Representation of climate extreme indices in the coupled atmosphere-land surface model ACCESS1.3b

R. Lorenz\textsuperscript{1}, A. J. Pitman\textsuperscript{1}, M. G. Donat\textsuperscript{1}, A. L. Hirsch\textsuperscript{1}, J. Kala\textsuperscript{1}, E. A. Kowalczyk\textsuperscript{2}, R. M. Law\textsuperscript{2}, and J. Srbinovsky\textsuperscript{2}

\textsuperscript{1}ARC Center of Excellence for Climate System Science and Climate Change Research Center, University of New South Wales, Australia
\textsuperscript{2}Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Australia

Correspondence to: R. Lorenz
(r.lorenz@unsw.edu.au)
Abstract

Climate extremes, such as heat waves and heavy precipitation events, have large impacts on ecosystems and societies. Climate models provide useful tools to study underlying processes and amplifying effects associated with extremes. The Australian Community Climate and Earth System Simulator (ACCESS) has recently been coupled to the Community Atmosphere Biosphere Land Exchange model. We examine how this model represents climate extremes derived by the Expert Team on Climate Change Detection and Indices and compare them to observational datasets using the AMIP framework. We find that the patterns of extreme indices are generally well represented. Indices based on percentiles are particularly well represented and capture the trends over the last 60 years shown by the observations remarkably well. The diurnal temperature range is underestimated, minimum temperatures ($T_{MIN}$) during nights are generally too warm and daily maximum temperatures ($T_{MAX}$) too low in the model. The number of consecutive wet days is overestimated while consecutive dry days are underestimated. The maximum consecutive 1-day precipitation amount is underestimated on the global scale. Biases in $T_{MIN}$ correlate well with biases in incoming longwave radiation, suggesting a relationship with biases in cloud cover. Biases in $T_{MAX}$ depend on biases in net shortwave radiation as well as evapotranspiration. The regions and season where the bias in evapotranspiration plays a role for the $T_{MAX}$ bias correspond to regions and seasons where soil moisture availability is limited. Our analysis provides the foundation for future experiments that will examine how land surface processes contribute to these systematic biases in the ACCESS modelling system.

1 Introduction

Climate extremes, including heat waves, heavy precipitation events or droughts have important effects on ecosystems and society (Easterling, 2000; Ciais et al., 2005; Pall et al., 2011). Many climate extremes are related to natural variability (Arblaster and Alexander, 2012; Seneviratne et al., 2012). However, climate change has the ability to modify the frequency, intensity, spatial extent, duration, and timing of climate extremes, and can lead to events unprecedented in the
historical record (Seneviratne et al., 2012). Given the impact of extremes, it is important to understand their causes, how they might change in the future and the role of potential interacting processes and feedbacks that might amplify them. This is urgent given some extremes appear to be increasing in frequency (Alexander et al., 2006; Coumou and Rahmstorf, 2012; Donat and Alexander, 2012; Perkins et al., 2012) and that extremes are considered to be a particularly challenging aspect of climate change adaptation (IPCC, 2012).

Extreme events can be directly influenced by land surface processes. Heat waves for example can be amplified by land surface processes including dryness and decreased vegetation (Zaitchik et al., 2006; Fischer et al., 2007; Koster et al., 2009; Hirschi et al., 2010; Stéfanon et al., 2012; Lorenz et al., 2013). Limited soil moisture availability and less active vegetation decreases evapotranspiration and, therefore, more energy is available for the sensible heat flux which increases temperatures (e.g., Seneviratne et al., 2010). Temperature variability is also affected by land surface processes (Seneviratne et al., 2006) and Jaeger and Seneviratne (2010) found a tendency for a greater impact of land surface processes on maximum temperatures, as distinct from minimum temperatures. The land-atmosphere coupling mechanisms for maximum and minimum temperatures differ. A clear relationship between maximum temperature and realistic soil moisture initialisation was found by Hirsch et al. (2014). This is in contrast to minimum temperatures where the influence of soil moisture is less clear given the role of net longwave emission in modulating nighttime temperatures. Interactions between the land surface and precipitation also exist, but the scale of the impact is less clear. It remains uncertain how soil moisture affects rainfall, with no agreement on the sign of the feedback (Findell and Eltahir, 2003; Ek and Holtslag, 2004; Taylor and Ellis, 2006). Pitman et al. (2012) analysed several global climate models to examine the impact of land use changes on temperature and precipitation extremes and found opposing effects to the impact of increasing CO$_2$ for some extreme indices and additive impacts for others. In short, to understand the changes in extremes linked with natural variability or increasing CO$_2$ we need to understand how land surface processes influence climate extremes.

Land-atmosphere feedbacks are difficult to investigate using observations alone because of the lack of suitable long-term datasets and the uncertainty regarding how feedbacks might
change in the future with climate change. Climate models are useful tools to investigate land-atmosphere feedbacks and their influence on extreme events (Fischer et al., 2007; Jaeger and Seneviratne, 2010; Lorenz et al., 2010). In Australia, a new global earth system model (ACCESS) has been developed (Bi et al., 2013; Kowalczyk et al., 2013). ACCESS1.0 compares well with other models in the Coupled Model Intercomparison Project, version 5 (CMIP-5) with regards to the representation of extremes (Sillmann et al., 2013). A more recent version of the model, ACCESS1.3, has also been used to run CMIP-5 simulations and performs similarly to ACCESS1.0 (Kowalczyk et al., 2013). Two major differences between ACCESS1.0 and ACCESS1.3 are the parameterization of the land surface and the cloud scheme. In ACCESS 1.0, the UK Meteorological Office Surface Exchange Scheme (MOSES) is used, but is replaced in ACCESS1.3 by the Community Atmosphere Biosphere Land Exchange (CABLE1.8) model. The two versions of the model were compared by Kowalczyk et al. (2013) although this analysis did not examine the representation of extreme events.

In this study, we undertake an analysis of the ACCESS1.3 model in terms of its ability to simulate a selection of extremes. We use ACCESS1.3, but replace CABLE1.8 with CABLE2.0 (the most recent released version) in an overall modelling system labelled ACCESS1.3b. We use an Atmospheric Model Intercomparison Project (AMIP) style experimental design (Gates, 1992) involving simulations over the period 1950–2012 with prescribed sea surface temperatures and sea ice concentration. The use of the AMIP experimental design decreases the uncertainty in terms of sea surface temperatures and associated teleconnections including the El Niño-Southern Oscillation. Our analysis focuses on the simulation of climate extreme indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) which are provided by two observational datasets by Donat et al. (2013b) and Donat et al. (2013a). Our goal is to assess the skill of ACCESS1.3b in simulating extremes and to identify systematic biases, strengths and weaknesses. This provides us with the foundation for future experiments aimed at resolving deficiencies, particularly where these relate to land surface processes. This study is organized as follows: Section 2 gives an overview of the model used and the datasets we used for evaluation. Section 3 provides our results. Section 4 provides a discussion and finally Section 5 concludes our study.
2 Data and Method

