Designing detection and attribution simulations for CMIP6 to optimize the estimation of greenhouse gas–induced warming

Aurélien Ribes, Nathan P. Gillett, & Francis W. Zwiers

2015

Pacific Climate Impacts Consortium (PCIC)

PCIC Publications

Original citation:

Ribes, A., Gillett, N. P., & Zwiers, F. W. (2015). Designing detection and attribution simulations for CMIP6 to optimize the estimation of greenhouse gas–induced warming. *Journal of Climate*, *28*(8), 3435–3438. https://doi.org/10.1175/JCLI-D-14-00691.1

Designing Detection and Attribution Simulations for CMIP6 to Optimize the Estimation of Greenhouse Gas–Induced Warming

AURÉLIEN RIBES

CNRM-GAME, Météo France–CNRS, Toulouse, France

NATHAN P. GILLETT

CCCma, Environment Canada, Victoria, British Columbia, Canada

FRANCIS W. ZWIERS

PCIC, University of Victoria, Victoria, British Columbia, Canada

(Manuscript received 13 October 2014, in final form 13 January 2015)

ABSTRACT

Climate change detection and attribution studies rely on historical simulations using specified combinations of forcings to quantify the contributions from greenhouse gases and other forcings to observed climate change. In the last CMIP5 exercise, in addition to the so-called all-forcings simulations, which are driven with a combination of anthropogenic and natural forcings, natural forcings–only and greenhouse gas–only simulations were prioritized among other possible experiments. This study addresses the question of optimally designing this set of experiments to estimate the recent greenhouse gas–induced warming, which is highly relevant to the problem of constraining estimates of the transient climate response. Based on Monte Carlo simulations and considering experimental designs with a fixed budget for the number of simulations that modeling centers can perform, the most accurate estimate of historical greenhouse gas–induced warming is obtained with a design using a combination of all-forcings, natural forcings–only, and aerosol forcing–only simulations. An investigation of optimal ensemble sizes, given the constraint on the total number of simulations, indicates that allocating larger ensemble sizes to weaker forcings, such as natural-only, is optimal.

Climate change detection and attribution (D&A) studies rely on historical simulations using specified combinations of forcings from a broad set of climate models to quantify the greenhouse gas and other contributions to observed climate change. In the last CMIP5 exercise (Taylor et al. 2012), in addition to the all-forcing simulations (ALL), natural-only (NAT), and greenhouse gas–only (GHG) simulations were prioritized among other possible experiments. However, the question of optimally designing this set of experiments to tackle some specific scientific questions has not been explicitly addressed.

Here we compare four possible strategies: ALL+GHG+NAT experiments as in CMIP5, ALL+AER+NAT, ALL+AER+GHG, and AER+GHG+NAT, where AER represents the response to aerosols. To make this comparison, we focus on how accurately each strategy will enable us to estimate the past GHG-induced warming, which is very closely related to the question of providing an observationally based constraint on the transient climate response (TCR).

Most D&A studies rely on a regression-based statistical model (Allen and Stott 2003), where observations Y are regressed onto the responses to individual forcings X_i^*,

$$Y = \beta_{\mathrm{AER}} X_{\mathrm{AER}}^* + \beta_{\mathrm{GHG}} X_{\mathrm{GHG}}^* + \beta_{\mathrm{NAT}} X_{\mathrm{NAT}}^* + \varepsilon_Y, \quad (1)$$

where ε_Y represents internal (unforced) variability in the climate system, and is assumed to follow a Gaussian $N(0, \Sigma)$ distribution. It is then assumed that climate model simulations provide noise contaminated realizations X_i

Corresponding author address: Aurélien Ribes, CNRM-GAME, Météo France–CNRS, 42 Avenue Gaspard Coriolis, Toulouse, France.
E-mail: aurelien.ribes@meteo.fr

DOI: 10.1175/JCLI-D-14-00691.1

of X_i^*, where the "noise" reflects internal variability within the climate model

$$X_i = X_i^* + \varepsilon_{X_i}. \qquad (2)$$

Similarly to ε_Y, ε_{X_i} is assumed to follow a Gaussian $N(0, \Sigma/N_i)$ distribution, where N_i is the number of simulations with forcing i. Note that in (2), the simulated responses to the individual forcings may be inferred using any of the four sets of simulations discussed above under the assumption of linear additivity (i.e., $X_{\mathrm{ALL}}^* = X_{\mathrm{AER}}^* + X_{\mathrm{GHG}}^* + X_{\mathrm{NAT}}^*$); for our sensitivity analysis we assume that forcings other than GHG, AER, and NAT have little impact. Within this statistical framework, the estimation of the GHG-induced warming is related to the estimation of the scaling factor β_{GHG}. Its accuracy depends on the extent to which internal variability affects the simulated responses in (2). The design of the simulation exercise may thus be optimized to alleviate this issue. Note that the discussion to this point does not reflect differences in model response due to differences in model implementation. This will be discussed further below.

