PRECIPITATION DOWNSCALING UNDER CLIMATE CHANGE: RECENT DEVELOPMENTS TO BRIDGE THE GAP BETWEEN DYNAMICAL MODELS AND THE END USER

D. Maraun,[1,2] F. Wetterhall,[3] A. M. Ireson,[4] R. E. Chandler,[5] E. J. Kendon,[6] M. Widmann,[7] S. Brienen,[8,9] H. W. Rust,[10] T. Sauter,[11] M. Themeßl,[12] V. K. C. Venema,[8] K. P. Chun,[4] C. M. Goodess,[2] R. G. Jones,[6] C. Onof,[4] M. Vrac,[10] and I. Thiele-Eich[8]

[1] Precipitation downscaling improves the coarse resolution and poor representation of precipitation in global climate models and helps end users to assess the likely hydrological impacts of climate change. This paper integrates perspectives from meteorologists, climatologists, statisticians, and hydrologists to identify generic end user (in particular, impact modeler) needs and to discuss downscaling capabilities and gaps. End users need a reliable representation of precipitation intensities and temporal and spatial variability, as well as physical consistency, independent of region and season. In addition to presenting dynamical downscaling, we review perfect prognosis statistical downscaling, model output statistics, and weather generators, focusing on recent developments to improve the representation of space-time variability. Furthermore, evaluation techniques to assess downscaling skill are presented. Downscaling adds considerable value to projections from global climate models. Remaining gaps are uncertainties arising from sparse data; representation of extreme summer precipitation, subdaily precipitation, and full precipitation fields on fine scales; capturing changes in small-scale processes and their feedback on large scales; and errors inherited from the driving global climate model.

1. INTRODUCTION

[2] *Global climate models* (GCMs) are the primary tool for understanding how the global climate may change in the future. (Italicized terms are defined in the glossary, after the main text.) However, these currently do not provide reliable information on scales below about 200 km [*Meehl et al.*, 2007] (for an illustration, see Figure 1). Hydrological processes typically occur on finer scales [*Kundzewicz et al.*, 2007]. In particular, GCMs cannot resolve circulation patterns leading to hydrological extreme events [*Christensen and Christensen*, 2003]. Hence, to reliably assess hydrological impacts of climate change, higher-resolution scenarios are required for the most relevant meteorological variables.

[3] Downscaling attempts to resolve the scale discrepancy between climate change scenarios and the resolution required for impact assessment. It is based on the assumption that large-scale weather exhibits a strong influence on local-scale weather but, in general, disregards any reverse effects from local scales upon global scales. Two approaches to downscaling exist. *Dynamical downscaling* nests a *regional climate model* (RCM) into the GCM to represent the atmospheric physics with a higher grid box resolution within a

---

[1]Department of Geography, University of Giessen, Giessen, Germany.

[2]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.

[3]Department of Geography, King's College London, London, UK.

[4]Environmental and Water Resources Engineering Research Group, Imperial College London, London, UK.

[5]Department of Statistical Science, University College London, London, UK.

[6]Met Office Hadley Centre, Exeter, UK.

[7]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.

[8]Meteorological Institute, University of Bonn, Bonn, Germany.

[9]Now at Deutscher Wetterdienst, Offenbach, Germany.

[10]Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CNRS, Centre d'Etude de Saclay, Gif-sur-Yvette, France.

[11]Department of Geography, RWTH Aachen University, Aachen, Germany.

[12]Wegener Center for Climate and Global Change, University of Graz, Graz, Austria.
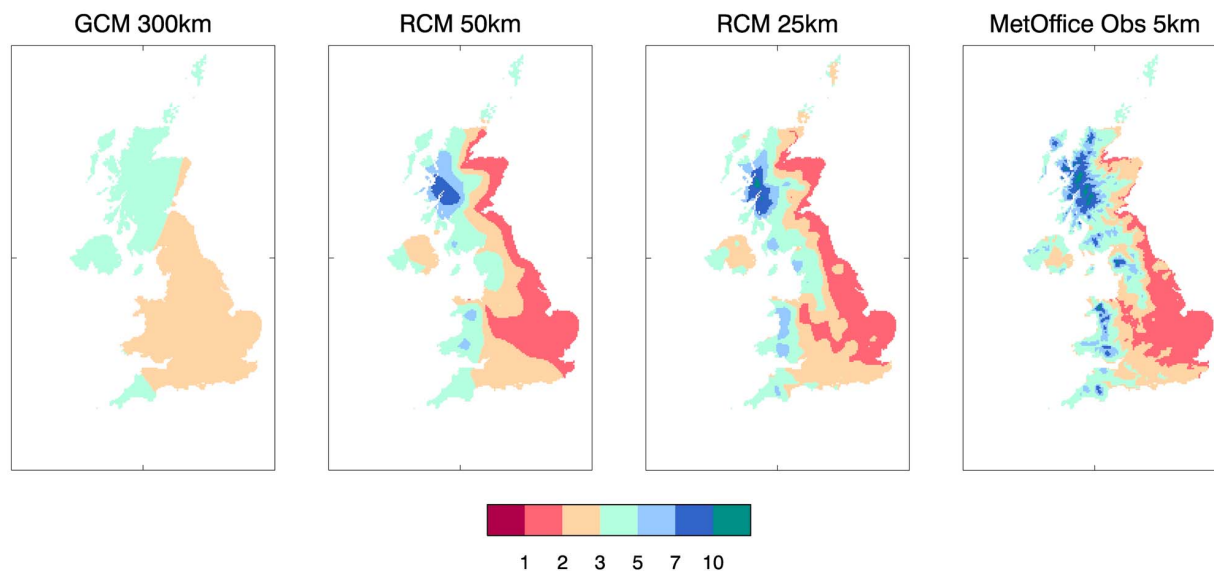

Figure 1. Average UK winter precipitation (mm/d) for 1961–2000 simulated by the Hadley Centre global climate model (GCM) HadCM3 and the regional climate model (RCM) HadRM3 at 50 and 25 km resolutions compared with gridded observations (E. Buonomo et al., unpublished data, 2009). The GCM does not provide regional precipitation information. The RCM reproduces basic regional structure but is limited in mountain areas (western UK); in addition, this particular RCM exaggerates the rain shadow effect (east Scotland).

limited area of interest. *Statistical downscaling* establishes statistical links between large(r)-scale weather and observed local-scale weather.

[4] During the last 2 decades, extensive research on downscaling methods and applications has been carried out. For a comprehensive overview of applications, see *Christensen et al.* [2007]; see also *Prudhomme et al.* [2002] and *Fowler et al.* [2007a], who focus on hydrology. Several reviews of downscaling methods have been published [e.g., *Hewitson and Crane*, 1996; *Zorita and von Storch*, 1997; *Wilby and Wigley*, 1997; *Xu*, 1999a; *Hanssen-Bauer et al.*, 2005]. In addition to updating these methodological reviews, this paper aims to integrate different perspectives on precipitation downscaling, in particular, from meteorologists, climatologists, statisticians, and impact modelers such as hydrologists. As such, we focus on laying out concepts and discussing methodological advances.

[5] In general, the most relevant meteorological variables for hydrological impact studies are precipitation and temperature [*Xu*, 1999b; *Bronstert et al.*, 2007]. For freshwater resources in particular, precipitation is the most important driver [*Kundzewicz et al.*, 2007], though it is considerably more difficult to model than temperature mostly because of its high spatial and temporal variability and its nonlinear nature. The overall objective of this paper is to define a set of generic end user needs (in particular, for impact modelers) for downscaled precipitation and then to discuss how these needs are met by various downscaling approaches and what gaps are remaining.

[6] Statistical downscaling has received considerable attention from statisticians. Their contributions have, however, largely been unrecognized by the climate community, although they attempt to address important end user needs. An essential part of this paper is therefore to review recent statistical models that have been developed to improve the representation of spatial-temporal variability and extremes. We attempt to bring these recent approaches together with classical statistical downscaling methods and discuss differences and similarities between individual methods and approaches, as well as their advantages and drawbacks.

[7] Traditionally, statistical downscaling has been seen as an alternative to dynamical downscaling. With the increasing reliability and availability of RCM scenarios, recent work on statistical downscaling has aimed to combine the benefits of these two approaches. Under the name *model output statistics* (MOS), gridded RCM simulations are statistically corrected and downscaled to point scales. We describe MOS approaches in detail and discuss their relation to other statistical downscaling approaches.

[8] To seriously evaluate the skill of downscaling approaches to meet the end user needs, a quantitative evaluation is necessary. Therefore, an important part of the paper is a review of validation techniques.

[9] In section 2 we identify a set of generic end user needs. The state of the art in dynamical and statistical downscaling is presented in sections 3 and 4, respectively, and in section 5, validation techniques are introduced. Finally, in section 6 we discuss how the approaches presented in sections 3 and 4 meet the specific needs identified in section 2. In particular, section 6 seeks to address the following questions: How does dynamical downscaling address a particular end user need? How can MOS improve the RCM simula-

tions and close potential gaps? How does statistical downscaling perform as alternative to dynamical downscaling? What are the remaining gaps? Sections 3–5 are quite technical in nature, while sections 2 and 6 are written to be accessible to the nonexpert.

2. NEEDS OF THE END USER

[10] Downscaling precipitation, in most cases, is not an end in itself but provides a product (in the form of data or information) to an "end user." Their goal may be, for example, to understand and potentially act upon the impacts that are likely to be caused by a localized climate extreme or by a future change in the climate. End users range from policy makers, through planners and engineers, to impact modelers. As well as the product, the end user might also require a clear statement of the assumptions involved and limitations of the downscaling procedure, a transparent explanation of the method, a description of the driving variables used in the downscaling procedure and their source, a clear statement of the validation method and performance, and some characterization of the uncertainty or reliability of the supplied data. *Fowler et al.* [2007a] note that very few downscaling studies consider hydrological impacts and those that do seldom provide any consideration of how results might enable end users to make informed, robust decisions on adaptation in the face of deep uncertainty about the future. To be able to successfully make such a decision, nonspecialist end users (e.g., the policy maker) might benefit from including social scientists with experience in translating between nonspecialists and natural scientists [*Changnon*, 2004; *Gigerenzer et al.*, 2005; *Pennesi*, 2007]. This communication process can ensure that the downscaled product can, in fact, be used as intended and is understood correctly. This paper mainly addresses the hydrological impact modeler, but sections 6 and 7, especially, provide useful information for other types of end user.

[11] In hydrological impact studies, whether using observed or simulated precipitation, assumptions about the spatial and temporal distribution of precipitation are required, and the pertinent question is what assumptions are appropriate given the nature of the specific problem being addressed. Hydrological impact analyses can have different objectives and hence focus on different components of the hydrological cycle. They are applied in differing environments (e.g., different climates, land use, and geology), and it is essential that the processes and pathways involved in a particular study area are well understood and represented in the model. Furthermore, they employ models of varying complexity and temporal resolution, depending on their purpose and model availability (e.g., empirical models on an annual base, "water balance models" on a monthly base, "conceptual lumped parameter models" on a daily base, and "process-based distributed parameter models" on an hourly or finer base [*Xu*, 1999a]). Therefore, the objective, study area characteristics, and type of model used will determine the sensitivity of the system to different precipitation character-

istics (spatial and temporal distribution) and the form of the precipitation required (e.g., continuous time series, seasonal averages, and annual extremes).

[12] It is well established that the minimum standard for any useful downscaling procedure is that the historic (observed) conditions must be reproducible [*Wood et al.*, 1997], but it is also necessary that the simulated conditions are appropriate for the particular hydrological problem being addressed. This can be achieved using a hydrological evaluation step in the downscaling procedure [*Bronstert et al.*, 2007], whereby the usefulness of the climatic data to the hydrological impact analysis is assessed. *Fowler et al.* [2007a] suggest using a sensitivity study to define the climatic variables that need to be accurately downscaled for each different impact application. This should apply not only to different variables but also to different characteristics of particular variables, i.e., different precipitation indices.

[13] In sections 2.1–2.6, a set of generic end user needs is identified, giving specific examples. The skill of the various downscaling methods to meet these needs is described in section 6.

2.1. Regional and Seasonal Needs

[14] The needs of the end user will vary regionally and seasonally as a function of socioeconomic needs and pressures, land use, and the climatological context. Depending on the particular end user, in some regions it may therefore be important to provide reliable precipitation characteristics for a particular season. In monsoonal climates, such as the Indian subcontinent [*Zehe et al.*, 2006] and West Africa [*Laux et al.*, 2008], the prediction of the onset and strength of monsoon rainfall is critical for management of water resources and agriculture. In temperate climates there is much less of a seasonal pattern in rainfall, though seasonal evaporation can significantly impact the water cycle. For instance, groundwater resources in southeast England are recharged (i.e., replenished with water originating from precipitation, infiltrating and percolating through the overlying rock) primarily in the winter months when precipitation exceeds evaporation, whereas during the summer much of the precipitation is lost to evaporation. Therefore, under current climate conditions, resource availability is considered primarily a function of winter precipitation (see, e.g., the recharge models discussed by *Ragab et al.* [1997]). *Herrera-Pantoja and Hiscock* [2008] have suggested that under climate change potential winter recharge will increase, while summer recharge will reduce (reflecting changes in both precipitation and potential evaporation). Therefore, the impact in terms of flood or drought risk will depend on a more complicated balance of these two seasonal components. Mean summer rainfall is an important control on agricultural yield, while extreme rainfall events, especially during the summer, can damage crops, reduce pesticide efficiency, erode soil, and cause flooding, all of which have a negative impact on crop yield [*Rosenzweig et al.*, 2001]. Therefore, agricultural impacts require reliable predictions of summer average and extreme rainfall conditions.

2.2. Event Intensity

[15] Many hydrological applications require continuous simulation and as such have a requirement for reliable precipitation intensities, from light to heavy events. Intensities are often characterized by their return level and return period. The return level is defined as the event magnitude which, in a stationary climate, would be expected to occur on average once within the return period. In this paper we refer to heavy precipitation as events having a return period of the order of months or a few years. The intensities of events with return periods of decades or centuries are rarely observed and probably exceed the range of observed intensities. To correctly assess such rare events, extreme value theory [e.g., *Coles*, 2001; *Katz et al.*, 2002; *Naveau et al.*, 2005] is necessary. We will refer to such events as extreme precipitation. In particular, extreme precipitation intensities are required for the design of urban drainage networks. The UK Department for Environment, Food and Rural Affairs sets a target of a 100 year return period protection for urban areas, prioritized on cost/benefit grounds [*Wheater*, 2006].

2.3. Temporal Variability and Time Scales

[16] Different temporal characteristics of precipitation are important depending on the catchment characteristics. The flooding in Boscastle, southwest England, in August 2004 was caused by 181 mm of rain which fell in 5 h [*Wheater*, 2006]. By contrast, groundwater flooding in Chalk catchments in Hampshire and Berkshire, south England, in 2001 was caused by the highest 8 month total precipitation in a record starting in 1883 [*Marsh and Dale*, 2002]. Daily precipitation totals during this period were unexceptional and not in themselves "flood producing" [*Department for Environment, Food and Rural Affairs*, 2001]. These are both examples of different types of extreme precipitation. In the case of Hampshire and Berkshire, it is necessary that the statistics of extremely long duration (up to 8 month) precipitation totals are projected reliably, while daily precipitation totals are much less important. In order to project the statistics of future flood events similar to the 2004 Boscastle flood, the downscaler should be able to supply reliable estimates of daily or even subdaily extreme precipitation. Another example where rainfall intensity over short durations is highly important is urban flooding [*Cowpertwait et al.*, 2007].

2.4. Spatial Coherence and Event Size

[17] In principle, downscaling can provide point scale, areal average, or spatially distributed precipitation fields, though the latter is challenging. Which of these is required by the end user will depend on the extent to which the spatial structure of precipitation is likely to affect the response of a system under study. For example, in the context of rainfall-runoff modeling there is evidence that spatial structure is important for small, rapidly responding catchments and for catchments that are larger than the scale of typical precipitation events [*Ngirane Katashaya and Wheater*, 1985; *Michaud and Sorooshian*, 1994; *Singh*, 1997; *Segond et al.*,

2007; *Wheater*, 2008], but other factors, such as catchment geology, may serve to damp out the effects [*Naden*, 1992].

2.5. Physical Consistency

[18] Many hydrological responses are affected by variables other than precipitation, notably evaporation and snowmelt (exceptions are short time scale responses to large rainfall events). Ignoring the coherence of these variables, i.e., treating them as though they were independent, may in some circumstances be inappropriate. In certain regions warmer winters might mean that precipitation falls as rainfall rather than snowfall, leading to lower snowmelt, lower spring-summer runoff, and hence potential drought risk [*Rosenzweig et al.*, 2001].

2.6. Downscaling for Future Climate Change

[19] End user needs for future scenarios fall into two categories: *projections* of the long-term (several decades to 100 years) trend and *predictions* of variability over the next 1 or 2 decades. The long-term trend is important for design of flood defenses and general infrastructure, as well as strategic planning regarding agriculture, water resources, and water-related hazards. The prediction of shorter-term climate variability has more immediate applications, such as predicting crop yields.

3.   HOW FAR HAVE RCMs COME?

[20] RCMs contain the same representations of atmospheric dynamical and physical processes as GCMs. They have a higher resolution (50 km or less) but cover a subglobal domain (e.g., Europe). Because of their higher resolution, RCMs typically require a reduced model time step (5 min or less) compared to GCMs (typically 30 min time step) to maintain numerical stability, although semi-Lagrangian semi-implicit RCMs such as the Canadian Regional Climate Model are able to use time steps as large as GCMs. RCMs are driven by winds, temperature, and humidity imposed at the boundaries of the domain and sea surface temperatures, supplied by the global model, which usually leads to large-scale fields in the RCM being consistent with the driving GCM.

[21] In general, the larger the domain size, the more the RCM is able to diverge from the driving model [*Jones et al.*, 1995]. The consistency of large-scale features can be further increased by forcing the large-scale circulation within the RCM domain to be in close agreement with the global model [*von Storch et al.*, 2000]. In these one-way nesting approaches there is no feedback from the RCM to the driving GCM [*Jones et al.*, 1995].

[22] Because of their higher spatial resolution, RCMs provide a better description of orographic effects, land-sea contrast, and land surface characteristics [*Jones et al.*, 1995; *Christensen and Christensen*, 2007]. They also give an improved treatment of fine-scale physical and dynamical processes and are able to generate realistic mesoscale circulation patterns which are absent from GCMs [*Buonomo et al.*, 2007]. They provide data that are coherent both spatially and temporally and across multiple climate variables, con-

sistent with the passage of weather systems. The fact that RCMs can credibly reproduce a broad range of climates around the world [*Christensen et al.*, 2007] further increases our confidence in their ability to realistically downscale future climates.

[23] Climate models need to represent processes at scales below those that they can explicitly resolve, such as radiation, convection, cloud microphysics, and land surface processes. This is done using parameterization schemes, which represent a simplification of the real world and hence lead to inherent modeling uncertainty. For example, the simulation of precipitation in an RCM is divided into a large-scale scheme, accounting for clouds and precipitation which result from atmospheric processes resolved by the models (e.g., cyclones and frontal systems), and a convection scheme describing clouds and precipitation resulting from subgrid-scale convective processes. For example, a convection scheme may model convective clouds in a grid box as a single updraft, with the amount of convection determined by the rate of uplift at the cloud base. Convective activity is restricted to a single time step, and thus, there is no memory of convection in previous time steps. In addition, there is no horizontal exchange regarding convective activity in neighboring grid boxes.

[24] There are many different RCMs currently available, for various regions, developed at different modeling centers around the world. The different RCMs produce different high-resolution scenarios for a given boundary forcing [e.g., *Buonomo et al.*, 2007], due to differences in model formulation, but also due to small-scale internal variability generated by the RCM. There has been considerable international effort recently to quantify uncertainty in regional climate change through the intercomparison of multiple RCMs, for example, the *Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects* (PRUDENCE) [*Christensen and Christensen*, 2007] and *ENSEMBLES* [*Hewitt and Griggs*, 2004; *van der Linden and Mitchell*, 2009] projects for Europe and the North American Regional Climate Change Assessment Program project (http://www.narccap.ucar.edu/) [*Mearns et al.*, 2009] for North America. The recent *Coordinated Regional Climate Downscaling Experiment* (CORDEX) initiative from the World Climate Research Program promotes running multiple RCM simulations at 50 km resolution for multiple regions.

[25] The typical grid size of RCM simulations to date has been 25 or 50 km. However, recently, a few RCM simulations with grid scales below 20 km have become available for Europe: the REMO-UBA (10 km) and the CLM (18 km) simulations of the Max Planck Institute for Meteorology and the HIRHAM (12 km) simulations of the Danish Meteorological Institute [*Dankers et al.*, 2007; *Früh et al.*, 2010; *Hollweg et al.*, 2008; *Tomassini and Jacob*, 2009]. In addition, RCMs with grid sizes of 5 km or less are being developed at several modeling centers. For example, a 5 km RCM has been developed over Japan [*Kanada et al.*, 2008]. Also, preliminary results using cloud-resolving models on climate time scales spanning small domains are becoming

available, e.g., for the Alpine region at a grid scale of 2.2 km [*Hohenegger et al.*, 2008].

3.1. Skill of RCMs to Downscale Precipitation

[26] Precipitation is one of the climate variables most sensitive to model formulation, being strongly dependent on several parameterization schemes and their interplay with the resolved model dynamics. For this variable, it has been shown that RCMs are able to contribute significant added value compared to the driving GCMs [e.g., *Durman et al.*, 2001; *Frei et al.*, 2006; *Buonomo et al.*, 2007].

[27] Compared to the driving GCM, RCMs produce an intensification of precipitation [*Durman et al.*, 2001], leading to an improved representation of the daily precipitation distribution, including extreme events [*Christensen and Christensen*, 2007]. Also, RCMs can reproduce many features of the precipitation distribution over regions of complex topography not resolved in the GCM [*Frei et al.*, 2006]. Significant biases in the simulation of mean precipitation on large scales can be inherited from the driving GCM [*Durman et al.*, 2001]. To provide a clearer assessment of the performance of an RCM, it can be driven by *reanalysis data* (see also section 5). These provide quasi-observed boundary conditions and allow RCM downscaling skill to be isolated [*Frei et al.*, 2003]. Reanalysis-driven RCM simulations not only exclude systematic biases in the large-scale climate but, in contrast to standard simulations, also are able to reproduce the actual day-to-day sequence of weather events, which allows for a more comprehensive and exact assessment of the downscaling skill. For instance, the ENSEMBLES project provides a set of *European Centre for Medium-Range Weather Forecasts 40 Year Reanalysis* (ERA40)-driven RCMs. Recent work within this project has shown that 25 km RCMs driven by ERA40 boundary conditions give a good representation of rainfall extremes over the UK, with model biases of a similar order to the differences between the 25 km ENSEMBLES and 5 km Met Office gridded observational data sets (E. Buonomo et al., manuscript in preparation, 2010).

