Skill, Correction, and Downscaling of GCM-Simulated Precipitation

JONATHAN M. EDEN AND MARTIN WIDMANN

School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom

DAVID GRAWE

Meteorological Institute, KlimaCampus, University of Hamburg, Hamburg, Germany

SEBASTIAN RAST

Max Planck Institute for Meteorology, Hamburg, Germany

ABSTRACT

The ability of general circulation models (GCMs) to correctly simulate precipitation is usually assessed by comparing simulated mean precipitation with observed climatologies. However, to what extent the skill in simulating average precipitation indicates how well the models represent temporal changes is unclear. A direct assessment of the latter is hampered by the fact that freely evolving climate simulations for past periods are not set up to reproduce the specific evolution of internal atmospheric variability. Therefore, model-to-real-world comparisons of time series of daily, monthly, or annual precipitation are not meaningful. Here, for the first time, the authors quantify GCM skill in simulating precipitation variability using simulations in which the temporal evolution of the large-scale atmospheric state closely matches that of the real world. This is achieved by nudging the atmospheric states in the ECHAM5 GCM, but crucially not the precipitation field itself, toward the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40). Global correlation maps between observed and simulated seasonal precipitation allow areas in which simulated future precipitation changes are likely to be meaningful to be identified. In many areas, correlations higher than 0.8 are found.

This means also that in these regions the simulated precipitation is a very good predictor for the true precipitation, and thus a statistical correction of the simulated precipitation, which can include a downscaling component, can provide useful estimates for local-scale precipitation. The authors show that a simple scaling of the simulated precipitation performs well in a cross validation and thus appears to be a promising alternative to standard statistical downscaling approaches.

1. Introduction

Increased atmospheric concentrations of greenhouse gases are associated not only with rising global temperatures but also are expected to lead to considerable changes in global precipitation patterns (Trenberth et al. 2007). Precipitation is characterized by large spatial and temporal variability, and estimating future precipitation is a major challenge in climate modeling (Trenberth et al. 2007; Randall et al. 2007). To simulate precipitation, atmospheric general circulation models (GCMs) need to represent condensation, evaporation, and other cloud-microphysical processes, which in turn depend on vertical air velocities related to large-scale uplift or convection. The latter are influenced by the large-scale meteorological situation and also by orographic effects. Many of these processes operate at spatial scales smaller than a typical GCM gridcell and are included in GCMs through approximations known as parameterizations. Although large-scale mean precipitation patterns are reasonably well captured by GCMs, regional details are often poorly represented, and not much is known about how well temporal variability is simulated.

When considering simulated precipitation over a historical period, it is possible to make direct comparisons

Corresponding author address: Jonathan M. Eden, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom.
E-mail: j.m.eden@bham.ac.uk

with real-world observations. Groisman et al. (2005) have shown that the multimodel mean global precipitation trend patterns during the twentieth century do not agree very well with observed trends but found better agreement for changes in extreme precipitation. It should be noted, however, that the differences might be partly related to random atmospheric variability rather than to problems in the models. When making inferences about the simulation of future climates, for which observational data are obviously not available, model consensus is an accepted indicator of skill. Multimodel mean trends for the twenty-first century were calculated in the most recent Intergovernmental Panel on Climate Change (IPCC) Assessment Report, accompanied by an analysis of the areas over which the models agree with each other with respect to the sign of the change (Randall et al. 2007). Again, it is partly unclear which of these differences are due to genuine differences in the model's response to greenhouse gas forcing and which are due to random, unpredictable differences in atmospheric variability. Assessing the anthropogenic influence on global precipitation is made troublesome by the averaging of varying regional responses, which may weaken the overall global signal (Allen and Ingram 2002; Held and Soden 2006), and successful quantification of anthropogenic signals must usually consider changes in particular regions or latitudinal bands (e.g., Zhang et al. 2007). Moreover, it is problematic to associate areas of high model agreement with a high confidence in predictions because simulations in such areas may still be wrong for a common reason.

Systematic biases and uncertainty about the skill of the simulated temporal variability means that simulated precipitation is often disregarded and estimates of small-scale precipitation changes are instead based on either high-resolution regional climate models (RCMs) driven by GCMs or on statistical downscaling approaches (Wilby et al. 2004; Christensen et al. 2007; Fowler et al. 2007; Maraun et al. 2010). The latter usually derive statistical relationships between precipitation observations and observed large-scale predictors (e.g., atmospheric circulation, temperature, and humidity) and then apply these relationships to simulated predictors for the future. Because this approach is only justified for realistically simulated predictors, it is known as "perfect-prog(nosis)" downscaling (Kalnay 2003; Zwiers and Von Storch 2004; Wilks 2006). To obtain meaningful results it is crucial that the predictors capture the aspects of climate change that affect precipitation. Assessing whether this is the case can be difficult. In contrast, the response of hydrological processes, including moisture transport and evaporation as well as precipitation, to large-scale warming has been shown to be robust in many GCM simulations

(Held and Soden 2006). Discarding simulated precipitation is thus possibly excluding critical information that may not be fully captured by perfect-prog downscaling.

It is the purpose of this paper to clarify that there are three sources of error for simulated precipitation and to demonstrate that over large areas it is possible to approximately remove the first two errors and thus to isolate the third error. This error is directly related to the skill of a GCM in reproducing temporal precipitation variability given realistic large-scale weather states. We then assess whether there are areas where this skill is sufficiently high to allow for a meaningful statistical correction of the simulated precipitation. This so-called model output statistics (MOS) approach can combine a correction and a downscaling component and offers an alternative to the perfect-prog methods that have traditionally formed the basis of statistical downscaling methodologies in climate research.

2. Sources of model error

A strict assessment of model skill requires the distinction between three sources of errors in simulated precipitation (Fig. 1). If the simulated large-scale atmospheric conditions differ from reality, the simulated precipitation will be different from observations even if the convection and precipitation parameterizations in the model are perfect. There are two possible causes for such differences. First, a GCM may have a systematic bias in the large-scale atmospheric states or an unrealistic large-scale response to climate forcings, both of which we term a "type 1" error. Second, internally generated variability is unpredictable and will be different from the real world. We term this error "type 2" but note that this mismatch between simulated and observed variability is not a model deficiency but a consequence of a freely evolving GCM. Standard GCM simulations for historical periods yield the climatic response to forcings such as solar variability and changes in atmospheric concentrations of greenhouse gases and of anthropogenic and volcanic aerosols. These simulations are not constrained by historic meteorological observations and thus, because of the chaotic nature of the climate system, the circulation and temperature fields differ from reality with respect to the random, internally generated variability component. This random component dominates daily to interannual time scales and is still substantial on decadal time scales. Although the type 2 error can be ameliorated to some extent by temporal averaging and is usually approximately quantified by initial value ensemble simulations, understanding the causes of and quantifying the type 1 error is highly important and at the core of many model
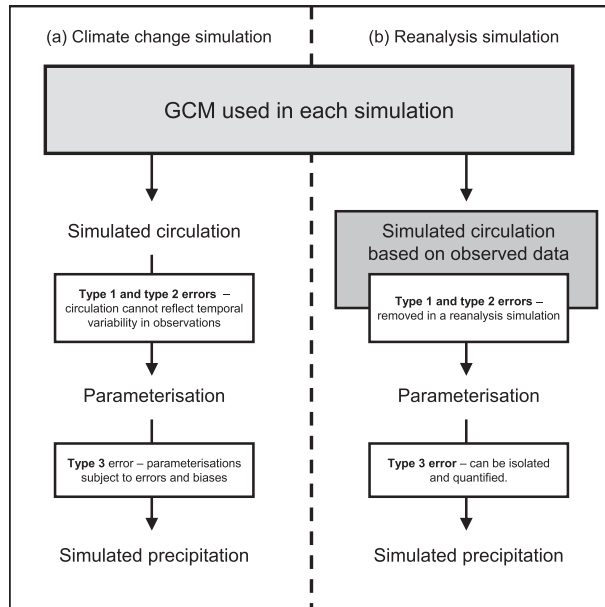
FIG. 1. (a) Climate change and (b) reanalysis GCM simulations of precipitation. Large-scale circulation fields simulated in the reanalysis are forced to real-world observations. Type 3 parameterization error still exists but is now isolated and can be quantified.

validation studies (Jansen et al. 2007; Randall et al. 2007) and of model development.

