AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2014JD022358

This article is a companion to *Levy et al.* [2014] doi:10.1002/2014JD022357.

Key Points:
- Detection and attribution of precipitation has proved difficult
- Location biases in GCMs make comparisons with observations difficult
- Removing these biases aids the detection of an external forcing

Correspondence to:
A. A. L. Levy,
levy@atm.ox.ac.uk

Citation:
Levy, A. A. L., M. Jenkinson, W. Ingram, F. H. Lambert, C. Huntingford, and M. Allen (2014), Increasing the detectability of external influence on precipitation by correcting feature location in GCMs, *J. Geophys. Res. Atmos.*, *119*, 12,466–12,478, doi:10.1002/2014JD022358.

Received 28 JUL 2014
Accepted 16 OCT 2014
Accepted article online 21 OCT 2014
Published online 18 NOV 2014

Increasing the detectability of external influence on precipitation by correcting feature location in GCMs

Adam A. L. Levy[1], Mark Jenkinson[2], William Ingram[1,3], F. Hugo Lambert[4], Chris Huntingford[5], and Myles Allen[1,6]

[1]Department of Physics, University of Oxford, Oxford, UK, [2]Oxford Centre for Functional MRI of the Brain, University of Oxford, Oxford, UK, [3]Met Office Hadley Centre, Exeter, UK, [4]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK, [5]Centre for Ecology and Hydrology, Wallingford, UK, [6]School of Geography and the Environment, University of Oxford, Oxford, UK

Abstract Understanding how precipitation varies as the climate changes is essential to determining the true impact of global warming. This is a difficult task not only due to the large internal variability observed in precipitation but also because of a limited historical record and large biases in simulations of precipitation by general circulation models (GCMs). Here we make use of a technique that spatially and seasonally transforms GCM fields to reduce location biases and investigate the potential of this bias correction to study historical changes. We use two versions of this bias correction—one that conserves intensities and another that conserves integrated precipitation over transformed areas. Focussing on multimodel ensemble means, we find that both versions reduce RMS error in the historical trend by approximately 11% relative to the Global Precipitation Climatology Project (GPCP) data set. By regressing GCMs' historical simulations of precipitation onto radiative forcings, we decompose these simulations into anthropogenic and natural time series. We then perform a simple detection and attribution study to investigate the impact of reducing location biases on detectability. A multiple ordinary least squares regression of GPCP onto the anthropogenic and natural time series, with the assumptions made, finds anthropogenic detectability only when spatial corrections are applied. The result is the same regardless of which form of conservation is used and without reducing the dimensionality of the fields beyond taking zonal means. While "detectability" is dependent both on the exact methodology and the confidence required, this nevertheless demonstrates the potential benefits of correcting location biases in GCMs when studying historical precipitation, especially in cases where a signal was previously undetectable.

1. Introduction

To evaluate the true impacts of climate change, we must understand how different regions of the globe are affected at different seasons by various climate variables—most crucially temperature and moisture availability. While past and projected changes in temperature have been robustly evaluated, it has been more difficult to quantify changes in precipitation and, consequently, other properties of the hydrological cycle.

In particular, it has proved difficult to detect an external effect on precipitation, let alone determine whether it is attributable to human influence. There are a number of reasons for this. Precipitation distributions vary over much shorter length scales than temperature. Unlike temperature, which in the past century has increased almost everywhere [*Hartmann et al.*, 2013, Figure 2.21], climate change is expected to increase the contrast between wet and dry regions [*Chou et al.*, 2009] as well as shifting climatological features [*Scheff and Frierson*, 2012], and so any forced changes in precipitation will also vary over small length scales. Exacerbating this problem is the limited observational record for precipitation. Studies aiming to detect and attribute an external influence on historical precipitation generally have to choose between relatively sparse land-based data sets [e.g., *Polson et al.*, 2013a; *Zhang et al.*, 2007] or relatively short satellite-based observations that began in 1979 [*Huffman et al.*, 1997] as used by *Marvel and Bonfils* [2013] and *Polson et al.* [2013b].

Precipitation also has much larger internal (i.e., unforced) variability than temperature, with large fluctuations up to decades in length [*Dai*, 2013], dominated by the El Niño–Southern Oscillation (ENSO). As ENSO

©2014. American Geophysical Union. All Rights Reserved.

will not be correlated between observations and the coupled general circulation models (GCMs) used to simulate climate, on decadal time scales, any signal can be masked by noise. To increase the signal-to-noise ratio, ENSO can be removed, especially from observations [*Polson et al.*, 2013b] where ENSO indexes have been defined [*Wolter and Timlin*, 2011], and ensemble means can be taken across GCM simulations to smooth out the variability [*Zhang et al.*, 2007; *Polson et al.*, 2013a, 2013b]. It is also possible to use fingerprints onto which ENSO projects poorly [e.g., *Marvel and Bonfils*, 2013]. Other studies have detected intensification of the hydrological cycle indirectly, as in *Durack et al.* [2012] where sea surface salinities are used as a proxy for the time-integrated difference between precipitation and evaporation.

Comparisons between simulated and observed precipitation are not only limited by uncorrelated variability. GCMs also have large and diverse biases in their simulated mean climate [*Dai*, 2013], which can create artificially large disagreements with each other and with observations on precipitation changes [*Allen and Ingram*, 2002]. Further, while GCM simulations of the mean climatology improve with each phase of the Coupled Model Intercomparison Project (CMIP) [*Taylor et al.*, 2011], as shown in *Knutti et al.* [2013], this has not given rise to better agreement on precipitation changes [*Knutti and Sedlacek*, 2013].

When comparing GCM-simulated precipitation changes, to other GCMs or to observations, it is therefore potentially valuable to remove underlying biases, where possible. While many techniques exist to correct simulated intensity distributions for each grid point [e.g., *Piani et al.*, 2010], such local techniques are limited where a climatological feature is simulated not just with the wrong intensity but also at the wrong location or season [*Haerter et al.*, 2011; *White and Toumi*, 2013]. To circumvent such disagreements in spatial and seasonal location of features, previous studies have averaged over latitude bands and analyzed each season independently [*Zhang et al.*, 2007; *Polson et al.*, 2013a].

It is possible, however, that image transformation tools could be used to correct the location and season of climatological features in GCMs. Such tools have been used to study atmospheric variables previously, though these applications have until recently been limited to numerical weather prediction, specifically validation [*Gilleland et al.*, 2009] and data assimilation [e.g., *Nehrkorn et al.*, 2013]. As such techniques are designed for regional scales, they tend not to be tailored to fields of the same complexity as climatological precipitation or the correct geometry (e.g., longitudinal and seasonal periodicity). As a result, we are only aware of one attempt to use such techniques to study a climate-scale precipitation feature—the South-Pacific Convergence Zone [*Brown et al.*, 2012].

