https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Globally Applicable Discharge Reanalysis using SWOT observations

Izzy Probyn[1,2], Jeffrey Neal[1], Stephen J. Chuter[2], and Paul Bates[1,2]

[1]School of Geographical Sciences, University of Bristol, Bristol, UK
[2]Fathom, Floor 2, Clifton Heights, Bristol, UK

Correspondence: Izzy Probyn (izzy.probyn@bristol.ac.uk)

Abstract. Global river discharge observations are critical for climate and hydrological research, used for water resource management, risk mitigation, infrastructure and environment conservation, among many other areas. However, existing observations are limited in availability, accuracy and spatial extent, with uneven geographical distribution and high levels of uncertainty often recorded. The NASA SWOT mission provides the opportunity to fill this gap, providing water level observations for all rivers exceeding $100\,\mathrm{m}$ in width worldwide. Combined with the daily discharge model GRADES-hydroDL, we present a statistical method to produce a set of global virtual gauging stations with the same temporal resolution as the satellite observations. The high level of accuracy that the SWOT water surface elevation observations allows us to improve the dynamics of discharge estimation over the input discharge model. For our validation gauges, where we compared modelled discharge to observed values, 66% of Pearson R values are larger than 0.9, showing the method's skill at replicating temporal patterns in recorded flow. The median Root Mean Square Error, $RMSE$, of the input discharge set is $36.3\,\mathrm{m}^3\,\mathrm{s}^{-1}$ larger than that of our proposed method, and the NSE 0.3 lower. However, our method fails to improve the bias; the median absolute normalised bias is 0.10 (10%) higher in our model than the input model, indicating that while the inclusion of Earth Observation can greatly improve the discharge time series dynamics, the trade-off is an increased bias in the distribution. Nevertheless, as the record of SWOT observations increases in length, these results should improve and the margins of bias difference decrease. The result is a globally applicable reanalysis dataset of discharge, which will complement existing physically based models and ground observations, and can be widely used within global hydrological models.

1 Introduction

Accurate estimates of river discharge are critical for global hydraulic models (Hunger and Döll, 2008), as well as a variety of other applications such as flood forecasting and water resources management (Gourley et al., 2013; Bates et al., 2021). Many models exist to predict river discharge on a global scale (Yang et al., 2025; Alfieri et al., 2013; Sutanudjaja et al., 2018; Schmied et al., 2021), using hydrological and machine learning methods to simulate the flow of water through predefined river networks. However, analysis of these methods shows large uncertainty, from a variety of sources including process simplification and reliance on prior assumptions and input data (Ward et al., 2015; Sampson et al., 2015; Teng et al., 2017).

A key restriction in estimating discharge is the limited availability of observations from gauging stations (Galelli et al., 2025; Do et al., 2018). In particular, they are unequally distributed, concentrated heavily in parts of Europe and North America, while much of Asia and Africa remain largely unmonitored, or have data that is not publicly available. Estimates of hydrological

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

variables such as water level and width are more routinely monitored on a global scale (Allen and Pavelsky, 2018) using altimeters and imagery respectively; therefore, many methods aim to utilise these variables to calculate discharge (Hao et al., 2024; Du et al., 2023; Durand et al., 2023; Gleason and Smith, 2014; Mengen et al., 2020). However, calculating discharge from water level or width is an ill-posed and unconstrained problem, and there is no closed solution to the relevant hydrodynamical equations. Therefore, any model to estimate discharge, whether physical or statistical, is always going to need additional prior data to converge on an accurate time series. Examples of this are prior discharge, river geomorphology such as width, slope or roughness coefficient, hydro-meteorological forcing data such as rainfall or potential evapotranspiration, and catchment characteristics such as drainage area (Du et al., 2023; Mengen et al., 2020; Munier and Decharme, 2022; Bahekar et al., 2021; Rakhymbek et al., 2025; Coxon et al., 2019). These can be used as initial conditions, constraints, or variables within the discharge model to simplify the equations and aid convergence to a realistic result. The current lack of consistency of this data, especially spatially, thus presents a serious barrier to producing accurate global estimates of discharge, as does the uncertainty and inconsistency of the measurements, since we rely on many prior assumptions that are variable and often at sub-optimal resolution. The availability and reliability of accurate discharge distributions, particularly in data-scarce regions, could be substantially improved by increasing hydrological observations, which Earth Observation (EO) has the potential to provide (Riggs et al., 2023; Gleason and Durand, 2020).

The Surface Water and Ocean Topography (SWOT) mission (Biancamaria et al., 2015), launched in December 2022, provides an opportunity for an alternative approach to producing discharge simulations, offering the potential to calibrate models with high resolution, high accuracy observations on an unprecedented scale. Providing water surface elevation (WSE) measurements for all rivers in the predefined river hydrography SWORD (SWOT River Database) (Altenau et al., 2023, 2021) exceeding 100 m (and in many cases 50 m) in width, with a repeat cycle of 21 days, SWOT allows the incorporation of WSE and river width observations into discharge models as, for example, forcing or prior information. SWOT not only complements existing gauges, but expands the spatial coverage of discharge observations uniformly across the globe. The assimilation of virtual gauging stations from altimetry data has become more prevalent with the increasing quantity and skill of satellite missions (Riggs et al., 2023; Gleason and Durand, 2020; Sahoo et al., 2020; Dijk et al., 2016). The motivation here is to use these observations to progress towards an improved method of river discharge estimation that can be applied to the majority of river reaches seen by SWOT, and hence a model that performs consistently across previously unmodelled and data scarce regions and highly monitored regions alike.

The SWOT Discharge Algorithm Working Group (DAWG) has proposed a collection of algorithms to calculate discharge from these observations (Durand et al., 2023; Andreadis et al., 2025). These include both Calibration (MOMMA) and Inverse Problem (MetroMan, geoBAM, HiVDI, SIC4DVar, SADS) algorithms at reach scale, though some are averaged from node scale outputs, as well as a basin scale Mean Optimization Integrator algorithm (MOI) which brings together all aforementioned algorithms into a physically realistic network with flow conservation at a larger scale. The calibration approach is a well-constrained problem but is limited by the accuracy of the calibration data; hence many ungauged reaches rely on estimates of bankfull discharge to estimate flow law parameters. The inverse-problem approach uses Bayesian principles to solve the under-constrained inverse problem where gauged data is unavailable or poor. This involves modelling uncertainty of all input

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

data and relies heavily on the physical flow laws but also the quality of the a priori information. Within the DAWG algorithms, both data assimilation and mass-conserved flow law inversion methods are applied to this problem. This is the first attempt to estimate discharge on such a large scale. However, the physical flow laws require many parametrisations and variables, such as width and slope from observations, of which the quality is less consistent and in turn reduces the number of locations at which the algorithms can run. All of these algorithms also require input priors, one of which is GRADES-hydroDL (Yang et al., 2025), a model-derived daily discharge database spanning 1980 to the present day. These can also be computationally expensive to run repeatedly as the temporal length of the observation record increases. An alternative approach using only the WSE observations with a statistical method could therefore complement these existing approaches, by providing an algorithm that is computationally cheaper and that produces discharge at a greater number of locations. With different strengths within a combination of these approaches, a global view of discharge could be understood on an unprecedented scale.

Here we present a statistical approach to modelling discharge, leveraging observations from SWOT to produce level-duration curves and integrating these with global flow-duration curves formed from GRADES-hydroDL (Yang et al., 2025), the same discharge dataset used as a prior for the DAWG algorithms. The union of these two curves defines a rating curve relating discharge to water level for each location. The idea of using rating curves to model discharge is well-established (Gleason and Durand, 2020; Paris et al., 2016; Nathanson et al., 2012), and this method can provide a baseline for expected performance compared to more complex algorithms. The high accuracy observations from SWOT could also produce a modelled discharge time series with a particular focus on the dynamics and timings. This would be a key improvement to current global discharge models, as analysis of the correlation between modelled and observed discharge suggests that their dynamics and hydrographs may lack the precision required for many uses (Senent-aparicio et al., 2021; Liu et al., 2011). The timing and rate of variability of flows is vital for a variety of applications, from sediment transportation and ecological conservation (Poff et al., 1997; Phillips et al., 2018), to understanding runoff and tributary interaction (Ternynck et al., 2016; Wendi et al., 2019; Pattison et al., 2014).

We do this analysis on the SWORD network at reach scale, and the result is a set of global virtual gauging stations with temporal resolution of the same order as the SWOT mission. In doing so we also create a simple method for computing rating curves calibrated to SWOT observations. We will assess whether the proposed reanalysis method can be used to create discharge time series with a greater accuracy than our input dataset GRADES-hydroDL, and therefore whether SWOT measurements can be used to calibrate existing discharge models to improve the overall accuracy. We will explore the necessity of the restriction to exclude higher flows, and the sensitivity of the results to the minimum exceedance probability value for which our model applies; we label this value p_{min}.

Ultimately, the reanalysis method we present in this paper provides a baseline for discharge algorithms utilising observations from the SWOT satellite and assesses whether these can be used to improve upon existing discharge models. The output is a global collection of virtual gauging stations which can be used to complement existing ground observations, particularly in data-scarce areas, and therefore help to improve our understanding of global hydraulics.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

2 Data

Conceptually, our model requires input data of water elevation levels and discharge values globally. In this section, we set out the datasets we have used in our model, as well as the gauge datasets we use for validation.

2.1 SWOT Data Product

The SWOT mission, launched on 16th December 2022, is a joint project from NASA, CNES, UKSA and CSA. The satellite observes the location and elevation of surface water on a global scale, at a resolution and accuracy that has never been achieved before by a single satellite (Biancamaria et al., 2015). It utilises radar interferometry techniques to measure water surface elevation (WSE), with twin Ka-band synthetic aperture radars separated by a 10 m baseline, an instrument named KaRIn (Ka-band Radar Interferometer). Ka band is the highest frequency that has been used on previous altimetry missions, the most common of which are the lower-frequency Ku-band and C-band (Grgić et al., 2021; Li et al., 2022). Ka band has minimal penetration and therefore increased backscatter intensity from water surfaces, enabling their identification more easily. The use of the short wavelengths also allows for higher spatial resolution than longer wavelengths. The Wide-Swath Altimeter results in 120 km swaths in which SWOT can observe water levels, making it stand out from previous missions, and the 20 km nadir gap is partially covered by an additional Nadir Altimeter taking observations along the path of the satellite. SWOT pixel cloud observations have a resolution of roughly 15–25 m cross-track and 5–10 m along-track, covering over 90% of the Earth's surface between $\pm 78°$ latitude on a 21-day repeat cycle (Desai, 2018). For hydrological purposes, this means that all rivers larger than 100 m wide, and many down to 50 m wide, are seen by the satellite with a mean elevation measurement error of 0.15 m globally, with significant improvements over this value in certain regions (Yu et al., 2024); an unprecedented scale and precision of global observations.

We use the Level 2 KaRIn High Rate River Single Pass Vector Product (L2_HR_RiverSP) (SWOT, 2024), which aggregates the raw pixel cloud observations to the SWORD hydrography. SWORD (Altenau et al., 2023, 2021) is a topographical network produced specifically for the SWOT mission, based on optical remote sensing of permanent water from GRWL (Global River Widths from Landsat) (Allen and Pavelsky, 2018), and defines river nodes (spaced at roughly 200 m intervals) and reaches (a group of nodes with an average length of 10 km) over all global rivers of width no less than 30 m. We use Version C SWOT observations, which is based on SWORD Version 16. Various hydrological attributes are included, such as water surface elevation (WSE), water surface slope (for reaches only) and river width, calculated from the observed resolution L2_HR_PIXC product.

All attributes are also assigned quality flags to indicate conditions that may have impacted the observations, either from characteristics of the radar return or from adjustments in the processing chain. These are an internal indicator only and are not a representation against external auxiliary datasets. The most basic level of filtering summarises the overall quality of an observation, marking it as good, suspect, degraded or bad. There are also more expert quality flags which detail the causes of possible impacts, as well as markers of other potential influences, such as dark water fraction, ice cover and, for reaches, the fraction of nodes with valid observations. We apply the so-called Permissive Relaxed configuration for quality filtering, taken

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

from the Confluence Prediagnostics module (SWOT-Confluence) used to run the DAWG algorithms, the full specifications of which, for reach-level observations, can be found in Table 1. It includes, but is not limited to, a prior width of at least 60 m and length at least 5 km, a maximum of 60% dark water, and a minimum of 40% of nodes contributing valid WSE levels, as well as outlier filtering via the Tukey method.

Table 1. Permissive Relaxed configuration filters for quality filtering reach-level SWOT WSE observations.

Variable	SWOT Attribute Name	Filtering Type	Value
Cross track distance	xtrk_dist	Min	10,000m
Cross track distance	xtrk_dist	Max	60,000m
Climatological ice cover	ice_clim_f	Max	1
Fractional area of dark water	dark_frac	Max	0.6
Quality of cross-over calibration	xovr_cal_q	Max	1
Prior reach width	p_width	Min	60m
Prior reach length	p_length	Min	5,000m
Fraction of nodes with a valid WSE	obs_frac_n	Min	0.4
Bitwise quality indicator	reach_q_b	Bitwise flags to exclude	23 26 27 28 29 (507510784 binary)
Tukey number	wse	Outlier values to NA	1.5

The data are accessed through Hyrdocron (v1.6.4) (Greguska et al., 2024). On average, each reach is observed by the satellite 2-3 times in each 21-day cycle (though this ranges between 1 and 7 according to latitude and satellite path). From the start of the Science Orbit in July 2023, this would average at roughly 100-150 observations per reach over the mission lifetime of 3 years prior to filtering.

We consider the time period between 21st July 2023 and 6th May 2025, the full span of the SWOT Science Phase with data collection under Version C. Of the 158,942 river reaches in the SWORD database, 156,255 have WSE observations from SWOT, all of which still have observations when filtered by the bitwise flag in Table 1, which is the equivalent of removing all observations labelled as "bad" by the mission. These have an average of 78 observations, with a maximum of 578 at the highest latitudes. Applying all of the remaining filters in Table 1, there are only 93,815 reaches remaining with observations: 59% of defined reaches. It is worth noting that only 97,662 reaches satisfy the prior width and length conditions, and therefore this is the maximum number of reaches that could remain after applying this filtering method, since these quantities are constant and pre-defined by the hydrography.

For these 93,815 reaches with filtered observations, an average of 45% of observations per reach are kept by this filtering method, which results in a reach average of 32 observations, with a maximum of 127.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

2.2 GRADES-hydroDL

GRADES-hyrdoDL (Version 2.0) (Yang et al., 2025), hereafter referred to as GRADES, is a model-derived dataset of daily discharge values spanning 1980 to the present day. This uses an LSTM (Long Short-Term Memory) model trained on discharge observations from a selection of basins, applied to global 0.25 degree grids, using a network forced by sets of year-long daily meteorological features as well as fixed basin attributes. It is calculated for MERIT-Hydro reaches but has recently been mapped to SWORD using the MERIT-SWORD data product (Wade et al., 2025, 2023). Of the 158,942 river reaches in SWORD, 156,823 have a corresponding GRADES time series (98.7%). We use this dataset for our daily discharge time series to create flow-duration curves (Sect. 3.3). Any model could be used for this distribution, and indeed an ensemble could be a possible progression of this step. We chose this particular model because of its long record, expansive global coverage, and the widely used and well-reviewed skill of the techniques used (Liu et al., 2024; Feng et al., 2021, 2020; Shen, 2018; Shen et al., 2023; Liu et al., 2022).

2.3 Reaches of Application

Of the 93,815 river reaches defined by SWORD with filtered SWOT WSE observations, 92,955 have GRADES data associated with them, all which fit positive parameters for the level-duration curve (Sect. 3.2) and therefore satisfy the a priori conditions of our model. This is 58% of all global SWORD river reaches but importantly is over 95% of reaches which satisfy the topographical filters (minimum prior reach width and length) described in Sect. 2.1, a chosen requirement for any discharge algorithm utilising SWOT WSE levels with the Permissive Relaxed Configuration method of filtering.

