Line 128 In cases of simultaneous acquisitions from multiple resources. Was the information coincident? For example, did information from traditional measurement stations coincide with google images? Was the 'flow condition information the same?

By grouping the data by reach and by date, we observe that there is simultaneity in only 0.26% of cases on average for the 3 catchments (Albarine: 83 cases of simulatneity over 28852 total cases, Genal: 16 cases over 7146, Lepsamanjoki: 12 cases over 6307).
The small amount of data observed on the same day on the same stretch of river can be explained by the complementary nature of the different sources, which each focus on different areas and periods.
Of the 111 cases of simultaneity, the different sources give the same state of flow in 88% of cases. Figure A shows the number of cases for which the sources agree or disagree according to the number of sources.

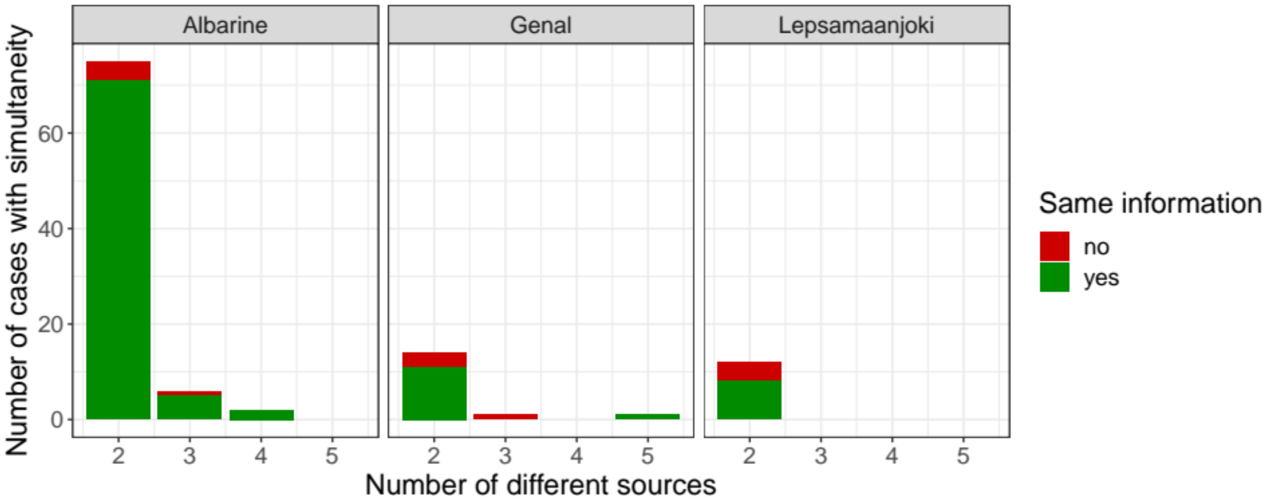

Figure A: Number of cases (reach and dates) with simultaneous observations

Disagreements between data sources may be due to :
- local phenomena: it is possible to observe different states of flow along the same reach;
- the subjectivity of observers during data acquisition (particularly in cases where there is stagnant water or very low flow) ;
- errors in assigning the flow class.

Line 200 I suggest you use accuracy, sensitivity, specifcity, and true skill statistic to validated RF

As you suggested we computed accuracy, sensitivity, specifcity, and true skill statistic (TSS) to validate RF. Sensitivity is équivalent to POD (probability of prediction of a drying event) which was originaly used in the manuscript. Along with Sensitivity, Specificity, Accuracy and TSS, we kept the False Alarm ratio (FAR) which indicates the probability of predicting a false drying event. Figure B shows that Specificity is above 0.99 for the 3 catchments, which means that the RF model predicts flowing events almost perfectly. Accuracy is also very close to 1 (> 0.98), this is due to the fact that flowing events are much more represented than drying events in the observed dataset, so prediction errors for dry events are negligible compared with the near-perfect predictions of flowing events. TSS (which is computed as Sensitivity + Specificity -1) is very similar to Sensitivity as Specificity is very close to 1.

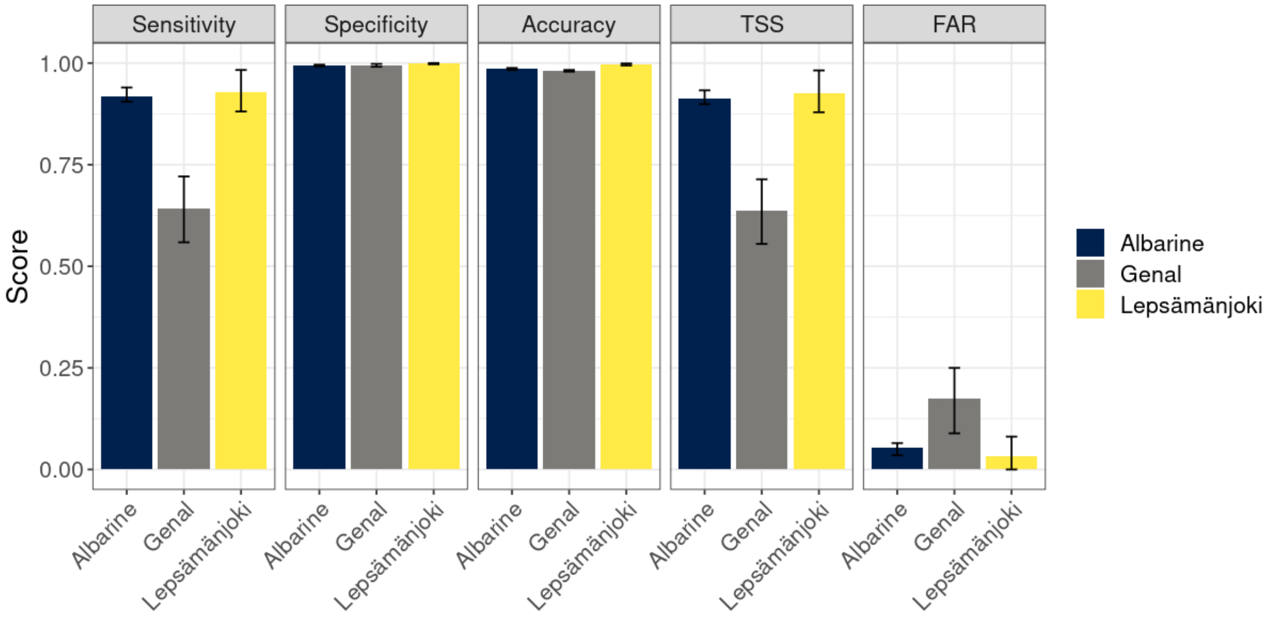

Figure B :Performance of the RF model when the model is trained with 75% of observed data and tested on the remaining 25% (configuration 1). TSS: True Skill Statistic, FAR: False Alarm Ratio (this figure will replace Fig.7 in the manuscript)

These new elements will be integrated to the revised version of the manuscript.