Similar image transformation techniques are also used in brain image registration, where anatomical features in magnetic resonance imaging (MRI) images of different subjects' brains are aligned, so that comparisons can be made. While these techniques also tend to operate in a Cartesian geometry, the complexity of the fields studied is more comparable to climatological precipitation. Brain image registration operates by transforming, or "warping," input brain images so that they closer resemble a reference image [*Ashburner*, 2007; *Andersson et al.*, 2010]. *Levy et al.* [2013] applied these tools to the study of precipitation, deriving transformations by comparing simulated precipitation to reanalysis of observations and applying these same transformations to projected changes under a particular forcing scenario. While this study demonstrated that a group of CMIP5 GCMs were brought into closer agreement using this technique, analysis of idealized projections does not allow for the validation of results, as the "true" changes cannot be known. Further, the investigation made use of medical registration software that is designed for transforming 3-D MRI images and so was not tailored to the specifics of precipitation climatology.

We have now developed a tool specifically for the correction of precipitation climatology [*Levy et al.*, 2014]. Not only does it respect the geometry of the problem (spatially spherical with cyclic seasons to allow for bias correction of features' timing) but it also allows for either grid point intensities or integrated amounts over transformed areas to be conserved upon warping. These two approaches are equivalent to preserving precipitation fluxes, or mass, upon warping.

We now investigate the potential of this technique to improve simulated historical changes in precipitation. We derive a set of transformations for 21 CMIP5 GCMs to correct errors of location relative to an observational data set. We then apply these transformations to the GCMs' historical changes and investigate the effect that this has on agreement with observed changes. While the optimal transformation will vary with time, if GCM location biases are large compared to the shifts in the climatology over time, these transformations will tend to reduce model error. By decomposing the GCM historical simulations into anthropogenic

©2014. American Geophysical Union. All Rights Reserved.

and natural time series, we then analyze whether observed changes can be attributed to either of these external forcings and the impact that warping has on this analysis. We repeat all studies with both the intensity and the integrated precipitation conserving transformations. We do not reduce the dimensionality of the monthly precipitation fields, except to match resolution to observations and to take zonal means.

This paper is organized as follows. Section 2 outlines the process of deriving and selecting transformations, and section 3 describes the application of these transformations to historical CMIP5 simulations. Section 4 discusses the results of this study.

2. Deriving and Selecting Transformations

We wish to remove spatial and seasonal biases from GCM-simulated precipitation so that we can evaluate the impact that this has on agreement with observed changes. To do this, we apply the tool described in [*Levy et al.*, 2014], which is based on medical image registration (or "warping") techniques.

In general, there are two approaches used to warp MRI brain images—the small and large deformation frameworks [*Sotiras et al.*, 2013]. Within the small deformation framework, intensities are translated only once, and so deformations must be "small" in order to avoid introducing tears or folds to an image [*Andersson et al.*, 2010]. Within the large deformation framework, it is common to displace intensities through a succession of small vectors that effectively comprise a velocity field [*Ashburner*, 2007; *Christensen et al.*, 1994]. This allows more accurate transformations but more computational resources to calculate.

The technique used here operates within the small deformation framework for simplicity [*Levy et al.*, 2014]. Though a tool developed in the large deformation framework could offer further improvements, [*Levy et al.*, 2014] found that the tool built in the small deformation framework provides the potential to remove more than half of climatological error. Within this framework, transformations are generally found by minimizing an objective function:

$$O = C(W(I), R) + \bar{\lambda} K(W) \tag{1}$$

where I and R are the input and reference images, respectively, and W is the warp. C is a cost function that evaluates how similar the warped input image is to the reference image, and K is a regularization term that penalizes warps that are less spatially and seasonally smooth. Here C is a sum of squared differences, and K is based on "linear elastic energy" [*Atanackovic and Guran*, 2000]; $\bar{\lambda}$ is a tuneable parameter that determines how smoothly the warp vectors vary. Although there are other parameters internal to K and C, we keep them fixed here, having found they make little difference to the goodness of fit achieved [*Levy et al.*, 2014].

When warping MRI images of brains, image intensities are conserved, and so integrated image brightness is not. However, it is not clear whether this is the most appropriate property to conserve when warping precipitation. The technique therefore includes an option to conserve total integrated precipitation upon warping [*Levy et al.*, 2014]. Preserving total integrated precipitation may seem the more appropriate option in that it maintains consistency with the energy and moisture budgets. However, as it changes precipitation flux values, it is further from the basic intention of merely correcting errors of location. We therefore use both algorithms to compare their abilities to improve simulated historical precipitation changes, regardless of how flux conservation may affect consistency with other simulated fields.

As we aim to remove location biases from each GCM's mean climate, we use the GCM's mean monthly climatology as the input "image," I. The satellite-based observation data set GPCP [*Huffman et al.*, 1997] is used as the reference "image," R, as it provides global, gridded precipitation values. The GCM's climatology is derived from its historical simulation, and only the years where the model and observations overlap (1979–2005) are used to form monthly means. Where more than one historical simulation is available for a GCM, we take a mean across these ensemble members.

Warps are derived with both forms of conservation and with values of $\bar{\lambda}$ ranging from 0.25 to 100 $(\text{mm/d})^2$, which includes warps that strike a balance between goodness of fit and smoothness [*Levy et al.*, 2014]. Generally, goodness of fit improves as $\bar{\lambda}$ decreases, except at very low values. Warps that are selected purely on their goodness of fit will have little applicability, however, as they will take no account of how physical the warp process is [*Rohlfing*, 2012]. We therefore select warps based both on their goodness of fit and a measure of how severely they distort the GCM's field.

©2014. American Geophysical Union. All Rights Reserved.

Table 1. List of GCMs Used in Study, With the Number of Ensemble Members for Each

Model Name(s)	Modeling Group	No. of Runs
BCC	Program for Beijing Climate Center	6
CCSM4	University Corporation for Atmospheric Research	6
CNRM-CM5	Centre National de Recherches Meteorologiques	10
CSIRO-ACCESS 1.0	Commonwealth Scientific and Industrial Research Organisation	1
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory	5
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory	1
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory	1
GISS-E2-R 1.0	National Aeronautics and Space Administration	6
HadGEM2-AO	Met Office Hadley Centre	1
HadGEM2-CC	Met Office Hadley Centre	1
HadGEM2-ES	Met Office Hadley Centre	4
INM-CM4	Institute for Numerical Mathematics	1
IPSL-CM5A-LR	Institut Pierre-Simon Laplace	4
IPSL-CM5A-MR	Institut Pierre-Simon Laplace	1
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology	1
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology	3
MIROC5	Japan Agency for Marine-Earth Science and Technology	3
MPI-ESM-LR	Max Planck Institute for Meteorology	3
MPI-ESM-P	Max Planck Institute for Meteorology	2
MRI-CGCM3	Meteorological Research Institute	5
NorESM1-M	Norwegian Climate Centre	3

To quantify deformations in the warp, we use its Jacobian determinants. The Jacobian determinant quantifies the extent to which volumes change upon warping, and so a very high or low Jacobian implies the introduction of folds or tears into the field. A Jacobian determinant of 1 implies no compression or expansion.

Levy et al. [2013] selected warps to maximize the improvements seen, while also requiring that all the values of the Jacobian lay between 0.1 and 10. While this requirement prevents extreme distortions, it does not take into account the mean effect of the warp (only the extremal values). In order to allow a more complete diagnostic, we require warps to satisfy

$$\left[\frac{\sum_i A_i \max\left(\frac{1}{J_i}, J_i\right)^n}{\sum_i A_i}\right]^{\frac{1}{n}} \leq X \tag{2}$$

where the sum is over all grid points and A_i and J_i are the area weighting and Jacobian determinant of a grid point, i. This diagnostic imposes a limit on a root mean nth power of the Jacobian, and the higher we choose n, the more we emphasize extremal values. If we set our threshold, X, to 10 and take the limit $n \to \infty$, this formula reduces to the requirement in previous work. A threshold of 1 requires a transformation consisting solely of a uniform translation or rotation.

