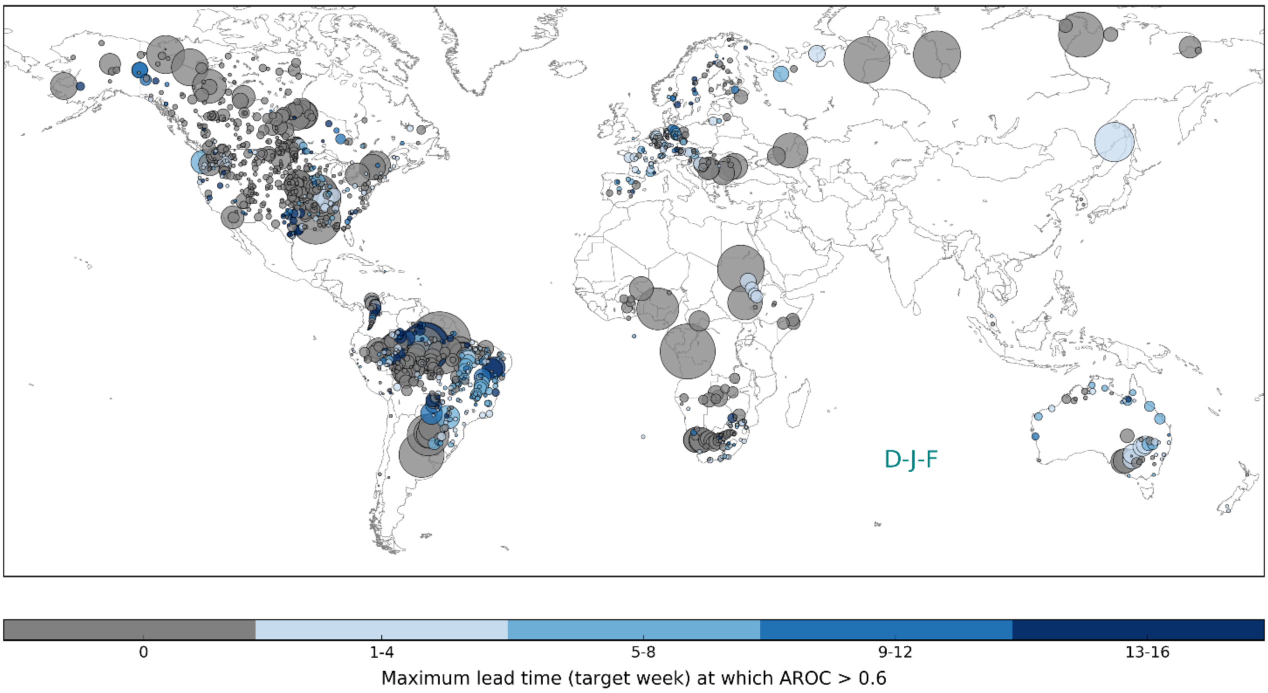
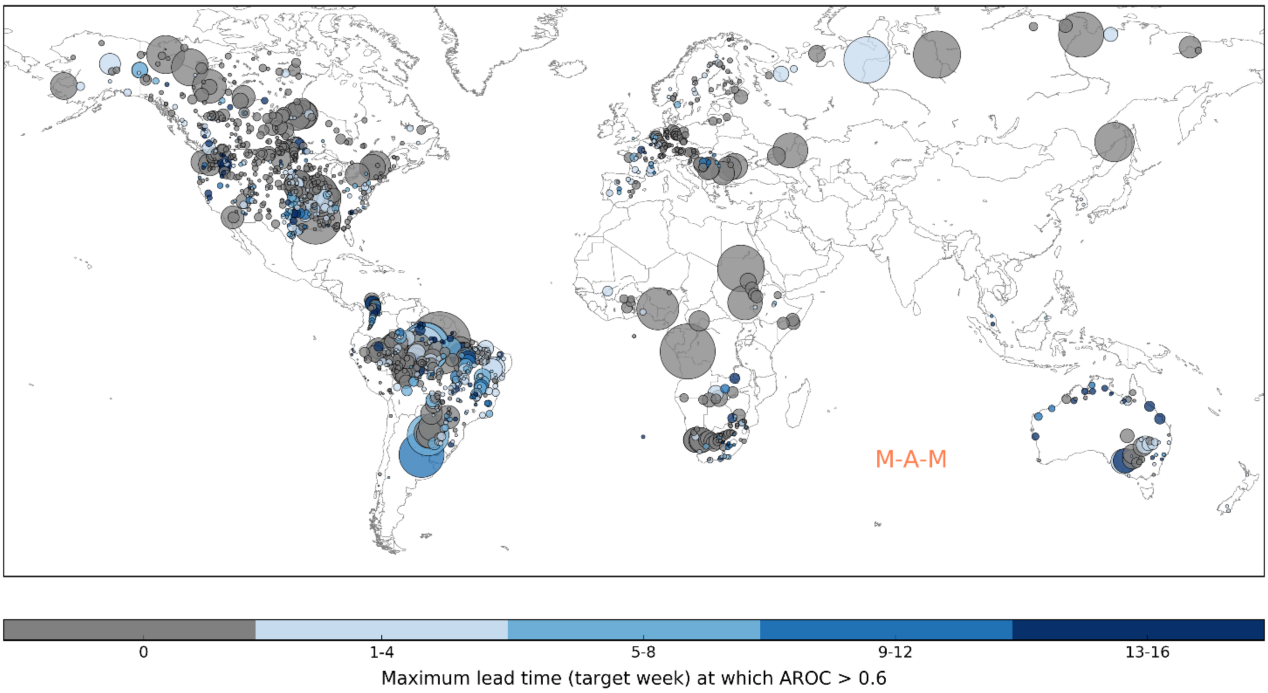
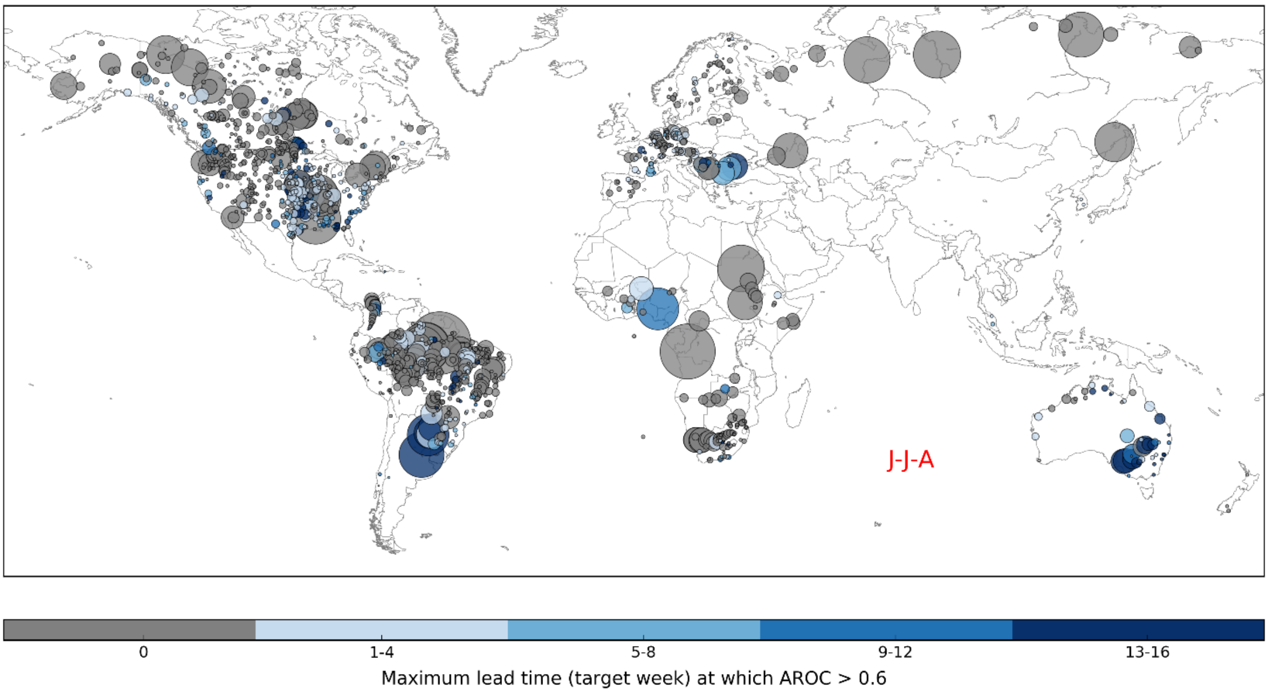