[28] There is evidence that RCM skill in simulating the spatial pattern and temporal characteristics of precipitation increases with increasing model resolution. Improved skill may result from the improved representation of complex topography and the resolution of fine-scale dynamical and physical processes and also through the sensitivity of physical parameterization to model grid size [*Giorgi and Marinucci*, 1996]. A recent study by *Rauscher et al.* [2010] compared the downscaling skill of RCMs at 25 and 50 km grid spacings over Europe. They found improved skill at higher resolution during summer, although not in winter in some regions. However, this apparent geographic dependence in the sensitivity to model resolution may, in part, reflect regional variations in observational station density.

[29] For a given RCM, downscaling skill has been shown to depend on the region, season, intensity, and duration of the precipitation event considered. In general, RCMs show better downscaling skill in winter than in summer and for moderate compared to very heavy precipitation. We will

discuss these issues in detail in section 6, where we compare the skill of RCMs with statistical downscaling approaches.

[30] We note that in the context of climate change projections, the effects of model biases may be reduced. In particular, biases in RCM precipitation may, in part, cancel out on taking differences between the control and future scenarios. For example, *Buonomo et al.* [2007] find that two RCMs give similar precipitation changes, despite significant differences in model biases for the present day. However, recent work by *Christensen et al.* [2008] suggests that biases may not be invariant in a warming climate. In particular, models tend to show a greater warm bias in those regions that are hot and dry, while wet (dry) months tend to show a greater dry (wet) bias.

3.2. Limitations of RCMs

[31] RCMs only provide meaningful information on precipitation extremes on the scale of a few grid cells, with considerable noise on the grid cell scale [*Fowler and Ekström*, 2009]. Thus, for RCMs with a typical grid spacing of 25–50 km, this equates to providing information on scales of ~100 km (although this also depends on other factors such as season and topography). Spatial pooling, whereby daily precipitation data from neighboring grid cells are concatenated to give one long time series, is effective at improving the signal to noise ratio and thus provides improved statistics of local heavy precipitation [*Kendon et al.*, 2008]. We note, however, that this technique is only applicable where neighboring grid cells are effectively sampling from the same precipitation distribution and also that spatial dependence needs to be accounted for when assessing uncertainties. As RCMs with grid scales of less than 20 km become available [e.g., *Dankers et al.*, 2007; *Hollweg et al.*, 2008], the spatial scale on which meaningful information is provided will decrease. Nevertheless, a discrepancy will remain between the spatial scale of RCM precipitation, which should be interpreted as areal average values [*Chen and Knutson*, 2008], and site-specific data needed for many impacts studies.

[32] Linked to the spatial resolution of RCMs, there is also a minimum temporal scale on which RCMs can provide meaningful information. In particular, current RCMs show skill in capturing statistics of the daily precipitation distribution but do not well represent subdaily precipitation and the diurnal cycle of convection [*Brockhaus et al.*, 2008; *Lenderink and van Meijgaard*, 2008]. As the spatial resolution of RCMs increases and, in particular, convection-resolving scales are achieved, models give an improved representation of the diurnal cycle [*Hohenegger et al.*, 2008] and may provide meaningful information on hourly time scales. It should be noted, however, that a 30 year RCM integration just represents one possible 30 year realization of the climate and not the actual sequence of weather events. In particular, natural variability on daily to decadal time scales is a key source of uncertainty when estimating precipitation extremes.

[33] A key source of model deficiencies in the simulation of precipitation is the convective parameterization. In particular, many of the parameterization schemes used in RCMs may not be appropriate, having been developed for coarser-resolution GCMs and tropical regions [*Hohenegger et al.*, 2008]. This is particularly likely to be an issue in summer, when rainfall is predominantly convective in nature, and on subdaily time scales, when the highest precipitation intensities are usually related to convective showers [*Lenderink and van Meijgaard*, 2008].

[34] Moreover, the simulation of precipitation in RCMs is also highly sensitive to other aspects of the model formulation, including the grid resolution, the numerical scheme, and other physical parameterizations [*Fowler and Ekström*, 2009]. A number of parameters in the model physics are not well constrained, and varying these parameters within reasonable bounds leads to differences in the simulated precipitation [*Bachner et al.*, 2008; *Murphy et al.*, 2009]. RCMs developed at different modeling centers around the world use different formulations, leading to differences in downscaling skill. There is some evidence that regions and seasons showing the greatest model biases in the simulation of precipitation are also those with the greatest intermodel differences [*Frei et al.*, 2006; *Fowler et al.*, 2007b]. Past experience has shown that no single RCM is best for all climate variables and statistics considered [*Jacob et al.*, 2007; *Christensen and Christensen*, 2007], and it is not trivial to develop an objective scheme for weighting different RCMs. Indeed, it has been argued that when using multiple outputs from climate models, it is necessary to develop methodologies that exploit each model predominantly for those aspects where it performs competitively [*Leith and Chandler*, 2010].

4. METHODS TO BRIDGE THE GAP: STATISTICAL DOWNSCALING

[35] There are many statistical approaches to bridge the gap between GCM or RCM outputs and local-scale weather required to assess impacts. In the simplest form, the idea of statistical downscaling comprises some kind of mapping between a large- (or larger-) scale predictor X and the expected value of a local-scale predictand Y,

$$E(Y|X) = f(X, \beta), \tag{1}$$

where β represents a vector of unknown parameters that must be estimated to calibrate the downscaling scheme. More advanced downscaling approaches may also explicitly model variability that is not explained by the dependence of Y upon X, as a random variable η.

[36] *Wilby and Wigley* [1997] classified statistical downscaling into regression methods, weather type approaches, and stochastic *weather generators* (WGs). As an alternative classification, *Rummukainen* [1997] suggested a categorization based on the nature of the chosen predictors, which distinguished between *perfect prognosis* (PP; also referred to as "perfect prog") and MOS. To integrate these sugges-
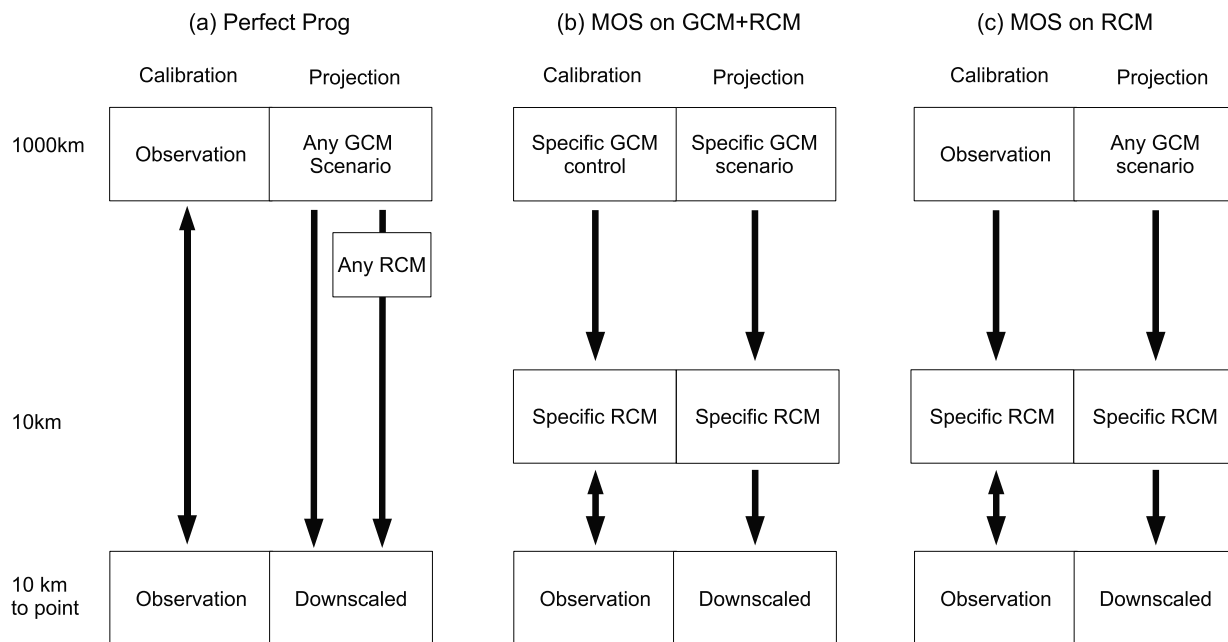
Figure 2. Statistical downscaling approaches. Two-headed arrow refers to a calibration, and regular arrow refers to a downscaling step. (a) Perfect prognosis (PP) is calibrated on large-scale and local-scale observations. For the projection, large-scale predictors are simulated by a GCM or RCM. Model output statistics (MOS) calibrates model output against observations. (b) The whole model (GCM+RCM) is corrected. Therefore, the same GCM and RCM have to be used in the projection. In this setting, the calibration is based on the distributions of model output and observations only. (c) Only the RCM output is corrected. In the projection, an arbitrary GCM can be used (this is a PP step). This setting allows for a calibration based on the whole time series of model output and observations. MOS can also be applied directly to GCMs, e.g., in a forecasting situation. Here the GCM is forced to closely follow observational data for the calibration. Conditional weather generators can be used either in a PP setting (Figure 2a) by using large-scale predictors or in a MOS setting (Figures 2b and 2c) by using change factors.

tions, we classify statistical downscaling approaches into PP, MOS, and WGs. This classification should only be seen as a means to sensibly structure sections 4.1–4.3.

[37] Classical statistical downscaling approaches, which include regression models and weather pattern–based approaches, establish a relationship between observed large-scale predictors and observed local-scale predictands (see Figure 2a). Applying these relationships to predictors from numerical models in a weather forecasting context is justified if the predictors are realistically simulated, and thus, these methods are known as perfect prognosis downscaling [e.g., *Klein et al.*, 1959; *Kalnay*, 2003; *Wilks*, 2006]. In the context of climate change projections PP methods are based on the assumption that the simulated large-scale predictors represent a physically plausible realization of the future climate. Common to these downscaling approaches, the weather sequences of the predictors and predictands can directly be related to each other event by event.

[38] PP approaches establish statistical relationships between variables at large (synoptic) scales and local scales. Physical processes on intermediate scales are usually ignored. With the increasing skill of RCMs and the availability of RCM scenarios (see section 3), alternative statistical downscaling approaches that make use of simulated mesoscale weather are becoming popular. These approaches are known

as MOS. The idea of MOS is to establish statistical relationships between variables simulated by the RCM and local-scale observations to correct RCM errors (see Figures 2b and 2c).

[39] WGs are statistical models that generate local-scale weather time series resembling the statistical properties of observed weather. In their most basic unconditional form, WGs are calibrated against observations on local scales only and are hence not downscaling approaches. Historically, the most common way of using such unconditional WGs in conjunction with climate change scenarios was to apply so-called change factors, derived from regional climate models [e.g., *Kilsby et al.*, 2007]. This approach can be considered as simple MOS (see Figure 2b). Other WGs condition their parameters on large-scale weather [*Wilks and Wilby*, 1999]. Such weather generators are thus hybrids between unconditional weather generators and PP statistical downscaling (Figure 2a).

4.1. Perfect Prognosis Statistical Downscaling

[40] This section reviews statistical downscaling approaches that establish links between observed large-scale predictors and observed local-scale predictands. The large-scale observations are often replaced by surrogate observational data such as those obtained from reanalysis products. For a dis-

cussion of problems related to observational data, refer to section 5. Many state-of-the-art PP approaches are used in a weather generator context. These specific applications will then be discussed in section 4.3.

[41] In a PP framework, equation (1) defines a relationship between a large-scale predictor X and the expected value of a local-scale predictand Y for times t, with some noise η not explained by the predictor. This is often achieved by regression-related methods, in which case the predictors X are also called covariates. Since for every observed large-scale event, there is a corresponding observed local-scale event, the calibration can be done event-wise, i.e., relating the time series of predictors and predictands to each other in sequence rather than only relating the distribution of predictors and predictands to each other.

[42] The model shown by equation (1) can be used to generate local-scale time series, by predicting $Y(t)$ from observed or simulated predictors $X(t) = (X_1(t), X_2(t), \ldots)$. Simple PP approaches disregard any residual noise term η_i, whereas state-of-the-art PP approaches explicitly provide a noise model to represent variability and extremes. The former are often called deterministic, and the latter are often called stochastic.

[43] The construction of the actual downscaling scheme can be divided into two steps: first, the selection of informative large-scale predictors and second, the development of a statistical model for the link between large-scale predictors and local-scale predictand (i.e., the $f(.)$ in equation (1)). Often, the first step also requires transformation of the raw predictors into a useful form. To avoid both overfitting or ignoring valuable information, a model selection according to statistical criteria should be carried out.

4.1.1. Predictor Selection

[44] The selection of suitable predictors is crucial when developing a statistical downscaling model. The most basic requirement for a predictor is that it is informative; that is, it has a high predictive power. Informative predictors can be identified by statistical analyses, typically by correlating possible predictors with the predictands. Various predictors representing the atmospheric circulation, humidity, and temperature have been used to downscale precipitation. According to *Charles et al.* [1999], measures of relative humidity (e.g., dew point temperature depression) are more useful than measures for specific humidity. In general, the predictor choice depends on the region and season under consideration [*Huth*, 1996, 1999; *Timbal et al.*, 2008a].

[45] In a climate change context, predictors that capture the effect of global warming [*Wilby et al.*, 1998] are necessary. In particular, measures of humidity are necessary to capture changes in the water-holding capacity of the atmosphere under global warming [*Wilby and Wigley*, 1997], whereas temperature adds little predictive power to predict long-term changes in precipitation. Suitable predictors need to be reasonably well simulated by the driving dynamical models (PP assumption), and the relationship between predictors and predictands needs to be stationary, i.e., temporally stable.

[46] These requirements are summarized in the *Statistical and Regional Dynamical Downscaling of Extremes for European Regions* (STARDEX) project [*Goodess et al.*, 2010]. A list of predictors used for precipitation downscaling is given by *Wilby and Wigley* [2000], along with a comparison of observed and simulated predictors and a stationarity assessment. A comparison of predictors for different regions is given by *Cavazos and Hewitson* [2005].

4.1.2. Predictor Transformation

[47] Raw predictors are generally high-dimensional fields of grid-based values. Moreover, the information at neighboring grid points is not independent. It is thus common to reduce the dimensionality of the predictor field and to decompose it into modes of variability.

[48] Principal component analysis (PCA) [*Preisendorfer*, 1988; *Hannachi et al.*, 2007] is the most prominent method for dimensionality reduction. It provides a set of orthogonal basis vectors (empirical orthogonal functions) allowing for a low-dimensional representation of a large fraction of the variability of the original predictor field [e.g., *Huth*, 1999]. PCA, however, does not account for any information about the predictands, and the predictor/predictand correlation might thus not be optimal. Different in this respect is canonical correlation analysis or maximum covariance analysis. These methods simultaneously seek modes of both the predictor and the predictand field (e.g., a set of rain gages), such that their temporal correlation or covariance is maximal [*Bretherton et al.*, 1992; *Huth*, 1999; *von Storch and Zwiers*, 1999; *Widmann*, 2005; *Tippett et al.*, 2008].

[49] Physically motivated transformations of the raw predictor field can provide predictors that are easily interpretable and influence the predictands in a straightforward way. For instance, *Wilby and Wigley* [2000] have used airflow strength and direction instead of the zonal and meridional components of the wind field. In a similar manner, airflow indices (strength, direction, and vorticity), derived from sea level pressure [*Jenkinson and Collison*, 1977; *Jones et al.*, 1993], have been used to downscale and model UK precipitation [*Conway and Jones*, 1998; *Maraun et al.*, 2010b]. Also, the North Atlantic Oscillation index is a transformation of the North Atlantic pressure field.

[50] Weather types (circulation patterns/regimes) can be considered as another meteorologically motivated predictor transformation. The large-scale atmospheric circulation is mapped to a usually small and discrete set of categories [*Michelangeli et al.*, 1995; *Stephenson et al.*, 2004; *Philipp et al.*, 2007]. Weather types are a straightforward way to allow for nonlinear relations between the raw predictors and predictands; the price paid is a potential loss of information due to the coarse discretization of the predictor field. Typical examples are patterns defined for geopotential heights [*Vautard*, 1990], sea level pressure [*Plaut and Simonnet*, 2001; *Philipp et al.*, 2007], or wind fields [*Moron et al.*, 2008a]. The number of types can range from small values (e.g., 4 in the case of North Atlantic circulation patterns [*Vautard*, 1990; *Plaut and Simonnet*, 2001]) to almost 30 (Großwetterlagen [*Hess and Brezowsky*, 1977]). A European cooperation in science and technology action has

been initiated to compare different weather types (http://www.cost733.org).

[51] Weather types can be defined subjectively by visually classifying synoptic situations or objectively using clustering and classification algorithms. The latter can be based on ad hoc or heuristic methods such as k means [*MacQueen*, 1967; *Plaut and Simonnet*, 2001], hierarchical clustering [*Ward*, 1963; *Casola and Wallace*, 2007], fuzzy rules [*Bárdossy et al.*, 2005], or self-organizing maps (SOMs) [*Kohonen*, 1998; *Wehrens and Buydens*, 2007; *Leloup et al.*, 2008]. Also, a variant of PCA, the T mode PCA, can be used for weather typing [e.g., *Jacobeit et al.*, 2003]. A relatively new and promising approach is model-based clustering, such as mixtures of Gaussian distributions to model the state space probability density function [*Bock*, 1996; *Fraley and Raftery*, 2002; *Vrac et al.*, 2007a; *Rust et al.*, 2010]. Many of these approaches have been compared with respect to circulation clustering by *Huth* [1996].

4.1.3. Statistical Models for PP

[52] In sections 4.1.3.1–4.1.3.6, we will describe a range of statistical models that are commonly used for PP statistical downscaling.

4.1.3.1. Linear Models

[53] One of the most widely used methods for statistical downscaling is linear regression. Here the relationship in equation (1) between the predictor X and the mean μ of the predictand Y, e.g., local-scale precipitation, is written as a linear model,

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots, \qquad (2)$$

where β_i represents the strength of the influence of X_i. In general, the predictors X explain only part of the variability of the predictands Y; thus, early downscaling approaches, which modeled the predictands according to equation (2), generally underrepresented the local-scale variance. *Karl et al.* [1990] suggested "inflating" (i.e., to scale) the modeled variance to match the observed. As noted by *von Storch* [1999], however, inflation fails to acknowledge that local-scale variation is not completely explained by the predictors; instead, it is preferable to randomize the predictand, i.e., to add an explicit noise term η, as in the methods that follow. In a standard linear regression framework, the unexplained variability η is assumed to be Gaussian distributed. Thus, the predictand Y is itself Gaussian, with mean μ and some variance representing the unexplained variability.

4.1.3.2. Generalized Linear and Additive Models

[54] The Gaussian assumption might be feasible for precipitation accumulated to annual totals. However, on shorter time scales, precipitation intensities become more and more skewed, and daily precipitation is commonly modeled with a gamma distribution [e.g., *Katz*, 1977]. A framework that extends linear regression to handle such situations is the generalized linear model (GLM) [e.g., *Dobson*, 2001]. Here the predictand Y is no longer assumed to be Gaussian distributed but may follow a wide range of distributions, e.g., a gamma distribution. The conditional mean μ of the chosen distribution, i.e., the expected value of Y, is still modeled as

a linear function of a set of predictors, but by contrast to a linear model, μ may be transformed by a link function $g(.)$ to a scale where the influence of the predictors X on μ can be considered linear:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \qquad (3)$$

Simulation of downscaled time series is achieved by drawing random numbers from the modeled distribution of Y, thus intrinsically representing the unexplained variability.

[55] In the context of precipitation downscaling, most applications of GLMs are effectively weather generators; see section 4.3. An extension of the GLM is the generalized additive model (GAM) [*Hastie and Tibshirani*, 1990], where the linear dependence is replaced by nonparametric smooth functions. The nonparametric framework generally requires more data for accurate estimation of relationships, however. GAMs have been employed, in a paleoclimate context, with large-scale data and geographical characteristics as predictors to downscale climatological monthly temperature and precipitation representative of the Last Glacial Maximum [*Vrac et al.*, 2007b]. GAMs in the context of weather generators will be discussed in section 4.3.

4.1.3.3. Vector Generalized Linear Models

[56] GLMs are capable of describing the mean of a wide class of distributions conditional on a set of predictors. In some situations, especially when studying the behavior of extreme events, one is additionally interested in the dependence of the variance or the extreme tail on a set of predictors. For instance, *Maraun et al.* [2009] and *Rust et al.* [2009] have shown that the annual cycles of location and scale parameters of monthly maxima of daily precipitation in the UK are slightly out of phase and are better modeled independently. For this purpose, vector generalized linear models (VGLMs) have been developed [*Yee and Wild*, 1996; *Yee and Stephenson*, 2007]. Instead of the conditional mean of a distribution only, a vector of parameters $p = (p_1, p_2, \ldots)$ of a distribution is predicted:

$$g_k(p_k) = \beta_{k,0} + \beta_{k,1} X_1 + \beta_{k,2} X_2 + \ldots \qquad (4)$$

The vector p could, for instance, include the mean $p_1 = \mu$ and the variance $p_2 = \sigma$ of a distribution. In extreme value statistics, these models have long been used when modeling the extreme value parameters dependent on covariates [*Coles*, 2001]. VGLMs have recently been applied to downscale precipitation occurrence in the United States [*Vrac et al.*, 2007d], and a VGLM developed to model UK precipitation extremes [*Maraun et al.*, 2010a] could easily be adopted to downscaling.

4.1.3.4. Weather Type–Based Downscaling

[57] The popular approach to condition local-scale precipitation on weather types can be thought of as a special case of a linear model. Instead of a continuous predictor field, a set of categorical weather types X_k are used to predict the mean of local precipitation:

$$\mu = \mu(X_k), \qquad (5)$$

where k gives the index of the actual weather type and $\mu(X_k)$ is the mean rainfall in this weather type. As in the case of standard linear regression, weather type approaches can, in principle, be extended to model an additional noise term η_k, such as generalized linear models and vector generalized linear models do. Weather types are mostly applied to condition weather generators; for examples, see section 4.3.

4.1.3.5. Nonlinear Regression

[58] There are also models available that aim to capture nonlinear and nonadditive relationships between the predictors and predictands. For instance, *Biau et al.* [1999] basically used a nonlinear regression to model the link between North Atlantic sea level pressure in winter and precipitation across the Iberian peninsula. Another nonlinear regression technique that has been applied in statistical downscaling is the artificial neural network (ANN). ANNs have, for instance, been used to downscale precipitation over South Africa [*Hewitson and Crane*, 1996], Japan [*Olsson et al.*, 2001] and the UK [*Haylock et al.*, 2006].

