Bias-corrected data sets of climate model outputs at uniform space–time resolution for land surface modelling over Amazonia

Article *in* International Journal of Climatology · February 2017

DOI: 10.1002/joc.4728

8 authors, including:

Ke Zhang
Hohai University; University of Oklahoma;
85 PUBLICATIONS **1,998** CITATIONS
SEE PROFILE

Ardeshir M. Ebtehaj
University of Minnesota Twin Cities
29 PUBLICATIONS **159** CITATIONS
SEE PROFILE

R. L. Bras
Georgia Institute of Technology
391 PUBLICATIONS **14,860** CITATIONS
SEE PROFILE

Jingfeng Wang
Georgia Institute of Technology
59 PUBLICATIONS **748** CITATIONS
SEE PROFILE

INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4728

RMetS
Royal Meteorological Society

Bias-corrected data sets of climate model outputs at uniform space–time resolution for land surface modelling over Amazonia

Sanaz Moghim,[a]* Shawna L. McKnight,[a] Ke Zhang,[b] Ardeshir M. Ebtehaj,[c] Ryan G. Knox,[d] Rafael L. Bras,[a] Paul R. Moorcroft[e] and Jingfeng Wang[a]

[a] *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, USA*
[b] *The Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK, USA*
[c] *Utah Water Research Laboratory, Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA*
[d] *Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[e] *Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA*

ABSTRACT: Developing high-quality long-term data sets at uniform space–time resolution is essential for improved climate studies. This article processes the outputs from two global and regional climate models, the Community Climate System Model (CCSM3) and the Regional Climate Model driven by the Hadley Centre Coupled Model (RegCM3). The results are bias-corrected time series of atmospheric variables corresponding to Intergovernmental Panel on Climate Change (IPCC's) historical (20C3M) and future (A2) climate scenarios over the Amazon Basin. We use a series of simple but effective interpolation approaches to produce hourly climate data sets at 1° by 1° grid cells. A quantile-based mapping approach is used to reduce the biases of temperature and precipitation in CCSM3 and RegCM3. Adjustments are also made on specific humidity and downwelling longwave radiation to avoid inconsistency between those variables and bias-corrected temperature values. We also interpolated an already bias-corrected Parallel Climate Model data set (PCM1) from 3-hourly to the hourly resolution. The final climate data sets can be used as forcing of ecosystem and hydrologic models to study climate changes and impact assessments over the Amazon Basin.

KEY WORDS General Circulation Models; bias correction; climate data sets; climate simulations; climate change; Amazon Basin

Received 22 May 2014; Revised 2 February 2016; Accepted 7 March 2016

1. Introduction

The Earth's climate has shaped the history of biological and cultural evolution and the geographical distribution of humans. Almost all of the human-made and natural systems related to water resources, agriculture, forestry, fishery, industry, energy, and financial services are impacted by natural and human-induced climate change (Harding *et al.*, 2011). As observations are limited in space and time, numerical simulations and predictions are necessary to understand and project the consequences of any potential changes in climate.

General Circulation Models (GCMs) are commonly used to simulate Earth's climate. However, current computation and storage restrictions result in a coarse space–time resolution of the GCM outputs, often limiting their ability to capture fine-scale regional climate variability that is important for natural hazard assessment and mitigation plans (Mearns *et al.*, 2009).

This paper uses linear, stochastic, and deterministic statistical downscaling methods to produce a uniform space–time resolution for all model outputs. Statistical downscaling approaches are based on empirical relationships between regional and large-scale variables. Although this family of downscaling approaches cannot completely address the underlying physics and variability of climate predictors, they have considerable computational advantages. Their performance may vary with models, seasons, and geographical locations (Murphy, 1999). For a complete explanation and a review of statistical downscaling techniques, the reader is referred to Fowler *et al.* (2007), Wilby and Wigley (1997), and Xu (1999), among many others.

Climate models are imperfect and their outputs have errors. The ability of the climate models to represent climate variables over large regions at decadal time scales is still uncertain. Johnson *et al.* (2011) evaluated the capability of the GCMs to represent the variability of El Niño-Southern Oscillation (ENSO) phenomenon. They showed that the performance of the Community Climate System Model (CCSM3) and the Hadley Centre Coupled Model (HadCM3) in simulating the spatial precipitation pattern and variability, SST, and surface pressure is better

* Correspondence to: S. Moghim, School of Civil and Environmental Engineering, Georgia Institute of Technology, North Ave., NW, Atlanta, GA 30332, United States. E-mail: smoghim@gatech.edu

at inter-annual scale than at sub-annual and inter-decadal ones. Many studies have reported the varying performance of GCMs (van Oldenborgh *et al.*, 2005; Coelho and Goddard, 2009). For instance, de Szoeke and Xie (2008) illustrated that HadCM3 outperforms CCSM3 and the Parallel Climate Model (PCM1) in simulating seasonality of SST and rainfall over the equator. The poor performance in CCSM3 and PCM1 may be caused by the errors in the simulated zonal and meridional winds, which affects the ENSO simulation (van Oldenborgh *et al.*, 2005; Guilyardi, 2006).

Despite the fact that the performance of the GCMs have been improving over the past years (Reichler and Kim, 2008), many studies have reported that biases still exist in the model outputs (e.g. Dai, 2001a, 2001b, 2006; Solomon *et al.*, 2007, among others), which can severely impact long-term climate hazard assessments and mitigation plans (Lettenmaier and Rind, 1992; Stamm *et al.*, 1994; Conway, 1998; Rheenen *et al.*, 2004; Wood *et al.*, 2004; Tanaka *et al.*, 2006; Maurer *et al.*, 2007; Georgakakos *et al.*, 2012). Among these, for example, an overestimation in the frequency of light precipitation and an underestimation in the intensity of heavy precipitation in GCMs and RCMs are reported by Sun *et al.* (2005) and Murphy (1999). Wood *et al.* (2004) studied the effect of bias correction on a downscaled PCM1 data set. They used statistical downscaling methods including linear interpolation, inverse distance squared weighting technique for spatial disaggregation and a quantile-mapping approach for bias correction. They concluded that the downscaled data without bias correction produces notable biases in hydrologic simulations and bias correction is a necessary step to produce reliable hydro-climatological variables.

Several attempts have been made to correct the biases and downscale the outputs of climate models. One simple bias correction approach, commonly used in climate studies, is called the delta-change method. This method shifts or re-scales the mean of the modelled data based on the mean of the observations (Ines and Hansen, 2006). The delta-change method is able to adjust only the first order moment (mean) without any adjustment to the higher order moments.

Another widely used statistical bias correction method is the quantile-based mapping approach, which adjusts not only the mean, but all moments of the distribution. In this method, a transfer function is derived from the cumulative distribution functions (CDFs) of the modelled variables and observations to adjust the entire distribution of the climate variable of interest (Ines and Hansen, 2006; Piani *et al.*, 2010). Distribution mapping methods have also been compared with other bias correction methods in various studies (Chen *et al.*, 2013; Lafon *et al.*, 2013). The results indicate that although all bias correction methods can improve modelled outputs to some extent, CDF matching methods outperform most of the available techniques by adjusting the higher order error statistics in a distribution sense. While statistical methods often use a stationary assumption for bias correction, which may not always hold true (e.g. Solomon *et al.*, 2007; Milly *et al.*, 2008; Li *et al.*, 2010 among others), cross-validation studies have shown that distribution mapping methods are best suited for correcting non-stationary biases (Teutschbein and Seibert, 2013).

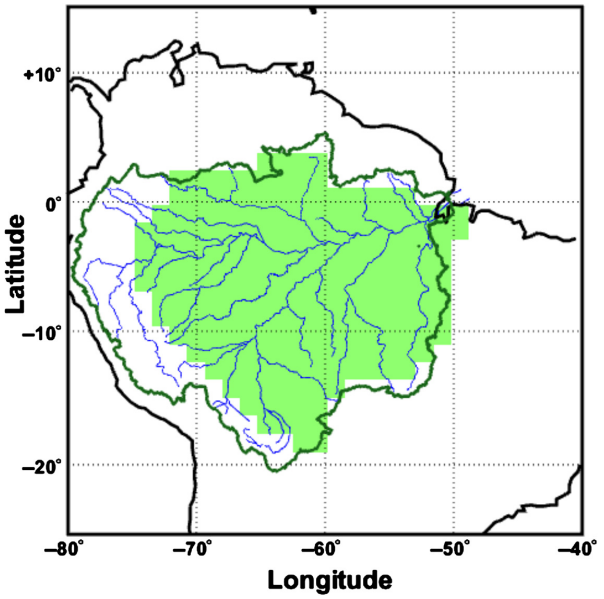

Figure 1. The Amazon Basin is bounded by the green line, and the study domain is shaded by green colour.

This study focuses on the Amazon Basin extending from 74.5° to 49.5°W and from 17.5°S to 3.5°N (Figure 1). The Amazon Basin has been experiencing severe droughts, forest fires, and deforestation events in the past decades that may change the structure of the entire Amazonian ecosystem (e.g. Aragao *et al.*, 2008; Bush *et al.*, 2008). Climate and land cover changes in the Amazon Basin, the largest tropical forest and reservoir of freshwater and carbon in the world, can influence the entire world's ecosystems (Cochrane and Barber, 2009). Most climate models show that rainfall is decreasing and the length of the dry season is increasing over this region (Christensen *et al.*, 2007; Malhi *et al.*, 2009; Costa and Pires, 2010). Different studies have investigated resilience of Amazonia against climate change, fires, droughts, and deforestation (Betts *et al.*, 2004; Saleska *et al.*, 2007; Cox *et al.*, 2013) and called for further research for better understanding and preserving the ecosystem of this critical region of the world.

This article uses historical and predicted time series from two models including CCSM3 and RegCM3, a RCM driven by HadCM3 to produce high-resolution unbiased data sets required for driving terrestrial biosphere models of land surface processes to study Amazonia's response to projected climate and land-use changes. This article also interpolates the 3-hourly, bias-corrected PCM1 data set (Sheffield *et al.*, 2006; Li *et al.*, 2010) to hourly resolution and compares the results with the other bias-corrected products. Simple but effective linear and stochastic interpolation is used to increase the resolution of model outputs to desired horizontal grid dimensions and time steps. A modified version of the CDF method, the equidistant CDF (EDCDF) method, was used for bias correction of

temperature and precipitation fields (Li *et al.*, 2010). The EDCDF method accounts for information from model projections, partly allowing for the non-stationarity of higher order moments of the climate variables (see Section 4 for a detailed description of the EDCDF method). Then the specific humidity and downwelling longwave radiation are adjusted to be physically consistent with the bias-corrected temperature values. Section 2 describes the study domain.

Data sets of the model outputs are described in Section 2. Section 3 describes the interpolation methods used, and the bias correction approach is explained in Section 4. The results are presented in Section 5, and discussion and conclusions are pointed out in Section 6.

