

# Multiscale Performance of the ALARO-0 Model for Simulating Extreme Summer Precipitation Climatology in Belgium

ROZEMIEN DE TROCH

*Royal Meteorological Institute, Brussels, and Department of Physics and Astronomy, Ghent University, Ghent, Belgium*

RAFIQ HAMDANI AND HANS VAN DE VYVER

*Royal Meteorological Institute, Brussels, Belgium*

JEAN-FRANÇOIS GELEYN

*Czech Hydrometeorological Institute, Prague, Czech Republic, and Department of Physics and Astronomy, Ghent University, Ghent, Belgium*

PIET TERMONIA

*Royal Meteorological Institute, Brussels, and Department of Physics and Astronomy, Ghent University, Ghent, Belgium*

(Manuscript received 30 November 2012, in final form 8 May 2013)

## ABSTRACT

Daily summer precipitation over Belgium from the Aire Limitée Adaptation Dynamique Développement International (ALADIN) model and a version of the model that has been updated with physical parameterizations, the so-called ALARO-0 model [ALADIN and AROME (Application de la Recherche à l'Opérationnel à Meso-Echelle) combined model, first baseline version released in 1998], are compared with respect to station observations for the period 1961–90. The 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40) is dynamically downscaled using both models on a horizontal resolution of 40 km, followed by a one-way nesting on high spatial resolutions of 10 and 4 km. This setup allows us to explore the relative importance of spatial resolution versus parameterization formulation on the model skill to correctly simulate extreme daily precipitation. Model performances are assessed through standard statistical errors and density, frequency, and quantile distributions as well as extreme value analysis, using the peak-over-threshold method and generalized Pareto distribution. The 40-km simulations of ALADIN and ALARO-0 show similar results, both reproducing the observations reasonably well. For the high-resolution simulations, ALARO-0 at both 10 and 4 km is in better agreement with the observations than ALADIN. The ALADIN model consistently produces too high precipitation rates. The findings demonstrate that the new parameterizations within the ALARO-0 model are responsible for a correct simulation of extreme summer precipitation at various horizontal resolutions. Moreover, this study shows that ALARO-0 is a good candidate model for regional climate modeling.

## 1. Introduction

Extreme precipitation events have a large impact on societies through damage caused by floods, landslides, and snow events. Precipitation is thus an important meteorological variable in weather prediction and climate studies. Herrera et al. (2010) studied the ability of regional climate models (RCMs) to reproduce the mean and extreme

precipitation regimes over Spain using a state-of-the-art ensemble of RCM simulations. The RCMs show good agreement with the observed mean precipitation regime, but for the extreme regimes the models reveal important limitations.

As described in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), the model skill to simulate realistic extreme daily precipitation strongly depends on the spatial resolution and convective parameterization of the model (Randall et al. 2007). However, it is not straightforward to quantify the relative contribution of an increase in spatial resolution

---

*Corresponding author address:* Rozemien De Troch, Royal Meteorological Institute, Ringlaan 3, B-1180 Brussels, Belgium.  
E-mail: rozemien.detroch@meteo.be

versus an improvement in physical parameterization of deep convection on the overall performance of the model.

On the other hand, precipitation is one of the most sensitive quantities in the different parameterization schemes of the climate models and to their interplay with the dynamics of the atmosphere represented in the models. For this variable it has been shown that RCMs are able to add significant information to the driving global simulations, both in space and time (e.g., Jones et al. 1995; Durman et al. 2001; Jones et al. 2004). In general terms, the RCMs produce an intensification of precipitation with respect to the driving global climate model (GCM), related to the intensification of the hydrological cycle (Jones et al. 1995; Durman et al. 2001; Buonomo et al. 2007). Lynn et al. (2010) tested a regional climate model with different physics components at two different spatial resolutions. Their results demonstrated a sensitivity of the RCM to the choice of the convective parameterization, leading to significantly different summer precipitation outcomes. The authors conclude that these differences are due to differences in the convective parameterizations and not because of the change in spatial resolution of the model.

The aim of the present paper is to elaborate on the relative importance of resolution versus parameterization formulation on the model skill to simulate realistic extreme daily precipitation. This is achieved by comparing at varying horizontal resolutions the Aire Limitée Adaptation Dynamique Développement International (ALADIN) model with a version of the model that has been updated with physical parameterizations, the so-called ALARO-0 model [ALADIN and AROME (Application de la Recherche à l'Opérationnel à Meso-Echelle) combined model, first baseline version released in 1998]. The ALADIN model is the limited area model (LAM) version of the Action de Recherche Petite Echelle Grande Echelle Integrated Forecast System (ARPEGE-IFS) (Bubnová et al. 1995; ALADIN International Team 1997). Since the 1990s the model has been widely used by the numerical weather prediction (NWP) community and, more recently, in regional climate modeling (e.g., Radu et al. 2008; Skalák et al. 2008). Furthermore, the model uses a diagnostic-type deep convection and microphysics parameterization based on Bougeault (1985) with upgrades from Gerard and Geleyn (2005). The new physical parameterizations within the ALARO-0 model, as proposed by Gerard et al. (2009), were specifically designed to be used from mesoscale to the convection-permitting scales (so-called gray-zone scales) and are centered around an improved convection and cloud scheme. For this study we use the version of the ALARO-0 model that was adopted for the operational applications in the Royal Meteorological Institute (RMI)

of Belgium in 2010. Since then this model has undergone systematic verification with respect to observations at 7-km resolution. Gerard et al. (2009) tested the new parameterizations within the ALARO-0 model in a 1-day case study over Belgium, which was characterized by heavy convective precipitation. From this study an improvement of ALARO-0 at varying horizontal scales has been demonstrated.

Basically, the "nesting" strategy, or climate downscaling technique, in which a LAM or RCM is driven by either a GCM or by analyses of observations, is the most widely used strategy to produce high resolution over a region of interest (Denis et al. 2002). Hence, limiting the geographical domain of these atmospheric models reduces the total number of grid points and allows one to perform simulations at high resolutions with an affordable computational cost. Because of the ability of these high-resolution LAMs or RCMs to reproduce meaningful small-scale features over a limited region (Denis et al. 2002; Giorgi et al. 2004), they have become a popular tool in both the NWP and the climate community for studying extreme events at regional and local scales (e.g., Jones and Reid 2001; Buonomo et al. 2007; Déqué and Somot 2008; Dulière et al. 2011).

However, studies show that RCMs do not necessarily improve their driving GCM simulations or global reanalyses (e.g., Castro et al. 2005; Jacob et al. 2007; Sylla et al. 2010). The use of nested LAMs or RCMs as a climate downscaling technique, indeed, involves a number of issues, one of which is related to the lateral boundary conditions (LBCs) (Giorgi and Mearns 1999; Denis et al. 2002). This drawback of RCMs is related to the fact that one is obliged to impose imperfect LBCs, inducing various errors at the boundaries (e.g., Warner et al. 1997; Termonia et al. 2009). Despite this, past and current applications with RCMs have shown that the one-way nesting strategy is a workable solution (Giorgi and Mearns 1999). To minimize the effects of the LBC problem, Giorgi and Mearns (1999) recommend to first validate the model for the current climate using analyses of observations, that is, the so-called perfect boundary conditions.

Interesting work has been carried out by de Elía et al. (2002) and Denis et al. (2002) with a perfect-model approach, showing that, in a downscaling with a one-way nesting, a LAM or RCM is able to regenerate the correct amount of variability at the scales smaller than the ones of the driving model in which the high-resolution variability had been removed by filtering. However, de Elía et al. (2002) found that the LAM is not capable of reproducing the correct details with sufficient precision required by the rms errors (RMSEs), that is, that the variables locally in space and time do not fully reproduce the ones of the perfect model run. Whereas de Elía et al. concentrated on the short-term evolution of weather

systems and quantified the models' ability to simulate the data in a deterministic day-by-day basis by means of RMSEs, Denis et al. focused on climate time scales and demonstrated the ability of high-resolution RCMs to gain accuracy in a climatic–statistical sense.

Therefore, for studying the climate of weather extremes it is rather the statistics of the extremes that are important, provided the large-scale evolution is consistent with the large-scale flow of the driving model. This is an important additional criterion in deciding to use RCMs with respect to global ones.

For long-range runs at temporal scales of multiple decades, there is also the problem that the internal climate can start to diverge from the climate of the global model (Nicolis 2003; Qian et al. 2003; Nicolis 2004). One can deal with this by either (i) interrupting the model runs of the LAM after a few days and restarting them, while allowing a spinup period so that the physics can adjust, or (ii) carrying out uninterrupted model runs over long periods, allowing the LAM to find its own climate equilibrium (Qian et al. 2003). In the second case, one can for instance apply a spectral nudging of the large scales to the large scale of the driving global model. In the present paper, we will also study whether the internal climate variability generated by the higher resolution of the RCM and its model physics, as identified by Denis et al. (2002) and de Elía et al. (2002), reproduces the correct statistics. For this we want to avoid imposing an upper-air spectral nudging; hence, we will merely carry out a pure downscaling with reinitializations using a one-way nesting approach. Lucas-Picher et al. (2013) demonstrated that dynamical downscaling with reinitializations has lower systematic errors than with a standard continuous model configuration.

The 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005) is used as large-scale coupling data to drive the coupled models, ALARO-0 and ALADIN. As suggested by Giorgi and Mearns (1999), atmospheric reanalyses, such as ERA-40, can be used in climate studies to provide the “perfect boundary conditions” for RCMs (e.g., Csima and Horányi 2008; Déqué and Somot 2008; Skalák et al. 2008; Heikkilä et al. 2011; Hamdi et al. 2012). These reanalyses are produced by means of data assimilation methods in order to find optimal estimates for past atmospheric states that are consistent with meteorological observations and the model dynamics.

In a recent study of Hamdi et al. (2012) the use of high-resolution dynamical downscaling of ALARO-0 at 4-km horizontal resolution is explored by means of the summer maximum surface air temperature over Belgium. Our study extends the work of Hamdi et al. in the sense that, instead of temperature, precipitation is now

analyzed. Daily summer precipitation from different model runs is compared with respect to station observations, with an emphasis on extreme precipitation. This approach by which model output is directly compared against station observations can be motivated by the fact that the station-level observations provide the closest representation of extreme events (Dulière et al. 2011). Furthermore, the motivation for only considering summer precipitation is threefold: (i) other regional climate studies (e.g., Caldwell et al. 2009; Soares et al. 2012a,b) show difficulties of RCMs to simulate summer precipitation; (ii) the new parameterization scheme within ALARO-0 mostly modifies convection, which is the process most relevant for (extreme) precipitation events in summer (Kyselý and Beranová 2009; Soares et al. 2012a); and (iii) the relatively small scale on which these convective processes often occur better corresponds to the high-resolution ALARO-0 simulation (Kyselý and Beranová 2009).

We add to our evaluation the ALADIN-Climate model developed by the Centre National de Recherches Météorologiques (CNRM), which took part in the European ENSEMBLES project ([www.ensembles-eu.org](http://www.ensembles-eu.org)). The ALADIN-Climate model is an ALADIN model version that is specifically used for regional climate modeling. The Ensemble-Based Predictions of Climate Changes and their Impacts (ENSEMBLES) project was finished near the end of 2009 and is aimed to develop an ensemble climate forecast system to produce probabilistic scenarios of future climate so as to provide detailed, quantitative, and policy-relevant information to the European society and economy. Several experiments were performed with some 10 state-of-the-art European and Canadian high-resolution, global, and regional climate models. The ENSEMBLES ALADIN-Climate/CNRM simulations use a long uninterrupted model run, which is a different setup than our ALADIN and ALARO-0 simulations. Hence, a direct comparison with the ALADIN-Climate/CNRM simulation is not possible, and these uninterrupted climate runs are merely added as a reference for regional climate modeling in order to make the present paper complete.

The model simulations, experimental design, and observational data used in this study are described in the next section. Section 3 gives a description of the applied methods, and the results are discussed in section 4. The results are summarized in the conclusions in section 5.

## 2. Model description and data

### a. Experimental design

The experimental design is summarized in Table 1. The ERA-40 reanalysis data (Uppala et al. 2005) are

TABLE 1. Overview of the experimental design.

