Remote sensing the gas exchange velocity and carbon efflux using the SWOT Satellite

CB Brinkerhoff, CJ Gleason, PA Raymond, MH Harlan

2021-01-28

## 1 Introduction

Greenhouse gas (GHG) emissions from rivers and lakes are significant (Cole et al. 2007). Rivers are usually supersaturated with dissolved GHGs and present a net flux of these gases from water to air (Cole & Caraco 2001; Jones et al. 2003). This flux is calculated, with in situ knowledge of the gas concentration gradient between the water and the air and the gas exchange velocity , as equation 1.

(1)

k is largely a function of water column turbulence, at least in fluvial settings. This is because k scales with diffusivity as normalized by the boundary-layer thickness, and in rivers turbulence defines the thickness of the boundary-layer (Hall & Ulseth, 2020 for a thorough review of gas exchange velocity in rivers). Due to its reliance on diffusivity (as well as kinematic viscosity), k is defined differently for different GHGs (Raymond et al. 2012; Wanninkof 1992). It is also often measured using different gases than those being studied, so a variant of k that is normalized to a Schmidt number of 600 () is generally used to be GHG and temperature agnostic (Hall & Ulseth, 2020). This is synonymous with a k for carbon dioxide () at 20 degrees Celsius.

It is infeasible to measure directly when working at the river network scale and so ‘upscaling functions’ are used that predict from readily available river geomorphology (e.g. O’Connor & Dobbins, 1958; Palumbo & Brown, 2014; Raymond et al. 2012). Following the theory above, many upscaling functions are based on the basic concept that k is empirically correlated with the turbulent energy dissipation e at the air-water interface (Zappa et al. 2007). This is convenient for upscaling because in rivers and streams e can be equated to the turbulent energy dissipation rate eD (Ulseth et al. 2019; Raymond et al. 2012), which is estimated using easily calculated hydraulic parameters (eD≈gSV where g is gravitational acceleration, S is channel slope, and V is average flow velocity- Tsivoglou & Neal, 1976). It has been observed that this scaling relationship is particularly strong in higher energy streams where bubble-mediated gas exchange occurs, though the general scaling function holds across all rivers and streams (Ulseth et al. 2019). Despite reasonable predictive accuracy, these upscaling models suffer from residual variation indicating large predictive error (Wallin et al. 2011). Still, these are the current best options for estimating in rivers. It has been suggested that exploring at-a-station hydraulic variation (and thus variation) is a promising approach to reducing the residual variation in upscaling models (Hall & Ulseth, 2020).

Because channel slope is readily available in any hydrographic data product, most efforts to upscale to river networks are limited by estimates of the reach-averaged flow velocity of the river channel. Therefore, network-scale GHG flux estimates exist in some part at the mercy of the quality of average flow velocity data. Ideally, a worker can directly measure velocity at some permanently or temporarily gauged river in their system and build velocity~discharge rating curves to upscale to the entire river network. Rating curves are also referred to as hydraulic geometry (HG) models: simple power law relations between hydraulic properties and discharge (Leopold & Maddock, 1953). However, this is not feasible in ungauged basins (whether physically or politically ungauged- Gleason & Durand, 2020) where no in situ hydraulic information is available and fieldwork is impractical. Thus, in ungauged settings, present understandings of GHG fluxes from rivers are likely limited by our access to velocity and/or discharge data.

To circumvent this problem in ungauged basins/at massive spatial scales, researchers have used ‘global-scope’ HG models that use a single set of model parameters trained on large hydraulics datasets. This is the approach used in upscaling average flow velocity to continental/global river networks and has proven quite useful in those settings (e.g. Horgby et al. 2019; Borges et al. 2015; Lauerwald et al. 2015; Raymond et al. 2013). However, it has long been established that HG parameters exhibit high spatial variability across landscapes and vary significantly from river to river (Park, 1977). Because of this, a large body of geomorphology work has attempted to parse out process-based explanations for HG parameters (e.g. Dingman 2007; Parker, et al. 2007; Singh, 2003; Ferguson, 1986) and an open research question is how to best prescribe a set of HG parameters to a given river (particularly at the global scale). A potential alternative to this approach, in ungauged basins at least, is to directly estimate a river’s hydraulic properties from remote sensing (RS) data.

The problem of hydraulic estimation in ungauged basins is itself a burgeoning subfield within fluvial hydrology and an analogue to the problem presented here is the ungauged remote sensing of river discharge (RSQ- see Gleason & Durand, 2020 for a thorough review). A specific subset of these ungauged RSQ methods are termed ‘McFLIs’ or ‘Mass Conserved Flow Law Inversion’ algorithms (e.g. Andreadis et al. 2020; Brinkerhoff et al. 2020a; Durand et al. 2014; Garambois & Monnier, 2015; Hagemann et al. 2017). They use basic geomorphic theories, rather than full hydrologic and/or hydraulic models, and the concept of ‘prior river knowledge’ to estimate discharge from RS data where not all hydraulic terms are RS-able (Gleason & Durand, 2020; Gleason et al. 2017). McFLIs have largely been developed in the context of the NASA/CNES/UKSA/CSA Surface Water and Ocean Topography (SWOT) mission, which will launch in 2022 and provide the world’s first global measurements of water surface extent and elevation at novel temporal resolutions (Biancamaria et al. 2016). On average SWOT will sample rivers every 11 days and provide a unique opportunity to measure river and lake hydraulics at the global scale. While the raw performance of McFLIs is lower than a more sophisticated approach, they are globally flexible, easily implementable in any river that SWOT can sample, and improve our understandings of ungauged rivers when little to no information was previously available (Brinkerhoff et al. 2020a; Durand et al. 2016). Recently, McFLIs have also shown promise in providing additional and beneficial information to traditional hydrologic modeling via data assimilation (Ishitsuka et al. 2020).

McFLIs often employ Bayesian inference. Bayesian techniques have been previously used to estimate concurrently with stream metabolism from dissolved oxygen (O2) datasets (Appling et al. 2018; Grace et al. 2015; Holtgrieve et al. 2010). While this approach requires detailed field data and is thus not applicable to estimation in ungauged rivers, it does suggest that Bayesian inference could be useful for estimating when parameter equifinality is a problem (at it is with McFLIs).

Within this context, we sought to explore estimating in ungauged basins with no in situ information. We hypothesize that remote sensing is possible by combining 1) hydraulic geometry theory, 2) upscaling theory, 3) Bayesian inference (in a manner similar to McFLI), and 4) SWOT data. More specifically, this chapter aims to answer the following questions: 1) Is upscaling the gas exchange velocity with no a priori knowledge of eD possible? 2) Is ungauged RS of gas exchange velocity possible? 3) How might this algorithm influence upscaling of riverine carbon efflux?

