SPi	Journal Code:				Article ID	Dispatch: 05.04.18	CE: MAEE
	G	R	L		57104	No. of Pages: 7	ME:

Revised proofs are sent only in the case of extensive corrections upon request.

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL076649

Key Points:
- There are significant discrepancies in historical global and regional transient climate sensitivity between temperatures and CMIP5 runs
- Temperature responses to anthropogenic forcings are approximately linear in CMIP5 scenario runs, and their biases in TCS are thus projected
- Alternative historical-based methods are required to improve our confidence in regional climate projections and correct biases in GCMs

Supporting Information:
- Supporting Information S1

Correspondence to:
R. Hébert,
raphael.hebert@mail.mcgill.ca

Citation:
Hébert, R., & Lovejoy, S. (2018), Regional climate sensitivity- and historical-based projections to 2100. *Geophysical Research Letters*, *45*. https://doi.org/10.1002/2017GL076649

Received 11 DEC 2017
Accepted 11 MAR 2018
Accepted article online 13 MAR 2018

Regional Climate Sensitivity- and Historical-Based Projections to 2100

Raphaël Hébert[1] and **Shaun Lovejoy**[1]

[1]Department of Physics, McGill University,

Abstract Reliable climate projections at the regional scale are needed in order to evaluate climate change impacts and inform policy. We develop an alternative method for projections based on the transient climate sensitivity (TCS), which relies on a linear relationship between the forced temperature response and the strongly increasing anthropogenic forcing. The TCS is evaluated at the regional scale (5° by 5°), and projections are made accordingly to 2100 using the high and low Representative Concentration Pathways) emission scenarios. We find that there are large spatial discrepancies between the regional TCS from 5 historical data sets and 32 global climate model (GCM) historical runs and furthermore that the global mean GCM TCS is about 15% too high. Given that the GCM Representative Concentration Pathway scenario runs are mostly linear with respect to their (inadequate) TCS, we conclude that historical methods of regional projection are better suited given that they are directly calibrated on the real world (historical) climate.

Plain Language Summary In this paper, we estimate the transient climate sensitivity, that is, the expected short-term increase in temperature for a doubling of carbon dioxide concentration in the atmosphere, for historical regional series of temperature. We compare our results with historical simulations made using global climate models and find that there are significant regional discrepancies between the two. We argue that historical methods can be more reliable, especially for the more policy-relevant short-term projections, given that the discrepancies of the global climate models directly bias their projections.

1. Introduction

Reliable climate projections are needed for informing policy, and at the moment, there is only one technique for projecting the climate to 2050 or 2100—general circulation models (GCMs). The difficulty is that each GCM makes significantly different projections leading to large uncertainties. Over the coming decades, these uncertainties are sufficiently large that widely divergent scenarios of economic development lead to significantly overlapping projections: both high- and low-emission pathways have nonnegligible probabilities of staying within the agreed 2 °C threshold (e.g., Andersen and Bows, 2011). Policies ranging from very low (Intergovernmental Panel on Climate Change [IPCC] Representative Carbon Pathways 2.6 [RCP2.6] scenario) to very high (IPCC RCP8.5 scenario) emissions—a range of more than a factor of 3—could all be said to be compatible with this threshold albeit over different parts of the projected probability distribution. As policy choices (emissions) do not uniquely map onto outcomes (temperatures), choices that ought to be clear become muddled. In addition to the uncertainty provided by the GCM results, there is a methodological, epistemological uncertainty: all IPCC projections rely on the same modeling approach. This reliance is a weakness that can only be addressed through the development of qualitatively different approaches.

The alternatives exploit the linearity of the climate response; see, for example, the multiparameter regression of the temperature response as a function of various forcing sources (Lean & Rind, 2009), or the Effective Climate Sensitivity (EffCS) to the observed CO_2 radiative forcing, which is used as a surrogate for the total anthropogenic forcing (thereby avoiding the uncertainty associated with aerosol forcing) (Lovejoy, 2014). Another approach found that the temperature increase is approximately linear to the cumulative carbon emissions, both in observations and in GCMs (Matthews et al., 2009), and this was even shown to apply at the regional scale (Leduc et al., 2016), but there are also evidences of local nonlinearity in GCM simulations (cf. Good et al., 2015).

