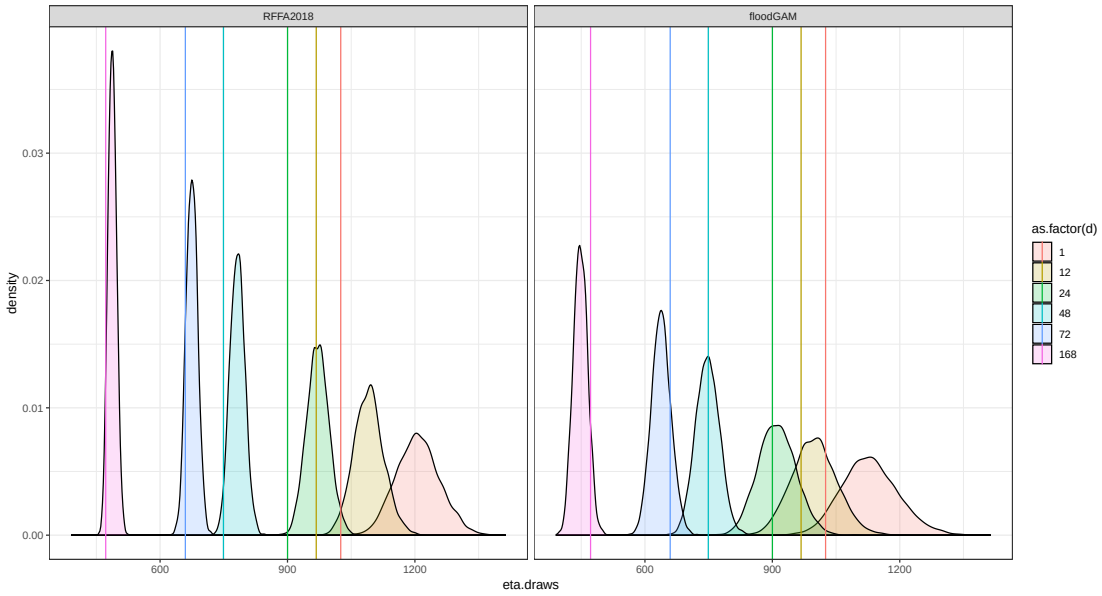
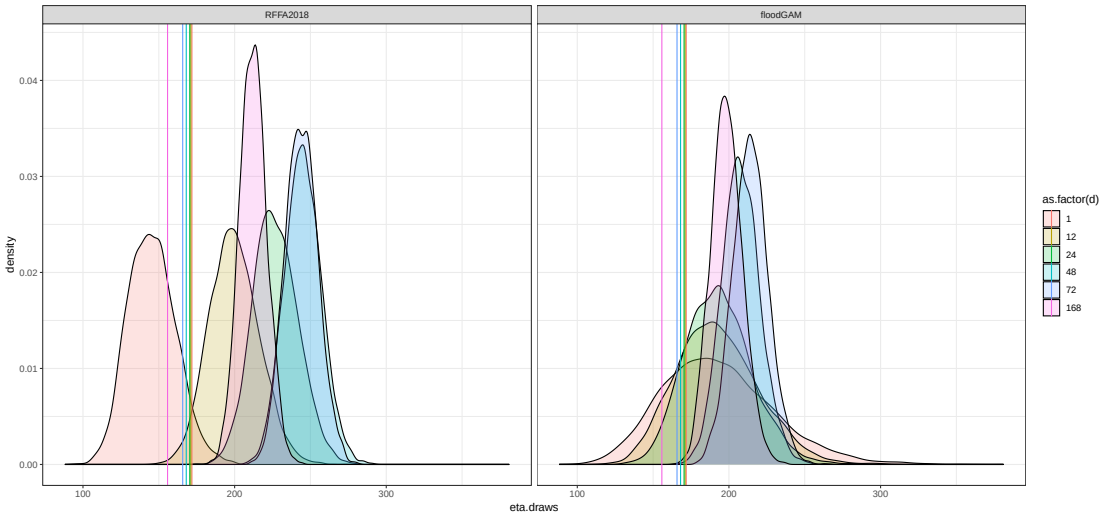
A fully duration-consistent station looks like this:



(the vertical lines are the observed median flood, the distributions are 5000 draws from the posterior. out of sample predictions).