2.4 Gauge Observations

Validation of this model is done by comparison to in-situ gauging stations which record both discharge and water level values that coincide with the collected SWOT observations, in the time interval between 21st July 2023, the beginning of SWOT's Science Phase Orbit, and 6th May 2025, the final date of data collection under Version C. These are sourced from the US Geological Survey (USGS), the National River Flow Archive (NRFA) from the UK Centre for Ecology and Hydrology (UKCEH), the Australian Bureau of Meteorology (BoM), Eau France (EAU), Environment and Climate Change Canada (ECC) and the Brazilian National Water Agency (ANA). We use a variety of techniques to collect these observations, using both inbuilt APIs (NFRA, EAU, ECCC, ANA) and the prebuilt python packages dataretrieval (USGS) and pybomwater (BOM). For each gauge, we find the nearest SWORD node (within 0.01°radius) and link it to the associated SWORD reach. Where multiple gauges have been linked to the same reach, we choose the gauge with the longest time series of flow observations within our observation window. We then supplement these gauges first with data pulled from the R package RivRetrieve, which collects discharge and water level observations from multiple of the agencies listed above, and then those from the SWOT Validation Set for Global River Discharge Evaluation (SVS) (Saemian et al., 2026) at gauges for which water level data could also be sourced using the aforementioned techniques, since this dataset only provides flow values.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Overall, this results in gauge observations at 1,928 unique river reaches. Of these, 119 are linked to gauges on which GRADES is trained, which we therefore remove to avoid any bias this may cause in the results. Finally, to ensure that there are sufficient data points for statistical tests while retaining as many gauges as possible, we discard any gauges where there are less than 10 intersecting gauge and SWOT/modelled discharge observations.

This results in a final set of 941 reaches for which we have both model results and gauge observations to validate against. Of course, this limits our evaluation to certain climates and regions, largely focused on North America and Europe, but this is due to gauge availability and their uneven distribution.

2.4.1 Chosen Gauges

There is uncertainty involved in measuring water level and discharge at gauging stations (de Oliveira and Vrugt, 2022; Levin et al., 2023), as well as in our method of mapping gauges to SWORD reaches. Because of this, we filter the gauges by monotonicity of water level observations to the SWOT WSE values. The aim of this is to remove reaches where either the gauged or SWOT observations are dominated by outliers or errors, or where a gauge may have been mapped to the wrong river reach. We only apply low thresholds to retain as many gauges as possible. This adds an additional, and significant, filter to the quality of the SWOT observations used, in addition to the initial configuration detailed in 1. We take our collected gauge observations for the 1,600 river reaches for which both filtered SWOT observations (Sect. 2.1) and GRADES data (Sect. 2.2) also exist, and applied the following filters:

$$R_S > 0.8$$

$$\tau_k > 0.7$$

for measures of monotonicity R_S the Spearman R value, and τ_k the Kendall Tau value. This leaves us with 399 reaches, shown in Fig. 1, at which to complete our analysis. Overall results using all 941 gauges can be seen in Appendix Sect. D. In summary, this section emphasises the need for quality filtering the SWOT observations, and the improvements still needed in this area to ensure consistency on a global scale.

3 Methodology

With the datasets described in Sect. 2, the following method is applied independently for each river reach in the SWORD hydrography. We produce two curves, level-duration and flow-duration, per reach, which are then combined via their shared variable, probability of exceedance. We name this method SWOT Discharge Reanalysis, henceforth referred to as SWOT-DR. This section explains our proposed method: how we calculate each of these curves in Sects. 3.2 and 3.3 respectively, and then how they are fitted together in Sect. 3.4. A flowchart explaining the process can be seen in Fig. 2.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

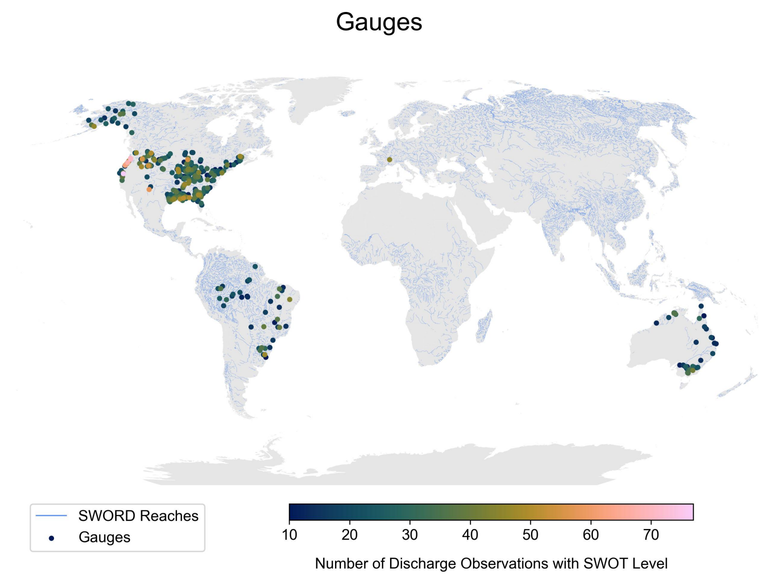

Figure 1. Location of the reaches used for analysis, sized and coloured by n, the number of observations to evaluate.

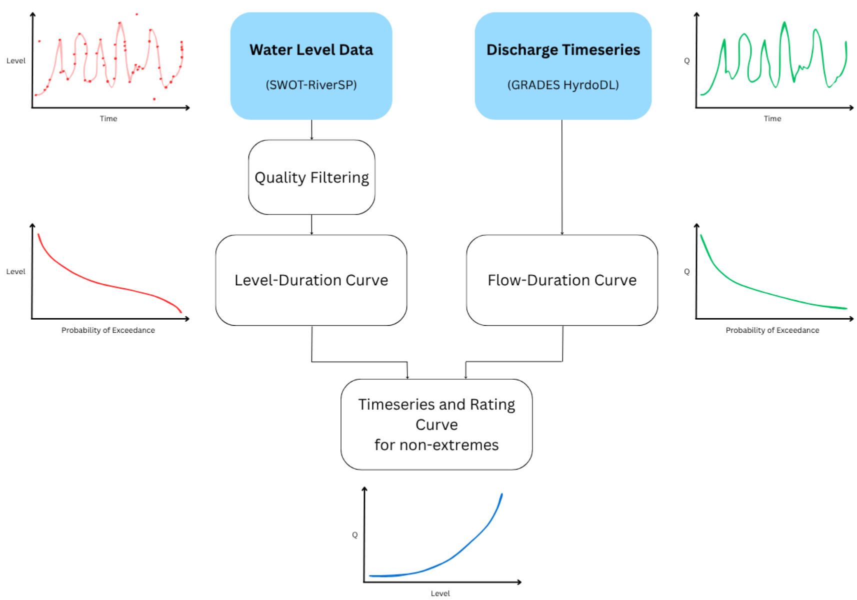

Figure 2. SWOT-DR Methodology Flow Chart.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

3.1 Minimum Exceedance Probability

We propose SWOT-DR to be used for non-extreme events only, due to short time series of the input SWOT WSE observations, and its relative temporal sparsity compared to the often daily resolution of gauges, which could affect its ability to capture extreme flows over the rarest events. The nature of extreme events means that they are poorly modelled by statistical methods, and the short record of SWOT emphasises this effect.

We investigated both the necessity and impact of this restriction by performing sensitivity analysis on the minimum probability exceedance we choose to model, which we will label p_{min}, for which we ran SWOT-DR at 21 different values equally spaced in the interval $p_{min} \in [0, 0.2]$, chosen to represent possible extreme event ranges, at each validation reach. We find that increasing p_{min} has little effect on the lowest 90% of the SWOT-DR discharge values, but the higher end of the discharge distribution has improved accuracy. However, it also means that this discharge estimation method can only be applied to a smaller proportion of observations, so these need to be counterbalanced. For our purposes, we have chosen $p_{min} = 0.05$, and further details of this analysis can be found in Appendix Sect. A.

3.2 Level-Duration Curves

The probability of exceedance of WSE is calculated empirically from observations. That is, for a time series of WSE observations W, each value w has quantile p such that

$$p(w) = P(W \geq w) \tag{1}$$

We then discard any values where $p < p_{min}$.

The accuracy of SWOT-DR should be improved by a longer record of WSE levels, which is shown in Appendix Sect. B, and indeed this may impact the choice of p_{min}, discussed in Sect. 3.1, as might a study focused on a smaller region.

3.3 Flow-Duration Curves

Based on the approach of Quimpo et al. (1983), we obtain a frequency distribution of discharge by fitting an exponential distribution to a discharge time series. That is, we assume that the discharge q has the form

$$q(p) = \lambda e^{(-\sigma p)} \tag{2}$$

as a function of the probability of exceedance p for undefined, positive parameters λ, σ, which are estimated via regression from the aforementioned time series. We utilise all values in the time series between the 0.01 and $1 - p_{min}$ fractional percentiles to fit these parameters; Eq. (2) is not applicable for extremes, and SWOT-DR will only be used to model exceedance probabilities of $p < p_{min}$, so this curve therefore more accurately represents our discharge distribution.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

3.4 Rating Curves

For a recorded observation of WSE, we calculate its probability of exceedance from the relevant level-duration curve, from which we calculate the associated discharge value from the flow-duration curve. This two-part method allows us to enforce a positive correlation between the two variables, WSE and discharge, in a single rating curve for each reach.

We considered the possible advantages of matching the time periods of these two datasets before applying the steps above for consistency of event occurrence. We performed sensitivity analysis on the input discharge dataset record length by calculating parameters λ, σ for 50 different, randomly chosen discharge time series of SWOT length and evaluating the results of SWOT-DR for each at the validation gauges. Our tests showed a large parameter variation and therefore the discharge time series are heavily dependent on this choice. The SWOT time period unsurprisingly produced the most accurate discharge when compared to gauge observations within the same period, but results from using the full GRADES time series before SWOT launch are comparable, and show skill towards the upper end of those seen from the SWOT-length inputs. The median Nash-Sutcliffe Efficiency (NSE) is 0.28 when using the full GRADES time series before the SWOT Science Orbit as opposed to the SWOT-overlap time period only, whereas our random samples of SWOT-length time periods have a median NSE of 0.11. We therefore decided it would be best to use the whole GRADES record for the use of SWOT-DR to new SWOT values, in order to represent events larger than may have been seen since the SWOT Science Orbit. Further details of this can be found in Appendix Sect. C.

4 Results

We use a variety of metrics to assess model performance of both GRADES and SWOT-DR. In this paper, we analyse correlation using the Pearson R (R), Spearman R (R_S) and R^2 metrics, to assess dynamics and temporal patterns. To evaluate bias and errors, we look at quantitative bias ($Bias$) and percent bias ($pBias$) as a ratio, with particular focus on their absolute values, as well as root mean square error ($RMSE$) and normalised $RMSE$ ($nRMSE$). Finally, we calculate the NSE, log-natural NSE ($LNSE$), normalised NSE ($nNSE$) and Kling–Gupta efficiency (KGE), which assess overall skill, influenced by both of the previously named factors. We focus in particular on R, $pBias$, $nRMSE$ and NSE, as these are the metrics commonly used to assess discharge in literature (Kalin et al., 2010; Brighenti et al., 2019; Jimeno-Sáez et al., 2018; Gupta et al., 1999).

For many of these validation measures, we look at the distribution of metrics over the chosen 399 evaluation reaches (Sect. 2.4.1). However, it can be helpful to visualise time series and distribution comparisons of WSE and discharge at individual locations, for which we use the same 8 reaches in each case (Figs. 4, 6, 8, 10). These example sites were picked at random, but they cover a wide range of the distribution of each validation metric, seen in Fig. 7, and the reach id, location and properties of each of these reaches can be found in Table 2.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Table 2. Location and physical characteristics of the example SWORD v16 river reaches shown in this paper.

Label	Reach ID	Lon	Lat	River Name	Length (m)	Width (m)	Slope	Max Flow Acc. (km^2)
a	21602600141	5.16	45.79	Canal de Miribel	7915.29	151	0.57	20140.17
b	56299500141	142.35	-11.15	Jardine River	19017.56	63	0.39	1562.24
c	74266700171	-87.37	39.79	Wabash River	18416.23	108	0.08	29787.04
d	74270800061	-90.73	38.47	Meramec River	11654.35	63	0.55	7945.28
e	75165000271	-96.82	31.13	Brazos River	11355.15	77	0.26	90763.85
f	77480100021	-123.98	41.51	Klamath River	10238.10	150	1.07	33818.23
g	78210000081	-122.91	46.27	Cowlitz River	9503.92	123	0.58	5925.79
h	81130400021	-151.08	60.48	Kenai River	18859.66	124	0.74	4589.82

4.1 WSE Time Series

For evaluation purposes, we compare only relative, rather than absolute, water levels due to datum differences in the measurements. We shift the gauge water level observations by a constant to enforce an equal median water level to the SWOT observations at each reach. Summary metrics over the reaches with gauged water level are shown in Fig. 3, and individual reach plots are shown in Fig. 4.

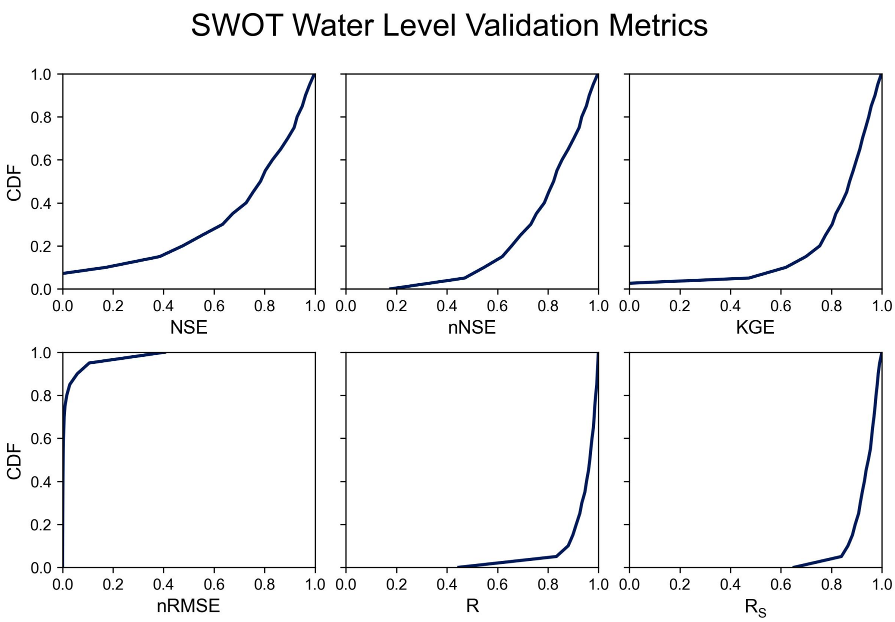

Figure 3. Summary metrics showing the cumulative distribution function of 6 metrics. The metrics are obtained from comparing relative timestamped SWOT WSE levels to gauged water levels (after shifting the gauge observations to equate the median values).

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

89% of reaches have a $RMSE$ of less than 1 m, and 94% have an $nRMSE$ of less than 0.1. 85% of R values are greater than 0.9 with 62% over 0.95, and 78% of NSE values exceed 0.5. Due to varying datums, and the fact that we only use water level observations for quantiles, evaluating the bias of SWOT WSE levels is beyond the scope of this paper, but many studies are focusing on this and the effect it has in other situations. Patidar et al. (2025) observe a relative error of 0.18 m in node WSE measurements for rivers wider than 100 m across India, and Neal et al. (2025) find an agreement of 0.16 m on the 40-70 m wide River Severn in the UK, demonstrating the high levels of accuracy in SWOT measurements even on rivers smaller than mission's specifications.

Figure 4 shows the high levels of accuracy in the dynamics and range of water levels. Only shifted by a constant to fix the median values, these time series plots show that the relative differences in SWOT water levels very closely follow the gauges observations; the maximum $|pBias|$ of these 8 reaches is 0.03 (3%) at reach (b) while the other 7 are all below 0.0063 (0.063%) and the minimum R_S value is 0.65.

4.2 Flow-Duration Curves

To evaluate the discharge distributions defined by our flow-duration curves in Sect. 3.3, we compare modelled discharge from SWOT-DR and GRADES to gauge observations at probability of exceedance between $p = p_{min} = 0.05$ and $p = 1$ at intervals of 0.01. Summary statistics are shown in Fig. 5 using cumulative frequency curves, and examples at individual sites can be seen in Fig. 6.

Our hypothesis is that our parameterised distributions should be of similar skill to GRADES, by all metrics, for the range of interest of exceedance probability $p \geq p_{min}$. However, we do not expect to see any significant improvement since our discharge distributions are fitted to the GRADES time series'. Our chosen distribution, defined by Quimpo et al. (1983), is well reviewed in literature but we investigate here how well it applies to the GRADES dataset. The shape of our chosen exponential distribution should result in a smoother curve, though the differences should be minimal, in the event of a good fit.

