
Copyright: The definitive version of this article, published by Wiley, 2009 is available at:

http://dx.doi.org/10.1002/joc.1807

Always use the definitive version when citing.

Further information on publisher website: http://onlinelibrary.wiley.com/

Date deposited: 29th October 2013

Version of file: Published

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License
Observed and modelled influence of atmospheric circulation on central England temperature extremes

S. Blenkinsop, P. D. Jones, S. R. Dorling and T. J. Osborn

ABSTRACT: The regional atmospheric circulation is a major driver of climate for western Europe and so the ability of climate models to accurately simulate its characteristics forms an important part of the rigorous testing of their performance. In this paper we examine the skills of seven regional climate models (RCMs) to reproduce mean daily temperatures for the central England region. Their ability to reproduce observed characteristics of the circulation is then tested using three airflow indices. This is achieved by comparing: (1) the frequency distribution of each index, (2) the relationships between the daily airflow indices and temperature, including daily extremes and (3) the ability of models to reproduce the observed persistence of specific flow regimes and the temperature response to this persistence.

It is demonstrated that RCM selection introduces uncertainty into temperature simulations and that there is no single model which outperforms the others. Although the models qualitatively reproduce the observed distributions of the airflow indices reasonably well, most models produce distributions that are statistically significantly different from the observations. In particular, all models overestimate the frequency of westerly flow, and biases in the relationships between circulation and temperature are also noted for extreme values of the airflow indices. The persistence of flow regimes is shown to have a significant effect on the observed temperature, though models have difficulty in reproducing the magnitude of the response to this persistence for some flow types.

The results presented here not only form a useful model validation exercise but also further highlight the need for the use of multi-model ensembles in the generation of future climate scenarios. Furthermore, they suggest that the use of atmospheric circulation in statistical downscaling methods might be enhanced by the inclusion of persistence, particularly when producing scenarios of temperature extremes. Copyright © 2008 Royal Meteorological Society

KEY WORDS: temperature; RCMs; atmospheric circulation; CET; validation; persistence

Received 7 November 2007; Revised 8 October 2008; Accepted 11 October 2008

1. Introduction

Climate models are important tools in our ability to assess the likely trajectory of future climate change and to form the basis of impact assessments which determine adaptation and mitigation strategies. However, limitations in our knowledge of atmospheric processes and the parameterization of small-scale physics mean that uncertainties are inherent in simulations produced by climate models. Therefore, rigorous validation of all aspects of climate model output forms an important feature of work in the climate community.

One of the key determinants of climate for the western Europe region is the large-scale atmospheric circulation. Its observed variability has been extensively studied (e.g. Jones et al., 1993; Esteban et al., 2006; James, 2007) as has its influence on temperature (e.g. Osborn et al., 1999; Brabson and Palutikof, 2002) and precipitation (e.g. Goodess and Jones, 2002; McGregor and Phillips, 2004). The atmospheric circulation has also been identified as a key determinant of extreme events such as heat waves (Della-Marta et al., 2007), droughts (Fowler and Kilsby, 2002) and floods (Bárdossy and Filiz, 2005). The role of circulation in influencing temperature trends during the 20th century has been examined (Osborn and Jones, 2000; van Oldenborgh and van Ulden, 2003) and may be useful for the attribution of observed climate change. The ability of individual models to reproduce observed characteristics of regional circulation and its relationship with surface climate has also been examined (e.g. Osborn et al., 1999; James, 2006).

However, more recently, increased emphasis has been placed on appreciating uncertainties in climate model simulations, and examining atmospheric circulation is an important part of understanding the ability of models to reproduce physical atmospheric processes. Jacob et al. (2007) have investigated biases of variables including mean sea level pressure (MSLP) fields in an ensemble of regional climate model (RCM) simulations from the PRUDENCE project (Christensen et al., 2007). Biases in
RCM simulations of wintertime MSLP were shown to be largely dependent on the driving general circulation model (GCM) boundary conditions, but in summertime smaller scale processes make differences between GCM and RCM biases more important. van Ulden and van Oldenborgh (2006) examined simulations of two geostrophic flow indices which correspond to westerly and southerly flow components, and a third representing geostrophic vorticity for 23 global coupled climate models. Many showed serious biases at mid-latitudes, particularly in the simulation of spatial variance which produced a poor simulation of circulation indices over central Europe. van Ulden et al. (2007) also compared the simulation of the same indices over central Europe for three GCMs and nine PRUDENCE RCMs. Furthermore, relationships between monthly temperature and rainfall statistics and westerly flow were compared for two of the regional models noting HadRM3H to be poor in reproducing the observed relationship between summer temperature and westerly flow.

Here, we use daily MSLP data to derive daily circulation indices for central England, which are used to examine the relationship between the atmospheric circulation and daily temperature means and extremes. This approach has been used to assess the output from an individual GCM (Osborn et al., 1999) and RCM (Turnpenny et al., 2002). However, both studies as well as only considering single model simulations only examined the relationships with mean daily temperature. We compare the ability of seven RCMs in simulating observed daily mean temperature and six of those in simulating daily extreme temperatures for central England as well as comparing their simulations of daily airflow indices. These models are described briefly in Section 2. We then outline the scheme used to represent the regional atmospheric circulation over the UK in Section 3. In Section 4, the ability of models to reproduce observed mean temperatures for central England is assessed. Their skill in reproducing observed frequencies of circulation regimes is examined along with the relationships between those regimes and near-surface temperature. The ability of models to reproduce observed characteristics of the persistence of different circulation regimes is also tested and will enable a more rigorous evaluation of RCM performance, offering a greater insight as to the nature of model biases in temperature simulations while the use of multiple climate models provides an indication of the uncertainty range in RCM simulations.