4.1.3.6. Analog Method

[59] The analog method has been developed for short-term weather forecasting [*Lorenz*, 1969]. In statistical downscaling, the large-scale weather situation is compared with the observational record. According to a selected metric (e.g., Euclidean distance), the most similar large-scale weather situation in the past is identified, and the corresponding local-scale observations $Y(\ )$ are selected as prediction for the desired local-scale weather [*Zorita and von Storch*, 1999]:

$$Y_i = Y(\text{analog}(X_i)). \qquad (6)$$

Lall and Sharma [1996] proposed not to select the most similar historic situation but to randomly choose between the k most similar ones. Potential limitations of the resampling scheme have been extensively discussed in the literature [e.g., *Young*, 1994; *Yates et al.*, 2003; *Beersma and Buishand*, 2003]. In particular, the standard analog method does not produce precipitation amounts that have not been observed in the past. Therefore, *Young* [1994] proposed a perturbation of observed values to overcome this problem. It is also pointed out that daily standard deviations of variables are underestimated because of the so-called "selection effect," a systematic underselection of certain days.

4.1.4. Model Selection

[60] In general, a range of physically plausible models exists for a given model structure (e.g., linear regression and GLM). For example, multiple variables exist that can be employed as predictors, but in many cases it is a priori not clear which of these are informative and which predictor transformation best conveys the information for the prediction. Taking too many predictors into account would lead to overfitting and would decrease the predictive power. Considering too few predictors would ignore valuable information. To objectively select a model, various statistical criteria have been developed. They are based on the likelihood of the model and assess whether an improvement in likelihood justifies an increased model complexity. Examples are likelihood ratio statistics and information criteria, such

as the Bayes and Akaike information criteria [see, e.g., *Davison*, 2003]. Once an appropriate model has been selected, a model validation (section 5) assesses the skill of this model to predict certain desired properties of the process under consideration.

4.2. Model Output Statistics

[61] As precipitation simulated in RCMs and GCMs is partly unrealistic (section 3, see also Figure 1) and represents areal means at the model resolution rather than local values, it cannot be directly used in many impact studies. The potential deviations from real precipitation make it unsuitable as a predictor in a PP context because it does not satisfy the crucial "perfect prognosis" assumption. However, despite potential errors, simulated precipitation may contain valuable information about the real precipitation.

[62] Statistical models that link simulated precipitation to local-scale, real precipitation have been developed recently for applications to RCMs, and there are also some feasibility studies for GCMs. Such methods are a form of so-called MOS models, which have been applied in numerical weather forecasting for a long time [e.g., *Glahn and Lowry*, 1972; *Klein and Glahn*, 1974; *Carter et al.*, 1989; *Kalnay*, 2003; *Wilks*, 2006]. In contrast to PP methods, the statistical relationship between predictors and predictands is calibrated not using observed predictors and predictands but using simulated predictors and observed predictands. In principle, predictors and predictands can be on the same spatial scale, in which case MOS would constitute a mere correction for a numerical model, but in most applications the predictand is local-scale precipitation, which means that MOS combines a correction and a downscaling step. The MOS corrections are specific to the numerical model for which they have been developed and cannot be used with other numerical models.

[63] Depending on the type of simulations used for MOS calibration the predictors can either be simulated precipitation time series or properties of the simulated intensity distribution (see Figure 2). Similarly, predictands can either be local precipitation series or properties of the local-scale intensity distribution. MOS can be used to transform deterministic predictors into probabilistic predictands (which is also possible with PP, see section 4.1.3). More general versions of MOS that link simulated and observed variables of different types are also conceivable [e.g., *Themeßl et al.*, 2010]; for such versions, the model structure should carefully be selected according to statistical criteria (see section 4.1.4). However, most examples in climate applications employ simulated precipitation to predict precipitation.

[64] If the MOS calibration is based on an RCM driven by a standard GCM simulation for the recent climate, in which the link to the real climate is established only via the external forcings (such as insolation and concentrations of greenhouse gases and aerosols), the observed and simulated day-to-day weather sequences are not related, and thus, MOS can only be used to link distributions of simulated and observed precipitation. The same is true when using standard GCM-simulated precipitation as predictors. In such a

setting there is a risk that differences in simulated and observed distributions, for instance, biases, are falsely attributed to model errors and thus falsely modified by the MOS approach, when they are actually caused by the random differences in the simulated and observed distribution of large-scale weather states on long time scales.

[65] If, however, the RCM is driven by an atmospheric reanalysis [*Kalnay et al.*, 1996; *Kistler et al.*, 2001; *Uppala et al.*, 2005] or GCMs forced toward observations are used, there is a direct correspondence between simulated and observed weather states, and thus, simulated and observed precipitation time series can be directly related, for instance, through regression techniques as discussed in section 4.1.3. Regional or global short-range weather forecast simulations also fall in this category as the synoptic-scale meteorological situation is usually well predicted, and thus, simulated and observed precipitation for individual days can be statistically linked. This setting does not apply for standard GCM simulations. This fact explains why MOS has first been developed in weather forecasting, has recently seen increasing popularity applied to RCMs, and is only in the development phase for GCMs.

4.2.1. Methods for MOS

[66] Most of the examples of MOS applied to RCMs are based on reanalysis-driven RCMs. The simplest method assumes that the scenario precipitation y_{i+T}^f at a time $i + T$ in the future can be represented by (observed) precipitation (or corrected RCM simulations [see *Lenderink et al.*, 2007]) $x_{\text{obs},i}^p$ at time i in the observational record, corrected by the ratio of the mean simulated future precipitation $\bar{y}_{\text{mod}}^f$ and the mean control run (or reanalysis-driven run) precipitation $\bar{x}_{\text{mod}}^p$:

$$y_{i+T}^f = x_{\text{obs},i}^p \frac{\bar{y}_{\text{mod}}^f}{\bar{x}_{\text{mod}}^p}. \qquad (7)$$

This method is sometimes misleadingly called the delta method because it was developed for temperature, where the change is additive rather than multiplicative. A mathematically similar but conceptually different approach is the so-called scaling method [e.g., *Widmann and Bretherton*, 2000; *Widmann et al.*, 2003]. Here the corrected scenario precipitation y_i^f at a time i in the future is represented by the (modeled) future scenario $x_{\text{mod},i}^f$ at time i, scaled with the ratio of the mean observed precipitation $\bar{y}_{\text{obs}}^p$ and the mean control run (or reanalysis driven) precipitation $\bar{x}_{\text{mod}}^p$:

$$y_i^f = x_{\text{mod},i}^f \frac{\bar{y}_{\text{obs}}^p}{\bar{x}_{\text{mod}}^p}. \qquad (8)$$

This method is sometimes called the direct approach [e.g., *Lenderink et al.*, 2007] and has been applied to GCMs [*Widmann and Bretherton*, 2000; *Widmann et al.*, 2003] and RCMs [e.g., *Leander and Buishand*, 2007; *Graham et al.*, 2007b; *Engen-Skaugen*, 2007]. *Schmidli et al.* [2006] further extended the approach by using a separate correction for precipitation occurrence and precipitation intensity. The aforementioned methods correct mean and variance by the same factor, such that the coefficient of variation (the ratio of the two) is unchanged.

[67] A generalized approach is quantile mapping, which considers different intensities individually [e.g., *Panofsky and Brier*, 1968; *Hay and Clark*, 2003; *Dettinger et al.*, 2004; *Wood et al.*, 1997; *Ines and Hansen*, 2006; *Déqué*, 2007; *Piani et al.*, 2009]. For the calibration period, the cumulative distribution function of simulated precipitation is adjusted to match the cumulative distribution function of observed precipitation. The mapping is usually done between empirical quantiles or quantiles of gamma distributions fitted to the observed and modeled precipitation. For modeling of values beyond the observed range, *Boé et al.* [2007] extrapolated the correction function by using a constant correction, using the correction of the highest quantile from the control simulation. This assumption, however, is, in general, not valid for the extreme tail of the precipitation distribution. A possible solution could be to adapt the mixture model by *Vrac and Naveau* [2007] (first developed by *Frigessi et al.* [2002] for temperature data) to shift between a gamma distribution for the core and an extreme value distribution for the tail.

[68] All of these methods can account for the annual cycle, for example, by applying them to individual months or seasons separately. As they calibrate only distributions but disregard any pairwise relationships between predictor and predictand, we refer to these methods as distribution-wise.

4.2.2. MOS for GCMs

[69] Most publications using MOS in a climate change context are related to correcting RCM output, while MOS for GCM-simulated precipitation is still in the development stage. MOS on GCMs might be very useful in areas where no RCM output is available. The potential usefulness of MOS corrections for GCMs was demonstrated by *Widmann et al.* [2003], who used the *National Centers for Environmental Prediction and National Center for Atmospheric Research* (*NCEP-NCAR*) *reanalysis* [*Kalnay et al.*, 1996] as an example for a GCM in which the synoptic-scale circulation is in agreement with reality because of the assimilation of meteorological data such as pressure, wind speeds, and temperature but in which the precipitation is still calculated according to model physics.

[70] The corrections for the NCEP-NCAR reanalysis model cannot be transferred to other GCMs, and thus, the development of MOS corrections for GCMs used for climate change experiments is difficult. The GCM simulations for the 20th and 21st century do not represent the real temporal evolution of large-scale weather states in the past. As a consequence only distribution-wise MOS would be possible, but it is difficult to assess whether the simulated precipitation is actually a skillful predictor. For this reason, MOS has been applied so far to nonreanalysis GCMs only in the context of seasonal prediction [*Landmann and Goddard*, 2002; *Feddersen and Andersen*, 2005; *Shongwe et al.*, 2006], where the simulated and true atmospheric circulation partly match, and in the simple form of climatology-based local debiasing of precipitation over the Alps for climate change simulations [*Schmidli et al.*, 2007].

[71] In order to provide the foundation for comprehensive MOS for future precipitation, J. Eden et al. (Reassessing the skill of GCM-simulated precipitation, submitted to *Journal of Climate*, 2010) nudged the European Center/Hamburg (ECHAM5) GCM toward the circulation and temperature in the ERA40 reanalysis and showed that the correlation of simulated and observed monthly mean precipitation over large parts of the Earth is larger than 0.8. This suggests that MOS corrections would provide precipitation estimates with small errors.

4.3. Weather Generators

[72] Weather generators, such as WGEN [*Richardson*, 1981; *Richardson and Wright*, 1984] and EARWIG [*Kilsby et al.*, 2007], are statistical models that generate random sequences of (usually several) weather variables, with statistical properties resembling those of observed weather. At the core of most weather generators is a precipitation generator, with any remaining variables usually simulated conditional on the generated precipitation.

[73] The general motivations for using weather generators are their capacity to provide synthetic series of unlimited length [*Hulme et al.*, 2002], the possibility of infilling missing values by imputation (i.e., sampling missing observations from their conditional distribution given the available observations [see *Yang et al.*, 2005]), and their computational efficiency [*Semenov et al.*, 1998] that allows for multimodel probabilistic projections or other impact assessments [*Jones et al.*, 2009]. The early weather generators (e.g., WGEN) were originally developed for providing surrogate climate time series to agricultural and hydrological models in case weather observations are too short or have quality deficiencies.

[74] In previous studies [e.g., *Fowler et al.*, 2007a; *Wilks and Wilby*, 1999; *Semenov et al.*, 1998], weather generators are distinguished on the basis of the implemented parameterization, the assumed distributions, and the suitability for particular application. Here, however, because of the importance of a proper representation of spatial rainfall [*Segond et al.*, 2007] and the limitations of spatial consistency associated with many weather generators [e.g., *Jones et al.*, 2009], we distinguish two groups of precipitation generators: single-station generators and multistation generators. In addition, weather generators have been developed that attempt to model a full precipitation field in continuous space. However, these methods have only recently been extended into a downscaling context. We will therefore present these methods as a brief outlook.

[75] Pure PP and MOS approaches do not explicitly model either temporal or spatial correlations; any structure is imposed by correlations present in the predictors. Weather generators explicitly aim to generate time series or spatial fields with the observed temporal or spatial structure.

4.3.1. Single-Station Generators
4.3.1.1. Unconditional Weather Generators

[76] Unconditional weather generators are calibrated to local observations only; that is, they do not directly use large-scale conditions from RCMs or GCMs. As discussed in section 4.1.3, at finer (e.g., daily) time scales, the distribution of precipitation tends to be strongly skewed toward low values, with a generally high number of zero values representing dry intervals. Moreover, precipitation sequences usually exhibit temporal dependence, particularly in the sequence of wet and dry intervals. Early weather generators treated single-site precipitation as a two-component process, describing precipitation occurrence and precipitation intensity separately. In the simplest case, introduced by *Gabriel and Neumann* [1962], the wet day occurrence is modeled as a two-state first-order Markov process. This structure implies that the occurrence or nonoccurrence of precipitation is only conditioned on the occurrence of precipitation on the previous day. Letting $I(t)$ denote the binary occurrence event (wet or dry) on day t, the transition probabilities $p_{ij}(t)$ are defined as

$$p_{ij}(t) = Pr(I(t) = j | I(t - 1) = i), \; i,j = 0, 1. \qquad (9)$$

The first-order Markov chain has been widely used as a simple model for rainfall occurrence [*Katz*, 1977; *Wilks*, 1998; *Wilks and Wilby*, 1999]. However, first-order models usually underrepresent long dry spells, and this has led to the use of more complex higher-order models [*Mason*, 2004; *Stern and Coe*, 1984].

[77] To model the skewed distribution of rainfall intensities, the two-parameter gamma distribution is often used [*Katz*, 1977; *Vrac et al.*, 2007d], although this is not the only choice; for example, *Wilks* [1998] uses a mixture of two exponential distributions. In the simplest daily weather generators, nonzero intensities are sampled independently for each wet day. To incorporate seasonality in these weather generators, parameters are typically estimated separately for each month or season.

[78] As an alternative to the separate modeling of precipitation occurrence and intensity, some authors have proposed modeling the two components together. The most common way of achieving this is using a power-transformed and truncated normal distribution [e.g., *Bárdossy and Plate*, 1992]. For example, if Y_t is the rainfall at time t, then a common family of transformations is

$$Y_t = \begin{cases} Z_t^{\beta} & Z_t > 0 \\ 0 & \text{otherwise} \end{cases}, \qquad (10)$$

where Z_t is a Gaussian random variable and β is a transformation parameter. *Glasbey and Nevison* [1997] and *Allcroft and Glasbey* [2003] use a more complex transformation in an attempt to reproduce the rainfall distribution more closely, but the power transformation equation (10) is by far the most widely used. More recently, innovative distributions such as those in the Tweedie family have been suggested as an alternative to transformed Gaussian variables [*Dunn*, 2004].

[79] The weather generators reviewed above take as their starting point a distribution of precipitation in each time interval. An alternative starting point is to consider explicitly the temporal structure of precipitation within a time interval: this forms the basis of cascade models, which are

used for subdaily downscaling because they are able to model correlated rain [*Olsson*, 1998; *Marani and Zanetti*, 2007]. As with other weather generators, the simplest way to incorporate seasonality is to calibrate the models separately for each month or season [e.g., *Furrer and Naveau*, 2007].

[80] A final class of precipitation generators is based on Poisson cluster processes [e.g., *Rodriguez-Iturbe et al.*, 1987, 1988; *Cowpertwait*, 1991]. This class again attempts to characterize the temporal structure of precipitation sequences but now by explicitly considering the mechanisms of precipitation generation in a simplified stochastic framework: a precipitation time series is considered as a sequence of "storms" (rain events), each consisting of a collection of "rain cells" with random intensity and duration. The models are parameterized using physically interpretable quantities such as storm arrival rate, mean cell intensity, and mean number of cells per storm and have been found to provide useful simulations of precipitation sequences at time scales down to hourly. For reviews of these models, see *Onof et al.* [2000] and *Wheater et al.* [2005].

4.3.1.2. Weather Generators and Downscaling

[81] A simple way to use unconditional weather generators for climate change scenarios is to perturb the parameters by so-called change factors [e.g., *Kilsby et al.*, 2007]: in a pair of RCM simulations, one of present day and one of the future climate, the change of the weather generator parameters (e.g., mean temperature or precipitation) from present to future is calculated for the grid box containing the location of the weather station of interest. These so-called change factors (usually differences for temperature and ratios for precipitation) are then used to modify the observed parameters for a future climate. Once these change factors are calculated, no large-scale drivers are needed to generate weather time series. A prominent example for change factor conditioned weather generators are the regional scenarios from the *UK climate projections* (UKCP09) [*Jones et al.*, 2009]. Deriving change factors for the statistical properties between the RCM control and scenario runs and applying these change factors to the statistical properties of the weather generator is mathematically equivalent to deriving a correction factor between the statistical properties of the RCM control run and the weather generator and then applying this correction factor to correct the statistical properties in the RCM scenario run. In that sense, change factor conditioned weather generators can be seen as a simple MOS (section 4.2).

[82] However, such weather generators often underestimate the interannual variability (overdispersion) and the frequency of extremes [e.g., *Katz and Parlange*, 1998] because the climatic processes influencing local weather exhibit longer-term variability, which is not captured by stationary low-order Markov models. A possible solution to the overdispersion problem is to condition specific parameters on covariates [*Katz and Parlange*, 1993; *Wilks*, 1989] controlling the low-frequency variability of the local weather, e.g., the large-scale atmospheric circulation. Such weather generators can be considered as PP (section 4.1). Besides large-scale weather predictors, transformations of lagged rainfall values, representations of seasonality, and topographic controls may also be used as covariates. Interaction terms can also be used in situations where one covariate modulates the effect of another [e.g., *Chandler*, 2005].

[83] One way to incorporate covariates into stochastic weather generators is based on GLMs (section 4.1.3). GLMs for rainfall usually use logistic regression to model the changing probability of rainfall occurrence and then consider nonzero rainfall intensities to be drawn from gamma distributions with means that are related (usually via a log link function) to linear combinations of covariates. In their simplest form, such GLMs can be regarded as extensions of the Markov Chain [see, e.g., *Coe and Stern*, 1982; *Grunwald and Jones*, 2000]. GLMs are being used increasingly for the analysis and downscaling of precipitation sequences [e.g., *Fealy and Sweeney*, 2007; *Furrer and Katz*, 2007], as are GAMS [e.g., *Hyndman and Grunwald*, 2000; *Beckmann and Buishand*, 2002; *Underwood*, 2009]. For parameter estimation of these models software routines are freely available, for example, in the stats package of the R software environment [*R Development Core Team*, 2008].

[84] Another way of incorporating large-scale information is via weather typing (see section 4.1.3). For example, *Hewitson and Crane* [2002] used SOMs to define a collection of weather states on the basis of January sea level pressure spatial fields for the northeast United States and, for each state, determined the mean and variance of daily rainfall for a gage in the center of the region. As another example of this kind of approach, *Fowler et al.* [2000] present a Poisson cluster model (see section 4.3.1.1) in which the parameters for each day are conditional on the particular weather state observed on that day.

4.3.2. Multistation Generators

[85] Multisite generation is challenging, essentially because of the need to model the joint (i.e., multivariate) distribution of precipitation simultaneously at all sites. Relatively few tractable models are available for multivariate distributions; hence, many approaches to multisite precipitation generation are based, at some level, on transformations of the multivariate Gaussian distribution. The use of transformed and truncated Gaussian distributions to model single-site precipitation has been discussed in section 4.3.1; the extension to the multisite setting is accomplished by specifying an intersite correlation structure for the Gaussian variables at each location. Generation of a multisite rainfall sequence therefore proceeds at each time instant by sampling a correlated vector of Gaussian variables (there is a standard algorithm for this [see, e.g., *Monahan*, 2001, section 11.3]) and back-transforming according to equation (10). The multisite generator of *Wilks* [1998] operates on a similar principle except that here the transformation to Gaussianity is determined by an assumption that the nonzero rainfall amounts at each site follow mixed exponential distributions.

[86] In a downscaling context, dependence on predictors can be incorporated into such models as discussed in section 4.1, either in a regression-like framework as by *Sansó and Guenni* [2000] or in conjunction with a weather typing
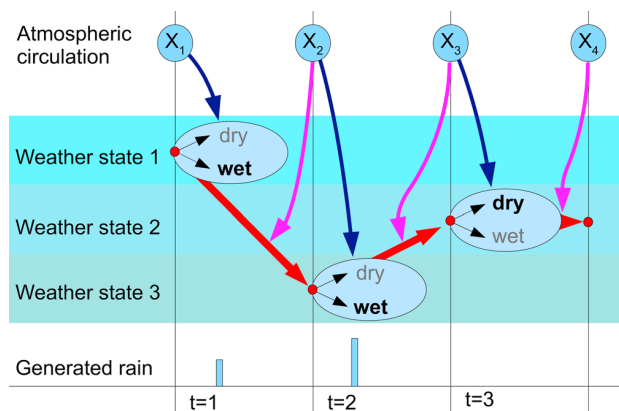
Figure 3. State-of-the-art weather generator using weather states [after *Vrac and Naveau*, 2007]. Weather time series are generated as follows: at each time step, the weather jumps into a specific weather state (red dots, spatial rain pattern); the transition probability from state to state is given by the state at the previous time step (red arrows, hidden Markov model) and depends on the large-scale atmospheric circulation at the particular time step (magenta arrows; this makes the Markov model nonhomogeneous). Furthermore, the atmospheric circulation determines the probability of having a dry or a wet day (blue arrows, logistic regression). If a wet day is generated, the actual amount of rain is generated from a distribution dependent on the weather state.

scheme whereby different sets of model parameters are used at each time instant, depending on the underlying sequence of weather states [*Bárdossy and Plate*, 1992; *Stehlík and Bárdossy*, 2002; *Ailliot et al.*, 2009; *Moron et al.*, 2008b]. In early applications of this type of methodology, weather types were typically defined solely in terms of the predictor variables.

[87] However, the more recent literature tends to focus on model variants employing so-called weather states: here precipitation patterns themselves are allowed to influence the weather state definitions, so that the resulting weather classifications can be interpreted as corresponding to distinct rainfall regimes. This includes a growing body of work based on nonhomogeneous and hidden Markov models, in which the link between weather states and predictors is probabilistic rather than deterministic [e.g., *Hughes et al.*, 1999; *Bellone et al.*, 2000; *Charles et al.*, 2004; *Vrac and Naveau*, 2007; *Vrac et al.*, 2007d]. For a schematic of such a weather generator, see Figure 3. In early versions of this type of model, the underlying weather states were considered to be entirely responsible for intersite dependence so that precipitation can be sampled independently at each site given the weather state. However, this may be inadequate at smaller spatial scales in particular, and this has led to the development of more complex models [e.g., *Ailliot et al.*, 2009; *Vrac et al.*, 2007c]. Bayesian hierarchical models also open a promising way forward here [e.g., *Cooley et al.*, 2007]. For all of the approaches outlined above, model calibration can be a challenging task that is nowadays accomplished most easily

using computationally intensive Bayesian methods (available software packages are WinBUGS [*Lunn et al.*, 2000], OpenBUGS [*Thomas et al.*, 2006] (software available at http://mathstat.helsinki.fi/openbugs), and BayesX (C. Belitz et al., BayesX—Software for Bayesian inference in structured additive regression models, version 2.0.0, available at http://www.stat.uni-muenchen.de/bayesx)). The appropriate use of such methods can require considerable technical expertise, however. Thus, there is arguably a market for simpler methods that are suitable for routine implementation.

