Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2009), 61A, 357–370

TELLUS

Evaluation of a regional climate model using in situ temperature observations over the Balkan Peninsula

By E. KOSTOPOULOU[1]*, K. TOLIKA[2], I. TEGOULIAS[2], C. GIANNAKOPOULOS[1], S. SOMOT[3], C. ANAGNOSTOPOULOU[2] and P. MAHERAS[2], [1]*Institute for Environmental Research and Sustainable Development, National Observatory of Athens, I. Metaxa & V. Pavlou, GR-15236 Palaia Pendeli, Athens, Greece;* [2]*Department of Meteorology and Climatology, Aristotle University of Thessaloniki, University Campus GR-54124, Thessaloniki, Greece;* [3]*Centre National de Recherches Météorologiques, GAME/CNRM (Météo-France, CNRS), 42 avenue G. Coriolis, 31057 Toulouse cedex 1, France*

(Manuscript received 2 April 2008; in final form 4 December 2008)

ABSTRACT

Climate model data provide large, dense coverage and long time-series, characteristics that are advantageous in the study of climate. However, it is not recommended that such data be used in any region without prior evaluation of their reliability based on comparisons with in situ observations. In this study, the accuracy of maximum and minimum temperature data from the ALADIN-Climate regional climate model (with a 25-km horizontal resolution driven at the lateral boundaries by the ERA-40 reanalysis) have been assessed for 53 stations across the Balkan Peninsula. The model temperatures corresponding to each station were extracted from their nearest land grids. The model data were first compared with observations and subsequently examined for their ability to identify extreme temperature events. In general, the model was found to be quite accurate in describing the seasonal cycle, as well as simulating the spatial distribution of temperature. Simulations were more realistic for stations along coastlines, highlighting the constraints of the topographic forcing in the simulations. Assessing the performance of the model to determine extremes (warm and cold spells), it was found to be better at detecting cold spells and has a tendency toward overestimating the frequency of occurrence of warm spells, particularly in summer.

1. Introduction

In the past, climate research suffered from limitations imposed by data availability, and research was limited to the national scale due to data exchange constraints. In recent decades, significant progress has been made towards the construction of gridded data sets with high spatial and temporal resolutions. The wide availability of global and regional gridded climate data has facilitated climate research and applications and made these data sets popular within the research community. The two most widely used long historical atmospheric analyses are those implemented at the European Centre for Medium-Range Weather Forecasts (ECMWF) and at the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR). The ECMWF reanalysis project, known as ERA-40 (Simmons and Gibson, 2000; Uppala et al., 2005), uses the T159L60 version of the Integrated Forecasting System to produce analyses every 6h. The horizontal resolution of the NCEP–NCAR assimilation model is 210 km (a T62 global spectral model was used) and produces 28 unequally spaced sigma-levels in the vertical (Kalnay et al., 1996; Kistler et al., 2001).

The construction of these data sets depends largely on numerous parametrizations of multidecadal sequences of past meteorological observations. Consequently, problems often arise regarding data reliability due to biases in the atmospheric assimilation models (Simmons et al., 2004; Zhao and Fu, 2006), changes in the observational system (Jones, 1999; Bengtsson et al., 2004a; Bromwich and Fogt, 2004), jumps in the historical records, as well as the sensitivity of certain meteorological variables in physical parametrizations (Rusticucci and Kousky, 2002). Validation of model estimates with observed data is thus considered essential to assess the reliability of modelled data at regional scales. Several studies have attempted to assess the skill of the reanalyses and have revealed that accuracy improves with the incorporation of satellite observations (Kistler et al., 2001; Bengtsson et al., 2004b; Simmons et al., 2004). The reanalyses seem to better represent climatological features over extended

*Corresponding author.
e-mail: ekosto@meteo.noa.gr
DOI: 10.1111/j.1600-0870.2009.00389.x

areas (Hertzog et al., 2006), whereas the accuracy of data at regional scales may vary from one variable to another (Reid et al., 2001).

Moreover, it is common to use a regional climate model (RCM) to increase the spatial resolution of a reanalysis data set, which is used to define the boundary conditions. This is different to other common RCM data sets used for climate change studies, where boundary information is derived from a global climate model (GCM) or an atmosphere–ocean general circulation model (AOGCM), and so, the temporal evolution (year-by-year) of the control period cannot be compared directly.

However, a large amount of uncertainty can be introduced into model estimates. Therefore, it is recommended that models are run for the recent past and validated against observed meteorological station data. Additionally, it is of great interest to evaluate climate extremes as simulated by models, as they are important aspects of climate (Kharin and Zwiers, 2000). In particular, the effects of anthropogenic emissions on the magnitude and the frequency of extreme events are associated with socioeconomic and environmental consequences and thus constitute the study of extreme climate conditions equally important to the study of changes of the mean climate. Principal aim of the current study is to compare modelled maximum and minimum temperature RCM data with observational data at 53 sites in the eastern Mediterranean basin, as well as extreme climate indices calculated by both data sets. The outline of the paper is as follows: in Section 2, the data and method are introduced; Section 3 presents results of the comparisons between model and observed temperature data; Section 4 investigates the ability of the RCM to reproduce extreme temperature events and analyses the trends of extreme temperature indices as derived by both model and station data sets; Section 5 discusses the intercomparison results and, finally, Section 6 summarizes the main findings of the study.

2. Data and method

The RCM data utilized in this study, consist in a limited-area simulation at 25-km horizontal resolution driven at the lateral boundaries by ERA-40 reanalysis data (Uppala et al., 2005). This RCM, called Aire Limitée Adaptation Dynamique Développement International (ALADIN)-Climate, has been developed at Météo-France/Centre National de Recherches Météorologiques (CNRM) and is described in Déqué and Somot (2007) and Radu et al. (2008). Its physical parametrizations (Radu et al., 2008) is common with the Action Recherche Petite Echelle Grande Echelle (ARPEGE)-Climate GCM, commonly used in a coupled mode by the IPCC or in an atmosphere stretched-grid mode for regional climate modelling. The ERA40-driven ALADIN-Climate simulation analysed in the current study has been generated within the framework of the EU-Project ENSEMBLES. It spans the 44-year ERA40 period and its geographical domain extends from Iceland to Egypt, covering the whole of Europe with 213 × 213 gridpoints, including Scandinavia and the Mediterranean Sea. The area studied here is situated within the southeastern corner of the domain but not too close to any lateral boundary.

The current validation study was undertaken to (1) compare RCM daily maximum (T_X) and minimum temperature (T_N) simulations against observed temperature data from 53 meteorological stations distributed throughout the Balkan Peninsula and western Turkey and (2) assess the ability of the model to simulate climate extreme indices. Figure 1 shows the geographical locations of the stations and the corresponding gridpoints used.

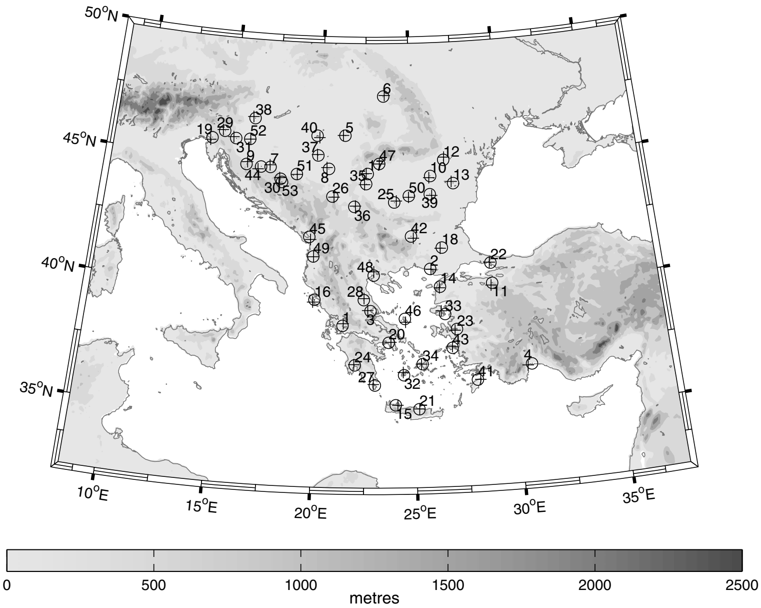

Fig. 1. Topographic map of the study area showing the locations of the meteorological stations (circles) and the corresponding gridpoints (crosses) used. Each station is assigned a map index number (see Appendix A).