Here we use Monte Carlo simulations (MCSs) to address this question. While the real world has only one realization, such MCSs allow us to generate a large sample of virtual observations and model responses under controlled assumptions. These are used to assess the accuracy of the estimation of β_{GHG} under each possible strategy. MCSs are implemented as follows. First, we fix realistic values of the parameters X_i^* and Σ; the values used are discussed below. Second, based on (1), we generate a large sample of virtual observations Y. Based on (2), we also generate a large sample of virtual responses X_i, which is a way to emulate climate models. Third, we apply standard D&A analysis to estimate the scaling factor β_{GHG} from these simulated Y and X_i values. Fourth, we measure how close these estimates are to its *true* value, which is known because it was set to unity in the MCSs.

Realistic values of the *true* responses to individual forcings X_i^* as well as the *true* covariance matrix of internal variability Σ are required to run such MCSs. For the former, we use the mean responses simulated by an ensemble of two models, CanESM2 and CSIRO Mk3.6.0. We use these models because they provided ensembles (five members) of ALL, GHG-only, AER-only, and NAT-only simulations over the considered period, which is 1901–2010. These responses are shown in Fig. 1 in terms of global mean temperature. Note that these two models are more sensitive to AER forcings than the CMIP5 multimodel mean. We use a covariance matrix inferred from a broad set of both control simulations and ensembles of historical runs with perturbed initial conditions, as in Ribes et al. (2013). This provides an estimate of Σ with lower uncertainty than if only based on control simulations from CanESM2 and CSIRO Mk3.6.0. The resulting multimodel estimate of internal variability has been shown to be suitable for D&A even if used with many different response patterns (Ribes and Terray 2013). We ran $N = 2000$ MCSs for each considered configuration of ALL, GHG, AER, and NAT ensembles.

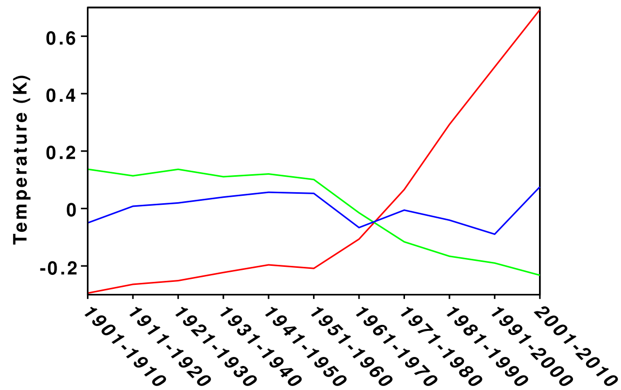

FIG. 1. Mean response to GHG-only (red), AER-only (green), and NAT-only (blue) forcings as simulated by the CanESM2 and CSIRO Mk3.6.0 models (multimodel average), in terms of global mean temperature averaged over decades (anomalies with respect to 1901–2010).

Our analysis is based on decadal averages of 1901–2010 global temperature projected onto T2-spherical harmonics. The covariance matrix of internal variability is assumed to be known (i.e., uncertainty in its estimation is not considered). A more detailed description of such MCSs is provided in Ribes et al. (2013; see their Fig. 1, black lines). Finally, we assume a phase 6 of CMIP (CMIP6) experimental design with a "budget" of 25 historical climate change simulations that can be allocated to different numbers of simulations with different combinations of forcings. We assume that such a design would require modeling groups to perform 10 ALL simulations, as has been proposed, and that a total number of 15 additional experiments could be requested with different combinations of forcings, consistent with the number of historical simulations that individual modeling centers were able to provide for CMIP5.

