

**Rainfall and River Flow Ensemble Verification: Phase 2**  
**G2G rainfall source comparison**

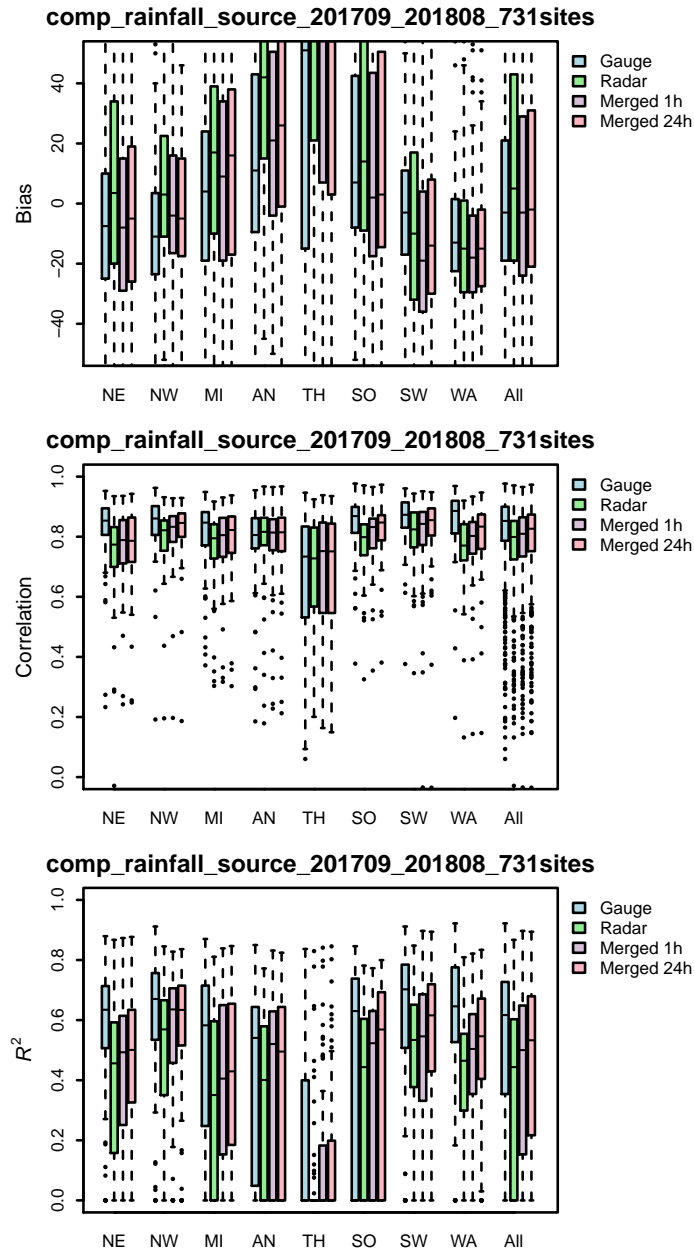
**Verification period:** 1 September 2017 to 31 August 2018 (Phase 2 Year 2 period)