2. Data and models

2.1. Data

Two ‘observed’ temperature data sets are used here. Firstly, the Climatic Research Unit (CRU) TS 2.0 data set (Mitchell et al., 2004) is used to enable a direct comparison between models and observations using identical spatial domains. This is a gridded global series of monthly climate means for the land surface for the period 1901–2000 and was constructed by the interpolation of station data onto a 0.5° grid and is an updated version of earlier datasets (New et al., 1999a,b). This series will be referred to hereafter as CRU. Secondly, the Central England Temperature (CET) series is used. This was first published as a series of monthly mean temperatures from land surface for the period 1901–2000 and was constructed by the interpolation of station data onto a 0.5° grid and is an updated version of earlier datasets (New et al., 1999a,b). This series will be referred to hereafter as CRU. Secondly, the Central England Temperature (CET) series is used. This was first published as a series of monthly mean temperatures from 1698 (Manley, 1953) and later extended back to 1659 (Manley, 1974), and subsequently a daily temperature series has been constructed back to 1772 (Parker et al., 1992) from various combinations of stations throughout the period of record due to station closures and missing data. These stations represent Manley’s concept of central England and try to combine stations with long, unbroken records and those which avoid urban influences. This daily series is used to derive observed relationships between circulation and temperature and the RCM grid cells used to represent the CET region are shown in Figure 1.

Figure 1. Grid points used to construct the airflow indices for the British Isles. The solid circles represent the grid points used in this study; the hollow circles represent those traditionally used when all grid points are available. The shaded boxes are the grid cells used to define the central England region on the CRU and RCM grids.
Finally, to calculate the airflow indices, a gridded 2.5° × 2.5° daily MSLP re-analysis series is used from the National Centers for Environmental Prediction (NCEP) re-analysis project described by Kalnay et al. (1996). An observed daily series of MSLP from the UK Meteorological Office (UKMO) data set (Jones, 1987) was considered; however, a coarser resolution of 5° latitude × 10° longitude resulted in one of the gridpoints used to calculate the airflow indices outside the domain of the RCMs. Potential inhomogeneities in the NCEP series have been identified over Greenland and the high altitude regions of southern Europe. Also, prior to the mid-1960s, NCEP annual MSLP is lower than the UKMO data at many grid points due to a data input error. The former problem is outside the area of interest of this study, while the latter should not affect the calculation of airflow indices as they are not dependent on absolute values, but the relative difference between selected grid points. Reid et al. (2001) indicate an excellent agreement between monthly mean MSLP obtained from NCEP and UKMO over northwest Europe while Blenkinsop (2005) indicated a good agreement in the derived airflow indices from the two datasets for the UK region.

Daily temperature and pressure data for the period 1961–1990 were extracted from each dataset to correspond with the control period used in the RCM simulations which are described in the following section.

2.2. Regional climate models

RCM output from the European Union Fifth Framework Programme (FP5) PRUDENCE project (Christensen et al., 2007) provides a series of high-resolution simulations of European climate for the control period 1961–1990 for a large range of climatic variables, using RCMs driven by boundary conditions from different GCMs. These ‘time-slice’ simulations are representative of a stationary climate over a 30-year period and may be compared with observations as part of model verification (Räisänen, 2007). Here the model selection was made to examine the uncertainty in RCM output due to the bounding GCM and that due to the choice of RCM. The contribution of these two sources to RCM uncertainty is tested by using a selection of models which investigate the role of:

- same bounding GCM in combination with different RCMs (e.g. HIRHAM H v RCAO H; Table I);
- same RCM in combination with different bounding GCMs (e.g. HIRHAM H v HIRHAM E; Table I)

A list of models and their acronyms used in this study is provided in Table I. The HIRHAM model is a version of HIRHAM4 (Christensen et al., 1996, 1998) updated to incorporate high-resolution physiographical data of surface topography and land use classification (Christensen et al., 2001; Hagemann et al., 2001). The RCAO model is composed of an atmospheric part RCA2 (Jones et al., 2004a) and an ocean model RCO (Meier et al., 2003) while HadRM3P (Jones et al., 2004b) is an updated version of HadRM3H (Hudson and Jones, 2002). Finally, ARPEGE-IFS is a global operational forecast model which may be run at a variable horizontal resolution and is nested directly within HadCM3. This is an updated version of that described by Déqué et al. (1998) to reflect changes in the radiation and cloud-precipitation-turbulence schemes. Model simulations of daily minimum, mean and maximum temperature and daily MSLP are available for control integrations for the period 1961–1990, though daily minima and maxima are not available for HIRHAM E. Therefore, an additional simulation of HIRHAM using boundary conditions derived from the ECHAM5 GCM is also examined. The main changes in ECHAM5 include an updated longwave radiation scheme, to that used in ECHAM4, new cloud microphysics and changes in the representation of land surface processes (Roeckner et al., 2003). Temperature and MSLP data for each of these simulations were re-gridded onto the CRU 0.5° grid to allow direct comparison with the CET and NCEP series respectively.

### Table I. Selection of PRUDENCE RCMs used for this study. Different acronyms are adopted here to provide an easier understanding of the format of each model run. The first part of each acronym refers to the RCM and the second to the GCM data used to provide the boundary conditions. Models are run for a 30-year control period (1961–90) except for HadRM3P which is run for a total of 31 years (1960–90).