To answer these questions, we build a RS algorithm that ingests SWOT data and produces estimates (with fully quantified uncertainty within a Bayesian framework) that requires no in situ information about a river’s hydraulic properties. We validate the algorithm on simulated SWOT data for 22 rivers. Finally, we compare our algorithm against established methods: we use previously published carbon dioxide () data to compare the bulk carbon efflux (via evasion) from the 22 rivers as calculated using both our new algorithm and established velocity rating curves and in situ discharge records.

## 2 Methods

We name the RS of algorithm BIKER, or the ‘Bayesian Inversion/Inference of the Rate’ algorithm. To build BIKER, two steps are taken: first, we develop a upscaling model that does not rely on a priori knowledge of eD (section 2.1) and then we use that upscaling model to build the actual RS algorithm (section 2.2). Following these methods, the validation setup is described (section 2.3), as is the comparison of estimated bulk carbon effluxes from a suite of average flow velocity models (section 2.4). A flowchart detailing the complete workflow is included (Figure 1).

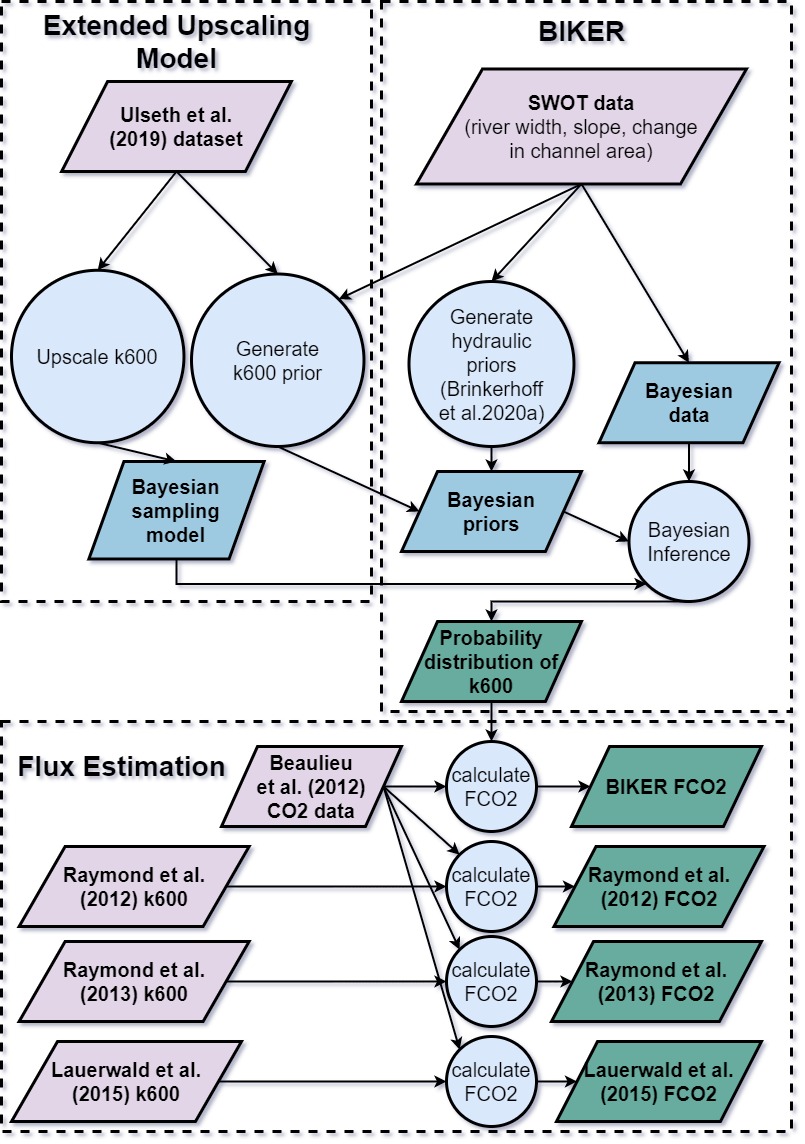


Figure 1. Flowchart of methodology developed in this chapter. ‘Extended Upscaling Model’ is detailed in section 2.1. ‘BIKER’ is detailed in section 2.2 and ‘Flux Estimation’ is detailed in section 2.4. See section 2.3 for the validation setup.

### 2.1 Developing an RS-able upscaling model

Raymond et al. (2012) first used river hydraulics and the theoretical relationship between the turbulent dissipation rate (eD) and gas exchange to upscale from average flow velocity and channel slope, extending earlier work predicting the reaeration coefficient (e.g. O’Connor & Dobbins, 1958; Tsivoglou & Neal, 1976). Later, Ulseth et al. (2019) built upon Raymond et al. (2012)’s work by expanding their dataset to include measurements from steeper rivers and finding that two distinct upscaling regimes exist in low and high energy rivers. More specifically, Ulseth et al. (2019) scale using eD, where the resulting upscaling model parameters are significantly different whether eD is high or low. Their upscaling function is reprinted as equation 2.

(2)

While equation 2 would be ideal for the purposes outlined for this manuscript and provides best-to-date predictive accuracy for upscaling , it requires an a priori knowledge of eD to assign its differential model parameters. We do not have this a priori knowledge in an ungauged setting and only have access to SWOT-observed river widths and heights (and therefore water surface slopes). Thus, we need to extend the Ulseth et al. (2019) model to assign the two model parameters using only RS-able hydraulics. To do this, we use the basic premise that the scaling relationship is different for different river sizes. We use river width as a proxy for river size and train different upscaling functions for different groups of data with similar river sizes (Figure 2, trained on the Ulseth et al. 2019 dataset). In doing so, we find that narrow rivers (< 10m wide) need to be further broken into low energy (slope < 0.05) and high energy (slope > 0.05) regimes in order to meet the assumptions necessary to fit a linear regression model (specifically, normally distributed residuals of the predictions).