	Reference	Daily cumulated precipitation	Model	Daily cumulated precipitation
1) Effect of downscaling	Station observations	0800 LT (day) → 0800 LT (day + 1)	ERA-40 ALR40 ALD40	0600 UTC (day) → 0600 UTC (day + 1)
2) Multiscale performance of ALARO-0	Station observations	0800 LT (day) → 0800 LT (day + 1)	ALR40 ALD40 ALR10 ALD10 ALR04	0600 UTC (day) → 0600 UTC (day + 1)
3) Reference for regional climate modeling	Station observations	0800 LT (day) → 0800 LT (day + 1)	CNRM	mean (0000–2400 UTC)

dynamically downscaled using the limited-area models ALADIN and ALARO-0.

The physics parameterization package of the ALARO-0 model has been specifically designed to be run at convection-permitting resolutions. The key concept behind the package lies in the precipitation and cloud scheme called Modular Multiscale Microphysics and Transport (3MT) developed by Gerard and Geleyn (2005), Gerard (2007), and Gerard et al. (2009). With mesh sizes mostly below the Rossby radius of deformation for convective phenomena, the parameterization schemes must take into consideration that the return current from updrafts is happening in a multitude of grid boxes. Therefore, each individual grid-box realization of the parameterization has a statistical view of the “compensating subsidence” happening inside its area. As long as the updraft computation can also be considered as statistical with respect to its population of updrafts of various depths and sizes, it seems not to matter much that the compensating subsidence is computed on the basis of a purely local closure. But when mesh sizes become so small that only a few updraft realizations happen inside each grid box, and with area fractions that cease to be negligible with respect to “one,” the whole concept of “classical” convective parameterization schemes collapses. In the 3MT scheme this problem is addressed by combining three key features of the scheme: (i) the separately computed deep convective condensation and large-scale condensation are merged as single input for a “prognostic–geometric” set of microphysical computations (sedimentation, auto-conversion, collection and melting–evaporation during fall); (ii) the convective detrainment is not diagnosed independently but becomes the result of the combined computations of closure, entrainment, and condensation; and (iii) the closure assumption (core of the physics–dynamics coupling) is a prognostic-type one with memory of the updraft area fraction and of the updraft vertical velocity of previous time steps. These

three interrelated characteristics of 3MT induce a good multiscale performance of 3MT, in particular in the gray zone. The latter can be defined as the range of horizontal mesh sizes for which the precipitating convection is partly parameterized and partly simulated by the resolved motions of the model. If nothing specific is done (i.e., using the classical diagnostic-type schemes of, e.g., ALADIN at gray-zone scales), this ambivalence results in double-counting or double-void situations, leading to several negative “gray-zone syndromes.” In convective situations drizzle appears nearly everywhere, and the precipitation maxima are too intense and too scattered. This happens especially over mountainous areas.

The multiscale performance of 3MT has been validated in a numerical weather prediction context up to a spatial resolution of 4 km (see Gerard et al. 2009). The ALARO-0 model utilizes 1) the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) Calcul Radiatif avec Nebulosité (ACRANEB) scheme for radiation (Ritter and Geleyn 1992, recast in a Net Exchanged Rate framework), 2) a semi-Lagrangian horizontal diffusion scheme (SLHD) (Vána et al. 2008), 3) some pseudo-prognostic turbulent kinetic energy (pTKE) scheme (i.e., a Louis-type scheme for stability dependencies, but with memory, advection, and autodiffusion of the overall intensity of turbulence), and 4) a statistical sedimentation scheme for precipitation within a prognostic-type scheme for microphysics (Geleyn et al. 2008). The physics package of the ALARO-0 model is coupled to the dynamics of the ALADIN model (Bubnová et al. 1995) via a physics–dynamics interface based on a flux-conservative formulation of the equations proposed by Catry et al. (2007).

For the present study, the same land surface model—Interactions between Soil, Biosphere, and Atmosphere (ISBA) (Noilhan and Planton 1989)—is used in both the ALARO-0 and ALADIN models. Furthermore, both models can be run with different schemes to impose the lateral-boundary conditions (Davies 1976; Radnóti 1995;

— 40 km domain    - - - 25 km and 4 km domain    ···· 10 km domain



FIG. 1. Domains corresponding to the different simulations at 40-, 25-, 10-, and 4-km horizontal resolution.

Termonia et al. 2012). For this study, the version of Radnóti (1995) is used in both models.

The ALARO-0 model runs operationally in a number of countries of the ALADIN and High-Resolution Limited-Area Model (HIRLAM) consortia (Austria, Belgium, Czech Republic, Croatia, Hungary, Norway, Portugal, Romania, Sweden, Slovenia, Slovakia, and Turkey) for the national NWP applications, the first of them already since 2008. More recently, the model is also used for climate runs. The ALARO-0 model is developed and maintained mainly through a collaboration between the RMI of Belgium and the Regional Cooperation for Limited Area Modelling for Central Europe (RC LACE). The developments of the ALARO-0 model (intentionally targeted at the gray-zone scales) are centered around the 3MT basic concept, which means that many other parameterization schemes must be adapted to the use of 3MT, but also sometimes the reverse. Thus, a rather wide international effort is needed.

As the first step of this study, the improvement of the downscaling by means of the ALADIN and ALARO-0 models is examined. This is done by comparing recent past (1961–90) summer precipitation data from an ALARO-0 and ALADIN simulation performed at 40-km spatial resolution (ALR40 and ALD40) (Fig. 1) with summer precipitation from the driving ERA-40 reanalysis data (Uppala et al. 2005).

Despite the fact that reanalysis data products are more continuous in space and time than station data, they

inevitably contain biases. A number of evaluations for ERA-40 reanalysis precipitation have been performed (e.g., Zolina et al. 2004; Ma et al. 2009). The ERA-40 precipitation has distinct regional limitations: most of them are generally related to the coarse horizontal resolution of the ERA-40 model, on one hand, and to its strong model dependency, on the other (Ma et al. 2009). All physical parameterizations within ERA-40, including those of precipitation, were run on a spatial resolution of about 125 km (Zolina et al. 2004; Ma et al. 2009). The model diagnostics precipitation in ERA-40 is produced by parameterized microphysical processes in clouds, which are formed at supersaturation by convective or large-scale processes (Ma et al. 2009). Total precipitation is then simply the sum of the convective precipitation generated by convective clouds and large-scale stratiform precipitation, associated with frontal or dynamical systems (Zolina et al. 2004). Hence, ERA-40 precipitation is a pure model product. Due to the poor skill of operational NWP models to account for all important physical mechanisms that affect the atmospheric water cycle, it appears to be one of the most uncertain forecasted parameters in the reanalysis (Zolina et al. 2004; Ma et al. 2009; Heikkilä et al. 2011). The 6-hourly forecasts from the ERA-40 reanalysis are used to calculate daily cumulated summer precipitation between 0600 and 0600 UTC of the next day. For coupling to the regional model we use a linear interpolation in time. This may produce errors at the lateral boundaries on our small domains (Fig. 1) but, as shown by Termonia et al. (2009), such errors only occur very rarely, and the impact on the statistics of extreme precipitation should be very minor.

To explore further the multiscale performance of ALARO-0, as found by Gerard et al. (2009) but now for climate time scales, we evaluate in a second step recent past simulations (1961–90) of the ALADIN and ALARO-0 models at varying horizontal resolutions against different station datasets.

- (i) and (ii) The ALADIN and ALARO-0 models are driven by ERA-40 and run at a horizontal resolution of 40-km spatial resolution with  $69 \times 69$  grid points on a domain that encompasses most of western Europe (ALD40 and ALR40, respectively; Fig. 1). These 40-km outputs are then used to perform a one-way nesting on a domain centered on Belgium (Fig. 1) using the following spatial resolutions:
- (iii) and (iv) 10-km spatial resolution on a  $67 \times 67$  grid (ALD10 and ALR10) and
- (v) 4-km spatial resolution on a  $181 \times 181$  grid (ALR04). That we did not run any ALD04 configuration is obviously linked to the corresponding gray-zone resolution, where the diagnostic

parameterization of convection would have become completely irrelevant (see section 4 for the first syndromes already noticeable in ALD10).

Finally, we also include ALADIN-Climate/CNRM simulations within our analysis so as to provide a reference for regional climate modeling. One part of the performed experiments within the ENSEMBLES project aimed to validate the models for the recent past climate. The results from this experiment, including 40 years of 25-km resolution ALADIN-Climate/CNRM simulations driven by the ERA-40 reanalysis (hereafter denoted as CNRM), are used in our analysis for the period 1961–90. From the ENSEMBLES data archive we have only selected the CNRM precipitation data for the grid points that coincide with the ALR04 domain (Fig. 1). The precipitation data correspond to daily means calculated for the interval 0000–2400 UTC. As mentioned in section 1, the model setup of CNRM and our simulations are different. The number of vertical levels that is used in our runs with the ALADIN and ALARO-0 models is 46 with a model top that extends up to 72 km. The CNRM simulations from ENSEMBLES have used 31 vertical levels. Furthermore, the CNRM simulations use a long-term and free run setup. Our procedure is to interpolate the original ERA-40 files to 40-km resolution. These 6-h files serve as initial and boundary conditions for 48-h ALD40 and ALR40 runs. These are started at 0000 UTC every day. The (3 h) output from these first runs serves as input for the high-resolution 10- and 4-km runs (ALD10, ALR10, and ALR04). However, to exclude spinup problems, the first 12 h are not taken into account. So we have 36 h of data left for the 4- and 10-km runs (which thus start at 1200 UTC). Finally, we again dismiss the first 12 h of the runs, to arrive at 24 h of output at 4- and 10-km resolution, and then integrate/reinitialize over each subsequent 24-h period during the summer period of June–August, 1961–90.

### b. Observations

The observation dataset comprises 93 climatological stations with daily accumulated precipitation, selected from the climatological network of the RMI of Belgium. The data have undergone a manual quality control by operators, and the stations were chosen so that continuous data for the 30-yr study period (1961–90) are available. The stations cover all of Belgium, thus representing conditions of coastal, inland, and higher orographic locations (Fig. 2).

## 3. Methods

### a. Data processing and analysis

Model validation against observations can either be done with station data or gridded station data. Both

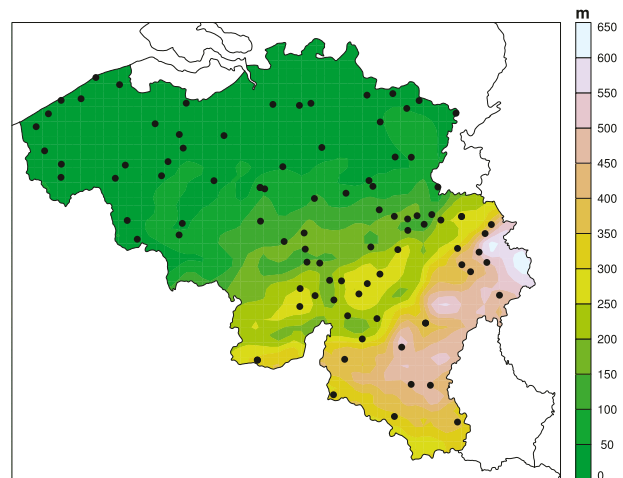


FIG. 2. Topography (m) of Belgium showing the location of the 93 selected climatological stations (black dots).

validation methods have their disadvantages (Hofstra et al. 2010). Model evaluation against observations at station level often raise issues related to the scale difference between the model and observation field (Tustison et al. 2001; Dulière et al. 2011). The model grid cell values correspond to spatially averaged values representing the area of the whole grid cell. Furthermore, the spatial variability of these averaged model fields will always be lower than the one of the observation field. These differences in spatial variability depend on the area of the grid cell as well as on the inherent variability of the field variable. Precipitation, for example, is known to have a relatively high spatial variability. To illustrate the differences in spatial variability in this study, Fig. 3 shows the different grid cell areas of the models together with the 93 climatological stations (i.e., observation points). The grid cell areas in this study range from 1600 km<sup>2</sup> for the 40-km horizontal resolution to 16 km<sup>2</sup> for the 4-km horizontal resolution (Fig. 3). Hence, reducing those spatially averaged model values with an originally greater heterogeneity to a single station point value leads to an inconsistent comparison. However, for long time periods, such as 30 years, we can assume that the spatial variability within a grid cell would be reduced in such a way that the spatial variability of both model and observation fields tends to converge (Dulière et al. 2011).