<table>
<thead>
<tr>
<th>RCM</th>
<th>Driving GCM</th>
<th>PRUDENCE acronym</th>
<th>Modified acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish Meteorological Institute (DMI)</td>
<td>HIRHAM</td>
<td>HadAM3H</td>
<td>HC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECHAM4/OPYC</td>
<td>Eccitr</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECHAM5</td>
<td>ECC</td>
</tr>
<tr>
<td>Swedish Meteorological and Hydrological Institute (SMHI)</td>
<td>RCAO</td>
<td>HadAM3H</td>
<td>HCCTL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECHAM4/OPYC</td>
<td>MPICTL</td>
</tr>
<tr>
<td>Hadley Centre–UK Met Office</td>
<td>HadRM3P</td>
<td>HadAM3P</td>
<td>Adeha</td>
</tr>
<tr>
<td>Météo-France, France</td>
<td>ARPEGE-IFS</td>
<td>HadCM3</td>
<td>DA9</td>
</tr>
</tbody>
</table>

Copyright © 2008 Royal Meteorological Society

DOI: 10.1002/joc
a useful means of objectively classifying daily circulation types affecting the UK (Jones et al., 1993), producing strong correlations with the subjective Lamb (1972) scheme in terms of circulation type frequencies and their relationship with seasonal temperature and precipitation. Three indices; flow direction (DIR), flow strength (STR) and vorticity (VORT) may be derived from the grid-ded MSLP at the points shown in Figure 1 using the equations shown in the appendix. For STR, each unit is equivalent to approximately 0.6 ms\(^{-1}\), while for VORT, which reflects the rotation of an air mass, 100 units are equivalent to 0.46 times the Coriolis parameter at 55°N. Negative (positive) VORT values are indicative of anti-cyclonic (cyclonic) rotation of an air mass. The DIR index is simply expressed in degrees from north (0–360°).

This automated typing scheme has also been applied to studies in Scandinavia (Chen, 2000; Linderson, 2001) and the Iberian Peninsula (Goodess and Jones, 2002). As in Osborn et al. (1999) each index is divided into 20-equal sized bins with each day subsequently classified according to the bin into which the index value for that day falls. The resultant distribution is used to describe the frequency of occurrence of different flow regimes and the climate characteristics on a given day. These indices have been used to assess the ability of the single GCM (Osborn et al., 1999) and RCM (Turnpenny et al., 2002) simulations to reproduce the observed relationships between daily circulation and surface climate. Here, we extend the analysis to examine multiple RCM simulations in order to obtain an estimate of the uncertainties in reproducing regional atmospheric circulation. Using the same indices to compare RCM simulations with those derived from the NCEP re-analysis data provides a direct comparison of model performance.

4. Results

4.1. RCM simulations of central England temperature

The average monthly CRU mean (\(T_g\)), minimum (\(T_n\)) and maximum (\(T_x\)) temperatures for 1961–1990 are compared with the model control simulations in Figure 2. While all the models reproduce the annual cycle of temperature reasonably well, there are some significant errors in the RCM simulations. Mean temperatures are overestimated during winter with the largest errors for most models observed during February [the largest being RCAO_E and RCAO_H (+2.6 °C and +2.3 °C respectively; Figure 2(a)]. However, HIRHAM_E5 produces the largest overestimates during late spring and early summer (up to +1.4 °C) while the largest errors for ARPEGE_H are underestimates during the same period (up to −0.9 °C) and also during October (−1.8 °C). During summer, for most simulations the magnitude of the errors is smaller with four models overestimating temperature (HIRHAM_E5, HIRHAM_H, RCAO_H and HAD_H_3P) and another three underestimating temperature (HIRHAM_E, RCAO_E and ARPEGE_H), indicating a seasonal difference in relative importance of the contribution of RCM and GCM choices to simulation biases.

Considering the daily extremes, the simulations of \(T_n\) (Figure 2(b)) indicate that the models are poor in reproducing the monthly minima throughout the year. The largest errors occur during winter with RCAO_E and RCAO_H overestimating mean February minima by 4.0 °C and 3.7 °C respectively, again demonstrating the choice of RCM as a significant source of error during this time of the year. However, much smaller errors of less than 1 °C are produced by HAD_H_3P and ARPEGE_H during spring and summer. For \(T_x\) (Figure 2(c)), all RCMs overestimate temperatures during January and February with the RCAO simulations again producing...
the largest errors, but in all cases the magnitude of the errors is smaller than for $T_n$. Throughout the rest of the year there is a general underestimation of $T_x$ resulting in a smaller than observed amplitude of the annual cycle. Consequently, most models fail to accurately reflect the daily temperature range (not shown), especially the two RCAO simulations which underestimate the mean temperature range by up to 4.2°C during late summer due to the simulation of minima that are too warm and maxima that are too cool. In contrast, HAD$_H$3P produces the smallest errors in temperature range, less than 1°C for 9 months of the year.

The root mean square error (RMSE) statistic was calculated to quantify these model biases based on the differences between simulated and CRU seasonal mean temperatures for each grid cell over the central England region. Figure 3 indicates that for $T_n$ the largest errors during autumn and winter are produced by the RCAO simulations with the smallest produced by HAD$_H$3P and ARPEGE$_H$ respectively. In contrast, during summer, the RCAO simulations produce the lowest RMSE statistic. For temperature extremes, some differences are observed. For $T_n$, HAD$_H$3P and ARPEGE$_H$ produce the smallest errors throughout the year while the largest RMSE statistics are generally produced by the RCAO simulations, particularly during autumn and winter. For $T_x$, ARPEGE$_H$ is relatively poor in all seasons except winter when it produces relatively small errors, although

![Figure 3. RMSE statistics for all grid cells in the CET region relative to CRU. Results are presented seasonally and separately for $T_n$ (left), $T_g$ (centre) and $T_x$ (right). The first row is winter (DJF) with subsequent rows spring (MAM), summer (JJA) and autumn (SON).](image-url)
there is little difference between the models in this season. The RCAO and ARPEGE_H simulations produce the largest errors during summer when the range of model results is largest; this range suggesting model parameterization has the greatest effect on warm extremes during summer months. Considering the RMSE statistics in conjunction with Figure 2 indicates that in general terms, RCAO has the greatest problems in the simulation of extremes with minima that are too low throughout the year and maxima that are too low during summer. These models would thus provide a poor representation of events such as summer heat waves and cold spells throughout the year. In contrast, HAD_H3P performs reasonably well in minimizing errors in winter $T_n$ and summer $T_x$.