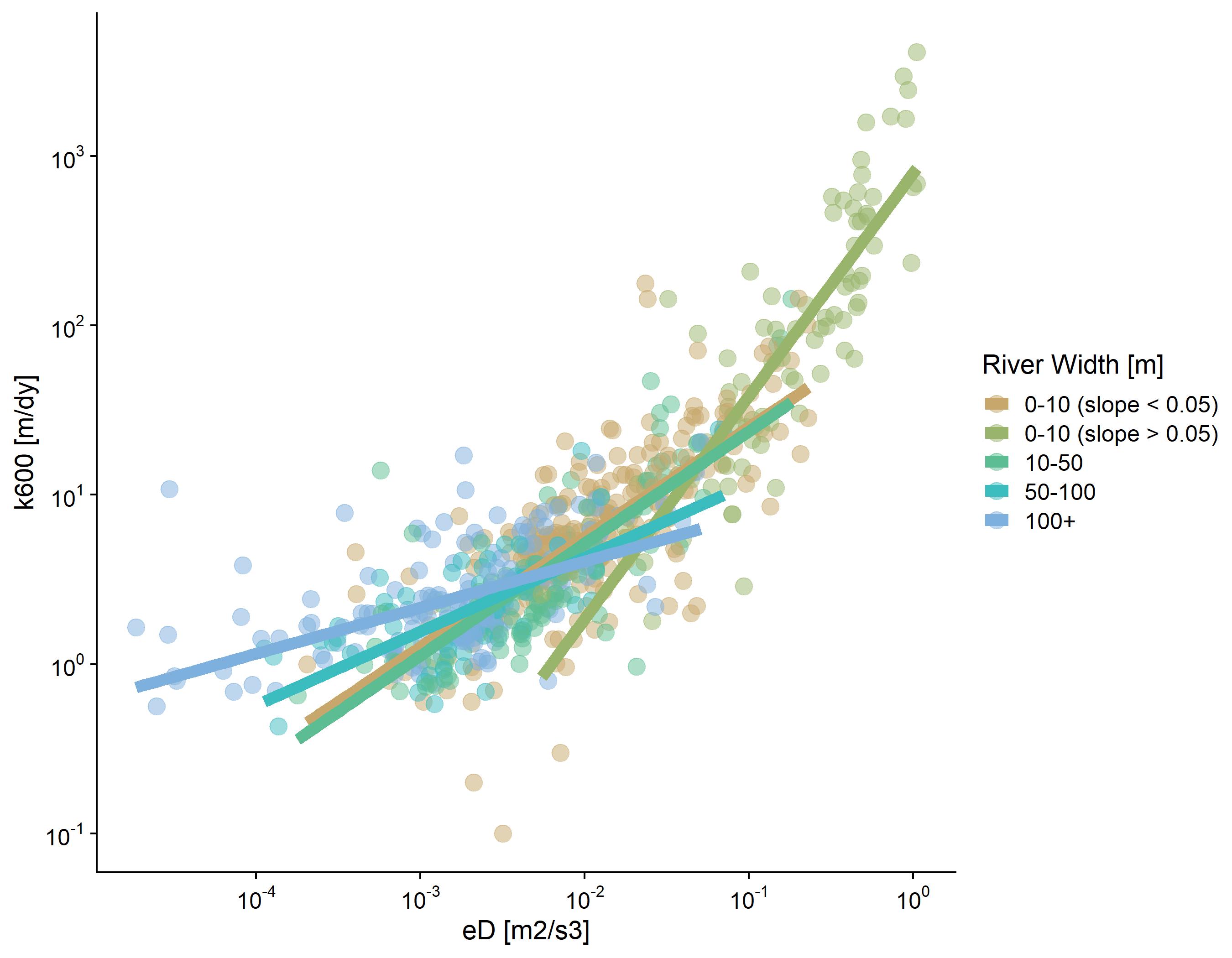


Figure 2: Theoretical basis for the RS-able extension of the Ulseth et al. (2019) k600 upscaling model. Data from Ulseth et al. (2019).

We combined these five regression models into a ‘rule-based regression model’ (equation 3) for river size i and model parameters a and b. Per Figure 2, there are five possible sets of a and b parameters, as defined for rivers 10-50m wide, 50-100m wide, 100+m wide, low energy rivers 0-10m wide (slope < 0.05), and high energy rivers 0-10m wide (slope > 0.05).

(3)

This model therefore assigns ai and bi a priori using just river width and slope to define i. Another implication of this extension of the Ulseth et al. (2019) model is that we now use a and b parameters that implicitly reflect the distinct geomorphology of different sized rivers. Finally, we tested whether this model can reasonably estimate relative to the original Ulseth et al. (2019) model, with the goal of at least replicating their performance. We compared it against the various eD-based upscaling models from Raymond et al. (2012) that are commonly used in the literature. This was done by training all upscaling functions on 80% of the Ulseth et al. (2019) data and comparing their performance on the remaining 20% of the data. Because the Raymond et al. (2012) models were originally trained on a subset of this dataset, we deemed it fairer to refit their parameters to this expanded dataset (but keep model structure identical).

### 2.2 Developing the BIKER algorithm

With a validated upscaling model that requires no a priori knowledge of eD (results presented in section 3.1) we turn to developing the algorithm to RS . The approach used here is heavily informed by Hagemann et al. (2017)’s algorithm for ungauged RSQ via Bayesian inference. Consult that paper for the logic behind conceptualizing the equation for discharge as a Bayesian inference problem, which we largely follow here to conceptualize upscaling as a Bayesian inference problem.

SWOT will provide river width and height (and thus water surface slope ) at nodes/cross-sections along a set of river reaches. Thus, our ‘data’ for the Bayesian inference are river width and water surface slope while all other terms are treated as statistical parameters. We also obtain a ‘change in channel area’ from the river height and width and assuming a rectangular channel geometry.

In order to use Bayesian inference to solve for , we need to derive a sampling model/likelihood function. Following Tsivoglou & Neal (1976), for unsteady flow conditions (channel slope - see text S1 for this model derivation and why assuming non-uniform flow is convenient for our application). Using this, as well as Manning’s equation to relate average flow velocity to channel geometry, we rewrite equation 3 as equation 4. In equation 4, we use the Manning’s formulation from Hagemann et al. (2017) which formulates depth via river width and area (in order to use SWOT-observables). is now defined as a function of , channel roughness n , channel width W [L], and channel area for the change in channel area dA). It is also defined by the upscaling parameters and .

(4)

Equation 4 is then reworked into a Bayesian sampling model with the ‘data’ on the left-hand side sampled from the unknown model parameters on the right-hand side (equation 5). Note that in the formal likelihood specification each term in equation 5 is log-transformed such that equation 5 is a lognormal distribution, however they are written here without a transformation for succinctness. refers to the uncertainty inherent in the estimates themselves (equation 4).

(5)

Equation five necessitates that Bayesian priors be assigned to model parameters n, , and . Bayesian priors formularize the a priori estimates (and uncertainties) for these parameters. Or more intuitively, they represent our ‘prior river knowledge’ of what n, A0, and probably are for some river given that they cannot be remotely sensed. All priors are formalized within the model as truncated normal distributions of the ln-transformed terms such that for λ-4.634$

Finally, we estimated using Monte Carlo (MC) methods to approximate total uncertainty from equation 4. Uncertainty in equation 4 can stem from two sources: 1) error baked into the upscaling model (ai and bi parameter uncertainty) and 2) error in Manning’s approximation of the average flow velocity. To do this, we ran MC simulations on 8,000 sets of field measurements of river channel hydraulics from the dataset built by Brinkerhoff et al. (2019). Each MC simulation was itself 10,000 runs, sampling from the normal distributions for , , and ln-transformed velocity. Estimate uncertainty for velocity followed Hagemann et al. (2017), who estimated ln-transformed Manning’s equation error to be 0.25. This resulted in 8,000 discrete distributions for from which the average uncertainty term is extracted and used for .

