation for the evaluation period of from 1 to 31 December 2009 (745 h of simulation) for the three WRF model experiments: RunRef, RunObsN and RunGridN.

Threshold ($mm\,h^{-1}$)	WRF experiments	*B*	PC (%)	POD	*F*	ETS	ME ($mm\,h^{-1}$)	MSE $(mm\,h^{-1})^2$	RMSE ($mm\,h^{-1}$)	SS
$t = 0.1$	RunRef	1.2	77	0.7	0.2	0.3	0.0	6.9	2.6	−1.7
	RunObsN	1.2	77	0.7	0.2	0.3	0.0	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

HESSD
10, 1423–1463, 2013

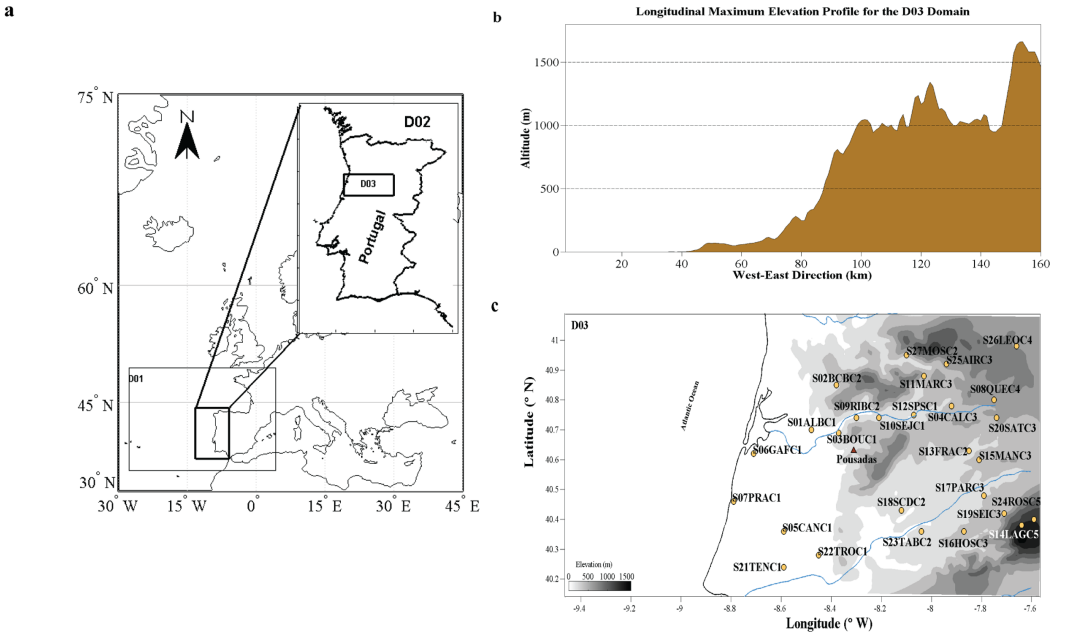


Fig. 1. Study area and stations location over the study area. **(a)** Nested domains for the WRF model experiments showing the outermost domain (D01) with 25-km of resolution, the middle domain (D02) with 5 km of resolution and the innermost domain (D03) with 1-km of resolution. The D03 frame mark the 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 (yellow triangle). Representing an area of about 16 800 km^2. The shaded areas show the model topographic heights in meters.