4.2. RCM simulations of regional atmospheric circulation

Figure 4 compares model simulations of the frequencies of the airflow indices with those derived from the NCEP data. Chi-squared goodness-of-fit tests were applied to test each of the modelled seasonal distributions shown against observations (NCEP) and indicated that the seasonal RCM distributions for each airflow index are significantly different from the corresponding NCEP distributions at the 95% level with the exception of the ARPEGE_H autumn STR distribution. For DIR, all models overestimate the frequency of westerly flow during winter at the expense of easterly and southerly regimes, while during spring HAD_H3P, RCAO_H and HIRHAM_H capture the frequency of flow regimes reasonably well whereas HIRHAM_E, RCAO_E and particularly ARPEGE_H again overestimate the westerly flow. This is consistent with results obtained by van Ulden et al. (2007) who noted that ARPEGE produced a strong positive bias in a zonal flow index during winter over central Europe. The overestimation of westerly flows by GCMS and their inability to produce sufficient blocking types in the Northern Hemisphere has been noted previously (D’Andrea et al., 1998; Pelly and Hoskins, 2003; James, 2006; van Ulden and van Oldenborgh, 2006). Nonetheless, the comparison of a common circulation index for a selection of PRUDENCE models is useful as it is indicative of the range of errors in simulations. Qualitatively, observed summer and autumn flow regimes are captured much better, though the two ECHAM-driven simulations both have a significant westerly bias, whereas HAD_H3P has an easterly bias. In all cases, the importance of GCM boundary conditions is noted with the paired simulations driven by ECHAM and HadAM3H both producing similar distributions. This conditioning of RCM projections by the behaviour of the driving GCMS has been demonstrated to be important when model outputs are used to generate scenarios of climate change impacts (e.g. Wilby and Harris, 2006; Fronzek and Carter, 2007).

For STR, most models simulate the observed winter distribution reasonably well, though most underestimate the frequency of low STR days to the preference of moderate strength flow days. However, ARPEGE_H again produces poor results, along with HAD_H3P significantly overestimating the frequency of strong flow days during winter, consistent with their simulation of too frequent westerly flow days. Throughout the rest of the year, most models tend to underestimate the frequency of high STR days, producing more frequent days with moderate STR values.

For VORT the greatest spread of results is again obtained for winter, but throughout the year most simulations underestimate the number of anti-cyclonic days (VORT $<0$) with too many moderately cyclonic days, particularly for the two ECHAM-driven simulations during summer and autumn. During spring and autumn, most models also tend to underestimate the frequencies of both extreme cyclonic and anti-cyclonic days.

4.3. Model simulations of the relationships between circulation and temperature

To examine the relationship between circulation and temperature, the daily temperature values were converted into anomalies. These were obtained by first calculating the annual cycle over the baseline period of 1961–90 by averaging the 30 values for each calendar date. An 11-term binomial filter was next applied to the resulting annual averages to produce a temperature cycle less strongly influenced by random variations while still maintaining genuine features of annual temperature. The annual cycle obtained from the observations and from each model was then subtracted from the corresponding daily temperature series to derive the anomalies.

Interdependencies in the relationship between temperature and airflow indices have previously been identified by considering a bivariate analysis of combinations of indices (Osborn et al., 1999). However, one of the disadvantages associated with the use of these indices is the loss of information when considering marginal distributions of bivariate pairs relative to those of the zonal and meridional flow components. Here, this problem is apparent due to the small sample sizes for some combinations of regimes, e.g. strong easterly flow, for the 30-year time slices and prohibits such an analysis here. Therefore, here relationships between daily temperature anomalies and the three airflow indices are determined and compared independently. Previous analysis of these relationships (Osborn et al., 1999; Turnpenny et al., 2002) indicated that mean daily temperature anomalies are most strongly dependent upon DIR, but also upon STR and vorticity during winter when stronger flow and cyclonic conditions are associated with milder temperatures. Some differences in relationships are observed when considering observed minimum and maximum temperatures. Figure 5 shows that DIR exerts a strong influence on minimum, mean and maximum daily temperature throughout the year but that this influence is greatest on maximum temperature and least on minima. This is most clearly demonstrated in summer when for maximum temperatures the amplitude of the influence of DIR is approximately $5^\circ$C but for minima it is less than $2^\circ$C.
The influence of STR on temperature is strongest during winter for both extremes when strong (weak) flows tend to produce large positive (negative) temperature anomalies. This effect is weaker during autumn and STR is less important still for the remainder of the year. During summer, however, maximum temperatures exhibit a relationship which is the reverse of that occurring during winter. This negative relationship with maximum temperatures may be related to different heating mechanisms at this time of year. During the other seasons, advection of heat from other regions is likely to be important but in summer direct heating from the warmer land surface is of increasing importance and stronger flow tends to be associated with increased advection of relatively cool and moist air.

