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

K. Foster[1, 2], C.B. Uvo[2], J. Olsson[1]

[1]Research & Development (hydrology), Swedish Meteorological and Hydrological Institute, 601 76 Norrköping, Sweden
[2]Department of Water Resources Engineering, Lund University, Box 118, 221 00 Lund, Sweden

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

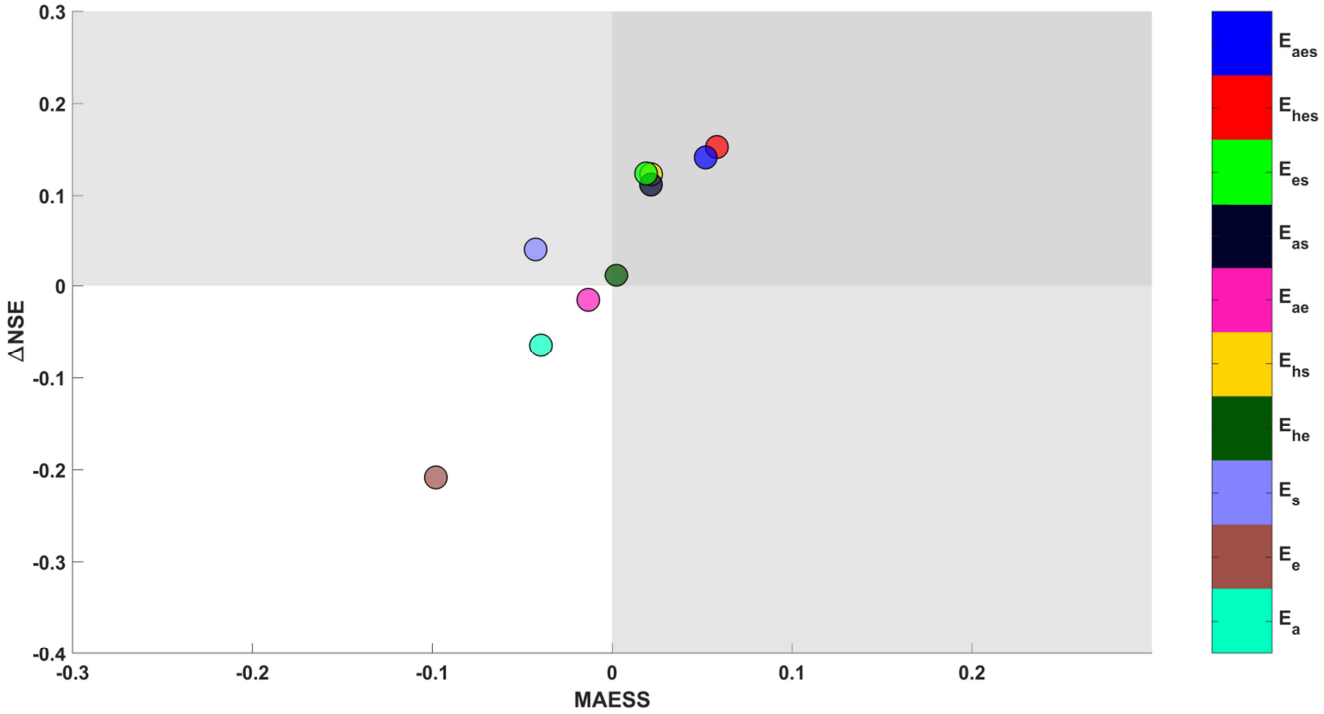


**Figure 1. A scatterplot of the MAESS vs ΔNSE scores for the different combinations of the different individual modelling chains. The shaded quadrants denote the respective areas of skill for the different scores.**