2. Data sets

Climate models are key tools to study short- and long-term climate changes under different scenarios. The Special Report on Emission Scenarios (SRES) by the IPCC Working Group III (Nakicenovic *et al.*, 2000) identifies different greenhouse gas emissions scenarios that are commonly used for future forcing of the GCMs. This work uses one emission scenario 'A2', which is based on a 'worst-case emissions scenario' with increases of CO_2 by factors of four to five over 2000–2099 (Li *et al.*, 2010). This scenario results in a high rate of warming based on model ensembles (Meehl *et al.*, 2007) and has been widely used for different climate change assessments including the North American Regional Climate Change Assessment Program, NARCCAP (Mearns *et al.*, 2009).

This work selects two climate models, CCSM3 and RegCM3, which are widely used in numerous climate studies (Shin *et al.*, 2003; Hayhoe *et al.*, 2004; Burke *et al.*, 2006; Meehl *et al.*, 2006; Reboita *et al.*, 2014). Proper parameterization and realistic characterization of the underlying physics have led to several successful results in simulation of long-term climatic changes by these models (Randall *et al.*, 2007). Table 1 summarizes general information of the models used in this study. The results of this effort are hourly bias-corrected data sets of temperature (T), precipitation (P), specific humidity (Q), and downwelling longwave radiation (LW_d) on a 1° by 1° horizontal grid. The resulting data sets also include interpolated values of incoming shortwave radiation (SW_d), horizontal wind velocities (u, v), and surface pressure (PS) on a 1° by 1° horizontal grid and hourly time step.

2.1. Community Climate System Model

The CCSM3 was developed by the University Corporation for Atmospheric Research (UCAR) and is maintained by the National Center for Atmospheric Research (NCAR). CCSM3 is a well-established model and has been successfully used for simulating the impacts of land-use change on climate (Collins *et al.*, 2006). This model includes the Community Land Model (CLM), which has dynamic vegetation growth, death, and succession (Bonan and Levis, 2006). The temporal resolutions of the available data are 6-hourly, daily, monthly, and annually from CCSM3 IPCC AR4 simulations (publicly available at www.cgd.ucar.edu/ccr/strandwg/ccsm_6hr_data.html). Forty-two variables are available at 1–26 vertical pressure levels with spatial resolution of about 1.4°. Among these variables, the 6-hourly instantaneous T, P, Q, LW_d, SW_d, u, v, and PS are used in this study for two time periods: historical period 1940–1999 (20th century), and SRES A2 future projection for 2000–2099.

2.2. Regional climate model driven by HadCM3

One of the coupled Atmosphere–Ocean General Circulation Models (AOGCM) used in the IPCC Third and Fourth Assessment Reports is the UK Met Office Hadley Centre Coupled Model (HadCM3) developed by the Hadley Centre in the United Kingdom. This model can properly simulate the effects of natural and anthropogenic forcings on the historical climate (Stott *et al.*, 2000). HadCM3 is designed to simulate climate without using artificial flux adjustments, which is often required to modify unrealistic climate states (Reichler and Kim, 2008). The global resolution of the model is 2.5° latitude by 3.75° longitude at 19 vertical pressure layers. The regional Eta-Centro de Previsão de Tempo e Estudos Climãticos (CPTEC) Model (RegCM3) nested in HadCM3 is used to provide the data over South America (Chou *et al.*, 2012). The standard (unperturbed) HadCM3 was used as boundary conditions of RegCM3 at every 6 h and updated linearly along the boundaries for each time step. RegCM3 data with 40-km horizontal resolution, 38 vertical layers are available for the following time periods: 1960–1990, 2010–2040, 2040–2070, and 2070–2099. RegCM3 may simulate mesoscale patterns and topographic effects of the local region better than HadCM3 (Chen *et al.*, 2013; Reboita *et al.*, 2014). For more detailed descriptions of the model, the reader is referred to the papers by Chou *et al.* (2012) and Marengo *et al.* (2012).

2.3. Climate Research Unit data set

We use the Climate Research Unit data set (CRU) as a reference data set (publicly available at http://www.cru.uea.ac.uk/data). The data are at monthly scale for the entire period of 1901–2009 at spatial resolution of 0.5°. Note that, in this article, to unify the spatial scales of this data set with the spatial resolution of the studied model outputs, the monthly CRU data with original resolution of 0.5° are averaged onto 1° resolution. The years of the observations are selected based on the available years of the historically modelled data (1940–2009 for CCSM3 and 1960–1990 for RegCM3).

3. Interpolation methodologies

Simple but effective interpolation approaches are used to produce variables at the desired horizontal grid and time step. In particular, 1.4° by 1.4° CCSM3 and 0.4° by 0.4° RegCM3 output variables are interpolated into 1° by 1° resolution using bilinear interpolation. For temporal disaggregation of different models' output variables,

Table 1. Spatial and temporal coverage of the models (CCSM3 and RegCM3).

Model	Variable	Spatial resolution	Temporal coverage
CCSM3	T, LW_d, SW_d, Q, u, v, PS, P	$1.4^\circ \times 1.4^\circ$ (T85)	1940–2099, 6-hourly instantaneous
RegCM3	LW_d, SW_d, T_d, u, v, PS	$0.4^\circ \times 0.4^\circ$	1960–1990, 2010–2099, 6-hourly average
	T	$0.4^\circ \times 0.4^\circ$	1960–1990, 2010–2099, 6-hourly instantaneous
	P	$0.4^\circ \times 0.4^\circ$	1960–1990, 2010–2099, 6-hourly accumulated

two statistical approaches, explained in the following subsections, are used.

3.1. Disaggregation of precipitation

CCSM3 generates and stores precipitation data instantaneously every 6 h, while RegCM3 produces 6-hourly accumulated precipitation values. Therefore, we use two different methods for hourly disaggregation of precipitation obtained from these models. Linear temporal interpolation is used for disaggregation of 6-hourly instantaneous CCSM3 precipitation, while a stochastic approach developed for the disaggregation of accumulated precipitation is applied to the RegCM3 precipitation. This stochastic method relies on random sampling from an estimated exponential distribution that allows the reproduction of a realization of high-temporal resolution rainfall data (hourly) from a given coarse resolution data set (Lammering and Dwyer, 2000; Knox, 2013). Specifically, this approach relates the distribution of disaggregated hourly precipitation P to the accumulated 6-hourly RegCM3 precipitation values P_{MOD} through the following exponential probability distribution function, $f(P)$:

$$f(P) = \left(\frac{\mu}{P_{\mathrm{MOD}}}\right) \exp\left[-\frac{\mu P}{P_{\mathrm{MOD}}}\right] \tag{1}$$

where μ denotes the fraction of wet periods (rainfall fraction). This parameter is typically considered as a climatological parameter that needs to be estimated from statistics of hourly rain gauge time series P_g (Eltahir and Bras, 1993) as follows:

$$\mu = \frac{\mathbb{E}\left[P_g\right]}{\rho} \tag{2}$$

Here, $\mathbb{E}\left[P_g\right]$ is the expected value of rainfall intensity considering both wet and dry periods, while ρ is the rainfall density only over the wet periods, which can be computed as follows:

$$\rho = \frac{\sum_{i=1}^{N} P_g(i)}{\sum_{i=1}^{N} a(i)} \tag{3}$$

where N denotes all wet and dry sampling times and $a(i)$ is an indicator function for rainfall occurrence as follows:

$$a(i) = \begin{cases} 1 & \text{if } P_g(i) > 0 \\ 0 & \text{if } P_g(i) = 0 \end{cases}$$

This disaggregation method produces hourly realizations influenced by the rainfall fraction μ and modelled precipitation (P_{MOD}). To estimate the rainfall fraction parameter μ in Equation (1), we use Santarem KM67 rain gauge station (latitude 2.86°S, longitude 54.96°W, kilometre 67 at Tapajos in Para, Brazil). The station is the only continuous and reliable hourly record in the region. Other stations in the study region have many missing data, which precludes the usage of their data to correctly calculate the rainfall fractions. Missing data are filled by the data from nearby stations (see, Knox, 2013). The statistics of this rain gauge are obtained from the large-scale biosphere–atmosphere experiment in Amazonia during the model inter-comparison project (LBA-MIP) from 2002 to 2005 (de Goncalves *et al.*, 2013).

Sampling from the CDF in Equation (1) over the time intervals when 6-hour RegCM3 precipitation is non-zero produces realizations of hourly precipitation (P_s^h). To conserve the original monthly precipitation (P^m), the hourly sampled precipitation (P_s^h) is scaled uniformly at the end of each month as follows:

$$P^h = P_s^h \times \frac{P^m}{P_s^m} \tag{4}$$

where P^h is the scaled hourly precipitation and P_s^m is the monthly accumulation of the sampled precipitation.

It needs to be noted that, although not ideal, using only one rain gauge for obtaining μ is the best choices that can be made given the lack of a dense and reliable long-term ground-based observation network in the study area. Clark and Buarque (2013) discuss several precipitation stations, however, the gauges described in that paper only have daily records. The method used here requires both continuous and hourly rainfall records. Furthermore, we find that the sensitivity to the parameter μ is small. Knox (2013) reported the μ values for 9 spatially varied stations within the basin, for each 6-h period, and for both the dry and wet season. All reported values vary between near 0 and 0.1 with somewhat consistent diurnal and seasonal variation between rain gauges. As is typical of the region, the majority of rainfall occurs from midday to late afternoon.

To explore the sensitivity of the algorithm to the diurnal cycle of μ, we have performed a test using a synthetic example. In this case, μ is varied diurnally for each 6-h period in three different ways with $\mu_a = [0.025; 0.05; 0.075; 0.1]$, $\mu_b = [0.05; 0.075; 0.1; 0.025]$, and μ_c left constant at 0.05. A 30-day hourly resolution time series of precipitation is generated using a stochastic weather generator (Ivanov *et al.*, 2007). The synthetic 1-h data is then aggregated to 6-h accumulated values. A total of

1000 realizations of hourly precipitation are produced using each μ regime. A linearly downscaled product is included for comparison. The ensemble mean 6-h diurnal cycles, as well as the 10th and 90th percentiles are shown in Figure 2(a). CDFs of hourly rainfall are shown in Figure 2(b). The diurnal cycle and CDFs are fairly insensitive to the choice of μ. Most importantly, the stochastic disaggregation provides a more realistic rainfall product that can exhibit extreme rainfall events while preserving the mean. Figure 3 shows a typical realization from the sensitivity studies showing the variability of disaggregated precipitation. It is important to note that the insensitivity to the parameter μ is partly due to the fact that what is being disaggregated is 6-hourly model produced accumulated precipitation which is further normalized to preserve monthly accumulations.

Although the stochastic method is the preferred method to disaggregate 6-hourly accumulated precipitation, it is not applicable to the 6-hourly instantaneous CCSM3 values. To use the stochastic method, we must have a volume to disaggregate and preserve. Interpolating between instantaneous values to produce a spurious (generally small) volume and then to disaggregate it, is not an effective approach. The stochastic downscaling method is meant to disaggregate accumulated values, not interpolate them. Having said that whenever reasonable, the stochastic disaggregation should be the preferred method.

