Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-74-AC3, 2016

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Detection and Attribution Model Intercomparison Project (DAMIP)" *by* Nathan P. Gillett et al.

Nathan P. Gillett et al.

nathan.gillett@canada.ca

Received and published: 16 August 2016

We thank the reviewer for his or her helpful comments. We have revised the manuscript accounting for these comments.

General comments. I fear, with respect to resources required, that the proposal is somewhat over ambitious. Just for Tier 1 the number of experiments and initial condition ensembles required, >1870 model years, may substantially limit the number of models participating in DAMIP. ScenarioMIP (O'Neill GMDD [2016]) say "the success of ScenarioMIP lies in the broad participation of the CMIP6 modelling groups in Tier 1 experiments...". Should the DAMIP plan have the same ambition? A much smaller population of experiments/ensembles for a 'Tier 0', to focus on a few important scientific questions, could encourage as

wide a range of models to take part in DAMIP as possible. The remaining tiers can then be populated by institutions with more resources. It would be a shame to miss the opportunity to design experiments that would encourage greater institution involvement than there was for CMIP5 detection analyses. A 'lessons learned' exercise and finding out why some institutions didn't produce 'detection' experiments for CMIP5 might have been helpful.

We respect the reviewer's opinion, but in our view including at least the three Tier 1 experiments described is a minimum requirement for participation in DAMIP. Ensembles of size one are of limited use for attribution, and we really feel that histNAT, histGHG and histAER are required at Tier 1. It is more important to have a complete set of simulations from a slightly more limited set of models, than to have small ensembles and individual experiments from a broader range of models. Initial consultations with modelling groups indicated anticipated participation from at least as many modelling groups as carried out equivalent simulations in CMIP5.

There are several experiments that were not included in the original DAMIP proposal circulated within the CMIP community i.e., tier 3 experiments - histCO2, histSOL and ssp245NAT. The inclusion of these experiments were also not discussed with other scientists at the IDAG (International Detection and Attribution Group) meeting held in February this year. The motivation for including these experiments should have wider community discussions, as it is not really clear how useful they are [See below specific comments about those experiments]. They may be done at the expense of more useful experiments.

The document referred to above was a proposal, and the CMIP6 panel rules indicated that additional experiments could be added at Tiers 2 or 3 prior to preparation of the MIP description paper. Note that the experiments described above were all added at the lowest priority Tier 3 level, so whether or not to carry out these experiments is at the discretion of the modelling groups, and their presence is unlikely to discourage a modelling group from taking part in DAMIP. Taking each of the additional experiments referred to in turn: histSOL was added to DAMIP after SolMIP was merged with DAMIP by the CMIP6 panel. ssp245NAT was added to the experimental design when it became clear that ScenarioMIP simulations would include time-varying solar and volcanic forcings (O'Neill et al., 2016). We had previously assumed that natural forcings would be held constant in the SSP simulations. These simulations are required in order to separate the effects of projected changes in anthropogenic emissions on future climate from the effects of projected changes in natural forcings. They are also required for observationally-constrained projections using the Allen, Stott and Kettleborough approach in which natural forcings are treated separately. histCO2 simulations were added after consultation among the authors in order to improve constraints on TCR and TCRE as described. HistSOL was presented at IDAG, though the referee is correct that the other two experiments were added after the IDAG meeting.

More is needed to be said about the general type of analyses expected. In particular what is required for analyses using multi-model mean is not the same as what is needed for analyses on individual models. The ensemble size, especially for forcing factors with relatively weak response patterns, is much more important for analyses on individual models. As mentioned below, the recommendation of at least 3 ensemble members for histNAT, histSOL and histVOL etc. is not ideal for many types of analyses on individual models. However it may be more than sufficient for a multi-model mean analysis, where the total number of ensemble members will be much larger (e.g., Gillett et al, Journal of Climate [2013] and Jones et al, Journal of Geophysical Research [2013] both included both types of analysis using CMIP5 models).