2.1 Australian Community Climate and Earth System Simulator (ACCESS1.3b)

The Australian Community Climate and Earth System Simulator (ACCESS) has been developed at the Centre for Australian Weather and Climate Research (CAWCR) to provide the Australian climate community with a state-of-the-art fully coupled climate model as well as weather prediction model (Puri et al., 2013). ACCESS1.3b consists of the atmospheric Unified Model (UM7.3, UK Meteorological Office), the Community Atmosphere Biosphere Land Exchange land surface model (CABLE2.0, CSIRO), the Modular Ocean Model (NOAA/GFDL), the Los Alamos sea ice model CICE (LANL) and the coupling framework OASIS (CERFACS) which couples the ocean and sea ice to the atmosphere (Bi et al., 2013). We use ACCESS1.3b in an AMIP-style configuration with prescribed sea surface temperatures and sea ice fractions. These were sourced from the Program for Climate Model Diagnosis and Comparison (Taylor et al. (2000), http://www-pcmdi.llnl.gov/projects/amip/AMIP2EXPDSN/BCS/amipbc_dwnld.php) and regridded and converted to the UM’s data format at the UK Meteorological Office. We performed simulations at 1.25° latitude × 1.875° longitude resolution (N96 resolution), 38 vertical levels, and a 30 minute time step. The simulation covers the period 1950–2012, the first year is used as spin-up period and not included in the analysis. Orography in ACCESS1.3b are derived from the 30” GLOBE dataset (GLOBE Task Team and others, 1999). However, since this dataset has deficiencies over Australia (Bi et al., 2013) it is improved for the Australian region using the Geoscience Australia high quality dataset (Hilton et al., 2003).

2.1.1 The atmosphere: Unified Model (HadGEM3)

The atmospheric model in ACCESS1.3b is the Unified Model developed at the UK Meteorological Office (Davies et al., 2005; Martin et al., 2006). Atmospheric dynamics in the UM are non-hydrostatic, fully compressible and the advection scheme is semi-lagrangian. The vertical coordinates are height based and follow the terrain and a regular Arakawa C grid is used in the horizontal. The radiation scheme is a general two-stream scheme developed by Edwards and
Slingo (1996) but was improved in terms of pressure and temperature scaling. In addition, the tripleclouds scheme of Shonk and Hogan (2008) was included to improve the representation of horizontal cloud inhomogeneity. Calculation of the radiation scheme is performed eight times per day (3-hourly). Convection is parameterized by a modified mass flux scheme based on Gregory and Rowntree (1990). The boundary layer mixing scheme represents nonlocal mixing in unstable layers and an explicit entrainment parameterization (Lock et al., 2000). The cloud microphysics scheme contains water vapour, total cloud fraction, cloud liquid water and cloud ice as prognostic variables and we use the prognostic condensate scheme PC2 described in Wilson et al. (2008). Atmospheric chemistry includes the aerosols sulphate, soot, biomass, dust (from IGBP soils, although values are very low), sea salt, and biogenic (climatology only) aerosols. Aerosol emissions are prescribed by monthly climatologies, and aerosols can be advected and deposited. More details can be found in Hewitt et al. (2011) and Bi et al. (2013).

2.1.2 The land surface: Community Atmosphere Biosphere Land Exchange (CABLE2.0)

Land surface models simulate biogeophysical and biogeochemical processes and handle the exchange of surface fluxes between the land surface and the atmosphere. Since the extremes explored in this paper are intimately associated with how the land surface is parameterized we provide some detail on how CABLE represents terrestrial processes. Further detailed descriptions of CABLE1.4 can be found in Wang et al. (2011) and CABLE1.8 in Kowalczyk et al. (2013).

CABLE consists of three submodels (1) canopy processes, (2) soil and snow, and (3) carbon pool dynamics and soil respiration. Canopy processes are simulated by a one layer two-leaf canopy scheme, distinguishing between sunlit and shaded leaves for the calculation of photosynthesis, stomatal conductance and leaf temperature (Wang and Leuning, 1998). The vegetation is placed above the ground, which allows for aerodynamic and radiative interactions between the ground and the canopy. CABLE includes a sub-grid tiling approach at the surface, meaning that several surface types can exist within a grid-cell (ten vegetation types and three non-vegetated types are distinguished, up to five tiles can be used within each grid cell). The soil model has six layers and the Richards equation is solved for soil moisture while soil temperature is calculated
from the heat conduction equation. The snow model has three snowpack layers and calculates
temperature, density and thickness of the snow. The carbon pool model used is simple, net pri-
mary productivity is calculated from the annual carbon assimilation corrected for respiratory
losses (carbon fluxes are not assessed in this study). The differences between CABLE1.8 used
in Kowalczyk et al. (2013) and CABLE2.0 used here are small. They include bug fixes and
updated optical leaf properties (transmission and reflectance) that are better calibrated for the
snow-free soil albedo used by ACCESS.

CABLE has been extensively evaluated (Abramowitz et al., 2008; Wang et al., 2011) and
an earlier version was used in the Land Use Change IDentification of robust impacts (LUCID)
project (Pitman et al., 2009; de Noblet-Ducoudré et al., 2012). Further, Mao et al. (2011) doc-
uments the performance of a low-resolution GCM of intermediate complexity, CSIRO Mk3L,
coupled to an earlier version of CABLE (version 1.4b) with a focus on terrestrial quantities. This
analysis provides strong evidence that the coupled model produces a reasonable large-scale cli-
matology. More recently, Zhang et al. (2013) ran CABLE2.0 offline with GSWP2 (Global Soil
Wetness Project) forcing and compared it with other participating land surface models in GSWP
and gridded observations. They found that whilst global mean evapotranspiration (\(ET\)) simu-
lated by CABLE agreed well with other land surface models and observations, CABLE under-
estimated \(ET\) in the tropics and had significant runoff errors. In addition, CABLE showed a
large sensitivity to soil and vegetation parameters in tropical rainforests and mid-latitude forest
regions.

2.2 ETCCDI Indices and Datasets

The Expert Team on Climate Change Detection and Indices (ETCCDI) defined a set of 27 in-
dices calculated from daily maximum (\(T_{MAX}\)) and minimum temperature (\(T_{MIN}\)) and daily
precipitation (http://www.climdex.org/indices.html). These indices were developed to investi-
gate changes in intensity, duration and frequency of extreme climate events. Most of these
indices describe moderate extremes with return periods of a year or shorter. We calculate all in-
dices using freely available software (http://www.climdex.org/climdex_software.html) and com-
pare the indices from our simulations to the HadEX2 dataset (Donat et al., 2013b). Only a
subset of indices are analysed in detail (see Table 1). We chose four indices that examine the frequency of high (warm days TX90p, warm nights TN90p) and low (cool days TX10p, cool nights TN10p) temperature extremes and one temperature index investigates the amplitude between the coldest and hottest temperature per day (diurnal temperature range DTR). Two of the chosen indices examine wet precipitation extremes (maximum 1-day precipitation amount Rx1day, consecutive wet days CWD) and one index looks at dryness (consecutive dry days CDD). For temperature extremes, we chose mainly indices based on percentiles which are relative to the base period 1961–1990 because they are applicable over all climate zones and show robust trends in observational datasets (Zhang et al., 2011; Donat et al., 2013b). The precipitation indices were chosen to examine several aspects of precipitation extremes, high precipitation amounts in Rx1day, high precipitation frequency in CWD and low precipitation frequency and drought in CDD.