The VORT index demonstrates the most varied responses for daily extremes with significant differences to the relationships with daily mean temperature observed.
Figure 5. Mean daily temperature anomaly for $T_n$, $T_g$ and $T_x$ on days falling into each index bin over the period 1961–1990. Calculations were made using CET temperature data and airflow indices calculated using the NCEP data. The bin sizes are as defined in Figure 4 and shading as in Figure 2. Means are only calculated for bins with a sample of at least 20 days throughout the period to ensure representativeness.

by Osborn et al. (1999). Whereas they found that vorticity only exerts a strong, positive influence over mean temperatures in winter, minimum temperature anomalies display a consistent, positive relationship with vorticity throughout most of the year, anti-cyclonic conditions producing cool/cold nights while cyclonic conditions result in warmer temperatures, as would be expected. However, the relationship is strongest when VORT <0, whereas for VORT >0 the curves flatten indicating that the influence on temperature anomalies is not as strong under cyclonic conditions. In contrast, for maxima the relationship with VORT reverses in sign during the year. In spring and summer VORT has a negative relationship with maximum temperatures (anti-cyclonic: warm, cyclonic: cool) but this changes to a weak relationship during autumn and a positive one (anti-cyclonic: cool, cyclonic: warm) during winter. These relationships with VORT reflect seasonal surface
radiation budgets, with summer daytime temperatures more strongly influenced by inward short-wave radiation than is the case in winter. These relationships may be related to cloud cover, as generally maximum temperatures decrease when cloud cover is above average and minimum temperatures increase. Relationships between cloudiness and temperature extremes have been identified (Plantico et al., 1990; Karl et al., 1993) and would be expected to affect extremes more than means (Campbell and Vonder Haar, 1997). The effect of cloud cover on maximum temperatures has been observed to be greater than on minima (Campbell and Vonder Haar, 1997; Dai et al., 1999), particularly for low-based cloud (Dai et al., 1999), which may be a mechanism for the increased response of maximum temperatures to certain circulation conditions. However, the role of clouds is not fully understood, with other factors such as particle size and cloud type also being important (Arking, 1991).

The same relationships were examined for the RCMs and are shown in Figures 6–8. The models generally capture the relationships between mean temperature and atmospheric circulation. In particular, they reproduce the seasonal changes in the relationships but also the different relationships between maxima and minima. However, specific circulation regimes are not well represented by either some or all of the models. In particular, all models underestimate the relative warmth of winter westerly flow while HAD.H.3P is relatively cold on northerly and easterly flow days. In contrast, during summer, the two HIRHAM simulations are relatively warm on such days. Relationships with STR are generally reproduced, though in winter most models tend to overestimate the magnitude of negative anomalies of $T_g$ and $T_m$ on low STR days (Figures 6 and 7). Some relationships show a clear influence of the RCM selection, for example between summer $T_x$ and STR (Figure 8). HIRHAM,E5 suggests a stronger negative relationship than observed, while RCAO,E simulates a weaker relationship. On the other hand, others are more strongly related to the driving GCM such as the significantly higher temperature anomalies on low STR days exhibited by HIRHAM,H and RCAO,H in autumn (Figures 6 and 8). During all seasons except summer, the models tend to overestimate the influence of low VORT values producing larger negative temperature anomalies than observed. This is most noticeable during winter with $T_m$ anomalies of over $-4^\circ$C simulated by HIRHAM,H and HAD.H.3P for the most extreme anti-cyclonic bin compared with an observed anomaly of $-2^\circ$C. The lack of skill in the reproduction of such extremes is likely in part due to the poor resolution by the models of strong near-ground gradients which occur in a stable, stratified atmosphere and which produce low temperatures under such conditions. During summer the relationships with temperature, though qualitatively realistic, show a larger range of amplitudes, providing further evidence of less reliable representation of extreme circulation conditions by RCMs.

As noted above, it would be desirable to examine bivariate relationships between pairs of airflow indices and temperature. Osborn et al. (1999) indicated some non-linear temperature responses when compared with those based on individual airflow indices. While this would be interesting here, using time-slice model simulations covering 30 years would not yield a sufficient sample size for a robust statistical analysis of some combinations, for example, strong flow on easterly days, particularly given the underestimation of some directional frequencies by the climate models. However, in order to further explore why DIR has a greater influence on $T_x$, a simple examination of bivariate relationships was undertaken by repeating the analysis of DIR but further dividing days into those where either of the conditions VORT $<0$ or VORT $>0$ was satisfied. This indicated that the positive enhancement of maximum temperature anomalies on warm DIR days (except winter) is produced when VORT is negative, i.e. when flow is anti-cyclonic, with no significant enhancement of maxima for the same DIR bins on cyclonic days. Conversely, the larger negative anomalies for maximum temperature on cold DIR days are generally more pronounced if the flow is cyclonic (except for winter easterlies). Such temperature differences are likely to be related to radiation budgets. Negative vorticity associated with high pressure at the surface and subsidence from above is more likely to produce clear skies. On days with warm advection, daytime temperatures are thus likely to be further increased by incoming solar radiation. Similarly, increased cloud cover on cyclonic days would be a mechanism for increasing (decreasing) the magnitude of the negative anomaly for maximum (minimum) temperatures when there is advection of air from the cold flow direction.

4.4. Persistence of circulation indices

The association between the persistence of particular flow regimes and surface temperature has been observed in central Europe with implications for more pronounced extremes under more persistent circulation regimes (Kyselý, 2007). Persistence in circulation regimes has been cited as responsible for recent extreme conditions over Europe (e.g. Black et al., 2004; van Oldenborgh, 2007), and the increased persistence of cold circulation types may also have been sufficient to negate the effect of warming on the occurrence of cold spells (Kyselý, 2007). Kyselý and Domonkos (2006) report a significant increase in the persistence of all atmospheric circulation types over Europe from the late 1970s to the 1990s. They indicate that the decrease in cyclonic activity over the North Atlantic mid-latitudes and the northward shift of storm tracks which are likely to be associated with anthropogenic climate change support more stable conditions over central Europe. If such changes in the persistence of circulation types could arise as a consequence of climate change, then models should realistically reproduce observed persistence characteristics. The validation of climate models should therefore test their ability to reproduce such persistence relationships.