D-J-F
0
1-4
5-8
9-12
13-16
Maximum lead time (target week) at which AROC > 0.6


M-A-M
0
1-4
5-8
9-12
13-16
Maximum lead time (target week) at which AROC > 0.6



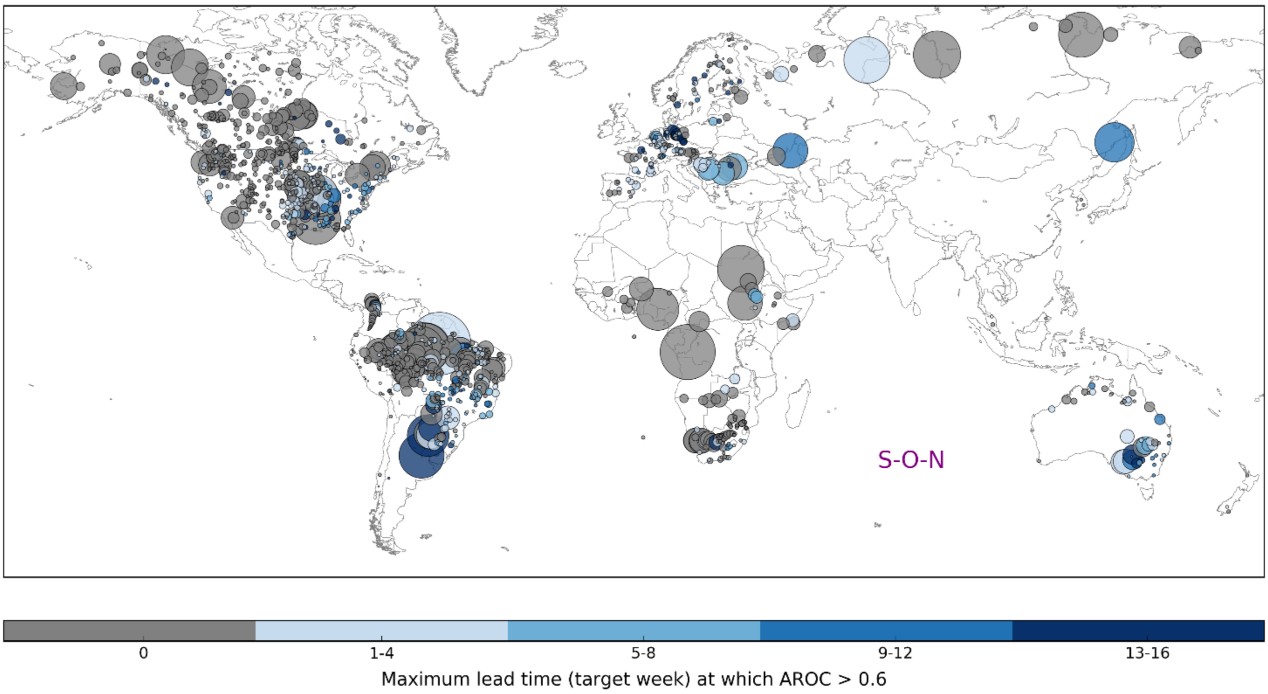
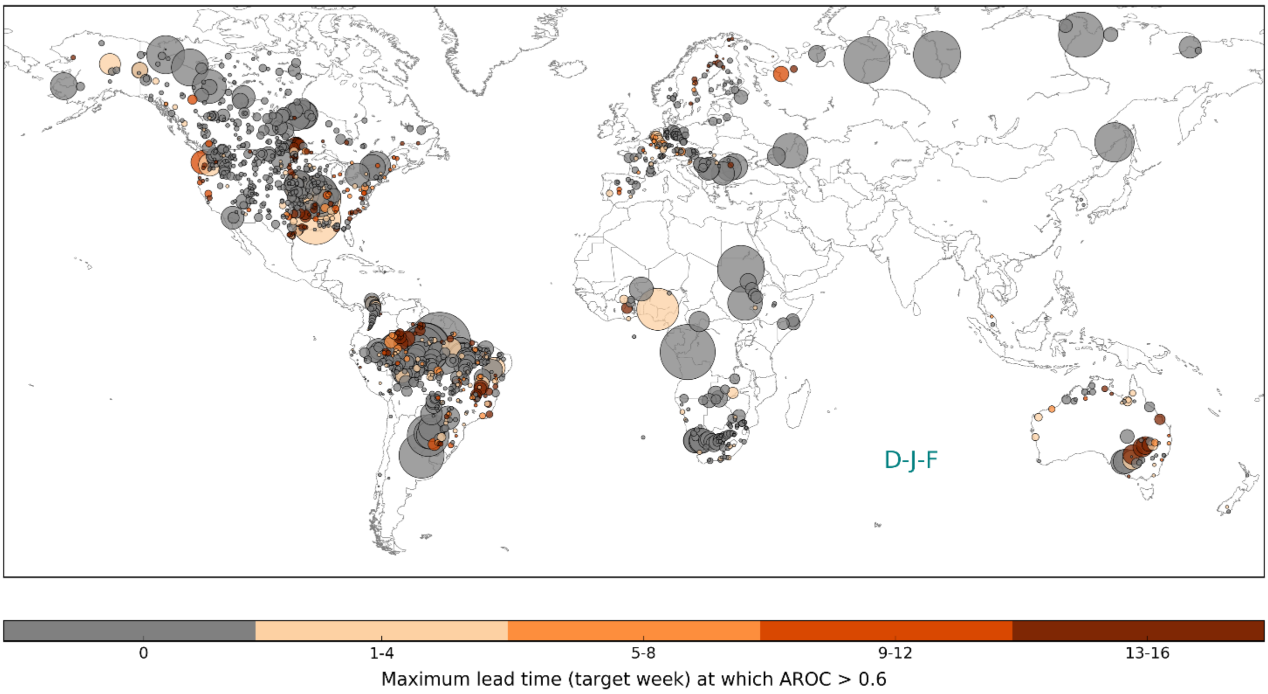
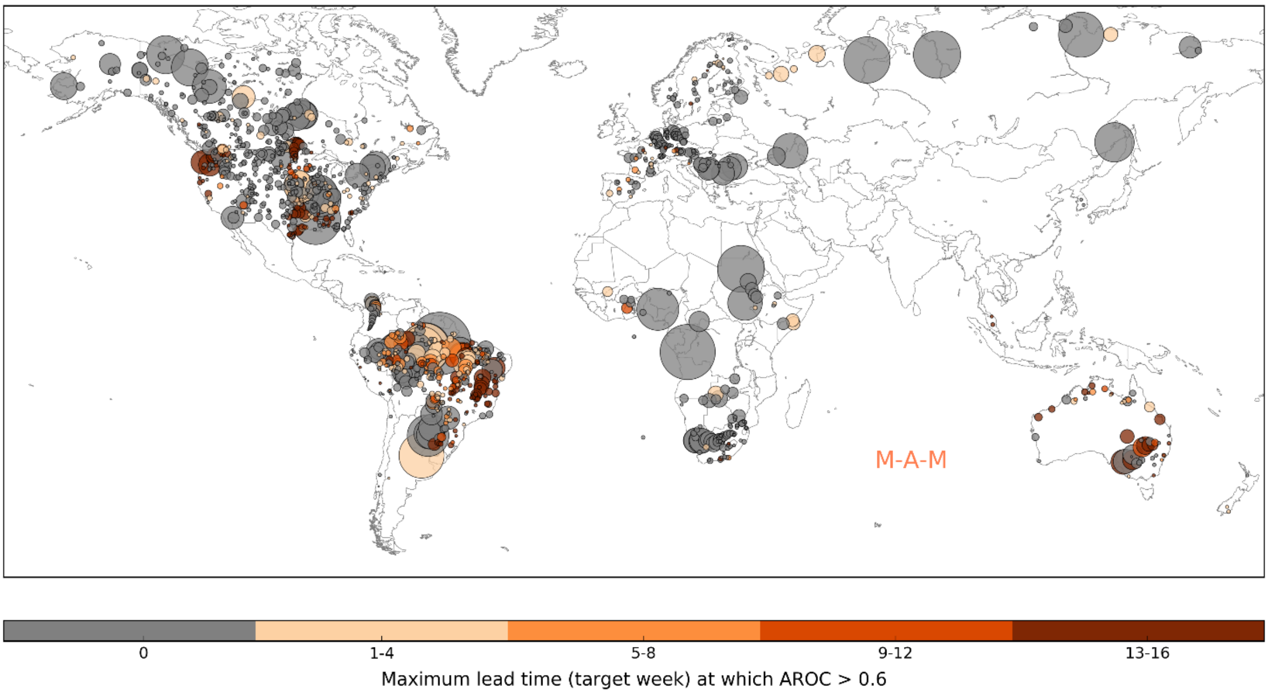