Here we set n to 6 and the limit, X, to 1.5. This puts some emphasis on extreme areas of distortion, while requiring that the overall deviation from a zero transformation is not large. While any choice is somewhat subjective, we found that this choice of n and X effectively ruled out transformations with visible discontinuities and those where $\bar{\lambda}$ was low enough that transformations were unconstrained and goodness of fit was sacrificed. Further, we found that varying the limits did not qualitatively change the nature of the warps, as transformations of the same model with varying values of $\bar{\lambda}$ transformed in qualitatively similar ways [*Levy et al.*, 2014].

We derive and select warps in this way for 21 GCMs from 13 different institutions (Table 1). The warps selected reduced RMS error in the GCMs' climatologies by an average of 44% and 55% in the intensity and integrated precipitation conserving cases, respectively. This has the effect of reducing RMS error in the intermodel ensemble means by 31% and 47%, respectively, relative to the GPCP climatology.

Figure 1 shows the ensemble mean spatial component of the transformation process for 1 month (January). The warp vectors (shown in Figures 1c and 1d) vary smoothly, and both transformations reduce the error in the ensemble mean effectively (as shown by the reduction in color in Figures 1g and 1h, relative

©2014. American Geophysical Union. All Rights Reserved.

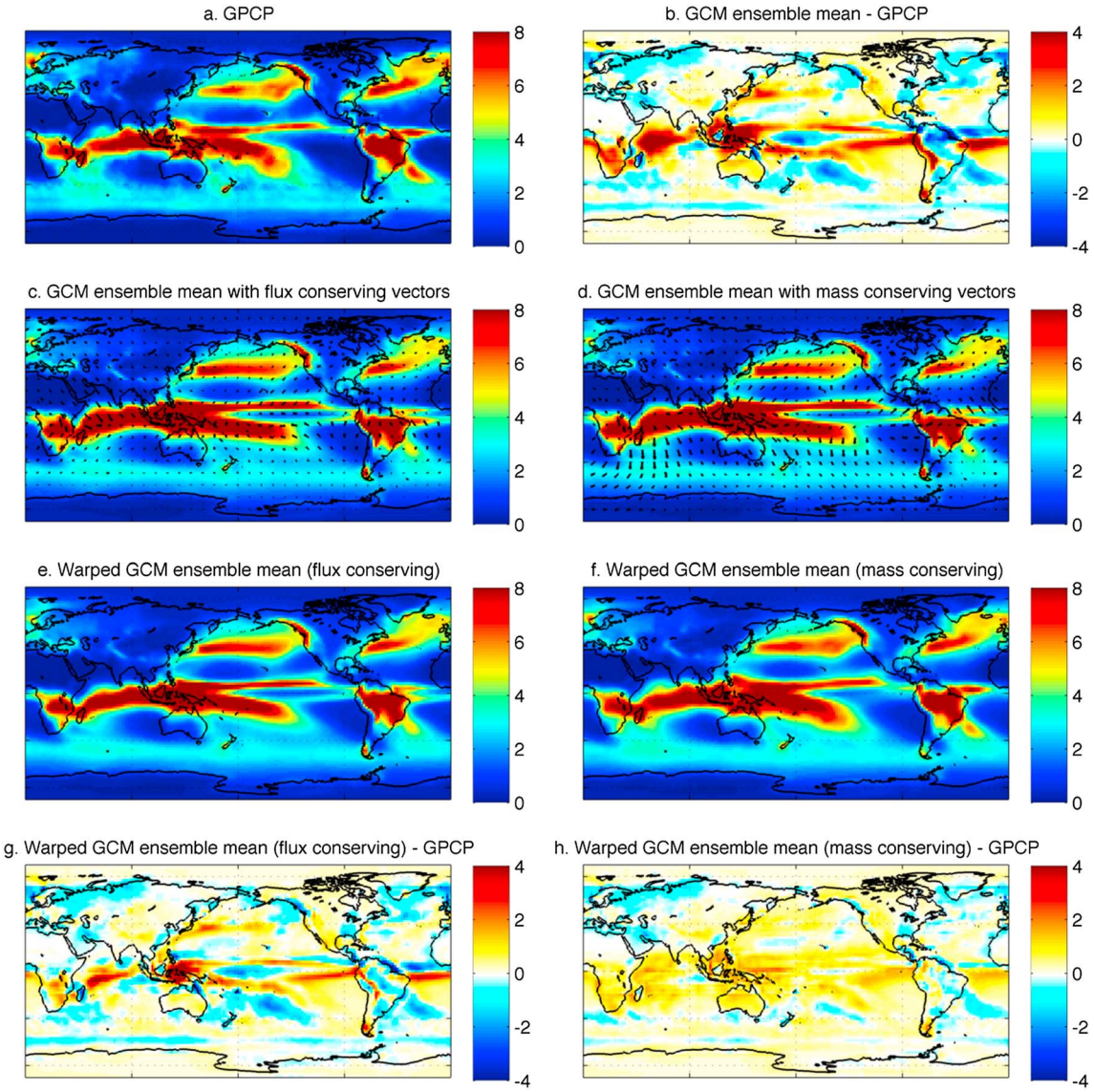

Figure 1. Figures showing mean historical precipitation climatology for January. They show the precipitation patterns (a) for GPCP and (c and d) for the GCM ensemble means before and (e and f) after warping. To highlight changes, difference plots are shown (b) before and (g and h) after warping. For each GCM, we derive transformations to remove spatial and seasonal biases relative to the GPCP climatology. We derive several transformations each with different values of the smoothness parameter $\bar{\lambda}$ (while conserving either intensities or integrated precipitation amounts, i.e., fluxes or mass), and an optimal warp is selected for each model under both regimes using equations (1) and (2). The ensemble mean vectors (Figures 1c and 1d) are displayed above for both forms of warp. All units are in mm d^{-1}, and ensemble means are across 21 GCMs. Historical means are taken over the period of overlap of the CMIP5 historical simulations and GPCP (1979–2005).

to Figure 1b). The integrated precipitation conserving transformations are able to reduce more of the small-scale errors than the intensity conserving case, especially for superfluous features, such as the spurious southern Intertropical Convergence Zone (ITCZ), which the integrated precipitation conserving transformation is able to dilute through expansion.

To explore how sampling uncertainty affects the warps derived, we examine differences in the derived warps when different ensemble members of the same GCM (IPSL-CM5A-LR) are used as input fields. These different ensemble members constitute different samplings of the GCM's variability. This allows us to estimate the sensitivity of the technique both to initial conditions (i.e., the input fields) and to limitations with regards to optimization, which result from a finite number of iterations and becoming "stuck" in local rather than global minima. We find the mean lengths of the three vector components (longitude, latitude, and season) and the mean standard deviation in these three directions. For the flux conserving warps, these

©2014. American Geophysical Union. All Rights Reserved.

are $\left(0.22 \pm 0.02, 0.49 \pm 0.03, (7.9 \pm 0.8) \times 10^{-3}\right)$, and for integrated precipitation conserving warps we find $\left(1.4 \pm 0.1, 1.2 \pm 0.1, (4.6 \pm 0.5) \times 10^{-2}\right)$, with units of the first two dimensions in degrees and the third in months. The large size of the vector lengths relative to the uncertainty demonstrates the robustness of the derived warps to sampling uncertainty both in the input fields and the algorithm.