**Observed river flows:** release to UKCEH ending 30 September 2018

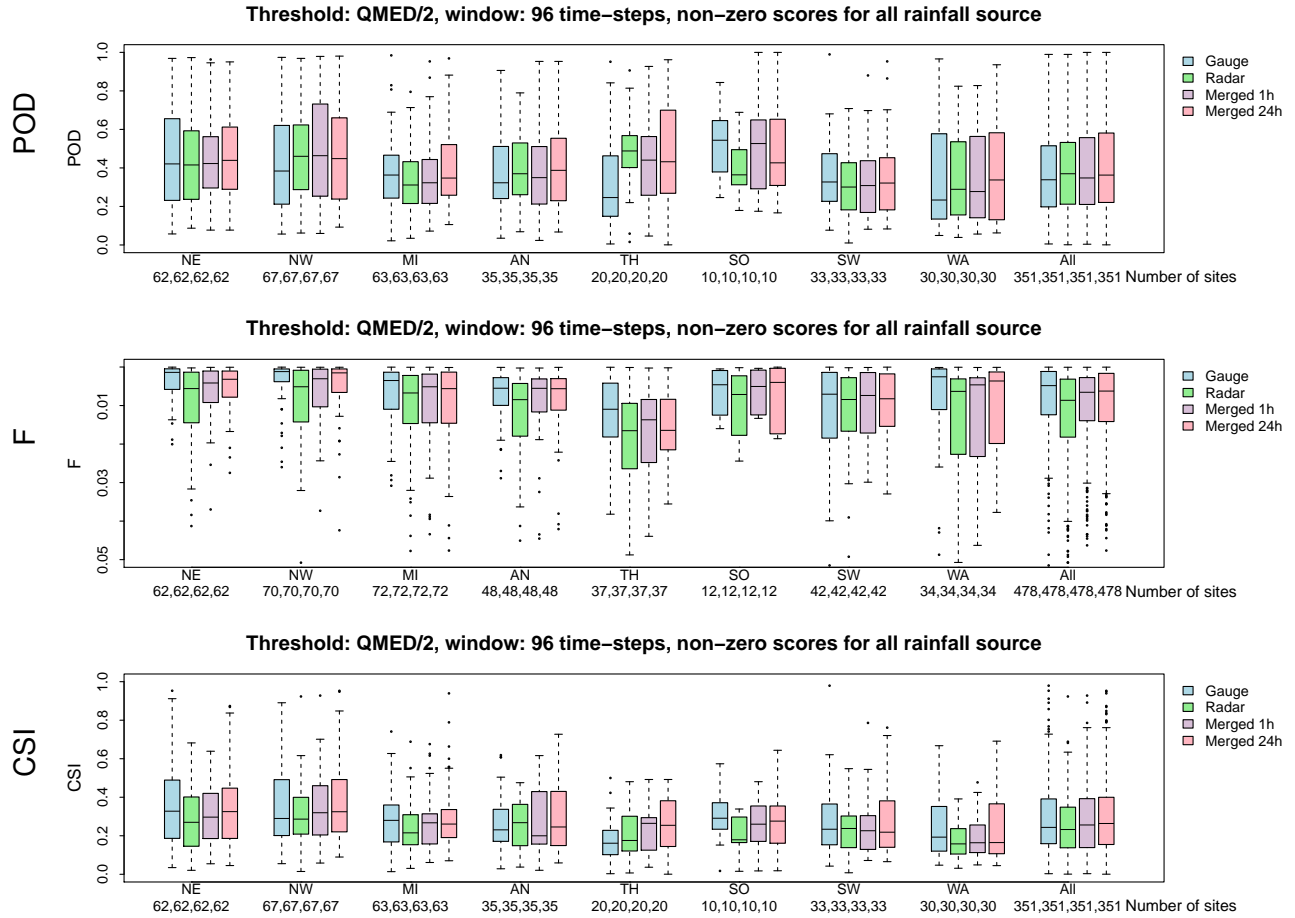
**Number of catchments:** 731

**Domain:** England & Wales

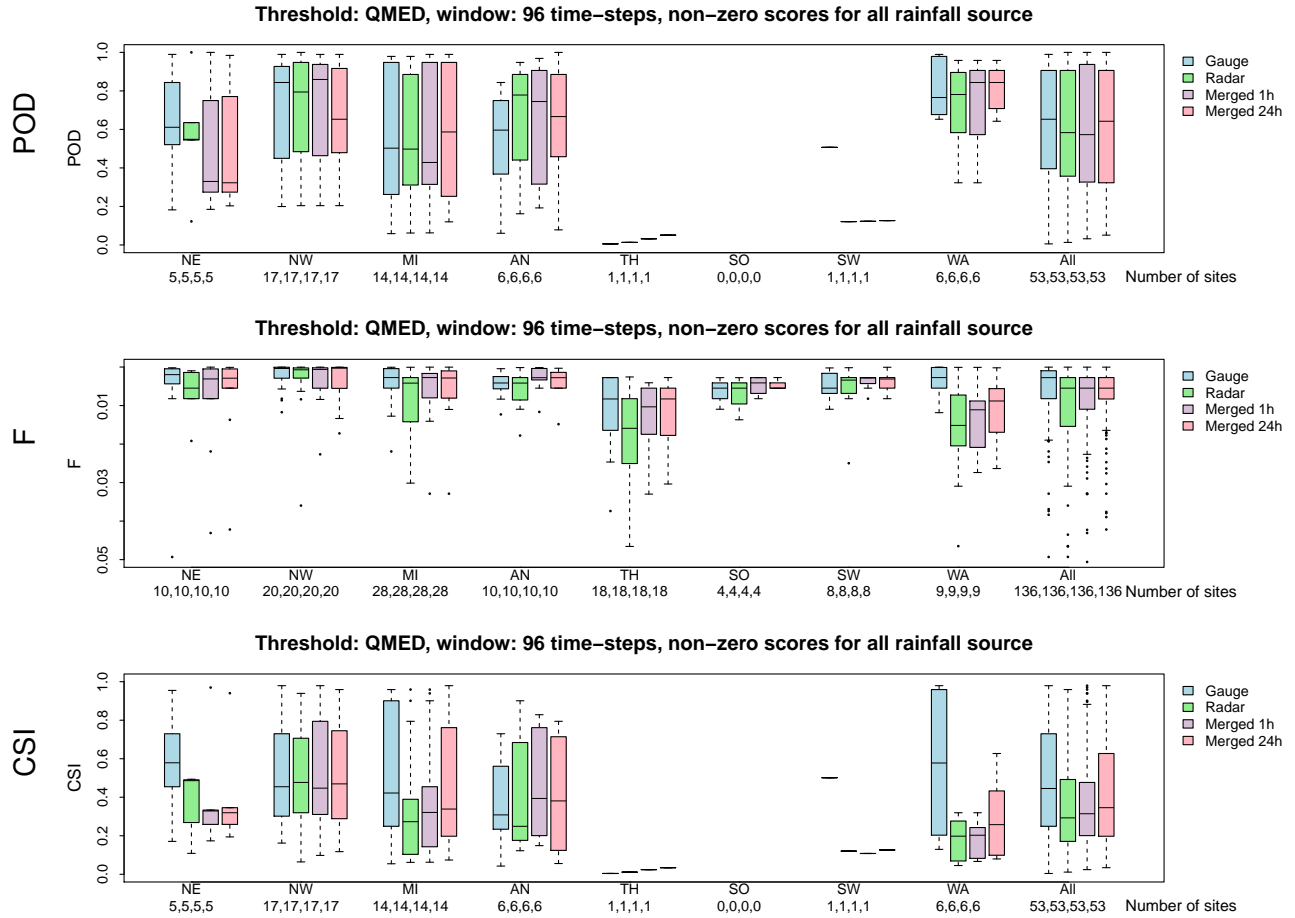
**Version 1.1:** 14 January 2021



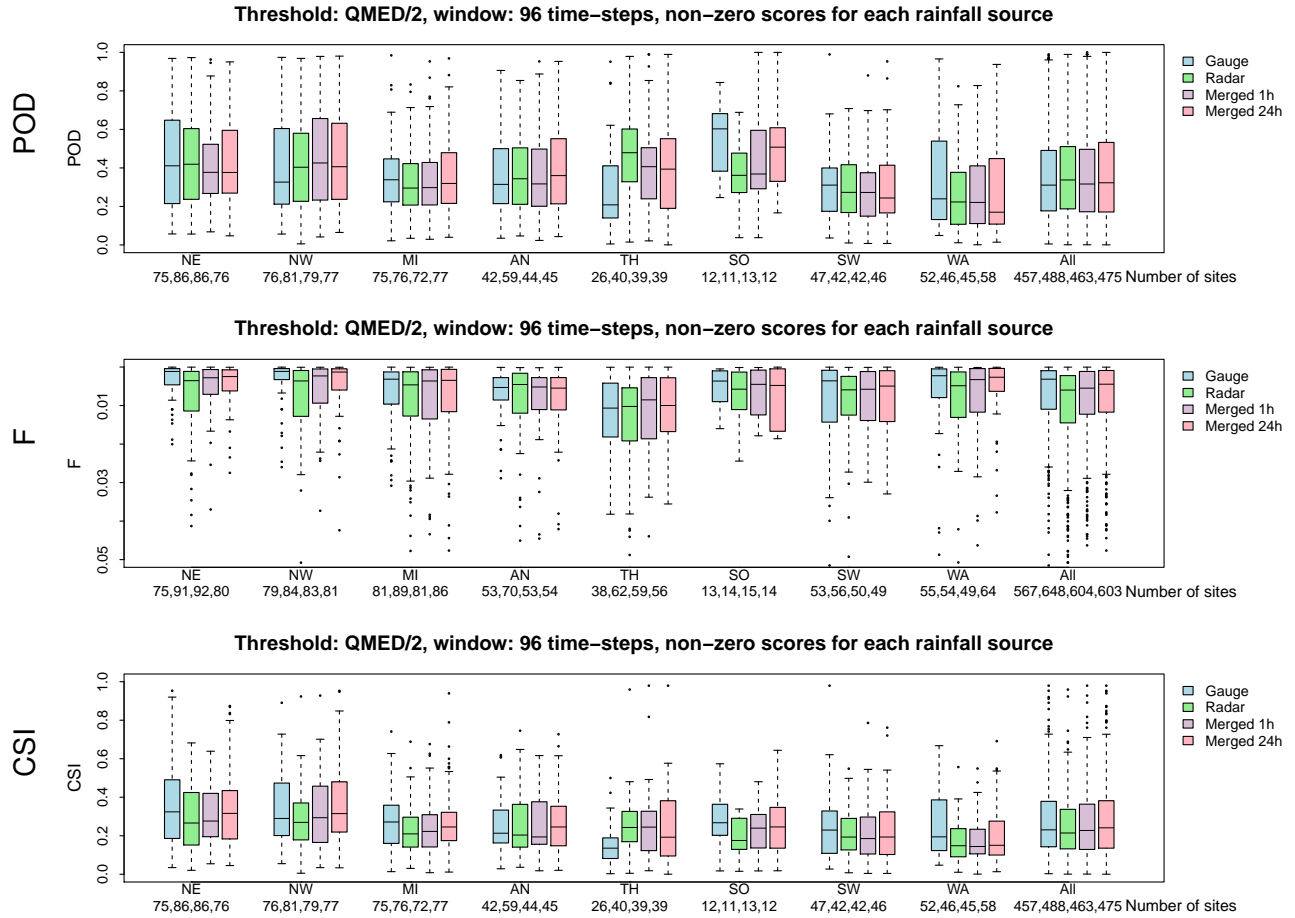
**Figure 1** Box Plots comparing the performance of G2G river flow simulations using different observed precipitation sources as input. The Bias (top), Correlation (middle), and  $R^2$  Efficiency (bottom) goodness-of-fit statistics are shown. Bars are for each grouping of catchments considered: each region in England, for Wales, and for all catchments in England & Wales. Each set of bars contains (from left to right) results for G2G simulations using Gauge, Radar, Merged 1h and Merged 24h observed precipitation data as input. Each bar shows the median (solid line) and interquartile range (coloured box) of the distribution of statistics over the set of catchments. Dashed lines extend to 1.5 times the interquartile range from the box, and indicate the typical range of the data. Outlying points are shown by black dots.



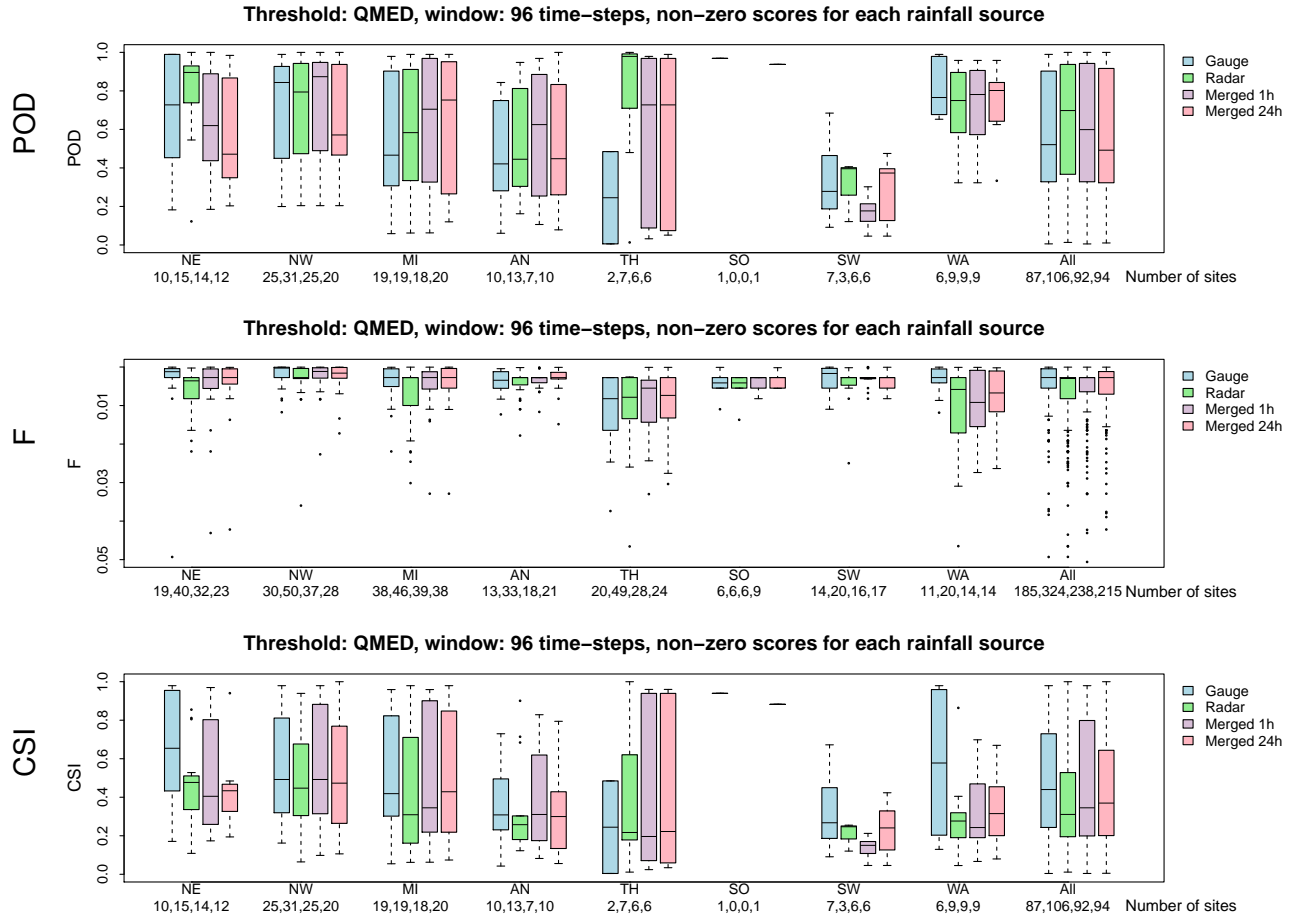
**Figure 2** Box Plots comparing the performance scores (POD, F, CSI) of G2G river flow simulations - using different observed precipitation sources as input - for the Q(2)/2 threshold and 24h moving window. Bars are for each grouping of catchments considered: each region in England, for Wales, and for all catchments in England & Wales. Each set of bars contains (from left to right) results for G2G simulations using Gauge, Radar, Merged 1h and Merged 24h observed precipitation data as input. Each bar shows the median (solid line) and interquartile range (coloured box) of the distribution of scores over the group of catchments. Only catchments with non-zero scores for all precipitation sources (Method 1) are included (the number of catchments is indicated beneath the bars). Dashed lines extend to 1.5 times the interquartile range from the box, and indicate the typical range of the data. Outlying points are shown by black dots.