2.2.1 HadEX2 dataset

The HadEX2 dataset, described in detail by Donat et al. (2013b), contains 17 temperature and 12 precipitation indices. These are derived from daily maximum and minimum temperature and precipitation observations for the period covering 1901 to 2010. The indices were calculated for each station and then the monthly and annual indices were interpolated onto a 2.5° latitude × 3.75° longitude grid. Donat et al. (2013b) derived linear trends from the gridded fields and tested these trends for statistical significance. The high quality in-situ observations were primarily sourced from the European Climate Assessment and Data set (ECA&D) and associated datasets in south-east Asia and Latin America, GHCN-Daily (USA-only), ETCCDI regional workshops and individual researchers. As a result, the spatial availability of HadEX2 data varies with time. Trend estimates can be influenced by the number of stations included in a dataset. To compare time series and trends of model and observations we calculated global averages. We apply a time independent masking of the model data and HadEX2, only including grid points where more than 50 years of observational data (out of 60) are available. This minimises the deteriorating effect of variable spatial coverage on the trend calculations. The spatial coverage of the Rx1day index is larger in monthly than annual fields because the decorrelation length scale is larger for
monthly compared to yearly extreme precipitation indices. To obtain a spatial coverage that is as good as possible, we calculate the annual maximum Rx1day amounts from the maximum of the monthly Rx1day fields provided by HadEX2. This increases data availability in data sparse regions (e.g. tropics), however, it needs to be taken into account that this may include stations which are less representative for a certain grid point. Therefore, grid points around areas with missing data need to be interpreted with care.

When comparing models and gridded observational datasets for extremes, it needs to be kept in mind that scaling effects likely play a role. That is, the gridded observational dataset is derived from annual extremes at each station whereas the models represent a grid point average for each day. Therefore, annual maxima from climate models are expected to be lower in intensity, especially for precipitation based indices (Kiktev et al., 2003; Tebaldi et al., 2006).

2.2.2 Other Datasets

We use the HadGHCND gridded daily temperature dataset (Caesar et al., 2006) derived from near-surface maximum and minimum temperature observations. It covers the period from 1951 to the present on a 2.75° latitude × 3.75° longitude grid. It was designed for the analysis of climate extremes and the evaluation of climate models. Note that the data coverage is varying with time.

We also use the Global Precipitation Climatology Project (GPCP) Version-2 precipitation (http://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html) dataset. This is derived from a combination of satellite and rain-gauge measurements (Adler et al., 2003). GPCP is available as global, monthly analysis of surface precipitation at 2.5° × 2.5° resolution from 1979 to the present (we use 1979/12–2012/11 here). GPCP has been shown to agree well with ground-based observations (Ma et al., 2009; Pfeifroth et al., 2013).

The NASA "Clouds and the Earth’s Radiant Energy System” (CERES EBAF Surface Ed2.7) dataset provides satellite-based estimates of surface radiative fluxes. This data set was specifically created for evaluation of climate models (http://ceres-tool.larc.nasa.gov). It includes surface downwelling shortwave and longwave radiation, surface upwelling shortwave and longwave radiation and estimates for clear-sky radiation from 2001 to 2009. Kato et al. (2013)
found that the biases over land were, on average, 21.7 W m$^{-2}$ for downward shortwave and 21.0 W m$^{-2}$ for downward longwave radiation. Therefore, biases between ±10 W m$^{-2}$ are not taken into account in our analysis.

For evapotranspiration, we use the GLEAM (Global Land-surface Evaporation: the Amsterdam Methodology, Miralles et al. (2011)) global evapotranspiration dataset and the LandFlux-EVAL dataset (Mueller et al., 2013). GLEAM is derived from various satellite products within a Priestley-Taylor framework (Priestley and Taylor, 1972). It estimates daily evaporation at a global scale and a 0.25° spatial resolution and is available from 1984 to 2007. GLEAM uses microwave-derived soil moisture, land surface temperature and vegetation density, as well as the detailed estimation of rainfall interception loss. GLEAM has been found to have a low average bias (<5%, Miralles et al. (2011)). LandFlux-EVAL is a merged multi-data benchmark product. It contains data from different sources including satellites, in situ observations, land surface models and reanalysis. We use the longer period data product (1989–2005) where 14 different datasets, including GLEAM, were used.

Given the sparse coverage and limited availability of flux observations, satellite estimates provide the “next-best approximation”. Although these are strictly models, and not true observations, the algorithms are usually constrained by as much data as possible (e.g., the GLEAM ET product is driven with gridded precipitation observations), and hence, these products have a well-defined accuracy and are, therefore, useful to compare against global climate models (GCMs), which have much larger degrees of freedom. For the calculation of the biases we used the coarsest grid involved, either interpolating the model output to the coarser grid of the observational dataset or interpolating the observations to the model resolution. Table 2 summarizes the datasets used for evaluation.

2.3 Statistical Methods

2.3.1 Statistical significance testing

We perform a modified t-test, as described in Zwiers and von Storch (1995), to indicate which biases between the model run and observations are statistically significant for $T_{MAX}$, $T_{MIN}$
and total precipitation ($P_{TOT}$). This modified t-test accounts for autocorrelation within the time-series. Total precipitation, $T_{MAX}$ and $T_{MIN}$ are robust observations for long timeseries and are the underlying data of the extreme indices.

### 2.3.2 Probability density functions and skill score

We calculate probability density functions (PDFs) for $T_{MIN}$ and $T_{MAX}$ from the model and observations to investigate which part of the distributions are most important for the biases. The PDFs are based on $T_{MAX}$(time,lat,lon), $T_{MIN}$(time,lat,lon) containing monthly means for the timeperiod 1951–2011 for the corresponding season and region. We use R’s kernel density function, using the default Gaussian smoothing kernel and a bandwidth estimated via Normal Reference Distribution to plot the lines.

We use a skill score defined in Perkins et al. (2007) which measures the overlap between two PDFs by summing up the cumulative minimum probability at each bin. A perfect skill equals one whereas values close to zero indicate a poor agreement. We use a bin width of 0.5 °C for the calculation of the histograms as in Perkins et al. (2007).

### 3 Results

First we present seasonal averages of $T_{MAX}$, $T_{MIN}$ and $P_{TOT}$. Daily $T_{MAX}$, $T_{MIN}$ and total precipitation data form the basis for the calculation of the ETCCDI indices. The seasonal averages are calculated over December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). Then we present biases in several (annual) ETCCDI indices before investigating the causes of the differences between model and observations.