[88] One such method uses GLMs to model precipitation sequences at individual sites (see sections 4.1.3 and 4.3.1), in conjunction with appropriately defined spatial dependence structures that enable the simulation of multisite sequences with realistic joint distributional properties. The potential for dependence between sites raises statistical issues when fitting models, however; for a review of these and straightforward ways of overcoming them, see *Chandler* [2005] and *Chandler and Bate* [2007]. The GLIMCLIM software package [*Chandler*, 2002] incorporates all of these features, as well as the possibility to include large-scale atmospheric variables as predictors and to handle missing data. These ideas are illustrated in the applications of GLMs to multisite rainfall simulation by *Yang et al.* [2005].

[89] A further approach to generate multisite weather is to apply the analog method (see section 4.1.3) in a weather generator context. For instance, *Buishand and Brandsma* [2001] proposed a nearest-neighbor resampling scheme conditioned on current large-scale atmospheric circulation patterns in order to derive local weather observations. To improve the temporal structure, some implementations of the analog method compare not only the large-scale weather situation at one point in time with historical weather but also the weather on preceding days. For a more realistic chronology of events, *Orlowsky et al.* [2008] suggested the resampling of time blocks instead of single events. Because multisite time series are sampled simultaneously, spatial correlations between stations are preserved and physically consistent. In this context no assumptions about the distribution and spatial correlations are necessary. However, in addition to the resampling of intensities, as discussed in section 4.1.3, spatial patterns are also resampled as a whole, and no unobserved patterns are generated.

[90] Most of the papers cited above focus on the generation of rainfall sequences at a daily time scale, which is considered adequate for many climate impact studies. However, in some specialized applications (for example, urban flooding studies and radio telecommunication links), data may be required at finer time scales. Models for the generation of single-site subdaily rainfall have been reviewed in section 4.3.1. At present, there are few extensions of these that provide for the generation of multisite subdaily sequences in a downscaling context. *Fowler et al.* [2005] describe one possibility in which a spatial-temporal Poisson cluster model is used as the basic multisite generator, with different parameters corresponding to distinct weather states. By contrast, *Segond et al.* [2006] suggested that subdaily sequences could be generated by first generating multisite daily sequences using one of the
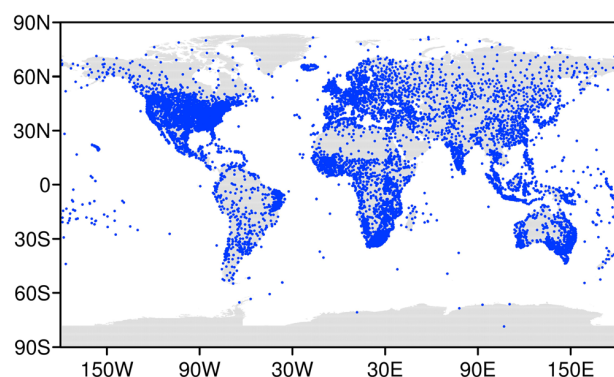
Figure 4. Rain gages used in the monthly CRU TS data set [e.g., *Mitchell and Jones*, 2005], which have been in situ for at least 40 years.

many available methods and then disaggregating the daily totals to the time scale of interest.

4.3.3. Full-Field Generators

[91] An important area of investigation in rainfall modeling is the development of models able to simulate a field of precipitation at any required fine scale and thereby provide inputs to distributed hydrological models. Currently, a number of techniques are available for such unconditional full-field simulation. They generally fall into one of the following three categories (see *Ferraris et al.* [2003] for a comparison). These are models based upon transformed Gaussian processes [*Guillot and Lebel*, 1999], point process models [*Wheater et al.*, 2005; *Cowpertwait et al.*, 2002; *Northrop*, 1998], and spatial-temporal implementations of multifractal cascade models [*Lovejoy and Schertzer*, 2006; *Marsan et al.*, 1996; *Over and Gupta*, 1996].

[92] Currently, aside from the simple scaling model of spatial rainfall fluctuations by *Perica and Foufoula-Georgiou* [1996], there are, in the literature, no implementations of such approaches for the downscaling of climate model output. But the potential of the existing methodologies is very clear. Multifractal representations of rainfall fields are well suited to downscaling implementations as they are simulated through cascade models [*Deidda*, 2000]. Similar but also allowing for nonfractal subgrid-scale structures is a downscaling algorithm based on spectral methods, for which an implementation for cloud fields already exists [*Venema et al.*, 2010]. Disaggregation methods using point process approaches [*Koutsoyiannis and Onof*, 2001] could, in principle, be extended to the spatial dimension. Transformed Gaussian processes can be conditioned by the average areal rainfall [*Onibon et al.*, 2004].

5. EVALUATION TECHNIQUES FOR DOWNSCALING METHODS

[93] Here we review methods which can be used to validate the performance of downscaling approaches to simulate specific characteristics of precipitation. These are often called metrics and are related to the end user needs, which we have introduced in section 2, and, in principle, form the basis of the discussion of downscaling skill to meet these end user needs in section 6.

[94] Any validation method ultimately relies upon the quality and quantity of observational data. Typical quality problems are inhomogeneities, outliers, and biases due to wind-induced undercatch (i.e., precipitation is underestimated by the rain gage because a nonnegligible amount of rain is blown over the gage). Inhomogeneities may induce spurious trends [e.g., *Yang et al.*, 2006] and increase uncertainty and may potentially weaken predictor/predictand relationships. Estimates of extreme events are particularly sensitive to outliers and inhomogeneities. For an appropriate signal to noise ratio, sufficiently long time series are needed, in particular, to reliably estimate extremes and infer trends. The validation of how natural variability is represented is limited by the length of observational records, typically a few decades. Furthermore, a sparse rain gage network limits the possibility for validation or may even render it impossible. For this reason, high-resolution data sets have been developed in some regions [e.g., *Haylock et al.*, 2008]. For an impression of the global rain gage network, see Figure 4. Data are particularly sparse in the high latitudes, deserts, central Asian mountain ranges, and large parts of South America.

[95] Reanalysis data, such as NCEP/NCAR [*Kalnay et al.*, 1996] or ERA40 [*Uppala et al.*, 2005], are frequently used as surrogates for observational data for validation of large-scale processes. Such data are basically interpolations of observational data based on a dynamical model (so-called data assimilation) and are therefore complete and physically consistent. However, they are subject to model biases and can significantly deviate from real weather. Precipitation is a variable which is generally not assimilated but completely generated by the parameterizations in the model, which may induce considerable biases in some locations [*Zolina et al.*, 2004]. Furthermore, the resolution of reanalysis data is too low to resolve local-scale precipitation. Therefore, NCEP/NCAR has developed the North American Regional Reanalysis [*Mesinger et al.*, 2006] that assimilates, among other variables, precipitation in order to provide a more realistic regional hydroclimatology.

[96] Reanalysis data are used to drive RCMs for validation purposes. First, this setting isolates the RCM model bias from any possible GCM bias [e.g., *Sanchez-Gomez et al.*, 2009; *Prömmel et al.*, 2009; *Vidale et al.*, 2003; *Jaeger et al.*, 2008]. Second, this setting accounts for natural variability. As discussed in the context of MOS calibration (section 4.2), the output of a GCM-driven RCM represents just one possible realization of the climate. Discrepancies might simply result from differences between the realization and the observed weather on long time scales rather than model errors. In a reanalysis-driven RCM, however, the sequence of synoptic weather in the RCM will be the same as observed. A remaining issue, though, is small-scale variability generated by the RCM that might be different from observed variability. In particular, if validating precipitation extremes, these may differ between the RCM and observations just because of natural variability.

5.1. Evaluated Metrics

[97] Depending on the application of the impact study, different metrics, or indices, of the downscaled precipitation may be of interest, including intensity metrics and temporal and spatial characteristics as well as metrics characterizing relevant physical processes.

[98] Metrics regarding precipitation intensity are mean, variance, and quantiles (i.e., return levels [*Frei et al.*, 2006; *Halenka et al.*, 2006; *May*, 2007; *Friederichs and Hense*, 2007; *Fowler and Ekström*, 2009; *Maraun et al.*, 2010a]) or parameters of the precipitation distribution. A typical metric for heavy precipitation is the 90th percentile of precipitation on wet days [*Goodess et al.*, 2010; *Haylock et al.*, 2006]. Validation of extreme precipitation intensities (e.g., 50 or 100 year return levels), which are perhaps beyond the range of observed values, should be carried out on the basis of extreme value theory [e.g., *Coles*, 2001; *Katz et al.*, 2002; *Naveau et al.*, 2005]. Studies applying this framework are still rare; for some notable exceptions in a model intercomparison context, see *Frei et al.* [2006], *Beniston et al.* [2007], and *Kendon et al.* [2008].

[99] Temporal metrics are the autocorrelation function, the annual cycle, interannual and decadal variability [*Maraun et al.*, 2010b] and trends, or metrics focusing on the precipitation occurrence such as wet day probabilities, transition probabilities (wet-wet), and the length of wet and dry spells [e.g., *May*, 2007; *Semenov et al.*, 1998]. Extremal measures for temporal metrics are, e.g., the maximum number of consecutive dry days. Spatial characteristics are spatial correlations [*Rauscher et al.*, 2010; *Achberger et al.*, 2003], cluster sizes, or spatial patterns [*Bachner et al.*, 2008].

[100] In addition, it is important to assess whether the processes leading to long-term changes in local precipitation are well captured by the models, in order for their projections of future change to be reliable [e.g., *Kendon et al.*, 2009; D. Maraun et al., manuscript in preparation, 2010]. This may be examined through the validation of process-based metrics, for example, relationships of precipitation with the large-scale circulation or with temperature [e.g., *Lenderink and van Meijgaard*, 2008; Maraun et al., manuscript in preparation, 2010] or the mechanisms of soil-precipitation feedback [*Schär et al.*, 1999].

[101] There have been several attempts to standardize indices; see the Expert Team on Climate Change Detection and Indices [e.g., *Peterson et al.*, 2001; *Nicholls and Murray*, 1999] and STARDEX project [*Goodess et al.*, 2010] for a full overview. Furthermore, a set of metrics and criteria has been defined in the ENSEMBLES project [*van der Linden and Mitchell*, 2009] in order to evaluate different aspects of the downscaling model. These aspects are (1) large-scale circulation and weather regimes, (2) temperature and precipitation mesoscale signal, (3) probability distribution functions of daily precipitation and temperature, (4) temperature and precipitation extremes, (5) temperature trends, and (6) temperature and precipitation annual cycle for RCMs and additionally of the stability of the predictor-predictand relationships for statistical downscaling (see a forthcoming special issue

in *Climate Research* (E. Kjellström et al., manuscript in preparation, 2010)).

[102] The simulated and observed characteristics need to enter the validation procedure on comparable spatial scales; that is, point observations may need to be averaged to represent areal means. Scale mismatches, typically occurring when comparing areal model outputs with point measurements, might induce representativeness errors because of the lower variance of averaged values [*Ballester and Moré*, 2007; *Tustison et al.*, 2001; *Ivanov and Palamarchuk*, 2007]. This is especially important for RCMs since not only is the grid point average smoothed over a large area, but neighboring grid points are also more correlated than in reality [*Déqué*, 2007].

5.2. Validation Measures

[103] Downscaling models (either dynamical or statistical) might be driven by GCM simulations or by observational data (often reanalysis data as surrogates). Validation for these two settings is fundamentally different. In the former case, simulated and observed weather are independent. Therefore, validation is limited to evaluating the distribution of precipitation over long periods in a particular grid box or the spatial structure of the climatology (section 5.2.2). In the latter case (called "perfect boundary conditions" in the case of dynamical downscaling), simulated and observed weather events can directly be related to each other. Here, in addition to validating the simulated distributions, validation techniques which have been developed for forecast verification can be applied. These techniques use the simulated time series as a prediction of the observed time series and assess the quality of the prediction (section 5.2.3). First, we will present measures that can be applied for the evaluation of both settings.

5.2.1. General Performance Measures

[104] Simple performance measures that can be applied to time series as well as to distributions and spatial patterns are bias, correlation, mean absolute error, and (root-) mean-square error. To visualize pattern correlation, root-mean-square error, and ratio of standard deviations simultaneously, Taylor diagrams have been introduced [*Taylor*, 2001] (see Figure 5 for an example where time series are compared). To assess the significance of discrepancies, statistical tests such as Student's t test may be carried out. For precipitation, nonparametric alternatives based on bootstrap resampling [*Efron and Tibshirani*, 1993; *Davison and Hinkley*, 1997] might prove useful [e.g., *Bachner et al.*, 2008]. A complex validation diagnostic for spatial characteristics is SAL, which considers aspects of structure (S), amplitude (A), and location (L) of precipitation in a certain region [*Wernli et al.*, 2008].

5.2.2. Measures to Validate Distributions

[105] A framework to compare the distributions of simulated and observed precipitation characteristics consists of statistical tests, such as the χ^2 test or the Kolmogorov-Smirnov test [e.g., *Semenov et al.*, 1998; *Bachner et al.*, 2008]. Another more graphical technique, especially for the
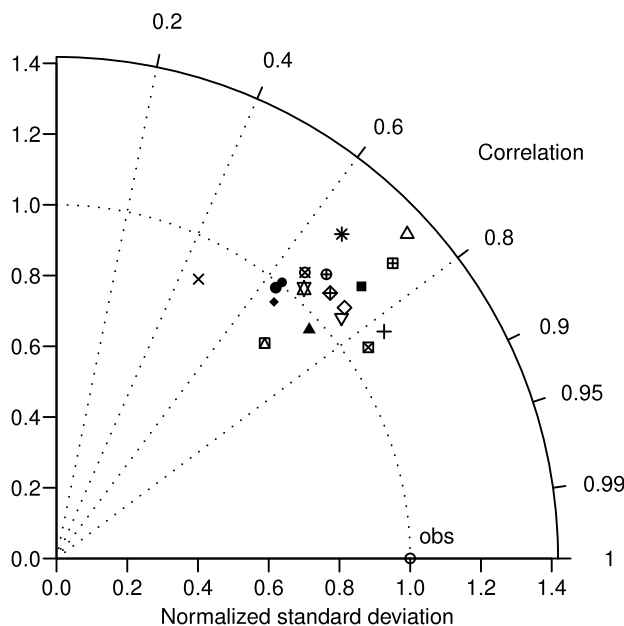
Figure 5. Taylor diagram showing the performance of 18 RCMs to simulate annual precipitation over the Thames catchment, UK. The 18 RCMs are driven with ERA40 reanalysis data, such that observed and simulated time series represent the same weather sequence and can be directly compared. The angle is given by the correlation between simulated and observed times series, and the norm is given by the ratio of simulated and observed standard deviation. The distance between the observation point (1, 1) and a model point gives the root-mean-square error between observed and modeled time series, normalized with the observed standard deviation (F. Wetterhall, unpublished data, 2009).

validation of the extreme tail, consists of (quantile-)quantile plots [e.g., *Déqué*, 2007; *Coles*, 2001], where observed and predicted quantiles are plotted against each other. For simple validation methods based on quantiles, see *Ferro et al.* [2005]. Validation of extremal properties (such as return levels) may be done parametrically, i.e., by fitting a generalized extreme value distribution to block maxima or by fitting a generalized Pareto distribution to threshold excesses [*Coles*, 2001].

5.2.3. Measures to Validate Time Series

[106] Typical measures to compare simulated binary events (e.g., wet/dry) with the actual observed outcome are hit rate, false alarm rate, frequency bias, and log odds ratio [e.g., *Jolliffe and Stephenson*, 2003; *Wilks*, 2006; *Stephenson*, 2000]. For these measures, the simulated weather sequence needs to correspond to the observed weather sequence; therefore, the downscaling model needs to be driven by observed (or surrogate) large-scale weather. Continuous events can be considered, e.g., by defining a threshold. These measures can be displayed in 2 × 2 contingency tables. A powerful tool to evaluate them graphically is the two-dimensional relative operating characteristics diagram, which displays the hit rate against the false alarm rate.

[107] Some of the downscaling approaches discussed in section 4 predict distributions rather than individual values.

Here classical measures comparing actual values are not directly applicable. Performance measures for such purpose are probability scores. The classical probability score to validate binary events, e.g., precipitation occurrence, is the Brier score [*Brier*, 1950]. To validate continuous events (e.g., precipitation amount) the (continuous) ranked probability score [*Hersbach*, 2000; *Jolliffe and Stephenson*, 2003] and the quantile verification score (see *Friederichs and Hense* [2007] and *Maraun et al.* [2010a] for examples) have been developed.

[108] Absolute values of performance measures are often not meaningful and are therefore compared with scores of reference predictions, such as the climatological mean. When developing a new downscaling method, a sensible reference prediction would be the best previously available downscaling. When assessing the predictive power of a certain predictor, a reference prediction would be the statistical model without this particular predictor. Relative measures of performance are skill scores, which can be derived from all of the aforementioned performance measures. Further skill scores are the Heidke skill score or the equitable threat score [see, e.g., *Jolliffe and Stephenson*, 2003; *Wilks*, 2006; *Stephenson*, 2000].

[109] To assess the performance of a downscaling approach on different time scales, *Maraun et al.* [2010b] applied the squared coherence [*Brockwell and Davis*, 1991]. They have investigated the performance of a statistical downscaling model on subannual, interannual, and decadal scales.

[110] To ensure robust results, any meaningful validation of time series needs to be carried out as cross validation; that is, the data used for the validation need to be independent of the data used for the model calibration. To this end, the data set is divided into a training subset and a validation subset. Splitting can be done either in time, by leaving out a certain time period for the validation, or in space, i.e., by leaving out a certain rain gage. Often, all disjunct subsets are successively left out.

5.3. Pseudorealities for Validation

[111] To overcome limitations in observational data and to better isolate different sources of error, validation in a pseudoreality has been suggested. Often, RCM validation is limited because of too sparse an observational network. Furthermore, it is difficult to isolate the contributions of the different components in the whole simulation to discrepancies between simulated and observed local variables. This problem can partly be overcome by driving the RCM with reanalysis data. However, even in this setting, errors caused by the nesting (i.e., the actual downscaling step) and the imperfection in the RCM itself cannot be discriminated. To address these issues, *Denis et al.* [2002] suggested what they call the "Big Brother Experiment": a model world is created by a high-resolution large-area RCM simulation ("Big Brother"). A perfect prognosis large-scale representation of this pseudoreality is then created by spatially filtering the high-resolution field, which is then used as boundary conditions for the same RCM but on a smaller domain ("Little Brother"). Because of the perfect prognosis construction, the

discrepancies between the Big Brother (pseudo-observed) and Little Brother (modeled) variables can exclusively be attributed to errors in the downscaling itself.

[112] Given the limited availability of long observational time series, each validation is limited by the simple fact that the time scales of interest are longer than the maximum available calibration period. This is crucial especially for statistical downscaling because stationarity issues are potentially more serious for statistical models than for models based on physical relationships. To address this potentially serious disadvantage, *Vrac et al.* [2007e] proposed a general method to validate statistical downscaling for future climate change in a model world. In addition to validating the statistical downscaling method against observations, they suggest evaluating whether the GCM driven statistical method is able to simulate realistic statistics. Furthermore, they suggest calibrating the statistical downscaling method on pseudo-observations from an RCM, driven by a GCM control run, and evaluating whether this calibrated statistical downscaling model performs well in a future scenario simulated with the same GCM and RCM.

6. SKILL OF DOWNSCALING APPROACHES TO MEET THE END USER'S NEEDS

[113] In sections 3 and 4, we have presented the state of the art in regional climate modeling and statistical downscaling. Here we discuss the extent to which the different approaches are able to meet the end user needs defined in section 2. In each case, we first present the performance of RCMs and then discuss MOS as a method of closing potential gaps between RCM output and the end user need. We then consider the skill of PP approaches and weather generators as stand-alone alternatives to dynamical downscaling.

[114] A recurring element in the discussion of downscaling skill is the difference between frontal and summertime convective precipitation. The former usually is quite homogeneous over large spatial and temporal scales, with moderate intensities. The latter is of a fine spatial-temporal structure, often with very high intensities. For an illustration, see Figure 6.

[115] In this section we first discuss the performance of downscaling approaches for different regions and seasons. We then discuss skill to simulate particular characteristics of precipitation related to the end user needs defined in section 2. We finally address the need for approaches to function in a changed climate.

6.1. Dependence of Downscaling on Region and Season

6.1.1. Regional Dependence of Downscaling Skill

[116] When assessing the potential to downscale precipitation it is important to first assess the performance of GCMs over the region of interest. For example, the GCMs in the latest Intergovernmental Panel on Climate Change report have biases in important large-scale circulation patterns like the El Niño–Southern Oscillation [e.g., *Latif et al.*, 2001; *Leloup et al.*, 2008], blocking (blocking occurs when large-

scale high-pressure systems persist in a stable state for several days, effectively "blocking" or redirecting cyclones [e.g., *Hinton et al.*, 2009] (see also Figure 7), monsoonal circulation, and tropical and extratropical cyclones [*Meehl et al.*, 2007]. These deficiencies will affect the ability to downscale precipitation locally. However, even in these areas the value added by downscaling in comparison with precipitation directly taken from GCMs is still substantial [e.g., *Christensen et al.*, 2007; *Schmidli et al.*, 2006]. Global maps of correlations between gridded observations and seasonal precipitation in a GCM (ECHAM5) in which the large-scale atmospheric states have been nudged toward a reanalysis indicate for all seasons a high skill of rescaled (i.e., MOS corrected) GCM precipitation over most parts of the Northern Hemisphere midlatitudes, relatively low skill over Africa and parts of South America, and moderate or seasonally dependent skill elsewhere (Eden et al., submitted manuscript, 2010).

[117] RCMs have been developed for many regions of the world and, in principle, are transferable to other regions. However, when transferring RCMs to very different climates, parameterizations may have to be adapted and the validation might be limited by data sparsity. Statistical downscaling can technically be performed in any part of the world, limited only by the requirement for sufficient data to calibrate and validate the model (see Figure 4).

[118] The number of downscaling studies varies regionally; a rough estimate from a search on the Web of Science (20 March 2010, keywords "Statistical Downscaling" and region, and "Dynamical Downscaling" or "Regional Climate Model" and region) indicates that most studies have been carried out for Europe and North America. There is also a difference in the relative number of studies applying dynamical and (in general, PP) statistical downscaling. For Europe and North and South America there are roughly 4 times as many studies using RCMs than PP, whereas for Africa and Asia there are over 10 times as many, and for Australia the ratio is nearly 1. These differences can partly be explained by large initiatives such as PRUDENCE or ENSEMBLES (which also provides simulations for northern Africa) and by the availability of reliable and dense observational data.