3.2. Shortwave radiation

To disaggregate 6-hourly shortwave radiation data to hourly values, a method that incorporates the solar zenith angle is used. The diurnal variation of incoming shortwave radiation can be determined by the solar zenith angle (Z) as a function of solar declination δ, latitude Φ in radians, and hour angle τ for every geographic location on the Earth's surface as

$$\cos(Z) = \max\{\sin(\delta)\sin(\Phi) + \cos(\delta)\cos(\Phi)\cos(\tau), 0\} \quad (5)$$

In this approach, the diurnal pattern of the shortwave radiation is generated by calculating the solar zenith angle. Then the calculated pattern is re-scaled to conserve the original 6-h averaged radiation energy provided by the modelled outputs (Knox, 2013). Using the solar zenith angle for disaggregation of downwelling shortwave radiation results in a realistic diurnal cycle of shortwave radiation.

3.3. Other hydrometeorological variables

The other hydrometeorological variables in CCSM3 and RegCM3 (temperature, specific humidity, longwave radiation, horizontal wind, and pressure) are interpolated linearly from 6-h to hourly temporal resolution. The 3-h bias-corrected PCM1 by Sheffield *et al.* (2006) and Li *et al.* (2010) are also interpolated to the hourly resolution.

4. Bias correction method

The nonlinear circulation and energy transfers in land surface–atmosphere interactions cause model errors to propagate easily across different spatiotemporal scales (Dai, 2006). In the following sections, we describe the approach we used to mitigate biases in temperature, precipitation, specific humidity, and downwelling longwave radiation for all chosen model outputs over the study domain. Note that, although physically based treatments of potential model biases are fundamentally preferred over statistically driven methods, the latter is often the only viable option for large-scale climate studies due to their effectiveness and computational tractability.

4.1. Temperature and precipitation

One approach to bias correction is to develop a statistical relationship between modelled and observed variables based on past experimental knowledge. Then the established relationship can be used to correct model biases not only for the past simulations but also for future predictions, assuming that the relationship remains almost stationary in time. One common approach is the so-called quantile-based method that maps the CDF of the biassed model outputs $\mathrm{CDF}_{\mathrm{MODh}}$ onto the distribution of unbiased observations $\mathrm{CDF}_{\mathrm{OBS}}$ (Panofsky and Brier, 1968; Hayhoe *et al.*, 2004; Cayan *et al.*, 2008; Maurer and Hidalgo, 2008).

An extended version of the CDF method is called the EDCDF matching method (Li *et al.*, 2010). This approach attempts to partially account for the distribution of the model projections in the CDF matching process. The EDCDF method not only uses the distributions of the historical observations (OBS) and model outputs (MODh) but also exploits the CDF of model projection (MODp). The idea is to find a transfer function based on the difference between the distribution of historically modelled and observed data (Δ) and then apply it to reduce the biases of the modelled values during the entire historical and future periods (see Figure 4(b)). The EDCDF method also assumes that the transfer function linking observations and model outputs remains time invariant and uses the incremental differences between the CDFs of the observations and the historical model outputs at each percentile (Δ) for bias correction of the entire period. More specifically, assuming the biassed climate variable of interest is denoted by X, then each percentile adjustment in the EDCDF method is computed as

$$\Delta = F^{-1}_{\mathrm{OBS}}\left(F_{\mathrm{MODp}}(X)\right) - F^{-1}_{\mathrm{MODh}}\left(F_{\mathrm{MODp}}(X)\right) \quad (6)$$

and thus

$$\widetilde{X} = X + \Delta \quad (7)$$

where $\widetilde{X}$ is an adjusted or bias-corrected value, while $F(\cdot)$ denotes the CDFs of observations (OBS), model in the historical climate (MODh), and future projection period (MODp). Figure 4 illustrates the CDF and the EDCDF methods for correction of the bias at an arbitrary point ($X = 3.5$, solid circle, selected for illustration purposes). Notice that, when F_{MODp} is close to F_{MODh} (see Equation (6)), the results of the EDCDF method approach the CDF results. The EDCDF method incorporates the distribution of model projections, therefore, it may perform

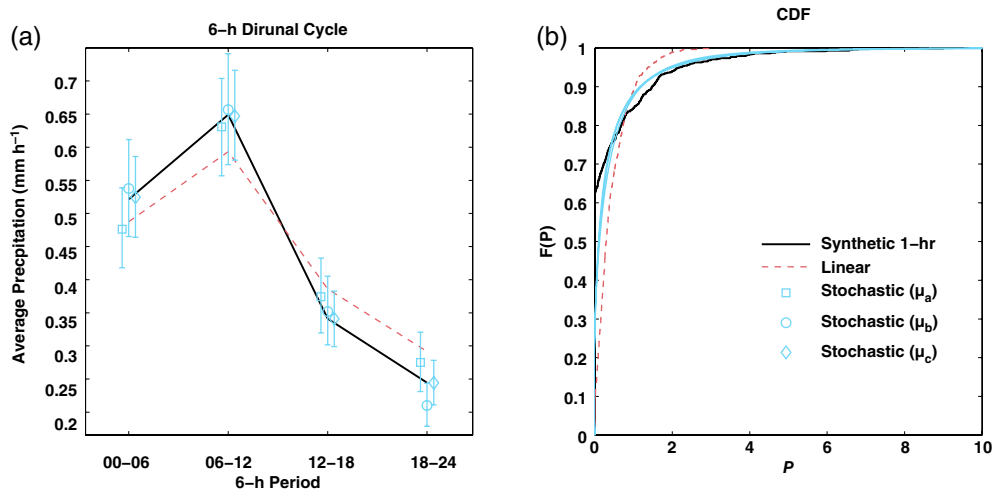

Figure 2. (a) Diurnal cycle and (b) CDFs of stochastically generated hourly rainfall (black solid line), linearly downscaled (red dashed line), and the ensemble mean of 1000 realizations of stochastically downscaled 6-h rainfall conditioned on the rainfall fraction (μ), where $\mu_a = [0.025; 0.05; 0.075; 0.1]$ (blue squares and lines), $\mu_b = [0.05; 0.075; 0.1; 0.025]$ (blue circles and lines), and μ_c left constant at 0.05 (blue diamonds and lines). The whiskers indicate the 10th and 90th percentile of the realizations.

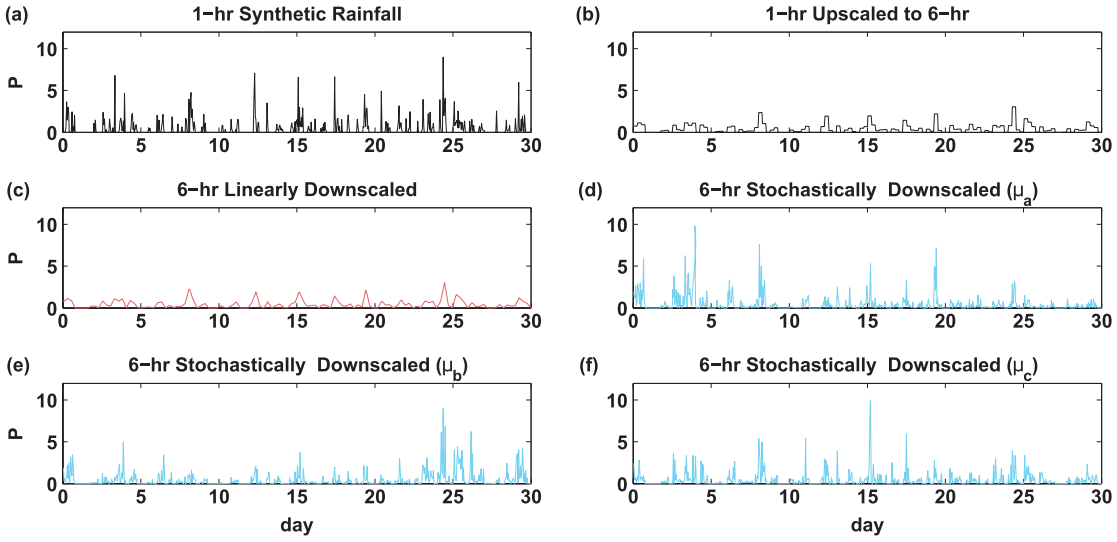

Figure 3. A 30-day time series of (a) synthetic hourly rainfall, (b) upscaled 6-h rainfall, (c) linearly downscaled rainfall, and (d–f) realizations of hourly rainfall conditioned on the rainfall fraction (μ), where $\mu_a = [0.025; 0.05; 0.075; 0.1]$, $\mu_b = [0.05; 0.075; 0.1; 0.025]$, and μ_c left constant at 0.05.

better than the CDF method, especially in capturing the projected climate variability (Li *et al.*, 2010).

In this study, the differences between monthly empirical CDFs of the observations and modelled outputs in the historical period at each percentile of the future projection (Δ) are used to correct biases of the monthly model projections (i.e. temperature and precipitation). Recall that the historical period for CCSM3 is 1940–2009, for RegCM3 is 1960-1990, and the future projection period is 2010–2099 (Table 1). The future projection is separated into 30-year periods (2010–2039, 2040–2069, 2070–2099) and CDF of MODp computed accordingly, to help minimize any non-stationarity in the future projection CDF.

For the temperature and precipitation variables, we first obtain bias-corrected monthly values using the EDCDF method. Then, for precipitation, we scale disaggregated hourly data (P^{h}) to conserve the bias-corrected monthly precipitation ($\widetilde{P}^{\mathrm{m}}$) as follows:

$$\widetilde{P}^{\mathrm{h}} = P^{\mathrm{h}} \times \frac{\widetilde{P}^{\mathrm{m}}}{P^{\mathrm{m}}} \tag{8}$$

where P^{m} is originally modelled monthly precipitation and $\widetilde{P}^{\mathrm{h}}$ denotes the bias-corrected hourly values. On the other hand, to correct biases in hourly temperature values (T^{h}), we use an additive adjustment scheme as

$$\widetilde{T}^{\mathrm{h}} = T^{\mathrm{h}} + \left(\widetilde{T}^{\mathrm{m}} - T^{\mathrm{m}}\right) \tag{9}$$

where T^{m} and $\widetilde{T}^{\mathrm{m}}$ are originally modelled and bias-corrected mean monthly values, while $\widetilde{T}^{\mathrm{h}}$ denotes the bias-corrected hourly temperatures. Notice that, analogous to precipitation, this adjustment scheme assures that the

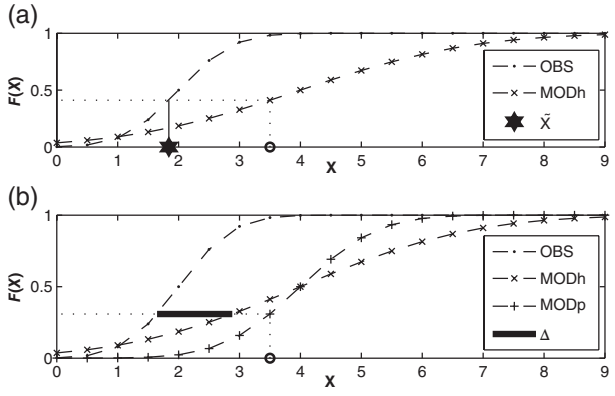

Figure 4. Illustration of the (a) CDF method and (b) EDCDF method for correction of the bias at an arbitrary point $X = 3.5$ (solid circle). Dashed line is the cumulative distribution function of the observation (CDF_{OBS}), cross-dashed line is the cumulative distribution function of the historical modelled variable (CDF_{MODh}), and plus-dashed line is the cumulative distribution function of the future modelled variable (CDF_{MODp}). Solid star in (a) is an adjusted value ($\widetilde{X}$) for the CDF method and thick solid line in (b) is an adjustment (Δ) for the EDCDF method (see Equation (6)).

bias-corrected mean monthly temperatures are properly conserved in hourly products.