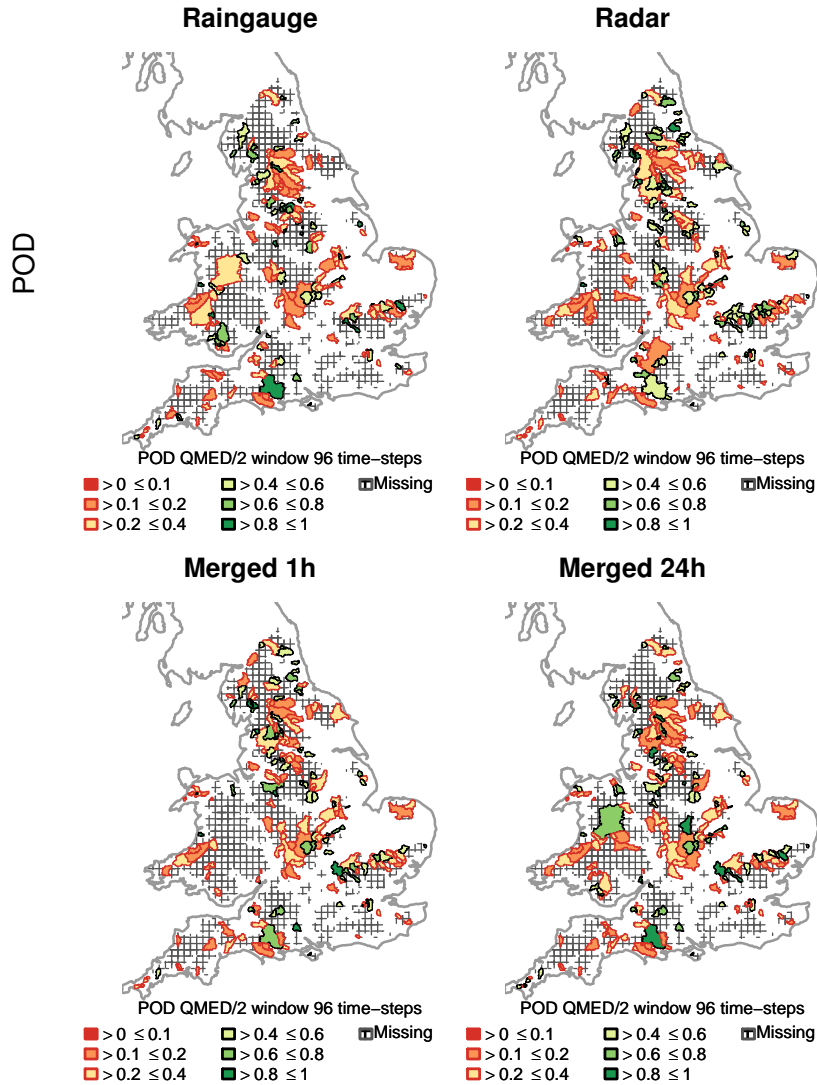
**Figure 3** Box Plots comparing the performance scores (POD, F, CSI) of G2G river flow simulations - using different observed precipitation sources as input - for the Q(2) threshold and a 24h moving window. Bars are for each grouping of catchments considered: each region in England, for Wales, and for all catchments in England & Wales. Each set of bars contains (from left to right) results for G2G flow simulations using Gauge, Radar, Merged 1h and Merged 24h observed precipitation data as input. Each bar shows the median (solid line) and interquartile range (coloured box) of the distribution of scores over the group of catchments. Only catchments with non-zero scores for all precipitation sources (Method 1) are included (the number of catchments is indicated beneath the bars). Dashed lines extend to 1.5 times the interquartile range from the box, and indicate the typical range of the data. Outlying points are shown by black dots.