When considering precipitation, there is also a "type 3" error, which is caused by the deficiencies in convection and precipitation parameterizations as well as by differences between the real orography and the model world. This error will lead to differences between simulated and real precipitation even if the large-scale atmospheric conditions are in agreement with reality. When the task is to assess the skill of simulated precipitation variability, it is in many cases useful to focus on the type 3 error and to ask the question, how well is precipitation simulated given realistic large-scale conditions? This is conceptually consistent with the validation of dynamical and perfect-prog statistical downscaling methods by using realistic predictors, for instance, from reanalyses. When applying downscaling methods to GCM simulations for future climates, the downscaling error derived in this way is the error component introduced by the downscaling. The total error also includes the type 1 and type 2 errors. Analogously, the total error in future GCM-simulated precipitation is a sum of all three error types. It is useful to separate the error contributions in this way because the three errors have fundamentally different reasons.

Although until now it has not been possible to isolate the type 3 error for GCMs used for climate change simulations, Widmann and Bretherton (2000) have shown

that this error can be quantified for the GCMs that are used in atmospheric reanalyses (Kalnay et al. 1996; Kistler et al. 2001; Uppala et al. 2005). This approach is possible because, due to the assimilation of meteorological measurements such as pressure, wind speeds, temperature and humidity, the large-scale states in a reanalysis are the best estimates for the state of the real atmosphere. They are consistent with both the assimilated observations and the physical laws that govern the atmosphere that are represented in the model. However, no precipitation observations are assimilated and precipitation is simulated using the parameterizations and the large-scale atmospheric states derived in the assimilation process (Fig. 1). Several studies have shown that reanalyses capture observed temporal precipitation variability well (Gutowski et al. 1997; Janowiak et al. 1998; Widmann and Bretherton 2000; Widmann et al. 2003; Bosilovich et al. 2008). Although the GCMs used for reanalyses are similar to those used for climate simulations, they differ in terms of resolution and parameterizations, and thus the reanalysis-based results cannot be directly transferred to other GCMs.

The present study extends the reanalysis-based studies and quantifies how well precipitation variability is simulated in a GCM (ECHAM5) used for climate prediction in the IPCC Fourth Assessment Report (AR4) (Randall et al. 2007; Trenberth et al. 2007) if the error components related to large-scale atmospheric states and internal variability are approximately removed. To obtain a simulation in which the variability in large-scale circulation and temperature is in agreement with that observed in the real world, a simulation with ECHAM5 is conducted in which the prognostic variables are forced toward reanalysis values for a historical period. The subsequently parameterized precipitation field is thus expected to represent observed temporal variability for regions where reanalysis fields are skillful.

3. Simulation and observations

A nudging technique (also known as Newtonian relaxation) (Hoke and Anthes 1976; Krishnamurti et al. 1991; Jeuken et al. 1996; Timmreck et al. 1999; Timmreck and Schulz 2004) was used to force the ECHAM5-simulated divergence, vorticity, temperature, and surface pressure fields to corresponding fields from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005) for the period 1958–2001. This method has been successfully used in previous studies to force the circulation in earlier ECHAM versions toward a reanalysis for short periods (Jeuken et al. 1996; Timmreck et al. 1999; Timmreck and Schulz 2004; Bauer et al. 2008), but this is the first

time it has been applied to ECHAM5 and for the entire reanalysis period, which is needed for our statistical analysis of model skill.

For the nudging procedure, an additional term $N \cdot (X^{\text{obs}} - X)$ is added to the tendency equations for each variable, which take the form

$$\frac{\partial X}{\partial t} = F_m(X_t) + N \cdot (X^{\text{obs}} - X), \qquad (1)$$

as described in Krishnamurti et al. (1991). Here, F_m represents the model tendency (the dynamical and physical processes that determine the temporal evolution of X) (Wilks 2006), N is the relaxation coefficient, and X^{obs} represents an observed value to which the nudging is aimed. In practice, the full integration is a two-step process, with the normal tendencies first carried out,

$$X^*_{t+\Delta t} = X_t + F_m(X_t) \cdot \Delta t, \qquad (2)$$

where $X^*_{t+\Delta t}$ represents a prenudging predicted value of X at time $t + \Delta t$. The nudging procedure is applied in a second step,

$$X_{t+\Delta t} = X^*_{t+\Delta t} + N \cdot (X^{\text{obs}}_{t+\Delta t} - X^*_{t+\Delta t}) \cdot \Delta t, \qquad (3)$$

resulting in a new value $X_{t+\Delta t}$. The relaxation e-folding time τ is equal to $1/N$. Our choice of τ is variable specific and based on the results of previous work (Jeuken et al. 1996; Kaas 2000) (Table 1). In each case, τ should be short enough that the observed fields exert the required influence but not so short that the relaxation term is dominant over the model tendencies (Kaas 2000). Unlike in a standard ECHAM5 simulation, in which the sea surface temperature (SST) field is prescribed using a representative climatology, the nudged simulation requires a temporally realistic SST field. In this case, we use the same observed, monthly varying SSTs as have been used in ERA-40.

It is important to consider an additional source of precipitation error that may be induced by the nudging procedure. In a normal (nonnudged) simulation, the dynamically resolved and parameterized fields are physically consistent but the addition of a nudging term in (1) inhibits the parameterized quantities to reach this balanced state (Jeuken et al. 1996). An implication is a modification of diabatic heating in the GCM and a possible violation of energy conservation, which in turn may result in spurious precipitation. Jeuken et al. (1996) state that such errors will be greatest where the nudging term is large relative to the other terms in each tendency equation and that the relaxation coefficient chosen for temperature must be small enough to prevent a distortion

TABLE 1. Relaxation coefficients used in the nudging procedure: based on simple nudging assimilations used at the Danish Meteorological Institute (DMI) (Kaas 2000).

Prognostic variable	$\tau(X)$ (h)	N (s^{-1})
Vorticity	6.00	4.63×10^{-5}
Temperature	24.00	1.16×10^{-5}
log(surface pressure)	24.00	1.16×10^{-5}
Divergence	48.00	0.58×10^{-5}

in the diabatic heating quantities. This nudging-induced error, which we term "type 4" in accordance with the definitions used thus far, contributes along with parameterization deficiencies (type 3 error) to form the total precipitation error after the type 1 and type 2 errors have been approximately removed. Although the set of coefficients in Table 1 has been previously shown to be sufficient to guarantee that the nudging term does not dominate the model tendency equations, the magnitude of a potential type 4 error is expected to be associated with considerable geographical variability.

The output from the nudged simulation is compared to that of a standard (nonnudged) free-running ECHAM5 simulation for the period 1958–2001. Model output is on a T63 Gaussian grid, which equates to $1.875° \times 1.875°$ or roughly 200-km latitude × 150-km longitude at 45°. All analysis is conducted using seasonal-mean precipitation, which is consistent with climate change projections and also allows for a broad understanding of the seasonal dependence of model skill and thus of the potential for a downscaling correction. Seasonal means from each simulation are compared to observations from the Global Precipitation Climatology Project (GPCP) dataset (Huffman et al. 1997; Adler et al. 2003). Version 2 of this dataset, described in Adler et al. (2003), provides gridded ($2.5° \times 2.5°$) monthly means based on satellite and rain gauge observations for the period 1979–2001. Consideration was also given to a similar merged dataset, the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) (Xie and Arkin 1997) that also provides monthly precipitation at the same spatial resolution. Gruber et al. (2000) compared CMAP and a previous version of the GPCP dataset (Huffman et al. 1997), generally finding good agreement between the datasets but also some acute differences. CMAP was shown to treat rain gauge measurements differently to the GPCP dataset, such as in the inclusion of atoll gauge data to calibrate oceanic satellite estimates. The GPCP dataset has been used in other GCM and reanalysis validation work (e.g., Janowiak et al. 1998; Trenberth and Shea 2005) and is considered most suitable for this analysis.

