Skill of 2-m Temperature Seasonal Forecasts over Europe in ECMWF and RegCM Models

MIRTA PATARČIĆ AND ĆEĐO BRANKOVIĆ
Meteorological and Hydrological Service (DHMZ), Zagreb, Croatia

(Manuscript received 28 April 2011, in final form 13 October 2011)

ABSTRACT

Various measures of forecast quality are analyzed for 2-m temperature seasonal forecasts over Europe from global and regional model ensembles for winter and summer seasons during the period 1991 to 2001. The 50-km Regional Climate Model (RegCM3) is used to dynamically downscale nine-member ensembles of ECMWF global experimental seasonal forecasts. Three sets of RegCM3 experiments with different soil moisture initializations are performed: the RegCM3 default initial soil moisture, initial soil moisture taken from ECMWF seasonal forecasts, and initial soil moisture obtained from RegCM3 ECMWF interim Re-Analysis (ERA-Interim)-driven integrations (RegCM3 climatology). Both deterministic and probabilistic skill metrics are estimated.

The better-resolved spatial scales in near-surface temperature by RegCM3 do not necessarily lead to the improved regional model skill in the regions where systematic errors are large. The impact of initial soil moisture on RegCM3 forecast skill is seen in summer in the southern part of the integration domain. When regional model soil moisture was initialized from ECMWF seasonal forecasts, systematic errors were reduced and deterministic skill was enhanced relative to the other RegCM3 experiments. The Brier skill score for rare cold anomalies in this experiment is comparable to that of the global model, whereas in other experiments it is significantly smaller than in global model.

There is no major impact of soil moisture initialization on forecast skill in winter. However, some significant improvements in RegCM3 probabilistic skill scores for positive anomalies in winter are found in the central part of the domain where RegCM3 systematic errors are smaller than in global model.

1. Introduction

Predictability of climate at seasonal time scales mostly depends on slowly varying lower boundary conditions, particularly oceans, with the El Niño–Southern Oscillation (ENSO) as the most important source of forcing. The influence of the tropical Pacific on other parts of the earth’s atmosphere is attained through ENSO teleconnections (Palmer and Anderson 1994). The interaction of the atmosphere and ocean together with ENSO teleconnections can be modeled by using the one-tier forecasting systems that consist of coupled atmosphere–ocean general circulation models (AOGCMs). Nonlinearity of climate system, parameterizations and discretizations in time and space in models, and errors in initial conditions cause uncertainties in seasonal predictability of atmospheric states. To represent some of these uncertainties, seasonal forecasts are normally run in ensemble mode starting from slightly different initial conditions.

Although AOGCMs realistically simulate large-scale flow at seasonal time scales (Goddard et al. 2001), because of their coarse horizontal resolution the surface fields that depend on orography, such as precipitation and surface air temperature, may not be adequately represented. Furthermore, end-user applications require better spatial and temporal resolutions than those provided by AOGCMs. One method of increasing the resolution of the fields of interest is dynamical downscaling of AOGCM output by using regional climate models (RCMs). The method was introduced by Dickinson et al. (1989) and Giorgi (1990), who have shown that RCMs can simulate large-scale circulation consistent with that from the driving AOGCMs. Because of the higher resolution and better resolved orography, land surface characteristics, and land–sea contrasts, RCMs are able to generate small-scale features with appropriate amplitudes and climate statistics (Laprise et al. 2008). Consequently, RCMs are used in climate research...
mainly to investigate regional climate and climate variability (e.g., Small et al. 1999; Seth et al. 2007) and for dynamical downscaling of AOGCMs’ climate change simulations (e.g., Giorgi et al. 2004; Jacob et al. 2007; Christensen and Christensen 2007).

In experiments dealing with downscaling of seasonal forecasts, the influence of ENSO on seasonal climate is transferred to an RCM via the boundary conditions provided by the driving AOGCM. ENSO forcing in an RCM is thus conditioned by the quality of driving field, the ENSO impact over a particular region, and the ability of RCM to reproduce large-scale fields supplied by global model. Other factors that influence atmospheric seasonal predictability are sea ice (Balmaseda et al. 2010) and land surface–related conditions such as snow cover (Shongwe et al. 2007) and soil moisture (e.g., Douville and Chauvin 2000; Douville 2010). Since these factors are of regional and local character, as opposed to remote ENSO forcing, they could bring some improvements in seasonal forecasts produced by a higher-resolution RCM. The downscaled seasonal forecasts could be applied to decision making models (e.g., crop yield prediction, early warning systems). In this context, seasonal predictions of surface parameters are very important, particularly forecasts of 2-m temperature (T2m), a parameter strongly influenced by small-scale variations in regional orography but also closely related to large-scale circulation.

The ability of RCMs to improve global model T2m seasonal forecasts was examined in several studies over different extratropical regions. Fennessy and Shukla (2000) reported a larger magnitude of regional model systematic errors in summer and smaller magnitude of systematic errors in winter when compared to global model alone, with a tendency of higher regional model anomaly correlations in summer and higher global model anomaly correlations in winter. Roads et al. (2003) found that their regional model’s seasonal near-surface temperature forecasts over the continental United States were better than that of the driving global model only in the colder regions related to high mountains, but the overall skill was degraded relative to the driving global model. Another study by Roads (2004) showed that the T2m skill of the regional model for the United States was comparable but not superior to the skill of the driving global model. The results of the abovementioned studies indicate that there is no clear-cut advantage of RCM seasonal forecasts over those made by AOGCMs. However, they also indicate a potential of seasonal forecasts at regional scales, conditioned by improvements in various components of the prediction system. The different performances of RCMs in seasonal forecast downscaling will depend on region examined and characteristics of regional model used.

In this study, we investigate whether the regional climate model RegCM3 (Pal et al. 2007) is able to make an improvement for the European region relative to the experimental seasonal forecasts generated by European Centre for Medium-Range Weather Forecasts (ECMWF) global model. Experimental seasonal forecasts were chosen because of a better availability of data required to drive regional models when compared to ECMWF operational seasonal forecasts. A previous study by Branković and Patarčić (2008), in which operational seasonal forecasts from ECMWF were downscaled by RegCM3, showed that the scarcity of the ECMWF model forcing data at the time (12-h frequency, six levels in the vertical) had a major detrimental effect on the regional model’s performance.

T2m seasonal forecasts of global and regional model ensembles are compared for the 11-yr period in terms of model bias and both deterministic and probabilistic skill measures. The three different ways of RegCM3 soil moisture initialization define three separate sets of RegCM3 experiments. This was motivated primarily because in the default version of RegCM3, the initial soil moisture is defined in a relatively crude way (Giorgi and Bates 1989), which may not be suitable for seasonal forecasts since several studies emphasized the relevance of realistic soil moisture conditions in seasonal climate predictions (e.g., Fennessy and Shukla 1999; Douville and Chauvin 2000; Kanamitsu et al. 2003; Douville 2010; van den Hurk et al. 2012). For this reason, two additional sets of experiments have been performed. In the first one, soil moisture from the driving ECMWF seasonal forecasts was used to initialize soil moisture in RegCM3. In the second, the “climate” soil moisture from the RegCM3 integrations driven by ECMWF interim Re-Analysis (ERA-Interim) data (Berrisford et al. 2009) were taken to initialize soil moisture in the RegCM3 seasonal hindcasts. The results of these three different sets of RegCM3 experiments are compared to the global model T2m seasonal forecasts.