Another common way to overcome this scale inconsistency is the use of gridded data. The Climate Research Unit (CRU) and the European ENSEMBLES project provide daily gridded observation datasets (Mitchell and Jones 2005; Haylock et al. 2008). However, these gridded datasets are in some regions constructed by interpolation or area-averaging of station observations from a small number of stations, which

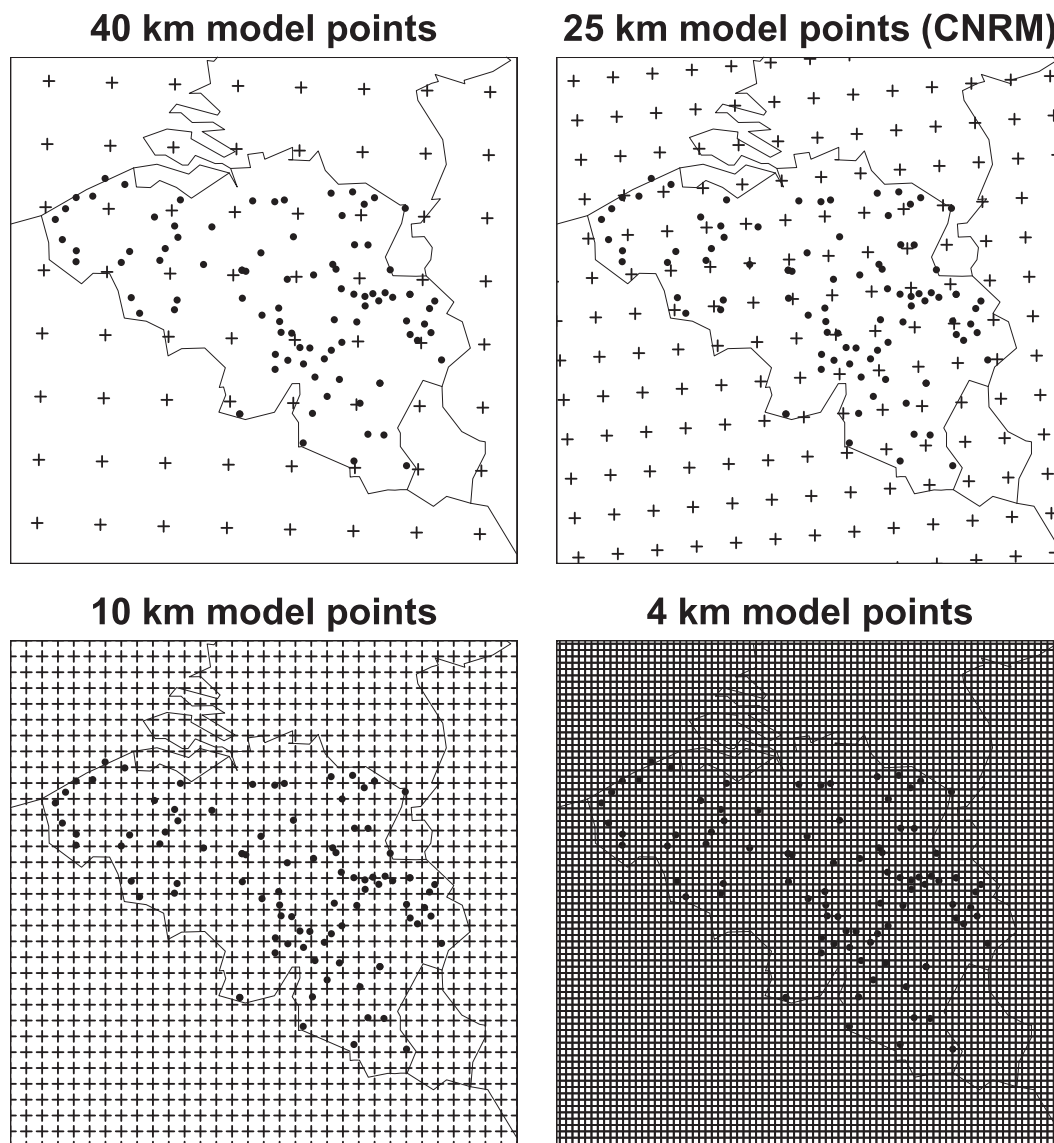


FIG. 3. Model grid points over Belgium for each of the horizontal resolutions for which the simulations are performed. The black dots represent the 93 climatological stations.

smooths and possibly affects the extreme values within the dataset (Hofstra et al. 2010). Since this study aims to examine extreme precipitation events, the models are evaluated against station observations. This is done by comparison of daily observed station-level precipitation with modeled daily precipitation of the nearest grid box over land. The 93 resulting precipitation time series selected from the model simulations are not corrected for topography with respect to altitude of the nearest station. It is difficult to apply such correction for precipitation because of its dependency on topography, humidity, buoyancy, and other local variables (Soares et al. 2012a).

Time discrepancy between computations of daily cumulated precipitation from station observations and model output is an important, but rarely highlighted, problem within precipitation evaluation studies. To deal with this problem, the error analysis can be performed on longer than daily time scales, such as monthly, seasonal, or annual time scales (Ma et al. 2009; Soares et al. 2012b). However, in this study the model evaluation is done on a daily basis, requiring a consistent calculation of the daily precipitation values. Daily observed precipitation corresponds to the total accumulated precipitation between 0800 and 0800 local time (LT) of the following day. Hence, the daily model values for all

simulations (ALR40, ALD40, ALR10, ALD10, and ALR04) have been calculated based on the definition of observed daily accumulation, which corresponds to 0600 and 0600 UTC of the following day (Table 1).

### b. Extreme value analysis and peak-over-threshold methods

The methods used for the modeling of extreme events are similar to those used in Hamdi et al. (2012). Threshold models and peak-over-threshold (POT) methods are useful tools for the modeling of extreme events. A well-known distribution that may describe the behavior of the excesses or POT events is the generalized Pareto distribution (GPD) (Coles 2001). Recently, several authors have modeled extreme precipitation with the GPD (e.g., Ribatet et al. 2009; Roth et al. 2012; Mailhot et al. 2013).

Consider a sequence of independent and identically distributed random variables  $X_1, X_2, \dots, X_i$  from an unknown distribution  $F$ . We are interested in the extreme events that exceed a certain high threshold  $u$ . The distribution function of such an extreme event  $X$  from the  $X_i$  sequence can then be defined as

$$F_u(y) = P\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad (1)$$

with  $y > 0$ . Equation (1) is the conditional probability that the threshold  $u$  is exceeded by no more than an amount  $y$ , given that the threshold  $u$  is exceeded. Given that  $X > u$ , the GPD of the excesses ( $X - u$ ) is then given by

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, \quad (2)$$

where  $\xi$  is the shape parameter and  $\sigma$  is the scale parameter. The GPD with parameters  $\xi$  and  $\sigma$  describes the limiting distribution for the distribution of excesses [Eq. (1)] and can be used to model the exceedances of a threshold  $u$  by a variable  $X$ . Thus, for  $x > u$ ,

$$P\{X > x | X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}. \quad (3)$$

It follows that

$$F(x) = P\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}, \quad (4)$$

where  $\zeta_u = P\{X > u\}$ . In this study the parameters of the GPD are estimated by the maximum-likelihood method, following the definitions of Stephenson (2002). The level

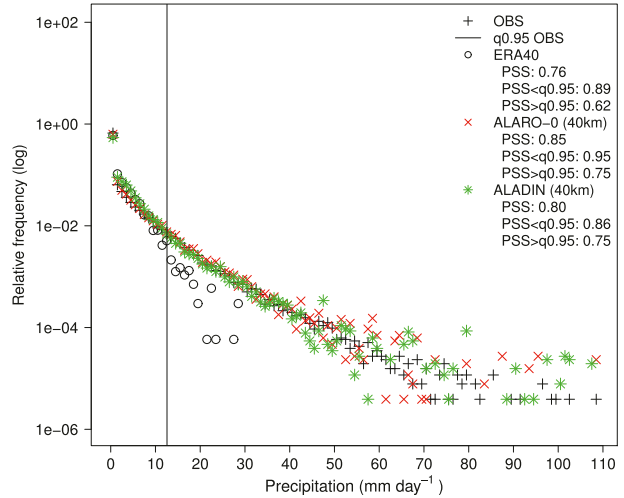


FIG. 4. Relative frequencies of observations, ERA-40, ALR40, and ALD40. Frequencies are computed with the 30-yr (1961–90) daily cumulated summer precipitation given for each station separately and are displayed on a logarithmic scale. Numbers for PSS correspond to the average of the Perkins skill score [Eq. (7)] calculated for precipitation amounts below and above the 0.95th quantile of the observations (PSS < q0.95 and PSS > q0.95). The black line indicates the 0.95th quantile of the observations.

$x_m$  that is on average exceeded once every  $m$  observations is the solution of

$$\zeta_u \left[1 + \xi \left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}. \quad (5)$$

The  $x_m$  return level, which gives the amount of extreme precipitation corresponding to a given number of observations  $m$ , is then given by

$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u)^\xi - 1]. \quad (6)$$

## 4. Results and discussion

### a. Effect of downscaling

As a first step we validate the effect of the downscaling of the ERA-40 with the ALADIN and ALARO-0 models. Figure 4 shows the relative frequencies calculated for daily precipitation amounts of ERA-40, ALR40, and ALD40, which are binned into bins of  $1 \text{ mm day}^{-1}$ . As a reference the relative frequencies of the observations are also shown. A logarithmic scale has been used for better representation of the extreme values. From both ERA-40 data and the ALR40 and ALD40 data 93 grid points, corresponding to the closest grid points to the observation stations, have been selected. It should be noted that the ERA-40 only has two grid points over



Belgium. For low precipitation amounts (i.e., <0.95th quantile of the observations) the ERA-40 as well as ALR40 and ALD40 coincide well with the observations. However, for the higher rainfall rates ERA-40 starts to diverge from the observations, while ALR40 and ALD40 still approach the observations. Both 40-km models are able to reproduce rainfall rates up to  $108 \text{ mm day}^{-1}$ , while the reanalysis does not capture the higher precipitation amounts due to the low spatial resolution of the ERA-40 data. To provide a measure of similarity between observed and modeled frequencies, the Perkins skill score (PSS) has been calculated (Perkins et al. 2007):

$$\text{PSS} = \sum_1^n \min(Z_1, Z_2), \quad (7)$$

where  $n$  is the number of bins and  $Z_{1,2}$  is the frequency of values in a given bin from the observation and model data, respectively. This metric measures how well the observations and modeled frequencies coincide, with a PSS ranging from zero for no overlap to a skill score of one for a perfect overlap. Similar to Boberg et al. (2010) and Domínguez et al. (2013), the PSS has been calculated for daily precipitation amounts going from  $0 \text{ mm day}^{-1}$  up to the 0.95th quantile of the observations ( $\text{PSS} < q_{0.95}$ ) and for precipitation amounts above the 0.95th quantile of the observations ( $\text{PSS} > q_{0.95}$ ). In this way, the skill score is to a larger extent influenced by the more extreme precipitation values (Boberg et al. 2010). The skill scores are calculated for each station separately. The final PSS is then simply the mean value of the average of  $\text{PSS} < q_{0.95}$  and  $\text{PSS} > q_{0.95}$  over the 93 stations. The 0.95th quantile of the observations, which is used as a threshold for the calculation of the modified PSS, is also shown in Fig. 4. The Perkins skill scores for ERA-40 are relatively low, and for the higher precipitation amounts ERA-40 has a much lower PSS ( $\text{PSS} > q_{0.95}$ : 0.62) than ALR40 and ALD40 ( $\text{PSS} > q_{0.95}$ : 0.75). ALR40 and ALD40 perform very similar with respect to the observations and have relatively high PSS, which are close to one. To summarize, the downscaling with the ALARO-0 and ALADIN models is significantly different from the driving ERA-40 and is closer to the observations. In particular, ALR40 and ALD40 produce more extreme precipitation than their driving ERA-40.

### b. Multiscale performance of ALARO-0

To investigate the multiscale performance of ALARO-0, 40-, 10-, and 4-km horizontal resolution simulations of ALARO-0 together with 40- and 10-km horizontal

resolution simulations of ALADIN are compared with respect to station observations.

#### 1) SPATIAL AND TEMPORAL DISTRIBUTION

Figure 5 shows the observed and simulated spatial distribution of the 30-yr-averaged summer precipitation. On top of each subfigure average values over the 93 stations for the cumulated summer precipitation are given. On average all models except for CNRM overpredict the observed cumulated summer precipitation. Both observation and simulation fields show a clear topographical dependency, with a gradual increase in precipitation going from the northwest (low altitudes) to the southeast (high altitudes) of the country. The ALARO-0 and ALADIN simulation at 40 km show a very similar distribution. Obviously, the precipitation fields for the simulations with low spatial resolution are less heterogeneous than the ones with high spatial resolution. However, the 25-km spatial resolution CNRM plot illustrates less variability than the 40-km simulations: also, the local maximum in the southeast cannot be seen on the CNRM plot. For the higher-resolution simulations ALARO-0 approaches much better the observations than ALADIN. For instance, ALD10 overpredicts cumulated summer precipitation with values that are, on average, over all stations almost 100 mm higher than observed. On the contrary, the average values for ALR10 and ALR04 differ only slightly from the observations, and the observed local maximum at the higher altitudes is very well simulated by both models.