The long-memory response that results from the thermal inertia of oceans, that is, the "warming in the pipeline" (Hansen et al., 2011), is well established, but during the last century, this delayed response to past forcing is essentially swamped by the fast increase in anthropogenic forcing, but if future emissions are to level off (RCP4.5), or even decrease (RCP2.6), the long-memory response will be important. However, it is generally agreed that the transient climate response, that is, the short-memory, or no, memory response, is more policy relevant for near-term assessment, and there have been numerous assessments of transient climate response at the global scale (Gregory & Forster, 2008; Otto et al., 2013; Padilla et al., 2012; Richardson et al., 2016; Schwartz, 2012).

To produce alternative regional climate projections, Mitchell et al. (1999) proposed the pattern scaling method in which warming patterns are derived and scaled proportionally to the global mean temperature. The approach was subsequently investigated and developed to take into account the impact of sulfate aerosol (Ishizaki et al., 2010; Mitchell et al., 2003; Schlesinger et al., 2000). More recently, Geoffroy and Saint-Martin (2014) have demonstrated that a simple two-layer global energy balance model and a two-pattern decomposition of the mean surface temperature were able to emulate GCM transient warming simulations.

In this paper, we restrict the scope of our analysis to a short-memory CRF in order to quantify the linearity assumption to anthropogenic forcing at the regional scale, and we therefore establish the basis to pursue a historical method for regional temperature projections. The first section of this paper defines the transient climate sensitivity (TCS) and applies the method to global temperature series. In the second section, we extend the TCS analysis to regional temperature series and identify significant discrepancies between TCS maps of historical observations and GCM historical runs. In the third section, we present and discuss global and regional TCS projections along the GCM scenario runs.

2. Global TCS

2.1. The Data

The parameter we chose to estimate regional climate sensitivity to anthropogenic impacts is the TCS (Bengtsson & Schwartz, 2013). TCS is more relevant to evaluate anthropogenic warming than simple warming trends since we take the temperature increase to be proportional to the increase in forcing rather than time. It is essentially equivalent to the EffCS method, but instead of CO_2, it uses CO_2 equivalent ($CO_{2\,EQ}$), which corresponds to the CO_2 concentration required to produce a forcing linearly equivalent to the total anthropogenic forcing (see the supporting information). The EffCS method implicitly assumes that all forcings—including aerosol forcings—are constant fractions of the total. While it is fairly accurate in the past, the RCP scenarios project that aerosols will be largely "cleaned up," which increases the uncertainty for future projections and implies that there may be significant short-term warming (Armour & Roe, 2011). We therefore use $CO_{2\,EQ}$ for which the ratio changes to estimate TCS and make projections. The same forcing is used for both observations and GCMs in order that the resulting differences sensitivity and warming be independent of the specific aerosol representation in each GCM.

Five observational records of surface air temperature are used for the analysis over the historical period 1880–2017: HadCRUT4 (Morice et al., 2011), the Cowtan and Way reconstruction version 2.0 (Cowtan & Way, 2014), Goddard Institute for Space Studies Surface Temperature Analysis (Hansen et al., 2010; GISTEMP Team, 2015), National Oceanic and Atmospheric Administration's merged global surface temperature anomaly (Smith et al., 2008), and Berkeley Earth Surface Temperature (Rohde et al., 2013). We also used a 32-member multimodel ensemble (MME) of selected GCMs—including different versions of the same GCMs—produced in accordance with the Coupled Model Intercomparison phase 5 (CMIP5) and which have historical simulation outputs available for the period from 1860 to 2005, and outputs of scenario runs over 2005–2100 for both RCP2.6 and RCP8.5 (see supporting information concerning the GCM details). For calculations over 1880–2016, the historical part was extended with the scenario run RCP2.6 (out of convenience since over 2005–2016 both RCP2.6 and RCP8.5 remain almost equal).

