Replies to referees for

Giot, O., Termonia, P., Degrauwe, D., De Troch, R., Caluwaerts, S., Smet, G., Berckmans, J., Deckmyn, A., De Cruz, L., De Meutter, P., Duerinckx, A., Gerard, L., Hamdi, R., Van den Bergh, J., Van Ginderachter, M., and Van Schaeybroeck, B.: Validation of the ALARO-0 model within the EURO-CORDEX framework, Geosci. Model Dev. Discuss., 8, 8387-8409, doi:10.5194/gmdd-8-8387-2015, 2015.

We would like to thank the two reviewers for their very positive comments and useful suggestions that have now improved the manuscript and have acknowledged their efforts in the manuscript.

A manuscript version with clear track changes is added below and line references **[l. linenumber]** refer to lines in this version.

Anonymous Referee #1

Received and published: 30 November 2015

Giot et al. present a short and concise work on the performance assessment of the ALARO-0 regional climate model operated according to the EURO-CORDEX experimental protocol at two horizontal resolutions of about 12.5 and 50 km and driven by the ERA-Interim reanalysis for the period 1979-2010. Model results are validated against the gridded EOBS reference and the obtained performance indicators are related to the work of Kotlarski et al. (2014) who evaluated a larger model ensemble of the EURO-CORDEX framework. The analysis indicates a reasonable performance of ALARO-0 which is comparable to the performance of most other EURO-CORDEX RCMs. Temperature biases are similar as those of the related ARPEGE model.

The presented work is of interest mainly for the ALADIN/ALARO community as it serves as a basic and rather technical reference for the performance of the newly developed ALARO-0 model. But also the wider EURO-CORDEX community is definitely interested in this work assessing the quality of a new ensemble member. As such, I consider the manuscript as relevant for a wider scientific community and, in general, suitable for publication in GMD. The paper is concise; methods, models and data are for most parts appropriately introduced. The content and quality of the figures are mostly appropriate as well. The conclusions are well justified by the results obtained. There are no language issues. Therefore, I could recommend a publication of this work after some minor issues (listed below) have been addressed.

With kind regards.

Remaining minor issues:

p 8389, l 10: I'd not speak of "feed-backs" here. Such feedbacks wouldn't be addressed by one-way coupled downstream models.

We agree and have changed [l. 31] "feed-backs" to "interactions" in the text.

p 8389, l 13-19: I'd suggest to mention here that also empirical-statistical downscaling is part of (EURO-)CORDEX.

Agreed, we have added the lines [l . 35]:

"The Coordinated Regional Climate Downscaling Experiment (CORDEX; Giorgi et al., 2009) aims to perform both empirical-statistical downscaling and regional climate simulations on different areas across the globe."

p 8389, l 23: "Limited Area Models" instead of "Local Area Models".

Thank you, we have changed this [l. 44].

p 8390, l 13-14: The term "scale awareness" remains obscure here. The following sentences are somehow related to it, but don't provide a clear picture. Can the authors

better specify what is meant here?

The details of the 'scale-awareness' are presented in De Troch et al., 2013 and Gerard et al., 2009. We have changed this part to [l. 62-66]:

"The main feature of 3MT is scale-awareness, i.e. the parameterization itself determines which processes are unresolved at the current resolution, in contrast to traditional parameterizations which are switched on or off or have different tuned parameter values at different resolutions. This allows 3MT to generate consistent results across scales, as shown by De Troch et al. (2013) in an extended downscaling experiment covering the period from 1961 to 1990."

p 8390, l 27-30: Unclear. What's the meaning of "uninterrupted" here?

The runs performed by De Troch et al. were re-initialized daily, i.e. the model fields were reset to the ERA-Interim values every 24 hours, followed by a 36-hour of which the last 24 hours were used for analysis. Another way to look at this, is that a 36-hour weather forecast was performed for every day in the ERA-Interim period, with ERA-Interim as initial and boundary conditions. The setup by De Troch et al. was chosen to capture afternoon summer convection for several model resolutions. By contrast, for the current study, initial conditions were taken only once from ERA-Interim (1st of January 1979) and then the simulation was only forced at the boundaries. To clarify this in the text we have added [l. 78-82]: "The model setup differs from the setup used in De Troch et al. (2013), since in the current study simulations are initialized on the 1st of January 1979, after which they are only forced at the boundaries by ERA-Interim. This allows the model and its surface fields in particular to become independent of the initial state."

p 8391, l 7: "The objective of the present work" instead of "The goal of the current text".

Thank you, we have changed this accordingly [l. 88].

p 8391, l 21: "were analyzed" instead of "are performed" (K14 only analyzed the ensemble results, but didn't carry out all the simulations).

Thank you, we have changed this accordingly [l. 102].

p 8392, l 8-28: The difference between ALADIN and ALARO-0 is not entirely clear to me. Please better clarify this aspect.

We added some clarifying sentences [l. 115-117]: "Essentially, ALARO-0 uses the dynamical core of ALADIN, but with different physics routines (e.g. for radiation, microphysics and convection, cloudiness, turbulence), which are designed to tackle the issues that arise when using resolutions of 1-15 km, which is known as the grey-zone for convection."

p 8392, l 8-24: The treatment of SSTs is not clear. According to my understanding of the current text, SSTs are only updated monthly. Is this really true? Furthermore, the authors speak of "interrupted" simulations, while before (page 8390) the present experiments were introduced as "uninterrupted" simulations. There seems to be some mismatch. Concerning the constant monthly fields (roughness length etc.): Are these sharply changed when reaching a new month (which I guess is not the case), or are they interpolated between the

centers-of-months?

Yes, indeed, SSTs are updated sharply every month. The reason for this is that ALARO-0 has been developed as a NWP model, for which over the course of a few days it is common practice to keep SSTs constant, especially for a domain of which only a small area consists of ocean. Therefore, technically changing SSTs during the simulations is not (yet) possible. To grasp at least the seasonal cycle of SSTs, runs were "interrupted", i.e. stopped and restarted with adjusted SSTs and some other climatological fields (mainly fields that are related to the yearly cycle of vegetation such as LAI, surface albedo and roughness lengths). However, all other (prognostic) fields are unchanged. As such, only parameters of the surface scheme and SSTs change instantly at the beginning of the month. We acknowledge this practice is not optimal. Indeed, interpolation between centers-of-months would be a first step in order to avoid introducing sharp changes. This is planned to be implemented in a new version. We however believe that the sharp changes introduced in this way do not lead to major issues or feed-back into climatological fields.

Together with the previous remark about the usage of "uninterrupted", we see that there is an inconsistent usage of the word throughout the manuscript. Therefore we have removed "uninterrupted" from the text and replaced it by 'a single initialization of all fields', as shown above for (p 8390, l 27-30) and also in the conclusion (p 8400, l 11-13) [l. 320]: "In this study, for the first time ever the ALARO-0 model was used to perform continuous climate simulations on a European scale for a 32-year period."

p 8393, l 11: Which version of EOBS has been used?

As in Kotlarski et al. version 7 was used. We have added this information in the text [l. 145].

p 8394, l 1: I'd suggest to rename this Section to "Analysis methods".

Agreed, we have changed this [l. 160].

p 8394, l 17: "for this purpose" instead of "for this end".

Thank you, this has been corrected accordingly.

p 8397, l 1-6: This results is very interesting (similar for precipitation later on). Do the authors have any explanation for the large confidence intervals for these scores?

The reason is given in the discussion section: p 8400 l4-5: "This does not hold for some RIAV and most of the TCOIAV scores due to the fact that these exactly assess interannual variability". Additionally, these scores are based on a sample of only 20.

p 8399, l 3: "bias patterns" instead of "bias pattern".