3. Historical Precipitation Changes

The usefulness of the derived transformations for understanding historical precipitation rests on the assumption that precipitation changes are tied to the underlying climatology so that correcting the seasonal and geographical location of climatological precipitation features will also correct the location of precipitation *changes*. This would be expected, as some climatological precipitation changes can be characterized as an intensification of the mean distribution [*Chou et al.*, 2009; *Held and Soden*, 2006]. Although shifts are also expected to take place over time [*Yin*, 2005], provided these are small relative to location biases, and the transformation is smooth, the derived warps should still serve to reduce errors. To evaluate the benefit of correcting mean feature location, we now apply the transformations to statistical measures that evaluate historical changes.

The simplest measure of change is the trend. We therefore create trend patterns by taking the trend of each grid point for each month of the year independently. These trend patterns are created for all the GCMs and for the observational data set, GPCP, for years with data from both the observations and the GCMs' historical simulations (1979–2005). We then apply our (intensity and integrated precipitation conserving) transformations to each GCM's trend pattern and investigate the effect on the RMS difference from the GPCP trend pattern. RMS error is reduced for all GCMs, using either form of transformation. On average, the error is reduced by 11% for either method of conservation.

While this confirms that location corrections can improve GCM agreement with observed precipitation changes, it is of limited use for detection and attribution of these changes. While the anthropogenic radiative forcing for this period was indeed close to linear, there has also been a substantial natural forcing (see Figure 2). These are most obvious in the form of volcanoes, which have strong time dependence and asymmetry over the time period of interest, and so will affect the trends.

We therefore decompose our GCM fields into two patterns, representing projections onto the natural and anthropogenic forcings. To achieve this, we perform a multiple regression [*Crooks and Gray*, 2005] of each GCM onto the two global annual mean radiative forcings [*Meinshausen et al.*, 2011] shown in Figure 2. As with the trend, this is repeated for each grid point, for each month of the year, independently. Though radiative forcings vary with both latitude and season, the responses to them are spread out by heat transport and storage by atmosphere and ocean and so depend more on the physics of the climate system than the distribution of the forcing [*Boer and Yu*, 2003]. Precipitation also depends on the surface temperature response to forcings, which will further reduce the temporal and spatial specificity of precipitation responses [*Lambert et al.*, 2011]. We therefore use annual global mean values for these forcings.

We can again apply our two types of conserving warps to these patterns. The ensemble means are displayed in Figure 3 for both the anthropogenic and natural patterns. Note that these ensemble means are taken across all GCMs, each of which is itself a mean over that model's available historical simulations. As can be seen, the GCMs' time variation projects much more strongly onto the anthropogenic forcing pattern (effectively a trend pattern). For both components, and for both types of transformation, warping reduces the small-scale features in the ensemble mean, while maintaining the broad structure of the original patterns. This is especially clear for the ITCZ and the South Pacific Convergence Zone, where the anthropogenic pattern is much smoother after warping.

3.1. Detection and Attribution Case Study

Detection and attribution studies aim to identify whether an external effect on a climatological field can be detected and if this effect can be attributed to a particular cause (e.g., anthropogenic or natural). This is achieved by projecting observed patterns onto expected patterns (fingerprints). Generally, these fingerprints are derived using GCMs. This process takes the form [*Allen and Stott*, 2003]

$$\boldsymbol{y} = \sum_i \left[\beta_i \left(\boldsymbol{X}_i + v_i\right)\right] + v_0 \tag{3}$$

©2014. American Geophysical Union. All Rights Reserved.

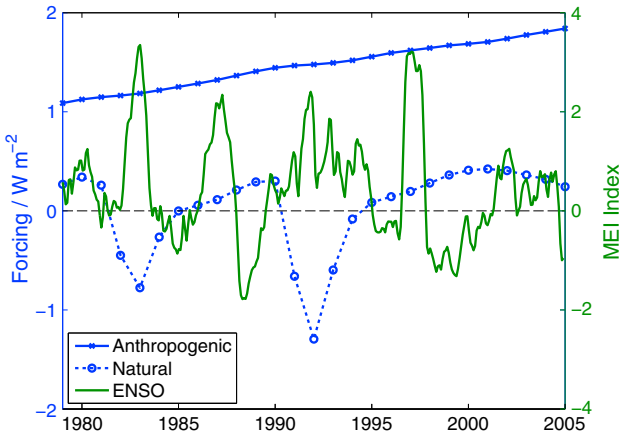

Figure 2. Annual mean, globally averaged radiative forcings, estimated for anthropogenic and natural components (blue). Also shown is the monthly MEI Index, which can be used to characterize ENSO [*Wolter and Timlin*, 2011]. The forcing time series are used to decompose GCM simulations into anthropogenic and natural components, and the MEI index is used to remove ENSO variability from observations (GPCP).

where $\boldsymbol{y}$ is the observed pattern and $\boldsymbol{X}_i$ is the ith fingerprint, v_0 and v_i represent estimates of the variability of the system and of the fingerprints, respectively, and β_i is the ith regression coefficient and so indicates a potential link between observations and the fingerprints. Characterizing the noise allows confidence limits to be estimated, and a signal is generally considered to be detected if zero lies below the 5–95% limits for an individual β_i. When more than one fingerprint is used, detectability for some combination of these fingerprints enables changes to be attributed to particular causes.

In many studies, the noise of the fingerprint is low, and so v_i can be neglected [*Stott et al.*, 2001]. As we take our ensemble mean over 21 GCMs, each of which averaged over the available runs (Table 1), the model contribution to internal variability will be small compared to that of the observations. While there will be a substantial contribution to v_i from differences between models, the effect of reducing this noise through warping is precisely what we wish to investigate. We therefore neglect fingerprint noise for simplicity. We note that this may bias the best estimate and upper bounds of β toward zero, though *Tett et al.* [1999] found that it did not affect the lower bounds.

In order to attribute observed changes to an external signal, we derive fingerprints from our simulated natural and anthropogenic patterns. To improve detectability, the component of internal variability due to ENSO can be removed before performing the detection study [*Polson et al.*, 2013b]. We remove ENSO variability from GPCP by performing a multivariate regression onto the anthropogenic and natural forcings, as well as the Multivariate ENSO Index (MEI) [*Wolter and Timlin*, 2011], shown in Figure 2. The outer product of the MEI time series and the MEI component of GPCP is then subtracted from GPCP's historical simulation, except when the result would imply negative precipitation (fewer than 1% of values). We use a multivariate (as opposed to single) regression in an attempt to reduce the impact of the ENSO removal on the natural and anthropogenic components. We refer to the resulting field as the "ENSO-free" GPCP precipitation field. While ENSO variability exists within the models, its impact will be minimal in the ensemble mean.