HESSD
10, 1423–1463, 2013

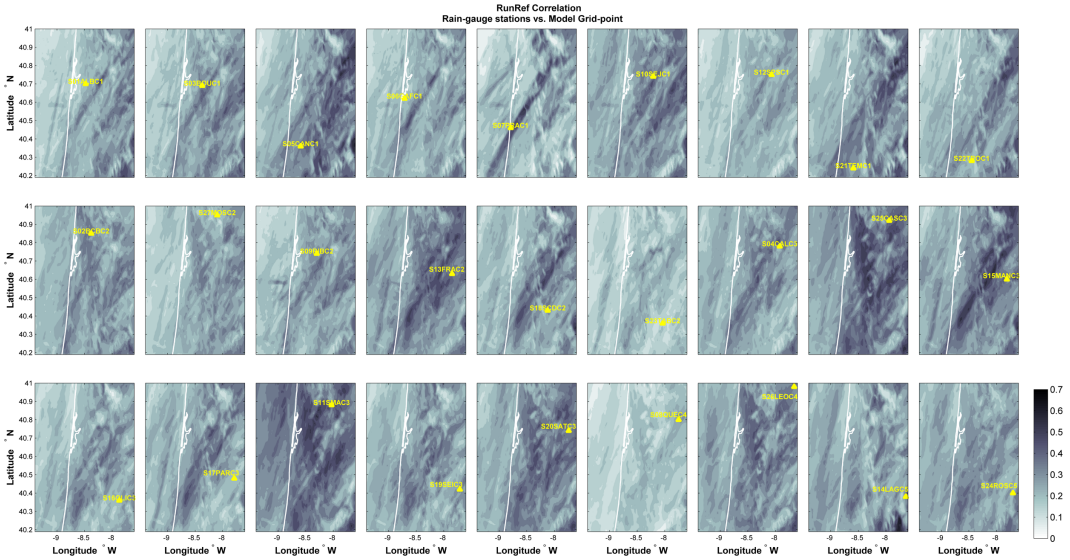


Fig. 2. Point correlations between a single station and all model grid-points for the 1-km domain during the period of evaluation (December 2009). Each map represents correlations for a particular station for the RunRef experiment. Lighter greys represent areas with low correlations and dark grey and/or black areas with high correlation values. Correlation values greater than 0.018 are significant at 1 % significance level.

HESSD
10, 1423–1463, 2013

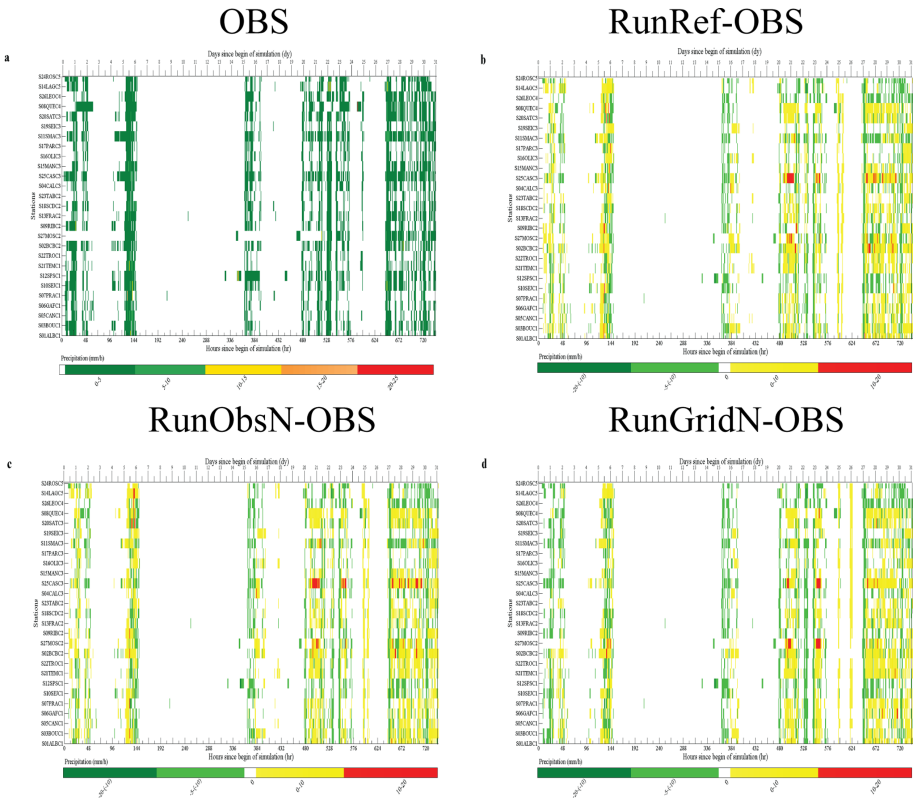


Fig. 3. Time series of hourly precipitation for the three experiments, for all the rain gauge stations scatter within the study area. The bottom x-axis represents hours since the simulation begin and the top x-axis the correspondent days. The vertical axis represents the stations and the different colours are related to the precipitation amount (mm h^{-1}). **(a)** Observed precipitation series; **(b)** precipitation series for the control run (RunRef); **(c)** precipitation series for the second experiment (RunObsN) and **(d)** precipitation series for the third experiment (RunGridN).