[119] An objective assessment of the downscaling skill depending on region is therefore not possible, but we will point out some general conclusions. We will mainly draw on results from the PRUDENCE [*Jacob et al.*, 2007; *Graham et al.*, 2007a] and ENSEMBLES [*van der Linden and Mitchell*, 2009] model intercomparison projects for Europe and the STARDEX project [*Goodess et al.*, 2010; *Haylock et al.*, 2006; *Schmidli et al.*, 2007] that compared several downscaling techniques in terms of their abilities to downscale high-precipitation events.

[120] Results over Europe show that the skill of RCMs is generally higher in the northern and western, wetter regions than in the drier, southern and eastern regions, but this varies from model to model [*Murphy*, 1999; *Jacob et al.*, 2007]. MOS techniques have the potential to increase the skill of RCM precipitation across Europe [e.g., *Boé et al.*, 2007;
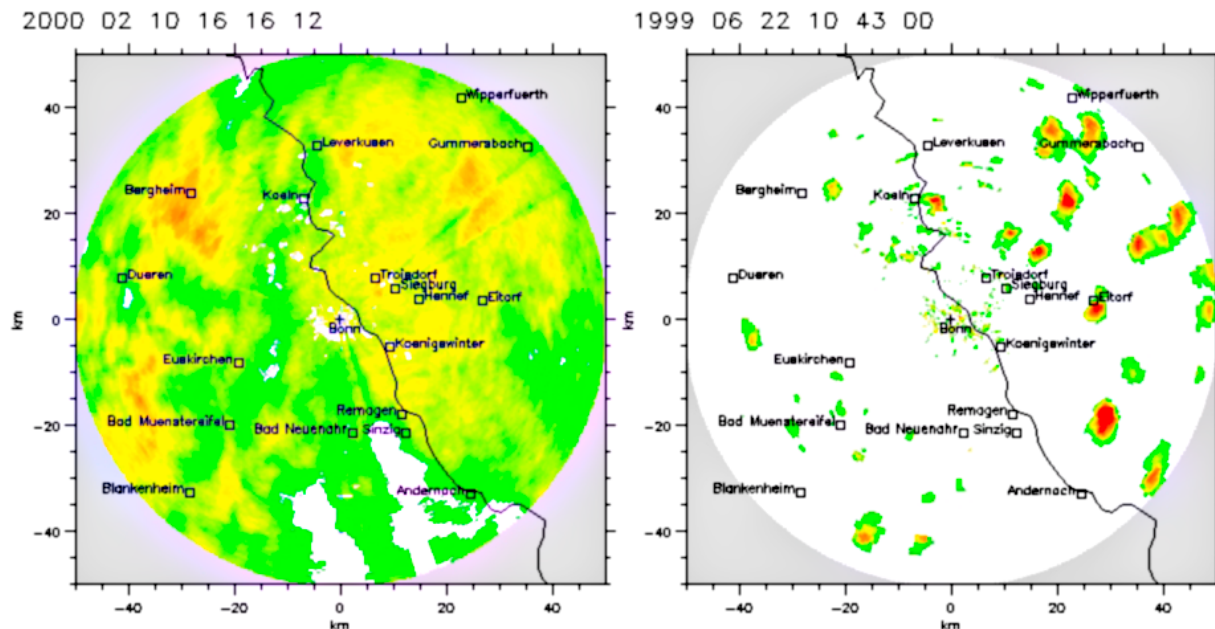
Figure 6. Radar images of the region around Bonn, Germany. White indicates no rain, green indicates light rain, and red indicates heavy rain. (left) Image from 10 February 2000, 1616 LT. A cold front crosses and causes a wide band of rain of moderate intensity. (right) Image from 22 June 1999, 1043 LT. Many small convective cells, some of high intensity, cross the Rhineland. Reprinted with kind permission from the Meteorological Institute, University of Bonn, Bonn, Germany (http://www.meteo.uni-bonn.de/forschung/gruppen/radar/index_en.htm).

Déqué et al., 2007; *Lenderink et al.*, 2007; *Yang et al.*, 2010; *Piani et al.*, 2009]. Results from the STARDEX project [*Goodess et al.*, 2010] indicate results for PP statistical downscaling that are similar to the results for RCMs: higher skill over northern Europe than over southern Europe, although the skill strongly depends on the method used.

[121] Over regions with high terrain, RCMs considerably reduce the precipitation bias compared to GCM-simulated precipitation [e.g., *Fowler et al.*, 2005; *Buonomo et al.*, 2007]. Although some of the remaining bias may be inherited from the lateral boundary conditions, a large fraction is likely to be attributable to RCM downscaling error. RCMs over the Alpine region are able to reproduce the most prominent
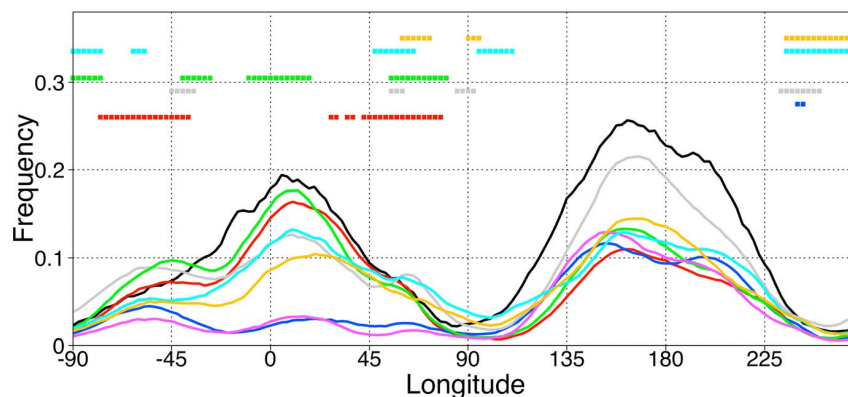

Figure 7. Mean blocking frequency. Black indicates ERA40 reanalysis, and colors indicate GCMs from the Development of a European Multi-model Ensemble System for Seasonal to Interannual Prediction (DEMETER) project. The dots indicate longitudes where the model climatology is not significantly different from the verification data. The underestimation in blocking frequency would, in turn, underestimate the occurrence of, e.g., heat waves or wet spells. Reproduced from *Palmer et al.* [2008, Figure 3].

features of the spatial pattern of precipitation, but they show a wet bias along the northwestern windward slopes and a dry bias along the southeastern leeward slopes; precipitation intensity and the frequency of heavy events are underestimated [*Frei et al.*, 2003, 2006].

[122]   *Salathe* [2003] has shown that to reduce the bias to a level that allows a reliable simulation of monthly flow in mountainous catchments, a resolution of 0.125° is needed. Studies by *Piani et al.* [2009] and *Themeßl et al.* [2010] suggest that MOS could correct bias in high-elevation regions in Europe, including the Alpine region. Following an idea by *Widmann et al.* [2003], *Schmidli et al.* [2006] applied MOS directly to ERA40 precipitation and showed the potential of directly correcting GCM simulated precipitation.

[123]   Regarding the representation of spatial precipitation variability in mountainous terrain, *Hellström et al.* [2001] and *Hanssen-Bauer et al.* [2003] concluded that PP statistical downscaling outperforms RCMs (with a spatial resolution of ~50 km). However, in a study of the Alps *Schmidli et al.* [2007] found that RCMs, in general, outperformed PP in winter but were on a par regarding the summer precipitation. With respect to the regional dependency of downscaling, the two major gaps are (1) limited representation of local-scale precipitation in areas where the large-scale modes of variability are insufficiently represented by GCMs and (2) the limited availability and/or accuracy of downscaled precipitation in data-sparse regions.

6.1.2.   Seasonal Dependence of Downscaling Skill

[124]   The assessment of 50 km resolution RCMs from the PRUDENCE project has shown that downscaling skill is generally better in winter than in summer across Europe [*Frei et al.*, 2006; *Jacob et al.*, 2007; *Fowler and Ekström*, 2009]. In winter, models tend to be too wet in northern Europe [*Christensen et al.*, 2007], and in summer, models tend to be too dry over southern and eastern Europe [*Jacob et al.*, 2007]. In the Alpine domain, biases of up to several tens of percent have been reported both for mean and for particularly extreme precipitation [e.g., *Frei et al.*, 2003, 2006]. Recent work within the ENSEMBLES project, however, has shown that 25 km RCMs driven by ERA40 boundary conditions give a good representation of rainfall extremes over the UK both in winter and in summer, indicating that higher model resolution might improve the representation of summer extremes (Buonomo et al., manuscript in preparation, 2010). By applying MOS on a seasonal basis, the representation of the annual cycle can be improved [*Boé et al.*, 2007; *Leander and Buishand*, 2007].

[125]   Like dynamical downscaling, statistical downscaling of precipitation shows greater skill in winter than in summer (for Sweden, see, e.g., *Wetterhall et al.* [2007]). Results from the STARDEX project [*Goodess et al.*, 2010; *Haylock et al.*, 2006] indicate the same seasonality in the skill to downscale heavy precipitation. However, for the UK *Maraun et al.* [2010a] found no seasonality in the skill to model the magnitude of monthly maxima of daily precipitation.

[126]   Both dynamical and statistical downscaling approaches show less skill in downscaling precipitation in summer, which may relate to the difficulty in modeling convective

precipitation. As such, providing accurate downscaled projections of precipitation in this season remains a challenge and potentially represents a remaining gap in meeting end user needs.

6.2.   Downscaling Skill to Model Precipitation Characteristics

6.2.1.   Event Intensity

[127]   Analysis of the PRUDENCE RCMs showed that models generally perform well for moderate precipitation intensities, with the greatest discrepancies for days with either light precipitation (<5 mm/d) or very heavy precipitation (>80 mm/d) [*Boberg et al.*, 2009]. Most RCMs tend to overestimate the occurrence of wet days ("drizzle effect") but underestimate heavy precipitation [*Murphy*, 1999; *Fowler et al.*, 2007b]. There is evidence that this tendency is not region specific, although to some extent, it varies between different RCMs [*Fowler et al.*, 2007b]. This tendency is also found to extend to RCMs with grid scales less than 20 km [*Früh et al.*, 2010].

[128]   Over the UK, for which there is a dense rain gage network, RCMs have been shown to realistically simulate extreme precipitation on an annual basis for return periods of up to 50 years [*Fowler et al.*, 2005; *Buonomo et al.*, 2007]. However, there is evidence that RCMs tend to underestimate extreme precipitation, in particular, where rainfall is heaviest [*Fowler et al.*, 2007b; Buonomo et al., manuscript in preparation, 2010] and for more intense events [*Buonomo et al.*, 2007]. On the 50 km grid scale model biases are highly spatially variable, ranging from −50% to +50% for 5 year return period events [*Fowler et al.*, 2005; *Buonomo et al.*, 2007], and also model dependent.

[129]   In general, high precipitation intensities occur in association with mesoscale convection or because of orographic enhancement. Thus, the tendency for RCMs to underestimate high-intensity events may be due to inadequate representation of convective processes. While over high terrain, model biases may be explained by inadequate resolution of the topography at the RCM grid scale.

[130]   The main rationale for using MOS is to correct RCM precipitation intensities, in particular, the drizzle effect and underestimation of heavy precipitation. A simple approach to correct the drizzle effect is to set all modeled precipitation values below a certain threshold to zero [e.g., *Hay and Clark*, 2003; *Schmidli et al.*, 2006; *Piani et al.*, 2009]. To improve the representation of precipitation intensities, different methods have been proposed (see section 4.2). Scaling precipitation corrects the mean and variance of precipitation by the same factor. This is generally a reasonable assumption for the core of the intensity distribution, but scaled precipitation might be biased for light and heavy precipitation. A more flexible tool is quantile mapping, which considers the whole frequency distribution of observed values. However, this approach does not explicitly consider the tail of the distribution, and extreme events might be misrepresented. A solution, which, to our knowledge, has not been applied in this context, might be the mixture model suggested by *Vrac and Naveau* [2007].
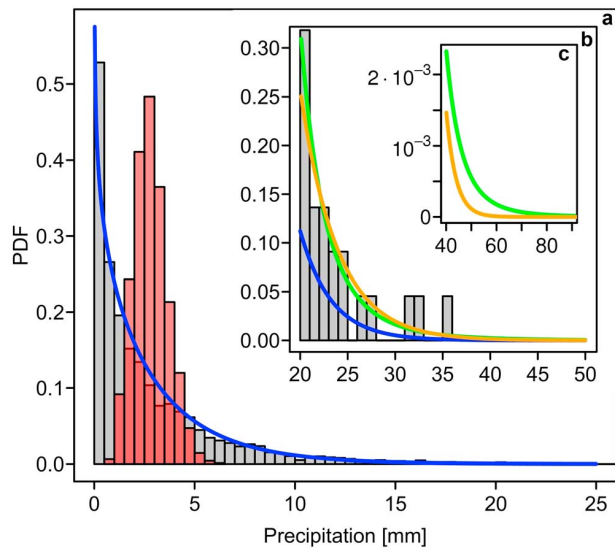
Figure 8. (a) Distribution of daily winter precipitation for Cambridge, Botanical Garden, 2 January 1898 to 31 December 2006. Grey histogram shows all observed wet day amounts. Red histogram shows amounts predicted by a simple multiple linear regression using airflow strength, direction, and vorticity as predictors. The variability is greatly underestimated and is not skewed. Blue line indicates gamma distribution, providing a suitable model for the core of the distribution. (b) The tail (>20 mm). Blue line indicates gamma distribution, which considerably underestimates the tail of the distribution. Orange line indicates exponential tail (or short/light tail), and green line indicates generalized Pareto (GP) distribution with a shape parameter of approximately 0.2 (heavy tail). For the plot, both extreme value distributions are rescaled to match the scale of the full distribution. The extreme value distributions suitably model the observed threshold exceedances, although further diagnostic plots (not shown) reveal a better fit of the GP distribution. (c) The exponential tail considerably underestimates the occurrence of extremes beyond the observed values.

[131] Early attempts at PP statistical downscaling have long been recognized to be oversimplistic in terms of representing the observed intensities: they ignored random variability (either completely or by using inflation; see section 4.1.3) and were generally unsuitable for modeling extremes (see Figure 8). *von Storch* [1999] therefore suggested to randomize the downscaled time series by adding noise realizations. *Haylock et al.* [2006] and *Goodess et al.* [2010] compared the performance of several downscaling approaches regarding the representation of different measures of precipitation intensity and found no single approach to perform systematically better than others. Not included in these intercomparison studies have been approaches based on GLMs. These models, in a simple PP setting (section 4.1) or incorporated into a stochastic weather generator (section 4.3), elegantly model the unexplained variability, commonly using a gamma distribution to generate random variability [e.g., *Yang et al.*, 2005; *Furrer and Naveau*, 2007] (see also Figure 8).

[132] Evaluation studies so far have focused on moderately heavy rain. For example, in their study on heavy precipitation over the United Kingdom, *Haylock et al.* [2006] choose the 90th percentile on wet days, roughly corresponding to subannual return levels. For many impact studies and design settings, however, much higher return levels of the order of decades or centuries are relevant. In general, there is no guarantee that statistical models for the core of the distribution will provide an adequate representation of extremes [*Wilks and Wilby*, 1999] (see also Figure 8). The distribution of precipitation tends to be heavy tailed [*Katz*, 1977], and statistical downscaling schemes that do not account for this are likely to be heavily biased for high extremes. Recently, statistical models based on extreme value theory have been developed for precipitation [*Maraun et al.*, 2010a, 2010b], which can easily be extended for downscaling. However, as these approaches only model the extreme tail but not the core of moderate precipitation, they are limited in their applicability. *Yang et al.* [2005] demonstrated that it is possible to obtain heavy-tailed distributions by incorporating nonlinear dependence structures into GLMs based on gamma distributions; however, at present, the conditions under which heavy-tailed distributions can be obtained from this kind of model are poorly understood. A possible alternative solution could be mixture models such as the one suggested by *Vrac and Naveau* [2007]. Here the authors combine a gamma distribution to model moderate precipitation and a generalized Pareto distribution to model extremes. The performance of these approaches has not yet been compared with standard statistical downscaling schemes. To summarize, downscaling has the potential to reliably simulate event intensities, in particular, when correcting RCM output by MOS or using PP methods to predict full distributions.

6.2.2. Temporal Variability and Time Scales

[133] Studies for the UK have shown that the extent to which model biases increase or decrease for longer-duration events depends on the region and the RCM [*Fowler et al.*, 2007b; *Fowler and Ekström*, 2009]. For Hadley Centre RCMs, *Buonomo et al.* [2007] find greater biases for longer-duration (30 day accumulation) extremes compared to 1 day events in regions of heavy precipitation but quite different behavior where long-duration extremes are strongly influenced by lighter precipitation events.

[134] There are relatively few studies to date examining RCM skill in simulating subdaily precipitation. A recent study by *Lenderink and van Meijgaard* [2008], however, shows deficiencies in the ability of the 25 km RACMO RCM to capture hourly precipitation for temperatures above 20°C. This deficiency is likely to be particularly important in summer months where convective processes may dominate and temperatures are high. *Hohenegger et al.* [2008] have shown that very high resolution (grid scale ≤5 km) climate modeling improves the diurnal cycle of convection. The representation of short-duration precipitation extremes is also significantly improved at high resolution [*Wakazuki et al.*, 2008]. These resolutions are now common practice in numerical weather prediction [*Roberts and Lean*, 2008]
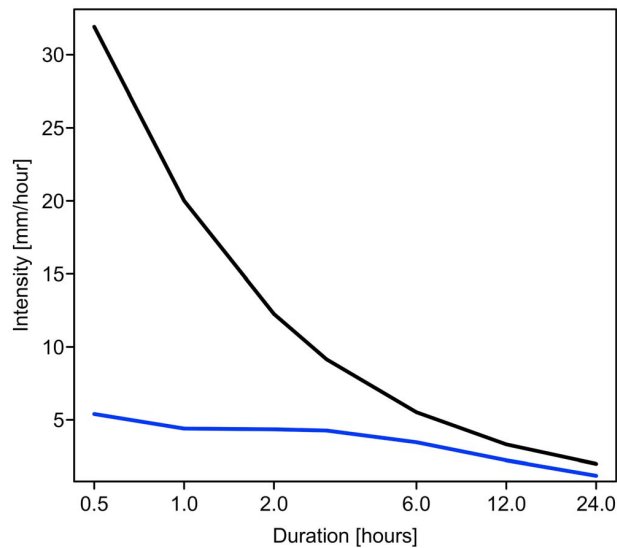
Figure 9. Intensity-duration plot: 5 year return period of precipitation intensities for subdaily durations, from Stockholm, Sweden. Black line indicates observed data, and blue line indicates regional climate model RCA driven by ERA40 reanalysis data.

but are computationally very expensive and thus are currently limited to either short time periods or small spatial domains. For an illustration of model deficiencies in simulating subdaily precipitation, see Figure 9.

[135] As MOS is designed to correct precipitation intensities, it does not improve the temporal structure. Even the adjustment of the number of wet days does not guarantee an improved representation of the lengths of dry and wet spells. However, the representation of seasonality can be improved by applying MOS to different seasons [*Boé et al.*, 2007] or months separately or to even shorter parts of the year (e.g., 5 day periods [*Leander and Buishand*, 2007]). If precipitation sums over longer time periods such as monthly totals are of interest, MOS could be applied to time aggregated precipitation.

[136] In PP statistical downscaling the temporal structure is not explicitly modeled. However, the large-scale predictors impose their time structure on the local-scale precipitation. For instance, *Haylock et al.* [2006] and *Goodess et al.* [2010] have shown that the maximum number of consecutive dry days is generally better modeled than the intensity of heavy rainfall, indicating that a reasonable fraction of the time dependency is captured by the predictors. *Maraun et al.* [2010b] found that predictors representing the large-scale atmospheric circulation explain a significant fraction of the monthly, interannual, and decadal variability of high precipitation intensities. Weather generators explicitly model the short-term day-to-day variability (see section 4.3) but require large-scale predictors to correctly simulate long-term variability [*Wilks and Wilby*, 1999].

[137] Weather generators, such as Poisson cluster models, can provide subdaily precipitation. They can, in principle, be implemented without subdaily data but perform better

when calibrated against subdaily data [*Cowpertwait et al.*, 1996]. For a reasonable calibration, at least 10 years of data are required; to calibrate the models for subdaily extreme precipitation, even longer time series are required. Furthermore, they are generally conditioned on daily RCM change factors and thus cannot provide subdaily information on climate change [*Jones et al.*, 2009].

[138] In summary, deficiencies remain in the ability of downscaling methods to generate local precipitation time series with the correct temporal variability. Many of these deficiencies are inherited from the driving GCMs, with deficiencies in the representation of blocking and tropical modes of variability [e.g., *Ringer et al.*, 2006] (see also section 6.1). RCMs and PP weather generators can "add value" in terms of the representation of short temporal variability.

6.2.3.   Spatial Coherence and Event Size

[139] In terms of spatial variability, two potential problems need to be considered: misrepresentation of event size, structure, and spatial coherence, e.g., by overestimating the extent of convective cells, and misplacement of precipitation events, e.g., due to orographic effects.

[140] RCMs tend to overestimate the spatial coherence of precipitation events. As discussed in section 3, convective events are difficult to model, and therefore, these events are often too low in intensity and extend over too large an area. This problem might be solved in the future with higher resolution and improved numerical schemes. Large-scale frontal precipitation is generally well simulated by RCMs, although the coarse orography, especially in mountainous regions, can cause erroneous spatial distributions of precipitation [*Frei et al.*, 2003]. In addition to improving subdaily precipitation representation, very high resolution climate modeling ensures more accurate localization of rainfall maxima over regions of complex topography [*Hohenegger et al.*, 2008].

[141] Most MOS approaches are not designed for correcting errors in spatial correlations since the predictand still inherits much of the spatial correlation structure of the simulated precipitation [*Boé et al.*, 2007]. However, *Widmann et al.* [2003] suggested a nonlocal MOS: they applied singular value decomposition to derive coupled spatial patterns of simulated and observed precipitation. These patterns can have a different structure with high values over different locations, such that this approach, in principle, can correct unrealistic aspects in the location and spatial structure of the simulated precipitation, which may be caused, for instance, by an unrealistic topography in a numerical model.

[142] Within individual grid boxes, *He et al.* [2009] have attempted to account for subgrid orography by distributing the simulated precipitation according to observed patterns. There are examples of MOS weather generators (using change factors derived from RCMs to represent climate change) that have been extended to a high-resolution grid (e.g., 5 km [*Jones et al.*, 2009]), but these are run independently for each grid point.