Thank you, this has been corrected accordingly [l. 285].

p 8399, l 6-7: Could the these low correlations partly be explained by the comparatively large model domain of ALARO-0 (weaker control of boundary forcing)?

Yes, this very possible. We have added this suggestion in the text [l. 288-289]: "Both spatial and temporal variability are very well reproduced by ALARO-0, while correlations are on the low side compared to other models. The latter could partly be explained by the comparatively larger domain of ALARO-0 which could imply a weaker control of the boundary forcing."

p 8400, l 13: "Within the framework".

Thank you, this has been corrected accordingly [l. 321].

Figures 2 and 4: The caption of these figures should additionally mention that RMIB-11 is shown.

Indeed, we have added this to the caption.

Figures 3 and 5: These figures need to be enlarged, there's a lot of detail here which is not really accessible. A legend should be introduced (meaning of markers and shadings). I'd also suggest to add a horizontal line above each "DJF" entry to better separate the individual regions from each other.

These figures were produced and submitted as a vector pdf and can therefore be enlarged without loss of quality. We will request the editor and typesetter to enlarge these figures as much as possible for publication (if possible put them on a separate full page and rotated 90 degrees). We have added a legend horizontal lines, which indeed allows for a better overview. See figures.

Availability of data: The authors should provide the information, if and where the ALARO-0 simulation results are available. Are they planned to be uploaded to the ESGF archive?

Yes, the uploading to ESGF is planned. We have added this in the last lines of the "Setup of the ALARO-0 model" section of the manuscript [l. 142-143]: "This model data will be uploaded to the Earth System Grid Federation (ESGF, website: esgf.llnl.gov/) data nodes."

Anonymous Referee #2

Received and published: 8 December 2015

This manuscript, as its title says, deals with model validation. This means here that temperature and precipitation climatologies of ALARO model are compared with an observed climatology. In addition, the differences between the two climatologies (aka model biases or systematic errors) are compared with differences obtained with other models in a similar exercise (EURO-CORDEX ERA-interim driven). One of these models, ARPEGE, shares with ALARO the same dynamical core (but uses a different grid, a different driving method, and different physical parameterizations). The evaluation is performed for different seasons and areas in Europe. It shows that ALARO is a state-of-the-art model. The manuscript is clear and corresponds to what one would expect from such a study. I should be accepted after a few minor adds or fixes:

1. p 8389 line 27 better predicted: What is improved? The chronology or the intensity? Getting a better chronology is not useful for climate application.

Indeed, chronology is not useful for climate applications, but intensity is. Although one could argue that a systematic error in chronology will eventually also result in a error in the intensity as the two are obviously interconnected through the physical processes involved. But in the end one does not expect climate models to get the timing right, but the statistics, which is the pdf of the variable at hand.
We have changed the manuscript to further clarify this and added other relevant effects (diurnal cycle, convection onset, less drizzle) [l. 44-53]: "Nowadays, NWP Limited Area Models (LAMs) are designed for resolutions down to a few kilometres, with adapted physics parametrisation schemes. At even higher resolutions, these models can (partly) resolve clouds and convective systems. Since a correct treatment of the cloud feedback is of critical importance for climate modelling (e.g., Sun et al., 2009; Lin et al., 2014), some of these NWP models have been used in climate mode: studies by De Meutter et. al 2015, Hohenegger et al., 2008, Kendon et al., 2012 and Chan et al., 2014, where models with resolution at the kilometer scale are used without convection parameterization, show a better representation of the intensity of extreme precipitation, the diurnal cycle, afternoon convection onset and less drizzle."

2. p 8393 lines 16-22: Explain why you interpolate differently temperature and precipitation. Is it to save the precipitation extremes?

Temperature and precipitation are interpolated in the same manner. First, from the ALARO-0 grid to the EUR-11 grid, using the closest grid point value. For temperature, here an additional height correction of 0.0064 K/m * (h_E – h_A) was applied, with h_A the ALARO_0 grid point height and h_E the EUR-11 closest grid point height. Then from EUR-11 to the E-OBS (.22°) grid, 2x2 grid box averages were taken for both precipitation and temperature. This method was adopted from K14 in order to be able to compare to those results.

We have changed the text to clarify this [l. 150-156]: "For the high-resolution simulations, first the values of the closest grid point were taken to go from the native Lambert ALARO-0 grid to the EUR-11 grid for both precipitation and temperature. For the latter, an additional height difference correction between the ALARO-0 and closest EUR-11 grid point was performed using the standard climatological lapse rate of 0.0064 K/m. Second, on this grid, for both precipitation and temperature two-by-two grid box averages were calculated to obtain an identical grid to the E-OBS dataset."

3. p 8394 lines 9-11: does it means that CRCO and ROYA are constant whatever the model ?

Not constant, but very similar. We refer to figures 13 and 14 in K14.

CRCO (climatological rank correlation) measures if the yearly cycle is captured or not, it is the normalized difference in the 12 ranked area averaged monthly means. For temperature this gives a value very close to 1 for all models, since this is strongly correlated with the yearly cycle of solar forcing. For precipitation the score very much depends on the (in)existence of a clear yearly cycle. When a certain region has a clear annual cycle, the score will be higher. And therefore it does not clearly measure model deficiency.

ROYA (ratio of yearly amplitudes) is the ratio of the amplitude of the yearly cycle based on monthly values (model max- model min)/(observed max – observed min). This score can effectively be read from the BIAS column in figures 3 for temperature. If JJA is cold (warm) biased and DJF warm (cold) biased, this score will be lower (higher) than 1.

For ALARO-0 we found the behavior of the scores to be similar to other models. Therefore, the additional information to K14 that would be provided by discussing these scores would seem minor and we chose not to include them for conciseness.

4. p 8396 last sentence: In fact 20 year is not a short period for such an exercise. When comparing two climate simulations which include interannual variability, even 30 year is short to draw conclusions. But here all simulations and observations follow the same chronology because of the common driving by ERA-interim. So, the signal is not blurred out by the noise of the interannual variability. This is why EURO-CORDEX is limited to the core period of ERA-interim, i.e. 1989-2008.

Indeed, but still the scores could be dependent on the state of the system. For example, a warm period could have a warm bias, while a cold period could have a cold bias. Therefore, it is still relevant to study if the provided scores for the different models are statistically significantly different from one another. Model scores could differ given a different period and model ranking, for example, would not be robust.

5. p 8398 lines 10-12: Indeed ALARO is a new comer in this community. But one should stress that the RCM community has trained his models with a 50 km resolution. A large EURO-CORDEX domain at 12 km resolution was a first attempt for most models, because of the computer cost. I do not believe that ALARO is compared with highly tuned climate models.

We agree and have changed this paragraph [l. 161-274]:

“This is the first time ALARO-0 was used for a climate experiment. Nevertheless, the performance of ALARO-0 on seasonal and yearly scales for both near-surface air temperature and precipitation is satisfactory. Generally ALARO-0 performs well, which is quantified by the large number of white boxes in Figs. 3 and 5 indicating that the ALARO-0 score lies within the existing K14 ensemble. For precipitation, ALARO-0 even outperforms all other models on numerous occasions. These results are encouraging, given that ALARO-0 does not yet have the experience in climate modelling that some of the other models of the K14 ensemble had, but was directly ported from its NWP setup. Although the 12.5-km resolution was also a novelty for the K14 models, their performance undoubtedly benefited from previous optimizations for climate experiments, albeit at a lower resolution of 50 km.”

