Week 3-4 Multi-Model Ensemble Subsampling: A Real-time Verification

Cory Baggett[1], Steven Simon[1,2], Michael Halpert[1]

[1] NOAA/NWS/NCEP/Climate Prediction Center, [2] Earth Resources Technology, Inc.

cory.baggett@noaa.gov / steven.simon@noaa.gov

1. Motivation

Forecasters at the Climate Prediction Center (CPC) produce probabilistic Week 3-4 outlooks for above or below normal temperature and precipitation. To produce these outlooks, a wide array of experimental tools are at their disposal. These experimental tools usually undergo a few years of real-time testing before they are moved into operational status. A key test an experimental tool must pass is that it adds skill over existing tools, thus improving CPC's Week 3-4 outlooks for stakeholders. One such experimental tool that forecasters have been utilizing since November 2020 is known as the Ensemble Subsampling Tool.

The goal of the Ensemble Subsampling Tool is to create a *subsample forecast* that improves upon the multi-model *all-member forecast* by objectively subsampling ensemble members in real-time.

2. Subsampling Method

The Ensemble Subsampling Tool makes a subsample forecast by objectively selecting ensemble members from a combined 185 ensemble members available from the following subseasonal dynamical models: the ECMWF, GEFSv12, CFSv2, ECCC, and JMA. The members are subsampled by choosing the members that have the highest pattern correlation between their Week 2 500-hPa height anomaly forecast and CPC's Week 2 500-hPa Autoblend height anomaly forecast, which is a simple multi-model ensemble weighted-mean of the ECMWF (50%), ECCC (25%), and GEFS (25%).

3. Example Forecast and Verification

As an example, Figure 1 displays the multi-model, 185-member all-member forecast issued 15-Feb-2022 for the Week 3-4 period valid 02-Mar to 15-Mar-2022. Figure 2 displays CPC's Week 2 500-hPa Autoblend for the period valid 23-Feb to 01-Mar-2022. In order to create a subsample forecast, the Week 2 500-hPa height anomaly forecast from each of the individual 185 members is compared to the Autoblend via pattern correlation analysis and are subsequently ranked from worst to best matching members.

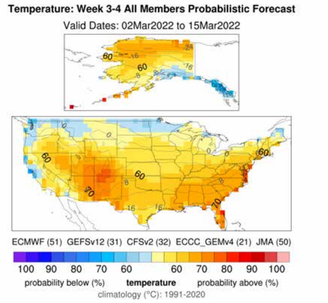

Figure 1. Multi-model, 185-member Week 3-4 probabilistic temperature all-member forecast issued on 15-Feb-2022.

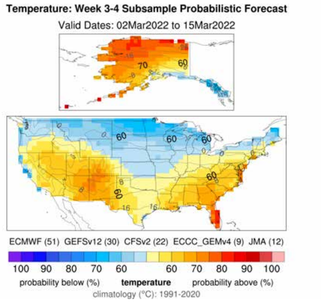

Figure 3. Multi-model, 124-member Week 3-4 probabilistic temperature subsample forecast issued on 15-Feb-2022.

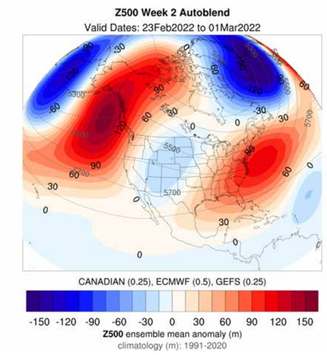

Figure 2. CPC's Week 2 500-hPa Autoblend height anomaly forecast issued on 15-Feb-2022. The Autoblend contains static weights: 25% ECCC, 50% ECMWF, and 25% GEFS.