Each station is represented by a map index number, and Appendix A provides a list of the stations, their identifiers in ascending numerical order and the latitude/longitude coordinates and elevation of each station.

As air temperature decreases with height, prior to further analysis the elevations of the two data sets were compared and where necessary a mean adjustment, using the constant lapse-rate of 6.5K km^{-1}, was applied to the extracted model data to attain better similarity between the assessed data sets. The data consists of daily records between January 1958 and December 1990. Although for many stations observational data were available for longer and more recent time periods, other stations suffered from large periods of missing data. This was particularly problematic for the newly formed Balkan countries where, during the political instability of the 1990s, many of the meteorological stations were forced to stop operating for long periods. The reliability of the observational data was examined in a previous study using basic quality tests (Kostopoulou and Jones, 2007b) and homogeneity tests based on the Alexandersson (1986) approach have been performed on a monthly basis for the Greek stations (Kostopoulou and Jones, 2007b). To obtain an even geographical distribution of stations throughout the study region (particularly stations located in relatively low-elevation regions) and to maximize the number of stations included in the analysis, we chose to truncate the study period to the 1961–1990 period. A period of 30yr constitutes a widely acceptable time period for climate research; in addition, the World Meteorological Organization recommends the use of the 1961–1990 reference period for climatological evaluations. This 30-year period is commonly used in climate models as the control run period (Giorgi et al., 2004; Déqué et al., 2005; Jacob et al., 2007) to calculate changes for future 30-year time slices. Moreover, 30yr of data could be considered short enough to not reveal warming trends caused by recent urban growth (Moberg and Jones, 2004).

The extraction of the model data was based on selecting the nearest RCM land gridpoint to each observation location. Another approach also taken, was to average the nine closest RCM points to each station, to prevent discontinuities caused by the heterogeneous terrain. However, when using the nine grid averages, the analysis tends to smooth extremes. Hence, since the goal of the current study is to assess the model's ability to simulate climate extremes, the nearest grid option is more appropriate.

3. Comparison of modelled and observed temperature data

3.1. Homogeneous zones of temperature

Homogeneous regions of surface air temperature can be delineated owing to distinct climatic and geographical conditions. Principal component analysis (PCA) was applied to both the observational and model data sets to identify regions with homogeneous temperature characteristics. The purpose of PCA is to extract the maximum amount of variance from the data set with each component and consequently to define homogeneous groups of stations (Kostopoulou and Jones, 2007a).

For the PCA to monitor the intraannual variability of the data, each season was treated separately, leading to four inputs per year and thus increasing the number of cases to 120 (4 seasons $\times$ 30yr) for each station. The analysis distributed the study sites between the first two principal components (PCs; approximately 53% and 46% explained variance) and divided the original area into two main subregions (subregion 1 and subregion 2 corresponding to PC1 and PC2, respectively, Fig. 2). The PCA based on station data defined subregion 1 as the northern part of the study area and includes all stations situated north of 42°N, whereas subregion 2 covers the southern part with stations located south of 42°N and mostly close to sea level. The classification for T_X and T_N revealed remarkably similar results, providing a clear division of the domain in terms of the temperature regimes. The PCA using RCM data also defined two subregions—one to the north and one to the south—except, in this case, the southern group was restricted to stations situated on islands or coastal areas, thus highlighting the important role of continentality.

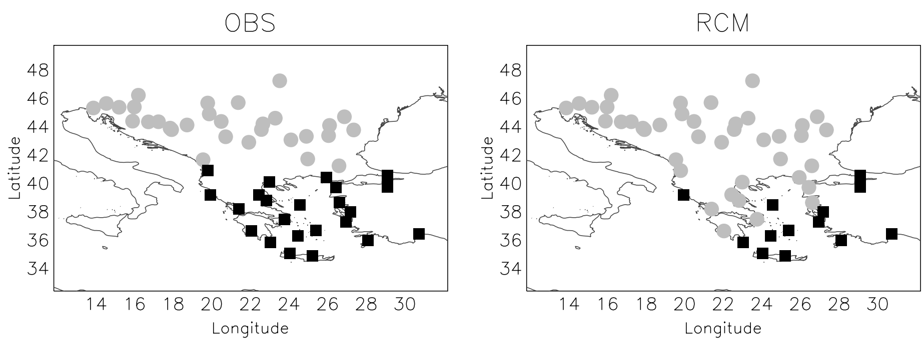

Fig. 2. The two subregions with homogeneous temperature characteristics as have been defined by PCA applied to observational (left-hand panel) and model (right-hand panel) data (subregion 1: filled grey circles, subregion 2: filled squares).

3.2. Temporal evaluation of the RCM

Seasonal temperature anomalies were calculated for each station and averaged over the two groups of stations (not shown), according to the classification derived from PCA. With respect to subregion 1, the correlation coefficients between the RCM and station data were quite weak, ranging between 0.26 and 0.67. Nevertheless, the RCM is still considered to provide an acceptable representation of the temperature conditions of the area (given the diversities of the large number of stations included in this group). Temperatures for this region were found to be better simulated for spring and autumn. Regarding subregion 2, station data are generally better reproduced (correlation coefficients range between 0.54 and 0.82), although maximum temperature exhibits stronger correlations between the two data sets. In general, this analysis demonstrated that winter T_N as well as both T_X and T_N for summer in subregion 1 are not well represented, highlighting that this region is one of the most difficult areas for an RCM to simulate with lateral boundary conditions situated in the Atlantic (Jacob et al., 2007). This fact also highlights the difficulty of reproducing extreme temperatures in this region. This issue will be further investigated in the second part of the paper.

The reproduction of seasonal cycles by the RCM was further examined for each station by calculating the mean value for each calendar day over the 30-year study period. Figure 3 shows the annual cycle of temperature for three selected stations, chosen because they represent various geographic locations and altitudes. Each graph contains three pairs of lines, with the middle pair representing the calendar day means. The upper and lower lines represent, respectively, the absolute maximum and the absolute minimum temperatures for each calendar day, during the period of study. Regarding T_X (left-hand panels of Fig. 3), the calendar day averages are well represented for most stations. In continental stations such as Turnu Magurele, the model overestimates the mean temperature during the cold season. This behaviour is more pronounced in the lowest recorded T_X values, and the positive temperature biases might be attributed to the strong continentality effect, which possibly is not very well simulated by the model. In contrast, the model captures quite well the intraannual variability of maximum temperatures at low-elevation sites (e.g. Thessaloniki). In high-elevation sites such as Godnje and Zenica (not shown), the simulation of the highest T_X values is rather poor, as individual warm and cold biases appear throughout the year and, in some cases, exceed $\pm5°$C. In stations located in low altitudes and close to the coast, for example, Thessaloniki or Istanbul (not shown), the model succeeds in reproducing T_X satisfactorily. Some notable cold biases are associated only with the highest recorded T_X, mainly during the cold months of the year.