4.2. Specific humidity

Adjustments of the temperature values can cause supersaturation and inconsistency between temperature and specific humidity. To resolve this inconsistency, specific humidity must be properly adjusted (Cosgrove *et al.*, 2003; Sheffield *et al.*, 2006). To this end, we express the hourly saturated vapour pressure $E^{\mathrm{h}}_{\mathrm{sat}}$ as a function of air temperature for pure water vapour as follows (Buck, 1981):

$$E^{h}_{sat} = a \times exp\left[\frac{\left(b - \frac{T^{h}}{d}\right) \times T^{h}}{\left(T^{h} + c\right)}\right], \tag{10}$$

where a, b, c, d are empirical parameters defined in Table 2, and T^{h} is the hourly modelled temperature before bias correction in degree Celsius. Buck (1981) also suggested that an enhancement factor f shall be used to account for moist air as follow:

$$f = 1 + \alpha + PS^{\mathrm{h}} \times \left[\beta + \gamma \times \left(T^{\mathrm{h}}\right)^{2}\right] \tag{11}$$

where PS^{h} is hourly modelled surface pressure (in millibar) and α, β, and γ are empirical constants reported in Table 2. Thus, the modified saturated vapour pressure ($E^{\mathrm{h}}_{\mathrm{sat,modif}}$) can be computed as

$$E^{\mathrm{h}}_{\mathrm{sat,modif}} = E^{\mathrm{h}}_{\mathrm{sat}} \times f \tag{12}$$

and the saturated specific humidity $Q^{\mathrm{h}}_{\mathrm{sat}}$ is

$$Q^{\mathrm{h}}_{\mathrm{sat}} = \frac{0.622 E^{\mathrm{h}}_{\mathrm{sat,modif}}}{\left(PS^{\mathrm{h}} - 0.378 E^{\mathrm{h}}_{\mathrm{sat,modif}}\right)} \tag{13}$$

As a result, the corresponding relative humidity RH^{h} is as follows:

$$RH^{\mathrm{h}} = \frac{E^{\mathrm{h}}}{E^{\mathrm{h}}_{\mathrm{sat,modif}}} = \frac{Q^{\mathrm{h}} \times \left(0.622 + 0.378 Q^{\mathrm{h}}_{\mathrm{sat}}\right)}{Q^{\mathrm{h}}_{\mathrm{sat}} \times \left(0.622 + 0.378 Q^{\mathrm{h}}\right)} \leq 1 \tag{14}$$

where Q^{h} denotes the hourly modelled specific humidity. To obtain the adjusted saturated specific humidity $Q^{\mathrm{h}}_{\mathrm{sat,adj}}$, bias-corrected hourly temperatures are used in all of the above equations. Then the hourly bias-corrected specific humidity $\widetilde{Q}^{\mathrm{h}}$ is expressed as

$$\widetilde{Q}^{\mathrm{h}} = Q^{\mathrm{h}}_{\mathrm{sat,adj}} \times RH^{\mathrm{h}} \tag{15}$$

The above steps are used to adjust specific humidity values for CCSM3. The RegCM3 provides dew point temperature and not the specific humidity. Therefore, one additional step is required to convert dew point temperature to specific humidity in this model's outputs. To this end, given the dew point temperature, the water vapour pressure E^{h} is calculated using Equation (10) and then is used in Equation (13) to obtain the specific humidity.

4.3. Downwelling longwave radiation

Changes in air temperature values influence downwelling longwave radiation. Although not explicitly bias corrected, we adjust the disaggregated longwave radiation using the modelled effective emissivity values, $\varepsilon_{\mathrm{eff}}$. The effective emissivity is computed using the hourly modelled longwave radiation ($LW^{\mathrm{h}}_{\mathrm{d}}$) and the original hourly temperature, and the Stefan–Boltzmann law as

$$\varepsilon_{\mathrm{eff}} = \frac{LW^{\mathrm{h}}_{\mathrm{d}}}{\sigma \left(T^{\mathrm{h}}\right)^{4}} \tag{16}$$

where σ is the Stefan–Boltzmann constant ($5.67 \times 10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}$). Then hourly bias-corrected downwelling longwave radiation ($\widetilde{LW}^{\mathrm{h}}_{\mathrm{d}}$) is identified using the emissivity values computed in Equation (16) and bias-corrected temperatures as,

$$\widetilde{LW}^{\mathrm{h}}_{\mathrm{d}} = \varepsilon_{\mathrm{eff}} \sigma \left(\widetilde{T}^{\mathrm{h}}\right)^{4} \tag{17}$$

It is important to note that changes in the temperature values play a more dominant role than changes in the emissivity values due to the fourth power temperature dependence in the Stefan–Boltzmann law. Thus, we adjust downwelling longwave radiation only based on the changes in temperature values. The reader is referred to Iziomon *et al.* (2003) for a review of these methods.

5. Results

To illustrate the spatial and temporal performance of the models, we use differences defined as the bias-corrected value minus the original model value. Hence, a negative value is an overestimation by the model, and a positive difference indicates an underestimation. The differences are shown for the historical and future periods. Figure 5

Table 2. Empirical parameters for calculation of vapour pressure as a function of temperature in Equations 10 and 11 (Buck, 1981).

Parameters	a	b	c	d	α	β	γ
T > 0 °C	6.1121	18.729	257.87	227.3	7.2×10^{-4}	3.2×10^{-6}	5.9×10^{-10}
T < 0 °C	6.1115	23.036	279.82	333.7	2.2×10^{-4}	3.83×10^{-6}	6.4×10^{-10}

illustrates the temporal average of estimated differences for CCSM3 and RegCM3 variables at 1° by 1° spatial resolution. In this figure, the rows from top to bottom show the spatial fields of the estimated differences over all grid cells for temperature (ΔT), precipitation (ΔP), specific humidity (ΔQ), and downwelling longwave radiation (ΔLW_{d}), respectively. The first two columns demonstrate the estimated differences in CCSM3 while the last two columns show the estimated differences for RegCM3 in the historical and future periods. The first row demonstrates that CCSM3 overestimates temperature in major parts of the domain, in particular over the western edge of the study domain. In contrast, RegCM3 underestimates temperature over most of those areas. The second row indicates that precipitation is generally underestimated by CCSM3, particularly over the west, north, and near the outlet of the basin. It is interesting that both CCSM3 and RegCM3 show a small underestimation of temperature by the models around the city of Manaus in the State of Amazonas. CCSM3 appears to overestimate precipitation near Manaus. A spatially similar anomaly is present in RegCM3, but it is shifted more to the northeast of Manaus. It appears that the two models, with their different parameterizations, are trying to reproduce unique circulation patterns of the region (and not doing that well). These circulation patterns respond to the fact that this is a very forested region of the basin interacting with a city with very different surface properties and energy fluxes.

The estimated differences in the precipitation fields are smaller in RegCM3 compared with CCSM3 over most of the study domain. Precipitation is underestimated by both models, but particularly by RegCM3 around the outlet of the basin. Poor performance of the models in simulating temperature and precipitation fields can be caused by inappropriate schemes and parameterizations used in the models over the above regions. Land cover representation can also contribute to the imperfect performance of the models. Differences between the land surface parameters of the models and land cover observations from remotely sensed moderate-resolution imaging spectroradiometer (MODIS) product are reported by Tian *et al.* (2004); Oleson *et al.* (2003), and Wang *et al.* (2004).

As Q and LW_{d} are adjusted based on the bias-corrected temperature values, the patterns of spatial differences for specific humidity and downwelling longwave radiation, shown in the third and the fourth rows, are similar to the patterns of the temperature differences shown in the first row. For instance, where ΔT is negative, the estimated difference for longwave radiation (ΔLW_{d}) is also negative and vice versa. We can also observe that the patterns of the differences in all fields for the historical and future periods are very similar, as expected.

Figure 6 illustrates the mean monthly domain average of the estimated differences between bias-corrected and the reference time series for CCSM3 and RegCM3 outputs in the historical and future periods. The figure illustrates the temporal performance of the models over the study domain. The first to fourth rows show estimated differences for temperature ΔT, precipitation (ΔP), specific humidity (ΔQ), and downwelling longwave radiation (ΔLW_{d}), respectively. The first column shows the results in the historical period and the second column is referred to the future period. From the first row, it can be inferred that CCSM3 overestimates temperature (negative difference) in all months except December where a very small underestimation occurs. RegCM3 underestimates temperature (positive difference) in all months except for August, September, and October (the dry-warm season). Both models tend to have a warm bias during this time. The second row indicates that, on average, the underestimation in the CCSM3 monthly precipitation fields is more significant than those by RegCM3 throughout the year. It is seen that the largest difference occurs during the rainy months in the Amazon (January to June) and the smallest difference is during the drier months (July through December) for both CCSM3 and RegCM3. As expected, the patterns of ΔQ and ΔLW_{d} in the third and fourth rows follow the same pattern as of temperature.

There are uncertainties regarding the predicted effects of climate change on the Amazon Basin and its ecosystem as a result of uncertainties in different GCMs' outputs. Therefore, comparing multiple models can perhaps shed more light on the existing uncertainties. For added comparison, we included disaggregated 3-hourly 1° by 1° bias-corrected PCM1 (publicly available at http://www.hydrology.princeton.edu/data.php) in the analysis. The downscaling methods used to create the data set are described by Sheffield *et al.* (2006), and the EDCDF method was used for bias correction with the CRU observations as the reference data set. To better understand the inter-annual trends of the domain averages of the climate variables of interest, Figure 7, from top to bottom, shows the annual bias-corrected values of CCSM3 and RegCM3 temperature, precipitation, specific humidity, and longwave radiation, respectively. The monthly CRU data are also shown for comparison purposes in Figure 7(a) and (b). To evaluate the statistical significance of the results, Table 3 presents the p-values of the linear trends in the modelled variables (T, P, Q, LW_{d}) for the historical and future periods. The measured p-value smaller than 0.05 or 0.01 is conventionally defined as 'statistically significant' or 'very significant', respectively. Figure 7(a) indicates that all models exhibit a long-term trend of rising temperature over the domain that is more statistically

