

Consistency of simulated and observed regional changes in temperature, sea level pressure and precipitation

J. Bhend · P. Whetton

Received: 31 July 2012 / Accepted: 27 December 2012 / Published online: 6 February 2013
© Springer Science+Business Media Dordrecht 2013

Abstract There is increasing pressure from stakeholders for highly localised climate change projections. A comprehensive assessment of climate model performance at the grid box scale in simulating recent change, however, is not available at present. Therefore, we compare observed changes in near-surface temperature, sea level pressure (SLP) and precipitation with simulations available from the Coupled Model Intercomparison Projects 3 and 5 (CMIP3 and CMIP5). In both multi-model datasets we find coherent areas of inconsistency between observed and simulated local trends per degree global warming in both temperature and SLP in the majority of models. Localised projections should thus take into account the possibility of regional biases shared across models. In contrast, simulated changes in precipitation are not significantly different from observations due to low signal-to-noise ratio of local precipitation changes. Therefore, recent regional rainfall change is likely not providing useful constraints for future projections as of yet. Comparing the two most recent sets of internationally coordinated climate model experiments, we find no indication of improvement in the models' ability to reproduce local trends in temperature, SLP and precipitation.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-012-0691-2) contains supplementary material, which is available to authorized users.

J. Bhend (✉)
CSIRO Marine and Atmospheric Research,
Private Bag 1, Aspendale, VIC 3195, Australia
e-mail: jonas.bhend@csiro.au

J. Bhend · P. Whetton
CSIRO Climate Adaptation Flagship, Aspendale, VIC 3195, Australia

1 Introduction

Over the past decade, demand for spatially explicit climate change information for impact and adaptation assessment has been steadily increasing. As many impact and adaptation assessments are carried out for specific small-scale regions, users of climate model output may find themselves relying entirely on the information from a single grid-box of the climate model output to produce what we call here *local* projections (e.g. Fronzek et al. 2010). The performance of climate models in simulating past observed change at these scales, however, has not been comprehensively assessed and hence trust in local climate projections mainly stems from skill in reproducing large-scale features (Randall et al. 2007). However, it is also known that climate model performance is scale-dependent (Masson and Knutti 2011b) and strongly varies across regions (Christensen et al. 2007). To address the above issues, we compare local observed climate change in near-surface temperature, sea level pressure (SLP) and precipitation for the second half of the twentieth century with corresponding simulations. This allows us to assess spatially varying model trend biases and identify regions where present-day climate models fail to reproduce recent climate change. Such information is of great practical relevance for users of local climate projections, as trust in future simulations would be significantly lowered if the majority of models fails to reproduce the past observed changes.

In the context of local climate projections, we are especially interested in assessing the climate model's ability to simulate the response to external forcing (mainly increasing greenhouse gas concentrations). So here we decompose the simulated and observed time series into deterministic climate change and variability. Local climate change is defined as the fraction of the variability in the time series that co-varies with global temperature. This maximises the signal-to-noise ratio of long-term changes such as the response to anthropogenic greenhouse gas emissions and compensates for varying climate sensitivity and differing external forcings across climate models. In addition, this pattern scaling approach is widely used to produce regional climate projections (Christensen et al. 2007).

A few pioneering studies have been carried out to analyse performance of climate models in simulating changes at the grid box scale, but have taken into account only anthropogenic forcing (Károly and Wu 2005), assume model independence (Räisänen 2007), and have considered a single region (van Oldenborgh et al. 2009; van Haren et al. 2012), or did not provide regional results (Yokohata et al. 2011). Our analysis extends the above by using simulations which include most relevant processes and forcing mechanisms and by carrying out a regionally specific but globally comprehensive assessment that allows us to exclude the influence of a selection bias that may impair regional analyses. Additionally, we also compare the results of the two latest generations of the World Climate Research Programme's (WCRP) Coupled Model Intercomparison Projects (CMIP3 and CMIP5) thus aiming to document similarities and potential progress in the two generations of multi-model datasets.

We assess individual models to identify regions and variables for which a majority of the models simulate recent local changes that are inconsistent with the observations. These shared inconsistencies may be indicative of systematic model trend biases. Furthermore, a spatially explicit analysis of inconsistencies may hint at processes that are misrepresented in the climate models leading to the identified

inconsistencies. To allow for observation errors we use both gridded observation datasets and reanalyses—model simulations with assimilated observations—for validation of the climate change simulations.

