nature climate change

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NCLIMATE1972

Overestimated global warming in the past 20 years

John C. Fyfe[1], Nathan P. Gillett[1], Francis W. Zwiers[2]

CMIP5 Models

The 37 CMIP5 models used, with the number of runs available in parenthesis, are: ACCESS1-0(1), ACCESS1-3(1), BNU-ESM(1), CCSM4(6), CESM1-BGC(1), CESM1-CAM5(3), CMCC-CM(1), CMCC-CMS(1), CNRM-CM5(10), CSIRO-Mk3-6-0(10), CanESM2(5), FGOALS-s2(3), FIO-ESM(3), GFDL-CM3(5), GFDL-ESM2G(1), GFDL-ESM2M(1), GISS-E2-H(5), GISS-E2-H-CC(1), GISS-E2-R(6), GISS-E2-R-CC(1),HadCM3(10), HadGEM2-AO(1), HadGEM2-CC(1), HadGEM2-ES(1), IPSL-CM5A-LR(6), IPSL-CM5A-MR(3), IPSL-CM5B-LR(1), MIROC-ESM(3), MIROC-ESM-CHEM(1), MIROC5(5), MPI-ESM-LR(3), MPI-ESM-MR(3), MRI-CGCM3(3), NorESM1-M(3), NorESM1-ME(1), bcc-csm1-1(3) and bcc-csm1-1-m(3). All the CMIP5 model output was formed by merging historical simulations up to 2005 with RCP4.5 simulations from 2006 to 2012. RCP4.5 simulations included greenhouse gas concentrations very similar to those observed over this period and aerosol precursor emissions broadly consistent with best observational estimates. Observed solar irradiance variations were specified up until 2008 followed by a repeating solar cycle which approximately reproduces observed solar irradiance variations over this period[1].

[1] Canadian Centre for Climate Modelling and Analysis, Environment Canada, Victoria, BC V8W 2Y2, Canada

[2] Pacific Climate Impacts Consortium, University of Victoria, Victoria, BC, V8W 3R4, Canada

Statistical Analysis

In our analysis two statistical representations for trends are estimated either from observations or from individual model simulations. The forms of these representations depend on the assumptions concerning the “exchangeability” of information between the models and the observations.

Trend representation assuming exchangeability between models

In this case a reasonable trend representation is:

(1) $M_{ij} = u^m + Eint_{ij} + Emod_i,\ i = 1,...,N^m,\ j = 1,...,N_i$ and

(2) $O_k = u^o + Eint^o + Esamp_k,\ k = 1,...,N^o$.

Here M_{ij} and O_k are trends calculated from single runs, or single bootstrap samples of the observations. u^m and u^o are the true, unknown, deterministic trends due to external forcing in the modelled and observed worlds, where u^m is the trend component that is common to all models (in the limit, as the collection of exchangeable models grows infinitely large). No prior assumption that $u^m = u^o$ is made. $Eint_{ij}$ and $Eint^o$ are perturbations to M_{ij} and O_k respectively due to internal variability. These are different for each model run, but are essentially identical for each resample of the observations. $Emod_i$ is the perturbation to M_{ij} that is introduced by model error in model *i*. We assume that these perturbations are exchangeable. $Esamp_k$ is the perturbation to O_k introduce by the *k*-th bootstrap resampling. N^m is the number of models, N_i is the size of the ensemble for model *i*, and N^o is the number of observed reconstructions.

Trend representation assuming exchangeability between models and observations

In this case a reasonable trend representation is:

(1)′ $M_{ij} = u^m + Eint_{ij} + Eexch_i$, $i = 1,...,N^m$, $j = 1,...,N_i$ and

(2)′ $O_k = u^o + Eint^o + Eexch_{N^m+1} + Esamp_k$, $k = 1,...,N^o$.

Here $Eexch_{N^m+1}$ is the deviation from the true, underlying (forced) trend in the observations that is exchangeable with the deviations $Eexch_i$, $i = 1,...,N^m$ of the models from the common component of the model trend. That is, the deviations are assumed to be exchangeable, but the two true trends u^m and u^o need not be the same. The other components are defined as above.

Null hypothesis

For either trend representation, a reasonable estimator of $u^o - u^m$ is $O_{.} - M_{..}$, where "." replacing a subscript indicates averaging over that subscript. In both cases, the null hypothesis H_0 : $u^m = u^o$ could be tested in a number of ways. One approach would be to make distributional and independence assumptions for the individual, non-deterministic components of (1) and (2), or (1)′ and (2)′, and subsequently derive a distribution for $O_{.} - M_{..}$ under H_0. Instead, we opt for a resampling approach, thereby avoiding distributional assumptions. In this approach, the equations above are used a guidance to ensure that the empirical distribution for $O_{.} - M_{..}$ includes the sources of uncertainty described in either equations (1) and (2), or (1)′ and (2)′.

