Weather and Climate Extremes 13 (2016) 35–43

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Comparing regional precipitation and temperature extremes in climate model and reanalysis products

Oliver Angélil [a,*], Sarah Perkins-Kirkpatrick [a], Lisa V. Alexander [a], Dáithí Stone [b], Markus G. Donat [a], Michael Wehner [b], Hideo Shiogama [c], Andrew Ciavarella [d], Nikolaos Christidis [d]

[a] Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, UNSW Australia, Sydney NSW 2052, Australia
[b] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[c] National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan
[d] Met Office Hadley Centre, Exeter EX1 3PB, UK

ARTICLE INFO

Article history:
Received 2 May 2016
Received in revised form
30 June 2016
Accepted 10 July 2016
Available online 12 July 2016

Keywords:
Extremes
Evaluation
Event attribution
CAM5.1
MIROC5
HadGEM3-A-N216

ABSTRACT

A growing field of research aims to characterise the contribution of anthropogenic emissions to the likelihood of extreme weather and climate events. These analyses can be sensitive to the shapes of the tails of simulated distributions. If tails are found to be unrealistically short or long, the anthropogenic signal emerges more or less clearly, respectively, from the noise of possible weather. Here we compare the chance of daily land-surface precipitation and near-surface temperature extremes generated by three Atmospheric Global Climate Models typically used for event attribution, with distributions from six reanalysis products. The likelihoods of extremes are compared for area-averages over grid cell and regional sized spatial domains. Results suggest a bias favouring overly strong attribution estimates for hot and cold events over many regions of Africa and Australia, and a bias favouring overly weak attribution estimates over regions of North America and Asia. For rainfall, results are more sensitive to geographic location. Although the three models show similar results over many regions, they do disagree over others. Equally, results highlight the discrepancy amongst reanalyses products. This emphasises the importance of using multiple reanalysis and/or observation products, as well as multiple models in event attribution studies.

1. Introduction

As the climate continues to change under the influence of anthropogenic emissions, there has been a growing interest in how the occurrence of extreme weather events fit within the climate change context (Seneviratne et al., 2012). A common method of characterising the anthropogenic contribution to extreme weather is to analyse the relative probabilities of exceeding an extreme threshold in two simulated distributions (Stone and Allen, 2005; Stott et al., 2004, 2013). These distributions can be constructed from two large ensembles of simulations generated by a dynamical climate model, each run under a different climate scenario: a historical 'real world' (RW) representative of recent observed climate, and a counter-factual 'natural world' (NAT) representative of a climate without human interference in the climate system. Purpose-built model evaluation should underpin the probabilistic event attribution framework used in these studies, whereby the probabilities of extremes are compared across the historical model output and a number of observation and/or reanalysis products. This is necessary as attribution statements are highly sensitive to the shapes of the tails from which they are calculated (Angélil et al., 2014b; Fischer and Knutti, 2015; Jeon et al., 2016). For example, the use of simulated RW and NAT distributions with shorter tails than those of observed distributions lead to exaggerated attribution statements – the shorter tails increase the relative strength of the anthropogenic signal from the noise of natural variability (Bellprat and Doblas-Reyes, 2016). In such an evaluation the use of multiple observation and/or reanalysis products must be considered, as their representation of extremes can differ remarkably (Donat et al., 2014).

Many event attribution studies however typically fail to incorporate multiple observation and/or reanalysis products to evaluate the extreme tails of simulated distributions (Stott et al., 2004; Pall et al., 2011; Peterson et al., 2012, 2013; Herring et al., 2014, 2015). One possible reason for the paucity of such evaluation is the lack of long (~50 years) historical simulations, and long

* Corresponding author.
E-mail address: oliver.angelil@student.unsw.edu.au (O. Angélil).

http://dx.doi.org/10.1016/j.wace.2016.07.001

spatially and temporally complete observational records required for the evaluation of extremes. For example, evaluating one-in-ten-year extremes with datasets ten years in length is both challenging and unreliable.

Using datasets 35 years in length (1979–2013), we evaluate the likelihood of exceeding (or falling below for cold events) one-in-one- and one-in-ten-year daily temperature and precipitation thresholds (defined according to a reference product) over land regions of the world, in ensembles of historical simulations generated by three Atmospheric Global Climate Models (AGCMs). The primary aim of this study is to explore observational uncertainties in model evaluation relevant for extreme event attribution, at the regional scale.

2. Data

2.1. Atmospheric Global Climate Model data

Output was generated by three AGCMs as part of the C20C+ Detection and Attribution Project (see http://portal.nersc.gov/c20c for more information, Folland et al. 2014). Since Pall et al. (2011) numerous event attribution studies have been published utilising output from AGCMs in order to produce the large ensembles needed to accurately resolve the statistics of rare weather events. Here we use CAM5.1, MIROC5 and HadGEM3-A-N216 ('HadGEM3' hereinafter), the first three AGCMs to have a sufficient number of simulations submitted to the C20C+ archive. As there are 10 ensemble members generated by MIROC5 (run at $\sim 1.4°$) which span a number of decades, we use the first 10 historical ensemble members from CAM5.1 and HadGEM3, run at $\sim 1°$ and $\sim 0.5°$ resolution respectively. The members in each ensemble differ from each other only in their initial conditions. Simulations from all AGCMs are roughly 50 years in length but have been trimmed to match availability of the AGCM and reanalyses products used.