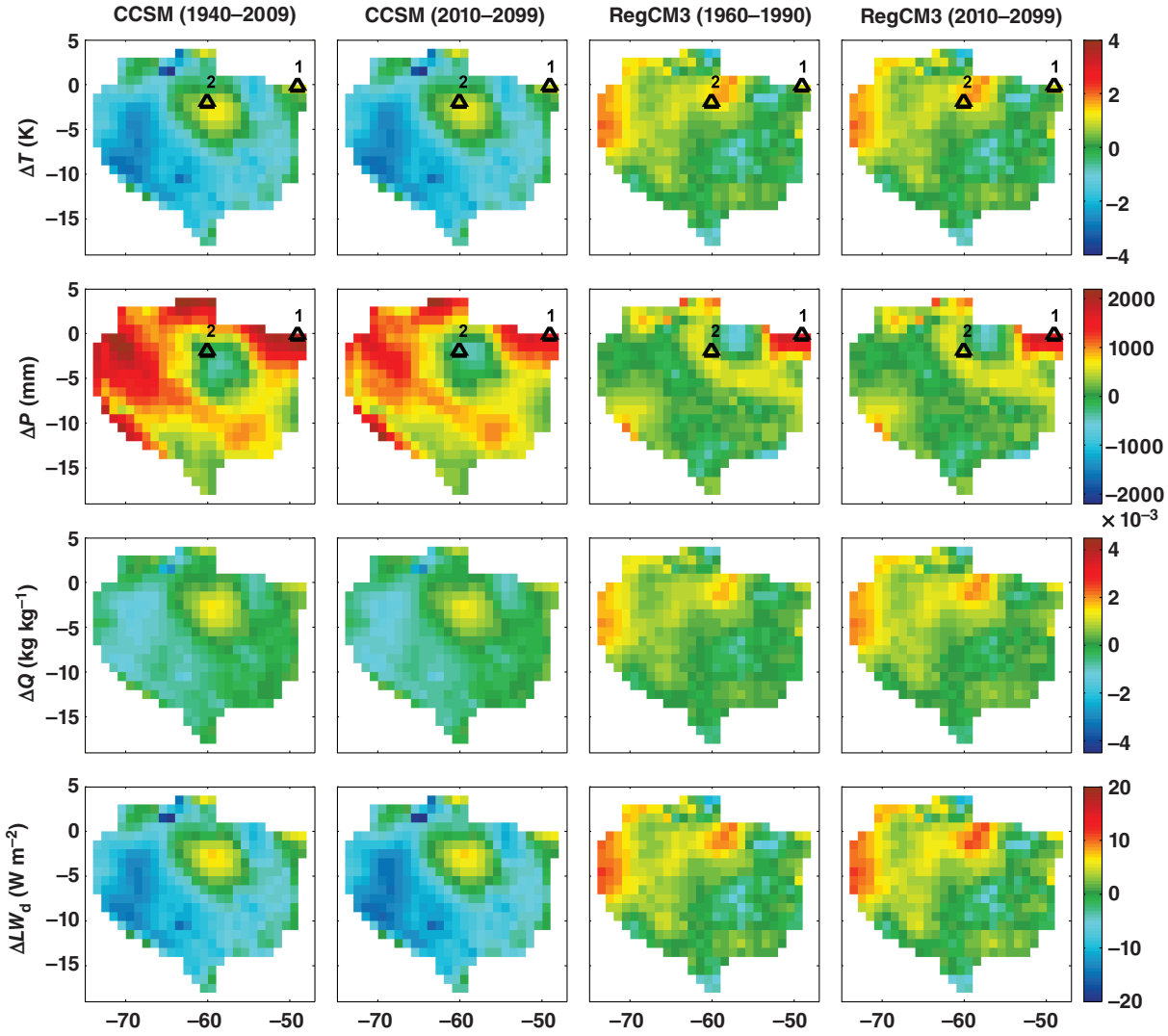

Figure 5. The temporal average of the estimated differences (bias-corrected minus original model value) of CCSM3 (first and second columns) and RegCM3 (third and fourth columns) in historical and future periods, respectively. From top to bottom: the estimated differences for temperature ΔT (K), precipitation ΔP (mm), specific humidity ΔQ (kg kg^{-1}), and downwelling longwave radiation. ΔLW_d (W m^{-2}), respectively. The outlet of the Amazon Basin and the city of Manaus are illustrated by numbers 1 and 2, respectively.

significant during the future period. The trend of rising temperature over the Amazon Basin has been previously reported (e.g. Malhi *et al.*, 2009). All models show similar performance representing the CRU reference temperature. For precipitation (Figure 7(b)), the monthly data from the Tropical Rainfall Measuring Mission (TRMM_3B43) product from 1998 to 2013 is also used as an independent data set for comparison purposes. The TRMM product (publicly available at http://trmm.gsfc.nasa.gov/) is known as one of the best estimates of precipitation, resulting from integrating multi-sensor precipitation data sets across different remote sensing platforms and ground-based rain gauges (Huffman *et al.*, 2007). The TRMM data at spatial resolution of 0.25° by 0.25° are averaged onto the horizontal grids used for the disaggregated GCM outputs. Figure 7(b) shows that the models agree well with CRU and TRMM precipitation when data are available. CCSM3 and PCM1 do not display a significant trend for precipitation in the historical period ($p = 0.29$ and 0.72, respectively). The upward trend in the TRMM and CRU precipitation is not statistically significant ($p = 0.24$ and 0.07, respectively), while RegCM3 shows a weak downward trend in the historical period ($p = 0.04$). Although CCSM3 and PCM1 show a significant increasing trend of precipitation for the future, RegCM3 displays a downward trend of precipitation. For specific humidity (Figure 7) the models exhibit upward trends, which are significant, with the exception of the historical period for RegCM3. The upward trends in specific humidity are consistent with the corresponding warming trends in atmospheric temperature that leads to increased moisture content. Similarly, the models exhibit significant upward trends in downwelling longwave radiation (Figure 7(d)), which are consistent with the model's warming trends. The upward trend of downwelling longwave radiation in CCSM3 and PCM1 is more significant than in RegCM3 during the historical time. It is also apparent that on average, RegCM3 shows a higher inter-annual variability

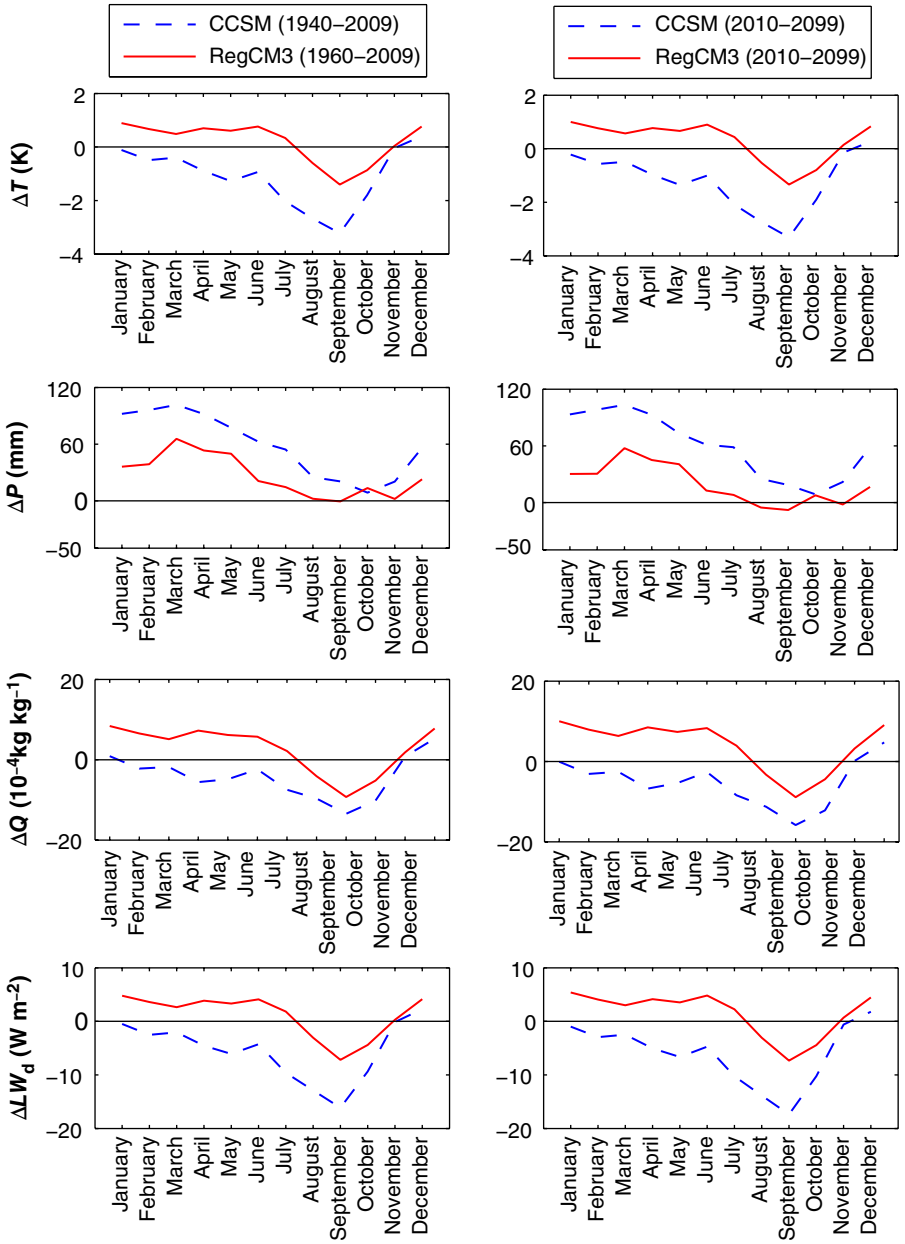

Figure 6. The mean monthly domain average of estimated differences (bias-corrected minus original model value) for CCSM3 and RegCM3 in the historical and future period. Dashed and solid lines are CCSM3 and RegCM3, respectively. The first to fourth rows are the estimated differences for temperature ΔT (K), precipitation ΔP (mm), specific humidity ΔQ (kg kg^{-1}), and downwelling longwave radiation ΔLW_d (W m^{-2}), respectively.

compared with the other models, particularly in future time periods.

Table 4 compares temporally averaged values of the annual domain mean and standard deviation of temperature, precipitation, specific humidity, and longwave radiation from the three models for the historical and future periods. In particular, on average, the mean temperatures increased by 2.22, 2.83, and 0.76°K, while comparing the historical years (1940–2009, 1960–1990, 1948–2009) with the future period (2010–2099) for CCSM3, RegCM3, and PCM1, respectively. Precipitation also increased 114.3, 116.9, and 116.8 mm in the bias-corrected data sets, respectively. Note that the average RegCM3 precipitation has increased from historical to future period conditioned on the fact that the data are missing from 1991 to 2009, which may affect the interpretation of the trends.