With the sampling model (equation 5) and prior distributions described, a joint posterior distribution conditional on the SWOT data is derived. To approximate this distribution, BIKER uses a Markov Chain Monte Carlo (MCMC) algorithm. Because it is written in the Stan probabilistic programming language, BIKER’s MCMC of choice is a Hamiltonian Monte Carlo which reduces computation time relative to other MCMC approaches (<https://mc-stan.org/docs/2_26/reference-manual/hamiltonian-monte-carlo.html>).

### 2.3 BIKER validation setup

We validated the BIKER algorithm on 22 rivers using observed average flow velocity and observed slope. First, we detail the 22 rivers, and then we detail how ‘observed’ is calculated.

Because SWOT has yet to launch, it is standard practice to benchmark SWOT-related algorithms on ‘SWOT-like data’. There are three current types of SWOT-like data: 1) AirSWOT, which is an airborne Ka-band inSAR currently limited to five rivers globally, 2) simulated rivers that mimic the type of data SWOT will provide, and 3) the SWOT simulator, which introduces errors to these simulated rivers (mostly from radar layover error and random noise- Oubanas et al. 2018a). Because we are principally interested in algorithm performance, we limit our validation setup to simulated rivers in order to benchmark across as many rivers as possible. These simulated rivers are simply reach-averaged hydraulic model outputs where the water surface heights and widths are labelled as ‘RS observations’ and are used as the sole inputs to the BIKER algorithm. Simulated rivers mimic perfect measurement conditions (i.e. no observation errors) and represent the best-case scenarios of what SWOT will provide to hydrologists. Here, we use the 22 rivers archived in Frasson et al. (2019) and used by Rodriquez et al. (2020) to develop reach-averaged Saint Vernant equations within the SWOT context. We omitted the Arial Khan river in Bangladesh due to known problems with that hydraulic model and we further sampled the models for only the observations every 11 days to mimic the average SWOT overpass frequency (this also approximately algins the SWOT observations with the bi-weekly CO2 data used later- Section 2.4). All told, this yielded 503 sets of SWOT observations to validate BIKER with.

Regardless of the validation data used, we do not have observed data for these rivers, and to our knowledge no field dataset of exists in the type of temporal and spatial frequency that SWOT (and therefore the BIKER algorithm) provides. Further, we are principally interested in our ability to upscale from RS observations to avoid relying on velocity rating curves. We are less concerend with the actual accuracy of the upscaling model itself (which can be validated using existing datasets). Therefore, we take the best performing upscaling model from sections 2.1 and 3.1 and use that to calculate the ‘observed’ that BIKER is validated against. This is calculated using equation 7, where is observed river discharge divided by observed channel area. With this setup, we can directly explore our ability to infer observed average flow velocity and from river width and height alone. It also means that, for a fair validation scheme, must be set to reflect only error from Manning’s equation (and not and ). Thus, is set to 0.25 for this validation. However in practice, it should reflect the total uncertainty calculated in section 2.2.

(7)

Validation metrics take two forms (and are detailed in Table 1). To validate across all rivers and timesteps, we used the coefficient of determination r2 and the root mean square error (RMSE). Four normalized metrics were used for by-river validation to compare across rivers: RRMSE and NRMSE are normalized root mean square errors that have been normalized by the observed value and the mean observed value, respectively, to compare across rivers. rBIAS is a measure of prediction bias that is normalized by the mean observed value to compare across rivers. KGE is a standard metric used in streamflow prediction with an intuitive basis: a value greater than -0.41 indicates a model outperforms a uniform prediction of the mean value (Knoben et al. 2019) and a value greater than 0 is often interpreted as a useful prediction in ungauged settings.

*Table 1: Validation metrics used in this study, where r is the correlation coefficient, Nt is number of observations and i is the specific observation. σ refers to the variance of the sample and μ refers to the mean of the sample.* |Description|Acronym|Definition|Ideal Score|Possible Range|Validation Scheme| |-----------|-------|----------|-----------|--------------|-----------------| |Coefficient of determination|| Eq|1|0 to 1| All rivers and all timesteps| |Root-mean-square-error|RMSE| Eq|0|0 to ∞|All rivers and all timesteps| |Relative root-mean-square error|RRMSE|Eq |0|0 to ∞|By river| |Normalized root-mean-square error|NRMSE|Eq|0|0 to ∞|By river| |Relative bias|rBIAS|Eq|0|-∞ to ∞|By river| |Kling-Gupta efficiency|KGE|Eq|1|-∞ to 1| By river|

### 2.4 Upscaling to CO2 evasion and bulk carbon efflux

It is one thing to be able to predict , but researchers are usually more interested in the gas fluxes from rivers and ultimately the bulk carbon efflux and not just the gas exchange velocity. Therefore, we explored 1) the ability of BIKER to reproduce hypothetical carbon dioxide evasion fluxes () from these 22 rivers, and 2) the sensitivity of the estimated bulk carbon efflux (via evasion) to the method used to estimate average flow velocity.

First, we calculated . To do this, we paired the modeled values (obtained from the estimates) with field-measurements of water-side concentrations. 29 bi-weekly samples were made by Beaulieu et al. (2012- Figure S1) at one location in the Ohio River. We paired these 29 values with each 11-day-repeat set of SWOT observations, ignoring the timesteps beyond 29 (only 7% of the SWOT observations were ignored here and this was deemed acceptable). This data is for the Ohio River only but was applied to all 22 rivers (which includes the Ohio River). Because we are exclusively interested in the relative differences between estimates and not the raw fluxes themselves, any data representative of SWOT-observable rivers was deemed acceptable. Atmospheric was assumed 390 uatm and water temperature was assumed 25 degrees Celsius. The Schmidt number, used to obtain kCO2 from , was calculated following Raymond et al. (2012) and Wanninkof et al. (1992).

Then, we calculated bulk carbon efflux using four models for average flow velocity: BIKER and three rating curves previously used for upscaling (Lauerwald et al. 2015; Raymond et al. 2013; Raymond et al. 2012- see Table S1 for their definitions). It is worth stressing that the Lauerwald, et al. (2015) model was developed as one of two components of the Raymond et al. (2013) velocity model, and so was actually developed by the latter’s authors. The names used here refer to the final implementation used in each study’s upscaling exercises. For this study, all rating-curve methods use the in-situ discharge record while BIKER does not. This allows us to explore the relative differences in bulk carbon efflux estimates if implementing a wholly ungauged method versus two gauged methods (that also use global-scope HG parameters). Finally, we express the bulk carbon efflux (via evasion) [M/T] as the average mass flow rate of carbon per year, i.e. multiplied by the surface area of all 22 rivers.

## 3 Results

First, we present the results from the upscaling experiment (section 3.1). We then take the best performing upscaling function, implement it within BIKER, and validate it on the 22 SWOT rivers (section 3.2). Finally, we use as modeled by both BIKER and gauge-based rating curves to compare bulk carbon efflux estimates from the 22 rivers (section 3.3).

