An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate

JAMES MURPHY

Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office, Bracknell, Berkshire, United Kingdom

(Manuscript received 6 November 1997, in final form 2 July 1998)

ABSTRACT

An assessment is made of downscaling estimates of screen temperature and precipitation observed at 976 European stations during 1983–94. A statistical downscaling technique, in which local values are inferred from observed atmospheric predictor variables, is compared against two dynamical downscaling techniques, based on the use of the screen temperature or precipitation simulated at the nearest grid point in integrations of two climate models. In one integration a global general circulation model (GCM) is constrained to reproduce the observed atmospheric circulation over the period of interest, while the second involves a high-resolution regional climate model (RCM) nested inside the GCM.

The dynamical and statistical methods are compared in terms of the correlation between the estimated and observed time series of monthly anomalies. For estimates of temperature a high degree of skill is found, especially over western, central, and northern Europe; for precipitation skill is lower (average correlations ranging from 0.4 in summer to 0.7 in winter). Overall, the dynamical and statistical methods show similar levels of skill, although the statistical method is better for summertime estimates of temperature while the dynamical methods give slightly better estimates of wintertime precipitation. In general, therefore, the skill with which present-day surface climate anomalies can be derived from atmospheric observations is not improved by using the sophisticated calculations of subgrid-scale processes made in climate models rather than simple empirical relationships. It does not necessarily follow that statistical and dynamical downscaling estimates of changes in surface climate will also possess equal skill.

By the above measure the two dynamical techniques possess approximately equal skill; however, they are also compared by assessing errors in the mean and variance of monthly values and errors in the simulated distributions of daily values. Such errors arise from systematic biases in the models plus the effect of unresolved local forcings. For precipitation the results show that the RCM offers clear benefits relative to the GCM: the simulated variability of both daily and monthly values, although lower than observed, is much more realistic than in the GCM because the finer grid reduces the amount of spatial smoothing implicit in the use of grid-box variables. The climatological means are also simulated better in the winter half of the year because the RCM captures some of the mesoscale detail present in observed distributions. The temperature fields contain a mesoscale orographic signal that is skillfully reproduced by the RCM; however, this is not a source of increased skill relative to the GCM since elevation biases can be largely removed using simple empirical corrections based on spatially averaged lapse rates. Nevertheless, the average skill of downscaled climatological mean temperature values is higher in the RCM in nearly all months. The additional skill arises from better resolution of local physiographical features, especially coastlines, and also from the dynamical effects of higher resolution, which generally act to reduce the large-scale systematic biases in the simulated values. Both models tend to overestimate the variability of both daily and monthly mean temperature. On average the RCM is more skillful in winter but less skillful in summer, due to excessive drying of the soil over central and southern Europe.

The downscaling scores for monthly means are compared against scores obtained by using a predictor variable consisting of observations from the nearest station to the predictand station. In general the downscaling scores are significantly worse than those obtained from adjacent stations, indicating that there remains considerable scope for refining the techniques in future. In the case of dynamical downscaling progress can be made by reducing systematic errors through improvements in the representation of physical processes and increased resolution; the prospects for improving statistical downscaling will depend on the availability of the observational data needed to provide longer calibration time series and/or a wider range of predictor variables.

1. Introduction

Global atmospheric general circulation models (GCMs) used in long-term climate change experiments are currently limited to a horizontal resolution of about 300 km (e.g., Johns et al. 1997). Results from GCMs are the primary source of information for assessments of the future impacts of climate change. For some types of impact assessment (e.g., risk of drought or flooding on large catchments) areally averaged quantites such as the grid-box variables output from a GCM may be sufficient; however, in many cases information will be required at a network of point locations (Robinson and Finkelstein 1991), implying the need to ''downscale''

Corresponding author address: Mr. James Murphy, Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY, United Kingdom.
E-mail: jmmurphy@meto.gov.uk

the GCM output. This is particularly true when the model results are required to drive impact models such as agricultural crop models (e.g., Mearns et al. 1997). This paper therefore addresses the problem of downscaling to individual sites, although the importance of validating GCMs at the grid-box scale (e.g., Osborn and Hulme 1997) is also recognized.

Any viable downscaling technique must take account of the regional forcings (arising from orography, coastlines, lakes, land surface characteristics, etc.) known to influence local climate. Three distinct methods can be identified. First, the local variable in question (e.g., surface air temperature T or precipitation P) can be predicted from values of the corresponding variable simulated at nearby GCM grid points, with empirical adjustments to allow for systematic simulation errors and unresolved subgrid-scale effects. Second, output from the GCM can be used to drive a nested high-resolution regional climate model (RCM), whereupon the prediction is derived from simulated values of the variable at nearby RCM points. Following previous authors these will be referred to as ''dynamical'' downscaling techniques. The third method is to use a fully statistical approach based on the development of relationships linking the local variable to atmospheric predictor variables [see Wilby and Wigley (1997) for a review]. The assumption underlying the statistical approach is that the atmospheric circulation in a GCM simulation is more likely to be reliable than the distributions of surface climate elements such as T or P (e.g., Karl et al. 1990), since these are more directly influenced by errors arising from the parameterization of subgrid-scale physical processes such as radiative transfer, cloud and precipitation formation, and turbulent transports in the boundary layer.

Computationally, the most expensive approach is the nested modeling method, in which sea surface temperatures (SSTs) and lateral boundary conditions (surface pressure and atmospheric winds, temperatures, and humidities) saved from the GCM integration are used to drive an RCM of resolution 50–100 km. Following initial experiments by Giorgi and coauthors (Giorgi and Mearns 1991), RCM integrations (sometimes driven by operational analyses rather than GCM output) have now been run by a number of groups for different parts of the world, including several multiannual integrations incorporating the seasonal cycle (Giorgi et al. 1993a; Giorgi et al. 1994; R. G. Jones et al. 1995, 1997; Hirakuchi and Giorgi 1995; Giorgi and Marinucci 1996). Results show that distributions of T and P contain a significant signal on scales not resolved by the global GCMs, whereas the large-scale circulation in the RCM generally follows that of its driving model, provided the domain is not too large. Thus the RCM acts, as required, as a physically based interpolator of the GCM output.

This intensive approach can only be justified if dynamical downscaling based on RCM output is found to be more skillful than either dynamical or statistical downscaling directly from GCM output. In order to address this question the three approaches are compared in this paper. Station observations of daily and monthly averaged T and P over Europe are estimated using the value of T or P simulated by each model at the nearest land point to the target station. These dynamical methods are compared against each other. For monthly mean data, they are also compared against statistical estimates obtained from regression equations linking the station observations to atmospheric predictor variables.

In this study the downscaling methods are assessed under present climate conditions. Use of the methods for climate change predictions will be discussed in a future paper. The required model integrations are performed using a GCM and RCM that are both configurations of the U.K. Meteorological Office Unified Forecast/Climate Model (UM; Cullen 1993). They possess identical subgrid-scale physics packages and the same distribution of vertical levels, differing only in horizontal resolution ($\sim$300 km vs 50 km) and related aspects (time step, diffusion, etc.). They are updated versions of the models used in the simulations of European climate described by R. G. Jones et al. (1995, 1997).

Calibration and verification of the downscaling methods is carried out using observations from June 1983 to February 1994. For each of the 129 months in the period regression relationships required for statistical downscaling are calibrated using parallel observed time series of the predictor and predictand (station) variables that exclude the target month. The relationships are then applied in the target month to generate a downscaling estimate of the observed station value. In order to obtain corresponding dynamical estimates, simulated fields of T and P must be generated that are consistent with the observed circulation over the same period. This is accomplished by forcing the GCM with a time series of operational analyses using the UM data assimilation code for numerical weather prediction (Noguer et al. 1998). The RCM is then integrated using a time series of atmospheric lateral boundary conditions saved from the GCM. During the RCM integration no attempt is made to assimilate analysis data in the interior of the domain. This is because only large-scale analyses were available, the assimilation of which would restrict the development of fine-scale features in the RCM. However, the RCM is integrated using a small domain so that the synoptic-scale circulation is constrained to follow that of the GCM (and hence observations) as closely as possible.

Note that by adopting this experimental design the influence of systematic errors in the GCM circulation is explicitly excluded, as in a similar study recently performed by Kidson and Thompson (1998). This is done to allow a clean comparison between the different approaches under idealized conditions. In practice the influence of circulation errors will degrade the skill of both the statistical and dynamical methods but will not necessarily change their relative performance.

The dynamical downscaling results for monthly means are compared using simple measures of skill:

1) the correlation between downscaled and observed variations in time; and
2) the magnitude of corrections to the mean and variance of the simulated distributions needed to reproduce the mean and variance of the observed distributions.

It is perfectly possible to obtain skillful predictions of (say) the phase of interannual anomalies from a model that gives poor predictions of the interannual variance, or vice versa. Unless a dynamical downscaling method shows skill according to both criteria 1 and 2, it is reasonable to conclude that the model does not adequately represent all the relevant physical processes. On the other hand, skillful performance under present conditions does not guarantee a skillful performance in predicting anthropogenic climate change scenarios, since the physical factors determining the response to changes in radiative forcing are in general different from those that influence the natural variability in the present climate.

The comparison between statistical and dynamical downscaling results for monthly means is based solely on criterion 1. Criterion 2 is not used because a) the statistical methods only predict anomalies from the climatological mean; and b) when calibrating specification equations based on statistical regression the required correction to the variance of the downscaling estimates is directly determined by the correlation between interannual anomalies of the predictor and predictand variables, hence assessing variance errors does not add any new information (in contrast to dynamical downscaling).

The dynamical downscaling results for daily values are compared by assessing the magnitude of corrections needed to reproduce the frequency distributions of the observations.

The models, integrations, and observed data are described in section 2. The results for downscaling of monthly mean and daily data are presented in sections 3 and 4, respectively. A concluding discussion follows in section 5.

2. Models, experiments, and observed data

a. Description of models

The RCM and GCM are both hydrostatic primitive equation models containing 19 levels described by a hybrid vertical coordinate that follows the terrain in the lowest layers. They are integrated on regular latitude–longitude grids of resolution $0.44° \times 0.44°$ and $2.5° \times 3.75°$, respectively, using time steps of 5 min (RCM) and 30 min (GCM). Quasi-uniform resolution is achieved in the RCM by shifting the coordinate pole so that its domain appears as a rectangular equatorial segment on a rotated grid. The state variables are surface pressure, the horizontal wind components, and generalized temperature and moisture variables designed to be conserved during cloud water phase changes (Smith 1990).

The representations of subgrid-scale physical processes are identical in both models. The radiation scheme includes the seasonal and diurnal cycles of insolation and computes fluxes that depend on temperature, cloud amount, water vapor, carbon dioxide, and ozone. Layer cloud cover and precipitation are calculated using a scheme based on an explicit cloud water variable (Smith 1990). The shortwave and longwave cloud radiative properties depend on cloud water path according to Slingo (1989) and Stephens (1978), respectively. Convection is parameterized using a penetrative mass flux scheme (Gregory and Rowntree 1990) including an explicit downdraft (Gregory and Allen 1991). Boundary layer transports are calculated from a first-order turbulent mixing scheme (Smith 1990). Over land, surface temperature, soil moisture content, and snow depth are computed. Evaporation from the soil is limited by stomatal resistance and is further reduced once the soil moisture falls below a critical value. Soil heat transfer is represented using a four-layer thermodynamic model, and a vegetative canopy intercepts and re-evaporates some of the incoming rainfall (Warrilow et al. 1986; Dolman and Gregory 1992). Vegetation and soil properties are prescribed from the $1° \times 1°$ dataset of Wilson and Henderson-Sellers (1985).

Certain improvements to the physics and dynamics have been implemented relative to a previous version of the model used by R. G. Jones et al. (1995, 1997). These changes are listed in appendix A.

b. Nesting technique

The RCM is driven using a time series of output saved from a previous GCM integration, that is, one-way nesting. The state variables are relaxed toward GCM values at each model level across a four-point boundary buffer zone. Orographic heights in the RCM are set equal to GCM values in the buffer zone and also in the four rows and columns immediately inside it. SSTs are prescribed from the values used to drive the GCM. See R. G. Jones et al. (1995) for further details.

c. Experiments

1) Assimilation of operational analysis data in the GCM

As discussed in the introduction, the GCM integration is required to reproduce the observed time-varying atmospheric circulation during the period used for calibration and verification of the downscaling methods. This was achieved by relaxing toward operational analysis fields using the repeated insertion data assimilation

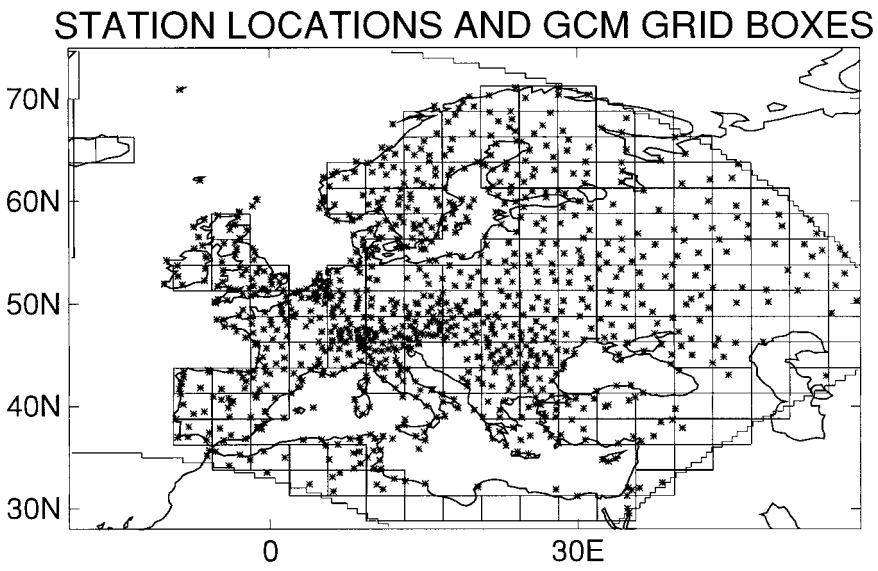

FIG. 1. Domain of RCM (excluding boundary buffer zone) and locations of observing stations. The locations of GCM grid boxes over land are also shown.

scheme employed in numerical weather prediction (NWP) at the U.K. Meteorological Office (UKMO; Lorenc et al. 1991). The required analyses were obtained from a twice-daily archive (0000 and 1200 UTC) of UKMO operational analysis fields interpolated to the GCM grid. Only fields available throughout the experimental period (May 1983–February 1994) were used, consisting of mean sea level pressure (PMSL); temperature on the pressure levels 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, and 50 mb; and horizontal wind components on the pressure levels 400, 300, 250, 200, 150, 100, 70, and 50 mb. See Noguer et al. (1998) for further details of the assimilation procedure. Note that specific humidity fields were not used because they were only available from 1986 and also because they contain systematic biases arising from the NWP model used to produce the analyses (e.g., McNally and Vesperini 1996). The influence of model biases is likely to be smaller for pressure, temperature, and wind fields, since observations of these quantities are much more plentiful than observations of humidity over the area of interest (Fig. 1). Even over Europe, however, the sequence of forcing analyses is likely to contain temporal inhomogeneities due to historical changes in the NWP model, analysis techniques, and observation sources. In future integrations of this type it is planned to minimize such problems by using fields obtained from the reanalysis project recently carried out at the European Centre for Medium-Range Weather Forecasts (Gibson et al. 1997).

2) INTEGRATIONS

The GCM was integrated from 1 May 1983 to 28 February 1994 using an initial state taken from a previous simulation. Atmospheric state variables were continuously relaxed toward observations as described above. The integration was also driven by observed SSTs and sea ice extents (Parker et al. 1995), using monthly mean fields interpolated in time every 5 days. Results from the GCM integration are discussed by Noguer et al. (1998). They find that the time-averaged general circulation remains close to observations, as required. For example, the rms error in climatological mean PMSL is 0.5 mb or less in each season. In addition, observed patterns of daily variance are almost perfectly reproduced by the GCM (not shown), confirming the effective operation of the assimilation procedure.

The RCM integration was started from 10 May 1983, using the European domain shown in Fig. 1. The integration was driven at its lateral boundaries by output from the GCM, as described in section 2b. The use of a small domain reflected the need to constrain the large-scale circulation to follow the GCM, and hence observations, as closely as possible. However, it was not possible to ensure complete correspondence by assimilating atmospheric observations within the interior of the RCM domain, because this would have led to undesirable damping of fine scale features in the RCM solution. Nevertheless, errors in the time-averaged circulation are small apart from in summer when the influence of the lateral boundary forcing is at its weakest, allowing excessively high continental surface temperatures to drive erroneous warming of the troposphere and errors in upper-air winds (Noguer et al. 1998). Daily variations in the RCM circulation follow those in the driving fields, although the correspondence is not perfect. For PMSL the average correlation between simulated and observed anomalies ranges from 0.91 in January to 0.70 in July. The effect on the downscaling results of circulation errors in the RCM will be discussed in section 3.

When referring to GCM output, values of variables that are assimilated directly from the driving analyses, such as mean sea level pressure or 850-mb temperature, will be referred to as ''observed'' values. Values of variables that are not directly assimilated, such as surface temperature or precipitation, will be referred to as ''simulated'' values. All values of RCM variables will be referred to as simulated values. Note, however, that the GCM and RCM integrations are not climate simulations in the normal sense, due to the use of relaxation terms or lateral boundary conditions designed to suppress the development of systematic errors in the atmospheric circulation.

d. Station observations

The required station observations were obtained from daily synoptic reports extracted from the Global Telecommunication System by the Climate Analysis Center (now renamed the Climate Prediction Center; Miskus et al. 1988) and archived at the National Center for Atmospheric Research. Daily values of maximum and minimum screen temperature and accumulated precipitation from 1 June 1983 to 28 February 1994 were taken from this archive for stations lying within the RCM domain. Only stations providing at least 200 valid daily reports for each variable in every month were considered. In order to be selected, a station was required to pass quality control checks applied to the precipitation and temperature data (see appendix B). Following quality control 976 stations were identified as suitable (Fig. 1).