In the following section the models, validation data, and experimental setup are described. In section 3 the models’ systematic errors and impact of soil moisture initialization are analyzed. The results on forecast quality measures are presented and discussed in section 4. A summary and conclusions are given in section 5.

2. Experiments and data

a. The models

The RegCM3 regional climate model was used to dynamically downscale ECMWF experimental seasonal forecasts, which were a part of the European Union (EU)-funded project Ensembles-Based Predictions of Climate
Changes and Their Impacts (ENSEMBLES; Hewitt and Griggs 2004; Doblas-Reyes et al. 2009). The output from ensemble hindcasts of the ECMWF AOGCM Integrated Forecast System (IFS)/Hamburg Ocean Primitive Equation (HOPE) model (Anderson et al. 2007) was used to create initial and boundary conditions for RegCM3. The horizontal resolution of the atmospheric component of ECMWF global model is $T_{85} = 2.85^\circ$ (approx. 200 km) with 40 levels in the vertical. The ocean model has the meridional horizontal resolution of 1° (enhanced to 0.3° near the equator) and 29 vertical levels. Seasonal forecasts were initiated twice a year, on 1 May and 1 November, and integrated for 7 and 14 months respectively during the 11-yr period from 1991 to 2001. Atmospheric and land initial conditions for the ECMWF model were taken from the 45-yr ECMWF Re-Analysis (ERA-40; Uppala et al. 2005) and ocean initial conditions from ocean reanalysis (Balmaseda et al. 2008). Each forecast comprises an ensemble of nine individual integrations and they are created from different oceanic initial conditions (Doblas-Reyes et al. 2009). The land surface scheme in the ECMWF forecasting system, the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) is used (Viterbo and Beljaars 1995; van den Hurk et al. 2000). Soil moisture is defined at four soil layers at the same horizontal resolution as the atmospheric fields. The depth of the first soil layer is 7 cm (0–7 cm), the thickness of the second layer is 21 cm (7–28 cm), the third layer is 72 cm thick (28–100 cm), and the fourth layer expands from 100 to 255 cm.

RegCM3 is the third generation of the RegCM, originally developed by Dickinson et al. (1989) and Giorgi (1989). It is a hydrostatic atmospheric model defined in the $\sigma$ vertical coordinate system. Model dynamics is similar to that of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5; Grell et al. 1995). Large-scale precipitation is parameterized by the Subgrid Explicit Moisture Scheme (SUBEX; Pal et al. 2000). The Grell convection parameterization (Grell 1993) can use either the Fritsch–Chappell (Fritsch and Chappell 1980) or the Arakawa–Schubert (Arakawa and Schubert 1974) closure schemes. Radiative transfer package is based on the NCAR Community Climate Model (CCM3; Kiehl et al. 1996), and for the representation of land surface processes the Biosphere–Atmosphere Transfer Scheme (BATS; Dickinson et al. 1993) is used. The land surface model has three soil layers: the surface soil layer (10 cm thick), the root zone layer (1, 1.5, or 2 m thick, depending on vegetation type), and the deep soil layer (3 m thick; Elguindi et al. 2007, unpublished manuscript (available online at http://users.ictp.it/RegCNET/regcm.pdf)). Further details of RegCM3 are given in Pal et al. (2007).

b. Experimental setup

Within the RegCM3 integration domain there are $80 \times 66$ grid points projected on the Lambert conformal coordinates with 50-km grid spacing. The domain is centered at 45°N, 12.5°E and covers central and southwestern Europe with the Mediterranean Sea and northern Africa included. Figure 1 shows the RegCM3 domain and topography for both models. The analysis of seasonal forecasts was made for three regions. They are marked by the rectangles in Fig. 1b indicating the analysis domain (region A: 30.75°–57.75°N, 7.75°W–33.25°E), the southern part of the domain (B: 30.75°–47.75°N, 7.75°W–33.25°E), and the central part of the domain (C: 36.75°–50.75°N, 3.75°–28.25°E). The maximum difference in topography between the two models is over the Alps, reaching about 800 m. The major orographic features of the domain are clearly visible at the higher resolution. In the vertical, there are 18 terrain-following levels, extending from the surface to 100 hPa. After performing a number of sensitivity tests, when RegCM3 was forced with ERA-40 data (Brankovic et al. 2004), we have chosen for our experiments the Arakawa–Schubert closure because RegCM3 simulations with this scheme best agreed with the observations.

Downscaling of ECMWF global seasonal forecasts has been carried out for all ensemble members over the 11-yr period for the summer [July–September (JAS)] and winter [January–March (JFM)] seasons, corresponding to forecast months 3–5 of the global model integrations. The spinup time for the integrations was 3 days (i.e., initial dates were 29 December for JFM and 28 June for JAS). The 6-hourly sea surface temperature, surface pressure, as well as temperature, humidity, and wind at 40 ECMWF model levels were retrieved and used as the initial and boundary conditions for RegCM3. Downscaling was performed for three different soil moisture (SM) initializations in RegCM3: 1) the default SM initialization (denoted RCM_DFSM), 2) the initial SM taken from the driving ECMWF model seasonal forecasts (denoted RCM_ECSM), and 3) the initial SM set to climatological values derived from ERA-Interim-driven RegCM3 integrations (denoted RCM_EISM). After the initialization, SM freely evolves until the end of simulation. The spinup time in RegCM3 experiments is needed because of a possible imbalance between land surface and atmospheric initial conditions (see the detailed discussion in section 3b).

In RCM_DFSM, the initial SM is defined as a function of land cover/soil type and vegetation type (Giorgi and Bates 1989). This is a rather crude way of defining the initial SM and it includes neither seasonal nor interannual variation. For the RCM_ECSM experiment,
initial SM fields were taken from ECMWF seasonal forecasts that correspond exactly to the initial dates of RegCM3 integrations (i.e., initial SM is consistent with the other initial data). Since there are five possible soil depths in the RegCM3 BATS scheme (see the previous section), soil water content from ECMWF four layers was first interpolated vertically to match vertical soil layers in BATS and then horizontally by using the bi-linear interpolation in each of the five layers to match RegCM3 horizontal grid. Finally, the root zone layer SM, with three possible soil depths, was obtained by merging those three layers into one field depending on the RegCM3 vegetation type. For the RCM_EISM experiment, RegCM3 was first continuously driven by ERA-Interim data for the same 11 years as ECMWF experimental seasonal forecasts, from 1991 to 2002. Similar to Moufouma-Okia and Rowell (2010), the first five years were allowed for SM to reach the balance with the model atmosphere. From the last six years, the average SM for each soil layer was calculated from 15 December to 15 January and from 15 June to 15 July for winter and summer seasons, respectively. These “climatological” mean SM fields were then associated with the integration initial dates, 29 December for JFM
and 28 June for JAS as initial SM. Thus, RCM_EISM has no interannual variability included in the initial SM (in this sense it is similar to RCM_DFSM), but it includes SM annual cycle. In total, 99 individual integrations were at our disposal for each experiment with different initial SM.

c. Validation dataset

The Climatic Research Unit (CRU; University of East Anglia) gridded global monthly means were used to validate T2m forecasts from both models. From the second version of CRU dataset (CRU TS 2.1; Mitchell and Jones 2005), the data for the period 1991–2002 were selected to validate both models. Models’ outputs were interpolated to the CRU regular 0.5° (latitude/longitude) grid and forecast quality is assessed over the land since CRU data are defined for land points only.