The right-hand panels of Fig. 3 present the corresponding T_N graphs, which generally show that minimum temperatures are not simulated as well as maximum temperatures. Minimum temperature in Thessaloniki is better reproduced for the warm season, with the exception of a small warm bias in the highest T_N records. In winter, the model exhibits cold biases in all time-series. Godnje and Zenica (not shown) are both situated at high altitudes (at 320 and 344 m, respectively) on the Dinaric Alps, and their graphs present similar curves, revealing that the model overestimates T_N substantially. Specifically

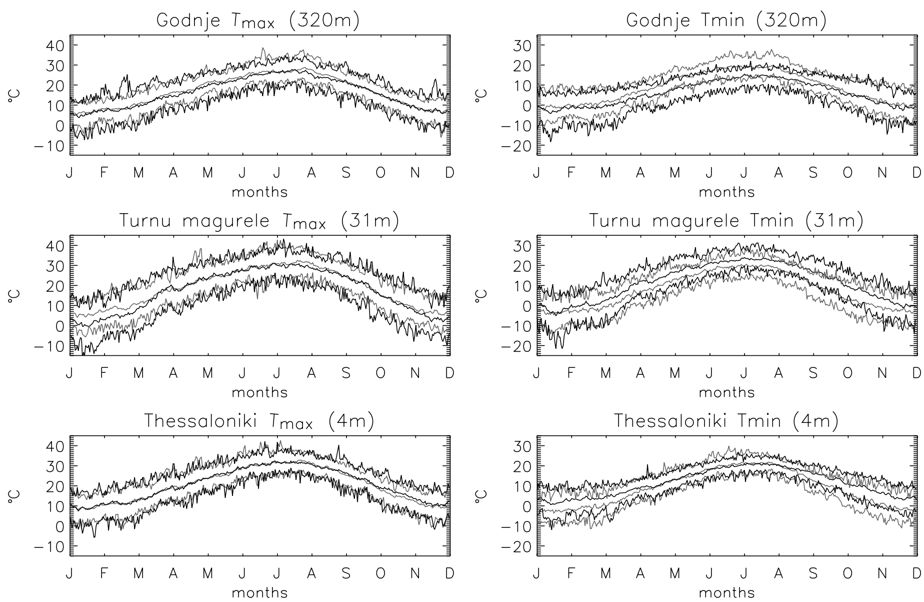

Fig. 3. Middle lines show the 30-year mean annual cycles of observed (black) and modelled (grey) T_{max} at three representative sites. The upper (lower) lines illustrate the absolute highest (lowest) T_{max} value for each calendar day during the study period.

during summer, the overestimation is so great that the mean model T_N is as high as the highest station T_N. Similarly, the lowest T_N of the model approaches the mean T_N values of the station. In addition, the model exhibits modest ability in representing exceptionally cold cases (lowest T_N records). The very low winter temperatures are generally associated with a surface temperature inversion, and such conditions, often caused by local features, are rather poorly simulated by a model. In addition, the surface snow scheme of the model could also be a source of error for T_N in winter (Christensen et al., 1997). This constitutes a weakness in climate models in general. In particular, the ALADIN snow scheme is based on the work described in Douville et al. (2000). This one-layer scheme includes prognostic equations for the snow mass, snow density and snow albedo. The ALADIN snow scheme is known to lead to too late melting of the snow in spring (Decharme and Douville, 2007), but too large or too small amount of snow in winter and spring could also be linked with bias in the atmospheric forcing in autumn and winter. Cold biases dominate the T_N curves of Turnu Magurele throughout the year, which in some cases can reach 4°C. These biases are possibly related to local effects (e.g. topographic relief, vegetation, infrastructure), which perhaps are not represented well in the model. In an evaluation study of the Hadley Centre HadRM3P model versus T_X and T_N observations, Moberg and Jones (2004) also found that Romania differed from its surrounding regions, as a coherent region in southeastern Europe, with particularly cold biases in T_N.

Overall, the previous analysis concludes that the model data are more reliable for T_X than T_N, and for the warm than the cold season. Therefore, model data for regions characterized by complex terrain must be treated with caution. Regarding the highest/lowest T_X and T_N values recorded over a course of years, the model was found to be more capable of simulating T_X extremes. Finally, the seasonal cycles for T_X and T_N of the two data sets using all-area averaged time-series were calculated (not shown). In this case, the simulations were in very good agreement with the observed data, except for a few minor biases in T_N. In particular, small cold biases were observed during the cold half of the year, whereas slight overestimations of T_N were found from May to August. The positive biases in summer in eastern-southeastern Europe are often associated with soils that are too dry (Moberg and Jones, 2004).

3.3. *Spatial evaluation of the RCM*

The accuracy of the model was initially assessed by correlation analyses undertaken between model and station observations using mean seasonal time-series. Subsequently, differences and standard deviations were computed and mapped. At this point, it is important to note that in this work, results referring to the sea are not neglected from the analysis, as this study focuses on a peninsula, and thus, the performance of the model over coastal regions and islands is equally meaningful. The reader, however, should bear in mind that only land data have been utilized in the evaluation analysis, and lapse-rate adjustments were made when necessary. The correlation results were examined for their statistical significance at the 0.05 level of confidence. Shaded maps present the correlation coefficients (r) between the data sets, with dark shading indicating a high degree of correlation. Dashed lines identify regions where correlations are statistically significant. With respect to maximum temperature (Fig. 4), during winter the largest part of the region reveals statistically significant correlation coefficients ranging between 0.5 and 0.7, whereas few stations located in the northwestern part of the area show coefficients below 0.5. In summer, significant correlations are shown in the western part of the region (r values in the range of 0.5–0.7), whereas sites situated in the north-northeastern sector of the domain showed low-to-moderate spatial correlations, with most correlation coefficients below 0.4. In these regions, convective precipitation episodes mainly occur during summer. This sort of meteorological condition often leads to a time decorrelation in RCMs, as convective situations are inherently associated with chaotic and unpredictable behaviour (Lucas-Picher et al., 2008). For the transition seasons (spring, autumn), the maps indicate that between the two data sets, the correlations generally increase with proximity to sea level. Specifically in spring, correlation coefficients are markedly higher over the northern Aegean Sea, with r exceeding 0.7. The autumn correlation coefficient pattern is similar to that for spring, as high correlation coefficients prevail across the Aegean Sea. In this case, the highest values (greater than 0.7) are aligned diagonally towards a southwesterly-northeasterly component. The weakest correlations for autumn, which nevertheless are approximately 0.5 and statistically significant, appear over the northeastern part of the domain, among stations in northern Bulgaria and Romania.

The correlation maps for winter T_N (Fig. 5) reveal fairly low correlations between model and observations in the northwestern part of the domain in the vicinity of the Dinaric Alps, whereas a cell of low correlations is also evident in northcentral Greece. Such low correlations ($r < 0.40$) in winter may be attributed to the surface snow scheme of the model. In the remaining region, statistically significant correlations are shown, with r values in the range of 0.5–0.7. The other three seasons demonstrate the strongest correlations around the Aegean Sea. In spring, the two data sets present moderate (r values between 0.5–0.6) but significant correlations north of 42°N, whereas over central Greece, the coefficients display small and insignificant values. In summer, weak and statistically insignificant correlations are identified across the Balkan Peninsula, which are probably due to the domination of convective weather patterns during the warm season. Statistically significant correlations exceeding 0.5 are found only across the Aegean Sea. In autumn, most of the correlations are found to be statistically significant, whereas high correlations ($r > 0.7$) are found again in low-elevation areas. Overall,