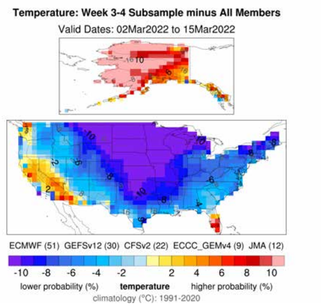

Figure 4. Difference in probabilities between the all-member and subsample forecasts issued on 15-Feb-2022. Positive values indicate that the subsample forecast has a higher probability of above normal temperatures than the all-member forecast, and vice versa.

3. Example Forecast and Verification <continued>

In Figure 3, the subsample forecast that retains the top 67% of the best matching members to the Autoblend is shown. Figure 4 highlights the difference in Week 3-4 temperature probabilities between the all-member forecast (Fig. 1) and the 67% subsample forecast (Fig. 3). For this forecast, the 67% subsample forecast has higher probabilities for below normal temperature across most of CONUS, such that its categorical forecast for the Northern Plains switches from above normal to below normal. Here, all 51 members of the ECMWF and 30 of 31 members of the GEFS were subsampled at the expense of fewer members being subsampled from the ECCC and JMA.

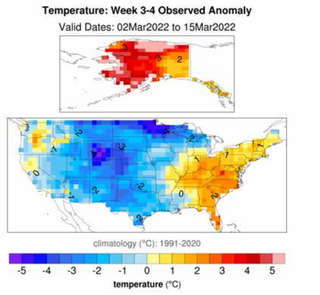

Figure 5. Observed anomalous temperature for the two-week period of 02-Mar to 15-Mar-2022, corresponding to the Week 3-4 forecast issued 15-Feb-2022.

Figure 5 provides the observed anomalous temperature for the forecast valid period, 02-Mar to 15-Mar-2022. Below normal temperatures across central CONUS verified, which resulted in the 67% subsample forecast having a Heidke Skill Score (HSS) of 19.8 in comparison to the all-member forecast having a worse HSS of -14.0.

4. Verification Statistics

Figures 6 and 7 show temperature verification statistics for the past 191 real-time Week 3-4 forecasts, issued on Tuesdays and Fridays, beginning 03-Nov-2020. Figure 6 provides the HSS as a function of the percentage of best members retained in the subsample. The all-member forecast has an average HSS of 29.4, the 67% subsample forecast has an HSS of 29.9, the 50% subsample forecast has an HSS of 29.4, and the best performing subsample forecast occurred at 59%, with an HSS of 30.2.

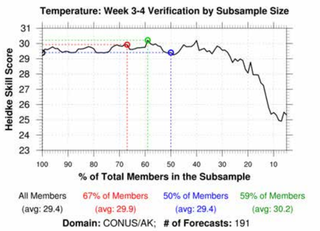

Figure 6. Averaged HSSs for Week 3-4 temperature forecasts as a function of the percentage of members retained in the subsample forecast. Here, 100% represents the all-member forecast.

Figure 7 assesses the statistical significance of the subsample forecasts versus the all-member forecasts via a sign test, displayed as a random walk. A random walk simply counts across forecast issuances the number of times the subsample forecast has a higher HSS or lower HSS versus the all-member forecast, assigning a +1 when it wins and a -1 when it loses. If the subsample forecast "walks" outside the gray areas, it is considered statistically distinguishable from the all-member forecast.

4. Verification Statistics <continued>

While the subsample forecasts performed significantly better early during the real-time period, they have not shown much of an advantage recently. However, the 67% subsample forecast appears to offer the most value during extended winter, as indicated by its positive slope during both 2020-21 and 2021-22.

Figures 8 and 9 provide similar random walk plots for precipitation and 500-hPa heights, respectively. For precipitation, the 81% subsample forecast has had the best performance, although it has performed poorly during recent verifications. For 500-hPa heights, the 69% subsample forecast has significantly outperformed the all-member forecast, having won 27 more times than losing.

