Water Resources Research

Supporting Information for

A framework for global-scale river discharge estimation by assimilating satellite altimetry

Menaka Revel[1], Daiki Ikeshima[2], Dai Yamazaki[1], and Shinjiro Kanae[2]

[1]Institute of Industrial Sciences, The University of Tokyo
[2] Department of Civil and Environmental Engineering, Tokyo Institute of Technology

Contents of this file

Introduction

In this supporting information, we provide the supplemental text, figures and table for the development of the framework for estimating river discharge by assimilating satellite altimetry. Here we introduce the generating input runoff, data assimilating procedure, empirical localization parameters, estimating SWOT observation error, upstream inflow correction, evaluation of river discharge accuracy using KGE, map of accumulated overpasses per SWOT cycle, variation of NSEAI with accumulated overpasses, and map of assimilation frequency.

Text S1. Generating input runoff forcing

We used HTESSEL (ECMWF) model runoff out form E2O WRR2 (Dutra et al., 2017) for true simulation. The original runoff was used for simulations in true simulation. The remaining runoff outs were perturbed by multiplying a random number to generate 18 ensembles. Table S1 presets the random values used for each LSM/GHM runoff output for generation ensembles for perfect model experiment and imperfect model experiment.

Table S1: Generation of input runoff forcing for perfect model experiment

Simulation	LSM/GHM	Random Value (Perfect Model experiment	Random Value (Imperfect Model experiment
True	HTESSEL	Original Runoff is used	Original Runoff is used
Corrupted/Assimilated	PCR-GLOBWB	0.93	0.95
		0.97	1.00
		1.04	1.08
	JULES	0.79	0.91
		0.96	0.96
		0.98	1.01
	LISFLOOD	0.88	0.91
		0.94	0.99
		1.08	1.02
	ORCHIDEE	0.81	0.93
		1.03	1.01
		1.12	1.06
	WaterGAP3	0.93	0.85
		0.97	0.97
		1.00	1.18
	W3	0.91	0.85
		0.99	1.01
		1.08	1.05

Text S2: Data assimilation procedure

The LETKF (Hunt et al., 2007; Ott et al., 2004) algorithm was used in this study to efficiently perform data assimilation in global scale. Here, we used the SWOT-observed water surface elevation as the 'observed variable' of the data assimilation procedure. The model forecasts were propagated using CaMa-Flood hydrodynamic model. Then the assimilated water state was diagnosed using LETKF algorithm and update the initial conditions for next days` simulation. The water state of the proceeding step (i.e., initial water storage) was computed with data assimilation using LETKF with Equation (A1):

$$x^a = \bar{x}^f + E^f \left[\tilde{P}^a (HE^f)^T \left({}^{R}/_{W} \right)^{-1} (y^o - H\bar{x}^f) + \sqrt{m-1} \left(\tilde{P}^a \right)^{\frac{1}{2}} \right] \quad \text{(S1)}$$

where x^a is the assimilated WSE; x^f is the forecasted WSE of each parallel CaMa-Flood ensemble, $\bar{x}^f$ is the mean forecasted value of ensemble members; E^f is the model forecast error covariance matrix, which consist of perturbations which calculated using;

$$E^f = x^f - \bar{x}^f \tag{S2}$$

$\tilde{P}^a$, and $\left(\tilde{P}^a\right)^{\frac{1}{2}}$ were calculated in Equations (S3) and (S4), respectively:

$$\tilde{P}^a = VD^{-1}V^T \tag{S3}$$

$$\left(\tilde{P}^a\right)^{\frac{1}{2}} = VD^{-\frac{1}{2}}V^T \tag{S4}$$

where,

$$VDV^T = (m-1)\,{}^{I}/_{\Delta} + (HE^f)^T\left({}^{R}/_{w}\right)^{-1}HE^f \tag{S5}$$

where m is the number of ensemble members (= 20), I is an identity matrix, Δ is the covariance inflation parameter (estimated adaptively using innovative statistics following Miyoshi, (2011), with background variance of 0.04^2), H is the observation operator which is linearly related to the observation and the state, R is the observation error covariance matrix, which is a diagonal matrix having observation error variances in the diagonal (explained in Appendix C); and w is the observation localization weightage (explained in Appendix B). LETKF applied using the equation (A1) to a 'empirical local patch' (explained in Appendix B), which is a small domain around each observation point where the observation has correlations with model state variables. The state variables are independently updated within each empirical local patch.