The potential impact of climate change varies over the region. To highlight the potential regional impacts of climate change, Figure 8 shows the difference between temporal averages of the bias-corrected modelled variables (T, P, Q, LW_{d}) in the future and historical periods (differences defined as future minus historical averages). This figure indicates that the potential impact of climate change on the differences is not evenly distributed over the study domain. All three models show warming over the eastern and the southern part of the Amazon Basin. The strongest, intermediate, and weakest warming

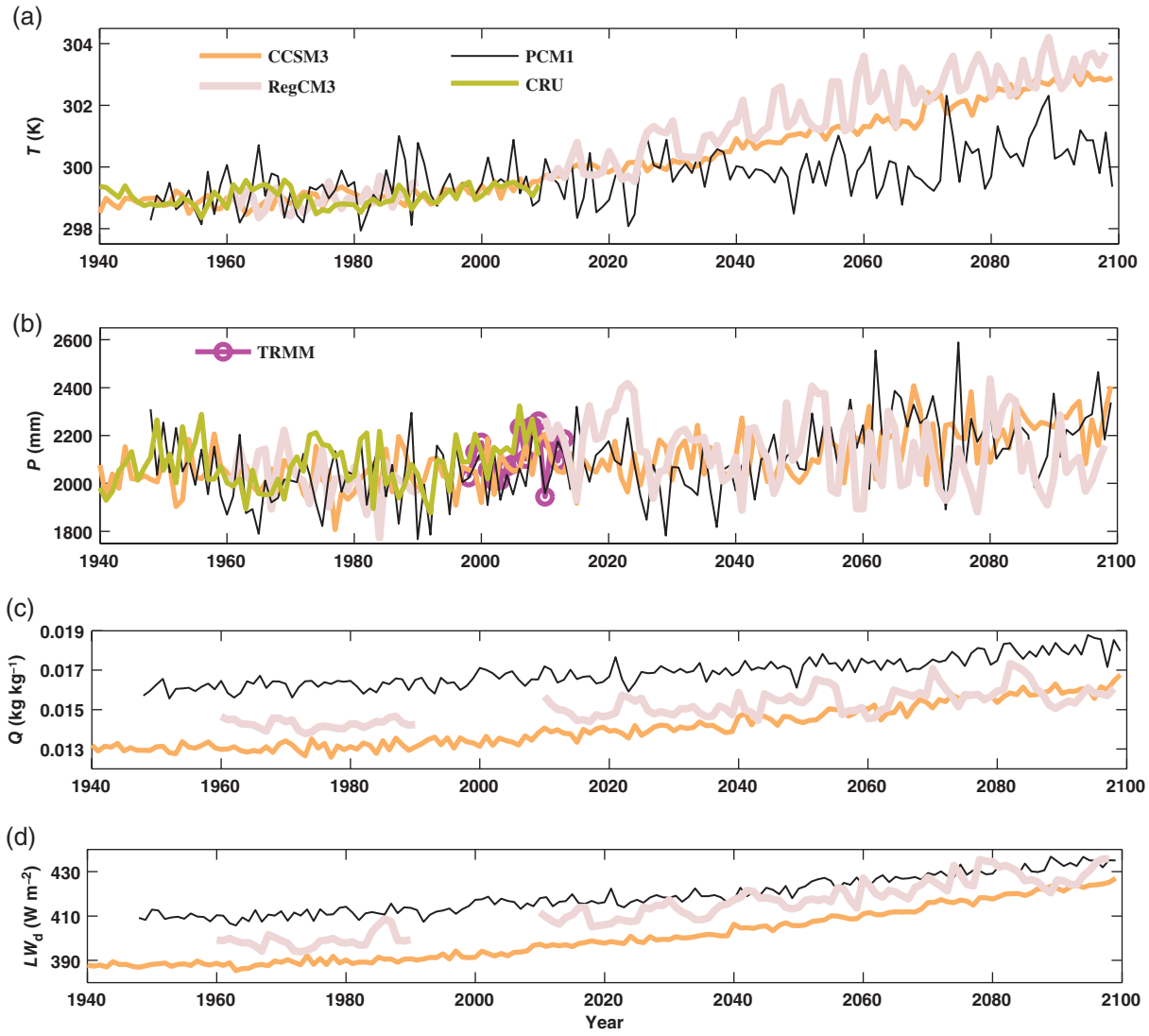

Figure 7. Annual domain average of bias-corrected (a) temperature T, (b) precipitation P, (c) specific humidity Q, and (d) downwelling longwave radiation LW_d for the models CCSM3, RegCM3, and PCM1.

Table 3. The statistical significance (p-value) of the linear trend in the modelled variables (T, P, Q, LW_d) for the historical and future periods.

Models		T	P	Q	LW_d
CCSM3	His	$p<0.0001$	$p=0.29$	$p<0.0001$	$p<0.0001$
	Fut	$p<0.0001$	$p<0.0001$	$p<0.0001$	$p<0.0001$
RegCM3	His	$p=0.02$	$p=0.04$	$p=0.75$	$p=0.04$
	Fut	$p<0.0001$	$p=0.02$	$p<0.0001$	$p<0.0001$
PCM1	His	$p=0.02$	$p=0.72$	$p<0.0001$	$p<0.0001$
	Fut	$p<0.0001$	$p<0.0001$	$p<0.0001$	$p<0.0001$

trends are in RegCM3, CCSM3, and PCM1, respectively, which is in agreement with previous IPCC AR4 model analysis of the region (Poulter *et al.*, 2009). The warming effect is reported to be related to the deforestation trend over those regions (Davidson *et al.*, 2012). While the models do not explicitly account for land cover change over time, CO_2 emissions related to deforestation and biomass burning are included in the A2 emission scenario. The areas affected by warming trends are smaller over the northwest of the study domain. The PCM1 results exhibit a cooling anomaly over the northern regions of the basin.

CCSM3 and RegCM3 indicate that precipitation may increase in the future period over the western parts of the study domain. Both models exhibit a decrease in precipitation over the eastern and central parts, in particular over the Bolivia and Mato Grosso (see Figure 8), where the vast areas of the Amazon rainforest are deforested (Bagley

Table 4. Comparison of the mean annual domain average of the statistics (mean μ and standard deviation σ) of the three modelled variables for the two historical and future periods.

Models		T (K)		P (mm)		Q (kg kg^{-1})		LW_d (W m^{-2})	
		μ	σ	μ	σ	μ	σ	μ	σ
CCSM3	His	299.01	0.28	2037.5	88.94	0.0132	2.60E-4	389.72	2.27
	Fut	301.23	1.05	2151.8	109.22	0.0149	8.21E-4	409.19	9.17
RegCM3	His	298.97	0.32	2019.6	91.90	0.0143	2.40E-4	398.96	3.48
	Fut	301.80	1.26	2136.5	143.13	0.0155	6.57E-4	419.29	8.30
PCM1	His	299.25	0.69	2029.7	125.81	0.0163	3.42E-4	411.63	2.98
	Fut	300.01	0.81	2146.5	148.66	0.0173	5.99E-4	424.39	6.39

et al., 2014). A decrease in precipitation over the eastern part of the Amazon Basin is also reported by Paiva and Clarke (1995); Reboita *et al.* (2014), and Bombardi and Carvalho (2009). The PCM1 and CCSM3 results show large precipitation decreases around the Belem and large precipitation increases around Manaus, which is consistent with the results by Chu *et al.* (1994). Spatially, the temperature and precipitation trends tend to be negatively correlated.

The spatial patterns of specific humidity trends closely follow those of precipitation in CCSM3 and RegCM3. This similarity may indicate that the higher water content of the air can increase the rainfall amount with the condition that the atmosphere has sufficient convective energy. The relatively large increase of LW_d seen in both CCSM3 and RegCM3 over the southwest (e.g. Bolivia) is concurrent with a strong warming trend in the region. This effect is less pronounced in the southeast where a significant drying of the region could lead to a decrease in cloud and air emissivity. The variability of specific humidity in PCM1 is much smaller when compared with the other two models, but similar spatial patterns emerge.

6. Discussion and conclusions

Efficient and effective long-term environmental management plans over the Amazon Basin require accurate and high-resolution climate data sets. This paper provides high-resolution bias-corrected climate data sets at a uniform space–time resolution, which can be used as forcing for ecosystem and land surface models over Amazonia.

We used a bilinear and linear interpolation approach to produce 1° by 1° and hourly variables from the outputs of CCSM3 and RegCM3. We also disaggregated 3-hourly PCM1 data to hourly resolution. For temporal disaggregation of shortwave radiation values, we used the solar zenith angle to account for the diurnal cycle of regional radiation regime. Although the bilinear interpolation may smooth out the space–time variability of the studied climate variables, this computationally efficient technique can provide realistic and sufficiently coherent disaggregated fields for the studied climate variables at the specified resolution.

To disaggregate accumulated RegCM3 precipitation data to hourly resolution, we used a stochastic approach that uses random samples of an exponential distribution to capture intermittency of precipitation. This method allows us to capture a more realistic variability of precipitation while preserving the total monthly values. A key parameter of this distribution represents the temporal average rate of wet periods over the disaggregation domain, which needs to be determined based on modelled and observed precipitation. To obtain this parameter, we used hourly precipitation from Santarem KM67 rain gauge during 2002–2005. It needs to be noted that although using only one rain gauge for obtaining this parameter was not a perfect choice, it was the best choice we could make given the lack of a dense and reliable long-term ground-based observation network in the study area. We also show that the sensitivity to the parameter is low. Linear interpolation was used to downscale the CCSM3 instantaneous precipitation product because the stochastic method is intended only for accumulated values. Users of this data set should note that linear interpolation smoothes the data and reduces the number of extreme rainfall values, which may be problematic for certain applications. The stochastic disaggregation is preferable when appropriate.

Limitations of most bias correction methods, including distribution mapping, are their statistical stationarity assumptions, sensitivities to selected training period, and the potential for overfitting. It is important to note that the EDCDF method cannot explicitly account for biases during particular years (e.g. El Niño or La Niña years). In other words, if the model is unable to properly simulate a specific phenomenon, or if the phenomenon is not present in the training period, the EDCDF method is not capable of correcting the associated biases. For instance, Johnson *et al.* (2011) evaluated the capability of GCMs to represent the variability of ENSO phenomenon. They showed that the performance of the CCSM3 and the HadCM3 in simulating spatial precipitation patterns and variability, SST, and surface pressure is better at inter-annual scale than at sub-annual and inter-decadal scale. It should be noted that the EDCDF method may not be capable of correcting the associated biases.

The use of empirical CDFs reduces the uncertainty associated with identifying and parameterizing a theoretical distribution; such approaches increase the risk of overfitting when applied outside the training period. Cross-validation tests have indicated that empirical quantile-mapping methods, including EDCDF, are not prone to this kind of over fitting provided sufficient