In the following section, we introduce the datasets used in this study. In Section 3, we describe how we extract the climate change signal from simulations and observations and how we assess consistency of simulated and observed local change. We present spatial patterns of significant inconsistencies along with global average statistics and a comparison of the CMIP3 and CMIP5 results in Section 4. Implications of the results and benefits and limitations of the approach are discussed in the final section.

2 Observations and climate model simulations

We analyse seasonal temperature, SLP, and precipitation for the most recent 50-year period common to observations and simulations. For near-surface temperature, we use the GISS surface temperature analysis from 1880 to present with 1200 km smoothing as our primary validation dataset (Hansen et al. 2010). The dataset—referred to here as GISTEMP—consists of interpolated station and ship observations. The primary SLP observations are from the gridded HadSLP2 dataset, which consists of local pressure observations aggregated to $5^\circ \times 5^\circ$ available for 1850–2004 (Allan and Ansell 2006). Precipitation observations before 1979 are only available over land. Here we use the Global Precipitation Climatology Centre's (GPCC) Variability Analysis of Surface Climate Observations (VASCLimO) version 1.1 consisting of quality-controlled rain-gauge observations on a $2.5^\circ \times 2.5^\circ$ grid for 1951–2000 (Beck et al. 2005).

In addition to the gridded station measurements, we also use data from the Twentieth Century Reanalysis Version 2 (20CRv2, Compo et al. 2011). With 20CRv2, only SLP data are assimilated and therefore the reanalysis is not subject to inhomogeneities related to the inclusion of satellite data after 1979 as other reanalyses (Sterl 2004). Changing observation density, however, still might affect the 20CRv2 (Ferguson and Villarini 2012). In addition to the assimilated SLP observations, gridded observed sea-surface temperatures constrain the atmospheric model at its lower boundary. In contrast, observed temperatures over land are not assimilated nor prescribed, but are simulated.

Here we analyse output stored in the two latest versions of the World Climate Research Programme's (WCRP) Coupled Model Intercomparison Projects (CMIP), termed CMIP3 and CMIP5 respectively. The simulations in CMIP3 (Meehl et al. 2007) represent the state of climate modelling in 2005. CMIP5 models have only recently been made available and more simulations will be added to the archive in the near future.

From the CMIP3 multi-model ensemble, we analyse 51 simulations from 24 models. These use observed external forcings for the twentieth century as boundary conditions (experiment 20c3m) and are extended into the twenty-first century with simulations according to the SRES A1B scenarios. Eleven models include both natural and anthropogenic forcings for the twentieth century and the remainder anthropogenic forcings only (Online Resource, Tab. S1). From 2000, all simulations are constrained by anthropogenic forcings scenarios only. This is expected to introduce

negligible errors due to the lack of major volcanic eruptions in the first five years of the twenty-first century, the inertia of the climate system, and the similarity of emission scenarios with observed emissions for this time period. From the CMIP5 multi-model ensemble, we analyse 93 simulations from 26 models using observed external forcings until 2005 as boundary conditions (Online Resource, Tab. S2). Initial condition ensemble means are computed from all the available simulations of each individual model.

All data are interpolated onto a common $1.5^\circ \times 1.5^\circ$ grid using conservative remapping and simulations are masked according to the observations. Seasonal averages are computed if all the corresponding monthly values were available. Trends are computed if at least 80 % of the seasonal averages are available. For precipitation, the analysis is carried out on relative anomalies with respect to the base period 1961–1990 and arid regions with a seasonal precipitation climatology of less than 20 mm are masked.

3 Analysis scheme

To assess the significance of the difference in observed and simulated local trends, we follow the approach as described by van Oldenborgh et al. (2009). We compute local climate change $A(x, y)$ by regressing the respective grid box time series $R(x, y, t)$ on the corresponding global temperature time series $T_{\text{global}}(t)$, smoothed with a 3-year running average to remove variability in global temperatures induced by El Niño–Southern Oscillation.

$$R'(x, y, t) = A(x, y) T'_{\text{global}}(t) + \epsilon(x, y, t) \quad (1)$$

The standardised difference z between simulated and observed changes at each grid box and for each model is then computed based on the respective trends per degree global warming A and their standard errors ΔA .

$$z = \frac{A_{\text{obs}} - \overline{A_{\text{mod}}}}{\sqrt{(\Delta A_{\text{obs}})^2 + (\Delta A_{\text{mod}})^2}/N} \quad (2)$$

The bar denotes the ensemble mean and N is the number of ensemble members (Online Resource, Tabs. S1 and S2). We compute the standard errors assuming *iid* residuals and correct these estimates for autocorrelated residuals using the effective sample size according to Wilks (1997). The standard errors describe the variability of trends in a stationary climate (Online Resource, Sec. S2).

