Anonymous Referee #1

The manuscript illustrates the calibration procedure of a semi-distributed rainfall-runoff model for flow prediction. The model is built based on the FLEXL model structure plus the Muskingum method for river routing, and it is applied to the Upper Ping catchment in Thailand. For this catchment there are 6 gauges and 32 sub-catchments are delineated. For flow prediction it is required to calibrate a large number of parameters for each sub-catchment; the calibration strategy relies on the observed discharge (between 2001 and 2016), the normalized difference infrared index (from 2002 to 2016) and the soil water Index for moisture conditions (from 2008 to 2016). Results are compared to those provided by a different modeling scheme, namely URBS model.

While the topic of this work is of interest for the scientific community, by providing additional developments for rainfall-runoff modeling, my general opinion is that the manuscript needs additional efforts from the Authors to be considered for publication in HESS. The main point here is that the research gaps motivating the present work and the innovative contribution to the literature are not clearly stated. As for results, a key issue is the estimation uncertainty, which should be quantified for model comparison in terms e.g. of prediction intervals. Due to the large number of calibration parameters, it is expected for the prediction intervals to be almost large. Hence, a fundamental aspect of this work should be how the information introduced here for calibration affects the prediction intervals (the estimation uncertainty) for different model structures. Note that only at the end of the manuscript (in the Conclusion Section) the Authors justify their work as a method to avoid uncertainty in runoff estimation (which is not avoidable in my opinion, but it can be reduced). Further, the text could be reorganized to be more concise and objective oriented, especially in the presentation of the methodology, yet not only. Finally, I suggest to revise figures to improve readability (e.g. remove "1 April" on x axis and use scientific notation in y-axis in figures 4 and A.1, increase text size in figures A.6-A.9).

Answer:

We really appreciate the detailed and constructive review given by referee#1. We would like to answer his/her concern on estimation uncertainty as follows.

FLEX-SD, FLEX-SD-$NDII_{Max-Min}$ and FLEX-SD-$NDII_{Avg}$ were calibrated (2001-2011) and validated (2012-2016) at P.1 station using 50,000 random parameter sets which were determined using the MOSCEM-UA algorithm by finding the Pareto-optimal solutions defined by three objective functions. These include the Kling-Gupta Efficiencies for high flows, low flows, and the flow duration (KGE_E, KGE_L and KGE_F) respectively. To evaluate estimation uncertainty, the 5% best-performing parameter sets were identified as feasible (Hulsman et al., 2019) and were utilized to evaluate model performance. All around 2,500 parameter sets were used to create the box plots of KGE_E, KGE_L and KGE_F at the calibrated station (P.1) and at 5 upstream stations (P.20, P.4A, P.21, P.75 and P.67) (see **Figure 1**). The box plots provided by all models at P.21, P.75, P.67 and P.1 are similar, while FLEX-SD-$NDII_{Max-Min}$ performed slightly better than FLEX-SD and FLEX-SD-$NDII_{Avg}$. However, the box plots of KGE_E and KGE_F contributed by FLEX-SD-$NDII_{Avg}$ at P.4A and P.20 (tropical forest catchments) are exceptionally better than FLEX-SD and FLEX-SD-

$NDII_{Max-Min}$. Observed and calculated hydrographs acquired from the 5% best performing parameter combinations using FLEX-SD-$NDII_{Avg}$ at P.4A and P.20 show a narrow band compared to other 2 models but very similar at other stations, since all 3 KGE values of all models are similar, as shown in **Figure 2**.

In the revised paper, we shall present and discuss the model uncertainty, as required.