2.2. The Linearity of the Global Response

In order to test the short-range linearity assumption over the historical period, we simply plot (Figure 1) the number of $CO_{2\,EQ}$ doublings (proportional to the $CO_{2\,EQ}$ forcing) against the observed temperature increase. From the figure, we see that the plot is very straight with fluctuations around the trend due to internal variability. Its slope is the TCS, and it is the observed sensitivity of the climate. While it does not take into account

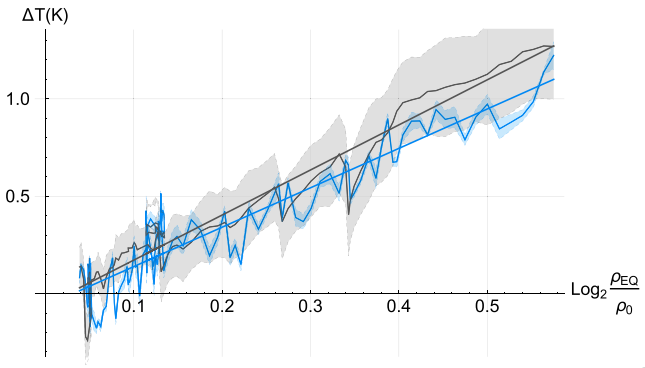

Figure 1. The global mean temperature of historical observations (blue) and of the multimodel ensemble of historical runs (black) are shown, and the standard deviation of each ensemble is given (shaded), as a function of the anthropogenic forcing associated to $CO_{2\,EQ}$. The slopes of the corresponding regression lines correspond to the TCS, calculated over 1880–2017.

the system memory, it nevertheless quantifies the historical increase in temperature due to the anthropogenic forcing. The anthropogenic transient temperature change ΔT_{TCS} is thus defined as

$$\Delta T_{\mathrm{TCS}}(t) = \lambda \log_2 \frac{\rho_{\mathrm{EQ}}(t)}{\rho_0} \quad (1)$$

where λ is the TCS, ρ_{EQ} is the global CO_{2EQ} atmospheric concentration, and ρ_0 is its preindustrial value (277 ppm). Subsequently, we refer to the TCS from observational data as TCS_{Obs} (or λ_{Obs}) and the one from the CMIP5 MME as TCS_{MME} (or λ_{MME}). In the supporting information, we discuss the sensitivity of the TCS to the measuring period and to the internal variability.

The $CO_{2\,EQ}$ forcing is the one used in the CMIP5 GCMs so that the linearity of Figure 1 confirms that the GCMs themselves are linear, although with a TCS that is about 15% higher than the historical observations. The global scale TCS calculated over the period 1880–2017 in historical observations ($\lambda_{Obs} = 2.0 \pm 0.2$ K) is significantly different ($p = 0.08$) from the TCS in the historical runs of the CMIP5 MME ($\lambda_{MME} = 2.3 \pm 0.5$ K), but the TCS of 12 of the 32 GCMs is not different ($p > 0.1$) from the observational TCS. The error on TCS_{MME} is representative of the spread within the ensemble.

3. RTCS

The TCS was calculated at a $5° \times 5°$ resolution for each of the five observational gridded data sets (for pixels where at least 40% of the temperature values were available) and for each of the 32 CMIP5 GCMs in the MME considered, resulting in maps of regional TCS (hereafter RTCS; see the supporting information for the RTCS of individual GCMs). The forcing used is a global series and therefore the same for all grid points. This is appropriate for well-mixed greenhouse gases, which constitute the dominant anthropogenic forcing, but not for aerosol forcing and land use changes. With this caveat in mind, it is informative to compare the resulting RTCS calculated from observation and simulations. Since we are a little past half-way to a doubling of $CO_{2\,EQ}$, the RTCS is roughly twice the amount of warming that has occurred since preindustrial times.

For statistical comparisons, the observational gridded data sets were taken to form one ensemble and the CMIP5 MME another. To assess whether the two ensembles yielded statistically equivalent means, a *t* test was performed after randomizing the TCS_{Obs} ensemble according to the uncertainty due to the internal variability (see supporting information). Similarly, the test was used to assess whether the mean RTCS of an ensemble at a given location was significantly different from zero. We concluded significantly different results when the test statistic $p < 0.1$, the alternative hypothesis was concluded when $p > 0.9$.

The RTCS is positive everywhere in the observational data sets (Figure 2), except for an area in the North Atlantic where deep water forms, but given that the trend is not significant, it is likely to be a transient phenomenon (see Rugenstein et al., 2016, for a discussion of nonlinearities in ocean warming).

