

Line-by-line comments

Wang et al. (2025) - Technical Note: Can Visual Gauges Trained on Biased Contact-based Gauge Data Accurately Estimate River Stage?

Major issues:

L. 152: If I am not totally mistaken, it's quite the opposite. True negative is the proportion of samples that are correctly judged as negative! The authors need to make sure this was only confused in the text and not also in their analysis.

L. 173: Why is there a months-long pause in 2020? I think the authors should address this.

L. 249 (Fig.8): what is "early stage" and "late stage"? This is not defined anywhere and does not seem obvious. I do not understand where the late stage in dataset 2019 and 2019-2020 come from if there was no gauge failure.

L. 259: in my opinion one cannot rely on such an extrapolation below gauge zero without any gauging data supporting this. This might work well in one location and completely fail in another one, solely based on river bed geometry. Just because the model provides values below zero does not mean one can trust them.

Minor issues:

L. 2: Title: I think the title could be chosen a bit more carefully to accurately reflect the content of this work: only 1 visual gauge is used and it's a case study with only one event. I propose something like: "Estimating river stage with a visual gauge trained on biased gauge data - a case study"

L. 50: "contact-based" does not seem like the right adjective for markers. They aren't contact-based sensors even though they are technically in contact with the water. I'd suggest simply going with "..reliance on markers placed within river..."

L. 88 ff: adding the word "sequence" after each data type seems unnecessary and confuses the reader (also applies to Fig.1).

L. 94: Since the term "gauge bias" is a core idea of this publication, it should be defined more precisely. If understood correctly the authors use this term to address singular events that significantly offset the stage measurements.

L. 108: the authors should add the abbreviation MTL in brackets right after mentioning Multi-task learning for the first time.

L. 213: I am missing a more detailed description of how the water segmentation was performed. This seems to be a core part of this work.

L. 220 ff: I suggest rephrasing this sentence and splitting it into several sentences. Shorter and more understandable would increase the readability significantly.

L. 234: To many people it is highly unclear whether 0:00 am means midnight or noon. Instead I'd suggest using 24h time format or at least noon/midnight.

L. 249 (Fig.8): the intuitive interpretation of the three graphs suggests that all three datasets are similar. Only when looking at the y-scale of the third plot one may notice that they are of different scale. The authors should consider using the same scale for all three plots.

Grammar, trivialities:

L. 11: I suggest using the plural "river cameras" here, as we are talking about a general type of instrumentation rather than one specific camera.

L. 286: "that that"

L. 302: here the authors could consider using "data" as a singular here to improve the readability of the sentence: "... even when training data contains errors..."