3. Systematic errors

a. Errors in seasonal mean

In this section, we present and discuss the T2m systematic errors for both global and regional model integrations. Validation is performed on seasonal-mean values for JAS and JFM and the model systematic errors are defined as the difference between the model and observed climatologies. Model climatology is calculated as the 11-yr mean from seasonal ensemble means, whereas observed climatology is defined as an average of seasonal means over the same 11-yr period.

Figure 2 shows systematic errors for the global model (Figs. 2a,b) and RCM_DFSM experiment (Figs. 2c,d) together with verifying CRU data (Figs. 2e,f). In winter, the model has a cold bias over much of Europe with the largest values between −2°C and −4°C over the Iberian Peninsula and up to −5°C over the northern Italy lowland (Fig. 2a). Systematic errors are very small in the northeastern part of the domain where cold temperatures prevail. A relatively large warm bias is generally found over the major mountainous regions (e.g., the Pyrenees, the Alps, the Apennines, the Dinaric Alps, and the Carpathian mountains). These errors range from 0.5°C to more than 5°C (over the Alps more than 7°C) and they are at least partly related to the coarse resolution in ECMWF model, which entails inadequate representation of orography. In RegCM3, positive errors over almost all the European mountains are reduced relative to the global model; this is primarily due to a better resolved orography in RegCM3. Cold bias dominates over much of the integration domain with amplitude generally between −1°C and −4°C. For example, large cold bias is seen over the Iberian Peninsula and over southern France (from −2°C to −4°C). However, when compared to global model, the magnitude of the RegCM3 cold bias is generally somewhat reduced, particularly in central and eastern Europe.

In summer, warm bias in ECMWF model is seen over western Europe (mainly between 0.5°C and 4°C), northern Africa (0.5°C–6°C), and again over mountains and high elevations (Fig. 2b). Central and northern Europe is, similar to JFM, cooler than observations. In northeastern Europe T2m predictions are very close to observations, except for the eastern boundary region where model exhibits warm bias. In RegCM3, cold bias is distributed all over Europe, with errors ranging between −2°C and −6°C over a large part of the domain (Fig. 2d). The amplitude of cold bias in RegCM3 is larger than in the global model.

The similarity of systematic errors in the regional model to those in the global model could be attributed, to a certain extent, to the impact of the global model boundary conditions on the regional model integrations. Such similarity is mostly confined to the boundary regions where the impact from the global model is strongest; however, it is beyond the scope of this study to elucidate the relative impact of the driving model errors on RegCM3 integrations. In summer, the RegCM3 error in the central part of the domain is less similar to ECMWF. This may be the consequence of an increased influence of the regional model physics when the advection from the lateral boundaries is relatively weak (Noguer et al. 1998). Prevailing RegCM3 cold bias in summer is also reported by Brankovic et al. (2012) in a downscaling study of long climate simulations, when RegCM3 was driven by a different AOGCM.

For RCM_ECSM and RCM_EISM experiments, spatial distributions of systematic errors in both seasons are very similar to those shown in Figs. 2c and 2d for RCM_DFSM (not shown). However, there are some differences as well—for example the area encompassed by the −2°C contour over Europe and northern Africa during summer is smaller in RCM_ECSM than in RCM_DFSM. This is seen as a reduction in the amplitude of the area-averaged absolute errors from 2.1 in RCM_DFSM to 1.6 in RCM_ECSM in the analysis domain (region A in Fig. 1b) and from 2.3 to 1.6 in its southern part (region B; see Table 1). For both region A and region B, medians and maximum values of absolute errors are also smaller in RCM_ECSM when compared to the other RegCM3 experiments. In JFM, on the other hand, the area-averaged absolute systematic errors for the analysis domain (A) reveal a somewhat increased bias in RCM_ECSM and RCM_EISM when compared to RCM_DFSM (Table 1), primarily owing to an increase of cold bias over northern Africa. In the central part of the domain (C), the average,
median, and maximum absolute systematic errors in all RegCM3 experiments are smaller than in global model.

b. Impact of soil moisture initialization

The impact of the different SM initializations on the T2m seasonal forecasts in RegCM3 is analyzed from the daily time-evolution diagrams in JAS (where daily averages were derived from the 3-hourly values) of the area-averaged ensemble-mean SM in the top soil layer (top 10 cm) and T2m in the driest (1991) and the wettest (2001) years (Fig. 3). The criteria for wettest–driest year were based on summer precipitation anomalies in all experiments in the southern part of the domain (region B in Fig. 1b). This area was specifically chosen because the reduction of summer T2m errors (in the RCM_ECSM experiment) was largest in region B, as discussed in the previous subsection (cf. Table 1). In both years, SM in RCM_DFSM is highest and T2m is lowest throughout the
RCM_ECSM is higher), in the same period T2m is about is higher than in RCM_EISM (because the initial SM in though in the first 10–15 days in July SM in RCM_ECSM Al-(i.e., T2m in our experiments is “inversely proportional” to the amount of SM in the uppermost soil layer). Al- though in the first 10–15 days in July SM in RCM_ECSM and observed T2m monthly means might point to either 2°C–3°C colder than CRU, suggesting that monthly mean SM is too high in this experiment in both years considered. Based on the above assessment we may conclude that the RCM_ECSM simulation, which includes interannual variation of the initial SM, could be considered as having the most realistic initialization of SM, which leads to more accurate predictions of T2m in JAS than in the other two RegCM3 experiments. The RCM_EISM experiment, which does not include interannual variation of the initial SM but has the initial SM defined in physically consistent way, closely follows RCM_ECSM. This agrees with some studies that emphasized that the use of realistic initial state of SM could enhance seasonal prediction of the near-surface temperature (Fennessy and Shukla 1999; Kanamitsu et al. 2003). However, according to Fig. 3, the impact of the initial SM in our experiments is diminished after about one month or so. Similar results were obtained by van den Hurk et al. (2012), who showed that ensemble T2m predictions over Europe are positively affected by realistic initial SM up to one month ahead. The increased discrepancy between the modeled and observed T2m monthly means might point to either some (systematic) deficiency in regional model or to possibly adverse impact of the lateral boundaries from the driving global model.

