

Ayala, A. I., Hinostroza, J. L., Mercado-Bettín, D., Marcé, R., Gosling, S. N., Pierson, D. C., and Sobek, S.: Integration of the Global Water and Lake Sectors within the ISIMIP framework through scaling of streamflow inputs to lakes, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2025-3126>, 2025.

Reviewer 1 (Miaohua Mao)

Summary:

This work integrates the stream flows from the nearby catchments into 71 lakes in Sweden, based on the scaling method of the global water and lake sector model. The model performances are compared with referenced model results and observed data from stations. The authors finally conclude that the updated model is satisfactory on modeling the streamflow. The authors have done a good work in explaining the workflow of their coded work, while the reviewers have some comments and suggestions needed to be clarified before it can be published after Minor Revision.

We thank the reviewer for the constructive comments and thoughtful suggestions, which have helped us improve the quality and clarity of the manuscript. Below, we provide a detailed response to each comment, and we indicate how the manuscript has been revised accordingly.

Comments:

Comment 1:

The reviewer's suggestion is avoiding using the specific values for the KGE in this Abstract section. Instead, this section should provide epitome of the entire work in a succinct and clear way.

Reply:

We have revised the Abstract by removing the specific *KGE* values rephrasing the content to provide a more general summary of main findings of the study. The part containing the *KGE* values was replaced by the following revised sentences: "The methodology was applied to 70 lakes across Sweden covering a wide range of sizes, hydrological settings and catchment characteristics. The estimated streamflow was validated against both the streamflow outputs from the hydrological model HYPE and observed data. The comparison demonstrated good agreement in terms of long-term streamflow mean and seasonal pattern, indicating that the proposed approach is capable of producing reliable streamflow estimates without requiring high-resolution local models."

Comment 2:

The authors have done a good work in introducing the previous study work and its research gap, and what they need to do to fill this research gap, i.e., develop the coupled streamflow and lake model via the various discharges (e.g., surface, subsurface, groundwater etc.)

Reply:

We thank the reviewer for this positive feedback. We are pleased that the research gap and motivation of our study were clear and well received.

Comment 3:

Material and methods. This section is generally well written and Fig. 2, 3, and 4 are nice figures to illustrate the procedure of the modeling frame well. Regarding the Section 2.5 Validation of streamflow at catchment scale, it is better by providing the range for the quality of Kling-Gupta efficiency (*KGE*) values. For example, in which ranges stand for model performance is excellent, good, poor etc., and this definition needs some references to support it. Another

comment is to define the CV_{sim} and CV_{obs}, which the reviewer considers as Coefficient of Variation.

Reply:

We appreciate the reviewer's positive feedback on the Material and Methods section and the figures illustrating the modeling framework.

Regarding the suggestion to provide interpretation range for the Kling-Gupta Efficiency (*KGE*), we have now included the following classification, based on Knoben et al. (2019):

$KGE=1$	Perfect
$0.75 \leq KGE < 1$	Very good
$0.5 \leq KGE < 0.75$	Good
$0.25 \leq KGE < 0.5$	Acceptable
$KGE < 0.25$	Poor

This classification has been added to Section 2.5 of the manuscript, along with the appropriate reference.

In addition, we clarify that the *KGE* was calculated using the *KGE()* function from the hydroGOF R package, with method="2012" to follow the revised formulation proposed by Kling et al. (2012). In this version, variability is represented by *KGE_g*, defined as the ratio of the coefficient of variation of the simulated values to the observed values:

$$KGE = \frac{CV_{sim}}{CV_{obs}} = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}$$

where σ and μ denote the standard deviation and mean of the simulated and observed time series respectively.

We also note that the definition of the coefficient of variation (CV) is already provided in Equation 3 of section 2.5.

The *KGE* classification has been incorporated into the Material and Methods section as follows: "Based on Knoben et al. (2019), *KGE* is interpreted as: *KGE*=1 perfect agreement, $0.75 \leq KGE < 1$ very good performance, $0.50 \leq KGE < 0.75$ good performance, $0.25 \leq KGE < 0.50$ acceptable performance and *KGE*<0.25 poor performance."

References:

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, 424–425, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>, 2012.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.

Comment 4:

Line 229: '..... we assume than lake evaporation' maybe changed to '..... we assume that lake evaporation

Reply:

Thank you for noticing this typographical error. It has been corrected as suggested in the revised manuscript.

Comment 5:

Line 223 and other places: The authors please make sure that whether 70 or 71 lakes in Sweden are studied. This needs to be consistent throughout the texts.