Figure 5 supports our hypothesis, showing only minor difference between the performance of the two distributions tested against gauged observations. Both distributions show a good representation of the gauged discharge for the range of interest; for SWOT-DR, 100% of R values are greater than 0.63 with 73% greater than 0.95, and 82% of NSE values are positive with 67% greater than 0.5.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

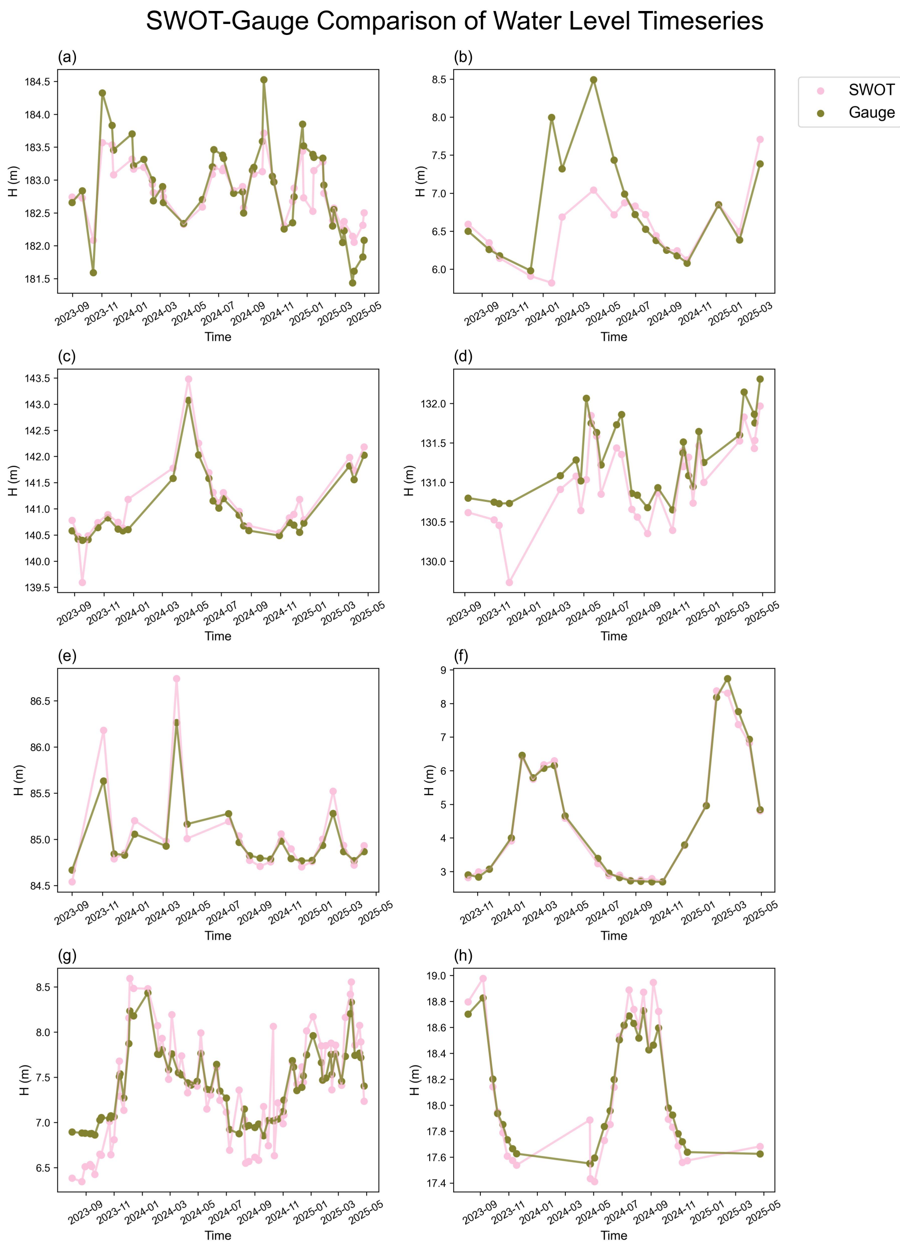

Figure 4. Relative water level time series for the Gauge (blue) and SWOT (pink) observations.

It is also worth noting that the $|pBias|$ is very similar; the mean values are 3.03 and 3.04 for SWOT-DR and GRADES respectively, with median values 0.23 and 0.20. This suggests that this parametrisation does not introduce significant bias in the discharge distribution, as we would hope for in the case of a good model fit.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

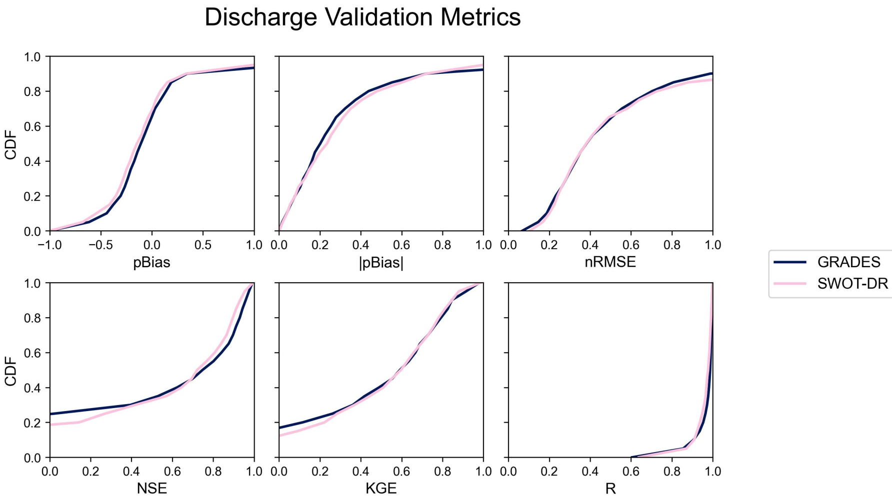

Figure 5. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculating by comparing modelled to gauged discharge distributions as described in Sect. 4.2.

We can see in Fig. 6 that the fitted exponential curves decrease the flow rates at the extremes. This is expected, as the significantly larger values at the extremes are very sparse within the 20 years of discharge data that we fit to, exemplifying our reason to parameterise. The higher values, unlikely to have been seen by SWOT due to its short record, can give unexpected results when combined with the level-duration curves. The parameterised form allows for the possibility of these higher values, unlike simply truncating the GRADES time series data, but strongly favours the more common flow conditions. For these higher probability flows, we can see that Eq. (2) is a good fit, justifying our use of this form.

4.3 Discharge Time Series

Here we compare the discharge time series of each model against the gauged discharge observations. This is where we see the biggest disparity between SWOT-DR and GRADES. As previously, we show both summary statistics (Fig. 7) and examples at individual sites (Fig. 8).

Figure 7 shows that SWOT-DR generates a good representation of gauged observations; at 64% of gauges SWOT-DR produces a positive NSE value, compared to only 52% for GRADES, and the 75th quantile of the NSE values is 0.67 for SWOT-DR compared to only 0.48 for GRADES. By formulation of SWOT-DR, all R values are predictably high, with 66% above 0.9, while only 20% of GRADES R values reach this mark. Similarly, 40% of reaches give a $nRMSE$ of less than 0.5 with SWOT-DR but only 23% for GRADES. These three metrics in particular show the improvement on the discharge time series compared to GRADES in the dynamics of the flow.

Further comparison of metrics shows that the median $RMSE$ of GRADES is $36.3\,\mathrm{m^3\,s^{-1}}$ larger than that of SWOT-DR, the median $nRMSE$ 0.13 larger, and the median NSE 0.30 smaller. This shows the benefits SWOT-DR can have on existing

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

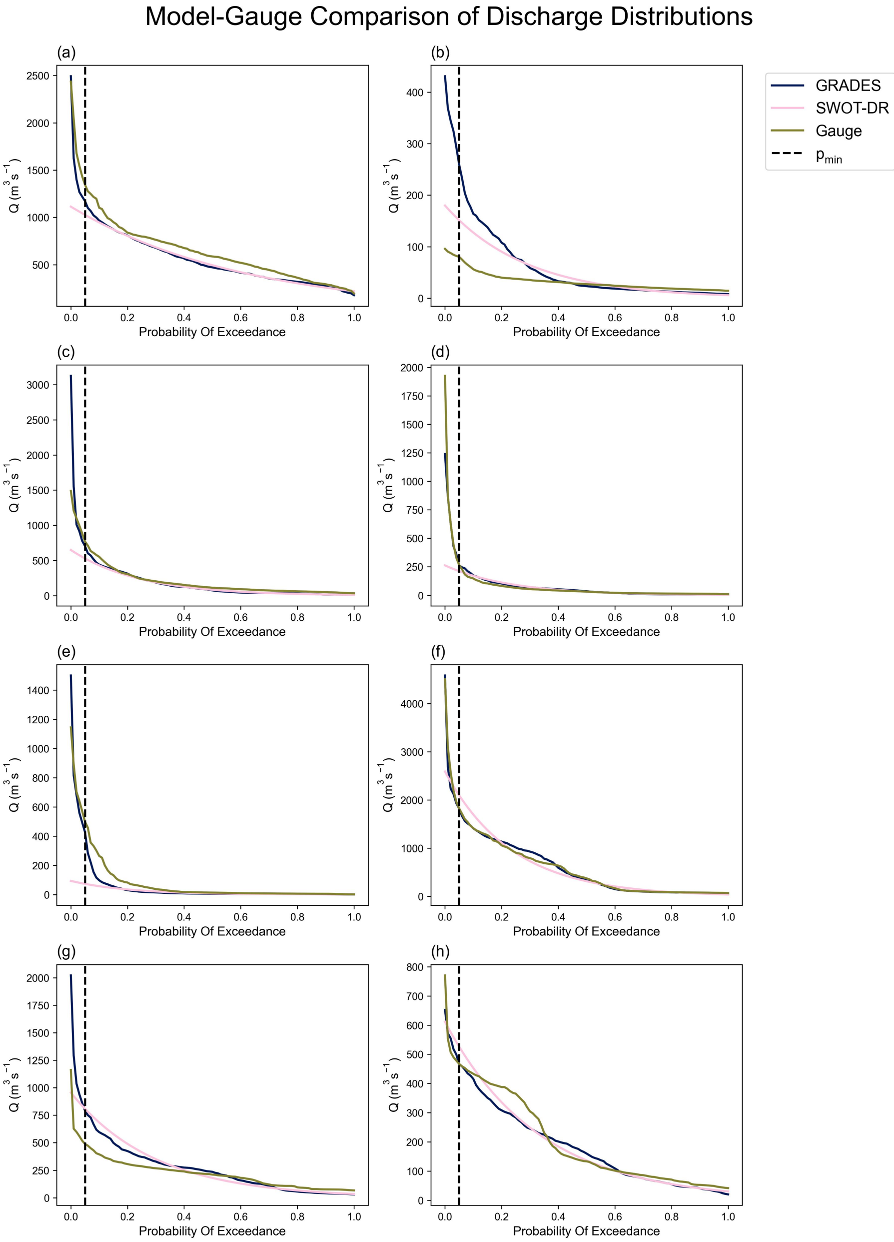

Figure 6. Discharge distributions for the Gauge (blue), GRADES (pink) and SWOT-DR (green) datasets. The dotted black line at $p = 0.05$ represents the minimum exceedance probability p_{min} we model with SWOT-DR.

discharge estimation methods. The bias is the area where GRADES excels, with a median $|pBias|$ of of 0.22 compared to 0.32 for SWOT-DR. Figure 5 showed that our flow duration curves have bias equivalent to that of GRADES, suggesting that this increased bias comes predominantly from the distribution of water level values and how they are matched to probability of

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

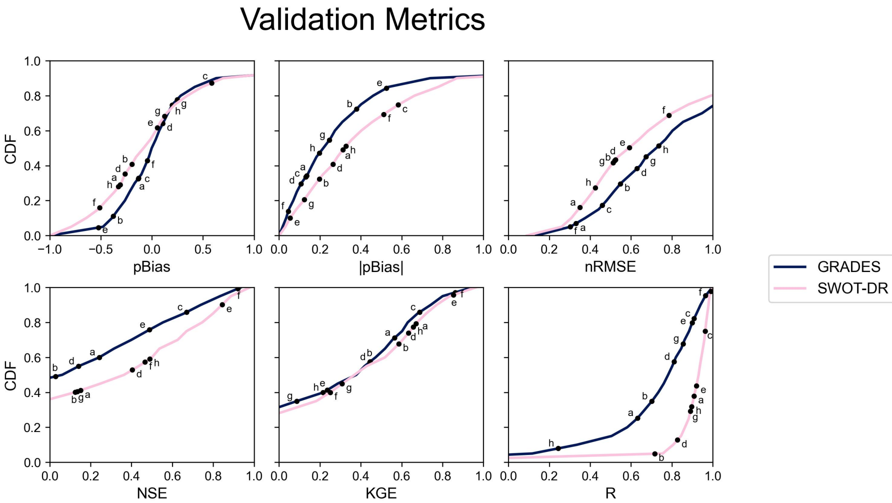

Figure 7. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series. The location of each individually plotted reach in Figs. 4, 6, 8 and 10 are marked on each plot.

exceedance. We theorise that this will therefore improve with an extended SWOT observation record, and explore this is Sect. B.

On examination of the difference in these metric values for each site between the SWOT-DR and GRADES flow values, we see that we make an overall improvement to R, R_S, NSE, $nNSE$, and $nRMSE$, while GRADES results in a lower $|pBias|$. These can be seen specifically in Fig. 9, where the plotted data values are $m_{SWOT-DR} - m_{GRADES}$ for each metric m; hence the first two rows of metrics are positive where we have made an improvement, while for the bottom row an improvement is shown by a negative value. We can see that the R value is improved by SWOT-DR over GRADES at 85% of reaches and the NSE at 64%, while the $|pBias|$ only at 37%.

We looked for spatial or reach attribute patterns in the results to look for possible correlation and causation, but the maximum absolute Spearman R correlation between the 6 metrics displayed in Fig. 9 and the attributes provided in SWORD - including, but not limited to, upstream accumulation, width and slope - is 0.25, indicating no clear links between any of these factors. A table of these values can be found in Supplementary Information Sect. S3 for interest.