HESSD
10, 1423–1463, 2013

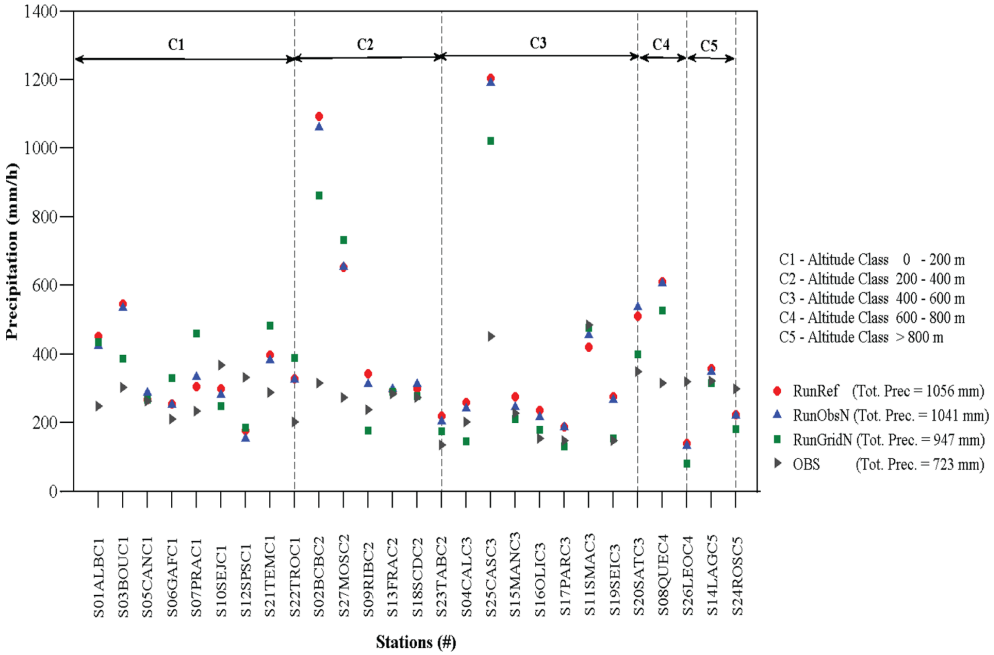


Fig. 4. Monthly accumulated precipitation for the simulation period (mm) over the study domain divided by altitude classes. The observed amounts are depicted as a grey triangle and for each of the model simulations as: RunRef (light blue circle), RunObsN (red triangle) and the RunGridN (dark blue square). The Caption box also shows the total rainfall observed and simulated by the model in each case.