For the development of a high-resolution scaling correction, a comparison is made with the Global Precipitation Climatology Center (GPCC) dataset, which is based only on interpolated rain gauge observations and covers only land areas but with a higher resolution of $0.5° \times 0.5°$ and for the period 1958–2001. Nudged ECHAM5 precipitation was linearly interpolated to the same $0.5° \times 0.5°$ grid in which the GPCC data are available. The correction was cross-validated using a leave-seven-out approach.

4. Results and discussion

a. Representation of interannual precipitation variability

It is first of all necessary to compare circulation and temperature variability in the nudged simulation with that in ERA-40 to assess how well the nudging procedure is able to bring the large-scale circulation and temperature into temporal phase with the reanalysis. Figure 2 shows correlation coefficients between simulated and observed (ERA-40) seasonal mean fields of geopotential height and temperature. Correlations are generally strong for both variables and are consistently higher than 0.95 across the majority of the extratropics. Tropical regions exhibit lower correlations, particularly in continental areas and during the boreal summer [June–August (JJA)]. Both simulated geopotential height and temperature at 850 hPa show noticeably low correlations with ERA-40 across much of northern South America (Figs. 2a–d). This may be attributable to a number of factors, including the relatively low number of upper-air observations over this region that are assimilated in ERA-40 and the important role of intense convective processes that prevail over the Amazon basin for much of the year. Both factors may lead to relatively large errors in the reanalysis and thus to target fields that are less likely to be consistent with the ECHAM5 dynamics and physics and that thus are less likely to be followed by ECHAM5. Correlations are generally higher at 500 hPa, reflecting the lower complexity of atmospheric fields at higher levels, although a tropical–extratropical difference is still apparent (Figs. 2e–h). In areas where ERA-40 represents the real-world large-scale atmospheric states well and where the correlations are high, the type 1 and type 2 errors are approximately removed and the remaining precipitation errors are a sum of the parameterization (type 3) and spurious precipitation (type 4) errors.

Simulated precipitation climatologies from both the nudged and standard (non-nudged) simulations are in reasonable agreement with observations in terms of large-scale features (Fig. 3). However, some key regional differences between the simulations are noted (Fig. 3, right), particularly lower precipitation in the western Pacific in the nudged simulation compared to the standard simulation. Other parts of the tropics also exhibit large differences, particularly Southeast Asia and the Amazon basin. These differences may be caused by either systematic circulation biases in the standard simulation or random differences in the internal variability, which may be relevant even for multidecadal averages.

Because the standard simulation is unable to represent observed interannual variability in seasonal mean precipitation, the correlation coefficients fluctuate randomly around zero (Fig. 4a). In contrast, seasonal-mean precipitation from the nudged simulation exhibits in many areas high correlations with observations (Figs. 4b–f), including the extratropics (30°–60°), especially over large parts of Northern Hemispheric landmass for all seasons with the exception of summer (JJA). Over the ocean, correlations fluctuate more, with the exception of the eastern equatorial Pacific, where they are consistently high; over tropical land areas, they are noticeably low. Indeed, agreement is considerably poorer over the majority of the African continent than any other region, tropical or extratropical. In general, correlations are low in regions of negligible precipitation, such as the maritime deserts of the subtropical Americas and southwestern Africa, which may be a consequence of poorer model performance or of larger errors in the observations.

This shows that given the correct large-scale atmospheric states ECHAM5 is over large areas able to successfully reproduce interannual variability of seasonal precipitation means. Many of the regions with low correlations are over regions where, according to Fig. 2, there is a low agreement between the large-scale atmospheric states in the nudged ECHAM5 simulation and in ERA-40. In addition, the ERA-40 atmospheric states may have considerable errors over some data-sparse regions such as the Southern Ocean or Antarctica. Thus, in these areas type 1 and type 2 errors are not fully eliminated and low correlations between simulated and observed precipitation do not necessarily indicate a large error for the sum of type 3 and type 4. In addition the low correlations over areas with sparse precipitation observations (e.g., over the oceans) are likely to include a contribution from errors in the gridded precipitation dataset.

Another potential reason for the lower correlations between the precipitation in the nudged ECHAM5 simulation and observations in the tropics is the dominance of convective processes in these regions. The
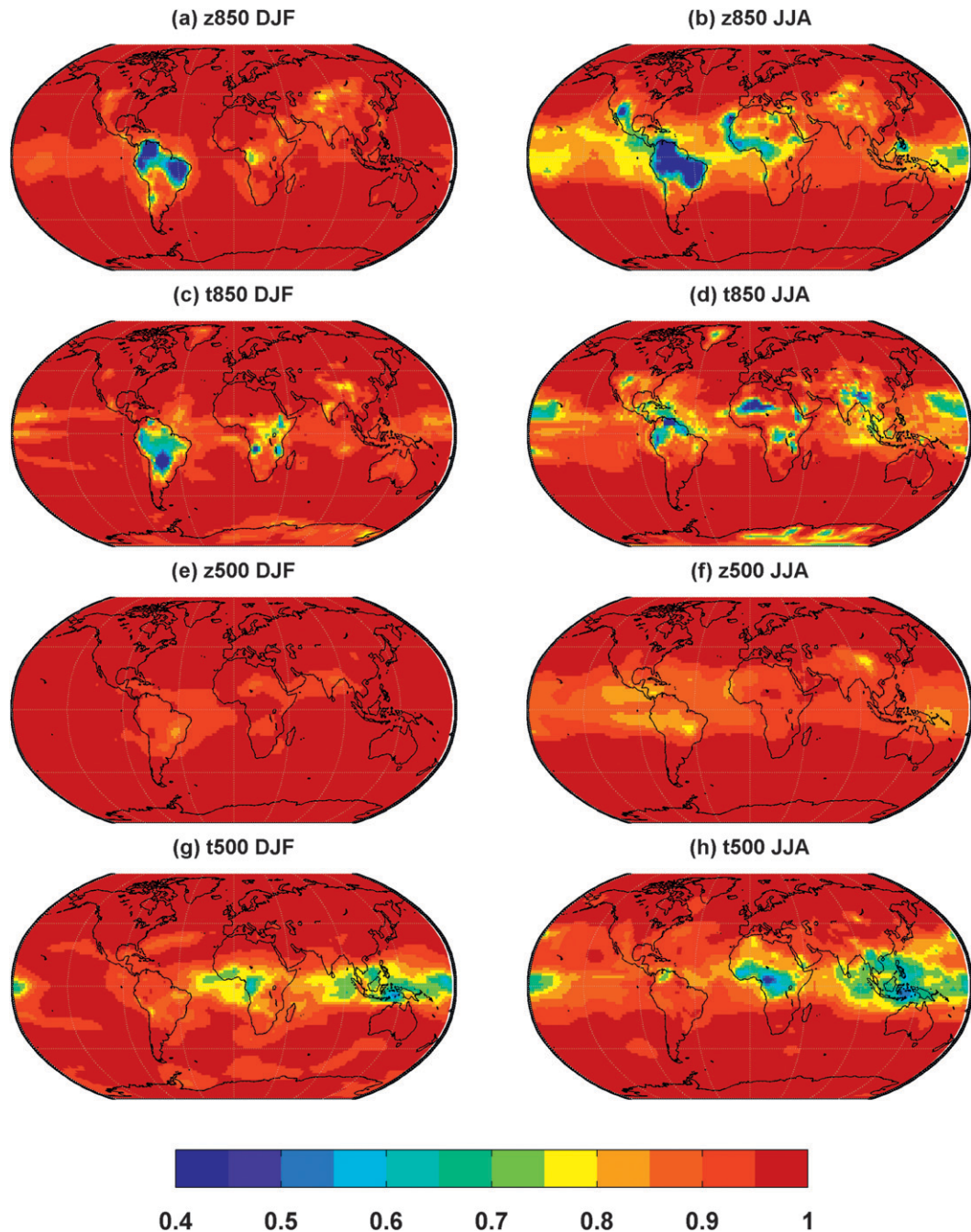
FIG. 2. Correlation between nudged ECHAM5 and ERA-40 seasonal mean geopotential height z and temperature t at 850 and 500 hPa for the period 1958–2001.