Due to SWOT's greater global coverage over any previous satellite based dataset, and therefore in particular the set of gauges used to train the GRADES model, as well as its high levels of accuracy, the shape of our time series follows that of the gauged discharge more accurately. This is shown both in individual locations in Fig. 8 and the summary metrics in Fig. 7, especially in the R value metric, a measure of correlation between the two datasets. While GRADES is trained on available data, details of which can be found at Yang and Pan (2025), the bulk of the input data is precipitation and other meteorological fields, and has a much lower resolution than SWOT. There has not been a dataset of gauged discharge values, physical or virtual, with such a wide and even global coverage as SWOT available until now, and hence many reaches will not have had discharge observations to be trained on explicitly. While the Long Short-Term Memory (LSTM) model used to create GRADES provides a good

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

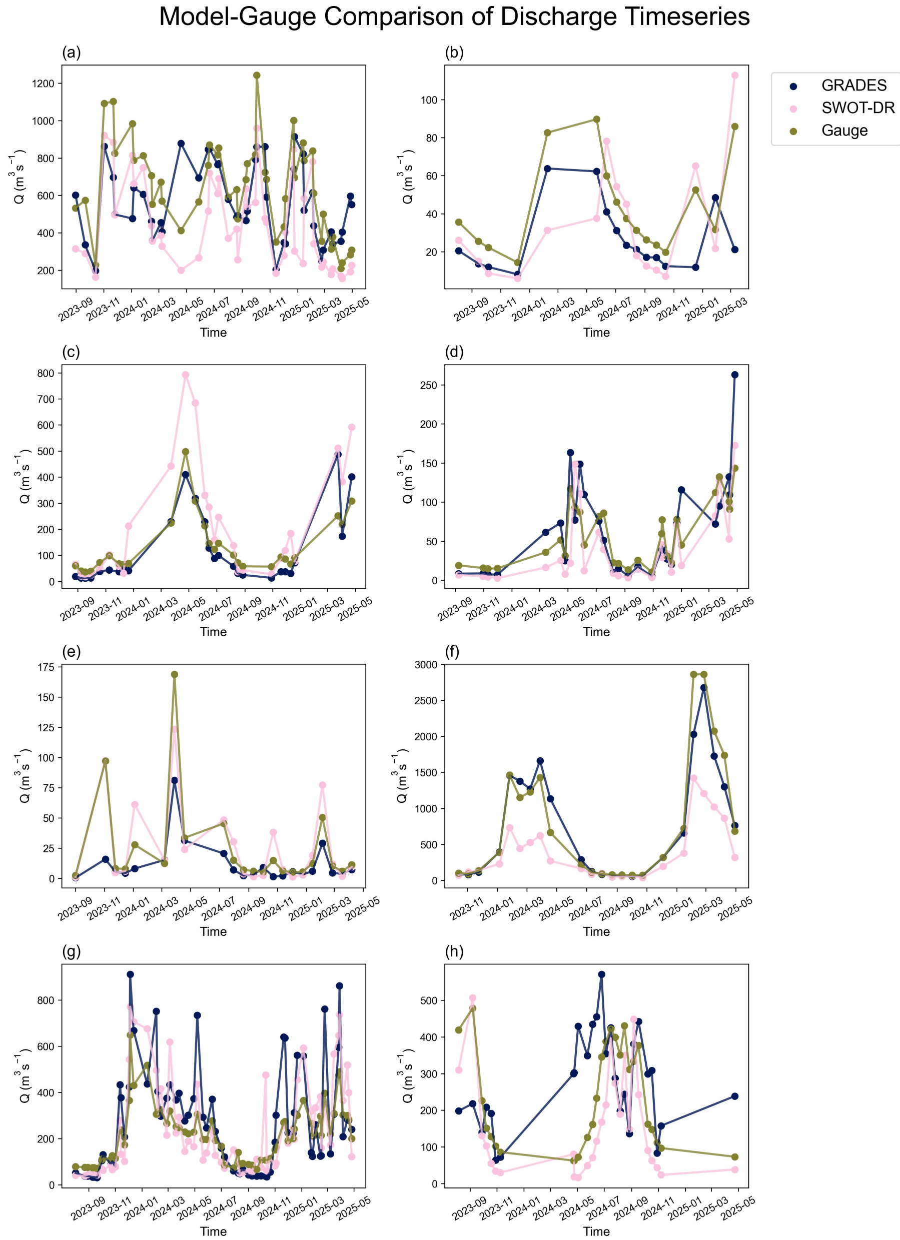

Figure 8. Discharge time series for the Gauge (blue), GRADES (pink) and SWOT-DR (green) datasets. These are only plotted on dates for which there was a valid SWOT WSE observation.

335 representation of the flow-duration curve at many reaches, the details of the dynamics, exact timing and shape of the flow are sometimes lost. By incorporating the SWOT water levels, we retain this level of detail and the subtleties of the dynamics of the discharge in time. However, it must be noted that it is not suggested to use SWOT-DR to replace GRADES, as its temporal

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

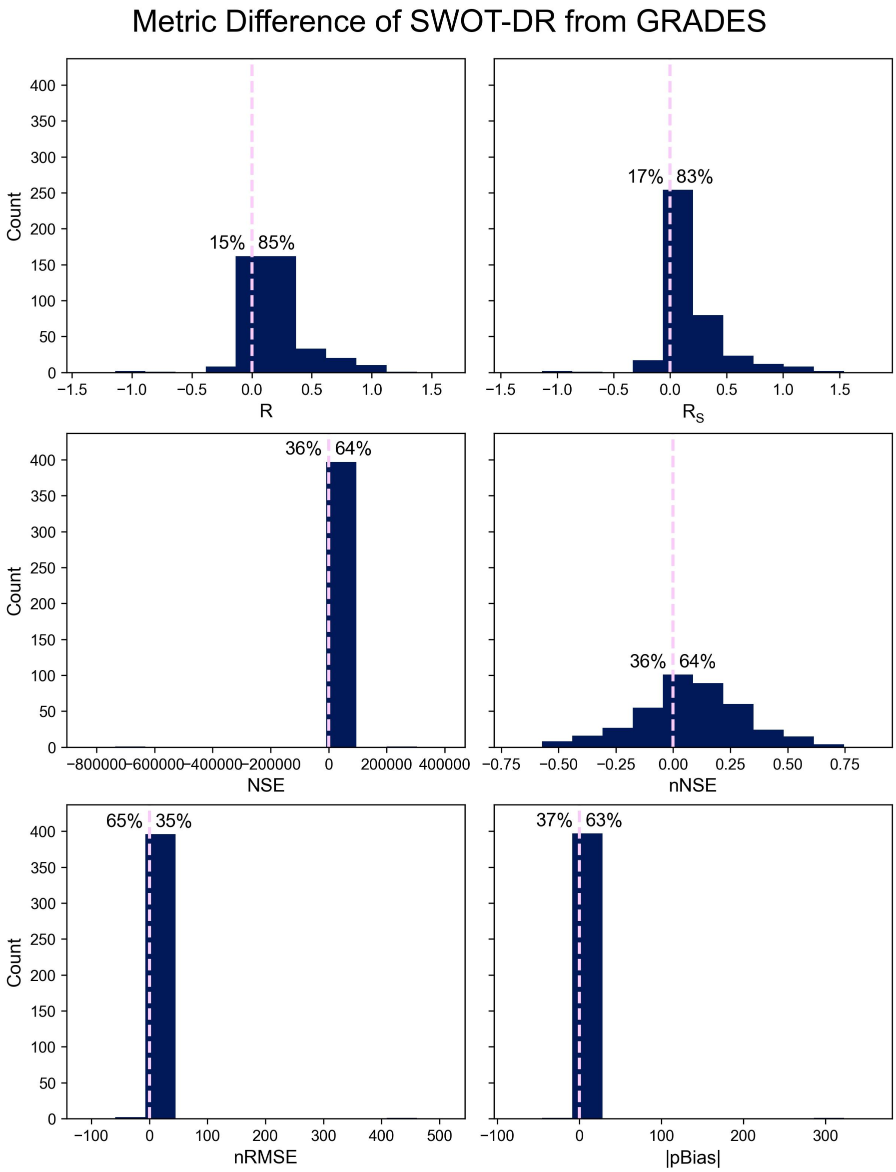

Figure 9. Histograms showing the difference between SWOT-DR and GRADES performance in a selection of metrics. The percentage of positive and negative values are shown in each instance.

resolution is much lower. Disaggregating SWOT observations to daily resolution is an active research field (Ke et al., 2025), but here we aim to create virtual gauging stations with SWOT's temporal resolution that have greater accuracy than using
340 timestamped GRADES values. For daily datasets, GRADES may still be a preferable dataset to use.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

4.4 Rating Curves

Due to the irregularity in the availability of gauged water level, and the variety of datums used in these datasets, the water level here is defined by SWOT for all curves, matched to gauged discharge observations by date. The rating curves for our example sites can be seen in Fig. 10. The overall trends, represented by exponential curves fitted to the data points, are good representations of the gauge observations in many locations; however, SWOT-DR provides a more consistent set of results when compared to GRADES. Even when the exponential curve applied to GRADES proves to be a good fit to the gauge observations, the individual points are a lot more scattered. Opposite biases often cancel out in summary metrics, resulting in unrepresentatively good results in Fig. 7, but looking at individual sites, as in Figs. 8 and 10, we can see the disparity. In contrast, due to the nature of SWOT-DR, the parameterised curves not only have a positive correlation, as we would expect in the majority of situations, but also prove to be a close representation of the gauge observations values both in dynamics and magnitude.

We can see in Fig. 10 that while the range of flow values is largely controlled by that of the GRADES distribution as expected, the correlation between water level and flow is more accurate in SWOT-DR. Reaches (a) and (h) are perfect examples of this; in the GRADES data, there is no clear trend at all between these variables. However, the accuracy of SWOT WSE levels has improved the dynamics, which has therefore reduced many of these errors.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

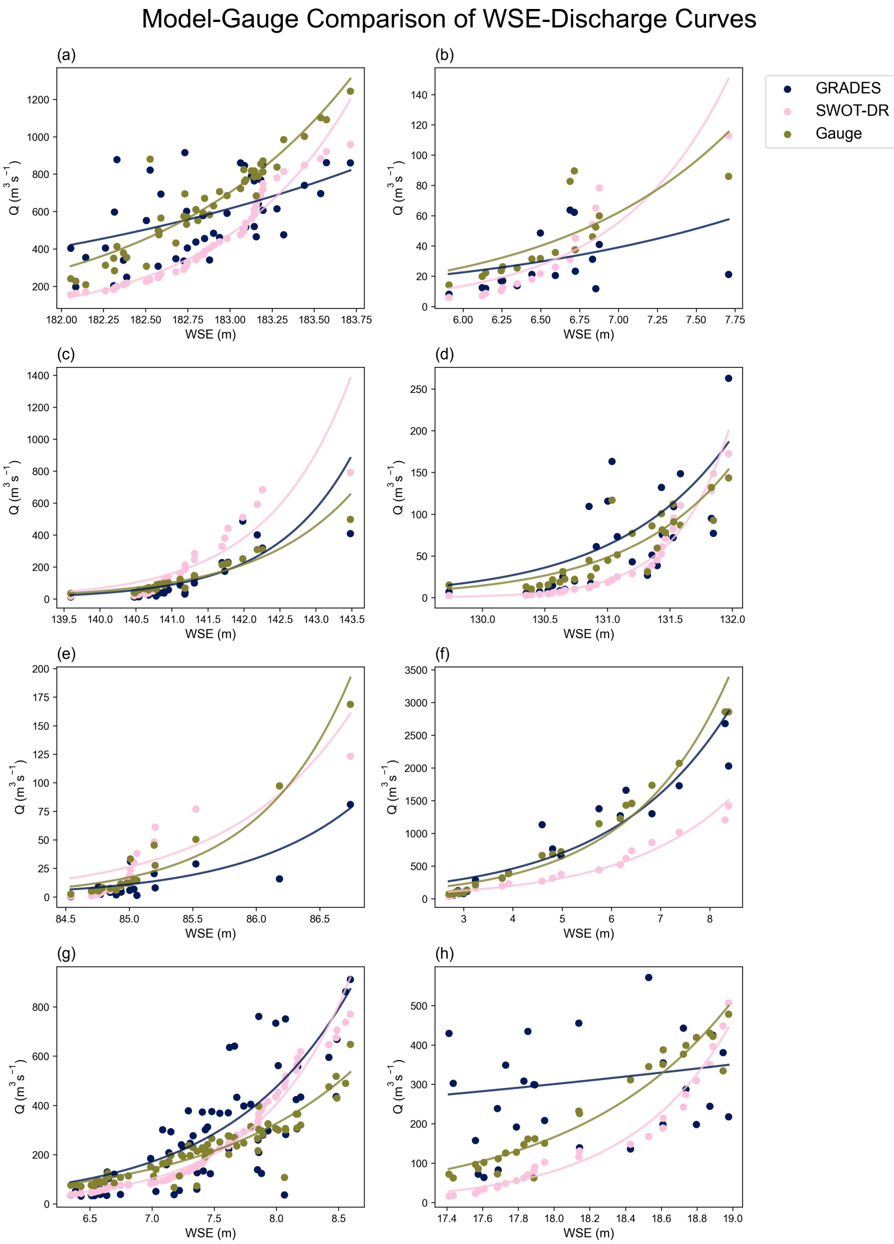

Figure 10. Rating curves for the Gauge (blue), GRADES (pink) and SWOT-DR (green) datasets. Individual data values are shown by points, while a fitted exponential curve is plotted for each method to show the overall trend.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

4.5 Performance Ratings

Performance ratings of streamflow at instantaneous time scales were proposed by Kalin et al. (2010), and can be used to classify model performance into four categories: unsatisfactory, satisfactory, good and very good, based on each of the NSE and $|pBias|$ metrics, and an overall classification min_class that is the minimum category of the two. We apply this acceptancy classification system to our outputs, and the results can be seen in Fig. 11. 213 of the 399 reaches (53%) are satisfactory or above by the overall classification for SWOT-DR, whilst for GRADES only 143 reaches (36%) meet these criteria. Many more reaches are at least satisfactory by the $|pBias|$ criterion; 331 (83%) in SWOT-DR and 357 (89%) in GRADES. The $|pBias|$ category is equal to or higher than that of the NSE category in 379 and 393 out of 399 reaches for SWOT-DR and GRADES respectively. This suggests that NSE is a more limiting factor in these models, and therefore while we do introduce bias, it is less significant than the changes we make to the NSE. That we make a larger improvement to the NSE is therefore a good indication that the implementation of SWOT observations are increasing the overall skill of the resultant model. A deeper analysis of these findings can be found in Supplementary Information Sect. S1.

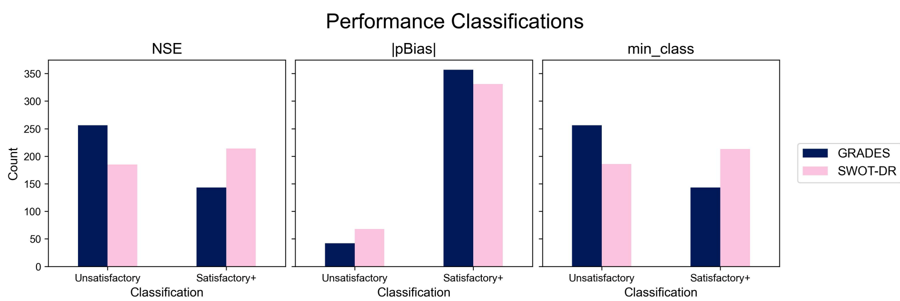

Figure 11. Distribution of binary performance rating classifications of 399 validation reaches, for each metric.

4.6 Correlation of GRADES Time Series

Gleason et al. (In Review) show that the correlation between GRADES discharge and observed water level, and therefore the dynamical skill of GRADES, is variable across different reaches. We investigate whether this effects SWOT-DR's performance, with particular focus on the difference in skill between SWOT-DR and GRADES. This is important, because a model that makes improvements in locations where GRADES has lowest accuracy would be most beneficial to expanding existing datasets and our understanding of discharge.

For each of our validation reaches, we calculated the Spearman R value between SWOT WSE levels and GRADES discharge, as per Gleason et al. (In Review), which we label R_G. We split the reaches into 5 groups defined by the R_G quintiles, with boundaries therefore at -0.72, 0.61, 0.75, 0.81, 0.88 and 0.98, and summarise group performance. The aim here is to see whether there is any correlation between R_G and the accuracy of SWOT-DR.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

The most notable impact is that on the performance ratings, as in Sect. 4.5, focusing on the overall classification min_class. This is shown in Fig. 12. Predictably, both GRADES and SWOT-DR mostly perform better at higher R_G, with SWOT-DR doing comparatively better for lower R_G. This shows that SWOT-DR, and the integration of SWOT observations, is of particular importance in order to develop a global discharge dataset that performs consistently spatially. A deeper analysis can be found in Supplementary Information Sect. S2.

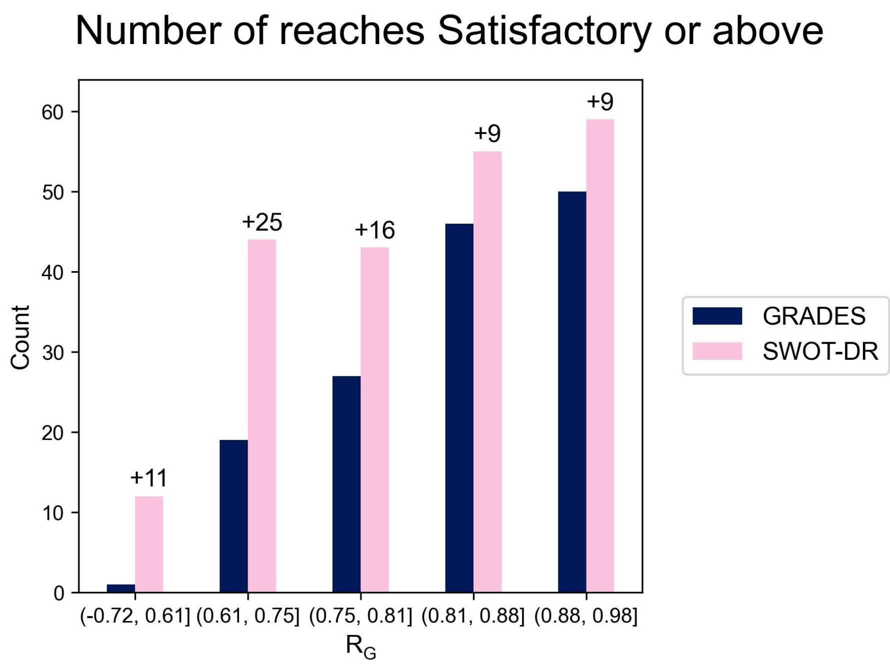

Figure 12. Distribution of performance rating classifications for each R_G group. The difference between counts for SWOT-DR and GRADES are labelled above the pink SWOT-DR bars.