Previously, internal variability and GCM biases have meant that studies that aim to detect and attribute an external signal in precipitation changes have been required to greatly reduce the dimensionality of the field. Approaches include taking annual or seasonal means, latitude bands [*Allen and Ingram*, 2002; *Zhang et al.*, 2007; *Polson et al.*, 2013a], and characterizing changes in terms of shifts and intensifications of extremal zonal values [*Marvel and Bonfils*, 2013]. While studies that attempt to attribute changes of a particular nature (e.g., intensification of the contrast between wet and dry regions) have been able to make use of shorter satellite-based observational data sets [*Polson et al.*, 2013b; *Marvel and Bonfils*, 2013], studies that attempt to identify a signal in the entire field generally require longer time series and so have been limited to precipitation over land.

Here we attempt to detect and attribute a signal using the entire year, without limiting our findings to land. It remains, however, desirable to reduce the dimensionality of the field partially to reduce the impact of internal variability. As we wish to determine the potential impact of warping techniques on such studies, we first investigate whether reducing the dimensionality in each dimension impacts on the improvements we see from warping. Naturally, if we took a global, annual mean, the effect of the warp would be reduced to near zero, as there can be no location biases in such a field.

©2014. American Geophysical Union. All Rights Reserved.

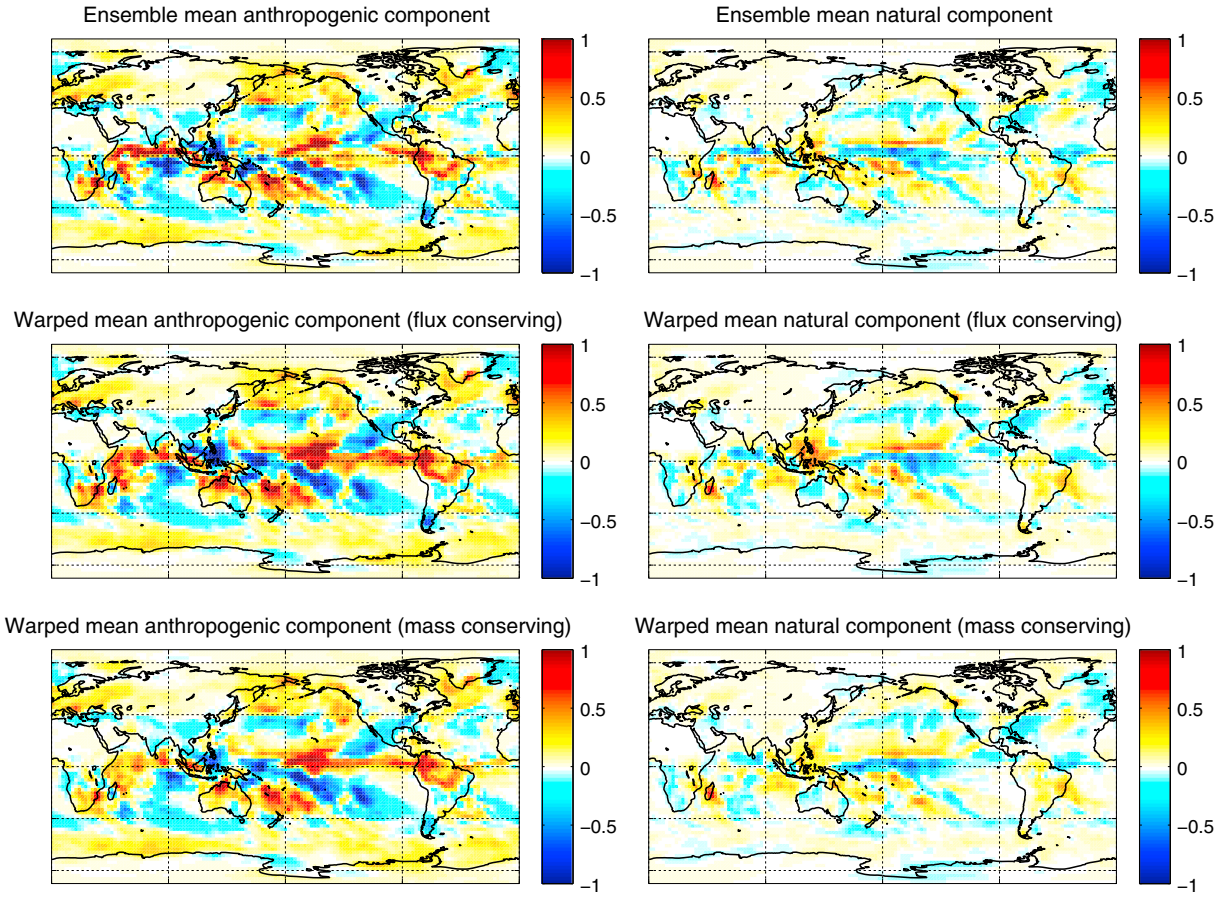

Figure 3. Plots showing ensemble mean regressed patterns for January, with (left) anthropogenic and (right) natural components. These patterns are derived by performing a multiple ordinary least squares regression of each GCM's historical run (for 1979–2005) onto global annual mean radiative forcings, independently for each grid point, for each month of the year. The ensemble means are taken of the 21 GCMs before and after applying intensity or integrated precipitation conserving warps (labeled flux and mass above, respectively). Units are in mm $d^{-1}(W\ m^{-2})^{-1}$.

While smoothing east-west does not substantially sacrifice improvements seen from warping, north-south or seasonal smoothing greatly reduces the effect of the warp. This is to be expected, as there is far more structure in precipitation climatology in the latitudinal and seasonal directions. We therefore take zonal means for this study, without reducing dimension in the seasonal or latitudinal dimensions, except to match the GCMs' latitudinal resolution to that of GPCP (2.5° spacing).

For the observed climate pattern, therefore, we simply take the zonally averaged ENSO-free GPCP time series, with the monthly mean subtracted, so that it forms a time series of monthly anomalies. We project this onto anthropogenic and natural fingerprints. These are composed by taking the outer product of the GCM ensemble mean coefficient patterns (described above and shown in Figure 3) with the relevant forcing time series (Figure 2). The fingerprints are therefore a simulated estimation of the natural and anthropogenic components of the historical time series.

Performing an ordinary least squares regression requires an estimate of the external noise. Here we derive 90% confidence intervals by replacing the GPCP field with sections of preindustrial control runs of the same length. We use the control runs of the same 21 GCMs used to simulate historical precipitation, which provide us with a distribution of 463 correlation coefficients. This large sample size allows us to approximate the distribution as Gaussian and so can derive confidence bands from the standard deviation of these coefficients. Note that by using climate model control simulations in computing confidence intervals, the covariance structure of internal climate variability will be taken into account.

©2014. American Geophysical Union. All Rights Reserved.

