

Climatic Change (2014) 122:539–554
DOI 10.1007/s10584-013-1021-z

Interrogating empirical-statistical downscaling

B. C. Hewitson · J. Daron · R. G. Crane ·
M. F. Zermoglio · C. Jack

Received: 4 September 2013 / Accepted: 25 November 2013 / Published online: 17 December 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The delivery of downscaled climate information is increasingly seen as a vehicle of climate services, a driver for impacts studies and adaptation decisions, and for informing policy development. Empirical-statistical downscaling (ESD) is widely used; however, the accompanying responsibility is significant, and predicated on effective understanding of the limitations and capabilities of ESD methods. There remain substantial contradictions, uncertainties, and sensitivity to assumptions between the different methods commonly used. Yet providing decision-relevant downscaled climate projections to help support national and local adaptation is core to the growing global momentum seeking to operationalize what is, in effect, still foundational research. We argue that any downscaled climate information must address the criteria of being plausible, defensible and actionable. Climate scientists cannot absolve themselves of their ethical responsibility when informing adaptation and must, therefore, be diligent in ensuring any information provided adequately addresses these three criteria. Frameworks for supporting such assessment are not well developed. We interrogate the conceptual foundations of statistical downscaling methodologies and their assumptions, and articulate a framework for evaluating and integrating downscaling output into the wider landscape of climate information. For ESD there are key criteria that need to be satisfied to underpin the credibility of the derived product. Assessing these criteria requires the use of appropriate metrics to test the comprehensive treatment of local climate response to large-scale forcing, and to compare across methods. We illustrate the potential consequences of methodological choices on the interpretation of downscaling results and explore the purposes, benefits and limitations of using statistical downscaling.

B. C. Hewitson (✉)
Climate System Analysis Group, University of Cape Town, Private Bag X3, Rondebosch, Western Cape,
South Africa
e-mail: hewitson@csag.uct.ac.za

J. Daron · C. Jack
Climate System Analysis Group, University of Cape Town, Rondebosch, Western Cape, South Africa

R. G. Crane
Alliance for Education, Science, Engineering and Development in Africa, University of Cape Town,
Rondebosch, Western Cape, South Africa

M. F. Zermoglio
AFR, The World Bank, 1818 H Street N.W., Washington, DC, USA

1 Introduction

Anthropogenic climate change is, in many respects, a problem of ethics. The challenges only exist because of the choices humankind has made, and our future depends on the choices we will make – choices that are now, in part, predicated on projected climate change. At the regional¹ scale our choices are complicated by the large uncertainties in the degree and rate of change and our incomplete knowledge of how human and physical systems will respond. The recognition that change is already happening has accelerated investment in adaptation, and central for many adaptation projects is the use of climate model downscaling to project future regional change.

Downscaling seeks to inform decision making by adding information to Global Climate Model (GCM) products. This approach is becoming mainstreamed as evidenced for example, by the CORDEX² activity of the World Climate Research Program (WCRP³). However, a broad range of difficult context questions hover in the background: Do the scientists engaged in downscaling trust the results? Does downscaling raise more questions than answers? Does downscaling add real information? Does the higher resolution of downscaling imply greater accuracy?

The social landscape of decision making is increasingly confused by the growing number of data sources from a range of climate services, portals, and agencies—all raising contradictions that impact the user community's confidence in the regional climate projections. As a result, there is a need to foster understanding of the issues between the user community and the scientists so that the data can be mutually examined in the appropriate context.

This paper also highlights the ethical questions inherent in producing downscaled information for use in impact and adaptation projects. In this context we consider information as referring to the regional climate projections of the response to anthropogenic forcing. Recognizing that the use of these projections has real societal consequence, we also want to raise awareness of the principal underlying assumptions, limitations and challenges involved in producing and interpreting downscaled products.

We explore the underlying assumptions that relate to the implementation of empirical-statistical downscaling (ESD), particularly where the intention is to inform policy decisions and adaptation actions. In addition we outline some guidelines for users by indicating questions that need to be asked of the methodologies and data. Following this we demonstrate the challenges by showing an ESD example where a choice within the implementation translates to problematic results.

2 Context

ESD builds on the premise that the larger scale atmospheric processes are dominant drivers of local scale climate (Hewitson and Crane 1996; Wilby et al. 1998). However, many assumptions and choices made in the implementation of ESD have the potential to substantially alter the results and so impact decision outcomes. In general ESD relates large scale atmosphere drivers (predictors) to a local scale variable of interest (predictand). The means employed to represent the relationship carry assumptions that if left unexamined can lead to artifacts in the result. For example, assumptions about spatial and temporal distributions (Maraun et al. 2010).

The bulk of the downscaling literature is focused on two broad areas: the development of methods (e.g. Mahmood and Babel 2012; Stoner et al. 2012), or on applications using one or

¹ The term “regional” is used here in a loose sense, as related to spatial scales of vulnerability, impact, and adaptation.

² <http://wcrp-cordex.ipsl.jussieu.fr/>

³ <http://www.wcrp-climate.org/index.php/key-deleverables/regional-climat6>

more ESD methods to providing information to end users (e.g. Pierce et al. 2012; Jeong et al. 2012). In terms of the value of downscaling, some are pessimistic (e.g. Pielke and Wilby 2012) while others view the current status in a more optimistic light (e.g. Maraun et al. 2010). These differing perspectives of downscaling are problematic in the face of a growing move to operationalize what is, in effect, still foundational research. For example, the uncertainty in weather forecasting is well characterized, making this a valuable operational activity (Palmer 2000; Slingo and Palmer 2011). By contrast, on longer time scales of seasons to decades the ranges of projections are poorly constrained in terms of the possible limits and distribution of outcomes. Yet societal needs pressure the scientific community to operationalize such research.

This push toward operational products is manifest in a proliferation of climate services activities, accelerated by initiatives such as the Global Framework for Climate Services (GFCS, Hewitt et al. 2012). Additional pressure comes from development banks and international aid organizations as they seek to inform national adaptation strategies. At the national, sub-national and local scale there are likewise project-driven downscaling activities (e.g. Janetos et al. 2010).