As described at the end of Section 1, we do not include a complete analysis plan because the detection and attribution community is already well-established, and its activities are already coordinated through IDAG, and to address the long-standing interest in attribution in the IPCC reports. Nontheless, anticipated analyses are described in paragraphs 4-6 of Section 1. Additional details on the usage of the histAER simula-

tions, and the usage of DAMIP simulations for event attribution have now been added to these paragraphs. We anticipate that DAMIP simulations will be used both for individual model attribution studies and multi-model attribution studies. The referee is correct that large ensemble sizes are more important for single model analyses than those based on multiple models. Nonetheless, approaches which attempt to account for model uncertainty would benefit from the availability of initial condition ensembles in order to more robustly separate uncertainties in the response patterns associated with internal variability from those associated with model uncertainty (e.g. Hannart et al., Geophys. Res. Lett., 2014). We thus prefer to retain a minimum ensemble size of three for the historical simulations.

Specific comments Page 3 Line 10 and elsewhere. As explained in the manuscript, specific analysis plans are not given, but it would be helpful to indicate how the contributions to observed climate changes from well mixed greenhouse gases and also aerosols can be extracted with the proposed tier 1 experiments. It is mistakenly stated that previous studies separated contributions from WMGHGs and aerosols (Page 3 lines 12-13 and elsewhere). Those studies actually separated contributions from WMGHGS and other (non WMGHG) anthropogenic factors, as well as natural influences. It may seem pedantic, but in fact the contributions from ozone and land use changes are potentially substantial important for some diagnostics and should not be excluded if possible.

To address the reviewer's comment regarding how the Tier 1 simulations may be used, the following text has been added on page 3: 'Such aerosol-only simulations may be used together with historical simulations including all forcings and historical simulations with natural forcings only to estimate attributable contributions to observed changes due to natural forcings, due to aerosols, and due to the combined effects of well-mixed greenhouse gases, ozone and land-use changes. Since some part of the greenhouse gas changes is associated with land-use change, and since ozone is a greenhouse gas, grouping these forcing together arguably makes more sense than grouping ozone and land-use change with aerosols.' As stated in the description of histGHG, 'histALL, histNAT and histGHG will allow the attribution of observed climate change to natural, greenhouse gas and other anthropogenic components.' Note that using all four Tier 1 experiments described here it would also be possible to carry out an attribution analysis of observed changes to natural forcings, to WMGHGs, to aerosols, and to ozone and land-use change combined. However, we expect that regression coefficients would likely be unconstrained in such a four-pattern regression, therefore we do not propose it here.

The text referred to on page 3 has been revised to read: 'While some earlier studies were able to clearly separate the influences of greenhouse gases and other anthropogenic forcings on observed temperature changes using individual models (Stott et al., 2006), more recent studies using newer models and a longer period of observations have identified substantial uncertainties in the separate estimation of greenhouse gas and other anthropogenic contributions, where the other anthropogenic contribution is dominated by aerosols but also includes the response to ozone changes and land use changes in most models (Jones et al., 2013; Gillett et al., 2013; Ribes and Terray, 2015)'.

Page 3 Lines 15-17. I think the wrong Ribes paper has been referenced here in two places. Ribes and Terray, Clim. Dynam. [2013] is much more appropriate, as it has a detection study on observations and examines differences when using different models. Ribes et al. [2015] does neither of those things.

The reference has been replaced as recommended.

Page 3 lines 20-23. If ozone and land use are not included with aerosols, what should they be included with? Are the authors proposing ignoring these forcing factors?

In this approach ozone and land use change are grouped with WMGHGs. This comment has been addressed by the insertion of the following two sentences: 'Such

aerosol-only simulations may be used together with historical simulations including all forcings and historical simulations with natural forcings only to estimate attributable contributions to observed changes due to natural forcings, due to aerosols, and due to the combined effects of well-mixed greenhouse gases, ozone and land-use changes. Since some part of the greenhouse gas changes is associated with land-use change, and since ozone is a greenhouse gas, grouping these forcing together arguably makes more sense than grouping ozone and land-use change with aerosols.'