#### 3.1 Minimum and maximum temperature and total precipitation

We calculate seasonal averages from daily $T_{MIN}$ and $T_{MAX}$ for the period 1951–2011 from ACCESS1.3b (Figures 1a and 2a) and compare them to gridded observations from HadGHCND
(Figures 1b and 2b). The overall seasonal patterns are reproduced reasonably well by ACCESS. $T_{MAX}$ shows a negative bias in most regions except North America and parts of South-East Europe and Africa in JJA (Fig. 1c) whereas $T_{MIN}$ shows a positive bias almost globally (Fig. 2 c), except for the Arctic and Himalayas. Since HadGHCND does not have a complete coverage in all grid boxes over the whole time period we analysed, regional biases can be influenced by temperature trends, e.g. in East Africa where there is only data between $\sim$1960–1990 (Caesar et al., 2006). The opposing $T_{MAX}$ and $T_{MIN}$ biases commonly lead to a good simulation of the mean temperature (Kowalczyk et al., 2013, Fig. 10). Only North America and South-East Europe show a positive bias in boreal summer (JJA) in $T_{MAX}$ as well as $T_{MIN}$, which results in a pronounced positive bias in JJA mean temperature especially in North America (not shown). Positive temperature biases in central Eurasia and North America were also found by Kowalczyk et al. (2013) who linked them to underestimation of precipitation. The North American bias in JJA has been previously associated with an underestimation of clouds in the area by up to 30% (Franklin et al., 2013a). Overall, however, Figure 1 suggests a cold bias in $T_{MAX}$ in most regions and in most seasons commonly reaching 4°C and regionally exceeding 7°C. The key exception to this is in JJA where there is a warm bias in the northern hemisphere mid-latitudes of $\sim$2°C, exceeding 5°C over North America. Fig. 2, in contrast, suggests a warm bias of $\sim$2°C in $T_{MIN}$ almost everywhere, exceeding 5°C over North Asia in DJF, and North America in JJA. Global patterns of $P_{TOT}$ are well represented in ACCESS1.3b compared to GPCP during the time period 1980–2012 (Figures 3a and 3b). ACCESS1.3b tends to overestimate total seasonal precipitation (Fig. 3 c) in most regions, although there is a small underestimate over Europe in most seasons. The wet precipitation bias is largest in the tropics (exceeding 5 mm d$^{-1}$) but elsewhere it is generally small ($<1$ mm d$^{-1}$). A large negative precipitation bias exists in India in JJA and SON where the monsoon is displaced. This bias has previously been reported by Kowalczyk et al. (2013) and Bi et al. (2013). These biases are statistically significant, indicated by stippling, in most regions.
3.2 ETCCDI Indices

Results for ETCCDI indices in ACCESS1.3b are compared to the HadEX2 dataset. Since the indices are calculated for station data in HadEX2 and then gridded, one would expect that model output might appear smoother with less extremes than the observational dataset (Donat et al., 2013b). In particular, one would expect precipitation based extremes estimates calculated from station-based observations to be more intense. However, we did not find a general underestimation of the variability in the extreme indices in the model.

The percentile-based indices are expected not to differ much from HadEX2, since the percentiles are calculated from the model data and are 10% on average during the base period (1961–1990) per definition. Hence, differences between ACCESS and HadEX2 are mainly driven by different trends and do not depend on biases in absolute values of $T_{MAX}$ and $T_{MIN}$. The two indices that examine cold extremes, cool nights (TN10p) and cool days (TX10p) are shown in Fig. 4. The ACCESS1.3b model represents the global patterns of both TN10p (Fig. 4 e) and TX10p (Fig. 4 f) reasonably well. ACCESS1.3b also captures the decreasing trends in TN10p (Fig. 4 g) and TX10p (Fig. 4 h) over the period 1951 to 2010. ACCESS underestimates the total trend in TN10p because values are underestimated by $\sim 2\%$ at the beginning of the simulation and the model data show an increasing trend until the 1980's. After $\sim 1980$, the trends are captured well such that the model simulates a close value to the observations of $\sim 6\%$ of days by 2010. The trend in TX10p is similar to TN10p, but differences are smaller between ACCESS and HadEX2. Both show a decrease from $\sim 11\%$ to $7\%$ of days. The global patterns in the two indices representing hot extremes (warm nights, TN90p and warm days, TX90p) are also captured well by ACCESS1.3b (Fig. 5). The regional differences for hot extremes are larger than for cold extremes. There is a large overestimate in the occurrence of TN90p in the southern hemisphere, particularly over South America but this difference also affects North America, Australia and southern Africa (Fig. 5e). Similar regions are affected by an overestimation of TX90p (Fig. 5 f). Despite these regional differences, ACCESS1.3b estimates the global increasing trends in both TN90p (Fig. 5 g) and TX90p (Fig. 5h) remarkably well. We note that this might be because data availability in some of the regions with large differences is
too low to pass the requirement of 50+ years of data in HadEX2 to be included in the global average. Also note the close agreement in interannual variability between ACCESS and HadEX2, suggesting a strong influence from sea surface temperatures, which are prescribed here, on hot extremes. The final temperature index is the diurnal temperature range (DTR, Fig. 6). This is simulated poorly by ACCESS1.3b and is globally underestimated by up to 4°C (Fig. 6c). This result is anticipated given the seasonal overestimation of $T_{MIN}$ and underestimation of $T_{MAX}$. The underestimation of DTR is shown clearly in the global time series (Fig. 6d). We put this large underestimation into context with CMIP-5 simulations and reanalysis in the discussion in Section 4. There is no clear trend in DTR in the model and the trend in HadEX2 is very hard to see because of the scale of the figure. However, DTR decreases from 11.2°C to 11°C in HadEX2 and was shown to have a significant decreasing trend by Donat et al. (2013b) of $\sim$0.05°C per decade.

The annual maximum consecutive 1-day precipitation (Rx1day) shows both, regions of over and underestimation (Fig. 7). The pronounced underestimation over India is clearly related to the missing monsoon (Fig. 7c). The smaller underestimation in North America is related to the underestimation of summer rainfall. Central Eurasia also shows an underestimation of Rx1day due to the underestimation in total precipitation during summer, autumn and winter. Overall, ACCESS1.3b underestimates Rx1day (Fig. 7d) by $\sim$2–5mm ($\sim$5–10%). However, ACCESS1.3b clearly captures some elements of Rx1day and some of the variability between 1951–2010. We also analysed Rx5day (maximum annual consecutive 5-day precipitation) that showed an overestimation from ACCESS on global average (not shown). However, this is an artefact of how we calculate these indices in HadEX2, as maxima out of the monthly maxima, which have a better coverage than the annual maximum (see Section 2.3) and the lack of observational data in the tropics. This problem is less pronounced for Rx1day, however, biases around the areas with missing values in HadEX2 have to be taken with care. Consecutive wet days (CWD) are clearly overestimated over the Northern Hemisphere (which is where CWD can be derived due to the low coverage in the southern hemisphere), while consecutive dry days (CDD) are underestimated (Fig. 8). There is no clear overall trend in the time series of CWD and CDD (Figures 8g, 8h). Overall, there is a clear picture of ACCESS1.3b heavily overesti-
mating consecutive wet days, and underestimating consecutive dry days in those regions where
the observations are complete enough to derive these indices.