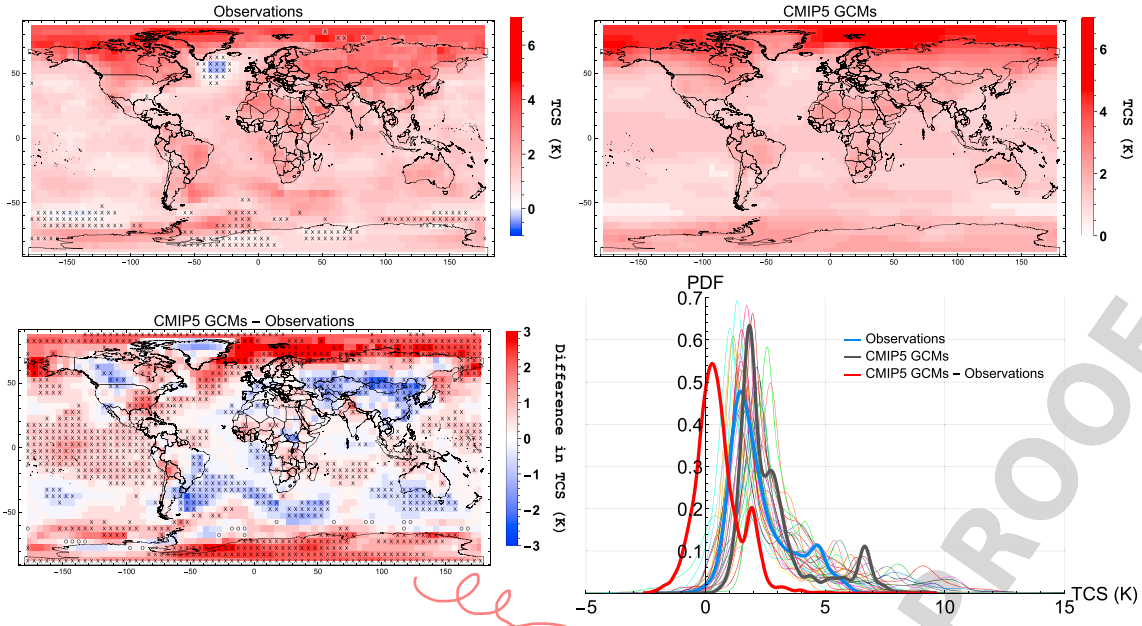

Figure 2. (top left) The mean $RTCS_{Obs}$ of the five observational data sets considered is evaluated and shown at a 5° × 5° resolution . "X" markers indicate pixels where $RTCS_{Obs}$ is not different from zero ($p > 0.1$). (top right) Same as top left but for the mean $RTCS_{MME}$ (32 models in the multimodel ensemble [MME]); all pixels showed regional transient climate sensitivity (RTCS) different from zero ($p < 0.1$). (bottom left) Difference between the maps of $RTCS_{Obs}$ and $RTCS_{MME}$ shown above. Red areas indicate that the $RTCS_{MME}$ is greater than the $RTCS_{Obs}$, and blue the converse. "X" markers indicate that the mean of the two ensembles are significantly different ($p < 0.1$), whereas the "O" markers indicate they were statistically equal ($p > 0.9$). (bottom right) The probability distribution functions (PDFs) for all the pixels of $RTCS_{Obs}$ (thick, blue), $RTCS_{MME}$ (thick, black), and their difference (thick, red) are shown. The thin lines show the PDFs for the RTCS of the 32 individual Coupled Model Intercomparison phase 5 (CMIP5) global climate models (GCMs) in the MME.

The highest warming is observed over land at high latitudes, particularly northwestern Canada and central Asia, and strong warming is also observed over the Arctic Ocean, although it is mostly the result of interpolation from neighboring land.

The ensemble mean of $RTCS_{MME}$ is (significantly) positive everywhere on the globe, even in the North Atlantic (Figure 2). Individually, many GCMs show regional cooling: about a third of them show a cooling in the North Atlantic and about half showed at least some cooling in the Southern Ocean. All of them show very strong warming over the Arctic Ocean, but only half show a similar polar amplification over Antarctica (see the supporting information for details).

The differences between the ensemble mean maps of $RTCS_{Obs}$ and $RTCS_{MME}$ are shown in Figure 2. Overall, the distribution of the differences has a mean of 0.3 K, which is significantly different from zero ($p < 0.01$), and the mean root-mean-square error is 1.1 K. The RTCS is generally higher in the MME over oceans, and it is significantly so over vast parts of the Pacific Ocean, the North Atlantic Ocean, and the Indian Ocean, while the opposite is true for the South Atlantic Ocean and the part of the Indian Ocean south of Australia. The RTCS is also significantly underestimated by the MME over northwestern Canada and central Asia, and a seasonal analysis reveals that this discrepancy is due to the winter months that show higher sensitivity, while the summer months show decreased sensitivity (see the supporting information).