formation of convective precipitation acts on small scales and has a large random component for a given large-scale atmospheric state. This means that relatively low correlations can be expected even if the nudging successfully controls the large-scale states and the precipitation parameterizations perform well. In contrast, the spatial and temporal distribution of frontal precipitation is strongly constrained by large-scale atmospheric states, and over areas dominated by frontal precipitation low correlations indicate problems with the precipitation parameterizations. The prevalence of convective precipitation in the tropics, as well as in extratropical regions during the summer (e.g., in southern Europe), is consistent with the tropical–extratropical difference in the
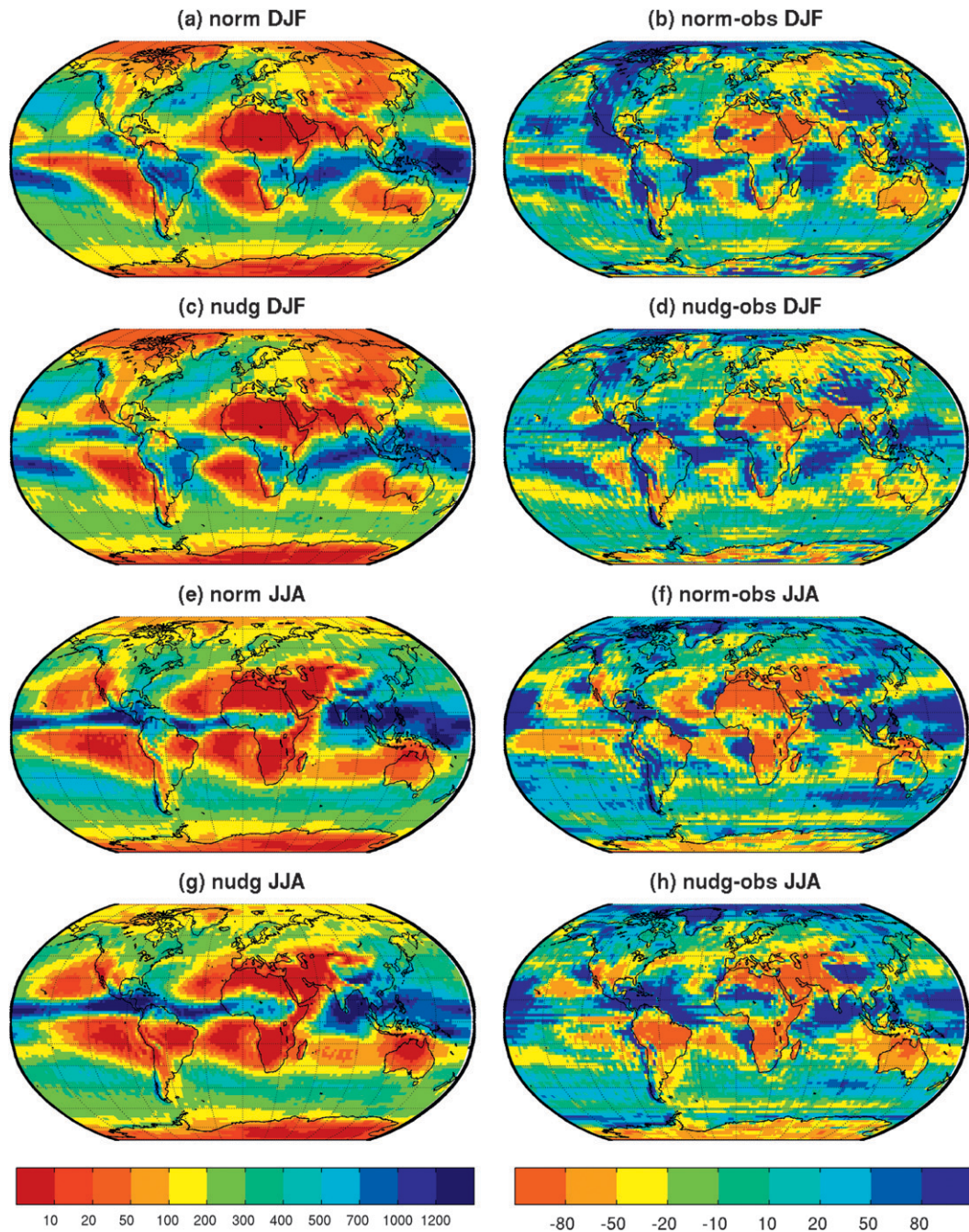
FIG. 3. (left) Seasonal precipitation in the standard (norm) and nudged (nudg) ECHAM5 simulations (mm) and (right) respective differences between simulated and GPCP observed precipitation (%) for the period 1979–2001.

correlations. Thus, potential difficulties in removing the type 1 and type 2 errors, in addition to the randomness of convective precipitation, may contribute to the low correlations in the tropics, but it is unclear whether these factors fully explain the low correlations or whether there is a substantial contribution from a genuine parameterization error.

Because type 1 and type 2 errors, in addition to errors in the gridded precipitation dataset, cannot be fully eliminated, the correlation maps, which quantify the total error, provide an upper estimate for the sum of the parameterization error (type 3) and the nudging-induced error (type 4). Because the type 4 error is very unlikely to compensate the type 3 error, the correlations also
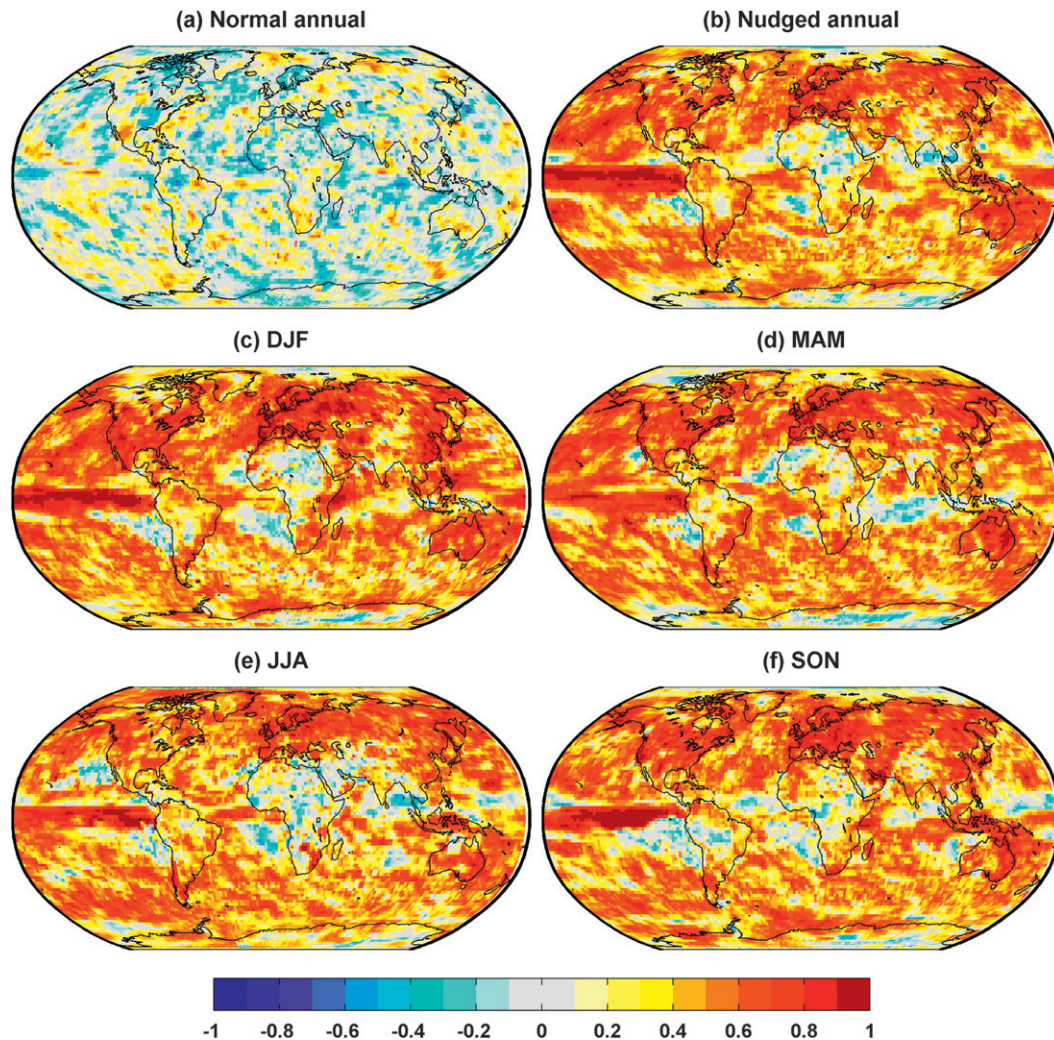
FIG. 4. Correlation of ECHAM5 precipitation and GPCP observations (1979–2001). Annual correlation of (a) normal and (b) nudged ECHAM5 precipitation and GPCP observations. (c)–(f) Seasonal mean correlation of nudged ECHAM5 precipitation and GPCP observations.

provide an upper estimate for the type 3 error. It is beyond the scope of this study to separate the individual error components. Our objective is to identify regions where the total precipitation error in the nudged ECHAM5 simulation is small and where we can therefore conclude that the type 3 error is small. In these areas, the temporal precipitation variability is realistic given the simulated large-scale atmospheric states.

b. Potential for correction and downscaling

Even in areas where the correlations of simulated and observed precipitation are high, the simulated values may still systematically under- or overestimate the real values. If these scaling errors in simulated precipitation are stationary in time, they can be corrected (Widmann et al.