Figure S1: Maximum forecast lead time at which the area under the ROC curve (AROC) is greater than 0.6 for high flow events (flow exceeding the 80th percentile of climatology), at each observation station, for forecasts started in each season. This is used to indicate the maximum lead time at which forecasts are skilful. Grey dots indicate that forecasts started in that season have an AROC < 0.6 at all lead times.


D-J-F
0
1-4
5-8
9-12
13-16
Maximum lead time (target week) at which AROC > 0.6


M-A-M
0
1-4
5-8
9-12
13-16
Maximum lead time (target week) at which AROC > 0.6

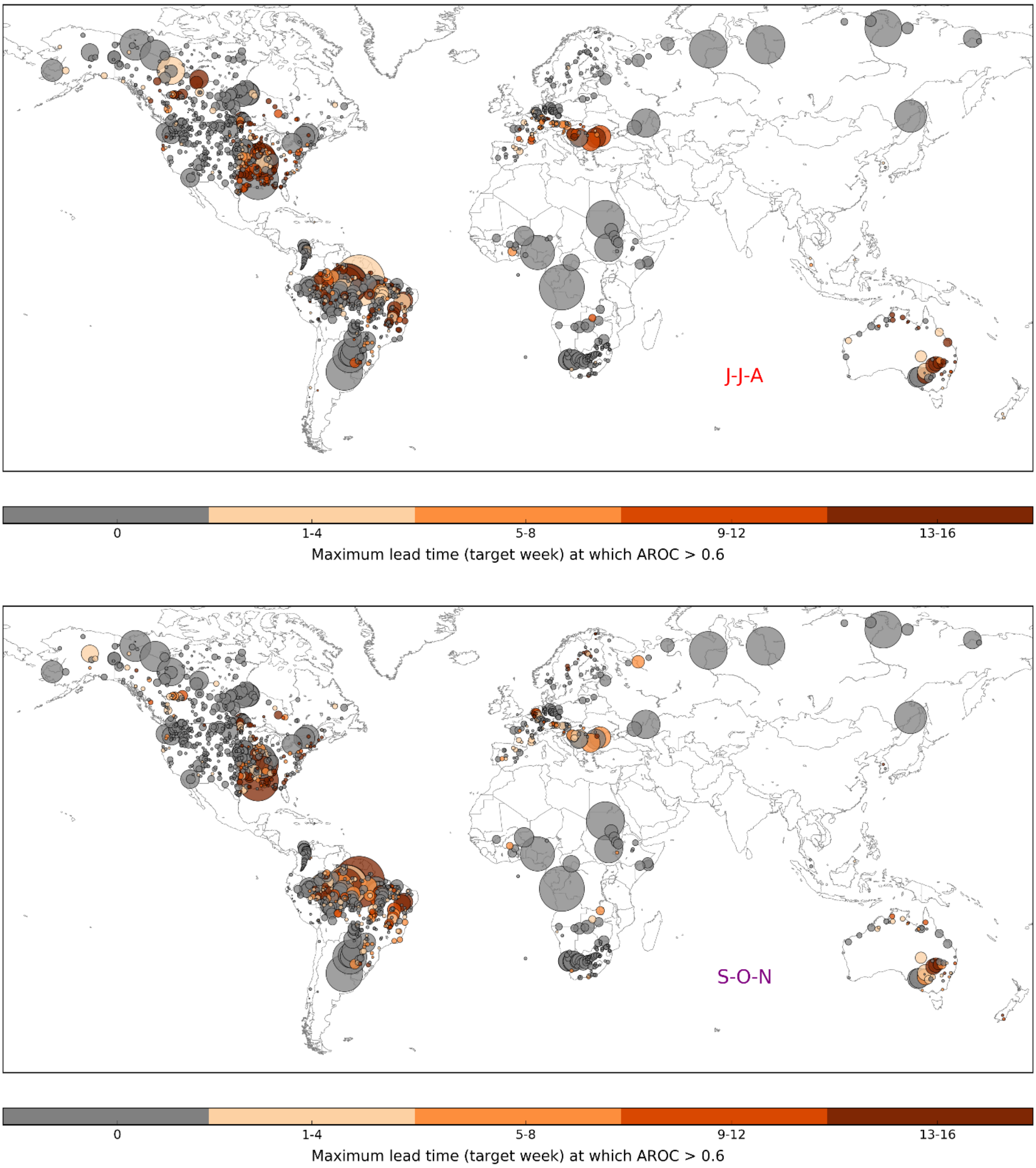


Figure S2: Maximum forecast lead time at which the area under the ROC curve (AROC) is greater than 0.6 for low flow events (flow below the 20th percentile of climatology), at each observation station, for forecasts started in each season. This is used to indicate the maximum lead time at which forecasts are skilful. Grey dots indicate that forecasts started in that season have an AROC < 0.6 at all lead times.