The rapid drift in SM during the first 10 days or so in RCM_DFSM (Figs. 3b, d) could be explained by unrealistically high initial SM magnitude which is due to the absence of seasonal variation in this experiment. The drift in RCM_ECSM is more difficult to explain, but somewhat higher initial SM values might have been generated by the interpolation procedures used to scale the original SM from ECMWF soil layers to the different soil scheme layout in RegCM3. For those two experiments, spinup is needed for SM to reach balance with atmospheric data. This seems to be indirectly confirmed by the absence of the drift in RCM_EISM where the initial SM was derived from RegCM3 integrations (i.e., when the identical model setup was used). Although in the RCM_ECSM experiment initial forcing season (black solid line in Fig. 3) when compared to the other two experiments. Although SM in RCM_DFSM drops sharply in about 10–15 days, for the most part of model integrations it stays comparably high; it might be even argued that such a high SM is partly responsible for the RegCM3 cold bias in summer (Fig. 2d). The initial difference in SM between RCM_ECSM and RCM_EISM decreases with time and, after approximately 15 days, the gray and dotted curves in Figs. 3a and 3c remain close to each other for the rest of integrations. The differences among RegCM3 experiments in the evolution of T2m (Figs. 3b, d) are the largest in the first month, then gradually reduce in the second month and come close to each other in the last month of the integrations. Irrespective of the year, the warmest temperatures are in RCM_ECSM, and the coolest temperatures are in the default experiment RCM_DFSM. This is the opposite to the time evolution of SM in Figs. 3a and 3c (i.e., T2m in our experiments is “inversely proportional” to the amount of SM in the uppermost soil layer). Although in the first 10–15 days in July SM in RCM_ECSM is higher than in RCM_EISM (because the initial SM in RCM_ECSM is higher), in the same period T2m is about 0.5°C warmer in the RCM_ECSM than in RCM_EISM. Based solely on SM and T2m data, we cannot offer a plausible explanation for this “exception,” but it seems that some other process(es) involved in the land–atmosphere interaction might be affecting the magnitude of T2m in the first few weeks of integrations. The daily variation in all three experiments is almost identical, indicating that the lateral boundary conditions from the driving global model dominate the evolution of both SM and T2m.

In Figs. 3b and 3d, the T2m monthly mean values for all three RegCM3 experiments are compared with CRU data. The RCM_ECSM T2m (open triangles) shows the best agreement with the CRU monthly means (solid circles), particularly in the first month of the JAS season. In the second and third months, the amount of SM in the top layer is continuously increasing and the T2m monthly means in RegCM3 become significantly cooler than CRU, but again RCM_ECSM is the closest to observations. In the default experiment RCM_DFSM, monthly mean T2m values (open circles) are for about 2°C–3°C colder than CRU, suggesting that monthly mean
fields, which are taken from the global model, are in balance, the drift in T2m is still possible (Figs. 3b,d). This is again because of the interpolation of SM from ECMWF to RegCM3 land surface layout and, more generally, because of the different physical parameterizations in the two models.

Since the climatic characteristics of northern Africa are generally different from those in Europe, the analysis of SM and T2m time evolutions was repeated for the same region B but with northern Africa excluded from the calculation (with the southern boundary defined at 37.25°N). For brevity, we discuss the results without showing figures. When compared to the area averages in Figs. 3a and 3c, the area averages without northern Africa indicate somewhat larger differences in the magnitude of SM between RCM_EISM and RCM_ECSM (over 1 mm) in the first two months, but with the relative positioning of the three curves remaining the same. Likewise, the time evolution of T2m is very similar to that in Figs. 3b and 3d with the warmest T2m in RCM_ECSM, which best agrees with the CRU monthly means. Based on this analysis, we decided to use regions A and B in Fig. 1b (with northern Africa included) for the calculation of scores discussed in the next sections.

4. Forecast quality measures

Different aspects of forecast performance are assessed by applying several statistical methods. In the previous section it was shown that there is a nonnegligible bias in all integrations for both seasons; this suggests that the verification methods based on forecast anomalies should be used. Thus, for ensemble-mean seasonal forecasts, deterministic skill of anomalies is assessed by using anomaly correlation coefficients (ACCs). Since we deal with ensembles of seasonal forecasts, probabilistic verification is also evaluated. For the comparison of global and regional model forecasts a statistical inference is applied (Jolliffe 2007; Doblas-Reyes et al. 2009) and a bootstrap method that includes a resampling of the forecast–verification pairs (1000 times with replacement) is used. New scores for each sample were calculated, ranked, and the 2.5 and the 97.5 percentiles of the bootstrap distribution were used as the limits of the 95% confidence interval. The above bootstrap score

FIG. 3. Time series of the ensemble-mean area averaged (region B) RegCM3 upper soil layer (top 10 cm) (left) soil moisture and (right) T2m for the (top) driest and (bottom) wettest years according to precipitation anomaly in summer. Open circles, triangles, and diamonds indicate the RCM_DFSM, RCM_ECSM, and RCM_EISM experiments ensemble-mean monthly means (MM) respectively, and solid circles are verifying CRU monthly means. Vertical lines denote the beginning of the month in season.
estimates determine confidence intervals for the differences in scores between global and regional model.

a. Deterministic skill score

If seasonal forecasts are to be skillful, it is important that they capture the observed interannual variability. Figure 4 shows the T2m area-averaged anomalies for global and regional ensembles for land points only and verifying CRU dataset in the analysis domain. For ECMWF and all RegCM3 experiments, model anomalies were calculated for seasonal ensemble mean and each ensemble member of each year relative to the mean over 11 years. CRU anomalies were obtained by removing CRU climatology for the same 11 years from observed seasonal means. Vertical bars denote the range of anomalies within ensembles (i.e., the largest and
smallest anomalies from individual ensemble members; they indicate ensemble spread for individual years.

For the given 11-yr period, interannual variation of the observed T2m seasonal anomalies, as revealed by solid circles in Fig. 4, is larger in winter than in summer: it varies between +1.0° and −2.0°C in winter (Fig. 4a) and between +1.0° and −1.0°C in summer (Fig. 4b). For all integrations, interannual variability of ensemble averages broadly follows observed anomalies; however, it is smoother than observed. The comparison of vertical bars and solid circles in Fig. 4 reveals that there are only few years when the observed anomaly is outside the range of ensemble members (two in winter and two in summer). This essentially implies that, in the 11-yr period, observed T2m anomalies (irrespective of their sign) are rarely underestimated by all ensemble members of global and all regional model experiments. Furthermore, a relatively close positioning of ensemble averages implies no major differences between the models. Different initial SM in each RCM_ECSM ensemble member contributes to the increased variability of summer T2m anomalies, seen from the larger extent of vertical bars in Fig. 4b.