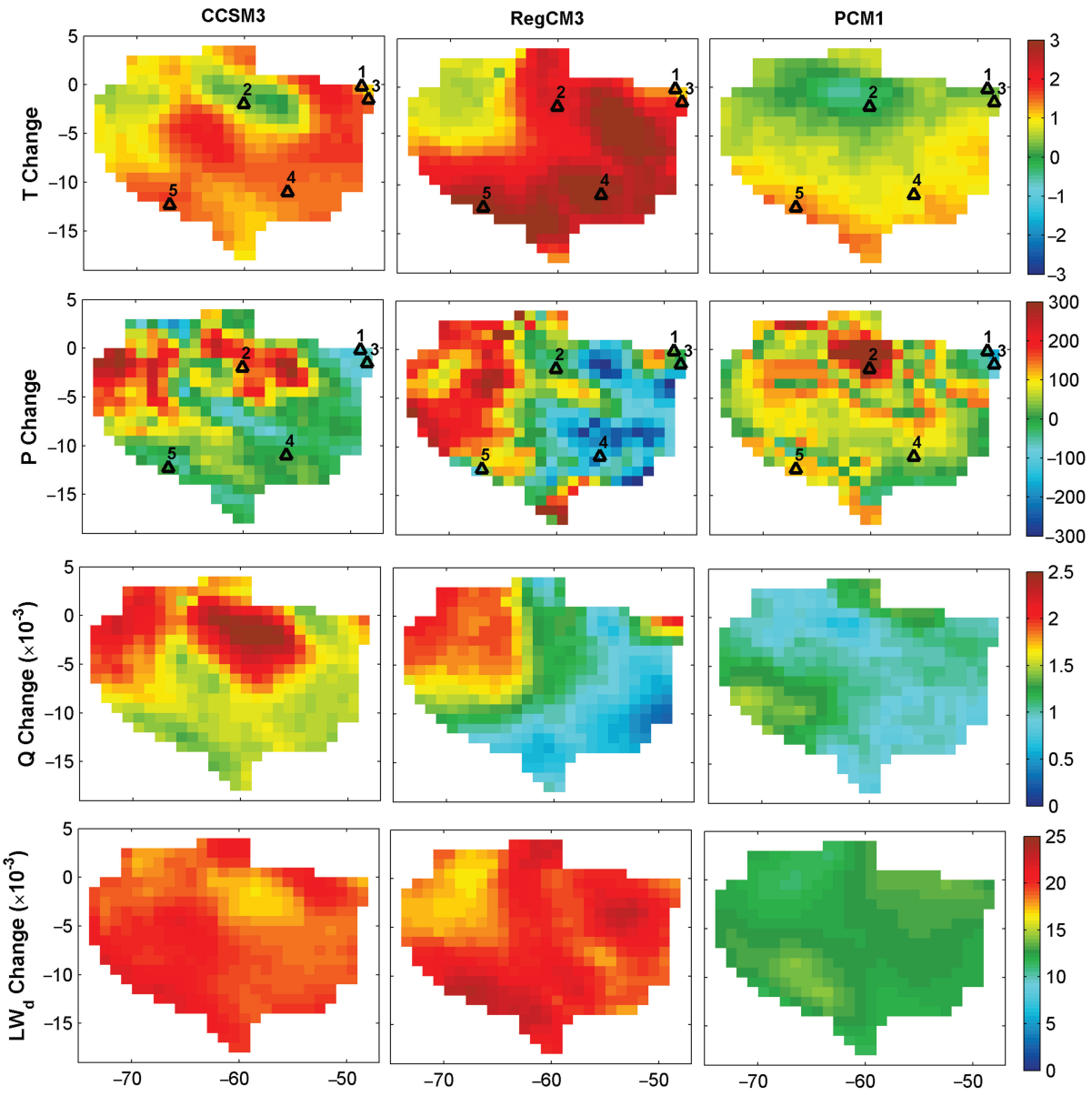

Figure 8. The difference of temporal averages of the models' variables in the future (2010–2099) from the corresponding ones in the historical period (1940–2009, 1960–1990, 1948–2009 for CCSM3, RegCM3, and PCM1, respectively). From top to bottom: changes in temperature (T change), precipitation (P change), specific humidity (Q change), and downwelling longwave radiation (LW_d change). From left-to-right columns: changes in CCSM3, RegCM3, and PCM1. The outlet of the Amazon Basin, Manaus, Belem, Mato Grosso, and Bolivia are illustrated by numbers 1, 2, 3, 4, and 5, respectively.

training periods (Lafon *et al.*, 2013; Teutschbein and Seibert, 2013; Bennett *et al.*, 2014). The advantage of using the EDCDF method is that it incorporates information from the projection period of the models to at least partially accommodate some features of potential future changes in first and higher order moments of climate variables. CDF matching attempts to fit the entire distribution and thus all moments are subject to change, including, but not limited to, the trends and variability of the data. The EDCDF method has been shown to perform well in periods with changed variability (Li *et al.*, 2010). Furthermore, the use of a long record of observation (CRU) from the historical period does add robustness to the method.

In general, CCSM3 tends to have a dry-warm bias in most parts of the domain, while RegCM3 mostly shows a tendency to have a dry-cool bias. These tendencies imply the presence of systematic errors in both models. The inadequate spatial resolution, physics schemes, and parameterizations of the models may influence their performances over different months and locations. On average, the corrected biases (estimated differences) in RegCM3 are smaller than those of CCSM3, for temperature and precipitation. RegCM3's higher resolution may better resolve scale-dependent processes. Further improvements might be due to a better parameterization, for this domain, of clouds, and other related processes in RegCM3 compared with CCSM3 (Smith, 1990; Milton

and Wilson, 1996; Gregory *et al.*, 1998). A better simulation of mesoscale patterns and topographic effects on precipitation in RegCM3, compared with other GCMs, is reported by many studies (e.g. Frei *et al.*, 2003, 2006; Buonomo *et al.*, 2007; Reboita *et al.*, 2014, among others).

Bias-corrected annual CCSM3, RegCM3, and PCM1 temperature showed a clear warming trend under the A2 scenario. The results also exhibit an upward trend in specific humidity and downwelling longwave radiation, which is consistent with the warming trend of temperature. Although no significant trend of precipitation is observed in the historical period, the RegCM3 results show a slight downward trend in precipitation, while the CCSM3 and PCM1 results show a more significant upward trend. The CCSM3 and PCM1 results mostly show lower inter-annual variability than the RegCM3 outputs, especially in the future period.

Regional analysis comparing the historical and future period revealed that the eastern and southern part of the region, where deforestation rate and risk of fire are high, experience a warmer and drier atmosphere under the A2 scenario. The results, especially those of CCSM3 and RegCM3, indicate that north and northwestern parts of the region may experience an increasing trend in precipitation, which is strongly correlated with the observed increases in the air humidity.

Climate predictions models are subject to model error due to the uncertainties in parameterizations of the underlying physics. As a result, bias correction and efficient downscaling methods are of value to climate research community. The bias-corrected projections presented in this study will allow for detailed studies of potential impacts of future climate change on the Amazon region.

Acknowledgements

This work was supported by the Andes-Amazon Initiative of the Gordon and Betty Moore Foundation. Additional support was provided to Prof. Bras by the K. Harrison Brown Family Chair. We are thankful to all sponsors. GCMs data sets for this study are provided by the UCAR, the Land Surface Hydrology Research Group in Princeton University, and CPTEC/INPE. The CRU data were created by the Climatic Research Unit, University of East Anglia (UK). We also appreciate Drs. Gaj Sivandran, Gautam Bisht, Anthony Parolari, and Liao-fan Lin for quality control of the data. We thank Drs Justin Sheffield, Marcos Heil Costa, and Adan Juliano de Paula Silva for providing some of the data sets used in this study.

References

Aragao L, Malhi Y, Barbier N, Anderson L, Saatchi S, Shimabukuro E. 2008. Interactions between rainfall, deforestation and fires during recent years in the Brazilian Amazonia. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**: 1779–1785.

Bagley JE, Desai AR, Harding KJ, Snyder PK, Foley JA. 2014. Drought and deforestation: has land cover change influenced recent precipitation extremes in the Amazon? *J. Clim.* **27**: 345–361.

Bennett JC, Grose MR, Corney SP, White CJ, Holz GK, Katzfey JJ, Post DA, Bindoff NL. 2014. Performance of an empirical bias-correction of a high-resolution climate dataset. *Int. J. Climatol.* **24**: 2189–2204.

Betts RA, Cox PM, Collins M, Harris PP, Huntingford C, Jones CD. 2004. The role of ecosystem-atmosphere interactions in simulated Amazonian precipitation decrease and forest dieback under global climate warming. *Theor. Appl. Climatol.* **78**: 157–175, doi: 10.1007/s00704-004-0050-y.

Bombardi RJ, Carvalho LMV. 2009. IPCC global coupled model simulations of the South America monsoon system. *Clim. Dyn.* **33**: 893–916.

Bonan GB, Levis S. 2006. Evaluating aspects of the Community Land and Atmosphere Models (CLM3 and CAM3) using a dynamic Global Vegetation Model. *J. Clim.* **19**: 2290–2301, doi: 10.1175/JCLI3741.1.

Buck AL. 1981. New equations for computing vapor pressure and enhancement factor. *J. Appl. Meteorol. Climatol.* **20**: 1527–1532.

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

Burke EJ, Brown SJ, Christidis N. 2006. Modeling the recent evolution of global drought and projections for the twenty-first century with the Hadley Centre Climate Model. *J. Hydrometeorol.* **7**: 1113–1125.

Bush MB, Silman MR, McMichael C, Restrepo-Correa A, Urrego DH, Correa A, Saatchi S. 2008. Fire, climate change, and biodiversity in Amazonia: a late-holocene perspective. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**: 1795–1802.

Cayan DR, Maurer EP, Dettinger MD, Tyree M, Hayhoe K. 2008. Climate change scenarios for the California region. *Clim. Change* **87**: S21–S42, doi: 10.1007/s10584-007-9377-6.

Chen J, Brissette FP, Chaumont D, Braun M. 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.* **49**: 4187–4205.

Chou SC, Marengo JA, Lyra AA, Sueiro G, Pesquero JF, Alves LM, Kay G, Betts R, Chagas DJ, Gomes JL, Bustamante JF, Tavares P. 2012. Downscaling of South America present climate driven by 4-member HadCM3 runs. *Clim. Dyn.* **38**: 635–653.

Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held R, Jones R, Kolli RK, Kwon WK, Laprise R, Rueda VM, Mearns L, Menendez CG, Räisänen J, Rinke A, Sarr A, Whetton P, Arritt R, Benestad R, Beniston M, Bromwich D, Caya D, Comiso J, de Elia R, Dethloff K. 2007. Regional climate projections, climate change, 2007. The physical science basis, *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Chu P-S, Yu ZP, Hastenrath S. 1994. Detecting climate change concurrent with deforestation in the Amazon basin: which way has it gone? *Bull. Am. Meteorol. Soc.* **75**: 579–583.

Clark R, Buarque D. 2013. Statistically combining rainfall characteristics estimated from remotely sensed and rain gauge data sets in the Brazilian Amazon-Tocantins basins. *J. Geophys. Res.* **118**: 7467–7480.

Cochrane MA, Barber CP. 2009. Climate change, human land use and future fires in the Amazon. *Glob. Chang. Biol.* **15**: 601–612, doi: 10.1111/j.1365-2486.2008.01786.x.

Coelho CAS, Goddard L. 2009. El Niño-induced tropical droughts in climate change projections. *J. Clim.* **22**: 6456–6476.

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, Mckenna DS, Santer BD, Smith RD. 2006. The Community Climate System Model version 3 (CCSM3). *J. Clim.* **19**: 2122–2143.

Conway D. 1998. Recent climate variability and future climate change scenarios for Great Britain. *Prog. Phys. Geogr.* **22**: 350–374.

Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock A, Marshall C, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD, Meng J. 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.* **108**: 8842, doi: 10.1029/2002JD003118.

Costa MH, Pires GF. 2010. Effects of Amazon and Central Brazil deforestation scenarios on the duration of the dry season in the arc of deforestation. *Int. J. Climatol.* **30**: 1970–1979.

Cox PM, Pearson D, Booth BB, Friedlingstein P, Huntingford C, Jones CD, Luke CM. 2013. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**: 341–344.

Dai A. 2001a. Global precipitation and thunderstorm frequencies. Part I: seasonal and interannual variations. *J. Clim.* **14**: 1092–1111.

Dai A. 2001b. Global precipitation and thunderstorm frequencies. Part II: diurnal variations. *J. Clim.* **14**: 1112–1128.

