AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL076297

Key Points:
- Internal variability at regional scales can be characterized consistently with a large ensemble of one model or multiple small ensembles
- Estimates of uncertainties due to model disagreement are improved when multiple ensemble members are available for each model and scenario

Supporting Information:
- Supporting Information S1

Correspondence to:
N. Goldenson,
ngoldens@uw.edu

Citation:
Goldenson, N., Mauger, G., Leung, L. R., Bitz, C. M., & Rhines, A. (2018). Effects of ensemble configuration on estimates of regional climate uncertainties. *Geophysical Research Letters*, *45*, 926–934. https://doi.org/10.1002/2017GL076297

Received 3 NOV 2017
Accepted 28 DEC 2017
Accepted article online 4 JAN 2018
Published online 17 JAN 2018

Effects of Ensemble Configuration on Estimates of Regional Climate Uncertainties

N. Goldenson[1], G. Mauger[2], L. R. Leung[3], C. M. Bitz[1], and A. Rhines[1]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [2]Climate Impacts Group, University of Washington, Seattle, WA, USA, [3]Pacific Northwest National Laboratory, Richland, WA, USA

Abstract Internal variability in the climate system can contribute substantial uncertainty in climate projections, particularly at regional scales. Internal variability can be quantified using large ensembles of simulations that are identical but for perturbed initial conditions. Here we compare methods for quantifying internal variability. Our study region spans the west coast of North America, which is strongly influenced by El Niño and other large-scale dynamics through their contribution to large-scale internal variability. Using a statistical framework to simultaneously account for multiple sources of uncertainty, we find that internal variability can be quantified consistently using a large ensemble or an ensemble of opportunity that includes small ensembles from multiple models and climate scenarios. The latter also produce estimates of uncertainty due to model differences. We conclude that projection uncertainties are best assessed using small single-model ensembles from as many model-scenario pairings as computationally feasible, which has implications for ensemble design in large modeling efforts.

1. Introduction

Quantifying regional uncertainty in climate projections is important for understanding the forced response to anthropogenic activities and planning for climate change. Compared to global changes, regional changes can have higher uncertainty in projections, not only because of the relatively coarse resolution of global models and the uncertainties in model physics parameterizations but because internal variability can have a greater impact at the regional scale. To help quantify internal variability, several modeling centers have devoted substantial resources to creating large ensembles based on a single-model configuration (Deser et al., 2012, 2014; Kay et al., 2015). These simulations represent different realizations of possible outcomes due to internal variability in the model climate. As such, a large ensemble can be used to isolate the role of internal variability in the model climate.

At the regional scale, future trends can vary widely among ensemble members even when produced by the same model and when the projected global mean near-surface air temperature trends of each ensemble member are nearly the same (Deser et al., 2016). Additionally, differences among climate models can be amplified at the regional scale because of inadequacies in how processes are represented. For example, changes to the large-scale circulation patterns of the atmosphere are more challenging to predict than global mean temperature and yet are more directly relevant to understanding the implications of climate change at the regional scale (Seager et al., 2014; Xie et al., 2015). An ability to distinguish between uncertainty due to model representation and that due to internal variability can inform both risk assessment and further model development.

To evaluate uncertainties due to model disagreement in projections of climate change, a multimodel ensemble is needed. The Coupled Model Intercomparison Project (CMIP) has developed the experimental protocol and organized data for numerous multimodel climate studies. CMIP archives climate projections provided by modeling centers from around the world, following multiple emissions scenarios based on different socioeconomic development pathways. Additionally, some centers provide small ensembles, with up to about 10 members from the same model. Henceforth, we refer to these as single-model ensembles. Many studies of the CMIP projections focus on intermodel differences only, with more limited use of the single-model ensembles.

Few studies comprehensively evaluate both model disagreement and internal variance in uncertainty estimates. Here we address whether it is necessary to create a large single-model ensemble to quantify internal variance and where resources might best be allocated to constrain the range of possible future climate regionally, by identifying the greatest sources of uncertainty. We focus on regions that span the midlatitudes in an area strongly influenced by large-scale circulation. The relative importance of internal variability and the sensitivity to greenhouse warming are different for each region, as well as for different seasons. This allows for a more thorough comparison of approaches. We also compare internal variance with other sources of uncertainty in projections and how the balance among these changes on decadal time scales. We use the terms variance component and uncertainty component interchangeably, although we acknowledge that these have distinct meanings in other contexts.