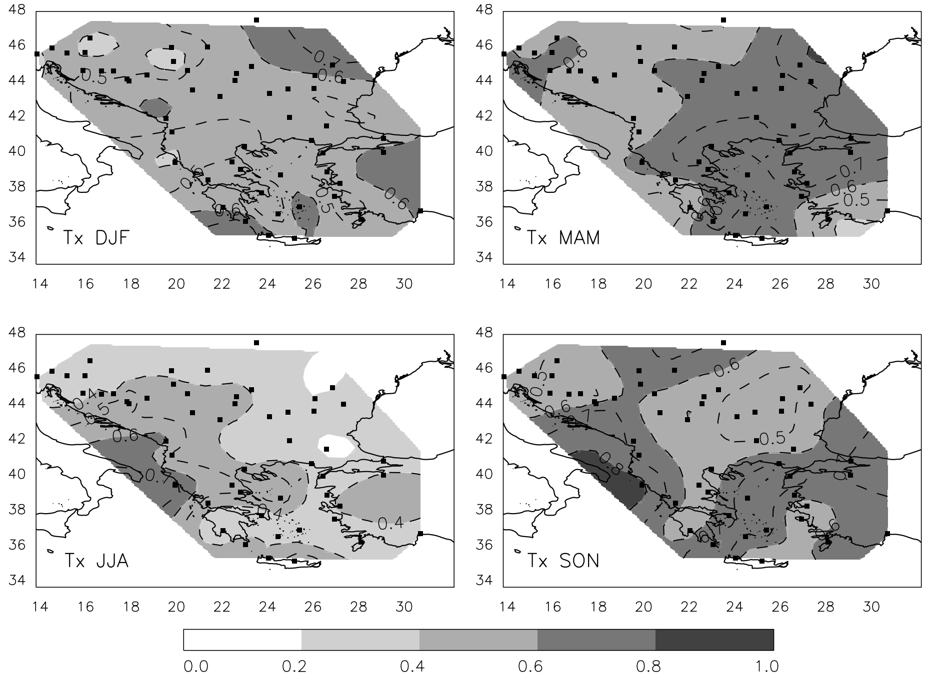

Fig. 4. Spatial correlation maps between model and station T_X data for the four seasons. Regions delimited by a dashed line indicate correlations that are significant at the 95% confidence level.

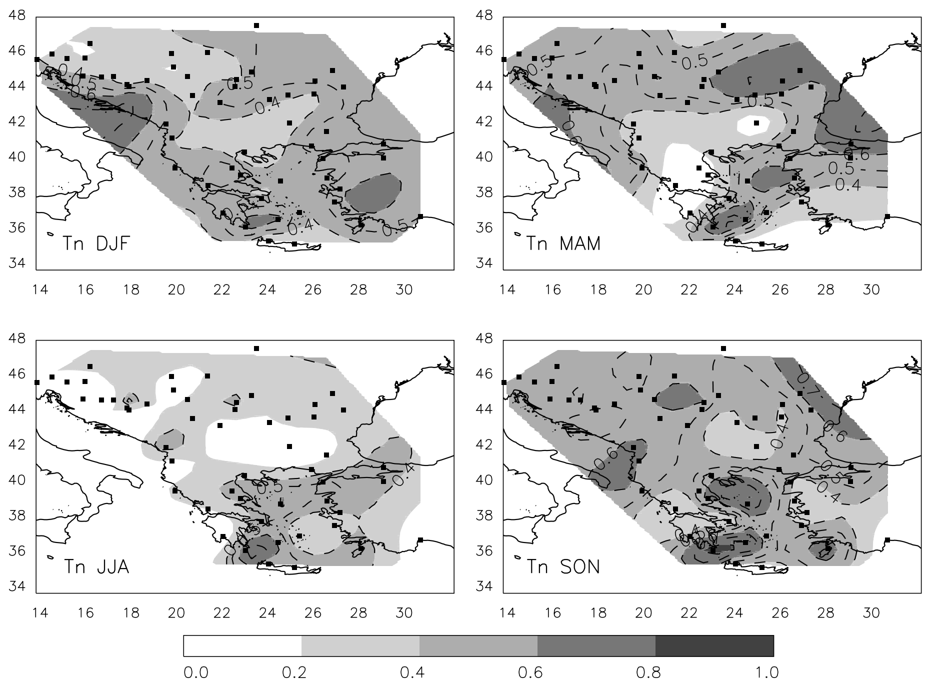

Fig. 5. As in Fig. 4 except for T_N.

the correlation analysis revealed that the correspondence between the two data sets (for all seasons) is best in stations close to sea level, whereas poor agreement between model and station data is found in high-altitude sites and particularly for winter T_N at stations close to the Dinaric Alps.

The seasonal differences (model minus observed) for T_X are illustrated in Fig. 6. The results were also subjected to a Student's *t*-test, and those found statistically significant are shown in shaded regions on the maps. The seasonal difference maps reveal a pattern characterized by a north-to-south gradient, with

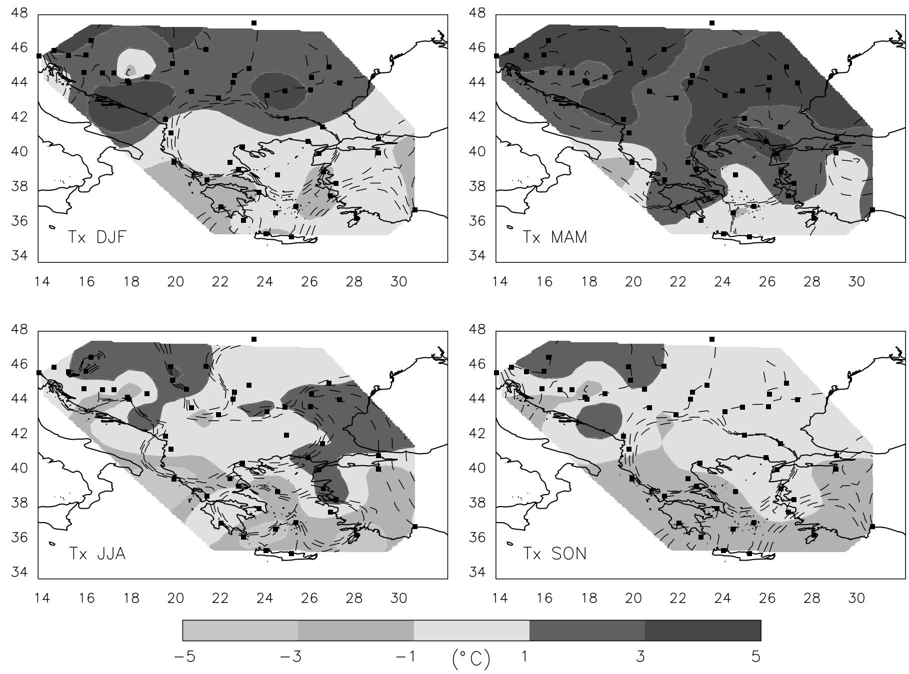

Fig. 6. Difference maps between model and station T_X data for the four seasons. The dashed line delineates regions of significant difference, as assessed using the Student's *t*-test.

positive values in the north and negative in the south of the study region. Overall, these maps show that the model tends to overestimate T_X in subregion 1 and underestimate it in subregion 2. In winter, the largest positive differences (up to 3°C) are observed in stations situated north of 43°N, especially in the eastern part of the domain. In contrast, underestimated T_X is observed along the southern part of the study region. Minimal differences are found over the Aegean Sea, which support the findings of the correlation analysis (Fig. 4), highlighting the ability of the model to better reproduce T_X in low altitude regions. A similar pattern is found in spring, with large positive differences reaching up to 4°C, marked along the northwest edge of the domain. The summer and autumn patterns reveal smaller differences (indicating that T_X is better reproduced in these seasons), which are statistically significant in the southern part of the study region. In summer, positive differences in the north are in the range of 1–2°C, whereas some large negative differences (up to −3°C) are only found in isolated cases such as in the island of Kythira. The results for autumn are more encouraging, as differences in the largest part of the study region are in the order of ±1°C between the compared seasonal means. The relative seasonal differences for T_N are presented in Fig. 7. In winter and autumn, the majority of differences are negative (ranging from −1 to −5°C) and statistically significant, indicating the tendency of the model to underestimate T_N. The spring and summer patterns, which are similar to one another, present large negative differences in the northeastern part of the region, reaching −5°C in spring at Romanian stations. This is in agreement with the findings of Moberg and Jones (2004), who stated that Romania stands out as a coherent region with large negative T_N biases. In summer, negative biases also appear at the Romanian stations, whereas across the remainder of the region, the model overestimates T_N. The largest positive biases exceed 3°C and occur at sites along the Dinaric Alps.