The climate scientist is drawn into this landscape of competing needs—an ethical swamp where scientists face serious questions about our responsibilities when delivering uncertain data and information, and so potentially precipitating real world actions. In many cases the challenge is even more fundamental: do we know when we have information that is “good enough” for informing decision making? Conversely, do we have a responsibility to withhold information because of the uncertainty—noting that by so doing we possibly contribute to amplifying consequences by preventing society from responding to climate change in a timely and appropriate manner?

Central to this dilemma is the credibility of regional climate change messages relevant to decision making. The user faces a proliferation of portals and data sets, developed with mixed motivations, poorly articulated uncertainties and weakly explained assumptions. Commonly the data are implied as information, poorly conveyed, often hard to find or access, and communicated in opaque language by ‘interface organizations’. A weak capacity to understand the information limitations leads to a propensity to over-interpret the robustness of climate information, ignores contradictions, and avoids the complications of uncertainty.

It is thus imperative that downscaling and the implications of methodological limitations are carefully considered. The consequences of the ESD assumptions and choices lead to contradictions between data sets that are not well explained (and also apply to dynamical downscaling). This is perhaps especially relevant for ESD where there is the large diversity of methods often applied for custom purposes in a specific region (e.g. Bedia et al. 2013; Goubanova et al. 2010).

3 Criteria of regional climate downscaling information

In ESD there is both implied and real information, and the final robustness of any message is subject to the interpretation of the scientists and users. Context is highly relevant, and a product can be robust at larger scales but questionable at small scales. In addition, in a quasi-deterministic system uncertainty in the range of outcomes is inherent, and thus one may identify three criteria against which ESD results should be assessed—is it: plausible, defensible, and actionable (PDA)?

- Plausible. The results are consistent with the known dynamics of the physical system.
- Defensible. There is a physical basis that can explain the ESD results. For example, a regional climate projection may show a decrease in rainfall, and there is evidence of a decrease in frontal intensity.

- **Actionable.** Defined as evidence strong enough to guide real-world decisions in the context of an accepted level of risk. Risk is necessarily subjective and context dependent; however, this does not absolve the scientist of responsibility to at least consider whether the data are robust within their own personal risk framework.

There is the temptation to try and define PDA criteria of what is / is not acceptable for a delivered data product. The reality is that the reliability of any product is inherently dependent on the intended application. In some cases it is enough for a user to know the regional sign of the average projected change (a tractable challenge), while another may require detailed specifics about the daily magnitudes (very uncertain). Additionally, the variable in question imposes constraints. For example, downscaling by bias correcting a model's output that has too few rain days is clearly an arbitrary adjustment at best (for there is no clear basis on how to add rain days), and hence gives rise to questionable solutions. Conversely, our process-based knowledge of the change in temperature is far more robust, allowing for more definitive conclusions to be drawn.

While some generalizations may be drawn (precipitation is less robust than surface temperature), there are no definitive answers to the PDA criteria, and the responsibility falls to both provider and user to understand the limits. Actionable criteria are naturally scale dependent, and to robustly address this requires substantial interaction between the climate science and user communities; the scientist needs to be fully cognizant of the user context so that they may mutually assess the degree of “actionable” information. While some sectors of society, such as the insurance industry, are proficient at making decisions with uncertain information, the delivery of downscaled data is most often not communicated with articulated limits, and there is a need to communicate how an ESD product fits the landscape of information for climate change, and the possible/probable contradictions that may exist with other products.

4 What downscaling ideally seeks to achieve

The information required is highly dependent on the application, the relevant scales in time and space, and the user's risk exposure. Figure 1 encapsulates this. For a given spatial scale the y-axis represents some measure of information, and the x-axis is the range of relevant time scales. The curves are speculative representations of the actual and required levels of information, and the theoretical limit. The nature of the information changes across the time scales from a prediction of absolute values at short lead times, to a projection of some probable mean system state or derived statistics at long lead times. For example, at weather forecasting lead times the target is the state of the atmosphere at a particular location (e.g. daily temperature) while on longer “climate” prediction lead times the target is the distribution of states (e.g. means, variances).

At any time scale there is a level of information content useful for decision making. For example, faced with the need to plan for a wet tomorrow, is the weather forecast “good enough” to predict rain or no rain? Or, at the 3 month lead time, what risk does one take by selecting a crop planting date based on the seasonal forecast. In other words, what information is “good enough” to manage the risk? By contrast is the actual skill of a forecast/projection. The current skill of short term weather forecasts may be “good enough”. However, on the 2–3 week lead time the skill is low, there is variable skill on seasonal time scales, uncertain skill on intra-decadal scales, and possibly good skill at multi-decadal scales for the statistics at some measure of spatial and temporal aggregation.

The objective with downscaling is to move the “actual skill” curve toward the curve of “good enough” information where the risk inherent in a decision becomes acceptable. However, a third curve introduces a critical issue; that of the limit to predictability. In a quasi-deterministic system there is a limit to the predictive skill (in terms of knowledge and/or tools