As z approximately follows a standard normal distribution for trends derived from long time series, we can test for consistency of observed and simulated trends by testing the null hypothesis H_0 : “The standardised difference z is drawn from a standard normal distribution”. Tests on the local significance are performed with probability of a type-1 error α of 10 %. To address the problem of multiple testing, we adjust the critical p-values of local tests by controlling the ‘False Discovery Rate’ (Ventura et al. 2004).

Additionally we infer the distribution of z given we had a perfect model. For each model with multiple ensemble members, we use each individual simulation in turn as pseudo-observation and compare its trends with the trends of the ensemble mean

of the remaining simulations. We then pool the results from all CMIP3 and CMIP5 models and we estimate the distribution of resampled z and derived statistics given a perfect model from these 126 independent estimates from 36 models.

4 Results and discussions

The observed trend (per degree global warming) in seasonal near-surface temperature shows a warming which is strongest over northern hemisphere continents in boreal winter (DJF) and least over oceans, where the warming is generally less than the global average warming (Fig. 1a, b). Patches of local cooling can be found in the southern oceans and North Pacific. In contrast, most climate model simulations show less spatial diversity in the warming over the past 50 years (not shown).

Significant inconsistencies can be found in the majority of CMIP3 and CMIP5 models in the Indian Ocean and Indonesia, the Arctic, and north-western Africa and south-western Europe in boreal summer (JJA), and central Asia in DJF where models underestimate the observed warming (Fig. 1c–f). In addition, models do not reproduce the regional cooling or lack of warming over parts of the southern Ocean

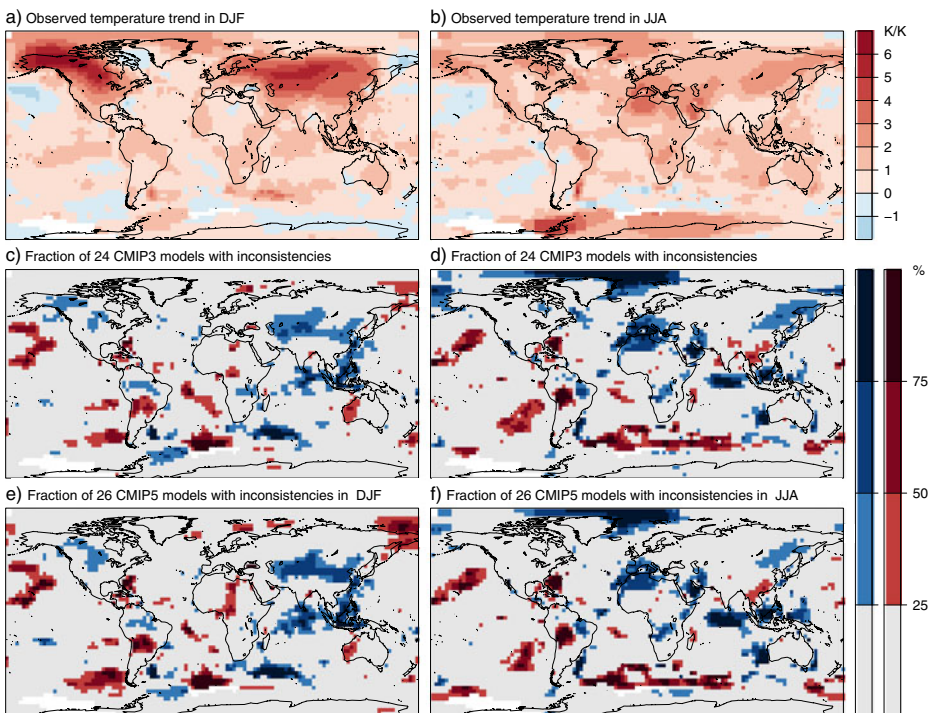


Fig. 1 Observed trend in DJF (a) and JJA (b) temperature per degree global warming (computed for 1956–2005) according to GISTEMP. Fraction of models with significant inconsistencies at the 10 % level for the CMIP3 (c, d) and CMIP5 (e, f) multi-model ensemble. The colour in c–f indicates the sign of the deviation of the majority of the models; *blue* corresponds to simulated warming smaller than the observed