Text S3: Empirical Localization Parameters

Empirical localization parameters were derived using the spatial auto-correlation of simulated WSEs adaptively. We developed physically-based local patch using CaMa-Flood modelled WSE using runoff simulated by minimal advanced treatments of surface interaction and runoff (MATSIRO: Takata et al., 2003) LSM forced by S14 (Iizumi et al., 2017). The empirical local patches where derived by defining a threshold to the spatial dependency weights calculated by conducting semi-variogram analysis on transformed WSE data. Transformation of WSEs involved three steps: (1) removing trends, (2) removing seasonality, and (3) standardizing. Then, we derived the observation localization weights using Gaussian function using localization lengths corresponds to the threshold defined to the spatial dependency weight to bound the empirical local patch. For further information on deriving physically based adaptive empirical localization parameters, please refer to Revel et al., (2019, 2018b).

Text S4: The SWOT observation error

The SWOT mission sets a goal of 10cm accuracy for water area **>**= 1km^{2}at the WSE measurement. However, the actual accuracy of future distributing observation data is unclear since it varies with river width, river length, surrounding topography (Durand et al., 2010) or even distance from the satellite track (varies between 4~10cm) (Desai et al., 2018). In this study, we modelled observation error to be normally distributed with zero mean and variance of σ_h.

$$\sigma_h = \begin{cases} \frac{1}{WL}0.10 & ,WL \geq 1.0\ km^2 \\ \frac{1}{WL}0.25 & ,1.0\ km^2 > WL \geq 0.625\ km^2 \\ 0.25 & ,WL < 0.625\ km^2 \end{cases} \quad \text{(S6)}$$

where W and L are river width and river length, respectively. We adopt L to be 1km as we assume only the observations near the outlet of the unit-catchment can be used for data assimilation because CaMa-Flood unit-catchments show internal variability in WSE, especially in steep upstream reaches. We used $\sigma_h$as the diagonal components in the observation error covariance matrix in LETKF.

Figure S1 presents the global map of observation error variance calculated using the equation (S6). Most of the upstream reaches where $W < 625\ m$ are having observation error covariance of 0.25m. Downstream of large rivers such as Amazons, Congo, Ob, Lena, etc. show smaller variances below 0.10m. Therefore, the observation error variance demonstrates a spatial variability.

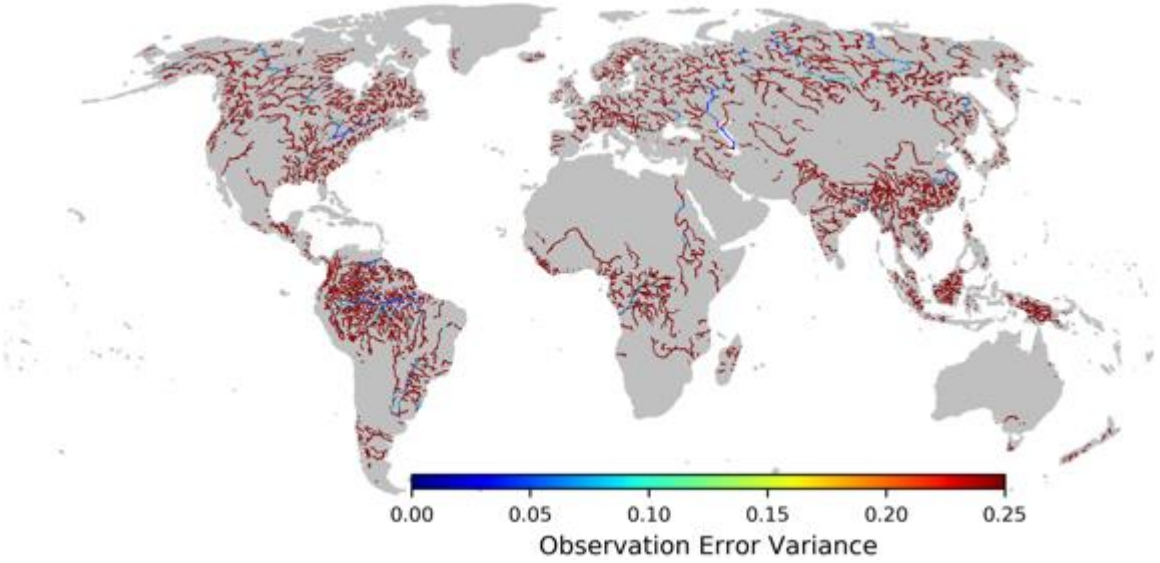

Figure S1. Observation error variance calculated using the equation (S6). Pixels with mean discharge > 100m3/s were used for visualization purposes.