Page 4 line 13 and elsewhere. The use of 'early 21st century slowdown' should be discouraged. Like other terms attempting to describe this period, e.g., 'hiatus', it is ill defined at best and inaccurate and misleading at worst. The scientific literature has tied itself up in knots about misleading names for this period. The authors should be really clear about what they mean.

Two instances of 'slowdown period' replaced by 'early 21st century' and one instance of 'early 21st–century slowdown of climate warming' replaced by 'period of reduced warming in the early 21st–century' to address the reviewer's comment.

Page 5 Lines 26-29 This is an excellent recommendation. It was generally difficult to deduce what models did follow after 2005 for the GHG experiments. Proposing a specific SSP will help to maintain consistency in the future simulations.

Noted – we are pleased the reviewer supports this recommendation.

Page 6 Lines 1-5 Presumably (as is stated later) the natural radiative forcings for the 2015-2020 period will follow those as planned to be used by ScenarioMIP. It should be mentioned that future volcanic forcing is planned to be "ramped up" [O'Neill et al GMDD 2016] between 2015 and 2025 to control levels and that future solar irradiance will contain a repeat of past solar irradiance variations. It might actually make sense to push back the ramping up (or down) of the volcanic aerosol until after 2020, to avoid introducing an artificial signal into the models at the end of the historical period. Something to liaise with O'Neill et al [2016]

about?

A description of the evolution of natural forcings over the 2015-2020 period in histALL has now been added to the description of histALL 'Time-evolving solar forcing, and stratospheric aerosol ramping up towards the piControl background level should be prescribed over the 2015-2020 period as specified by ScenarioMIP (O'Neill et al., 2016).', including a reference to O'Neill et al. (2016). We have requested that ScenarioMIP move the ramp up of aerosols to after 2020 as suggested, but do not yet know whether this change will be made.

Page 6 Lines 12-13 Is it wise to put simulations covering the period 2015-2020 in CMIP6 with the label 'SSP2-4.5'? That is inconsistent with what is required by ScenarioMIP for that label, i.e., they expect simulations to cover the 2015-2100 period. This could cause confusion.

For groups who are running an ensemble of SSP2-4.5 simulations of the same size as the ensemble of historical simulations this will not be a problem. We have asked ScenarioMIP whether they are concerned about this in the case that some modelling groups participate in DAMIP, but do not run the same size ensemble of SSP2-4.5 simulations. They have indicated that this is not a concern.

Page 6 lines 14-17 There may be some diagnostics or filtering combinations it will be ok, but for characterising multi-decadal near surface temperature patterns of change in individual models, 3 ensemble is likely to be insufficient, e.g., Ribes et al [2015].

We agree with the reviewer that an ensemble size for historicalNAT of greater than three is desirable, but as the reviewer states elsewhere, setting the minimum ensemble size too high could discourage participation from modelling groups. Clearly we need to set a balance. It is expected that some modelling groups will run ensembles larger than the minimum requested size of three.

Page 6 Lines 21-25 It is excellent to see this recommendation. It is also worth giving a recommendation for land use changes. For instance at least one CMIP5 model (IPSL-CM5A-LR) may have included historic land use change in their historicalGHG experiment.

Thanks for the positive comment. As requested by reviewer 1, we now include a full list of forcings to be perturbed in each experiment in Table 1, clearly indicating that land use change is changed in histALL, but not in histGHG.

Page 8 lines 10 onwards It is not clear how useful a small ensemble of histSOL will be for many detection studies. The authors are rather over confident when they say (line 24) that "unambiguous characterization of each model's solar signal" can be deduced and "separate clearly solar and volcanic effects". Having only three initial condition ensemble members will make this very difficult for many diagnostics and filtering choices (Ribes [2015]). If analysts are interested in impacts of just the solar cycle, then better experiments could be designed. The ssp245Nat for instance may provide the required data for those needs.

This experiment was adopted from SolarMIP. As noted in the text, if analysts are interested in identifying the historical simulated response to solar variations, then simulations with solar forcing only are required. ssp245Nat could be used to identify a model's response to solar forcing variations, but this would be for projected future variations, rather than observed historical variations. As noted previously three is a minimum ensemble size, and modelling groups interested in separating the solar signal could run a larger ensemble. Nonetheless, in response to the reviewer's comment, 'umambiguous characterisation' and 'separate clearly' were moderated so that the text now reads 'The proposed histSOL experiment will facilitate the characterization of each model's solar signal and allow the separation of solar and volcanic effects over the historical period.'