Dai A. 2006. Precipitation characteristics in eighteen Coupled Climate Models. *J. Clim.* **19**: 4605–4630.

Davidson EA, de Arajo AC, Artaxo P, Balch JK, Brown IF, Bustamante MMC, Coe MT, DeFries RS, Keller M, Longo M, Munger JW, Schroeder W, Soares-Filho BS, Souza CM, Wofsy SC. 2012. The Amazon basin in transition. *Nature* **481**: 321–328.

Eltahir E, Bras RL. 1993. On the response of the tropical atmosphere to large-scale deforestation. *Q. J. R. Meteorol. Soc.* **119**: 779–793.

Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.* **27**: 1547–1578, doi: 10.1002/joc.1556.

Frei C, Christensen JH, Deque M, Jacob D, Jones RG, Vidale PL. 2003. Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. *J. Geophys. Res.* **108**: 4124.

Frei C, Scholl R, Fukutome S, Schmidli J, Vidale PL. 2006. Future change of precipitation extremes in Europe: intercomparison of scenarios from regional climate models. *J. Geophys. Res.* **111**: D06105.

Georgakakos AP, Yao H, Kistenmacher M, Georgakakos KP, Graham NE, Cheng F-Y, Spencer C, Shamir E. 2012. Value of adaptive water resources management in Northern California under climatic variability and change: reservoir management. *J. Hydrol.* **412-413**: 34–46.

de Goncalves LGG, Restrepo-Coupe N, da Rocha HR, Saleska SR, Stockli R. 2013. LBA-ECO CD-32 LBA Model Intercomparison Project (LBA-MIP) Forcing Data. Data set. Oak Ridge National Laboratory Distributed Active Archive Center: Oak Ridge, TN, doi: 10.3334/ORNLDAAC/1177.

Gregory D, Shutts GJ, Mitchell JR. 1998. A new gravity wave drag scheme incorporating anisotropic orography and low level wave breaking: impact upon the climate of the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **124**: 463–493.

Guilyardi E. 2006. El Niño mean state-seasonal cycle interactions in a multi-model ensemble. *Clim. Dyn.* **26**: 329–348.

Harding R, Best M, Blyth E, Hagemann S, Kabat P, Tallaksen LM, Warnaars T, Wiberg D, Weedon GP, Lanen HV, Ludwig F, Haddeland I. 2011. WATCH: current knowledge of the terrestrial global water cycle. *J. Hydrometeorol.* **12**: 1149–1156, doi: 10.1175/JHM-D-11-024.1.

Hayhoe K, Cayan D, Field CB, Frumhoff PC, Maurer EP, Miller NL, Moser SC, Schneider SH, Cahill KN, Cleland EE, Dale L, Drapek R, Hanemann RM, Kalkstein LS, Lenihan J, Lunch CK, Neilson RP, Sheridan SC, Verville JH. 2004. Emissions pathways, climate change, and impacts on California. *Proc. Natl. Acad. Sci. USA* **101**: 12422–12427, doi: 10.1073/pnas.0404500101.

Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF. 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**: 38–55.

Ines AVM, Hansen JW. 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteorol.* **138**: 44–53, doi: 10.1016/j.agrformet.2006.03.009.

Ivanov V, Bras R, Curtis D. 2007. A weather generator for hydrological, ecological, and agricultural applications. *Water Res. Res.* **43**(10): W10406, doi: 10.1029/2006WR005364.

Iziomon M, Mayer H, Matzarakis A. 2003. Downward atmospheric longwave irradiance under clear skies: measurement and parameterization. *J. Atmos. Sol.-Terr. Phys.* **65**: 1107–1167.

Johnson F, Westra S, Sharma A, Pitman AJ. 2011. An assessment of GCM skill in simulating persistence across multiple time scales. *J. Clim.* **24**: 3609–3623.

Knox RG. 2013. Land Conversion in Amazonia and Northern South America: Influences on Regional Hydrology and Ecosystem Response. PhD thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology: Cambridge, MA.

Lafon T, Dadson S, Buys G, Prudhomme C. 2013. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *Int. J. Climatol.* **33**: 1367–1381.

Lammering B, Dwyer I. 2000. Improvement of water balance in land surface schemes by random cascade disaggregation of rainfall. *Int. J. Climatol.* **20**: 681–695.

Lettenmaier DP, Rind D. 1992. Hydrological aspects of global climate change – preface. *J. Geophys. Res.* **97**: 2675–2676.

Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.* **115**: D10101, doi: 10.1029/2009JD012882.

Malhi Y, Aragao LEOC, Galbraith D, Huntingford C, Fisher R, Zelazowski P, Sitch S, McSweeney C, Meir P. 2009. Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proc. Natl. Acad. Sci. USA* **106**: 20610–20615.

Marengo JA, Chou SC, Kay G, Alves LM, Pesquero JF, Soares WR, Santos DC, Lyra AA, Sueiro G, Betts R, Chagas DJ, Gomes JL, Bustamante JF, Tavares P. 2012. Development of regional future climate change scenarios in South America using the Eta CPTEC/HadCM3 climate change projections: climatology and regional analyses for the Amazon, Sao Francisco and the Parana River basins. *Clim. Dyn.* **38**: 1829–1848.

Maurer EP, Hidalgo HG. 2008. Utility of daily *vs*. monthly largescale climate data: an intercomparison of two statistical downscaling methods. *Hydrol. Earth Syst. Sci.* **12**: 551–563.

Maurer EP, Brekke L, Pruitt T, Duffy PB. 2007. Fine-resolution climate projections enhance regional climate change impact studies. *EOS* **88**: 504, doi: 10.1029/2007EO470006.

Mearns LO, Gutowski WJ, Jones R, Leung LY, McGinnis S, Nunes AMB, Qian Y. 2009. A regional climate change assessment program for North America. *EOS* **90**: 311–312.

Meehl GA, Teng H, Branstator G. 2006. Future changes of El Niño in two global coupled climate models. *Climate Dynam.* **26**: 549–566.

Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ, Zhao ZC. 2007. Global climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge, UK and New York, NY.

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ. 2008. Stationarity is dead: Whither water management? *Science* **319**: 573–574, doi: 10.1126/science.1151915.

Milton SF, Wilson CA. 1996. The impact of parameterized sub-grid scale orographic forcing on systematic errors in a global NWP model. *Mon. Weather Rev.* **124**: 2023–2045.

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

Nakicenovic, N., O. Davidson, G. Davis, A. Grübler, T. Kram, E. L. L. Rovere, B. Metz, T. Morita, W. Pepper, H. Pitcher, A. Sankovski, P. Shukla, R. Swart, R. Watson, and Z. Dadi (2000), Emissions Scenarios. *A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, IPCC [Summary for Policymakers].

van Oldenborgh GJ, Philip SY, Collins M. 2005. El Niño in a changing climate: a multi-model study. *Ocean Sci.* **65**: 1107–1116.

Oleson KW, Bonan GB, Schaaf C, Gao F, Jin Y, Strahler AH. 2003. Assessment of global climate model land surface albedo using MODIS data. *Geophys. Res. Lett.* **30**: 1443.

Paiva EMCD, Clarke RT. 1995. Time trends in rainfall records in Amazonia. *Bull. Am. Meteorol. Soc.* **76**: 2203–2209.

Panofsky HA, Brier GW. 1968. *Some Application of Statistics to Meteorology*. University Park: PA, 224.

Piani C, Haerter JO, Coppola E. 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* **99**: 187–192, doi: 10.1007/s00704-009-0134-9.

Poulter B, Aragao L, Heyder U, Gumpenberger M, Heinke J, Langerwisch F, Rammig A, Thonicke K, Cramer W. 2009. Net Biome production of the Amazon Basin in the 21st century. *Global Change Biol.* **16**: 2062–2075, doi: 10.1111/j.1365-2486.2009.02064.x.

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer R, Sumi A, Taylor K. 2007. Climate models and their evaluation. In *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge and New York, NY.

Reboita MS, da Rocha RP, Dias CG, Ynoue RY. 2014. Climate projections for South America: RegCM3 driven by HadCM3 and ECHAM5. *Adv. Meteorol.* **2014**: 1–17.

Reichler T, Kim J. 2008. How well do coupled models simulate today's climate. *Bull. Am. Meteorol. Soc.* **89**: 303–311, doi: 10.1175/BAMS-89-3-303.

Rheenen NTV, Wood AW, Palmer RN, Lettenmaier DP. 2004. Potential implications of PCM climate change scenarios for Sacramento-Jan

Joaquin River Basin hydrology and water resources. *Clim. Change* **62**: 257–281.
Saleska SR, Didan K, Huete AR, da Rocha HR. 2007. Amazon forests green-up during 2005 drought. *Science* **318**: 612.
Sheffield J, Goteti G, Wood EF. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**: 3088–3111.
Shin SI, Liu Z, Otto-Bliesner B, Brady EC, Kutzbach JE, Harrison S. 2003. A simulation of the last glacial maximum climate using the NCAR-CCSM. *Clim. Dynam.* **20**: 127–151.
Smith RNB. 1990. A scheme for predicting layer clouds and their water content in a general circulation model. *Q. J. R. Meteorol. Soc.* **116**: 435–460.
Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor M, Miller HL (eds). 2007. Climate change 2007: the physical science basis. In *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC WGI Fourth Assessment Report. Cambridge University Press: Cambridge and New York, NY.
Stamm JF, Wood EF, Lettenmaier DP. 1994. Sensitivity of a GCM simulation of global climate to the representation of land-surface hydrology. *J. Clim.* **7**: 1218–1239.
Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ. 2000. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**: 2133–2137.
Sun Y, Solomon S, Dai A, Portmann R. 2005. How often does it rain. *J. Clim.* **19**: 916–934.
de Szoeke SP, Xie SP. 2008. The Tropical Eastern Pacific seasonal cycle: assessment of errors and mechanisms in IPCC AR4 coupled ocean-atmosphere general circulation models. *J. Clim.* **21**: 2573–2590.
Tanaka SK, Zhu T, Lund JR, Howitt RE, Jenkins MW, Pulido MA, Tauber M, Ritzema RS, Ferreira IC. 2006. Climate warming and water management adaptation for California. *Clim. Change* **76**: 361–387, doi: 10.1007/s10584-006-9079-5.
Teutschbein C, Seibert J. 2013. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.* **17**(12): 5061–5077.
Tian Y, Dickinson RE, Zhou L, Myneni R, Friedl M, Schaaf C, Carroll M, Gao F. 2004. Land boundary conditions from MODIS data and consequences for the albedo of a climate model. *Geophys. Res. Lett.* **31**: L05504.
Wang Z, Zeng X, Barlage M, Dickinson RE, Gao F, Schaaf C. 2004. Using MODIS BRDF and albedo data to evaluate global model land surface albedo. *J. Hydrometeorol.* **5**: 3–14.
Wilby RL, Wigley TML. 1997. Downscaling general circulation model output: a review of methods and limitations. *Phys. Geogr.* **21**: 530–548, doi: 10.1177/030913339702100403.
Wood AW, Leung LR, Sridhar V, Lettenmaier DP. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Change* **62**: 189–216.
Xu C. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modeling approaches. *Phys. Geogr.* **23**: 229–249, doi: 10.1177/030913339902300204.