Text S5: Importance of upstream inflow correction

To evaluate the necessity of upstream inflow correction, we performed a 'partially observed experiment,' simulating a situation where only part of the observation is available. In this experiment the inflow correction from the upstream was not corrected at all. Here, we used the settings similar to perfect model experiment and set the whole Amazon River basin as an experimental target but disabled the observations westward (upstream) of the midstream location Y (Figure S2b). Therefore, the location Y received local state correction due to its local SWOT observation, but the inflow from upstream was not corrected (Figure S2b). Aside from the observation area, the data assimilation process was identical to that of the perfect model experiment for the whole Amazon River basin. In the partially observed experiment, the improvement by data assimilation at the midstream location Y was mostly lost. In the partially observed experiment, the assimilated discharge was similar to the corrupted discharge at the

location X (Figure S2a). The assimilated discharge was slightly improved at the location Y due to the local assimilations (Figure S2b). The AI was very low, reaching only ~0.5 even on days with local observations. On the other hand, the assimilation in the location Z was very successful, because that location is situated sufficiently far away from the unobserved area and the local patch is large enough to receive local observations every day. The spatial distribution of the NSEAI showed that decrease in assimilation efficiency up river reaches around 100km downstream of location Y. But the far downstream reaches were not affected by the unavailability of the observations in upstream. This suggests the propagation of corrected discharges from upstream pixels (i.e., upstream inflow correction) is important. Data assimilation should be applied to the entire upstream region to achieve reasonable estimations of discharge in continental-scale rivers with large drainage areas.