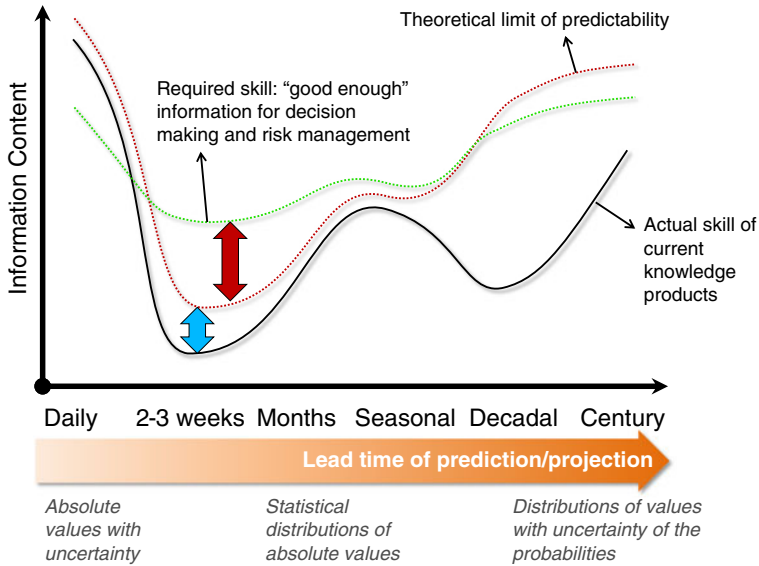


Fig. 1 Idealized representation of conceptual information issues in relation to using climate information for a given scale, variable, metric, and application. The curves are hypothetical, and in practice each line is a zone of gradation, but is represented here as a simple line for clarity (after Landman et al. 2010)

and methods). If this limit is less than the required skill, then there is a fundamental constraint to developing actionable climate information.

A point on uncertainty bears mentioning here. Uncertainty on the shorter lead times can be quantified in probabilistic terms, whereas uncertainty on the longer lead times is an issue of uncertainty about the probabilities. For example, we are less certain about the absolute global mean temperature 100 years from now, than we are about the global mean temperature 20 years from now. However, we have more confidence that the global mean temperature in 100 years, as opposed to 20 years, will be warmer than at present. If our decisions are sensitive to knowledge regarding *changes* to climate rather than absolute values of the climate state (such as the annual mean temperature) then we have more actionable information on the longer prediction lead times.

5 Frameworks for using regional downscaling information

There is a temptation to adopt a single source of information thus limiting the spread of possibilities to consider. This may be due to limited accessibility to information or a preference for a favourite model, or most importantly the challenge of understanding and distilling clear messages from a diverse and contradictory evidence base. From a physical science perspective it is clear that no one source can provide the definitive message. GCMs provide skill on synoptic circulation and larger scale processes that are informative of regional change, but are weak at representing local scale surface climate (e.g. Räisänen 2007; Radić and Clarke 2011). Moreover, GCMs produce area averages and these are fundamentally different to surface observations. Likewise the historical record provides information about local trends, but extrapolation of these trends is problematic. Thus, drawing on multiple sources is imperative for achieving robust understanding of local change and even then is no guarantee that actionable information will be achieved.

The Intergovernmental Panel on Climate Change (IPCC) 4th Assessment report (AR4) suggests a fourfold approach for assessing regional change (Christensen et al. 2007): Atmosphere-Ocean Global Climate Models (AOGCMs); downscaling of AOGCMs; physical understanding of the circulation changes; and recent historical climate change. While the literature at that time did not support a comprehensive implementation of this approach, it captures the essential elements needed to form a clear regional message. Figure 2 represents the AR4 approach and expands on this to suggest how it may be best applied.

The foundations of this assessment framework are:

- A historical record that is uncertain: The spatial and temporal paucity of observations has led to multiple data products that agree on the essentials of historical change, but deviate amongst each other at local scales. For example, the TRMM satellite offers global observations of rainfall, but contains regional biases compared to other observational sources (e.g. Gopalan et al. 2010).
- System dynamics and process understanding: The global modes of variability, teleconnections, and synoptic scale dynamics directly inform our understanding of how local and regional climates are established and may change.
- AOGCM data: Simulations with the coupled AOGCMs give a view of the large-scale changes. The local scale skill is poor, particularly with regards to the surface diagnostic variables (such as the temporal and spatial distribution of precipitation). However, at aggregate scales of time and space, the models do provide a coherent, large-scale picture of change, albeit with uncertainty in absolute magnitudes.
- Downscaled data: This source promises a region-specific high resolution and scale relevant data product conditioned by the AOGCM simulations, and is often the source for driving impacts modeling and adaptation decision making. In practice there remains great disparity between the results from different methods and approaches (e.g. Wilby et al. 2000).

None of these sources are without error, but collectively they represent a means for developing defensible information and regional integrated messages. These may be storylines of change based on the qualitative assessment of multiple lines of evidence, or numerical data

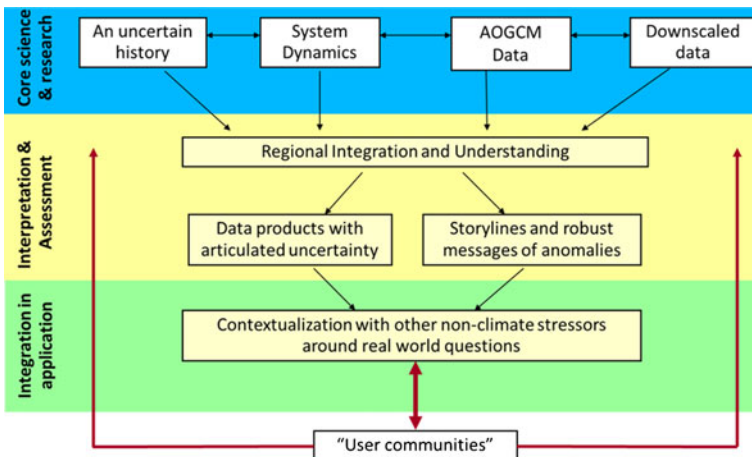


Fig. 2 A conceptual framework to consider the development of regional scale climate change messages and data

for use with impact models. Interfacing this with the concerns of the user community adds the context and relevance, and feeds back to inform continuing research.

6 Downscaling approaches

There is a wide diversity of techniques termed downscaling, although with varied degrees of adherence to the concept of a cross-scale transformation that adds information. Some typologies have been proposed, for example, a three-way categorization by Wilby and Wigley (1997) and a different three-way split by Maraun (2013). However, the reality is that the intermingling of concepts, and the imperative to capture stochastic and deterministic variance, leads to no neat classification. Table 1 shows the general components in downscaling methods (which in any given application may be used in some combination, e.g. Hewitson and Crane 2006; Jeong et al. 2012) and the associated issues. Reviews of the methods outlined in Table 1 may be found in, for example, Bürger et al. 2012; Gutiérrez et al. 2013; Christensen et al. 2007).

