Choice of baseline climate data impacts projected species' responses to climate change

Running head: Baseline climate data uncertainty

Baker, D.J.[1], Hartley, A.J.[2], Butchart, S. H. M.[3,4] & Willis, S.G.[1]

[1] School of Biological & Biomedical Sciences, Durham University, Mountjoy Site, Durham, DH1 3LE, UK
[2] Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB, UK
[3] BirdLife International, David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK
[4] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB23EJ, UK

Corresponding author: David J. Baker
Email: d.j.baker@durham.ac.uk
Tel: +44 1913 341 327

Keywords: Species Distribution Model, Satellite Remote Sensing, Change Factor Method, Downscaling, General Circulation Models, Important Bird and Biodiversity Areas, Sub-Saharan Africa

Type of paper: Primary Research Article

Abstract

Climate data created from historic climate observations are integral to most assessments of potential climate change impacts, and frequently comprise the baseline period used to infer species-climate relationships. They are often also central to downscaling coarse resolution climate simulations from General Circulation Models (GCMs) in order to project future climate scenarios at ecologically relevant spatial scales. Uncertainty in these baseline data can be large, particularly where weather observations are sparse and climate dynamics are complex (e.g. over mountainous or coastal regions). Yet, importantly, this uncertainty is almost universally overlooked when assessing potential responses of species to climate change. Here we assessed the importance of historic baseline climate uncertainty for projections of species' responses to future climate change. We built species distribution models (SDMs) for 895 African bird species of conservation concern, using six different climate baselines. We projected these models to two future periods (2040-2069, 2070-2099), using downscaled climate projections, and calculated species turnover and changes in species-specific climate suitability. We found that the choice of baseline climate data constituted an important source of uncertainty in projections of both species turnover and species-specific climate suitability, often comparable with, or more important than, uncertainty arising from the choice of GCM. Importantly, the relative contribution of these factors to projection uncertainty varied spatially. Moreover, when projecting SDMs to sites of biodiversity importance (Important Bird and Biodiversity Areas), these uncertainties altered site-level impacts, which could affect conservation prioritisation. Our results highlight that projections of species' responses to climate change are sensitive to uncertainty in the baseline climatology. We recommend that this should be considered routinely in such analyses.

Introduction

The effects of climate change on the distribution and abundance of species are already being observed (Chen *et al.*, 2011; VanDerWal *et al.*, 2013), with increasing evidence of long-term climate trends driving changes in populations across a range of ecological systems (Cahill *et al.*, 2013). Climate change, along with changes in patterns of land use, is likely to be a major driver of biodiversity loss over the coming centuries. Species with narrow climate tolerances and low capacity to adapt to novel conditions are likely to be particularly affected (Foden *et al.*, 2013). Consequently, a major conservation priority is to develop an understanding of how populations are affected by climate variability and long-term change (Ockendon *et al.*, 2014; Pearce-Higgins *et al.*, 2015), and to develop modelling frameworks to predict potential climate change impacts on biodiversity in order to inform conservation management (e.g. Kearney & Porter, 2009; Dullinger *et al.*, 2012; Foden *et al.*, 2013).

The dominant methodological approach used to assess potential climate change impacts on species has been the development of statistical models that aim to describe a species' relationship (in terms of, for example, distribution or abundance) to climate (Pacifici *et al.*, 2015). These correlative species distribution models (SDMs) can be used in conjunction with simulated future climate data to project likely responses to climate change (Elith & Leathwick, 2009). Such models are primarily aimed at assessing likely geographical shifts in climate suitability (Willis *et al.*, 2015) and, in their simplest form, do not directly consider species-specific traits that might affect climate change vulnerability (e.g. dispersal ability or demography). As a consequence, other approaches have been developed that attempt to directly incorporate demographic processes (Dullinger *et al.*, 2012), physiological limits (Kearney *et al.*,

2008) and species-specific traits (Foden *et al.* 2013) into assessments of future climate change impacts on species.

When evaluated, uncertainty in projected species responses to climate change (e.g. range shifts, changes in abundance) tend to be high, with the dominant sources of uncertainty including variability among future climate projections, modelling methodologies, choice of climate predictor variables and the underlying biodiversity data (Dormann *et al.*, 2008; Buisson *et al.*, 2010; Synes & Osborne 2011; Cheaib *et al.*, 2012; Bagchi *et al.*, 2013). Most studies quantify uncertainty from choices made during the modelling processes, for example, by using future climate projections derived from several different General Circulation Models (GCMs) and using multiple SDM techniques (Araújo *et al.*, 2011; Garcia *et al.*, 2012). The range of responses that might result from different future greenhouse gas emissions scenarios is frequently assessed using data from GCMs run under multiple scenarios. These projection ensembles can then be used to estimate the likely range of species or community responses to climate change across the range of known uncertainty (Araújo & New, 2007; Bagchi *et al.*, 2013; Baker *et al.*, 2015). However, much methodological and data uncertainty remains unaccounted for in such model ensembles, including biological effects (Willis *et al.*, 2015). The influence of this uncertainty on projections of species' responses to climate change and the effectiveness of conservation planning is itself uncertain (Carvalho *et al.*, 2011; Kujala *et al.*, 2013).

Historic gridded climate data, often referred to as 'observational' data, are central to many ecological studies, for example, to assess the importance of climate variability on population dynamics (e.g. Gregory *et al.*, 2009) or for building models to project future impacts (e.g. Bagchi *et al.*, 2013). Historic gridded climate data are also often central to the process of downscaling coarse resolution climate simulations from GCMs (typically available at 100-300km resolution) to scales of ecological relevance (typically 50km resolution or finer). The simplest and most frequently used downscaling approaches (e.g. statistical downscaling and the change factor method, CFM) apply change in a given variable simulated by the GCM, e.g. temperature or precipitation, to a finer resolution baseline climate (Wilby & Wigley, 1997; Tabor & Williams, 2010). This results in climate data with a higher spatial resolution than the GCM, although the underlying simulation of climate change is influenced only by coarse-scale output from the GCM. Thus, local scale (i.e. sub-GCM grid cell resolution) climatic characteristics are entirely dependent upon the baseline climatology used in the downscaling. If uncertainty in the baseline climatology is high, this can lead to erroneous realisations of the climatic landscape, which could, in turn, affect assessments of climate change impact on species. Few studies have even noted the potential importance associated with uncertainty among baseline climate datasets (Parra & Monahan, 2008; Roubicek *et al.*, 2010; Watling *et al.*, 2014) and no studies have yet incorporated this source of uncertainty into a regional climate change impact assessment (although, Baker *et al.* (2015) used multiple modelled baselines simulated in a regional climate model).

The most commonly used baseline climate data are derived from observation records, usually in a gridded format that represent area-based averages across grid cells, with the spatial extent of cells typically between 1km^2 and 2500km^2 (Hijmans *et al.*, 2005; Haylock *et al.*, 2008). Both ground-based and satellite observation data are used to construct these climate baselines, though ground-based observations provide the only source for long-running (pre-1970s) reconstructions. Converting these observations into a coherent gridded climate product requires considerable data processing (Haylock *et al.*, 2008; Hofstra *et al.*, 2009; Isotta *et al.*, 2014). Uncertainty in ground-based observations may arise from differences in the density of observation stations, interpolation methodology, or simple recording errors (Efthymiadis *et al.*, 2006; Hofstra *et al.*, 2010). For satellite-derived observations of the climate, uncertainties may be introduced when converting the retrieved electromagnetic signal to a physical parameter (e.g. precipitation) or by atmospheric factors that affect the signal retrieved by the satellite (Tapiador *et al.*, 2012). The methodological choices and assumptions made during the downscaling process, along with error and bias in the original observation data, often results in datasets that contain much uncertainty.