To evaluate the deterministic skill of forecasts, ACCs, which measure the association between gridpoint values of forecast and observed fields, are computed as follows (e.g., Déqué 1997; Doblas-Reyes et al. 2000):

\[
    \text{ACC} = \frac{\sum [(F_m - \overline{F}_m)(O_m - \overline{O}_m)]}{\left[\sum (F_m - \overline{F}_m)^2 \sum (O_m - \overline{O}_m)^2\right]^{1/2}},
\]

where \( F_m \) is forecast, \( O_m \) observation, and \( \overline{F}_m \) and \( \overline{O}_m \) are model and observed climatologies respectively at the grid point \( m \). The summation in (1) could be either over time (11 yr) or over grid points and it is weighted by the cosine of the gridpoint latitude.

Time series of ensemble-mean ACCs for T2m in both seasons is shown in Fig. 5, for the same region (A) as in Fig. 4. Clearly, in the period 1991/92–2001/02 models show large interannual variation in the ensemble-mean skill, with generally higher skill in summer than in winter. For example, in JAS the ECMWF ACC is positive in 8 out of 11 years and at least equal to 0.5 in 5 years. If we focus on positive ACCs only, the RegCM3 ACCs tend to follow relatively closely those of the global model (more so in winter than in summer), and in most cases the 95% confidence intervals between ECMWF and RegCM3 overlap (Figs. 5a,b).

For the years with relatively high skill (with ACC over 0.5 in at least one of the ensembles), relatively narrow confidence intervals are seen, indicating a good correspondence between modeled and observed anomalies over the analysis domain A (i.e., low variability or robustness of ACCs calculated from bootstrap samples). A bootstrap method was applied to determine whether the values of RegCM3 and global model ACCs differ significantly. In the winters 1995 and 2002, the differences between all RegCM3 experiments and ECMWF ACCs are not statistically significant. In the summer 1996, the differences between global model and RegCM3 are statistically significant, while in the summer 1999 models do not differ significantly in statistical sense (i.e., the regional model skill is comparable to the skill of the global model). In the summers 1991 and 1993, the RCM_DFSM and RCM_EISM ACCs are comparable to those in global model, and in the summer 2001, only RCM_ECSM (open triangle) does not significantly differ from the ECMWF ACC.

When the summation in (1) is performed over time, a geographical distribution of ACCs is obtained. In Fig. 6, the geographical distribution for the global model ACCs is compared with RCM_ECSM only, since all the other RegCM3 experiments have distributions similar to RCM_ECSM (not shown). Statistical significance of anomaly correlations at the 95% confidence level is estimated by using the one-tailed Student’s \( t \) test against the null hypothesis of zero correlation. The black solid contours (corresponding to the ACC value 0.52) in Fig. 6 enclose the area with statistically significant ACCs, and the percentage of the area with statistically significant ACCs, relative to the total RegCM3 land area, is shown in the top left corner of Figs 6a–d and in Table 2. Statistical significance of the differences in ACCs between ECMWF and RCM_ECSM was assessed by using the two-tailed \( t \) test for the difference between two nonindependent correlations (Howell 2009) at the 95% confidence level. Although there are regions where correlation coefficients between ECMWF and RCM_ECSM differ significantly (denoted by black dots in Figs. 6e,f), they mostly do not coincide with the regions of significant skill.

Since in winter there is a weak statistically significant skill (around 5% of the total area in both models), we discuss only the summer distribution of ACCs. The area of a relatively high skill extends over the southern part of the domain (Figs. 6b,d) where warm temperatures prevail (cf. Fig. 2f). These results agree with those from Branković and Palmer (2000), who found a higher level of predictability in ECMWF summer seasonal predictions of the low-troposphere temperature in southern (south of 50°N) than in northern Europe. Some studies relate this kind of seasonal predictability to the impact of SM anomalies. For example, Kanamitsu et al. (2003) have demonstrated that, because of a generally low
predictability of precipitation, SM is more predictable in the arid and semi-arid regions than in the wet areas. This is related to the SM memory in warm climates, which is longer in dry than in wet conditions (Wu and Dickinson 2004). It could be asserted therefore that a part of the higher T2m forecast skill over southern Europe in summer is associated with a relatively low SM content that brings about an improved predictability of SM anomalies. In high latitudes, the shorter time scale of SM memory is coupled with snowmelt or soil water phase change (Wu and Dickinson 2004), which might negatively affect the T2m ACCs in this region.

The area of statistically significant summer ACCs in RCM_DFSM and RCM_EISM experiments is reduced when compared to RCM_ECSM (Table 2), particularly in the Mediterranean and the adjacent regions (not shown), indicating the impact of initial SM on the T2m seasonal forecasts skill. Conil et al. (2007) argued that the realistic SM leads to a better representation of T2m anomalies and their spatial structure in Europe during summer, which in turn is relevant for predictive skill of
T2m (Douville 2010). Based on consideration that the SM initializations in RCM_ECSM has the most beneficial impact on systematic errors in the RegCM3 T2m (cf. the previous section), we may infer that the initial SM in the RCM_ECSM experiment is more realistic compared to the other two RegCM3 experiments, thus enhancing the T2m skill. However, even in RCM_ECSM the area with statistically significant skill is smaller than in global model.

b. Verification of probability forecasts

In addition to providing deterministic skill, ensembles of (seasonal) forecasts enable the usage of more complex
probabilistic information. Ensemble predictions are transformed into probabilities by calculating a fraction of ensemble members predicting a particular event. Probabilistic verification is essentially based on the comparison of forecast probabilities of an event with observed frequencies of the same event. For this purpose, forecasts are categorized into probability bins and the number of bins usually equals to the number of ensemble members plus one. Doblas-Reyes et al. (2008) have shown that any smaller number of bins would reduce the forecast Brier skill score (BSS; Wilks 2006) with respect to the BSS obtained by using all possible probability values determined by the ensemble size. The increased number of categories enhances the resolution of forecasts, which may contribute to a higher skill score. Thus, for our seasonal integrations consisting of nine ensemble members, ten probability bins were formed.

1) BRIER SKILL SCORE AND RELIABILITY DIAGRAMS

Probabilistic skill of an ensemble of forecasts can be determined by using BSS, which in turn is based on the Brier score (BS) as the basic accuracy measure (Brier 1950; Wilks 2006):

\[
BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)^2, \tag{2}
\]

where \( p_i \) is forecasts probability of a given event, \( o_i \) is observation of the same event (\( o_i = 1 \) if event occurred, otherwise \( o_i = 0 \)), and \( N \) is the number of forecast–event pairs. As in the above discussion of deterministic skill, “observed” events were derived from CRU data. The BS can be decomposed into three terms (Murphy 1973; Wilks 2006):

\[
BS = \frac{1}{N} \sum_{i=1}^{I} N_i(p_i - \bar{\sigma}_i)^2 - \frac{1}{N} \sum_{i=1}^{I} N_i(\bar{\sigma}_i - \bar{\sigma})^2 + \bar{\sigma}(1 - \bar{\sigma}), \tag{3}
\]

where \( I \) is the number of probability bins, \( N_i \) is the number of forecasts in the \( i \)th bin, \( \bar{\sigma}_i \) is the relative frequency of observations in the \( i \)th bin, and \( \bar{\sigma} \) is the overall relative frequency of the event (sample climatology). The first two terms in (3) determine reliability \( B_{rel} \) and resolution \( B_{res} \) of forecasts respectively, and the last term is Brier score of a climatological forecast \( B_{cli} \).

