Sea breeze thunderstorms in the eastern Iberian Peninsula. Neighborhood verification of HIRLAM and HARMONIE precipitation forecasts

Cesar Azorin-Molina[1,*], Sander Tijm[2], Elizabeth E. Ebert[3], Sergio M. Vicente-Serrano[1], María J. Estrela[4]

Surnames (or family names) are underlined

1. *Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Departamento de Procesos Geoambientales y Cambio Global, Avda. Montañana 1005, 50059-Zaragoza, Spain*

2. *Royal Netherlands Meteorological Institute (KNMI), PO Box 201, NL-3730 AE De Bilt, The Netherlands*

3. *Centre for Australian Weather and Climate Research (CAWCR), Bureau of Meteorology, GPO Box 1289, VIC 3001 Melbourne, Australia*

4. *Laboratory of Meteorology-Climatology, Mixed Unity CEAM-UVEG, Department of Physical Geography, University of Valencia, Avda. Blasco Ibáñez 28, 46010-Valencia, Spain*

* Corresponding author at: Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Departamento de Procesos Geoambientales y Cambio Global, Avda. Montañana 1005, 50059-Zaragoza, Spain. Tel.: (+34) 97.671.60.34 ; fax: (+34) 97.436.32.22

E-mail address: cazorin@ipe.csic.es (C. Azorin-Molina)

Manuscript published in Atmospheric Research

ABSTRACT

In this study we investigated sea breeze thunderstorms with intense convective activity (i.e., heavy rainfall, hail and gusty winds) that occurred over the eastern Iberian Peninsula (Spain) and were missed by the operational HIRLAM model. We used two grid-spacing setups (5.0-km and 2.5-km) of the hydrostatic HIRLAM model, and the non-hydrostatic spectral HARMONIE suite (2.5-km), to simulate isolated convection associated with sea breezes. The overall aim is to estimate the ability of these three experimental setups, in particular the HARMONIE model as the forthcoming operational Numerical Weather Prediction in most European Weather Services, to correctly simulate convective precipitation associated with sea breezes. We evaluated high-resolution gridded precipitation forecasts from HIRLAM and HARMONIE suites for 15 sea breeze thunderstorms against high-density gridded raingauge measurements applying different neighborhood verification techniques. The results indicate that higher horizontal resolutions of HIRLAM and HARMONIE models succeeded in predicting the occurrence of these missed sea breeze thunderstorms, the HARMONIE suite being the most capable of providing good estimates of accumulated precipitation in convective events in terms of space and time. Advances in quantitative precipitation forecasting of locally driven convection could have practical applications for nowcasting dangerous sea breeze convective phenomena.

Keywords: sea breeze thunderstorm; operational forecasting; HIRLAM; HARMONIE; neighborhood verification; Iberian Peninsula.

1. Introduction

The effect of sea breezes in triggering deep convection has been noted in numerous coastal areas around the world (Simpson, 1994). Many numerical modelling and observational studies, particularly for the subtropical Florida Panhandle (e.g., Pielke, 1974; Pielke and Cotton, 1977; Pielke and Mahrer, 1978; Blanchard and Lopez, 1985; Nichols et al., 1991; Wakimoto and Atkins, 1994), found that low-level sea breeze convergence and consequently sea breeze front (SBF) development enhance planetary boundary layer (PBL) air parcels to lift up to the level of free convection (LFC) (Wilson, 2008; Muppa et al., 2012). Convective initiation can also occur from horizontal convective rolls (HCR) ahead of the advancing SBF (Atkins et al., 1995). Both SBF and HCR updrafts merge, promoting deep convective sea breeze thunderstorms that can sometimes be extraordinarily severe, causing significant rainfall, hail and gusty winds along the frontal boundary (Dailey and Fovell, 1999; Fovell and Dailey, 2001; Fovell, 2005).

Convection resulting from horizontal low-level convergence along SBFs is intensified by several factors (Simpson et al., 2007). The most relevant are as follows: (a) anabatic valley wind circulations on heated south-facing mountain slopes play a major role in strengthening uplift processes and developing deep convection by combining with moist sea breezes that transport water vapor to the coastal mountains in the daytime (Millan et al., 2005); (b) convergence and SBF intensification is also enhanced by frictional effects (upslope orographic lifting) produced in coastal areas of complex terrain (Petterssen, 1956; Pérez-Landa et al., 2006; Papanastasiou et al., 2010); (c) strong low-level sea breeze convergence also occurs on convex coastlines (e.g.,

81 peninsulas, capes, points, and also islands) contributing to increased upward vertical
82 motion, *Cumulus* (Cu) and *Cumulonimbus* (Cb) activity and thunderstorms along the
83 frontal area (Neumann, 1951; Pielke, 1974; Purdom, 1976; Strickler, 2003); (d) a
84 sharply defined discontinuity and convergence intensification is observed inland along
85 the SBF under offshore large-scale synoptic flows (Bechtold et al., 1991; Atkins and
86 Wakimoto, 1997), and light to moderate winds aloft (<5.0 m s^{-1}) result in more clouds at
87 the leading edge of sea breezes (Azorin-Molina et al., 2009); and (f) regions of high soil
88 moisture expect heavy precipitation along the SBF (Baker et al., 2001). Most of the
89 aforementioned factors interact over the complex eastern façade of the Iberian Peninsula
90 (Spain), where thermally induced local circulations interact.

91 Short-term forecasting of the timing, location and intensity of isolated sea breeze
92 thunderstorms represents a challenging task in numerical weather prediction (NWP),
93 mainly due to the uncertainties in the initial conditions, limited knowledge about the
94 cloud microphysical processes, and difficulties in resolving low-level sea breeze
95 convergence and convection with fairly coarse horizontal resolution operational NWP
96 models (Mazarakis et al., 2009). Furthermore, isolated sea breeze thunderstorm cells
97 with severe weather can develop unexpectedly under weakly defined synoptic-scale or
98 mesoscale precursor disturbances (Wilson, 2008), and may be missed by forecasts.
99 Strong low-level sea breeze boundaries can deliver enough energy to overcome the
100 stable cap of the PBL and generate local showers and thunderstorms unexpectedly. For
101 instance, deep convection associated with sea breezes can occur even when sounding
102 indices indicate stable weather conditions. This is because sounding indices do not
103 consider layers below 850 hPa, where strong low level convergence may accumulate
104 lower tropospheric moisture (Pielke et al., 1971) or capping of inversions can occur,
105 enhancing or weakening lifting mechanisms, respectively.

Low-level sea breeze convergence occurs preferentially during the warm season (May-September), but also in the transition months of April and October, over the complex eastern façade of the IP. Its time of occurrence is nearly always in the mid-afternoon. Despite the high level of occurrence of thermally-driven winds (sea breezes blow two out of three days of the year, Azorin-Molina and Martin-Vide, 2007) and also that sea breeze convection brings an average of 100-125 mm yearly to inland areas during the summer dry season (Millan et al., 2005), there is little knowledge concerning the important role that sea breezes play in convection initiation in the eastern coast of Spain. A review of these few investigations was presented by Azorin-Molina et al. (2009), who used high-resolution cloud frequency composites derived from NOAA-AVHRR data to identify the location of five preferential sea breeze convergence zones (SBCZ; hot spots) in the Iberian Mediterranean area and the Isle of Mallorca. The current study is focused on two of them, i.e., the SBCZ2 (eastern region of the Iberian system mountains; 1000-1900 m) and the SBCZ3 (Prebetic mountain ranges; 1000-1600 m). These regions correspond to the east of the Iberian Peninsula, an orographically highly complex area (Figure 1a).

The main goal of this study is to estimate the ability of NWP to correctly simulate convective precipitation associated with sea breezes. Two different setups (5.0-km and 2.5-km horizontal grid-spacing) of the operational HIRLAM model and the non-hydrostatic spectral HARMONIE suite (2.5-km horizontal grid-spacing; Hirlam Aladin Regional/Meso-scale Operational NWP In Europe) are evaluated. Three different neighborhood (also known as 'fuzzy') verification techniques are applied here in order to measure the strength of the HIRLAM and HARMONIE agreement with the observations. The article is structured as follows. In section 2, we summarize the model description, set up and initialization of HIRLAM and HARMONIE suites, present the

sea breeze thunderstorms simulated and observed precipitation data, and briefly describe the neighborhood verification methods applied in this study. In section 3, the performance of HIRLAM and HARMONIE gridded precipitation forecasts is evaluated against gridded precipitation observations. In section 4, the sea breeze thunderstorm occurred on 7 August 2008 is analyzed. Finally, a summary and discussion of the findings from this study are presented in section 5.

2. Data and methods

2.1. HIRLAM and HARMONIE model description, set up and initialization

The NWP systems used here are the three-dimensional hydrostatic grid-point model version 7.2.2 of HIRLAM (Undén et al., 2002), and the non-hydrostatic spectral model version 36h1.2 of HARMONIE (Seity et al., 2011). The HIRLAM short-range forecasting model was chosen for this research because it is currently employed as one of the most important operational NWP system at the AEMET, and by eight other European Weather Services: Denmark, Estonia, Finland, Iceland, Ireland, the Netherlands, Norway and Sweden. The HARMONIE limited area model is also used in this study because it will be implemented shortly as the operational NWP in most of the European Weather Services. Actually, it is currently being evaluated as a next-generation replacement for HIRLAM (van der Plas et al., 2012; de Brujin and de Rooy, 2012).

The dynamical core of HIRLAM model is based on a semi-implicit semi-Lagrangian discretization of the multi-level primitive equations, employing a hybrid coordinate in the vertical. The comprehensive set of physical parameterization schemes

154 selected in the HIRLAM model suite in order to take into account a variety of sub-
155 gridscale physical processes include: (a) a radiation scheme (Savijärvi, 1990), (b) an
156 adapted Rasch-Kristjansson condensation (Zhang et al., 2003; Ivarsson, 2007) and a
157 Kain-Fritsch mass-flux convection scheme with CAPE closure (Kain, 2004; Calvo,
158 2007) , (c) a prognostic moist turbulent kinetic energy (TKE) parameterization (Tijm
159 and Lenderink, 2003), (d) a tiled surface approach distinguishing seven surface types
160 (Interaction Soil-Biosphere-Atmosphere, ISBA surface scheme, Noilhan and Planton,
161 1989; snow and forest, Gollvik, 2002, 2004; lake, Mironov, 2008, Kourzeneva et al.,
162 2008), and (e) a mean and subgrid-scale orography parametrization (Rontu et al., 2002).
163 The HIRLAM system also includes a 4D-VAR data assimilation system (Huang et al.,
164 2002) with analysis of conventional and non-conventional observations. More detailed
165 descriptions about the dynamical and numerical features of the HIRLAM short-range
166 forecasting model can be found in Undén et al. (2002) and at the following URL:
167 http://hirlam.org/index.php?option=com_content&view=article&id=64:general-
168 description-of-the-hirlam-model&catid=48:synoptic-scale-model-hirlam&Itemid=101
169 (last accessed 1 November 2013).

170 The HARMONIE model suite is designed in a cooperation between the
171 HIRLAM and ALADIN (*Aire Limitée Adaptation dynamique Développement*
172 *InterNational*) consortia. The dynamical core of the mesoscale model HARMONIE is
173 developed by ALADIN high precision short-range NWP model and based on a two-
174 time level semi-implicit semi-Lagrangian discretization of the fully elastic equations,
175 also employing a hybrid coordinate in the vertical. ALADIN is a spectral limited-area
176 model developed at Météo-France and based on the ARPEGE/IFS model. The
177 HARMONIE mesoscale option consists of a combination of ALADIN's non-hydrostatic
178 dynamics and Mesoscale Non-Hydrostatic atmospheric model (NH;

179 http://mesonh.aero.obs-mip.fr/mesonh410; last accessed 1 November 2013) physics.
180 The parametrizations of a variety of sub-gridscale physical processes are the same as
181 those adopted of the AROME physics (Seity et al., 2011). The HARMONIE system
182 includes a 3D-VAR data assimilation system. A more detailed description of the
183 HARMONIE model can be found at this URL:
184 http://hirlam.org/index.php?option=com_content&view=article&id=65&Itemid=102
185 (last accessed 1 November 2013).