HESSD
10, 1423–1463, 2013

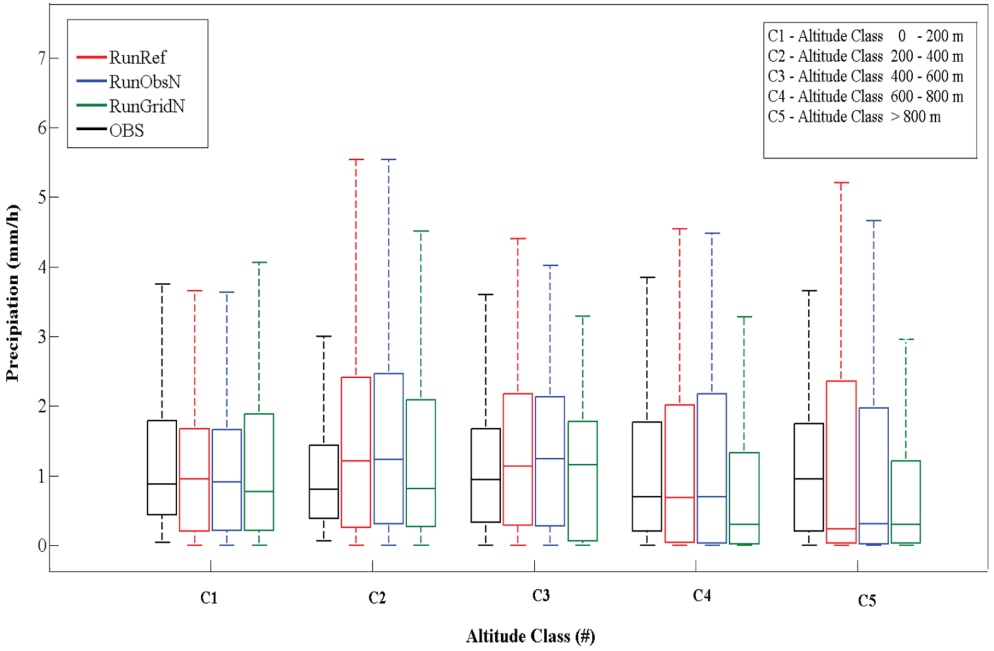
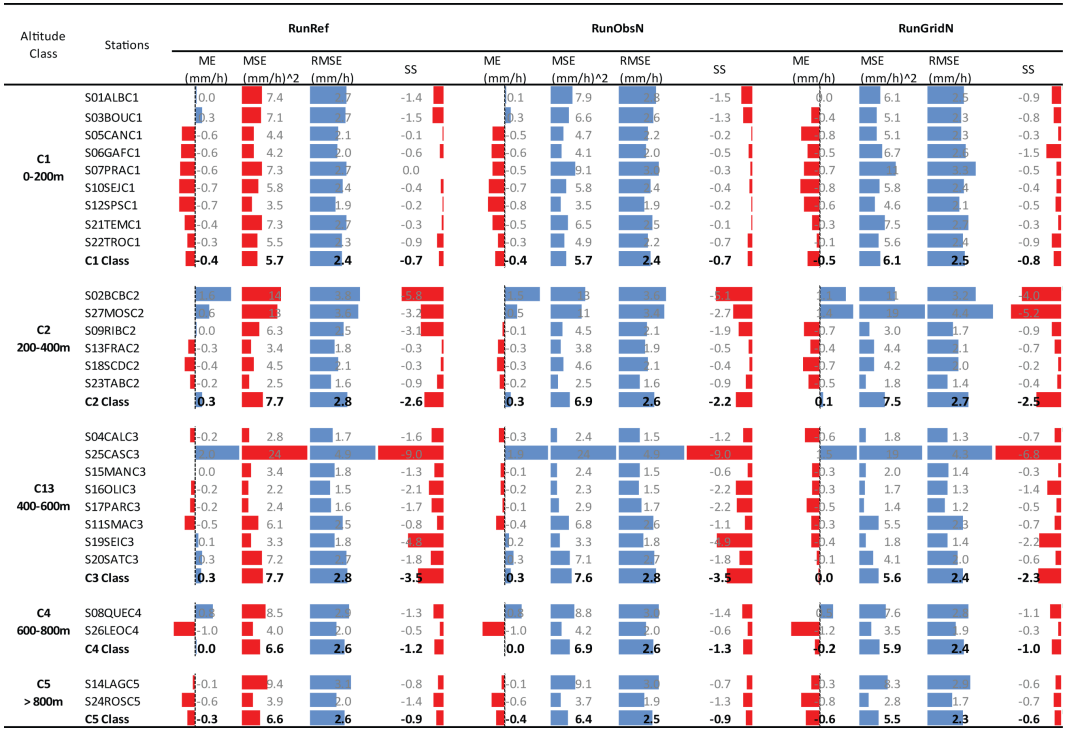


Fig. 5. Box plot for the observations and for the three experiments. The horizontal box line represents the median (50th percentile), the lower line the 25th percentile, the upper line the 75th percentile. The dashed lines represents 1.5 times IQR.