4. CMIP5 Projections to 2100

Following Figure 1, we obtain Figure 3, which compares the TCS method with the MME projections for the global averaged temperature. The figure shows both the means and the spreads (one standard deviation σ) of each ensemble (the ensemble of the TCS method comes from the use of the five different temperature data sets). We can draw several conclusions. The first is that the mean of the TCS method applied to the data is within one σ of the MME. The second is that the mean MME projections are about 15% (RCP8.5) to 40% (RCP2.6) higher than the TCS projections, and the third, that the uncertainty of the historical method is reduced even when the uncertainty due to the natural variability is taken into account.

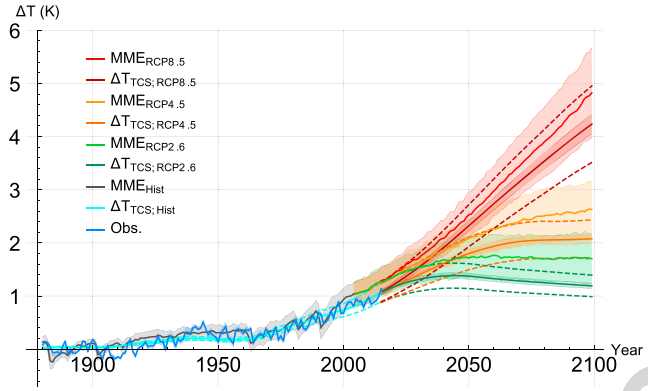

Figure 3. A global surface temperature comparison of Coupled Model Intercomparison phase 5 global climate model historical runs (MME_{Hist}, black) with historical observations (Obs., blue) since 1880 and the anthropogenic temperature variability reconstructed using equation (1) ($\Delta T_{TCS;\,Hist}$, teal), both with projections into the future (past 2005 and 2016, respectively). The global climate model scenario runs are shown for scenarios Representative Carbon Pathway 2.6 (RCP2.6; $MME_{RCP2.6}$, green), RCP4.5 ($MME_{RCP4.5}$, orange), and RCP8.5 ($MME_{RCP8.5}$, red), along with the transient climate sensitivity (TCS) projections for the corresponding scenarios, that is, RCP2.6 ($\Delta T_{TCS;\,RCP2.6}$, dark green), RCP4.5 ($\Delta T_{TCS;\,RCP4.5}$, dark orange), and RCP8.5 ($\Delta T_{TCS;\,RCP8.5}$, dark red). The shaded envelopes correspond to the one standard deviation structural uncertainty in each ensemble, and the dashed lines around the mean TCS projections correspond to the uncertainty on the TCS due to the natural variability. MME = multimodel ensemble.

The TCS method can also be applied at regional scales and compared with MME projections (Figure 4; see the supporting information for a discussion of the method applied to the MME). The discrepancies between the two are in fact a reflection of the discrepancy in their historical RTCS (Figure 2, bottom left). While it may appear limited that the RTCS (linear) projection is approximately the RTCS map rescaled by the global forcing, we see that in fact the same limitation is true for the MME; the $RTCS_{MME}$ map has a high correlation coefficient (ρ) with both projection maps, for RCP2.6 (ρ=0.98) and for RCP8.5 (ρ=0.95), which are also highly correlated together (ρ = 0.98).

The pattern of warming of the MME is in fact nearly stationary over the 21st century, even for RCP2.6, showing practically no evolution between 2000 and 2100. To show this, we calculated the correlation coefficient (ρ) between an initial temperature map averaged over 2000–2020 with successive 20-year periods up to