186 Here, we are interested in evaluating the precipitation performance of the
187 HIRLAM 0.05º (4.3-km length; hereafter HIR-D5) horizontal grid-spacing against a
188 higher horizontal grid-resolution of HIRLAM 0.025º (2.5-km length; hereafter HIR-
189 D2.5) and HARMONIE 0.025º (2.5-km length; hereafter HAR-D2.5), for those sea
190 breeze thunderstorm episodes described in section 2.2. On the one hand, the HIRLAM
191 model is set up with unique domains with an integration area of 300 x 306 horizontal
192 grid points (latitude by longitude) and the size of 1495 km x 1525 km for the HIR-D5
193 experiment, and 400 x 406 horizontal grid points and 997.5 km x 1012.5 km for the
194 HIR-D2.5 experiment. Increased boundary layer vertical resolution can provide
195 improved forecast performance (Ries and Heinke-Schlunzen, 2009), and therefore the
196 vertical domain for the two set ups of the HIRLAM contains 60 sigma levels with
197 enhanced resolution (27 levels in the lowest 3 km) in the PBL. The model top is at 10
198 hPa and the lowest model level is around 32 m. On the other hand, the HAR-D2.5
199 experiment is set up with an integration area of 400 x 400 horizontal grid points
200 (latitude by longitude) and the size of 1000 km x 1000 km, also employing a vertical
201 domain of 60 hybrid levels with enhanced resolution (27 levels in the lowest 3 km) in
202 the PBL. The HARMONIE model top is at 1 hPa and the lowest model level is around
203 10 m. The ECMWF archive supplies the operational boundaries data for both HIRLAM

204 and HARMONIE suites with a horizontal grid-spacing of 0.5º x 0.5º and a temporal
205 resolution of 6 h. These boundaries are also applied as initial conditions. Lateral
206 boundaries are overspecified, all variables being externally prescribed by the nesting
207 model. A relaxation zone of 10 grid points is normally adopted and boundary relaxation
208 is performed after the horizontal diffusion. Additionally, at the upper boundary a
209 condition of zero vertical velocity is imposed. The dynamics time step is 120 s for the
210 HIR-D5, and 60 s for the HIR-D2.5 and the HAR-D2.5 experiments. All the
211 simulations were initialized at 1200 UTC the day before the sea breeze thunderstorm
212 event, and integrated for 48 h until 1200 UTC the day after the episode, with model
213 outputs at intervals of 1 h. The domains for each NWP system are represented in Figure
214 1b.

215

216 *2.2. Simulated sea breeze thunderstorms and observed precipitation data*

217 We exclusively focused on 15 deep convective sea breeze thunderstorms that
218 were missed by the operational HIRLAM model; i.e., 'observed but not forecasted
219 episodes'. A detailed verification of 'forecasted but not observed events' (i.e., false
220 alarms) is also an interesting topic for future work, but is out of the scope of this study
221 aimed in analyzing the ability of hydrostatic and non-hydrostatic convection-permitting
222 models in resolving these sea breeze thunderstorms.

223 These thunderstorms corresponded to typical sea breeze fronts that developed
224 explosively under weakly defined synoptic-mesoscale forcing; i.e., weak surface
225 pressure gradient in the western Mediterranean basin. Some of these storms were
226 severe, causing significant amounts of rainfall, hail and gusty winds. This synoptic

227 pattern allowed the inflow of moisture driven by well-developed local thermal
228 circulations, which mainly triggered these storms at the low-level convergence zone of
229 sea breezes. All these convective events were isolated and stationary as was revealed in
230 the geostationary and polar satellite images, Doppler radar, and lightning data. Table I
231 summarizes the list of sea breeze thunderstorms simulated, the number of raingauge
232 station points used for the neighborhood verification, and the maximum of precipitation
233 recorded for each episode at 06 UTC.

234 With the aim to evaluate the performance of HIRLAM *versus* HARMONIE
235 precipitation forecasts of these sea breeze thunderstorm episodes, this research used 24
236 h-precipitation accumulations (from 0800 UTC till 0800 UTC) from a high-density
237 network of Hellman-type raingauges covering the eastern region of the IP (Figure 1b).
238 The observed precipitation data have been supplied mainly by the AEMET, but we also
239 used a small number of observations from automatic weather stations (AWS; tipping
240 bucket raingauges) collected from the Mediterranean Center for Environmental Studies
241 (CEAM; http://www.ceam.es/index_i.htm; last accessed 1 November 2013), the
242 Automatic System of Hydrological Information of the Ebro basin management agency
243 (SAIH; http://www.saihebro.com/saihebro/index.php; last accessed 1 November 2013),
244 the Valencian Institute for Agriculture Research (IVIA; http://www.ivia.es/; last
245 accessed 1 November 2013), the Agro-climatic Information System for Irrigation
246 (SIAR, http://crea.uclm.es/siar/; SIAM, http://siam.imida.es/; last accessed 1 November
247 2013) and the METEOCLIMATIC network (http://www.meteoclimatic.com/; last
248 accessed 1 November 2013). The high-density raingauge network consists of a
249 minimum of 304 and a maximum of 533 observation points depending on each episode
250 and covers an area bounded between 37º00'N and 41º00'N and 3º00'W and 1º00'E, i.e.
251 both sea breeze convergence zones (SBCZ2 and SBCZ3; see Fig. 1a) with intensified

252 convective activity. Since data were obtained from various institutions, the 24-h raw
253 accumulated precipitation passed quality control (Aguilar et al., 2003) based on
254 removing gross errors due to archiving, transcription and digitalization. This basically
255 consisted on screening out suspicious precipitation values (outliers), such as extreme
256 and large values for 10-minute, 30-minute or daily measurements.

257 The precipitation forecasts and observations were gridded onto a regular grid
258 with dimensions of 81x81 grid points for the 5-km runs, and 161x161 for the 2.5-km
259 runs, resulting in a grid forecast and grid observation with the same dimensions for each
260 episode and model run. For gridding the observations into a regular grid we placed the
261 point observation values at the grid points closest to the observation points. When more
262 than one observation was present in a grid box of the single intercomparison grid, which
263 occurred for only two grid points in the total 15 sea breeze thunderstorms, the highest
264 value was chosen. Here we compared 24 h-precpitation accumulations (from 08:00
265 UTC till 08:00 UTC) from the gridded simulated and observed precipitation boxes.

266

267 *2.3. Neighborhood or 'fuzzy' verification methods*

268 The statistical verification of high-resolution gridded precipitation forecasts such
269 as those from HIRLAM and HARMONIE model suites against raingauge data is
270 increasingly conducted by applying neighborhood or 'fuzzy' methods. We decided to
271 use this approach because traditional verification metrics (e.g., mean square error, root
272 mean square error, etc.) typically give poor and uninformative values for high resolution
273 forecasts of small-scale features (e.g., low-level convergence of sea breezes triggering
274 isolated convection) since exact matches between forecasts and observations are quite

275 difficult to achieve (Ebert, 2008). However, neighborhood methods use a spatial
276 window or neighborhood surrounding the forecast and observed points and therefore
277 relax the condition for exact matches between forecasted and observed precipitation in
278 space, time and intensity.

279 Three of the single observation-neighborhood forecast approaches are used here
280 for evaluating the precipitation performance of the HIRLAM and the HARMONIE
281 simulations during the 15 sea breeze thunderstorm events ('observed but not forecast
282 episodes'). The three methods compare a neighborhood of forecast values to the
283 observation in the centre of the neighborhood. This mimics how a forecaster might
284 interpret high resolution model outputs when making a prediction for a particular
285 location. These particular methods were selected to assess the frequency of the forecasts
286 predicting appropriately high rainfall near the observed location, and to evaluate the
287 spatial scales over which a forecaster might consider the model output to be useful.

288 The first neighborhood technique corresponds to the 'multi-event contingency
289 table' proposed by Atger (2001), which measures the closeness of the forecasted
290 precipitation to the raingauge observations for multiple decision intensity, spatial and
291 temporal thresholds. Basically, the 'multi-event contingency table' builds on the
292 traditional 2x2 contingency table, in which hits, misses, false alarms, and correct
293 rejections are counted for forecasted and observed occurrences of an event (for
294 precipitation an event is usually defined as rain exceeding a given intensity threshold).
295 The method combines the customary intensity dimension with additional thresholds
296 such as spatial and temporal closeness. The method is based on the criterion that a
297 forecast is useful if the event is forecasted in at least one grid box within a specified
298 distance *s* of an observed event. If that is the case, then the forecast is counted as a hit; if

not, then the forecast is a miss. Forecasts for rain within a distance *s* of a location where rain was not observed are counted as false alarms.

The Hanssen and Kuipers (HK) score is computed in order to measure the success of the precipitation forecasts. The HK score (also known as true skill statistics - TSS- and Peirce´s skill scoce -PSS-) represents a balance between the hit rate (H) and false alarm rate (F), i.e., HK=H-F, with a perfect value of 1:

$$H = \frac{hits}{hits + misses} \qquad (1) \qquad F = \frac{false\ alarms}{correct\ rejections\ +\ false\ alarms} \qquad (1)$$

The second neighborhood method applied here corresponds to the pragmatic technique defined by Theis et al. (2005) which uses the observed precipitation in the central grid box as a value to verify the model performance against the forecast precipitation probability within a neighborhood. A useful forecast has skill at detecting precipitation events and non-events, that is, it has high probabilities in the vicinity of observed events and low probabilities in the vicinity of observed non-events. The probabilistic verification metric corresponds to the Brier skill score (BSS) with respect to the sample climatology. The reference forecast (i.e., the forecast to beat) is the fraction of observations over the whole domain with rain meeting or exceeding the threshold.

$$BSS = 1 - \frac{\frac{1}{N}\sum_{N}(\langle P_y \rangle_s - I_x)^2}{\frac{1}{N}\sum_{N}(\overline{P_x} - I_x)^2} \qquad (2)$$

where N is the number of observation locations, I_x is an indicator (1=yes, 0=no) for an observed event at each location, $\langle P_y \rangle_s$ is the fraction coverage of forecast events within

the neighborhood of radius s, and $\overline{P_x}$ is the domain mean value of I_x. BSS varies between minus infinity and 1, with larger values indicating more skillful forecasts.

The third and last neighborhood verification technique applied here is the conditional square root of ranked probability score (CSRR), defined by Germann and Zawadzki (2004). The definition of a useful forecast according to this metric is one with a high probability of matching the observed precipitation value. This method is based on the computation of the ranked probability score (RPS), which uses the frequencies of forecast precipitation values in logarithmically increasing intervals. The RPS quantifies the sum of squared differences in cumulative probability space for a multi-category probabilistic forecast (e.g., Wilks, 2011):

$$RPS = \frac{1}{M-1}\sum_{m=1}^{M}(CDF_{y,m} - I_{x,m})^2 \tag{3}$$

where M is the number of forecast categories, $\mathrm{CDF}_{y,m}$ is the cumulative probability of the forecast exceeding the threshold for category m, and $I_{x,m}$ is 1 if the observed rainfall meets or exceeds the threshold for category m and 0 if not. The square root of the RPS is the standard error of the forecast probability in probability space, and therefore values closer to 0 indicate better forecasts. In order to compare different rainfall episodes the square root of the RPS is normalized by the observed rain fraction.

$$CSRR = \frac{\sqrt{RPS}}{\overline{P_{x>0}}} \tag{4}$$

All three fuzzy methods were applied to the individual gridded forecasts. For each method and model-observation category, the results were aggregated over all cases

and plotted both as a function of neighborhood size and as quilt plots. The window sizes were 1x1, 3x3, 5x5, up to 25x25, only plotting results for spatial scales from 2.5 up to 97.5-km. The rainfall thresholds used to define events and categories (needed by the various scores) were 0.2, 1.0, 2.0, 5.0, 10.0, 20.0, and 50.0 mm, where an event is defined as the rain equal to or exceeding the threshold. A more detailed description of these three neighborhood verification techniques and metrics can be found in the review article of Ebert (2008).