and western Atlantic and the north-eastern and south-eastern Pacific. Shading in Fig. 1 starts at 0.25, corresponding to at least 6 (7) of the 24 (26) CMIP3 (CMIP5) models. Assuming that the models are independent draws from the real climate system—which they are most likely not, see Masson and Knutti (2011a) for a discussion of dependence in the CMIP3 multi-model ensemble—the likelihood of getting at least 6 significant (at the 10 % level) inconsistencies out of 24 random tries is 0.0075. Assuming that there are in fact only four independent models, the probability of getting at least 2 significant results (i.e. 50 %) is 0.05. Hence the results presented here are likely not random.

The identified patterns of inconsistencies are robust to the choice of time period. Such analysis indicated that the last decades dominate our findings, which is to be expected in the presence of emerging anthropogenic climate change.

The pattern of SLP changes in boreal winter (DJF, Fig. 2a) is almost zonally symmetric, with strong increases in high latitudes, strong decreases in the subtropics, and weakly increasing SLP in the tropics. In boreal summer (JJA, Fig. 2b), the area of decreasing SLP in the southern hemisphere extends further equatorwards and the

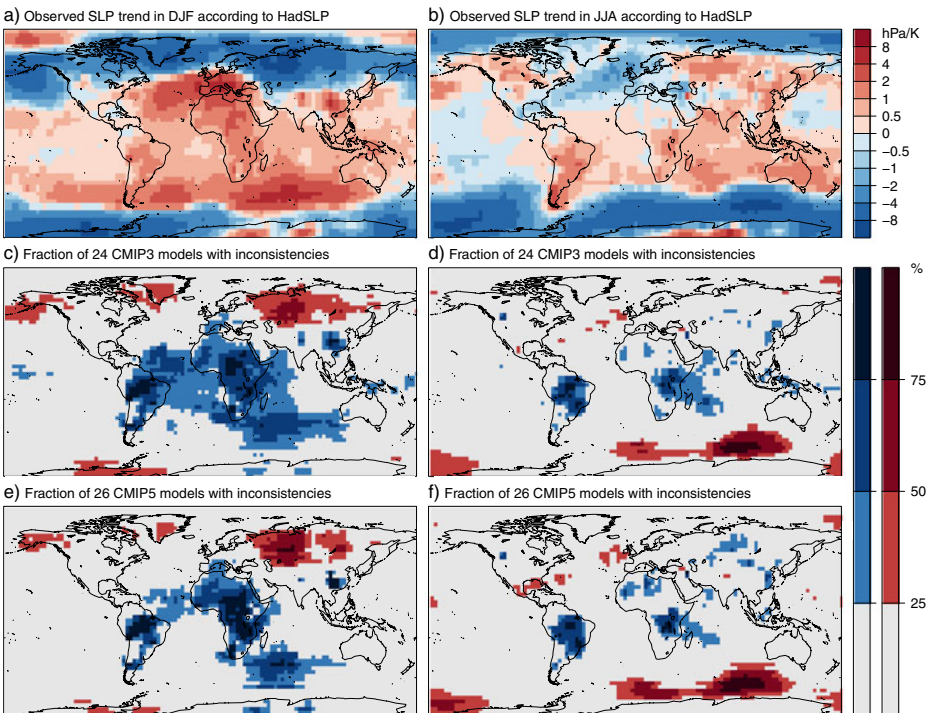


Fig. 2 Observed trend in SLP per degree global warming (computed for 1955–2004) in boreal winter (DJF, **a**) and summer (JJA, **b**) along with the fraction of models with significant inconsistencies in seasonal SLP trends for the CMIP3 (**c**, **d**) and CMIP5 (**e**, **f**) multi-model ensemble. The colour in **c**–**f** indicates the sign of the deviation of the majority of the models; *blue* corresponds to simulated changes smaller than the observed. The models with significant differences underestimate the magnitude of the observed trend

increases in the subtropics are less pronounced. In comparison, the models show less spatial variability and generally weaker trends (not shown). Consequently, the majority of the models significantly underestimate the magnitude of the observed decrease in SLP in parts of the high latitudes in the respective winter months (Fig. 2c–f). Additionally, most of the models underestimate the magnitude of the observed increase over Africa and tropical South America in DJF, and a smaller fraction of models also in the tropical Atlantic and the eastern Indian Ocean (Fig. 2c, e).