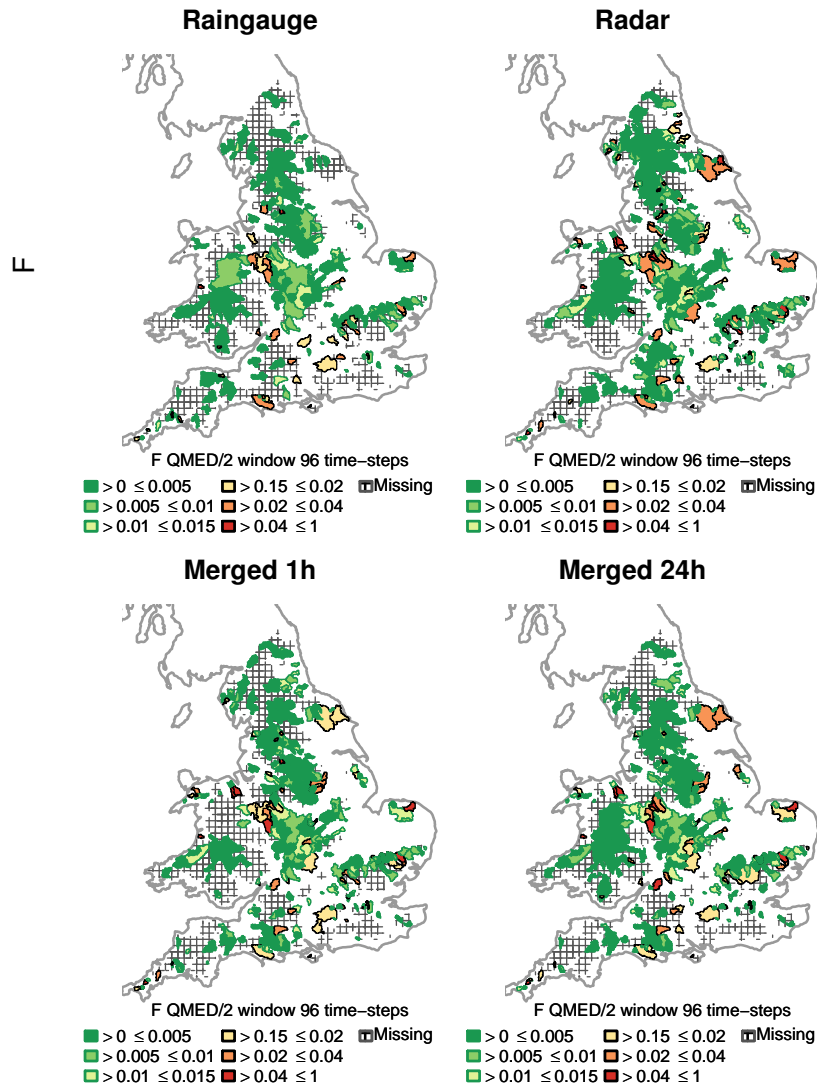
**Figure 4** Box Plots comparing the performance scores (POD, F, CSI) of G2G river flow simulations - using different observed precipitation sources as input - for the Q(2)/2 threshold and a 24h moving window. Bars are for each grouping of catchments considered: each region in England, for Wales, and for all catchments in England & Wales. Each set of bars contains (from left to right) results for G2G simulations using Gauge, Radar, Merged 1h and Merged 24h observed precipitation data as input. Each bar shows the median (solid line) and interquartile range (coloured box) of the distribution of scores over the set of catchments. Catchments with non-zero scores for each precipitation source (Method 2) are included (the number of catchments included is indicated beneath the bars). Dashed lines extend to 1.5 times the interquartile range from the box, and indicate the typical range of the data. Outlying points are shown by black dots.



**Figure 5** Box Plots comparing the performance of G2G river flow simulations - using different observed precipitation sources as input - for the Q(2) threshold and a 24h moving window. Bars are for each grouping of catchments considered: each region in England, for Wales, and for all catchments in England & Wales. Each set of bars contains (from left to right) results for G2G simulations using Gauge, Radar, Merged 1h and Merged 24h observed precipitation data as input. Each bar shows the median (solid line) and interquartile range (coloured box) of the distribution of scores over a grouping of catchments. Catchments with non-zero scores for each precipitation source are included - the number of catchments included is indicated beneath the bars. Dashed lines extend to 1.5 times the interquartile range from the box, and indicate the typical range of the data. Outlying points are shown by black dots.