No attempt was made to correct the station time series for errors arising from a) inhomogeneities caused by changes in location, instrumentation, or observing practices, or b) systematic undercatch of precipitation caused by obstruction of the wind by the rain gauge, wetting losses, or evaporation from the gauge. This is because the information required to calculate corrections was not available in most cases. The error arising from undercatch increases with wind speed and is larger for snow than for rain. On average it amounts to about 10% (Legates and Willmott 1990).

3. Downscaling monthly mean data

This section describes downscaling results for station precipitation (P) and diurnally averaged surface air temperature (T), taken as the average of the daily maximum and minimum temperatures.

a. Simulated and observed climatologies

In order to interpret the downscaling results it is first necessary to compare simulated and observed (OBS) climatologies of P and T over the available years (i.e., from 1983 to 1993 for June–December, from 1984 to 1994 for January and February, and from 1984 to 1993 for March–May). We consider multiannual monthly means, denoted by $\langle\ \rangle$, and the interannual standard deviation of monthly values, denoted by σ, concentrating mainly on January and July. In all cases the simulated value for a given station is taken from the nearest land point. The biases in the simulated multiannual means for January and July are generally similar to the seasonal biases for winter and summer discussed by Noguer et al. (1998), whose analysis will be used to help explain the results presented in section 3a(1).

1) MULTIANNUAL MEANS

Figure 2a shows annual cycles of $\langle P\rangle$ averaged over all stations for the RCM, GCM, and OBS. The RCM mean consistently exceeds that of the GCM due to greater large-scale precipitation caused by stronger vertical motion (R. G. Jones et al. 1995). The difference is smallest in summer when the bulk of precipitation is convective. The RCM also overestimates the OBS value, by amounts ranging from 10% to 50%. The OBS values will themselves be too low due to rain gauge undercatch (section 2d); however, this is unlikely to account fully for the differences from the RCM. Noguer et al. (1998) found that part of the precipitation bias in the RCM arises from an excessive supply of moisture from the lateral boundaries: the specific humidities inherited from the GCM are typically 5%–10% too high. In the GCM itself, mean precipitation is closer to OBS but is still too high apart from in the autumn.

The OBS spatial distributions of $\langle P\rangle$ are illustrated in Figs. 3a,d for January and July by plotting the values on a $0.5^\circ \times 0.5^\circ$ grid, where the gridded values are the mean of the nearest five stations weighted by inverse distance.[1] In January (Fig. 3b) the GCM tends to underestimate OBS in west-facing coastal regions due to poor resolution of coastlines and the effects of orographic uplift on the moist climatological westerlies. The GCM overestimates OBS almost everywhere else except over North Africa, although here the density of observing stations is very low (Fig. 1). In the RCM the tendency to underestimate OBS near coasts is removed; however, the overestimate of OBS over central and eastern parts of the domain is enhanced relative to the GCM.

In July $\langle P\rangle$ is too large over much of northern, western, and central Europe in the GCM, with errors exceeding 1 mm day^{-1} in many places (Fig. 3e). On the other hand, there is too little precipitation over southeastern parts of Europe, probably due to excessive dry-

[1] These gridded distributions are provided to show errors in model simulations of station values. In areas such as North Africa and southeastern Europe, where the station network is sparse, the error distributions will not capture local variations in skill. Similarly, the OBS distributions will not show the fine scale signal present in observed climatologies based on a greater density of stations and more sophisticated interpolation techniques (e.g., Hulme et al. 1995).

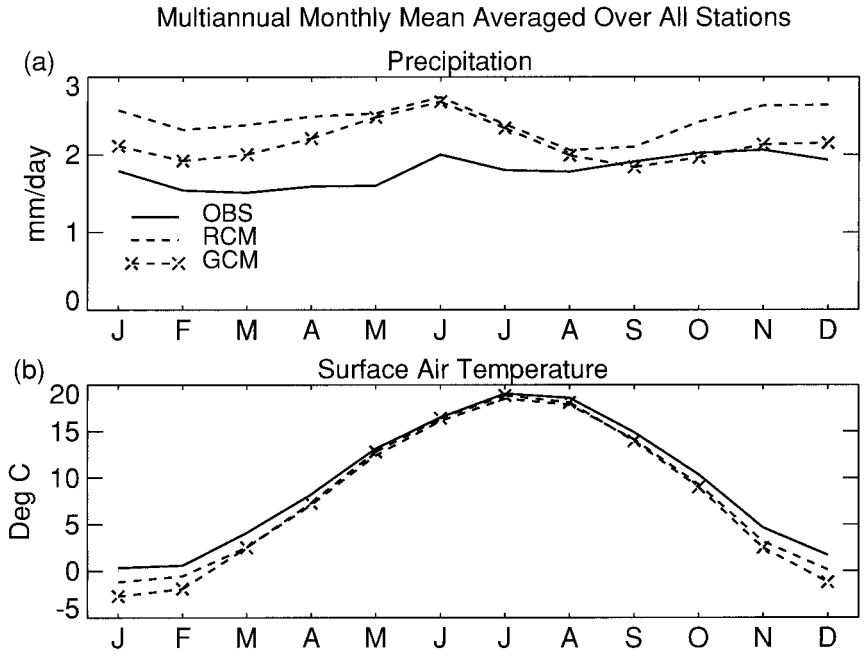

FIG. 2. Multiannual monthly means for 1983–93 (Jun–Dec), 1984–94 (Jan–Feb), or 1984–93 (Mar–May) of (a) precipitation (mm day^{-1}) and (b) diurnal mean surface air temperature (°C) averaged over all 976 observing stations for the observed station value (OBS) and the values simulated at the nearest regional model (RCM) and global model (GCM) land points.

ing of the soil resulting from insufficient cloud cover. Over much of western and central Europe cloud cover, soil moisture, and precipitation are lower in the RCM than in the GCM (Noguer et al. 1998), although both models overestimate the OBS precipitation values (Figs. 3e,f). The RCM also produces too much precipitation adjacent to its northern and eastern outflow boundaries, probably due to enhanced convergence of moisture caused by the application of the boundary relaxation terms (Giorgi et al. 1993b).

Both models slightly underestimate the mean value of $\langle T \rangle$ throughout the year (Fig. 2b), mainly because the mean station elevation is about 200 m lower than that of the relevant model grid points. This can be demonstrated by performing a multiple linear regression of observed station temperature against latitude, longitude, and elevation for each month: the resulting lapse rates suggest a mean elevation bias of -0.9°C for the RCM and -1.3°C for the GCM. However, this cannot fully account for the cooling found in the GCM in winter: the mean temperature for December–February is 1.4°C lower than in the RCM and 2.9°C lower than observed.

In Figs. 4a–c distributions of simulated and observed January temperature are compared after correcting for elevation bias using a lapse rate determined as described above. The excessive winter cooling in the GCM is found to be widespread (Fig. 4b), whereas errors in the RCM are generally smaller (Fig. 4c). According to Noguer et al. (1998) the errors in the GCM in coastal areas are associated with inadequate resolution of coastlines, while in the southern half of the domain the RCM is warmer than the GCM due to stronger solar heating, although this is achieved at the expense of larger errors in cloud cover. Over northern Europe the smaller errors in the RCM appear to arise from the dynamical effects of finer resolution. First, the frequency of low wind speeds is smaller due to the presence of mesoscale circulations not resolved by the GCM. This reduces the frequency of extreme low values of nighttime minimum temperature (see also section 4). Second, stronger vertical motion in the RCM promotes more efficient mixing through the boundary layer. Since there is a temperature inversion in the boundary layer in the northern half of the domain, this implies stronger warming of the surface in the RCM.

In summer the simulated temperatures are too low over most of northern Europe in both models (Figs. 4e,f). This bias is probably associated with excessive evaporation, while the additional bias to the east of the Baltic Sea in the GCM is caused by insufficient surface solar heating (Noguer et al. 1998). In the southern half of the domain excessive solar heating leads to positive errors, with the exception of some areas where precipitation is too high. In the RCM the area affected by warming errors extends farther north and west due to drier soil.

2) INTERANNUAL VARIABILITY OF MONTHLY MEANS

Figure 5a shows annual cycles of $\sigma(P)$ averaged over all stations. The GCM values are consistently too small,

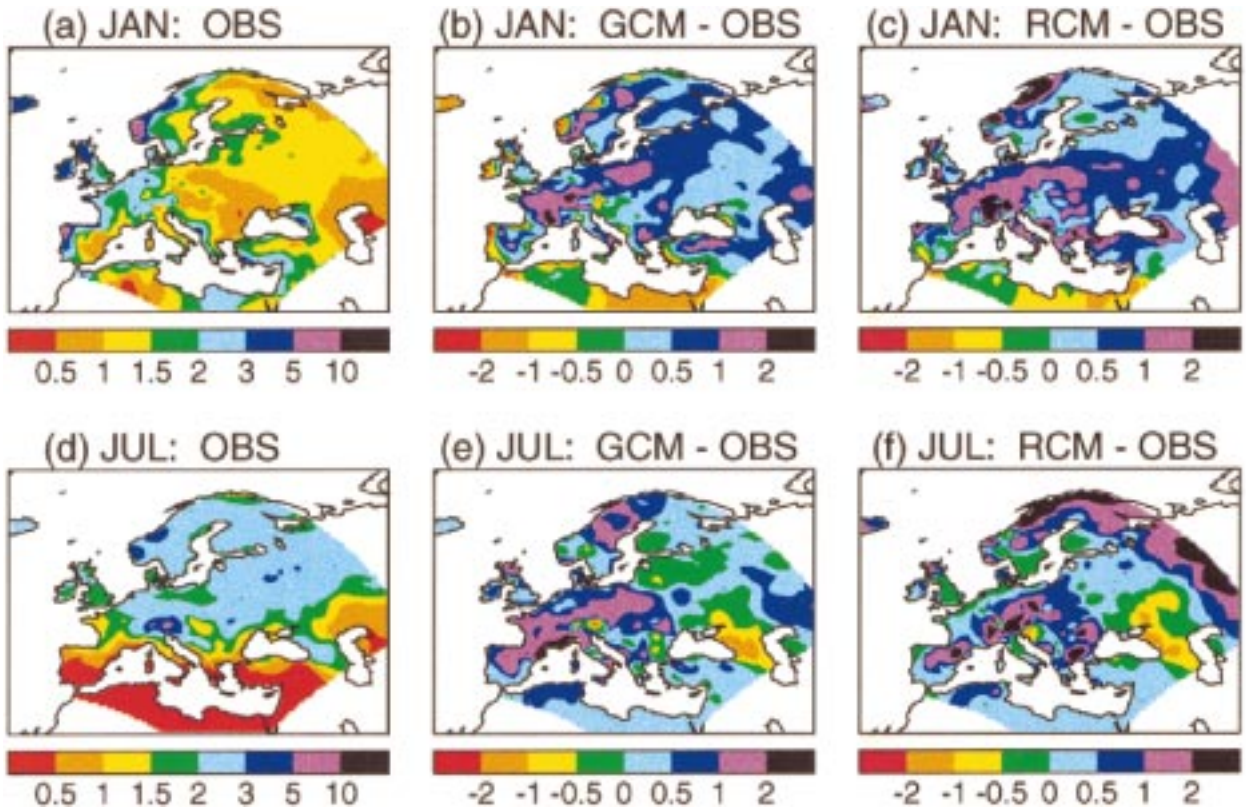

FIG. 3. Spatial distributions of multiannual monthly mean precipitation in mm day^{-1} for Jan [(a)–(c)] and Jul [(d)–(f)]. OBS is observed station values, (a) and (d); GCM − OBS is difference from OBS of value simulated at nearest land point in global model, (b) and (e); RCM − OBS is difference from OBS of value simulated at nearest land point in regional model, (c) and (f). The distributions are plotted on a 0.5° × 0.5° grid, where the values at each grid point are inverse-distance weighted means of the nearest five stations.

whereas the RCM values are quite close to OBS except in winter when they are too large. The first step in understanding these results is to recognize that much of the variability on this timescale arises from the time-averaged characteristics of the daily precipitation distributions (Gregory et al. 1993). In particular, errors in $\sigma(P)$ will be strongly linked to errors in wet day probability (p_w) and errors in the average intensity per precipitation event (i_w). This in turn implies a link between errors in $\sigma(P)$ and errors in $\langle P \rangle$, since $\langle P \rangle = p_w i_w$. In the present study errors in p_w or i_w may arise from the spatial smoothing implicit in the use of grid-box variables (e.g., Mearns et al. 1995), from errors in the physical representation of processes leading to precipitation or (in the case of the RCM) from errors in the supply of moisture from the lateral boundaries. In general the biases in $\sigma(P)$ and $\langle P \rangle$ can be addressed by changing the simulated values of p_w, i_w or both. Here we make an empirical estimate of the impact of changing i_w by scaling each precipitation event in the relevant model by the ratio of the observed and simulated values of $\langle P \rangle$ averaged over all stations. This allows us to estimate the potential of each model to reproduce the observed interannual variability once tuned to simulate the correct level of space- and time-averaged precipitation.

Following application of this scaling both models are now found to show too little variability, but the RCM is always closer to OBS than the GCM (Fig. 5b). This is consistent with Figs. 17 and 18, which show that both models underestimate the variance of daily precipitation (i.e., p_w too large and i_w too small), with the RCM being closer to OBS than the GCM. These errors arise because the RCM and GCM grid-box variables effectively represent spatial means whereas the station variables represent point locations (see section 4). By comparing Figs. 5a and 5b it is clear that the good fit between the mean RCM and OBS values in Fig. 5a results from a cancellation of errors: the tendency to underestimate $\sigma(P)$ arising from spatial smoothing is compensated by an increase in the mean intensity of precipitation events arising from an overactive hydrological cycle.

For bias-corrected precipitation the underestimation of observed variability is geographically widespread for both models. In January (Figs. 6b,c) the only areas where the simulated variance exceeds the observed value are those where $\langle P \rangle$ is significantly too high. In July (Figs. 6e,f) the shortfalls in variance are larger, especially for the GCM. This is because the errors in p_w and i_w are larger than in winter (cf. Figs. 18 and 17), reflecting the seasonal shift from large-scale to convective

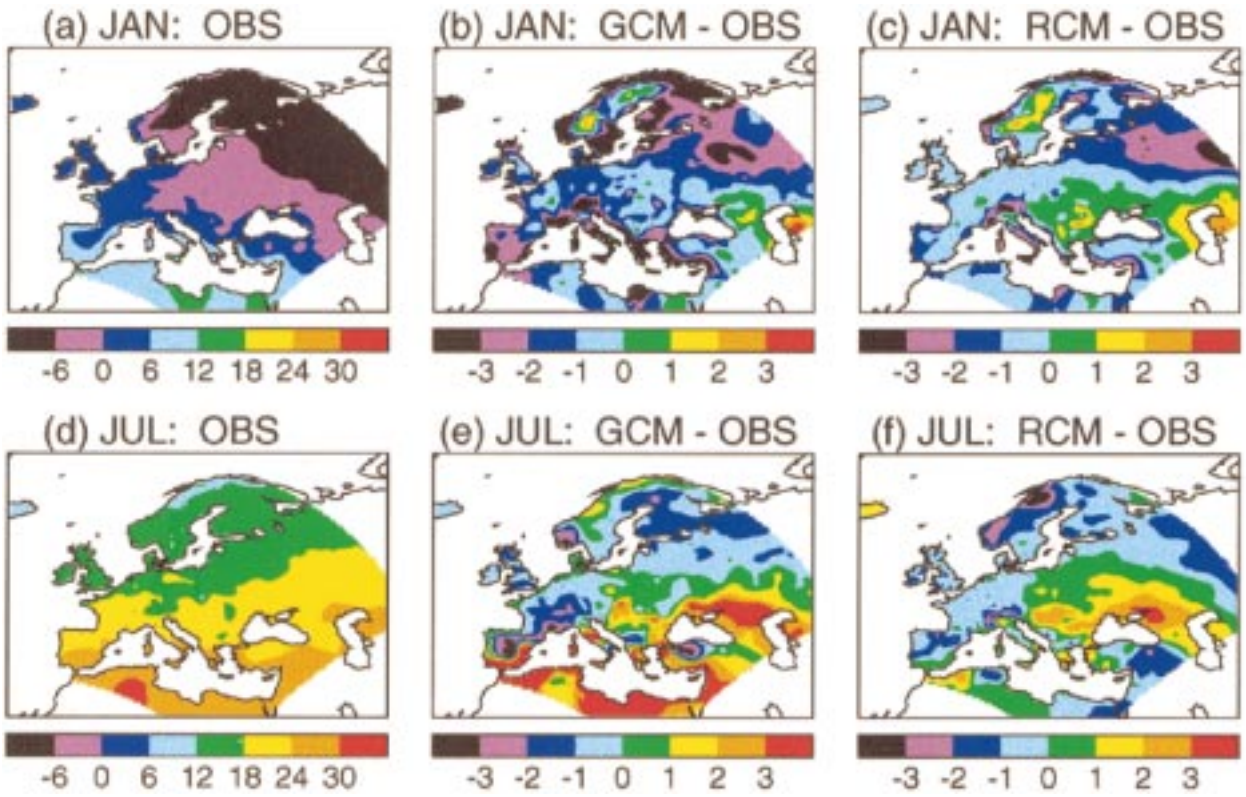

FIG. 4. As in Fig. 3 for surface air temperature (°C). The simulated temperatures have been corrected for elevation bias.

precipitation. The models simulate too little variability everywhere, except in parts of North Africa and southern Europe where the OBS variance is very small and near the outflow boundaries in the case of the RCM.

Both models slightly overestimate the mean value of $\sigma(T)$ in each month (Fig. 5c). The RCM is closer to OBS during the winter while the GCM is closer in summer. In winter (e.g., Fig. 7a) the variability is larger in the northern half of the domain for several reasons, including stronger variability in the synoptic circulation, stronger east–west gradients in climatological mean temperature, and reduced evaporative damping of surface temperature anomalies. The GCM agrees qualitatively with the broad-scale pattern in OBS, but overestimates the OBS value over most of the domain (Fig. 7b). The errors are relatively large along parts of the Atlantic coast, probably due to poor resolution of the modification of mild westerly airstreams by coastal land. Errors in coastal areas are generally smaller in the RCM (Fig. 7c). Some other factors influencing the errors shown in Figs. 7b and 7c are discussed in the context of daily temperature distributions in section 4.