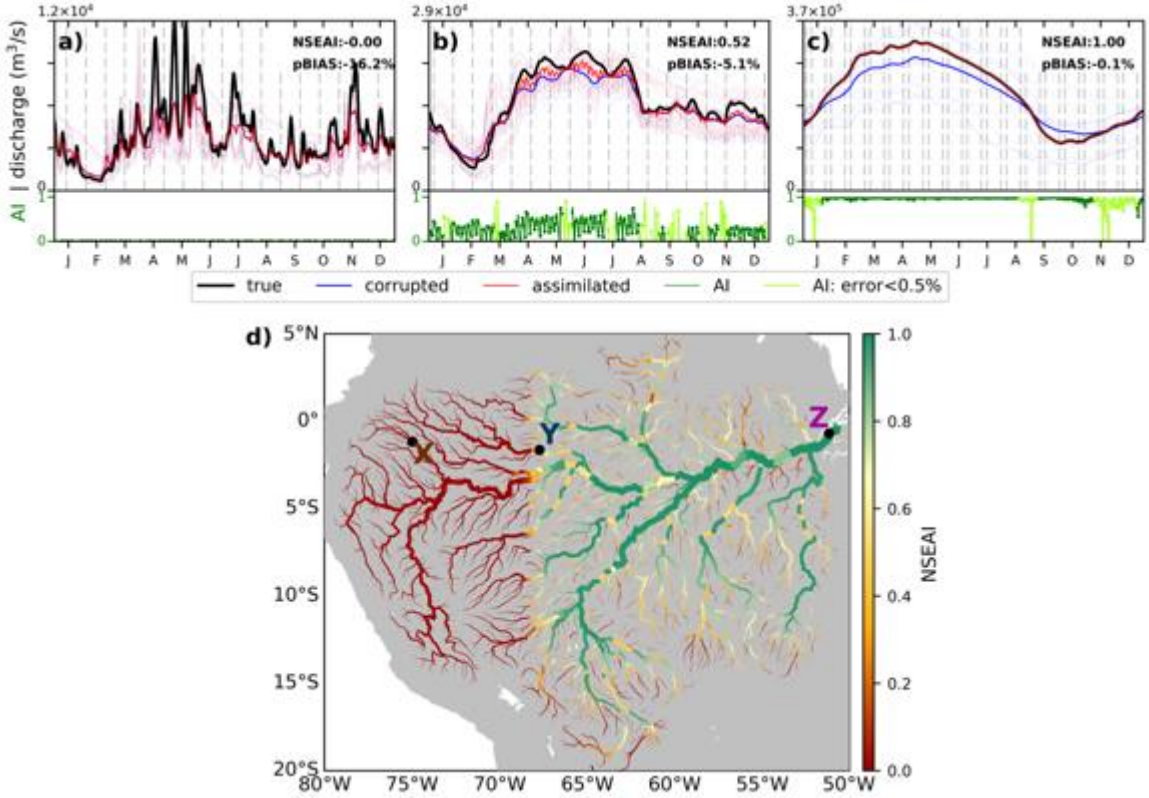

Figure S2. a)-c) Hydrograph for locations X, Y, and Z as in section 4.1.1; and d) NSEAI map for partial observed experiment. Rules are similar to Figure 5**.**

Text S6. Global River Discharge Estimation Accuracy

a) Perfect model Experiment

To further evaluate the assimilation effectiveness, we compare the KGE metric of assimilated and corrupted simulations at global scale (Figure S3). KGE offers diagnostic insights into the performance of our assimilation framework on estimating river discharge. KGE is a combination of correlation, relative bias, and variability which presets the ability to reproduce of temporal dynamics with preserving flow durations. KGE of assimilated simulation (Figure S3a) results are similar to the global map of NSEAI (Figure 7m), large rivers in low latitudes (i.e. Amazons, Congo, Nile, Mekong, Niger, Mississippi) and rivers in higher latitudes show higher KGE values (>0.8). But the relatively small rivers in south-east Asia, Europe and East Coast of Northern America shows

slightly lower KGE values (0.8~0.6). On the other hand, KGE values of the corrupted simulation has values around 0.4~0.6 due to the difference of the runoff forcing from the true simulation. Almost all the global rivers are having >0.1 difference in KGE between assimilated and corrupted simulation. The KGE difference is also similar to NSEAI (Figure 7m) large low latitude rivers and higher latitude rivers demonstrate large difference (≥0.4) and smaller rivers in low latitudes shows lower KGE difference (≈0.1). The high KGE difference here means that the data assimilation can benefit the hydrodynamic model, under the assumption that core hydrodynamic model has the correct water physics and river routing system. Therefore, assimilated river discharge shows better simulation efficiency than the non-assimilated simulation (corrupted simulation) in most of the global rivers.

b) Imperfect model Experiment

The KGE-statistic was calculated for evaluating the results of imperfect model experiment and illustrates the insights into the performance of our assimilation framework. Figure S4 shows the global extent of the KGE at imperfect model experiment; KGE of river discharge at assimilated simulation (Figure S4a), corrupted simulation (Figure S4b), and the difference between two simulations (Figure S4c) are presented. Similar to the result of NSEAI (Figure 13), the KGE in the assimilated simulation was large at midstream and downstream locations of the large-scale rivers. Although the upstream locations had a small KGE value, the value rises in the downstream and it becomes almost 1.0 in the downstream. Furthermore, KGE of assimilated simulation are higher than that of corrupted simulation in almost all the global rivers (Figure S4c), difference of KGE was positive at most locations. This denotes that data assimilation of SWOT observation has the potential to correct the simulation even when the model has erroneous geographical parameters (i.e., Manning's Coefficient) and inaccurate runoff forcing. However, an important characteristic of this result must be pointed out: Some locations near river mouth of large rivers had a high KGE value even in the corrupted simulation (Figure S4b). The high KGE values at downstream of the large rivers are due to the coincidence of the true and corrupted discharge well agrees with each other (true and corrupted) (i.e., Congo). In addition, the KGE evaluates the prediction power of model, by focusing on seasonal variation in terms of correlation, relative bias, and variability. As a consequence, KGE was able to remain high at downstream reaches of larges rivers. Those locations tend to have similar seasonal trend (i.e. high-water season happened in the same time) between true and assimilated/corrupted simulation, or have a long period when seasonal trend is almost the same (i.e. discharge at winter season was almost same). Therefore, the high KGE here only means how the data assimilation can benefit the model, under the assumption that core hydrodynamic model error is included in Manning's Coefficient. Hence, the data assimilation is very effective to improve global river discharge under such assumptions. To make data assimilation effective under the real operation of SWOT satellite, hydrodynamic model uncertainties need be decreased.