Here, we demonstrate variation among different historic climate baselines and explore how this uncertainty affects species-climate relationships and, consequently, how this impacts projections of species'(Nakicenovic *et al.*, 2000) responses to climate change. We explore climate data and produce models across sub-Saharan Africa, a region with a large spatial extent, that experiences a range of climate phenomena, and where uncertainty in the historic climate record is high (Sylla *et al.*, 2013). Within this region we examine spatial patterns in baseline climate uncertainty, where uncertainty is a measure of variation among different baseline climate datasets. We evaluate the consequences of this uncertainty on projections of climate change impacts for birds of conservation concern (birds being the best-studied class of organisms, with moderately high resolution distribution maps available for all species across the region). We evaluate the impact of baseline climate variability on projections of species-climate relationships, and we contrast the resultant uncertainty in model projections of climate impacts on species with other measures of uncertainties that are now routinely incorporated into species-climate modelling. These other sources of model uncertainty arise from the choice of GCMs and SDMs, and the uncertainty associated with using subsets of data for model fitting (which we term ‘blocks’). Currently, considerable effort is made to describe uncertainty in species-climate projections arising from, for example, GCM and SDM choices (Garcia *et al.*, 2012; Bagchi *et al.*, 2013), but variation in the baseline climate products used to construct such models has, to date, been overlooked. We contextualise the importance of baseline climate uncertainty to biodiversity conservation by evaluating projected changes to avian diversity across the network of Important Bird and Biodiversity Areas (IBAs; the largest global network of systematically identified sites that are significant for the persistence of biodiversity; BirdLife International, 2014) across sub-Saharan Africa. Finally, we discuss approaches for incorporating uncertainty associated with historic climate data into assessments of climate change impacts for biodiversity.

Materials & methods

General circulation model ensemble

GCMs are not equally capable of representing key regional climate phenomena, such as the spatial and temporal patterns of precipitation. Here, we select an ensemble of GCMs for downscaling, based on the assumption that models capable of simulating past climates with some accuracy are the ‘best candidates’ for predicting future climates (Stott & Kettleborough, 2002; Rowlands *et al.*, 2012). Thus, we selected a five-member subset of a 17-member Perturbed Physics Ensemble (PPE) of the Hadley

Centre GCM (Gordon *et al.*, 2000; Pope *et al.*, 2000). A PPE explores uncertainty in the parameterisation of the GCM by varying uncertain model parameters systematically. The five-member ensemble was selected from the PPE based on the criteria of realistically simulating the main features of the regional climate, and of capturing a range of plausible climate outcomes (McSweeney *et al.*, 2012; Buontempo *et al.*, 2014). The models were run over the global domain for the SRES A1B scenario (Nakicenovic *et al.*, 2000).

Historic gridded climate baselines (c. 1979-2009)

We selected six 'observed' gridded datasets as the historic baselines for SDM fitting and for the GCM downscaling; these represented products derived from ground observations, satellite observations and hybrid products (see Table 1 for full details). For each dataset we obtained the mean monthly temperature (T_{mean}) and the total monthly precipitation (P_{total}). In addition to readily available datasets, we also combined *TRMM*, a satellite precipitation product, with the T_{mean} from CRU TS3.1, to create a hybrid dataset, here named CRU.TRMM. We extracted baseline climate data for the period 1979-2009, where available, in observed datasets; this permitted the inclusion of satellite products. WorldClim data was only available for the period 1950-2000 (as a pre-processed product) and TRMM satellite data only for the period 1998-present. However, it was important to include these data in this analysis despite the temporal mismatch due to the popularity of WorldClim and the importance of TRMM for tropical precipitation monitoring. WorldClim shows very similar trends and magnitudes across different regions to CRU, WFDEI.CRU and WFDEI.GPCC data (Fig. 1). The magnitude of TRMM P_{total} tends to be lower than the other datasets, but this is consistent with prior evaluation (e.g. comparision with CRU, Mariotti *et al.*,

2014) and unlikely to be due to the temporal mismatch. Each gridded climate baseline was resampled onto the same grid; African CORDEX domain (longitude range = -24.64, 60.28; latitude range = -45.76, 42.24; Giorgi *et al.*, 2009) at a 0.44^{o} spatial resolution (c. 50km resolution).

Figure 2 shows the spatial patterns of the uncertainty in climate observation datasets for each season, for precipitation and temperature. Temperature uncertainty is shown by the range of values (degrees Celsius) across the climate observations. Precipitation uncertainty is shown by the coefficient of variation, calculated on the observed climate datasets. For the latter, we excluded areas where the total seasonal precipitation was less than 30mm. The 30mm threshold was intended to remove very arid areas, which may have only 1 or 2 short duration, but intense, rain events per year. In these locations, the 3-hourly repeat cycle of TRMM may be insufficient to identify the rainfall event. This means that there is a greater chance of the satellite not capturing the climate correctly in comparison with rain gauge measurements that capture the accumulated precipitation over one hour. Such a discrepancy creates an unrealistically high standard deviation in the mean, which no longer reflects differences in the observed quantities.

For modelling species distributions, we derived four bioclimate variables that showed low colinearity (correlation coefficients <0.7) and that have been related previously to species distributions (Barbet-Massin & Jetz, 2014). Bioclimatic variables are widely used in SDM analyses, and aim to describe biologically important aspects of climatic variation (Busby, 1991). We calculated the mean of T_{mean} and P_{total} for each month across the 30-year baseline time period, and used these to calculate the four bioclimate variables: annual total precipitation (annual sum of P_{total}); annual mean temperature (annual mean of T_{mean}); precipitation seasonality (coefficient of variation of P_{total}); and temperature seasonality (standard deviation of T_{mean} x 100). From this point, we refer to these climate baseline datasets as CLIM.

Downscaling GCM simulations

The five GCM simulations were each downscaled using the CFM following the method of Tabor & Williams (2010), in which the monthly absolute anomaly for each variable (from GCMs) was calculated between the baseline period (c. 1979-2009) and the two future focal periods (2040-2069; 2070-2099). This anomaly was then regridded to 0.44 degrees spatial resolution using cubic spline interpolation, and added to the observed T_{mean} and P_{total} for the baseline period to produce projections of future climate. This process was carried out using each of the six gridded CLIM datasets, to produce 30 climate projections (5 GCMs x 6 CLIM) of future climates for each time period. The bioclimate variables described above were then calculated for each future time period and projection.

Species distribution modelling

For the analysis, we selected bird species of conservation concern (BirdLife International, 2015) that have their entire breeding range within the African CORDEX domain (n = 925). The latter criterion ensured that we could model the entire species-climate response. Species of conservation concern include those classified as threatened, restricted-range, biome-restricted or congregatory species (those that trigger criteria for identifying IBAs; BirdLife International, 2014). These species were included in order to be representative of those typical of impact assessments (e.g.

Hole *et al.*, 2009; Bagchi *et al.*, 2013). Species distribution data were derived from refined species distribution maps from BirdLife International & NatureServe (2013). These distributions were gridded onto a regular grid across Africa, to match the resolution of the climate data (0.44 degree resolution). A species was considered to occur in a cell if the distribution polygon overlapped ≥10% of the cell, which is a liberal threshold that helps ensure that species with restricted ranges are represented. Due to a lack of true absence data, and because all areas beyond the range extent are extremely unlikely to contain false absences, for modelling we consider all cells beyond the range to be true absences.

We used a jack-knife approach to model the distribution of each species, that aimed to capture the contribution of several sources of uncertainty in projected species' responses to future climate conditions, and closely follows Bagchi *et al.* (2013) and Baker *et al.* (2015). The principle of the approach is to build a model using different combinations of data and modelling techniques (i.e. potential sources of uncertainty), and to use the variability in the resulting models to identify the contribution of each potential source of uncertainty to assessments of species' responses to climate change.

Firstly, the region was divided into six spatially disaggregated blocks for model building/testing (see Bagchi *et al.* 2013 for full description of blocking method). In brief, the blocking approach involves dividing the region into small subunits and then grouping these into six spatially disaggregated blocks, such that the mean and variance of each bioclimatic variable was approximately equal across the blocks (using Blocktools package in R). In model building/testing, models were built on each combination of five blocks and tested on the omitted block. This protocol: (1) reduces the potential confounding effect of spatial autocorrelation in both cross-validation and the assessment of model performance (unlike random k-fold partitioning); (2) maintains similar parameter space (e.g. the numeric range of climatic variables) in all model building and testing procedures; and (3) can be used to assess the effect of spatial autocorrelation on projected impacts.