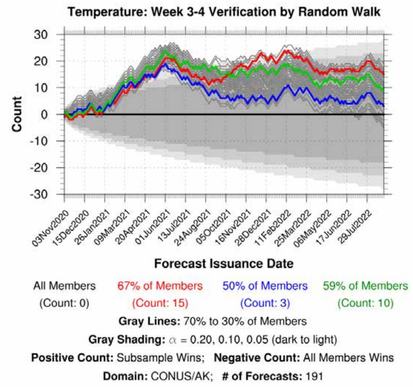

Figure 7. Sign test, or random walk, of the Week 3-4 temperature subsample forecasts versus the all-member forecast.

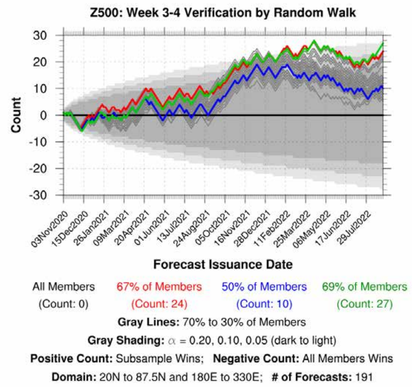

Figure 9. As in Fig. 7, but for subsample forecasts of 500-hPa heights.

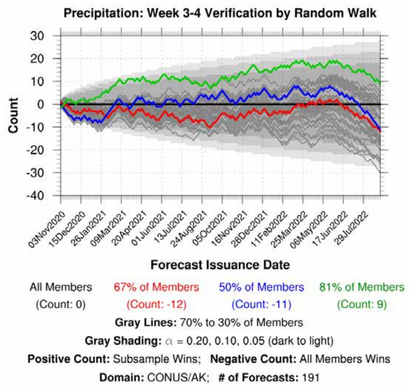

Figure 8. As in Fig. 7, but for subsample forecasts of precipitation.

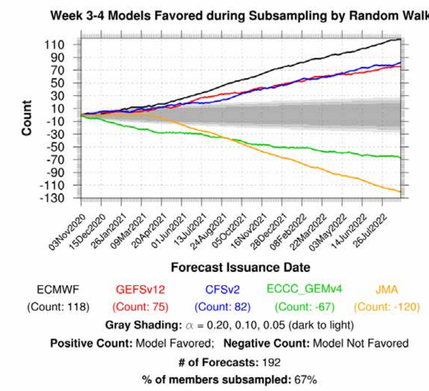

Figure 10. Sign test, or random walk, of which dynamical models are favored during subsampling.

5. Discussion, Conclusion, and Future Work

One question that often arises is how often a particular dynamical model has its ensemble members subsampled. In Fig. 10, we count across forecast issuances how often a given dynamical model is favored when creating a 67% subsample forecast. Here, if more than 67% of a given model's ensemble members are subsampled then it is considered favored (+1) and if less than 67% are chosen, then it is not favored (-1). The subsampling process clearly favors the ECMWF, GEFSv12, and CFSv2 while often rejecting members from the JMA and ECCC. While not shown, the JMA and ECCC are rejected most on Tuesday issuances of the Week 3-4 forecast. This occurs because the JMA and ECCC are only initialized once per week, and it happens that Tuesday issuances of the Week 3-4 forecast do not have recent initializations of these models available.

While the verification results are not entirely conclusive that the subsample forecast outperforms the all-member forecast, the Ensemble Subsampling Tool is nonetheless moving into operational status during 2023. Week 3-4 forecasters at CPC have regularly used the experimental tool to inform their forecasts, as it often provides an important link between CPC's Week 2 and Week 3-4 products. Further, it seemingly adds the most value during extended winter, which is often challenging due to high, intrinsic variability.

The Ensemble Subsampling Tool will also be maintained in an experimental mode 1) to serve as a back-up should the operational tool fail, 2) to convert the tool from 2-category to 3-category forecasts, 3) to determine the optimal subsample size, and 4) to test additional methods to achieve subsampling.