Manuscript prepared for Geosci. Model Dev.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 February 2016

Validation of the ALARO-0 model within the EURO-CORDEX framework

Olivier Giot[1,2], Piet Termonia[1,3], Daan Degrauwe[1], Rozemien De Troch[1,3], Steven Caluwaerts[1,3], Geert Smet[1], Julie Berckmans[1,2], Alex Deckmyn[1], Lesley De Cruz[1], Pieter De Meutter[1,3], Annelies Duerinckx[1,3], Luc Gerard[1], Rafiq Hamdi[1], Joris Van den Bergh[1], Michiel Van Ginderachter[1,3], and Bert Van Schaeybroeck[1]

[1]Royal Meteorological Institute, Brussels, Belgium
[2]Centre of Excellence PLECO (Plant and Vegetation Ecology), Department of Biology, University of Antwerp, Universiteitsplein 1, B-2610, Wilrijk, Belgium
[3]Department of Physics and Astronomy, Ghent University, Ghent, Belgium

Correspondence to: Olivier Giot (olivier.giot@meteo.be)

Abstract. Using the regional climate model ALARO-0, the Royal Meteorological Institute of Belgium ~~has~~ and Ghent University have performed two simulations of the past observed climate within the framework of the Coordinated Regional Climate Downscaling Experiment (CORDEX). The ERA-Interim reanalysis was used to drive the model for the period 1979-2010 on the EURO-CORDEX domain with two horizontal resolutions, .11 and .44 degrees. ALARO-0 is characterised by the new microphysics scheme 3MT, which allows for a better representation of convective precipitation. In Kotlarski et al. (2014) several metrics assessing the performance in representing seasonal mean near-surface air temperature and precipitation are defined and the corresponding scores are calculated for an ensemble of models for different regions and seasons for the period 1989-2008. Of special interest within this ensemble is the ARPEGE model by the Centre National de Recherches Météorologiques (CNRM), which shares a large amount of core code with ALARO-0.

Results show that ALARO-0 is capable of representing the European climate in an acceptable way as most of the ALARO-0 scores lie within the existing ensemble. However, for near-surface air temperature some large biases, which are often also found in the ARPEGE results, persist. For precipitation, on the other hand, the ALARO-0 model produces some of the best scores within the ensemble and no clear resemblance to ARPEGE is found, which is attributed to the inclusion of 3MT.

Additionally, a jackknife procedure is applied to the ALARO-0 results in order to test whether the scores are robust, by which we mean independent of the period used to calculate them. Periods of 20 years are sampled from the 32-year simulation and used to construct the 95% confidence interval

for each score. For most scores these intervals are very small compared to the total ensemble spread, implying that model differences in the scores are significant.

1 Introduction

The climate projections used in the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC, 2013) are based on the set of Global Climate Model (GCM) simulations performed within the fifth Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2011). The horizontal resolution of the contributing GCMs is limited to typically 1° to 2° by computational constraints. For many local climate impact studies Regional Climate Models (RCMs; Giorgi and Mearns, 1999) are needed to reveal the fine-scale details of potential climate change (Teutschbein and Seibert, 2010). In addition, specific downstream models which simulate processes such as vegetation ~~feed-backs~~interactions, urban effects (e.g., Hamdi et al., 2015) or extreme hydrological events in river catchments often require high-resolution (both in time and space) forcing data from atmospheric models.

The Coordinated Regional Climate Downscaling Experiment (CORDEX; Giorgi et al., 2009) aims to perform both empirical-statistical downscaling and regional climate simulations on different areas across the globe using an ensemble of RCMs. By prescribing several integration domains and resolutions a direct quantitative comparison between the participating models' performances and projections is feasible. The domain of interest ~~here~~in this study, is the EURO-CORDEX domain shown in Figure 1 (inner orange box). Several RCM groups have performed simulations on this domain with horizontal resolutions of both .11 and .44 degrees.

All RCMs have a history in Numerical Weather Prediction (NWP) and often consist of a modified NWP code which is further developed separately from or parallel to the NWP code, borrowing for example its dynamical core but using different physics parametrisations or surface schemes (Dudhia, 2014). Nowadays, NWP ~~Local~~ Limited Area Models (LAMs) are designed for resolutions ~~up~~ down to a few kilometres, with adapted physics parametrisation schemes. At even higher resolutions, these models can (partly) resolve clouds and convective systems~~and evidence has been presented that in this case extreme precipitation events are better predicted (e.g., De Meutter et al., 2015)~~. Since a correct treatment of the cloud feedback is of critical importance for climate modelling (e.g., Sun et al., 2009; Lin et al., 2014), some of these NWP models have been used in climate ~~studies (Hohenegger et al., 2008; Kendon et al., 2012; Chan et al., 2014)~~ mode: studies by De Meutter et al. (2015) , Hohenegger et al. (2008) , Kendon et al. (2012) and Chan et al. (2014) , where models with resolution at the kilometer scale are used without convection parameterization, show a better representation of the intensity of extreme precipitation, the diurnal cycle, afternoon convection onset and less drizzle. For instance, ALADIN-CLIMATE of the Centre National de Recherches Météorologiques (CNRM;

Spiridonov et al., 2005) is a climate version of the ALADIN limited area model that has been developed in the context of the international ALADIN consortium (ALADIN international team, 1997).

Over the past decade, within the context of the ALADIN consortium, a physics parametrisation scheme called 3MT (Modular Multiscale Microphysics and Transport) has been developed and used as the central feature of a new NWP model, ALARO-0 (Gerard and Geleyn, 2005; Gerard, 2007; Gerard et al., 2009). It is based on a parametrisation of deep convection and optimally adapted to be used at resolutions in the so-called ~~grey~~grey-zone. Several countries have used and tested the model for operational weather forecasting and regional climate studies. ~~Its main feature~~ The main feature of 3MT is scale-awareness, ~~which was~~ i.e. the parameterization itself works out which processes are unresolved at the current resolution, in contrast to traditional parameterizations which are switched on or off or have different tuned parameter values at different resolutions. This allows 3MT to generate consistent results across scales, as shown by De Troch et al. (2013) in an extended downscaling experiment covering the period from 1961 to 1990. ~~For~~ In their study, for every day, short-term runs were performed at different horizontal resolutions between 40 km and 4 km. Both the initial and lateral boundary conditions were provided by either the ERA-40 reanalysis (Uppala et al., 2005) or model runs at lower resolution in a double nesting procedure. Given the large amount of required computing resources for such a simulation, this type of validation is rather unusual for NWP models. The results showed that extreme precipitation values are correctly and consistently reproduced for all horizontal resolutions by a model version including 3MT, ~~but that~~ whereas extreme precipitation was progressively overestimated when increasing the resolution by a model version without 3MT.

In the present study the ALARO-0 model has been used to perform the EURO-CORDEX validation simulations, i.e. the ERA-Interim reanalysis (Dee et al., 2011) is used as lateral boundary conditions allowing for a direct comparison to observations. The model setup differs from the setup used in De Troch et al. (2013), since in the current study simulations are ~~uninterrupted to allow~~ initialized on the 1st of January 1979, after which they are only forced at the boundaries by ERA-Interim. This allows the model and ~~more specifically~~ its surface fields ~~to find their equilibrium~~ in particular to become independent of the initial state. Results are then compared to an ensemble of 17 other EURO-CORDEX experiments which have been evaluated in (Kotlarski et al., 2014, which we will refer to as K14 from now on). In K14, seasonal means of near-surface air temperature and precipitation amounts are compared to observations using several metrics which quantify the spatiotemporal performance of the ensemble. In their article, Kotlarski et al. evaluate the rather short 20-year period 1989-2008, while for this study the 32-year period 1979-2010 was simulated.

The ~~goal of the current text~~ objective of the present work is (1) to quantify the performance of the ALARO-0 model within the existing K14 ensemble and (2) to assess the robustness of the calculated scores given the rather short 20-year period used in K14.