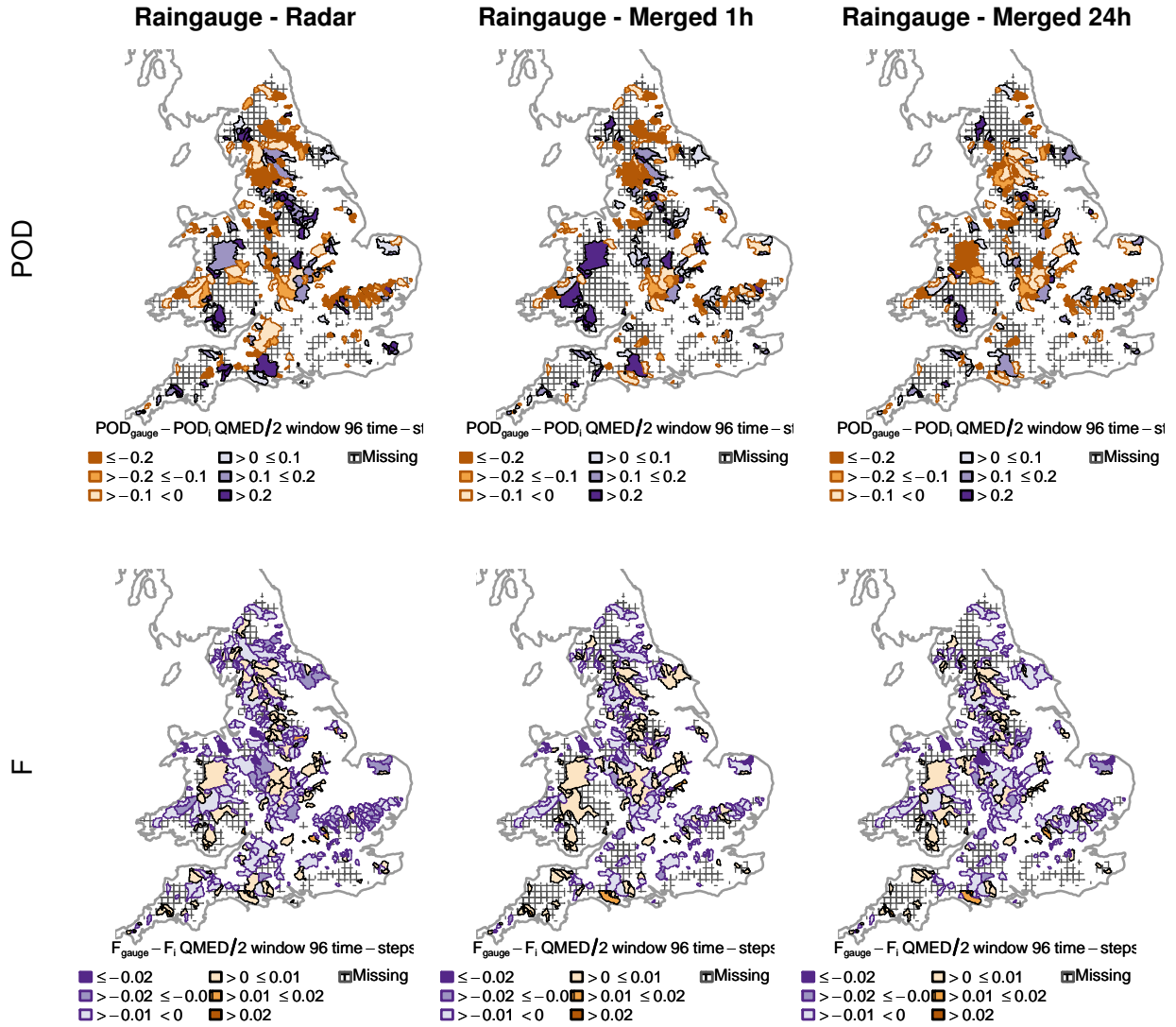


**Figure 6 Maps of POD scores calculated for the Q(2)/2 threshold and a 24h moving window. POD scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.**

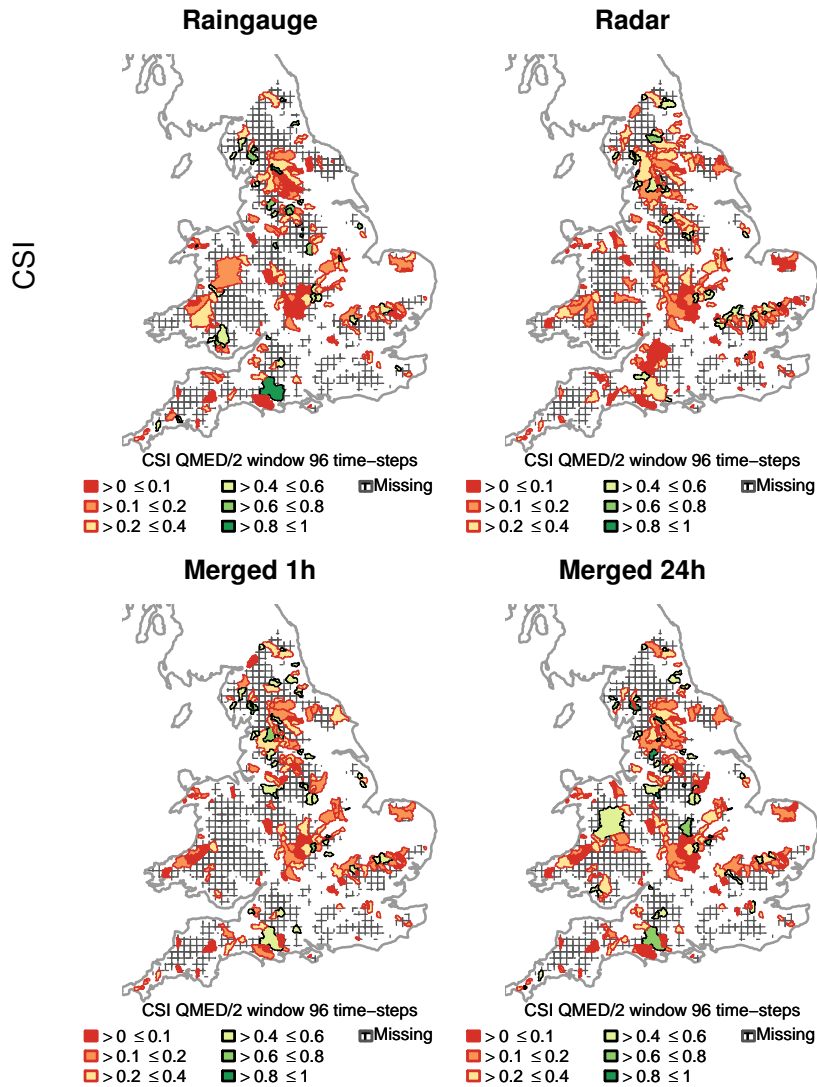


**Figure 7 Maps of F scores calculated for the Q(2)/2 threshold and a 24h moving window. F scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.**