For each species, we modelled the statistical relationship between the species' distribution and the four bioclimate variables, calculated for each of the six CLIM datasets, using each of four SDM techniques (Generalised Linear Models, GLMs; Generalised Additive Models, GAMs; Generalised Boosted Models, GBMs; Random Forests, RFs) for each of the six combinations of five blocks. For each species, a maximum of 144 models could be built, with each jack-knife combination of GCM, SDM, CLIM and block. The median area under the receiver operating characteristic curve (Area Under Curve; AUC) from across the six blocks was used to assess final model accuracy for each species, SDM, GCM and CLIM combination. The median AUC was consistently high (0.98; 95% quantiles = 0.84, 1.00). The model cross-validation procedures used to optimise each model follows Bagchi *et al.* (2013). Models were not run for a species where an excluded block contained no presences, which meant this block could not be used for cross-validation (see Bagchi *et al.* 2013). All species with breeding ranges occupying fewer than 10 cells were also omitted from the analysis due to difficulties in modelling such sparse data. From the 925 species of conservation concern, 895 had sufficiently large range extents to be included in the final analysis.

Projections of contemporary climate suitability across the entire region were made for each species and each model, by applying models to the same CLIM dataset as used for training. Projections were made to the baseline period, so that future suitability could be assessed relative to the modelled baseline suitability for consistency. Each model was used to project future suitability for a species, applying the model to the future climate projection downscaled using the same baseline climate data used in model building. For each species and time period, this resulted in a maximum of 720 future projections (CLIM [6] x GCM [5] x SDM [4] x block [6]).

The importance of baseline climate uncertainty to projected impacts

We assessed the importance of baseline climate (CLIM) uncertainty to overall uncertainty in the context of two commonly employed metrics of climate change impacts on species: species turnover and change in species-specific climate suitability. We calculated the projected species turnover in each cell for each projection combination using the Bray-Curtis index, a measure of dissimilarity between two communities. Species turnover is commonly used in climate change impact studies as a way of representing projected change in community composition through time (Hole *et al.* 2009; Buisson *et al.* 2010; Bagchi *et al.* 2013). Species turnover ($T_j[t_f]$) for each cell j was calculated between the t_0 = baseline (c. 1979-2009) and t_f = future (2040-69 or 2070-2099) from projected climate suitabilities as:

$$T_j\left[t_f\right] = \frac{\sum_{k=1}^{s} |P_{jk}[t_f] - P_{jk}[t_0]|}{\sum_{k=1}^{s} P_{jk}[t_0] + \sum_{k=1}^{s} P_{jk}[t_f]} \qquad \text{Eq. 1}$$

where, P_{jk} = suitability of species k in cell j, and s is the total number of species. This resulted in 720 projections of species turnover for each cell and time period.

The variability in projected turnover was partitioned out between the potential sources of uncertainty (GCM, SDM, CLIM and block) by modelling projected species turnover (values bounded between 0 and 1) within each cell as a function of the four potential sources of uncertainty, using generalised linear models with binomial error distribution and logistic link function. We then dropped each factor in turn from the full model and assessed the contribution of each factor to overall uncertainty (Buisson *et al.*, 2010) as:

$$P_f = \frac{D_f - D_1}{D_0} \times 100 \qquad \text{Eq. 2}$$

where, P_f = percentage of deviance explained by factor f, D_1 = deviance of full model, D_f = deviance of full model minus factor f, and D_0 = deviance of null model (intercept only). Deviance is approximately equivalent to sums of squares for generalised linear models. The percentage of deviance explained by each factor in each cell was mapped and the results presented in Fig. 3.

The change in the summed climate suitability (S_k) for each species (k) between the baseline period and each future period, which provides an index of overall change in suitability for a species, was calculated separately for each of the (max.) 720 species-specific future projections. The change in climate suitability for each of these projections is simply the summed climate suitability across all cells for the future period, minus the summed climate suitability across all cells for the appropriate baseline projection. The variability in projected S_k was partitioned following the above approach, but using a general linear model, assuming Gaussian errors, and with an identity link (Fig. 4).

Baseline climate uncertainty in a conservation context

To contextualise the contribution of baseline climate (CLIM) uncertainty to climate change impact assessments, we projected species turnover within African sub-Saharan IBAs for the 2070-2099 period. We used an approach that aims to avoid the high uncertainty that occurs when climate data is downscaled to very high resolutions for assessment of climate change impact in small spatial areas, such as protected areas (Hole *et al.*, 2009; Bagchi *et al.*, 2013). Thus, we use species-specific climate suitabilities at the resolution of the climate projections, here 50km, and assume that the suitability within an IBA is broadly characterised by the suitability of the cell(s) in which the IBA is embedded. The methodology follows that of Baker *et al.* (2015). Thus, species turnover was calculated (using Eq. 1, but redefining j = IBA) for each IBA using a weighted mean of the species-specific climate suitability for the cell(s) that are intersected by the IBA, with weights equal to the percentage of the IBA's extent that overlaps the cell(s). Turnover was calculated separately for each of the 720 jack-knifed combinations, and then the ensemble mean species turnover for each IBA was calculated for each set of projections built using the same CLIM. This is similar to Hole *et al.* (2009), where the mean across climate projections was calculated.

We use the ensemble mean projected turnover for models built using CRU climate data as a reference projection with which to compare turnover from the other ensemble projections made using different CLIM datasets. To visualise the impact of using different CLIM datasets to create a projected ensemble turnover estimate, for all six CLIM ensembles we assign turnover values for IBAs to one of five percentile categories (0-20%, 21-40%, 41-60%, 61-80%, 81-100%). We then summarise turnover differences between projections based on the CRU baseline dataset and those based on each of the other CLIM datasets, in terms of shifts between turnover categories. This is important because climate change impacts are often presented in such a categorical or absolute fashion (i.e. without a measure of uncertainty), but shifts between categories due to underlying uncertainty could drastically alter perceptions of climate change vulnerability.

Results

Variability among historic gridded baseline climate datasets

The observed annual cycle for the period c. 1979-2009 varied considerably between observational datasets for P_{total} (Fig. 1); however, similar variations were not found for T_{mean}. For precipitation, all observational datasets in all regions showed agreement on the timing of precipitation peaks, but the P_{total} varied considerably between datasets. This difference was most pronounced for the TRMM and UDEL datasets in the West Sahel, West Tropical and Southern Africa regions, although P_{total} for the CRU, WorldClim and WFDEI datasets were very similar. The spatial distribution of disagreement between precipitation datasets (Fig. 2a), shown by the coefficient of variation, revealed that the main locations of disagreement between precipitation datasets are in the Sahel between September and November, and southern and eastern Africa in March to May and September to November.

The spatial distribution of disagreement between temperature observations (Fig. 2b), shown by the T_{mean} range between observations, did not show large differences between seasons. In this case, the locations of large disagreement tended to be confined to small areas in the Namib Desert, semi-arid savannahs, or East African montane environments. This highlights the potential for considerable variability between baseline climatologies.

The importance of baseline climatology in climate change impacts assessments

Uncertainty in species turnover attributable to choice of baseline climate data (CLIM) was high for both time periods (Fig. 3; median 2040-2069 = 15.9%; 2070-2099 = 16.5%), and was comparable in magnitude and importance to GCM choice (19.1%; 22.2%). The largest source of uncertainty in species turnover across the region was attributable to SDM choice (37.5%; 31.4%), the importance of which decreased by 2070-2099, but remained dominant. In both time periods all three main sources of uncertainty (SDM, CLIM, GCM) affected species turnover estimates. Uncertainty in projected turnover attributed to variability associated with using different data subsets (blocks) was consistently low.