### 3.1 Validating the upscaling model

Figure 3 plots the validations for the five upscaling functions tested. Both the original Ulseth et al. (2019) model, and our RS-able extension, perform the best on this independent set of validation data, with near identical performance. The RS-able Ulseth et al. (2019) model achieves an of 0.83 and an RMSE of 1.98 m/day, which is just marginally better than the original Ulseth et al. (2019) model (0.82 and 2.03 m/day, respectively). The two power-law based Raymond models (Raymond et al. (2012) equations 4 and 3) achieve slightly worse performance, with in the 0.70s and their RMSE scores slightly higher (2.31 and 2.29 m/day, respectively). Raymond et al. (2012) equation 5, which is linear in structure, worked well on their dataset but fails to capture the non-linear relationship evident in this expanded and more geomorphically diverse dataset. In fact, it predicts negative values and error metrics cannot be calculated because the prediction residuals are not normally distributed. Because we successfully reproduced the Ulseth et al. (2019) model, the rest of this manuscript deals exclusively with this RS-able variant of the model.

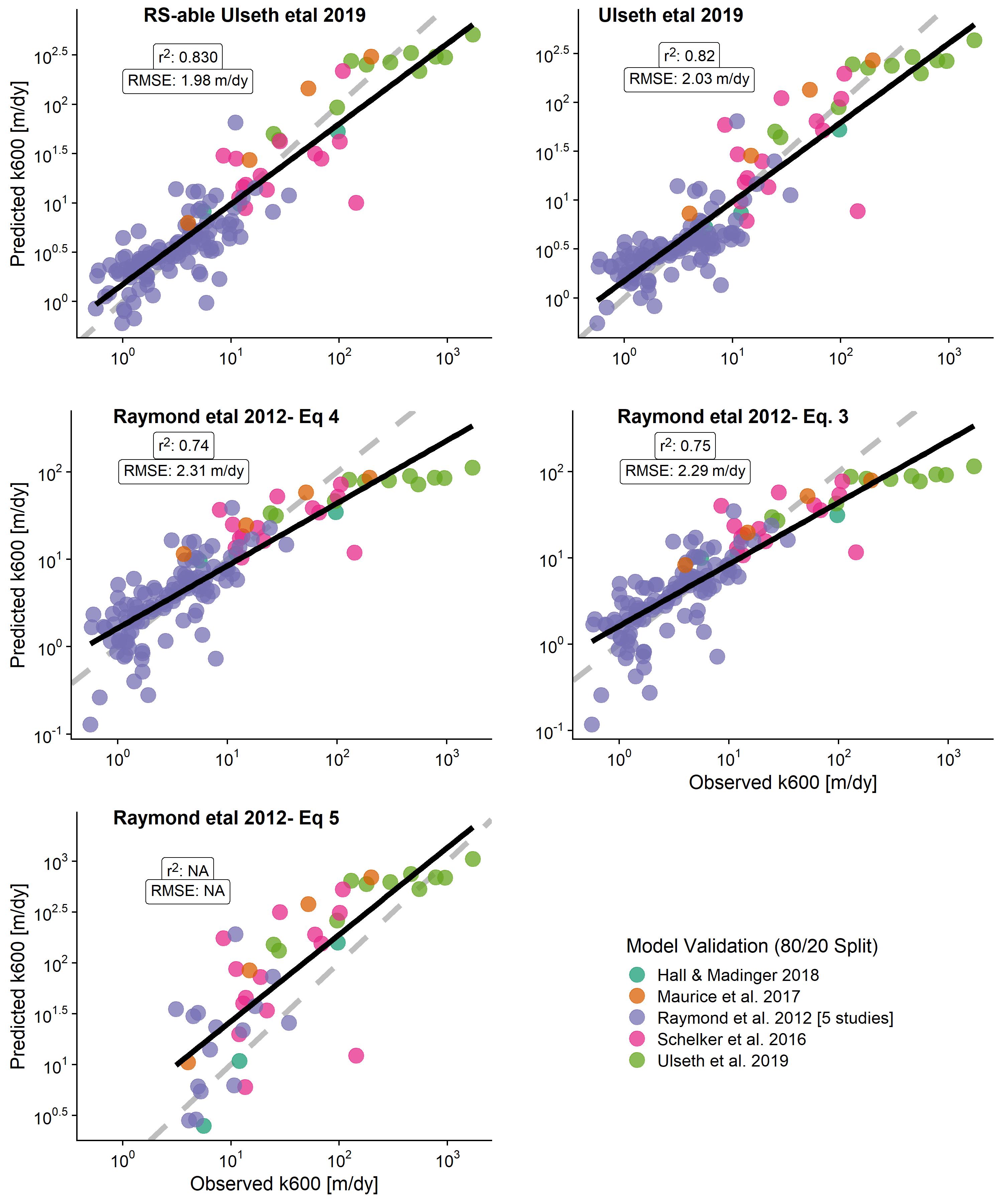


Figure 3. Validation of five k600 upscaling models on 20% of the Ulseth et al (2019) dataset, withheld for independent testing. Raymond et al. (2012) Eq. 5 predicts negative k600 values, which are not plotted here and highlight a limitation of using a linear model for a non-linear problem (though it did work well on their smaller dataset).

Uncertainties were then propagated through the RS-able Ulseth et l. (2019) model via MC simulations (section 2.2) for the 8,000 hydraulic measurements mapped in Figure 4c. Figure 4a plots a histogram of the 8,000 uncertainty terms extracted from those distributions (Figure 4b plots three of the distributions used to build Figure 4a). Across the 8,000 tests, we quantified average ln-transformed uncertainty in equation 4 to be 1.31 (~3.72 m/day- dashed blue line in Figure S2a). Uncertainty solely from the upscaling model is 1.30, confirming that using Manning’s equation to approximate average flow velocity is introducing very little additional uncertainty to the model (Figure S2). This also confirms the large residuals inherent in upscaling models as described by Hall & Ulseth, (2020).

