Abstract:
This talk discusses new methods for diagnosing differences in skill and predictability between forecasts in multi-model ensembles, particularly the North American Multi-model Ensemble (NMME). We first discuss an objective, rigorous method for deciding whether the multi-model mean forecast has significantly larger skill than an individual forecast. Applying this method to the NMME demonstrates that the skill of the NMME is significantly higher than that of any single model for most lead times and target months. We further clarify that this improvement comes from including different signals from different models, rather than from the decrease in noise due to increasing the ensemble size. We also discuss rigorous methods for comparing noises and signals across model forecasts, which provide the basis for quantifying differences in skill and predictability between forecasts. These results are used to isolate models that perform significantly different from other models. Identification of significant differences between forecasts may help forecasters judge which models to trust in conflicting situations, and to assist model development by clarifying shortcomings or advantages of particular model forecasts.

End