The AGCMs are forced under observed boundary conditions. These boundary conditions include greenhouse gases, tropospheric aerosols, volcanic aerosols, ozone concentrations, solar luminosity, sea surface temperature (SST), sea ice coverage (SIC), and land cover. In CAM5.1, prescribed SSTs up to 1982 are an adjusted version of the HadISST1 dataset (Rayner et al., 2003), after which the NOAA-OI.v2 dataset is used (Hurrell et al., 2008). The HadGEM3 (Christidis et al., 2013) and MIROC5 (Shiogama et al., 2013, 2014) prescribed monthly SST and SIC were taken from the HadISST1 dataset.

2.2. Reanalyses

We compare the probabilities of daily extremes in the three AGCMs with four reanalysis products (results using two additional reanalyses products can be found in the Supplementary Material). We firstly examine the ECMWF Interim Reanalysis (ERA-Interim, Dee et al. (2011)) as it has been found that temperature extremes in ERA-Interim correlate more strongly with gridded observations than a selection of other reanalysis products (Donat et al., 2014). Because there is some uncertainty in the representation of extreme weather between observations and reanalyses products (Donat et al., 2014), we complement ERA-Interim with three additional products from the current state-of-the-art generation (Rienecker et al., 2011). These are: NCEP Climate Forecast System Reanalysis (CFSR, Saha et al., (2010)); National Aeronautics and Space Administration (NASA)'s Modern-Era Retrospective Analysis for Research and Applications (MERRA, Rienecker et al., (2011)); and most recently available, the Japanese 55-year Reanalysis (JRA-55, Kobayashi et al. (2015)).

As they are still widely used products, results using the National Centers for Environmental Prediction/National Centre for Atmospheric Research (NCEP/NCAR) Reanalysis 1 (NCEP1, Kalnay et al., (1996)) and NCEP Department of Energy (DOE) Reanalysis 2 (NCEP2, Kanamitsu and Ebisuzaki, (2002)) are included in the Supplementary Material. Despite NCEP1 being shown to perform poorly relative to other reanalyses and observation products for temperature extremes (Donat et al., 2014), it has been widely used in recent event attribution studies (Herring et al., 2014, 2015).

HadGHCND (Caesar et al., 2006) – the only quasi-global long-running in situ-based observation product consisting of daily temperature fields, was excluded from this study not only because it is spatially and temporally incomplete, but also as it is developed at coarse resolution ($3.75° \times 2.5°$) relative to other products in this study. As all data in this study are remapped to the resolution of the coarsest product, we have opted for high resolution analysis over using HadGHCND. All AGCM and reanalysis data have been interpolated to the NCEP1/NCEP2 grid (192×94 grid; 1.9°), using a first-order conservative remapping technique (Jones, 1999).

Since reanalyses are different from observations as they are essentially an assimilation of observations through an atmospheric model, we use gridded observations of daily temperature and precipitation over Australia, from the Australian Water Availability Project (AWAP, Jones et al., (2009)). Observations over only Australia are used because existing gridded observations of daily temperature and precipitation are spatially incomplete. Hot, cold, and wet extremes over three Australian regions are compared between AWAP and ERA-Interim (see Fig. S7).

It should be noted however, that caution should be taken when comparing gridded observations with models due to the "issue of scale" (Avila et al., 2015), which leads to a mismatch between the two types of products. Gridded observations represent regularly spaced values derived from point locations, while output from models represent area averages. There is an additional issue at play in gridded observations such as HadEX2 and GHCNDEX: the order of operations applied to calculate extremes differ from products that provide daily grids of temperature and precipitation, such as climate models and reanalyses. Extremes are first calculated at point locations and then gridded, while in models, extremes are calculated from the gridbox average. This creates a systematic bias where the difference in hot and cold extremes in models are smaller than those found in GHCNDEX and HadEX2.

3. Method

For the evaluation of extremes, thresholds of one-in-one-year ($\frac{1}{365}$ chance of occurrence) and one-in-ten-year ($\frac{1}{3650}$ chance of occurrence) hot, cold, and wet days occurring at the grid and regional scales have been defined from daily anomalies in ERA-Interim, with the base period being the 1979–2013 climatology at each grid cell or region. ERA-Interim serves as our reference product in order to clearly demonstrate differences amongst all AGCMs and reanalyses products. Although perhaps less relevant for extreme event attribution, the selection of the one-in-one-year thresholds allows us to examine extreme anomalies for which sampling should not be problematic considering the length of the period examined. When the desired percentile was between two data points, the nearest point to a linearly interpolated value between the two points was chosen.

The regions used are demarcated by the 58 regions (see Fig. 1 and Angélil et al., (2014b)) in the Weather Risk Attribution Forecast (WRAF, http://web.csag.uct.ac.za/~daithi/forecast/). Each region, roughly $2 \cdot 10^6$ km^2, is based on political-economic borders, and omits regions dominated by small islands (for which the statistical characteristics of extreme atmospheric weather will be