This paper is organised as follows. In Section, 2 the existing K14 ensemble, details on the setup of ALARO-0 and the methods used to attain the goals of this paper are discussed. In Section 3, results are presented for ALARO-0 and compared to the K14 ensemble, followed by a discussion in Section 4. Finally, in Section 5, we come back to the goals that were set, formulate conclusions and present an outlook.

2 Data and methods

K14 ensemble

The CORDEX community prescribes two European integration grids which differ only in resolution. The low-resolution EUR-44 domain's grid points are .44 degrees apart on a rotated lat-lon grid limited to Europe (see inner orange box in Fig. 1, 106x103 grid boxes). For the high-resolution EUR-11 experiment each EUR-44 grid box is divided into sixteen .11 degrees-wide grid boxes. In K14, a total of 17 experiments ~~are performed~~ were analyzed by 9 different research groups. Eight groups performed both the EUR-11 and EUR-44 simulations, one group only EUR-11, and three groups used the same model (WRF) but with different physics parametrisations. All models are forced directly by ERA-Interim except for the experiment performed by CNRM. This group set up the global model ARPEGE (version 5.1) to be strongly nudged towards ERA-Interim outside of the CORDEX domain, but allowed the model to evolve freely inside of it. Further details on all models can be found in Table 1 of K14.

The main conclusions of K14 were that the higher resolution simulations did not perform significantly better and the models in the ensemble generally had a cold and wet bias, except for summers in Southern Europe which are commonly warm and dry biased.

Setup of the ALARO-0 model

The ALARO-0 model used for this study is the identical configuration of the ALADIN system (ALADIN international team, 1997) described in detail and validated by De Troch et al. (2013). Essentially, ALARO-0 uses the dynamical core of ALADIN, but with different physics routines (e.g. for radiation, microphysics and convection, cloudiness, turbulence), which are designed to tackle the issues that arise when using resolutions of 1-15 km, which is known as the grey-zone for convection. Here, we only describe the EURO-CORDEX specific setup of the model, which is the coupling to the boundary conditions and the definition of the integration grids.

Similar to all other models in K14 (except for the global CNRM model), ALARO-0 is coupled to ERA-Interim by the classical Davies procedure (Davies, 1976). The relaxation zone consists of 8 grid points irrespective of resolution, and new boundary conditions are provided every six hours. No further nudging or relaxation towards the boundary conditions was done inside of the domain. Some fields in ALARO-0 are constant during run-time, most notably Sea Surface Temperatures (SSTs).

Simulations are however interrupted and restarted monthly to allow for SSTs to be updated. Other fields that have monthly updates, but are constant during any given month are surface roughness length, surface emissivity, surface albedo and vegetation parameters. All other variables were computed continuously from 1 January 1979 to 31 December 2010 and thus, in contrast to De Troch et al. (2013), no daily restarts were done.

It would preferable to use the exact rotated lat-lon grids defined by the CORDEX community for the simulations. However, ALARO-0 does not support this projection but instead uses a conformal Lambert projection. Following the CORDEX guidelines, two new grids with a 12.5 km and 50 km resolution were defined for the ALARO-0 simulations. Figure 1 shows the bounding boxes of the low-resolution (full green lines) and high-resolution (dashed green lines) ALARO-0 Lambert domains. The outer boxes show the complete domain, while the inner boxes exclude the relaxation zone. The grids were chosen such that the common EURO-CORDEX analysis domain (inner orange box in Fig. 1) is completely included in the non-coupling zone. The low-resolution Lambert domain consists of 139-by-139 grid points, while the high resolution domain consists of 501-by-501 grid points (both including 8 coupling grid points at every boundary). In both simulations the number of vertical levels was 46. Following K14, we will refer to the results with the acronym of the institute performing the simulations, yielding ~~RMIB-11 and RMIB-44~~RMIB-UGent-11 and RMIB-UGent-44, for the high- and low-resolution simulations respectively. This model data will be uploaded to the Earth System Grid Federation (ESGF, website: esgf.llnl.gov/) data nodes.

Data

As observational reference set, the E-OBS dataset version 7 was used (Haylock et al., 2008). The E-OBS dataset has a .22° rotated lat-lon version (outer orange box in Fig. 1) which encompasses the complete EURO-CORDEX domain. In the overlapping area, each E-OBS grid box contains four grid boxes of the EUR-11 domain and by consequence each EUR-44 box contains four E-OBS boxes.

In order to effectively compare model and observations, both need to share a common grid. The same approach as in K14 was taken to interpolate all data to a common grid. For the high-resolution simulations, first the values of the closest grid point were taken to go from the native Lambert ALARO-0 grid to the EUR-11 grid ~~. For temperature, a height correction~~ for both precipitation and temperature. For the latter, an additional height difference correction between the ALARO-0 and closest EUR-11 grid point was performed using the standard climatological lapse rate of 0.0064 K/m. Second, on this grid, for both precipitation and temperature two-by-two grid box averages were calculated to obtain an identical grid to the E-OBS dataset.

For the low-resolution simulations, again a closest grid point mapping from the native grid to the EUR-44 grid and temperature-height correction was performed. Then, the E-OBS dataset was averaged over two-by-two grid boxes that are in every EUR-44 grid box and used as reference.

Analysis methods

In K14, model performance is quantified for several metrics in different regions and seasons based on seasonal mean values of near-surface air temperature (or simply temperature from now on) and precipitation. All considered regions and their acronyms are shown in Fig. 1 and details on the definition of the different metrics can be found in K14, more specifically in Appendix A. Here, we only consider mean bias (BIAS), 95th percentile of the absolute grid point differences (95 %-P), ratio of spatial variability (RSV), pattern correlation (PACO), ratio of interannual variability (RIAV) and temporal correlation of interannual variability (TCOIAV). The climatological rank correlation (CRCO) and ratio of yearly amplitudes (ROYA) were not considered here, since these metrics showed very similar performance for all other models.

All scores in K14 are calculated based on the 20-year period 1989-2008 and they therefore state that the 'short evaluation period, leading to a sample size of only 20 seasonal/annual means, also hampers a sound analysis of statistical robustness'. The 32-year long integration period of ALARO-0 allows us to quantify the robustness of the scores by calculating how they change for a different analysis period. A jackknife procedure was applied for this ~~end~~purpose: let $\mathcal{I} = \{1979,\ldots,2010\}$ be the set of 32 years for which the ALARO-0 simulations were performed and I a random subset of length 20 of $\mathcal{I}$. We write the score for the metric s for a certain subregion j and season k based on the set of years I as

$$s_{jk}(I)$$

with $j \in$ {BI, IP, FR, ME, SC, AL, MD, EA}, $k \in$ {DJF, MAM, JJA, SON, YEAR}. For example, in K14, values for s_{jk} are calculated based on $I_{K14} = \{1989,\ldots,2008\}$. To study the robustness of s_{jk} we study the distribution of $s_{jk}(I)$ for all possible I. The number of possible 20-year subsets from 32 years without repetition and ordering is given by the binomial coefficient: 32!/(20!(32-20)!) = 225792840. It is however not feasible to perform the calculations for all possible combinations and therefore only 1000 random sequences were chosen. The width of the 95% confidence interval, limited by the 25th and 975th value of the ordered series of s_{jk}, then quantifies the robustness of the score.

3 Results

3.1 Temperature

Figure 2 shows the spatial distribution of the daily mean temperature ~~RMIB-11~~ RMIB-UGent-11 BIAS in winter (DJF, left) and summer (JJA, right) for the years in I_{K14}. Compared to Fig. 2 from K14, the spatial bias of ~~RMIB-11~~ RMIB-UGent-11 in winter looks very similar to CNRM-11. Both models show a general cold bias in Southern Europe, a warm bias in North-Eastern Europe and a large east-west bias gradient linked to orography in Scandinavia. Compared to CNRM-11, the cold

biases in mountainous regions are smaller for ~~RMIB-11~~RMIB-UGent-11. In summer, again CNRM-11 and ~~RMIB-11~~ RMIB-UGent-11 share some biases although the difference is larger than in winter and again the orographic forcing of the bias of CNRM-11 is more pronounced. Generally we find a cold bias, except in Southern Europe where a warm bias is present.