5 Discussion

In Sect. 4 we have evaluated the performance of our suggested discharge estimation method, SWOT-DR, against gauged values and compared it to the daily discharge input dataset, GRADES. While the volume of water and bias in discharge values are controlled and therefore somewhat bounded by GRADES, the introduction of SWOT observations produces a discharge time series that is more consistently accurate when also considering the dynamics of the time series. For our set of 399 validation gauges, 66% of R values produced by SWOT-DR are above 0.9, showing it's particular skill at replicating temporal patterns. We have shown that the 75th quantile of the NSE values is 0.67, which is 0.19 higher than that for the GRADES flow time series, and that the $nRMSE$ is less than 0.5 in 40% of reaches with SWOT-DR while only in 23% of reaches for GRADES. The GRADES time series also has a median $RMSE$ that is $36.3\,\mathrm{m^3\,s^{-1}}$ larger, median $nRMSE$ that is 0.13 larger and

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

median NSE that is 0.30 smaller than those for SWOT-DR. However, SWOT-DR has a median $|pBias|$ that is 0.10 larger than that for GRADES, showing that SWOT-DR is limited in its ability to stabilise or correct any bias. This is expected, as the only additional constraint on the GRADES distribution that we implement in SWOT-DR is the exponential fit; a form that is shown to have good accuracy for non-extremes both in literature (Quimpo et al., 1983; Castellarin et al., 2004) and in these results. This bias must be considered for any use case; this model has no benefit, in fact it decreases accuracy, for long term water resources applications that prioritise stable water accessibility and volume characteristics. However, for applications that rely on the characterisation of flow dynamics through time, such as habitat mapping, sediment transportation and tributary interaction, this model provides a much needed insight with significant improvement to the relevant flow characteristics. Despite the degradation of bias performance, performance classification ratings defined to assess modelled discharge suitability show SWOT-DR to be at least satisfactory for $|pBias|$ in 83% of reaches, suggesting that bias is not, in most cases, the limiting factor. Using the combined min_class criteria, SWOT-DR is at least satisfactory in 53% of reaches, compared to only 36% of reaches for the GRADES dataset. This shows that the inclusion of SWOT WSE levels, and the impact this has on the flow dynamics, improves on GRADES as a suitable method to create a virtual gauging station dataset of (non-daily) timestamped flow values, with particular improvement at locations where GRADES performs poorly, as defined by the Spearman Correlation R_S, shown in Sect. 4.5. This is important because it is an area of discharge prediction that current models struggle to represent accurately (Senent-aparicio et al., 2021; Liu et al., 2011; Gleason et al., In Review).

Studies both pre- and post-launch have suggested that bias will be the largest source of error in discharge calculation (Durand et al., 2023; Andreadis et al., 2025), as it is controlled by the bias in prior estimates used in the models. In our case, using the GRADES dataset, this is consistent with our results. Our median $|pBias|$ of 0.32 (32%) is lower than the first post-launch discharge evaluation by Andreadis et al. (2025), which showed a median of 56%, though these are both higher than the pre-launch estimates of 15% (Durand et al., 2023). This early study also had a median NSE of less than zero, while SWOT-DR shows a median NSE of 0.36, despite the bias in both cases. It is worth noting that this initial study was conducted over only 65 reaches, using only the highest quality data based on different filtering methods. The more recent versions of the DAWG discharge algorithms perform substantially better than the initial release, with significantly larger spatial scale, but no paper on these results has been published at the time of writing. The advent of improved hydrography with SWORD Version 17, upgraded Version D SWOT observations, and improved algorithms will likely continue to improve these statistics across all methods. The DAWG algorithms, all dependent on input datasets, could also be further improved with a more refined prior, and these results suggest that SWOT-DR, a model that enhances the current prior, GRADES, with SWOT observations, could provide such a development. The combination of these approaches inherits the benefits of both the statistical and physical models, calibrating lower flows seen by SWOT whilst allowing for extrapolation to larger events.

Additionally to the improvements in accuracy, we also increase the spatial scope of discharge results from initial studies. Andreadis et al. (2025) models discharge for 11,389 reaches where SWOT observations are considered to be of the highest quality, stating this as a benchmark for future models to improve upon. Due in combination to the increased data length (21 months in our study over the 16 months in Andreadis et al. (2025)), improved understanding of quality filtering allowing for more relaxed constraints, and the simplicity of SWOT-DR, we increase this number to 92,955; an increase of 816% in the

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

number of reaches globally where we can estimate discharge from SWOT observations. However, we also suggest that further filtering is applied, detailed further in Appendix Sect. D, in order for the SWOT observations to be of high quality; one that is impossible to quantify without global gauged water levels. Of our collected gauged reaches, 42% were kept by this method; if this is assumed to be representative of global reaches, which of course we cannot say without sufficient gauge data, this would indicate that SWOT-DR still improves upon the spatial scale with this additional filter with an increase of over 300% in the number of reaches we can apply the method. Nevertheless, SWOT-DR is applicable in many more reaches, and this shows the progress in the understanding of SWOT data and in algorithm development over only a year and suggests that the uses of this work, for all developed algorithms, will be improved over time, aided by the increased number and quality of observations this will provide.

As the satellite continues to collect measurements, we may also start to predict more extreme events. We have suggested a minimum exceedance probability of $p_{min} = 0.05$, above which SWOT can be directly used by this statistical method to estimate discharge. Due to SWOT's short record and relatively sparse temporal resolution in comparison to a daily or sub-daily dataset, a physical model would likely produce more accurate discharge than this statistical approach at less frequent events. Key discharge statistics calculated from the SWOT-DR could provide valuable prior information to a model such as LISFLOOD-FP (Bates and Roo, 2000; Neal et al., 2012), to simulate out-of-bank flows cohesive with the results presented in this paper. Such a complementary set of physical models influenced by the characteristics of these non-extreme distributions could be produced to create a set of consistent and continuous rating curves to span the full discharge distribution. However, this value of p_{min} still allows us to produce a discharge value at 95% of (filtered) timestamps observed by SWOT, at 95% of topographically filtered SWORD river reaches; a view of discharge at a large spatial scale.

6 Conclusions

In this paper, we have suggested a reanalysis method of estimating discharge from SWOT WSE values, which we call SWOT-DR, creating a global set of virtual gauging stations with the same temporal resolution as the SWOT satellite on an unprecedented spatial scale with accuracy higher than previous models. We have shown that due to the accuracy of the observations, the dynamics of the resultant modelled discharge time series' are more accurate than the daily discharge model, GRADES-hydroDL, that was used as an input to SWOT-DR. These are not suggested as a replacement for GRADES, however, as the temporal resolution is much lower. This method also produces a set of global rating curves for non-extreme events, a relationship that has previously been difficult to calibrate to observations on this spatial scale. Whilst more accurate discharge values and rating curves may be generated on local scales using available observations and models, the capacity to produce globally consistent virtual gauging stations and rating curves in both highly-monitored and data-scarce regions alike could prove valuable to numerous applications hydrological research. This is the purpose of our investigations, as SWOT presents a new opportunity to enhance current methods on this larger spatial scale, both in coverage and number of locations, over previous altimetry missions or individual gauging-station agencies. Ultimately, discharge estimation forced by observations with the high precision and accuracy SWOT aims for could aid the analysis of discharge patterns and dynamics to improve our understanding

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

of freshwater movement, modelled, at a minimum, in all rivers that exceed 100 m in width globally; knowledge that will only continue to grow as the SWOT observation record increases in length. This data is essential to a variety of applications and ultimately to both utilising the natural water cycle and transportation of water, and to protecting the population from future flood risk.

Code and data availability. All data used in this study are freely available. The SWOT L2_HR_RiverSP product can be accessed from the NASA Earthdata portal at https://doi.org/10.5067/SWOT-RIVERSP-2.0. The GRADES-hydroDL daily discharge dataset can be accessed from the GRADES-hydroDL website at https://www.reachhydro.org/home/records/grades-hydrodl. The validation gauge observations was compiled from a variety of sources, as described in Sect. 2.4, and referenced as USGS; NRFA; UKCEH; BoM; EAU; ECC; ANA. The data can be accessed respectively at https://waterdata.usgs.gov/, https://nrfa.ceh.ac.uk/, https://www.ceh.ac.uk/, https://www.bom.gov.au/water/hrs/, https://www.eaufrance.fr/, https://www.canada.ca/en/environment-climate-change.html and https://www.gov.br/ana/en/monitoring. The processing scripts and the resultant datasets used in this study are available upon request from the corresponding author, and will be published on Zenodo upon final publication of this paper.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Appendix A: Choice of Minimum Exceedance Probability

We present a method to be used for non-extreme events only, due to characteristics of the input SWOT WSE observations. In this section, we will show why this is necessary and our reasoning for the choice of $p_{min} = 0.05$. Sensitivity analysis on p_{min} showed relatively minimal effect on the resultant discharge time series, but we present our findings here.

There are 2 steps of SWOT-DR which rely on p_{min}; firstly, the subset of the input WSE values at which we produce a modelled discharge (Sect. 3.2), and secondly, the subset of the input discharge time series that is used to estimate the parameters in Eq. (2) (Sect. 3.3). The former of these will evolve in time, but the latter will produce fixed results from which we can continue to use the same output (unless an updated model is released). At each of our 399 validation reaches, we run SWOT-DR at 21 different p_{min} values equally spaced in the interval $p_{min} \in [0, 0.2]$, chosen to represent possible extreme events ranges.

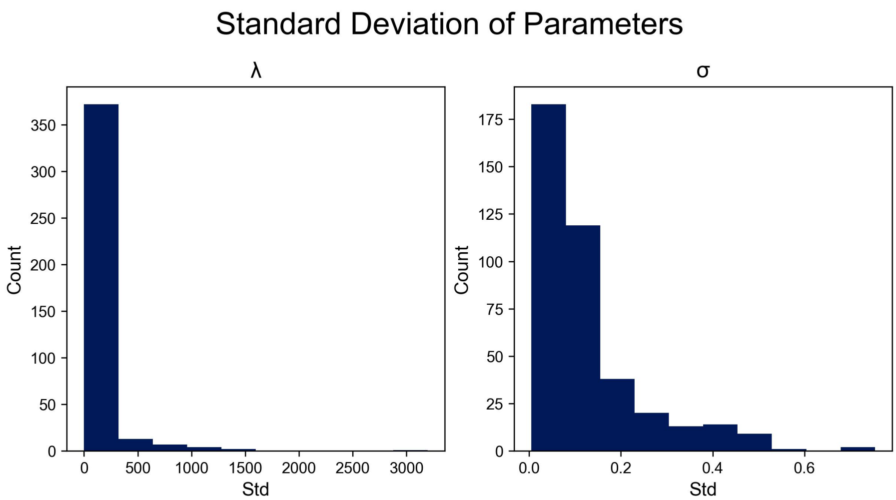

Figure A1. Histogram of the standard deviation in the derived flow-duration curve parameters over the different values of p_{min} at each reach.

Figure A1 shows that there is relatively little deviation in the parameter values derived using different p_{min} values. The scale, λ, naturally shows a larger variation, as this directly correlates with removing the largest scale events, but on the whole the difference is fairly minimal. The median standard deviation of λ is $37 m^3 s^{-1}$ and of σ parameter is 0.09. The relevance of these differences is more apparent when comparing the result modelled discharges to those of the gauged observations.

Figure A2 shows the variation in the cumulative distribution of 6 metrics chosen to evaluate SWOT-DR against gauged discharge. The primary conclusion we draw from this is that the difference in these curves is relatively small, indicating that the impact of p_{min} on overall model performance is low. Increasing p_{min} has the added effect of reducing the number of modelled values, so the combination of both variables must be considered to diagnose the cause of any differences. Increasing p_{min} introduces a negative $pBias$ to SWOT-DR, but the absolute values are very stable. The $nRMSE$ is fractionally better for higher p_{min}, whereas NSE and R seem to perform slightly better at lower values of p_{min}. It must be noted that due to the

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

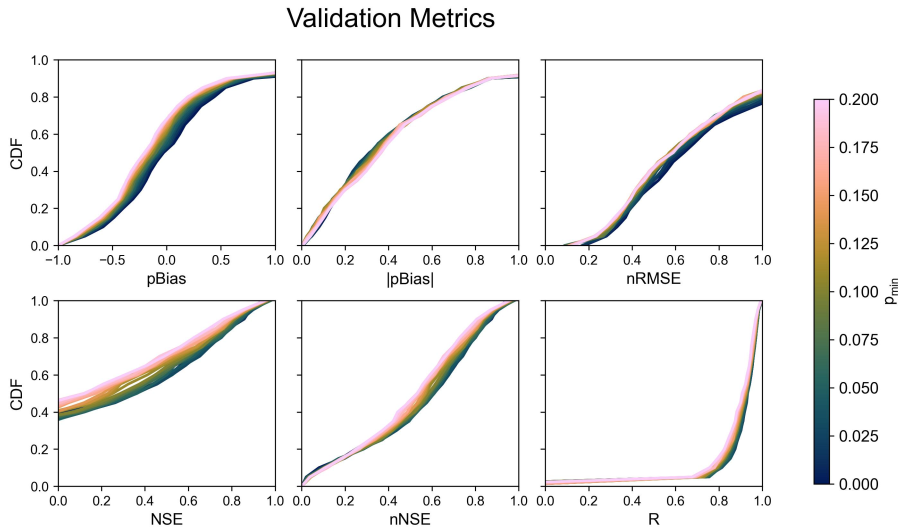

Figure A2. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series, using parameters derived using different p_{min} values.

inclusion of higher discharge events here, any variability within the data appears less significant than it would for a set of more constrained data values from a higher p_{min}, and hence conclusions drawn from these metrics need to be considered carefully.

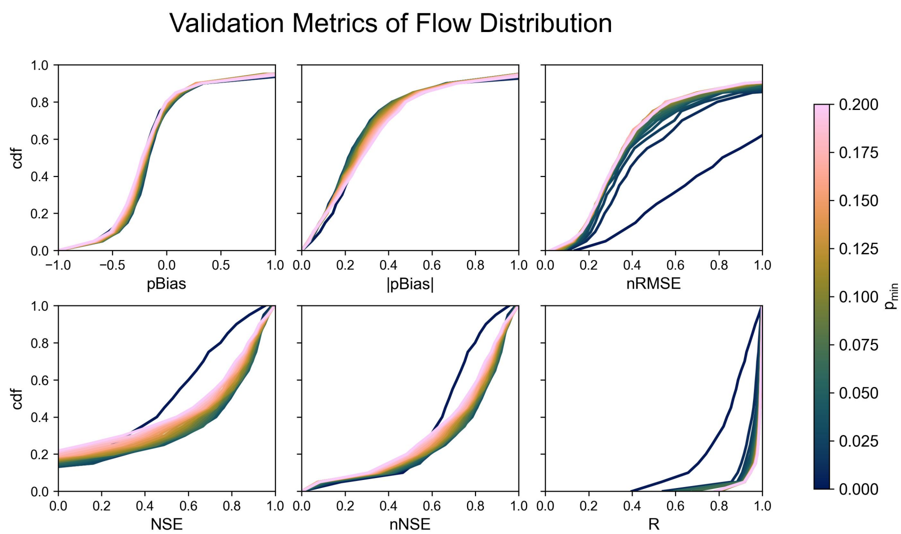

Figure A3. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series, using parameters derived using different p_{min} values.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

A larger difference can be seen on the distributions of discharge created different values of p_{min}, seen in Fig. A3. This clearly shows improved performance for higher values of p_{min} as defined by R and $nRMSE$, while the opposite is (mostly) true for NSE. However all plots emphasise the need for a non-zero p_{min}, and also to balance these conflicting effects carefully in order to produce the most reliable and consistent results.