Reply:

A total of 71 lakes in Sweden were initially considered. However, streamflow simulations for lake 149288 were included in the validation against observations, but not in the validation against reference values due to data limitations. This discrepancy caused some confusion in the lake count in the different sections of the text. We have now removed lake 149288 from the analysis entirely, as it was only partially included in the original validation. As a result, the total number of lakes studied is now 70. We have carefully reviewed and revised the manuscript to ensure that this number is consistent throughout the text and have removed any reference to the previously included lake 149288.

Comment 6:

Line 248-249: 'For all study sites, the KGE exceeded -0.41, indicating that the simulated streamflow provided added value compared to using long-term mean values.' The reviewer is a little bit confused that a negative value of KGE (e.g., -0.41) means this revision provides added value.

Reply:

Negative *KGE* values generally indicate poor model performance, the original intent of the sentence was to highlight that the model still performed better than a simple baseline (e.g. using the mean observed discharge as prediction).

To avoid confusion, we have revised the sentence for clarity: "For all study sites, the *KGE* exceeded -0.41, indicating that the simulated streamflow provided added value compared to simple prediction based on the long-term mean streamflow"

Comment 7:

In the Results Section. The authors have compared their model with reference results and observed data, respectively. The reviewer considers that would that also necessary to compare the reference results with the observed data. By doing so, the reader could have better understanding that whether the developed model improves its performance or not, compared with the traditional model (i.e., the reference results).

Reply:

Lake 149288, which had streamflow observations available but lacked reference data, was excluded from the original validation analysis to maintain consistency throughout the text. To ensure comprehensive comparison, additional lakes with both streamflow observations and reference data were included in the analysis. As a result, streamflow simulations were compared with observations for a total of 10 lakes.

Additionally, in the revised results sections, we have incorporated a comparison between the reference and observed streamflow, using the same performance metrics. Note that, while observations were available for 10 lakes, comparison were conducted for only 9 lakes, as reference streamflow data are available from 1981-2010. A new table (Table S4) presenting performance metrics has been added to the supplementary material and a brief description of the performance has been added to the result section.

Brief description of the performance added to the result section: "Finally, a further evaluation was conducted by comparing reference and observed streamflow for 9 study sites (note that the reference and observations datasets cover different time periods, which limited direct comparability in the 10 study sites for which observations were available) (Table S4). At the monthly scale, the average *KGE* was 0.44 ± 0.44 (with *KGE_r* of 0.65 ± 0.23 , *KGE_b* of 1.12 ± 0.34 , *KGE_g* of 1.13 ± 0.46), indicating on average acceptable agreement with substantial inter-site differences. At the yearly scale, performance improved to *KGE* of 0.55 ± 0.26 (with *KGE_r* of 0.78 ± 0.12 , *KGE_b* of 1.12 ± 0.34 , *KGE_g* of 0.77 ± 0.19). Overall, these results demonstrate that the

scaling method provides added value, improving the simulations of streamflow compared with standard catchment-scale hydrological models.”

Table S4 of supplementary material:

Lake	Name	Latitude	Longitude	<i>MBE</i>	<i>RMSE</i>	<i>NRMSE</i>	<i>NSE</i>	<i>KGE</i>	<i>KGE_r</i>	<i>KGE_b</i>	<i>KGE_g</i>	Frequency
102	Mälaren	59.49	16.79	31.66	119.55	0.16	0.24	0.56	0.64	1.19	0.83	monthly
102	Mälaren	59.49	16.79	32.17	38.46	0.22	0.13	0.73	0.87	1.19	0.86	yearly
104	Vättern	58.33	14.49	3.03	25.64	0.31	-1.25	0.24	0.39	1.07	1.45	monthly
104	Vättern	58.33	14.49	2.94	9.43	0.22	0.33	0.47	0.63	1.07	0.64	yearly
105	Vänern	58.88	13.55	-25.90	136.73	0.14	0.60	0.71	0.79	0.95	0.81	monthly
105	Vänern	58.88	13.55	-25.32	57.00	0.10	0.80	0.86	0.92	0.95	0.89	yearly
12423		62.05	14.15	-51.94	77.45	0.23	-1.20	-0.34	0.53	0.57	2.18	monthly
12423		62.05	14.15	-51.70	53.28	0.55	-4.30	0.56	0.90	0.57	0.97	yearly
12791		60.07	15.57	4.17	17.94	0.17	-0.53	0.48	0.65	1.19	1.33	monthly
12791		60.07	15.57	4.24	5.76	0.32	-0.13	0.60	0.72	1.20	0.78	yearly
12809	Erken	59.84	18.60	0.69	1.27	0.40	-2.72	-0.14	0.27	1.86	0.84	monthly
12809	Erken	59.84	18.60	0.70	0.73	0.57	-5.03	-0.01	0.75	1.87	0.55	yearly
12965	Roxen	58.49	15.63	14.20	31.24	0.20	0.41	0.62	0.75	1.18	0.78	monthly
12965	Roxen	58.49	15.63	14.86	22.19	0.29	0.08	0.35	0.74	1.18	0.43	yearly
142240		66.66	22.22	1.20	7.08	0.08	0.85	0.88	0.93	1.08	0.95	monthly
142240		66.66	22.22	1.18	3.07	0.21	-0.02	0.59	0.60	1.08	0.99	yearly
152977	Hasselasjön	62.08	16.78	-0.31	2.78	0.06	0.88	0.92	0.94	0.96	1.04	monthly
152977	Hasselasjön	62.08	16.78	-0.18	1.10	0.11	0.78	0.79	0.89	0.98	0.82	yearly