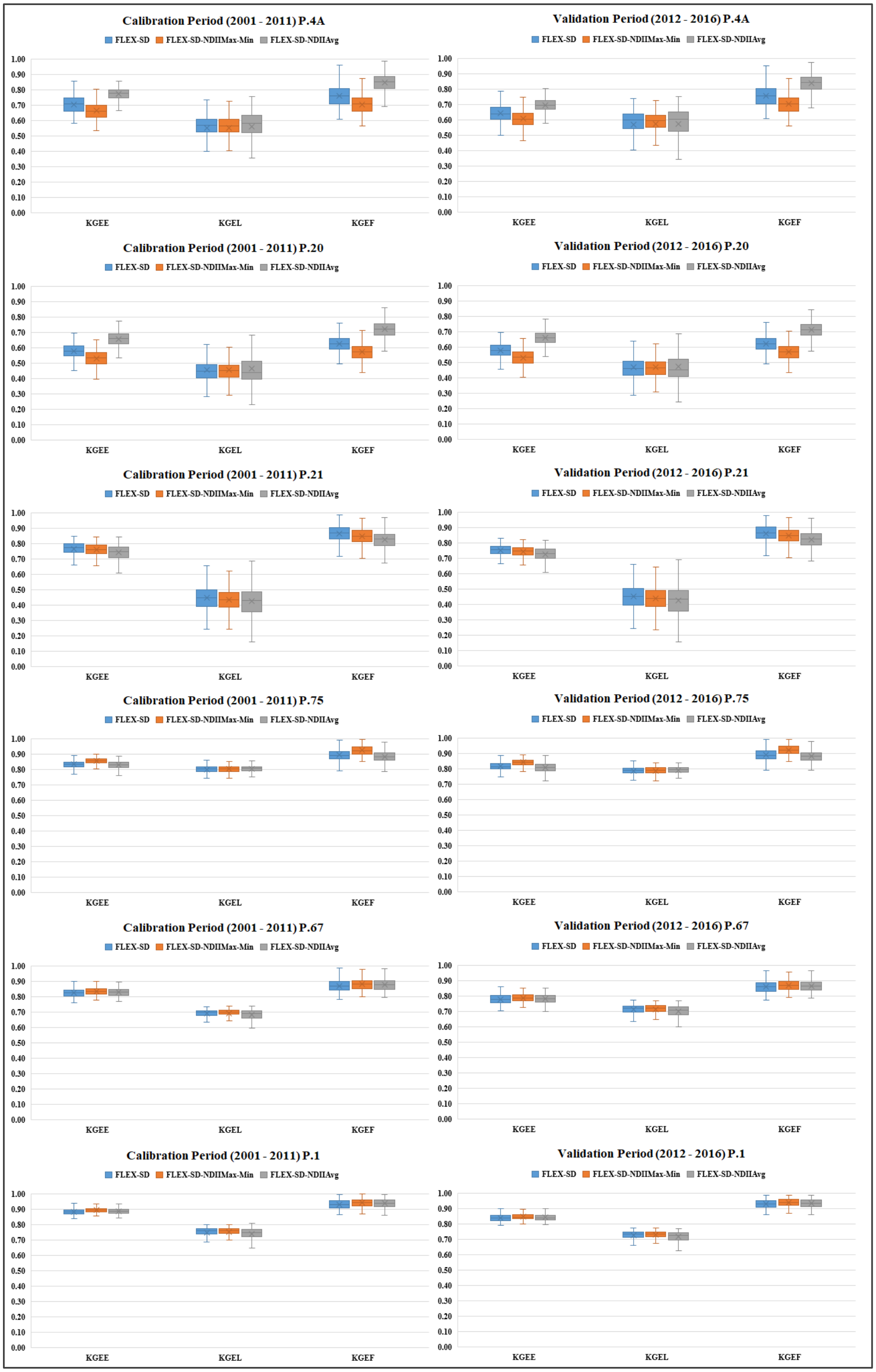

Figure 1: Comparison of box plots of the KGE_E, KGE_L and KGE_F at 6 gauging stations provided by 3 FLEX-SD models using 5% best-performing parameter sets

Figure 2: Comparison of the hydrographs at 6 gauging stations provided by 3 FLEX-SD models using 5% best-performing parameter sets