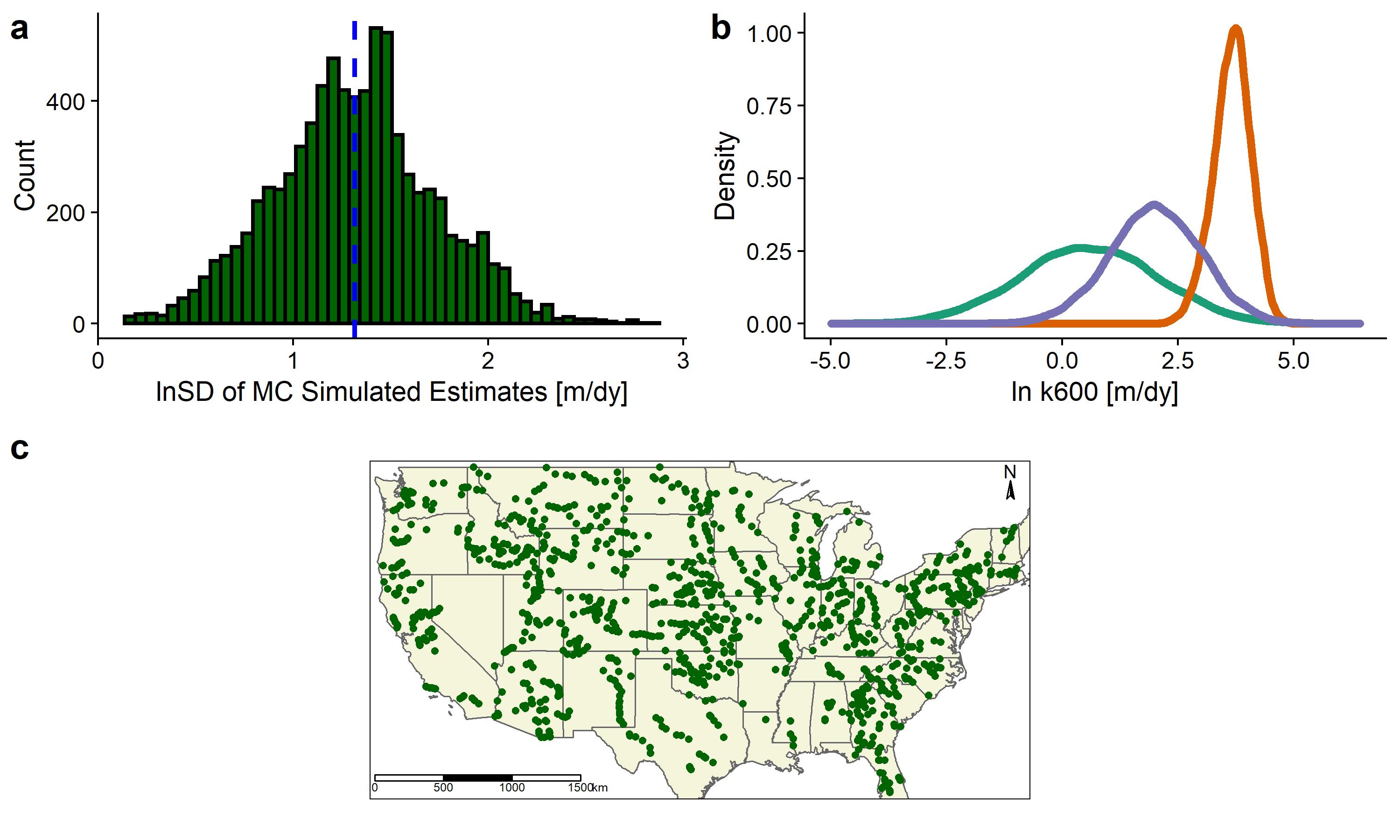


Figure 4 Model uncertainty estimates: a: histogram of SDs for all 8,000 MC simulations with blue dashed line denoting the mean (1.32). b: Three example MC simulation results of 10,000 samples each. c: Map of the 8,000 sets of hydraulic measurements used for this analysis.

### 3.2 Validating the BIKER algorithm

Figure 5a plots the validation results for across all 22 rivers and all timesteps. For BIKER, the points are the posterior means while the black lines are the 95% confidence intervals (CIs) for the predictions. is strongly correlated with the predicted by BIKER, capturing the general magnitude of the predictions and with most points falling on or near the 1:1 line. However, the 95% prediction intervals (dashed grey lines) highlight relatively large residuals for many predictions. Regardless, the RMSE for the BIKER predictions is only 1.38 m/day across all predictions. This is substantially less than the RMSEs for the upscaling models (Figure 3) and the MC-simulated uncertainties (Figure 4a), indicating that recovering upscaled from SWOT data is much more accurate than the upscaling process itself (which is logical).

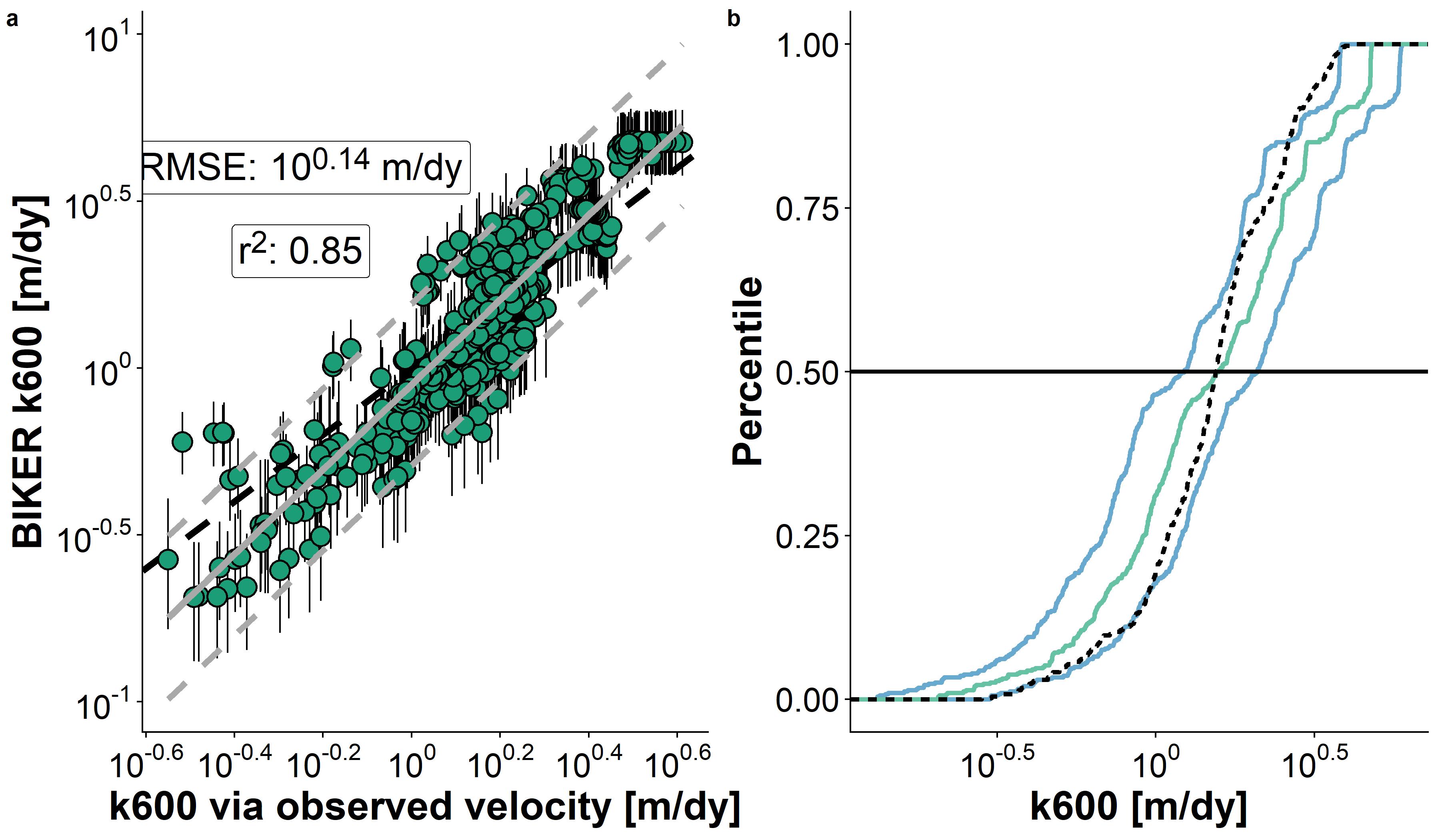


Figure 5. a: Validation of remote sensing algorithm for 22 rivers. Black bars are 95% CIs for the modeled values. Grey line is linear regression (and 95% prediction intervals are dashed) and dashed black line is 1:1 line. b: Cumulative density functions (CDFs) of the same results: dashed black line are observed values, green line are BIKER posterior means, and blue lines are BIKER 95% CIs.