[143] Standard PP statistical downscaling is facing a dilemma: in a "deterministic" context, i.e., without explicitly

adding noise to the downscaled variables, the predictors impose a strong spatial coherence. Yet randomization in the form of adding uncorrelated noise might weaken the spatial coherence too much. The same holds for weather generators based on weather states, which themselves induce intersite correlations. At large spatial scales, it may be reasonable to consider that all of the intersite dependence is captured. Often, however, and particularly at smaller spatial scales, the induced dependence is weaker than that found in observations. A way out is the explicit modeling of spatial dependence, i.e., using multisite weather generators (section 4.3.2) or full-field weather generators (section 4.3.3). The analog method, either in a simple PP setting (section 4.1.3) or extended to a weather generator (section 4.3.2), provides an easy way to simulate spatially coherent and realistic fields. However, this method cannot simulate unobserved weather patterns which might emerge because of changes in the atmospheric circulation. Therefore, its use for climate change projections is limited, especially in simulating fields of extreme precipitation. All PP methods, including PP weather generators, can, in principle, correctly represent orographic influences as their calibration intrinsically accounts for the interplay between the large-scale atmospheric circulation and the orography, such as lee and rain shadow effects.

[144] The representation of spatial variability is limited by the density of the rain gage network. Still unresolved is the issue of full-field precipitation, i.e., the provision of downscaled precipitation between rain gages. Often, this problem is addressed by interpolation from neighboring sites. However, such techniques are a form of smoothing that leads to underestimation of rainfall variability, especially on short time scales and for extremes [e.g., *Hofstra et al.*, 2008]. This is particularly serious in mountain areas, where the relationships between orography and precipitation are very complex and the rain gage network is generally sparse compared to the high spatial variability (for a notable exception, see *Frei and Schär* [1998]).

6.2.4. Physical Consistency

[145] RCMs model the full atmospheric state and therefore intrinsically address physical coherence. However, small temperature biases might lead to considerable biases in impact models when temperature and precipitation are required. *Yang et al.* [2010] showed that a MOS correction of temperature and precipitation bias could improve the simulation of river discharge in spring. In general, however, it should be noted that MOS may disrupt internal consistency between weather variables, especially between temperature and precipitation.

[146] Pure PP statistical downscaling does not, in general, explicitly model physical coherence between variables unless, for example, large-scale temperature is used as predictor for precipitation [e.g., *Chun et al.*, 1999]. This is, however, problematic since high summer temperatures may be a consequence of dry conditions (i.e., due to clear skies) or a cause of convective wet conditions, so the correlations are difficult to interpret [*Wilby and Wigley*, 2000]. Unlike other PP approaches, the analog method intrinsically captures physical coherence.

[147] Most weather generators attempt to model the relationships between relevant variables, mostly by regressing other variables on the generated precipitation [*Kilsby et al.*, 2007; *Jones et al.*, 2009]. An advancement of this approach based on GLMs was developed by *Furrer and Naveau* [2007]. In these methods, the other variables are derived from the downscaled precipitation without referencing the actual variable (e.g., temperature) in the driving GCM.

6.3. Downscaling for Future Climate Change

[148] Downscaling of climate change scenarios requires the chosen methodology to function in a perturbed climate, i.e., under conditions different from those for which it was developed [*Huth and Kyselý*, 2000]. Therefore, skill for the present-day climate, although necessary, may not be a sufficient indicator of skill for the future climate [e.g., *Charles et al.*, 1999; *Christensen and Christensen*, 2007]. It is also difficult to objectively quantify model skill as different models perform better for different variables and processes.

[149] When discussing skill to downscale future climate scenarios, two points affecting the skill have to be addressed, both for dynamical and statistical downscaling. First, stationarity of the physical and statistical relationships has to be established, and second, the driving GCM simulation needs to be informative for the downscaled variable. Closely connected with downscaling of future scenarios is the question of predictability and uncertainty. Often, model consensus is taken as evidence for robust skill. This assumption will be critically reviewed.

6.3.1. Model Consensus as a Measure of Skill

[150] Model consensus does not imply reliability since there may be missing processes or deficiencies common to all models. An understanding of the underlying processes and mechanisms of change, and their evaluation in models, is key to assessing reliability. Modeling, theory, and observational studies suggest that increases in extreme precipitation are reliable, at least on large scales, since they are dominated by increases in atmospheric moisture with warming [*Allen and Ingram*, 2002; *Allan and Soden*, 2008; *Kendon et al.*, 2009]. However, for local precipitation extremes, small-scale dynamics of clouds and the subcloud layer and cloud microphysics as well as changes in precipitable water may play an important role [*Lenderink and van Meijgaard*, 2008]. These small-scale processes are not well represented in current RCMs, as evident from deficiencies in the simulation of high precipitation intensities for the present-day climate. The same holds for statistical downscaling as predictors used in different approaches are often similar, if not identical, and all approaches ultimately rely on a small number of driving GCMs.

[151] Some degree of confidence might be gained from comparing dynamical and statistical downscaling techniques [e.g., *Murphy*, 1999; *Haylock et al.*, 2006]. For other model comparison examples, see *Semenov et al.* [1998], *Zorita and von Storch* [1999], *Schmidli et al.* [2007], and *Timbal et al.* [2008b]. In fact, dynamical downscaling and statistical

downscaling can be used to mutually validate one another. For instance, an RCM pseudoreality can be used to validate statistical downscaling approaches (section 5.3) [*Vrac et al.*, 2007e], and statistical downscaling can be used to validate physical relationships in the RCM (section 5.1) [*Kendon et al.*, 2009; Maraun et al., manuscript in preparation, 2010].

6.3.2. Stationarity

[152] In the case of dynamical downscaling, assumptions need to be made for RCM parameterizations to be valid in a perturbed climate. This may be a significant issue for RCMs that have been developed for a specific region. For RCMs that have been shown to perform well for multiple regions, there is greater confidence in the applicability of the parameterization schemes in future climates [*Christensen et al.*, 2007].

[153] When correcting the RCM output, the stationarity issue might become more serious. Most MOS methods described in section 4.2 correct the distribution of modeled precipitation, estimated over a long time interval. However, this distribution is, in fact, a mixture of various other distributions, depending on the different weather conditions. Since the relative frequency of different weather conditions might change in a future climate, the resulting mixed distribution might also change, such that the correction function is potentially not valid under climate change. For instance, *Christensen et al.* [2008] suggest that biases may not be invariant in a warming climate. This argument holds, in particular, for methods that scale observed or control run precipitation, which do not account for possible dynamic changes in temporal variability, for instance, in the frequency of circulation patterns [e.g., *Lenderink et al.*, 2007].

[154] The stationarity issue is also significant for PP statistical downscaling. The more heuristic and less physical the predictor/predictand relationship, the less confident one can be that the relationship might remain stable under climate change. A way to gage the transferability of statistical relationships into the future is to use a sensitivity analysis when calibrating a statistical downscaling method [*Frías et al.*, 2006]. One way is to build the model on data from the coldest (driest) years and then validate it on data from the warmest (wettest) years, thus testing the scheme on two different climate situations. The model can also be tested against extreme years in order to test the stability [*Wilby*, 1994]. If the time series used for calibration are long enough, it is reasonable to believe that they are representative of those situations that will be more frequent in a future climate [*Zorita and von Storch*, 1999]. Confidence in the approach is highest if it can model such situations and if the range of variability of the large-scale variable in a future climate is of the same order as today.

[155] Sometimes, nonstationarity in the relationships is only an artifact because the chosen predictors do not convey enough information about long-term variability. *Wilby and Wigley* [1997] showed that certain changes in the relationship between weather types and precipitation in the UK could be explained by a modulating effect of the central England temperature. Therefore, it is necessary to identify all predictors informative for climate change and to incorporate them in a multivariate approach. A similar issue is discussed by *Wilby et al.* [2004] regarding nondynamical shifts of predictors due to climate change. Spurious effects on rainfall could be corrected by subtracting the mean shift from the predictors.

6.3.3. Capturing Climate Change

[156] For reliable simulations of future climate, the mechanisms of future change in precipitation need to be represented [e.g., *Kendon et al.*, 2009]. Thus, it is important that the processes leading to long-term changes in local precipitation, such as relationships of precipitation with the large-scale circulation or with temperature [e.g., *Lenderink and van Meijgaard*, 2008] or the mechanisms of soil-precipitation feedback [*Schär et al.*, 1999], are well captured by the models.

[157] Biases in the GCM-simulated large-scale atmospheric circulation might considerably bias the RCM simulation. For instance, *Leander et al.* [2008] noted that the representation of extreme precipitation events is potentially sensitive to the driving GCM, limiting the overall possibility to correctly downscale high-intensity rainfall.

[158] A similar argument holds for MOS applications, only on smaller scales. Any correction yields meaningful results only if the temporal variability or the long-term changes in the simulated precipitation are good predictors for the changes in the real world. In the case of MOS calibrated on the basis of reanalysis-driven RCMs or GCMs nudged toward reanalyses this can be assessed directly by comparing the simulated and observed changes in the past, whereas in control run calibrated setups that allow only distribution-wise MOS it is difficult to judge whether the application of MOS corrections is justified. Where the simulated precipitation has simply no skill the application of distribution-wise MOS would not be justified, even if the corrected and observed precipitation intensity distributions could be brought into perfect agreement.

[159] In PP statistical downscaling, the choice of predictors is crucial to capture climate change (see section 4.1). Predictors that are informative on relatively short time scales might not capture long-term variability and, in particular, trends induced by global warming. PP statistical downscaling approaches also rely on the skill of the driving GCM to correctly simulate the relevant predictors. A predictor that is characterized as informative might be of little use if it cannot be assumed to be reliably modeled in the GCM/RCM (in particular, moisture-related quantities are generally considered problematic [*Cavazos and Hewitson*, 2005]).

6.3.4. Uncertainty and Predictability

[160] An important aspect in assessing predictability is the quantification of the total uncertainty of the downscaled result and the sources that contribute to it. For predictability, the main sources of uncertainty are model formulation, which includes the numerical schemes, parameterizations, and resolution; uncertainty in anthropogenic climate forcing factors; and natural variability [*Palmer*, 1999; *Hawkins and Sutton*, 2009], which includes internal variability of the

chaotic climate system dependent on initial conditions and natural forced variability due to, e.g., solar forcing.

[161] The range of uncertainty due to model formulation in general, parameterizations in particular, and natural variability can be assessed by ensemble simulations based on different GCMs and RCMs (multimodel ensembles), perturbed parameterizations (perturbed physics ensembles), and different initial conditions. Notable initiatives are the PRUDENCE, ENSEMBLES, and CORDEX projects, which study the uncertainty due to structural errors of different GCMs and/or RCMs. For the development of the probabilistic UKCP09 national climate change projections, a large GCM ensemble with perturbed physics parameterizations was used to drive the Hadley Centre regional climate model HadRM3 [*Murphy et al.*, 2009].

[162] The relative roles of these different sources of uncertainty depend on the time scales under consideration. On decadal time scales, the climate change signal is small compared to natural variability, such that uncertainty caused by initial conditions and natural forcing dominates. Memory, and thus predictability, of natural variability on decadal time scales is generated by the oceans. However, because of limited availability of (deep) ocean data to initialize the prediction, predictability is, in practice, limited. Research on decadal climate predictions is just emerging [e.g., *Collins et al.*, 2006; *Smith et al.*, 2007; *Keenlyside et al.*, 2008], and no regional climate predictions on decadal scales exist. As natural decadal variability increases with decreasing spatial scale, the extent to which regional decadal predictions are possible is largely unknown.

[163] On longer time scales, the signal to noise ratio between climate change signal and natural variability increases, and uncertainty due to model formulation becomes dominant. For instance, results from the PRUDENCE project suggest that GCM uncertainty dominates in the case of changes in seasonal mean climate [*Rowell*, 2006; *Déqué et al.*, 2007], and variations in RCM formulation are important at fine scales and for changes in precipitation extremes, particularly in summer [*Frei et al.*, 2006]. However, recent studies [*Kendon et al.*, 2009; *Kendon et al.*, 2010] suggest a still dominant role of natural variability for summertime precipitation and precipitation extremes, such that a single 30 year climate projection is not robust. It should be noted that a climate projection represents just one possible realization of the future climate, conditional on a given scenario of natural and anthropogenic forcing.

7. CONCLUSIONS AND OUTLOOK

[164] Reliable downscaling for precipitation is needed, independent of region and season. Depending on the application, generic needs are the correct representation of (1) intensities, (2) temporal variability, (3) spatial variability, and (4) consistency between different local-scale variables, and these are required for future scenarios.

[165] To meet these specific needs, there have been considerable efforts to further develop dynamical and statistical downscaling. We reviewed several recent developments in statistical downscaling, which have not yet received much attention in the climate community. These developments focus on capturing intensities, especially extremes, and the representation of spatial-temporal variability. However, there are still major gaps which currently are not resolved by downscaling:

[166] 1. Downscaling in regions with sparse data is still highly uncertain, mainly in remote areas or developing countries (see Figure 4). RCMs can, in principle, be set up in these regions, but they may not correctly represent region-specific processes. With data sparsity, their validation is limited. Statistical downscaling is even more restricted in such regions, especially to assess precipitation extremes and spatial variability. This problem will make it harder for end users operating in these countries to make optimal planning decisions in all areas, e.g., from water resources, to flood risk management, to urban design, to agricultural activities.

[167] 2. The performance of both dynamical and statistical downscaling schemes is currently better for synoptic and frontal systems than for convective precipitation. End users that are adversely affected by this limitation would be the flood risk managers in arid regions subject to flash flooding or in temperate regions subject to summer flooding. In these cases, improvements in the representation of heavy, localized convective precipitation are needed (see also Figure 6).

[168] 3. Representation of subdaily rainfall is still poor, especially for extremes, both by RCMs and by statistical downscaling. Furthermore, few statistical models are currently available that attempt to capture subdaily information on climate change. The end user community most seriously impacted by this limitation consists of urban planners since runoff generation from largely impermeable urban areas occurs rapidly and is highly sensitive to the fine temporal scale distribution of precipitation (see also Figures 6 and 9).

[169] 4. Downscaling to a fully distributed spatial field at scales smaller than RCM grid size is still unresolved. Full-field weather generators are under development but have not yet been implemented for downscaling. One end user affected by this limitation is the hydrological impact modeler using a spatially distributed model for areas sensitive to the spatial distribution of precipitation, such as small catchments or catchments with an impermeable underlying geology (see also Figure 6).

[170] 5. Changes in small-scale processes (on sub–RCM grid scales) and their feedback on the large scale are not adequately captured in projections of precipitation change. Currently, it is difficult to identify how significant this shortcoming may be and, indeed, which end users may be more affected. For instance, the importance is likely to be seasonally and regionally dependent. This shortcoming remains a challenge for climate modelers.

[171] 6. All downscaling approaches inherit errors in the representation of temporal variability from the driving GCM. Examples are blocking over Europe (see Figure 7) and tropical modes of variability. The former example is especially relevant for agriculture as blocking strongly influences the length of dry spells. Summer drought often

comes along with heat waves and thus affects health authorities as well.

[172] These gaps are caused by poor data availability, process understanding, and quality of the GCMs and limitations with the downscaling procedure itself. In the following, we lay out directions in research to address the remaining gaps.

[173] In regions with sparse rain gage networks, the installation of new gages will improve the situation in the long run. However, in many regions networks do exist, but the data have not been made available by the responsible institutions such as national weather services. Here efforts should be undertaken to make these data readily available and to assemble high-resolution gridded data sets as input for hydrological models (where these require spatially averaged rainfall inputs) or for climate model validation; see *Haylock et al.* [2008] for an example in Europe. Furthermore, digitizing handwritten reports can help to extend the daily database [*Moberg and Jones*, 2005]. Especially in urban areas, denser networks of subdaily data need to be set up.

[174] The quality of GCM climate projections is constantly improving, and the latest generation of models shows better representation of climate variability [*Shaffrey et al.*, 2009].

[175] In terms of future RCM development there are two competing strands. The first concentrates on developing multimodel ensemble systems, including multiple RCMs as well as multiple GCMs, to quantify modeling uncertainty. Performance-based weighting of different RCMs could add value [e.g., *Fowler and Ekström*, 2009], although model weighting is a nontrivial task. The second aims to improve the simulation of local processes through the development of RCMs of increasing resolution (which includes improvements in the parameterizations). This is expected to lead to improvements not only in terms of the spatial scale on which meaningful information is provided but also the accuracy of subdaily precipitation.

[176] A key feature of statistical downscaling is the ability to generate complete distributions. They can be used to randomize the downscaled result and thus better represent local variability and extremes. These techniques should be used by default, in particular, when downscaling of extremes is required. However, these methods mostly involve a considerable statistical and computational knowledge; therefore, especially for multistation downscaling, accessible implementations suitable for routine use by researchers and practitioners are needed.

[177] Because of the still limited understanding of multivariate extreme value statistics [e.g., *Coles*, 2001], multistation weather generators have not yet been extended to explicitly capture extremes. The characterization and modeling of spatial extremes is currently an active area of statistical research.

[178] A promising direction of research is the application of MOS to correct climate model output. Currently, the proposed methods almost exclusively use modeled precipitation as predictors and mostly correct distributions only.

None of the approaches explicitly account for extremes. It has been shown that MOS could be applied to directly correct GCM simulations [*Widmann et al.*, 2003]. This approach might prove useful for regions where no RCM simulations are available.

[179] We presented the potential usefulness of full-field weather generators for hydrological modeling. The complexity of existing full-field spatial-temporal models may suggest that it is not currently a realistic research aim. However, rather than add complexity to a spatial-temporal model, conditioning upon climate model outputs may provide useful information for the difficult task of representing advection. Research into linking parameter models with climatological information should be seen as a first step in this direction.

[180] Providing probabilistic climate projections is a key challenge. Initiatives such as PRUDENCE, ENSEMBLES, CORDEX, UKCP09, and *Climate Prediction Net* provide a first step toward probabilistic climate projections. They have generated a wealth of information about uncertainty in model formulation, but they still do not cover the full plausible range of model uncertainty and do not sufficiently address uncertainty due to natural variability. In particular, on decadal time scales, probabilistic predictions are needed because the anthropogenic climate change signal is still low compared to natural variability. We note that while GCM and downscaling uncertainties can partly be reduced in the future, the internal variability leads to fundamental limitations of predictability, which can be expected to strongly depend on the location and on the precipitation properties under consideration.

[181] In almost all forms of downscaling today, the coarse-scale conditions given by the GCM are taken as fixed. However, this does not reflect the reality of the real climate system in which there are feedbacks between coarse and fine scales. This has been noted by *Wilby et al.* [2004] as a limitation of statistical downscaling schemes, but of course, it applies equally to RCMs. To represent these feedbacks in any climate simulation will require coupled runs of the coarse- and fine-scale models, and although the implications for impacts applications are unknown at present, this represents an exciting challenge for a future generation of downscaling techniques.

GLOSSARY

Climate Prediction Net: Initiative to enable probabilistic predictions of future climate conditional on a scenario [*Stainforth et al.*, 2005]. A GCM is run on thousands of home computers to create a large ensemble of future projections, each of which is given a certain likelihood given observational data. (See http://climateprediction.net/.)

Coordinated Regional Climate Downscaling Experiment (CORDEX):   Recent initiative from the World Climate Research Program for running multiple RCM simulations at 50 km resolution for multiple regions. (See http://copes.ipsl.jussieu.fr/RCD_CORDEX.html.)

Dynamical downscaling: Nests a high-resolution regional climate model into a lower-resolution global climate model to represent the atmospheric physics with a higher grid box resolution within a limited area of interest.

ENSEMBLES: Project of the European Union 6th framework program. The project created ensembles of general circulation models and regional climate models for Europe and North Africa, developed statistical downscaling models and tools, and constructed a high-resolution gridded validation data set. (See http://ensembles-eu.metoffice.com; RCM data are available at http://ensemblesrt3.dmi.dk/.)

European Centre for Medium-Range Weather Forecasts 40 Year Reanalysis (ERA40): Six hourly reanalysis of the European Centre for Medium-Range Weather Forecasts, September 1957 to August 2002. Basic resolution of $2.5° × 2.5°$, full resolution of $1.125° × 1.125°$. (See http://www.ecmwf.int/research/era/.)

Global climate model (GCM): The acronym GCM usually stands for general circulation model, but it is often, as in this paper, also used for global climate model. A general circulation model is a dynamical model that numerically integrates the Navier-Stokes equations for either atmosphere or ocean across the globe, typically of a resolution of 100–200 km. Atmosphere and ocean general circulation models are key components of global climate models, which, in general, additionally include sea ice and land surface components.

Model output statistics (MOS): A statistical downscaling approach that corrects dynamical model simulations. The statistical model is calibrated against simulated predictors and observed predictands. Therefore, the statistical model is only valid for the dynamical model it was calibrated with.

National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis: Six hourly reanalysis of the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), 1948 to present. Available at $2.5° × 2.5°$ and $1.875° × 1.875°$. (See http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml.)

Perfect prognosis (PP): a statistical downscaling approach that assumes that the predictor variables are perfectly modeled by the dynamical model used. The statistical model is calibrated against large-scale and local-scale observed data and then is transferred to an arbitrary dynamical model that is assumed to fulfill the PP assumption.

Prediction: An estimate of a future climate state (or a range of states) that is assigned a certain probability (which might be low or subjective) to occur. Climate predictions are possible only for relatively short time scales (seasons to decades) because beyond these time scales the influence of different emission scenarios begins to dominate (see projection).

Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE): Project of the European Union 5th framework program. (See http://prudence.dmi.dk/.)

Projection: A simulation of the response of the future climate to a forcing scenario that is not assigned a certain probability. A projection is therefore only a plausible state of the future climate.

Reanalysis data: Combination of observational data and the forecast of a high-resolution global climate model to build a best estimate of a consistent global weather state. They fill gaps in observational data and provide estimates of nonobserved variables.

Regional climate model (RCM): High-resolution dynamical climate model, typically of a resolution of 25–50 km, though some recent models provide a resolution of 10 km or less. Usually a limited area model nested into a GCM over a specific region.

Statistical and Regional Dynamical Downscaling of Extremes for European Regions (STARDEX): Project of the European Union 5th framework program. (See http://www.cru.uea.ac.uk/projects/stardex/.)

Statistical downscaling: Establishes statistical links between large-scale weather and observed local-scale weather. Either PP or MOS.

UK climate projections (UKCP09): A project funded by the Department for Environment, Food and Rural Affairs to create regional probabilistic climate projections and a weather generator for the United Kingdom. (See http://ukclimateprojections.defra.gov.uk/.)

Weather generator: A stochastic model to create random time series which resemble the observed weather statistics (marginal distribution, short-term temporal variability, and sometimes spatial dependence between multiple sites) at a certain point. To account for variability on longer time scales, weather generators can be run in a downscaling context, either PP or MOS.

REFERENCES

Achberger, C. M., M. L. Linderson, and D. Chen (2003), Performance of the Rossby Centre regional atmospheric model in Southern Sweden: Comparison of simulated and observed precipitation, *Theor. Appl. Climatol.*, 76, 219–234.

Ailliot, P., C. Thompson, and P. Thomson (2009), Space-time modelling of precipitation using a hidden Markov model and censored Gaussian distributions, *Appl. Statist.*, 58, 405–426.

Allan, R. P., and B. J. Soden (2008), Atmospheric warming and the amplification of precipitation extremes, *Science*, 321, 1481–1484.