In summer the OBS variances (Fig. 7d) are smaller than in winter due to reduced synoptic variability. However, both models still simulate too much variability in most locations (Figs. 7e,f). Errors exceed 0.5°C in the southeastern part of the domain; in the RCM the region affected by errors of this magnitude extends farther north and west. These errors occur in areas where the soil moisture content drops below the critical value at which evaporation becomes water limited, suggesting that summer drying is too strong in both models, especially the RCM.

b. Downscaling equations

Let X and X_{OBS} represent specified and observed values of monthly precipitation or temperature for a particular station. In general the climatological mean ($\langle\ \rangle$) and variance (σ) of X will differ from the climatological mean and variance of X_{OBS}, so it is necessary to remove these biases empirically when converting X into an optimum downscaling estimate X_{DOWN}, that is,

$$X_{\mathrm{DOWN}} = [\sigma(X_{\mathrm{OBS}})/\sigma(X)](X - \langle X\rangle) + \langle X_{\mathrm{OBS}}\rangle. \quad (1)$$

From Eq. (1) it follows that $\langle X_{\mathrm{DOWN}}\rangle = \langle X_{\mathrm{OBS}}\rangle$ and that $\sigma(X_{\mathrm{DOWN}}) = \sigma(X_{\mathrm{OBS}})$. The magnitudes of the required adjustments to the mean and variance are taken as indicators of the reliability of the downscaling method (see the introduction). A further indicator is the correlation in time between ΔX_{DOWN} and ΔX_{OBS}, the downscaled and observed anomalies in a given year. We thus define three measures of downscaling quality:

1) Scaled error in climatological mean, ERRM(X, X_{OBS}), given by

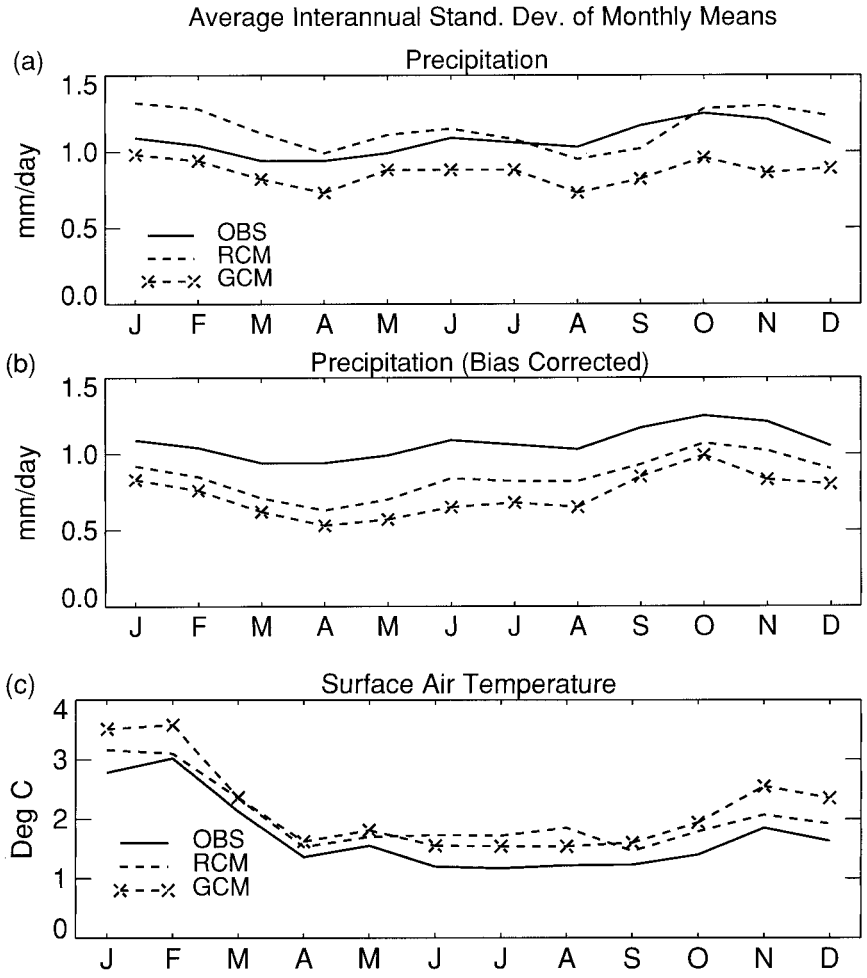

FIG. 5. Interannual standard deviation of monthly means averaged over all 976 observing stations, for the observed station value (OBS) and the values simulated at the nearest regional model (RCM) and global model (GCM) land points. (a) Precipitation (mm day^{-1}), (b) precipitation with simulated values scaled to remove the space- and time-averaged bias relative to OBS, and (c) surface air temperature (°C).

$$\mathrm{ERRM} = |\langle X\rangle - \langle X_{\mathrm{OBS}}\rangle|/\{\tfrac{1}{2}(s(\langle X\rangle) + s(\langle X_{\mathrm{OBS}}\rangle))\}; \quad (2)$$

2) Scaled error in interannual variability, ERRV (X, X_{OBS}), given by

$$\mathrm{ERRV} = |\sigma(X) - \sigma(X_{\mathrm{OBS}})|/\{\tfrac{1}{2}[\sigma(X) + \sigma(X_{\mathrm{OBS}})]\}; \quad \text{and} \quad (3)$$

3) Correlation between the time series of interannual variations, $\mathrm{COR}(X_{\mathrm{DOWN}}, X_{\mathrm{OBS}})$, given by

$$\mathrm{COR} = \langle \Delta X_{\mathrm{DOWN}} \Delta X_{\mathrm{OBS}}\rangle/\{\sigma(X_{\mathrm{DOWN}})\sigma(X_{\mathrm{OBS}})\}. \quad (4)$$

Equations (2) and (3) are used to assess downscaling estimates derived from the simulated value of X at the nearest land point in either model, denoted by X_{GCM} and X_{RCM}. In Eq. (2) $s(\langle X\rangle)$ and $s(\langle X_{\mathrm{OBS}}\rangle)$ refer to the standard deviations of $\langle X\rangle$ and $\langle X_{\mathrm{OBS}}\rangle$ over all stations, thus the simulation error is expressed as a fraction of the mean spatial variability of the predictor and predictand. This scaling allows a fair comparison between downscaling estimates obtained from variables based on different degrees of spatial smoothing.

Equation (4) is applied to X_{GCM} and X_{RCM}, and also to various statistical estimates X_{STAT}. The statistical estimates are based on atmospheric predictor variables obtained from the GCM output. The predictors considered are listed in Table 1 alongside the station variable(s) they are used to estimate. CIRC is derived from the PMSL field; WSPD, UWND, VWND, and VORT are all obtained from the 10-m wind field; and T850 and Q850 are temperature and specific humidity at 850 mb. KIND is the K-index, a stability index known to be related to convective precipitation (Peppler and Lamb 1989). It is defined as

$$\mathrm{KIND} = (\mathrm{T850} - \mathrm{T500}) + \mathrm{TD850} - (\mathrm{T700} - \mathrm{TD700}), \quad (5)$$

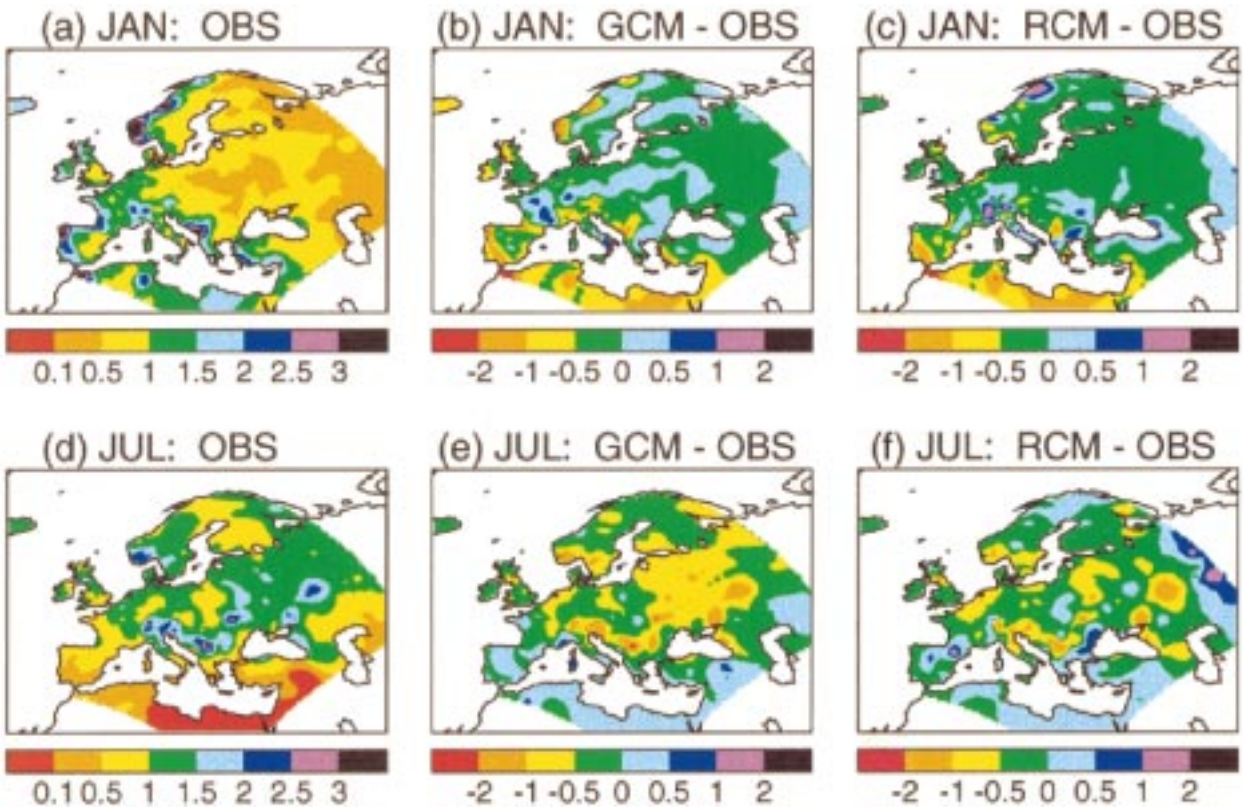

FIG. 6. Spatial distributions of the interannual standard deviation of monthly precipitation in mm day^{-1} for Jan [(a)–(c)] and Jul [(d)–(f)]. OBS is observed station value, (a) and (d); GCM − OBS is difference from OBS of value simulated at nearest land point in global model, (b) and (e); RCM − OBS is difference from OBS of value simulated at nearest land point in regional model, (c) and (f). Simulated precipitation values have been scaled to remove the space- and time-averaged bias as in Fig. 5b. Distributions gridded for plotting as in Fig. 3.

where T700 and T500 are the temperatures at 700 and 500 mb, and TD850 and TD700 are the dew points at 850 and 700 mb.

For present purposes values of the predictors are interpreted as observations even though they are actually values output from an assimilation integration of the GCM. The CIRC and T850 predictors are obtained from fields assimilated directly from analyses (see section 2c) and are essentially unaffected by systematic biases of the GCM. Near-surface winds were not assimilated directly in the GCM integration; however, the values are strongly constrained by the assimilation of the PMSL field so the interannual anomalies of WSPD, UWND, VWND, and VORT considered here are unlikely to be seriously degraded by model biases. The Q850 and KIND predictors are more likely to be biased because they rely on moisture variables that are (a) not assimilated directly and (b) not necessarily strongly constrained by the assimilation of pressures, winds, and temperatures due to the influence of subgrid-scale moist processes in the GCM physics. Hence the anomalies of Q850 and KIND derived from the GCM are only estimates of the observed values of Q850 and KIND.

Apart from CIRC all the predictors in Table 1 are intended to capture regional relationships and consist simply of the value of the relevant quantity at the GCM land point nearest to the target station. CIRC is included to capture relationships between local temperature or precipitation and the large-scale flow. Principal component (PC) analysis of the PMSL field is used to capture the variability of the flow in a few key modes [see, e.g., Hewitson and Crane (1992), Zorita et al. (1992), and Corte-Real et al. (1995)], following which the specified anomalies ΔCIRC are formed from linear combinations of the leading five PCs, that is,

$$\Delta\text{CIRC} = \sum_{k=1}^{5} a_k \text{PC}_k, \tag{6}$$

where the a_k are regression coefficients. See section 3c for more details.

The skill of each of the individual predictors listed in Table 1 is assessed in section 3d. In addition, the skill of various linear combinations of the predictors is also assessed. In general,

$$\Delta X_{\text{STAT}} = \alpha \Delta X_1 \quad \text{or}$$

$$\Delta X_{\text{STAT}} = \alpha \Delta X_1 + \beta \Delta X_2 + \gamma \Delta X_3 + \ldots, \tag{7}$$

where ΔX_1, ΔX_2, ΔX_3, etc., are predictors from Table 1 and α, β, γ, etc., are regression coefficients. The cal-

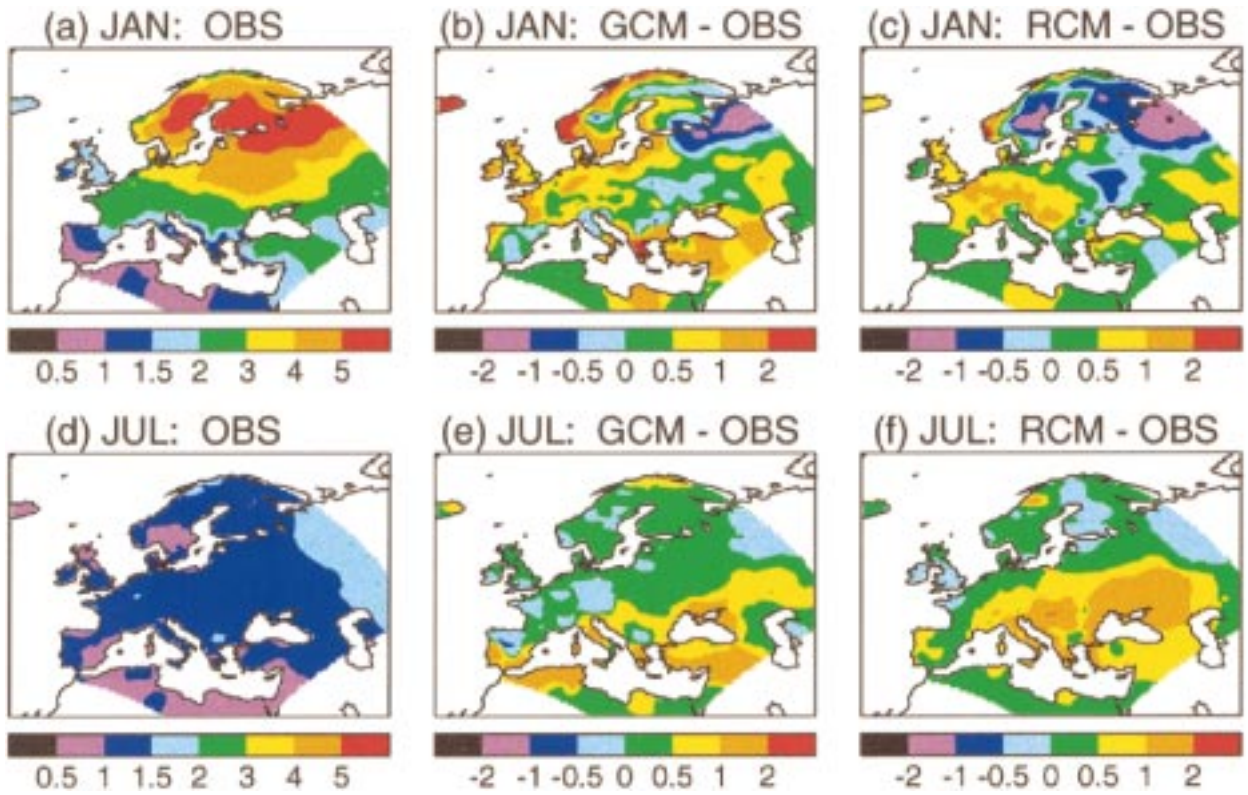

FIG. 7. As in Fig. 6 for surface air temperature (°C).

culation of the regression coefficients employed in Eqs. (6) and (7) is described in the following section.

c. Calibration and verification

When validating downscaling techniques based on regression relationships such as Eqs. (6) or (7) it is normal to split the data time series into independent calibration and verification samples in order to avoid overestimating the skill of the method (e.g., Wigley et al. 1990). Unfortunately, our time series are not long enough to allow this; however, independent verification can still be performed by the following procedure. For a given target month (say, December 1983) calibration is done using data from all the Novembers, Decembers, and Januarys in the dataset apart from December 1983. The calibrated equations are then used to estimate ΔX_{OBS} in December 1983. The process is then repeated for each December in turn to perform the downscaling for each year. Following this the correlation between the downscaled and observed anomalies, hereafter denoted by $\mathrm{COR}_{\mathrm{IND}}$, is obtained by applying Eq. (4). $\mathrm{COR}_{\mathrm{IND}}$ can be compared against $\mathrm{COR}_{\mathrm{CAL}}$, the average of the correlations between the time series of downscaled and observed anomalies found in each calibration sample. Verification results are produced for each month of the year by this method. In each case the months adjacent to the target month are included in the calibration samples, the size of which varies from 29 to 32 monthly observations. (Increasing the calibration sample by including five months instead of three yielded inferior results in independent verification.)

In order to calibrate Eq. (6) the dataset of (29–32) PMSL anomaly fields is expressed in terms of its q (=28–31) empirical orthogonal functions (EOFs). The spatially orthogonal and normalized EOFs are the eigenvectors of the covariance matrix of the dataset. Each EOF possesses an associated time-varying PC representing the loading on the relevant pattern, for example,

TABLE 1. Predictors used for statistical downscaling.