The scatterplots presented in Fig. 6 are consistent with the spatial distributions shown in Fig. 5. Each point in the scatterplots represents the summer cumulated precipitation for each year in the 30-yr period averaged for the 93 stations. The linear regression line (solid line) and its determination coefficient ( $R^2$ ) is also presented for each of the five models. Except for ALD10, summer precipitation is relatively well simulated by all models. The ALD10 model shows again a clear overestimation of observed summer precipitation. This is an indirect confirmation that, with 10-km mesh sizes, the syndromes linked to the gray-zone performance are already present (see section 2a).

#### 2) ERROR STATISTICS

The previous analysis showed the ability of the models to represent the spatial and temporal pattern of mean annual summer precipitation. To quantify this ability we have computed some important error statistics. Figure 7 shows the spatial distribution of the 30-yr average summer biases of the daily cumulated precipitation, as well as the mean bias over the 93 climatological stations.

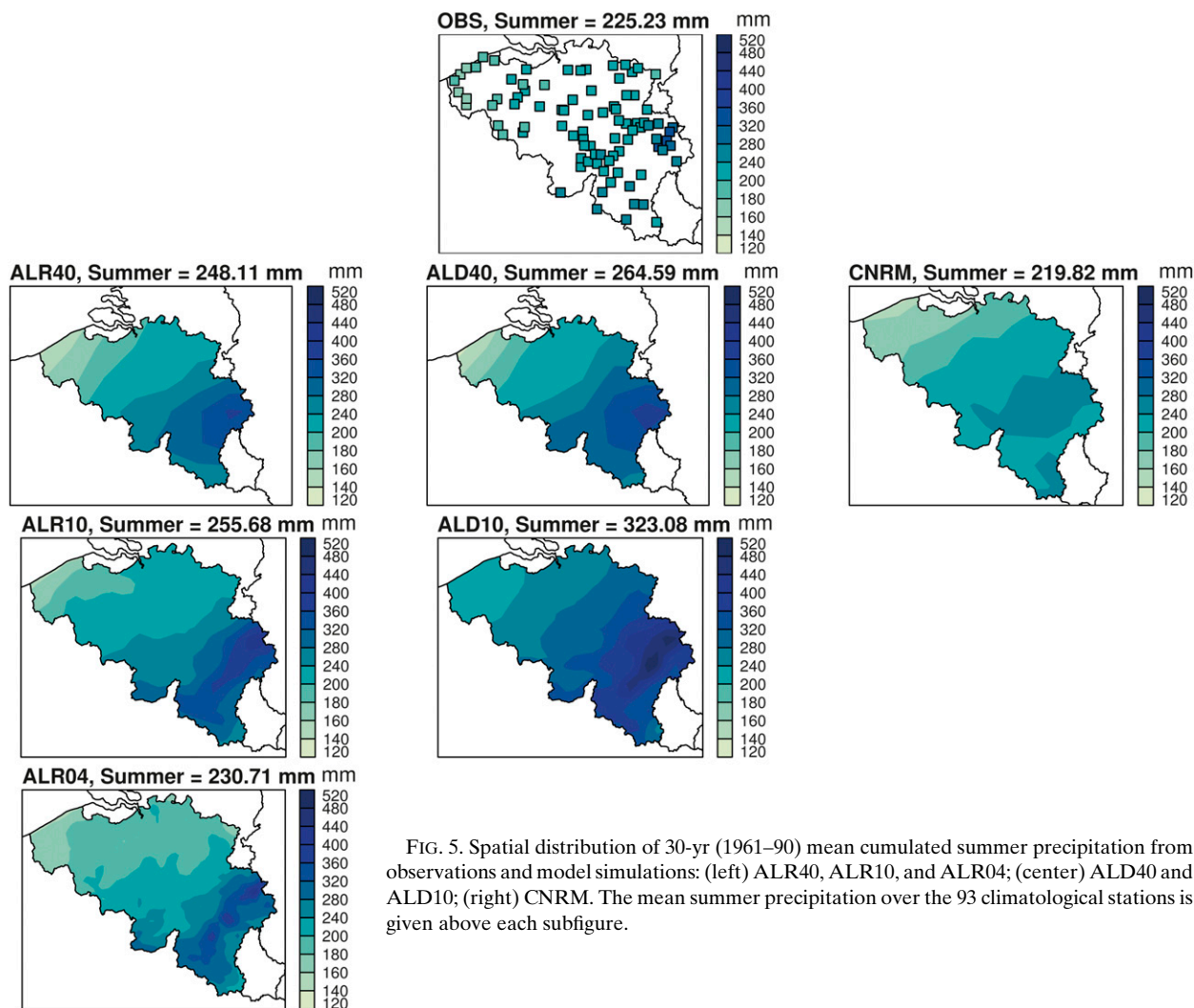


FIG. 5. Spatial distribution of 30-yr (1961–90) mean cumulated summer precipitation from observations and model simulations: (left) ALR40, ALR10, and ALR04; (center) ALD40 and ALD10; (right) CNRM. The mean summer precipitation over the 93 climatological stations is given above each subfigure.

Average values over the 93 stations of other 30-yr mean summer statistics are also given: the RMSE and the mean absolute error (MAE). The statistics are calculated with daily values for each station separately. Both 40-km simulations ALR40 and ALD40 again perform similar. Overall, the biases are remarkably lower for ALARO-0 than for ALADIN. The bias over the 93 climatological stations between model simulations and observations is  $0.25 \text{ mm day}^{-1}$  for ALR40,  $0.43 \text{ mm day}^{-1}$  for ALD40,  $-0.06 \text{ mm day}^{-1}$  for CNRM,  $0.33 \text{ mm day}^{-1}$  for ALR10,  $1.06 \text{ mm day}^{-1}$  for ALD10, and  $0.06 \text{ mm day}^{-1}$  for ALR04. The error statistics for all three ALARO-0 simulations show a similar improvement, suggesting a multiscale performance of ALARO-0. However, one should also keep in mind that error statistics are not entirely fair when validating models with different spatial resolution. Small displacements of precipitation maxima and minima in higher-resolution models

are highly penalized by error statistics because of the so-called double penalty effect (Soares et al. 2012a).

The aforementioned underestimation by CNRM is confirmed by the spatial distribution of its bias. Furthermore, the coastal precipitation is by all other models generally better simulated than the inland precipitation (Fig. 7). The larger and positive differences at the higher elevations can partly be assigned to higher uncertainties in the measurements of the observations due to rain gauge undercatchment (Buonomo et al. 2007). However, this overestimation, which is pronounced more strongly for ALD10 (Fig. 7), can also be attributed to the model or the driving ERA-40 data. All three ALARO-0 simulations (40-, 10-, and 4-km horizontal resolution) produce the lowest deviations from the observations, with a tendency to slightly overestimate (underestimate) in the southern (northern) part of the country. ALARO-0 values for RMSE and MAE lie in the same range as those for

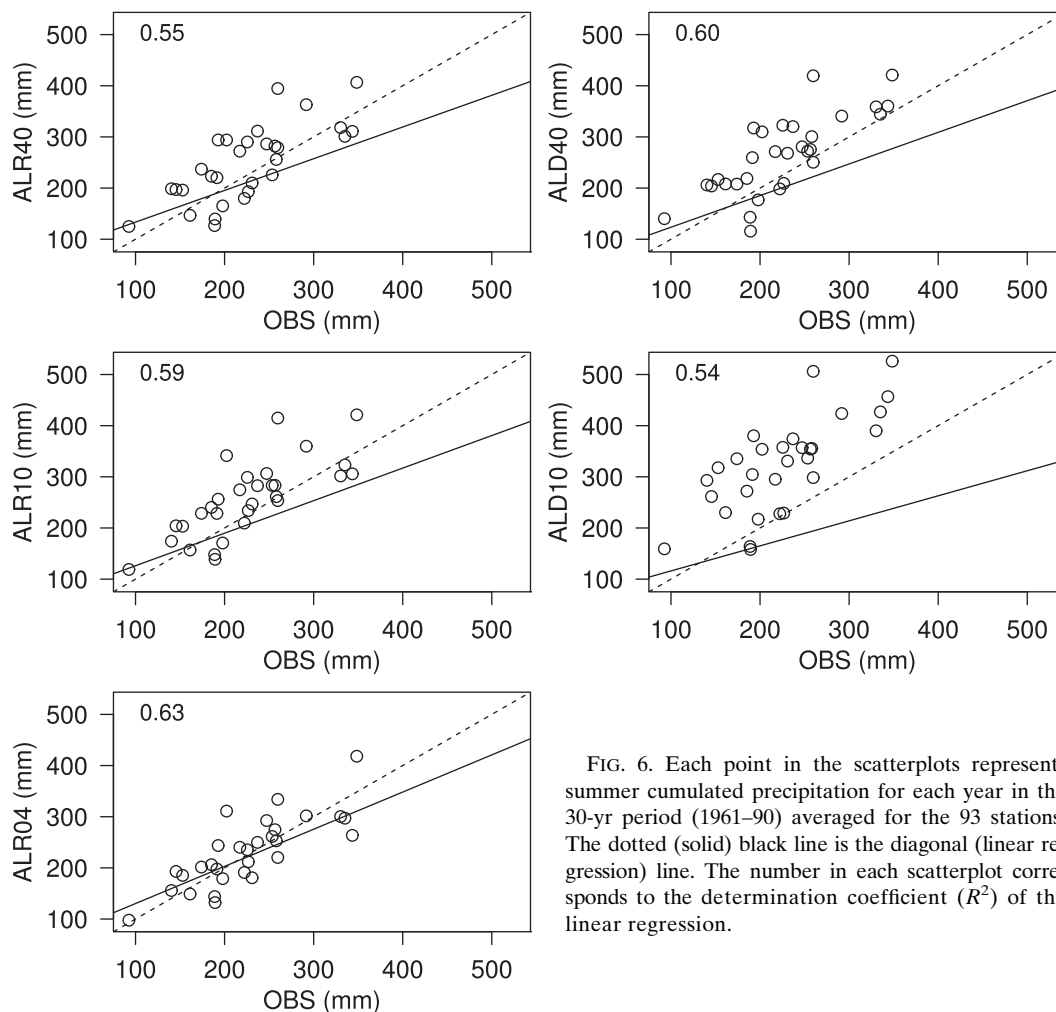


FIG. 6. Each point in the scatterplots represents summer cumulated precipitation for each year in the 30-yr period (1961–90) averaged for the 93 stations. The dotted (solid) black line is the diagonal (linear regression) line. The number in each scatterplot corresponds to the determination coefficient ( $R^2$ ) of the linear regression.

ALADIN, indicating that the low mean biases of ALARO-0 are possible owing to cancellation effects arising from the bias computation. Nevertheless, the overall errors of the ALARO-0 simulations are still smaller than those of ALD10.

To get an understanding of the trend of frequency and intensity of extreme precipitation, density curves and frequency and quantile distributions of all six simulations have been created (Figs. 8–10). The densities in Fig. 8 have been calculated with the square root of the daily precipitation since the majority of the precipitation rates are less than  $10 \text{ mm day}^{-1}$ . All models tend to overestimate the amount of “drizzle” and low precipitation (i.e.,  $<1 \text{ mm day}^{-1}$ ). In the  $1\text{--}2 \text{ mm day}^{-1}$  range, both ALADIN simulations as well as CNRM overestimate the observed density almost by 2 times, while ALARO-0 starts to approach closely the observed density (Fig. 8, center). The latter continues to do this up to the right-end tail of the observed density curve (Fig. 8,

right). Perkins et al. (2007) use probability density functions (PDFs) for the evaluation of simulated daily precipitation over Australia from 14 different climate models. Similarly to the density curves of ALADIN and CNRM, the PDFs in Perkins et al. show for all models an overestimation of “drizzle,” with most models overestimating the observed density of rainfall in the  $1\text{--}2 \text{ mm day}^{-1}$  range by 2–3 times.