Each method's characteristics can influence, to a greater or lesser degree, the robustness of the downscaled outcomes. For ESD and dynamical downscaling, stationarity (or non-stationarity) is a key issue that needs careful consideration and refers to whether the relationship between the predictand and the predictors stay constant in time (e.g. Zhang et al. 2013; Schmith 2008). Climate change is manifest principally through changes in persistence, intensity, frequency, and recurrence intervals of synoptic weather events, modulated by other factors such as feedbacks or land use dynamics. Thus ESD, properly formulated, should accommodate most of the climate-related changes so long as the future synoptic forcing stays within the bounds of the training data. However, a difficult issue for ESD to accommodate is the impact of enhanced greenhouse gases (GHGs) on temperature. If source region air mass characteristics change, some ESD approaches will not reflect this, depending on the predictors used (perhaps most relevant to the analog approach). Second, the direct radiative forcing from increased GHGs may be excluded by a method's choice of predictors. This refers to the fact that aside from changes in the large scale circulation features, there is an expected increase in the base temperature of a feature due to enhanced GHG forcing. This change may not be reflected in the common choice of predictors.

The assumptions underlying ESD approaches, and the choices made in the implementation, thus have the potential to introduce significant biases and errors in the resultant downscaled product in a way that is subtle, often unrecognized, and challenging to manage. Consequently, in choosing, evaluating and implementing a downscaling method it is important to understand how the method handles attributes that may be of critical importance to subsequent use of the data in impacts and adaptation work.

Table 2 outlines a series of key questions that users should consider if they are to be confident in the outcomes. Of a more philosophical, as well as decision relevant nature, is the question of whether a method should attempt to correct GCM predictor bias error, or instead convey the error in the downscaled solution? The predictors are intimately coupled to the internal dynamics and feedbacks of the host GCM, and thus not independent of the multi-scale system processes. ESD methods that seek to mask or correct the predictor bias ignore the underlying interdependency of the variables in the GCM dynamics. For example, in a quantile mapping bias correction approach the GCM precipitation is quantile mapped to the observed data. Yet the GCM precipitation is inextricably bound to the complexity of the driving GCM physics, dynamics, and parameterizations. As such, when applied in the context of a future climate projection, the mapping function assumes stationarity of both the scale relationship and the sources of the GCM internal errors. Thus bias correction approaches can at best be considered an adjustment of uncertain veracity for future climate projections. Similarly,

Table 1 General methods in approaches commonly referred to as downscaling, and important characteristics associated with each

Method	Characteristics
Weather generator	<ul style="list-style-type: none"> • Needs long observation time series for training. • Parameters can be conditioned by a GCM. • There are potential issues of stationarity in the model parameters. • May have issues with low frequency variability. • There are challenges in handling direct radiative forcing of enhanced GHGs.
Transfer functions	<ul style="list-style-type: none"> • Can use shorter time series if the data spans the full range of possible atmospheric states at that location. • There are potential stationarity issues in the transfer function parameters. • Usually underestimates extremes and high frequency variance. • There are challenges in handling direct radiative forcing of enhanced GHGs.
Index approaches	<ul style="list-style-type: none"> • Vulnerable to non-stationarity of the index. • Not directly tied to physically interpretable predictors. • There are challenges in handling direct radiative forcing of enhanced GHGs.
Analogue patterns	<ul style="list-style-type: none"> • Not strictly downscaling. • Needs a long observation time series. • The method is constrained to only reproduce historical patterns. • There are challenges in handling direct radiative forcing of enhanced GHGs. • There are possible problems with differential spatial responses to global change.
Perturbed observed (Delta method)	<ul style="list-style-type: none"> • Not strictly downscaling. • Reflects the signal from a GCM grid cell diagnostic variable. • Does not add additional climate change signal information; only adds observational spatial detail.
Quantile mapping / bias correction	<ul style="list-style-type: none"> • Not strictly downscaling. • Reflects the signal from a GCM grid cell diagnostic variable. • Does not add additional climate change signal information; only adds observational spatial detail. • Difficult to justify the bias correction of some parameters, e.g. rain-day frequency. • Bias correction is only strictly true for historical period, and is a bias adjustment for the future.
Regional climate model	<ul style="list-style-type: none"> • Dependent on the quality of the lateral boundary conditions (LBCs). • Develops (mostly) its own internal domain climate. • Complicated by lateral boundary effects. • Sensitive to domain size, resolution, and boundary location. • Has dependency on the model's parameterization skill and stationarity, and possible parameterization compatibility issues with the driving GCM. • Output usually requires (problematic) bias correction.

Table 1 (continued)

Method	Characteristics
High resolution or variable resolution global model forced by Sea surface temperatures from a host AOGCM	<ul style="list-style-type: none"> • Not strictly downscaling, but a high resolution atmosphere-only GCM. • Forced only by sea surface temperatures from a coupled AOGCM (may include nudging). • Develops its own climate independent of the source GCM. • Has dependency on the model's parameterization skill and stationarity. • Output usually requires (problematic) bias correction.

consider a weather generator trained to represent the temporal sequence of the observed data, and then conditioned by a GCM to project the climate change. If the GCM temporal sequencing is markedly different from the observed data, even though the GCM mean and variability may be well simulated, then will the method hide the temporal inaccuracies of the GCM, or convey these inaccuracies, and which should it legitimately do?

We argue that for PDA purposes, the defensibility of a downscaled solution should be contingent on including and representing GCM error. Failing to do so communicates an unsubstantiated confidence in the strength of the product, and obscures additional sources of uncertainty.

