Testing the Stability of Cross-Validated Super-Ensemble Weights

A supplement to:

Superensemble Statistical Forecasting of Monthly Precipitation over the Contiguous United States, with Improvements from Ocean-Area Precipitation Predictors

Thomas M. Smith, Samuel S.P. Shen, and Ralph R. Ferraro

June 3, 2016

 Computation of the relative weights for super-ensemble members is described in the main text. Those weights are cross validated because they are based on cross-validated correlations. There is one set of single-cross-validated weights for each month and for each statistical model in the ensemble. Here the stability of those weights is shown by performing double-cross-validation of the weights. Double-cross-validation weights are computed for each year in the record by withholding the cross-validation results for the year of interest when computing correlations and weights for months in that year.

 Comparison of the cross-validation and double-cross-validation weights is used to evaluate stability of the weights. Since double-cross-validation weights for each year, we here use the mean and standard deviation of those weights over all of the years in the record to show their typical values and how much they may change from year to year.

 For this evaluation we use one statistical model from the 18-model ensemble: the CCA using tropical Pacific area SST predictors, here referred to as CCA(SST1). We show comparisons for January (Fig. s1) and July (Fig. s2) weights. For both months the mean of the double-cross-validation weights is essentially the same as the cross-validation weights. In addition, there is little variation among the double-cross-validation weights as shown by the low standard deviation. Some year-to-year variation in the double-cross-validation weights is expected from withholding one of the 18 cross-validation years in the record. A longer record would be better, but the satellite record length is limited. However, the low standard deviation in this test shows that the pattern of the cross-validation weights is stable for this record length. Since the super-ensemble forecast is based on these weights, the stability of the weights suggests that the forecast should also be stable. That basic stability is confirmed by the bootstrap analysis discussed in the main text.



Figure s1. January weights for the CCA(SST1) model from cross-validation (upper) and the average from double-cross-validation (middle). The standard deviation of the double-cross-validation weights is also shown (lower panel).



Figure s2. July weights for the CCA(SST1) model from cross-validation (upper) and the average from double-cross-validation (middle). The standard deviation of the double-cross-validation weights is also shown (lower panel).