In section 3 we present neighborhood verification results for all the 15 sea breeze thunderstorms and show an example of good high-resolution precipitation forecast performance for an isolated sea breeze thunderstorm developed over the Iberian System Mountains (ISM) on 7 August 2008.

3. Results

3.1. Rain distribution from gridpoint match-ups

The rain distribution plot displayed in Figure 2 shows that all three high-resolution gridded rainfall forecasts overestimate the observed point precipitation intensity distribution, with the HIR-D2.5 having a slightly greater over-prediction bias than the coarser HIR-D5 model. The HAR-D2.5 did a much better job of representing the observed rain amount distribution. The models also differed in their ability to predict the maximum rainfall. For instance, forecast rain accumulations exceeding 20 mm are uncommon in the HIR-D5 (sample maximum of 22 mm), whereas the higher resolution models produced much higher rain accumulations (sample maxima of 113 mm for HIR-D2.5 and 64 mm for HAR-D2.5). The observed maximum rain amount

362 was 90 mm (see Table 1). The overprediction bias has an impact on the neighborhood
363 scores discussed below.

364

365 *3.2. Multi-event contingency table method*

366 The multi-event contingency table technique, which checks whether a
367 precipitation forecast event is found near an observed value, suggest that the HIRLAM
368 model provided better forecasts than the HARMONIE model when low rain thresholds
369 were chosen, but for precipitation exceeding about 5 mm the HARMONIE model
370 performed more reasonably. This overall feature is clearly discernible looking at the
371 quilt plots shown in Figure 3 which display the HK metric for different rain intensity
372 thresholds (0.2, 1.0, 2.0, 5.0, 10.0, 20.0 and 50.0 mm; *x*-axes) as a function of various
373 window sizes (from 2.5 up to 97.5-km; *y*-axes). The magnitude of HK score is plotted
374 and the shade gives an idea of the model´s performance, i.e., darker (orange colors) and
375 lighter (blue colors) shades represent better and poorer skills, respectively. In addition,
376 an intensity dependent skillful scale can be obtained by drawing a line through the
377 better score in the verification HK results as shown in Figure 3. Table 2 compares the
378 skillful scales and accompanying HK scores for three models. From the quilt plots and
379 table we can conclude that for rain thresholds of 5.0 mm and greater, the optimal scale
380 for the HARMONIE model is smaller than for the HIRLAM suites, meaning that one
381 needs to look over a smaller distance to find a forecast matching the observation,
382 without having too many false alarms for non-events. For instance, the 0.2 mm rain
383 threshold is best achieved at grid scale by the HIR-D2.5 (2.5 km; HK 0.34; Figure 3a)
384 and the HIR-D5 (5.0 km; HK 0.31; Figure 3b), and at a slightly higher scale but with
385 greater accuracy by the HAR-D2.5 (7.5 km; HK 0.37; Figure 3c). Higher horizontal grid

386 resolution for both the HIR-D2.5 and the HAR-D2.5 is crucial for obtaining smaller
387 optimal scales for rain intensity thresholds equal or greater than 1.0 mm in comparison
388 to the coarser HIR-D5. The optimal scales and HK scores found for the rain threshold of
389 50.0 mm are not representative since there were only 3 observations exceeding this
390 intensity.

391 Comparing the 2.5-km models the HAR-D2.5 model outperformed the HIR-
392 D2.5 model for most rainfall thresholds and scales, with HIR-D2.5 only scoring better
393 than HAR-D2.5 at the finest 2.5-km grid scale and large scales for 10-20 mm rain
394 thresholds. These results confirm the improved ability of the forthcoming non-
395 hydrostatic convection-permitting HARMONIE model suite as an operational NWP
396 model to capture isolated convection associated with mesoscale / local winds.

397

398 *3.3. Pragmatic method*

399 The BSS is used as a metric for the pragmatic neighborhood verification
400 technique in order to test how useful forecasts are in detecting events and non-events.
401 The reference forecast (i.e., the forecast to beat) is the fraction of observations in each
402 case that are events (i.e., the sample climatology over the whole domain). Figure 4
403 shows the quilt diagrams where the magnitude of the BSS metrics is plotted as a
404 function of the aforementioned intensity precipitation thresholds (x-axes) and the range
405 of spatial scales (y-axes). In general, the pragmatic method suggests that the forecasts
406 had poor skill for most rain rates and scales, with the HARMONIE scheme performing
407 best according to the BSS scores. To resolve this apparent contradiction in comparison
408 to the reasonable HK scores shown in Figure 3, it is important to recognize that both

neighborhood methods are addressing different questions; the pragmatic approach is a much tougher test than the "at least one nearby forecast event". Because all three models over-predicted the number of rain events, it was difficult to beat the sample climatology and the skill was negative for all scales and thresholds as shown in Figures 4a (HIR-D2.5), 4b (HIR-D5) and 4c (HAR-D2.5), and summarized in Table 3. We can conclude that, according to this score, poorer performance occurred at small scales and low to moderate thresholds. This skill according to the BSS metric was poorer for HIR-D2.5 than for HIR-D5, while the HAR-D2.5 had the best performance due partly to its more appropriate rain intensity distribution.

3.4. Conditional square root of RPS

The conditional square root of RPS method, which is an error metric with a perfect value of 0, also showed the best performance for the HAR-D2.5, followed by the HIR-D5 and the HIR-D2.5. As with the pragmatic method, the CSRR scores improve when increasing to larger scales as shown in Figure 5. This method rewards forecasts that have a peak in the (neighborhood-based) probability density function that is close to the observed value. The threshold dependence is integrated into the overall score through the forecast PDF. The forecast precipitation for the HIR-D2.5 tended to be higher than for the HIR-D5 and the HAR-D2.5, which means that the forecast PDF peaked even further away from the observation (see Fig. 2).

4. Sea breeze front case study: 7 August 2008

4.1. Observational data

The case study corresponds to a typical SBF that was not forecasted by the operational HIRLAM numerical weather prediction and occurred over the Iberian system mountains (SBCZ2) on 7 August 2008. We chosen this particular event because the highest availability of grid observation points (i.e., 533 gauges; see Table 1) which helped for better evaluating the performance of HIRLAM and HARMONIE forecasts. This non-forecasted sea breeze thunderstorm produced severe convective activity with heavy showers, large hail up to 3.5 cm in diameter, and gusty winds, causing significant damage in some fruit groves and irrigated areas for vegetables, and also damaged vehicles. The precipitation map displayed in Figure 6 confirms that the study case corresponded to a very isolated sea breeze thunderstorm cell which accumulated significant rainfall over a small area in the SBCZ2. For instance, 40 out of the 533 rain-gauge points, i.e. 7.5%, measured precipitation ≥0.1 mm, whereas just 6 meteorological stations reported rainfall ≥10.0 mm: Catí-Meteoclimatic (45.0 mm), Vallivana-CEAM (27.8 mm), Morella-CHE (15.8 mm), Vallibona-CHJ (11.2 mm), Catí-CHJ (11.2 mm), and El Boixar (10.0 mm).

4.2. Operational weather forecast

For the 7 August 2008 experiment the operational HIRLAM model used at the AEMET was version 6.1.2 with a European domain of 0.16º (16-km length) to simulate the large scale meteorological conditions, and an inner domain with fine resolution of 0.05º (5-km length) for the Iberian Peninsula and the Balearic Islands; 40 full sigma levels in the vertical direction were used. The operational HIRLAM 6.1.2 version from the AEMET partly failed in predicting this sea breeze thunderstorm, as is shown for

455 both the 0.16º (Figure 7a) and the 0.05º (Figure 7b) domains, where the 0.16° run gave
456 some light showers and the high resolution run did not produce any precipitation at all.

457

458 *4.3. HIRLAM and HARMONIE performance*

459 Figure 8 shows 24-h accumulated precipitation forecasts for the same
460 integration time from the HIR-D5 (Fig.8a), the HIR-D2.5 (Fig. 8b), the HAR-D2.5
461 (Fig. 8c), and (d) the HAR-D2.5 (Fig. 8d) with a factor of 5 increase in precipitation
462 fall speed. Both the HIRLAM and the HARMONIE research simulations succeeded in
463 capturing the occurrence of convective showers over the SBCZ2 in comparison to the
464 operational HIRLAM model version 6.1.2. However, the HIR-D5 run predicted a weak
465 widespread precipitation signal and displaced far to the northwest; the HIR-D2.5 run
466 simulated heavier rainfall over a larger area located far to the north; and the HAR-D2.5
467 run forecasted the precipitation at the right place with a reasonable amount. For
468 instance, the precipitation maxima accumulated in 24 h was 6.6 mm for the HIR-D5,
469 92.5 mm for the HIR-D2.5, and 28 mm for the HAR-D2.5; the observed maximum was
470 45.0 mm in the Cati station. Forecast rain accumulations exceeding 45.0 mm were not
471 uncommon in the HIR-D2.5 model (seven grid-points sampled 24-h accumulated
472 precipitation greater than this threshold), whereas the maximum 24-h gridded rain
473 accumulations in the HIR-D5 model were less than 10 mm. In contrast, the HAR-D2.5
474 produced a reasonable number of grid-points with maximum rain accumulations similar
475 to the observations. Summarizing, the HIR-D2.5 version tended to strongly
476 overestimate precipitation at the rain-gauge points; the HIR-D5 version tended to
477 strongly underestimate rainfall at rain-gauge points; and the HAR-D2.5 version
478 performed better precipitation forecasts than both HIRLAM set-ups in terms of amount

and location, because it correctly reproduced the heavy rainfall along the sea breeze boundary. We also tested the hypothesis that a large part of the overestimation of HIR-D2.5 in comparison to HAR-D2.5 is associated with the fact that the fall speed is infinity in the HIRLAM model suite, i.e., precipitation is put on the ground directly, which is quite normal in hydrostatic models until recently. This was investigated by increasing the fall speed of the rain, snow and graupel by a factor of 5 for the HAR-D2.5 model suite. Fig. 8d shows a significant increase of precipitation by a factor of 2 or more over a large part of the precipitation covered area in Fig. 8c, with a maximum precipitation of 55 mm whereas the original run produced a maximum precipitation amount of 28 mm. This difference can be explained by the increase in fall speed as this reduces the water loading of parcels in the updraft of the showers, increasing the updraft wind speed and associated precipitation formation. Additionally, precipitation spends less time in unsaturated air beneath the cloud, causing less evaporation of this precipitation.

Examination of the wind field shown in Figure 9 reveals that the three model setups resolved the development of a well-defined low-level convergence of sea breezes (easterly-southeasterly flows from the Mediterranean sea) and synoptic flows (westerly-southwesterly flows over the lower plateau of the IP). This convergence zone played a major role in initiating severe weather on the afternoon of 7 August 2008. However the location and strength of the low-level convergence of sea breezes is determined by the model resolution, and this could hypothetically explain why the position of the precipitation is much better in the HAR-D2.5 than in HIR-D5 and HIR-D2.5.

The neighborhood verification metrics used here did not consider closeness in time (Weusthoff, 2011), but time-series plots represent a way to look at the agreement between HIRLAM and HARMONIE model results and the observations in terms of timing. However, for this case-study significant precipitation amounts (i.e., ≥5.0 mm) were available in only three AWS, which are shown in Figure 10. Here we compared the hourly observed precipitation against the maximum modelled precipitation in an area with radius of 5-km around the AWS. For the Vallivana-CEAM station (27.8 mm; Fig. 9a) both the HIR-D5 and the HIR-D2.5 did not capture the hourly maxima of precipitation that occurred at 15 UTC (15.0 mm) and 16 UTC (12.7 mm), whereas the HAR-D2.5 correctly predicted the timing and also the amount (16.5 and 14.2 mm at 15 and 16 UTC, respectively). For the Morella-CHE station (15.8 mm; Fig. 9b), the hourly maxima of precipitation occurring at 15 UTC (4.6 mm) and 16 UTC (11.2 mm) were fairly reproduced by the HIR-D5 (2.3 mm at 16 UTC); the HIR-D2.5 detected two peaks of precipitation at the time of this episode (4.7 and 3.4 mm at 15 and 16 UTC, respectively) but incorrectly predicted precipitation during the rest of the day with a maximum peak of rainfall at 22 UTC (6.5 mm). The HAR-D2.5 best represented this episode, but with a maximum of 7.9 mm at 15 UTC, i.e., showing a timing error of one hour before the observed maximum. Lastly, for the Michavilla-CHE station (8.2 mm; Fig. 9c) again the HAR-D2.5 performed reasonably well for the hourly maximum of precipitation observed at 15 UTC (5.8 mm against 7.6 mm for the HAR-D2.5); whereas the HIR-D2.5 overestimated the amount and timing with a peak of maximum precipitation of 73.4 mm at 17 UTC; and the HIR-D5 captured an hourly maximum of 4.5 mm but one hour later than the observed event. Therefore, we can conclude that, for this particular case, the HARMONIE was also able to perform better in terms of timing in comparison with the HIRLAM set-ups.