2003), and this simple correction has been used, for instance, as a reference method for statistical downscaling (Schmidli et al. 2006). Although such corrections are a form of statistical downscaling as they transform GCM output to small-scale estimates, the approach, which is generally known as MOS, is different from standard perfect-prog downscaling (Kalnay 2003; Zwiers and Von Storch 2004; Wilks 2006). Although perfect-prog downscaling is based on real-world-scale relations and is not designed for the purpose of correcting a particular model's inherent errors or biases, MOS combines a downscaling step with a model-specific correction step. MOS is routinely used in weather forecasting, and it has been shown that, in the context of global climate modeling, scaling or more sophisticated MOS methods are a promising alternative to perfect-prog downscaling

(Landman and Goddard 2002; Widmann et al. 2003; Feddersen and Andersen 2005; Schmidli et al. 2006).

However, so far it has only been possible to investigate this approach for GCMs used in reanalyses or for seasonal forecasts. There are two main reasons for this. First, the fitting of many MOS methods that are more advanced than a simple scaling (e.g., regression, multiple predictors, nonlocal methods based on coupled patterns) requires GCM simulations in which the simulated large-scale atmospheric states are in agreement with the real-world states. For standard GCM simulations for past climates this is not the case. Second, simple correction methods such as scaling that can in principle be fitted using standard simulations yield only meaningful results if the simulated precipitation is an informative predictor for the real-world precipitation for a given large-scale atmospheric state: in other words, if the type 3 error is small. This point is crucial as can be seen, for instance, from a hypothetical example in which the two hemispheres have been erroneously swapped in the output for the simulated precipitation. Obviously, scaling would be able to bring the corrected simulated precipitation climatologies in agreement with observations, but nevertheless the corrected time series would be meaningless. Our analysis of nudged simulations allows regions in which this condition is satisfied to be identified. Moreover, only simulations in which the large-scale atmospheric states are brought close to reality by nudging (or by direct data assimilation) allow a derivation of scaling factors that are free from errors caused by differing internal variability in the simulation and in the real world (type 2 error). Recently, MOS corrections have been successfully applied to RCMs (e.g., Leander and Buishand 2007; Graham et al. 2007; Themeβl et al. 2011). In this case, the MOS models can be fitted and validated using reanalysis-driven RCM hindcasts, in which the large-scale atmospheric states are close to the real-world states.

To assess whether different corrections should be used for wet and dry situations, scaling factors for ECHAM5 output were derived separately for the lower, middle, and upper terciles (t1, t2, and t3) of extratropical seasonal precipitation (Fig. 5). At this stage, scaling is conducted on the ECHAM5 grid ($1.87° \times 1.87°$) and thus does not include a downscaling step. Northern Hemisphere precipitation for the dry winter seasons [December–February (DJF)–t1] is slightly overestimated by the model over the Arctic, western Canada, and parts of Asia and slightly underestimated in most of the mid-latitudes. For the wet winter seasons (DJF–t3), under-/ overestimation is similar, with Arctic estimates very close to observations. Most of the areas with substantial scaling errors are very dry in winter, and even small absolute

errors lead to considerable scaling factors. In summer (JJA), the results are similar in the sense that scaling errors are mainly located over dry regions and that the t1 and t3 patterns are similar, whereas the magnitude of the scaling errors is larger than in winter. Thus, the model performs in both seasons similarly in wet and dry years, whereas scaling factors are regionally and seasonally dependent. In both seasons, the scaling factors are close to 1 over most areas of substantial precipitation, with East Asia in summer being the most important exception. In the Southern Hemisphere, a similar pattern emerges of underestimation (overestimation) of the wettest (driest) events. There is a more widespread underestimation of dry winter seasons (JJA–t1) compared to dry summer seasons (DJF–t1). Across the southern part of South America, all three terciles are consistently underestimated, with the exception of the western coastline, where precipitation is consistently overestimated.

This approach was then extended to derive a downscaled correction for ECHAM5 precipitation based on the high-resolution GPCC gridded precipitation dataset ($0.5° \times 0.5°$ grid). We focus on Europe, where correlations are generally high during most of the year. As mentioned above, the scaling correction of simulated precipitation would yield only meaningful results in areas where the type 3 error is small, and thus the factors are shown only for areas in which the correlations between seasonal precipitation means in the GPCC dataset and in the nudged ECHAM5 simulation are higher than 0.7 (Fig. 6). The GPCP- and GPCC-based correlation maps (not shown) are very similar to one another on the coarser GPCP grid, and GPCC-based correlations over data-dense regions are as high as 0.9. Considerable scaling errors include overestimation of spring and partly autumn precipitation over Scandinavia by a factor of 1.7 and underestimation in parts of southern Europe in all seasons up to a factor of 2. In autumn and winter, almost all of Europe exhibits correlations above the 0.7 threshold (with the Alps being a noticeable exception), whereas in spring and particularly in summer the temporal variability is not well enough simulated to make a correction of simulated precipitation meaningful.