Hawkins and Sutton (2009, hereafter HS09) introduced a method to quantify the portion of the uncertainty in climate change projections that can be attributed to internal variability relative to the uncertainty due to model disagreement and future greenhouse gas emissions. For internal variability, HS09 calculated the variance based on deviations from a fit to the climate variable of interest and hence can use a multimodel ensemble without any single-model ensembles. Methods like HS09 assume that all intermodel variance, after removing higher-frequency fluctuations, is due to model disagreement. While a valid approximation at the global scale, with regional differences averaged out, much of the intermodel variance at the regional scale can actually be explained by internal variability alone (Deser et al., 2012). As a result, the assumption of HS09 could lead to an overestimate of model uncertainty.

Subsequent studies have developed alternative approaches for partitioning the uncertainty into internal, model, and scenario sources. Yip et al. (2011) introduced a method based on the commonly used analysis of variance (ANOVA) statistical approach that makes use of the small single-model ensembles provided in the CMIP archive. This ensures more internal consistency than HS09; the total variance of a climate variable across the simulations will equal the sum of the variance contributed by each uncertainty source. Later work extended this approach to relax the requirement of a balanced data set (Northrop & Chandler, 2014, hereafter NC14); that is, the number of scenarios and ensemble members from each model does not need to be identical, so an entire CMIP database can be used, even if one model has more ensemble members than another.

This study compares approaches for estimating the internal variance and its contribution to the overall uncertainty in projecting regional climate changes. To address whether a large single-model ensemble is necessary to quantify internal variance, we combine a large single-model ensemble with a multimodel ensemble from CMIP, as described below. Our results can inform ensemble design for estimating internal variability, and hence significance of trends, in projected climate change at regional scale.

2. Data and Methods

Our primary region of interest is the northwestern United States, which is defined to include all model grid cells within 42°N to 49°N and 124°W to 111°W (hereafter "the Northwest"). Other regions along the West Coast of North America are defined similarly with their own latitude and longitude boundaries, as indicated. We have found that the variance estimates are within the confidence intervals whether we include or exclude the ocean grid cells, as time series that exclude them are highly correlated with those averaged over the entire region.

For a large single-model ensemble we use the Community Earth System Model large ensemble (CESM LE, Kay et al., 2015). We use 40 ensemble members, all following the RCP 8.5 emissions scenario (Riahi et al., 2011). For a multimodel, multiscenario ensemble of opportunity, we use the projections from phase 5 of the CMIP (CMIP5, Taylor et al., 2012) that span the entire 21st century. The CMIP5 ensemble contains small ensembles from many models across four scenarios based on different representative concentration pathways (RCPs, van Vuuren et al., 2011) for anthropogenic emissions. Additionally, we combine the two ensembles to consider the effect of adding a large single-model ensemble to the CMIP5 ensemble (hereafter the "combined ensemble"). In another case we use only one simulation per model/scenario pairing from CMIP5, combined with the CESM LE (hereafter the "reduced combined ensemble").

For each climate variable in our set of model projections, we first average monthly data over our region and then compute the time series of the annual or seasonal mean. Model grid spacing varies from as little as 70 km to as much as 350 km. Most model grids are in the range of 100–200 km across. We compute the anomaly for each projection by subtracting that model's historical ensemble mean (for 1970–1999) from all ensemble members being used by the method from the future time series. Then we derive variances as defined by each respective method.

Using the CESM LE alone, we apply a decadal running mean to the anomalies and then compute the variance across ensemble members as a function of time. Finally, we average the variance across time to get a single estimate of the internal variance.

To estimate multiple sources of uncertainty and their relative importance, we use multiple approaches, beginning with the HS09 method to calculate internal variability. The HS09 approach uses only a single ensemble member from each model in the CMIP5 ensemble for which all four scenarios are available. Internal variability is estimated from the residuals of a fourth-order polynomial fitted to the anomalies.The residuals are decadally smoothed, and the variance of the smoothed residuals is taken across all times and scenarios and finally averaged across models. The result is a time-invariant estimate for internal variance. We note that a fit to a fourth-order polynomial yields artificially high intermodel variance in the final decade due to a fitting artifact (see supporting information). This dispersion is more notable in time series with a lot of decadal variability and little trend but is less pronounced with a locally weighted scatterplot smoothing (LOWESS)-filtered time series in place of the polynomial fit because the LOWESS filter only includes local values in the fit. We adopt the LOWESS filtering method for this reason. Unlike HS09, we weight all models equally because assigning weights based on ability to simulate the historical period, when all models stray from reality, can be misleading (Stainforth et al., 2007). In another departure from HS09, we do not interpolate the projections to a common grid before calculating the regional means. Since our regions are relatively large, errors that may be introduced by differences in model grids should be small compared to the uncertainty in the variance estimates themselves.