The relative frequencies, shown in Fig. 9, are again calculated for daily precipitation amounts of the observations and model data, which are binned into bins of  $1 \text{ mm day}^{-1}$ . For the low precipitation rates all models manage to reproduce the observed frequencies relatively well. Once the 0.95th quantile of the observations (indicated by the vertical black line) is exceeded, CNRM shows an increasing departure from the observations with frequencies left shifted from the observations. ALARO-0 and ALADIN at 40-km horizontal resolution reveal again a similar result, while for the higher

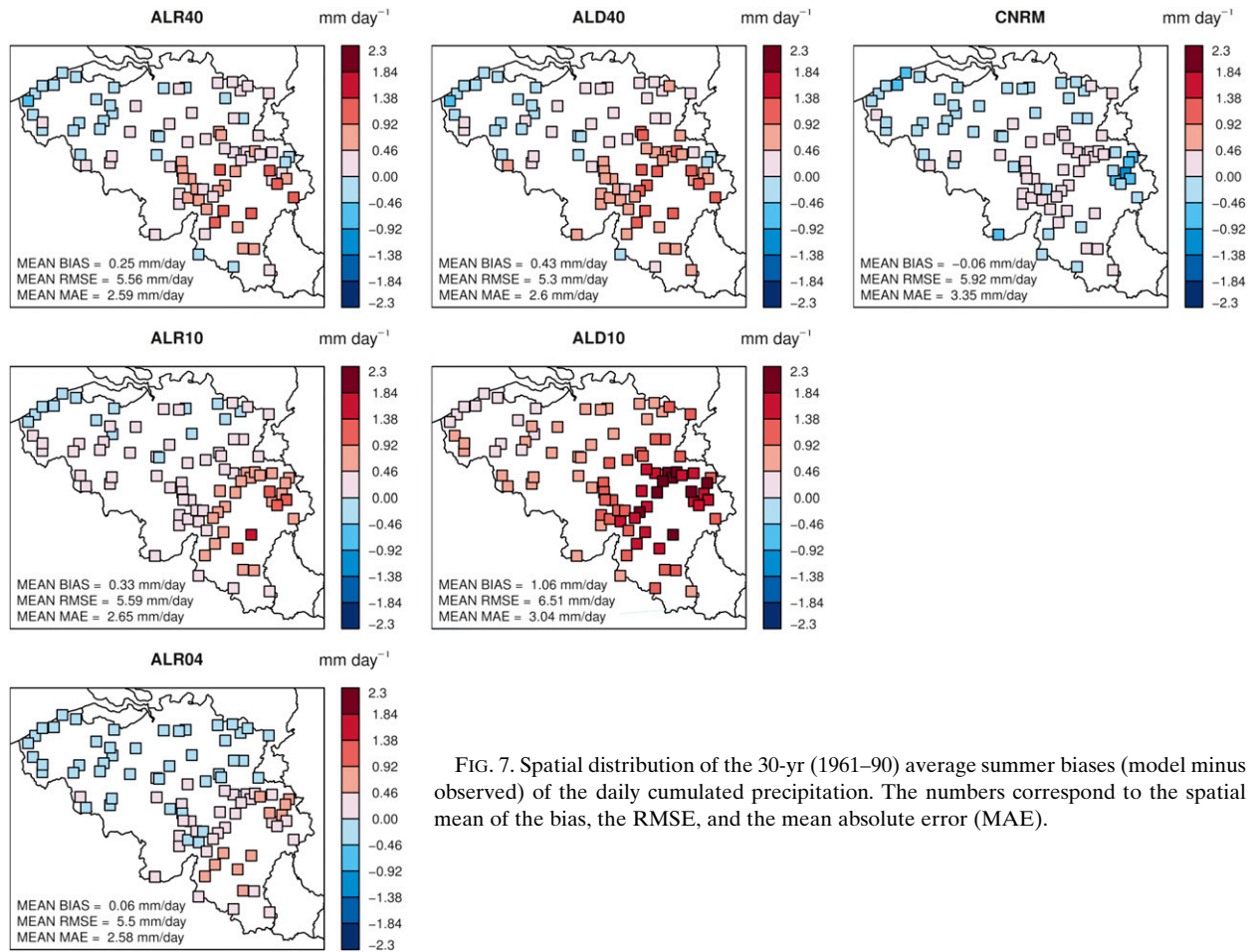


FIG. 7. Spatial distribution of the 30-yr (1961–90) average summer biases (model minus observed) of the daily cumulated precipitation. The numbers correspond to the spatial mean of the bias, the RMSE, and the mean absolute error (MAE).

10-km resolution a clear difference between both models is apparent. The small overestimation of ALD10 for the low precipitation rates persists and becomes larger for the higher rates. The model clearly rains too often, both with very small and very high quantities of rainfall. On the other hand, the frequencies of ALR04 and ALR10 nicely follow the observations, showing their ability to capture the occurrence of extreme and rare precipitation events, with values around 100 mm, quite well. As a measure for similarity between the observed and modeled frequencies, the PSS [Eq. (7)] are also given in Fig. 9. The overall PSS, as well as PSS for precipitation amounts below and above the 0.95th quantile of the observations, is higher for ALARO-0 than for ALADIN and CNRM.

The quantile distributions confirm the ability of ALR04, ALR10, and even ALR40 to reproduce extreme rainfall rates (Fig. 10). Only the highest 99.9 quantile (i.e., strongest events) is slightly overestimated by ALARO-0. It is evident that such events, which are situated in the very end of the distribution, might correspond to

outliers. Consistently with the frequency plots, the higher quantiles are over- and underestimated by ALD10 and CNRM, respectively.

Previous results can be qualified in the context of other regional downscaling studies; however, a direct comparison is difficult because of differences in study area and model design. Soares et al. (2012a) performed a dynamical downscaling of 20 years of the ECMWF Interim Re-Analysis (ERA-Interim) (1989–2008) for Portugal using the Weather Research and Forecasting (WRF) model. Two WRF high-resolution simulations (9 and 27 km) and ERA-Interim are compared with station observations. For summer precipitation, their results show a different frequency distribution for the 9- and 27-km simulation. The 9-km frequencies of summer precipitation follow well the observed frequencies and show a clear improvement compared to the driving reanalysis. Our results show a coherent performance of the ALARO-0 model across all resolutions and the good model performances as displayed in Figs. 8–10 can be practically attributed to the quality of the physics

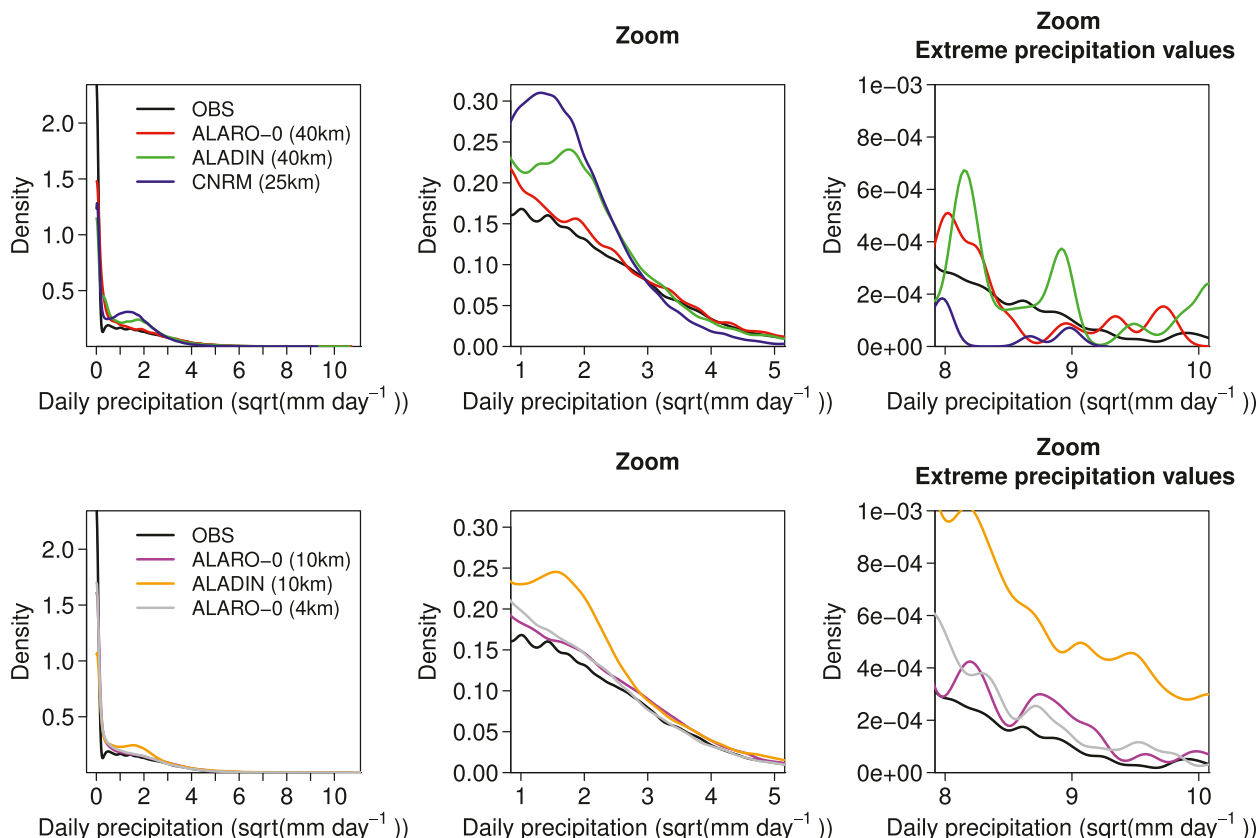


FIG. 8. Density curves of (top) observations, ALR40, ALD40, and CNRM and (bottom) observations, ALR10, ALD10, and ALR04. Densities are computed with the 30-yr (1961–90) daily cumulated summer precipitation given for each station separately. The  $x$  axes represent the square root of the daily precipitation since the majority of the precipitation rates are less than  $10 \text{ mm day}^{-1}$ .

parameterizations unrelated to the increase of the resolution. Finally, the persistent positive biases of the ALADIN model ALD10 are in accordance with other studies where recent past (1961–90) ALADIN simulations at 10-km horizontal resolution, driven by ERA-40 data, are validated against gridded observations (see Csima and Horányi 2008; Skalák et al. 2008). According to Skalák et al. (2008), these positive (summer) precipitation biases can be linked with the tendency of the model “to precipitate more often than in the station observations.”

### 3) EXTREME VALUE ANALYSIS

The extreme value analysis has been performed for each station separately, using the 30-yr daily summer data. The use of a generalized Pareto distribution as a model for threshold excesses assumes independent excesses (Coles 2001). In practice this is rarely the case. Exceedances over a certain threshold often occur in clusters. To account for these clusters of POT events, the data have been declustered by selecting the maximum value within each cluster. The independence of two clusters of POT events is determined by a combination

of the threshold and the separation time between both clusters. However, the choice of a suitable threshold and separation time is relatively arbitrary. The threshold has to be high enough in order to ensure *extreme* events and to avoid dependency between the events, but a threshold that is too high prevents statistical significance owing to a loss of information (Kysely and Beranová 2009; Heikkilä et al. 2011). Similar to the study of Heikkilä et al., the threshold has been defined for each station separately as the 0.95th quantile of daily summer precipitation so that spatial differences in the precipitation amount (see Fig. 5) are taken into account.

The results obtained by using cluster maxima defined with different separation times (e.g., 1, 2, or 4 days) do not differ much from the results when the original non-declustered data have been used (not shown). Hence, in accordance with another study on extreme precipitation of Kysely and Beranová (2009), two POT events are considered to be independent when the minimum separation time between both events is one day.