527

528 *4.4. Neighborhood verification of the HAR-D2.5 precipitation forecast*

529 Tartaglione et al. (2005) pointed out that single case-study verification by means
530 of non-parametric statistical methods may lead to unstable or poor results, due to the
531 paucity of the statistical sample. In addition, with high resolution models like the ones
532 applied here and with the operational forecasts increasing in resolution at the Weather
533 Services it is quite difficult to get good verification results, even when radar data are
534 available, due to the difficulty in matching between forecasts and observations.
535 However, here we show the neighborhood verification of the HAR-D2.5 as an example
536 of good forecast performance of the non-forecasted sea breeze thunderstorm that
537 occurred on 7 August 2008. The neighborhood verification of the HAR-D2.5
538 precipitation forecast against the maximum observed rain is shown in Figure 11. The
539 multi-event contingency table (Fig. 10a) displays strong HK scores for most rainfall
540 thresholds, and particularly for neighborhood sizes greater than 10-km. The pragmatic
541 approach (Fig. 10b) also reported positive BSS scores for most spatial scales and
542 rainfall thresholds. Lastly, the conditional square root of RPS method (Fig. 10c)
543 showed lower (better) than average values of CSRR. In general, the HARMONIE
544 forecast precipitation overlapped quite well with several observations of heavy rain,
545 demonstrating the ability of this model to reproduce isolated convection associated with
546 low-level convergence of sea breezes.

547

548 **5. Summary and discussion**

549

In this study we investigated the ability of both the three-dimensional hydrostatic grid-point HIRLAM model (5.0-km and 2.5-km grid-spacing set-ups) and the forthcoming operational non-hydrostatic spectral HARMONIE suite (2.5-km) in predicting sea breeze thunderstorms that occurred over preferential convergence zones in the eastern façade of the Iberian Peninsula, Spain. We simulated a set of 15 sea breeze thunderstorms with intense convective activity that were not forecasted by the operational HIRLAM system. Neighborhood or 'fuzzy' verification techniques were applied to this set of high resolution regional forecasts making use of high-density gridded raingauge measurements. The aim was to statistically compare the performance of HIRLAM and HARMONIE suites with different grid resolutions in predicting this dangerous sea breeze convective phenomena in the Mediterranean area of the Iberian Peninsula.

The overall results revealed that both the more recent high resolution HIRLAM and the HARMONIE model succeeded in capturing the occurrence of these convective showers, which were missed in the operational HIRLAM runs, with the HARMONIE model being most capable of delivering high quality forecasts of isolated convection associated with sea breezes in terms of amount, location, and also time. To summarize, it was found that the HIRLAM model at 5-km predicted light to moderate precipitation signals (weak overprediction bias) and a bit displaced in space; the HIRLAM model at 2.5-km simulated excessive amounts of rainfall (strong overprediction bias) for large areas, i.e., the precise location was not too accurate; whereas the HARMONIE suite at 2.5-km forecasted the precipitation in the right place at the right time with reasonable amount. It was therefore found that the performance of HARMONIE was better for the cases studied here as shown by the neighborhood verification metrics. Weusthoff et al. (2010) also evaluated by means of 'fuzzy' techniques that high-resolution models give

575 better results for more localized convective precipitation events, and Van der Plas et al.
576 (2012) found that higher precipitation intensities were better represented in the
577 HARMONIE model.

578 The completely missed or strongly underestimated convection of the HIRLAM
579 versions and the experimental HIRLAM model at 5-km forecast are hypothetically
580 caused by the coarser resolution of these HIRLAM versions, combined with the less
581 developed physics in the operational AEMET runs of that time. Differences in
582 predicting isolated convection found between the coarser HIRLAM versions and the
583 highest resolution HIRLAM and HARMONIE set-ups can be linked to the impact of
584 model grid resolution and therefore the ability to skillfully resolve sea breezes and local
585 winds (Ries and Schlünzen, 2009), in combination with the absence of a deep
586 convection parameterization in these cases. Different horizontal grid spacing schemes
587 result in a different placement and strength of the sea breeze front since the temperature
588 gradient is only a few kilometers wide (Atkins and Wakimoto, 1997). In theory, finer
589 resolution allows much stronger horizontal gradients and updrafts at the sea breeze
590 front and a quicker development of convective showers, but sometimes even a 2.5-km
591 grid spacing is too coarse to simulate convection associated with sea breezes. Another
592 effect of the higher resolution is that orography gradients become steeper, having a
593 direct impact on resolving the development of upslope winds. Furthermore, we found
594 that the overestimation bias for the HIRLAM forecasts can be partly explained by the
595 fact that the precipitation fall speed is infinity in the HIRLAM model suite, i.e.,
596 precipitation is put on the ground directly, which is quite normal in hydrostatic models.

597 To conclude, the HARMONIE model showed more realistic forecasts of
598 precipitation associated with sea breeze convergence than those given by the HIRLAM

599 model, since it is better able to provide good estimates of the precipitation
600 accumulations and location of isolated convective events. This overall result confirms
601 the potential capability and skill of the forthcoming HARMONIE model as operational
602 for most European Weather Services to capture isolated showers and thunderstorms
603 associated with mesoscale / local winds in areas of complex terrain. This advance in
604 quantitative precipitation forecasting has practical applications for very short range
605 forecasting of local dangerous convective phenomena, which cause significant
606 economic damage and flooding in the Mediterranean area of the Iberian Peninsula
607 (Sánchez et al., 2003).

608

609 **Acknowledgements**

610 The authors wish to acknowledge the editor and two anonymous reviewers for
611 their detailed and helpful comments to the original manuscript. This research was
612 undertaken in the frame of the BEST/2010/014 (GV), the JAE-DOC043 (CSIC;
613 European Social Fund, FSE) and the JCI-2011-10263 grants. This research was
614 supported by the projects CGL2011-27574-C02-02 and CGL2011-27536/HID,
615 financed by the Spanish Commission of Science and Technology, and FEDER;
616 ACQWA (FP7-ENV- 2008-1-212250), financed by the European Comission; "Efecto
617 de los escenarios de cambio climático sobre la hidrología superficial y la gestión de
618 embalses del Pirineo Aragonés", financed by "Obra Social La Caixa"; and CTTP1/12
619 "Creación de un modelo de alta resolución espacial para cuantificar la esquiabilidad y
620 la afluencia turística en el Pirineo bajo distintos escenarios de cambio climático",
621 financed by the "Comunidad de Trabajo de los Pirineos". The authors would like to
622 thank the AEMET, CEAM, CHE, CHJ, IVIA, METEOCLIMATIC, SIAM, SIAR and

SMC for raingauge data; and to José-Antonio García-Moya and Estrella Gutiérrez-Marco (AEMET) for providing the HIRLAM 6.1.2 outputs.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J., 2003. *Guidelines on Climate Metadata and Homogenization*. World Meteorol. Organ., 52 pp. [Available online at http://www.wmo.int/datastat/documents/WCDMP-53_1.pdf] (last accessed 1 November 2013)

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. Non. Proc. Geophy. **8,** 401-417.

Atkins, N.T., Wakimoto, R.M., Weckwerth, T.M., 1995. Observations of the sea-breeze front during CaPE. Part II: Dual-Doppler and aircraft analysis. Mon. Wea. Rev. **123**, 944-969.

Atkins, N.T., Wakimoto, R.M., 1997. Influences of the synoptic-scale flow on sea breezes observed during CaPE. Mon. Wea. Rev. **125**, 2112-2130.

Azorin-Molina, C., Martin-Vide, J., 2007. Methodological approach to the study of the daily persistence of the sea breeze in Alicante (Spain). Atmosfera **20**, 57-81.

Azorin-Molina, C., Guijarro, J.A., Baena-Calatrava, R., Jansà, A., 2007. Sea breeze convergence and convective cloud frequencies from AVHRR data over Mallorca Island. 7th EMS Annual Meeting/8th European Conference On Applications of

644 Meteorology, El Escorial, Madrid, Spain, EMS Annual Meeting Abstracts **4**,
645 EMS2007-A-00541

646 Azorin-Molina, C., Connell, B.H., Baena-Calatrava, R., 2009. Sea-breeze convergence
647 zones from AVHRR over the Iberian Mediterranean area and the Isle of Mallorca,
648 Spain. J. Appl. Meteor. Climatol. **48**, 2069-2085.

649 Baker, R.D., Lynn, B.H., Boone, A., Tao, W.-K., Simpson, J., 2001. The influence of
650 soil moisture, coastline curvature, and land-breeze circulations on sea-breeze
651 initiated precipitation. J. Hydromet. **2**, 193–211.

652 Bechtold, P., Pinty, J.-P., Mascart, P., 1991. A numerical investigation of the influence
653 of large-scale winds on sea breeze and land breeze type circulations. J. Appl.
654 Meteor. **30**, 1268-1279.

655 Blanchard D. O., López, R. E., 1985. Spatial patterns of convection in south Florida.
656 Mon. Wea. Rev. **113**, 1282–1299.

657 Calvo, J., 2007. Kain-Fritsch convection in HIRLAM. Present status and prospects.
658 Hirlam Newsletter **52**, 57-64.

659 Dailey, P. S., Fovell, R.G., 1999. Numerical simulation of the interaction between the
660 sea-breeze front and horizontal convective rolls. Part I: Offshore ambient flow.
661 Mon. Wea. Rev. **127**, 858-878.

662 de Brujin, E. I. F., de Rooy, W. C., 2012. Evaluation of HARMONIE in the KNMI
663 parameterisation testbed. Adv. Sci. Res. **8**, 167-170.

664 Ebert, E.E., 2008. Fuzzy verification of high-resolution gridded forecasts: a review and
665 proposed framework. Meteorol. Appl. **15**, 51-64.

666 Fovell, R.G., Dailey, P. S., 2001. Numerical simulation of the interaction between the
667 sea-breeze front and horizontal convective rolls. Part II: Alongshore ambient flow.
668 Mon. Wea. Rev. **129**, 2057-2072.

669 Fovell, R.G., 2005. Convective initiation ahead of the sea-breeze front. Mon. Wea. Rev.
670 **133**, 264-278.

671 Germann, U., Zawadzki, I., 2004. Scale dependence of the predictability of precipitation
672 from continental radar images. Part II: probability forecasts. J. Appl. Meteorol. **43**,
673 74-89.

674 Gollvik, S., 2002. A snow model intended for HIRLAM. Proceedings of the
675 SRNWP/HIRLAM Workshop on Surface Processes, Turbulence and Mountain
676 Effects. Spanish Meteorological Agency, Madrid (Spain).

677 Gollvik, S., 2004. Surface modelling in Northern Europe. HIRLAM Baltic Workshop
678 Report, pp. 52-60.

679 González-Márquez, J., Heredia, M. A., 2001. Sea breeze convection in Mallorca.
680 Proceedings, V National Symposium of Forecasting, Madrid, Spain, Spanish
681 Meteorological Agency, pp. 1-6. (in Spanish)

682 Huang, X.-Y., Yang, X., Gustafsson, N., Mogensen, K., Lindskog, M., 2002. Four-
683 dimensional variational data assimilation for a limited area. HIRLAM Technical
684 Report **57**, 44 pp.