The biases in extreme precipitation indices are largely influenced by the bias in total precip-
itation in ACCESS1.3b. This is not a surprise; climate models commonly rain too often, but as
low intensity precipitation ("drizzle problem", e.g., Dai, 2006). On a global scale, ACCESS1.3b
has too many consecutive wet days, so it rains too often, and underestimates consecutive 1-day
precipitation. The biggest bias identified is the underestimation of the diurnal temperature range,
due to an overestimation of $T_{MIN}$ and an underestimation of $T_{MAX}$. Therefore, the next section
focuses on the distributions of $T_{MIN}$ and $T_{MAX}$.

### 3.3 Probability density functions of $T_{MAX}$ and $T_{MIN}$

The probability density functions (PDFs) of $T_{MAX}$ and $T_{MIN}$ for ACCESS1.3b and the Had-
GHCND dataset are shown in Fig. 9 (DJF) and Fig. 10 (JJA). We restrict our analysis of the
PDFs to four regions with good data coverage in HadGHCND. These regions are defined in
Table 2 and correspond to Asia, Australia, Europe and North America. The results from the
PDFs are summarized in each panel using the skill score defined by Perkins et al. (2007) which
measures the overlap of the PDFs (perfect agreement is a skill score of 1.0).

For DJF, the three northern hemispheric regions (Asia, Europe, and North America) repro-
duce the PDF’s of the observational data set well. In Asia (Fig. 9a), the lower tail of the $T_{MAX}$
distribution is almost perfectly captured. The upper tail is also captured well although there is a
small deviation between 10–20 °C. In $T_{MIN}$ the upper tail is well reproduced by ACCESS1.3b
but the lower tail shows a bias of $\sim 5$°C with too frequent $T_{MIN}$ simulated around $-20$°C.
In North America (Figure 9d), the biases are the opposite; the upper tail of $T_{MAX}$ is better
captured than the lower tail, while the lower tail of the $T_{MIN}$ distribution is reproduced well.
For Europe (Fig. 9c), both $T_{MAX}$ and $T_{MIN}$ distributions only show small deviations from the
observations. For Australia (Figure 9b), the upper tail of $T_{MAX}$ is reasonably captured, but the
mean of $T_{MAX}$ is underestimated and the lower tail shows a bias of $\sim 5$°C. The lower tail of
$T_{MIN}$ in Australia is better captured than the upper tail, but the whole PDF of $T_{MIN}$ is shifted
to the right in the model. Overall the PDF for both $T_{MAX}$ and $T_{MIN}$ are simulated with a skill
score exceeding 0.8 for all regions except Australia in DJF. There is a clear problem with the PDF for $T_{MAX}$ in Australia in DJF linked to a large bias associated with the lower tail of the distribution.

In JJA (Fig. 10), Australia (Fig. 10b) reproduces the distributions of $T_{MIN}$ and $T_{MAX}$ better than the Northern Hemispheric regions Asia and North America. The lower tail of $T_{MIN}$ is almost perfect but the upper tail has a bias of $\sim 5^\circ C$. For $T_{MAX}$ the upper tail is reproduced well, but the lower tail is shifted to the left in the model by about $3^\circ C$. Overall, however, ACCESS1.3b captures the $T_{MIN}$ and $T_{MAX}$ for Australia in JJA with a skill score exceeding 0.8. The PDFs for the northern hemisphere region are less well captured than in DJF (Fig. 10, a, c, d) with half the skill scores below 0.8 for these regions. Europe (Fig. 10c) captures the lower tail for $T_{MIN}$ well, but the upper tail is slightly overestimated. For $T_{MAX}$, the PDF is shifted to the left in the model by $\sim 3^\circ C$ and the mean of the distribution is underestimated. However, for Europe the skill scores in JJA are still larger than 0.8. $T_{MAX}$ in Asia (Fig. 10a) shows a similar picture although the biases are larger than in Europe. $T_{MIN}$ is shifted to the right, especially the main peak that is also underestimated leading to a low overall skill score. In North America (Fig. 10d), only the lower tail of $T_{MIN}$ is reasonably reproduced by the model, the main peak is underestimated and the upper tail shifted to the right by $\sim 5^\circ C$. The lower tail for $T_{MAX}$ in North America in JJA is too low and the upper tail too high in ACCESS1.3b. Generally, the lower tail of $T_{MIN}$ is reproduced better than the upper tail, whereas the upper tail in $T_{MAX}$ is often reproduced better than the lower tail.

4 Discussion

The driver of temperatures at the Earth’s surface is the surface radiation balance, but different components of the radiation balance are associated with $T_{MIN}$ and $T_{MAX}$. The daily minimum temperature, which normally occurs just before sunrise, is mainly determined by longwave radiation ($LW$) because there is no shortwave ($SW$) radiation during the night. The magnitude of incoming longwave ($LW_{IN}$) depends on sky temperature and emissivity and is affected by cloud cover and humidity while outgoing $LW$ depends on the emissivity and temperature of
the earth’s surface. Maximum temperatures during the day are dependent on the incoming solar radiation ($SW$) and modulated by cloud cover and aerosols. Surface temperatures are also affected by the surface albedo, availability of soil moisture for evapotranspiration and stability conditions of the atmosphere. We examined the biases in net longwave ($LW_{NET}$) and net shortwave ($SW_{NET}$) from ACCESS1.3b to explain the biases in temperature. The satellite product CERES is used to estimate the biases in the radiative fluxes which has a well defined level of accuracy. When compared to CERES, ACCESS generally has excess amount of $SW$ absorbed at the surface (Fig. 11a) in all seasons. In the northern hemisphere this is small in DJF and largest in JJA where the bias exceeds 50 W m$^{-2}$ over Europe and North America. There are other regions with biases exceeding 50 W m$^{-2}$ including central Africa, India and the Amazon delta (Fig. 11a). The high bias in $SW_{NET}$ (Fig. 11a) is likely associated with a low cloud bias enabling excessive incoming $SW$. This is evident in JJA over India where the Indian monsoon is severely under-predicted, see Fig 3. The bias in $LW_{NET}$ is generally negative (Fig. 11b), especially in the arid and semi-arid areas, pointing to either outgoing $LW$ being overestimated or incoming $LW$ being underestimated. Outgoing $LW$ radiation is directly proportional to the surface temperature to the 4th power. In areas with positive biases in $T_{MIN}$ and $T_{MAX}$, ACCESS overestimates outgoing $LW$ which could explain the negative bias in $LW_{NET}$ radiation in central Eurasia and North America in JJA. Overall, the largest errors in $SW_{NET}$ are in JJA in the northern hemisphere and India as well as the Amazon delta in SON. While the largest biases in $LW_{NET}$ occur in warm seasons in the arid and semi–arid areas of North Africa, central Eurasia, middle east, India, North America and Australia. The biases in $LW$ and $SW$ lead to an overall overestimation of total net radiation in northern hemisphere spring and summer and most of the tropics (not shown).