To investigate if the underlying probability distribution of the (declustered) peak-over-threshold

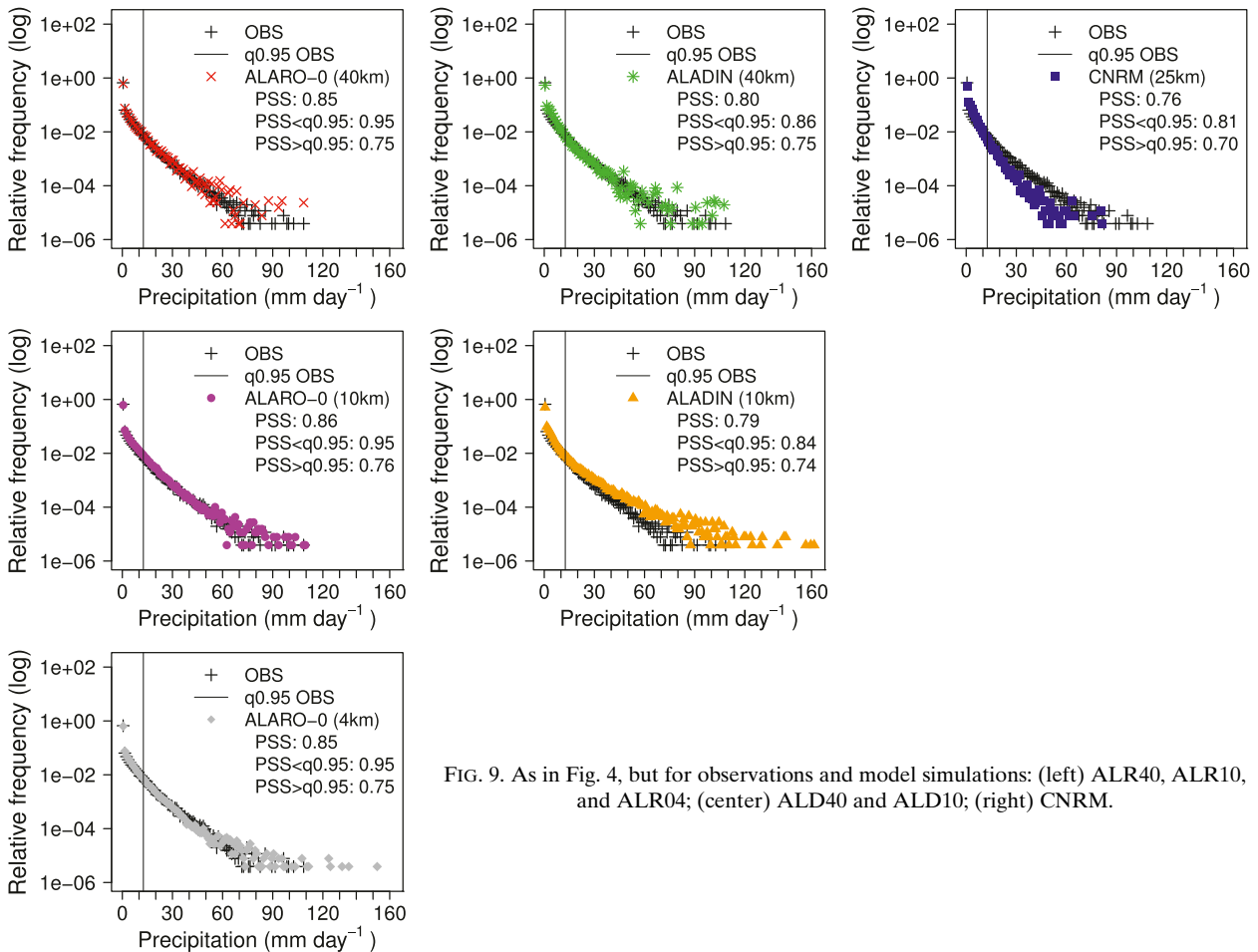


FIG. 9. As in Fig. 4, but for observations and model simulations: (left) ALR40, ALR10, and ALR04; (center) ALD40 and ALD10; (right) CNRM.

events of the observations and models significantly differs, a Kolmogorov–Smirnov test has been applied. The Kolmogorov–Smirnov test statistic is defined as the maximum absolute difference between two distribution functions:

$$D_{n_1, n_2} = \max |F_{n_1}(x) - F_{n_2}(x)|, \quad (8)$$

where  $F_{n_1}(x)$  and  $F_{n_2}(x)$  are the empirical distribution functions of the observations and the model, respectively, and  $n_i$  refers to the number of samples. The null hypothesis ( $H_0$ ) that the distribution of the observed POT events equals the distribution of the modeled POT events is rejected at significance level  $\alpha = 0.05$  if

$$K = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} D_{n_1, n_2} > K_\alpha, \quad (9)$$

where  $K_\alpha$  is the critical  $\alpha$  level of the Kolmogorov distribution:

$$\Pr(K \leq K_\alpha) = 1 - \alpha. \quad (10)$$

Figure 11 shows for each station the  $K$  statistic of the observations and models. In general, the  $K$  values for the ALARO-0 model at all three spatial resolutions are much smaller than ALADIN and CNRM.  $H_0$  is accepted at the 95% level at 35 and 16 stations for ALD40 and ALD10, respectively. For ALARO-0 at 40, 10, and 4 km,  $H_0$  is accepted at 46, 47, and 38 locations, respectively. Compared to ALD10, there are for ALARO-0 more stations at the high altitudes for which the distribution of the POT events equals the observed distribution of the POT events. This indicates that an increase in resolution does not necessarily contribute to a better representation of orographic precipitations. In the case of CNRM,  $H_0$  is rejected for all stations. Thus, consistent with the results from the frequency and quantile distributions, the Kolmogorov–Smirnov test confirms that the ALARO-0 simulations yield more reliable statistics of the extreme events.

The GPD equation [Eq. (2)] is then fitted through the selected cluster maxima of the observations and the six model simulations ALR40, ALD40, CNRM, ALR10, ALD10, and ALR04. The 5- and 20-yr return levels of

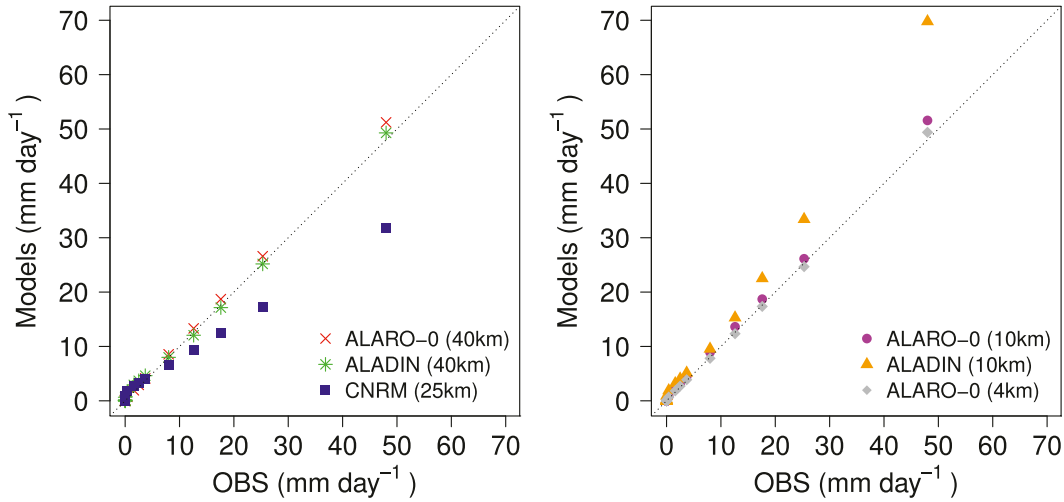


FIG. 10. Quantiles (2.5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 97.5, 99, and 99.9) of observations vs (left) the ALR40, ALD40, and CNRM models and (right) ALR10, ALD10 and ALR04 models. Quantiles are computed with the 30-yr (1961–90) daily cumulated summer precipitation given for each station separately.

the POT models for the observations and six simulations are shown in Figs. 12 and 13. The return levels  $x_m$  are calculated by Eq. (6) using the declustered data with 1-day separation time and a threshold  $u$ , defined as the 0.95th quantile. Since the return levels  $x_m$  are calculated on an annual basis, the value for  $m$  equals 92

observations, corresponding to the number of summer days within one year of the study period. The return levels for both return periods are generally larger at the higher elevations. The 95% confidence levels of the observed return levels are also indicated. It appears that for most stations the return levels of ALARO-0 lie

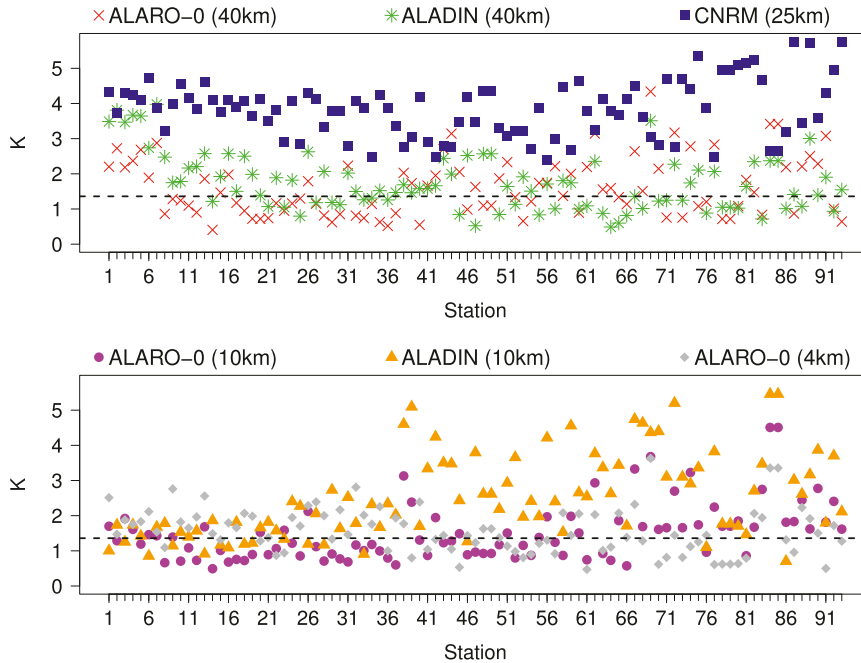


FIG. 11. The  $K$  statistic from a Kolmogorov–Smirnov test [Eq. (9)]. The 93 stations (abscissa) are shown by ascending altitude (from left to right). The test is performed on the POT events of the observations vs the (top) ALR40, ALD40, and CNRM and (bottom) ALR10, ALD10, and ALR04 model simulations. The horizontal dotted line represents the critical  $K$  level with significance  $\alpha = 0.05$ .

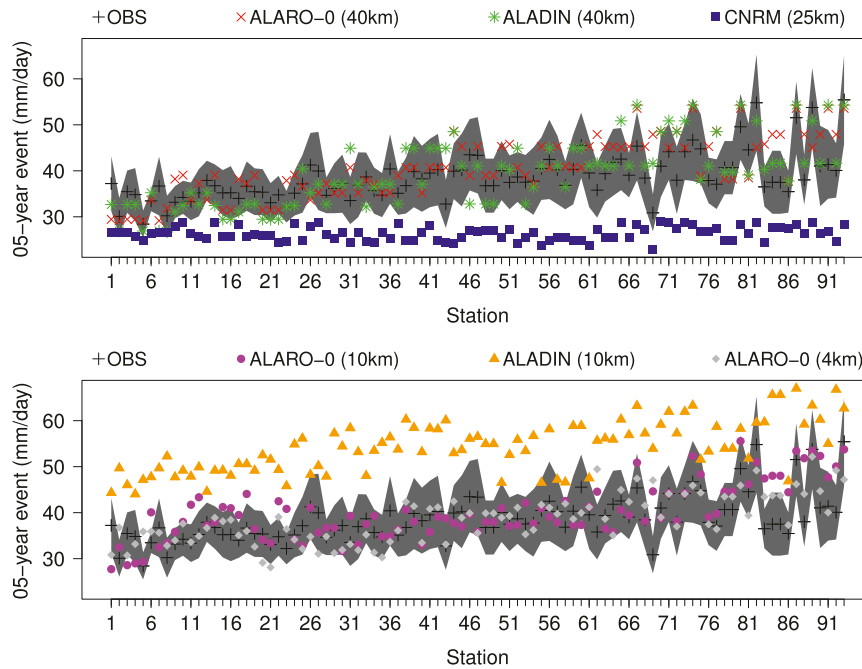


FIG. 12. The 5-yr return levels of the POT models for the observations and model simulations: (top) ALR40, ALD40, and CNRM and (bottom) ALR10, ALD10, and ALR04. The 93 stations (abscissa) are shown by ascending altitude (from left to right), and the shaded area represents the 95% confidence interval of observed return levels.

within the 95% confidence range of the observed return levels. In contrast to ALARO-0, ALD10 and CNRM are not able to produce the observed 5- and 20-yr return events. Their estimated return levels lie for a great

number of stations outside the observed confidence interval.