In addition, standard deviations of the examined data sets (modelled, observations) were calculated and their differences (*StDevRCM – StDevOBS*) displayed on maps (not shown). Generally, the seasonal patterns presented small differences, indicating that the model succeeds in capturing the natural variability of the station data. It is worth highlighting the substantial differences found between the standard deviations of the two data sets in summer T_X and winter T_N, particularly for stations located at northern latitudes. This is consistent with the results discussed in the previous section, reinforcing the conclusion that the model performs poorly in continental and high-altitude regions. However, one should bear in mind the large domain covered by the model, and thus, it is expected that the RCM would lose its time chronology in a region that lies so far from the model's western boundary.

4. Assessment of RCM performance in determining climate extremes

4.1. Analysis of warm and cold spells

Output from climate models is also used for studying climate extremes. To assess the ability of this RCM to simulate temperature extremes in the study region, cold and warm spells were

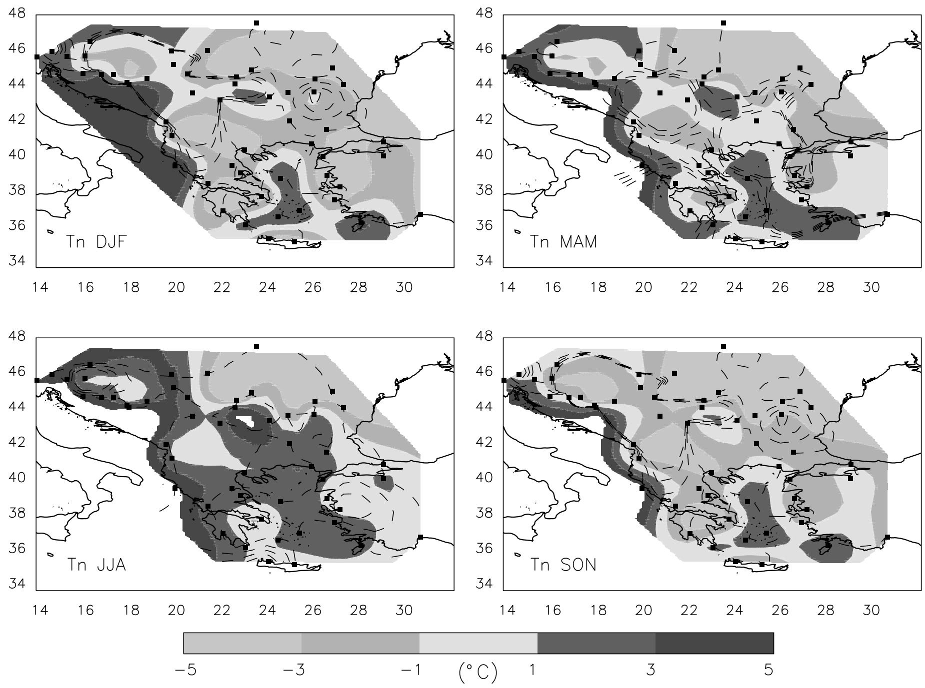

Fig. 7. As in Fig. 6 except for T_N.

detected and compared with those derived from the observational data sets. The spells are defined as sequences of 5, 6, 7, 8, 9 d or longer than 10 d duration with temperatures above/below the 90th/10th percentile (warm/cold spell), based on the temperature distribution of each series. To assess the ability of the model to simulate spells, the total number of warm and cold spells for all seasons, as calculated from model and observed data and examples, is presented in Fig. 8. An additional line on the graphs represents the number of common spells between the two data sets. Common spells are considered to be those with at least 2 d in common between the two data sets. As far as the warm spells are concerned, the model (dark-grey bars) tends to overestimate the number of events compared with those obtained by the station data (light-grey bars). Good representation can be seen for the spring warm spells at most sites and at a large number of stations for warm spells in autumn. Substantial biases are observed in summer warm spells, as the model generates much more warm spells, compared with those recorded at stations. Generally, the model better captures cold spells, although there is an overestimation in some cases. In particular, poor results were obtained for stations located in the northwestern part of the study region, which is consistent with the poor correlations illustrated in Fig. 5. Summarizing the overall findings of the spell analysis, it is found that slightly more cold than warm spells have occurred over the area during the study period. Regarding the frequency distribution of warm and cold spells for winter, the model performs better in detecting cold spells, particularly those lasting 6, 7 or 8 d. In contrast, the frequency distribution of spells for summer showed that the model overestimates the number of warm spells supporting the results discussed in the previous section. However, the model simulates warm spells of 7 and 8 d duration reasonably well. Focusing on the locations used, it was found that warm spells were better reproduced in stations located close to sea level, such as Alexandroupoli, Istanbul and Rhodes, and cold spells in higher elevated sites, such as Bursa, Calarasi and Novi Sad. The coincidence rate of spells for each location was then calculated, based on the number of common spells between the two data sets. The common spell lines (Fig. 8) reveal the modest ability of the model to capture the timing and occurrence of spells. In general, the model better simulates the occurrence of spells in the low-altitude sites of coastal Greece and Turkey.

4.2. Analysis of percentile-based indices (T_X90, T_N10)

Percentile-based climate indices are useful for the analysis of extreme temperatures, as they constitute conventional indicators of the mechanisms or the consequences of the temperature trends (Jones et al., 1999). Therefore, it is important for the models to be capable of reproducing temperature indices trends. A temperature trend analysis was thus applied, based on the trend coefficients of T_X90 and T_N10, separately for the station and model data, and their statistical significance was evaluated using the Kendall's τ-test (Kendall and Gibbons, 1990). A trend was considered to be significant if the probability value (p) was less than or equal to 0.05. All seasonal trends for model and station data were calculated for T_X90 and T_N10, and representative examples are shown in Fig. 9. Coincident positive

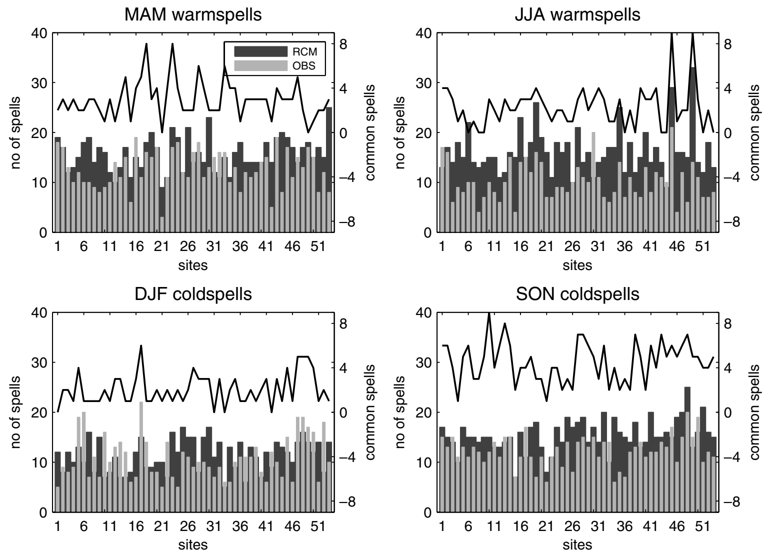

Fig. 8. Total number of warm and cold spells (events of more than five consecutive days duration) as determined by RCM (dark grey bars) and observational data (light grey bars). The additional line represents the number of common spells (with at least 2d in common) between the two data sets.