Predictor	Predictand	Description
CIRC	*T, P*	Linear combination of projections onto large-scale flow patterns
WSPD	*T, P*	Regional value of near-surface wind speed
UWND	*T, P*	Regional value of near-surface westerly wind
VWND	*T, P*	Regional value of near-surface southerly wind
VORT	*P*	Regional value of near-surface vorticity
T850	*T*	Regional value of temperature at 850 mb
Q850	*P*	Regional value of specific humidity at 850 mb
KIND	*P*	Regional value of stability index

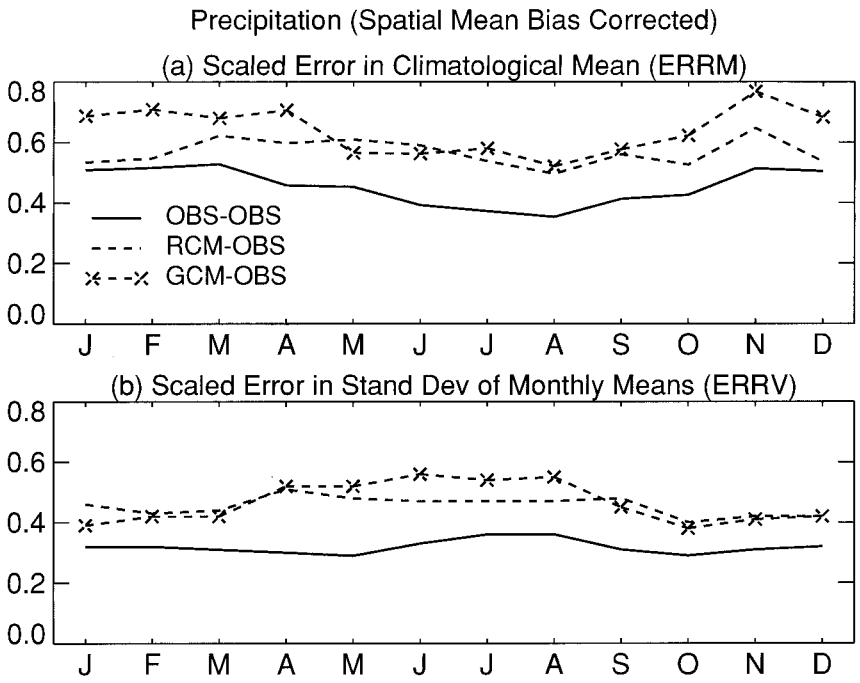

FIG. 8. (a) Scaled error in climatological mean, ERRM, averaged over all 976 stations for predictions of monthly mean precipitation. OBS − OBS are predictions based on nearest observing station to predictand station; RCM − OBS are predictions based on nearest regional model land point to predictand station; GCM − OBS are predictions based on nearest global model land point to predictand station. Simulated precipitation values have been scaled to remove the space- and time-averaged bias relative to OBS. (b) As in (a) for the scaled error in interannual variability, ERRV.

$$\Delta\text{PMSL}(x, y, t) = \sum_{k=1}^{q} \text{PC}_k(t)\text{EOF}_k(x, y). \qquad (8)$$

The PCs are ordered in terms of explained variance (1 = highest). The required regression coefficients are obtained by applying Eq. (6) to the calibration data with ΔX_{CIRC} replaced by ΔX_{OBS}. Since the PCs are temporally uncorrelated, each a_k is determined via an independent univariate least squares regression, hence

$$a_k = [\sigma(X_{\text{OBS}})/\sigma(\text{PC}_k)]\text{COR}(\text{PC}_k, X_{\text{OBS}}). \qquad (9)$$

In general up to q PCs can be retained in the downscaling equation; however, the best results in independent verification were obtained with five retained (sufficient to explain 84%–96% of the variance). The downscaling estimates are generated by applying Eqs. (6) and (1) to the target month. The values of $\sigma(X_{\text{CIRC}})$ and $\sigma(X_{\text{OBS}})$ required in Eq. (1) are determined from the calibration dataset.

Calibration of the specification equations [Eq. (7)] requires the determination of regression coefficients α, β, etc. This is achieved by applying Eq. (7) to the calibration data with ΔX_{STAT} replaced by ΔX_{OBS}. In cases involving the combination of two or more predictors (e.g., $\Delta X_1 \equiv \Delta X_{\text{CIRC}}$, and $\Delta X_2 \equiv \Delta X_{\text{WSPD}}$) it cannot be assumed that the predictors are independent so the multiple linear regression problem does not reduce to a superposition of independent univariate regressions [cf. Eq. (6)]. The downscaling estimates are then generated by applying Eqs. (7) and (1) to the target month.

For downscaling based on X_{GCM} and X_{RCM} no regression relationships are needed. The purpose of calibration is simply to calculate the means and variances of the observed and simulated values of X required to generate downscaling estimates via Eq. (1).

Values of ERRM, ERRV, and COR_{IND} were calculated for each station in each month. Results are presented in the following section.

d. Downscaling results

1) SPECIFICATION OF CLIMATOLOGICAL MEANS AND INTERANNUAL VARIABILITY

Figures 8a,b show seasonal cycles of ERRM and ERRV for precipitation, averaged over all stations. Scores are plotted for simulated values at the nearest land point in each model and also for the nearest observing station to the predictand station. The latter values (labelled OBS-OBS) give a simple estimate of the maximum skill likely to be achievable by improving the models in the future. The simulated precipitation values have been scaled to remove the spatially averaged bias (see discussion in section 3a). For ERRM the average

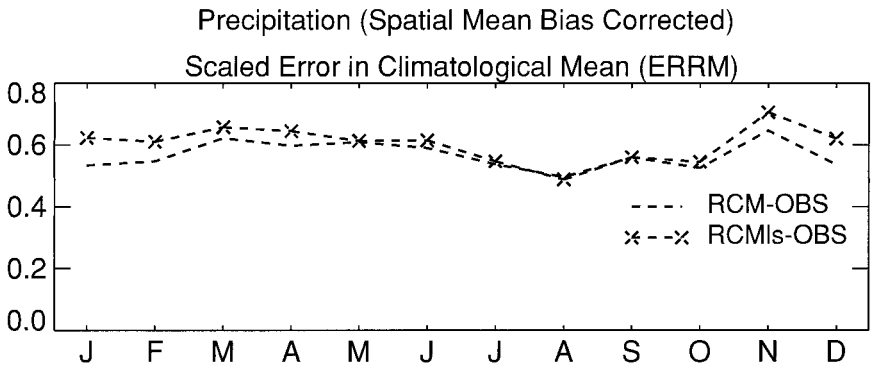

FIG. 9. Values of ERRM for predictions of monthly mean precipitation obtained from the regional model fields, repeated from Fig. 8a. These are compared against corresponding predictions obtained from the large-scale component of the fields (RCMls). The large-scale component is defined as the component resolved on the global model grid.

RCM scores are consistently better than those of the GCM from October to April. From May to September the scores for the two models are similar. The OBS − OBS scores are always better than those for either model. This indicates the presence of systematic biases throughout the annual cycle, although the gap between the OBS − OBS and RCM − OBS scores is encouragingly small in winter. In summer, however, the gap is larger. This is consistent with the results of Noguer et al. (1998), who found that the influence on precipitation of (imperfectly represented) regional physical processes relative to the influence of the (perfect) large-scale forcing supplied from the lateral boundaries was greater in summer than in winter.

In order to intepret the results of Fig. 8a the monthly precipitation distributions in the RCM are decomposed into large-scale and mesoscale components. The mesoscale component represents the signal on scales too fine to be resolved by the GCM grid. Following Noguer et al. (1998) the decomposition is performed as follows, using values at land points only: 1) interpolate the RCM fields to a regular $0.5^\circ \times 0.5^\circ$ grid; 2) average all the $0.5^\circ \times 0.5^\circ$ grid points lying within each GCM grid square (see Fig. 1) to obtain the large-scale component, hereafter RCMls; and 3) subtract the large-scale component from the value at each $0.5^\circ \times 0.5^\circ$ grid point to obtain the mesoscale component. Figure 9 compares values of ERRM for RCMls against those for the unfiltered RCM output, copied from Fig. 8a. Addition of the mesoscale component improves the RCM score between October and April, and has little effect between May and September. The results of Noguer et al. (1998) suggest two reasons for this: first, the mesoscale signal is more skillful in winter; second, the signal is more strongly correlated with orographic height in summer. The second factor will reduce the impact of the mesoscale signal on ERRM due to the underrepresentation of high-elevation sites in our station dataset. Between October and April the scores for RCMls are superior to those for the GCM, indicating that the influence of the mesoscale orographic forcing on precipitation in the RCM projects onto the large-scale component as well as the mesoscale component. This does not occur in summer, when the orographic forcing occurs mainly through convection driven by surface heating from the mountaintops, rather than through interactions with the large-scale flow.

For ERRV (Fig. 8b), the RCM consistently outperforms the GCM from May to August, while the scores for the two models are similar from September to April, apart from in January when the GCM is better. The model scores are significantly worse than the OBS − OBS scores, demonstrating that there is considerable scope for improving the simulation of interannual variability for precipitation. The inferior performance of the GCM in summer is due to the general reduction in variability relative to the RCM (Fig. 5b), which occurs because GCM grid points represent larger areas than do RCM grid points (see also section 4). Note, however, that the ERRV scores in Fig. 8b are influenced by the patterns of variability (e.g., Fig. 6) as well as the spatially averaged values; thus the RCM scores in winter are not better than the GCM scores even though the spatially averaged variances are still more realistic in the RCM (Fig. 5b).

Values of ERRM and ERRV for elevation-corrected temperature are given in Figs. 10a,b. The scores are consistently better than the corresponding values for precipitation (cf. Fig. 8). Nevertheless, the model values are always significantly worse than the OBS − OBS scores. For ERRM the RCM consistently shows better skill than the GCM. For ERRV the relative performance of the RCM and GCM closely follows that for the spatially averaged value of σ (cf. Figs. 5c and 10b). This result, which is not obvious, probably occurs because monthly temperature anomaly patterns are dominated by regional/continental scales rather than by mesoscale detail. This is supported by the OBS − OBS results, which show that highly skillful estimates of temperature variability can be made using the nearest observing sta-

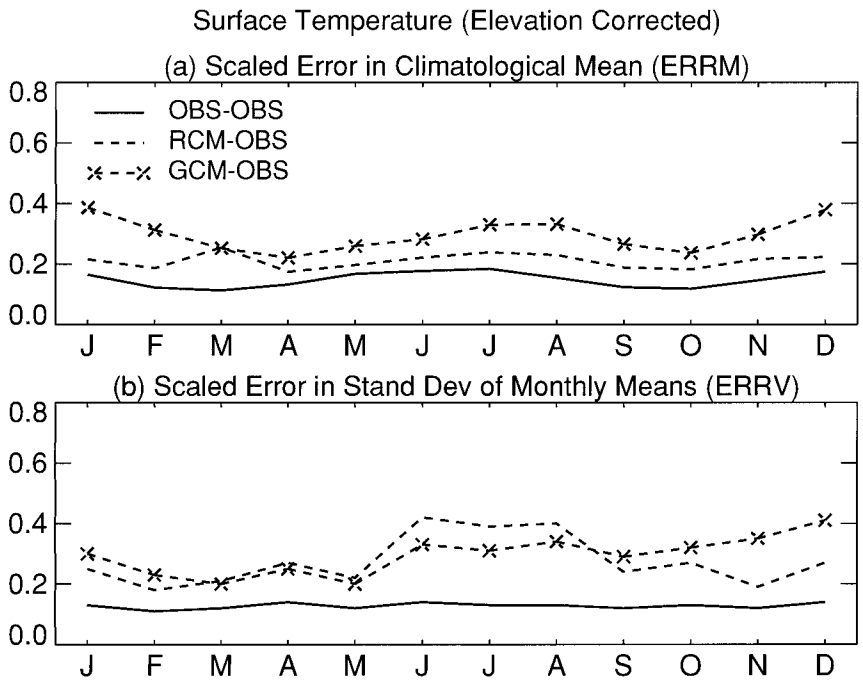

FIG. 10. As in Fig. 8 for surface air temperature. Following Fig. 4, the predictor temperatures have been corrected to allow for the difference in elevation between the predictor grid point (or station) and the predictand station using spatially averaged lapse rates that vary with the season.

tion, even though the mean distance between neighboring stations is 74 km. In order to reduce the RCM errors in summer it will be necessary to cure the excessive drying of the soil referred to in section 3a.

The elevation corrections applied in Fig. 10a are based on domain-averaged lapse rates that vary only with season. If local lapse rates depart significantly from the domain-averaged value, the elevation corrections will not be reliable. Unreliable elevation corrections would tend to reduce the skill of the GCM relative to the RCM because higher resolution leads to smaller differences in elevation between the nearest model grid point and the predictand station. In order to assess the elevation corrections the following procedure was used.

1) For a given month the distribution of climatological mean land temperature simulated by the GCM was interpolated to RCM grid points with elevation adjustments calculated using the same lapse rate employed in Fig. 10a. Note that these elevation ''adjustments'' are applied to allow for differences in elevation between GCM and RCM grid points, whereas the elevation ''corrections'' in Fig. 10a allowed for differences in elevation between model grid points (either RCM or GCM) and observing stations.
2) The mesoscale component from the elevation-adjusted GCM distribution was calculated and compared against the (unadjusted) mesoscale component of the climatological distribution simulated by the RCM.

The results for January are shown in Fig. 11. The RCM mesoscale signal is closely related to orographic height [cf. Figs. 11a and 11c; see also Noguer et al. (1998)]; however, the empirically created GCM signal (Fig. 11b) successfully reproduces the maxima and minima associated with local variations in elevation: the correlation between the distributions in Figs. 11a and 11b is 0.88. Similar results are obtained in other months, suggesting that elevation adjustments based on a spatially averaged lapse rate are sufficiently reliable to allow the orographic mesoscale component simulated by the RCM to be generated empirically from the GCM output. This implies that the corrections applied in going from model grid point to station elevations in Fig. 10a are probably also reliable, in which case alternative explanations are required for the superior ERRM scores for the RCM.

One possibility is that the additional skill arises from better resolution of land–sea contrast in coastal regions: roughly half of the 976 observing stations lie within 30 km of the coast (Fig. 1). For these stations, the mean magnitude of the difference between the simulated and observed temperatures is 1.5°C in January and 1.3°C in July for the RCM, compared with values of 2.3° and 1.9°C for the GCM. Thus the RCM does indeed perform better in coastal regions. However, it also performs better in inland regions; the mean error magnitudes for the RCM are 1.4°C in January and 0.9°C in July, compared to 2.8° and 1.4°C for the GCM. Thus the improved skill in the RCM is a widespread feature arising, to a sig-

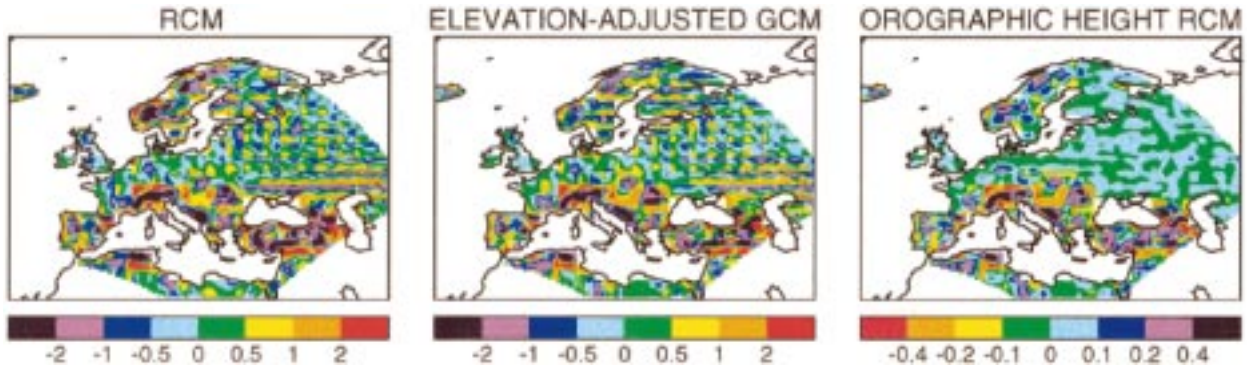

FIG. 11. (a) The mesoscale component (in °C) of the distribution of climatological mean surface air temperature simulated by the RCM in Jan. The mesoscale component is the component that cannot be resolved on the GCM grid. (b) A mesoscale component obtained from the corresponding GCM distribution after interpolating to the RCM grid with adjustments to allow for differences in elevation between the RCM and GCM grid points. The adjustments use the same spatially invariant lapse rate used for Jan in Figs. 4 and 10. (c) The mesoscale component of orographic height (km) in the RCM.

nificant degree, from the large-scale component of the simulated distributions. Differences between the large-scale distributions of the two models are associated with the consequences of enhanced horizontal resolution [see section 3a and Noguer et al. (1998)]. These include stronger surface winds, more efficient heat transfer through the boundary layer, reduced cloud cover, stronger surface solar heating, and lower soil moisture in summer. Although the impact of these differences on surface temperature is generally beneficial, this is sometimes achieved at the expense of larger errors in other aspects of the simulation (i.e., cloud cover, solar radiation, and soil moisture). On the other hand, biases in both the mean and variance of July temperature are larger in the RCM over parts of central and eastern Europe because increasing the resolution upsets an error balance operating in the GCM [i.e., excessive solar heating tends to deplete soil moisture, but this is opposed reduced drainage from the root zone caused by an underestimate of intense rainfall events (Noguer et al. 1998)]. Clearly, therefore, simultaneous optimization of the performance of RCMs and their driving GCMs will require careful study of the effects of varying the horizontal resolution and the implications for the formulation of subgrid-scale physical parameterization schemes.