The downscaling correction was cross-validated using a variant of the leave-one-out approach. This approach allows for seasonal precipitation for each year to be estimated independently using a scaling factor derived from simulated and observed data from all other years between 1958 and 2001. A leave-seven-out approach was used in which a scaling factor to estimate seasonal precipitation for a given year was derived using simulated and observed data from all other years aside from the 7-yr period centered on the year to be estimated. A period of 7 yr was chosen to account for the influence autocorrelation.
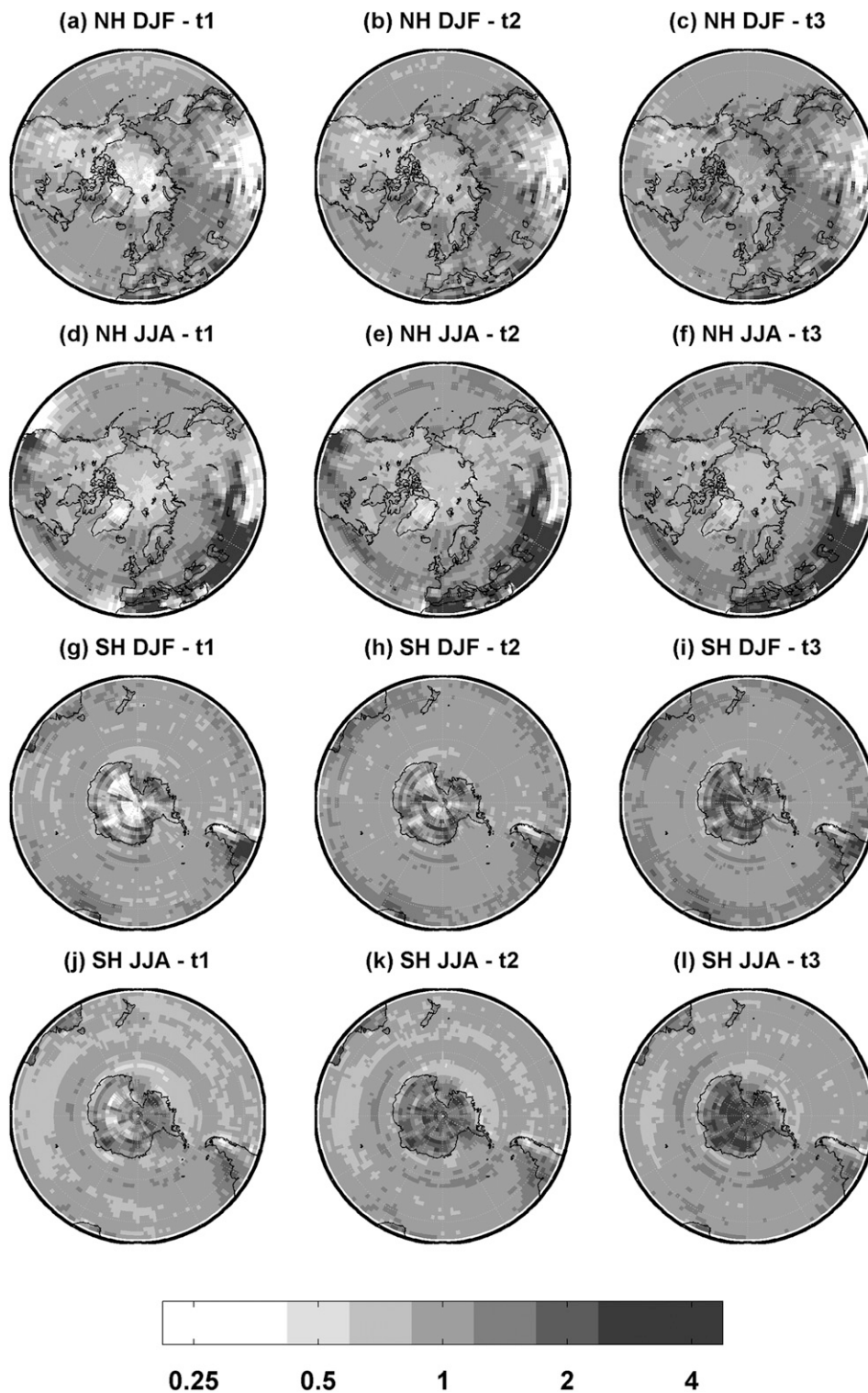
FIG. 5. Seasonal scaling corrections for (a)–(f) the Northern Hemisphere and (g)–(l) the Southern Hemisphere from 30° to 90° (1979–2001). Scaling correction is defined as GPCP observations divided by nudged ECHAM5 precipitation. DJF scaling corrections for the (a),(g) driest (t1), (b),(h) middle (t2), and (c),(i) wettest (t3) third of events.
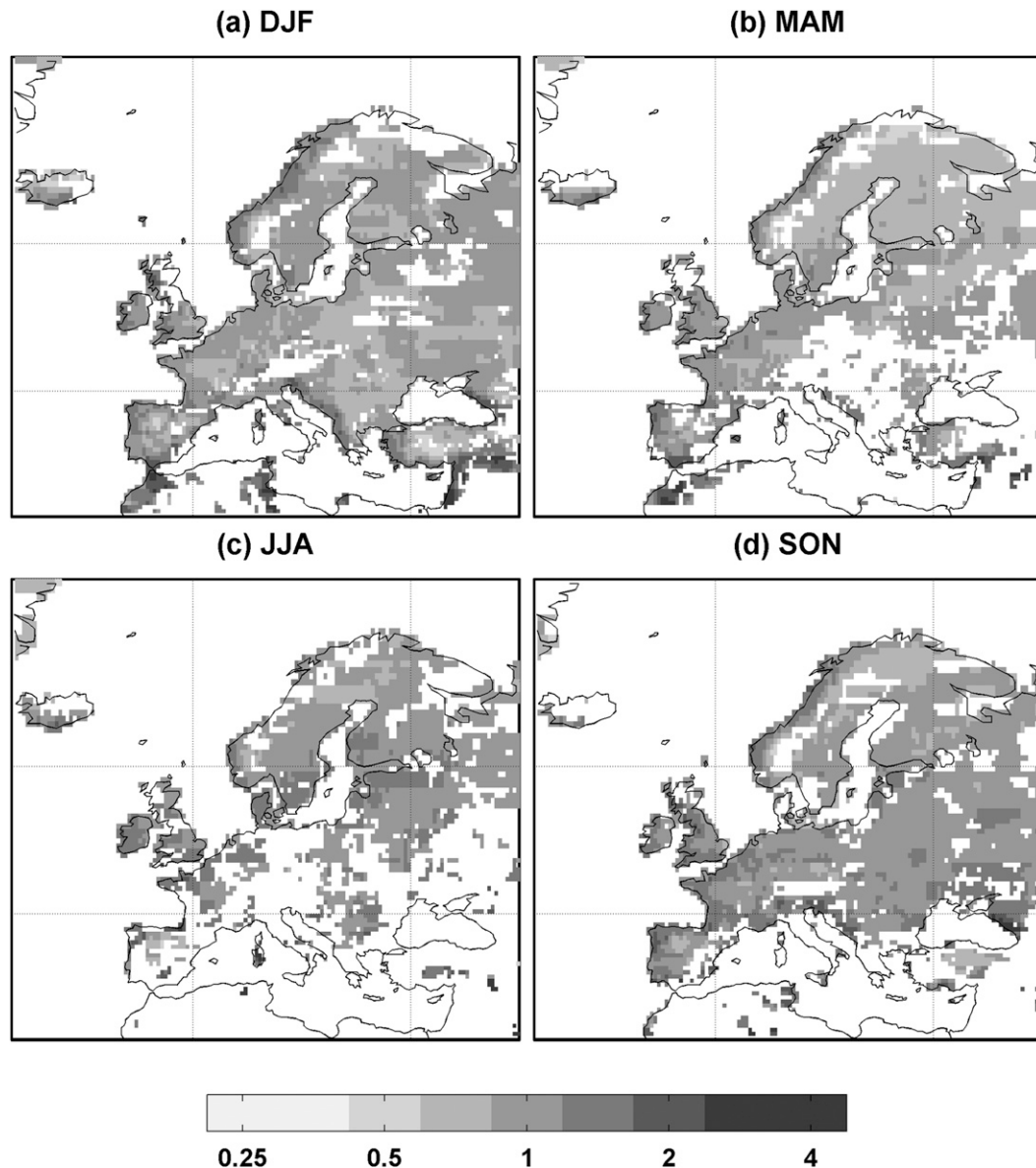
FIG. 6. High-resolution (0.5° × 0.5°) seasonal scaling corrections for Europe (1958–2001). Scaling correction is defined as GPCC observations divided by nudged ECHAM5 precipitation. (a) DJF, (b) March–May, (c) JJA, and (d) September–November. Because the scaling correction of simulated precipitation would yield only meaningful results in regions where the correlation between simulated and observed values is high, the factors are shown only for areas in which the correlations for seasonal means (using the GPCC data) are higher than 0.7.

Area means of the observed, the simulated, and the corrected precipitation are shown in Fig. 7 for three different regions over which the correlations between observations and simulated values are high. Over the Mediterranean basin and Australia, ECHAM5 consistently underestimates precipitation and the corrected values are a substantial improvement. Over southern Africa, the uncorrected precipitation has a less consistent bias and the local correlations with observations as well as the correlation of the area mean are lower than