Figure 5a suggests a flip in the bias of the predictions, which is confirmed in Figure 5b. Figure 5b plots the cumulative density functions (CDFs) of observed and predicted where the green line is the BIKER posterior mean and the blue lines are the 95% CIs for the BIKER posterior. We see, more clearly than in Figure 5a, that the posterior mean is systematically underestimated below the median values and systemically overestimated above the median However, the uncertainty in these estimates captures very well: nearly the entire observed CDF falls between, on, or just outside of the 95% CIs (Figure 5b). This highlights the benefit of using a Bayesian predictive framework to fully propagate all parameter uncertainties through to the posterior. While these CIs are large and thus BIKER estimates are reasonably uncertain, we do improve upon our baseline understanding of in these rivers. Put another way, we reasonably capture with no in situ information about the river while simultaneously and explicitly accounting for the large uncertainties inherent in our estimates.

Figure 6a plots validation metrics calculated for each river (i.e. the boxplots are composed of metric scores for the 22 rivers- see Table 1 for metric definitions). Median rBIAS is 0.02, indicating nearly no bias in most rivers’ predictions. However, some rivers are substantially biased in both directions, further supporting the visual evidence in Figure 5 that sometimes BIKER is substantially under/overestimating and that this is river-specific. Median KGE is 0.53, which is excellent given that absolutely no in situ information is being used to predict . NRMSE and RMSE have median scores of 0.31 and 0.29, respectively. While median KGE and rBIAS scores were strong, the ranges of these scores were somewhat large (standard deviation for KGE of 0.53 and for rBIAS of 0.34). This suggests that BIKER predicts very well in some rivers, but not well in other rivers.

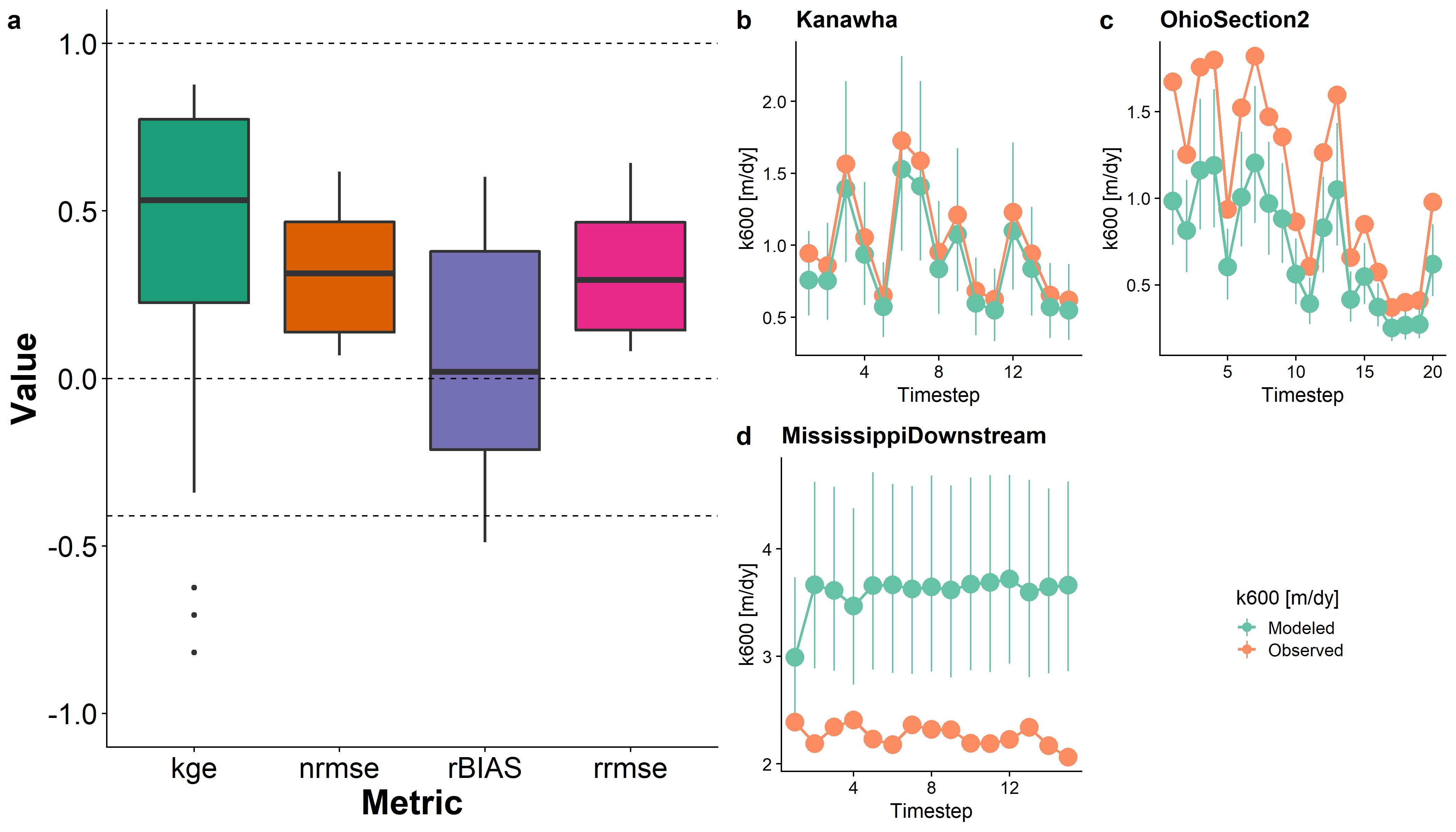


Figure 6. a: Performance metrics by river. See Table 1 for metric definitions. Dashed lines denote scores of 1, 0, and -0.41 for KGE (Knoben et al. 2019). b-d: validation timeseries for three rivers representative of good, reasonable, and poor performance: b was randomly selected from upper tertile of KGE scores, c was randomly selected from the middle tertile, and d from the worst tertile. Model results include the posterior means and 95% CIs.