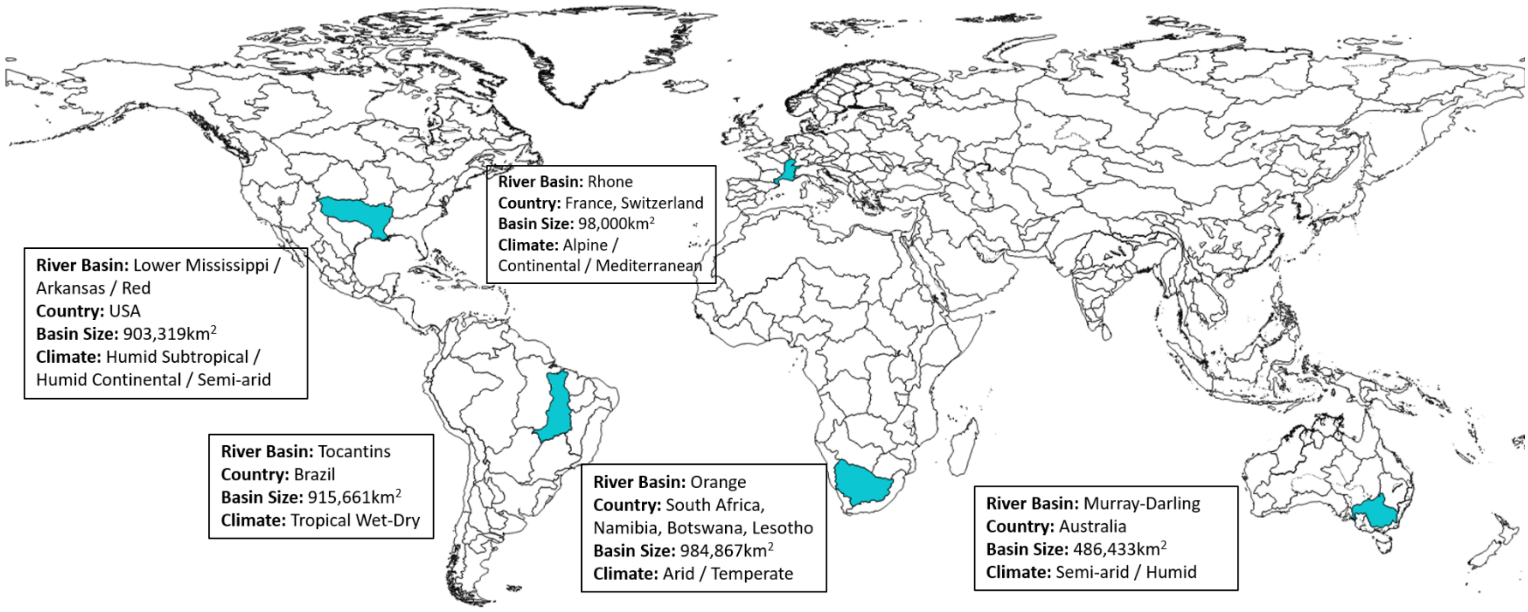


Figure S3: Map of the GloFAS-Seasonal major river basins, highlighting the river basins used for the forecast reliability evaluation.