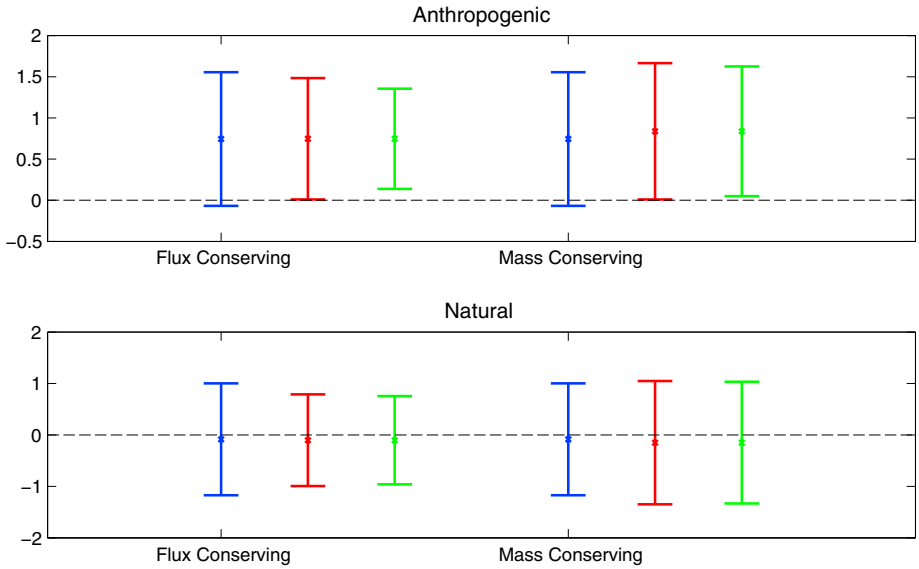

Figure 4. Regression coefficients ("β") from regressing GPCP anomalies onto (top) anthropogenic and (bottom) natural components of CMIP5 ensemble mean. The observed patterns are taken from GPCP, with ENSO variability removed by regression as described in the text (section 33.1). The error bars represent the 90% confidence range, derived using GCM control runs to capture the variability. The best guess values for the regression coefficients are shown by crosses. Shown are results without warping (blue), with warped historical runs (red), and with warped historical and control runs (green). Detection is marginally achieved in all anthropogenic cases where warping is applied but in none of the natural cases (where the best guess is slightly negative). The GCMs were warped using (left) intensity conserving and (right) integrated precipitation conserving warps.

We make no attempt to remove ENSO from these control runs, since each one will be different, and any attempt to derive a suitable index for each one would necessarily be both onerous and subjective. Thus, the model simulations that give our estimate of natural variability will contain extra noise, which will tend to make tests conservative. In order to test the impact of correcting location biases, we repeat this experiment with warped historical runs and with both historical and control runs warped, with both forms of conserving warp. By applying our warps to simulated internal variability as well as the fingerprint signals, we hope to address model-model differences in climate variability.

The results of these regressions are shown in Figure 4. As can be seen, with or without warping, no signal is detected when projecting onto the natural component. Indeed, in all cases the best guess is a slightly negative correlation coefficient. For the anthropogenic component, on the other hand, there is consistency with one in all cases. Before warping, there is marginal consistency with zero, indicating that a signal cannot be detected in this case. With either intensity or integrated precipitation conserving warps (with or without warped control runs), however, the 90% error bars no longer cross the *x* axis, indicating that the anthropogenic signal is detected at the 5% level by this particular analysis.

While 95% confidence bands are the standard used, they reflect an arbitrary choice. It is therefore more useful to consider the *change* in confidence of the detection rather than the binary of whether or not a signal has been detected under a particular criterion. Here an anthropogenic signal was detected with 93% confidence before warping. After warping both historical and control runs, the confidence in the result is increased to 98% and 96% in the intensity and integrated precipitation conserving cases, respectively.

To verify that these improvements are not an artefact of the removal of ENSO from GPCP, we also consider results where ENSO is not removed from GPCP. Figure 5 shows the results obtained through this process, and same results without ENSO removal from GPCP. While detection is not achieved with 95% confidence, we can see that warping again improves the detectability of the anthropogenic component. Here confidence is increased from 84% before warping to 90% and 89% in the intensity and integrated precipitation conserving cases, respectively. Again, warping does not substantially affect the detectability of the natural component, although not removing ENSO from GPCP provides better consistency with 1 for this component.

©2014. American Geophysical Union. All Rights Reserved.

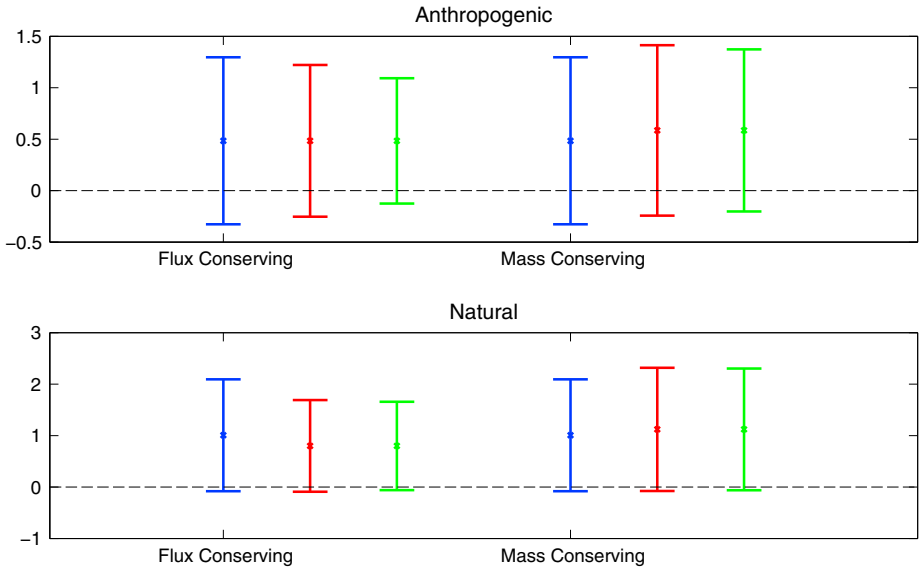

Figure 5. As with Figure 4 but without ENSO variability removed from GPCP. The error bars represent the 90% confidence range, derived using GCM control runs to capture the variability. Shown are results without warping (blue), with warped historical runs (red), and with warped historical and control runs (green). Detection at the 95% level is no longer achieved in the anthropogenic case, though warping the models still results in improved detectability. Improvements in the natural component are negligible.

4. Discussion and Conclusion

This result demonstrates that our ability to detect and attribute a human influence on historical precipitation changes can be enhanced by correcting the location of features in GCMs. In the case presented here, these changes are attributable only where location correction techniques are used. While this attribution relies on an arbitrary threshold, we have also demonstrated that the probability of a nonpositive regression coefficient is reduced from 7% to 2% with intensity conservation or to 4% with integrated precipitation conservation.

There are several reasons why correcting the location of features enhances the detectability of the signal. By transforming the GCMs' historical runs, features are lined up better before taking ensemble means so that the signal is not smoothed out before the fields are merged. More importantly, the ensemble mean changes are then better aligned with the observed patterns, improving our chances of detecting a signal. Further, correcting feature location in control runs allows a better estimate of the variability of the system to be made, which is not artificially inflated by climatological differences between GCMs. The improvements in detectability support our expectation that both intermodel agreement and agreement between models and observations on precipitation changes are partially masked by the various location biases in the GCMs' underlying climatologies.

This does not apply for the natural signal, where the best guess correlation coefficient is found to be small and negative in all cases. This is an artefact of the ENSO removal technique, as when this study is repeated without removing ENSO variability; the best guess, indicated by the crosses in Figure 5, is found to be close to one for both natural and anthropogenic coefficients (detection with a 95% confidence is lost for the anthropogenic correlation coefficient, however). Further, as shown in Figure 5, the result that warping improves detectability is resilient to whether or not ENSO is removed from the observations.