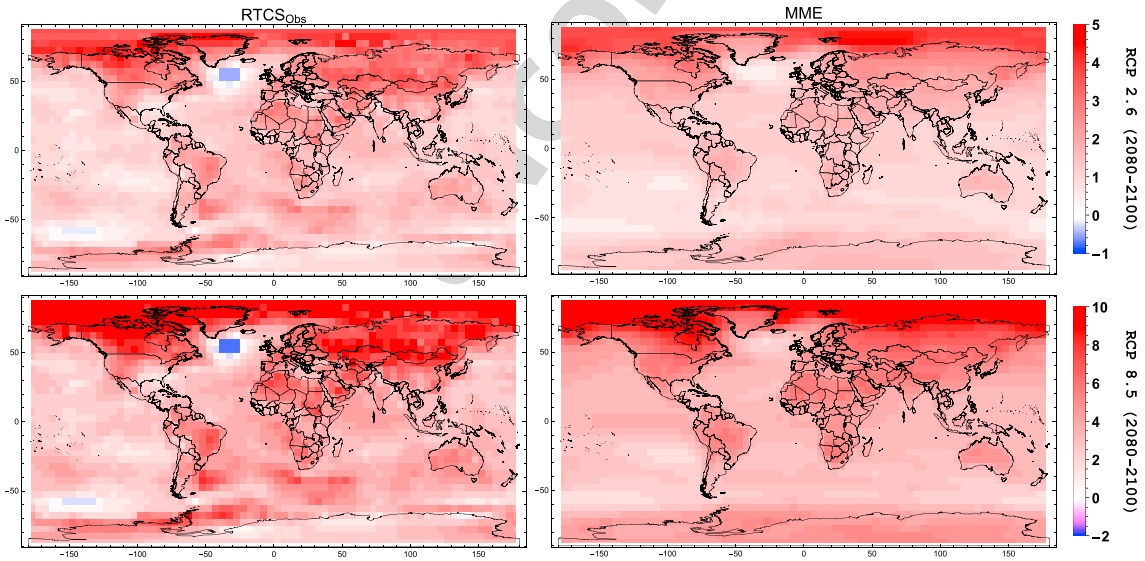

Figure 4. (top left) The surface temperature is projected using the regional transient climate sensitivity (RTCS) along Representative Carbon Pathway 2.6 (RCP2.6) for the period 2080–2100. (top right) Mean surface temperature of the multimodel ensemble (MME) for RCP2.6 scenario runs over the period 2080–2100. (bottom row) Same as top row but for RCP8.5. Notice that the temperature scale for RCP8.5 is twice as wide as that of RCP2.6.

2080–2100 (see the supporting information). For both RCP2.6 and RCP8.5, the correlation remains very close to unity ($\rho > 0.98$) for their MME mean. There is more variation for individual models, and especially for RCP2.6, where the correlation dips below 0.8 for 10 GCMs out of 32, while only 3 GCMs do the same for RCP8.5.

The GCMs have overestimated the warming since 1880 over most of the globe, but they also underestimated it over some significant regions such as northwestern Canada and central Asia. For 39% of the Earth's surface, the GCMs and the historical data are statistically in disagreement. The GCM regional projections are thus of limited utility since they poorly reproduced past regional variability and their projections are practically rescaled maps of their historical RTCS. It is more reliable to extrapolate the RTCS of the data, especially for short-term projections where the transient response is more important. For long-term projections, RTCS projections might underestimate the warming over some regions, due to the neglected long-memory response.

It may seem that regional projections using the historical RTCS method are of limited value because the warming patterns it yields are practically stationary and will not predict nonlinear regional climate changes. However, although GCMs have the ability to predict nonlinear evolutions, we found that the actual projections from individual GCMs mostly average each other out so that their MME mean regional projections are quite linear. Given the significant discrepancies between the RTCS of the MME mean with the actual historical temperature evolution, the MME projections will project the same discrepancies. In contrast, the observations-based RTCS will make projections forced to agree with the real—not model—climate.

5. Conclusion

Accurate climate projections are increasingly needed to inform policy. In this context, the large uncertainties implied by the wide dispersion of future projections is a major concern. At the moment, even widely diverse IPCC emission scenarios lead to probability distributions of temperature response that are largely overlapping. This implies that a policy decision to pursue a specific outcome (e.g., to remain within a warming of 2 °C), does not adequately constrain policy.

In this paper, we considered a simplified historical approach using the TCS both at global and at regional scales. Over the period 1880–2016, the global TCS was found to be generally too high in the MME, by about 15% more than in observations. Given that the same forcing series was used, this discrepancy could also be explained by errors in the internal representation of forcing in the GCMs.