We also use the method of NC14 to quantify internal variability using the CMIP5 ensemble, the combined ensemble, and the reduced combined ensemble. For any given time, the NC14 method treats each climate model value from each of the simulations as a sample belonging to a particular climate model and scenario, which are factors that control variance in the statistical model. Various statistical approaches are possible to partition the variance due to each source in some manner so that

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad (1)$$

is satisfied for each Y_{ijk}, where Y_{ijk} is the deviation from the mean across all samples (μ). The parameter α is the deviation from μ due to model i, β is the deviation from μ due to scenario j, and γ_{ij} is a nonlinear interaction term that depends on both model and scenario. The residuals (ϵ_{ijk}) that cannot be explained by any of the three previous factors are assumed to be due to internal variance.

Yip et al. (2011) were the first to take a more formal statistical approach than HS09, using a classical ANOVA. The ANOVA solution, however, requires that the data set be balanced — the same number of samples must be present for each model and scenario. The NC14 innovation is to use a random effects model (see, e.g., Searle et al., 2006) instead, which does not presume a balanced data set. While all of these approaches only approximate the true partitioning of uncertainties, we prefer the NC14 method because it can make use of all available information from all simulations that have been conducted across all models.

NC14 suggest two alternatives to estimate a solution to a random effects statistical model of the variance components. We adopt their restricted maximum likelihood (REML) statistical model to self-consistently estimate at once the variance due to internal variability, intermodel differences, scenario, and nonlinear interactions between model and scenario.

Maximum likelihood (ML) estimators assume that the available data are drawn from a specified probability distribution. The likelihood function is based on a multivariate normal probability density. The available data constrain the search for the variances that maximize this likelihood function. REML is a variant on ML that maximizes the likelihood after fixed effects are already included, to account for the degrees of freedom in the fixed effects. These types of estimation require an iterative search for coefficients for each of the variance components (e.g., model or scenario). In this application, the residual variance that cannot be explained by the fit is attributed to internal variance. Thus, all of the components of variance add up to the total variance, whereas there is no such constraint in HS09. The 95% confidence intervals are simulated by parametric bootstrapping — using simulated data based on the statistical model to recalculate the variance estimates 1,000 times.

The NC14 method includes a fourth nonlinear component for interactions between model and scenario, hereafter referred to as the "model-scenario interaction component." Physically, this can be interpreted in at least two ways. This component tends to be well defined (where the confidence intervals do not include 0) toward the end of the century. This could reflect nonlinearities in the response of the climate system itself to varying amounts of forcing from the different scenarios. In other words, model-dependent scenario differences, reflected in the potentially nonlinear model climate response to greenhouse gas forcing, would appear in the interaction component. The interaction component could also be interpreted as the component of model disagreement due to non–greenhouse gas forcing differences between the scenarios. Indeed, the scenarios represent different pathways with respect to greenhouse gas emissions, aerosol emissions, and land use-associated land cover change. The aerosol and land cover forcings do not increase monotonically from one scenario to the next like the greenhouse gas forcings do. Thus, differences in models corresponding to different aerosol-radiation or aerosol-cloud interactions and land-atmosphere interactions would appear as scenario-dependent model differences, indicated by the interaction term.

The NC14 method estimates variances after smoothing with a decadal running mean. When we apply the NC14 method to regional-scale precipitation time series, a decadal running mean is not sufficient to remove the substantial subdecadal variability that is present at the regional scale. As a result, we further smooth the time series of variance components at the end to improve readability of the plots. For this step, we apply a LOWESS filter, similar to that used in place of the fourth-order polynomial for the HS09 method. Versions of the figures without this extra smoothing are displayed in the supporting information along with confidence intervals.

The NC14 study used CMIP3 (Meehl et al., 2007) projections, with only three available scenarios, so they cautioned that an undersampling of socioeconomic scenarios could bias the results. The Bayesian alternative presented in NC14 specifies a weakly informative prior for the scenario uncertainty component, which we conclude is inappropriate for time series that exhibit little sensitivity of the trend to scenario, as is the case for precipitation in the regions we consider.