This negative effect on the natural component due to ENSO removal is due to the known negative correlation in recent decades between the natural forcing and ENSO variability, though there is no evidence that this relationship is causal [*Self et al.*, 1997]. For the time period used here (1979–2005), the correlation coefficient between these two time series is −0.4. As a result, in spite of removing the ENSO signal using a multiple regression that incorporates both the anthropogenic and natural forcings, we have reduced our ability to attribute a component of observed changes to natural forcing. While future work could attempt

©2014. American Geophysical Union. All Rights Reserved.

to better decouple the ENSO and natural signals, we are primarily concerned with investigating the effect of warping on the anthropogenic signal here.

It is important to note the sensitivity of the method presented here. The detection of the anthropogenic signal with 95% confidence is only marginal, as suggested by the loss of detectability when ENSO variability is not removed from GPCP. Although leaving ENSO variability in the control runs will have increased the size of our confidence bars, other factors could imply that the error bars are an underestimate. Most importantly, model-simulated variability may underestimate the true natural variability of precipitation in the climate system. Further, by using an ordinary least squares rather than a total least squares regression, we have neglected the noise internal to the climate models. Although taking an ensemble mean across 21 GCMs (most of which are comprised of several ensemble members) reduces the size of this noise, there remains model noise through differences in model climatologies and underlying physics, though this will primarily bias only the central and upper estimates [*Tett et al.*, 1999].

Nevertheless, the increases in confidence (with or without ENSO removal) demonstrate that correcting feature location reduces model error in simulated historical changes, which in turn improves the possibility of detecting and attributing changes to external forcings. We find that in spite of substantial differences in the behavior of the transformations, both the intensity and integrated precipitation conserving warps enhance detectability. The reduction of RMS error with the observed trend patterns, using either technique, also indicates that more than a tenth of the error in simulated historical precipitation changes can be attributed to errors in location, which in previous studies have gone uncorrected.

Further, in spite of the ability of the integrated precipitation conserving technique to better reduce errors in the climatology, both conservation approaches removed a similar amount of RMS error from GCM trend patterns. This indicates that the applicability of a particular transformation cannot be exclusively determined based on its success at removing errors from a GCM's mean historical climate. We note that the flux conserving technique improves the consistency of GCM-simulated precipitation changes with observations, in spite of introducing inconsistency between precipitation and other GCM-simulated fields. This provides evidence that bias correction of feature location is a valuable tool regardless of its precise form (though one form may be more appropriate than others for a particular application). Examining the warps derived for different ensemble members of the same GCM, we find that the standard deviations are substantially smaller than the mean vector lengths, demonstrating the transformations' robustness to sampling uncertainty. Uncertainty in the warping technique could be explored further in future work by, for example, using a different form for the regularization term (R in equation (1)) or criteria for selecting warps (see equation (2)).

Note that in this study, we have not considered errors in our observation data set, GPCP. Climatological biases in the observations will tend to be due to errors in precipitation intensity, not location [*Huffman et al.*, 2009]. However, land-sea contrast may be affected by differences in satellite retrieval algorithms for these two contexts, which may introduce biases into the warps derived for GCMs . Further, introductions of new satellites may introduce biases in the GPCP trend [*Polson et al.*, 2013b]. Any source of error that is uncorrelated between observations and GCMs will tend to obscure any common signal that is present and so serves to reduce the detectability of a signal, with or without warping. Future work could therefore seek to quantify these sources of uncertainty by using other precipitation observations, both in deriving warps and in comparing observations to fingerprints.

Similarly, we do not explicitly consider the effect of aerosols on precipitation, which are resolved to varying extents in modern GCMs. Errors in the mean response to aerosols are one of the sources of location biases that our technique hopes to reduce and so are dealt with implicitly. However, there will also be a precipitation response to aerosol forcing over the time period studied, though there remains substantial uncertainty both in the observed [*Stocker et al.*, 2013] and simulated [*Forster et al.*, 2013] radiative forcing. Aerosol microphysics may of course have further influence on precipitation than from their radiative effect, and GCMs' varying abilities to resolve these processes will further reduce our ability to detect a signal. The techniques used in this paper, however, could help quantify such effects by separating climatological and trend biases that result from limitations in the simulation of aerosols.

There have been several efforts previously to detect and attribute precipitation changes by increasing the signal to noise in the observations and fingerprints being studied. This may be achieved, for example, by reducing the dimensionality [*Marvel and Bonfils*, 2013] or by using a longer, land-limited observational data

©2014. American Geophysical Union. All Rights Reserved.

set [*Polson et al.*, 2013a]. By correcting spatial and seasonal biases in GCMs, we have demonstrated that the agreement between observed and fingerprint patterns can be improved, further enhancing our ability to attribute observed changes to external forcings. This allows more robust understanding of changes in observed precipitation records and builds confidence in model-simulated projections.

It would therefore be of interest to apply this technique where a signal has previously been undetectable or only marginally detectable. We hope to also make use of the location correction technique developed by [*Levy et al.*, 2014] to improve precipitation projections under various forcing scenarios. While *Levy et al.* [2013] found improvements in intermodel agreement on precipitation change, an idealized forcing experiment was used, and changes were evaluated over a long period to improve signal to noise. We therefore hope to investigate the potential of warping techniques to improve near-term precipitation projections.

For such analyses, further care must be taken to ensure that the transformation derived for the historical period still applies to the GCM's future climatology (i.e., shifts of features in the climatology are small relative to the scale of the error corrections applied). Indeed, in this study, the intensity conserving warps are able to remove slightly more error from the preindustrial control runs than the integrated precipitation conserving warps, despite correcting the historical (1979–2005) climatology less accurately.

In addition, techniques such as these could be applied to fields besides precipitation. Further, improved estimates of precipitation could be fed into other models to improve simulation of other hydrological fields, such as runoff. It is also our hope that transformations such as these could serve as a useful diagnostic tool for the evaluation of model errors. Through such feedback, we hope to assist in the development of future GCMs.

Acknowledgments

We acknowledge the World Climate Research Programme's Working Group on coupled modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This paper benefited greatly from the Oxford Advanced Research Computing department, and we are especially grateful for invaluable assistance from Albert Solernou. A.A.L.L., W.J.I., F.H.L., C.H., and M.J. were supported by NERC under contract NE/I00680X/1 (HYDRA). M.R.A. also received support from the NOAA/DOE IDAG project. Both the data from this study and the location correction tool used can be accessed through communication with the corresponding author, A.A.L.L. Please contact via email using levy@atm.ox.ac.uk.

References

Allen, M., and P. Stott (2003), Estimating signal amplitudes in optimal fingerprinting. Part I: Theory, *Clim. Dyn.*, *21*(5–6), 477–491, doi:10.1007/s00382-003-0313-9.

Allen, M. R., and W. J. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, *419*(6903), 224–232.

Andersson, J., M. Jenkinson, and S. Smith (2010), Non-linear registration, aka spatial normalisation, *FMRIB Tech. Rep.*, available from http://fsl.fmrib.ox.ac.uk/analysis/techrep/tr07ja1/.