7 Illustration of the impact of implementation choices

Contradictions between different ESD approaches are common and arise partly out of assumptions and choices in implementation. The contradictions are poorly explored by the Impacts, Adaptation and Vulnerability (IAV) community, and it is difficult to determine if one source, be it which model or method, is more reliable for projecting the future. There is thus a natural disinclination to use multiple information sources and this reflects a desire to avoid additional complications. Yet doing so greatly increases the risk of using an outlier or erroneous single source.

Here we show an example of how a simple choice in implementation can complicate the objective of reaching actionable information on regional scales. In each method there are numerous options available that combine at least some of the characteristics listed in Table 1. Common to all ESD methods is the use of an observational training data set, and we use this aspect to illustrate the way methodological choices can propagate and influence the results. The example here uses the downscaling method described in Hewitson and Crane (2006), which combines a weather classification approach with a stochastic generator to capture both the deterministic and stochastic variance.

Two gridded data sets of the recent historical climate are used for training two realizations of the same ESD method. Inherently, any gridded product of surface climate variables is a post-processed data set that draws on station data to constrain or inform the product. Here we use two of many gridded observational data sets. First is the WATCH WFDEI data set (Weedon et al. 2011), created in support of the hydrological and land surface modeling communities. The data are based on the ERA-Interim reanalysis with corrections applied using the CRU TS2.1 monthly gridded observations (Mitchell and Jones 2005). Second is the CFSR Reanalysis data set (Saha et al. 2010), a current-generation high resolution global reanalysis. One may debate the relative merits of using reanalysis-based predictands, however,

Table 2 Key questions of importance for impacts modeling and adaptation relevant to downscaling

Question	Concern
Does the method appropriately reflect low frequency variability (interannual to trend)?	A method may be designed to capture time scales of considered importance, for example, daily precipitation. In doing so it is possible to miss critical longer time scales of variability. For contexts where interannual variability is important, this may present a serious deficiency.
Does the method appropriately reflect high frequency variability (e.g. daily variability and extremes)?	Methods that generalize the relationship between large scale forcing and local response do not capture the high frequency variance, and especially extremes. If one is working with an impacts sector sensitive to the event nature of climate, this could be critical.
How does the method respond to non-stationarity in the predictors?	Downscaling is trained on the historical data, and under climate change the predictors will likely exceed the range of values used in the training. Methods may respond to non-stationarity in a number of ways: extrapolation (which carries low robustness), conservative estimation (where the sign of the change is reflected but the magnitude is underestimated), or with unpredictable values resulting from nonlinearities or the method being unable to accommodate the extended predictor range.
Is temporal autocorrelation of the predicted values representative of the predictors temporal sequencing (relates also to response to non-stationarity).	In many impacts related studies, the sequencing of weather events is of critical importance. For example, in hydrological modeling, the sequencing of rain events and the wet and dry spell duration has a critical impact on surface hydrology, regardless of whether the mean changes. Alternatively, the joint impact of temperature and precipitation event sequences can substantially impact agriculture. Likewise, recurrence intervals of extreme events are vital to disaster risk management.
Is spatial autocorrelation of the predicted values representative of the real-world spatial autocorrelation (relates also to response to non-stationarity).	In many impact studies the spatial relationships of the data strongly influences the response of the system being investigated. For example, water flows downhill through a catchment, and if the spatial pattern of downscaled precipitation is erroneous, the catchment flow as simulated by a hydrology impacts model will be wrong.
Do the predictors represent all sources of climate change forcing?	Predictor selection can introduce subtle influences. If selected on the basis of correlation and/or explained variance in the predictand, even perhaps using an optimization method, then implicitly the predictors are tuned to the current climate. Two challenges arise. First, the stationarity question of whether the predictors remain important under a future climate (especially if they are geographically remote to the target location). Second and more importantly, do the selected predictors include the ones that exhibit strong change into the future? Correlative and optimization procedures for the selection of predictors based on current climate relationships could exclude critical predictors of the future change signal.
How does the method respond to changes in the relative contribution of predictors under a future	The relative importance of predictors can change under a future climate. For example, the role of

Table 2 (continued)

Question	Concern
climate? (Especially if the predictors are optimized during training for each location).	atmospheric humidity: this is likely to increase in the future, and can reasonably be expected to have a non-linear response in importance compared to other predictors. The sensitivity of a method to such aspects thus needs to be assessed.
Does the method accommodate the local changes in direct response to the enhanced GHG radiative forcing, as well as the changes in the characteristics of non-local source air mass changes?	As discussed earlier, this is a subtle issue that can influence the downscaling of temperature. Understanding how a method accommodates this is critical to evaluating the robustness of temperature downscaling.

the surface variables are consistent with the atmospheric dynamics, and as such are legitimate predictands consistent with the synoptic predictors.

Figure 3 shows the mean differences between the two historical data sets for the period 1979–2009 (WFDEI-CFSR). This highlights that our understanding of the past is somewhat