Comment 8:

The current writing of this part could be improved in a more detailed way. For example, providing some skill metric values that are specified (e.g., various KGE values), so that this work could be better summarized in a more strict way.

Reply:

We have revised the Results section to improve the clarity and structure of the performance evaluation. These changes make the summary more quantitative and structured, as suggested, and improve the overall readability of the results. The changes can be found throughout the revised results section. Furthermore, as show in the Supplementary Material (Tables S2-S4) additional performance metrics including *MBE*, *RMSE*, *NRMSE* and *NSE* are already provided for each study site. We believe these revisions address your concerns and improve the overall presentation of the results.

Revised result section:

“The performance of the scaled streamflow simulations from grid cells to the catchment scale (hereafter referred to as simulations) was evaluated for monthly time series over the period 1981-2010 across 70 study sites (Figure 6A; Table S2). The average Kling-Gupta efficiency, *KGE*, was 0.59 ± 0.18 (mean $\pm$ standard deviation), with individual values ranging from -0.07 to 0.86. For all study sites, the *KGE* exceeded -0.41, indicating that the simulated streamflow provided added value compared to simple prediction based on the long-term mean streamflow.

To better diagnose performance, *KGE* was decomposed into its three components: correlation (*KGE_r*), bias (*KGE_b*) and variability (*KGE_g*). The average *KGE_r* was 0.79 ± 0.08 , suggesting generally strong agreement between reference and simulated streamflow timing. A total of 52 out of 70 sites (74 %) exhibited a *KGE_r* greater than 0.75, reflecting very good representation of seasonal and interannual flow dynamics. The bias component, *KGE_b*, averaged 1.06 ± 0.30 , was close to the optimal value of 1, indicating that the overall volume of streamflow was, on average, well captured. However, the relatively high standard deviation highlights substantial variability in bias among the study sites. Only 39 study sites (56%) had a *KGE_b* within the range of 0.75 to 1.25, indicating that for a significant number of study sites, deviations in simulated streamflow volumes were a key source of reduced performance. The variability component, *KGE_g*, averaged 0.88 ± 0.22 , indicating generally very good representation of streamflow variability, though with some underestimation of streamflow. Similar to *KGE_r*, 52 sites (72%) had *KGE_g* values within the range of 0.75 to 1.25. In summary, the simulations demonstrated generally very good performance in reproducing time and variability of monthly streamflow across study sites. However, discrepancies in the magnitude of the simulated streamflow, reflected in the higher variability of *KGE_b*, where the bias component more frequently deviated from its optimal range compared to the correlation and variability components (Figure 6A).

The inter-annual variability of streamflow was assessed by comparing the simulated and reference annual average streamflow (Table S2; Figure S2). The average values of the *KGE* components were *KGE_r* of 0.77 ± 0.14 , *KGE_b* of 1.06 ± 0.30 , *KGE_g* of 1.06 ± 0.31 , indicating an overall very good performance in responding differently to wet and dry years. The relatively high *KGE_r* suggest that the simulated streamflow timing was very well captured. However, the standard deviations of both *KGE_b* and *KGE_g* were relatively large, reflecting considerable variability in the ability to simulate annual streamflow volumes and variability. While the mean values of *KGE_b* and *KGE_g* were close to the optimal value of 1, these high standard deviations indicate that performance differed substantially among study sites, with some ties showing over- or underestimation of interannual streamflow characteristics. The combined *KGE* for interannual streamflow was 0.54 ± 0.23 , which is slightly lower but comparable to the *KGE*

(0.59 ± 0.18) for monthly streamflow, suggesting that the model maintained reasonable skill across both temporal scales.