Figure 3 shows all metrics in separate columns for all different domains and seasons for seasonal and yearly mean temperature. The scale is shown at the bottom of each column, the full grey line shows the 'optimal' score of the metric (0 K for BIAS and 95%-P, 1 for all others). The grey circles show the scores for the high-resolution K14 ensemble (9 models). For each season and region two transparent red bands are superimposed, which show the jackknife 95% confidence interval for the high-resolution (top band) and low-resolution (bottom band) simulations with ALARO-0. The vertical red dashes show the value of $s_{jk}(I_{K14})$, again for the high-resolution (top) and low-resolution (bottom) simulation. When the background colour is white, the ~~RMIB-11~~ RMIB-UGent-11 value of $s_{jk}(I_{K14})$ lies within the K14 high-resolution ensemble spread. If the background colour is yellow, this value lies outside and is 'worse' than the other members of the K14 ensemble. Worse here means that the absolute distance from the ~~RMIB-11~~ RMIB-UGent-11 value based on I_{K14} (top red dash) to the optimal value (grey line) is larger than that of any other K14 ensemble member. For example, the bias for the Iberian Peninsula in winter (in short written as as BIAS-IP-DJF) is more negative than any other model, and it is in absolute value the furthest from the optimal 0 K. If instead the background colour is green, this indicates again the value is outside of the K14 ensemble but not the furthest from the optimal value. This implies that either ~~RMIB-11~~ RMIB-UGent-11 outperforms all other models (e.g. RSV-AL-DJF) or is not the worst model as defined above (e.g. RSV-EA-DJF is outside of the K14 ensemble, but not as bad as models at the other end of the ensemble).

Overall, Fig. 3 shows that (i) ~~RMIB-11~~ RMIB-UGent-11 mostly falls within the K14 ensemble (white background colour), (ii) the jackknife confidence intervals are always much smaller than the total spread of the K14 ensemble, except for RIAV and TCOIAV where the intervals often cover half of the ensemble spread, (iii) the difference between the ~~RMIB-11~~ RMIB-UGent-11 (top red dash) and ~~RMIB-44~~ RMIB-UGent-44 (bottom red dash) scores is very small considering the total range covered by the ensemble and the calculated jackknife confidence intervals.

A more detailed analysis shows that for BIAS, ~~RMIB~~ RMIB-UGent is almost always on the 'cold side' of the K14 ensemble and even outside of its range on a fairly large amount of occasions. Especially for IP-DJF and SC-MAM, the cold bias is considerable. Also, ~~RMIB-44~~ RMIB-UGent-44 is slightly ($\sim$ .2 K) colder than ~~RMIB-11~~RMIB-UGent-11, which may be due to regridding and the resolution difference. For 95%-P, ~~RMIB-11~~ RMIB-UGent-11 is the worst model on four occasions among which most notably again IP-DJF and SC-MAM.

For spatial correlation (PACO) and variability (RSV) ~~RMIB-11~~ RMIB-UGent-11 performs better. Although in K14 these two metrics are plotted on a Taylor diagram, we choose to show them here separately in one figure for clarity and conciseness. RSV for ~~RMIB~~ RMIB-UGent is almost always

larger than 1, even where other models show less variability (e.g. ME). In the Alpine region (AL), ~~RMIB~~ RMIB-UGent seems to be able to grasp RSV well, but not at the right locations, as shown by the low PACO, especially in DJF. The jackknife confidence intervals are very small here, which indicates that both RSV and PACO produce very robust scores.

For RIAV and TCOIAV, ~~RMIB~~ RMIB-UGent again shows acceptable scores, being outside of the K14 ensemble in a limited amount of cases. More notably, the jackknife confidence intervals are relatively large for these scores and this questions the robustness of these metrics. For example, for FR-MAM the TCOIAV based on I_{K14} is 0.6, but the jackknife confidence interval extends from 0.6 to 0.8, covering all but two other models. For RIAV a similar situation for AL-JJA can be seen.

3.2 Precipitation

Figure 4 shows the spatial distribution of the relative seasonal precipitation BIAS (in %, (model-observed)/observed) for the winter and summer season for the years in I_{K14}. Comparison to Fig. 3 of K14 shows that in winter, like all other models, ~~RMIB-11~~ RMIB-UGent-11 generally overestimates precipitation amounts, except in Northern Africa. In contrast to temperature, ~~RMIB-11~~ RMIB-UGent-11 clearly differs from CNRM-11, with the latter showing large dry biases. In summer, ~~RMIB-11~~ RMIB-UGent-11 overestimates precipitation amounts, especially in the Mediterranean. Again, no clear resemblance to CNRM-11 is found.

Figure 5 is constructed in the same way as Fig. 3 and shows all precipitation scores for all different metrics, regions and seasons. Similar to the temperature scores, the results for precipitation reveal that the majority of scores lies within the K14 ensemble, no difference between ~~RMIB-11 and RMIB-44~~ RMIB-UGent-11 and RMIB-UGent-44 is found and the jackknife confidence intervals are much smaller than the total ensemble range except for RIAV and TCOIAV. However, there is a clear absence of yellow scores and an increased presence of green scores, indicating that ~~RMIB~~ RMIB-UGent precipitation scores are generally better than the temperature scores.

~~RMIB~~ RMIB-UGent has a wet BIAS for almost all regions and seasons. Remarkably, the best BIAS scores are obtained for SC-MAM and AL-DJF, where large temperature biases were found. Additionally, the corresponding 95%-P scores are also on the low side which shows that the good performance is not due to compensating biases.

For RSV, ~~RMIB~~ RMIB-UGent performs relatively well and for PACO it excels, with 10 out of 80 ~~regions/seasons~~ region-season combinations performing better than the complete K14 ensemble. Only in for AL-MAM its performance is not satisfactory, but remark that the actual score is an extreme outlier considering the jackknife confidence interval.

For RIAV, ~~RMIB~~ RMIB-UGent again performs consistently well, especially compared to the K14 ensemble which sometimes shows a large overestimation of interannual variability, i.e. very large values of RIAV. On the other hand, TCOIAV is mostly on the low side of the K14 ensemble, which shows that although ~~RMIB~~ RMIB-UGent gets the variability right, the actual temporal correlation is

not well grasped. As for temperature, the large jackknife confidence intervals question the robustness of the scores.

4 Discussion

~~Contrary to most of the other models in the K14 that have a long history in climate modelling and might have gone through a set of improvements and tuning throughout the years, this~~ This is the first time ALARO-0 was used ~~on such a large domain. Also, it was directly ported from its NWP setup to climate-scale simulations and therefore it was very possible that some problematic conclusions would arise from the output analysis. However~~for a climate experiment. Nevertheless, the performance of ALARO-0 on seasonal and yearly scales for both near-surface air temperature and precipitation is satisfactory. Generally ALARO-0 performs well, which is quantified by the large number of white boxes in ~~Fig~~Figs. 3 and 5 indicating that the ALARO-0 score lies within the existing K14 ensemble. For precipitation, ALARO-0 even outperforms all other models on numerous occasions. These results are encouraging, given that ALARO-0 does not yet have the experience in climate modelling that some of the other models of the K14 ensemble had, but was directly ported from its NWP setup. Although the 12.5-km resolution was also a novelty for the K14 models, their performance undoubtedly benefited from previous optimizations for climate experiments, albeit at a lower resolution of 50 km.