Page 8 lines 27 It is rather over confident to state "will allow the characterisation of and attribution to volcanic influences." Three initial condition ensemble

members are likely to be insufficient for multi decadal near surface temperature analyses on individual models (Ribes [2015]).

The reviewer is correct – we meant to indicate that the simulations would allow detection and attribution studies on volcanic influence, rather than guaranteeing that volcanic influence will be detectable. Revised to read 'The histVLC experiments will facilitate detection and attribution studies on volcanic influence.'

Page 8 Line 32 onwards the usefulness of an ensemble of histCO2 is not at all obvious and I wonder if it should have been more widely discussed within the detection and modelling communities before being proposed. The authors main reason for inclusion of this experiment seems to be helping to constrain TCRE (Eq 2 in Gillett [2013]). The TCRE (temperature change relative to cumulative emission of CO2) should be constant for a given model. So shouldn't the information provided by the 1experiment be sufficient? If delta T (1@ 2010)/E(2010) then the usefulness of TCRE is itself questionable. Thus eqn 2 in Gillett [2013] should just be TCRE = beta * TCR / E(2x) The authors also give another reason - that it would help to better characterise the uncertainties in TCR. I am not sure this makes sense. histCO2 would help in understanding uncertainties in the predictive power of TCR estimating GHG warming, not the other way round.

The reviewer misunderstands equation 2 in Gillett et al. (2013). One way of estimating TCRE from observations is to calculate the ratio of CO2-attributable warming to date, to an estimate of cumulative CO2 emissions to date. This calculation relies on TCRE being approximately constant as a function of cumulative emissions. In order to calculate CO2-attributable warming, simulations of the response to historical changes in CO2 only are required. The only other way to calculate CO2-attributable warming would be to multiply GHG-attributable warming by the ratio of CO2 to total GHG forcing (as in Matthews et al., Nature, 2009), but this calculation assumes that the efficacy of all GHGs is one, and it assumes that the temperature response to CO2 and the other GHGs is proportional to their present-day forcing, which is not generally true, given

that the time evolution of their radiative forcings has been different. While we could use simulations from a single model to estimate the ratio of CO2-attributable to GHG-attributable warming (as in Gillett et al., 2013), efficacies differ amongst models, as does the climate response, and thus a much more robust approach would be to use a multi-model ensemble to assess this ratio and its uncertainty.

The reviewer asks why eq (2) in Gillett et al. (2013) cannot be replaced with TCRE = beta * TCR / E(2x). Presumably in this equation the reviewer means that TCR is assessed from one or more models and E(2x) is assessed from the same models, while beta is the regression coefficient obtained from a detection and attribution analysis applied to global temperature. This would be one way to assess TCRE from observations, but it would only use temperature information from observations, and not use any observations of the carbon cycle. Equation 2 from Gillett et al. (2013) uses an estimate of actual emissions from observations (rather than diagnosed emissions from a model), and hence uses observations of both carbon cycle and physical climate to constrain TCRE.

Regarding TCR, approaches to estimating TCR from observations typically calculate a scaling factor on well-mixed greenhouse gases from a regression of observed temperature change onto the simulated response to WMGHGs, as well as other anthropogenic forcings and natural forcings. They then use this regression coefficient to scale a model estimate of TCR (i.e. the simulated response to 1PCTCO2 at doubled CO2), without accounting for the additional uncertainty introduced by this step. This approach would be valid if there were a perfect relationship between historical warming due to WMGHGs and TCR across models. But there is a pronounced spread in this ratio (Gillett et al., 2013; Figure 6). It is not clear whether the spread in this ratio comes mainly from differences in radiative forcings or efficacies between different GHGs across models, or from differences in the temporal response to different time evolution of the radiative forcings of the different GHGs. A multi-model ensemble of CO2-only simulations would allow us to address this issue. The reviewer notes such simulations could also be used to understand the relationship between TCR and predicted GHG-induced warming – this is correct – but the simulations would be just as useful for understanding the sources of uncertainty in observationally-constrained estimates of TCR.

