Review Comments: Streamflow data availability in Europe: a detailed dataset of interpolated flow-duration curves

Persiano et al have used a statistical procedure named Total Negative Deviation Top-Kriging to estimate flow duration curves for a large sample of river basins in Europe.

I will provide a brief summary of my understanding of the paper. Perhaps the authors would be able to correct me if I am wrong.

2484 stream gauges for which empirical FDCs can be calculated are used as input (training) data, and the target is to produce predicted FDCs for 32,960 nodes across the European river network (ungauged sites). No other information about catchment geology or other catchment characteristics is used as a covariate/independent variable, just the FDCs for the 2484 training gauges.

The top-kriging approach is applied and results are presented for all 32,960 prediction nodes (i.e. in Fig 4 the blue points are individual prediction nodes). Performance scores are estimated using a leave-one-out cross validation procedure to produce robust estimates of performance of the method.

Uncertainty is estimated using the predictor variance of the top-kriging estimator. This shows that as the LNSE of the training (gauged) sites increases, the predictor variance declines. This therefore shows that more accurate estimates are less uncertain (smaller variance).

The paper is well written and concise, and I commend the authors for the length of their manuscript. It was well received.

I believe that the manuscript can be recommended for publication subject to minor corrections.

Specific Comments:

L106--110: This may be naive, but based on Figure 1 in Pugliese et al 2014 (screenshotted below), it seems to me as if TND does not describe the whole curve, it captures the majority but fails to capture the low duration flood events (high flows). Is that an important aspect of the method? Does that have any impact on the estimation of FDCs?

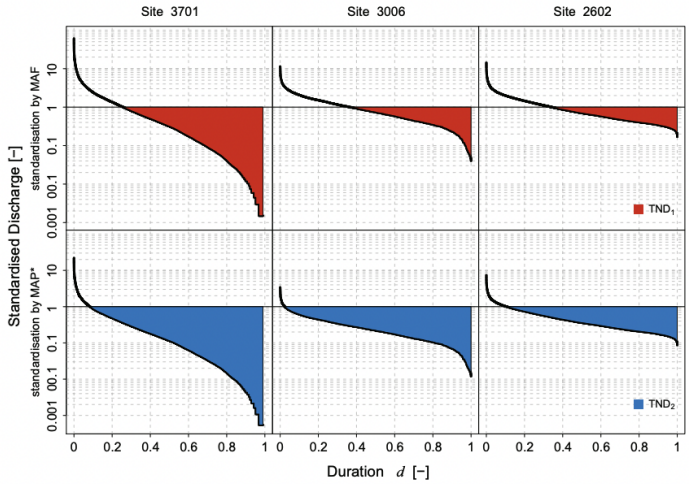

Figure 1. Total negative deviation (TND, filled area) for three catchments with different hydrological behaviours (see Sect. 4). Top panels: TND_1 (red area) for an empirical FDC (black thick line) standardised by mean annual flow (MAF); bottom panels: TND_2 (blue area) for an empirical FDC (black tick line) standardised by $MAP^* = MAP \cdot A \cdot CF$, where MAP is the mean annual precipitation, A is the drainage area, and CF is a unit-conversion factor.

L143–144 & Fig. 3:

L16 & L205: It is claimed that the data layer is available, however I had difficulty accessing the data on Pangea: "*Download Data (login required; moratorium until 2023-12-03)*". Does this mean that we are unable to access the data until the end of 2023?

Download Data (login required; moratorium until 2023-12-03)

Download dataset

L144-145: "*In particular, interpolation is performed only within watersheds including at least one measuring point of the DG-JRC dataset (see blue area in Fig. 3), whereas no FDC prediction has been performed for elementary catchment belonging to watershed without measuring points (see black area in Fig. 3).*"

The final portion of the sentence is a little unclear. I would suggest rewriting:
"*In particular, interpolation is performed only within watersheds including at least one measuring point of the DG-JRC dataset (see blue area in Fig. 3). Watersheds where there are no measuring gauges are excluding from our analysis (see black area in Fig. 3).*"

Fig 4 & L164–169: The results do look good! One thing that I would appreciate some insight into is the large difference between the Median and Mean NSE, this suggests highly skewed performances with long lower-LNSE tails.

Would it be possible to see a histogram of these performances? Or at least a comment on the characteristics of the catchments with worse performing predictions? You have mentioned that the performance is worse for lower-flow sections, so does that mean that for dryer (more arid) catchments the performance is worse?

L233-240: If possible, I believe that it would also be great for the community to be able to test the procedure, perhaps in a Rmarkdown or Jupyter notebook form. I believe that this would add extra weight to the paper, allowing for full reproducibility and transparency of the method. Does the code found here: https://github.com/SimonePersiano/TNDTK/tree/v1.0.0 provide a sufficient overview to reproduce the TNDK method for ourselves?

L230-231: Would it be possible to provide some further context of how this dataset might be used as a “benchmark for the development of hydrological macroscale models”?