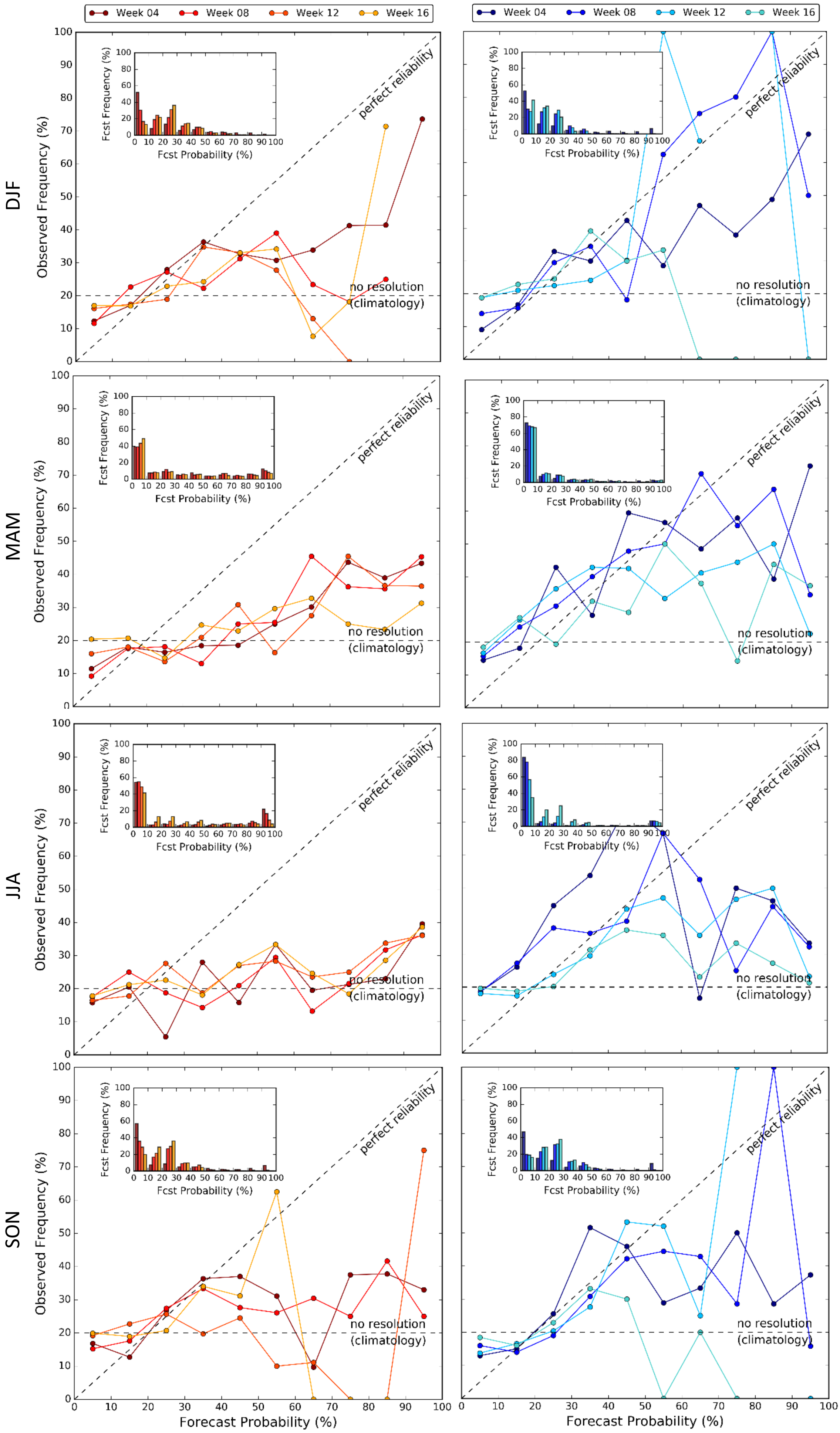


Figure S4: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Tocantins river basin (40 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

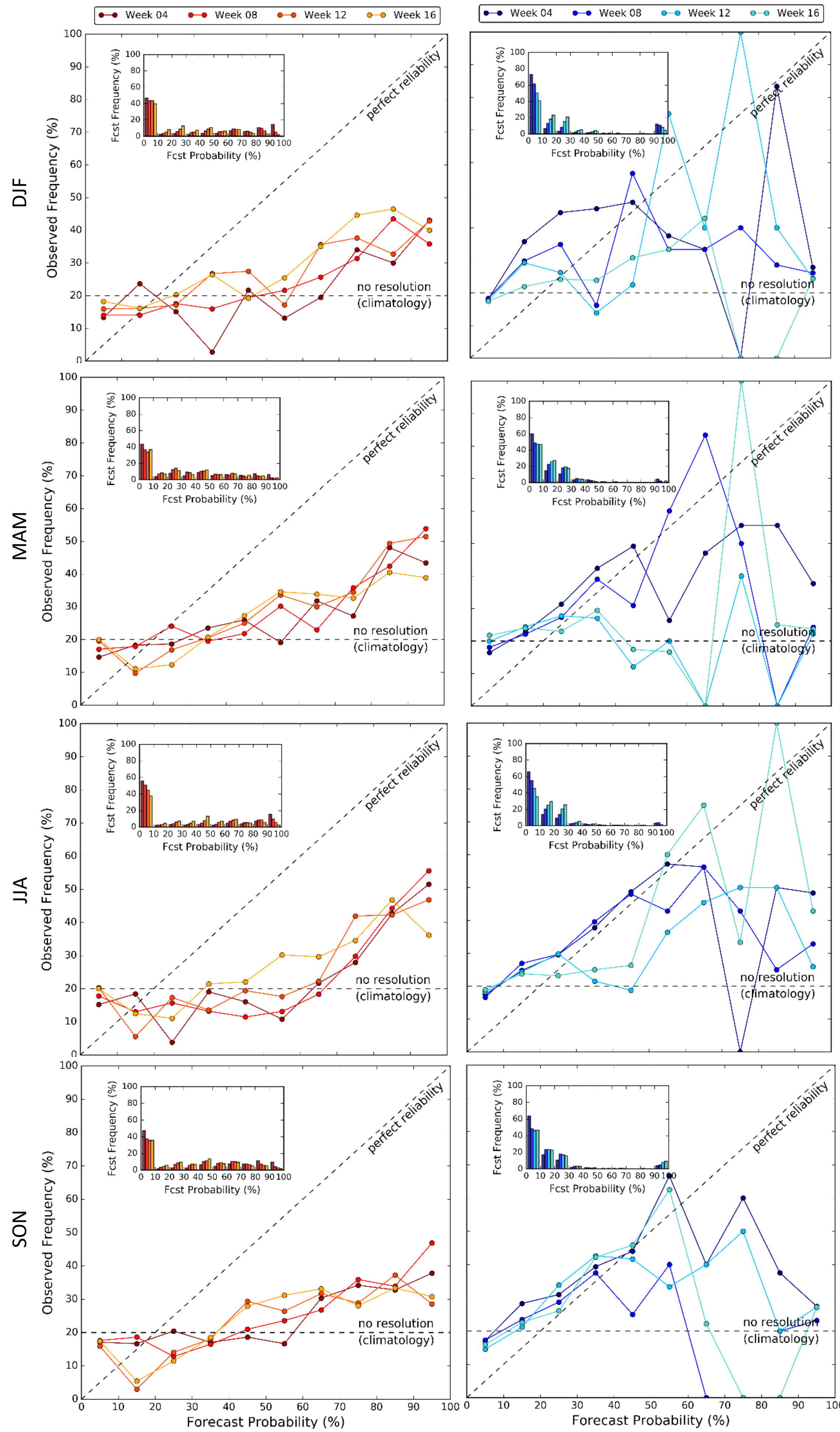


Figure S5: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Lower Mississippi river basin (35 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