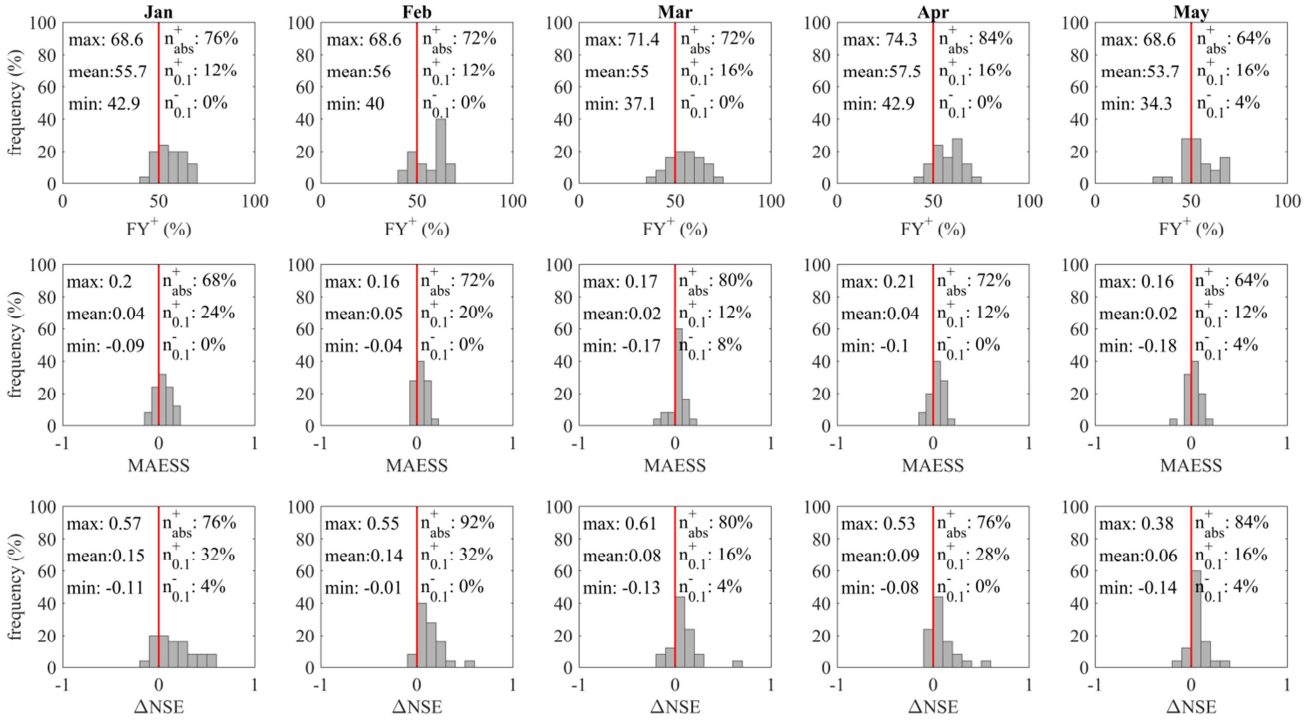


**Figure 2. Bootstrapped (N = 10000) FY+, MAESS, and ΔNSE scores for $ME_{hds}$ with respect to HE for all subbasins in the cluster $S^1$. Each subplot is a histogram of the medians of the bootstrapped validations scores for each initialisation month. Above the histograms are six related statistics: (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 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.**