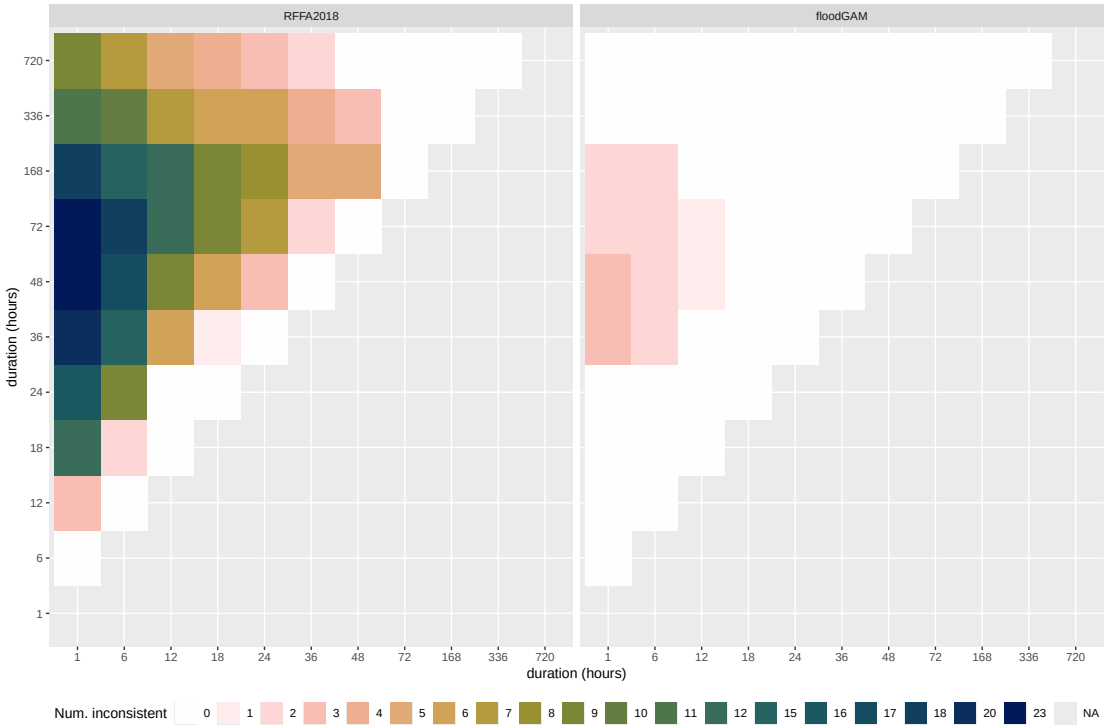
A duration-inconsistent station can look like this:



This is a difficult station for the models (note how close together the empirical medians are– not much difference between durations). Here, although the predictions are not ideal, floodGAM is not statistically significantly inconsistent. The predicted median flood distributions overlap. However, RFFA2018 is "statistically significantly duration inconsistent" at some durations. For example, the distribution for the 48 median flood is higher than the distribution for the 1 hour

median flood (inconsistent) and there is no overlap at the 5% level between the distributions (statistically significantly inconsistent). Note also how the distributions march upwards into inconsistency but then march back down (i.e. the 168 hour distribution is consistent with the 72 hour distribution, even though the 72 hour distribution is not consistent with the 12 or 1 hour distributions).

We check each duration for statistically significant inconsistencies with every other duration. This plot shows the number of statistically significant inconsistencies:



Two things to note from this plot:

- floodGAM has in total 30 statistically significant inconsistencies. RFFA2018 has in total 507.
- consecutive durations are never statistically significant inconsistent (the 6 hour duration never ssi with the 1 hour or 12 durations, etc).