In summary, the RCM yields improvements relative to the GCM in downscaling the mean and variance of monthly station temperatures and precipitation amounts. For precipitation the improvements stem directly from enhanced horizontal resolution, which allows the RCM to capture some of the mesoscale detail in observed climatological mean distributions and also reduces the underestimation of interannual variability because the gridpoint variables represent a smaller area than in the GCM. For surface temperature the improvements stem indirectly from enhanced resolution, through the effect of changes in the dynamics and the hydrological cycle on the large-scale component of the simulated distributions. The RCM is never less skillful than the GCM, apart from in the simulation of temperature variability in summer.

TABLE 2. COR_{IND} for winter and summer half-years averaged over all stations.

Predictor	*T*: Oct–Mar	*T*: Apr–Sep	*P*: Oct–Mar	*P*: Apr–Sep
CIRC	0.60	0.48	0.47	0.30
WSPD	0.37	0.31	0.33	0.22
UWND	0.38	0.31	0.29	0.21
VWND	0.46	0.29	0.27	0.22
VORT	n/a	n/a	0.36	0.32
T850	0.77	0.82	n/a	n/a
Q850	n/a	n/a	0.30	0.14
KIND	n/a	n/a	0.45	0.34

2) SPECIFICATION OF MONTHLY ANOMALIES

In this section we identify optimum statistical downscaling estimates from the range of possible predictors listed in Table 1 and then compare them against estimates based directly on model output.

Table 2 shows the average skill of each of the individual statistical predictors for winter and summer half-years (hereafter ''extended winter'' and ''extended summer''). Skill is measured by COR_{IND}, the correlation between estimated and observed monthly mean anomalies. For estimates of temperature, T850 explains 60% or more of the variance in either half-year and comfortably exceeds the mean score of any other predictor. This is consistent with results published recently by Kidson and Thompson (1998) who found that 1000–500-mb thickness was usually the leading predictor for screen temperature at stations in New Zealand. However, the skill varies substantially with location. In January, for example, COR_{IND} exceeds 0.9 ($>$80% of variance explained) over most of central and northern Europe but falls to 0.7 or less over western and eastern Europe. In these areas the large-scale circulation (CIRC)

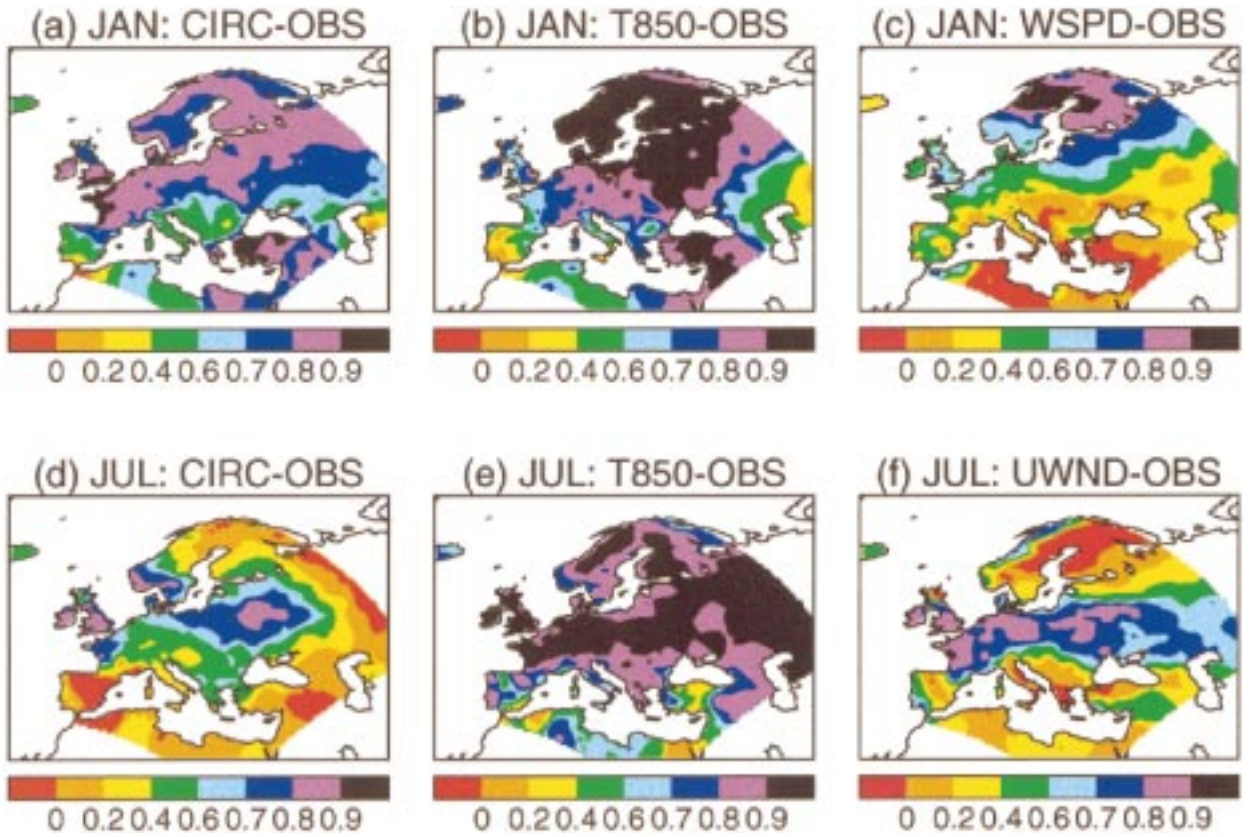

FIG. 12. Spatial distributions of the correlation COR_{IND} between estimated and observed anomalies of monthly mean surface air temperature in independent verification in Jan [(a)–(c)] and Jul [(d)–(f)]. Scores are given for certain statistical predictors selected from the list available in Table 1.

is often a better predictor (cf. Figs. 12a and 12b), particularly over western Europe where strong land–sea contrasts and high variability in the synoptic circulation enhance the importance of advection in determining local temperature. On average the skill of CIRC is higher than that of any of the regional wind predictors UWND, VWND, or WSPD (Table 2); however, over particular regions one of these predictors sometimes matches or exceeds the CIRC score. For example, WSPD is a highly skillful predictor over northern Europe in January (Fig. 12c), probably because the presence of a strong climatological temperature inversion creates favorable conditions for wind-speed-dependent variations in turbulent heat transport. In July the circulation generally explains less of the variation of temperature than in January (cf. Figs. 12d and 12a), although UWND is a skillful predictor at midlatitudes (Fig. 12f) due to the influence of variations in the strength of the climatological westerlies (stronger westerlies lead to lower temperatures by advecting cooler, cloudier air from sea to land).

For estimates of precipitation there is no single preferred predictor in terms of the average scores (Table 2). In both extended seasons the predictors with the highest scores are CIRC, VORT, and KIND. Each of these represents different atmospheric properties, all of which have been found to be related to precipitation in previous regional studies. For example, large-scale flow anomalies (represented here by CIRC) explain much of the variance of Iberian precipitation in winter (Zorita et al. 1992; Corte-Real et al. 1995); links between regional vorticity (VORT) and precipitation have been identified at sites in Europe and the United States (e.g., Wilby 1997; Wilby et al. 1998); regional stability (KIND), which is a plausible surrogate for either convective or synoptic-scale ascent, is known to be a useful indicator of convective precipitation over the United States in summer (Peppler and Lamb 1989). Here the distributions of COR_{IND} show wide variations in skill between these predictors (Fig. 13). In January CIRC often gives higher scores than VORT or KIND over western Europe, whereas VORT and KIND both explain more variance than CIRC over northern Russia and much of Turkey and the Balkan region. In July the skill is generally lower. This probably reflects the seasonal shift from stratiform to convective precipitation, the latter being a mesoscale phenomenon influenced by local forcings not represented in our set of predictors. Nevertheless, CIRC, KIND, and VORT all achieve correlations exceeding 0.6 at some locations in the northern half of the domain,

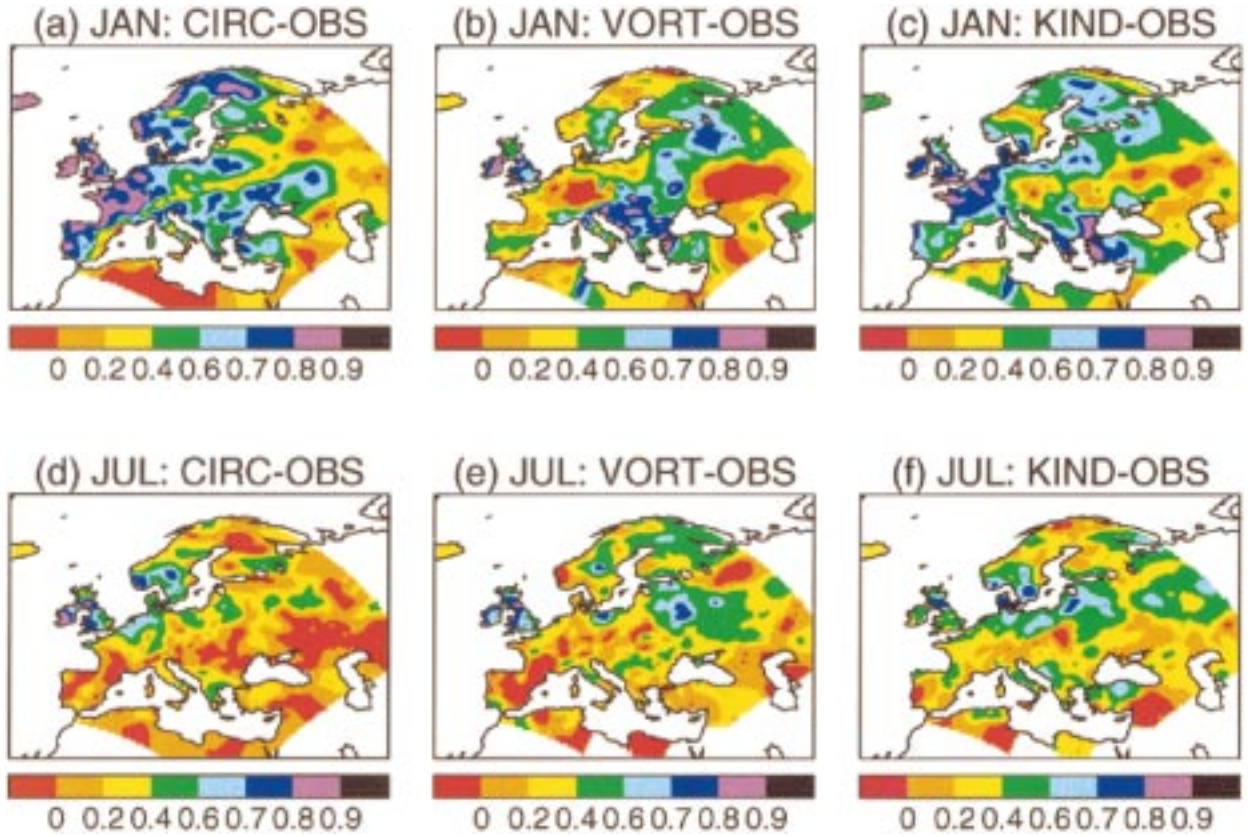

FIG. 13. As in Fig. 12 for estimates of precipitation.

although the scores rarely exceed 0.4 in the southern half. Of the remaining predictors, those based on regional winds explain less variance than CIRC or VORT on average (Table 2), although at individual stations UWND, VWND, or WSPD is sometimes the leading predictor. In general moisture (Q850) shows limited skill in extended winter but very little in extended summer. This is disappointing from the standpoint of climate change applications, since increases in atmospheric water vapor are expected to be an important factor in determining the response of precipitation to changes in greenhouse gases (Manabe and Wetherald 1975). The skill of Q850 in predicting present climate anomalies could probably be improved by using a moisture variable free from the influence of GCM biases (see section 3b); nevertheless, the present result illustrates the difficulties associated with using historical data to calibrate relationships intended for application in future climates.

TABLE 3. Annual average frequency of selection (%) for statistical predictor sets.

Temperature		Precipitation	
Predictor set	Frequency	Predictor set	Frequency
CIRC*	0.0	CIRC	1.2
UWND	0.1	UWND	6.6
CIRC-UWND*	0.0	CIRC-UWND	1.7
VWND	0.6	VWND	4.4
CIRC-VWND*	0.0	CIRC-VWND	2.7
WSPD	0.1	WSPD	8.9
CIRC-WSPD*	0.0	CIRC-WSPD	2.7
T850	52.1	Q850	2.1
CIRC-T850*	6.7	CIRC-Q850	0.9
UWND-T850*	2.2	VORT	23.9
VWND-T850*	1.1	CIRC-VORT	10.2
WSPD-T850*	2.5	KIND	23.4
CIRC-WSPD-T850*	9.4	CIRC-KIND	5.3
CIRC-UWND-VWND-T850*	25.2	KIND-VORT	0.1
		KIND-WSPD	1.2
		CIRC-KIND-VORT	6.1

* Methods involving linear combinations of two or more variables.

For each station in each month the best predictor set was selected from the alternatives listed in Table 3 by identifying the method giving the highest skill in calibration (COR_{CAL}). For predictor sets consisting of a single variable, COR_{CAL} is an unbiased estimator of COR_{IND}, because the correlation between the estimated and observed anomalies is not influenced by errors in the value of the regression coefficient. However, this is not the case for predictor sets based on more than one variable (i.e., those marked with an asterisk in Table 3), hence COR_{CAL} generally exceeds COR_{IND}. This is a well-known result (e.g., Wigley et al. 1990) that occurs be-

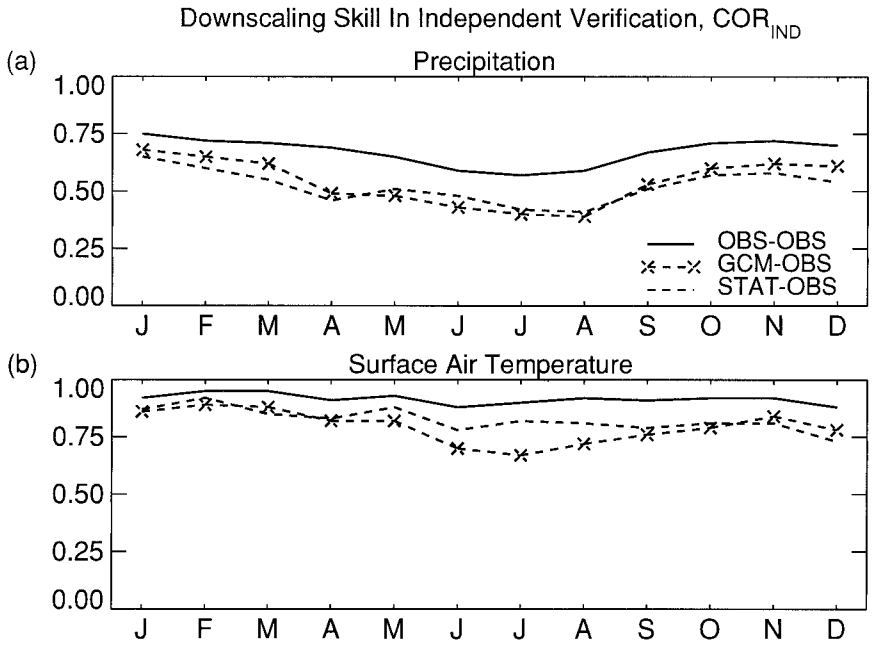

FIG. 14. Values for (a) monthly mean precipitation and (b) monthly mean surface air temperature of COR_{IND}, averaged over all 976 stations. OBS − OBS are estimates based on nearest observing station to predictand station; GCM − OBS are estimates based on nearest global model land point to predictand station; STAT − OBS are estimates based on atmospheric observations.

cause the regression coefficients are influenced by noise (i.e., variability arising from physical processes whose effects are not represented in the equations) due to the limited size of the calibration sample. It is therefore necessary to adjust the values of COR_{CAL} for multiple variable predictor sets in order to make an unbiased comparison with values for single variable predictor sets. This was done by reducing the value of COR^2_{CAL} at each station by an amount equal to the average over all stations of the difference between COR^2_{CAL} and COR^2_{IND} in the relevant month.[2]

The annual average frequency of selection for each predictor set is shown in Table 3. For estimates of temperature T850 is selected on about half of all occasions. A quadrilinear combination of T850 with CIRC, UWND, and VWND accounts for 25% of selections while the remaining 25% of selections is dominated by various combinations of T850 with the wind-related predictors. The frequency of selection of predictor sets excluding T850 is only 0.8%, reflecting the clear margin by which T850 exceeds the skill of other individual predictors (see Table 2). The dominance of T850 as the leading choice is most marked in summer when advection is generally at its weakest: between June and September, T850 is selected on 75% of occasions on average. For estimates of precipitation the selections are distributed more evenly among the predictor sets. An individual predictor is selected on 70% of occasions, with KIND and VORT accounting for almost 50% of all selections. CIRC appears in 31% of the selected predictor sets; however, it is rarely selected on its own, even though its average skill is comparable to that of KIND and VORT. This is because a combination of CIRC with one or more regional predictors can usually be found that explains more variance than CIRC alone, even in areas where CIRC is the leading individual predictor. In such areas it may be concluded that large-scale and regional variability in the atmosphere contribute (to some extent) independently to the observed variability in precipitation.

Figure 14 shows COR_{IND} averaged over all stations for statistical downscaling (STAT − OBS), where the estimate at a given station is obtained from the best predictor set determined as described above. The scores are compared against scores obtained by using the value of the predictand simulated at the nearest GCM land point (GCM − OBS), and the value observed at the nearest station (OBS − OBS). As before the OBS − OBS values are included to estimate the scope for improving skill by refining the downscaling methods in future. For precipitation (Fig. 14a) skill is highest in winter and lowest in summer when the influence of regional processes on precipitation is strongest. During

[2] The use of COR_{IND} in the selection procedure is allowed because our values of COR_{IND} are estimates of the skill in independent verification from the calibration sample, rather than values obtained from a true independent sample. This information would therefore be available a priori in a practical downscaling application involving predictions of future climates.