685 Ivarsson, K.I., 2007. The Rasch-Kristjansson large scale condensation; present status
686 and prospects. Hirlam Newsletter **52**, 50-56.

687 Kain, J.S., 2004. The Kain-Fritsch Convective Parameterization: An Update. J. Appl.
688 Meteor. **43**,170-181.

689 Kourzeneva, E., Samuelsson, P., Ganbat, G., Mironov, D., 2008. Implementation of
690 lake model FLake in HIRLAM. HIRLAM Newsletter **54**, 54-61.

691 Mazarakis, N., Kotroni, V., Lagouvardos, K., Argiriou, A.A., 2009. The sensitivity of
692 numerical forecasts to convective parameterization during the warm period and the
693 use of lightning data as an indicator for convective occurrence. Atmos. Res. **94**,
694 704–714.

695 Millán, M. M., Estrela, M. J., Miró, J., 2005. Rainfall Components: Variability and
696 spatial distribution in a Mediterranean area (Valencia region). J. Climate **18**, 2682-
697 2705.

698 Mironov, D. V., 2008. Parameterization of lakes in numerical weather prediction.
699 Description of a lake model. COSMO Technical Report, 11, Deutscher
700 Wetterdienst, Offenbach am Main, Germany, 41 pp.

701 Muppa, S.K., Anandan, V.K., Amit Kesarkar, K., Vijaya Bhaskara Rao, S., Narasimha
702 Reddy, P. 2012 Study on deep inland penetration of sea breeze over complex terrain
703 in the tropics. Atmos. Res. **104-105**, 209-216

704 Neumann, J., 1951. Land breezes and nocturnal thunderstorms. J. Meteor. **8**, 60-67.

Nicholls, M.E., Pielke, R.A., Cotton, W.R., 1991. A two-dimensional numerical investigation of the interaction between sea breezes and deep convection over the Florida Peninsula. Mon. Wea. Rev. **119**, 298-323.

Noilhan, J., Planton, S., 1989. A simple parameterization of land surface processes for meteorological models. Mon. Wea. Rev. **117**, 536-549.

Papanastasiou, D.K., Melas, D., Lissaridis, I., 2010. Study of wind field under sea breeze conditions; an application of WRF model. Atmos. Res. **98(1)**, 102-117.

Pérez-Landa, G., Ciais, P., Sanz, M.J., Gioli, B., Miglietta, F., Palau, J.L., Gangoiti, G., Millán, M.M., 2006. Mesoscale circulations over complex terrain in the Valencia coastal region, Spain, Part 1: simulation of diurnal circulation regimes. Atmos. Chem. Phys. Discuss. **6**, 2809-2852.

Petterssen, S., 1956. Weather Analysis and Forecasting, ed. McGraw-Hill Book Company, New York.

Pielke, R.A., Song, A., Michaels, P.J., Lyons, W.A., Arritt, R.W., 1971. The predictability of sea-breeze generated thunderstorms. Atmosfera **4**, 65-78.

Pielke, R.A., 1974. A three-dimensional numerical model of the sea breezes over South Florida. Mon. Wea. Rev. **102**, 115-139.

Pielke, R.A., Cotton, W.R., 1977. A mesoscale analysis over south Florida for a high rainfall event. Mon. Wea. Rev. **105**, 343-362.

724 Pielke, T.A., Mahrer, Y., 1978. Verification analysis of the University of Virginia three-
725 dimensional mesoscale model prediction over south Florida for July 1, 1973. Mon.
726 Wea. Rev. **106**, 1568-1589.

727 Purdom, J. F. W., 1976. Some uses of high-resolution GOES imagery in the mesoscale
728 forecasting of convection and its behavior. Mon. Wea. Rev. **104**, 1474-1483.

729 Ries, H., Heinke Schlünzen, K., 2009. Evaluation of a mesoscale model with different
730 surface parameterizations and vertical resolutions for the Bay of Valencia. Mon.
731 Wea. Rev. **137**, 2646-2661.

732 Rontu, L., Sattler, K., Sigg, R., 2002. Parametrization of subgrid-scale orography effects
733 in HIRLAM. Technical Report **56**, 46 pp.

734 Sánchez, J.L., Fernández, M.V., Fernández, J.T., Tuduri, E., Ramis, C., 2003. Analysis
735 of mesoscale convective systems with hail precipitation. Atmos. Res. **67-68**, 573-
736 588.

737 Savijärvi, H., 1990. Fast radiation parameterization schemes for mesoscale and short-
738 range forecast models. J. Appl. Meteor. **29**, 437-447.

739 Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C.,
740 Masson, V., 2011. The AROME-France Convective-Scale Operational Model. Mon.
741 Wea. Rev. **139**, 976–991.

742 Simpson, J.E., 1994. Sea breeze and local wind, ed. Cambridge University Press, New
743 York.

744 Simpson, M., Warrior, H., Raman, S., Aswathanarayana, P.A., Mohanty, U.C., Suresh,
745 R., 2007. Sea-breeze-initiated rainfall over the east coast of India during the Indian
746 southwest monsoon. Nat. Hazards **42**, 401-413.

747 Strickler, M. W., 2003. Sea-Breeze Fronts and Their Role in Convective Initiation.
748 North Carolina State University, Raleigh, North Carolina, 22 pp.

749 Tartaglione, N., Mariani, S., Accadia, C., Speranza, A., Casaioli, M., 2005. Comparison
750 of rain gauge observations with modeled precipitation over Cyprus using
751 Contiguous Rain Area analysis. Atmos. Chem. Phys. **5**, 2147-2154.

752 Theis, S.E., Hense, A., Damrath, U., 2005. Probabilistic precipitation forecasts from a
753 deterministic model: a pragmatic approach. Meteorol. Appl. **12**, 257-268.

754 Tijm, A.B.C, Lenderink, G., 2003. Characteristics of CBR and STRACO versions.
755 HIRLAM Newsletter **43**, 115-124.

756 Undén, P., Rontu, L., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-
757 Moya, J.A., Gustafsson, N., Jones, C., Järvenoja, S., Järvinen, H., Lynch, P.,
758 McDonald, A., McGrath, R., Navascues, B., Odegaard, V., Rodriguez, E.,
759 Rummukainen, M., Room, R., Sattler, K., Savijärvi, H., Sass, B.H., Schreur, B.W.,
760 The, H., Tijm, S., 2002. The HIRLAM-5 scientic documentation. HIRLAM-5
761 project. SMHI, S-60176, Norrköping, Sweden, 144 pp.

762 Van der Plas, E. V., Wichers Schreur, B., Kok, K., 2012. A quantitative evaluation of
763 the high resolution HARMONIE model for critical weather phenomena. Adv. Sci.
764 Res. **8**, 149-155.

765 Wakimoto, R. M., Atkins, N.T., 1994. Observations of the sea breeze front during
766 CAPE. Part I: Single Doppler, satellite, and cloud photogrammetry analysis. Mon.
767 Wea. Rev. **122**, 1092-1114.

768 Weusthoff, T., Ament, F., Arpagaus, M., Rotach, M.W., 2010. Assessing the Benefits of
769 Convection-Permitting Models by Neighborhood Verification: Examples from MAP
770 D-PHASE. Mon. Wea. Rev. **138**, 3418–3433.

771 Weusthoff, T., 2011. Neighbourhood verification in space and time. 5th Intl. Verification
772 Methods Workshop, 1-6 Dec. 2011, Melbourne, Australia.

773 Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences. 3rd Edition.
774 Elsevier Inc.

775 Wilson, P.H., 2008. Warm-season lake-/sea-breeze severe weather in the Northeast.
776 Master thesis, Department of Earth and Atmospheric Sciences, University at
777 Albany, State University of New York, 115 pp.

778 Zhang, M., Lin, W., Bretherton, C.S., Hack, J.J., Rasch, P.J., 2003. A modified
779 formulation of fractional stratiform condensation rate in the NCAR community
780 atmospheric model (CAM2). J. Geophys. Res. – Atmos. **108**, 4035,
781 doi:10.1029/2002JD002523.

782

783 **List of Tables**

784

785 Table 1. Sea breeze thunderstorms selected for verifying the precipitation performance
786 of the HIRLAM and the HARMONIE models suites. The number of raingauge points,
787 the maximum of precipitation and location of this event are tabulated for each episode.

788

Date	Nº Raingauges	MaximumPrecip. (mm)	Station
31/08/2001	450	57.5	Villena (A)
19/04/2002	380	22.2	Beneixama (A)
29/09/2002	420	31.0	Fontanars (V)
18/06/2003	423	13.9	Mosqueruela (T)
25/08/2003	424	3.1	Alcora (C)
02/06/2004	361	12.0	Talave (AB)
20/06/2004	363	90.0	Pina (C)
21/06/2004	361	36.0	Fredes (C)
12/08/2005	363	1.7	Ibi (A)
28/09/2006	356	25.0	Pinoso (A)
08/10/2007	350	35.0	Ibi (A)
01/08/2008	323	34.0	Villafranca (C)
07/08/2008	533	45.0	Cati (C)
08/08/2008	325	53.7	Relleu (A)
11/09/2009	304	43.0	Cati (C)

789

790 (A) Alicante province, (AB) Albacete province, (C) Castellon province, (T) Teruel province, and (V)
791 Valencia province. The (C) and (T) regions correspond to areas affected by the SBCZ2 (Iberian system
792 mountains) and the (A) and (V) regions by the SBCZ3 (Prebetic mountain ranges)

793

794

795

796 Table 2. Optimal scale (in km) and the HK score found for all three NWP models as a

797 function of the rain intensity thresholds (in mm).

Model suite		Rain threshold (in mm)						
		0.2	1.0	2.0	5.0	10.0	20.0	50.0*
HIR-D2.5	Scale	2.5	2.5	17.5	37.5	37.5	37.5	17.5
	HK	0.34	0.38	0.40	0.52	0.57	0.54	0.32
HIR-D5	Scale	5.0	25.0	55.0	65.0	65.0	85.0	---
	HK	0.31	0.40	0.45	0.49	0.35	0.16	0
HAR-D2.5	Scale	7.5	22.5	22.5	27.5	22.5	27.5	37.5
	HK	0.37	0.45	0.45	0.53	0.59	0.59	0.63

798 *There were only 3 observations of 50 mm or more.