Some issues ~~do however~~ still remain. Most notably, this study has revealed some large temperature biases in Scandinavia and Eastern Europe. The spatial pattern of the BIAS resembles CNRM's ARPEGE model (shown in Fig. 2 of K14). In winter, the common east-west bias gradient can possibly be attributed to the shared dynamical core and the strong synoptic scale forcing in winter. In NWP applications of the ALADIN system similar symptoms have been diagnosed and have been shown to be related to stable boundary layer issues. The dampened bias patterns for ~~RMIB-11~~ RMIB-UGent-11 compared to CNRM-11 in the Alps and other mountainous regions is probably due to the different surface and snow cover scheme that is used by both. In summer, ~~RMIB-11~~ RMIB-UGent-11 is generally cold biased, except in Southern Europe where it suffers from the common summer warm bias, probably due to soil moisture feedbacks. Also, the ~~RMIB-11~~ RMIB-UGent-11 and CNRM-11 bias ~~pattern~~ patterns are less alike than in winter, possibly due to the increased number of local processes that influence and feed back into the mean fields. Both spatial and temporal variability are very well reproduced by ALARO-0, while correlations are on the low side compared to other models. The latter could partly be explained by the comparatively larger domain of ALARO-0 which could imply a weaker control of the boundary forcing.

For precipitation, ALARO-0 performs very well. Aside from some large wet biases in summer for the Iberian Peninsula (IP) and the Mediterranean (MD), biases are almost always below 50%. Contrary to temperature, the precipitation bias pattern shows no resemblance to ARPEGE (shown in

Fig. 3 of K14). This can be attributed to the different microphysics and convection parametrisation schemes that are used by both models. A similar result was found in K14 about the three WRF experiments that were analysed in the K14 ensemble. These only differed in the parametrisation schemes used, but often covered the complete ensemble spread. Remarkably, in Scandinavia all precipitation scores are very good, although temperature scores are sometimes very bad. It is very possible that the two are linked and some compensating effects or feedbacks exist, which is an additional incentive for a more thorough study. The good scores for spatial variability (RSV) and correlation (PACO) show that ALARO-0 is capable of producing not only the right amount of precipitation, but also at the right locations. The common model overestimation of spatial variability is also present in the ~~RMIB~~ RMIB-UGent runs, but as stated in K14, this could be due to a smoothing of the reference E-OBS dataset. Temporal variability is very well reproduced, but correlations are again rather low.

Similarly to the conclusions in K14, no consistent difference between the low- and high-resolution simulations in the scores is shown. However, ~~we expect~~ based on preliminary results, we expect that at the sub-daily scale the timing of precipitation is better represented by the high-resolution simulation.

Finally, it is clear that the period I_{K14} (1989-2010) used in K14 is sufficient to produce robust scores for BIAS, 95%-P, RSV, PACO and partly RIAV. This is quantified by the fact that the jackknife intervals for these metrics are very small compared to the total ensemble spread and they therefore do not depend strongly on the period used to compute them. For example, temperature biases calculated for I_{K14} are mostly within .1 K of the jackknife mean. This does not hold for some RIAV and most of the TCOIAV scores due to the fact that these exactly assess interannual variability. For model intercomparison a larger period should be considered for these scores.

5 Conclusions

The ALARO-0 model has its origins in the general circulation model ARPEGE and mainly its limited area model ALADIN. The new microphysics and convection scheme 3MT was implemented in ALADIN to form ALARO-0, which is used operationally for daily weather forecasts at the Royal Meteorological Institute of Belgium (RMIB). In this study, for the first time ever the ALARO-0 model was used to perform ~~uninterrupted~~ continuous climate simulations on a European scale for a 32-year period. Within the ~~the~~ framework of the CORDEX project, one low- and one high-resolution simulation were done on the EURO-CORDEX domain for the period 1979-2010, using the ERA-Interim reanalysis as boundary conditions. The results are compared to an existing ensemble of 19 similar simulations using different models that were analysed in Kotlarski et al. (2014), referred to as K14 in this text. One of the models used in K14 is the ARPEGE model by the Centre National de Recherches Météorologiques (CNRM), which due to its relation to ALARO-0 serves as a first reference for the performed simulations.

Main conclusions are that (1) ALARO-0 is able to represent both seasonal mean near-surface air temperature and accumulated precipitation amounts well and (2) all scores computed in K14 are robust, except for RIAV and TCOIAV.

The first conclusion is founded by the fact that most of the ALARO-0 scores lie within the K14 ensemble, thus not performing worse or better than other models. This is qualified in Fig. 3 and 5 by a white background. For temperature, some clear cold biases remain, which will be the subject of a follow-up study. Also, for temperature ALARO-0 seems to share some large biases with ARPEGE, while for precipitation this is not the case due to the inclusion of the 3MT scheme in ALARO-0. For precipitation, ALARO-0 performs very well consistently for all scores, regions and seasons and is on several instances better than all other models in the K14 ensemble.

In the second conclusion, what is meant by robust is 'independent of the time period used to compute the scores'. The ~~RMIB~~ RMIB-UGent simulations span the 32-year period 1979-2010, which is longer than the 20-year period 1989-2008 used in K14. By taking 1000 random 20-year samples from the 32-year pool, we computed 95% confidence intervals for all scores. Figure 3 and 5 show that the confidence intervals (red transparent bands) are generally much smaller than the total ensemble spread. Assuming this also holds for other models, this shows that model differences are significant. For RIAV this does not always hold and a longer period should be taken into account to compute the scores. For TCOIAV the situation is even more problematic and scores or model ranking should not be interpreted too strictly.

The outcomes of this study confirm the potential of ALARO-0 as a climate model on European scales. Future work will focus on pinpointing the causes of some of the remaining biases and performing simulations in which ALARO-0 is driven by a GCM, rather than ERA-Interim.

Acknowledgements. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Hercules Foundation and the Flemish Government – department EWI. This work was financially supported by the Belgian Science Policy (BELSPO) within the ECORISK (SD/RI/06A) and the CORDEX.be (BR/143/A2) projects. We would also like to thank Sven Kotlarski and Klaus Keuler for providing the necessary data and the two anonymous reviewers for their comments and useful suggestions that have improved the manuscript.

References

ALADIN international team: The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research, WMO Bull., 46, 317–324, 1997.

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Roberts, N. M., and Ferro, C. A. T.: The Value of High-Resolution Met Office Regional Climate Models in the Simulation of Multihourly Precipitation Extremes, Journal of Climate, 27, 6155–6174, doi:10.1175/JCLI-D-13-00723.1, http://dx.doi.org/10.1175/JCLI-D-13-00723.1, 2014.

Christensen, J. and Christensen, O.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, doi:10.1007/s10584-006-9210-7, http://dx.doi.org/10.1007/s10584-006-9210-7, 2007.

Davies, H. C.: A lateral boundary formulation for multi-level prediction models, Quarterly Journal of the Royal Meteorological Society, 102, 405–418, doi:10.1002/qj.49710243210, http://dx.doi.org/10.1002/qj.49710243210, 1976.

De Meutter, P., Gerard, L., Smet, G., Hamid, K., Hamdi, R., Degrauwe, D., , and Termonia, P.: Predicting Small-Scale, Short-Lived Downbursts: Case Study with the NWP Limited-Area ALARO Model for the Pukkelpop Thunderstorm., Mon. Wea. Rev., 143, 742–756, doi:http://dx.doi.org/10.1175/MWR-D-14-00290.1, 2015.

De Troch, R., Hamdi, R., Van de Vyver, H., Geleyn, J.-F., and Termonia, P.: Multiscale Performance of the ALARO-0 Model for Simulating Extreme Summer Precipitation Climatology in Belgium, Journal of Climate, 26, 8895–8915, doi:10.1175/JCLI-D-12-00844.1, http://dx.doi.org/10.1175/JCLI-D-12-00844.1, 2013.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, http://dx.doi.org/10.1002/qj.828, 2011.