We calculate temporal correlations between the biases in $T_{MIN}$, $T_{MAX}$ and radiation per season on each longitude,latitude (using the NCAR Command Language ”NCL” function “esccor”) when biases are larger than $\pm 1 \degree C$ or $\pm 10$ W m$^{-2}$ respectively. The bias in $T_{MIN}$ correlates strongly with the bias in incoming $LW$ (Fig. 12a). This provides further evidence to associate this temperature bias with problems in the ACCESS1.3b simulated cloud cover. Franklin et al. (2013b) evaluated cloud fraction in ACCESS1.3 in detail. They found that clouds are
represented reasonably but found differences in the horizontal distribution. These include a tendency for too little clouds throughout the subtropics and trade wind regions, an underestimation of up to 25% in DJF over Russia as well as an underestimation by 30% in JJA over North America. The bias in $T_{MAX}$ correlates with the bias in $SW_{NET}$, but the correlation is weaker than for $T_{MIN}$ and $LW_{IN}$. For example, regions with large negative biases in $T_{MAX}$ (Figure 1c) in the Himalayas, Arctic, and southwest South America, which are persistent in all seasons, do not always correspond to a negative bias in $SW_{NET}$ (Fig. 11a). In the northern hemisphere summer, correlations between $SW_{NET}$ biases and $T_{MAX}$ biases are strong (Fig. 12b) and usually exceed $\sim 0.8$. However, in SON and MAM, and in particular in DJF in the northern hemisphere the correlation between $SW_{NET}$ biases and $T_{MAX}$ becomes weaker and even negative in some grid points (Fig. 12b). The weaker correlations between $SW_{NET}$ and biases in $T_{MAX}$ (Fig. 12b) points to other factors than atmospheric processes playing a role and these are likely to be linked to land processes.

ACCESS1.3b is generally lacking in its capacity to capture $T_{MAX}$. This was apparent in Figure 1c for $T_{MAX}$ and Fig. 6 for DTR. Reflecting on Figures 9 and 10, the simulation of $T_{MAX}$ was shifted to the left in the ACCESS1.3b model in both DJF and JJA in all four regions. It is noteworthy that the largest biases tended to be at the lower tail of the PDF for $T_{MAX}$ (only in Europe and Asia in DJF this was not true). The most straightforward explanation for this is linked with evapotranspiration. For instance, Watterson (1997) found close spatial correlations between DTR and $SW_{NET}$ minus the evaporative and sensible fluxes, or $LW_{NET}$. Examining how well a land surface model simulates evapotranspiration is challenging because a bias in this quantity can result from poor forcing (rainfall, $SW$ and $LW$), poor surface states (soil moisture) or poor parameterization of the relationship between the states and the fluxes. It is also challenging because there are considerable uncertainties in estimates of evapotranspiration from observations-based products. We use GLEAM (see Section 2.5) recognising that this product is a model-based estimate of evapotranspiration and that there are likely significant uncertainties associated with the estimates. To decrease these uncertainties we use a second $ET$ product, the multi-data synthesis LandFluxEVAL.
Fig. 13a shows the simulation of evapotranspiration in ACCESS1.3b compared with GLEAM, Fig. 13b shows the same for LandFluxEVAL. There is a systematic bias in simulated evapotranspiration, commonly reaching $30 \text{ W m}^{-2}$ and regionally exceeding $50 \text{ W m}^{-2}$. In almost all cases, ACCESS1.3b simulates excess evapotranspiration. This is in contrast to Zhang et al. (2013) who found an underestimation of $ET$ in the tropics in offline CABLE2.0 runs. There are, however, some important exceptions; there is too little $ET$ over the Indian subcontinent in JJA and SON linked with the failure of the monsoon in this model. There is also a lack of evapotranspiration over parts of North America, despite the excess $SW$, in JJA. However, the pattern of excess evapotranspiration shown in Fig. 13a and b is large-scale and systematic. The patterns of the evapotranspiration biases are dissimilar to the $LW_{NET}$ biases (Figure 11b) and are weakly linked to the biases in $SW_{NET}$ (Fig. 11a). The largest positive $ET$ biases occur in densely forested areas (e.g. tropics) and in the northern hemisphere in summer. As shown in Figure 1c, most of the biases in $T_{MAX}$ are small, or negative except over the mid-latitudes of the northern hemisphere in JJA which are closely linked to the negative rainfall bias in the model (Fig. 3 c). This general low bias in $T_{MAX}$ could be explained by the excessive evapotranspiration, which has been found in other climate models as well (Mueller and Seneviratne, 2014). Fig. 13c shows the temporal correlation (calculated as for temperature and radiation using NCL’s “escorc”) between biases in $T_{MAX}$ and biases in evapotranspiration (or latent heat flux, $LH$, in W m$^{-2}$). Small biases ($< \pm 1 ^\circ \text{C}$ for $T_{MAX}$ and $< \pm 10 \text{ W m}^{-2}$ for $LH$) are masked to focus on the correlation of significant biases. We expect a negative correlation in areas where either $ET$ is too low and $T_{MAX}$ is too high or $ET$ is too high and $T_{MAX}$ is too low. There are many regions where the biases in $LH$ and $T_{MAX}$ are negatively correlated. These include regions over the mid-latitudes of the northern hemisphere in JJA, Eurasia in SON and the southern hemisphere in DJF and JJA. There are also large areas where the correlation is positive including parts of the mid-latitudes of the northern hemisphere in SON, high latitudes in MAM and south-east Asia in MAM, JJA and SON. Unfortunately, using $LH$ to explain biases in $T_{MAX}$ is limited by major gaps in $T_{MAX}$ observations. Despite this, the regions where the clearest negative correlations are found, and the seasons they occur within are not unexpected. Areas with negative correlations correspond to areas where $ET$ is limited by soil moisture availability, areas where
the correlation between $T_{MAX}$ and $ET$ biases is positive relate to areas where $ET$ is limited by radiation/temperature (see Seneviratne et al., 2010; Jung et al., 2010; Wang and Dickinson, 2012). In regions where $ET$ is limited by soil moisture, a high influence from the land surface on temperature is expected due to strong land-atmosphere coupling (e.g., Seneviratne et al., 2010; Mueller and Seneviratne, 2012). These tend to be transitional regions between wet and dry climates during the summer season in both hemispheres. This link between biases and coupling is an area we will pursue in the future.

One question might be how the ACCESS1.3b simulation of the ETCCDI indices compares with other models. Our use of AMIP makes a direct comparison with other models infeasible. However, Sillmann et al. (2013) have provided an evaluation of climate extreme indices from CMIP-5 models for the present climate. In addition to HadEX2, they included four reanalysis datasets in their analysis. Some of the reanalysis also show large biases to the observations, partly due to different computational approaches when calculating indices from daily grid-point averages in comparison to grids of station extremes (Donat et al., 2014). Therefore, biases between reanalyses/model output and observations are expected to some degree because of scaling effects. Sillmann et al. (2013) concluded that CMIP-5 models are generally able to simulate climate extremes and their trend patterns in comparison to HadEX2. The percentile indices TN10p, TX10, TN90p and TX90p compare very well with CMIP-5 since they are calculated relative to their specific PDF, thus insensitive to biases in absolute temperature values. Sillmann et al. (2013) also found that models and reanalyses disagree with HadEX2 for DTR. HadEX2 shows much larger values for DTR than the median of the analysed CMIP-5 models and most reanalyses. The question might arise if the comparison of the models to HadEX2 is fair for DTR. As mentioned in section 2.5, extreme indices derived from model output are expected to be less intense than those derived from station observations. The spatial-scale mismatch between model and HadEX2 probably explains a small part of the bias. However, the spatial mismatch between models and observations plays less a role for indices based on monthly averages such as DTR. HadGHCND $T_{MAX}$ and $T_{MIN}$ seasonal averages also suggest an underestimation of the DTR. In addition, Lewis and Karoly (2013) also found deficiencies in the CMIP-5 models in simulating trends in DTR. Hence, the underestimation of DTR is a common problem in many climate
models although it remains possible that the model derived DTR is not directly comparable with
the observed derived value. Rx1day was not considered in Sillmann et al. (2013). For Rx5day,
ACCESS1.3b’s global mean is higher than the median of the CMIP-5 ensemble investigated
by Sillmann et al. (2013), CWD is at the lower end of the CMIP-5 models and CDD is also
lower than the CMIP-5 median. ACCESS1.0 was among the models that reproduces most tem-
perature and precipitation indices reasonably well in Sillmann et al. (2013). Therefore, overall,
ACCESS1.3b performs comparably to other CMIP-5 models for ETCCDI with some indices
simulated particularly well, and others in a more limited way.