799

800

801

802

803

804

805

806

807

808

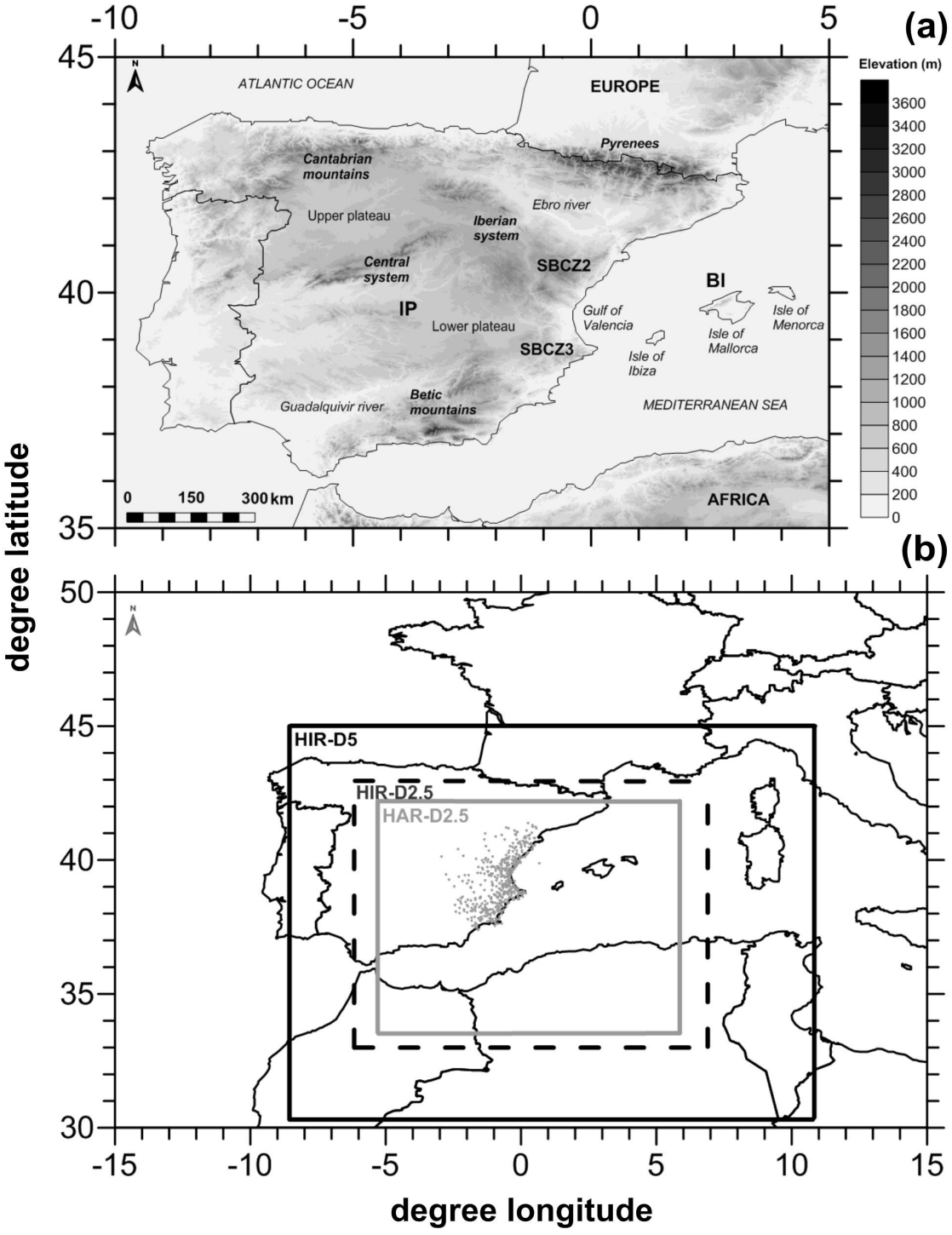
809

Table 3. As in Table 2 but for the BSS scores.

Model suite		Rain threshold (in mm)						
		0.2	1.0	2.0	5.0	10.0	20.0	50.0*
HIR-D2.5	Scale	97.5	97.5	97.5	97.5	97.5	92.5	97.5
	BSS	-5.11	-2.52	-1.69	-1.16	-0.59	-0.16	-0.03
HIR-D5	Scale	95	95	95	95	95	95	65
	BSS	-3.55	-1.40	-0.83	-0.34	-0.10	-0.01	0
HAR-D2.5	Scale	97.5	97.5	97.5	97.5	92.5	87.5	97.5
	BSS	-0.48	-0.18	-0.12	-0.15	-0.10	-0.02	-0.03

*There were only 3 observations of 50 mm or more.

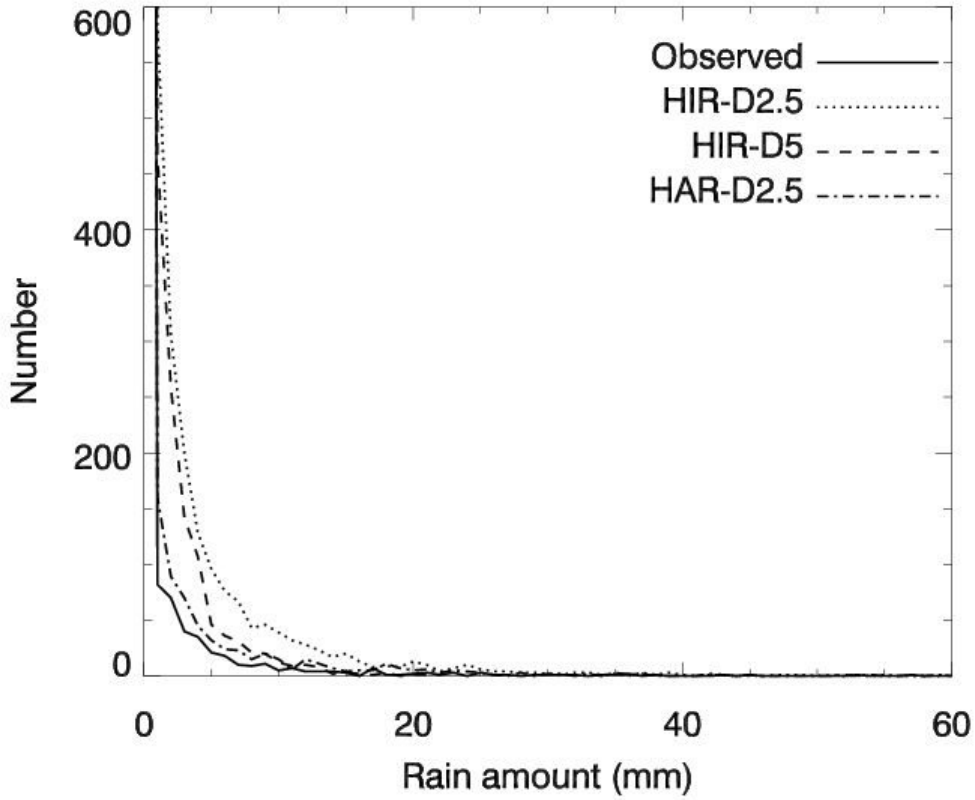
824 **List of Figures**



825

826 FIG. 1. (a) Terrain map of the Iberian Peninsula showing the complex topography of the
827 study area focussed on the steep Iberian System mountains (SBCZ2) and the Prebetic
828 mountains (SBCZ3). Specific locations cited to in the text and the most important

geographical features of the IP are shown on the map. (b) HIRLAM model configuration showing a largest (HIR-D5; black solid line) and smallest (HIR-D2.5; black dashed line) domain with horizontal grid lengths of 5.0-km and 2.5-km, respectively, and also displaying the HARMONIE model configuration with a domain of 2.5-km grid spacing (HAR-D2.5; grey solid line). The map shows location (grey dots) of the high-density rain-gauge network for verification model output.



848

849 FIG. 2. Rain distribution from gridpoint match-ups using maximum observation in grid

850 box, pooled over all 15 non-forecasted sea breeze thunderstorms in the dataset.