The identified inconsistencies in SLP are robust to choosing a different time period. Whereas significant underestimation of the observed magnitude of SLP change in high southern latitudes in austral winter (JJA) is found even in 30-year trends, inconsistencies in the tropics are mostly not significant for trend lengths shorter than 40 years. The general pattern of underestimation of the magnitude of the observed changes in SLP in the tropics and the high latitudes of the respective winter hemisphere, however, is consistent across time periods and in line with previous findings (Gillett and Stott 2009). In contrast to earlier studies, we find globally significant inconsistencies between simulated and observed changes in SLP.

The observed trend in seasonal precipitation from 1951 to 2000 according to VASCLimO is strongly variable in space and large in magnitude in some regions (more than ± 100 % per degree global warming). The simulated changes are considerably weaker but generally consistent with the observed change, except in boreal spring (MAM) when we find some coherent areas of inconsistencies shared across models (not shown). The lack of significant differences is mainly due to the low signal-to-noise ratio of externally forced changes in precipitation.

We compare the performance of individual models across regions using area-mean squared standardised differences z_{rms} . z_{rms} is a good proxy for the fraction of area with significant inconsistencies, with a correlation of 0.93 between the area fraction of significant inconsistencies and global z_{rms} for seasonal temperature across the CMIP3 models. In contrast to the area-fraction, z_{rms} also allows us to indicate whether the correspondence between simulations and observations is better than expected under H_0 resulting in $z_{\text{rms}} < 1$.

For seasonal temperature (Fig. 3a) and DJF SLP and precipitation changes (Fig. 3b), the global average z_{rms} for the majority of CMIP3 and CMIP5 models is larger than resampled z_{rms} (representative of z_{rms} given we had a perfect model) for the main observational dataset. The majority of the models is thus biased in its representation of regional changes at least in some regions. For SLP and precipitation changes in JJA (Fig. 3b, c) the global average z_{rms} based on the primary observation data set are not significantly different from resampled z_{rms} for most models in both multi-model datasets.

The respective ensemble means (Online Resource, Sec. S3) outperform the individual models in z_{rms} for temperature and precipitation. This is consistent with earlier findings that the ensemble mean performs better in a root mean square error sense (see Annan and Hargreaves 2011, for further discussion). For changes in SLP, on the other hand, we find that up to seven individual models outperform the respective ensemble mean (Fig. 3b).

The results for temperature and SLP are similar if 20CRv2 is used as observations, which is not surprising given that observed SLP is assimilated and observed sea surface temperature are used as boundary conditions in 20CRv2. For precipitation, however, global average z_{rms} based on 20CRv2 indicate strong trend biases in

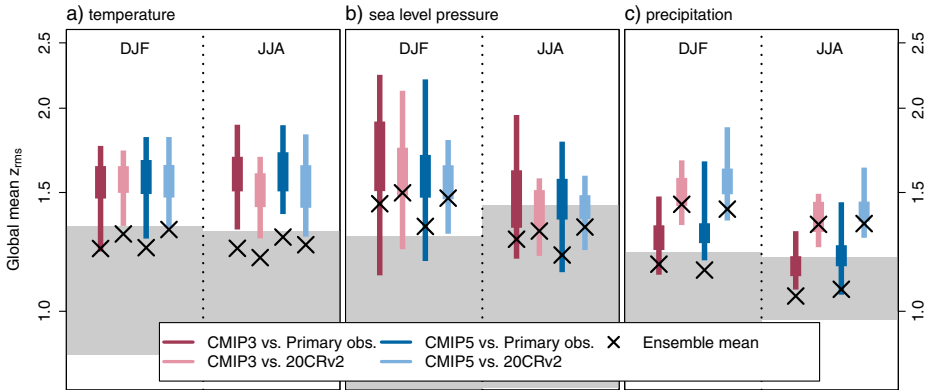


Fig. 3 Global average root mean squared standardised differences z_{rms} for near-surface temperature (a), SLP (b), and precipitation (c). Range (lines) and interquartile range (bars) of z_{rms} are shown in red for CMIP3 and blue for CMIP5 models. Results according to the primary observation dataset are shown in saturated colours, according to 20CRv2 in light colours. The light grey shaded areas denote the central 90th percentile of resampled z_{rms} and crosses denote z_{rms} for the respective multi-model ensemble mean

the models throughout the year. Precipitation in reanalyses is known to differ considerably from rain-gauge measurements (Pavelsky and Smith 2006), and the differences between the two observation-based datasets highlight the need to treat precipitation results presented here with caution.