Results from the MCSs are shown in Fig. 2a in terms of the square root of the 90% quantile of squared errors (Rq90SE) on β_{GHG}; these values can be regarded as the typical expected half-width of the 90% confidence intervals. The figure illustrates how each strategy and each possible allocation of the 15 single forcing runs make the estimation of β_{GHG} more or less accurate. The best (i.e., minimum) value indicates the optimal ensemble sizes,

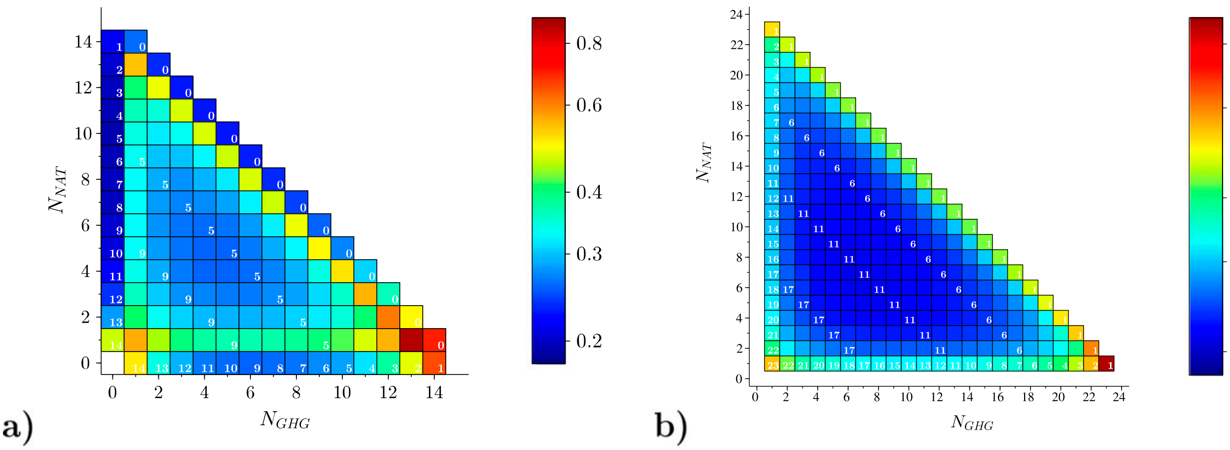

FIG. 2. Accuracy (Rq90SE) of the estimation of β_{GHG} as a function of the strategy and sizes of the GHG-only and NAT-only ensembles (N_{GHG} and N_{NAT}, respectively). (a) Here the total number of single-forcing experiments is fixed to $N_{\mathrm{AER}} + N_{\mathrm{GHG}} + N_{\mathrm{NAT}} = 15$, while $N_{\mathrm{ALL}} = 10$ ALL-forcing experiments are assumed to be available. The number of aerosol-only experiments N_{AER} is constant on diagonals, and is mentioned in a few particular boxes (white). The ALL+AER+NAT strategy (i.e., $N_{\mathrm{GHG}} = 0$) is shown on the left-side edge, the ALL+AER+GHG strategy (i.e., $N_{\mathrm{NAT}} = 0$) on the bottom edge, the ALL+GHG+NAT strategy (i.e., $N_{\mathrm{AER}} = 0$) on the main diagonal (hypotenuse), and the AER+GHG+NAT strategy (i.e., N_{AER}, N_{GHG}, $N_{\mathrm{NAT}} > 0$) on the interior of the triangle. In the latter case, only 15 runs are considered (i.e., the ALL runs are not used). (b) Here the total number of experiments is still fixed to 25, but $N_{\mathrm{ALL}} = 0$, which allows $N_{\mathrm{AER}} + N_{\mathrm{GHG}} + N_{\mathrm{NAT}} = 25$. Only the strategy AER+GHG+NAT is considered in this case.

either under a given strategy, or overall. The ALL+AER+GHG+NAT strategy is not considered here, as it cannot be treated with the standard statistical model described in (1) and (2). Therefore, in the AER+GHG+NAT strategy (interior of the triangle in Fig. 2a), ALL simulations are not used, so that only 15 runs are considered in the D&A analysis, while 25 runs are used in the three other strategies (edges of the triangle). For this reason, the comparison may be regarded as unequal. As historical simulations including ALL forcings will certainly be performed regardless of the design of simulations specific to D&A, this corresponds to a realistic situation. However, in order to also make a more equal comparison, we show in Fig. 2b results obtained with the AER+GHG+NAT strategy assuming no historical simulation is performed ($N_{\mathrm{ALL}} = 0$), and the same total number of experiments are available (i.e., $N_{\mathrm{AER}} + N_{\mathrm{GHG}} + N_{\mathrm{NAT}} = 25$).