Dudhia, J.: A history of mesoscale model development, Asia-Pacific Journal of Atmospheric Sciences, 50, 121–131, doi:10.1007/s13143-014-0031-8, http://dx.doi.org/10.1007/s13143-014-0031-8, 2014.

Gerard, L.: An integrated package for subgrid convection, clouds and precipitation compatible with the meso-gamma scales, Quart. J. Roy. Meteor. Soc., pp. 711–730, 2007.

Gerard, L. and Geleyn, J.-F.: Evolution of a subgrid deep convection parameterization in a limited area model with increasing resolution, Quart. J. Roy. Meteor. Soc., pp. 2293–2312, 2005.

Gerard, L., Piriou, J.-M., Brožková, R., Geleyn, J.-F., and Banciu, D.: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model, Mon. Wea. Rev., pp. 3960–3977, 2009.

Giorgi, F. and Mearns, L. O.: Introduction to special section: Regional Climate Modeling Revisited, Journal of Geophysical Research: Atmospheres, 104, 6335–6352, doi:10.1029/98JD02072, http://dx.doi.org/10.1029/98JD02072, 1999.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58, 175–183, 2009.

Hamdi, R., Giot, O., Troch, R. D., Deckmyn, A., and Termonia, P.: Future climate of Brussels and Paris for the 2050s under the {A1B} scenario, Urban Climate, 12, 160 – 182, doi:http://dx.doi.org/10.1016/j.uclim.2015.03.003, http://www.sciencedirect.com/science/article/pii/S2212095515000097, 2015.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres, 113, n/a–n/a, doi:10.1029/2008JD010201, http://dx.doi.org/10.1029/2008JD010201, d20119, 2008.

Hohenegger, C., Brockhaus, P., and Schär, C.: Towards climate simulations at cloud-resolving scales, Meteorologische Zeitschrift, 17, 383–394, doi:doi:10.1127/0941-2948/2008/0303, http://www.ingentaconnect.com/content/schweiz/mz/2008/00000017/00000004/art00004, 2008.

IPCC: Summary for Policymakers, book section SPM, p. 1–30, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.004, www.climatechange2013.org, 2013.

Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J.: Realism of Rainfall in a Very High-Resolution Regional Climate Model, Journal of Climate, 25, 5791–5806, doi:10.1175/JCLI-D-11-00562.1, http://dx.doi.org/10.1175/JCLI-D-11-00562.1, 2012.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geoscientific Model Development, 7, 1297–1333, doi:10.5194/gmd-7-1297-2014, http://www.geosci-model-dev.net/7/1297/2014/, 2014.

Lin, J.-L., Qian, T., and Shinoda, T.: Stratocumulus Clouds in Southeastern Pacific Simulated by Eight CMIP5-CFMIP Global Climate Models, Journal of Climate, 27, 3000–3022, doi:10.1175/JCLI-D-13-00376.1, http://dx.doi.org/10.1175/JCLI-D-13-00376.1, 2014.

Spiridonov, V., Déqué, M., and Somot, S.: ALADIN-CLIMATE: from the origins to present date, ALADIN Newsletter, 29, 2005.

Sun, D.-Z., Yu, Y., and Zhang, T.: Tropical Water Vapor and Cloud Feedbacks in Climate Models: A Further Assessment Using Coupled Simulations, Journal of Climate, 22, 1287–1304, doi:10.1175/2008JCLI2267.1, http://dx.doi.org/10.1175/2008JCLI2267.1, 2009.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, http://dx.doi.org/10.1175/BAMS-D-11-00094.1, 2011.

Teutschbein, C. and Seibert, J.: Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies, Geography Compass, 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, http://dx.doi.org/10.1111/j.1749-8198.2010.00357.x, 2010.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins,

B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, doi:10.1256/qj.04.176, http://dx.doi.org/10.1256/qj.04.176, 2005.

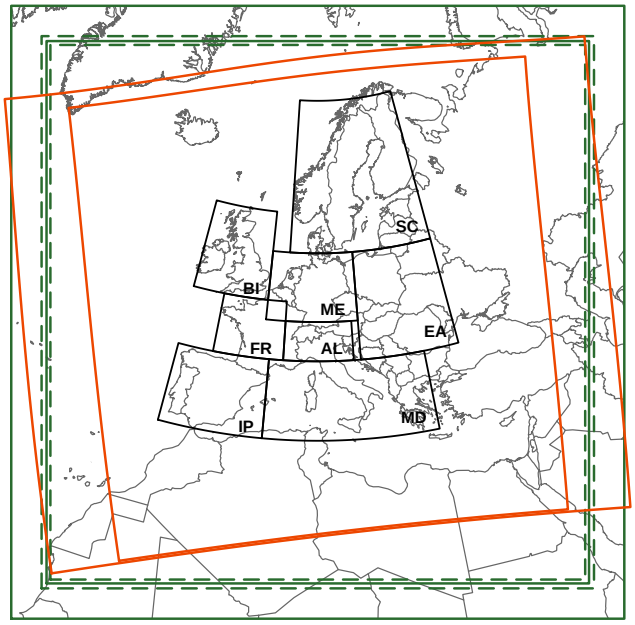

Figure 1. Domain boundaries of the used integration grids. The CORDEX community prescribes the rotated lon-lat EURO-CORDEX domain (inner orange box) which is completely encompassed by the E-OBS domain (outer orange box). The outer green boxes show the ~~RMIB-11~~ RMIB-UGent-11 (dashed lines) and ~~RMIB-44~~ RMIB-UGent-44 (full lines) conformal Lambert domain boundaries. The inner green boxes exclude the eight grid point Davies coupling zone. In black the different European climatic regions as defined in Christensen and Christensen (2007) are shown.

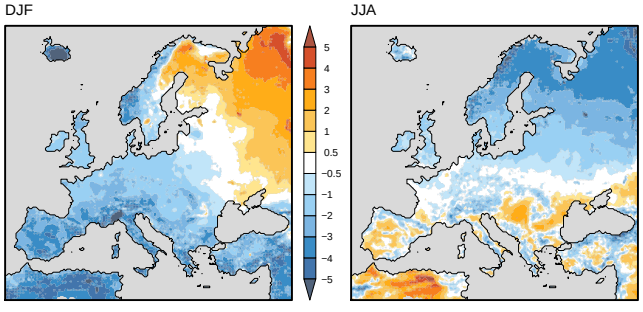

Figure 2. spatial BIAS of near-surface air temperature [K] over the sample I_{K14} for DJF (left) and JJA (right) for RMIB-UGent-11. Compare to Figure 2 of Kotlarski et al. (2014).

