Response to reviewers for egusphere-2024-2591: "Extended range forecasting of stream water temperature with deep learning models"

Referee #1:

Regarding Question 2: "Please provide common evaluation metrics such as correlation coefficient, NSE, and MAE," you only added MAE and stated that MAE is equivalent to CRPS. However, this does not explain why you disregarded the other metrics (correlation coefficient and NSE). If possible, please also include these metrics.

We do not consider it useful to additionally provide estimates of NSE/R2 in our study. These metrics are not designed to evaluate probabilistic forecasts nor to be assessed when dealing with an extended range forecast lead time. It is not clear to us what NSE/R2 values should we even compute. As opposed to the CRPS (and also MAE and MSE) where the error metric is defined only by each pair of observations and predictions (o_i and p_i), NSE/R2 additionally require the average of the observations ($\bar{o}$). If we were to compute NSE/R2 for each forecast and station separately, using the 32 points from the 32 days lead time to compute $\bar{o}$, it would be important to note that these results are not comparable to those of other studies using prediction models for one time-step ahead. On the other hand, we could extract only $o_i$and $p_i$from the first day lead time for all 90 forecasts and compute $\bar{o}$ from these 90 o_i values. In this case however, we would already expect high NSE/R2 values as we would be capturing the seasonal behavior of the water temperature timeseries.

Overall, we consider that the metrics already provided in the study are sufficient to convey our results. Adding NSE/R2 values might even create confusion for the reader.