Across the region, the dominant source of uncertainty in species turnover was highly spatially variable, and in many areas multiple sources were simultaneously important (Fig. 3). By the end-of-century, uncertainty associated with CLIM was highest in southern Africa, and in parts of the western Sahel (e.g. Senegal and Gambia). GCM uncertainty dominated across parts of Eastern Africa, and became more important in parts of the Sahel and montane areas by the end-of-century. Uncertainty in turnover associated with SDM choice was spatially distributed across much of the region in both time periods, and showed several areas where this source of uncertainty was overwhelmingly dominant.

Uncertainty in the change in climate suitability (S_k) for individual species attributable to the use of different CLIM datasets was, on average, almost twice as important as variation due to GCM choice (Fig. 4; median 2040-2069=12.1% vs. 6.5%; 2070-2099=10.4% vs. 4.6%), although there was considerable variation across species. The uncertainty attributed to SDM methodology was almost double that attributable to CLIM, and four times that attributable to GCM, in both time periods (median 2040-2069=22.4%; 2070-2099=24%), but CLIM remained a much more important source of uncertainty than GCM, or that associated with using different data subsets (block). The species for which CLIM is a dominant source of uncertainty (Fig. 5) occur principally in areas of where CLIM variability was indicated to be high (e.g. Atlantic coastal regions in the sub-tropical zone) and also areas that are remote and have few weather stations (e.g. Sahel; see Fig. 2).

The importance of baseline climatology in a conservation context

The choice of CLIM dataset can impact upon projected species turnover across areas of conservation value, substantially altering projected climate change impacts (Fig. 6). For example, changing the source of precipitation data from ground observations to satellite derived products (e.g. CRU [Fig. 6a] vs. CRU.TRMM [Fig. 6b]), but using the same temperature data, increases the severity of projected turnover across most of the continent. Conversely, species turnover derived from models built on WFDEI.CRU baseline data projected similar (or lower) species turnover than those made using CRU climate data. Comparing models derived from CRU versus UDEL data, less severe turnover was projected for some montane IBAs (e.g. Ethiopian Highlands) using UDEL data, but higher turnover was projected elsewhere. Thus, the choice of baseline climate data had a notable impact on projections of climate change impact for sites of conservation relevance.

Discussion

Here we have demonstrated that the choice of historic baseline climate data can have substantial and important impacts on projected responses of species and communities to future climate change. This is an almost universally overlooked source of uncertainty, but could severely affect projected responses of species to climate change, with significant consequences for conservation prioritisation and management. We found that the choice of baseline climate data affects the overall uncertainty in climate change impacts (measured as species turnover and change in species-specific climate suitability) to a degree comparable with the choice of GCM data. The literature on projecting species' responses to climate change is dominated by calls to consider uncertainty arising from the choices of GCMs, SDMs and climate predictors (Elith & Graham, 2009; Synes & Osborne, 2011; Garcia *et al.*, 2012), as well as spatial autocorrelation (Dormann *et al.*, 2008; Bagchi *et al.*, 2013). This has resulted in the widespread use of ensemble models to average across, or more recently, to quantify uncertainty (Bagchi *et al.*, 2013; Baker *et al.*, 2015). Yet here we provide evidence that the choice of baseline climate data (CLIM) is at least as important as previously identified sources of uncertainty. Importantly, the three dominant sources of uncertainty tested here show idiosyncratic spatial patterning in their importance. For example, uncertainty associated with CLIM was consistently highest in Southern Africa and parts of Western Africa.

In agreement with previous studies (e.g. Buisson *et al.*, 2010; Garcia *et al.*, 2012; Bagchi *et al.*, 2013), choices in SDM methodology and GCM data contribute substantially to the uncertainty in projected species turnover, dominating in many regions. Uncertainty due to differences in the modelled species-climate response using different SDM methodologies is a well-established source of uncertainty in SDM analyses (Elith & Graham, 2009), and is one of the primary reasons for using ensembles of models (Araújo & New, 2007). Without truly independent data for evaluating the predictive performance of these different modelling algorithms it is difficult to select a single best approach, and this makes it highly important that the uncertainty associated with these methodological choices is explored and quantified (Baker *et al.*, 2015). It should also be noted that overall uncertainty in projected impacts based on correlative models is likely to be overly narrow. In a meta-analysis of projected extinction risk, Urban (2015) found that mechanistic and correlative models projected the lowest extinction risk, while species-area relationship models and expert opinion had substantially higher extinction risks.

Figure 1 shows important differences between the observational datasets, especially with regard to precipitation. While the month of seasonal minima and maxima are generally in agreement across all datasets, there are considerable discrepancies between monthly precipitation totals – differences that are also supported by Nikulin *et al.* (2012). These differences are most evident between the TRMM, UDEL and other synoptically derived datasets (CRU, WFDEI.CRU, WFDEI.GPCC and WorldClim), and can be mostly explained by variations in sources of synoptic observations and methodological differences. To some extent these differences should not come as a surprise, as CRU, GPCC and WorldClim datasets have provenance from synoptic weather reports by National Meteorological and/or Hydrological Services (NMHSs) to the World Meteorological Organisation (WMO) Global Telecommunication System (GTS). While the UDEL dataset is also derived from synoptic observations, it has provenance from different databases (National Oceanic and Atmospheric Administration's (NOAA) Global Historical Climatology Network (GHCN) version 2, and the National Climate Data Center's (NCDC) Global Surface Summary of the Day (GSOD), as well as other national level data). The quality control, processing steps, interpolation methods and evaluation methods are all potentially additional factors that could explain the differences found between UDEL and the other synoptically derived datasets.

Variation in turnover projections associated with GCM uncertainty was important in some regions, particularly in the East Africa. However, it is surprising that the relative importance of GCM uncertainty was not higher, given the considerable variability in climate anomalies among these GCMs (Buontempo *et al.*, 2014), and the range of uncertainty associated with GCMs used in other assessments (e.g. Garcia *et al.*, 2012). The GCM ensemble used in this study was composed of a subset of models that were able to simulate well observed climate phenomena across the region, but was also selected to represent the range of responses to climate forcing, as found in a larger multi-model ensemble. Thus, despite the considerable uncertainty across this ensemble, the uncertainty is likely to be narrower than ensembles used in many studies, due to the omission here of regionally implausible GCMs (McSweeney *et al.*, 2014). Few studies justify the selection of GCMs and provide an assessment of their ability to capture the historic climatology of the focal region (Baker *et al.* 2015). More careful consideration of the GCMs selected might reduce unwarranted uncertainty (McSweeney *et al.*, 2012, 2014).

Here we demonstrate spatial heterogeneity in the importance of potential sources of uncertainty, and that no one source consistently dominates. This has important consequences for regionally focused impact studies, where regional characteristics associated with baseline climate data availability could alter the importance of different sources of uncertainty. Differences in baseline climatologies could arise from differences in the selection of climate data (e.g. difference subsets of stations, use of satellite data, different interpolation algorithms) and the spatial variability of climatic conditions (e.g. high variability over mountainous or coastal areas). The importance of this uncertainty for individual species might be dependent on the characteristics of their range (see Fig. 2 and Fig. 5). Species with ranges that encompass high orographic variation, and span areas with low densities of climate observation data, are likely to be particularly affected by baseline uncertainty (e.g. Hofstra *et al.*, 2010). Such regions in Africa might include topographically diverse regions such as the Albertine Rift Valley, and montane ecosystems such as the Ethiopian and Cameroon Highlands. Additionally, sparsely populated regions, such as the Saharan and Sahelian biomes, have low densities of weather observation records and are likely to be particularly affected by climate baseline uncertainty (e.g. Sylla *et al.*, 2013).

Several other sources of uncertainty that are not explored in this study are likely to be important for projecting species' responses to climate change. Most notably, the choice of climate predictor variables has been shown previously to have a large effect on the projected distribution (Synes & Osborne, 2011; Braunisch *et al.*, 2013). The relative importance of the choice of predictor variables is likely to be high, and in future should be assessed in the context of the wider uncertainty, as has been done here for baseline climate uncertainty. However, it is important to make sure that the range of uncertainty is realistic by including only biologically plausible combinations of climate predictors (Synes & Osborne, 2011).