3. Results

In section 3.1 we begin by evaluating uncertainty estimates for temperature and precipitation in the Northwest from internal variability and the other uncertainty sources. Then, adopting the NC14 method, in section 3.2 we assess the influence of the ensemble type on the estimates. Finally, in section 3.3 we consider differences due to season and region.

3.1. Variance Components Across Methods

The internal variance of surface temperature for a particular region can be estimated in a variety of ways, as described in section 2. Perhaps the most straightforward approach is to compute the variance of the CESM LE; since it is a single model and scenario, the variance is entirely due to internal variability. Estimates of internal variance for temperature averaged globally and for the Northwest and several other regions are summarized in Figure 1. We take the variance among the decadal running mean time series of temperature from the CESM LE (denoted as σ^2 in the Figure 1 legend). This is comparable to the estimates of internal variance arrived at with the HS09 method.

Estimates of the internal variance with the NC14 method vary with time. However, in Figure 1 we give just the average across the entire period of the projections, which provides the most stable estimate since there is currently no evidence that internal variance will systematically increase or decrease with time. For the estimates using the NC14 method Figure 1 shows the 95% confidence intervals. Using the NC14 method yields internal variance estimates that are consistent with the simpler estimate with the CESM LE. The HS09 method estimates are frequently outside the range of the NC14 confidence intervals for the case with the highest total number of simulations (*N*, reported in the supporting information). Other seasons and regions have lower HS09 method estimates (see also Figure S1), suggesting that the HS09 method is not systematically biased high relative to the NC14 method. The internal variance estimates for the global annual mean temperature are slightly higher than the published values in HS09 and NC14 based on the CMIP3 ensemble.

The corresponding results for all variance components, based on the HS09 and NC14 methods, are shown as a function of year in the supporting information (Figures S4 for temperature and S5–S7 for precipitation). Figure 2 shows the temporal evolution of the variance components for the Northwest precipitation.

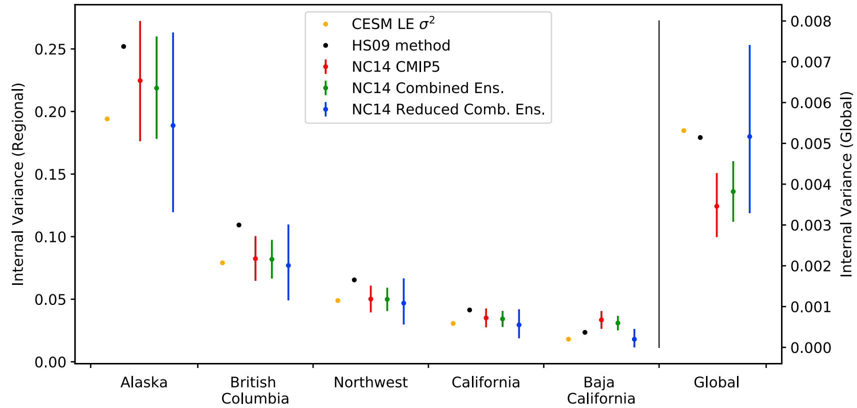

Figure 1. Internal variance estimates for various methods, data, and regions, as indicated, with 95% confidence intervals for the NC14 method.

In the Northwest there is very little trend in annual mean precipitation, so the scenario has very little effect, whereas internal variance is substantial.

In Figure 2 the variance by component is presented both as a fraction and in an absolute sense. Qualitatively, the HS09 method fractional variance in Figure 2e is reasonably similar to the NC14 method results in Figure 2f, but model disagreement is higher. For the HS09 method, some of the intermodel variance is almost certainly due to internal variance, as noted in section 1. With only one ensemble member per model/scenario combination, HS09 systematically conflates internal variance with model disagreement. This limitation does not affect the internal variance component in HS09 because it is estimated in an entirely independent manner.