In line with what Hamdi et al. (2012) found for summer maximum temperature, previous results from the

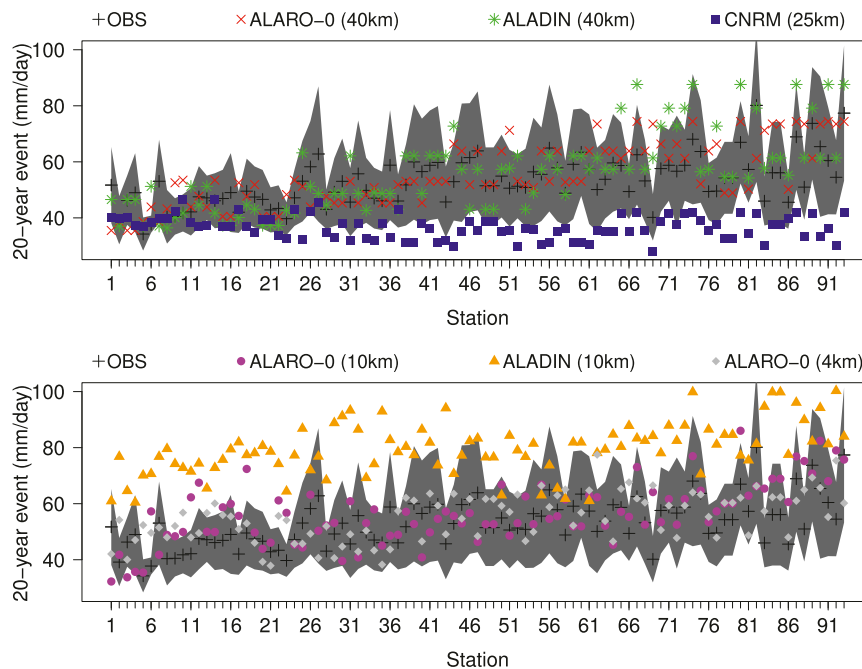


FIG. 13. As in Fig. 12, but for 20-yr return levels.



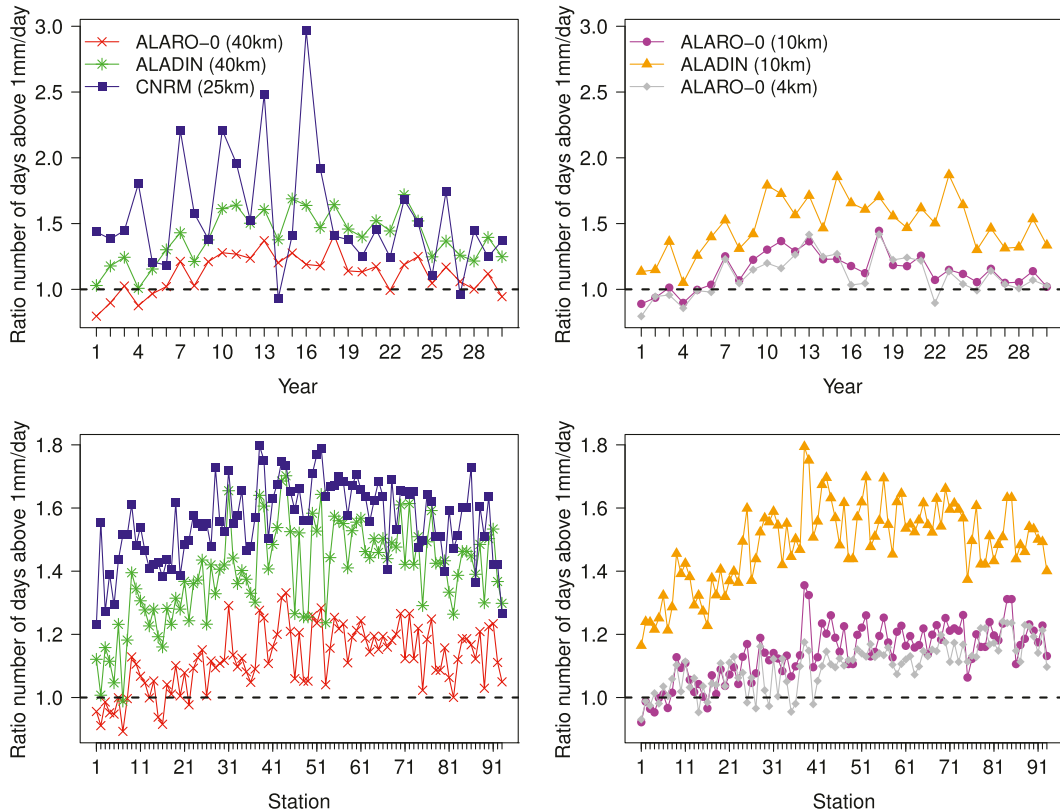


FIG. 14. (top) Spatial mean of ratio of number of days above 1 mm day<sup>-1</sup> (i.e., WD) in models to observations. (bottom) Temporal mean of ratio of number of days above 1 mm day<sup>-1</sup> (i.e., WD) in models to observations. Station sequence as in Fig. 12.

extreme value analysis show for ALARO-0 at the high resolutions of 4 and 10 km, as well as at 40-km horizontal resolution, a clear improvement in simulating extreme summer precipitation. Extreme events are also often investigated by means of climate indices (e.g., Herrera et al. 2010; Domínguez et al. 2013; Dulière et al. 2011; Soares et al. 2012b). To complete the extreme value analysis, two main precipitation indices have been calculated: the number of wet days and the number of very heavy precipitation days. Both indices are explained below and are calculated for each year (i.e., summer season) and each climatological station.

4) NUMBER OF WET DAYS

The number of wet days (WD) for the observations and models are defined as the annual count of days when precipitation is >1 mm. Figure 14 shows the ratio of WD in models to observations. As the model values represent a whole grid box, we could assume that the models, and especially the lower resolution models, will poorly reproduce the indices at the station points. However, the low-resolution ALR40 model (left) reproduces

relatively well the observed WD. On the other hand, ALADIN and CNRM show an overestimation for WD. This can be explained by the fact that precipitation may occur more systematically at the model grid box level, which gives rise to a WD even when no precipitation has been observed at the station location. Compared to ALADIN and CNRM, the ALARO-0 model (at 4-, 10-, and 40-km horizontal resolution) is able to better reproduce the number of wet days.

5) NUMBER OF VERY HEAVY PRECIPITATION DAYS

The number of very heavy precipitation days is derived by an annual counting of days with precipitation rates >20 mm. The temporal and spatial means of the number of very heavy precipitation days are consistent with the results from foregoing extreme value analysis. Overall, ALR04, ALR10, and ALR40 can reproduce the number of days with precipitation >20 mm day<sup>-1</sup> very well (Fig. 15). ALR04 and ALR10 have the highest correlations, and for three out of the 93 stations ALR10 predicts exactly the same number of days with heavy precipitation rates as have been observed.

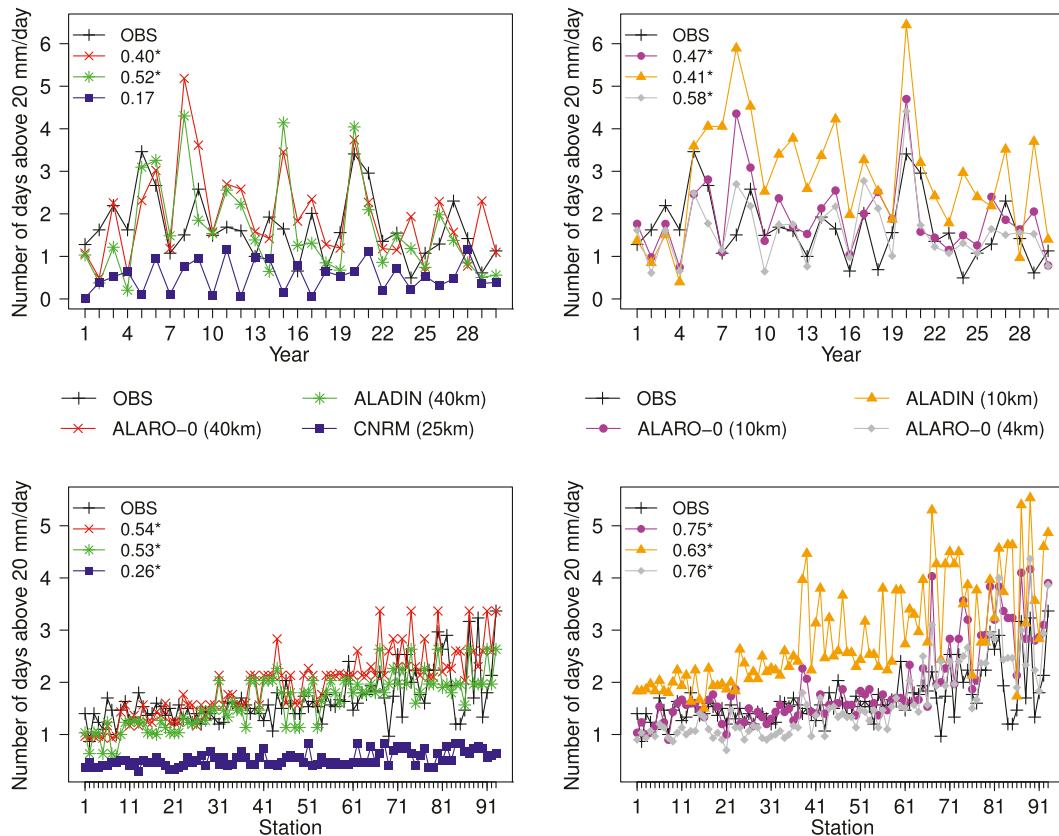


FIG. 15. (top) Spatial mean of number of days above 20 mm day<sup>-1</sup>. (bottom) Temporal mean of number of days above 20 mm day<sup>-1</sup>. The numbers correspond to Pearson correlation coefficients. Significant correlation coefficients at significance level 0.05 based on *t* statistics are indicated with an asterisk. Station sequence as in Fig. 12.

## 5. Conclusions

Extreme value analysis, using the peak-over-threshold method and generalized Pareto distribution, was performed in order to explore the relative importance of resolution versus parameterization formulation on the simulation of extreme daily summer precipitation. The results show that dynamical downscaling of the ERA-40 reanalysis using the ALARO-0 model adds value to the prediction of extreme daily summer precipitation when compared to the ERA-40 results. Hence, running a limited area model with the adapted parameterization, which was originally motivated to perform in the convection-permitting resolutions, statistically outperforms the global data in the output of extreme precipitation events of the ERA-40 reanalysis. The main strength of these tests is that, by the choice of the setup, we are considering the pure effect of the downscaling, without being obliterated by issues such as spectral nudging. Moreover, the model regenerates the precipitation instead of letting it evolve from its initial state. The regional nature keeps the computing cost within reach of

a typical small center, like the RMI, while reproducing the correct statistics of the extreme precipitation events consistently with the large-scale forcing imposed by the initial conditions and lateral boundaries. Furthermore, it should be stressed that the present model version has been developed and tuned in a context of NWP, is used as a 12-member component of the Grand Limited Area Model Ensemble Prediction System (GLAMEPS), and has been taken as such to downscale ERA-40 data. This can be seen as an extra indirect validation of the NWP applications running ALARO-0, in the sense that the model has a more correct climatology of convective rain. It is clear that there are several components, such as the physics–dynamics interaction, the interaction between model physics, and the numerics, that may influence the climatology of the precipitation. However, it is difficult to isolate the importance of these components, and it is beyond the scope of this study to address the relative impact of the different parameterization updates within ALARO-0. It should be kept in mind, though, that all of these factors play a crucial role in the model performance at gray-zone resolutions.

ALARO-0 simulations at 40-, 10-, and 4-km horizontal resolution with a new parameterization scheme of deep convection and microphysics and 40- and 10-km horizontal resolution output from the ALADIN model with an old parameterization scheme were compared with respect to station observation data. We find for ALARO-0 at high spatial resolutions of 10 and 4 km an improvement in the spatial distribution of summer precipitation, such that the distinct local maximum at the highest elevations is well resolved by the model, a feature strongly overestimated by the ALADIN model at 10-km resolution. Furthermore, the results from the extreme value analysis suggest that the new parameterization scheme of ALARO-0 contributes to the improvement in the modeling of extreme precipitation events at varying horizontal resolutions, rather than the increase in spatial resolution. Thus, the nature of the parameterization is more important than the resolution, which confirms the previous findings of Lynn et al. (2010) and Hamdi et al. (2012). As an outlook, the ALARO-0 model will be used to compute IPCC scenarios.