The RTCS was defined as the ratio of the observed regional warming to the global increase in anthropogenic forcing, for data sets of historical observations and for a MME of historical runs of CMIP5 GCMs. We have found several discrepancies in RTCS, particularly at high latitudes in the Northern Hemisphere where the RTCS in the GCMs is underestimated over northwestern Canada and central Asia and overestimated over the Arctic Ocean; GCMs have their own climate that is significantly different than that of historical observations.

Furthermore, the spatial warming patterns of the MME for the 2080–2100 period were shown to be highly correlated with their RTCS map from the historical runs and also mostly stationary over the 21st century with respect to the 2000–2020 period. Individual GCMs show greater evolution in their spatial warming patterns, but there are large variations in the maps of RTCS between models, and there is not one which clearly outperforms the mean MME.

Given these facts, it is questionable how accurate the mean projection of the MME will be given that its spatial pattern of warming evolves little over the 21st century and remains very close to its RTCS over the historical period, which significantly differs over most of the globe from the actual RTCS of historical observations. Therefore, the resulting MME projection will suffer from the same faults as the historical runs; that is, it will be a projection of the faulty GCM climate in historical runs.

GCMs are important research tools, but their regional projections are not yet reliable enough to be taken at face value. It is therefore required to develop further historical approaches that will thus project the real world climate rather than the GCM climates. An exciting possibility for further improvements in projections would be to use the strengths of the GCMs to improve the historical projections and the strengths of the latter to improve the former. We investigate such hybrid methods in a future publication.

The TCS method presented could be improved by using local forcing series and introducing a long-memory climate response function that would allow a delayed response to forcing. In particular, a detailed regional analysis of aerosol forcing, and the associated corrections in TCS, is in order to avoid bias in regions where

the contribution of aerosols to total forcing changes between the historical and projection period. This would yield an independent robust framework for climate projections that would be based on the real climate and the fact, observed in GCMs, that even the regional temperature response is approximately linear for long enough timescales. It would strengthen our confidence in regional climate projections and could then be used for impact assessments of future climate changes, therefore better informing policy makers.

Acknowledgments
Raphaël Hébert thanks a McGill Fessenden grant for partial support. The data used are listed in the references, tables, and supporting information. A more detailed discussion of the methodology can be found in the supporting information (Koh et al., 2012; Lovejoy et al., 2012; Meinshausen et al, 2012; Stevens, 2015).

References

Anderson, K., & Bows, A. (2011). Beyond "dangerous" climate change: Emission scenarios for a new world. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *369*(1934), 20–44. https://doi.org/10.1098/rsta.2010.0290.

Armour, K. C., & Roe, G. H. (2011). Climate commitment in an uncertain world. *Geophysical Research Letters*, *38*, L01707. https://doi.org/10.1029/2010GL045850.

Bengtsson, L., & Schwartz, S. E. (2013). Determination of a lower bound on Earth's climate sensitivity. *Tellus B*, *65*(21533), 1–16. https://doi.org/10.3402/tellusb.v65i0.21533.

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. https: //doi.org/10.1002/qj.2297.

Geoffroy, O., & Saint-Martin, D. (2014). Pattern decomposition of the transient climate response. *Tellus A: Dynamic Meteorology and Oceanography*, *66*, 1. https://doi.org/10.3402/tellusa.v66.23393.

GISTEMP Team (2017). GISS Surface Temperature Analysis (GISTEMP), http://www.data.giss.nasa.gov/gistemp/, NASA Goddard Institute for Space Studies. pp. 2017–09, https://doi.org/10.1029/2010RG000345.

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., et al. (2015). Nonlinear regional warming with increasing CO_2 concentrations. *Nature Climate Change*, *5*(2), 138–142. https://doi.org/10.1038/nclimate2498.

Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, *113*, D23105. https://doi.org/10.1029/2008JD010405.

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*, RG4004. https://doi.org/10.1029/2010RG000345.

Hansen, J., Sato, M., Kharecha, P., & Von Schuckmann, K. (2011). Earth's energy imbalance and implications. *Atmospheric Chemistry and Physics*, *11*(24), 13,421–13,449. https://doi.org/10.5194/ acp-11-13421-2011.

Ishizaki, Y., Shiogama, H., Emori, S., Yokohata, T., Nozawa, T., Takahashi, K., et al. (2013). Dependence of precipitation scaling patterns on emission scenarios for representative concentration pathways. *Journal of Climate*, *26*(22), 8868–8879. https://doi.org/10.1175/JCLI-D-12-00540.1.