Ashburner, J. (2007), A fast diffeomorphic image registration algorithm, *NeuroImage*, *38*(1), 95–113.

Atanackovic, T., and A. Guran (2000), *Theory of Elasticity for Scientists and Engineers*, Birkhäuser, Boston, Mass.

Boer, G., and B. Yu (2003), Climate sensitivity and response, *Clim. Dyn.*, *20*(4), 415–429, doi:10.1007/s00382-002-0283-3.

Brown, J. R., A. F. Moise, and F. P. Delage (2012), Changes in the South Pacific convergence zone in IPCC AR4 future climate projections, *Clim. Dyn.*, *39*(1–2), 1–19, doi:10.1007/s00382-011-1192-0.

Chou, C., J. D. Neelin, C.-A. Chen, and J.-Y. Tu (2009), Evaluating the "rich-get-richer" mechanism in tropical precipitation change under global warming, *J. Clim.*, *22*(8), 1982–2005, doi:10.1175/2008JCLI2471.1.

Christensen, G. E., R. D. Rabbitt, and M. I. Miller (1994), 3D brain mapping using a deformable neuroanatomy, *Phys. Med. Biol.*, *39*(3), 609–618.

Crooks, S. A., and L. J. Gray (2005), Characterization of the 11-year solar signal using a multiple regression analysis of the ERA-40 dataset, *J. Clim.*, *18*(7), 996–1015, doi:10.1175/JCLI-3308.1.

Dai, A. (2013), The influence of the inter-decadal Pacific oscillation on US precipitation during 1923–2010, *Clim. Dyn.*, *41*, 633–646, doi:10.1007/s00382-012-1446-5.

Durack, P. J., S. E. Wijffels, and R. J. Matear (2012), Ocean salinities reveal strong global water cycle intensification during 1950 to 2000, *Science*, *336*(6080), 455–458.

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka (2013), Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, *J. Geophys. Res. Atmos.*, *118*, 1139–1150, doi:10.1002/jgrd.50174.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert (2009), Intercomparison of spatial forecast verification methods, *Weather Forecasting*, *24*(5), 1416–1430, doi:10.1175/2009WAF2222269.1.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani (2011), Climate model bias correction and the role of timescales, *Hydrol. Earth Syst. Sci.*, *15*(3), 1065–1079, doi:10.5194/hess-15-1065-2011.

Hartmann, D., et al. (2013), Observations: Atmosphere and surface, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. Stocker et al., p. 193, Cambridge Univ. Press, Cambridge, U. K., and New York.

Held, I. M., and B. J. Soden (2006), Robust responses of the hydrological cycle to global warming, *J. Clim.*, *19*(21), 5686–5699, doi:10.1175/JCLI3990.1.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The global precipitation climatology project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, *78*, 5–20, doi:10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the global precipitation record: GPCP version 2.1, *Geophys. Res. Lett.*, *36*, L17808, doi:10.1029/2009GL040000.

Knutti, R., and J. Sedlacek (2013), Robustness and uncertainties in the new CMIP5 climate model projections, *Nat. Clim. Change*, *3*(4), 369–373.

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5 and how we got there, *Geophys. Res. Lett.*, *40*, 1194–1199, doi:10.1002/grl.50256.

©2014. American Geophysical Union. All Rights Reserved.

Lambert, F. H., M. J. Webb, and M. M. Joshi (2011), The relationship between land–ocean surface temperature contrast and radiative forcing, *J. Clim.*, *24*(13), 3239–3256, doi:10.1175/2011JCLI3893.1.

Levy, A. A. L., W. Ingram, M. Jenkinson, C. Huntingford, F. H. Lambert, and M. Allen (2013), Can correcting feature location in simulated mean climate improve agreement on projected changes?, *Geophys. Res. Lett.*, *40*, 354–358, doi:10.1029/2012GL053964.

Levy, A. A. L., M. Jenkinson, W. Ingram, and M. Allen (2014), Correcting precipitation feature location in General Circulation Models, *J. Geophys. Res. Atmos.*, doi:10.1002/2014JD022357.

Marvel, K., and C. Bonfils (2013), Identifying external influences on global precipitation, *PNAS*, *110*, 19,301–19,306, doi:10.1073/pnas.1314382110.

Meinshausen, M., et al. (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Clim. Change*, *109*(1–2), 213–241, doi:10.1007/s10584-011-0156-z.

Nehrkorn, T., B. Woods, T. Auligné, and R. N. Hoffman (2013), Application of feature calibration and alignment to high-resolution analysis: Examples using observations sensitive to cloud and water vapor, *Mon. Weather Rev.*, *142*(2), 686–702, doi:10.1175/MWR-D-13-00164.1.

Piani, C., J. Haerter, and E. Coppola (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.

Polson, D., G. C. Hegerl, X. Zhang, and T. J. Osborn (2013a), Causes of robust seasonal land precipitation changes, *J. Clim.*, *26*, 6679–6697, doi:10.1175/JCLI-D-12-00474.1.

Polson, D., G. C. Hegerl, R. P. Allan, and B. B. Sarojini (2013b), Have greenhouse gases intensified the contrast between wet and dry regions?, *Geophys.Res. Lett.*, *40*, 4783–4787, doi:10.1002/grl.50923.

Rohlfing, T. (2012), Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable, *IEEE Trans. Med. Imaging*, *31*(2), 153–163, doi:10.1109/TMI.2011.2163944.

Scheff, J., and D. Frierson (2012), Twenty-first-century multimodel subtropical precipitation declines are mostly midlatitude shifts, *J. Clim.*, *25*(12), 4330–4347, doi:10.1175/JCLI-D-11-00393.1.

Self, S., M. R. Rampino, J. Zhao, and M. G. Katz (1997), Volcanic aerosol perturbations and strong El Niño events: No general correlation, *Geophys. Res. Lett.*, *24*(10), 1247–1250, doi:10.1029/97GL01127.

Sotiras, A., C. Davatzikos, and N. Paragios (2013), Deformable medical image registration: A survey, *IEEE Trans. Med. Imaging*, *32*(7), 1153–1190, doi:10.1109/TMI.2013.2265603.

Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2013), Climate change 2013: The physical science basis, in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., 1535 pp., Cambridge Univ. Press, Cambridge, U. K., and New York.

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell (2001), Attribution of twentieth century temperature change to natural and anthropogenic causes, *Clim. Dyn.*, *17*(1), 1–21, doi:10.1007/PL00007924.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell (1999), Causes of twentieth-century temperature change near the Earth's surface, *Nature*, *399*(6736), 569–572.

White, R. H., and R. Toumi (2013), The limitations of bias correcting regional climate model inputs, *Geophys. Res. Lett.*, *40*, 2907–2912, doi:10.1002/grl.50612.

Wolter, K., and M. S. Timlin (2011), El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), *Int. J. Climatol.*, *31*(7), 1074–1087, doi:10.1002/joc.2336.

Yin, J. H. (2005), A consistent poleward shift of the storm tracks in simulations of 21st century climate, *Geophys. Res. Lett.*, *32*, L18701, doi:10.1029/2005GL023684.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa (2007), Detection of human influence on twentieth-century precipitation trends, *Nature*, *448*(7152), 461–465.

©2014. American Geophysical Union. All Rights Reserved.