Significant standardised differences between simulated and observed trends are due to either the simulated forced response differing from the true response, or the simulated variability being lower than the observed variability (and hence $\overline{(\Delta A_{\text{mod}})^2} < \overline{(\Delta A_{\text{obs}})^2}$). For temperature and SLP, models do not underestimate the observed variability and, consequently, we find as much or more inconsistency when replacing the simulated standard errors $\overline{(\Delta A_{\text{mod}})^2}$ with the observed $\overline{(\Delta A_{\text{obs}})^2}$ in Eq. 2. We find some indication that simulations tend to underestimate observed variability in precipitation, less so in CMIP5 compared to CMIP3.

With respect to the representation of the forced response, we analyse global mean z for all variables. We find no indication of significant global average trend biases except for a positive trend bias in DJF in the majority of models consistent with the large area of shared inconsistency in the tropics shown in Fig. 2. As in previous studies (Wentz et al. 2007; Zhang et al. 2007; Bhend and von Storch 2008; van Haren et al. 2012), our results indicate that simulations tend to underestimate the magnitude of the forced response in both SLP and precipitation as significant inconsistencies are found where models underestimate the magnitude of the observed change.

To assess whether CMIP3 or CMIP5—as an ensemble—better represent local trends in individual regions, we test the difference in the distribution of z_{rms} between CMIP3 and CMIP5 using the Wilcoxon rank sum test for 21 subcontinental regions (Fig. 4). We find no clear indication of difference in the ability of CMIP3 and CMIP5 to reproduce local observed trends. Similar but generally less significant differences are found between performance in CMIP5 and the smaller CMIP3 ensemble which includes both anthropogenic and natural forcing.

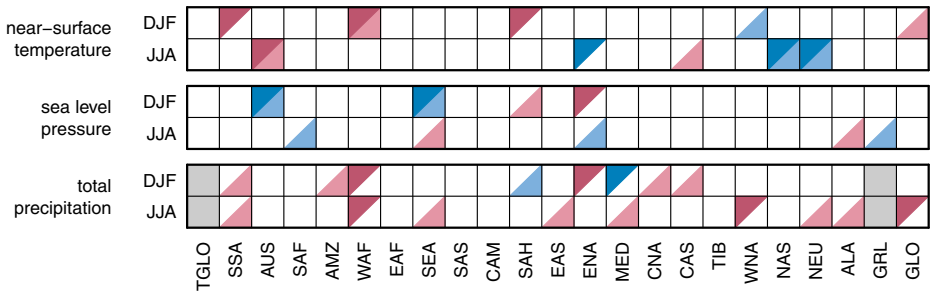


Fig. 4 Results of Wilcoxon rank sum tests on z_{rms} for various regions and variables and the two observational datasets. *Red* denotes cases in which CMIP3 significantly (two-sided test at the 10 % level) outperforms CMIP5, *blue* denotes cases in which CMIP5 significantly outperforms CMIP3. The *bold triangle* in the respective upper-left corner corresponds to the comparison with the primary observational dataset, the *light-shaded triangle* in the lower-right corresponds to the comparison with 20CRv2. The regions are global land and ocean (TGLO), 21 sub-continental regions (land only, see Giorgi and Francisco 2000) ordered from south to north and global land (GLO). There is no observation data for precipitation in Greenland (GRL) and over oceans

Even though the lack of progress from CMIP3 to CMIP5 might seem counter-intuitive at first, there are reasons why such an outcome could be expected. In comparison with CMIP3, models in CMIP5 are generally more complex in that a larger number of processes and forcing mechanisms are explicitly simulated (Taylor et al. 2011). Whether this contributes to more or less realistic representations of the local change is difficult to infer from the two ensembles at hand. To be able to attribute differences in the models’ ability to reproduce observed local trends to varying model complexity, however, controlled experiments with individual models are needed.

5 Conclusions and outlook

We present a globally comprehensive assessment of consistency of simulated and observed local trends in near-surface temperature, SLP, and precipitation per degree global warming. To avoid issues with model dependencies (Jun et al. 2008) and to be able to address inter-model differences, we chose to analyse the consistency of changes for individual models. Observational uncertainties are treated by conducting the analysis with two partly independent observation-based data sets, and we note that the observation errors might be large at high southern latitudes for all variables and globally for precipitation.