5 Conclusions

To provide a benchmark for how well the ACCESS1.3b climate model simulates extremes, we
undertook an AMIP-style simulation involving simulations over the period 1950–2012 with pre-
scribed sea surface temperatures and sea ice concentration. Our goal was to identify strengths
and weaknesses in the ACCESS1.3b modelling system to provide a basis for experiments and
model developments to resolve these weaknesses. Our analysis is founded on the capacity of
the model to simulate daily $T_{MAX}$, $T_{MIN}$ and precipitation. From these three variables we
calculated climate extremes derived by the Expert Team on Climate Change Detection and In-
dices (ETCCDI). This work builds on earlier analyses of the mean climate of the ACCESS1.3
model, that included CABLE1.8 rather than CABLE2.0 by Kowalczyk et al. (2013) and Bi et al.
(2013). These analyses showed that ACCESS1.3 captured the large-scale mean temperature and
precipitation well, and compared favourably with other climate models in CMIP-5. Our analy-
sis highlighted a large (2–6 °C) cold bias in the simulation of $T_{MAX}$ in all seasons and in all
regions except North America. We also showed a large positive bias (1–5 °C) in $T_{MIN}$ in all
seasons and in all regions. As a consequence, ACCESS1.3b fails to represent the diurnal tem-
perature range well in comparison with the HadEX2 data. However, the model captures patterns
in, and trends in, indices for cool nights (TN10p) and cold days (TX10p) extremely well, al-
though there is an overestimation in the change in both indices between ~1975 and 2010. Warm
nights (TN90p) and warm days (TX90p) are also captured well. ACCESS1.3b simulates rainfall
indices quite variably. Rainfall intensity (Rx1day) is simulated reasonably well but consecutive wet days are badly overestimated and consecutive dry days are badly underestimated in the model. The biases in temperature related indices are very likely associated with a large positive bias in net short wave radiation (Fig. 12a) and a large negative bias in net long wave radiation (Fig. 12b). Some of the precipitation biases are related to the common “drizzle” problem. Our results highlight challenges in simulating climate extremes in climate models, a result previously identified (Kiktev et al., 2003; Kharin et al., 2013; Sillmann et al., 2013). However, our results provide a benchmark from which we will now examine how land processes can be improved to capture these extremes. There are some clear ways forward for improving the model. Some of the biases are likely linked with a bias in simulating evapotranspiration and this will be a priority to resolve. For example, application of the GLACE methodology (Koster et al., 2006), could be used to quantify the degree of land-atmosphere coupling in ACCESS. Other biases might be linked with albedo, especially the correct parameterisation of snow albedo, which is a common challenge in land surface models. It will be more challenging to identify how to improve the cloud climatology but by identifying these biases, and the impact these have on extreme indices we provide a clear statement of the state of ACCESS1.3b and a benchmark from which the model can be improved.

Acknowledgements. We thank Dan Copsey from the UK Meteorological Office for producing ancillary files enabling our simulation of the AMIP period. We also thank Scott Wales for technical support and Lauren Stevens and Sophie Lewis for help with ACCESS. Dr Ian Watterson provided valuable comment on a draft of this paper. This work was supported by the Australian Research Council Centre of Excellence for Climate System Science grant CE110001028. The computational modeling was supported by the NCI National Facility at the ANU, Australia. The GPCP combined precipitation data were developed and computed by the NASA/Goddard Space Flight Center’s Laboratory for Atmospheres as a contribution to the GEWEX Global Precipitation Climatology Project. GPCP data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. NCL (2013) and R (2013) were used to draw the figures. We thank the referees, V. Brovkin and J. Mao for there valuable comments to improve this manuscript.
References