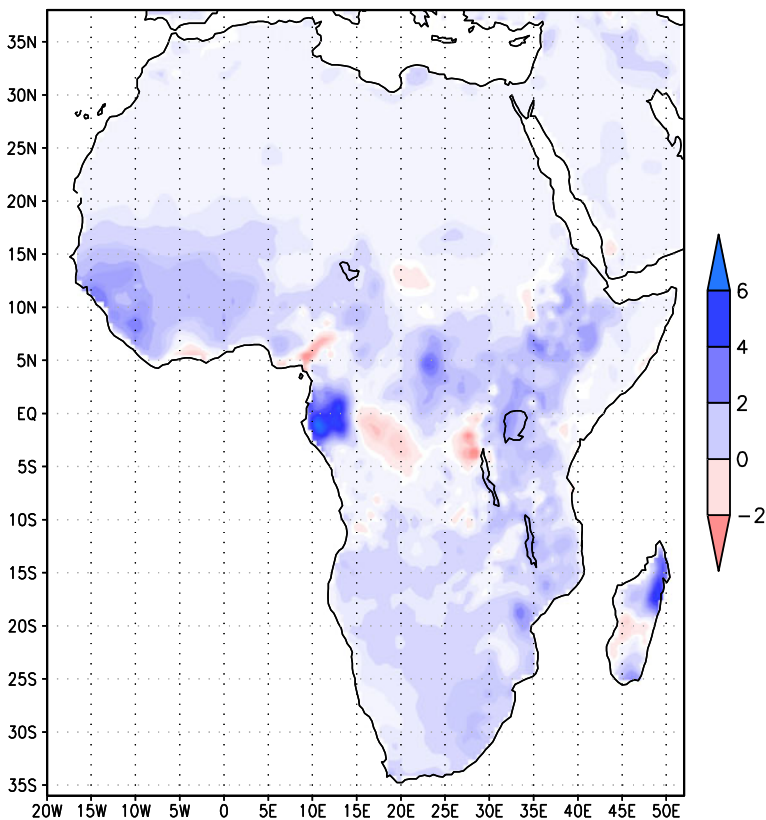


Fig. 3 Difference in the mean precipitation per day (mm/day) between the two historical data sets (WFDEI-CFSR) for the period 1979 to 2009

probabilistic—we do not have a “true” reference. Even dynamical downscaling is not immune to this problem, as a reference data set is needed to evaluate the RCMs.

The downscaling was trained for all terrestrial grid cells (0.5° degree resolution) across Africa, using daily precipitation for the period 1979–2009 as the predictand. Two separate trainings were completed, one with each of the WFDEI and the CFSR data. Following this, the differently trained downscaling is applied to a common GCM data set from the CMIP5 (Taylor et al. 2012) multi-model ensemble archive. The GCM used is the Canadian Centre for Climate Modelling and Analysis CanESM2, forced with the RCP 4.5 GHG concentrations scenario (Meinshausen et al. 2011).

Figure 4 shows the climate change results of the GCM-forced downscaling. This should not be interpreted as a robust regional projection for Africa and shows only the downscaling of one GCM by one method. The important result is that the differences are attributable only to the choice of predictand training data.

The results, at the large sub-continental scale are spatially similar, although magnitudes differ as a function of location. More concerning is that, for some locations, the sign of the projected change is reversed; most notably at a number of locations in coastal West Africa. Considering that downscaling is intended to inform local scale impacts and adaptation, these differences from the simple choice of training data are cause for concern. Additionally, for some impacts sectors where the system is strongly dependent on the spatial pattern, such as hydrology, the differences in the spatial variation of magnitude alone would lead to different impact consequences, especially as surface run-off magnifies changes in precipitation by factors of 2–5 or more (Schulze 2000).

8 Considerations for designing an idealized ESD

As the activities of climate services gain momentum there is a compelling need to address the issues raised here. For ESD two frameworks are needed: first are criteria for developing

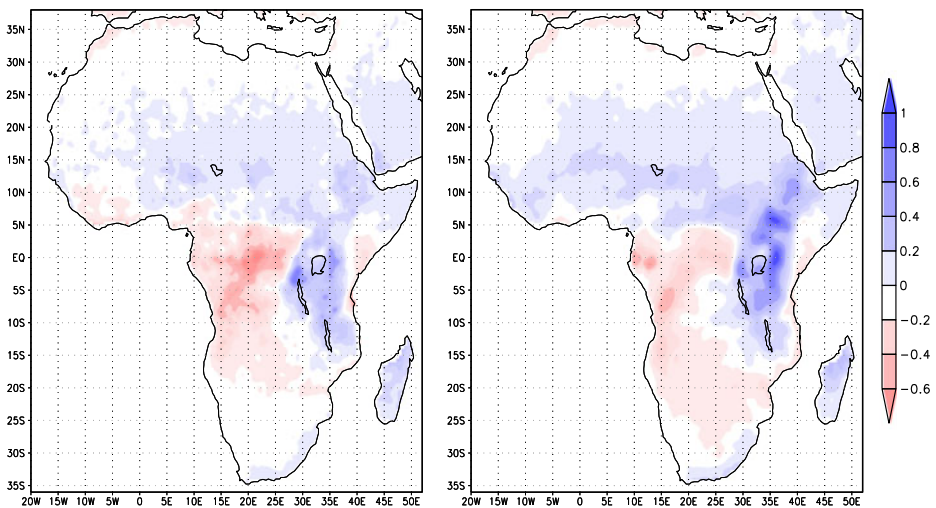


Fig. 4 Projected downscaled climate change anomaly of mean daily precipitation (mm/day) where the downscaling was trained using the CFSR (*left*) and WFDEI (*right*) gridded estimates of historical climate. The change is the difference between the mean of two 30 year periods: 2071–2100–1981–2010. The driving GCM is the CanESM2 forced by the RCP 4.5 GHG scenario from the CMIP5 archive

defensible downscaled solutions, and second is a set of standard metrics to form the basis of ESD assessment. For establishing optimal downscaling criteria, we recognize that the local climate is a result of the dynamics of the co-located synoptic weather systems. In this respect we may identify three issues. First, ESD predictors that do not fully incorporate the local atmospheric dynamics implicitly assume stationarity with whatever other indirect predictors are chosen. Second, the synoptic dynamics represent variance of short time scale processes, and any ESD that uses time-aggregate predictors (such as monthly means) assumes that the relation between the high frequency events and the time-aggregated predictor are stationary. Third, the low frequency modulation of local climate through seasonal forcing and natural interannual variability are an important source of variance, and an ESD insensitive to this will provide erroneous information on changes in the low frequency variability.

A good case may be made that these assumptions are likely to be non-stationary to some degree under a future climate. There is no obvious reason to expect teleconnections to remain spatially stable, that daily weather events will maintain the same distribution of frequency, persistence and intensity, or that the seasonal timing and interannual variance will remain the same.