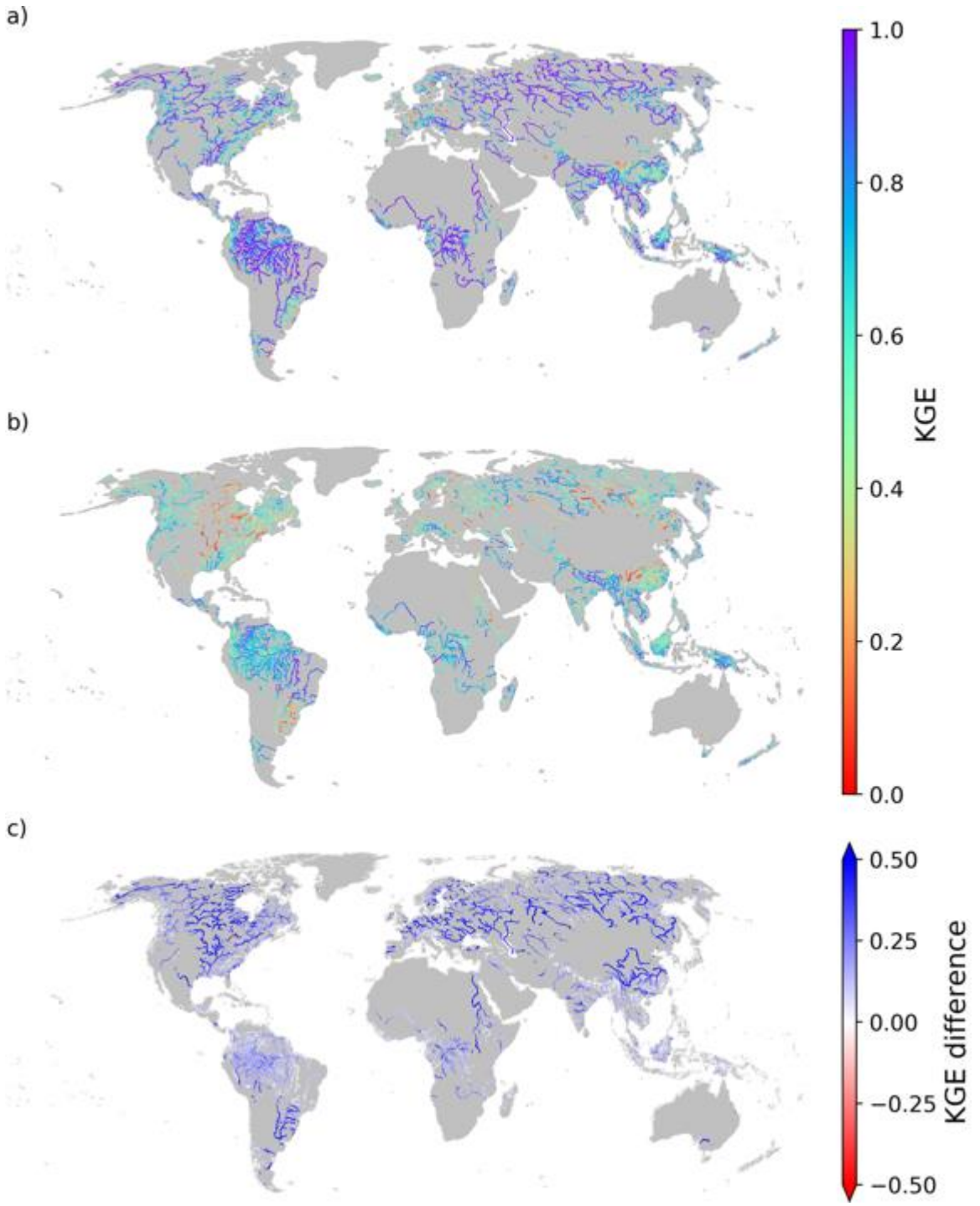

Figure S3. KGE coefficient coefficients of river discharge of a) assimilated and b) corrupted simulations. c) Difference in KGEs for assimilated and corrupted simulations for perfect model experiment. Pixels with mean discharge > 100 m3/s were shown for visualization purposes.