3.2. Variance Components by Ensemble Choice

We choose to adopt the NC14 method to compare, in a consistent manner, internal variance estimates for temperature and precipitation in the CMIP5 ensemble, the combined ensemble, and the reduced combined ensemble. Figure 1 shows that the estimates of internal variance for near-surface air temperature are similar

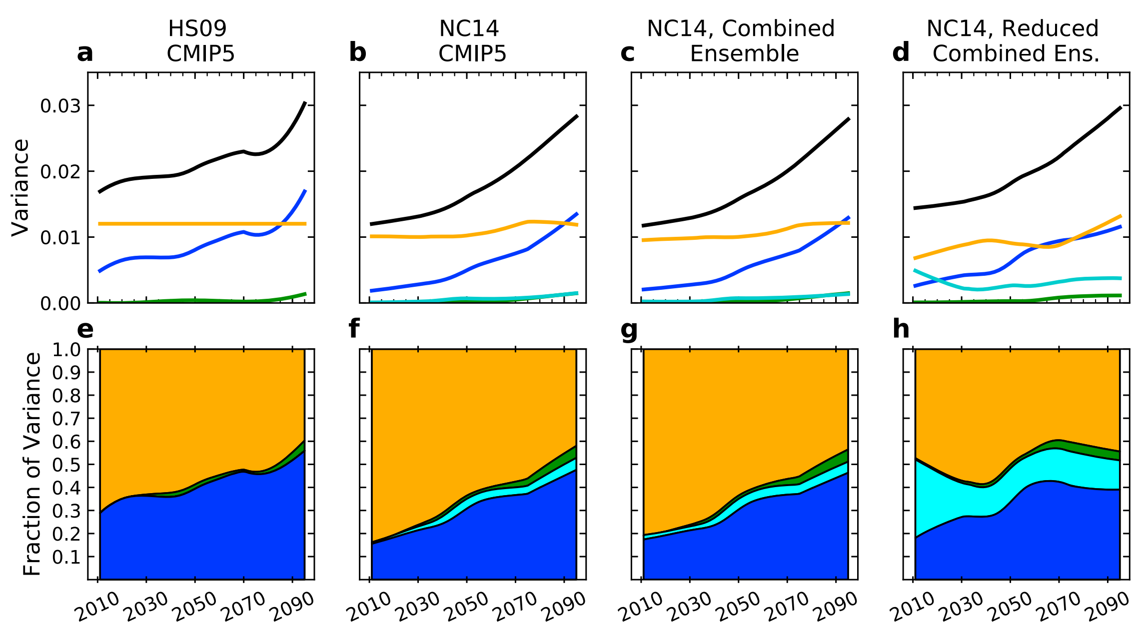

Figure 2. Comparison of methods for variance in model projections for the Northwest annual mean precipitation. Variance is presented (a–d) as total and by component in an absolute sense ((mm/day)^2) and (e–h) as a fraction of the total. Colors indicate component for model (blue), interaction (cyan), scenario (green), internal (orange), and total variance (black). Figures 2a and 2e show the modified HS09 method, Figures 2b and 2f show the NC14 method using CMIP5, Figures 2c and 2g show the NC14 method using the combined ensemble, and Figures 2d and 2h show the NC14 using the reduced combined ensemble.