Thus, in an idealized ESD, be it one that directly uses predictors (e.g. transfer functions) or one that is only conditioned by the predictors (e.g. weather generator), it may be postulated that an optimal approach will (see also, for example, the IPCC guidance Document by Wilby et al. 2004):

- Use predictors that clearly capture the local thermodynamic state of the atmosphere through which the climate change signals are communicated.
- Use predictors that are related to a GCMs scale of skill.
- Capture the range of frequency response from interannual variability to at least daily synoptic time scales.
- Be conservative under non-stationarity; that is, under non-stationarity the ESD would at worst underestimate the response, but not give a response physically inconsistent with the atmospheric predictors.

9 A framework of evaluation

With the broad range of ESD approaches it is exceptionally difficult to evaluate one set of results against another, or to assess whether contradictions are inherent representations of the climate system, or are artifacts of the methodology. There is thus a need for a set of metrics that serve the end user's interests, and in particular diagnostic metrics to evaluate:

- Performance across the range of temporal scales of variability—that is, does the method capture local climate response to the predictors from low frequency (inter-annual/trend) to high frequency (daily/extremes)?
- Comparison with RCMs. This is the generic issue of comparing apples and oranges—for example, in the case where ESD is used to downscale point locations in contrast to the RCM grid cell scale.
- Homogeneity in time. For example, it is not uncommon for ESD approaches to subset the annual cycle into discrete training periods for each sub-season. This introduces the potential for jumps in downscaled response through the seasonal cycle and questions around stationarity if the seasonal cycle changes.
- Spatial and temporal autocorrelation. For some impact modeling sectors this can be critical when translated into, for example, agricultural or hydrological processes.

- Sensitivity and response to non-stationarity in the predictor-predictand relationship.
- Representation of extremes in terms of events as well as the extreme duration of a climate state (e.g. drought, wet spells, heat waves).
- Covariate response. Does, for example, temperature and precipitation realistically correspond to the predictors?

The design and application of such a set of metrics requires ESD to produce data in a standard form comparable with dynamical downscaling. In the case of assessing the temporal variance of the ESD, the metrics also presuppose at least a daily temporal resolution to the downscaling.

10 Closing the gap of contradictions

There is a growing move toward populating multi-model matrices of GCM-downscaling combinations. For example, as is done in the PRUDENCE (Christensen et al. 2002), ENSEMBLES (Hewitt 2004), NARCCAP (Mearns et al. 2012), and CORDEX (Nikulin et al. 2012) projects. In its fullest expression the objective would be to complete a 4-dimensional matrix of the combination of GCMS, RCMs, ESD methods, and GHG emission scenarios. While such a matrix is far from complete for any region, there is a sufficiently large sample to reveal significant contradictions within and between the results for GCMs, RCMs and ESD. These contradictions are probably the single biggest factor undermining the confidence in regional and local scale projections.

In many cases, the end user is poorly equipped to evaluate these contradictions, which leads to either loss of confidence in any climate message, or the selection of a favored result from a trusted source. In actuality the contradictions are important for understanding the sources of uncertainty in regional projections. Contributing to this, the greatest confusion results from the diversity of ESD methods that downscale to an inconsistent set of variables, time scales and spatial resolutions. Consequently, it is imperative to evaluate ESD methods, and where ESD produces output similar to RCMs, to evaluate the differences between RCM and ESD approaches.

The effective and appropriate use of regional climate information to support adaptation decision making depends on first resolving these issues within the modeling community. Moving forward to develop rigor and robustness in regional projections requires a shift away from the ad-hoc approach to developing products to service finite project needs.

The current state of affairs is not unexpected; ESD has evolved in response to needs, and consequently lags the user expectations. Thus the adoption of data simply on the basis of access and availability is natural, but is clearly not sustainable as the issues of data integrity emerge. This argues for the need to formalize the development and delivery of ESD output, and invest in creating metrics and guidance as seen from a user's perspective.

The WCRP CORDEX activity likely offers the best opportunity to address these issues. With a global focus, consistent experimental framework, controlled vocabulary, and quality controlled publication of data on the Earth System Grid (Bernholdt et al. 2007), the CORDEX programme enables a rigorous evaluation of the results. The CORDEX opportunity builds on, but is different to, activities such as STARDEX⁴ which contributed to the growing understanding that no one method is best. CORDEX offers a well-defined and standardized experimental framework for method comparison, with side-by-side evaluation of RCMs and ESD, and which critically covers all terrestrial regions to enable assessment of methods across polar, mid-latitude, and tropical climate regimes.

⁴ http://www.cru.uea.ac.uk/projects/stardex/reports/STARDEX_FINAL_REPORT.pdf

Through CORDEX there is potential to obtain an in-depth understanding of the strengths and limits of different methods, and for the coherent development of defensible and actionable messages at the local scale drawing on the perspectives of a global range of users. The dynamical downscaling activity in CORDEX, at least in terms of undertaking simulations, is well advanced. The ESD component of CORDEX lags, but is gaining momentum. The critical gap is in comparative analysis of the results. Arguably the leading challenges are the design of metrics from the user perspective, comparison and evaluation of multi-method data, and the identification and articulation of signal versus noise as a function of scale. Achieving these objectives, or at a minimum making progress on understanding how to deal with these, is essential if downscaling is to defensibly contribute to actionable information at user relevant scales. User-driven collaboration between downscaling scientists and the various types of user communities is essential for a co-exploration of the data, not only to determine what information it is possible to derive, but also what the limits are to its utility under different risk scenarios. On the other hand, business as usual avoids responsibility, raises questions of accountability, and ignores ethical concerns.