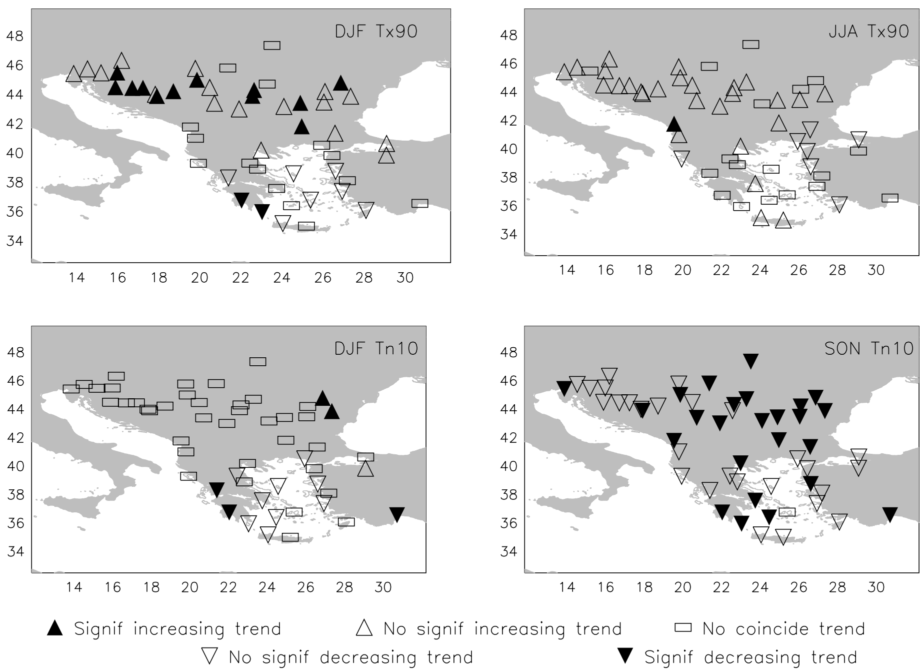

Fig. 9. Examples of seasonal trends of T_X90 and T_N10. Upward (downward) triangles denote positive (negative) trends, whereas solid black triangles indicate statistically significant trends. Empty rectangles define cases where the two trends do not coincide.

(negative) trends are marked by upward (downward) triangles, whereas solid black triangles indicate statistically significant trends. Empty rectangles define cases where the two trends do not coincide.

For the T_X90 index, both data sets present mainly positive trends in all seasons, over the northern continental part of the Balkan Peninsula. In particular, the winter pattern (Fig. 9) presents positive trends, which were found to be statistically significant, especially, in sites that lie above 41°N. These northern sites reveal positive trends also in the other three seasons, although they do not appear to be statistically significant. Some negative trends for the T_X90 index are found at a few stations in southern Greece and the eastern Aegean border. As far as the T_N10 index is concerned, few coincident trends are found for winter and spring. Nevertheless, negative trends are notable for stations in the southern part of the study region. In summer, both model and station data reveal positive trends in the northwestern part of the study domain and a few negative trends in southern Greece. The most striking results were found for the autumn T_N10 trends (Fig. 9). In this case, the vast majority of

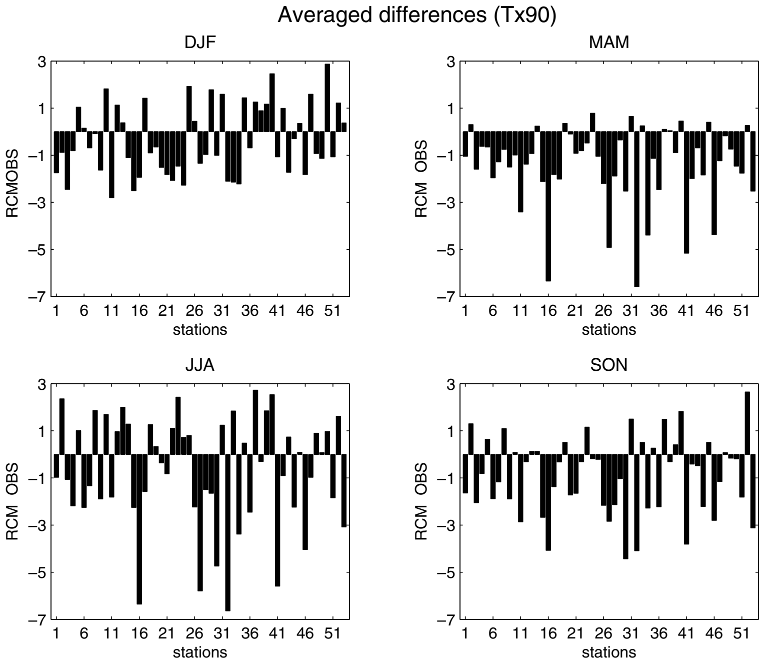

Fig. 10. Seasonal differences (RCM–OBS) of T_X90 averaged over the period 1961–1990.

trends are negative, revealing statistically significant coincident negative trends for at least half of the stations analysed.

Finally, differences between the 30-year averages of the model and station T_X90 and T_N10 indices were calculated. Regarding T_X90 (Fig. 10), negative differences were mainly derived, reaching higher values in spring and summer. This indicates the model's tendency to underestimate the T_X90 index in all seasons. In winter and summer, positive differences are also found, yet these do not exceed 3°C and occur mainly at continental stations situated in Romania, Serbia and Bulgaria. Regarding T_N10 (not shown), both positive and negative differences are found. Specifically in winter and summer, the model tends to overestimate the index (positive differences) at stations situated in the northwestern portion of the study region, namely in the vicinity of Serbia, Slovenia, Bulgaria and Albania. In contrast, negative differences are found in stations in the southern part of the area. The graphs of the transition seasons (MAM, SON) are remarkably similar with the index showing negative differences for most stations, implying that the simulated T_N suggests cooler conditions. At this point, it is worth noting that probability density function (PDF) correction methods exist to better compare model and observation extremes. It is recommended that such correction methods are applied to data prior to conducting climate change scenario studies (Déqué, 2007).

5. Discussion

Projecting climate change at regional and local scales, based on the use of a RCM is a topic of great interest in current climatological studies, especially as far as extreme climate events and their temporal variability are concerned. However, when using models, uncertainty is introduced related to a number of significant temporal and spatial scaling issues of the input data. This study examined the accuracy of simulated maximum and minimum temperature from the ALADIN-Climate RCM, over the Balkan Peninsula, which is a region of variant climate features and contrasting topography. The computational domain of this RCM covers the European continent, from Iceland to the southeastern corner of the Mediterranean Sea. This is a large domain for an RCM, and it is known that RCMs lose their time chronology (they are able to diverge slightly from their lateral boundary conditions) across regions far from their western boundary (as the atmosphere flux derives from the west in the Northern Hemisphere). With this setup, the Balkan and the Greek areas are understandably hard to reproduce. Moreover, this is a very difficult area for the models due to complex topography, complex coastline, air–sea interaction and local convective situations in summer. In such areas, it is common to find discrepancies between the compared model and observational data sets, which are related to elevation (Jones and Lister, 2007) or local inaccurate behaviours of the physical parametrizations in the model. The problem of simulating the southeastern part of Europe, with the state of the art of the European RCM, has also been emphasized in the PRUDENCE project (Jacob et al., 2007).

Principal component analysis was employed to establish a regionalization of the temperature regimes, and this exercise revealed good agreement between the two data sets. Minor discrepancies were found for a few stations in north and central Greece, indicating that the station grouping based on the model data is dependent more on topography than on latitudinal location, and that the model data are associated with enhanced topographical forcing in the RCMs. The influence of topography on the model simulations was also found in the temporal evaluation of the two data sets. The correspondence between the two compared series was generally better for the southern region (subregion 2), which refers mainly to coastal and low-altitude stations. It seems that the lower boundary conditions from the eastern Mediterranean Sea, which are taken from observations, play a more important role for subregion 2 than for the continental subregion 1, where unpredictable atmospheric processes govern maximum and minimum temperatures. Moreover, the reproduction of annual cycles at a subset of stations showed that the model simulates the observed T_X and T_N much more realistically during the warm season. Noticeable biases were revealed in winter T_N, which are associated with warm (cold) biases at stations located at high (low) altitudes.