We find that all of the models analysed show significant inconsistencies between the observed and simulated changes in temperature and SLP. Each of the models therefore fails to reproduce the observed trend at least in some areas. Also areas with significant inconsistencies are congruent across a majority of models and the inconsistencies are robust to using two different observation-based datasets as the reference. This suggests that the climate models share biases in their representation of recent regional changes. Therefore, efforts which use past local changes to constrain climate projections should address biases shared across simulations (e.g. Sexton et al. 2011).

Simulated changes in precipitation, on the other hand, are generally not significantly different from the observed changes due to the low signal-to-noise ratio of externally forced changes in precipitation. However, the differences in trends across the two observation-based datasets used in this study are large. These results have two important implications for regional climate projections. First, quick and qualitative approaches to putting climate projections and past observed changes in precipitation into context will be misleading in most cases, as apparent differences between observed and projected precipitation changes might be due to internal variability. Second, the observed recent change in precipitation will likely not provide useful constraints for future projections as of yet and constraints for regional precipitation projections will have to come from other variables (e.g. regional rainfall climatology such as discussed in Whetton 2007 and applied in Smith 2010).

Reichler and Kim (2008) document considerable improvement from CMIP1 to CMIP3 when analysing the climatological mean state across different variables. Considering measures of climate change such as climate sensitivity, Andrews et al. (2012) find little change between CMIP3 and an early version of the CMIP5 ensemble even though improvements have been documented for individual models (Watanabe et al. 2010). For local changes, differences between CMIP3 and CMIP5—or the lack thereof—has not been documented so far. Here we find no improvement from CMIP3 to CMIP5 with respect to consistency of simulated local trends per degree warming in near-surface temperature, SLP, and precipitation with the observed change. That is, recent model development has not significantly altered our understanding and description of long-term regional change in these variables. Increasing complexity and realism of the models, on the other hand, seem to mainly affect the representation of the variability in the models which might change the significance of the results. Indeed progress in the description of variability has been documented already for ENSO (Guilyardi et al. 2012). Our analysis reveals, however, that the representation of interannual variability in the two multi-model ensembles has not changed sufficiently (and/or sufficiently systematically) to affect the consistency of observed and simulated local changes.

We emphasise that this consistency assessment is only one step in understanding regional changes. Here, we diagnose inconsistencies between simulated and observed local changes, but process studies and a more careful analysis of individual models and intermodel differences will be needed to explain why these inconsistencies occur. Furthermore, future studies will need to more rigorously address observation uncertainties as these contribute considerably to the overall uncertainties in some cases. Finally, it should be noted that the analysis has been performed on a subset of the eventual CMIP5 ensemble. Even though the complete set of models may offer additional insight into differences between CMIP5 and CMIP3, we do not expect the general conclusions and results presented here to change.

Acknowledgements We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling 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 (PCMDI) provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment

(DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office.

We thank David Kent and Tim Erwin for software to facilitate the analysis of large multi-model datasets, and we thank Janice Bathols for help with downloading the CMIP5 data.