This study has focused on a region that, overall, has a low density of weather observations (Sylla *et al.*, 2013), especially when compared to parts of, for example, Europe and North America (for example, Hijmans *et al.* 2005). However, weather stations across the globe are patchily distributed, and typically reach the highest densities in areas of importance for human populations. Thus, in more remote locations, many of which are likely to be of higher importance for biodiversity, weather observations densities are likely to be low. Even where high densities of weather stations occur, careful consideration should be given to the variation in local climate. Factors that are likely to reduce the correlation in observations between neighbouring weather stations, such as topographically complex terrain or coastal features, might suggest that baseline uncertainty should be considered. The conclusions of this study are likely to apply broadly to most ecological studies involving climate data, but will be most problematic in areas where the density of observation data is low compared to the spatial variability of the local climate.

Correlative species distribution models are frequently used to assess the potential impacts of climate change in networks of protected areas or sites of biodiversity importance (Hole *et al.*, 2009; Araújo *et al.*, 2011; Bagchi *et al.*, 2013; Baker *et al.*, 2015). Adaptation plans are informed by such projections and, consequently, it is important to understand the sensitivity of projected impacts to choices made during the modelling process. As we have shown, projected impacts across IBAs can be altered considerably by simply choosing different baseline climatology to correlate with a species' distribution. Rather than simply mapping impacts based on the ensemble average (which is common practice: e.g. Hole *et al.*, 2009; Araújo *et al.*, 2011; Bagchi *et al.*, 2013), we advocate representing the degree of uncertainty in spatial maps to better communicate the degree of confidence in projected impacts (Baker *et al.*, 2015).

Uncertainty in baseline climate data has relevance beyond species distribution modelling. It will be important in any situations where uncertainty in the historic record has the potential to undermine inferences, such as studies analysing ecological responses to inter-annual climate variability (VanDerWal *et al.*, 2013), phenological studies (Phillimore *et al.*, 2012) and climate impact indicators (Gregory *et al.*, 2009). There are several ways to incorporate uncertainty in baseline climate into models of species-climate responses. Uncertainty can be explored, as here, by using multiple historic baselines, or alternatively, exploring the impact of uncertainty within a historic climate dataset using stochastic simulations that assume each variable has an associated random error (Folland *et al.*, 2001; Brohan *et al.*, 2006). Estimates of these error distributions are often generated along with the estimated climatology by, for example, leaving single observations out of the interpolation and assessing the difference between observed and predicted values (e.g. Hijmans *et al.* 2005). Some climate products consider a broad range of climate uncertainty, including measurement errors, homogenisation uncertainty and sampling errors (e.g. Brohan *et* *al.* 2006). Recent advances in modelling allow for the explicit inclusion of uncertainty associated with environmental predictor variables, and these approaches could be used to incorporate uncertainty in the climate data into modelled species-climate responses (Stoklosa *et al.*, 2015). Our findings should encourage greater consideration of uncertainty associated with historic baselines when assessing potential responses of species to climate change. Indeed, considerations of uncertainty in historic baseline data should become routine for all research incorporating such data (Parra & Monahan, 2008).

To conclude, we have shown that projected responses of species to climate change can be highly affected by uncertainty in the historic climate baseline data used to model species-climate relationship in SDM analyses and to downscale GCM data from coarse resolutions to ecologically relevant spatial scales. From our results it is evident that ecological studies should begin to routinely account for this source of uncertainty. Within a conservation context, this will facilitate better planning for targeting monitoring and adaptation interventions, and help strengthen conservation efforts in the face of a rapidly changing climate.

Acknowledgements

This project followed on from the PARCC West Africa projected funded by the Global Environment Facility (GEF). We are grateful to the comments of two anonymous reviewers, which helped greatly to improve the clarity of the manuscript.

References

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Araújo MB, Alagador D, Cabeza M, Nogués-Bravo D, Thuiller W (2011) Climate change threatens European conservation areas. *Ecology Letters*, **14**, 484–492.

Bagchi R, Crosby M, Huntley B et al. (2013) Evaluating the effectiveness of conservation site networks under climate change: accounting for uncertainty. *Global Change Biology*, **19**, 1236–1248.

Baker DJ, Hartley AJ, Burgess ND et al. (2015) Assessing climate change impacts for vertebrate fauna across the West African protected area network using regionally appropriate climate projections. *Diversity and Distributions*, **21**, 991–1003.

Barbet-Massin M, Jetz W (2014) A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, **20**, 1285–1295.

BirdLife International (2014) *Important Bird and Biodiversity Areas: A global network for conserving nature and benefiting people.* Cambridge, UK: BirdLife International.

BirdLife International (2015) *IUCN Red List for birds. Downloaded from http://www.birdlife.org.*

BirdLife International, NatureServe (2013) *Bird Species Distribution Maps of the World. Version 3.0.* BirdLife International and NatureServe, Cambridge, UK and Arlington, USA.

Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**, 971–983.

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research: Atmospheres*, **111**, D12106.

Buisson L, Thuiller W, Casajus N, Lek S, Grenouillet G (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.

Buontempo C, Mathison C, Jones R, Williams K, Wang C, McSweeney C (2014) An ensemble climate projection for Africa. *Climate Dynamics*, **44**, 2097–2118.

Busby J (1991) BIOCLIM-a bioclimate analysis and prediction system. *Plant Protection Quarterly (Australia)*, **6**, 8–9.

Cahill AE, Aiello-Lammens ME, Fisher-Reid MC et al. (2013) How does climate change cause extinction? *Proceedings of the Royal Society B: Biological Sciences*, **280**, 21231890.

Carvalho SB, Brito JC, Crespo EG, Watts ME, Possingham HP (2011) Conservation planning under climate change: Toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time. *Biological Conservation*, **144**, 2020–2030.

Cheaib A, Badeau V, Boe J et al. (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology letters*, **15**, 533–544.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, **333**, 1024–1026.

Dormann CF, Purschke O, Márquez JRG, Lautenbach S, Schröder B (2008) Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology*, **89**, 3371–3386.

Dullinger S, Gattringer A, Thuiller W et al. (2012) Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Clim. Change*, **2**, 619–622.

Efthymiadis D, Jones PD, Briffa KR et al. (2006) Construction of a 10-min-gridded precipitation data set for the Greater Alpine Region for 1800–2003. *Journal of Geophysical Research: Atmospheres*, **111**, D01105.

Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677.

Foden WB, Butchart SHM, Stuart SN et al. (2013) Identifying the World's Most Climate Change Vulnerable Species: A Systematic Trait-Based Assessment of all Birds, Amphibians and Corals. *PLoS ONE*, **8**, e65427.

Folland CK, Rayner NA, Brown SJ et al. (2001) Global temperature change and its uncertainties since 1861. *Geophysical Research Letters*, **28**, 2621–2624.

Garcia RA, Burgess ND, Cabeza M, Rahbek C, Araújo MB (2012) Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, **18**, 1253–1269.

Giorgi F, Jones C, Asrar GR (2009) Addressing climate information needs at the regional level : the CORDEX framework. **58**, 175–183.

Gordon C, Cooper C, Senior CA et al. (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16**, 147–168.

Gregory RD, Willis SG, Jiguet F et al. (2009) An Indicator of the Impact of Climatic Change on European Bird Populations. *PLoS ONE*, **4**, e4678.

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, **34**, 623–642.

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **113**, 2156–2202.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research: Atmospheres*, **114**, D21101.

Hofstra N, New M, McSweeney C (2010) The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, **35**, 841–858.

Hole DG, Willis SG, Pain DJ et al. (2009) Projected impacts of climate change on a continent-wide protected area network. *Ecology Letters*, **12**, 420–431.

Huffman GJ, Bolvin DT, Nelkin EJ et al. (2007) The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, **8**, 38–55.

Isotta FA, Frei C, Weilguni V et al. (2014) The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, **34**, 1657–1675.

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Kearney M, Phillips BL, Tracy CR, Christian KA, Betts G, Porter WP (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.

Kujala H, Moilanen A, Araújo MB, Cabeza M (2013) Conservation Planning with

Uncertain Climate Change Projections. *PLoS ONE*, **8**, e53315.