Seasonal spatial variations of T_X and T_N were analysed, and highest agreement (i.e. low differences and standard deviations and high correlation coefficients) was noted at stations situated close to sea level. Correlation coefficients decrease and differences increase significantly with elevation, indicating poor agreement between model and observations in northern and continental parts of the region due to topography, which is a significant factor affecting temperature variability. To assess the accuracy of the model in simulating extremes, the 90th upper and 10th lower percentiles were calculated for T_X and T_N, respectively, and the corresponding warm and cold spells were defined. The model did not appear very sensitive in detecting particular events of warm/cold spells. This divergence is explained by the fact that a large domain size in an RCM does not follow perfectly the chronology of the driver (ERA-40). Instead, it is designed to reproduce average climate variables and thus does not seem to have a good chronology for extreme events. Nevertheless, the model exhibited a remarkable ability in reproducing the seasonal trends of T_X90 and T_N10.

As a general summary assessment, the comparison of modelled and observed temperatures showed good agreement, although the model data exhibited some biases particularly relative to extreme maximum and minimum observed temperatures. It is important to emphasise that a 25×25 model cell was directly compared with single point station data situated over a complex terrain. The different characteristics of the two comparison data sets could be responsible for some of the RCM failures identified in this study. This is probably the reason for the large biases in winter minimum temperatures (RCM resolution cannot detect the existence of local thermal inversions) or large biases in continental areas (rugged terrain) compared with coastal areas. In addition, it should be stressed that the RCM was driven by the ERA-40 data set, which also inherits some bias compared with station data. The key findings of the study are summarized in the following section.

As an extension to this study, we plan to compare the various ERA40-driven simulations of the ENSEMBLES project for the same area (more RCM simulations will be available soon). The goal would be to find which part of the described errors is specifically due to model formulation and which is due to the ENSEMBLES RCM set-up. The climate of the area studied is also known to be strongly affected by the Black Sea and Mediterranean Sea air masses. Consequently, it would be interesting to compare other RCMs, as used in the ENSEMBLES project, with air–sea coupled RCMs (also called AORCM) as planned in the European CIRCE project. For the European climate, such models already exist with a local coupling for the Baltic Sea (Kjellström et al., 2005) and for the Mediterranean Sea (Somot et al., 2008).

6. Conclusions

Despite overall biases, the model was found to be quite accurate in simulating intraannual variability and the seasonal cycle, as well as simulating the spatial distribution of T_X and T_N. Not surprisingly, at the daily timescale extreme events such as warm and cold spells are out of phase between the model and observations. It is noteworthy that simulations for coastal stations were generally good, highlighting the constraints of the topographic forcing on the simulations. Climate models are particularly strong in simulating temperature patterns associated with specific circulation modes. However, when comparing modelled to observed data in continental or mountainous areas, station data may be affected by local feedbacks with land–surface conditions (Seneviratne et al., 2006; López-Moreno et al., 2008), which are not captured by the model. This is possibly caused by errors in the model's physical parametrizations for a certain region, and thus, the estimated temperatures are biased. Altogether, the model performs better for T_X than T_N and better for the transition seasons. In the absence of available station data in this region, it is essential for studies on climate extremes to use RCM data. However, it is recommended that results must be interpreted with caution. Correction for the daily distribution of the surface parameters can also be applied before using the RCM data (Déqué, 2007).

The main conclusions of the seasonal comparisons of simulated and observed temperatures can be summarized as follows:

(1) The comparison between simulated and observed maximum temperatures showed positive differences (overestimation) in the northern (continental) and negative differences (underestimation) in the southern parts of the domain. Specifically, the model underestimated mean and extreme T_X in western and southern stations except for coastal and island stations in the eastern Aegean Sea.

(2) The simulation of minimum temperatures yielded more complicated results. The model generally underestimated them in winter and autumn, describing colder conditions than those actually observed. In spring, T_N is overestimated (underestimated) in the western (eastern) sector, whereas in summer, the model generally overestimated the minimum temperatures except for the Romanian territory.

(3) The model generally reproduced, to a lesser degree, the extreme values of maximum temperature; however, the frequency of warm spells has been overestimated over the entire region in all seasons. This might be attributed to model's persistence to simulate very warm days in sequence.

(4) The frequency of winter cold spells is underestimated in the northeastern part of the domain. Specifically, fewer spells are found at Romanian stations, although the model's minimum temperature is mostly lower than observed. This implies that the model's extreme low minimum temperatures occur over individual days, which do not constitute spells.

(5) Across the northeastern region (Romania), the model overestimates the maximum and underestimates the minimum winter temperature, resulting in an expansion of the diurnal temperature range. Exceptionally low temperatures and an overestimation of the number of cold spells is found at a few stations across the southern Balkan Peninsula and might be associated with an uncommon prevalence of anticyclonic circulation at the surface, which favours decreases in minimum temperatures (Good et al., 2008).

7. Acknowledgements

We would like to express our gratitude to Meteo-France/CNRM who provided us with the RCM data. The data are maintained by Ole Bøssing Christensen at the DMI data server. We would also like to thank all the Meteorological Institutions that provided the station data. This study has been supported by the European Commission ENSEMBLES project (Contract number GOCE-CT-2003-505539).

8. Appendix A: Station information

No.	Country	Station	Lon.	Lat.	Alt.
1	Greece	Agrinio	21°23′	38°37′	47
2	Greece	Alexandroupoli	25°55′	40°51′	3
3	Greece	Anchialos	22°48′	39°13′	15
4	Turkey	Antalya	30°42′	36°53′	51
5	Romania	Arad	21°21′	46°08′	116
6	Romania	Baia Mare	23°30′	47°40′	216
7	Bosnia and Herzegovina	Banja Luka	17°13′	44°47′	153
8	Serbia	Beograd	20°28′	44°48′	132
9	Bosnia and Herzegovina	Bihac	15°53′	44°49′	246
10	Romania	Bucuresti	26°05′	44°31′	90
11	Turkey	Bursa	29°04′	40°11′	100
12	Romania	Buzau	26°51′	45°08′	97
13	Romania	Calarasi	27°20′	44°12′	18
14	Turkey	Canakkale	26°25′	40°09′	6
15	Greece	Chania	24°02′	35°30′	63
16	Greece	Corfu	19°55′	39°37′	4
17	Romania	Drobeta	22°38′	44°38′	77
18	Turkey	Edirne	26°34′	41°40′	51
19	Slovenia	Godnje	13°51′	45°45′	320
20	Greece	Helliniko	23°44′	37°54′	15
21	Greece	Heraklio	25°11′	35°20′	39
22	Turkey	Istanbul	29°05′	40°58′	33
23	Turkey	Izmir	27°10′	38°26′	25
24	Greece	Kalamata	22°01′	37°04′	8
25	Bulgaria	Kneja	24°05′	43°30′	117
26	Serbia	Kraljevo	20°42′	43°43′	215
27	Greece	Kythira	23°01′	36°17′	167
28	Greece	Larissa	22°25′	39°38′	74
29	Slovenia	Ljubljana	14°31′	46°04′	299
30	Bosnia and Herzegovina	Mostar	17°49′	44°20′	99
31	Slovenia	Murska Sobota	16°11′	46°39′	188
32	Greece	Milos	24°27′	36°43′	183
33	Greece	Mytilene	26°36′	39°04′	5
34	Greece	Naxos	25°23′	37°06′	9

No.	Country	Station	Lon.	Lat.	Alt.
35	Serbia	Negotin	22°33′	44°14′	42
36	Serbia	Nis	21°54′	43°20′	201
37	Serbia	Novi Sad	19°51′	45°20′	86
38	Slovenia	Novo Mesto	15°11′	45°48′	220
39	Bulgaria	Obraszov-Chiflic	26°02′	43°48′	157
40	Serbia	Palic	19°46′	46°06′	102
41	Greece	Rhodes	28°05′	36°24′	11
42	Bulgaria	Sadovo	24°57′	42°09′	155
43	Greece	Samos	26°55′	37°42′	7
44	Bosnia and Herzegovina	Sanski Most	16°42′	44°46′	158
45	Albania	Shkodra	19°32′	42°06′	43
46	Greece	Skyros	24°33′	38°54′	18
47	Romania	Tg Jiu	23°16′	45°02′	203
48	Greece	Thessaloniki	22°58′	40°31′	4
49	Albania	Tirana	19°47′	41°20′	89
50	Romania	Turnu Magurele	24°53′	43°45′	31
51	Bosnia and Herzegovina	Tuzla	18°42′	44°33′	305
52	Croatia	Zagreb	15°58′	45°49′	157
53	Bosnia and Herzegovina	Zenica	17°54′	44°13′	344

References

Alexandersson, H. 1986. A homogeneity test applied to precipitation data. *J. Clim.,* **6**, 661–675.

