



Supplement of

The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers

Kean Foster et al.

Correspondence to: Kean Foster (kean.foster@smhi.se)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.



Figure S1. A scatterplot of the MAESS vs Δ NSE scores for the different combinations of the different individual modelling chains. The shaded quadrants denote the respective areas of skill for the different scores.



Figure S2. Bootstrapped (N = 10000) FY+, MAESS, and Δ NSE scores for ME_{hds} with respect to HE for all subbasins in the cluster S¹. Each subplot is a histogram of the medians of the bootstrapped validations scores for each initialisation month. Above the

5 histograms are six related statistics: (left of the red line) the maximum, mean, and minimum of the validation scores shown in the histograms; (right of the red line) percentages of the subbasins where ME_{hds} performed better than HE (n_{abs}^+) , the percentage of subbasins where ME_{hds} performed better than HE $(n_{0.1}^-)$ at the significance level 0.1, and lastly the percentage of subbasins where ME_{hds} performed worse than HE $(n_{0.1}^-)$ at the 0.1 level.



Figure S3. Bootstrapped (N = 10000) FY+, MAESS, and Δ NSE scores for ME_{hds} with respect to HE for all subbasins in the cluster S². Each subplot is a histogram of the medians of the bootstrapped validations scores for each initialisation month. Above the histograms are six related statistics: (left of the red line) the maximum, mean, and minimum of the validation scores shown in the histograms; (right of the red line) percentages of the subbasins where ME_{hds} performed better than HE (n⁺_{abs}), the percentage of subbasins where ME_{hds} performed worse than HE (n⁻_{0.1}) at the 0.1 level.



Figure S4. Bootstrapped (N = 10000) AROCSS for the lower, middle, and upper terciles between the ME_{hds} and HE for subbasins in the cluster S¹. Each subplot is a histogram of the medians of the bootstrapped validation score's ensembles for each initialisation month. Above the histograms are six related statistics: (left of the red line) the maximum, mean, and minimum of the validation scores shown in the histograms; (right of the red line) percentages of the subbasins where ME_{hds} performed better than HE (n_{abs}^+) , the percentage of subbasins where ME_{hds} performed better than HE $(n_{0.1}^+)$ at the significance level 0.1, and lastly the percentage of subbasins where ME_{hds} performed worse than HE $(n_{0.1}^-)$ at the 0.1 level.



Figure S5. Bootstrapped (N = 10000) Δ ROCSS for the lower, middle, and upper terciles between the ME_{hds} and HE for subbasins in the cluster S². Each subplot is a histogram of the medians of the bootstrapped validation score's ensembles for each initialisation month. Above the histograms are six related statistics: (left of the red line) the maximum, mean, and minimum of the validation scores shown in the histograms; (right of the red line) percentages of the subbasins where ME_{hds} performed better than HE (n⁺_{abs}), the percentage of subbasins where ME_{hds} performed better than HE (n⁺_{0.1}) at the significance level 0.1, and lastly the percentage of subbasins where ME_{hds} performed worse than HE (n⁻_{0.1}) at the 0.1 level.