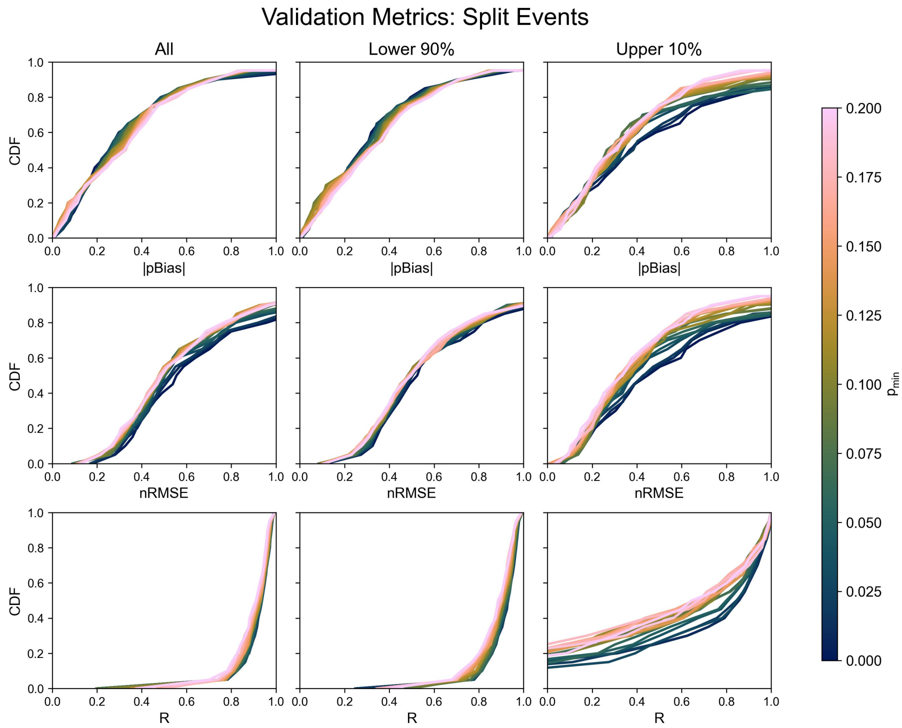

Figure A4. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge distribution, using parameters derived using different p_{min} values.

To explain why a different method is needed for extreme events, we evaluate the upper 10% and lower 90% of modelled events separately for each reach and p_{min} combination, the summaries of which can be seen in Fig. A4. For fair comparison, this is only done at the 169 reaches for which there are at least 3 observations (and hence defined metric values) in the upper 10% of WSE values that have a gauged discharge to evaluate against. Overall, we can see that the difference from changing p_{min} is linked to the more extreme events, as would be expected, but that while R is improved by a lower p_{min}, the opposite is true for both $|pBias|$ and $nRMSE$. These conclusions emphasise the need for a minimum exceedance probability which we chose to estimate, in order to produce consistently accurate results across the modelled range, as model performance is unreliable at more extreme events.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

A1 Optimal Value

As would be expected, the optimal value of p_{min} varies on a reach-by-reach basis, but we look for the most suitable value to apply globally, and therefore where to classify this threshold for the purposes of SWOT-DR.

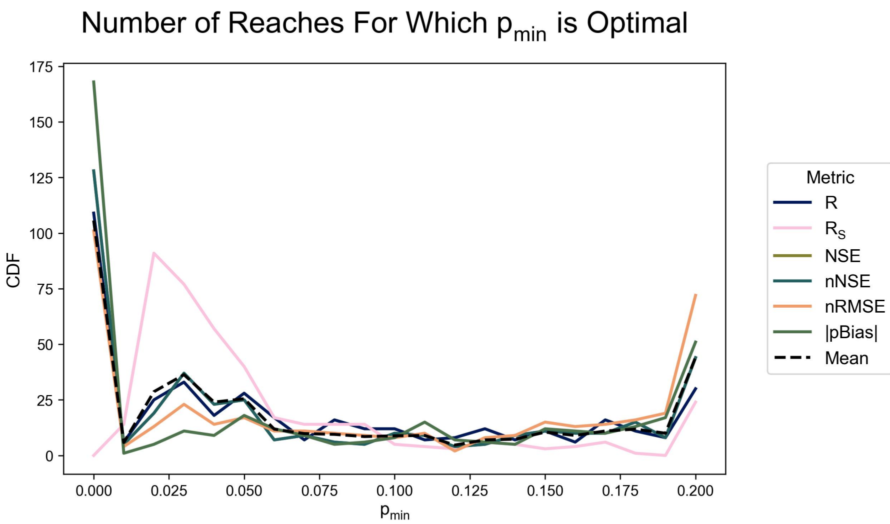

Figure A5. Count of reaches where p_{min} is the optimal value, defined by different metrics. This is the smallest value for $nRMSE$ and $|pBias|$, and the largest value for all other metrics.

Our first approach is to calculate the optimal value of p_{min} with respect to different metrics at each reach individually. For this, we discard completely unnormalised metrics. The reasoning behind this choice is to represent all attributes of a model's skill – scale, variation, distribution and dynamics – whilst attempting to limit the bias caused by higher values, which are by definition correlated with lower p_{min} values.

Figure A5 shows the raw counts of these calculations, both individually for each metric and the mean over all 6. Other than the expected maximum at $p_{min} = 0$ for most metrics, due to higher observation numbers, and then at $p_{min} = 0.2$, which excludes the most extreme events, the globally optimal value seems to lie within the range of $p_{min} \in [0.02, 0.07]$. Figure A6 shows the distribution of the modally optimal value for each reach (from the 6 metrics), and of the mean optimal value. Similarly, both of these plots suggest that a value of $p_{min} \in [0.02, 0.08]$ would be a suitable choice. These lower values also allow for SWOT-DR to run at a greater proportion of events and therefore increases the functionality of the model.

A2 Comparison to GRADES

Figure A7 plots the fraction of reaches where SWOT-DR outperforms GRADES as defined by each metric. As above, we see that R shows significant improvement from SWOT-DR over GRADES for all values of p_{min} though the $|pBias|$ remains higher. Also similarly to the previous section, the left hand plot clearly shows that the variation due to p_{min} is minimal, with

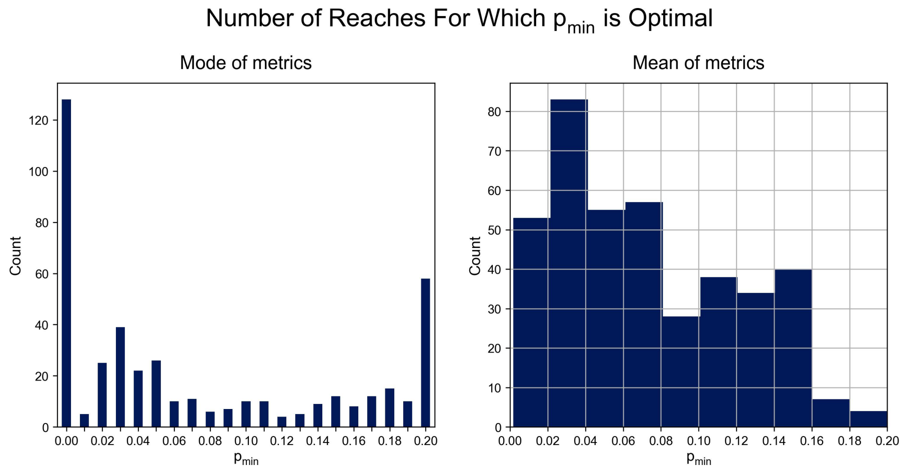

Figure A6. Count of reaches where p_{min} is the optimal value, as in Fig. A5, and then taking the modal and mean value respectively over the different metrics for each reach.

the range of these fractions over all p_{min} values all between 0.027 and 0.044 (2.7-4.4% of reaches) for all metrics. In the right hand plot, the metrics are separately normalised to $[0,1]$ to better compare dynamics along the p_{min} axis. This plot, along with the small change in the proportion of reaches improved upon, emphasises lack of any clear trends and the minimal difference made by changing p_{min}, and hence that a range of values will produce a reasonable result, and can therefore be changed to suit individual needs. It does suggest, however, that a non-zero p_{min} can improve results. The centre plot strengthens this conclusion, with an optimal choice of $p_{min} \in [0.02, 0.07]$, though it clearly shows the degradation in performance if p_{min} is increased too high.

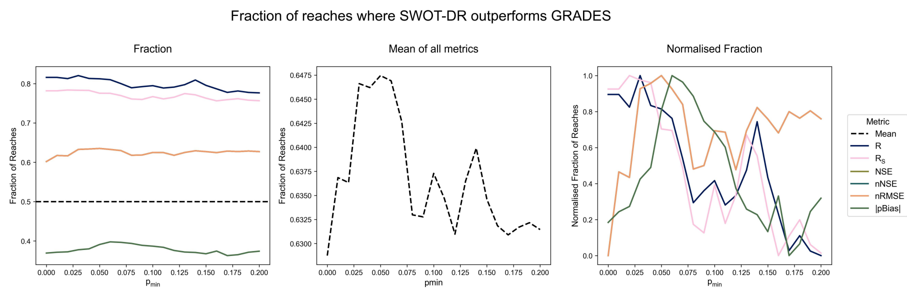

Figure A7. A plot to show how the fraction of reaches where SWOT-DR outperforms GRADES, as defined by different metrics, varies with choice of p_{min}. The left hand plot shows all metrics, the centre plot shows the mean, and the right hand graph shows the values when each metric is normalised to the range $[0,1]$, to compare shape more explicitly.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

A3 Summary

For our purposes, considering all of the above, we choose to use a value of $p_{min} = 0.05$ as a suitable mark of extreme events for SWOT-DR. Both Sects. A1 and A2 suggest this to be within optimal range, whilst retaining as much of the input data as possible in order to run SWOT-DR sufficiently well for a maximal number of SWOT observations.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Appendix B: Water Level Record Length

Figure 7 shows that while the dynamics and overall prediction accuracy of SWOT-DR are an improvement on GRADES, we do introduce an increased (mostly negative) bias. We hypothesised that this was due to the short record and sparse temporal resolution of SWOT measurements; in many reaches, the set of WSE levels seen by the satellite only represents a small fraction of the distribution seen in daily gauged observations. This is not an inaccuracy of SWOT; as seen in Sect. 4.1, when evaluated against timestamped gauged observations, SWOT shows a high level of skill. However, as measurements have a median frequency of 2 times every 21 days globally for a duration of 2 years, we miss a large proportion of the level distribution due to a relative scarcity of observation values. Filtering of observations only worsens this effect, with a global median of 1 valid filtered observation per cycle. We test this hypothesis by using gauged water levels in place of SWOT water levels, with observation records of up to 45 years.

We applied SWOT-DR using only gauged water levels as input to Sect. 3.2, replacing SWOT WSE levels for both input w and the distribution W, for varying input time periods. This is important, because we have theorised that SWOT's short record length is a major source of the errors we see from SWOT-DR, and that increasing this length will improve our results. We test all possible time periods of full calendar years starting between 1980 (the earliest GRADES record) and 2023 (the start of the SWOT Science Orbit) and ending before 2025 (the end of SWOT observation collection under Version C). This give us data lengths between 2 and 45 years, and we will call these model variants SWOT-DR$_{Gauge}$. We then performed the same tests as in Sect. 4.3, and the results are shown in Figs. B1 and B2.

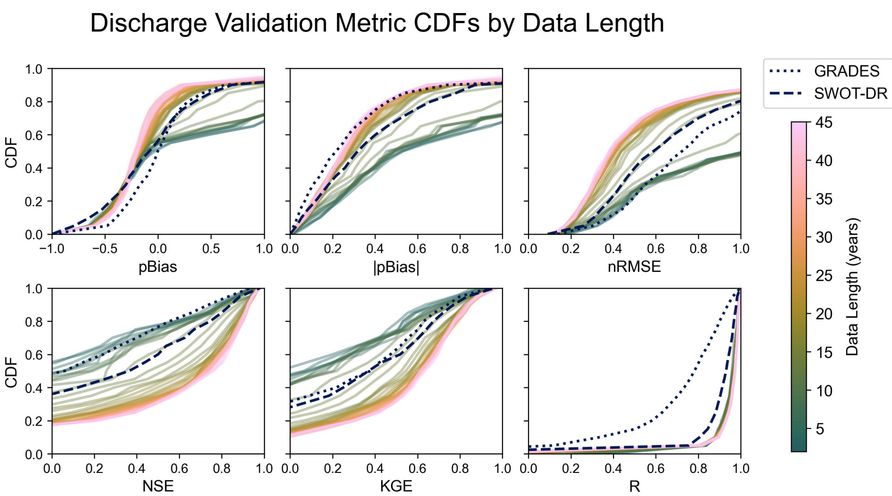

Figure B1. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series, for various data record lengths of gauged water levels as inputs to run SWOT-DR$_{Gauge}$.

Figure B1 clearly shows that a longer data record can significantly improve model performance under all of these metrics. Figure B2 shows that while the length of the data record has quite a significant impact on shorter record lengths, with large

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

improvements in all metrics, this effect reduces quickly. This all supports our theory that SWOT-DR will only continue to improve as SWOT continues to provide observations. For comparison, the original SWOT-DR and GRADES models are
555 plotted in Fig. B1. SWOT-DR is within the range seen by the SWOT-DR$_{Gauge}$ models, which is promising; the SWOT mission has been running for 2 years, the minimum data record length plotted in these figures. This shows that these results are actually a conservative estimate of the potential skill of SWOT-DR, based on the short time series so far.

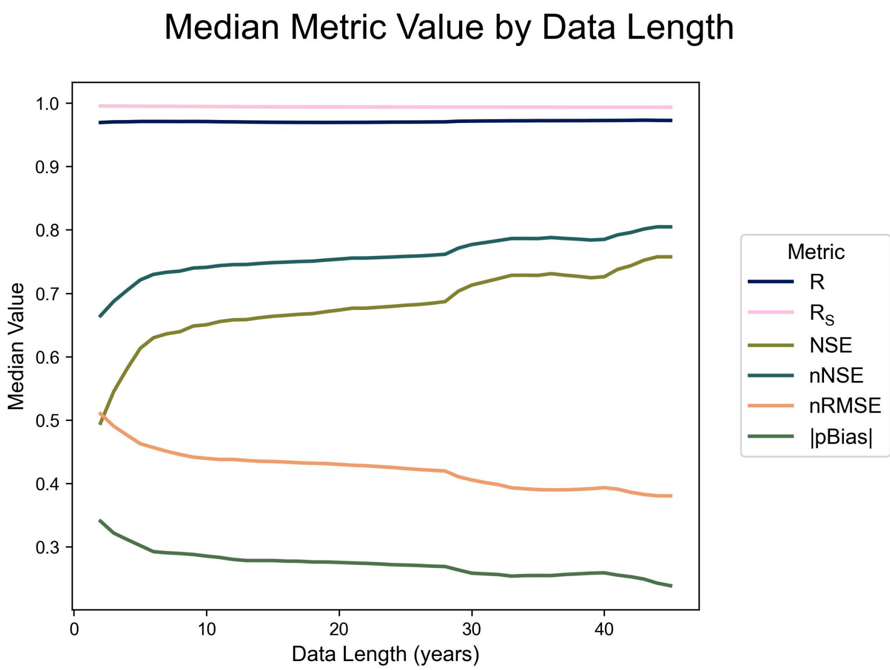

Figure B2. Summary plot showing the median value of various metrics over all reaches for SWOT-DR$_{Gauge}$, for each time period in Fig. B1.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Appendix C: GRADES Input Data

An initial decision to be made was whether to use only the GRADES data since the SWOT Science Orbit in July 2023, or whether to use the whole record, back to January 1980, to estimate the flow distribution parameters λ and σ defined in Eq. (2). An argument could be made to use only the period of time overlapping with SWOT, so that the level-duration and flow-duration curves from Sects. 3.2 and 4.2 are consistent temporally. However, for a future model, at unseen events and therefore exceedance probabilities, it could be beneficial to use the full GRADES record, the longer time period providing more insight into rarer events.

We experimented to see the effect this could have on our results. Firstly, we calculated the flow distribution (Sect. 4.2) of SWOT-DR for 50 different input time series of randomly chosen consecutive time periods, all of SWOT length, across the GRADES record. This gave us 50 different parameter estimates for λ and σ at each reach; the distribution of the standard deviation of these parameters for each reach can be seen in Fig. C1. There is clearly a large variation in these parameters; this suggests that the use of only roughly 2 years of data (the length of the SWOT Version C observation record) does not produce a stable set of parameters, and while they may represent the specific time period well, we cannot expect the same for future events.