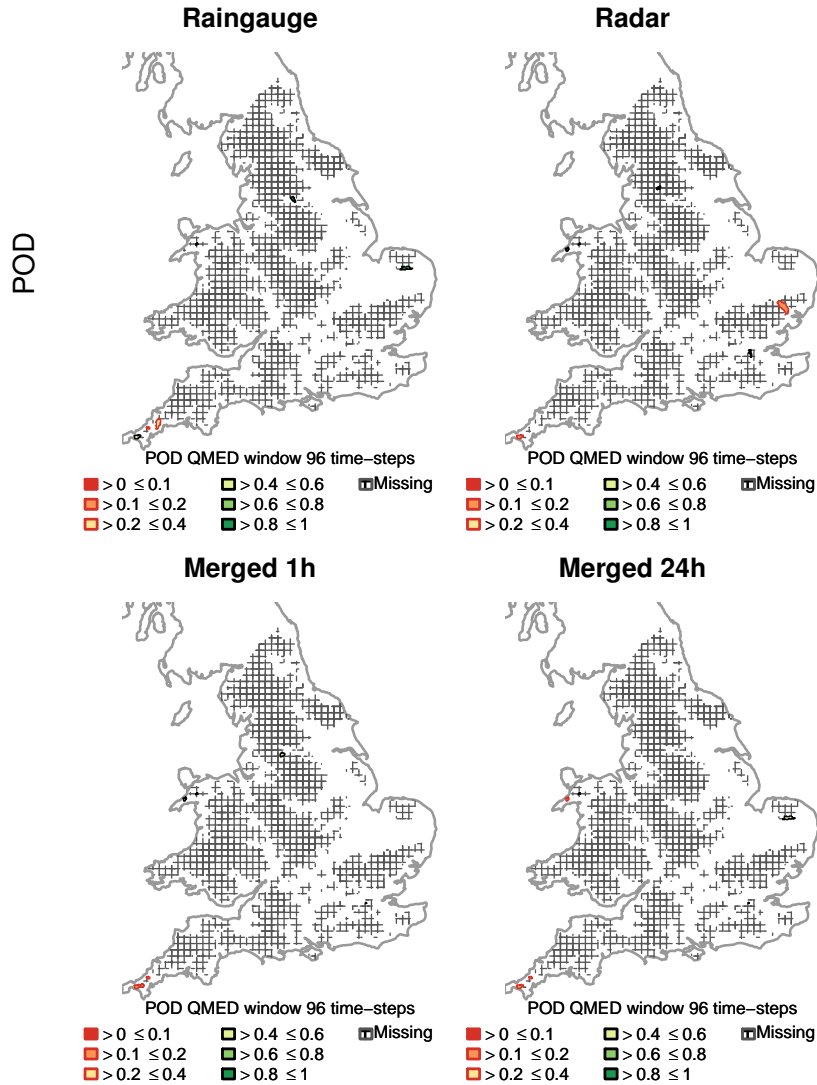




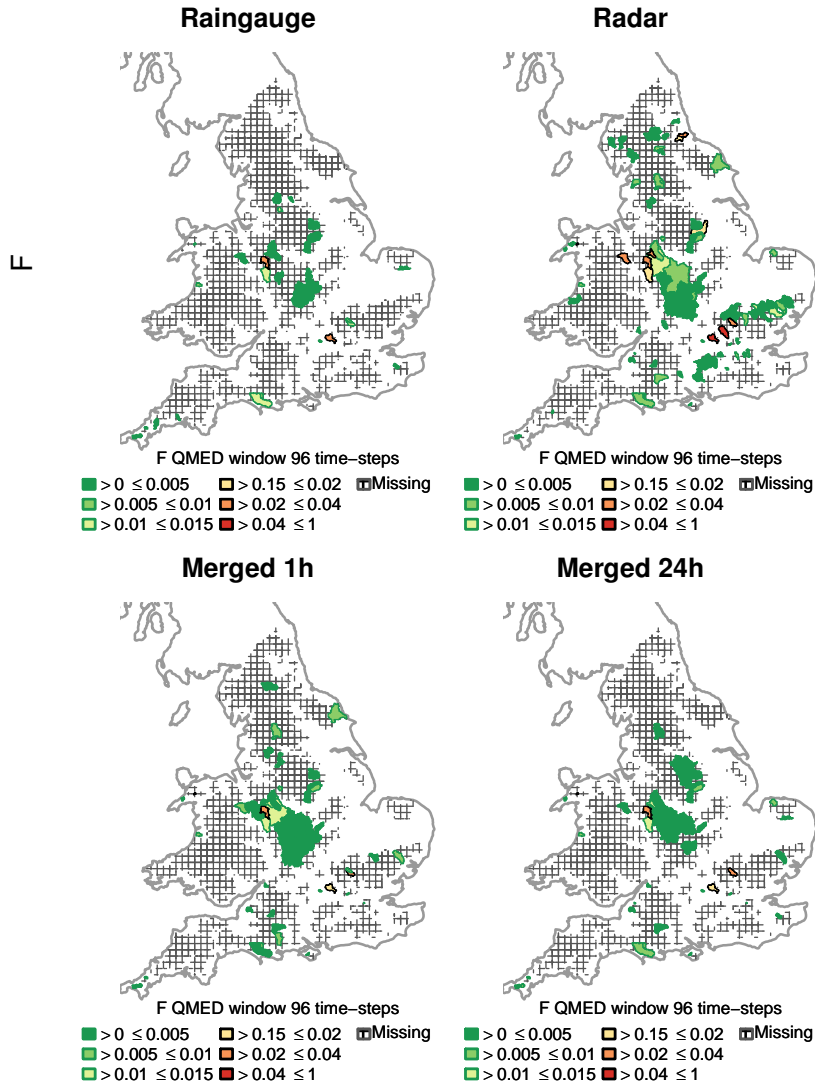
**Figure 8** Maps of the difference in POD and F scores calculated for the Q(2)/2 threshold and a 24h moving window. Gauge-Radar (left), Gauge-Merged 1h (middle) and Gauge-Merged 24h (right). Purple colours show Gauge performing better, orange colours show Gauge performing worse.