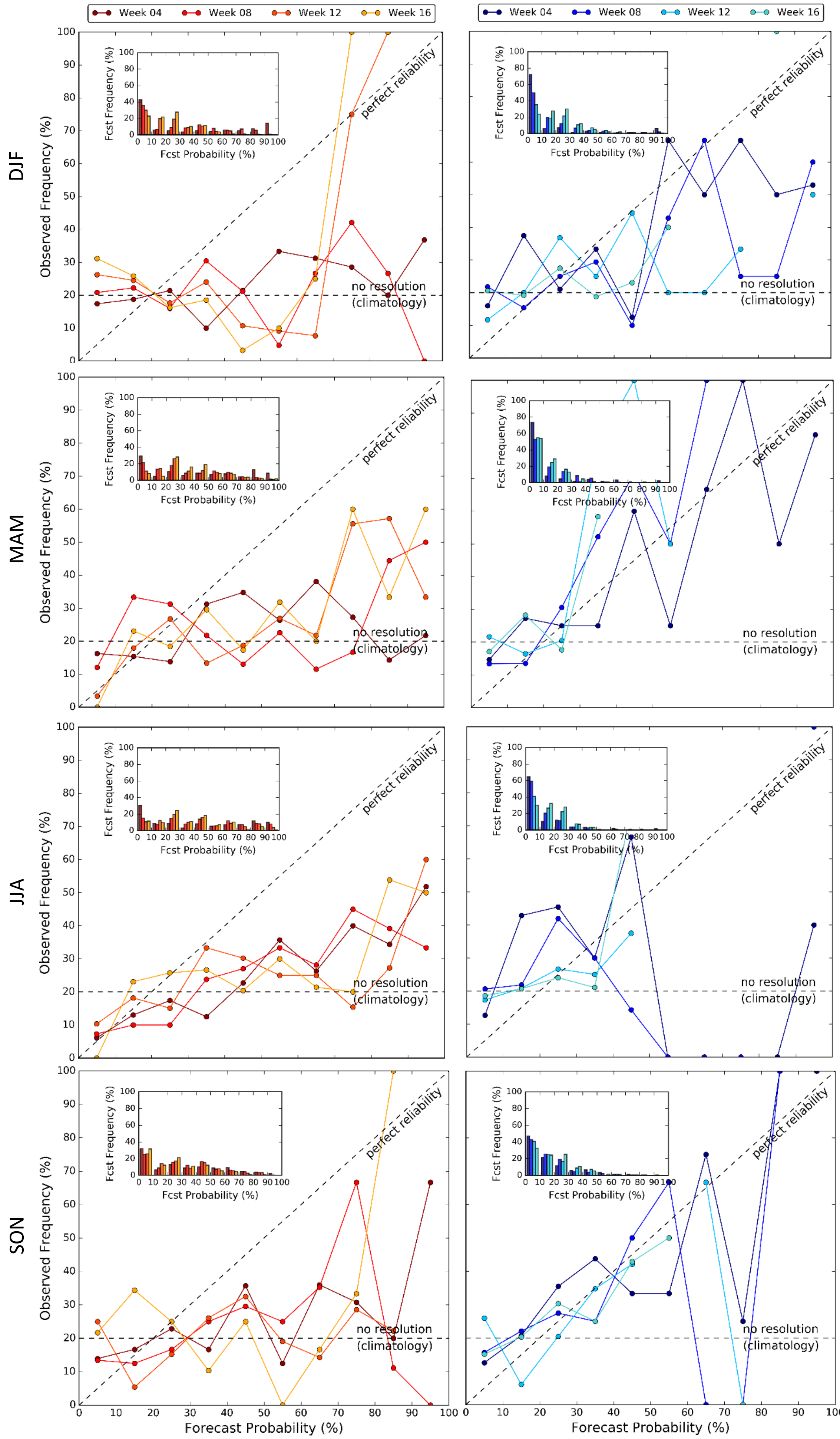


Figure S6: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Rhone river basin (8 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

