Reliability of multi-model and single-model ensembles

Tokuta Yokohata[†]; James Annan; Julia Hargreaves; Charles Jackson; Michael Tobis; Mark Webb; Mathew Collins
[†] National Institute for Environmental Studies, Japan
Leading author: yokohata@nies.go.jp

In order for our society to efficiently mitigate and adapt to climate change, it is necessary to have climate projections accompanied by assessments of the uncertainty in those projections using climate model ensembles. There are a large number of methods to evaluate the performance of model ensembles. These methods generally use one of two paradigms. One paradigm assumes that the truth should be close to the centre of the ensemble members (ie. close to the ensemble mean), and many studies of climate model evaluation in the literature have been based on this paradigm. On the other hand, an alternative paradigm is to consider the truth as being drawn from the distribution sampled by the ensemble. In this case, the model ensemble can be regarded as perfect if the ensemble members and the truth are "statistically indistinguishable". This idea is common in the field of weather forecasting but not common in the evaluation of climate model ensemble so far. In this paradigm the reliability of model ensembles can be evaluated through the "rank histogram" approach whereby the distribution of the observed occurrence of an event in the prediction ensembles is evaluated. Such an analysis can reveal if prediction ensembles are too narrow, too broad, or biased. In the present paper, we evaluated the reliability of multi-model ensembles (MME, we used the CMIP3 ensemble so far) through the rank histogram approach. While the MME samples uncertainties in model structure because it is consist of models developed by world research institute, each model has one parameter set, and thus it is often referred to as an "ensemble of opportunity". Therefore, there is a concern that the MME cannot have enough range of spread, and cannot provide a realistic probabilistic range of future climate change. For this reason, we compared the reliability of the MME to the other kind of ensemble, single-model ensemble (SME, sometimes also referred to a perturbed physics ensembles). In SMEs, a range of different parameter values are used for the generation of ensemble, and parameter values may be set to rather extreme values (compared to those used in the MME) so as to generate a wide range of responses. Here, we used four different SMEs (generated by HadCM3/SM3, NCAR CAM, and MIROC) in order to get general conclusions which are not dependent on the structures of models used in the SMEs. Our analysis reveals that, in the MMEs, climate variables are broadly reliable on the global scale, with a tendency towards overdispersion. On the other hand, in the SMEs, the reliability differs depending on the ensemble and variable field considered. In general, the mean state and historical trend of surface air temperature, and mean state of precipitation are reliable in the SMEs, but variables related to model dynamics, such as sea level pressure or clear-sky shortwave radiation, do not cover a sufficiently wide range. These results indicate that it is important to have various climate models to generate ensembles in order to get a realistic probabilistic range of future climate change, and also give us confidence in the research activities using the CMIP3 ensemble. We used the CMIP3 ensemble for the analysis but will include the CMIP5 ensemble in the presentation if possible.