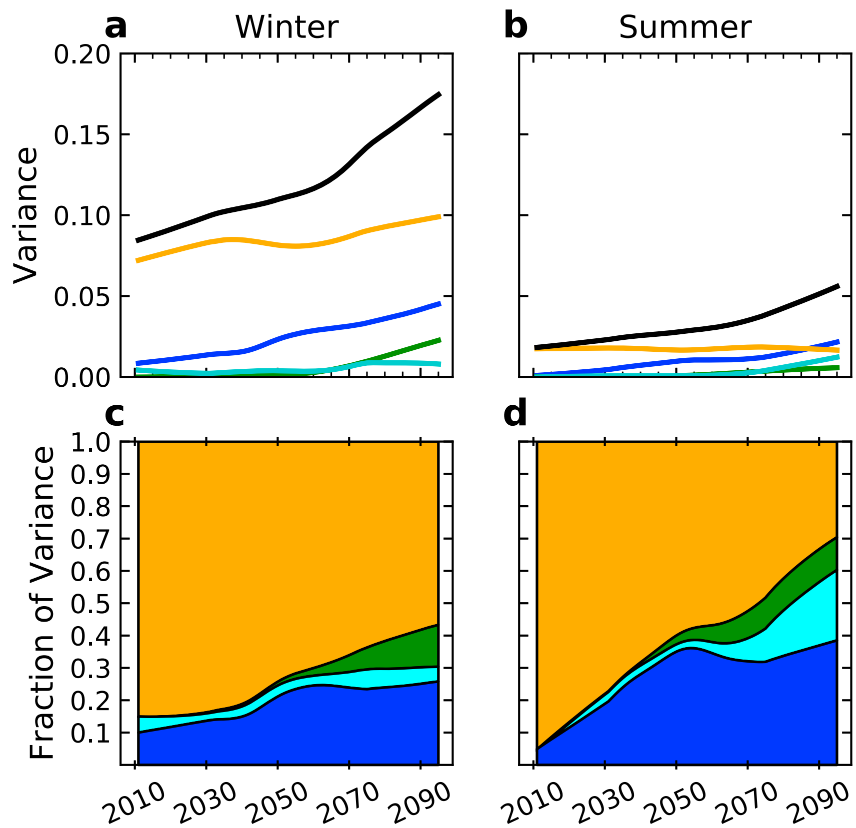

Figure 3. As in Figures 2c and 2g for the Northwest using the combined ensemble, but for (a, c) December–February mean precipitation and (b, d) June–August mean precipitation.

regardless of the ensemble(s) for the NC14 method, but the precision declines when the estimates are based only on the reduced combined ensemble.

In the NC14 method, adding the CESM LE to the CMIP5 ensemble makes little difference to the variance estimates. The internal variance is already well characterized on the basis of all of the small single-model ensembles in the CMIP5 ensemble, from which we have a total of 288 integrations. Including the CESM LE adds another 40, only ~14% more integrations; thus, the difference between Figures 2b and 2f and Figures 2c and 2g is small.

The variance estimates for precipitation provide a strong example of the importance of having small single-model ensembles. In the reduced combined ensemble, shown in Figures 2d and 2h, the differences from the other columns are notable. Because the behavior of each model is constrained by only one ensemble member, it cannot be determined whether the differences between models are due in part to internal variance. Because the reduced combined ensemble has only one ensemble member per model for every model except one, intermodel variance could at times be inflated. The HS09 method is also based on only one ensemble member per model for each scenario. Although the NC14 method has a more sophisticated way of calculating the variance components, it can still misattribute variance to model differences or internal variability if there are insufficient ensemble members to reliably distinguish between the two. Additionally, in Figures 2d and 2h we see an inflated interaction component in all years, which can be taken as an indicator that the reduced combined ensemble is too unbalanced in number of single-model ensemble members to produce accurate estimates of the partitioning of variance.

Further evidence of the inadequacy of the reduced combined ensemble is the noisier estimates for intermodel variance in the unsmoothed version of Figure 2 (Figure S5 in the supporting information). With only one ensemble member per model/scenario combination for most models, decadal fluctuations influence the model disagreement component. The small single-model ensembles of CMIP5, on the other hand, can ensure that decadal fluctuations are not misattributed as a part of the model's forced response. Thus, small ensembles