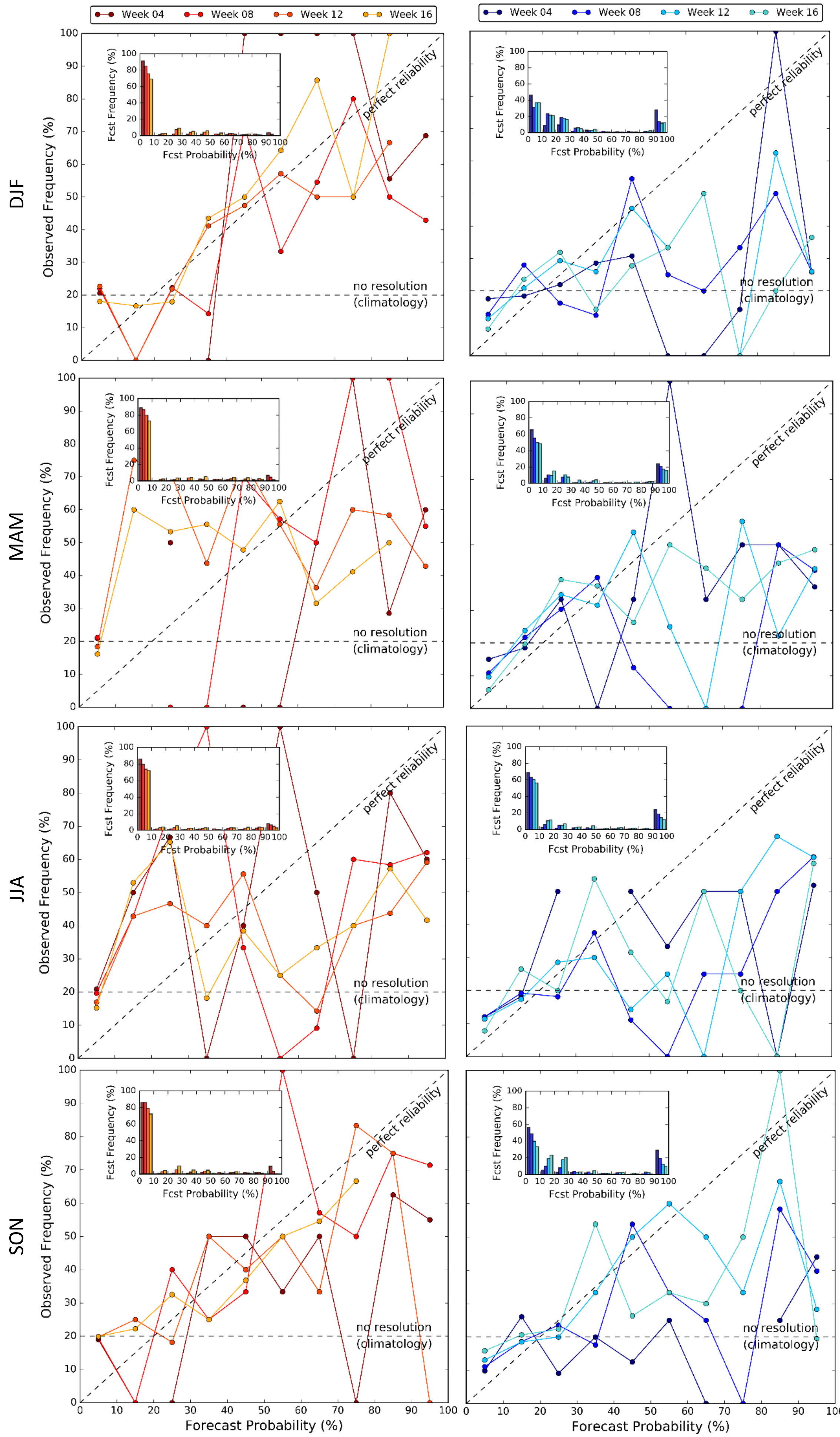


Figure S7: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Murray river basin (12 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

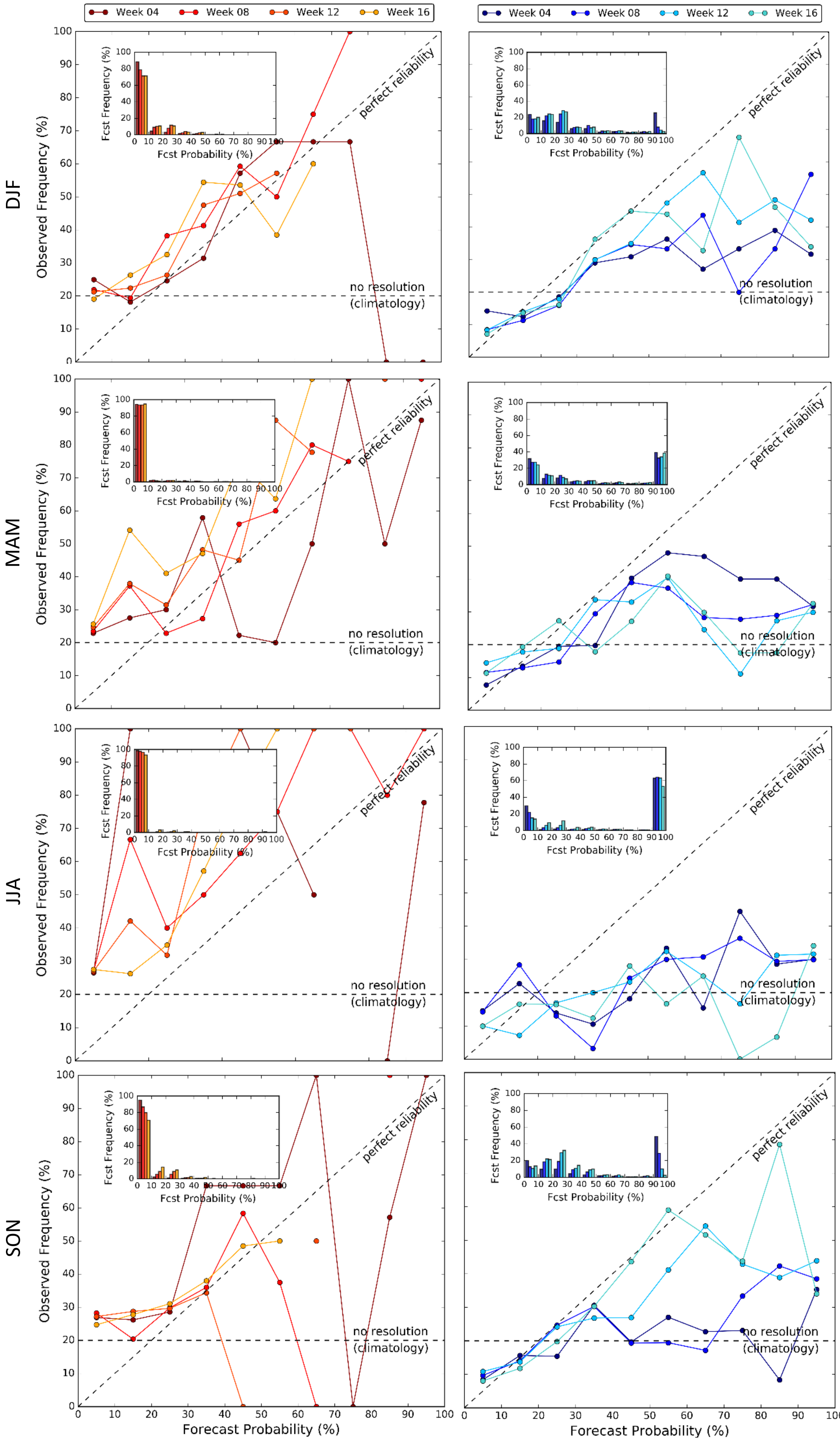


Figure S8: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Orange river basin (46 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.