for the other regions, which leads to a less successful correction. The quality of ERA-40 (to which the simulation is nudged) is problematic over the Southern Hemisphere in the presatellite era, which appears to be reflected in the better skill of the nudged simulation in the second half of the analysis period. In all three cases, good skill on interannual time scales is also associated with good reproduction of decadal variability. It is this property that makes the application of scaling correction factors to climate change simulations promising.
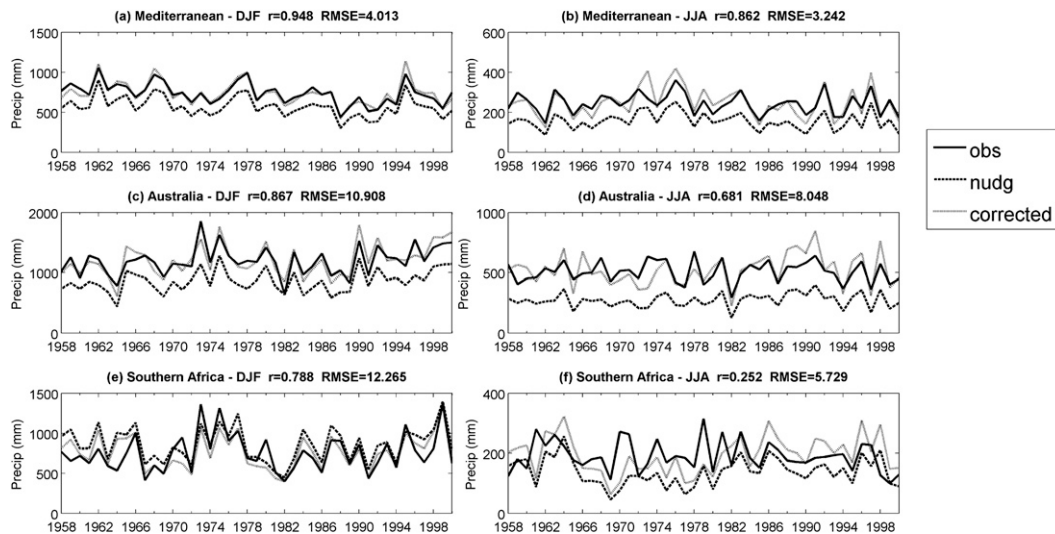
FIG. 7. Observed, simulated, and corrected DJF and JJA precipitation (1958–2001). (a),(b) Mediterranean basin, (c),(d) Australia, and (e),(f) southern Africa. GPCC observations are represented by the solid line, nudged ECHAM5 precipitation is represented by the dashed line, and the cross-validated correction of nudged ECHAM5 precipitation is represented by the dotted line. Correlation coefficient r and root-mean-square error (RMSE) of the observed and corrected time series are shown for each location.

To further demonstrate the potential of a downscaling correction for climate change simulations, the scaling factor method was applied to the ECHAM5/Max Planck Institute Ocean Model (MPI-OM) three-member ensemble mean projection for the twenty-first century for the Special Report on Emissions Scenarios (SRES) A1B scenario, which is included in the IPCC AR4. Raw and corrected projections for European mean seasonal precipitation for the period 2080–99 are shown in Fig. 8, with a "stippling" overlay indicating areas of good model skill (correlations higher than 0.7). During both winter (DJF) and summer (JJA), the downscaled and corrected precipitation captures key features that are not included in the original simulated precipitation. Heavy winter precipitation is far more concentrated along the coastline of western Europe after the correction. Differences in precipitation across the Alps, Pyrenees, and other regions of complex topography are also visible, although model skill tends to be lacking in these areas.

It is important to note that a meaningful application of the corrections developed here to simulations of future climate requires statistical relationships that are stationary in time. Although this is a key limitation of all statistical downscaling methods, it seems likely that MOS corrections are more stable than perfect-prog relationships. The latter would become unstable if the link between large-scale predictors, such as atmospheric circulation or temperature, with local precipitation changed for any reasons in a future climate. Many of the physical processes that may change typical perfect-prog

statistical links are however resolved by GCMs: for instance, changes in moisture availability that may affect circulation–precipitation links. Thus, MOS models that correct and downscale simulated precipitation are only affected by instabilities that affect the nonresolved processes represented by the parameterizations. One reason for such instabilities is a potential dependency of systematic biases caused by the parameterizations on the large-scale atmospheric state: for instance, through a dependency on the relative contributions of frontal and convective precipitation. For both perfect-prog and MOS relationships, the validity of the statistical links in different climates should therefore be thoroughly tested.

5. Summary and conclusions

The skill of GCMs in simulating precipitation is difficult to assess through simulation–observation comparison, because standard GCM simulations for historic periods do not represent random real-world temporal variability. Here, a distinction was made between three sources of error in GCM-simulated precipitation, and it was suggested that through removing the model's systematic large-scale circulation error (type 1) and the error due to random internally generated variability (type 2) it is possible to isolate the error caused by parameterization and by differences in the model and real-world orography (type 3).

The skill of ECHAM5 to simulate precipitation given realistic large-scale conditions has been investigated by
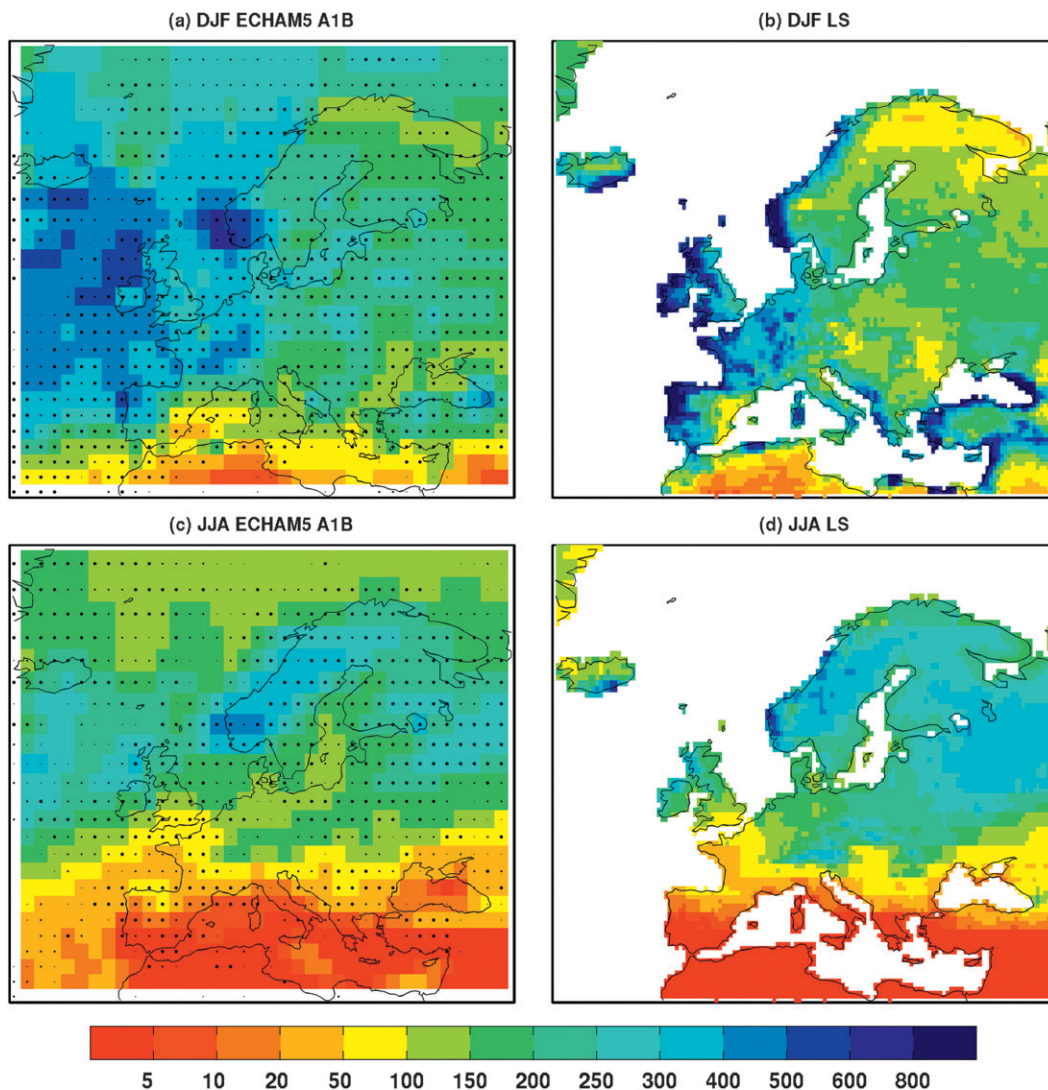
FIG. 8. (left) Seasonal precipitation projections (mm) for 2080–99 simulated by ECHAM5 according to the SRES A1B scenario and (right) subsequent downscaled corrections on a 0.5° × 0.5° grid. Stippled areas in the raw ECHAM5 projections indicate regions of good model skill, determined by the degree of correlation between simulated and observed precipitation between 1958 and 2001. Stipples are sized to reflect correlation coefficients of greater than 0.6 (small stipples), 0.7 (medium stipples) and 0.8 (large stipples).

using a nudged simulation for the period 1958–2001, in which the simulated circulation and temperature variables were forced toward corresponding fields from ERA-40. To the extent that the nudged simulation reproduces the temporal variability in the observed large-scale atmospheric states, the type 1 and type 2 errors can be eliminated and the type 3 precipitation error can be assessed. Because the type 1 and type 2 errors cannot be completely removed, the analysis of the remaining total error leads to an upper estimate for the type 3 error. As noted by Jeuken et al. (1996), the nudging may induce imbalances and spurious precipitation, and thus an additional (type 4) error may be introduced. Yet the

total precipitation error in the nudged simulations still yields an upper estimate for the type 3 error.