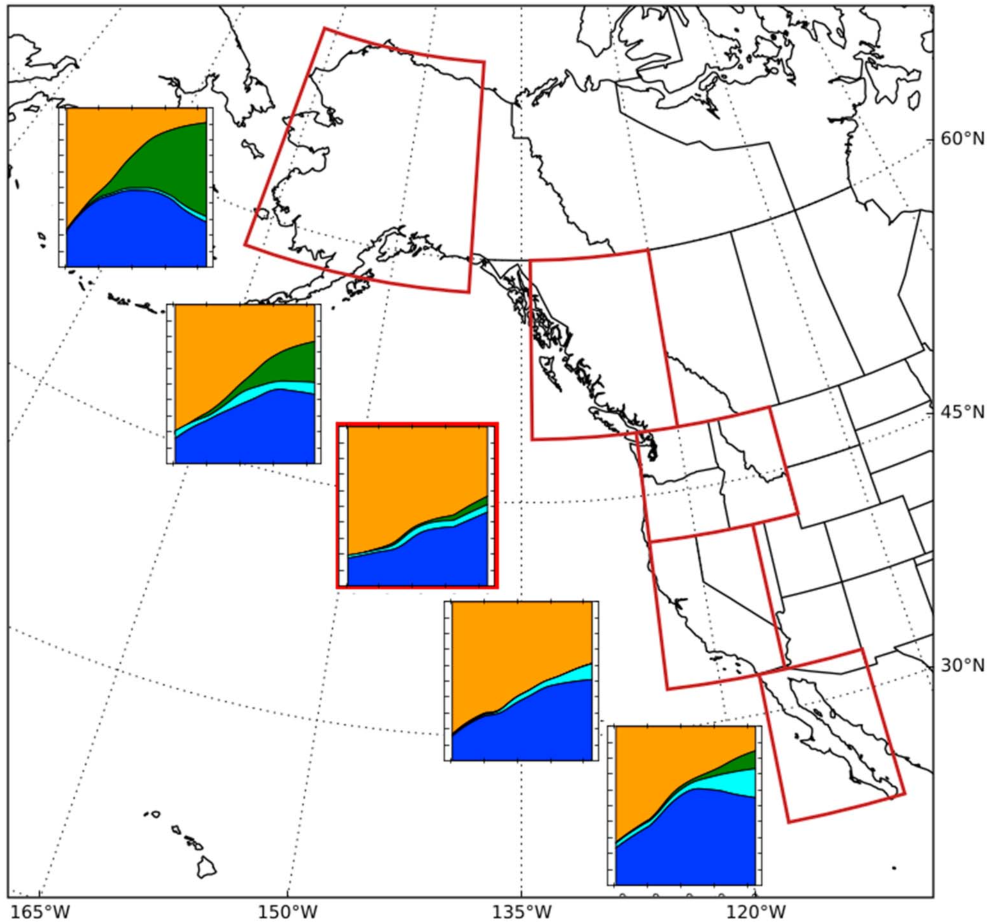

Figure 4. Fractional variance of regional annual mean precipitation for the five regions indicated by the latitude-longitude boxes shown in brown on the map. Fractional variance is plotted as in Figure 2g, to the left of the region to which each corresponds. The central plot (outlined in red) is a duplicate of Figure 2g for the Northwest. Axes of subplots as in Figures 2 and 3.

are most important for obtaining a robust estimate of the model disagreement component of uncertainty without convolving some internal variance into this estimate and also improving the accuracy of the internal variance estimate.

3.3. Season and Region

Precipitation trends in the Northwest are opposite in sign for winter (December–February) and summer (June–August): models generally show increases in winter and decreases in summer precipitation. Total precipitation, and its fluctuations, are much larger for winter precipitation. Figure 3, for the combined ensemble, has larger internal variance for winter. Although internal variance is lower in summer in an absolute sense, it occupies a similar fractional variance in both seasons and the annual mean. Total uncertainty is lower in an absolute sense in summer because both model disagreement and internal variance are lower. The larger contribution of the model-scenario interaction term for the Northwest during summer than winter could be related to the soil moisture-precipitation feedback in the eastern part of the region, which is a region of relatively high land-atmosphere coupling (Koster et al., 2004). The larger proportional variance due to scenario uncertainty and the model-scenario interaction term indicates that although projected trends in summer precipitation are small, they are just as important as internal variability by the end of the century.

To make sure that these results are robust across the topographically diverse western North America, in Figure 4 we compare two regions north of and two south of the Northwest. From north to south across this range of latitudes, the sign of the annual mean precipitation trend ranges from positive to negative. Regardless of the magnitude of the forced trend, the fractional contribution of internal variance to the overall uncertainty is large up and down the west coast of North America. Where the magnitude of the forced trend is larger, the scenario uncertainty occupies a larger fraction of the partitioned total variance. This is notable at the higher latitudes in Alaska and British Columbia where increases in precipitation are dominated by the "wet-get-wetter" thermodynamic response (e.g., Figure 3 of Seager et al., 2010) to warming, which varies with scenario. In the region that includes Baja California, the interaction term is larger in the second half of the century than at other latitudes, and we speculate that this could be due to model-dependent shifts in the Hadley circulation in response to the varying amounts of warming in each scenario.

4. Discussion and Conclusions

By applying the NC14 method to the CMIP5 ensemble, the combined ensemble, or the reduced combined ensemble, we were able to produce a summary of fractional contributions to the uncertainty in projections that properly accounts for the substantial internal variability at the regional scale. Noting the width of the confidence intervals on all of the variance components (see Figure 1 and supporting information), the estimates of internal variability produced by different methods are consistent with one another. If we only want to estimate internal variance, any of these would suffice. However, we caution that the fractional variance plots do not capture the uncertainty in these variance estimates and that it is substantial. Those data sets that include the small single-model ensembles (as in the CMIP5 and combined ensembles) do result in more precise estimates of model uncertainty and internal variance.

Our conclusions are based on a set of regions along the west coast of North America, where internal variability is a major source of uncertainty in temperature and precipitation projections. We emphasize the importance of a seasonal analysis where seasonal trends may have opposite signs. For example, during the Northwest summer, the projected trends are smaller than in winter, but the scenario and interaction terms become fractionally more important later in the century. The relative importance of each source of uncertainty also varies across the regions considered, but internal variance is important for all of them.