To reflect this, here we examine the skill of models in reproducing the observed distributions of circulation spell
Two types of spells are examined, firstly those of specific directional types based solely on DIR. Because flow from some directions occurs relatively infrequently, the daily DIR values were grouped into classes of several combined DIR bins in order to ensure a sufficient sample size for statistical analysis. Thus, for example, easterly flow days are less frequent than westerly days and so a greater range of bins is used to represent these days. The bin groupings used to define four directional types are described in the caption for Figure 9. Secondly, spells of anti-cyclonic and cyclonic regimes are also studied. These regimes are defined by the pure anti-cyclonic and cyclonic types of Jenkinson and Collison (1977). These types occur on days where $|\text{VORT}| > 2\text{STR}$. Pure anti-cyclonic (cyclonic) types are thus days where the above condition is satisfied and VORT is negative (positive). The position of each day within a sequence of each...
Figure 7. As in Figure 6 but using daily temperature anomalies of $T_n$.

directional regime was recorded and the frequencies of each spell position are shown in Figure 9. Again, the influence of the driving GCM is shown to be an important source of bias for most regimes, most clearly for spells of winter southerly flow where RCMs driven by HadAM3H have a tendency to simulate longer than observed spells, and summer easterlies where the same models provide a better simulation of the observed distribution than those driven by the ECHAM GCM. However, this is not always the case; for example, for summer westerly spells RCAO_E produces a distribution which is very similar to the observations but the less steep distribution of HIRHAM_E5 is indicative of a tendency to produce disproportionately longer spells. In general terms, the models tend to best reproduce observed distributions for northerly spells and for spells in summer, with the exception of easterly flow. With regards to specific direction types, given that all models underestimate the frequency of easterly flow days, the underestimated frequency of all easterly spell positions was expected,
though the steeper curves do indicate a tendency for them to be disproportionately short. Broadly, reasonably results are observed for spells of meridional flow with all models not only reproducing the absolute frequencies of northerly and southerly spells reasonably well but in most cases, also reproducing the distribution of spell frequencies. It is also worth noting that the overestimation of winter westerlies by the models is manifested in terms of much more frequent, short to moderate spell lengths but with fewer long spells by all models except HAD_H_3P. Given the large overestimation of these days, particularly by ARPEGE_H, it is somewhat surprising that this does not result in more frequent long spells and indicates that the model is not only poorly representing the frequency of zonal flow but also failing to capture some of its important characteristics.

The observed temperature effect of persistent circulation regimes is shown in Figure 10 with a significant effect demonstrated for some spell types. Longer spells of northerly flow, for example, produce lower temperatures...
in all seasons but there are seasonal differences in both the magnitude of the temperature response and the number of days required to produce the greatest response. In winter, by the third day of northerly flow, temperature anomalies have reached their minimum, but in summer the effect is smaller and is on an average reached after a sequence of four days. In contrast, spells of winter easterlies have a greater effect on temperature than northerly flows, but longer spells are required to produce these effects while the warming effect of summer easterlies is not significantly enhanced by persistent flows. Persistent southerly flow produces warmer temperatures throughout the year but the effect is again strongest during winter when temperature anomalies increase throughout the duration of a sequence of such days. Spells of westerlies also produce warmer temperatures throughout most of the year, particularly winter, except summer when the temperature is not sensitive to the persistence of this type of flow. The models have limited skill in capturing these persistence–temperature relationships,
generally reproducing the form but not always the magnitude of the observations. For example, the effects of the persistence of winter northerly flow are overestimated by most models, while the effects of winter westerly spells are underestimated. Other relationships are poorly represented by most models, particularly that of persistent winter southerly flow.

Persistent anti-cyclonic conditions are shown to have a negative effect on temperatures during winter and a positive effect in summer (Figure 11) while persistent cyclonic conditions have a negative effect in winter (not shown). The models generally capture the form of these relationships but fail to reflect the magnitude. All models simulate too strong a temperature response to persistent winter anti-cyclonic regimes, with temperatures decreasing at over 3 times the observed rate in most cases. In summer, RCAO_H, HAD_H_3P and ARPEGE_H do reasonably well in simulating the observed temperature
increase but the other models produce a stronger temperature response. During summer the observed response of maximum temperatures is greater than that of minimum temperatures (Figure 11, right). This may be due to feedbacks between soil moisture and the atmosphere. Brabson et al. (2005) demonstrated that an increase in hot spells may be partly due to extended periods of low soil moisture. Soil moisture is likely to decrease as dry anti-cyclonic conditions persist, especially in summer, resulting in the use of less heat for surface evaporation. Though not shown here, the models do generally reproduce this temperature response – though the magnitude again varies depending upon RCM selection.

5. Discussion and conclusions

This study used seven RCMs from the PRUDENCE suite of model simulations for the 1961–1990 control period and demonstrated that although they are able to reproduce the form and magnitude of the annual cycle of temperature for the central England region, significant biases in monthly means are apparent. All of the models examined here overestimate temperatures during winter while the RCAO and ARPEGE_H simulations are also unable to reproduce the magnitude of summer maxima. Only HAD_H_3P reasonably captures the mean temperature range. RCM selection thus introduces considerable uncertainties into regional temperature simulations. The RMSE statistics calculated for all grid cells in the region indicate that the relative size of model errors varies not only between models but also internally for each model throughout the year and between daily extremes. There is thus no model which fully reproduces all aspects of regional temperatures and could be described as ‘best’. Similar conclusions have been derived for mean UK precipitation. Blenkinsop and Fowler (2007) noted that the relative skill of six of the same RCMs varies throughout the year and also spatially across the UK. Furthermore, the models display different relative skills in capturing precipitation occurrence and intensity. Fowler et al. (2007a) have also noted that HAD_H which has a dry bias in mean precipitation has the greatest relative skill for extremes.