Allcroft, D. J., and C. A. Glasbey (2003), A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation, *Appl. Stat.*, 52, 487–498.

Allen, M. R., and W. J. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, 419, 224–232.

Bachner, S., A. Kapala, and C. Simmer (2008), Evaluation of daily precipitation characteristics in the CLM and their sensitivity to parameterizations, *Meteorol. Z.*, 17, 407–419, doi:10.1127/0941-2948/2008/0300.

Ballester, J., and J. Moré (2007), The representativeness problem of a station net applied to the verification of a precipitation forecast based on areas, *Meteorol. Appl.*, 14, 177–184.

Bárdossy, A., and E. Plate (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, 28, 1247–1259.

Bárdossy, A., I. Bogardi, and I. Matyasovszky (2005), Fuzzy rule-based downscaling of precipitation, *Theor. Appl. Climatol.*, 82, 119–129.

Beckmann, B.-R., and T. A. Buishand (2002), Statistical downscaling relationships for precipitation in the Netherlands and north Germany, *Int. J. Climatol.*, 22, 15–32.

Beersma, J. J., and T. A. Buishand (2003), Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation, *Clim. Res.*, 25, 121–133.

Bellone, E., J. P. Hughes, and P. Guttorp (2000), A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts, *Clim. Res.*, 15, 1–12.

Beniston, M., et al. (2007), Future extreme events in European climate: An exploration of regional climate model projections, *Clim. Change*, 81, 71–95.

Biau, G., E. Zorita, H. von Storch, and H. Wackernagel (1999), Estimation of precipitation by kriging in the EOF space of the sea level pressure field, *J. Clim.*, 12(4), 1070–1085.

Boberg, F., P. Berg, P. Thejll, W. Gutowski, and J. Christensen (2009), Improved confidence in climate change projections of precipitation evaluated using daily statistics from the PRUDENCE ensemble, *Clim. Dyn.*, 32(7), 1097–1106.

Bock, H.-H. (1996), Probabilistic models in cluster analysis, *Comput. Stat. Data Anal.*, 23(1), 5–28.

Boé, J., L. Terray, F. Habets, and E. Martin (2007), Statistical and dynamical downscaling of the Seine basin climate for hydrometeorological studies, *Int. J. Climatol.*, 27, 1643–1655.

Bretherton, C. S., C. Smith, and J. M. Wallace (1992), An intercomparison of methods for finding coupled patterns in climate data, *J. Clim.*, 5, 541–560.

Brier, G. (1950), The statistical theory of turbulence and the problem of diffusion in the atmosphere, *J. Atmos. Sci.*, 7(4), 283–290.

Brockhaus, P., D. Lüthi, and C. Schär (2008), Aspects of the diurnal cycle in a regional climate model, *Meteorol. Z.*, 17, 433–443.

Brockwell, P. J., and R. A. Davis (1991), *Time Series: Theory and Models*, Springer, Heidelberg, Germany.

Bronstert, A., V. Kolokotronis, D. Schwandt, and H. Straub (2007), Comparison and evaluation of regional climate scenarios for hydrological impact analysis: General scheme and application example, *Int. J. Climatol.*, 27(12), 1579–1594.

Buishand, T., and T. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling, *Water Resour. Res.*, 37(11), 2761–2776.

Buonomo, E., R. G. Jones, C. Huntingford, and J. Hannaford (2007), On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations, *Q. J. R. Meteorol. Soc.*, 133, 65–81.

Carter, G. M., J. P. Dallavalle, and H. R. Glahn (1989), Statistical forecasts based on the National Meteorological Center's numerical weather prediction system, *Weather Forecast.*, 4, 401–412.

Casola, J., and J. Wallace (2007), Identifying weather regimes in the wintertime 500-hPa geopotential height field for the Pacific–North American sector using a limited-contour clustering technique, *J. Appl. Meteorol. Climatol.*, 46, 1619–1630.

Cavazos, T., and B. Hewitson (2005), Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation, *Clim. Res.*, 28, 95–107.

Chandler, R. E. (2002), GLIMCLIM: Generalized linear modelling for daily climate time series (software and user guide), *Tech. Rep., 227*, Dep. of Stat. Sci., Univ. Coll. London, London, (Available at http://www.ucl.ac.uk/Stats/research/reports/index.html)

Chandler, R. E. (2005), On the use of generalized linear models for interpreting climate variability, *Environmetrics*, 16, 699–715.

Chandler, R. E., and S. M. Bate (2007), Inference for clustered data using the independence log-likelihood, *Biometrika*, 94, 167–183.

Changnon, D. (2004), Improving outreach in atmospheric sciences—Assessment of users of climate products, *Bull. Am. Meteorol. Soc.*, 85, 601–606.

Charles, S. P., B. C. Bates, P. H. Whetton, and J. P. Hughes (1999), Validation of downscaling models for changed climate conditions: Case study of southwestern Australia, *Clim. Res.*, 12, 1–14.

Charles, S., B. Bates, I. Smith, and J. Hughes (2004), Statistical downscaling of daily precipitation from observed and modelled atmospheric fields, *Hydrol. Processes*, 18, 1373–1394.

Chen, C. T., and T. Knutson (2008), On the verification and comparison of extreme rainfall indices from climate models, *J. Clim.*, 21(7), 1605–1621.

Christensen, J. H., and O. B. Christensen (2003), Climate modelling: Severe summertime flooding in Europe, *Nature*, 421, 805–806.

Christensen, J. H., and O. B. Christensen (2007), A summary of the PRUDENCE model projections of changes in European climate by the end of this century, *Clim. Change*, 81, 7–30.

Christensen, J. H., et al. (2007), *Climate Change 2007: The Physical Science Basis—Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, Cambridge, U. K.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (2008), On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophys. Res. Lett.*, 35, L20709, doi:10.1029/2008GL035694.

Chun, K., H. Wheater, and C. Onof (2009), Streamflow estimation for six UK catchments under future climate scenarios, *Hydrol. Res.*, 40, 96–112.

Coe, R., and R. D. Stern (1982), Fitting models to daily rainfall data, *J. Appl. Meteorol.*, 21, 1024–1031.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

Collins, M., et al. (2006), Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study, *J. Clim.*, 19(7), 1195–1203.

Conway, D., and P. D. Jones (1998), The use of weather types and air flow indices for GCM downscaling, *J. Hydrol.*, 212–213, 348–361.

Cooley, D., D. Nychka, and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Stat. Assoc.*, 102(479), 824–840.

Cowpertwait, P. S. P. (1991), Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resour. Res.*, 27, 1431–1438.

Cowpertwait, P., P. O'Connell, A. Metcalfe, and J. Mawdsley (1996), Stochastic point process modelling of rainfall. I. Single-site fitting and validation, *J. Hydrol.*, *175*(1–4), 17–46.

Cowpertwait, P. S. P., C. G. Kilsby, and P. E. O'Connell (2002), A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes, *Water Resour. Res.*, *38*(8), 1131, doi:10.1029/2001WR000709.

Cowpertwait, P., V. Isham, and C. Onof (2007), Point process models of rainfall: Developments for fine-scale structure, *Proc. R. Soc. A*, *463*(2086), 2569–2588.

Dankers, R., O. B. Christensen, L. Feyen, M. Kalas, and A. de Roo (2007), Evaluation of very high resolution climate model data for simulating flood hazards in the Upper Danube Basin, *J. Hydrol.*, *347*, 319–331.

Davison, A. (2003), *Statistical Models*, Cambridge Univ. Press, Cambridge, U. K.

Davison, A., and D. Hinkley (1997), *Bootstrap Methods and Their Application*, Cambridge Univ. Press, Cambridge, U. K.

Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1794.

Denis, B., R. Laprise, D. Caya, and J. Côté (2002), Downscaling ability of one-way nested regional climate models: The Big Brother Experiment, *Clim. Dyn.*, *18*(8), 627–646, doi:10.1007/s00382-001-0201-0.

Department for Environment, Food and Rural Affairs (2001), To what degree can the October/November 2000 flood events be attributed to climate change?, *Tech. Rep., Proj. FD2304*, London.

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

Déqué, M., et al. (2007), An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections, *Clim. Change*, *81*, 53–70.

Dettinger, M. D., D. R. Cayan, M. K. Meyer, and A. E. Jeton (2004), Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American river basins, Sierra Nevada, California, 1900–2099, *Clim. Change*, *62*, 283–317.

Dobson, A. J. (2001), *An Introduction to Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

Dunn, P. K. (2004), Occurrence and quantity of precipitation can be modelled simultaneously, *Int. J. Climatol.*, *24*, 1231–1239.

Durman, C. F., J. M. Gregory, D. C. Hassel, R. G. Jones, and J. M. Murphy (2001), A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates, *Q. J. R. Meteorol. Soc.*, *127*, 1005–1015.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, 436 pp., Chapman and Hall, New York.

Engen-Skaugen, T. (2007), Refinement of dynamically downscaled precipitation and temperature scenarios, *Clim. Change*, *84*(3), 365–382.

Fealy, R., and J. Sweeney (2007), Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach, *Int. J. Climatol.*, *15*, 2083–2094.

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

Ferraris, L., S. Gabellani, N. Rebora, and A. Provenzale (2003), A comparison of stochastic models for spatial rainfall downscaling, *Water Resour. Res.*, *39*(12), 1368, doi:10.1029/2003WR002504.

Ferro, C. A. T., A. Hannachi, and D. B. Stephenson (2005), Simple nonparametric techniques for exploring changing probability distributions of weather, *J. Clim.*, *18*, 4344–4354.

Fowler, H. J., and M. Ekström (2009), Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, *Int. J. Climatol.*, *29*(3), 385–416.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connell (2000), A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions, *Hydrol. Earth. Syst. Sci.*, *4*, 263–282.

Fowler, H. J., C. G. Kilsby, P. E. O'Connell, and A. Burton (2005), A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climate variability and change, *J. Hydrol.*, *308*, 50–66.

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

Fowler, H. J., M. Ekström, S. Blenkinsop, and A. P. Smith (2007b), Estimating change in extreme European precipitation using a multimodel ensemble, *J. Geophys. Res.*, *112*, D18104, doi:10.1029/2007JD008619.

Fraley, C., and A. E. Raftery (2002), Model-based clustering, discriminant analysis, and density estimation, *J. Am. Stat. Assoc.*, *97*, 611–631.

Frei, C., and C. Schär (1998), A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, *18*(8), 873–900.

Frei, C., J. H. Christensen, M. Déqué, D. Jacob, R. G. Jones, and P. L. Vidale (2003), Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, *J. Geophys. Res.*, *108*(D3), 4124, doi:10.1029/2002JD002287.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2006), Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, *111*, D06105, doi:10.1029/2005JD005965.

Frías, M., E. Zorita, J. Fernández, and C. Rodríguez-Puebla (2006), Testing statistical downscaling methods in simulated climates, *Geophys. Res. Lett.*, *33*, L19807, doi:10.1029/2006GL027453.

Friederichs, P., and A. Hense (2007), Statistical downscaling of extreme precipitation events using censored quantile regression, *Mon. Weather Rev.*, *135*, 2365–2378, doi:10.1175/MWR3403.1.

Frigessi, A., O. Haug, and H. Rue (2002), A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, *5*(3), 219–235, doi:10.1023/A:1024072610684.

Früh, B., H. Feldmann, H. Panitz, G. Schädler, D. Jacob, P. Lorenz, and K. Keuler (2010), Determination of precipitation return values in complex terrain and their evaluation, *J. Clim.*, *23*(9), 2257–2274, doi:10.1175/2009JCLI2685.1.

Furrer, E. M., and R. W. Katz (2007), Generalized linear modeling approach to stochastic weather generators, *Clim. Res.*, *34*, 129–144.

Furrer, R., and P. Naveau (2007), Probability weighted moments properties for small samples, *Stat. Probab. Lett.*, *77*(2), 190–195.

Gabriel, K. R., and J. Neumann (1962), A Markov chain model for daily rainfall in Tel Aviv, *Q. J. R. Meteorol. Soc.*, *88*, 90–95.

Gigerenzer, G., R. Hertwig, E. van den Broek, B. Fasolo, and K. Katsikopoulos (2005), A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts?, *Risk Anal.*, *25*(3), 623–629.

Giorgi, F., and M. R. Marinucci (1996), An investigation of the sensitivity of simulated precipitation to the model resolution and its implication for climate studies, *Mon. Weather Rev.*, *124*, 148–166.

Glahn, H. R., and D. A. Lowry (1972), The use of model output statistics (MOS) in objective weather forecasing, *J. Appl. Meteorol.*, *11*, 1203–1211.

Glasbey, C. A., and I. M. Nevison (1997), Rainfall modelling using a latent Gaussian variable, *Lect. Notes Stat.*, *122*, 233–242.

Goodess, C. M., et al. (2010), An intercomparison of statistical downscaling methods for Europe and European regions: Assessing their performance with respect to extreme weather events and the implications for climate change applications, *CRU Res.*

Rep., Clim. Res. Unit, Univ. of East Anglia, Norwich, U. K., in press.

Graham, L., S. Hagemann, S. Jaun, and M. Beniston (2007a), On interpreting hydrological change from regional climate models, *Clim. Change*, *81*, 97–122.

Graham, L. P., J. Andréasson, and B. Carlsson (2007b), 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, *Clim. Change*, *81*, 293–307.

Grunwald, G. K., and R. H. Jones (2000), Markov models for time series with mixed distribution, *Environmetrics*, *11*, 327–339.

Guillot, G., and T. Lebel (1999), Disaggregation of Sahelian meso-scale convective system rain fields: Further developments and validation, *J. Geophys. Res.*, *104*(D24), 31,533–31,551, doi:10.1029/1999JD900986.

Halenka, T., J. Kalvová, Z. Chládová, A. Demeterová, K. Zemánková, and M. Belda (2006), On the capability of RegCM to capture extremes in long term regional climate simulation: Comparison with the observations for Czech Republic, *Theor. Appl. Climatol.*, *86*, 125–145.

Hannachi, A., I. T. Jolliffe, and D. Stephenson (2007), Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol*, *27*, 1119–1152.

Hanssen-Bauer, I., E. Forland, J. Haugen, and O. Tveito (2003), Temperature and precipitation scenarios for Norway: Comparison of results from dynamical and empirical down-scaling, *Clim. Res.*, *25*(1), 15–27.

Hanssen-Bauer, I., C. Achberger, R. Benestad, D. Chen, and E. Forland (2005), Statistical downscaling of climate scenarios over Scandinavia, *Clim. Res.*, *29*(3), 255–268.

Hastie, T. J., and R. J. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall, London.

Hay, L. E., and M. P. Clark (2003), Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States, *J. Hydrol.*, *282*, 56–75.

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess (2006), Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios, *Int. J. Climatol.*, *26*(10), 1397–1415.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteorol. Soc.*, *90*(8), 1095–1107.

He, Y., F. Wetterhall, H. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor (2009), Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, *Meteorol. Appl.*, *16*(1), 91–101.

Hellström, C., D. Chen, C. Achberger, and J. Räisänen (2001), Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation, *Clim. Res.*, *19*(1), 45–55.

Herrera-Pantoja, M., and K. Hiscock (2008), The effects of climate change on potential groundwater recharge in Great Britain, *Hydrol. Processes*, *22*(1), 73–86.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, *15*, 559–570.

Hess, P., and H. Brezowsky (1977), Katalog der Großwetterlagen Europas (1861–1976), *Rep. 15*, Dtsch. Wetterdienstes, Offenbach am Main, Germany.

Hewitson, B., and R. Crane (1996), Climate downscaling: Techniques and application, *Clim. Res.*, *7*, 85–95.

Hewitson, B., and R. Crane (2002), Self-organizing maps: Applications to synoptic climatology, *Clim. Res.*, *22*, 13–26.

Hewitt, C. D., and D. J. Griggs (2004), Ensembles-based predictions of climate changes and their impacts, *Eos Trans. AGU*, *85*, 566.

Hinton, T., B. Hoskins, and G. Martin (2009), The influence of tropical sea surface temperatures and precipitation on north Pacific atmospheric blocking, *Clim. Dyn.*, *33*(4), 549–563.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, *113*, D21110, doi:10.1029/2008JD010100.

Hohenegger, C., P. Brockhaus, and C. Schär (2008), Towards climate simulations at cloud-resolving scales, *Meteorol. Z.*, *17*(4), 383–394.

Hollweg, H. D., et al. (2008), Ensemble simulations over Europe with the Regional Climate Model CLM forced with IPCC AR4 global scenarios, *MD Tech. Rep., 3*, Max-Planck-Inst., Hamburg, Germany.

Hughes, J. P., P. Guttorp, and S. Charles (1999), A nonhomogeneous hidden Markov model for precipitation, *Appl. Stat.*, *48*, 15–30.

Hulme, M., et al. (2002), *Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report*, Tyndall Cent. for Clim. Change Res., Sch. of Environ. Sci., Univ. of East Anglia, Norwich, U. K.

Huth, R. (1996), An intercomparison of computer-assisted circulation classification methods, *Int. J. Climatol.*, *16*, 893–922.

Huth, R. (1999), Statistical downscaling in central Europe: Evaluation of methods and potential predictors, *Clim. Res.*, *13*, 91–103.

Huth, R., and J. Kyselý (2000), Constructing site-specific climate change scenarios on a monthly scale using statistical downscaling, *Theor. Appl. Climatol.*, *66*(1), 13–27.

Hyndman, R. J., and G. K. Grunwald (2000), Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall, *Aust. N. Z. J. Stat.*, *42*, 145–158.

Ines, A. V. M., and J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies, *Agr. For. Meteorol.*, *138*, 44–53.

Ivanov, S., and J. Palamarchuk (2007), Fine-scale precipitation structure of a cold front and the problem of the representativeness error, *Adv. Geosci.*, *10*, 3–8.

Jacob, D., et al. (2007), An intercomparison of regional climate models for Europe: Model performance in present-day climate, *Clim. Change*, *81*, 31–52.

Jacobeit, J., H. Wanner, J. Luterbacher, C. Beck, A. Philipp, and K. Sturm (2003), Atmospheric circulation variability in the North-Atlantic-European area since the mid-seventeenth century, *Clim. Dyn.*, *20*(4), 341–352, doi:10.1007/s00382-002-0278-0.

Jaeger, E., I. Anders, D. Lüthi, B. Rockel, C. Schär, and S. Seneviratne (2008), Analysis of ERA40-driven CLM simulations for Europe, *Meteorol. Z.*, *17*(4), 349–367.

Jenkinson, A., and F. Collison (1977), An initial climatology of gales over the North Sea, *Synopt. Climatol. Branch Memo.*, *62*, Meteorol. Off., Bracknell, U. K.

Jolliffe, I. T., and D. B. Stephenson (Eds.) (2003), *Forecast Verification, A Practitioner's Guide in Atmospheric Science*, 240 pp., John Wiley, Chichester, U. K.

Jones, P. D., M. Hulme, and K. Briffa (1993), A comparison of Lamb circulation types with an objective classification scheme, *Int. J. Climatol.*, *13*(6), 655–663.

Jones, P. D., C. Harpham, C. Kilsby, V. Glenis, and A. Burton (2009), Projections of future daily climate for the UK from the Weather Generator, *Tech. Rep., 1*, Met Off., Exeter, U. K.

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, *Q. J. R. Meteorol. Soc.*, *121*, 1413–1449.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation, and Predictability*, 341 pp., Cambridge Univ. Press, New York.

Kalnay, E., et al. (1996), The NCEP/NCAR reanalysis project, *Bull. Am. Meteorol. Soc.*, 77, 437–471.

Kanada, S., M. Nakano, S. Hayashi, T. Kato, M. Nakamura, K. Kurihara, and A. Kitoh (2008), Reproducibility of maximum daily precipitation amount over Japan by a high-resolution non-hydrostatic model, *SOLA*, 4, 105–108.

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 observed local climate, part I: Seasonal statistics, *J. Clim.*, 3, 1053–1079.

Katz, R. (1977), Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, 16(7), 671–676.

Katz, R. W., and M. B. Parlange (1993), Effects of an index of atmospheric circulation on stochastic properties of precipitation, *Water Resour. Res.*, 29(7), 2335–2344.

Katz, R. W., and M. B. Parlange (1998), Overdispersion phenomenon in stochastic modeling of precipitation, *J. Clim.*, 11(4), 591–601.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287–1304.

Keenlyside, N., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, 453(7191), 84–88.

Kendon, E. J., D. P. Rowell, R. G. Jones, and E. Buonomo (2008), Robustness of future changes in local precipitation extremes, *J. Clim.*, 21, 4280–4297.

Kendon, E. J., D. P. Rowell, and R. G. Jones (2009), Mechanisms and reliability of future projected changes in daily precipitation, *Clim. Dyn.*, 35, 489–509, doi:10.1007/s00382-009-0639-z.

Kendon, E. J., R. G. Jones, E. Kjellström, and J. M. Murphy (2010), Using and designing GCM-RCM ensemble regional climate projections, *J. Clim.*, in press.

Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby (2007), A daily weather generator for use in climate change studies, *Environ. Modell. Software*, 22(12), 1705–1719.

Kistler, R., et al. (2001), The NCEP/NCAR 50-year reanalysis, *Bull. Am. Meteorol. Soc.*, 82(2), 247–268.

Klein, W. H., and H. R. Glahn (1974), Forecasting local weather by means of model output statistics, *Bull. Am. Meteorol. Soc.*, 55, 1217–1227.

Klein, W. H., B. M. Lewis, and I. Enger (1959), Objective prediction of 5-day mean temperatures during winter, *J. Meteorol.*, 16, 672–682.

Kohonen, T. (1998), The self-organizing map, *Neurocomputing*, 21(1–3), 1–6, doi:10.1016/S0925-2312(98)00030-7.

Koutsoyiannis, D., and C. Onof (2001), Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *J. Hydrol.*, 246(1–4), 109–122.

Kundzewicz, Z. W., et al. (2007), Freshwater resources and their management, in *Climate Change 2007: Impacts, Adaptation and Vulnerability—Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 174–210, Cambridge Univ. Press, Cambridge, U. K.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, 32(3), 679–693.

Landmann, W. A., and L. Goddard (2002), Statistical recalibration of GCM forecasts over southern Africa using model output statistics, *J. Clim.*, 15, 2038–2055.

Latif, M., et al. (2001), ENSIP: The El Niño simulation intercomparison project, *Clim. Dyn.*, 18(3), 255–276.

Laux, P., H. Kunstmann, and A. Bardossy (2008), Predicting the regional onset of the rainy season in West Africa, *Int. J. Climatol.*, 28(3), 329–342.

Leander, R., and T. A. Buishand (2007), Resampling of regional climate model output for the simulation of extreme river flows, *J. Hydrol.*, 332, 487–496.

Leander, R., T. Buishand, B. van den Hurk, and M. de Wit (2008), Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, *J. Hydrol.*, 351(3–4), 331–343.

Leith, N. A., and R. E. Chandler (2010), A framework for interpreting climate model outputs, *Appl. Stat.*, 59(2), 279–296.

