

REPLY

10.1002/2013JD020732

This article is a reply to comment by
Laprise [2014] doi:10.1002/
2013JD019945

Correspondence to:

D. T. Shindell,
drew.t.shindell@nasa.gov

Citation:

Shindell, D., P. Racherla, and G. Milly
(2014), Reply to comment of Laprise on
"The added value to global model
projections of climate change by
dynamical downscaling: A case study
over the continental U.S. using the
GISS-ModelE2 and WRF models,"
Racherla et al. (2012), *J. Geophys. Res.*
Atmos., 119, doi:10.1002/2013JD020732.

Received 12 AUG 2013

Accepted 17 FEB 2014

Accepted article online 19 FEB 2014

Reply to comment by Laprise on "The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models"

Drew Shindell¹, Pavan Racherla¹, and George Milly¹

¹NASA Goddard Institute for Space Studies, New York, New York, USA

In his comment, Laprise raises several points that we agree merit consideration. His primary critique is that our study [Racherla et al., 2012] tested the ability of the WRF regional climate model to reproduce historical temperature and precipitation change relative to the driving global climate model (GCM) using only a single simulation rather than an ensemble. He asserts that the observed changes are smaller than the internal variability in the climate system (i.e., not statistically significant) and that thus a single simulation should not necessarily be able to capture the observations.

Laprise points out that the statistical signal is reduced for a multi-decadal trend such as the one we analyzed in comparison with mean climatology and cites two studies showing that for particular climate parameters it can take many years for a signal to be discerned over internal variability. He states that "The results of the experiment as designed were strongly influenced by the presence of internal variability and sampling errors, which masked the rather small climate changes that may have occurred as a consequence of changes in forcing during the period considered." While Laprise discusses statistics in general terms at some length, for the actual climate trends examined in our study, he offers no evidence that the forced signal was small compared with internal variability. The two studies he cites [de Elía et al., 2013; Maraun, 2013] do not provide convincing evidence as they concern climate variables averaged over different times and areas. One in fact examines extreme precipitation events, which by definition are rare and thus have a lower significance level. We accept the general point that it is important to consider internal variability, and as noted in our paper we agree that an ensemble of simulations is in principle an optimal, though computationally expensive, approach. While we did not present the statistical significance of the observations in our original paper, we have now evaluated those for the regional temperature trends used in our study to evaluate the added value of WRF and thus can analyze data as to the magnitude of the trends with respect to internal variability.

We calculated the standard deviation in regional temperatures for the 11 regions used in our original paper from an ensemble of global climate models participating in the most recent worldwide intercomparison project (the Coupled Model Intercomparison Project phase 5; CMIP5 [Taylor et al., 2012]). We examined the change over equivalent length (11 years) periods separated by the same number of years (27 between the central years) in long (450–700 years) control runs submitted to the CMIP5 archive from seven, independent climate models (NorESM1-M, CCSM4, MRI-CGCM3, MIROC5, CanESM2, GISS-E2-R, and bcc-csm1-1). While the coupled models could potentially either underestimate or overestimate long-term variability, the observational record is too short to reliably constrain unforced variability on long timescales as it contains a very limited number of samples for long periods and includes the large forcings that have occurred during the industrial era. Hence, coupled ocean-atmosphere models represent the best estimates currently available for natural, unforced variability in regional US temperatures.

We compared that variability with the observed regional changes between the 1968–1978 and 1995–2005 periods as in our original study. We find that for the 11 regions and 4 seasons analyzed, the majority of the warming trends observed between these decades are in fact statistically significant at the 90% confidence level (28 of 44 points; with a 95% confidence level, 23 of 44 points are significant; Table 1). Note that most of the nonsignificant trends occur during boreal spring, and so results for many regions should be interpreted with great caution for that season. However, all regional trends are significant for the winter, and the majority

Table 1. Ratio of Regional Mean Observed 1995–2005 Versus 1968–1978 Surface Temperature Anomaly to the Standard Deviation of CMIP5 Control Runs^a

	DJF	MAM	JJA	SON
R1	1.67	1.52	1.52	1.82
R2	1.75	2.57	2.12	1.80
R3	2.31	0.98	1.15	1.74
R4	2.47	2.42	1.67	2.55
R5	2.43	0.07	0.54	1.21
R6	2.34	0.77	1.12	1.75
R7	2.24	0.43	1.15	1.02
R8	2.03	0.60	1.34	1.05
R9	1.70	1.01	2.31	1.33
R10	2.13	0.80	1.78	1.34
R11	2.04	0.86	2.15	1.05

^aValues exceeding 1.65 are significant at the 95% confidence interval, while values greater than 1.28 are significant at 90% and are shown in bold type.

of trends are significant for summer and fall. Restricting the original evaluation of the global and regional models to those regions and seasons where the observed changes are significant (at either the 90% or the 95% confidence level) has little effect on the conclusions of our paper regarding the ability of WRF to improve upon global model simulations of regional temperature change.