Model simulations of the frequencies of three airflow indices representing the atmospheric circulation produce distributions which qualitatively reflect those obtained from re-analysis data, including the changing seasonal distributions of flow regimes, but which are statistically significantly different. Of most concern are winter simulations, with all models overestimating the frequency of days with westerly/cyclonic flow characteristics. Applying the same indices to a suite of models has indicated the relative extent of this problem with ARPEGE_H noticeably worse than the other models. The model biases in winter temperature simulations are likely to be in part attributable to these errors in regional atmospheric circulation. However, errors in the simulation of DIR do not translate directly into errors in the simulation of surface temperature; despite the large overestimation of westerly flow days by ARPEGE_H, it has relatively small errors in the simulation of mean temperature compared to other models.

Examining the modelled relationships between the airflow indices and temperature indicates that further temperature biases arise from the RCM simulations. Although the models reproduce the general relationships between the atmospheric circulation and near-surface temperature, they have the most difficulty in reproducing the relationships at extreme values of the STR and VORT indices and the relationships with the DIR index during winter and summer. This could have a significant effect when considering the relationship between the atmospheric circulation and extreme temperature events. To demonstrate this, extreme cold events, as defined by those days where the temperature anomaly relative to the 30-year mean is less than the 10th percentile daily
OBSERVED AND MODELLED CIRCULATION–TEMPERATURE RELATIONSHIPS

temperature anomaly, were identified using the daily minimum series for both the observed data and HAD\_H\_3P simulation. This model was selected because it has large errors in the temperature relationships with DIR and VORT at the part of the distribution producing the lowest temperature anomalies. For this purpose, percentiles were calculated using the method described by Horton et al. (2001). Figure 12 indicates significant differences in the relationships between the DIR and VORT indices and the probability of an extreme cold event. The probability of such an event may be twice as likely on a northerly flow day with HAD\_H\_3P while the relationship on anticyclonic days is also poorly reproduced. This poorly represented relationship not only indicates deficiencies in model dynamics but could also limit the application of dynamically downscaled RCM output to studies of some climate change impacts. For example, given the more persistent nature of some anti-cyclonic systems, the greater probability of simulated extremes on these days may lead to biases in the simulation of cold spell characteristics.

The boundary conditions provided by adopting different driving GCMs are a major source of uncertainty in the simulation of the occurrence of circulation regimes. Biases in simulated monthly means, however, are more strongly influenced by RCM selection than by GCM-induced biases in circulation frequencies. The same is also true of the circulation–temperature relationship; for example, during summer there are large differences in the DIR- and STR-temperature relationships simulated by HIRHAM\_E and RCAO\_E.

This study has highlighted a number of implications for the application of statistical downscaling techniques in this region where such methods apply functions between the large-scale circulation and local-scale climate variables. In addition to biases in the frequencies of simulated flow regimes, the observed persistence of some circulation regimes has been demonstrated to exert a strong influence on the relationship between some directional types and temperature, particularly in winter. This may be due to the advection of cold/warm air which consequently leads to an increase in the magnitude of the temperature anomaly as those conditions persist. Failing to consider the effects of persistence could lead to an underestimation of the frequency and magnitude of temperature extremes and in turn to an underestimation of spells of such events, particularly if a change in the persistence of certain regimes is considered a potential symptom of climate change (Kyselý and Domonkos, 2006). Most of the RCM simulations examined here capture the form of these persistence relationships, though the magnitude of the relationship is not always accurately reproduced. This suggests the failure of climate models to represent important atmospheric processes which could be associated with significant future impacts such as heat waves.

Given the uncertainties in model simulation of these features, and of the other climatic properties examined here, the use of multi-model ensembles is considered essential in generating future scenarios of events such as heat waves. Any exercise that is based on only one climate model is constrained by its ability to reproduce observed circulation frequency and persistence. It is now widely appreciated that the production of climate change scenarios should not be restricted to only one model (e.g. Räisänen, 1997; Räisänen and Palmer, 2001; Fowler et al., 2007b; Tebaldi and Knutti, 2007) and multi-model ensembles are now being applied for use in climate change impacts studies (e.g. Fronzek and Carter, 2007). Such approaches enable the uncertainties inherent in the use of RCM output for generating climate change impacts scenarios to be highlighted. For example, Pryor et al. (2005) indicate that the differences between near-surface wind fields for the Baltic region derived in climate change projections are of a magnitude similar to the differences between RCAO fields and re-analysis data for the control period. Blenkinsop and Fowler (2007) use the same PRUDENCE simulations as this study to derive projections of future characteristics of meteorological drought occurrence across the UK, finding that for most

![Figure 12. Observed and modelled probabilities for HAD\_H\_3P of extreme cold events defined by 10th percentile minimum temperature anomalies. Relationships are shown with DIR (left) and VORT (right) for winter. The shaded areas represent the 95% confidence intervals as calculated in Figure 4.](image)

Copyright © 2008 Royal Meteorological Society

DOI: 10.1002/joc
areas the uncertainty in the frequency of long duration droughts encompasses the direction of change. Fowler et al. (2007a) undertook a comparable study of regional UK precipitation extremes but additionally produced multi-model ensemble projections of change to indicate the uncertainties in RCM simulations.