Over the four strategies considered, the highest accuracy (i.e., lowest Rq90SE) in estimating β_{GHG} is found under the ALL+AER+NAT strategy. This main result is robust to considering other preprocessing (e.g., spatial resolutions other than T2 spherical harmonics), or another criterion (e.g., the accuracy in estimating β_{AER} or β_{NAT} instead of β_{GHG}). The global minimum is reached for $N_{\mathrm{NAT}} = 10$ and $N_{\mathrm{AER}} = 5$, but the response function is quite flat from $N_{\mathrm{NAT}} = 7$ to $N_{\mathrm{NAT}} = 12$, suggesting some flexibility in the balance between NAT and AER simulations when allocating the 15 non-ALL forcing runs. Based on our criterion, ALL+AER+NAT is found to be 15% more accurate than the previously used ALL+GHG+NAT strategy, which represents a substantial improvement in accuracy. The two other strategies, ALL+AER+GHG and AER+GHG+NAT, are found to lead to lower accuracy than the ALL+GHG+NAT strategy. The accuracy obtained with the AER+GHG+NAT strategy is improved if 25 runs are used (Fig. 2b) instead of only 15 (Fig. 2a), but even so this strategy is less accurate than the ALL+AER+NAT strategy. As expected, the accuracy of the estimation of β_{GHG} strongly deteriorates when the size of at least one of the ensembles considered is small (e.g., 1 or 2 members only). The only exception to this rule is found in the upper-left corner of Fig. 2a, where the accuracy is still relatively high with $N_{\mathrm{NAT}} = 14$, $N_{\mathrm{ALL}} = 10$, and $N_{\mathrm{AER}} = 1$ or $N_{\mathrm{GHG}} = 1$. This result suggests that properly estimating the NAT response is a critical issue.

Overall, these results suggest that accurate estimation of the signals with the smallest amplitudes (primarily NAT-only, then AER-only) is critical for D&A, consistent with the fact that the estimation of smaller-amplitude signals is more affected by internal variability than that of larger signals. The conclusion that explicit modeling of the AER response should be prioritized over modeling of the GHG response is expected to be robust to uncertainty as to the *true* AER response, as we based our analysis on a simulated response to aerosols that was particularly strong. Considering a lower AER response could lead to a slightly increased N_{AER} (and decreased N_{NAT}).

This work suggests that an ALL+AER+NAT strategy be adopted for the upcoming CMIP6 D&A exercise. It also suggests that a large NAT ensemble is helpful: this would also be useful for studies that only seek to separate anthropogenic and natural responses. Nevertheless, a few limitations should be mentioned. First, our brief study is based on a single criterion (estimation of TCR) and a single method (standard D&A analysis). Results may also be influenced by the number of runs considered (here 25, with 10 runs allocated to ALL-forcing runs). Second, the study was conducted under a perfect model assumption, which means that modeling uncertainty was not taken into account. A more sophisticated statistical framework would have to be considered to account for such uncertainty. Third, forcings other than GHG, AER, and NAT have been neglected. Fourth, considerations other than D&A might be taken into account to design this set of experiments—for example, related to our understanding of the physical processes involved in the response to each forcing, our confidence in the specifications of the different forcings, or the technical difficulty in carrying out consistent individual forcing simulations in a range of models with different levels of complexity and different processes represented.

Acknowledgments. AR gratefully acknowledges support from the Canadian Sea Ice and Snow Evolution network (CanSISE; http://www.cansise.ca), PCIC, and CNRM that facilitated a recent research visit to CCCMA and PCIC. The authors are grateful to Slava Kharin and Qiuzi Wen for useful comments on a previous version of this manuscript.

REFERENCES

Allen, M., and P. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.,* **21,** 477–491, doi:10.1007/s00382-003-0313-9.

Ribes, A., L. Terray, and S. Planton, 2013: Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate Dyn.,* **41,** 2817–2836, doi:10.1007/s00382-013-1735-7.

——, and ——, 2013: Application of regularised optimal fingerprinting to attribution. Part II: Application to global near-surface temperature. *Climate Dyn.,* **41,** 2837–2853, doi:10.1007/s00382-013-1736-6.

Taylor, K., R. Stouffer, and G. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.