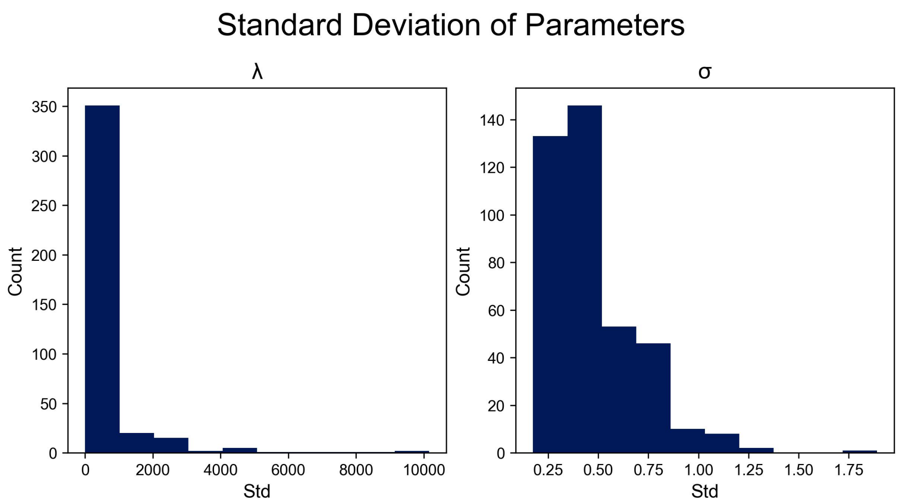

Figure C1. Histogram of the standard deviation in the derived flow-duration curve parameters at each reach across 50 different input data series, each of the same length as SWOT.

The effect on the resultant flow distribution can be seen in Fig. C2. Here, as in Sect. 4.2, we compare the modelled flow distribution to the gauge flow distribution at regularly spaced exceedance probabilities in the interval $[p_{min}, 1]$, evaluated only against gauge data within the SWOT science data collection period. The pink lines represent the model variants using the 50 parameter sets as described above, with roughly 2 years of GRADES data as input - labelled "SWOT Length". The dashed lines show the result for parameters derived from (a) all of GRADES - "All", (b) all of GRADES preceding the SWOT Science Orbit

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

- "Before SWOT", and (c) only GRADES proceeding the SWOT Science Orbit - "SWOT". All three of these distributions are at the upper end of skill across all metrics. There are 2 main conclusions to draw from this. Firstly, the SWOT time period derived parameters result in the most accurate flow distribution, as is expected, since both gauge and model data represent the same time period and therefore have the potential to have seen the same flows. Where possible, this suggests that we should only estimate our parameters based on data from the time period we are interested in. However, for the use of SWOT-DR as an alternative to future runs of GRADES, rather than a recalibration-after-the-fact, this would not be an option. The second observation is therefore that using the full length of GRADES is the best alternative. Using only GRADES from the SWOT time period ("SWOT" in Fig. C2) to estimate future events could be equivalent to using any of the generated 50 curves of the same length ("SWOT Length") to estimate events within the SWOT collection time period; i.e. any of the pink lines in Fig. C2. However, the model using only GRADES data from before the SWOT Science Orbit ("Before SWOT") is the equivalence of a model using all of GRADES ("All") to simulate future events (that are therefore unmodelled by GRADES). The latter option is clearly more consistent and reliable in it's performance. The distributions using all of GRADES and only that before SWOT have almost indistinguishable difference; these both perform well compared to the range of skill seen by the shorter input lengths and therefore justifies the use of the full time period in Sect. 4.2.

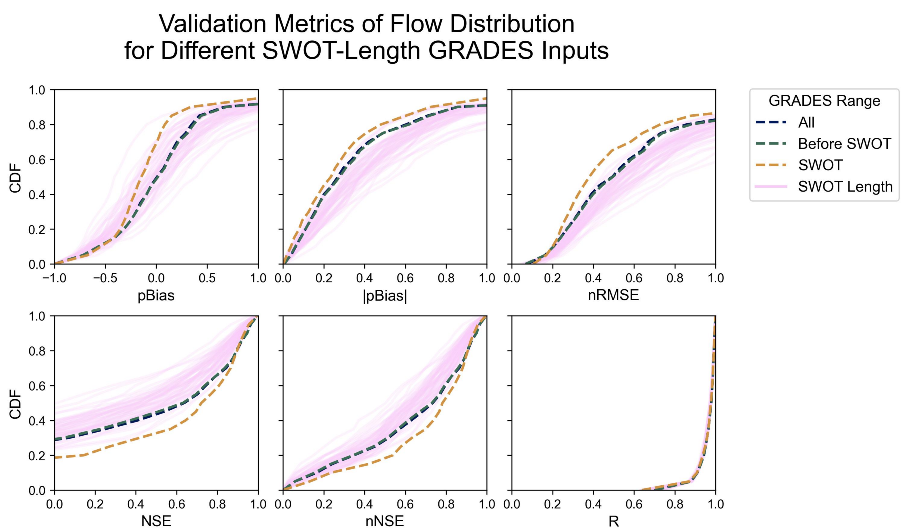

Figure C2. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge distribution, using different input GRADES time series.

To see the impact on our full time series, we plot the same metrics in Fig. C3, and draw the same conclusions. The application of both the flow-duration and level-duration quantile matching improves the comparable “All” variant relative to the “SWOT” variant, and similarly the comparable "Before SWOT" variant to the majority of the "SWOT Length" variants, supporting our decision to use the full GRADES time series further.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

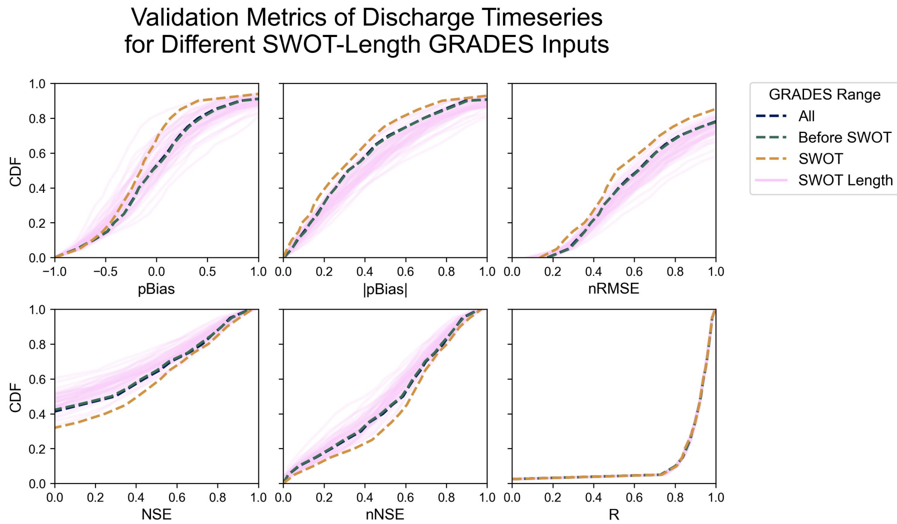

Figure C3. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series, using parameters derived using different input GRADES time series.

Comparing resultant discharge time series at our chosen evaluation reaches, as in Fig. C3, the median NSE is 0.28 for "Before SWOT", whereas our random samples of "SWOT Length" time periods have a median NSE of 0.11. The median $|pBias|$ of the former model is 0.32, less than only 0.002 below that of the "All" model, and only slightly larger than the "SWOT" exact time period model, which has a median value of 0.28; in comparison, the "SWOT Length" models have a median of 0.39. This shows that for future events, the accuracy is increased using maximal GRADES data, rather than using a SWOT-length dataset, when trained on data with a time period disjoint from the calculation, or validation, dataset time period.

Considering these experiments, we decided it would be best to use the whole GRADES record for the application of SWOT-DR to new SWOT values, in order to represent events outside of the range seen by SWOT so far. In the instance that GRADES has been run over all dates of interest, it would be beneficial to recalculate the parameters using only data since the SWOT Science Orbit, however our experiments have shown that using the full record of GRADES produces a model of comparable skill. This allows for SWOT-DR to run in future without running an extension on GRADES, creating a more time- and compute-efficient model, without compromising too much on accuracy.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Appendix D: Extended Gauges

In Section 2.4.1, we chose to filter the SWOT observations by evaluating the water levels against gauged water levels, specifically with regards to monotonicity of the two timeseries'. The filters we applied were as follows:

$$R_S > 0.8$$

$$\tau_k > 0.7$$

for measures of monotonicity R_S the Spearman R value, and τ_k the Kendall Tau value. We only compare monotonicity, rather than error, bias or scale of the distributions because of varying datums between different sources. Additionally, only the probability of exceedance of water level is used for our method, not absolute values; and therefore this is the important aspect to ensure the accuracy and quality of our input SWOT water levels. Here, we present our findings without this filter, and hence show how important the quality of the SWOT observations are to predicting discharge.

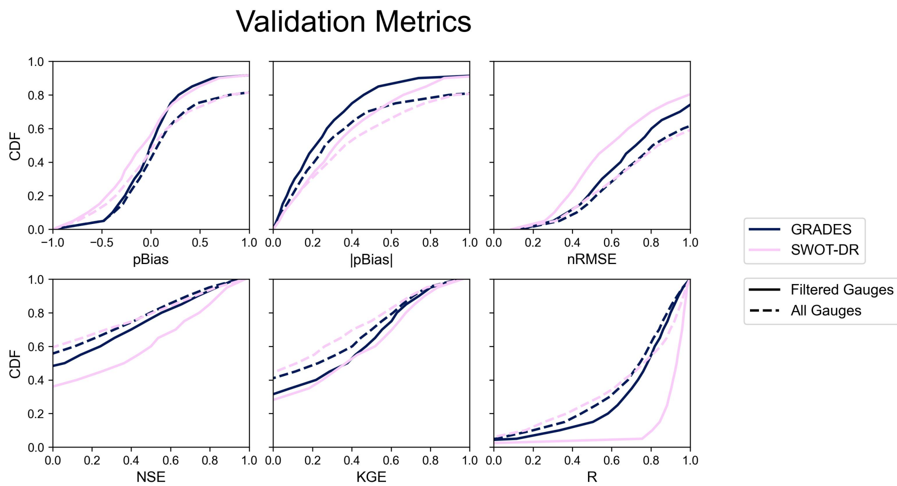

Figure D1. Cumulative distribution functions of $pBias$, $|pBias|$, $nRMSE$, NSE, KGE and R, calculated by comparing modelled to gauged discharge time series at all collected gauges.

Figure D1 shows the same metric distributions as Fig 7, but allows us to see the difference in model performance based on the accuracy of the input SWOT observations, by comparing the results for all 941 gauges to those with our filtered 399 gauges. We can see that SWOT-DR performs similarly to GRADES across all metrics shown, and that compared to our more filtered data, SWOT-DR produces less accurate results while the difference to the metric distributions for the GRADES model is therefore less significant, though still pronounced. For example, the median NSE value of SWOT-DR over all collected gauges is -0.34, 0.7 lower than that for the filtered gauges; for GRADES, this statistic only drops by 0.35 (from 0.06 to -0.29) when including all gauges. Similarly, the median R value across all gauges is 0.76 and 0.74 for SWOT-DR and GRADES respectively, while for the filtered gauges, these values are 0.93 and 0.78; a much bigger improvement in SWOT-DR by filtering, though the two

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

methods are similar for all gauges. This shows that accurate SWOT observations are essential to reliably predict discharge, and that while for some reaches the Permissive Relaxed filtering configuration in combination with Version C SWOT observations results in a sufficiently accurate representation of the local water level, this is not the case in many locations. With updated versions of SWOT processing, node-level observations and a greater understanding of different attributes and how to filter them, we can hope that more reaches will meet this threshold in future, and can therefore be used to improve discharge predictions on a wider scale.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Author contributions. Izzy Probyn: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original draft and Reviewing and Editing, Visualisation. Jeffrey Neal: Conceptualization, Methodology, Writing - Reviewing and Editing, Supervision. Stephen J. Chuter: Conceptualization, Methodology, Writing - Reviewing and Editing, Supervision. Paul Bates: Conceptualization, Methodology, Writing - Reviewing, Supervision.

Competing interests. The contact author has declared that none of the authors has any competing interests.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS-global ensemble streamflow forecasting and flood early warning, Hydrology and Earth System Sciences, 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013.

Allen, G. H. and Pavelsky, T. M.: Global extent of rivers and streams, Science, 361, 585–588, https://doi.org/10.1126/science.aat0636, 2018.

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., de Moraes Frasson, R. P., and Bendezu, L.: The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A Global River Network for Satellite Data Products, Water Resources Research, 57, e2021WR030 054, https://doi.org/10.1029/2021WR030054, 2021.

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., d. M. Frasson, R. P., and Bendezu, L.: SWOT River Database (SWORD), https://doi.org/10.5281/ZENODO.10013982, 2023.

ANA: Monitoring — Agência Nacional de Águas e Saneamento Básico (ANA), https://www.gov.br/ana/en/monitoring, last access: 27 June 2026.

Andreadis, K. M., Coss, S. P., Durand, M., Gleason, C. J., Simmons, T. T., Tebaldi, N., Bjerklie, D. M., Brinkerhoff, C., Dudley, R. W., Gejadze, I., Larnier, K., Malaterre, P. O., Oubanas, H., Allen, G. H., Bates, P. D., David, C. H., Domeneghetti, A., Elmi, O., Marc, L. F., de Moraes Frasson, R. P., Friedmann, E., Garambois, P. A., Gehring, J., Getirana, A., Hughes, M., Lee, J., Matte, P., Minear, J. T., Monnier, J., Muhebwa, A., Tourian, M. J., Pavelsky, T. M., Riggs, R. M., Rodríguez, E., Sikder, M. S., Smith, L. C., Stuurman, C., Taneja, J., Tarpanelli, A., Wang, J., Williams, B. A., and Yadav, B.: A First Look at River Discharge Estimation From SWOT Satellite Observations, Geophysical Research Letters, 52, e2024GL114 185, https://doi.org/10.1029/2024GL114185, 2025.

Bahekar, D., Ahmad, I., and Prasad, A. D.: A review of methodologies to estimate river discharge from satellite observations, Research Journal of Engineering Sciences, 10, 15–19, www.isca.me, 2021.

Bates, P. D. and Roo, A. P. D.: A simple raster-based model for flood inundation simulation, Journal of Hydrology, 236, 54–77, https://doi.org/10.1016/S0022-1694(00)00278-X, 2000.

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Schumann, G., Giustarini, L., Coxon, G., Porter, J. R., Amodeo, M. F., Chu, Z., Lewis-Gruss, S., Freeman, N. B., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K. E., Emanuel, K., Ferreira, C. M., Khalid, A., Haigh, I. D., Couasnon, A., Kopp, R. E., Hsiang, S., and Krajewski, W. F.: Combined Modeling of US Fluvial, Pluvial, and Coastal Flood Hazard Under Current and Future Climates, Water Resources Research, 57, e2020WR028 673, https://doi.org/10.1029/2020WR028673, 2021.

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and Its Capabilities for Land Hydrology, Surveys in Geophysics 2015 37:2, 37, 307–337, https://doi.org/10.1007/s10712-015-9346-y, 2015.

BoM: Hydrologic Reference Stations: Water Information: Bureau of Meteorology, https://www.bom.gov.au/water/hrs/, last access: 27 June 2026.

Brighenti, T. M., Bonumá, N. B., Grison, F., de Almeida Mota, A., Kobiyama, M., and Chaffe, P. L. B.: Two calibration methods for modeling streamflow and suspended sediment with the swat model, Ecological Engineering, 127, 103–113, https://doi.org/10.1016/j.ecoleng.2018.11.007, 2019.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A.: Regional flow-duration curves: reliability for ungauged basins, Advances in Water Resources, 27, 953–965, https://doi.org/10.1016/J.ADVWATRES.2004.08.005, 2004.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J., Howden, N. J., Quinn, N., Wagener, T., and Woods, R.: DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology, Geoscientific Model Development, 12, 2285–2306, https://doi.org/10.5194/gmd-12-2285-2019, 2019.

de Oliveira, D. Y. and Vrugt, J. A.: The Treatment of Uncertainty in Hydrometric Observations: A Probabilistic Description of Streamflow Records, Water Resources Research, 58, e2022WR032 263, https://doi.org/10.1029/2022WR032263, 2022.

Desai, S.: Surface Water and Ocean Topography Mission (SWOT) Project Science Requirements Document, Rev B, JPL D-61923, Tech. rep., Jet Propulsion Laboratory, https://swot.jpl.nasa.gov/system/documents/files/2176_2176_D-61923_SRD_Rev_B_20181113.pdf, 2018.