Since the last term in (3) is not related to forecasting system, the Brier skill score and its components, reliability \( B_{cli} \) and resolution \( B_{res} \), can be calculated relative to \( B_{cli} \) from BS, \( B_{rel} \), and \( B_{res} \), respectively (Palmer et al. 2000), thus yielding

\[
BSS = 1 - \frac{BS}{B_{cli}}, \tag{4}
\]

\[
B_{rel} = 1 - \frac{B_{rel}}{B_{cli}}, \tag{5}
\]

\[
B_{res} = 1 - \frac{B_{res}}{B_{cli}}, \tag{6}
\]

For perfect deterministic forecasts, each individual measure—BSS, \( B_{rel} \), and \( B_{res} \)—is equal to 1.

Since the ACCs analysis in section 4a indicates that the skill of both models is relatively higher in JAS, we show and discuss the above probability skill measures for three events when T2m anomalies in JAS are above normal (0.0°C), above 0.5°C, and below −0.5°C in the analysis domain (region A) and in its southern part (region B). This is essentially indicated in Fig. 4b where ensemble-mean anomalies in the analysis domain range between −0.5°C and 0.5°C and anomalies above 0.5°C and below −0.5°C are considered as rare events. Since the thresholds for rare events may vary regionally and seasonally and they depend on the model and observations, we have chosen the above thresholds because they represent rare anomalies in the large part of the analysis domain.

The graphical representation of reliability and resolution is given by reliability diagrams (e.g., Wilks 2006). Reliability diagrams are normally accompanied by forecast frequency histograms that determine the sharpness of the forecasts. The calibration function plotted on the diagrams in Fig. 7 indicates conditional observed probability of the event as a function of forecast probability. Forecast probabilities from all land grid points in the analysis domain over 11 years were used (36 718 cases in total). Here, we compare reliability diagrams for each of the three RegCM3 experiments with that for global model. For positive anomalies (left column in Fig. 7), calibration functions in reliability diagrams are similar to each other, revealing conditional biases typical of seasonal forecasts (Mason and Stephenson 2008). For example, small forecast probabilities are underforecast; that is, forecast probabilities are too small relative to the event

<table>
<thead>
<tr>
<th></th>
<th>ECMWF</th>
<th>RCM_DFSM</th>
<th>RCM_ECSM</th>
<th>RCM_EISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>JFM</td>
<td>5.0</td>
<td>4.9</td>
<td>5.1</td>
<td>4.6</td>
</tr>
<tr>
<td>JAS</td>
<td>33.1</td>
<td>20.9</td>
<td>27.9</td>
<td>17.8</td>
</tr>
</tbody>
</table>

TABLE 2. Percentage of the area, relative to the RegCM3 land points, with statistically significant ACC for ECMWF model and three regional model experiments with different soil moisture initializations.

For positive anomalies (left column in Fig. 7), calibration functions in reliability diagrams are similar to each other, revealing conditional biases typical of seasonal forecasts (Mason and Stephenson 2008). For example, small forecast probabilities are underforecast; that is, forecast probabilities are too small relative to the event
FIG. 7. Reliability diagrams and frequency histograms (gray bars) in JAS in the analysis domain (region A) for ECMWF and RegCM3 T2m anomaly (a)–(d) >0.0°C, (e)–(h) >0.5°C, and (i)–(l) <0.5°C. The horizontal dashed line corresponds to the observed climatological frequency of the event. The diagonal line represents perfect reliability.
Anomaly, 2

Although reliability

However, B

correspond to the larger and smaller observed relative frequencies, respectively.

For both models, reliability diagrams exhibit poor resolution $B_{\text{res}}$ because frequencies of occurrence do not depend strongly on forecasts probability and tend to equal climatological probability of the event (denoted by the horizontal dashed line in reliability diagrams). This is also pronounced for the anomaly greater than 0.5°C where the outcome relative frequencies for large probabilities tend, more in RegCM3 experiments than in the global model, to equal climatological probability of the event (Fig. 7, middle column). The lower resolution in RegCM3 is confirmed in Table 3 by consistently smaller $B_{\text{res}}$ than that for ECMWF for all events considered.

Although reliability $B_{\text{rel}}$ in all RegCM3 experiments for positive thresholds is close to that of ECMWF (Table 3), because of the better resolution, the global model BSSs are higher than those for regional model. Some BSS values in RCM_DFSM and RCM_EISM are negative, implying that the forecasts are worse than climatology.

For the negative threshold (right column in Fig. 7), small forecast probabilities are underforecast, similar to the results for positive anomalies discussed above. Visual inspection indicates that RegCM3 forecasts are better calibrated than ECMWF forecasts; this is also confirmed by slightly higher values of $B_{\text{rel}}$ for RegCM3 than for ECMWF in Table 3. A good agreement between observed frequencies of this event and the RegCM3 forecast probabilities is most noticeable for large probabilities. Such high probability forecasts are very rare—only a few cases over all grid points in 11 years—and are also seen in frequency histograms in the bottom row of Fig. 7 (see also the discussion below). However, $B_{\text{res}}$ for RegCM3 in Table 3 are smaller when compared to the global model forecasts, eventually affecting the skill (BSS) of the regional model to be smaller than that of ECMWF.

Frequency histograms in Fig. 7 generally exhibit poor sharpness because forecast probabilities peak near climatological probability (sharpness normally refers to the degree by which forecasts depart from climatology). Most of the histograms in Fig. 7 indicate small samples for small and high probability bins, whereas perfectly sharp forecasts tend to forecast probabilities close to 0% and 100% (the U-shaped frequency histogram).

Since Figs. 6b and 6d indicate that statistically significant ACCs in summer are found over the southern part of the analysis domain, we assess models’ performance in terms of probabilistic verification for region B. Probabilistic forecasts of both models for this subdomain (Fig. 8; Table 4) are more reliable and have a better resolution and skill than for the analysis domain A (Fig. 7; Table 3). This is in agreement with the results of Kirtman and Pirani (2008), who found statistically significant BSS in the Mediterranean region in an ensemble of seasonal forecasts from the Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER) project (Palmer et al. 2004). Although the BSS values for all RegCM3 experiments are now better than climatology, the improvement for the southern part of the domain is more noticeable in the global than in the regional model. Despite the improvement in RegCM3 for positive anomalies, large probabilities are still overforecast and low probabilities are underforecast, although to a lesser extent than for the whole analysis domain (cf. the left panels in Figs. 7 and 8).