References

- Allan R, Ansell T (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J Climate* 19(22):5816–5842
- Andrews T, Gregory JM, Webb MJ, Taylor KE (2012) Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys Res Lett* 39(9):L09,712
- Annan JD, Hargreaves JC (2011) Understanding the CMIP3 multimodel ensemble. *J Climate* 24(16):4529–4538
- Beck C, Grieser J, Rudolf B (2005) A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. *Climate Status Report 2004*, German Weather Service, pp 181–190
- Bhend J, von Storch H (2008) Consistency of observed winter precipitation trends in northern Europe with regional climate change projections. *Clim Dyn* 31(1):17–28
- Christensen J, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli R, Kwon WT, Laprise R, Magaña Rueda V, Mearns L, Menéndez C, Räisänen J, Rinke A, Sarr A, Whetton P (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chap Regional Climate Projections, pp 847–940
- Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Mauerer M, Mok HY, Nordli O, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The twentieth century reanalysis project. *Q J R Meteorol Soc* 137(654):1–28
- Ferguson C, Villarini G (2012) Detecting inhomogeneities in the twentieth century reanalysis over the central United States. *J Geophys Res* 117:D05123. doi:10.1029/2011JD016988
- Fronzek S, Carter T, Räisänen J, Ruokolainen L, Luoto M (2010) Applying probabilistic projections of climate change with impact models: a case study for sub-arctic Palsa Mires in Fennoscandia. *Climatic Change* 99:515–534. doi:10.1007/s10584-009-9679-y
- Gillett NP, Stott PA (2009) Attribution of anthropogenic influence on seasonal sea level pressure. *Geophys Res Lett* 36(23):L23,709
- Giorgi F, Francisco R (2000) Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Clim Dyn* 16:169–182. doi:10.1007/PL00013733
- Guilyardi E, Bellenger H, Collins M, Ferrett S, Cai W, Wittenberg A (2012) A first look at ENSO in CMIP5. *CLIVAR Exchanges* No 58 17(1):29–32
- Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. *Rev Geophys* 48(4):RG4004
- Jun M, Knutti R, Nychka D (2008) Spatial analysis to quantify numerical model bias and dependence. *J Am Stat Assoc* 103(483):934–947
- Karoly DJ, Wu QG (2005) Detection of regional surface temperature trends. *J Climate* 18(21):4337–4343
- Masson D, Knutti R (2011a) Climate model genealogy. *Geophys Res Lett* 38(8):L08,703
- Masson D, Knutti R (2011b) Spatial-scale dependence of climate model performance in the CMIP3 ensemble. *J Climate* 24(11):2680–2692
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE (2007) The WCRP CMIP3 multimodel dataset—a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394
- Pavelsky TM, Smith LC (2006) Intercomparison of four global precipitation data sets and their correlation with increased Eurasian river discharge to the Arctic Ocean. *J Geophys Res* 111(D21):D21,112
- Randall D, Wood R, Bony S, Colman R, Fichetef T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer R, Sumi A, Taylor K (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental*

- Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chap Climate Models and Their Evaluation, pp 589–662
- Räsänen J (2007) How reliable are climate models? *Tellus A* 59(1):2–29
- Reichler T, Kim J (2008) How well do coupled models simulate today's climate? *Bull Am Meteorol Soc* 89(3):303–311
- Sexton D, Murphy J, Collins M, Webb M (2011) Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Clim Dynam* 38(11–12):2513–2542. doi:[10.1007/s00382-011-1208-9](https://doi.org/10.1007/s00382-011-1208-9)
- Smith I, Chandler E (2010) Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance. *Climatic Change* 102:377–393. doi:[10.1007/s10584-009-9757-1](https://doi.org/10.1007/s10584-009-9757-1)
- Sterl A (2004) On the (in)homogeneity of reanalysis products. *J Climate* 17(19):3866–3873
- Taylor KE, Stouffer RJ, Meehl GA (2011) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93(4):485–498
- van Haren R, van Oldenborgh G, Lenderink G, Collins M, Hazeleger W (2012) SST and circulation trend biases cause an underestimation of European precipitation trends. *Clim Dynam* 40(1–2):1–20. doi:[10.1007/s00382-012-1401-5](https://doi.org/10.1007/s00382-012-1401-5)
- van Oldenborgh GJ, Drijfhout S, van Ulden A, Haarsma R, Sterl A, Severijns C, Hazeleger W, Dijkstra H (2009) Western Europe is warming much faster than expected. *Clim Past* 5:1–12
- Ventura V, Paciorek CJ, Risbey JS (2004) Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *J Climate* 17(22):4343–4356
- Watanabe M, Suzuki T, Oishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T, Hasumi H, Tatebe H, Kimoto M (2010) Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. *J Climate* 23(23):6312–6335
- Wentz FJ, Ricciardulli L, Hilburn K, Mears C (2007) How much more rain will global warming bring? *Science* 317(5835):233–235
- Whetton P, Macadam I, Bathols J, O'Grady J (2007) Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophys Res Lett* 34(14):L14701. doi:[10.1029/2007GL030025](https://doi.org/10.1029/2007GL030025)
- Wilks DS (1997) Resampling hypothesis tests for autocorrelated fields. *J Climate* 10(1):65–82
- Yokohata T, Annan J, Collins M, Jackson C, Tobis M, Webb M, Hargreaves J (2011) Reliability of multi-model and structurally different single-model ensembles. *Clim Dynam* in print, pp 1–18. doi:[10.1007/s00382-011-1203-1](https://doi.org/10.1007/s00382-011-1203-1)
- Zhang X, Zwiers F, Hegerl GC, Lambert FH, Gillett NP, Solomon S, Stott PA, T N (2007) Detection of human influence on twentieth-century precipitation trends. *Nature* 448:461–465