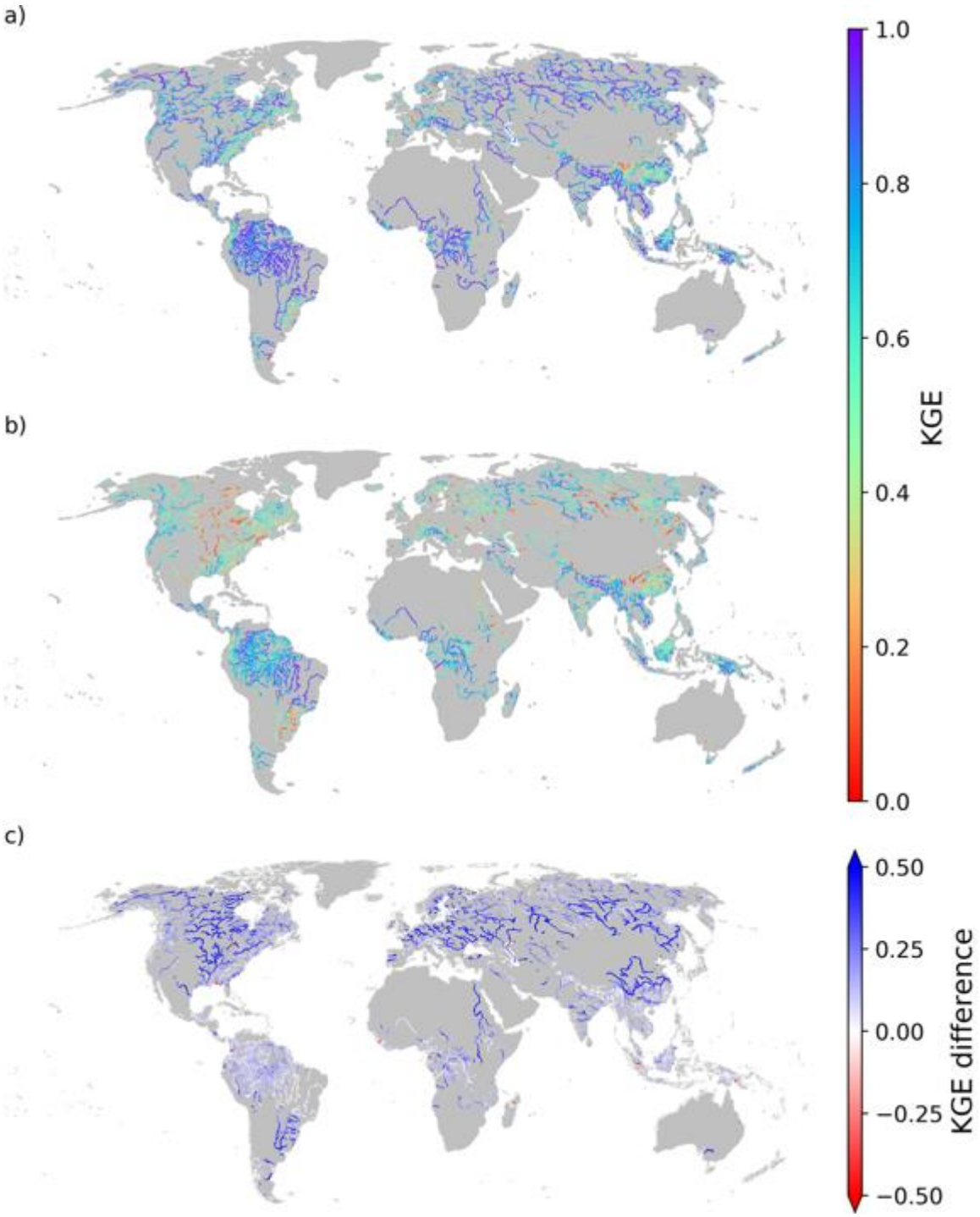

Figure S4. KGE coefficient coefficients of river discharge of a) assimilated and b) corrupted simulations. c) Difference in KGEs for assimilated and corrupted simulations for imperfect model experiment. Pixels with mean discharge > 100 m3/s were shown for visual purposes.