To summarize, only for negative anomalies is the BSS for RCM_ECSM in the analysis domain (region A) equal to that of global model and nearly equal in the southern portion of the domain (region B). For this anomaly, the differences in BSS between ECMWF and RCM_ECSM are not statistically significant. Otherwise, in most combinations of the thresholds and regions, the BSS values of the global model are higher than those of RegCM3, and these differences are statistically significant. The only significantly higher quantity in RegCM3 than in ECMWF is reliability in RCM_ECSM for positive anomalies in the

| Table 3. Brier skill scores for T2m anomaly >0.0°C, >0.5°C and <−0.5°C in global and regional model integrations in summer for the RegCM3 analysis domain (region A). Statistically significantly higher RegCM3 scores are in bold. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | BSS             | B_{\text{rel}}  | B_{\text{res}}  |                 |
|                | ECMWF           | RCM_DFSM        | RCM_ECSM        | RCM_EISM        | ECMWF           | RCM_DFSM        | RCM_ECSM        | RCM_EISM        |
| Anomaly > 0.0  | 0.04            | −0.01           | 0.02            | 0.00            | 0.95            | 0.94            | 0.96            | 0.95            |
| Anomaly > 0.5  | 0.05            | −0.02           | 0.01            | −0.03           | 0.95            | 0.93            | 0.95            | 0.93            |
| Anomaly <−0.5  | 0.04            | 0.01            | 0.04            | 0.01            | 0.95            | 0.96            | 0.96            | 0.96            |

Downloaded from http://journals.ametsoc.org/doi/pdf/10.1175/MWR-D-11-00104.1 by guest on 15 June 2020
FIG. 8. As in Fig. 7, but for the southern part of the analysis domain (region B).
analysis domain. For resolution, RegCM3 is comparable to ECMWF only in the RCM_ECSM experiment for negative anomalies. Thus, in terms of probabilistic verification RCM_ECSM emerges as the best of all RegCM3 experiments, but overall it is hardly better than ECMWF.

2) RELATIVE OPERATING CHARACTERISTIC

The relative operating characteristic (ROC) curve aggregates in a single graph the relationship between the hit rate $HR_n$ and the false alarm rate $FAR_n$ for accumulated probability bins (i.e., for each probability threshold $n$). They are defined as (WMO 2005)

$$HR_n = \frac{\sum_{i=1}^{I} O_i}{\sum_{i=1}^{I} O_i},$$ (7)

$$FAR_n = \frac{\sum_{i=1}^{I} NO_i}{\sum_{i=1}^{I} NO_i},$$ (8)

where $i$ is the bin number, $O_i$ is the number of observed occurrences, and $NO_i$ is the number of observed non-occurrences in the $i$th probability bin. The ROC diagram indicates the ability of probability forecasts to discriminate occurrences from nonoccurrences. The area below the ROC curve (ROC score) is a measure of potential forecast skill (i.e., the skill of correctly calibrated forecasts; Wilks 2006) with a perfect value being 1, while for the no-skill forecasts (climatological frequency) the hit rate and false alarm rate are equal (and lie on the diagonal line of ROC diagram), yielding a ROC score equal to 0.5.

For T2m anomalies in JAS discussed above, the characteristics of ROC scores are similar to those for reliability diagrams: forecasts of both models show a better discrimination for the southern part of the analysis domain than for the whole analysis domain (Table 5). For all thresholds (and in both regions), the ROC scores for global model are higher than those for regional model and the differences between them are statistically significant, except for the negative anomaly in RCM_ECSM where the ROC score is comparable to that of ECMWF.

By computing the area below the ROC curve for each grid point, the spatial distribution of ROC score can be obtained. It gives an insight into the spatial variation of probabilistic skill and may indicate regions with comparably high potential skill. In summer, geographical distribution of ROC score is consistent with the spatial distribution of ACCs (i.e., with the deterministic skill; not shown, but cf. Fig. 6). For both models, ROC scores greater than the no-skill value of 0.5 are found mostly in the area to the south of 50°N.

We also consider ROC scores for winter for the same thresholds as defined for JAS. That is, from Fig. 4a the T2m anomalies above 0.5°C and below −0.5°C can be assumed to be rare events in winter as well. The spatial distribution of ROC scores for the winter positive anomalies (>0.0°C) in global model and RCM_ECSM (Fig. 9) also corresponds to that for the ACCs’ distribution (cf. Figs. 6a,c). Relatively high ROC scores are mostly in the central part of the domain (in Fig. 9 only grid points with ROC area over 0.5 are shaded). RCM_ECSM shows improvement over ECMWF in eastern France and the Alps, Apennines, southern Balkans, and Turkey (Fig. 9c). Spatial distribution of the differences between the ECMWF and the RCM_ECSM ROC scores indicates only few grid points where these differences are

### Table 4. As in Table 3, but for the southern part of the RegCM3 analysis domain (region B).

<table>
<thead>
<tr>
<th>Anomaly</th>
<th>ECMWF</th>
<th>DFSM</th>
<th>ECSM</th>
<th>EISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.0</td>
<td>0.21</td>
<td>0.13</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>0.14</td>
<td>0.07</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>&lt;−0.5</td>
<td>0.17</td>
<td>0.12</td>
<td>0.16</td>
<td>0.13</td>
</tr>
</tbody>
</table>

### Table 5. Area under the ROC curve in summer for T2m anomaly >0.0°C, >0.5°C, and <−0.5°C in global and regional model integrations over regions A and B.

<table>
<thead>
<tr>
<th>Region</th>
<th>ECMWF</th>
<th>DFSM</th>
<th>ECSM</th>
<th>EISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;0.0</td>
<td>0.663</td>
<td>0.617</td>
<td>0.642</td>
<td>0.627</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>0.670</td>
<td>0.619</td>
<td>0.642</td>
<td>0.608</td>
</tr>
<tr>
<td>&lt;−0.5</td>
<td>0.666</td>
<td>0.628</td>
<td>0.663</td>
<td>0.631</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;0.0</td>
<td>0.761</td>
<td>0.703</td>
<td>0.727</td>
<td>0.709</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>0.730</td>
<td>0.683</td>
<td>0.693</td>
<td>0.663</td>
</tr>
<tr>
<td>&lt;−0.5</td>
<td>0.763</td>
<td>0.731</td>
<td>0.763</td>
<td>0.733</td>
</tr>
</tbody>
</table>
However, if the ROC score is calculated from all grid points in the central part of the domain (depicted by the rectangle in Fig. 9b), for anomalies greater than 0.8°C higher in all RegCM3 experiments than in ECMWF and the differences between them are statistically significant (Table 6, first row).

This improvement of the RegCM3 winter skill for positive anomalies in the central part of the domain is also seen in Brier skill score and resolution (12,826 forecast–event pairs in total; Table 7); BSS and $B_{eq}$ in RegCM3 experiments are higher than in global model and the differences between them are all statistically significant. Reliability in all RegCM3 experiments, similar to what was found in summer for the analysis domain and for the southern part (Tables 3 and 4), remains comparable to that in global model. For the anomaly greater than 0.5°C, reliability in the central part of the domain is statistically significantly higher in all RegCM3 experiments than in ECMWF, mostly because observed frequencies correspond better to small and large forecast probabilities (not shown). For winter rare cold events, BSS is negative (i.e., forecasts of both models are worse than climatology). In summary, in the central part of the domain, the RegCM3 winter skill scores for positive anomalies improve over ECMWF in terms of both BSS and ROC. This improvement is seen in all RegCM3 experiments, indicating that the SM initialization in winter has no major effect on probabilistic skill.