Legates DR, Willmott CJ (1990) Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, **10**, 111–127.

Mariotti L, Diallo I, Coppola E, Giorgi F (2014) Seasonal and intraseasonal changes of African monsoon climates in 21st century CORDEX projections. *Climatic Change*, **125**, 53–65.

McSweeney CF, Jones RG, Booth BBB (2012) Selecting Ensemble Members to Provide Regional Climate Change Information. *Journal of Climate*, **25**, 7100–7121.

McSweeney CF, Jones RG, Lee RW, Rowell DP (2014) Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, **44**, 3237–3260.

Nakicenovic N, Alcamo J, Davis G et al. (2000) *Emissions scenarios*. Cambridge University Press, Cambridge.

Nikulin G, Jones C, Giorgi F et al. (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *Journal of Climate*, **25**, 6057–6078.

Ockendon N, Baker DJ, Carr JA et al. (2014) Mechanisms underpinning climatic impacts on natural populations: altered species interactions are more important than direct effects. *Global Change Biology*, **20**, 2221–2229.

Pacifici M, Foden WB, Visconti P et al. (2015) Assessing species vulnerability to climate change. *Nature Climate Change*, **5**, 215–224.

Parra JL, Monahan WB (2008) Variability in 20th century climate change reconstructions and its consequences for predicting geographic responses of California mammals. *Global Change Biology*, **14**, 2215–2231.

Pearce-Higgins JW, Ockendon N, Baker DJ et al. (2015) Geographical variation in species' population responses to changes in temperature and precipitation. *Proc. R. Soc. B*, **282**, 20151561.

Phillimore AB, Stålhandske S, Smithers Richard J., Bernard R (2012) Dissecting the Contributions of Plasticity and Local Adaptation to the Phenology of a Butterfly and Its Host Plants. *The American Naturalist*, **180**, 655–670.

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, **16**, 123–146.

Roubicek AJ, VanDerWal J, Beaumont LJ, Pitman AJ, Wilson P, Hughes L (2010) Does the choice of climate baseline matter in ecological niche modelling?

Ecological Modelling, **221**, 2280–2286.

Rowlands DJ, Frame DJ, Ackerley D et al. (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geosci*, **5**, 256–260.

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2014) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, **115**, 15–40.

Stoklosa J, Daly C, Foster SD, Ashcroft MB, Warton DI (2015) A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution*, **6**, 412–423.

Stott PA, Kettleborough JA (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 723–726.

Sylla MB, Giorgi F, Coppola E, Mariotti L (2013) Uncertainties in daily rainfall over Africa: assessment of gridded observation products and evaluation of a regional climate model simulation. *International Journal of Climatology*, **33**, 1805–1817.

Synes NW, Osborne PE (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography*, **20**, 904–914.

Tabor K, Williams JW (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications*, **20**, 554–565.

Tapiador FJ, Turk FJ, Petersen W et al. (2012) Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, **104–105**, 70–97.

Urban MC (2015) Accelerating extinction risk from climate change. *Science* , **348** , 571–573.

VanDerWal J, Murphy HT, Kutt AS, Perkins GC, Bateman BL, Perry JJ, Reside AE (2013) Focus on poleward shifts in species' distribution underestimates the fingerprint of climate change. *Nature Clim. Change*, **3**, 239–243.

Watling JI, Fletcher RJ, Speroterra C et al. (2014) Assessing Effects of Variation in Global Climate Data Sets on Spatial Predictions from Climate Envelope Models. *Journal of Fish and Wildlife Management*, **5**, 14–25.

Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, **50**, 7505–7514.

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, **21**, 530–548.

Willis SG, Foden W, Baker DJ et al. (2015) Integrating climate change vulnerability assessments from species distribution models and trait-based approaches. *Biological Conservation*, **190**, 167–178.

Willmott CJ, Robeson SM (1995) Climatologically aided interpolation (CAI) of terrestrial air temperature. *International Journal of Climatology*, **15**, 221–229.

Table 1. Historic gridded climate datasets (CLIM) used in SDM fitting and downscaling GCM simulations of future climates using the change factor method (CFA). The TRMM precipitation dataset was combined with the temperature variables from CRU TS3.1 to create CRU.TRMM used in the analysis.

Dataset	**Variables**	**Time period**	**Spatial resolution**	**Description**	**References**
CRU TS3.1	T_{mean} P_{total}	1900-2012	0.5^0	Time series of spatially interpolated monthly observations from meteorological stations	(Harris *et al.*, 2014)
TRMM	P_{total}	1998-present	0.25^0	Satellite observations, calibrated using rain gauge data	(Huffman *et al.*, 2007)
WFDEI.CRU	T_{mean} P_{total}	1979-2012	0.5^0	ERA-Interim reanalysis data elevation and bias corrected using CRU TS3.1	(Weedon *et al.*, 2014)
WFDEI.GPCC	T_{mean} P_{total}	1979-2012	0.5^0	ERA-Interim reanalysis data elevation and bias corrected using GPCC	(Schneider *et al.*, 2014; Weedon *et al.*, 2014)
WorldClim	T_{mean} P_{total}	1950-2000	0.167^0	Spatially interpolated monthly mean observations from meteorological stations for 1950-2000	(Hijmans *et al.*, 2005)
UDEL	T_{mean} P_{total}	1900-2012	0.5^0	Time series of spatially interpolated monthly observations from meteorological stations	(Legates & Willmott, 1990; Willmott & Robeson, 1995)

Figure 1. Summary of mean monthly total precipitation (P_{total}, mm/month) and monthly mean temperature (T_{mean}, ^{o}C) for the period c. 1979-2009 for six regions of sub-Saharan Africa (demarcated by solid lines) derived from six baseline climate products (see legend and Table 1). The mean climate for each focal region is calculated from the mean monthly 30-year (although see methods for details on temporal periods) average of the variables in each of the cells across the region. The x-axis tick marks represent the twelve months of the year, ordered from January to December.

Figure 2. Spatial patterns of the uncertainty in climate observation datasets for precipitation (a) and temperature (b), shown for each season (DJF: December, January, February; MAM: March, April, May; JJA: June, July, August; SON: September, October, November). Precipitation uncertainty is shown by the coefficient of variation, calculated on the observed climate datasets (n=6). Areas of total seasonal precipitation less than 30mm were removed from the analysis in order to exclude areas where the standard deviation was much greater than the mean (see Methods for full details). Temperature uncertainty is shown by the range of values in degrees Celsius, across the climate observations (n=4).

Figure 3. The percentage of the total variability (deviance) in species turnover explained by each uncertainty factor - a measure of the relative importance of each factor to the overall uncertainty in projected climate change impacts across the region. Changes are measured between the baseline period of c. 1979-2009 and each of two focal time periods, 2040-2069 (top) and 2070-2099 (bottom). SDM = species distribution model; CLIM = climate baseline data; GCM = general circulation model;

block = uncertainty due to using different data subsets. Note: block was included in the analyses but its uncertainty was not mapped due to its minimal effect.

Figure 4. The percentage of the total variability (sums of squares) in species-specific change in climate suitability, measured between the baseline period of c. 1979-2009 and each of two focal time periods, 2040-2069 (top) and 2070-2099 (bottom), explained by each uncertainty factor (the latter as in Figure 3). The boxplots summarise the importance of each sources of uncertainty across all species included in the analysis *(n* = 895).

Figure 5. The pattern of species richness for species where climate baseline data uncertainty (CLIM) was a dominant source of uncertainty (upper 5th percentile of CLIM affected species, *n* = 48). The patterns reflect known areas of climatic complexity (see Fig. 2).

Figure 6. The ensemble mean projected species turnover (by 2070-2099) for the region's Important Bird and Biodiversity Areas (IBAs) for: (a) projections derived from models built using CRU climate baseline data; and (b-f) the number of turnover categories (percentage species turnover: 0-20%, 21-40%, 41-60%, 61-80%, 81-100%) by which the projections are shifted when projections are derived from models built on one of the other five climate baseline (CLIM) datasets. Thus, IBAs in plots b-f that are coloured green decrease one category, and are therefore projected to have lower species turnover in comparision to CRU-based projections. Similarly, IBAs in plot b-f coloured red or dark red increase one or two categories, respectively, and are

therefore projected to have higher species turnover in comparision to CRU-based projections.