Although a more detailed analysis of the different error components is in principle possible, it is beyond the scope of this study. We focused on the identification of areas with a small total error in the nudged simulation. Over these areas, the type 3 error is small and thus the temporal variability of the simulated precipitation is consistent with the simulated large-scale atmospheric states.

Until now, the confidence one should have in GCM-simulated precipitation has typically been assessed from the level of agreement in multimodel ensembles (e.g.,

Randall et al. 2007) and by focusing on the representation of large-scale modes of variability (i.e., ENSO). Such comparisons cannot distinguish between the three types of error that may be responsible for the differences. With our analysis of type 3 errors, it is possible to identify regions where agreement among different models is backed up by good parameterization skill in ECHAM5.

Using correlation maps between observed (based on rain gauges and satellites) and simulated seasonal precipitation, it was shown that over many regions of the world the precipitation variability in the nudged simulation is in excellent agreement with the observations, and thus over these areas the type 3 error is small. The model skill in representing interannual precipitation variability, given the correct large-scale atmospheric states, was shown to be particularly high in the extratropics and over the Northern Hemispheric land surface. In particular, over Africa and South America the remaining total error was still high, which may be caused by large type 3 errors but potentially also by remaining type 1 and type 2 errors in the nudged simulation or by errors in the precipitation observations.

For regions with a small type 3 error, a statistical correction of the simulated precipitation was proposed, which can include a downscaling component and is conceptually a model output statistics (MOS) approach. MOS is regularly used in weather forecasting but has so far not been used for climate simulations because the fitting and validation of the statistical correction usually requires nudged GCM simulations. A simple scaling-factor correction and cross-validation was used to demonstrate that simulated precipitation is an excellent predictor for local precipitation in many regions. We note that through the nudging systematic biases in the simulated large-scale atmospheric states are approximately removed and therefore the application of the MOS corrections to future simulations does not correct for type 1 errors in these simulations. Whether it is the purpose of statistical downscaling to correct for systematic errors in the large-scale GCM states is arguable. For large errors, the answer is clearly no, because a fundamentally low-performing GCM cannot be corrected by statistical methods that are substantially less complex than the GCM itself. Our approach is fully consistent with standard perfect-prog downscaling approaches, which also do not remove type 1 errors because the statistical models are based on observations, as well as with regional modeling, where the large-scale states in the regional model are constrained by the driving GCM. In all these cases, the downscaling methods attempt to find local variables that are consistent with the large-scale GCM states.

Being able to make reliable estimates of future precipitation continues to be a major challenge to global climate modeling. At the local scale, methods of downscaling GCM output continue to form the basis for these estimates. The success of a relatively simple MOS downscaling approach suggests there is potential to exploit more sophisticated techniques that use simulated precipitation as a predictor variable. It is anticipated that future work will develop a range of MOS methods and evaluate their success against traditional perfect-prog methods and that skillful downscaling methods will be applied more comprehensively to climate change simulation than in this study.

REFERENCES

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature,* **419,** 224–232, doi:10.1038/nature01092.

Bauer, H.-S., V. Wulfmeyer, and L. Bengtsson, 2008: The representation of a synoptic-scale weather system in a thermodynamically adjusted version of the ECHAM4 general circulation model. *Meteor. Atmos. Phys.,* **99** (3–4), 129–153, doi:10.1007/s00703-007-0275-2.

Bosilovich, M. G., J. Y. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.,* **47,** 2279–2299.

Christensen, J. H., and Coauthors, 2007: Regional climate projections. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 847–940.

Feddersen, H., and U. Andersen, 2005: A method for statistical downscaling of seasonal ensemble predictions. *Tellus,* **57A,** 398–408.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.,* **27,** 1547–1578.

Graham, L. P., J. Andreasson, and B. Carlsson, 2007: Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—A case study on the Lule River basin. *Climatic Change,* **81** (Suppl.), 293–307, doi:10.1007/s10584-006-9215-2.

Groisman, P. Ya., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. A. N. Razuvaev, 2005: Trends in intense precipitation in the climate record. *J. Climate,* **18,** 1326–1350.

Gruber, A., X. J. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two merged rain gauge–satellite precipitation datasets. *Bull. Amer. Meteor. Soc.,* **81,** 2631–2644.

Gutowski, W. J., Y. B. Chen, and Z. Otles, 1997: Atmospheric water vapor transport in NCEP–NCAR reanalyses: Comparison with river discharge in the central United States. *Bull. Amer. Meteor. Soc.,* **78,** 1957–1969.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate,* **19,** 5686–5699.

Hoke, J. E., and R. A. Anthes, 1976: Initialization of numerical-models by a dynamic initialization technique. *Mon. Wea. Rev.,* **104,** 1551–1556.

Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bull. Amer. Meteor. Soc.,* **78,** 5–20.

Janowiak, J. E., A. Gruber, C. R. Kondragunta, R. E. Livezey, and G. J. Huffman, 1998: A comparison of the NCEP–NCAR reanalysis precipitation and the GPCP rain gauge-satellite combined dataset with observational error considerations. *J. Climate,* **11,** 2960–2979.

Jansen, E. J., and Coauthors, 2007: Palaeoclimate. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 433–497.

Jeuken, A. B. M., P. C. Siegmund, L. C. Heijboer, J. Feichter, and L. Bengtsson, 1996: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *J. Geophys. Res.,* **101** (D12), 16 939–16 950.

Kaas, E., Ed., 2000: Project on Tendency Evaluations using New Techniques to Improve Atmospheric Long-Term Simulations. EU-Commision POTENTIALS Final Rep., 93 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

——, and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Krishnamurti, T. N., J. S. Xue, H. S. Bedi, K. Ingles, and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over the tropics. *Tellus,* **43A,** 53–81.

Landman, W. A., and L. Goddard, 2002: Statistical recalibration of GCM forecasts over southern Africa using model output statistics. *J. Climate,* **15,** 2038–2055.

Leander, R., and T. A. Buishand, 2007: Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrol.,* **332** (3–4), 487–496, doi:10.1016/j.jhydrol.2006.08.006.

Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.,* **48,** RG3003, doi:10.1029/2009RG000314.

Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 589–662.

Schmidli, J., C. Frei, and P. L. Vidale, 2006: Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods. *Int. J. Climatol.,* **26,** 679–689.

Themeβl, M. J., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.,* **31,** 1530–1544.

Timmreck, C., and M. Schulz, 2004: Significant dust simulation differences in nudged and climatological operation mode of the AGCM ECHAM. *J. Geophys. Res.,* **109,** D13202, doi:10.1029/2003JD004381.

——, H. F. Graf, and J. Feichter, 1999: Simulation of Mt. Pinatubo volcanic aerosol with the Hamburg climate model ECHAM4. *Theor. Appl. Climatol.,* **62** (3–4), 85–108.

Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Lett.,* **32,** L14703, doi:10.1029/2005GL022760.

——, and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 235–336.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Widmann, M., and C. S. Bretherton, 2000: Validation of mesoscale precipitation in the NCEP reanalysis using a new grid-cell dataset for the northwestern United States. *J. Climate,* **13,** 1936–1950.

——, ——, and E. P. Salathe, 2003: Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor. *J. Climate,* **16,** 799–816.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. IPCC Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA) Rep., 27 pp.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier, 627 pp.

Xie, P. P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature,* **448,** 461–465, doi:10.1038/nature06025.

Zwiers, F. W., and H. Von Storch, 2004: On the role of statistics in climate research. *Int. J. Climatol.,* **24,** 665–680.