Empirical distribution assuming exchangeability between models

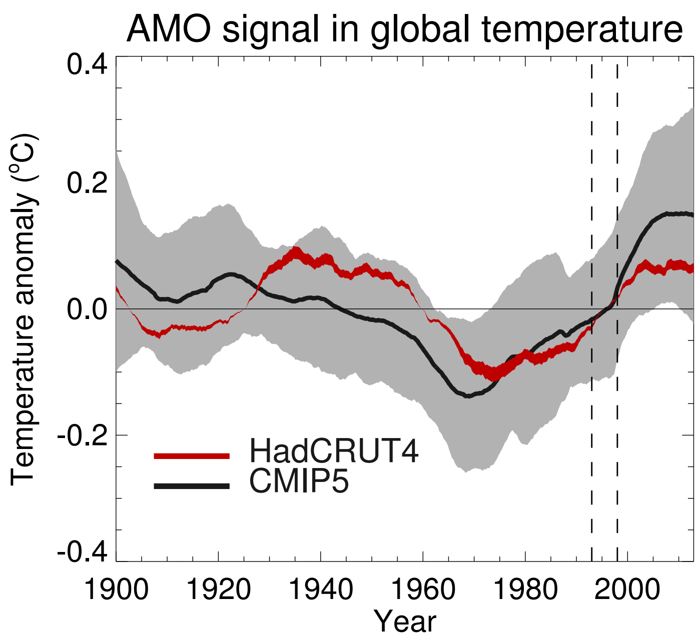
In this case an empirical estimate of the distribution for $O_{.} - M_{..}$ under H_0 is constructed as follows: a) Select a sample of N^o observed reconstructions with replacement and average to obtain a $O_{.}$ realization. b) Select a sample of N^m models with replacement and for each selection, draw one run at random from that model's available ensemble of simulations, and then average over those N^m runs to obtain a $M_{..}$ realization. c) Select, at random, a single model *i* from models with multi-run ensembles, and then select, at random, a single run *j* from that model's ensemble. Calculate the difference $M_{ij} - M_{i.}$ between the trend in that single run and the mean of the trends from that model's ensemble. This difference is an estimate of the deviation in the *j*-th trend for model *i* that is induced by internal variability. Since the model *i* ensemble is generally small, the deviations are smaller than would be representative of an infinitely large replication of runs for model *i,* and so to compensate for that loss of variance, multiply the difference $M_{ij} - M_{i.}$ by $[N_i/(N_i - 1)]^{1/2}$. Finally, calculate *a-b+c* computed in the steps above and repeat the resampling procedure many times to build a distribution for *a-b+c* under the null hypothesis that $u^o - u^m = 0$ and the assumption of model exchangeability. From this distribution we compute p-values where a p-value is the probability of occurrence of a trend at least as large as that found under the null hypothesis of equal underlying trends. Here we note that the smaller the p-value the stronger the evidence against the null hypothesis.

The rationale for this procedure is a follows: 1) Variations in $O_{.}$ from step a) represent the effect of sampling uncertainty in the observational estimate of u^{o} . That is, the uncertainty that results from the $Esamp_k$ term in equation (2). 2) Variations in $M_{..}$ from step b) represent the uncertainty in the multimodel estimate of u^m that arises from the choice of exchangeable models used to obtain $M_{..}$ [i.e., from the $Emod_i$ term in equation (1)] and from internal variability [i.e., from the $Eint_{ij}$ term in equation (1)]. Step (c) is used to estimate the uncertainty in $O_{.}$ that arises from internal variability. Since a single realization of internal variability is confounded with $O_{.}$, c) is constructed by estimating single realizations of internal variability as they were realized in models. This can only be done using models with multi-member ensembles. An implicit assumption is that sampling uncertainty in $O_{.}$ is independent of uncertainty due to internal variability and also independent of uncertainty in $M_{..}$.

Trend distribution assuming exchangeability between models and observations

In this case an empirical estimate of the distribution for $O_{.} - M_{..}$ under H_0 is constructed as follows: a) As above. b) As above. c) Select a sample of $N^m + 1$ entities with replacement from the pool of $N^m + 1$ entities consisting of N^m models plus the observations (as the additional entity in the pool). For each member of the sample of entities, draw an ensemble member at random from that entity's available ensemble. From these entities $P_i, i = 1,...,(N^m + 1)$ calculate $P_{.}$ and then select a single P_i at random and calculate $P_i - P_{.}$. Finally, calculate *a-b+c* as computed in the steps above and repeat the resampling procedure many of times to build a distribution for *a-b+c* under the null hypothesis that $u^o - u^m = 0$ and the assumption of exchangeability between models and observations. The rationale for this approach is the same as above, except component c) now includes uncertainty from two sources. From this distribution we again compute p-values that assess how unusual the discrepancy is between the observed and mean of model trends.

Note that the p-values are necessarily dependent upon the assumptions that are used to construct reference distributions for $O_{.} - M_{..}$ under the null hypothesis. The assumption of exchangeable between models and observations, which is a stronger assumption than model exchangeability, sets a more stringent criterion for rejecting the null hypothesis at a given significance level (i.e., the discrepancy between observed and mean model trends needs to be larger) than the model exchangeability assumption.



Supplementary Figure 1 | Atlantic Multidecadal Oscillation (AMO) signal in simulated and observed global mean surface temperature. The AMO index was computed from monthly mean sea surface temperature averaged over the North Atlantic (i.e. 25°N-60°N and 75°W-7°W) with 1900-2012 trends removed and smoothed with a sliding 121-month average. ENSO, COWL and volcanic signals[2-3] were removed from the global mean surface temperature, which was then regressed against the AMO index to give the timeseries shown above. The 2.5-97.5% range of observed estimates is shown with red shading and the 2.5-97.5% range of simulated estimates is shown with grey shading. The years 1993 and 1998 are indicated with vertical dashed lines.

References

1 Van Vuuren, D.P., Edmonds, D., Kainuma, J., Riahi, M. & Weyant, J. *Climatic Change*, **109**, doi: 10.1007/s10584-011-0148-z (2011).
2 Thompson, W.J., Wallace, J.M., Jones, P.D. & Kennedy, J.J. *J. Climate*, **22**, 6120–6141 (2009).
3 Fyfe, J.C., Gillett, N.P. & Thompson, D.W.J. *Geophys. Res. Lett.*, **37**, L16802, doi: 10.1029/2010GL044255 (2010).