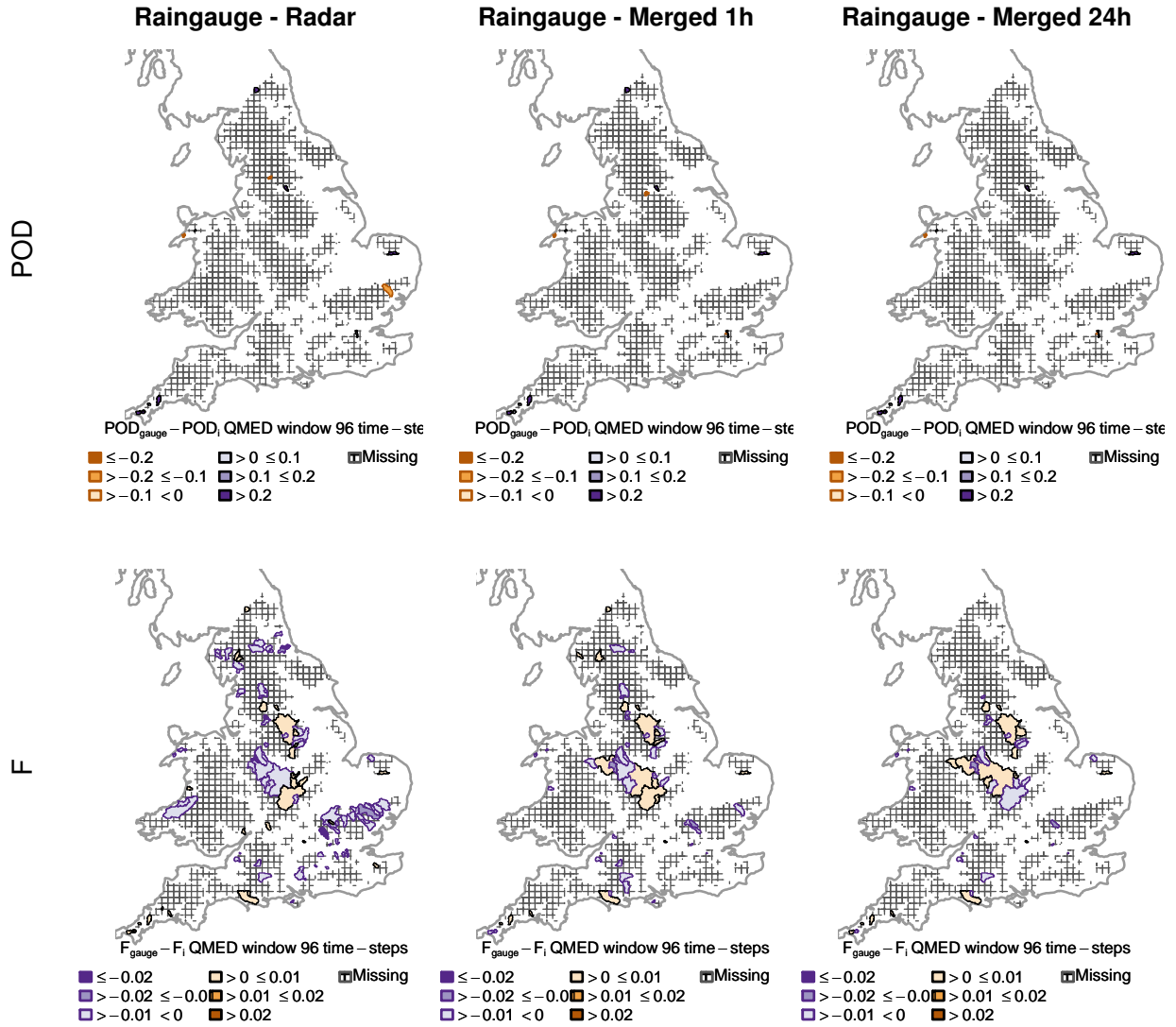
**Figure 9** Maps of CSI scores calculated for the Q(2)/2 threshold and a 24h moving window. CSI scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.



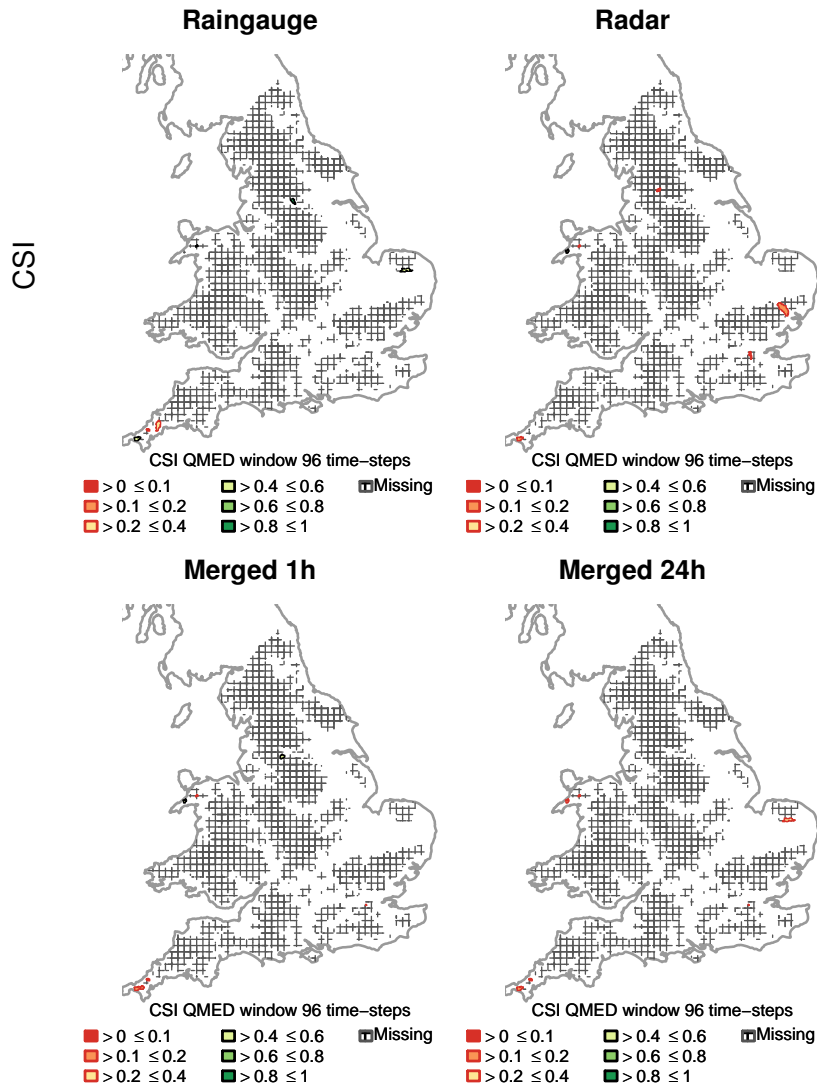
**Figure 10** Maps of POD scores calculated for the Q(2) threshold and a 24h moving window. POD scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.



**Figure 11** Maps of F scores calculated for the Q(2) threshold and a 24h moving window. F scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.



**Figure 12** Maps of the difference in POD and F scores calculated for the Q(2) threshold and a 24h moving window. Gauge-Radar (left), Gauge-Merged 1h (middle) and Gauge-Merged 24h (right). Purple colours show Gauge performing better, orange colours show Gauge performing worse.



**Figure 13** Maps of CSI scores calculated for the Q(2) threshold and a 24h moving window. CSI scores are shown from red with no outline (poor) to green with outline (good) for the G2G river flow simulations using Gauge, Radar, Merged 1h and Merged 24h precipitation data as input.