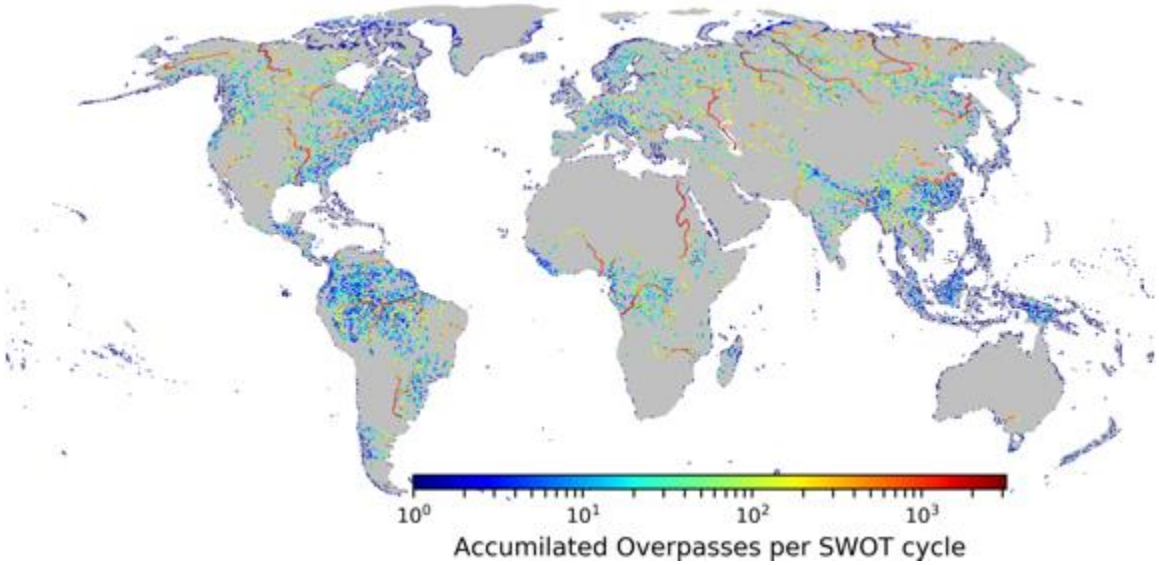

Figure S5: Accumulated Overpasses per SWOT cycle. Pixels with mean discharge > 100 m3/s were shown for visual purposes. Color bar is presented in log scale.

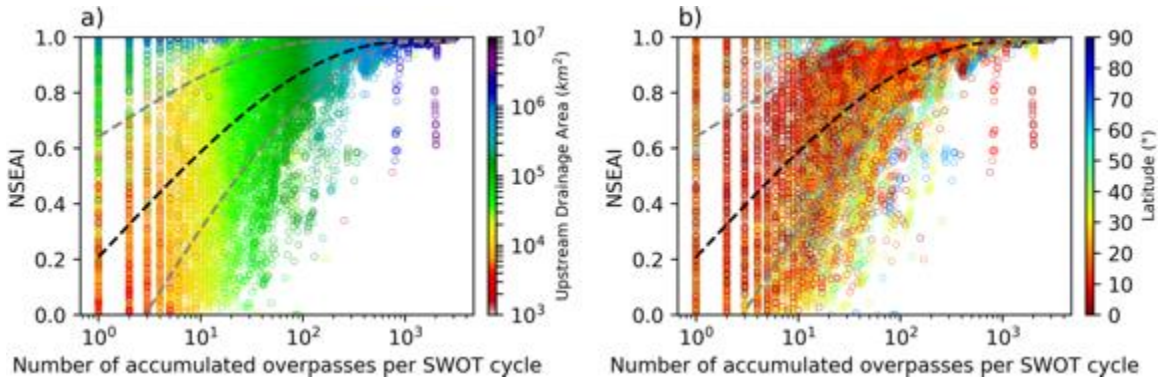

Figure S6: Relationship between accumulated overpasses with NSEAI according to a) upstream drainage area and b) latitude. b) Upstream drainage area with number of accumulated overpasses according to the latitude. The colors represent the upstream drainage area of each pixel in a) and the latitude of each pixel in b) and c).