Performance was further analysed based on the streamflow scaling approach. Of the 70 study sites, 68 were analysed using Approach I ($A_{lake} \leq A_{grid}$), with 39 study sites following Approach I.a ($N \leq 1$) and 29 study sites following Approach I.b ($N > 1$). The average KGE was 0.56 ± 0.15 for Approach I.a and 0.60 ± 0.21 for Approach I.b, indicating similar performance across these two subcategories. In 5 of the Approach I.b study sites, an additional comparison was made between counting all grid cells at the last level versus only those with the steepest slope. In both cases, the performance was acceptable, and the differences between KGE and its components were marginal. When all grid cells were counting at the last level, the KGE was 0.49 ± 0.31 , with KGE_r of 0.76 ± 0.05 , KGE_b of 0.87 ± 0.17 , KGE_g of 1.23 ± 0.49 ; when only the steepest grid cells were counted the KGE was 0.48 ± 0.31 , with KGE_r of 0.74 ± 0.07 , KGE_b of 0.85 ± 0.15 , KGE_g of 1.22 ± 0.49 . These small differences suggest that the method is robust to the choice of how grid cells are selected at the last level.

The Approach II ($A_{lake} > A_{grid}$) was applied to the two largest lakes in this study: Vänern (105) and Vättern (104), the performance was very good in both Vänern and Vättern (Figure 6), with a KGE of 0.77 (KGE_r of 0.85, KGE_b of 0.97, KGE_g of 1.17) and 0.79 (KGE_r of 0.79, KGE_b of 0.97, KGE_g of 1.00), respectively. Lake Mälaren (102), the third largest lake in Sweden, extends over 9 grid cells (Figure S1); however, its A_{lake} (of 1083 km^2) does not exceed the A_{grid} of 1580 km^2 due to its irregular and branched shape. Scaling streamflow Approach I.b ($A_{lake} \leq A_{grid}$ for $N > 1$) and Approach II ($A_{lake} > A_{grid}$) were tested (Figure 2). For Approach I.b, simulated streamflow showed good performance at the seasonal scale, with KGE of 0.71 (KGE_r of 0.72, KGE_b of 1.04, KGE_g of 1.06); however, errors in reproducing the timing of flow reduced the overall performance. In Approach II, the simulated seasonal streamflow was less accurate, with a KGE of 0.47 (KGE_r of 0.52, KGE_b of 0.98 and KGE_g of 0.80). The errors were caused by either a reduced ability to accurately reproduce the timing of flow increases and decreases; and an underestimation of the magnitude of the variability, although it was still reasonably good.

In addition, the performance of simulated streamflow was assessed by comparing simulations with observations for 10 study sites, which are both representative lakes in the ISMIP3 Global Lake Sector and for where observations are available (Figure 6B; Table S3). At the seasonal scale, the average KGE was 0.46 ± 0.21 , with KGE_r of 0.65 ± 0.12 , KGE_b of 1.10 ± 0.20 , KGE_g of 1.07 ± 0.40 . Overall performance was acceptable but was primarily limited by mismatches in flow timing. At the annual scale, the performance of the scaling streamflow from grid cells to catchment scale was good (KGE of 0.70 ± 0.15 , with KGE_r of 0.85 ± 0.05 , KGE_b of 1.10 ± 0.20 , KGE_g of 0.98 ± 0.20), indicating strong agreement in timing, bias and variability across study sites (Figure S2).

Finally, a further evaluation was conducted by comparing reference and observed streamflow for 9 study sites (note that the reference and observations datasets cover different time periods, which limited direct comparability in the 10 study sites for which observations were available) (Table S4). At the monthly scale, the average KGE was 0.44 ± 0.44 (with KGE_r of 0.65 ± 0.23 , KGE_b of 1.12 ± 0.34 , KGE_g of 1.13 ± 0.46), indicating on average acceptable agreement with substantial inter-site differences. At the yearly scale, performance improved to KGE of 0.55 ± 0.26 (with KGE_r of 0.78 ± 0.12 , KGE_b of 1.12 ± 0.34 , KGE_g of 0.77 ± 0.19). Overall, these results demonstrate that the scaling method provides added value, improving the simulations of streamflow compared with standard catchment-scale hydrological models.

We conclude that the overall performance of the scaled streamflow simulations matched satisfactorily to both reference (derived from the hydrological model HYPE) and observed streamflow (Figure 7; Figures S3-S11).”