851

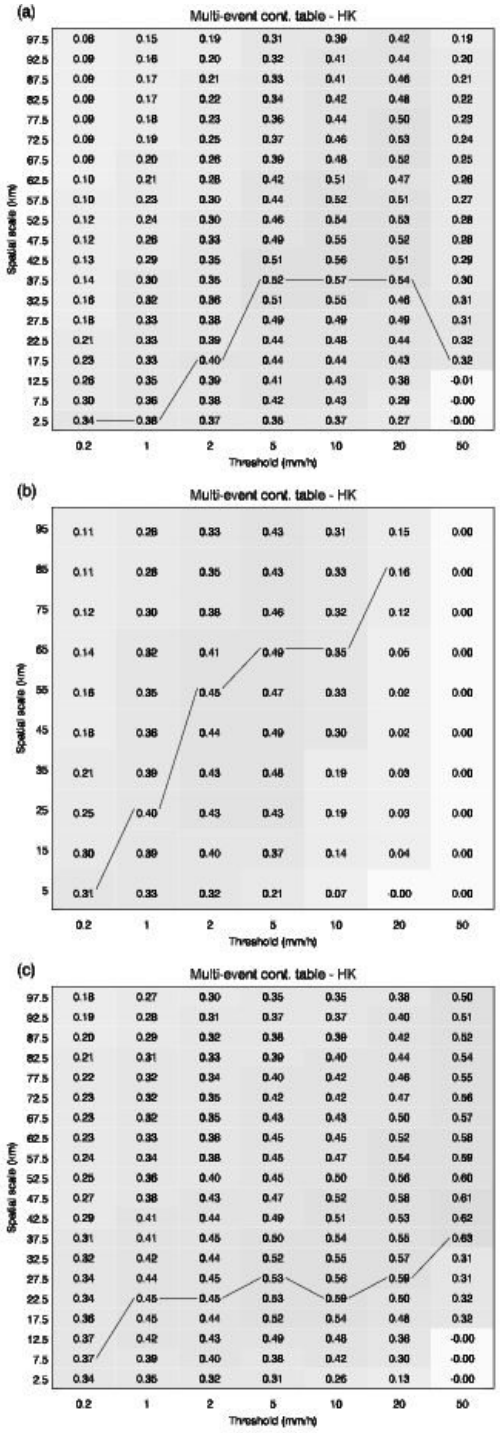
852

853

854

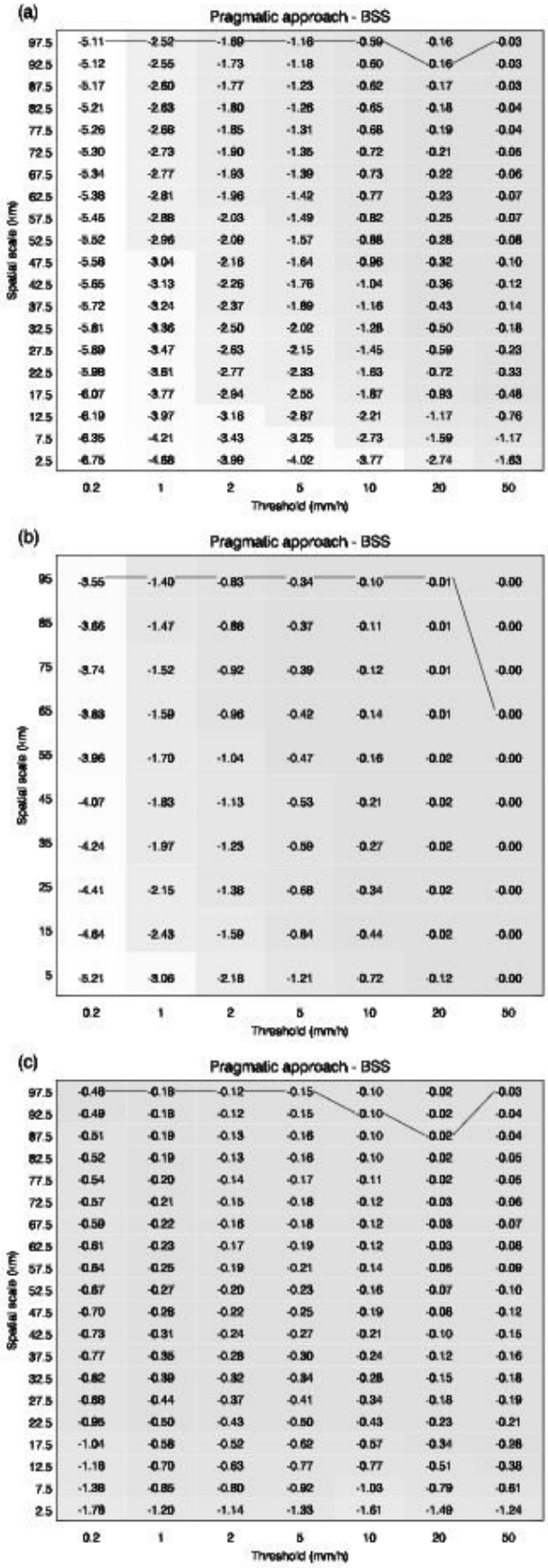
855

856



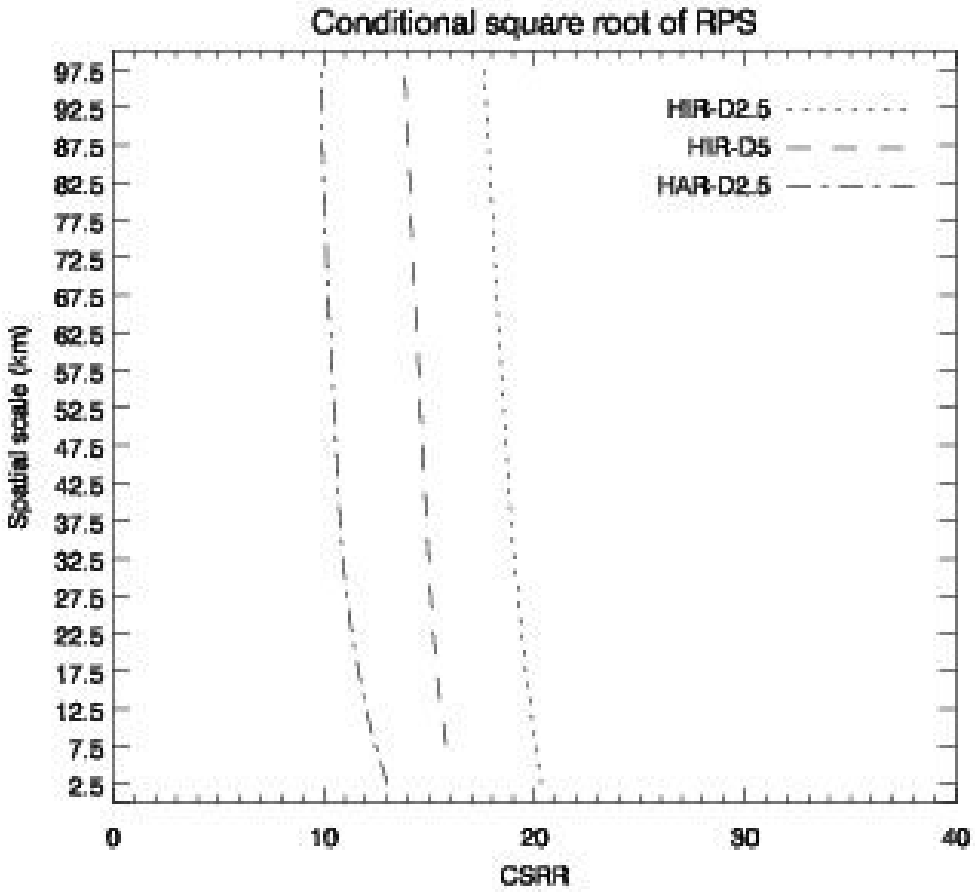
857

858 FIG. 3. Quilt plots of the HK scores as a function of different rain intensity threshold
859 (0.2, 1.0, 2.0, 5.0, 10.0, 20.0 and 50.0 mm) and neighborhood sizes (from 2.5 up to
860 97.5-km) for (a) the HIR-D2.5, (b) the HIR-D5 and (c) the HAR-D2.5. The darker and
861 lighter shades show good or poor model performance, respectively. The lines indicate
862 the better scale at each rain threshold.



863

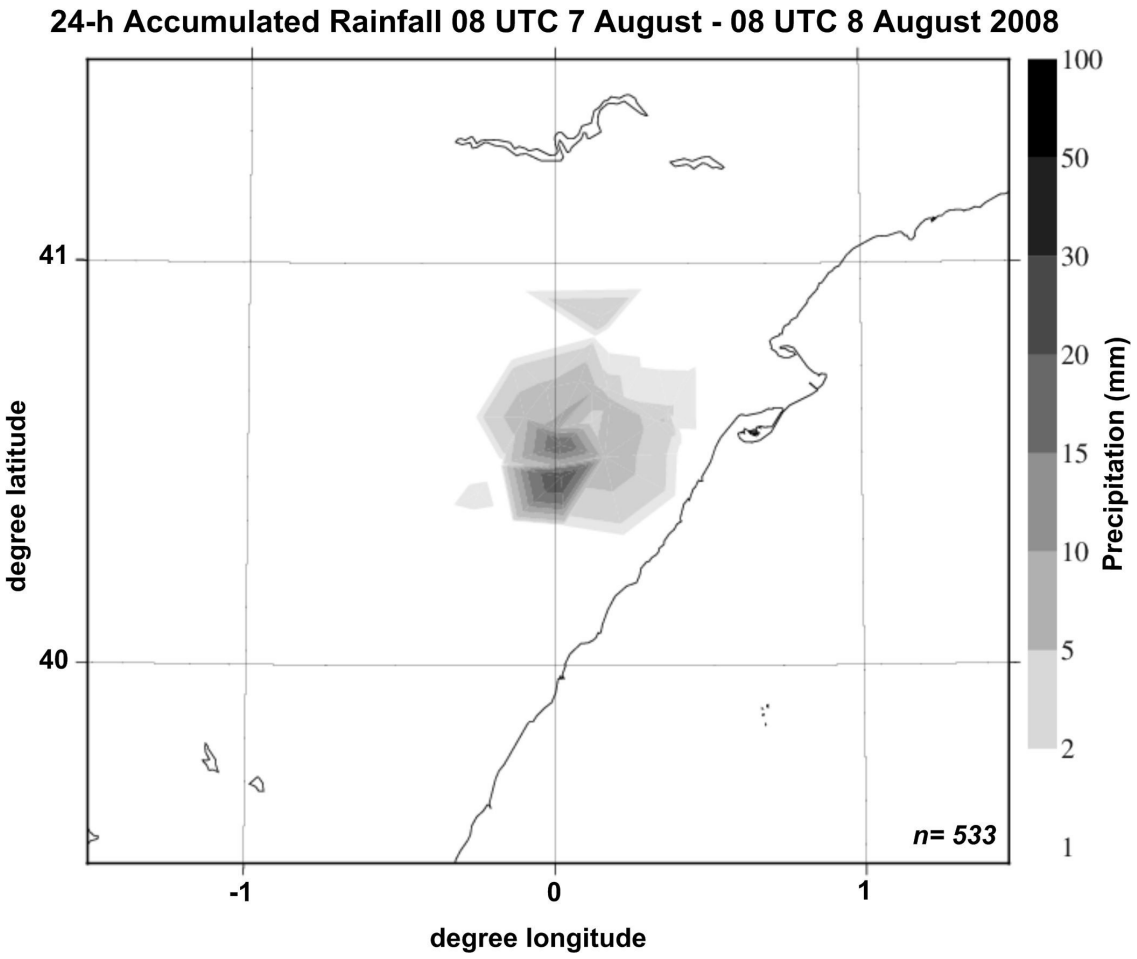
864 FIG. 4. As in Figure 3 but for the BSS computed for the pragmatic method.



865

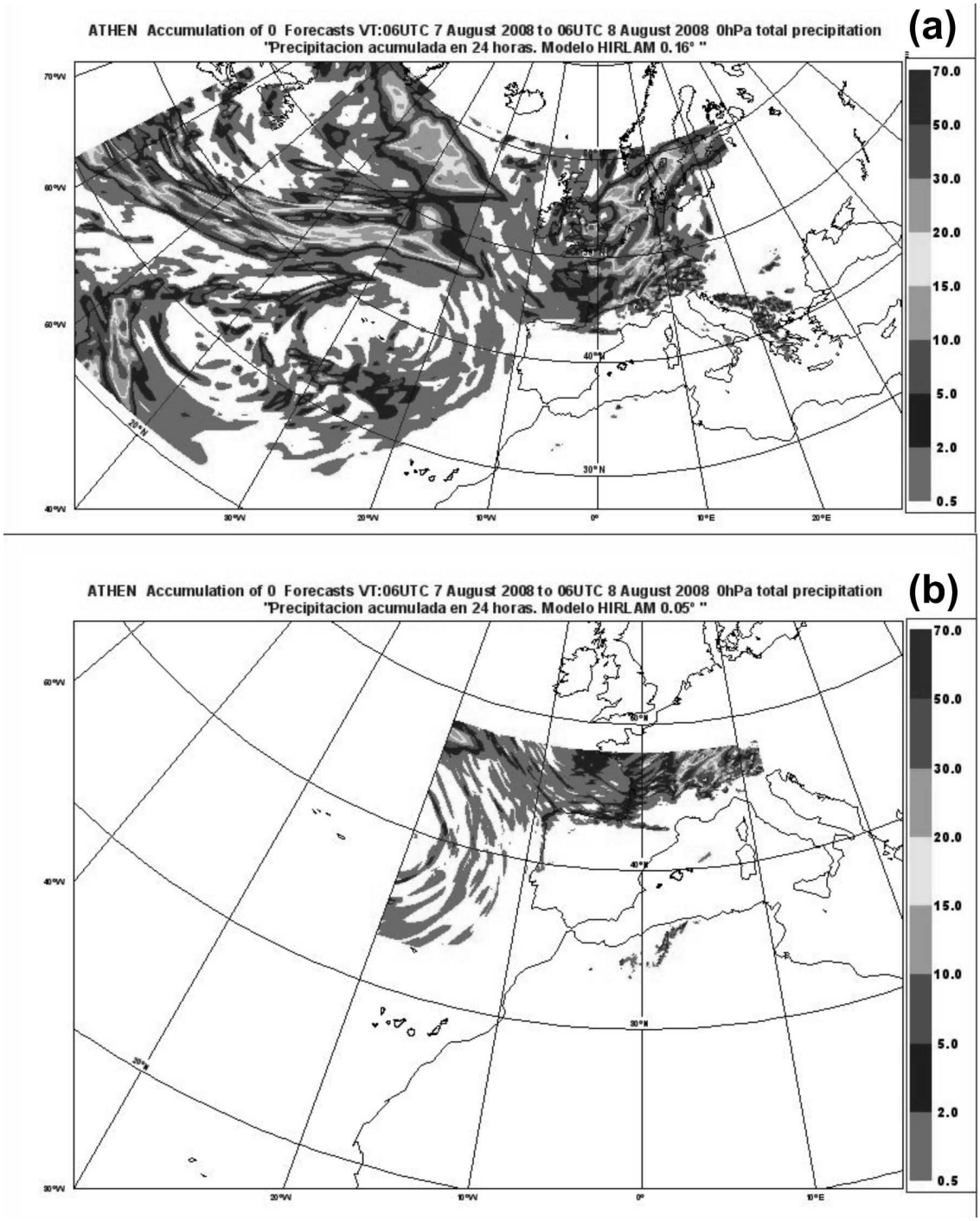
866 FIG. 5. Conditional square root of the RPS showing the CSRR metric as a function of

867 different window sizes (from 2.5 up to 97.5-km).



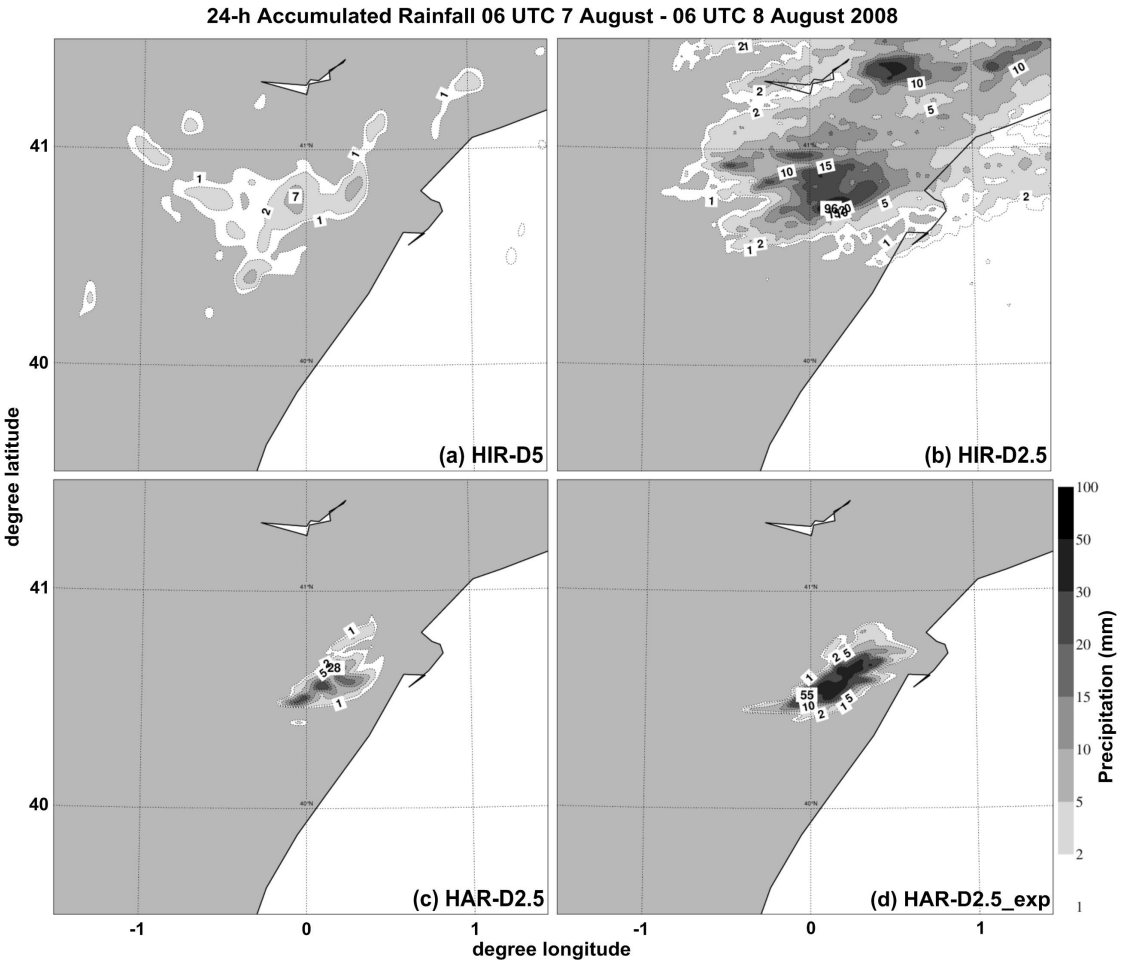
868

869 FIG. 6. Accumulated rainfall recorded between 08 UTC 7 August 2008 till 08 UTC 8
870 August 2008. The high spatial density of measurements (*n*= sample size) was obtained
871 by combining data from Hellmann (SYNOPS) and tipping bucket rain-gauges
872 (automatic weather station; AWS). The map was produced using a Kriging
873 interpolation.