Fig. 1. Daily $T_{\text{MAX}}$ for ACCESS1.3b (a), HadGHCND (b) and bias (c). Unshaded areas in (b) and (c) reflect missing observation data. The considered time period is the overlapping time period between model run and observations, 1951–2011. Stippling in (c) indicates areas where bias is statistically significant based on the modified t-test.
Fig. 2. Daily $T_{MIN}$ for ACCESS1.3b (top row), HadGHCND (middle row) and bias (bottom row). Unshaded areas in (b) and (c) reflect missing observation data. The considered time period is the overlapping time period between model run and observations, 1951–2011. Stippling in (c) indicates areas where bias is statistically significant based on the modified t-test.
Fig. 3. Total precipitation for ACCESS1.3b (a), GPCP (b) and bias (c). The considered time period is the overlapping time period between model run and observations, 1980–2012. Stippling in (c) indicates areas where bias is statistically significant based on the modified t-test.
Fig. 4. Annual Cool nights (TN10p, left column) and cool days (TX10p, right column) for ACCESS1.3b (a, b), HadEX2 (c, d), bias (e, f) and the global average timeseries showing trends (g, h). Unshaded areas reflect missing observation data. ACCESS data were masked depending on the availability of HadEX2 before calculating the global average. We applied a static masked only including grid point were more than 50 years of data (out of 60) were available to calculate the global average. The solid lines in panel g and h show the timeseries smoothed by a 21 year gaussian filer while the dashed lines show the yearly values.
Fig. 5. Annual warm nights (TN90p, left column) and warm days (TX90p, right column) for ACCESS1.3b (a, b), HadEX2 (c, d), bias (e, f) and the global average timeseries showing trends (g, h). Unshaded areas reflect missing observation data. We applied a static masked only including grid point were more than 50 years of data (out of 60) were available to calculate the global average. The solid lines in panel g and h show the timeseries smoothed by a 21 year gaussian filer while the dashed lines show the yearly values.
Fig. 6. Annual diurnal temperature range (DTR) for ACCESS1.3b (a), HadEX2 (b), bias (c) and the global average timeseries showing trends (d). Unshaded areas reflect missing observation data. We applied a static masked only including grid point were more than 50 years of data (out of 60) were available to calculate the global average. The solid lines in panel g show the timeseries smoothed by a 21 year gaussian filter while the dashed lines show the yearly values.
Fig. 7. Annual maximum consecutive 1-day precipitation (Rx1day) for ACCESS1.3b (a), HadEX2 (b), bias (c) and the global average timeseries showing trends (d). Unshaded areas reflect missing observation data. We applied a static masked only including grid points were more than 50 years of data (out of 60) were available to calculate the global average. The solid lines in panel g show the timeseries smoothed by a 21 year gaussian filter while the dashed lines show the yearly values.
Fig. 8. Annual Consecutive wet days (CWD, left column) and consecutive dry days (CDD, right column) for ACCESS1.3b (a, b), HadEX2 (c, d), bias (e, f) and the global average timeseries showing trends (g, h). Unshaded areas reflect missing observation data. We applied a static masked only including grid point were more than 50 years of data (out of 60) were available to calculate the global average. The solid lines in panel g and h show the timeseries smoothed by a 21 year gaussian filter while the dashed lines show the yearly values.
Fig. 9. $T_{MAX}$ and $T_{MIN}$ probability density functions for ACCESS1.3b and HadGHCND over Asia, Australia, Europe and North America for DJF (1951–2011) and skill scores for PDF’s of $T_{MAX}$ and $T_{MIN}$. The skill score is calculated by summing up the lower probability at each bin of two given PDFs, $S_{score} = \sum_{1}^{n} minimum(Z_m, Z_o)$ where n is the number of bins used to calculate the PDFs and $Z_m$ and $Z_o$ correspond to the frequency of values in a bin from the model and the observations, respectively.
Fig. 10. $T_{MAX}$ and $T_{MIN}$ probability density functions for ACCESS1.3b and HadGHCND over Asia, Australia, Europe and North America for JJA (1951–2011) and skill scores for PDF’s of $T_{MAX}$ and $T_{MIN}$ (as in Fig. 9).
Fig. 11. Biases between ACCESS1.3b and CERES in net shortwave radiation (a) and net longwave radiation (b). The considered time period is the overlapping time period between model run and observations, 2001–2009.
Fig. 12. Correlations for biases between ACCESS1.3b and observations in $T_{MIN}$ and $LW_{IN}$ (a), $T_{MAX}$ and $SW_{NET}$ (b) in all seasons. Biases smaller than $\pm 1^\circ C$ and $\pm 10\,\text{W}\,\text{m}^{-2}$ were masked before the calculation of the correlations and are represented by white.
Fig. 13. Biases between ACCESS1.3b and GLEAM data (a) respectively LandFluxEVAL (b) for latent heat flux and correlations for biases between ACCESS1.3b and observations in $T_{MAX}$ and LH (c) in all seasons. The considered time period is the overlapping time period between model run and observations, 1984–2007 (a) resp. 1985–2005 (b,c). Biases smaller than $\pm 1^\circ$C and $\pm 10$ W m$^{-2}$ were masked before the calculation of the correlations and are represented by white.
Table 1. Extreme temperature and precipitation indices used in this study. First 4 indices measure frequency of temperature extremes, DTR measures the amplitude between coldest and hottest daily temperatures and last 3 indices measure rainfall extremes (two wet extremes, last one dry extremes). Percentiles are calculated over reference period 1961–1990.

<table>
<thead>
<tr>
<th>Index</th>
<th>Indicator Name</th>
<th>Definition</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX10p</td>
<td>Cool days</td>
<td>Percentage of time when daily max temperature $&lt; 10^{th}$ percentile</td>
<td>%</td>
</tr>
<tr>
<td>TN10p</td>
<td>Cool nights</td>
<td>Percentage of time when daily min temperature $&lt; 10^{th}$ percentile</td>
<td>%</td>
</tr>
<tr>
<td>TX90p</td>
<td>Warm days</td>
<td>Percentage of time when daily max temperature $&gt; 90^{th}$ percentile</td>
<td>%</td>
</tr>
<tr>
<td>TN90p</td>
<td>Warm nights</td>
<td>Percentage of time when daily min temperature $&gt; 90^{th}$ percentile</td>
<td>%</td>
</tr>
<tr>
<td>DTR</td>
<td>Diurnal temperature range</td>
<td>Annual mean difference between daily max and min temperature</td>
<td>°C</td>
</tr>
<tr>
<td>Rx1day</td>
<td>Max 1-day precipitation</td>
<td>Annual maximum consecutive 1-day precipitation amount</td>
<td>mm</td>
</tr>
<tr>
<td>Rx5day</td>
<td>Max 5-day precipitation</td>
<td>Annual maximum consecutive 5-day precipitation amount</td>
<td>mm</td>
</tr>
<tr>
<td>CWD</td>
<td>Consecutive wet days</td>
<td>Maximum number of consecutive days when precipitation $\geq 1$mm</td>
<td>days</td>
</tr>
<tr>
<td>CDD</td>
<td>Consecutive dry days</td>
<td>Maximum number of consecutive days when precipitation $&lt; 1$mm</td>
<td>days</td>
</tr>
</tbody>
</table>
Table 2. Datasets used for evaluation.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description/Name</th>
<th>Time</th>
<th>Resolution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HadEX2</td>
<td>dataset containing ETCCDI indices</td>
<td>1901–2010</td>
<td>$2.5^\circ \times 3.75^\circ$</td>
<td>Donat et al. (2013b)</td>
</tr>
<tr>
<td>HadGHCND</td>
<td>gridded daily temperature dataset</td>
<td>1951–now</td>
<td>$2.5^\circ \times 3.75^\circ$</td>
<td>Caesar et al. (2006)</td>
</tr>
<tr>
<td>GPCP</td>
<td>Global Precipitation Climatology Project Version-2</td>
<td>1979–now</td>
<td>$2.5^\circ \times 2.5^\circ$</td>
<td>Adler et al. (2003)</td>
</tr>
<tr>
<td>CERES</td>
<td>Clouds and the Earth’s Radiant Energy System</td>
<td>2001–2009</td>
<td>$1^\circ \times 1^\circ$</td>
<td>Kato et al. (2013)</td>
</tr>
<tr>
<td>GLEAM</td>
<td>Global Land-surface Evaporation: the Amsterdam Methodology</td>
<td>1984–2007</td>
<td>$0.25^\circ \times 0.25^\circ$</td>
<td>Miralles et al. (2011)</td>
</tr>
<tr>
<td>LandFlux-EVAL</td>
<td>Land evapotranspiration multi-data synthesis</td>
<td>1985–2005</td>
<td>$1^\circ \times 1^\circ$</td>
<td>Mueller et al. (2013)</td>
</tr>
</tbody>
</table>
Table 3. Subregions used for probability density functions

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition (Latitude [°], Longitude[°] of region corners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>(20N, 145E) (20N, 100E) (60N, 100E) (60N, 145E)</td>
</tr>
<tr>
<td>Australia</td>
<td>(50S, 155E) (50S, 110E) (10S, 110E) (10S, 155E)</td>
</tr>
<tr>
<td>Europe</td>
<td>(30N, 75E) (30N, 10W) (70N, 10W) (70N, 75E)</td>
</tr>
<tr>
<td>North America</td>
<td>(25N, 60W) (25N, 130W) (60N, 130W) (60N, 60W)</td>
</tr>
</tbody>
</table>