Koh, T. Y., Wang, S., & Bhatt, B. C. (2012). A diagnostic suite to assess NWP performance. *Journal of Geophysical Research*, *117*, D13109. https://doi.org/10.1029/2011JD017103.

Lean, J. L., & Rind, D. H. (2009). How will Earth's surface temperature change in future decades?. *Geophysical Research Letters*, *36*, L15708. https://doi.org/10.1029/2009GL038932.

Leduc, M., Matthews, H. D., & de Elía, R. (2016). Regional estimates of the transient climate response to cumulative CO_2 emissions. *Nature Climate Change*, *6*, 474–478. https://doi.org/10.1038/nclimate2913.

Lovejoy, S. (2014). Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming. *Climate Dynamics*, *42*(9–10), 2339–2351. https://doi.org/10.1007/s00382-014-2128-2.

Lovejoy, S., Del Rio Amador, L., & Hébert, R. (2015). The ScaLing Macroweather Model (SLIMM): Using scaling to forecast global-scale macroweather from months to decades. *Earth System Dynamics*, *6*(2), 637–658. https://doi.org/10.5194/esd-6-637-2015.

Matthews, H. D., Gillett, N. P., Stott, P. A., & Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature*, *459*(7248), 829–832. https://doi.org/10.1038/nature08047.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L., Lamarque, J., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1), 213–241. https://doi.org/10.1007/s10584-011-0156-z.

Mitchell, J. F. B., Johns, T. C., Eagles, M., Ingram, W. J., & Davis, R. A. (1999). Towards the construction of climate change scenarios. *Climatic Change*, *41*(3–4), 547–581. https://doi.org/10.1023/A:1005466909820.

Mitchell, T. D. (2003). Pattern scaling. An examination of the accuracy of the technique for describing future climates. *Climatic Change*, *60*(3), 217–242. https://doi.org/10.1023/A:1026035305597.

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*, D08101. https://doi.org/10.1029/2011JD017187.

Olson, R., Sriver, R., Goes, M., Urban, N. M., Matthews, H. D., Haran, M., & Keller, K. (2012). A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *Journal of Geophysical Research*, *117*, D04103. https://doi.org/10.1029/2011JD016620.

Olson, R., Sriver, R. L., E., Vallis, G. K., & Rowley, C. W. (2011). Probabilistic estimates of transient climate sensitivity subject to uncertainty in forcing and natural variability. *Journal of Climate*, *24*(21), 5521–5537. https: //doi.org/10.1175/2011JCLI3989.1.

Richardson, M., Cowtan, K., Hawkins, E., & Stolpe, M. B. (2016). Reconciled climate response estimates from climate models and the energy budget of Earth. *Nature Climate Change*, *6*(10), 931–935. https: //doi.org/10.1038/nclimate3066.

Rohde, R., Muller, R., Jacobsen, R., Muller, E., Groom, D., & Wickham, C. (2012). A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinformatic & Geostatistics: An Overview*, *1*(1), 1–7. https://doi.org/10.4172/gigs.1000101.

Rugenstein, M. A., Sedláček, J., & Knutti, R. (2016). Nonlinearities in patterns of long-term ocean warming. *Geophysical Research Letters*, *43*, 3380–3388. https://doi.org/10.1002/2016GL068041.

Schlesinger, M. E., Malyshev, S., Rozanov, E. V., Yang, F., Andronova, N. G., De Vries, B., et al. (2000). Geographical distributions of temperature change for scenarios of greenhouse gas and sulfur dioxide emissions. *Technological Forecasting and Social Change*, *65*, 167–193. https://doi.org/10.1016/S0040-1625(99)00114-6.

Schwartz, S. E. (2012). Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: Strong dependence on assumed forcing. *Surveys in Geophysics*, *33*(3-4), 745–777. https: //doi.org/10.1007/s10712-012-9180-4.

Smith, T. M., Reynolds, R. W., Peterson, T. C., & Lawrimore, J. (2008). Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *Journal of Climate*, *21*(10), 2283–2296. https://doi.org/10.1175/2007JCLI2100.1.

Stevens, B. (2015). Rethinking the lower bound on aerosol radiative forcing. *Journal of Climate*, *28*(12), 4794–4819. https://doi.org/10.1175/JCLI-D-14-00656.1.