5. Summary and conclusions

The regional climate model, RegCM3, has been used to dynamically downscale the ECMWF experimental seasonal forecasts over Europe. Downscaling was performed on nine-member winter and summer ensembles for the 11-yr period 1991 to 2001. Three downscaling experiments with the different soil moisture (SM) initialization were carried out. In the default experiment (RCM_DFSM), initial SM is defined as a function of land cover/soil type and vegetation type. This type of initialization includes no seasonal or interannual variations of SM. The second set of experiments (RCM_ECSM) used initial SM from ECMWF seasonal forecasts at appropriate initial times, and in the third experiment (RCM_EISM) the initial SM was calculated as winter and summer climatology derived from RegCM3 integrations that were

<table>
<thead>
<tr>
<th>Anomaly</th>
<th>ECMWF</th>
<th>RCM_DFSM</th>
<th>RCM_ECSM</th>
<th>RCM_EISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.0</td>
<td>0.661</td>
<td>0.679</td>
<td>0.690</td>
<td>0.685</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>0.734</td>
<td>0.724</td>
<td>0.718</td>
<td>0.711</td>
</tr>
<tr>
<td>&lt;−0.5</td>
<td>0.584</td>
<td>0.560</td>
<td>0.573</td>
<td>0.565</td>
</tr>
</tbody>
</table>
driven by ERA-Interim data. The 2-m temperature forecasts from global model and all RegCM3 experiments have been assessed and compared using various forecast quality measures: systematic errors, anomaly correlations, and verification of probability forecasts.

In both seasons, ECMWF exhibits cold bias whereas, because of the coarse resolution, warm bias in T2m is related to high altitudes (e.g., the Alps and the Carpathian mountains). The smallest T2m bias is found over the northeastern part of the domain where in winter cold temperatures prevail. Although RegCM3 gives more detailed structure of the temperature field than ECMWF, it underestimates T2m in both seasons, particularly in summer. In comparison with global model, systematic errors in RegCM3 are reduced in winter over central and eastern Europe, and in summer in the southern part of the domain (region B in Fig. 1b) in RCM_ECSM.

In summer, the RCM_ECSM T2m is closest to the observations of all RegCM3 experiments. This experiment, which includes interannual variation in SM consistent with the other initial data, represents the most realistic initialization of SM, thereby leading to an improved prediction of T2m. The largest improvement in RCM_ECSM is seen in summer over the relatively dry southern part of the domain. For RCM_DFSM and RCM_EISM, the results are poorer than for RCM_ECSM, but RCM_EISM is generally better than the default experiment because its initial SM was derived in a physically consistent way with the other initial data (as seasonally varying “climatology”). In winter, when, on average, wet soil conditions prevail, soil moisture initialization makes no or very little effect on the RegCM3 T2m errors and predictive skill.

Time series of the T2m anomaly correlation coefficients (ACCs) for both models reveal a large interannual variability. In winter, high RegCM3 ACCs are not significantly different from those of global model; in summer, in most cases the ACC from at least one of the RegCM3 experiments is comparable to the ACC in global model. The geographical distribution of ACCs indicates that, irrespective of the model, the total area with statistically significant skill is much smaller in winter than in summer, when relatively high ACCs are located in the Mediterranean region and southern Europe. For this skill metric the global model in summer generally outperforms RegCM3, but of all the RegCM3 experiments, the largest area with significant ACC and the best predictive skill of T2m is found in RCM_ECSM.

In terms of probabilistic verification, for positive summer anomalies small probabilities over the analysis domain (region A in Fig. 1b) are underforecast and, conversely, large probabilities are overforecast, irrespective of the model considered. These biases are more pronounced in RegCM3 than in the global model and contribute to its overall lower (and in some cases even negative) Brier skill score (BSS) than in ECMWF. However, for rare cold events (T2m anomaly less than \(-0.5^\circ\text{C}\), RegCM3 high probability forecasts are better calibrated than in the global model and contribute to a positive BSS in all RegCM3 experiments. If we focus on the southern part of the domain, the BSS is increased in both models when compared to the analysis domain, indicating a higher predictability for this subdomain than for the entire analysis domain. Of all RegCM3 experiments, the best BSS is obtained with RCM_ECSM for all thresholds and in both regions considered. However, even then the RCM_ECSM probabilistic skill remains lower than that for global model (except for the rare cold anomalies), mainly because of significantly lower resolution.

Geographical distribution of the potential probabilistic skill, measured by the area under the ROC curve, is for both models in both seasons consistent with the distribution of the deterministic skill measured by ACCs. In both models, the ROC score in summer is better in the southern part of the domain and in winter in its central part (region C in Fig. 1b) relative to the whole analysis domain. However, this summer improvement in the southern areas is more evident in ECMWF than in RegCM3, except for the rare cold anomalies. For the winter anomalies greater than \(0.0^\circ\text{C}\), probabilistic skill of all RegCM3 experiments is significantly better than ECMWF in the central part of the domain (which is the region with the smallest RegCM3 systematic errors in T2m).

We have demonstrated that the better-resolved details in near-surface temperature by RegCM3 do not
necessarily lead to the improved regional model skill (i.e., regional model skill is closely related to the model’s systematic errors). Although the T2m errors in RegCM3 are generally not much reduced when compared to global model, in regions and seasons where such a reduction of errors is seen the determinist and probabilistic skill of regional model is improved relative to the global counterpart. A better exploitation of regional model potential for seasonal ensemble forecasts would therefore require a further reduction of the model’s systematic errors at small scales. However, this may not be an easy exercise to implement because some RegCM3 errors are inadvertently transmitted from the global model via the lateral boundary conditions.

The impact of initial SM on regional model’s T2m predictive skill is seen mainly in summer in a relatively dry part of the domain. In this region, realistic seasonally and annually varying initial SM contributes to the reduction of systematic errors, enhancing the model’s deterministic and probabilistic skill of the T2m temperature forecasts.

Acknowledgments. The authors are grateful to two anonymous reviewers for their valuable suggestions and comments that improved the original version of the manuscript. This study was partly supported by the Ministry of Science, Education and Sports of the Republic of Croatia (Project 004-1193086-3035).

REFERENCES


——, and M. Patarčić, 2008: Downscaling of ECMWF seasonal integrations by RegCM. Proc. ECMWF Workshop on Ensemble Prediction, ECMWF, Reading, United Kingdom, 141–150.


——, ——, and L. Srnec, 2004: Seasonal dynamical downscaling of the PRUDENCE model project ions of changes in European climate by the end of this century. Climate Change, 28, 125–145.