Räisänen (2007), in reviewing assessments of model reliability, raises the questions of how well a model should mimic reality to be believed in and which aspects are most important. He indicates that the lack of a generally accepted figure of merit for measuring model performance makes the ranking of models difficult. This is almost certainly the case but any ranking of models should not only be based on reproducing climatic characteristics important for the application but should also be used with care and not for the selection of one model. Rather, such metrics would be best employed as tools for the weighting of climate models in the generation of probabilistic future scenarios. It may be argued that for climate variables that are strongly coupled to the regional atmospheric circulation, any skill in reproducing observed statistics that is not supported by skill in reproducing the observed circulation regime is achieved without adequately capturing the physical mechanisms that determine the climate. Thus, simulations of future climate could not be viewed with a great degree of confidence. In generating climate change scenarios for the Netherlands, van den Hurk et al. (2007) eliminated GCMs which displayed systematic biases in surface pressure and circulation patterns (van Ulden and van Oldenborgh, 2006). However, Räisänen (2007) also highlights the lack of evidence that models which simulate present-day climate poorly produce outliers in terms of future temperature change. The exclusion of models on the basis of their simulation of current climate is therefore likely to lead to an underestimation of uncertainty in future scenarios. This is particularly important if using some representation of the atmospheric circulation given that the relationship between large-scale circulation patterns and regional climates are characterized by substantial non-stationarity (Slonosky et al., 2001; Beck et al., 2007). Nonetheless, consideration of how well models reproduce atmospheric circulation regimes and their relationships with surface climate offers potential in not only understanding the source of errors in those models but also provides insights into the appropriate choice of predictors for statistical downscaling, most notably, the inclusion of persistence when considering impacts associated with extreme events.

Acknowledgements

Some of this work was part of a PhD studentship funded by the School of Environmental Sciences, University of East Anglia (UEA), Norwich. Thanks for supervision are extended to Phil Jones, Tim Osborn and Steve Dorling of the UEA, and to Hayley Fowler of Newcastle University for useful comments and the development of ideas within the AquaTerra project. The NCEP re-analysis data are available to download from the NOAA-CIRES Climate Diagnostics Center website at http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml. PRUDENCE models data may be obtained from http://prudence.dmi.dk/. CRU dataset TS 2.0 was made available by the Climatic Research Unit, University of East Anglia. Further details may be obtained from http://www.cru.uea.ac.uk/~timm/grid/CRU_TS_2.0.html.

A1. Appendix

A1.1. Airflow indices

Each of the indices is calculated using a network of 16 grid points as shown in Figure 1.

- The westerly flow \( w \) is the westerly (zonal) component of the geostrophic surface wind calculated as the pressure gradient between 50°N and 60°N.

\[
w = \frac{1}{2}(12 + 13) - \frac{1}{2}(4 + 5) \quad (A1)
\]

- The southerly flow \( s \) is the southerly (meridional) component of the geostrophic surface wind and is the pressure gradient between 10°W and 0°.

\[
s = 1.74 \left[ \frac{1}{4}(5 + 2 \times 9 + 13) - \frac{1}{4}(4 + 2 \times 8 + 12) \right] \quad (A2)
\]

- The resultant flow \( f \) strength is the total resultant westerly and southerly flow.

\[
f = (s^2 + w^2)^{\frac{1}{2}} \quad (A3)
\]

- The westerly shear vorticity \( zw \) is the difference of the westerly flow between 45°N and 55°N minus that between 55°N and 65°N and represents the meridional gradient of \( w \).

\[
zw = 1.07 \left[ \frac{1}{2}(15 + 16) - \frac{1}{2}(8 + 9) \right] - 0.95 \left[ \frac{1}{2}(8 + 9) - \frac{1}{2}(1 + 2) \right] \quad (A4)
\]

- The southerly shear vorticity \( zs \) is the difference of the southerly flow between 10°E and 0° minus that between 10°W and 20°W and is the zonal gradient of \( s \).

\[
zs = 1.52 \left[ \frac{1}{4}(6 + 2 \times 10 + 14) - \frac{1}{4}(5 + 2 \times 9 + 13) - \frac{1}{4}(4 + 2 \times 8 + 12) + \frac{1}{4}(3 + 2 \times 7 + 11) \right] \quad (A5)
\]

- The total shear vorticity \( z \) is the sum of the westerly and southerly vorticity.

\[
z = zw + zs \quad (A6)
\]
The constants used in these equations reflect the differing sizes of the grid cells at each latitude. The STR and VORT indices referred to in the paper are calculated from Equations (A3) and (A6) respectively. The direction of flow (DIR) is calculated as \( \tan^{-1}(w/s) \) with 180° added if \( w \) is positive.

A1.2. Confidence intervals

Confidence intervals (CI) of observed mean statistics are calculated using the method described by Osborn et al. (1999). As sample means of daily temperatures are approximately normally distributed, the Student’s \( t \) distribution may be used to calculate CI for each flow bin using the following equation:

\[
CI = \bar{x} \pm t_{\alpha/2,df} \times \frac{s}{\sqrt{n}}
\]

(A7)

where \( \bar{x} \) is the sample mean, and \( t_{\alpha/2,df} \) is the \( t \) value associated with the required confidence level and bin sample size (von Storch and Zwiers, 1999). The sample standard deviation is denoted by \( s \), and the bin sample size by \( n \).

CI for proportions are calculated using the Gaussian approximation to the binomial distribution. The CI is calculated as described by Wilks (1995):

\[
CI = p \pm z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}
\]

(A8)

where \( p \) is the sample proportion which is the best estimate of the binomial event probability \( \pi \), \( z_{\alpha/2} \) is the \( z \) value associated with the desired confidence level, and \( n \) is the sample size.

References


Copyright © 2008 Royal Meteorological Society