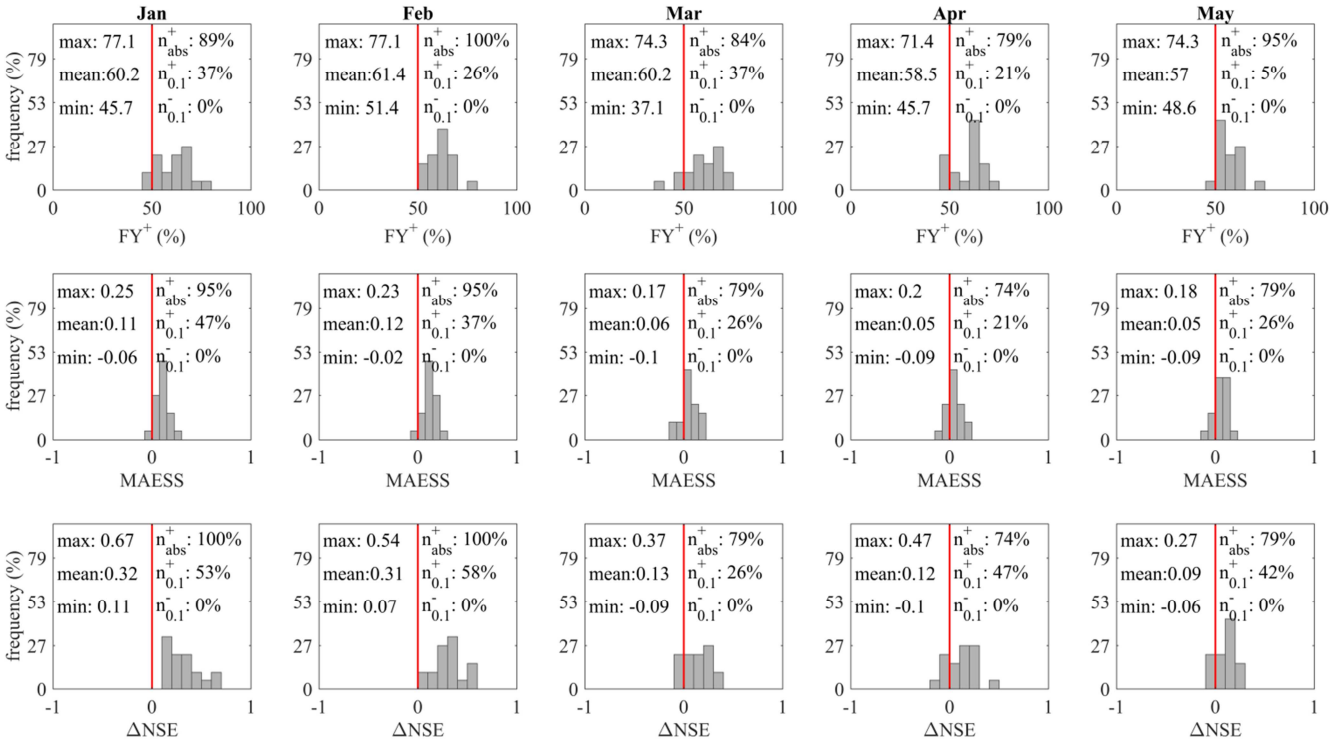


**Figure 3. Bootstrapped (N = 10000) FY+, MAESS, and ΔNSE scores for $ME_{hds}$ with respect to HE for all subbasins in the cluster $S^2$. Each subplot is a histogram of the medians of the bootstrapped validations scores for each initialisation month. Above the histograms are six related statistics: (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 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.**