Leloup, J., M. Lengaigne, and J.-P. Boulanger (2008), Twentieth century ENSO characteristics in the IPCC database, *Clim. Dyn.*, 30, 277–291.

Lenderink, G., and E. van Meijgaard (2008), Increase in hourly precipitation extremes beyond expectations from temperature changes, *Nat. Geosci.*, 1, 511–514.

Lenderink, G., A. Buishand, and W. van Deursen (2007), Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach, *Hydrol. Earth Syst. Sci.*, 11(3), 1145–1159.

Lorenz, E. N. (1969), Atmospheric predictability as revealed by naturally occuring analogs, *J. Atmos. Sci.*, 26, 639–646.

Lovejoy, S., and D. Schertzer (2006), Multifractals, cloud radiances and rain, *J. Hydrol.*, 322(1–4), 59–88.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000), WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility, *Stat. Comput.*, 10, 325–337.

MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, Univ. of Calif. Press, Berkeley.

Marani, M., and S. Zanetti (2007), Downscaling rainfall temporal variability, *Water Resour. Res.*, 43, W09415, doi:10.1029/2006WR005505.

Maraun, D., H. W. Rust, and T. J. Osborn (2009), The annual cycle of heavy precipitation across the United Kingdom: A model based on extreme value statistics, *Int. J. Climatol.*, 29(12), 1731–1744.

Maraun, D., H. W. Rust, and T. J. Osborn (2010a), Synoptic airflow and UK daily precipitation extremes: Development and validation of a vector generalised linear model, *Extremes*, 13, 133–153, doi:10.1007/s10687-010-0102-x.

Maraun, D., H. W. Rust, and T. J. Osborn (2010b), The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships, *Clim. Dyn.*, doi:10.1007/s00382-009-0710-9, in press.

Marsan, D., D. Schertzer, and S. Lovejoy (1996), Causal space-time multifractal processes: Predictability and forecasting of rain-fields, *J. Geophys. Res.*, 101(D21), 26,333–26,346.

Marsh, T., and M. Dale (2002), The UK floods of 2000–2001: A hydrometeorological appraisal, *Water Environ. J.*, 16(3), 180–188.

Mason, S. J. (2004), Simulating climate over western North America using stochastic weather generators, *Clim. Change*, 62(1), 155–187.

May, W. (2007), The simulation of the variability and extremes of daily precipitation over Europe by the HIRHAM regional climate model, *Global Planet. Change*, 57, 59–82.

Mearns, L. O., et al. (2009), A regional climate change assessment program for North America, *Eos Trans. AGU*, 90(36), 311.

Meehl, G. A., et al. (2007), Global climate projections, in *Climate Change 2007: The Physical Science Basis—Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 747–845, Cambridge Univ. Press, Cambridge, U. K.

Mesinger, F., et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, 87(3), 343–360.

Michaud, J., and S. Sorooshian (1994), Effect of rainfall-sampling errors on simulations of desert flash floods, *Water Resour. Res.*, 30, 2765–2775.

Michelangeli, P. A., R. Vautard, and B. Legras (1995), Weather regimes: Recurrence and quasi-stationarity, *J. Atmos. Sci.*, 52, 1237–1256.

Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*(6), 693–712.

Moberg, A., and P. D. Jones (2005), Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99, *Int. J. Climatol.*, *25*, 1149–1171.

Monahan, J. F. (2001), *Numerical Methods of Statistics*, Cambridge Univ. Press, Cambridge, U. K.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye (2008a), Weather types and rainfall over Senegal. Part I: Observational analysis, *J. Clim.*, *21*, 266–287, doi:10.1175/2007JCLI1601.1.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye (2008b), Weather types and rainfall over Senegal. Part II: Downscaling GCM simulations, *J. Clim.*, *21*, 288–307, doi:10.1175/2007JCLI1624.1.

Murphy, J. M. (1999), An evaluation of statistical and dynamical techniques for downscaling local climate, *J. Clim.*, *12*, 2256–2284.

Murphy, J. M., et al. (2009), UK climate projections science report: Climate change projections, technical report, Met Off. Hadley Cent., Exeter, U. K.

Naden, P. (1992), Spatial variability in flood estimation for large catchments: The exploitation of channel network structure, *Hydrol. Sci. J.*, *37*, 53–71.

Naveau, P., M. Nogaj, C. Ammann, P. Yiou, D. Cooley, and V. Jomelli (2005), Statistical methods for the analysis of climate extremes, *C. R. Geosci.*, *337*(10–11), 1013–1022.

Ngirane Katashaya, G., and H. Wheater (1985), Hydrograph sensitivity to storm kinematics, *Water Resour. Res.*, *21*, 337–345.

Nicholls, N., and W. Murray (1999), Workshop on indices and indicators for climate extremes, Asheville, NC, USA, 3–6 June 1997: Breakout group B: Precipitation, *Clim. Change*, *42*, 23–29.

Northrop, P. (1998), A clustered spatial-temporal model of rainfall, *Proc. R. Soc. London, Ser. A*, *454*, 1875–1888.

Olsson, J. (1998), Evaluation of a scaling cascade model for temporal rainfall disaggregation, *Hydrol. Earth Syst. Sci.*, *2*, 19–30.

Olsson, J., C. Uvo, and K. Jinno (2001), Statistical atmospheric downscaling of short-term extreme rainfall by neural networks, *Phys. Chem. Earth, Part B*, *26*(9), 695–700.

Onibon, H., T. Lebel, A. Afouda, and G. Guillot (2004), Gibbs sampling for conditional spatial disaggregation of rain fields, *Water Resour. Res.*, *40*, W08401, doi:10.1029/2003WR002009.

Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham (2000), Rainfall modelling using Poisson-cluster processes: A review of developments, *Stochastic Environ. Res. Risk Assess.*, *14*, 384–411.

Orlowsky, B., F.-W. Gerstengarbe, and P. C. Werner (2008), A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM, *Theor. Appl. Climatol.*, *92*, 209–223, doi:10.1007/s00704-007-0352-y.

Over, T., and V. Gupta (1996), A space-time theory of mesoscale rainfall using random cascades, *J. Geophys. Res.*, *101*(D21), 26,319–26,331.

Palmer, T. N. (1999), A nonlinear dynamical perspective on climate prediction, *J. Clim.*, *12*(2), 575–591.

Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell (2008), Toward seamless prediction. Calibration of climate change projections using seasonal forecasts, *Bull. Am. Meteorol. Soc.*, *89*(4), 459–470.

Panofsky, H. W., and G. W. Brier (1968), *Some Applications of Statistics to Meteorology*, 224 pp., Pa. State Univ. Press, University Park.

Pennesi, K. (2007), Improving forecast communication: Linguistic and cultural considerations, *Bull. Am. Meteorol. Soc.*, *88*, 1033–1044.

Perica, S., and E. Foufoula-Georgiou (1996), Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, *J. Geophys. Res.*, *101*(D21), 26,347–26,361.

Peterson, T. C., C. Folland, G. Gruza, W. Hogg, A. Mokssit, and N. Plummer (2001), Report on the activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998–2001, *WMO-TD 1071*, World Meteorol. Organ., Geneva, Switzerland.

Philipp, A., P. M. Della-Marta, J. Jacobeit, D. R. Fereday, P. D. Jones, A. Moberg, and H. Wanner (2007), Long term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering, *J. Clim.*, *20*(16), 4065–4095.

Piani, C., J. O. Haerter, and E. Coppola (2009), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, *99*, 187–192.

Plaut, G., and E. Simonnet (2001), Large-scale circulation classification, weather regimes, and local climate over France, the Alps and western Europe, *Clim. Res.*, *17*, 303–324.

Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, Amsterdam.

Prömmel, K., B. Geyer, J. M. Jones, and M. Widmann (2009), Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature, *Int. J. Climatol.*, *30*, 760–773, doi:10.1002/joc.1916.

Prudhomme, C., N. Reynard, and S. Crooks (2002), Downscaling of global climate models for flood frequency analysis: Where are we now?, *Hydrol. Processes*, *16*(6), 1137–1150, doi:10.1002/hyp.1054.

Ragab, R., J. Finch, and R. Harding (1997), Estimation of groundwater recharge to chalk and sandstone aquifers using simple soil models, *J. Hydrol.*, *190*, 19–41.

Rauscher, S., E. Coppola, C. Piani, and F. Giorgi (2010), Resolution effects on regional climate model simulations of seasonal precipitation over Europe, *Clim. Dyn.*, in press.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.

Richardson, C. W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, *17*(1), 182–190.

Richardson, C., and D. A. Wright (1984), WGEN: A model for generating daily weather variables, *Rep. 8*, 83 pp., Agric. Res. Serv., U.S. Dep. of Agric., Washington, D. C.

Ringer, M., et al. (2006), The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model (HadGEM1). Part II: Aspects of variability and regional climate, *J. Clim.*, *19*(7), 1302–1326.

Roberts, N. M., and H. W. Lean (2008), Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Mon. Weather Rev.*, *136*, 78–97.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London, Ser. A*, *410*, 269–288.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1988), A point process model for rainfall: Further developments, *Proc. R. Soc. London, Ser. A*, *417*, 283–298.

Rosenzweig, C., A. Iglesias, X. Yang, P. Epstein, and E. Chivian (2001), Climate change and extreme weather events: Implications for food production, plant diseases, and pests, *Global Change Human Health*, *2*(2), 90–104.

Rowell, D. P. (2006), A demonstration of the uncertainty in projections of UK climate change resulting from regional model formulation, *Clim. Change*, *79*, 243–257.

Rummukainen, M. (1997), Methods of statistical downscaling of GCM simulations, *Rep. Meteorol. Climatol.*, *80*, Swed. Meteorol. and Hydrol. Inst., Norrköping.

Rust, H. W., D. Maraun, and T. J. Osborn (2009), Modelling seasonality in extreme precipitation: A UK case study, *Eur. Phys. J. Spec. Top.*, *174*, 99–111, doi:10.1140/epjst/e2009-01093-7.

Rust, H. W., M. Vrac, M. Lengaigne, and B. Sultan (2010), Quantifying differences in circulation patterns based on probabilistic

models. IPCC-AR4 multi-model comparison for the North Atlantic, *J. Clim.*, in press.

Salathe, E. (2003), Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin, *Int. J. Climatol.*, 23(8), 887–901.

Sanchez-Gomez, E., S. Somot, and M. Déqué (2009), Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000, *Clim. Dyn.*, 33(5), 723–736, doi:10.1007/s00382-008-0502-7.

Sansó, B., and L. Guenni (2000), A nonstationary multisite model for rainfall, *J. Am. Stat. Assoc.*, 95(452), 1089–1100.

Schär, C., D. Lüthi, U. Beyerle, and E. Heise (1999), The soil-precipitation feedback: A process study with a regional climate model, *J. Clim.*, 12, 722–741.

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.

Schmidli, J., C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith (2007), Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, *J. Geophys. Res.*, 112, D04105, doi:10.1029/2005JD007026

Segond, M. L., C. Onof, and H. S. Wheater (2006), Spatial-temporal disaggregation of daily rainfall from a generalized linear model, *J. Hydrol.*, 331(3–4), 674–689.

Segond, M., H. Wheater, and C. Onof (2007), The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK, *J. Hydrol.*, 347, 116–131.

Semenov, M., R. Brooks, E. Barrow, and C. Richardson (1998), Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates, *Clim. Res.*, 10, 95–107.

Shaffrey, L., et al. (2009), UK HiGEM: The new UK high-resolution global environment model—Model description and basic evaluation, *J. Clim.*, 22(8), 1861–1896.

Shongwe, M. E., W. A. Landman, and S. J. Mason (2006), Performance of recalibration systems for GCM forecasts for southern Africa, *Int. J. Climatol*, 26, 1567–1585.

Singh, V. (1997), Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrol. Processes*, 11, 1649–1669.

Smith, D., S. Cusack, A. Colman, C. Folland, G. Harris, and J. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, 317(5839), 796–799.

Stainforth, D., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, 433(7024), 403–406.

Stehlík, J., and A. Bárdossy (2002), Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, *J. Hydrol*, 256, 120–141.

Stephenson, D. B. (2000), Use of the "odds ratio" for diagnosing forecast skill, *Weather Forecast.*, 15, 221–232.

Stephenson, D. B., A. Hannachi, and A. O'Neill (2004), On the existence of multiple climate regimes, *Q. J. R. Meteorol. Soc.*, 130, 583–605, doi:10.1256/qj.02.146.

Stern, R., and R. Coe (1984), A model fitting analysis of daily rainfall data, *J. R. Stat. Soc. A*, 147, 1–34.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106(D7), 7183–7192.

Themeßl, M. J., A. Gobiet, and A. Leuprecht (2010), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *Int. J. Climatol.*, doi:10.1002/joc.2168, in press.

Thomas, A., B. O'Hara, U. Ligges, and S. Sturtz (2006), Making BUGS open, *R News*, 6, 12–17.

Timbal, B., E. Fernandez, and Z. Li (2008a), Generalization of a statistical downscaling model to provide local climate change projections for Australia, *Environ. Modell. Software*, 24, 342–358.

Timbal, B., P. Hope, and S. Charles (2008b), Evaluating the consistency between statistically downscaled and global dynamical model climate change projections, *J. Clim.*, 21, 6052–6059.

Tippett, M., T. DelSole, S. Mason, and A. Barnston (2008), Regression-based methods for finding coupled patterns, *J. Clim.*, 21(17), 4384–4398.

Tomassini, L., and D. Jacob (2009), Spatial analysis of trends in extreme precipitation events in high-resolution climate model results and observations for Germany, *J. Geophys. Res.*, 114, D12113, doi:10.1029/2008JD010652.

Tustison, B., D. Harris, and E. Foufoula-Georgiou (2001), Scale issues in verification of precipitation forecasts, *J. Geophys. Res.*, 106(D11), 11,775–11,784.

Underwood, F. M. (2009), Describing long-term trends in precipitation using generalized additive models, *J. Hydrol.*, 364, 285–297.

Uppala, S. M., et al. (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc*, 131, 2961–3012.

van der Linden, P., and J. F. B. Mitchell (2009), ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project, technical report, Met Off. Hadley Cent., Exeter, U. K.

Vautard, R. (1990), Multiple weather regimes over the North Atlantic: Analysis of precursors and successors, *Mon. Weather Rev.*, 118, 2056–2081.

Venema, V., S. Gimeno García, and C. Simmer (2010), A new algorithm for the downscaling of cloud fields, *Q. J. R. Meteorol. Soc.*, 136, 91–106, doi:10.1002/qj.535.

Vidale, P. L., D. Lüthi, C. Frei, S. I. Seneviratne, and C. Schär (2003), Predictability and uncertainty in a regional climate model, *J. Geophys. Res.*, 108(D18), 4586, doi:10.1029/2002JD002810.

von Storch, H. (1999), On the use of "inflation" in statistical downscaling, *J. Clim.*, 12(12), 3505–3506.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, Cambridge, U. K.

von Storch, H., H. Langenberg, and F. Feser (2000), A spectral nudging technique for dynamical downscaling purposes, *Mon. Weather Rev.*, 128(10), 3664–3673.

Vrac, M., and P. Naveau (2007), Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, 43, W07402, doi:10.1029/2006WR005308.

Vrac, M., K. Hayhoe, and M. Stein (2007a), Identification and intermodel comparison of seasonal circulation patterns over North America, *Int. J. Climatol.*, 27, 603–620, doi:10.1002/joc.1422.

Vrac, M., P. Marbaix, D. Paillard, and P. Naveau (2007b), Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe, *Clim. Past*, 3(4), 669–682.

Vrac, M., P. Naveau, and P. Drobinski (2007c), Modeling pairwise dependencies in precipitation intensities, *Nonlinear Processes Geophys.*, 14, 789–797.

Vrac, M., M. Stein, and K. Hayhoe (2007d), Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Clim. Res.*, 34, 169–184.

Vrac, M., M. L. Stein, K. Hayhoe, and X. Z. Liang (2007e), A general method for validating statistical downscaling methods under future climate change, *Geophys. Res. Lett.*, 34, L18701, doi:10.1029/2007GL030295.

Wakazuki, Y., M. Nakamura, S. Kanada, and C. Muroi (2008), Climatological reproducibility evaluation and future climate projection of extreme precipitation events in the Baiu season using a high-resolution non-hydrostatic RCM in comparison with an AGCM, *J. Meteorol. Soc. Jpn.*, 86(6), 951–967.

Ward, J. H. (1963), Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58(301), 236–244.

Wehrens, R., and L. M. C. Buydens (2007), Self- and super-organizing maps in R: The Kohonen package, *J. Stat. Software*, 21(5), 1–19.

Wernli, H., M. Paulat, M. Hagen, and C. Frei (2008), SAL—A novel quality measure for the verification of quantitative precipitation forecasts, *Mon. Weather Rev.*, *136*, 4470–4487.

Wetterhall, F., S. Halldin, and C. Xu (2007), Seasonality properties of four statistical-downscaling methods in central Sweden, *Theor. Appl. Climatol.*, *87*(1), 123–137.

Wheater, H. (2006), Flood hazard and management: A UK perspective, *Philos. Trans. R. Soc. A*, *364*(1845), 2135–2145.

Wheater, H. (2008), Modelling hydrological processes in arid and semi-arid areas, in *Hydrological Modelling in Arid and Semi-Arid Areas*, pp. 1–20, Cambridge Univ. Press, Cambridge, U. K.

Wheater, H. S., R. E. Chandler, C. J. Onof, V. S. Isham, E. Bellone, C. Yang, D. Lekkas, G. Lourmas, and M.-L. Segond (2005), Spatial-temporal rainfall modelling for flood risk estimation, *Stochastic Environ. Res. Risk Assess.*, *19*, 403–416.

Widmann, M. (2005), One-dimensional CCA and SVD, and their relationship to regression maps, *J. Clim.*, *18*(14), 2785–2792.

Widmann, M., and C. Bretherton (2000), Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States, *J. Clim.*, *13*(11), 1936–1950.

Widmann, M., C. S. Bretherton, and E. P. Salathé (2003), Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor, *J. Clim.*, *16*(5), 799–816.

Wilby, R. (1994), Stochastic weather type simulation for regional climate change impact assessment, *Water Resour. Res.*, *30*(12), 3395–3403.

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

Wilby, R. L., and T. M. L. Wigley (2000), Precipitation predictors for downscaling: Observed and general circulation model relationships, *Int. J. Climatol.*, *20*, 641–661.

Wilby, R., H. Hassan, and K. Hanaki (1998), Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, *205*(1–2), 1–19.

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, technical report, Intergovt. Panel on Clim. Change, Geneva, Switzerland.

Wilks, D. (1989), Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resour. Res.*, *25*(6), 1429–1439.

Wilks, D. (1998), Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Academic, Amsterdam.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Prog. Phys. Geogr.*, *23*(3), 329–357.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier (1997), Non-stationarity in daily precipitation series: Implications for GCM down-scaling using atmospheric circulation indices, *Int. J. Climatol.*, *17*, 439–454.

Xu, C. (1999a), From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches, *Prog. Phys. Geogr.*, *23*(2), 229–249.

Xu, C.-Y. (1999b), Climate change and hydrologic models: A review of existing gaps and recent research developments, *Water Resour. Manage.*, *13*, 369–382.

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater (2005), Spatial-temporal rainfall simulation using generalized linear models, *Water Resour. Res.*, *41*, W11415, doi:10.1029/2004WR003739.

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater (2006), Quality control for daily observational rainfall series in the UK, *Water Environ. J.*, *20*, 185–193.

Yang, W., J. Andréasson, L. P. Graham, J. Olsson, J. Rosberg, and F. Wetterhall (2010), Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, *Hydrol. Res.*, *41*(3–4), 211–229.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest-neighbor algorithm, *Water Resour. Res.*, *39*(7), 1199, doi:10.1029/2002WR001769.

Yee, T. W., and A. G. Stephenson (2007), Vector generalized linear and additive extreme value models, *Extremes*, *10*, 1–19.

Yee, T. W., and C. J. Wild (1996), Vector generalized additive models, *J. R. Stat. Soc., Ser. B*, *58*, 481–493.

Young, K. (1994), A multivariate chain model for simulating climatic parameters from daily data, *J. Appl. Meteorol.*, *33*(6), 661–671.

Zehe, E., A. Singh, and A. Bardossy (2006), Modelling of monsoon rainfall for a mesoscale catchment in North-West India I: Assessment of objective circulation patterns, *Hydrol. Earth Syst. Sci.*, *10*(6), 797–806.

Zolina, O., A. Kapala, C. Simmer, and S. Gulev (2004), Analysis of extreme precipitation over Europe from different reanalyses: A comparative assessment, *Global Planet. Change*, *44*(1–4), 129–161.

Zorita, E., and H. von Storch (1997), A survey of statistical downscaling techniques, *GKSS Rep.*, *97/E/20*, GKSS Res. Cent., Geesthacht, Germany.

Zorita, E., and H. von Storch (1999), The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *J. Clim.*, *12*, 2474–2489.

S. Brienen, Deutscher Wetterdienst, Frankfurter Str. 135, D-63067 Offenbach, Germany. (Susanne.Brienen@dwd.de)

R. E. Chandler, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK. (richard@stats.ucl.ac.uk)

K. P. Chun, A. M. Ireson, and C. Onof, Environmental and Water Resources Engineering Research Group, Imperial College London, Skempton Building, South Kensington Campus, London SW7 2AZ, UK. (kwok-pan.chun05@imperial.ac.uk; andrew.ireson@imperial.ac.uk; c.onof@imperial.ac.uk)

C. M. Goodess, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. (c.goodess@uea.ac.uk)

R. G. Jones and E. J. Kendon, Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK. (richard.jones@metoffice.gov.uk; elizabeth.kendon@metoffice.gov.uk)

D. Maraun, Department of Geography, University of Giessen, Senckenbergstr. 1, D-35390 Giessen, Germany. (douglas.maraun@geogr.uni-giessen.de)

H. W. Rust and M. Vrac, Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CNRS, Centre d'Etude de Saclay, Orme des Merisiers, Bat. 701, F-91191 Gif-sur-Yvette, France. (henning.rust@lsce.ipsl.fr; mathieu.vrac@lsce.ipsl.fr)

T. Sauter, Department of Geography, RWTH Aachen University, D-52056 Aachen, Germany. (tobias.sauter@geo.rwth-aachen.de)

M. Themeßl, Wegener Center for Climate and Global Change, University of Graz, Leechgasse 25, A-8010 Graz, Austria. (matthias.themessl@uni-graz.at)

I. Thiele-Eich and V. K. C. Venema, Meteorological Institute, University of Bonn, Auf dem Huegel 20, D-53121 Bonn, Germany. (imteich@uni-bonn.de; victor.venema@uni-bonn.de)

F. Wetterhall, Department of Geography, King's College London, Strand, London WC2R 2LS, UK. (fredrik.wetterhall@kcl.ac.uk)

M. Widmann, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK. (m.widmann@bham.ac.uk)