Acknowledgments We are grateful for helpful comments received from three anonymous reviewers, and from W. Gutowski and R. Benestad on the final draft of the paper. We also acknowledge the WCRP's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the Canadian Centre for Climate Modelling and Analysis for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Bedia J, Herrera S, Martín DS et al (2013) Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. *Climate Change*. doi:10.1007/s10584-013-0787-3
- Bernholdt D, Bharathi S, Brown D et al (2007) The Earth system grid: Supporting the next generation of climate modeling research
- Bürger G, Murdock TQ, Werner AT et al (2012) Downscaling extremes—an intercomparison of multiple statistical methods for present climate. *J Clim* 25:4366–4388. doi:10.1175/JCLI-D-11-00408.1
- Christensen JH, Carter TR, Giorgi F (2002) PRUDENCE employs new methods to assess European climate change. *EOS Trans Am Geophys Union* 83:147. doi:10.1029/2002EO000094
- Christensen JH, Hewitson B, Busuioic A, et al (2007) Regional climate projections. *Clim Chang* 2007 Phys Sci Basis. Contrib. Work. Gr. I to Fourth Assess. Rep. Intergov. Panel Clim. Chang
- Gopalan K, Wang N-Y, Ferraro R, Liu C (2010) Status of the TRMM 2A12 land precipitation algorithm. *J Atmos Ocean Technol* 27:1343–1354. doi:10.1175/2010JTECHA1454.1
- Goubanova K, Echevin V, Dewitte B et al (2010) Statistical downscaling of sea-surface wind over the Peru-Chile upwelling region: diagnosing the impact of climate change from the IPSL-CM4 model. *Clim Dyn* 36:1365–1378. doi:10.1007/s00382-010-0824-0
- Gutiérrez JM, San-Martín D, Brands S et al (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *J Clim* 26:171–188. doi:10.1175/JCLI-D-11-00687.1
- Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. *Clim Res* 7:85–95
- Hewitson BC, Crane RG (2006) Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa. *Int J Climatol* 26:1315–1337
- Hewitt CD (2004) Ensembles-based predictions of climate changes and their impacts. *EOS Trans Am Geophys Union* 85:566. doi:10.1029/2004EO520005
- Hewitt C, Mason S, Walland D (2012) The global framework for climate services. *Nat Clim Chang* 2:831–832. doi:10.1038/nclimate1745

- Janetos AC, Collins W, Wuebbles D et al (2010) Climate change modeling and downscaling: issues and methodological perspectives for the U.S. national climate assessment. NCA Rep Ser 7:50
- Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P (2012) Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator. *Climate Change* 114:567–591. doi:10.1007/s10584-012-0451-3
- Landman W, Engelbrecht F, Landman S (2010) I get all the news I need on the weather report. Pretoria: Council for Scientific and Industrial Research. Retrieved from <http://hdl.handle.net/10204/6959>
- Mahmood R, Babel MS (2012) Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theor Appl Climatol*. doi:10.1007/s00704-012-0765-0
- Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J Clim* 26: 2137–2143. doi:10.1175/JCLI-D-12-00821.1
- Maraun D, Wetterhall F, Ireson AM et al (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48, RG3003. doi:10.1029/2009RG000314
- Meams LO, Arritt R, Biner S et al (2012) The North American regional climate change assessment program: Overview of phase I results
- Meinshausen M, Smith SJ, Calvin K et al (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climate Change* 109:213–241. doi:10.1007/s10584-011-0156-z
- Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* 25:693–712. doi:10.1002/joc.1181
- Nikulin G, Jones C, Giorgi F et al (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *J Clim* 25:6057–6078. doi:10.1175/JCLI-D-11-00375.1
- Palmer TN (2000) Predicting uncertainty in forecasts of weather and climate. *Rep Prog Phys* 63:71–116. doi:10.1088/0034-4885/63/2/201
- Pielke RA Sr, Wilby RL (2012) Regional climate downscaling: what's the point? *EOS (Washington DC)* 93:52–53. doi:10.1029/2012EO050008
- Pierce DW, Das T, Cayan DR et al (2012) Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Clim Dyn* 40:839–856. doi:10.1007/s00382-012-1337-9
- Radić V, Clarke GKC (2011) Evaluation of IPCC Models' performance in simulating late-twentieth-century climatologies and weather patterns over North America. *J Clim* 24:5257–5274. doi:10.1175/JCLI-D-11-00011.1
- Räisänen J (2007) How reliable are climate models? *Tellus A* 59:2–29. doi:10.1111/j.1600-0870.2006.00211.x
- Saha S, Moorthi S, Pan H-L et al (2010) The NCEP climate forecast system reanalysis. *Bull Am Meteorol Soc* 91:1015–1057. doi:10.1175/2010BAMS3001.1
- Schmith T (2008) Stationarity of regression relationships: application to empirical downscaling. *J Clim* 21:4529–4537. doi:10.1175/2008JCLI1910.1
- Schulze R (2000) Transcending scales of space and time in impact studies of climate and climate change on agrohydrological responses. *Agric Ecosyst Environ* 82:185–212
- Slingo J, Palmer T (2011) Uncertainty in weather and climate prediction. *Philos Trans A Math Phys Eng Sci* 369: 4751–4767. doi:10.1098/rsta.2011.0161
- Stoner AMK, Hayhoe K, Yang X, Wuebbles DJ (2012) An asynchronous regional regression model for statistical downscaling of daily climate variables. *Int J Climatol*. doi:10.1002/joc.3603
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498. doi:10.1175/BAMS-D-11-00094.1
- Weedon GP, Gomes S, Viterbo P et al (2011) Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J Hydrometeorol* 12:823–848. doi:10.1175/2011JHM1369.1
- Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Prog Phys Geogr* 21:530–548. doi:10.1177/030913339702100403
- Wilby RL, Wigley TML, Conway D et al (1998) Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour Res* 34:2995–3008. doi:10.1029/98WR02577
- Wilby RL, Hay LE, Gutowski WJ et al (2000) Hydrological responses to dynamically and statistically downscaled climate model output. *Geophys Res Lett* 27:1199–1202. doi:10.1029/1999GL006078
- Wilby R, Charles S, Zorita E et al (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods
- Zhang X-C, Chen J, Garbrecht JD, Brissette FP (2012) Evaluation of a weather generator-based method for statistical downscaling non-stationary climate scenarios for impact assessment at a point scale. *Trans ASABE* 55:1745–1756