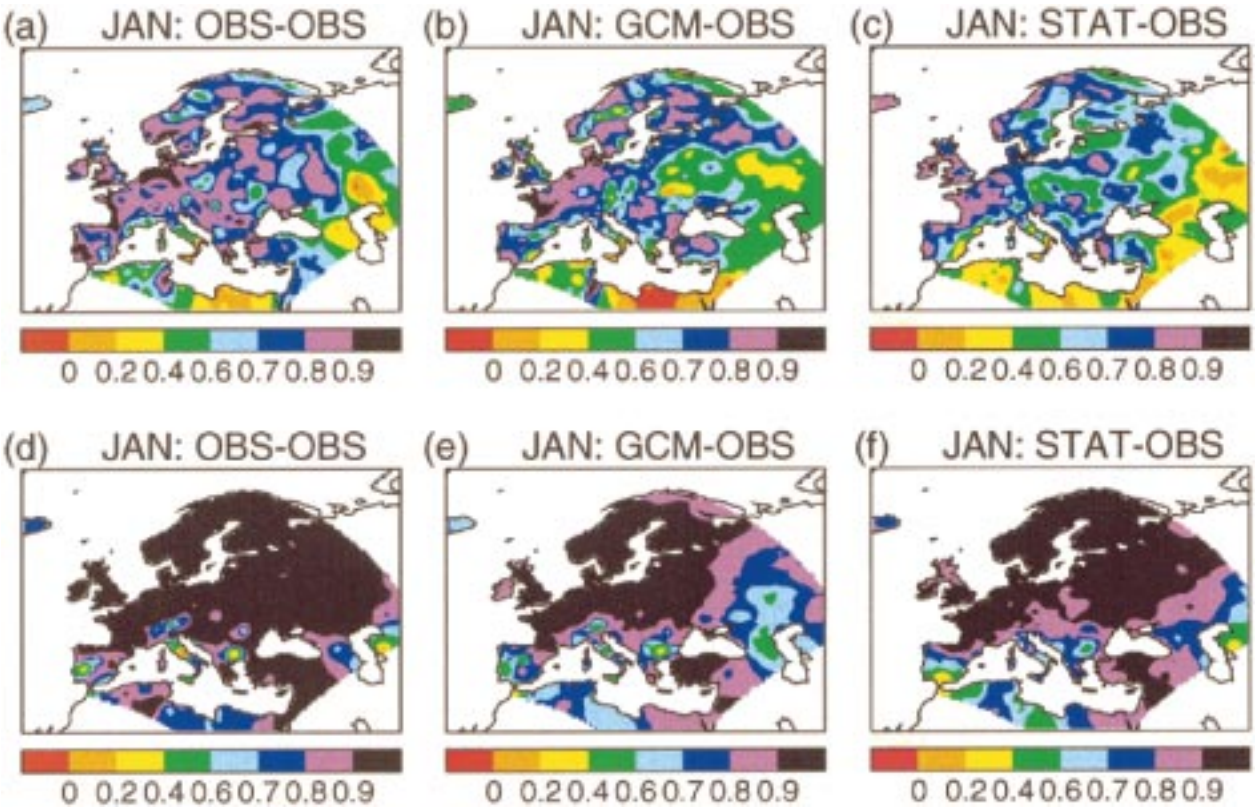

FIG. 15. Spatial distributions of COR_{IND} in Jan for precipitation [(a)–(c)] and surface air temperature [(d)–(f)]. OBS − OBS are estimates based on nearest observing station to predictand station, (a) and (d); GCM − OBS are estimates based on nearest global model land point to predictand station, (b) and (e); STAT − OBS are statistical estimates based on atmospheric observations.

extended winter the average GCM score is consistently a little higher than that for STAT, the difference typically amounting to ~5% of the predictand variance. From May to August STAT is slightly more skillful than the GCM. In January the spatial variations in COR_{IND} are broadly similar: both GCM and STAT explain more than half the variance over much of western, central, and northern Europe (Figs. 15b,c). On the other hand, skill is lower over the drier regions to the south and east, where the monthly precipitation totals are generally accumulated from a small number of individual events (see section 4). In July skill again tends to be higher over western and central Europe than elsewhere (Figs. 16b,c); however, the explained variance exceeds 50% only in a few isolated areas. Skill is low over most of eastern and southern Europe and north Africa, although the statistical estimates perform somewhat better than the GCM over eastern Europe at midlatitudes.

The mean scores for temperature are better than for precipitation in each month (cf. Figs. 14b and 14a). During extended winter STAT and GCM achieve similar levels of average skill; however, from April to September the STAT score consistently exceeds that of the GCM. The largest differences occur between June and August, where STAT explains 65% of the variance on average, compared to 50% for the GCM. In January COR_{IND} exceeds 0.8 in most areas for both methods (Figs. 15e,f). The areas where skill falls below this level in the GCM are usually areas where STAT also shows reduced skill, indicating that the forcing of surface conditions by atmospheric winds and temperatures is weaker than average. In such areas the key to achieving better dynamical downscaling probably lies in improving the simulation of other factors influencing surface temperature, such as soil moisture, snow cover, cloud cover, surface fluxes, etc. In July the superiority of the STAT estimate is widespread. Over most of the eastern half of the domain the inferior performance of the GCM appears to be due to an inability to reproduce the strength of the observed link between surface and lower-tropospheric temperature; that is, the correlations between the GCM surface temperature anomalies and T850 (not shown) are considerably lower than those shown in Fig. 12e.

The mean OBS − OBS score is always greater than that for either GCM − OBS or STAT − OBS (Fig. 14). This supports the evidence from Figs. 8 and 10 that the model scores can be improved by reducing systematic simulation errors, while the statistical estimates could be improved in regions where longer time series are available for calibration (e.g., Corte-Real et al. 1995). For precipitation the large-scale spatial variations in the

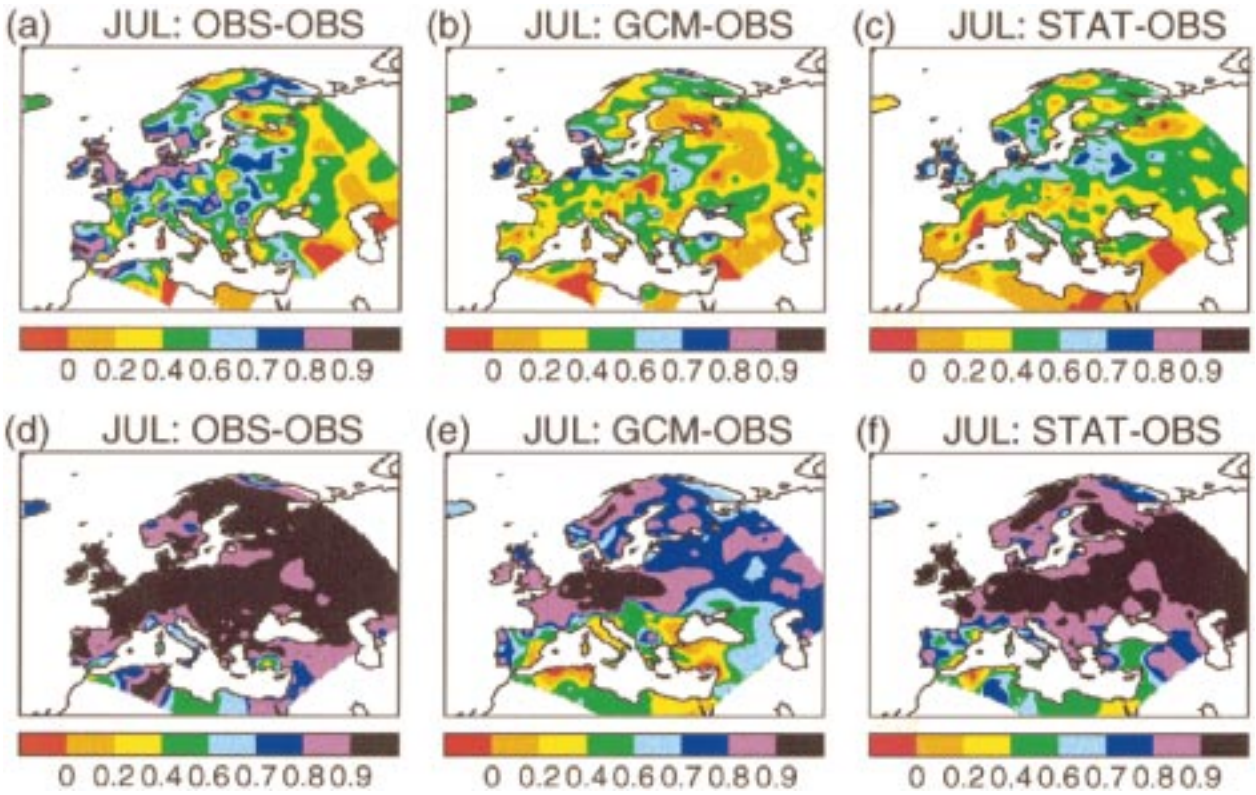

FIG. 16. As in Fig. 15 for Jul.

OBS − OBS score are similar to those for GCM − OBS and STAT − OBS (Figs. 15a–c, 16a–c). In January the GCM and STAT scores may be quite close to the theoretical maximum skill in some areas, notably over much of northwestern Europe, although this is not the case over central parts of the domain. In July the gap between OBS − OBS and the GCM − OBS and STAT − OBS scores is generally larger. The OBS − OBS scores suggest that the maximum theoretical skill over much of the eastern half of the domain is quite small, particularly in summer. Note, however, that the OBS − OBS scores might be somewhat higher in these areas if the observational network was less sparse (Osborn and Hulme 1997).

For temperature the OBS − OBS score exceeds 0.8 over the vast majority of stations (Figs. 15d, 16d) and shows less spatial variation than the scores for the GCM and STAT. The results suggest that there is considerable scope for improving the GCM score in those areas where it is currently low.

COR_{IND} scores for the RCM are not shown in Figs. 14–16. This is because the time series of circulation anomalies in the RCM does not follow the GCM, and hence observations, exactly (section 2c). For example in January and July COR_{PMSL} (RCM − GCM), the average correlation between time series of daily PMSL anomalies in the two models, is 0.91 and 0.70, respectively. It is therefore unfair to compare the RCM − OBS scores directly against those for GCM − OBS and STAT − OBS, since the latter are generated using methodologies that guarantee that the results are not degraded by circulation errors. However, for any given month the mean score that the RCM would achieve with a perfect circulation, say, $COR_{IND}(RCM^{perfect} - OBS)$, can be estimated by assuming that $COR_{IND}(RCM - OBS) \approx COR_{IND}(RCM^{perfect} - OBS) \times COR_{PMSL}(RCM - GCM)$, where $COR_{IND}(RCM - OBS)$ is the mean score achieved in practice. Values of $COR_{IND}(RCM^{perfect} - OBS)$ deduced from this formula are found to be very close to those for the GCM shown in Fig. 14. This is not too surprising, given that the phase variations in typical monthly anomaly patterns tend to be determined by large-scale features rather than by mesoscale signals, especially for temperature. Thus ERRM and ERRV are more effective than COR_{IND} in distinguishing between the performance of the RCM and the GCM.

In summary, both dynamical and statistical downscaling generally achieve a high level of skill in reproducing observed interannual variations of surface temperature. For precipitation, skill is generally lower, especially in summer; however, both downscaling methods explain 20% or more of the predictand variance at most locations in winter, and at about 50% of locations in summer. During extended winter dynamical downscaling performs slightly better than statistical downscaling for precipitation and at a similar level for tem-

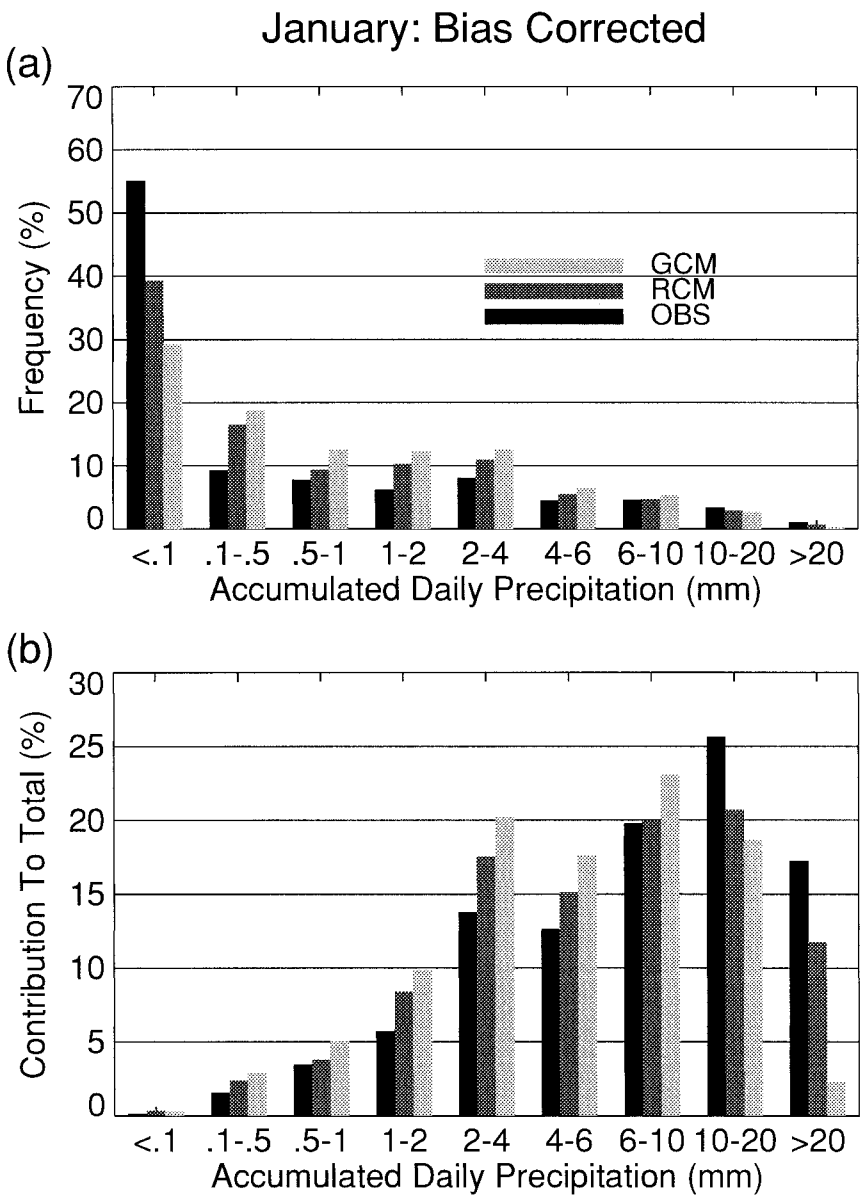

FIG. 17. (a) Frequency of daily accumulations of precipitation in various ranges, averaged over 976 observing stations and 11 Januarys, for OBS and values simulated at the nearest RCM and GCM land point to each station. Simulated precipitation values have been scaled to remove the space- and time-averaged bias relative to OBS. (b) Contribution of each bin to total precitation.

perature. During extended summer statistical downscaling performs significantly better for temperature and marginally better for precipitation. Overall, the dynamical and statistical methods show comparable skill, in agreement with the results of Kidson and Thompson (1998). The skill is usually below the level obtained by using observations from the nearest neighbor station, suggesting that there is scope to increase skill by improving the model physics or by using longer calibration time series (where available) in the case of the statistical estimates.

4. Downscaling daily data

This section describes estimates of daily distributions of station temperature and precipitation made using the values simulated at the nearest land point by the RCM and the GCM. For the sake of brevity only January and July are considered.

a. Precipitation

Figures 17 and 18 show histograms of simulated and observed daily precipitation amounts for January and

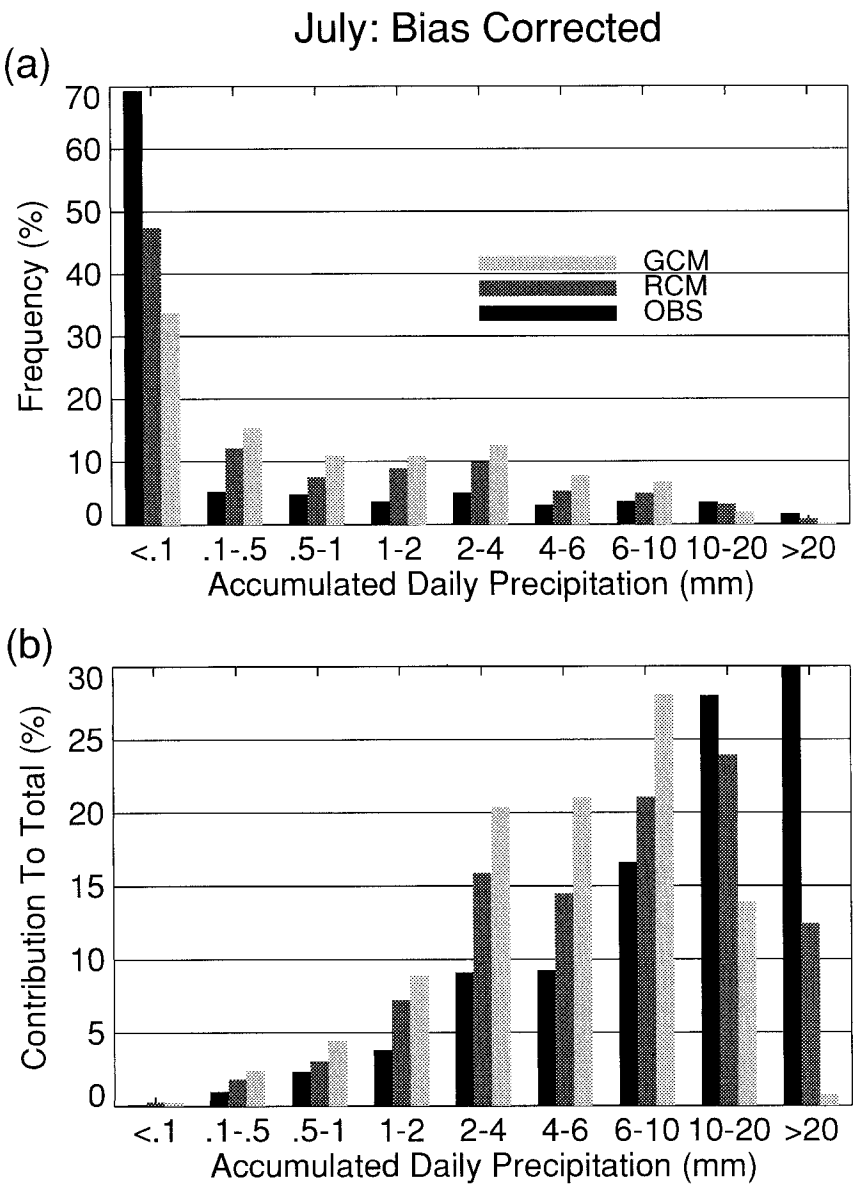

FIG. 18. As in Fig. 17 for Jul.