Figure 1

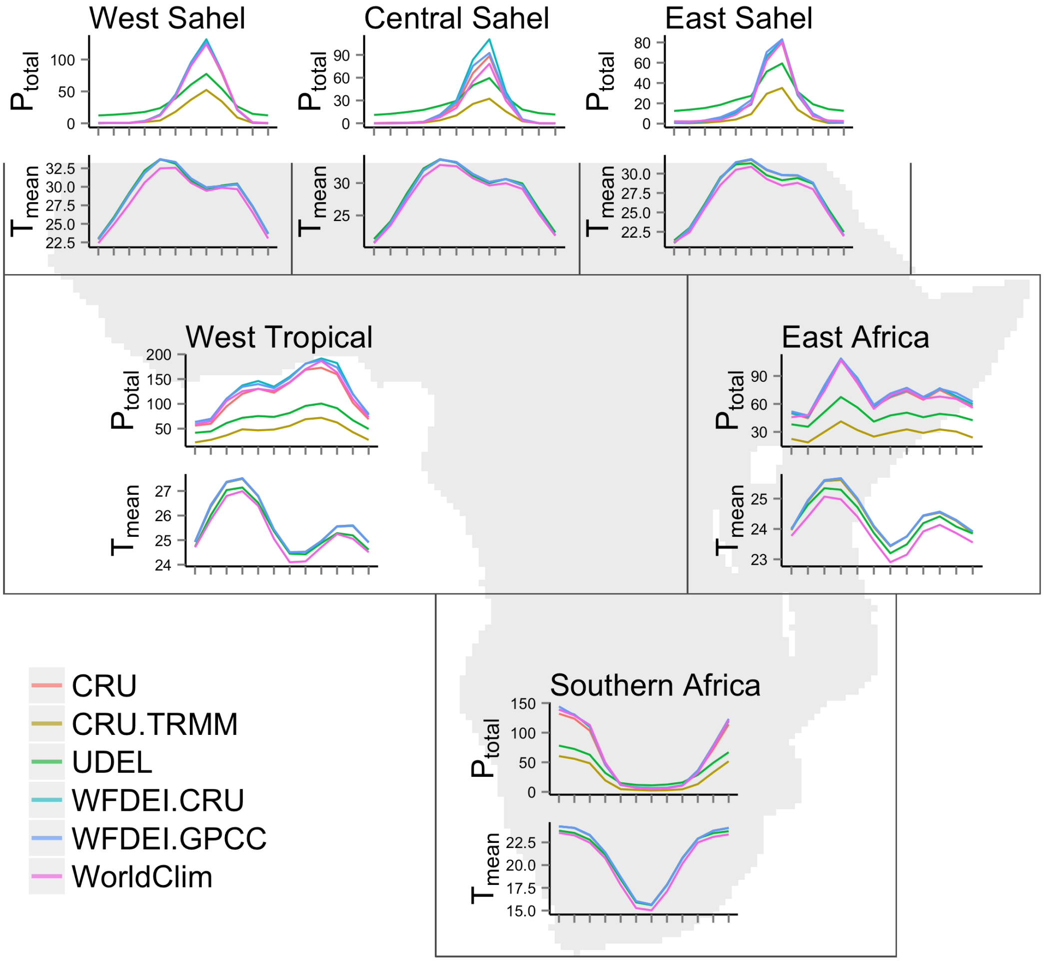

West Sahel
Central Sahel
East Sahel
West Tropical
East Africa
Southern Africa
P_{total}
T_{mean}
100
50
0
90
60
30
80
40
20
32.5
30.0
27.5
25.0
22.5
30
25
200
150
27
26
24
23
20.0
17.5
15.0
CRU
CRU.TRMM
UDEL
WFDEI.CRU
WFDEI.GPCC
WorldClim

Figure 2

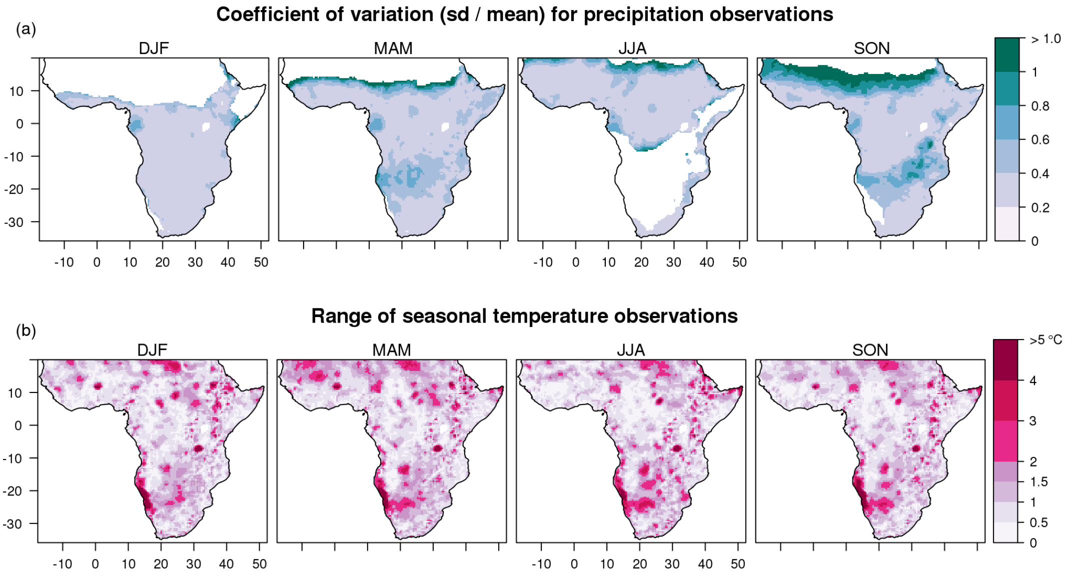
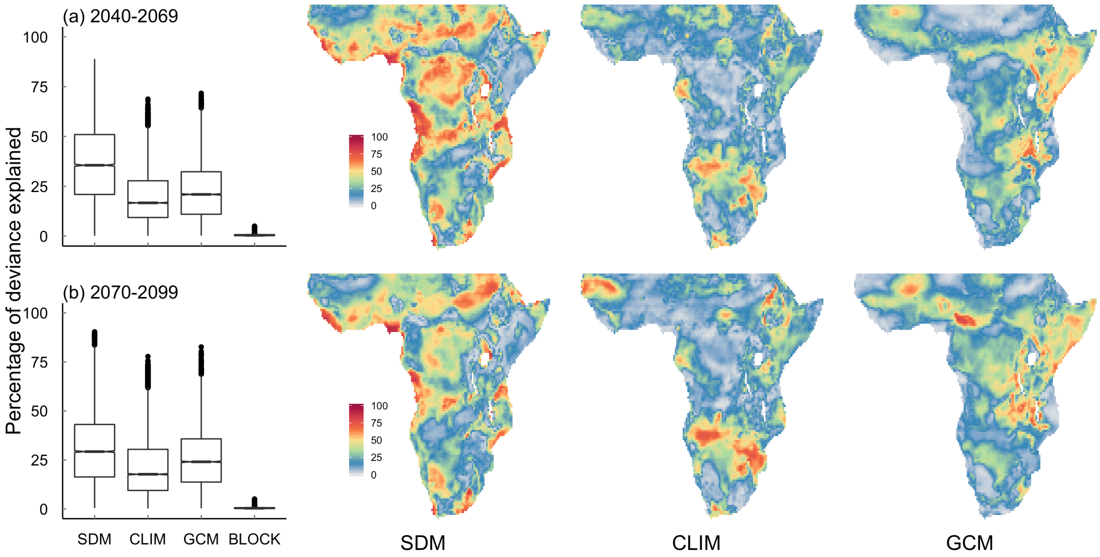
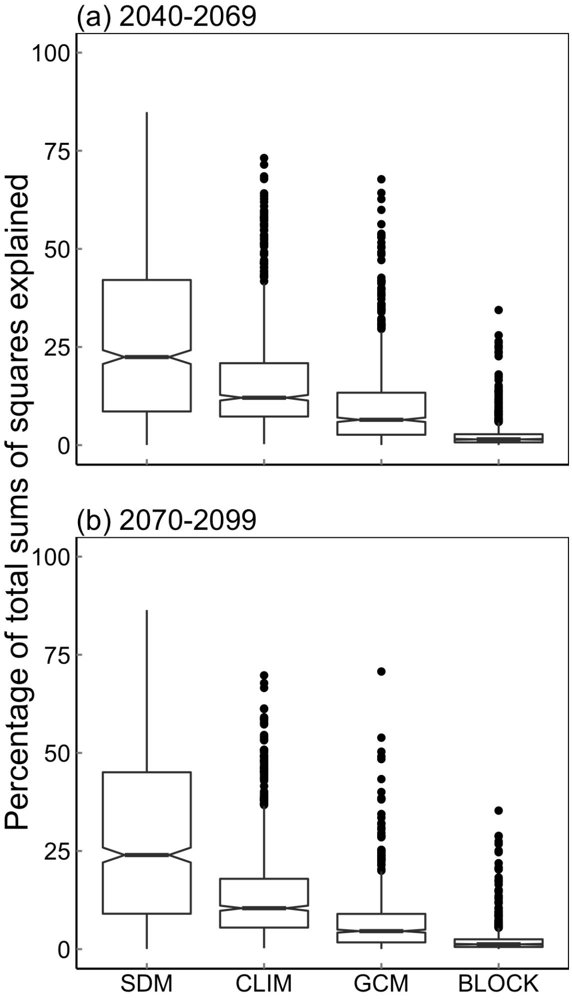