874

875 FIG. 7. 24-h accumulated rainfall between 06 UTC 7 August 2008 till 06 UTC 8 August
876 2008 simulated from (a) the HIR-D16 (0.16º, i.e., 16-km length) and (b) the HIR-D5
877 (0.05º, i.e., 5-km length) operational HIRLAM 6.1.2 version from the AEMET.



878

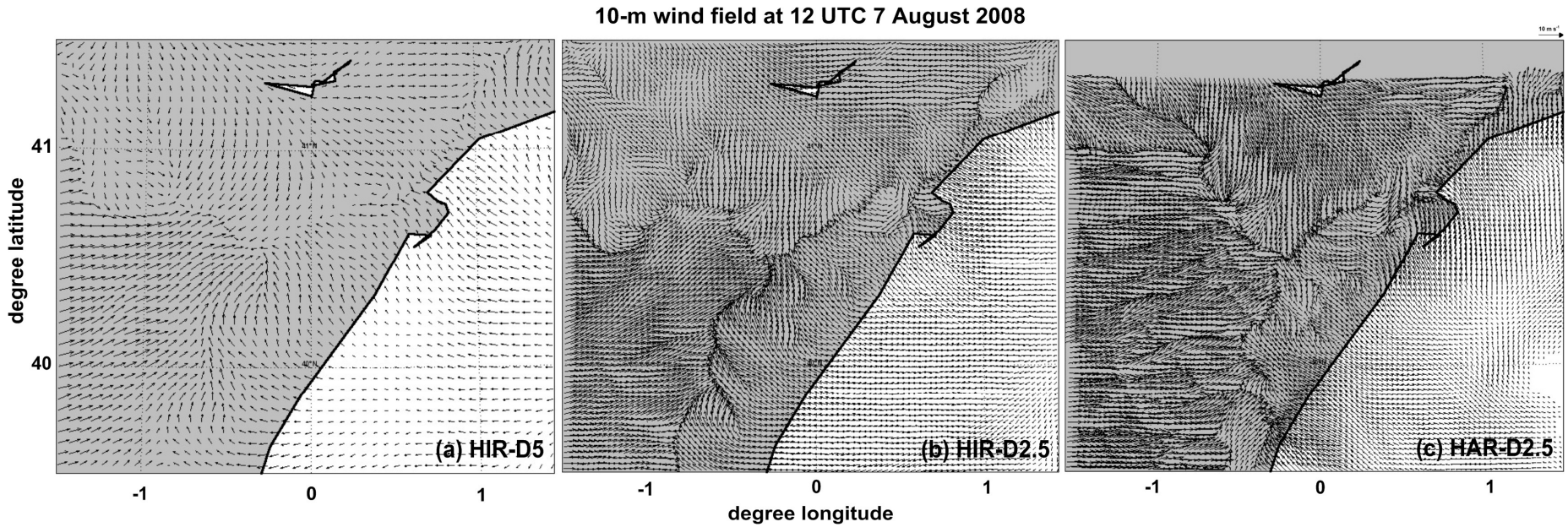
879 FIG. 8. 24-h accumulated rainfall between 06 UTC 7 August 2008 till 06 UTC 8 August
880 2008 simulated from (a) the HIR-D5, (b) the HIR-D2.5, (c) the HAR-D2.5, and (d) the
881 HAR-D2.5_exp with an increase in fall speed of precipitation by a factor of 5. Note that
882 forecasted maps are plotted using the same boundaries and colors for comparison
883 against the observed rainfall map shown in Figure 6.

884

885

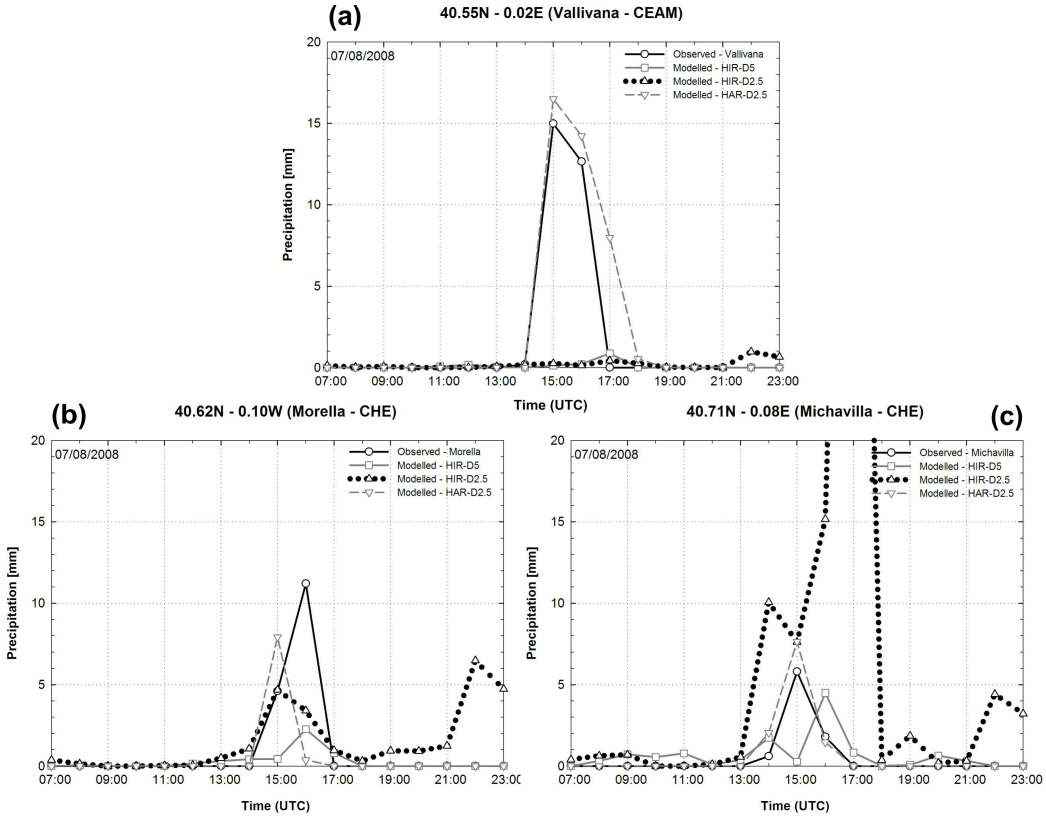
886

887



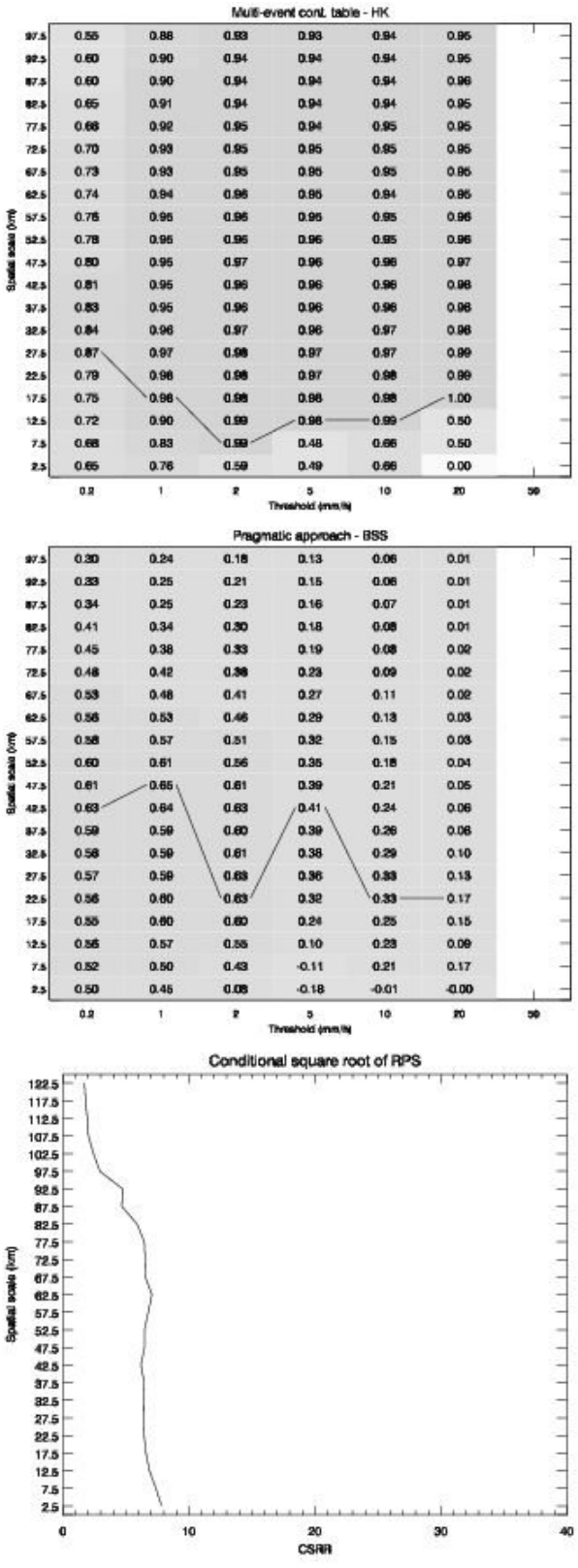
888

889 FIG. 9. 10-m wind vector simulated by (a) the HIR-D5, (b) the HIR-D2.5, and (c) the HAR-D2.5 at 12 UTC on 7 August 2008.



890

891 FIG. 10. Time series of observed and modelled precipitation in (a) Vallivana-CEAM,
892 (b) Morella-CHE and (c) Michavilla-CHE from 0700 till 2300 UTC on 7 August 2008.
893 The modelled values shown here correspond to the maximum precipitation in area with
894 radius of 5-km around raingauge station. Note that modelled precipitation for HIR-D2.5
895 exceeds the maximum y-axis for the Michavilla-CHE station.



896

897 FIG. 11. Neighborhood verification of the forecast performed by the HAR-D2.5 on 7

898 August 2008 for (a) the HK scores and (b) the BSS metrics as a function of different

899 rain intensity threshold (0.2, 1.0, 2.0, 5.0, 10.0, 20.0 and 50.0 mm) and neighborhood
900 sizes (from 2.5 up to 97.5-km), and (c) the CSRR scores as a function of different
901 window sizes (from 2.5 up to 122.5-km). The quilt plots of the HK and the BSS metris
902 show in darker and lighter shades good or poor model performance, respectively. The
903 lines indicate the better scale at each rain threshold.

904

905

906

907

908

909

910