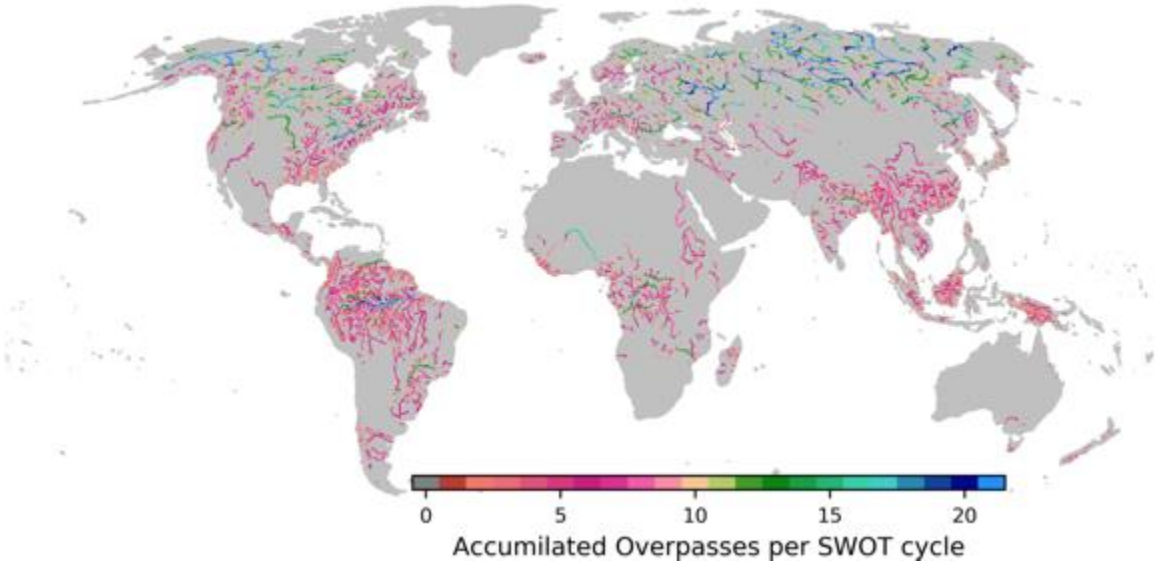

Figure S7: Assimilation Frequency per SWOT cycle. Pixels with mean discharge > 100 m3/s were shown for visual purposes.

Reference

1. Desai, S., Rodriguez, E., Fernandez, D. E., Peral, E., Chen, C. W., Bleser, J.-W. De, & Williams, B. (2018). *Surface Water and Ocean Topography Mission (SWOT) Science Requirements Document. Rev B*. California. Retrieved from https://swot.jpl.nasa.gov/docs/D-61923_SRD_Rev_B_20181113.pdf
2. Durand, M., Fu, L. L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., & Esteban-Fernandez, D. (2010). The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, *98*(5), 766–779. https://doi.org/10.1109/JPROC.2010.2043031
3. Dutra, E., Gianpaolo, B., Jean-Christophe, C., Munier, S., Burke, S., Fink, G., et al. (2017). *Report on the improved water resources reanalysis Deliverable*. Retrieved from http://earth2observe.eu/files/Public Deliverables/D5.2 - Report on the Improved Water Resources Reanalysis (WRR tier 2).pdf
4. Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008
5. Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N., & Nishimori, M. (2017). Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes. *Journal of Geophysical Research*, *122*(15), 7800–7819. https://doi.org/10.1002/2017JD026613
6. Miyoshi, T. (2011). The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. *Monthly Weather*

Review, *139*(5), 1519–1535. https://doi.org/10.1109/ICCSCE.2012.6487150
7. Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., et al. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *56*(5), 415–428. https://doi.org/10.3402/tellusa.v56i5.14462
8. Revel, Ikeshima, Yamazaki, & Kanae. (2019). A Physically Based Empirical Localization Method for Assimilating Synthetic SWOT Observations of a Continental-Scale River: A Case Study in the Congo Basin. *Water*, *11*(4), 829. https://doi.org/10.3390/w11040829
9. Revel, M., Yamazaki, D., & Kanae, S. (2018). Model Based Observation Localization Weighting Function for Amazon Mainstream. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, *74*(5), I_157-I_162. https://doi.org/10.2208/jscejhe.74.5_I_157
10. Takata, K., Emori, S., & Watanabe, T. (2003). Development of the minimal advanced treatments of surface interaction and runoff. *Global and Planetary Change*, *38*(1–2), 209–222. https://doi.org/10.1016/S0921-8181(03)00030-4