July, based on frequencies averaged over all 976 stations. Both models significantly underestimate the frequency of dry days; however, the errors are much larger in the GCM. The models also underestimate the frequency of heavy precipitation events. For example, events in excess of 20 mm contribute 17% of total precipitation in January (Fig. 17b), and 30% in July (Fig. 18b). The contribution from such events is much smaller in the RCM, especially in July, and is negligible in the GCM. The errors identified above are found in almost all parts of the domain. In July, for example, the wet day probability p_w is much too large everywhere apart from North Africa (Figs. 19a–c), whereas the probability of accumulations in excess of 10 mm is too small (Figs. 19d–f), apart from in a few places where time-averaged precipitation is much too high (cf. Figs. 3d–f). Results of a similar nature are found in January. Previous work (Mearns et al. 1995; Gregory and Mitchell 1995) suggests that the model errors reflect the effect of spatial averaging: the observations are from point locations, whereas the grid-box variables represent areas of 2.5×10^3 km^2 (RCM) and $\sim 8 \times 10^4$ km^2 (GCM), and are thus equivalent to means over a number of stations. This is confirmed by Osborn and Hulme (1997), who show that the dry day probability and standard deviation of observed daily precipitation are both strongly reduced when station distributions are aggregated to form areal means representative of a model grid box.

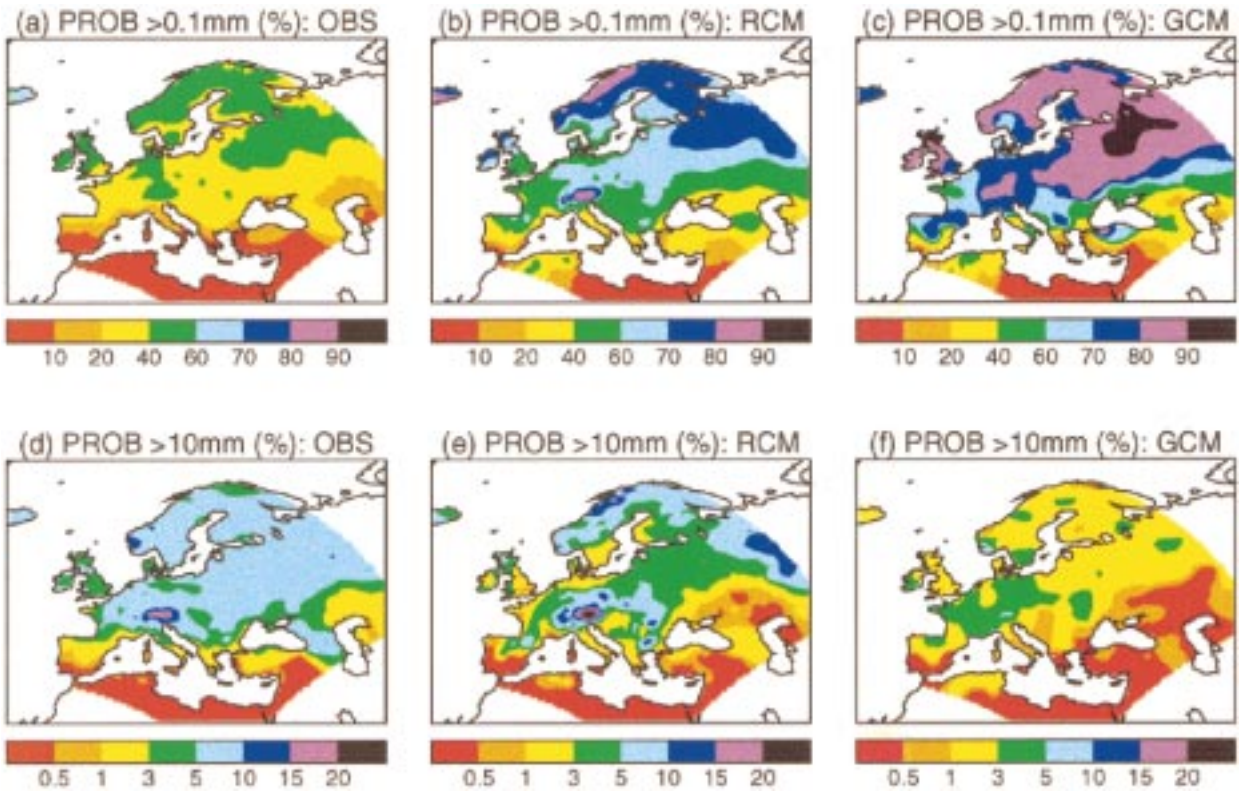

FIG. 19. Jul distributions of wet day probability (threshold 0.1 mm) [(a)–(c)] and probability of a daily precipitation total exceeding 10 mm [(d)–(f)]. OBS are station observations, (a) and (d); RCM are simulated values at the nearest regional model land point to each station, (b) and (e); GCM are simulated values at the nearest global model land point to each station, (c) and (f). Simulated precipitation values have been scaled to remove the space- and time-averaged bias relative to OBS.

These results do not preclude the use GCM or RCM gridpoint values in impact studies requiring knowledge of local precipitation statistics; however, they do imply the need to make significant empirical adjustments to the simulated distributions. These will be sensitive to the scale of individual precipitation events (Osborn 1997), so adjustments calculated for present climate could become invalid if, for example, the balance between convective and large-scale precipitation alters in future. However, the fact that the adjustments are smaller for the higher-resolution model encourages the hope that further increases in resolution could eventually remove the need for such adjustments and thus circumvent this problem.

b. Surface air temperature

In January both models overestimate the mean frequency of large daily temperature anomalies (Fig. 20a); however, the errors are larger in the GCM. In the observations the probability of large daily anomalies increases with latitude (Fig. 21a), as for the monthly means discussed earlier (Fig. 7a). The models capture the general pattern correctly (Figs. 21b,c), but overestimate the variability in most regions. The inferior performance of the GCM relative to the RCM arises mainly from northern, central, and eastern parts of the domain. A number of factors appear to contribute to this, including (a) an enhanced frequency of low surface wind speeds in the GCM, which increases the probability of anomalously low nighttime minimum temperatures, especially over northern Europe; (b) lower time-averaged temperatures than in the RCM (see section 3a), reducing evaporative damping of temperature anomalies; and (c) slightly greater variability in cloud cover, which exerts a strong influence on the surface radiative fluxes. Another factor increasing variability in the GCM is poor resolution of the modifying influence of coastal land on maritime airstreams; note the lack in Fig. 21c of the fringes of reduced variability near coastlines apparent in Fig. 21a and, to a lesser extent, Fig. 21b.

In July both models again overestimate the average level of observed variability (Fig. 20b), the magnitudes of the errors in the GCM and the RCM being similar. The distributions of large anomaly frequencies (where ''large'' is defined as $\pm 7^{\circ}$C; cf. $\pm 10^{\circ}$C for January) show that both models possess too much variability in areas to the north of the Black and Caspian Seas (Figs. 21d–f). In the RCM the area of excessive variability also extends westward into central Europe. These errors in daily variability correspond quite closely to those in the variance of monthly means (Figs. 7e,f) and are prob-

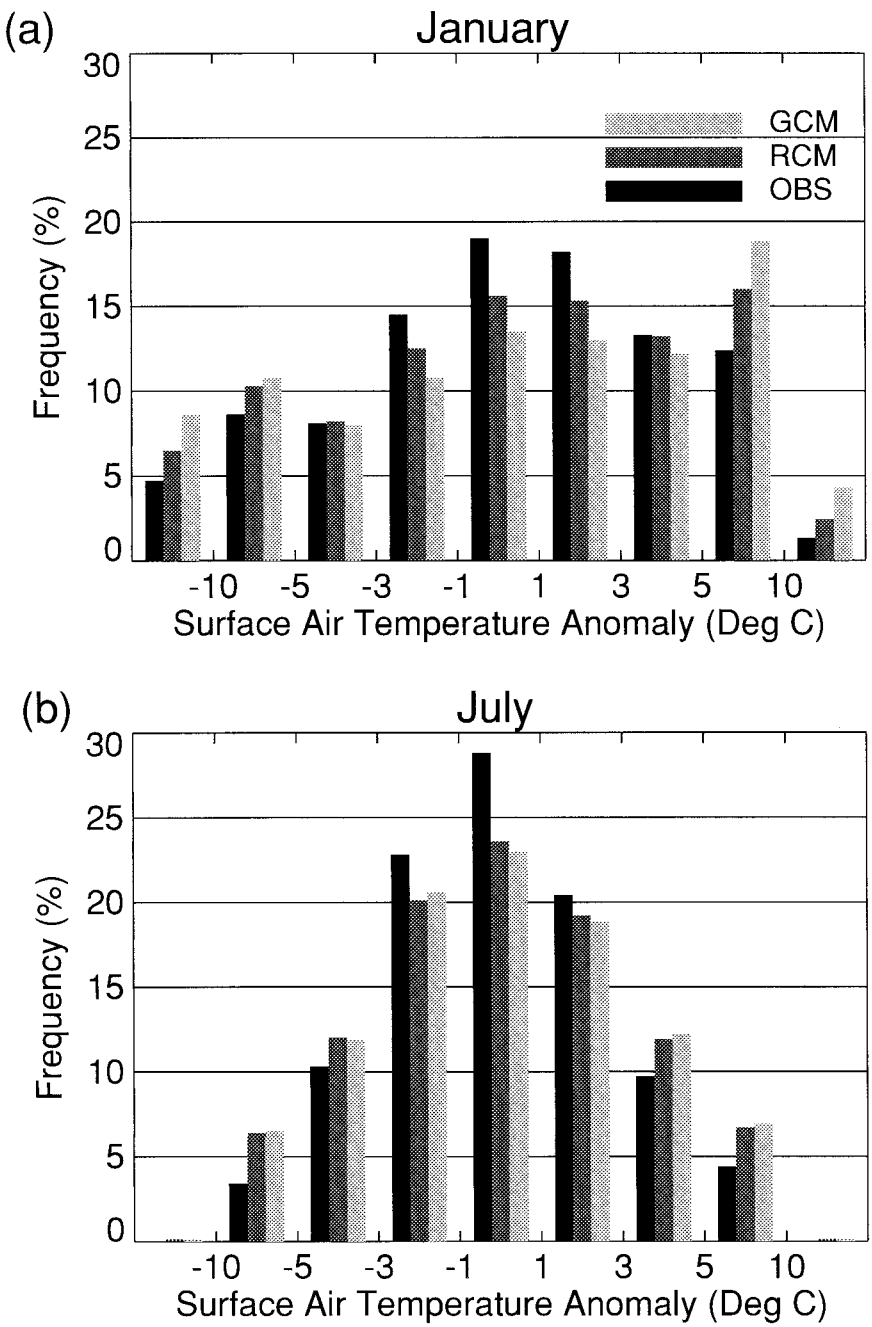

FIG. 20. Frequency of daily surface air temperature anomalies in (a) Jan and (b) Jul, averaged over 976 stations and 11 years for OBS and values simulated at the nearest RCM and GCM land point to each station.

ably explained by the same mechanism, that is, insufficient evaporative damping of temperature anomalies in areas where the soil has become excessively dry.

5. Summary and concluding remarks

Statistical downscaling estimates of local surface air temperature and precipitation derived from atmospheric observations have been compared with dynamical downscaling estimates involving the use of temperature or precipitation simulated by two climate models at the nearest grid point to the target location. The models considered are a global general circulation model (GCM) of horizontal resolution ~300 km and a high-resolution regional model (RCM) of resolution 50 km nested inside the GCM. Downscaling estimates for 976

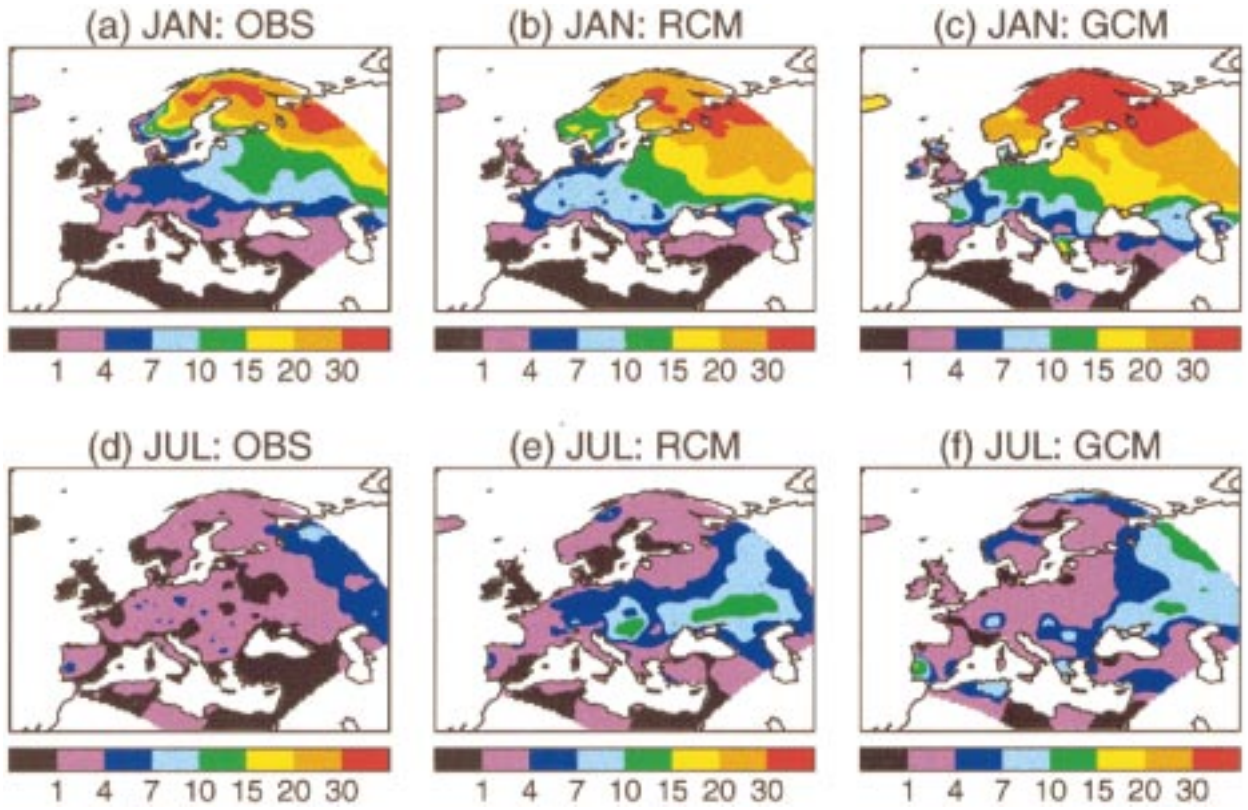

FIG. 21. Distributions of the probability (%) of large daily surface air temperature anomalies in Jan [(a)–(c)] and Jul [(d)–(f)]. A ''large'' anomaly is defined as one exceeding ±10°C in Jan, and ±7°C in Jul. OBS are station observations; RCM are values simulated at nearest regional model land point to each station; GCM are values simulated at nearest global model land point to each station.

European stations from June 1983 to February 1994 are assessed. The statistical method involves choosing optimum predictor sets for each station from a range of candidates comprising eigenvectors of the mean sea level pressure field plus regional values of various quantities, including lower-tropospheric temperature and moisture, vertical stability, and properties of the near-surface wind field. In order to make a fair comparison of the statistical and dynamical approaches, the GCM estimates are derived from an integration constrained to reproduce the observed circulation over the relevant period. The nested RCM simulation cannot be constrained in the same way; however, the domain used is small enough to ensure that differences from the GCM (and hence observed) circulation are quite small except in summer, when the influence of the lateral boundary forcing is less dominant than in other seasons.

Overall, the dynamical and statistical methods perform with similar skill in downscaling observed monthly mean anomalies (see also Kidson and Thompson 1998). For the GCM the average anomaly correlation with observations in a given month varies from 0.67 to 0.89 for temperature and from 0.39 to 0.68 for precipitation, with the highest values occurring in winter. Similar values are found for the RCM once the effect of circulation errors is accounted for. For precipitation the average scores for the statistical method are slightly lower in winter and slightly higher in summer; for temperature they are similar in winter and higher in summer (minimum average value 0.78 cf. 0.67). In the latter case the inferior performance of the models is explained by a failure to reproduce the strength of the observed relationship between surface and lower tropospheric temperature over much of the continental interior. For precipitation, skill is generally highest over western and northern Europe but is low over eastern and southern Europe and North Africa, particularly in summer when precipitation arises mainly from infrequent localised convective events. For temperature both dynamical and statistical downscaling explain well over half of the predictand variance at most locations in the northern half of the domain; the lowest skill occurs around the Mediterranean fringes, particularly for dynamical downscaling in summer.

Downscaling estimates of monthly means from the two models are compared by quantifying the magnitude of corrections required to reproduce the mean and variance of the predictand time series. Small corrections suggest that the simulations successfully account for the physics controlling the observed climatological distributions, whereas large corrections imply missing or inadequately represented processes, in which case the

model predictions cannot be assumed to be reliable under conditions of climate change.