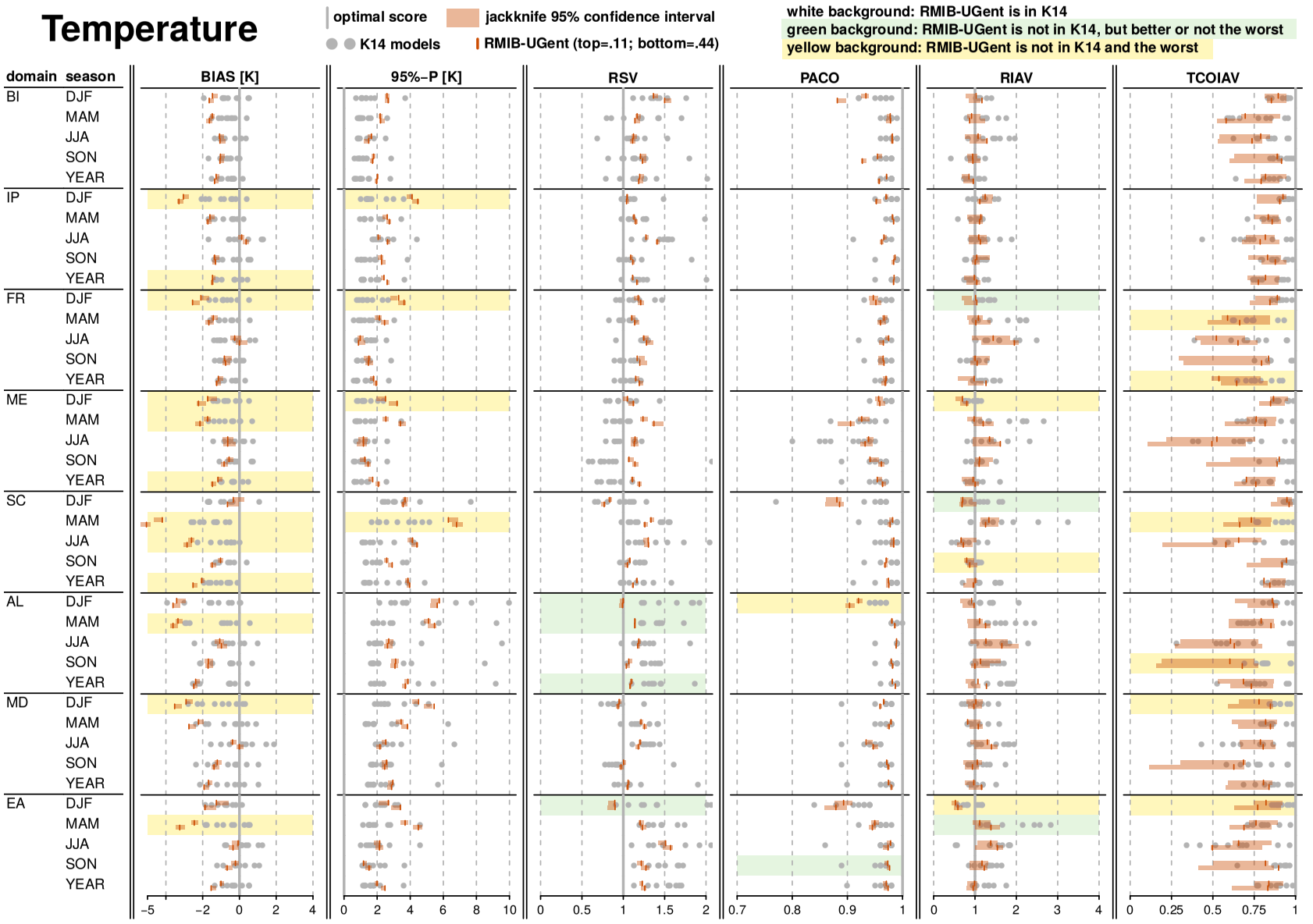

Figure 3. Scores for near-surface air temperature for all ~~domain~~ domains (first column), seasons (second column) and metrics.

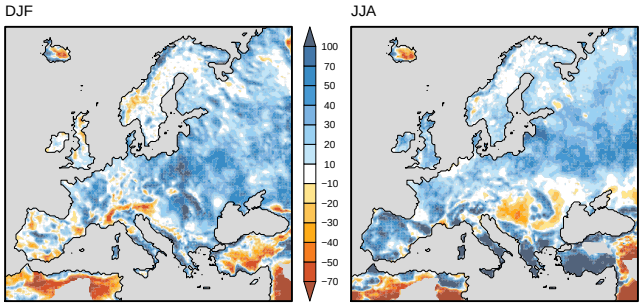

Figure 4. spatial BIAS of precipitation [%] over the sample I_{K14} for DJF (left) and JJA (right) for RMIB-UGent-11. Compare to Figure 3 of Kotlarski et al. (2014).

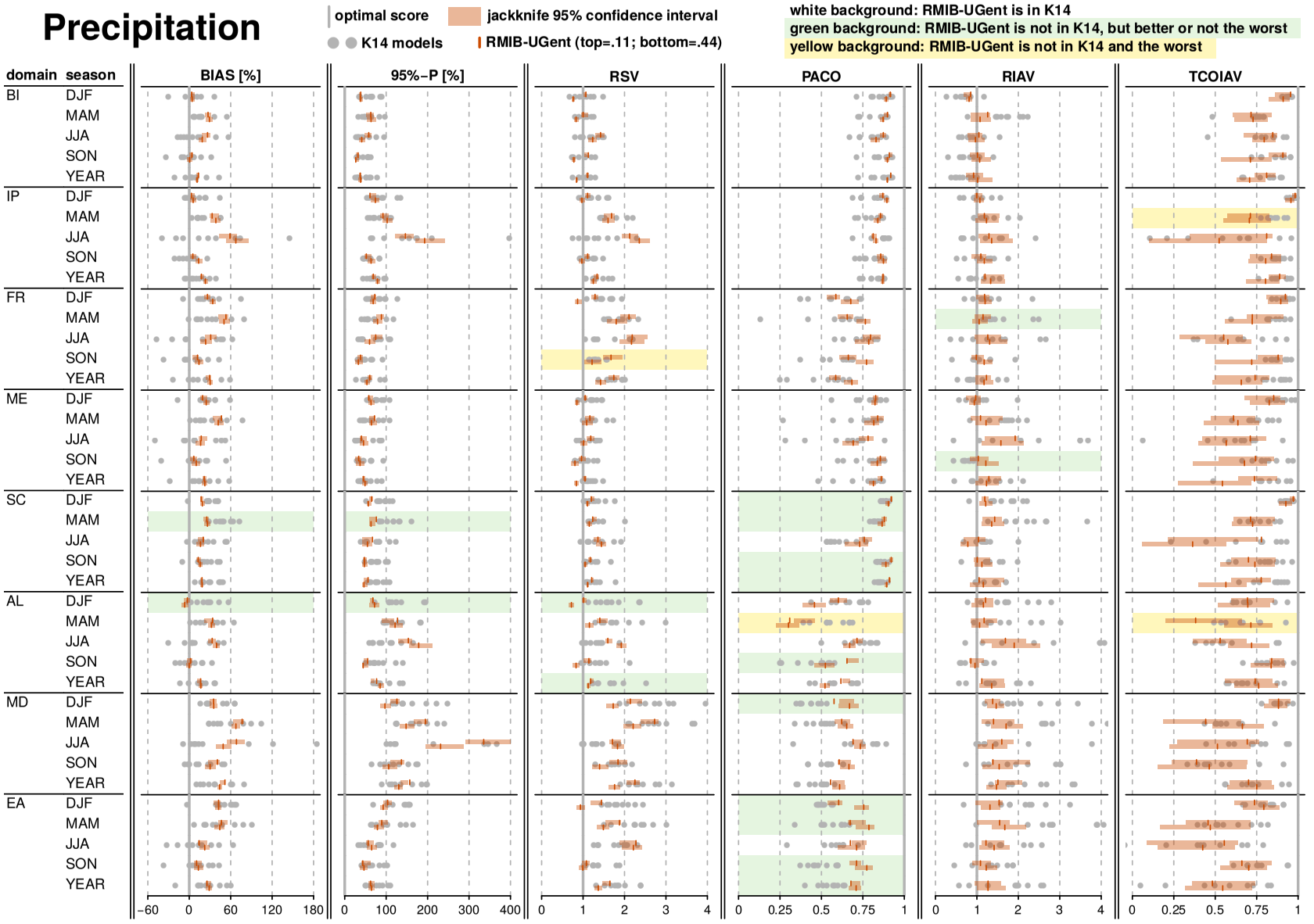

Figure 5. Scores for precipitation for all ~~domain~~ domains (first column), seasons (second column) and metrics.