HESSD
10, 1423–1463, 2013



Altitude Class	Stations	RunRef ME (mm/h)	RunRef MSE (mm/h)^2	RunRef RMSE (mm/h)	RunRef SS	RunObsN ME (mm/h)	RunObsN MSE (mm/h)^2	RunObsN RMSE (mm/h)	RunObsN SS	RunGridN ME (mm/h)	RunGridN MSE (mm/h)^2	RunGridN RMSE (mm/h)	RunGridN SS
C1 0-200m	S01ALBC1	0.0	7.4	2.7	-1.4	0.1	7.9	2.8	-1.5	0.0	6.1	2.5	-0.9
	S03BOUC1	0.3	7.1	2.7	-1.5	0.3	6.6	2.6	-1.3	0.4	5.1	2.3	-0.8
	S05CANC1	-0.6	4.4	2.1	-0.1	-0.5	4.7	2.2	-0.2	-0.8	5.1	2.3	-0.3
	S06GAFC1	-0.6	4.2	2.0	-0.6	-0.6	4.1	2.0	-0.5	-0.5	6.7	2.6	-1.5
	S07PRAC1	-0.6	7.3	2.7	0.0	-0.5	9.1	3.0	-0.3	-0.7	8.1	2.8	-0.5
	S10SEJC1	-0.7	5.8	2.4	-0.4	-0.7	5.8	2.4	-0.4	-0.8	5.8	2.4	-0.4
	S12SPSC1	-0.7	3.5	1.9	-0.2	-0.8	3.5	1.9	-0.2	-0.6	4.6	2.1	-0.5
	S21TEMC1	-0.4	7.3	2.7	-0.3	-0.5	6.5	2.5	-0.1	-0.3	7.5	2.7	-0.3
	S22TROC1	-0.3	5.5	2.3	-0.9	-0.3	4.9	2.2	-0.7	-0.1	5.6	2.4	-0.9
	C1 Class	**-0.4**	**5.7**	**2.4**	**-0.7**	**-0.4**	**5.7**	**2.4**	**-0.7**	**0.5**	**6.1**	**2.5**	**-0.8**
C2 200-400m	S02BCBC2	-1.8	14	3.8	-4.8	-1.8	13	3.6	-4.1	-1.4	11	3.3	-4.0
	S27MOSC2	-1.5	13	3.6	-3.2	-1.5	11	3.4	-2.7	-1.4	19	4.4	-5.2
	S09RIBC2	0.0	6.3	2.5	-3.1	-0.1	4.5	2.1	-1.9	0.7	3.0	1.7	-0.9
	S13FRAC2	-0.3	3.4	1.8	-0.3	-0.3	3.8	1.9	-0.5	0.4	4.4	2.1	-0.7
	S18SCDC2	-0.4	4.5	2.1	-0.3	-0.3	4.6	2.1	-0.4	0.7	4.2	2.0	-0.2
	S23TABC2	-0.2	2.5	1.6	-0.9	-0.2	2.5	1.6	-0.9	0.5	1.8	1.4	-0.4
	C2 Class	**0.3**	**7.7**	**2.8**	**-2.6**	**0.3**	**6.9**	**2.6**	**-2.2**	**0.1**	**7.5**	**2.7**	**-2.5**
C13 400-600m	S04CALC3	-0.2	2.8	1.7	-1.6	-0.3	2.4	1.5	-1.2	0.6	1.8	1.3	-0.7
	S25CASC3	-2.8	24	4.9	-9.0	-2.8	24	4.9	-9.0	-2.5	19	4.3	-6.8
	S15MANC3	0.0	3.4	1.8	-1.3	-0.1	2.4	1.5	-0.6	0.3	2.0	1.4	-0.3
	S16OLIC3	-0.2	2.2	1.5	-2.1	-0.2	2.3	1.5	-2.2	0.3	1.7	1.3	-1.4
	S17PARC3	-0.2	2.4	1.6	-1.7	-0.1	2.9	1.7	-2.2	0.5	1.4	1.2	-0.5
	S11SMAC3	-0.5	6.1	2.5	-0.8	-0.4	6.8	2.6	-1.1	0.3	5.5	2.3	-0.7
	S19SEIC3	0.1	3.3	1.8	-1.8	0.2	3.3	1.8	-1.9	0.4	1.8	1.4	-2.2
	S20SATC3	0.3	7.2	2.7	-1.8	0.3	7.1	2.7	-1.8	-0.1	4.1	2.0	-0.6
	C3 Class	**0.3**	**7.7**	**2.8**	**-3.5**	**0.3**	**7.6**	**2.8**	**-3.5**	**0.0**	**5.6**	**2.4**	**-2.3**
C4 600-800m	S08QUEC4	0.3	8.5	2.9	-1.3	0.3	8.8	3.0	-1.4	0.2	7.6	2.8	-1.1
	S26LEOC4	-1.0	4.0	2.0	-0.5	-1.0	4.2	2.0	-0.6	-1.2	3.5	1.9	-0.3
	C4 Class	**0.0**	**6.6**	**2.6**	**-1.2**	**0.0**	**6.9**	**2.6**	**-1.3**	**-0.2**	**5.9**	**2.4**	**-1.0**
C5 > 800m	S14LAGC5	-0.1	9.4	3.1	-0.8	-0.1	9.1	3.0	-0.7	0.3	8.3	2.9	-0.6
	S24ROSC5	-0.6	3.9	2.0	-1.4	-0.6	3.7	1.9	-1.3	-0.8	2.8	1.7	-0.7
	C5 Class	**-0.3**	**6.6**	**2.6**	**-0.9**	**-0.4**	**6.4**	**2.5**	**-0.9**	**-0.6**	**5.5**	**2.3**	**-0.6**

Fig. 6. Continuous verification measurements of the hourly precipitation (above 0.1 mm h^{-1}) for the each of the stations divided by altitude class. The red bars denote negative values of the measurements and the positive are presented by the blue bars.

HESSD
10, 1423–1463, 2013

Simulation of a persistent medium-term precipitation event

S. C. Pereira et al.

Fig. 7. Categorical verification measurements for the hourly precipitation (above 0.1 mm h^{-1}) for each of the stations divided by altitude class. The blue bars represented the measurements that are unity oriented (perfect score = 1). The green bar stands for the PC measurement and the orange bars for measurements that the perfect score is zero.

HESSD
10, 1423–1463, 2013