For precipitation the RCM shows improvements in skill relative to the GCM for both the climatological mean and interannual variability. These improvements, which occur in the winter half of the year for the climatological mean and in summer for interannual variability, stem directly from the impact of enhanced horizontal resolution. First, the RCM is able to capture some of the observed mesoscale structure in climatological mean spatial distributions, which cannot be resolved by the GCM. Second, although both models underestimate the variance of observed distributions due to the spatial smoothing implicit in the use of grid-box variables, the errors are smaller in the RCM, especially in summer when precipitation is mainly convective in origin.

For climatological mean temperature the RCM simulates a realistic orographic signal on scales too fine to be resolved by the GCM. The lack of this signal does not impair the skill of the GCM predictions, however, since simple empirical corrections are effective in removing biases arising from differences in elevation between model grid points and observing stations. Nevertheless, throughout the year the RCM estimates of the climatological mean are more skillful than those of the GCM at most locations. This arises partly from better resolution of local physiographical features such as coastlines, islands, and lakes, and partly from the influence of higher resolution on the large-scale component of the simulated distributions, which occurs via changes in the dynamics and the hydrological cycle relative to the GCM. Both models tend to overestimate the interannual variability of temperature. The RCM performs better than the GCM in the winter half-year, especially in coastal regions. The GCM is better in summer due to excessive drying of the soil in the RCM, although the superior performance of the GCM arises from a compensation of errors (the effect of underestimating the mean intensity of rainfall events offsets that of excessive solar heating).

The ability of the models to predict daily distributions of precipitation and temperature is also assessed. Errors in daily variability generally show similar patterns to errors in the variability of monthly means. Both models underestimate the frequency of dry days and intense precipitation events due to the use of grid-box variables representing areas of 2.5×10^3 km^2 (RCM) and $\sim 8 \times 10^4$ km^2 (GCM) to estimate distributions at point locations. However, the errors are much smaller in the RCM and could be reduced in the future by refining its resolution, thus providing the prospect of reliable predictions of greenhouse-gas-driven changes in the daily distributions. For temperature, the models overestimate the frequency of large anomalies. In winter the errors are smaller in the RCM, due to changes in the dynamics associated with finer resolution and also to better resolution of local physiography, especially coastlines; in summer the largest overestimates occur in areas where the soil is too dry.

In summary, the RCM offers potential benefits relative to the GCM in downscaling local climate. In order to realize these benefits, systematic errors in variables influencing surface temperature, precipitation, and other elements of surface climate must be reduced. In the present models, for example, excessive moisture contents in the atmosphere contribute to overestimates of domain-averaged precipitation, which were corrected empirically prior to the assessment of local downscaling errors discussed above. In order to optimize the physics for both the RCM and the GCM it may be necessary to introduce explicit scale dependences in some areas (Noguer et al. 1998). In the case of the RCM there is also the prospect of improvements in skill from future increases in resolution. For precipitation, in particular, finer resolution should allow more accurate reproduction of the mesoscale signal in climatological mean fields and better simulations of both daily and interannual variability.

The potential for improving both the statistical and dynamical downscaling methods is assessed by comparing the scores for estimates of monthly means against those obtained by using observations from the nearest station to the predictand station. The spatially averaged downscaling scores are always lower than those based on nearest-neighbor observations. In the case of dynamical downscaling, closing the gap in skill will require action to reduce systematic simulation errors (see above), while the statistical methods could be improved by using longer calibration time series or a wider range of predictors, assuming that the necessary observed data are available. Note, however, that achieving skillful downscaling under present climate conditions does not guarantee that the methods will perform equally well when used to predict future conditions. For example, statistical relationships between the predictand and the predictors are not necessarily invariant in time (Wilby 1997), while a demonstration that a model can successfully capture natural interannual variations does not guarantee that it will successfully simulate the feedbacks that occur in response to (say) changes in greenhouse forcing.

Although this study has presented statistical and dynamical downscaling as distinct alternatives, in practice dynamical downscaling requires a statistical element because even the horizontal resolution of an RCM is not fine enough to capture local forcings on scales of a few kilometers. In the present study statistical processing of the simulated gridpoint values took the form of simple empirical adjustments; however, it is likely that more sophisticated diagnostic techniques based on physical principles (e.g., Sinclair 1994) will provide a better method of accounting for unresolved local effects.

Acknowledgments. This work would not have been possible without the efforts of Drs. Maria Noguer and

Richard Jones, who designed, set up, and ran the model integrations used in this study. The reviews of Dr J. Kidson and an anonymous referee led to considerable improvements to the original manuscript. I also wish to thank the Climate Prediction Center, Washington, D.C., and the National Center for Atmospheric Research, Boulder, Colorado, for making available the observed station data. This work was supported by the Department of the Environment, Transport, and the Regions under Contract PECD/7/12/37.

APPENDIX A

Recent Improvements Included in the GCM and the RCM

This appendix lists improvements incorporated in the GCM and the RCM since the previous study of R. G. Jones et al. (1995, 1997).

1) The parameterization of gravity wave drag has been modified to include the effects of anisotropic orography, trapped lee waves, and and a better representation of high drag states in which flow blocking and hydraulic jumps occur. Also, roughness lengths for momentum have been changed to include the effect of orographic form drag (see Milton and Wilson 1996).
2) The temperature range for partitioning mixed phase clouds into ice and water has been changed from 0° to −15° to 0° to −9°C, based on observational evidence (Gregory and Morris 1996).
3) The effective radius for cloud water droplets (r_e) now varies with cloud water content instead of taking a fixed value of 7 μm. The parameterization is as described by A. Jones et al. (1995) except that the number concentration of activated droplets is prescribed as 6×10^8 m^{-3} over land and 1.5×10^8 m^{-3} over the sea. For deep convective clouds (>500 m thick) r_e is now prescribed as 9.5 μm over land points and 13 μm over sea points.
4) The calculation of surface soil moisture is now based on a four-layer soil hydrology scheme instead of a single moist layer. This allows the models to represent recharge of the root zone from below during dry periods.
5) Horizontal diffusion has been reduced in the GCM; this improves the simulated cloud and precipitation fields and reduces the moist bias in the extratropical troposphere (Hall and Stratton 1994). Second, diffusion is now switched off (in both the GCM and RCM) where the model coordinate surfaces are steep, in order to prevent excessive precipitation over mountains.
6) Negative humidities caused by finite differencing errors in regions of sharp gradient are now corrected by taking moisture from neighboring points only, rather than from all points in the relevant layer. Also, the vertical advection of moisture is now calculated using a weighted finite difference scheme instead of the simple centered scheme used previously. These changes improve the simulation of stratospheric moisture.
7) The calculation of large-scale cloud and precipitation has been altered to allow precipitating ice to fall through more than one model layer in a timestep, dependent on fall speed. Also, the timestep for this calculation in the GCM has been shortened from 30 to 5 min to be consistent with the corresponding calculation in the RCM. These changes reduce an implicit timestep dependence that previously caused differences in cloud cover between the RCM and GCM (R. G. Jones et al. 1995).

APPENDIX B

Quality Control of Station Observations

a. Precipitation

Two statistical tests were applied to determine whether the pattern of missing precipitation reports was such as to bias the time-average precipitation and the dry day probability. This was done by using estimated precipitation (EP) values. The EP was set equal to reported precipitation (RP) when a reported value was available; otherwise it was derived by assigning precipitation amounts to past and present weather codes with adjustments dependent on latitude, season, and geographical location (Thomas and Patterson 1983). For each season the time-averaged values ⟨RP⟩ and ⟨EP⟩ were compared using a *t*-test, forming the required sample variances from all the constituent daily values (assumed to be statistically independent from one another). If the means were different at the 1% level in at least three of the four seasons ⟨RP⟩ was deemed to be significantly different from ⟨EP⟩. However, this alone was not sufficient to justify rejecting a station since ⟨EP⟩ may itself be biased by the use of imperfect estimation algorithms. A second test was therefore applied to the dry day probability, on the assumption that this was less likely to be biased by the use of estimated values. A chi-square test with 2 degrees of freedom was performed on the 2×2 contingency table whose categories were EP present/absent and RP present/absent. Expected frequencies were based on the null hypothesis that the dry day probability does not depend whether or not a reported precipitation value was available. The chi-square test was failed if the null hypothesis was rejected at the 0.001% level. The station was rejected if it failed both the *t*-test and the chi-square test. Fifty-four stations were rejected by applying this criterion.

b. Temperature

Following May et al. (1992) a test for unreliable data was performed on diurnally averaged temperature, taken

here as the average of the daily maximum and minimum values. For each day of the year the interannual variance was calculated. If the variance was based on fewer than four observations it was interpolated from the values on adjacent dates. A smoothed annual cycle of variance was then created using Fourier analysis. Six harmonics were retained, with harmonics 4–6 given weights of $\frac{5}{6}$, $\frac{1}{2}$, and $\frac{1}{6}$ the weight of harmonics 1–3. On a particular day of the year an outlier will considerably increase the unsmoothed variance but will have much less effect on the smoothed value. Outliers were detected using a chi-square test with $N - 1$ degrees of freedom, where N is the number of observations from which the unsmoothed variance was calculated. A station was rejected as unreliable if it failed the test at the 0.01% level on two or more days of the year. Otherwise, the station was passed; however, if the test was failed at the 0.0001% level on one day of the year a subjective search was made for a rogue observation, which was set to missing data if found. In the event, 15 stations were rejected, and 13 erroneous observations were identified and removed in passed stations.

REFERENCES

Corte-Real, J., X. Zhang, and X. Wang, 1995: Downscaling GCM information to regional scales: A non-parametric multivariate regression approach. *Climate Dyn.,* **11,** 413–424.

Cullen, M. J. P., 1993: The unified forecast/climate model. *Meteor. Mag.,* **122,** 81–94.

Dolman, A. J., and D. Gregory, 1992: The paramaterization of rainfall interception in GCMs. *Quart J. Roy. Meteor. Soc.,* **118,** 455–467.

Gibson, R. K., P. Kållberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Reanalysis Project Rep. Series No. 1, 8 pp.

Giorgi, F., and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: A review. *Rev. Geophys.,* **29,** 191–216.

——, and M. R. Marinucci, 1996: Improvements in the simulation of surface climatology over the European region with a nested modelling system. *Geophys. Res. Lett.,* **23,** 273–276.

——, G. T. Bates, and S. Nieman, 1993a: The multiyear surface climatology of a regional atmospheric model over the western United States. *J. Climate,* **6,** 75–95.

——, M. R. Marinucci, G. T. Bates, and G. de Canio, 1993b: Development of a second generation regional climate model (RegCM2). Part II. Convective processes and assimilation of lateral boundary conditions. *Mon. Wea. Rev.,* **121,** 2814–2832.

——, C. Brodeur, and G. T. Bates, 1994: Regional climate change scenarios over the United States produced with a nested regional climate model. *J. Climate,* **7,** 375–399.

Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. *Mon. Wea. Rev.,* **118,** 1483–1506.

——, and S. Allen, 1991: The effect of convective scale downdrafts upon NWP and climate simulations. Preprints, *Ninth Conf. on Numerical Weather Prediction,* Denver, CO, Amer. Meteor. Soc., 122–123.

——, and D. Morris, 1996: The sensitivity of climate simulations to the specification of mixed phase clouds. *Climate Dyn.,* **12,** 641–651.

Gregory, J. M., and J. F. B. Mitchell, 1995: Simulation of daily variability of surface temperature and precipitation over Europe in the current and $2\times CO_2$ climates using the UKMO climate model. *Quart. J. Roy. Meteor. Soc.,* **121,** 1451–1476.

——, T. M. L. Wigley, and P. D. Jones, 1993: Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. *Climate Dyn.,* **8,** 299–310.

Hall, C. D., and R. A. Stratton, 1994: The effect of changing horizontal diffusion in the atmospheric version of the Unified Climate Model. Climate Research Tech. Note CRTN 44, Bracknell, United Kingdom, 11 pp. [Available from National Meteorological Library, Meteorological Office, Bracknell, United Kingdom.]

Hewitson, B., and R. G. Crane, 1992: Regional climates in the GISS global circulation model: Synoptic-scale circulation. *J. Climate,* **5,** 1002–1011.

Hirakuchi, H., and F. Giorgi, 1995: Multiyear present-day and $2\times CO_2$ simulations of monsoon climate over eastern Asia and Japan with a regional climate model nested in a general circulation model. *J. Geophys. Res.,* **100,** 21 105–21 125.

Hulme, M., D. Conway, P. D. Jones, T. Jiang, E. M. Barrow, and C. Turney, 1995: Construction of a 1961–90 European climatology for climate change impacts and modelling applications. *Int. J. Climatol.,* **5,** 1333–1363.

Johns, T. C., C. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean–atmosphere GCM: Model description, spinup and validation. *Climate, Dyn.,* **13,** 103–134.

Jones, A., D. L. Roberts, and A. Slingo, 1995: A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols. *Nature,* **370,** 450–453.

Jones, R. G., J. M. Murphy, and M. Noguer, 1995: Simulation of climate change over Europe using a nested regional climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Quart. J. Roy. Meteor. Soc.,* **121,** 1413–1449.

——, ——, ——, and A. B. Keen, 1997: Simulation of climate change over Europe using a nested regional climate model. II: Comparison of driving and regional model responses to a doubling of carbon dioxide. *Quart. J. Roy. Meteor. Soc.,* **123,** 265–292.

Karl, T. R., W.-C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, 1990: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *J. Climate,* **3,** 1053–1079.

Kidson, J. W., and C. S. Thompson, 1998: A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate,* **11,** 735–753.

Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected global precipitation. *Int. J. Climatol.,* **10,** 111–127.

Lorenc, A. C., R. S. Bell, and B. Macpherson, 1991: The Meteorological Office 'analysis correction' data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **117,** 59–89.

Manabe, S., and R. T. Wetherald, 1975: The effects of doubling the CO_2 concentration on the climate of a general circulation model. *J. Atmos. Sci.,* **32,** 3–15.

May, W., D. J. Shea, and R. A. Madden, 1992: The annual variation of surface temperatures over the world. NCAR Tech. Note TN-372+STR, 135 pp. and 17 microfiche.

McNally, A. P., and M. Vesperini, 1996: Variational analysis of humidity information from TOVS radiances. *Quart. J. Roy. Meteor. Soc.,* **122,** 1521–1544.

Mearns, L. O., F. Giorgi, L. McDaniel, and C. Shields, 1995: Analysis of daily variability of precipitation in a nested regional climate model: Comparison with observations and doubled CO_2 results. *Global Planet. Change,* **10,** 55–78.

——, C. Rosnezweig, and R. Goldberg, 1997: Mean and variance change in climate scenarios: Methods, agricultural applications, and measures of uncertainty. *Climate Change,* **35,** 367–396.

Milton, S. F., and C. Wilson, 1996: The impact of parametrized subgrid-scale orographic forcing on systematic errors in a global NWP model. *Mon. Wea. Rev.,* **124,** 1746–1766.

Miskus, D., J. D. Laver, and F. G. Finger, 1988: The Climate Analysis Center's Interactive Climate Assessment Data Base. Preprints,

Fourth Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Anaheim, CA, Amer. Meteor. Soc., 217–223.

Noguer, M., R. G. Jones, and J. M. Murphy, 1998: Effect of systematic errors in the lateral boundary forcing on the climatology of a nested regional climate model over Europe. *Climate Dyn.,* **14,** 691–712.

Osborn, T. J., 1997: Areal and point precipitation changes: Implications for the application of climate models. *Geophys. Res. Lett.,* **24,** 2829–2932.

——, and M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Climate,* **10,** 1885–1908.

Parker, D. E., C. K. Folland, M. N. Ward, A. Bevan, M. Jackson, and K. Maskell, 1995: Marine surface data for analysis of climatic fluctuations on interannual to century time scales. *Natural Climate Variability on Decade-to-Century Time Scales*, D. G. Martison et al., Eds., National Research Council, 222–228 and 241–250.

Peppler, R. A., and P. J. Lamb, 1989: Tropospheric static stability and central North American growing season rainfall. *Mon. Wea. Rev.,* **117,** 1156–1180.

Robinson, P. J., and P. L. Finkelstein, 1991: The development of impact-oriented climate scenarios. *Bull. Amer. Meteor. Soc.,* **72,** 481–490.

Sinclair, M. R., 1994: A diagnostic model for estimating orographic precipitation. *J. Appl. Meteor.,* **33,** 1163–1174.

Slingo, A., 1989: A GCM parametrization for the shortwave radiative properties of water clouds. *J. Atmos. Sci.,* **46,** 1419–1427.

Smith, R. N. B., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Quart. J. Roy. Meteor. Soc.,* **116,** 435–460.

Stephens, G. L., 1978: Radiation profiles in extended water clouds. II: Parametrization schemes. *J. Atmos. Sci.,* **35,** 2123–2132.

Thomas, A. R., and V. L. Patterson, 1983: A reliable method for estimating precipitation amounts. Preprints, *Fifth Symp. on Meteorological Observations and Instrumentation,* Toronto, ON, Canada, Amer. Meteor. Soc., 554–561.

Warrilow, D. A., A. B. Sangster, and A. Slingo, 1986: Modelling of land surface processes and their influence on European climate. Met O 20 DCTN 38, Bracknell, United Kingdom, 90 pp. [Available from National Meteorological Library, Meteorological Office, Bracknell, United Kingdom.]

Wigley, T. M. L., P. D. Jones, K. R. Briffa, and G. Smith, 1990: Obtaining sub-grid-scale information from coarse-resolution general circulation model output. *J. Geophys. Res.,* **95,** 1943–1953.

Wilby, R. L., 1997: Non-stationarity in daily precipitation series: Implications for GCM down-scaling using atmospheric circulation indices. *Int. J. Climatol.,* **17,** 439–454.

——, and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.,* **21,** 530–548.

——, ——, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.,* **34,** 2995–3008.

Wilson, M. F., and A. Henderson-Sellers, 1985: A global archive of land cover and soils data for use in general circulation models. *J. Climatol.,* **5,** 119–143.

Zorita, E., V. Kharin, and H. von Storch, 1992: The atmospheric circulation and sea surface temperature in the North Atlantic area in winter: Their interaction and relevance for Iberian precipitation. *Climate Dyn.,* **5,** 1097–1108.