Bengtsson, L., Hagemann, S. and Hodges, K. I. 2004a. Can climate trends be computed from reanalysis data? *J. Geophys. Res.,* **109**, D11111, doi:10.1029/2004JD004536.

Bengtsson, L., Hodges, K. I. and Hagemann, S. 2004b. Sensitivity of the ERA40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations. *Tellus* **56A**, 456–471.

Bromwich, D. H. and Fogt, R. L. 2004. Strong trends in the skill of the ERA-40 and NCEP–NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001. *J. Clim.*, **17**, 4603–4619.

Christensen, J. H., Machenhauer, B., Jones, R. G., Schär, C., Ruti, P. and co-authors. 1997. Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. *Clim. Dyn.* **13**, 489–506.

Decharme, B. and Douville, H., 2007. Global validation of the ISBA sub-grid hydrology. *Clim. Dyn.*, **29**, 21–37, doi:10.1007/s00382-006-0216-7.

Déqué, M. 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Glob. Planet. Change*, **57**, 16–26.

Déqué, M. and Somot, S. 2007. Variable resolution versus limited area modelling: perfect model approach. In: *Research Activities in Atmospheric and Oceanic Modelling* Report No.37. April 2007, 3.03–3.04, CAS/JSC Working Group on Numerical Experimentation. Available online at: http://collaboration.cmc.ec.gc.ca/science/wgne/index.html.

Déqué, M. Jones, R. G., Wild, M., Giorgi, F., Christensen, J. H. and co-authors. 2005. Global high resolution versus Limited Area Model climate change projections over Europe: quantifying confidence level from PRUDENCE results. *Clim. Dyn.* **25**, 653–670.

Douville, H., Planton, S., Royer, J. F., Stephenson, D. B., Tyteca, S. and co-authors. 2000. The importance of vegetation feedbacks in doubled-CO2 time-slice experiments. *J. Geophys. Res.*, **105**, 14 841–14 861.

Giorgi, F., Bi, X. and Pal, J. S. 2004. Mean interannual variability and trends in a regional climate change experiment over Europe, I: present-day climate (1961–1990). *Clim. Dyn.* **22**, 733–756.

Good, P., Giannakopoulos, C., Flocas, H., Tolika, K., Anagnostopoulou, C. and co-authors. 2008. Significant changes in the regional climate of the Aegean during 1961–2002. *Int. J. Climatol.,* **28**, 1735–1749.

Hertzog, A., Basdevant, C. and Vial, F. 2006. An assessment of ECMWF and NCEP-NCAR Reanalyses in the southern hemisphere at the end of the presatellite era: results from the EOLE experiment (1971–72). *Mon. Wea. Rev.*, **134**, 3367–3383.

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H. and co-authors. 2007. An inter-comparison of regional climate models for Europe: model performance in Present-Day Climate. *Clim. Change*, **81**, 31–52, doi:10.1007/s10584-006-9213-4.

Jones, P. D. 1999. The instrumental data record: its accuracy and use in attempts to identify the ‘CO2 signal’. In: *Analysis of Climate Variability* (eds. H. von Storch and A. Navarra). Springer-Verlag, Berlin, 53–76.

Jones, P. D. and Lister, D. H. 2007. Intercomparison of four different Southern Hemisphere sea level pressure datasets. *Geophys. Res. Lett.* **34**, L10704, doi:10.1029/2007GL029251.

Jones, P. D., Horton, E. B., Folland, C. K., Hulme, M., Parker, D. E., and co-authors. 1999. The use of indices to identify changes in climatic extremes. *Clim. Change*, **42**, 131–149.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. and co-authors., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471.

Kendall, M. and Gibbons, J. D. 1990. *Rank Correlation Methods*. 5th Edition. Oxford University Press, New York, 260 pp.

Kharin, V. V. and Zwiers, F. W., 2000. Changes in the extremes in an ensemble of transient climate simulation with a coupled atmosphere-ocean GCM. *J. Clim.*, **13**, 3760–3788.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G. and co-authors. 2001. The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**, 247–267.

Kjellström, E., Döscher, R. and Meier, H. E. M. 2005. Atmospheric response to different sea surface temperatures in the Baltic Sea: coupled versus uncoupled regional climate model experiments. *Nord. Hydrol*. **36**, 397–409.

Kostopoulou, E. and Jones, P. D. 2007a. Comprehensive analysis of the climate variability in the eastern Mediterranean, part I: map pattern classification. *Int. J. Climatol.* **27**, 1189–1214.

Kostopoulou, E. and Jones, P. D. 2007b. Comprehensive analysis of the climate variability in the eastern Mediterranean, part II: relationships between atmospheric circulation patterns and surface climatic elements. *Int. J. Climatol.* **27**, 1351–1371.

López-Moreno, J. I., Goyette, S. and Beniston, M. 2008. Climate change prediction over complex areas: spatial variability of uncertainties and expected changes over the Pyrenees from a set of regional climate models. *Int. J. Climatol.* **28**, 1535–1550.

Lucas-Picher, P., Caya, D., de Elia, R. and Laprise, R. 2008. Investigation of regional climate models internal variability with a ten-member ensemble of 10-year simulations over a large domain. *Clim. Dyn*. **31**, 927–940.

Moberg, A. and Jones, P. D. 2004. Regional Climate Models simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data for 1961–1990. *Clim. Dyn.* **23**, 695–715.

Radu, R., Déqué, M. and Somot, S. 2008. Spectral nudging in a spectral regional climate model. *Tellus* **60A**, 885–897. doi:10.1111/j.1600-0870.2008.00343.x.

Reid, P. A., Jones, P. D., Brown, O., Goodess, C. M. and Davies, T. D. 2001. Assessments of the reliability of NCEP circulation and relationships with surface climate by direct comparisons with station based data. *Clim. Res.* **17**, 247–261.

Rusticucci, M. M. and Kousky, E. 2002. A comparative study of maximum and minimum temperatures over Argentina: NCEP-NCAR reanalysis versus station data. *J. Clim.* **15**, 2089–2101.

Seneviratne, S. I., Lüthi, D., Litschi, M. and Christoph Schär, C., 2006. Land–atmosphere coupling and climate change in Europe. *Nature* **443**, 205–209.

Simmons, A. J. and Gibson, J. K. 2000. The ERA-40 Project Plan. In: *ERA-40 Project Rep* Series 1. ECMWF, Shinfield Park, Reading, United Kingdom, 63 pp.

Simmons, A. J., Jones, P. D., da Costa Bechtold, V., Beljaars, A. C. M., Kållberg, P. W., and co-authors. 2004. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP–NCAR analyses of surface air temperature. *J. Geophys. Res*. **109**, D24115, doi:10.1029/2004JD005306.

Somot, S., Sevault, F., Déqué, M. and Crépon, M. 2008. 21st century climate change scenario for the Mediterranean using a coupled atmosphere-ocean regional climate model. *Global Planet. Change* **63**, 112–126, doi:10.1016/j.gloplacha.2007.10.003

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., and co-authors. 2005. The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

Zhao, T. and FU, C. 2006. Comparison of products from ERA-40, NCEP-2, and CRU with station data for summer precipitation over China. *Adv. Atmos. Sci*. **23**, 593–604.