Dijk, A. I. V., Brakenridge, G. R., Kettner, A. J., Beck, H. E., Groeve, T. D., and Schellekens, J.: River gauging at global scale using optical and passive microwave remote sensing, Water Resources Research, 52, 6404–6418, https://doi.org/10.1002/2015WR018545, 2016.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata, Earth System Science Data, 10, 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018.

Du, B., Jin, T., Liu, D., Wang, Y., and Wu, X.: Accurate Discharge Estimation Based on River Widths of SWOT and Constrained At-Many-Stations Hydraulic Geometry, Remote Sensing 2023, Vol. 15, Page 1672, 15, 1672, https://doi.org/10.3390/rs15061672, 2023.

Durand, M., Gleason, C. J., Pavelsky, T. M., de Moraes Frasson, R. P., Turmon, M., David, C. H., Altenau, E. H., Tebaldi, N., Larnier, K., Monnier, J., Malaterre, P. O., Oubanas, H., Allen, G. H., Astifan, B., Brinkerhoff, C., Bates, P. D., Bjerklie, D., Coss, S., Dudley, R., Fenoglio, L., Garambois, P. A., Getirana, A., Lin, P., Margulis, S. A., Matte, P., Minear, J. T., Muhebwa, A., Pan, M., Peters, D., Riggs, R., Sikder, M. S., Simmons, T., Stuurman, C., Taneja, J., Tarpanelli, A., Schulze, K., Tourian, M. J., and Wang, J.: A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission, Water Resources Research, 59, e2021WR031 614, https://doi.org/10.1029/2021WR031614, 2023.

EAU: Eaufrance | The public water information service, https://www.eaufrance.fr/, last access: 27 June 2026.

ECC: Environment and Climate Change Canada - Canada.ca, https://www.canada.ca/en/environment-climate-change.html, last access: 27 June 2026.

Feng, D., Fang, K., and Shen, C.: Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales, Water Resources Research, 56, e2019WR026 793, https://doi.org/10.1029/2019WR026793, 2020.

Feng, D., Lawson, K., and Shen, C.: Mitigating Prediction Error of Deep Learning Streamflow Models in Large Data-Sparse Regions With Ensemble Modeling and Soft Data, Geophysical Research Letters, 48, e2021GL092 999, https://doi.org/10.1029/2021GL092999, 2021.

Galelli, S., Turner, S. W., Pokhrel, Y., Ng., J. Y., Castelletti, A., Bierkens, M. F., Pianosi, F., and Biemans, H.: Advancing the Representation of Human Actions in Large-Scale Hydrological Models: Challenges and Future Research Directions, Water Resources Research, 61, e2024WR039 486, https://doi.org/10.1029/2024WR039486, 2025.

Gleason, C. J. and Durand, M. T.: Remote Sensing of River Discharge: A Review and a Framing for the Discipline, Remote Sensing 2020, Vol. 12, Page 1107, 12, 1107, https://doi.org/10.3390/rs12071107, 2020.

Gleason, C. J. and Smith, L. C.: Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry, Proceedings of the National Academy of Sciences of the United States of America, 111, 4788–4791, https://doi.org/10.1073/pnas.1317606111, 2014.

Gleason, C. J., Bates, P. D., Durand, M. T., Feng, D., Getirana, A., Lin, P., Oubanas, H., Pavelsky, T. M., Shen, C., Smith, L. C., Wang, J., Yamazaki, D., Yang, Y., An, H., Bennitt, F. B., Chuter, S. J., Fang, C., Flores, J. A., Friedmann, E., Jarugula, S., Ji, H., Langhorst, T., Long, D., Vu, D. T., Yin, Z., Allen, G. H., Andreadis, K. M., Biancamaria, S., Brown, C. M., David, C. H., Pan, M., Neal, J. C., Coss, S. P., Harlan, M. E., Simmons, T. T., and Tebaldi, N.: SWOT, empiricism, and river modelling, Geophysical Research Letters, In Review.

Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P.-E., Clark, E., Krajewski, W. F., Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P.-E., Clark, E., and Krajewski, W. F.: A Unified Flash Flood Database across the United States, Bulletin of the American Meteorological Society, 94, 799–805, https://doi.org/10.1175/BAMS-D-12-00198.1, 2013.

Greguska, F., Tebaldi, N., McDonald, V., Vanesa, Nickles, C., and Wood, J.: podaac/hydrocron, https://doi.org/10.5281/ZENODO.15831782, 2024.

Grgić, M., Bašić, T., Grgić, M., and Bašić, T.: Radar Satellite Altimetry in Geodesy - Theory, Applications and Recent Developments, Geodetic Sciences - Theory, Applications and Recent Developments, https://doi.org/10.5772/intechopen.97349, 2021.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration, Journal of Hydrologic Engineering, 4, 135–143, https://doi.org/10.1061/(asce)1084-0699(1999)4:2(135), 1999.

Hao, Z., Xiang, N., Cai, X., Zhong, M., Jin, J., Du, Y., and Ling, F.: Remote Sensing of River Discharge From Medium-Resolution Satellite Imagery Based on Deep Learning, Water Resources Research, 60, e2023WR036 880, https://doi.org/10.1029/2023WR036880, 2024.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrology and Earth System Sciences, 12, 841–861, https://doi.org/10.5194/hess-12-841-2008, 2008.

Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., and Pulido-Velazquez, D.: A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain, Water 2018, Vol. 10, Page 192, 10, 192, https://doi.org/10.3390/w10020192, 2018.

Kalin, L., Isik, S., Schoonover, J. E., and Lockaby, B. G.: Predicting Water Quality in Unmonitored Watersheds Using Artificial Neural Networks, Journal of Environmental Quality, 39, 1429–1440, https://doi.org/10.2134/jeq2009.0441, 2010.

Ke, S., Tourian, M. J., Sneeuw, N., de Moraes Frasson, R. P., Paiva, R. C., Durand, M., Gleason, C., Elmi, O., Malaterre, P. O., and David, C.: A Kalman Filter Approach for Estimating Daily Discharge Using Space-Based Discharge Estimates, Water Resources Research, 61, e2024WR037 667, https://doi.org/10.1029/2024WR037667, 2025.

Levin, S. B., Briggs, M. A., Foks, S. S., Goodling, P. J., Raffensperger, J. P., Rosenberry, D. O., Scholl, M. A., Tiedeman, C. R., and Webb, R. M.: Uncertainties in measuring and estimating water-budget components: Current state of the science, Wiley Interdisciplinary Reviews: Water, 10, e1646, https://doi.org/10.1002/wat2.1646, 2023.

Li, W., Xie, X., Li, W., van der Meijde, M., Yan, H., Huang, Y., Li, X., and Wang, Q.: Monitoring of Hydrological Resources in Surface Water Change by Satellite Altimetry, Remote Sensing, 14, 4904, https://doi.org/10.3390/rs14194904, 2022.

Liu, J., Rahmani, F., Lawson, K., and Shen, C.: A Multiscale Deep Learning Model for Soil Moisture Integrating Satellite and In Situ Data, Geophysical Research Letters, 49, e2021GL096 847, https://doi.org/10.1029/2021GL096847, 2022.

Liu, J., Bian, Y., Lawson, K., and Shen, C.: Probing the limit of hydrologic predictability with the Transformer network, Journal of Hydrology, 637, 131 389, https://doi.org/10.1016/j.jhydrol.2024.131389, 2024.

Liu, Y., Brown, J., Demargne, J., and Seo, D. J.: A wavelet-based approach to assessing timing errors in hydrologic predictions, Journal of Hydrology, 397, 210–224, https://doi.org/10.1016/j.jhydrol.2010.11.040, 2011.

Mengen, D., Ottinger, M., Leinenkugel, P., and Ribbe, L.: Modeling River Discharge Using Automated River Width Measurements Derived from Sentinel-1 Time Series, Remote Sensing 2020, Vol. 12, Page 3236, 12, 3236, https://doi.org/10.3390/rs12193236, 2020.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Munier, S. and Decharme, B.: River network and hydro-geomorphological parameters at 1/12 ° resolution for global hydrological and climate studies, Earth System Science Data, 14, 2239–2258, https://doi.org/10.5194/essd-14-2239-2022, 2022.

Nathanson, M., Kean, J. W., Grabs, T. J., Seibert, J., Laudon, H., and Lyon, S. W.: Modelling rating curves using remotely sensed LiDAR data, Hydrological Processes, 26, 1427–1434, https://doi.org/10.1002/hyp.9225, 2012.

Neal, J., Schumann, G., and Bates, P.: A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas, Water Resources Research, 48, https://doi.org/10.1029/2012WR012514, 2012.

Neal, J. C., Chuter, S. J., Archer, L., Hawker, L. P., Bates, P. D., Savage, J., Wilkinson, H., Rougier, J., Neal, J., Chuter, S. J., Hawker, L., and Bates, P.: Flood Inundation Modelling using the Surface Water Ocean Topography Mission, https://doi.org/10.22541/ESSOAR.174319735.54049202/V1, 2025.

NRFA: National River Flow Archive | National River Flow Archive, https://nrfa.ceh.ac.uk/, last access: 27 June 2026.

Paris, A., de Paiva, R. D., da Silva, J. S., Moreira, D. M., Calmant, S., Garambois, P. A., Collischonn, W., Bonnet, M. P., and Seyler, F.: Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin, Water Resources Research, 52, 3787–3814, https://doi.org/10.1002/2014WR016618, 2016.

Patidar, G., Indu, J., and Karmakar, S.: Performance Assessment of Surface Water and Ocean Topography (SWOT) Mission for WSE Measurement Across India, Geophysical Research Letters, 52, e2025GL115 804, https://doi.org/10.1029/2025GL115804, 2025.

Pattison, I., Lane, S. N., Hardy, R. J., and Reaney, S. M.: The role of tributary relative timing and sequencing in controlling large floods, Water Resources Research, 50, 5444–5458, https://doi.org/10.1002/2013WR014067, 2014.

Phillips, C. B., Hill, K. M., Paola, C., Singer, M. B., and Jerolmack, D. J.: Effect of Flood Hydrograph Duration, Magnitude, and Shape on Bed Load Transport Dynamics, Geophysical Research Letters, 45, 8264–8271, https://doi.org/10.1029/2018GL078976, 2018.

Poff, N. L. R., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C.: The Natural Flow Regime, BioScience, 47, 769–784, https://doi.org/10.2307/1313099, 1997.

Quimpo, R. G., Alejandrino, A. A., and McNally, T. A.: Regionalized Flow Duration for Philippines, Journal of Water Resources Planning and Management, 109, 320–330, https://doi.org/10.1061/(asce)0733-9496(1983)109:4(320), 1983.

Rakhymbek, K., Mukanova, B., Bondarovich, A., Chernykh, D., Alzhanov, A., Nurekenov, D., Pavlenko, A., and Nugumanova, A.: LSTM-Based River Discharge Forecasting Using Spatially Gridded Input Data, Data 2025, Vol. 10, Page 122, 10, 122, https://doi.org/10.3390/data10080122, 2025.

Riggs, R. M., Allen, G. H., Wang, J., Pavelsky, T. M., Gleason, C. J., David, C. H., and Durand, M.: Extending global river gauge records using satellite observations, Environmental Research Letters, 18, 064 027, https://doi.org/10.1088/1748-9326/acd407, 2023.

Saemian, Peyman, Tourian, and J., M.: SVS: SWOT Validation Set for Global River Discharge Evaluation, https://doi.org/10.18419/DARUS-5843, 2026.

Sahoo, B., Sahoo, D. P., Tiwari, M. K., Sahoo, B., Sahoo, D. P., Tiwari, M. K., Sahoo, Bhabagrahi, Sahoo, P., Debi, Tiwari, and Kumar, M.: A Hydrodynamics and Remote Sensing based Framework for Establishing Virtual Streamflow Measurement Stations in Scantily-gauged River Reaches, EGUGA, p. 18311, https://doi.org/10.5194/egusphere-egu2020-18311, 2020.

Sampson, C. C., Smith, A. M., Bates, P. B., Neal, J. C., Alfieri, L., and Freer, J. E.: A high-resolution global flood hazard model, Water Resources Research, 51, 7358–7381, https://doi.org/10.1002/2015WR016954, 2015.

Schmied, H. M., Caceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C. E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geoscientific Model Development, 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

Senent-aparicio, J., Blanco-gómez, P., López-ballesteros, A., Jimeno-sáez, P., and Pérez-sánchez, J.: Evaluating the Potential of GloFAS-ERA5 River Discharge Reanalysis Data for Calibrating the SWAT Model in the Grande San Miguel River Basin (El Salvador), Remote Sensing 2021, Vol. 13, Page 3299, 13, 3299, https://doi.org/10.3390/rs13163299, 2021.

Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resources Research, 54, 8558–8593, https://doi.org/10.1029/2018WR022643, 2018.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song, Y., Beck, H. E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Mohanty, B., Roy, T., Xu, C., and Lawson, K.: Differentiable modelling to unify machine learning and physical models for geosciences, Nature Reviews Earth & Environment 2023 4:8, 4, 552–567, https://doi.org/10.1038/s43017-023-00450-9, 2023.

Sutanudjaja, E. H., Beek, R. V., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., Ent, R. J. V. D., Graaf, I. E. D., Hoch, J. M., Jong, K. D., Karssenberg, D., López, P. L., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F.: PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model, Geoscientific Model Development, 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

SWOT: SWOT Level 2 River Single-Pass Vector Data Product, NASA Physical Oceanography Distributed Active Archive Center, https://doi.org/10.5067/SWOT-RIVERSP-2.0, 2024.

SWOT-Confluence: SWOT-Confluence: Prediagnostics, https://github.com/SWOT-Confluence/prediagnostics, last access: 09 June 2026.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modelling & Software, 90, 201–216, https://doi.org/10.1016/j.envsoft.2017.01.006, 2017.

Ternynck, C., Alaya, M. A. B., Chebana, F., Dabo-Niang, S., and Ouarda, T. B. M. J.: Streamflow Hydrograph Classification Using Functional Data Analysis, Journal of Hydrometeorology, 17, 327–344, https://doi.org/10.1175/JHM-D-14-0200.1, 2016.

UKCEH: UK Centre for Ecology & Hydrology, https://www.ceh.ac.uk/, last access: 27 June 2026.

USGS: USGS Water Data for the Nation, https://waterdata.usgs.gov/, last access: 27 June 2026.

Wade, J., David, C., Altenau, E., Collins, E., Coss, S., Cerbelaud, A., Tom, M., Durand, M., and Pavelsky, T.: MERIT-SWORD: Bidirectional Translations Between MERIT-Basins and the SWOT River Database (SWORD), https://doi.org/10.5281/ZENODO.13152826, 2023.

Wade, J., David, C. H., Altenau, E. H., Collins, E. L., Oubanas, H., Coss, S., Cerbelaud, A., Tom, M., Durand, M., and Pavelsky, T. M.: Bidirectional Translations Between Observational and Topography-Based Hydrographic Data Sets: MERIT-Basins and the SWOT River Database (SWORD), Water Resources Research, 61, e2024WR038 633, https://doi.org/10.1029/2024WR038633, 2025.

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., Groeve, T. D., Muis, S., Perez, E. C. D., Rudari, R., Trigg, M. A., and Winsemius, H. C.: Usefulness and limitations of global flood risk models, Nature Climate Change 2015 5:8, 5, 712–715, https://doi.org/10.1038/nclimate2742, 2015.

Wendi, D., Merz, B., and Marwan, N.: Assessing Hydrograph Similarity and Rare Runoff Dynamics by Cross Recurrence Plots, Water Resources Research, 55, 4704–4726, https://doi.org/10.1029/2018WR024111, 2019.

Yang, Y. and Pan, M.: Global Daily Discharge Estimation Based on Grid Long Short-Term Memory (LSTM) Model and River Routing, https://doi.org/10.5281/ZENODO.15644728, 2025.

Yang, Y., Feng, D., Beck, H. E., Hu, W., Abbas, A., Sengupta, A., Monache, L. D., Hartman, R., Lin, P., Shen, C., and Pan, M.: Global Daily Discharge Estimation Based on Grid Long Short-Term Memory (LSTM) Model and River Routing, Water Resources Research, 61, e2024WR039 764, https://doi.org/10.1029/2024WR039764, 2025.

https://doi.org/10.5194/egusphere-2026-3843
Preprint. Discussion started: 31 July 2026

Yu, L., Zhang, H., Gong, W., and Ma, X.: Validation of Mainland Water Level Elevation Products From SWOT Satellite, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 13 494–13 505, https://doi.org/10.1109/JSTARS.2024.3435363, 2024.