*Acknowledgments.* The authors thank the anonymous reviewers for their suggestions, which greatly helped to improve this manuscript. We are very grateful to BELSPO (Belgian Science Policy) by who this work is funded (Reference IRM-act2-11-01). We also wish to acknowledge the contribution of the ALADIN Partners inside the collaboration for the development and maintenance of ALARO, mostly with the support of RC LACE and of the Czech Grant GACR P209-11-2405. The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (Contract 505539), whose support is gratefully acknowledged. All the data processing and analyses have been performed with the statistical computing program R (R Development Core Team 2011).

## REFERENCES

- ALADIN International Team, 1997: The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. *WMO Bull.*, **46**, 317–324.
- Boberg, F., P. Berg, P. Thejll, W. J. Gutowski, and J. H. Christensen, 2010: Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. *Climate Dyn.*, **35**, 1509–1520, doi:10.1007/s00382-009-0683-8.
- Bougeault, P., 1985: A simple parameterization of the large-scale effects of cumulus convection. *Mon. Wea. Rev.*, **113**, 2108–2121.
- Bubnová, R., G. Hello, P. Bénard, and J.-F. Geleyn, 1995: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/ALADIN NWP system. *Mon. Wea. Rev.*, **123**, 515–535.
- Buonomo, E., R. Jones, I. C. Huntingford, and J. Hannaford, 2007: On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations. *Quart. J. Roy. Meteor. Soc.*, **133**, 65–81, doi:10.1002/qj.13.
- Caldwell, P., H.-N. S. Chin, D. C. Bader, and G. Bala, 2009: Evaluation of a WRF dynamical downscaling simulation over California. *Climatic Change*, **95**, 499–521, doi:10.1007/s10584-009-9583-5.
- Castro, C. L., R. A. S. Pielke, and G. Leoncini, 2005: Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *J. Geophys. Res.*, **110**, D05108, doi:10.1029/2004JD004721.
- Catry, B., J.-F. Geleyn, M. Tudor, P. Bénard, and A. Trojáková, 2007: Flux-conservative thermodynamic equations in a mass-weighted framework. *Tellus*, **59A**, 71–79, doi:10.1111/j.1600-0870.2006.00212.x.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 210 pp.
- Csima, G., and A. Horányi, 2008: Validation of the ALADIN-Climate regional climate model at the Hungarian Meteorological Service. *Idojaras*, **112**, 155–177.
- Davies, H. C., 1976: A lateral boundary formulation for multilevel prediction models. *Quart. J. Roy. Meteor. Soc.*, **102**, 405–418.
- de Elía, R., R. Laprise, and D. Bertrand, 2002: Forecasting skill limits of nested, limited-area models: A perfect-model approach. *Mon. Wea. Rev.*, **130**, 2006–2023.
- Denis, B., R. Laprise, D. Caya, and J. Côté, 2002: Downscaling ability of one-way nested regional climate models: The Big-Brother Experiment. *Climate Dyn.*, **18**, 627–646, doi:10.1007/s00382-001-0201-0.
- Déqué, M., and S. Somot, 2008: Analysis of heavy precipitation for France using high resolution ALADIN RCM simulations. *Idojaras*, **112**, 179–190.
- Domínguez, M., and Coauthors, 2013: Present climate precipitation and temperature extremes over Spain from a set of high resolution RCMs. *Climate Res.*, in press.
- Dulière, V., Y. Zhang, and E. P. J. Salathé, 2011: Extreme precipitation and temperature over the U.S. Pacific Northwest: A comparison between observations, reanalysis data, and regional models. *J. Climate*, **24**, 1950–1964.
- Durman, C. F., J. M. Gregory, D. C. Hassell, R. G. Jones, and J. M. Murphy, 2001: A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Quart. J. Roy. Meteor. Soc.*, **127**, 1005–1015, doi:10.1002/qj.49712757316.
- Geleyn, J.-F., B. Catry, Y. Bouteloup, and R. Brožková, 2008: A statistical approach for sedimentation inside a microphysical precipitation scheme. *Tellus*, **60A**, 649–662, doi:10.1111/j.1600-0870.2008.00323.x.
- Gerard, L., 2007: An integrated package for subgrid convection, clouds and precipitation compatible with the meso-gamma scales. *Quart. J. Roy. Meteor. Soc.*, **133**, 711–730, doi:10.1002/qj.58.
- , and J.-F. Geleyn, 2005: Evolution of a subgrid deep convection parameterization in a limited area model with increasing resolution. *Quart. J. Roy. Meteor. Soc.*, **131**, 2293–2312, doi:10.1256/qj.04.72.
- , J.-M. Piriou, R. Brožková, J.-F. Geleyn, and D. Banciu, 2009: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model. *Mon. Wea. Rev.*, **137**, 3960–3977.

- Giorgi, F., and L. O. Mearns, 1999: Introduction to special section: Regional climate modeling revisited. *J. Geophys. Res.*, **104**, 6335–6352.
- , X. Bi, and J. S. Pal, 2004: Mean, interannual variability and trends in a regional climate change experiment over Europe. I. Present-day climate (1961–1990). *Climate Dyn.*, **22**, 733–756, doi:10.1007/s00382-004-0409-x.
- Hamdi, R., H. Van de Vyver, and P. Termonia, 2012: New cloud and microphysics parameterisation for use in high-resolution dynamical downscaling: Application for summer extreme temperature over Belgium. *Int. J. Climatol.*, **32**, 2051–2065, doi:10.1002/joc.2409.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:10.1029/2008JD010201.
- Heikkilä, U., A. Sandvik, and A. Sorterberg, 2011: Dynamical downscaling of ERA-40 in complex terrain using WRF regional climate model. *Climate Dyn.*, **37**, 1551–1564, doi:10.1007/s00382-010-0928-6.
- Herrera, S., L. Fita, J. Fernández, and J. M. Gutiérrez, 2010: Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. *J. Geophys. Res.*, **115**, D21117, doi:10.1029/2010JD013936.
- Hofstra, N., M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dyn.*, **35**, 841–858, doi:10.1007/s00382-009-0698-1.
- Jacob, D., and Coauthors, 2007: An inter-comparison of regional climate models for Europe: Model performance in present-day climate. *Climatic Change*, **81**, 31–52, doi:10.1007/s10584-006-9213-4.
- Jones, P. D., and P. A. Reid, 2001: Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *Int. J. Climatol.*, **21**, 1337–1356, doi:10.1002/joc.677.
- Jones, R. G., J. M. Murphy, and M. Noguer, 1995: Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Quart. J. Roy. Meteor. Soc.*, **121**, 1413–1449, doi:10.1002/qj.49712152610.
- , M. Noguer, D. C. Hassell, D. Hudson, S. S. Wilson, G. J. Jenkins, and J. F. B. Mitchell, 2004: Generating high resolution climate change scenarios using PRECIS. National Communications Support Unit Workbook, Met Office Hadley Centre, Exeter, United Kingdom, 32 pp. [Available online at <http://www.uncclern.org/sites/www.uncclern.org/files/inventory/UNDP17.pdf>.]
- Kyselý, J., and R. Beranová, 2009: Climate-change effects on extreme precipitation in central Europe: Uncertainties of scenarios based on regional climate models. *Theor. Appl. Climatol.*, **95**, 361–374, doi:10.1007/s00704-008-0014-8.
- Lucas-Picher, P., F. Boberg, J. Christensen, and P. Berg, 2013: Dynamical downscaling with reinitializations: A method to generate fine-scale climate datasets suitable for impact studies. *J. Hydrometeorol.*, **14**, 1159–1174.
- Lynn, B. H., and Coauthors, 2010: Testing GISS-MM5 physics configurations for use in regional impacts studies. *Climatic Change*, **99**, 567–587, doi:10.1007/s10584-009-9729-5.
- Ma, L., T. Zhang, O. W. Frauenfeld, B. Ye, D. Yang, D. Yang, and D. Qin, 2009: Evaluation of precipitation from the ERA-40, NCEP-1, and NCEP-2 Reanalyses and CMAP-1, CMAP-2, and GPCP-2 with ground-based measurements in China. *J. Geophys. Res.*, **114**, D09105, doi:10.1029/2008JD011178.
- Mailhot, A., S. Lachance-Cloutier, G. Talbot, and A.-C. Favre, 2013: Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach. *J. Hydrol.*, **476**, 188–199, doi:10.1016/j.jhydrol.2012.10.036.
- Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712, doi:10.1002/joc.1181.
- Nicolis, C., 2003: Dynamics of model error: Some generic features. *J. Atmos. Sci.*, **60**, 2208–2218.
- , 2004: Dynamics of model error: The role of unresolved scales revisited. *J. Atmos. Sci.*, **61**, 1740–1753.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.
- Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate*, **20**, 4356–4376.
- Qian, J.-H., A. Seth, and S. Zebiak, 2003: Reinitialized versus continuous simulations for regional climate downscaling. *Mon. Wea. Rev.*, **131**, 2857–2874.
- R Development Core Team, cited 2011: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. [Available online at <http://www.R-project.org/>.]
- Radnóti, G., 1995: Comments on “A spectral limited-area formulation with time-dependent boundary conditions applied to the shallow-water equations.” *Mon. Wea. Rev.*, **123**, 3122–3123.
- Radu, R., M. Déqué, and S. Somot, 2008: Spectral nudging in a spectral regional climate model. *Tellus*, **60A**, 898–910, doi:10.1111/j.1600-0870.2008.00341.x.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.
- Ribatet, M., T. B. M. J. Ouarda, E. Sauquet, and J.-M. Gressillon, 2009: Modeling all exceedances above a threshold using an external dependence structure: Inferences on several flood characteristics. *Water Resour. Res.*, **45**, W03407, doi:10.1029/2007WR006322.
- Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.*, **120**, 303–325.
- Roth, M., T. A. Buishand, G. Jongbloed, A. M. G. K. Tank, and J. H. van Zanten, 2012: A regional peaks-over-threshold model in a nonstationary climate. *Water Resour. Res.*, **48**, W11533, doi:10.1029/2012WR012214.
- Skalák, P., P. Stepánek, and A. Farda, 2008: Validation of ALADIN-Climate/CZ for present climate (1961–1990) over the Czech Republic. *Idojaras*, **112**, 191–201.
- Soares, P. M. M., R. M. Cardoso, P. M. A. Miranda, J. Medeiros, M. Belo-Pereira, and F. Espirito-Santo, 2012a: WRF high resolution dynamical downscaling of ERA-Interim for Portugal. *Climate Dyn.*, **39**, 2497–2522, doi:10.1007/s00382-012-1315-2.
- , —, —, P. Viterbo, and M. Belo-Pereira, 2012b: Assessment of the ENSEMBLES regional climate models in the representation of precipitation variability and extremes over Portugal. *J. Geophys. Res.*, **117**, D07114, doi:10.1029/2011JD016768.
- Stephenson, A., 2002: evd: Extreme value distributions. *R News*, Vol. 2 (2), R Project, 31–32. [Available online at [http://cran.r-project.org/doc/Rnews/Rnews\\_2002-2.pdf](http://cran.r-project.org/doc/Rnews/Rnews_2002-2.pdf).]

- Sylla, M. B., E. Coppola, L. Mariotti, F. Giorgi, P. M. Ruti, A. Dell'Aquila, and X. Bi, 2010: Multiyear simulation of the African climate using a regional climate model (RegCM3) with the high resolution ERA-interim reanalysis. *Climate Dyn.*, **35**, 231–247, doi:10.1007/s00382-009-0613-9.
- Termonia, P., A. Deckmyn, and R. Hamdi, 2009: Study of the lateral boundary condition temporal resolution problem and a proposed solution by means of boundary error restarts. *Mon. Wea. Rev.*, **137**, 3551–3566.
- , F. Voitus, D. Degrauwe, S. Caluwaerts, and R. Hamdi, 2012: Application of Boyd's periodization and relaxation method in a spectral atmospheric limited-area model. Part I: Implementation and reproducibility tests. *Mon. Wea. Rev.*, **140**, 3137–3148.
- Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.*, **106**, 11 775–11 784.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.
- Váňa, F., P. Bénard, J.-F. Geleyn, A. Simon, and Y. Seity, 2008: Semi-Lagrangian advection scheme with controlled damping: An alternative to nonlinear horizontal diffusion in a numerical weather prediction model. *Quart. J. Roy. Meteor. Soc.*, **134**, 523–537, doi:10.1002/qj.220.
- Warner, T., R. Peterson, and R. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.
- Zolina, O., A. Kapala, C. Simmer, and S. K. Gulev, 2004: Analysis of extreme precipitation over Europe from different re-analyses: A comparative assessment. *Global Planet. Change*, **44**, 129–161, doi:10.1016/j.gloplacha.2004.06.009.