(a)
Coefficient of variation (sd / mean) for precipitation observations
DJF
MAM
JJA
SON
10
0
-10
-20
-30
-10
0
10
20
30
40
50
> 1.0
1
0.8
0.6
0.4
0.2
0
(b)
Range of seasonal temperature observations
DJF
MAM
JJA
SON
>5 °C
4
3
2
1.5
1
0.5
0

Figure 3

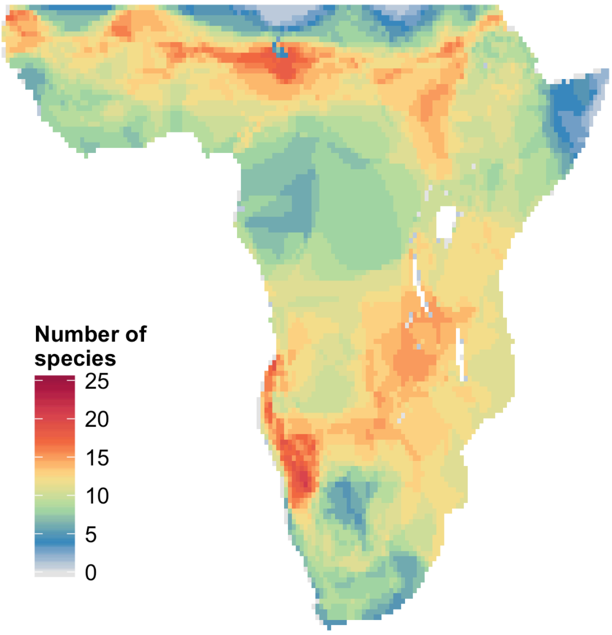

(a) 2040-2069
(b) 2070-2099
Percentage of deviance explained
100
75
50
25
0
SDM
CLIM
GCM
BLOCK
100
75
50
25
0
SDM
CLIM
GCM

872 Figure 4



873

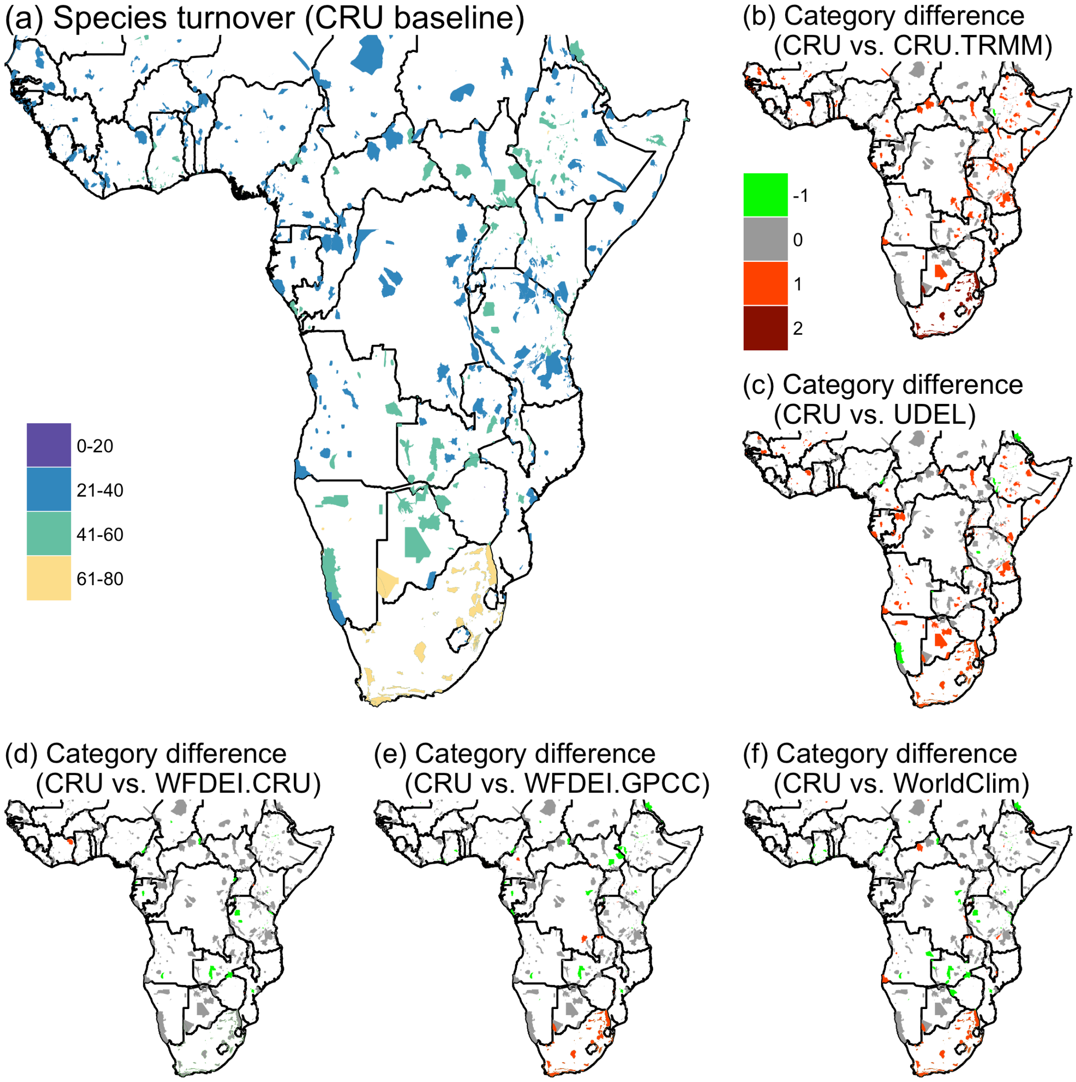
874 Figure 5



875

876

Figure 6


(a) Species turnover (CRU baseline)
0-20
21-40
41-60
61-80
(b) Category difference (CRU vs. CRU.TRMM)
-1
0
1
2
(c) Category difference (CRU vs. UDEL)
(d) Category difference (CRU vs. WFDEI.CRU)
(e) Category difference (CRU vs. WFDEI.GPCC)
(f) Category difference (CRU vs. WorldClim)