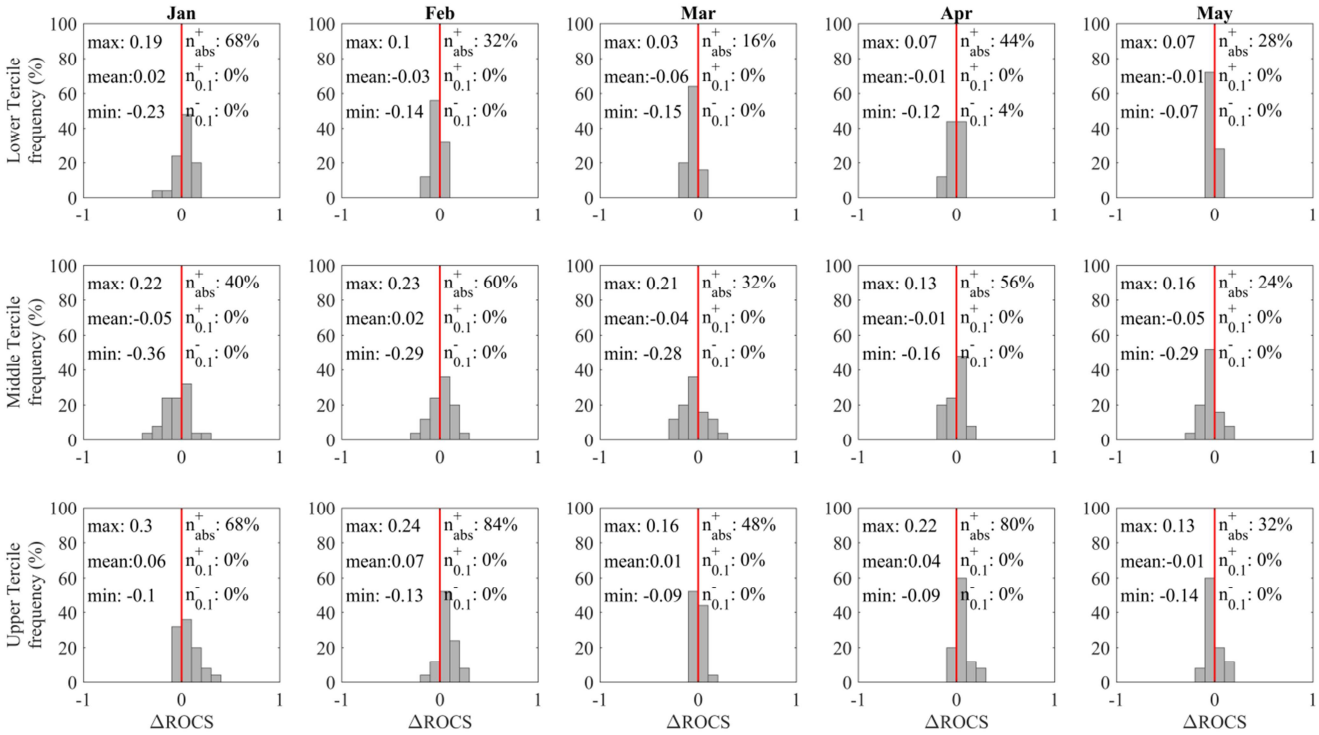


**Figure 4. Bootstrapped (N = 10000) ΔROCSS for the lower, middle, and upper terciles between the $ME_{hds}$ and HE for subbasins in the cluster $S^1$. Each subplot is a histogram of the medians of the bootstrapped validation score's ensembles for each initialisation month. Above the histograms are six 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.**

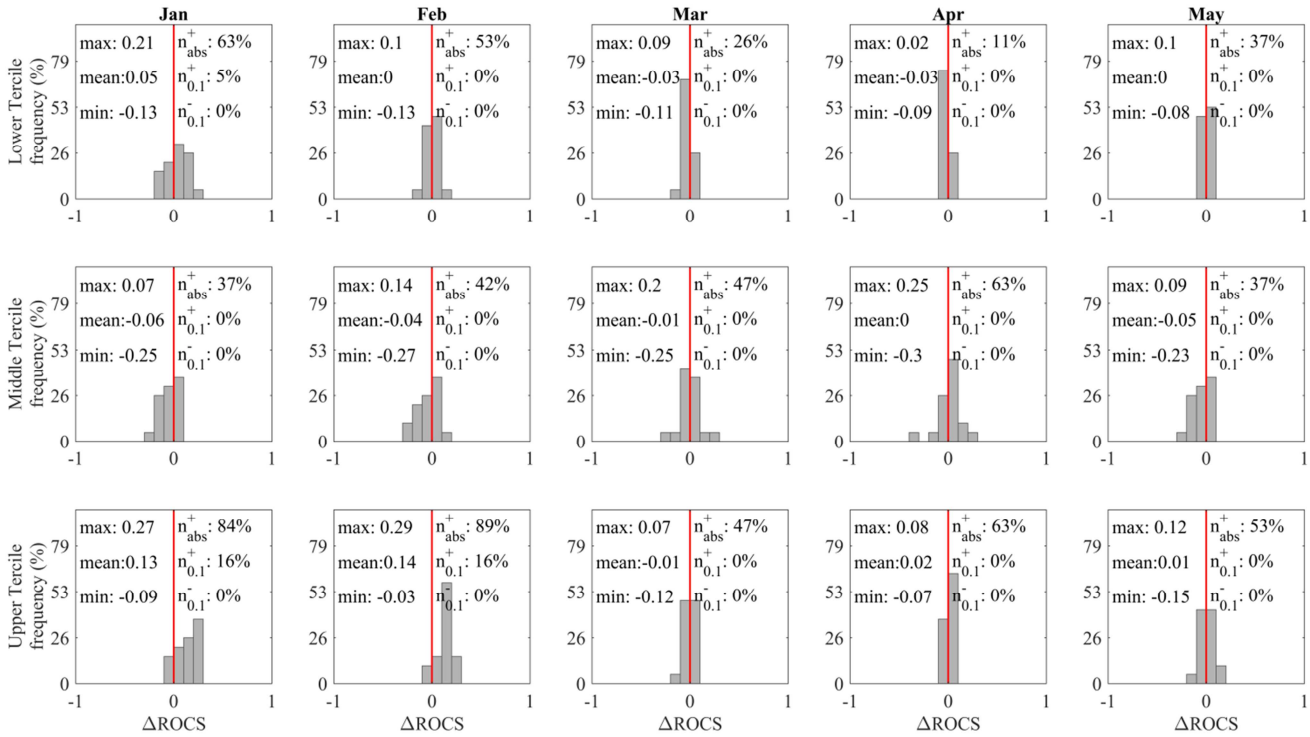


**Figure 5. Bootstrapped (N = 10000) ΔROCSS for the lower, middle, and upper terciles between the $ME_{hds}$ and HE for subbasins in the cluster $S^2$. Each subplot is a histogram of the medians of the bootstrapped validation score's ensembles for each initialisation month. Above the histograms are six 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.**