While we appreciate that Laprise raised this question and regret that we did not present this analysis in our original paper, we are heartened to find that the observed temperature changes are in fact highly significant (at least with respect to model estimated internal variability). Were they not, we would be left with a situation in which we were able to validate only the range of internal variability in regional models against observations. Without a demonstration that regional models can successfully capture the response to forced changes using the same setup as used for future

projections, as a global model can at the global scale (e.g., in response to the eruption of Mt. Pinatubo, or the trend over the full 20th century [Hegerl *et al.*, 2007]), we would have little reason to trust their ability to project the kind of future responses to forced changes that, as Laprise points out, are likely to be much larger than those of recent decades. Thus, in our opinion, it is a positive result for regional modeling that the observed changes can be attributed to external forcing as that raises the prospect of an eventual successful reproduction of those changes by a regional model. We note also that our analysis of statistical significance agrees well with the study of *de Elia et al.* [2013] cited by Laprise as a demonstration of the long timescales required for emergence of significant climate signals. That study reports an estimated year of emergence for the seasonal signal of temperature change of 10–20 years for North America as a whole, and 20–40 years for the average over individual grid points within North America. Given that our regional averages fall somewhere between these two in terms of spatial averaging, it seems wholly consistent that they have emerged in most, but not all, regions during the 37 year period we examined.

Beyond his primary critique of statistical non-significance, Laprise in many instances points out issues we have already discussed in the original paper and, in other instances, those that are not relevant to the paper's scope and objectives. Posing the question, "should climate models be expected to capture changes in surface air temperature and precipitation between two historical decades," he cites the latest evidence suggesting that initial conditions (of the ocean, cryosphere, and biosphere) play a role, albeit minor, in near-term climate predictions. Here he indirectly raises the issue of driving regional climate models (RCMs) or GCMs with observed sea surface temperatures and sea ice cover (as a way to obtain better historical accuracy), but such observations are obviously not available for the future. Hence, such a configuration does not provide a test of the method that must be used in future regional climate model projections and in fact highlights the discrepancy between the setup used for typical evaluations of historical downscaling and that used for providing future projections.

Laprise suggests ensembles of global model simulations as a way to minimize the effect of internal variability (we also discuss ensembles in the conclusions of the original paper, although in a different context) but does not dwell on the complexity of driving RCMs using such ensembles or the added computational expense so incurred (which is significant). Though internal variability would of course be reduced using an ensemble, we note that the observed temperature changes are statistically significant with respect to the internal variability in the GISS GCM alone for 29 of the 44 region-season pairs (similar to the results in Table 1). Hence, in principle, a single realization could be adequate for comparison with these temperature observations. Multi-decadal internal variability arises largely from the oceans, and thus is imposed upon WRF via the boundary conditions. Hence it seems unlikely that internal variability within the WRF model masked its added value over the GCM. Although ensembles should be helpful, it is not clear they are required or that they represent the optimum use of resources. To the extent that ensembles of GCM simulations are needed for RCMs to provide added value, then many prior RCM studies are inadequate, and future studies

would be extremely computationally expensive, and so we concur with Laprise that this is an important research question. An additional research question concerns the skill level of the driving GCM. Our study explored the added value of dynamic downscaling rather than focusing on the skill of the underlying GCM, and further research would be needed to determine the relationship between that skill and the ability of dynamic downscaling to improve the simulation (the GCM should not be too skillful at reproducing the quantity of interest or there will be little opportunity for added value, but beyond that the relationship is not obvious).

With regards to the second objective of our paper he asserts, "it is of course impossible now to know whether there exists a relation between RCMs' skill for the present-day climate and future climate-change projections," which is precisely why we examined whether or not such a relation exists in a historical context. He then notes, "the use of recent past climate changes to assess RCM performance has already been used to some extent," but then refers to a study by *Lorenz and Jacob* [2010] in which different RCMs are driven by reanalysis fields rather than fields from coupled ocean-atmosphere models. The new *de Elia et al.* [2013] study he also cites indeed includes recent past climate change simulations driven by GCM boundary conditions, though it neither addresses the added-value issue nor compares skill in capturing climatology versus climate change (we hope the model simulations presented in that study will be analyzed to address these issues in the future). Nonetheless, we are gratified to see that this study, which appeared after our work was published and so was not mentioned in our original paper, follows a method similar to the one we proposed.