Hence, CO2-only simulations would help us characterize and understand the uncertainties in observationally-constrained estimates of TCRE and TCR, both of which are important policy-relevant metrics, which were both extensively discussed in the Summary for Policymakers of the Working Group I contribution to the IPCC Fifth Assessment Report, for example.

Finally regarding the question of consultation – these simulations were added after the initial proposal was prepared, but were added with consultation of the DAMIP committee. Note that they are included only at the lowest Tier 3 priority, and therefore can be carried out at the discretion of modelling groups and are unlikely to dissuade modelling groups from carrying out DAMIP.

Page 9 lines 9-22 Depending on whether analyses using ssp245AER and ssp245NAT are using multi-model means or individual models, just asking for one ensemble member (Table 1) may not be sufficient to characterizes responses accurately.

Clearly there is a balance to be struck between asking for large ensembles to better account for internal variability, and asking for small ensembles to encourage the participation of the largest number of modelling groups. In an observationally-constrained projection exercise, large ensembles are less critical for the future simulations than for the historical portion of the simulations where the ensemble size may strongly influence the uncertainties on the regression coefficients on individual forcings. For this reason we chose to only request a minimum ensemble size of one for the future simulations. Note that this is a minimum ensemble size, and modelling groups are free to carry out larger ensembles.

We have added this final sentence to the introductory paragraph of Section 2.2 to make clear that larger ensembles are encouraged: 'Minimum ensemble sizes are three for the historical simulations and one for the future simulations, though modelling groups are encouraged to run larger ensembles if resources allow.'

Page 9 lines 31-34 It should be mentioned that ozone and land use forcing factors will be folded up in the examples given, contributing to the "aerosol forcing uncertainty" and "natural forcing uncertainty".

We are not proposing that alternative ozone or land use forcings are used in either histALLestAER2 or histALLestNAT2. If alternate estimate of these forcings were available it would be possible to investigate the role of these forcings in introducing uncertainty in attribution results. However, after consultation with the community, including at the IDAG meeting, we assess that uncertainties in aerosols and natural forcings have made the largest contribution to uncertainties in global temperature changes, and it was necessary to focus on a limited number of experiments here. Thus ozone and land use forcings are not contributors to the estimates of 'aerosol forcing uncertainty' and 'natural forcing uncertainty' that may be derived using these experiments.

Page 10 - section 3 Are there any MIPs that it would be vital to have involvement with? It is highlighted that RFMIP is close to DAMIP, but I wonder if it should be considered a vital MIP that should be done with the same model as used for DAMIP. There is a danger that some institutions will use one version of their model for some MIPs and other versions of their model for other MIPs.

Other than the DECK and CMIP6 historical simulation there are no other simulations that it is vital that participating groups carry out. ScenarioMIP is the most strongly-linked MIP, since it provides the SSP simulations which are needed for observationally-constrained projections carried out using some of our Tier 2 and 3 simulations. But given that these simulations are optional in DAMIP, participation in ScenarioMIP is not required. Participation of models in DAMIP and in the other MIPs included in Table 2 (including RFMIP) would have benefits as described there, but participation in these other MIPs is not vital.

Technical comments Page 8 lines 12-13 The past solar cycle in solar irradiance has had cycle lengths of between 9 and 13 years.

‘11-year cycle’ replaced with ‘approximately 11-year cycle’.

Page 13 Line 9 Gregory et al should be on a new line.

Thanks for pointing this out. Corrected.

line 14 page 15 Large font size

Thanks for pointing this out. Corrected.

line 21 page 14 Submitted to what? I guess GMD?

The journal was already specified – GMD. The reference has now been updated with a doi.

line 32 page 15 Submitted to what?

The journal the paper was originally submitted to did not publish it since it was judged out of scope. The paper is currently in preparation for submission to another journal, and the reference has been revised accordingly.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-74, 2016.