Using the CMIP5 ensemble alone can serve to quantify internal variance over spatially averaged domains as we have done here. To communicate the role of internal variance visually, however, large single-model ensembles are valuable because maps of results from individual ensemble members, which differ entirely due to internal variance, can be compared.

Compared to internal variance, model uncertainty estimates were more influenced by the method and ensembles used. We find that it is helpful to have small ensembles from as many models as possible to properly distinguish the forced response from decadal fluctuations, which the HS09 method and reduced combined ensemble fail to do. Properly characterizing disagreement is important to assess the current state of understanding and identify areas where additional model refinement is warranted.

Given improved process understanding and more accurate model representations, model disagreement is, in principle, reducible. While climate sensitivity explains intermodel differences in global mean temperature response, processes such as orographically forced mesoscale circulation can be more important at the regional scale. Varying model representations of El Niño and the jet stream should also affect model uncertainty in the midlatitude regions we selected. Additionally, models do not consistently predict how either of these climate features will change with global warming. Uncertainty in large-scale circulation features and cloud microphysical process representation could lead to significant model differences in large-scale precipitation projections as well.

For scenario uncertainty, we note that the NC14 method yields a greater fractional contribution of the emissions scenario to the total uncertainty than the HS09 method. The confidence intervals are wide for scenario uncertainty (see supporting information) because it is insufficiently constrained by only four discrete possibilities from the continuous distribution of possible emissions trajectories. It would also be interesting to consider the sensitivity of the scenario uncertainty estimate to additional scenarios that prescribe both higher and lower emissions, even if these are considered less likely trajectories. Projects like ScenarioMIP (O'Neill et al., 2016) could easily double the number of scenarios, offering the opportunity to narrow the range among estimates of scenario uncertainty.

In ensemble design there must be a balance between the computational expense of additional scenarios and model variants, and the number of realizations for each model/scenario pair. Stouffer et al. (2016) note that the number of ensemble members required varies with the field in question and is constrained by available computational and archiving capability. We have shown that to robustly quantify each component of the projection uncertainties in key temperature and precipitation metrics, ensemble design benefits from the inclusion of small single-model ensembles for many model/scenario combinations. Adopting a framework for quantifying uncertainty that can make use of all available information, like NC14, allows us to assess improvements in the state of knowledge as models develop and additional experiments are conducted.

Acknowledgments

This study was supported by the U.S. Department of Energy Office of Science Biological and Environmental Research (BER) as part of the Regional and Global Climate Modeling program. PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. We also thank Northrop and Chandler for making available their statistical model code. Python scripts used to process data and generate the plots are available at https://github.com/thenaomig/regionalVariance.

References

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, *2*(11), 775–779.

Deser, C., Phillips, A. S., Alexander, M. A., & Smoliak, B. V. (2014). Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *Journal of Climate*, *27*(6), 2271–2296.

Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, *29*(6), 2237–2258.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., … Yamada, T. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., … Mitchell, J. F. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, *88*(9), 1383–1394.

Northrop, P. J., & Chandler, R. E. (2014). Quantifying sources of uncertainty in projections of future climate*. *Journal of Climate*, *27*(23), 8793–8808.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., … Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., … Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1), 33–57.

Seager, R., Naik, N., & Vecchi, G. A. (2010). Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *Journal of Climate*, *23*(17), 4651–4668.

Seager, R., Neelin, D., Simpson, I., Liu, H., Henderson, N., Shaw, T., … Cook, B. (2014). Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over North America in response to global warming. *Journal of Climate*, *27*(20), 7921–7948.

Searle, S., Casella, G., & McCulloch, C. (2006). Variance components. Hoboken, NJ: John Wiley.

Stainforth, D. A., Allen, M. R., Tredger, E. R., & Smith, L. A. (2007). Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 2145–2161.

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., & Taylor, K. E. (2016). CMIP5 scientific gaps and recommendations for CMIP6. *Bulletin of the American Meteorological Society*, *98*(1), 95–105.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5–31.

Xie, S.-P., Deser, C., Vecchi, G. A., Collins, M., Delworth, T. L., Hall, A., … Watanabe, M. (2015). Towards predictive understanding of regional climate change. *Nature Climate Change*, *5*(10), 921–930.

Yip, S., Ferro, C. A. T., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate*, *24*(17), 4634–4643.