Discussing the skill of regional versus global models, Laprise also writes that "One may expect improvements only if there are improved representation of changes in regional forcings, such as aerosols and land-use changes." These were not included in our study, which imposed changes only in the boundary conditions. This was a deliberate choice since, as we discussed in our original paper, we were aware of only a single study examining dynamical-chemistry-aerosol downscaling, and so we instead examined the far more common case of purely dynamical downscaling [e.g., *de Elia et al.*, 2013]. Given the paucity of the type of study suggested by Laprise, in our opinion it remains premature to conclude that specification of regional forcings is required.

Laprise also suggests that the optimal configuration of WRF is the one that best reproduces observations when driven by reanalysis fields (accusing us of "admitting" we did not do this). Again, Laprise offers no evidence to support this assertion, and we do not believe that such a configuration is obviously the ideal one. If there were not substantial uncertainties in many of the physical processes being modeled, there would not be multiple configurations of WRF. It is entirely possible that WRF driven by reanalysis might produce a better match to observations for the wrong reasons since we cannot yet constrain the accuracy of the alternate physics packages. Thus, we believe that it is a valid decision to choose the physics version of WRF that produces a realistic climate when driven with fields from the global model we used, as clearly obtaining a realistic climatology is an important part of regional modeling. Our entire argument is not that climatology is not important, but rather that it is not a sufficient test of regional models.

While most prior regional climate modeling studies did not evaluate the ability of the models to reproduce forced responses, we are gratified to see that, as Laprise reports, some of the current regional modeling work is moving in this direction. We hope that future studies will test the issues raised in this comment and reply in detail, examining the impact of different regional forcings (aerosols, land use, etc.) and physical parameterizations in RCMs, testing the number of ensemble members needed to successfully capture changes for various climate parameters, and examining the significance of precipitation changes (which was beyond the resources available to us for this reply). Further work could also explore the ability of RCMs to improve on simulations of extreme events, a point raised both in our original paper and in the comment of Laprise, although in that case historical trends may very well not be significant, which would make it impossible to validate the response of RCMs to forced climate changes (Laprise states that "Previous studies have shown RCMs to improve not so much the mean climate but the frequency distribution and representation of extremes," citing the IPCC AR4, but again this evaluated climatology rather than climate change). Hence, both our original paper and this exchange highlight that a great deal of additional research remains in order to clearly determine when RCMs can provide added value for climate change projections and with what experimental setup they can best do so.

We emphasize, however, as in the original paper, that the time/resource-consuming nature of such research needs to be weighed against the fact that coupled ocean-atmosphere models are not only getting more and more sophisticated and holistic in their representation of the earth system, but closer in resolution to those used in typical RCM simulations. And because coupled ocean-atmosphere models are the primary source of information on future climate change, upon which we found WRF only modestly improves, high priority should be given within the climate modeling community to improving the long-range skill of global coupled ocean-atmosphere models upon which both global and regional modeling rests while continuing to investigate the added skill of RCMs.

Errata: In Table 1 of our original paper, the analysis-nudging column of simulation #2 should be *no*, whereas for simulation #3 it should be *yes*. Correspondingly, line #2 of paragraph #17 should read: "Note the use of analysis-nudging in simulations 3 and 4."

Acknowledgments

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, and we thank the climate modeling groups for producing and making available their model output.

References

- de Elía, R., S. Biner, and A. Frigon (2013), Interannual variability and expected regional climate change over North America, *Clim. Dyn.*, doi:10.1007/s00382-013-1717-9.
- Hegerl, G. C., F. W. Zwiers, P. Braconnot, N. P. Gillett, Y. Luo, J. A. Marengo-Orsini, N. Nicholls, J. E. Penner, and P. A. Stott (2007), Understanding and attributing climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 663–746, Cambridge Univ. Press, Cambridge, New York.
- Lorenz, P., and D. Jacob (2010), Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40, *Clim. Res.*, 44, 167–177, doi:10.3354/cr00973.
- Maraun, D. (2013), When will trends in European mean and heavy daily precipitation emerge?, *Environ. Res. Lett.*, 8, 7, doi:10.1088/1748-9326/8/1/014004.
- Racherla, P. N., D. T. Shindell, and G. S. Faluvegi (2012), The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models, *J. Geophys. Res.*, 117, D20118, doi:10.1029/2012JD018091.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An Overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1.