Figure 6b-d are representative timeseries plots of predicted and observed for three rivers chosen randomly from those with ‘good’ KGE scores (b), ‘okay’ KGE scores (c), and ‘bad’ KGE scores (d). See the Figure 5.6 caption for how this was calculated. For the Kanawha River, the entire timeseries of is correctly predicted, while in the Ohio Section 2 River there is a negative bias in the estimates. However, temporal dynamics are still correctly recovered. In the downstream Mississippi River, there is significant bias in the estimates as well as significant uncertainty (per the 95% CIs). The temporal dynamics are also quite off.

### 3.3 Implications for carbon cycle estimates

Finally, we explore our ability to 1) use BIKER-produced to estimate from rivers and then 2) to estimate bulk carbon efflux from evasion (section 2.4).

In Figure 7a, there is a strong fit to the observed data, with an RMSE of 1.41 . The is slightly lower than that for (Figure 7a) due to some abnormally large residuals on a handful of points. There is also a slight underestimation bias for the lowest of the observed , though this is smaller than the bias in the predictions (Figure 5a). prediction intervals also have a similar relative magnitude to those presented in Figure 5. Figure 7b-d includes subplots for the same rivers as Figure 4b-d, however with instead of There is functionally perfect recovery of in the Kanawha River, very good recovery (with some positive bias in peak evasion events) for the Ohio Section 2 River, and substantial negative bias in the Downstream Mississippi river. Again, the temporal dynamics are accurately modeled however there is some bias in the magnitude of the predictions. However, is so low in the Mississippi river (Figure 7d) that the absolute bias is not as large as it looks in these timeseries plots. This suggests that the BIKER model can easily estimate the changing dynamics of evasion but is at the mercy of the prior on to accurately estimate the magnitude of the evasion.

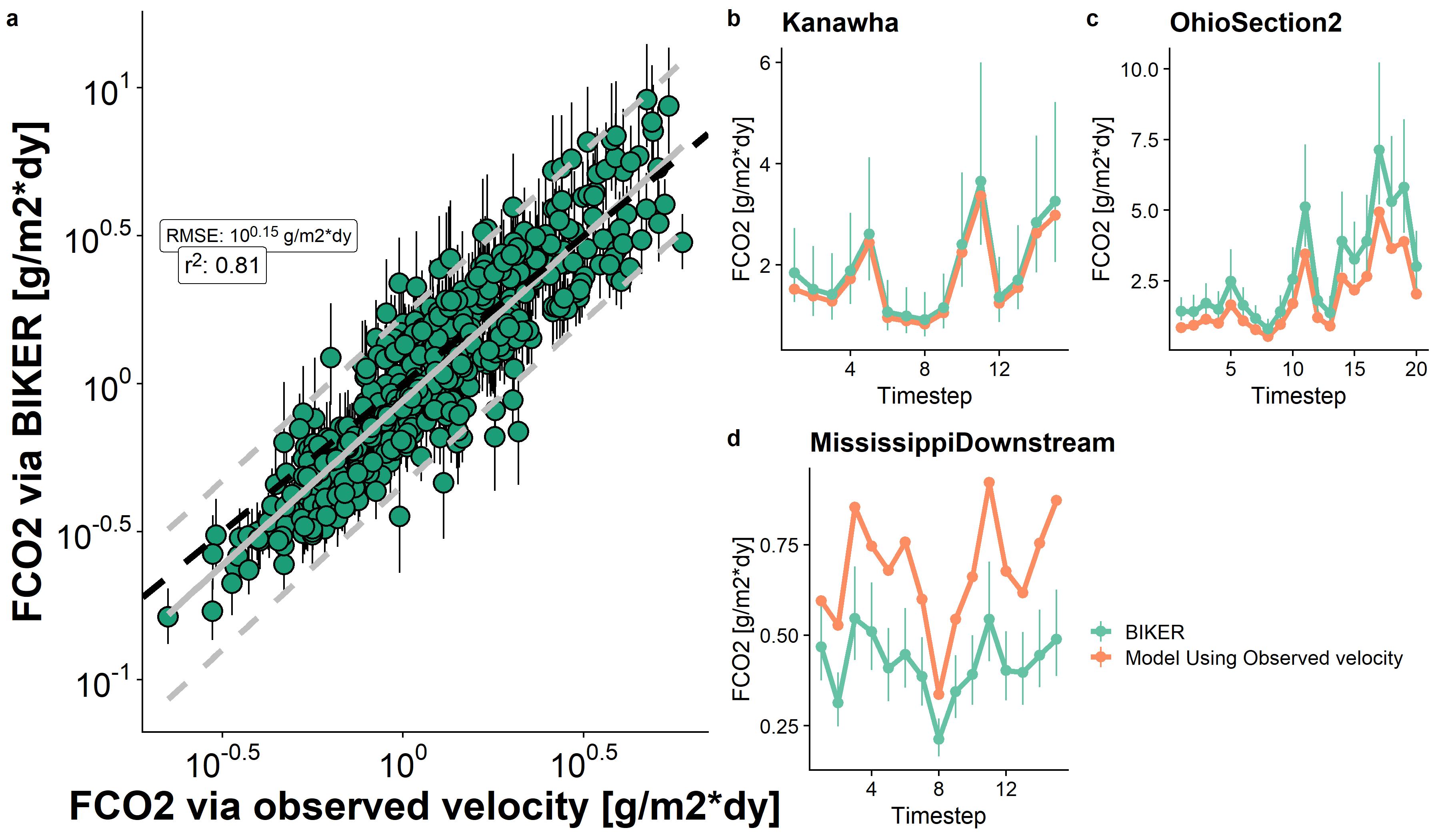


Figure 7. a: FCO2 from BIKER versus using observed average flow velocity for all timesteps for 22 SWOT rivers (grey lines are linear regression and 95% prediction intervals, while black dashed line is the 1:1 line). b-d: timeseries plots for three example rivers from Figure 5.6.

Compared against Figures 5a and 6b-d, there is an arguably stronger recovery of FCO2 than (Figure 7). This is presumably due to the structure of equation 1, of which is a function of both the observed concentration gradients and This means that the errors in the BIKER estimates are diminished in relative importance when multiplied by the concentration gradients.

Finally, we compare the bulk carbon efflux (via CO2 evasion) from the 22 rivers using BIKER posterior means and three gauge-based rating curves (Figure 8). In Figure 5.8a are barplots of the bulk carbon efflux across all rivers in gigagrams of carbon per year. The BIKER bulk carbon efflux (1113 gG-C/yr) is nearly identical to the bulk carbon efflux calculated using observed velocity (1119 gG-C/yr) while the rating curves underestimate the efflux (799, 837, and 930 gG-C/yr for ‘Raymond 2012’, ‘Raymond 2013’, and ‘Lauerwald 2015’, respectively). BIKER functionally perfectly captures the estimated efflux from these rivers purely from RS data. This is borne out in the cumulative density functions (CDFs) in Figure 8b, which plot the distribution of all FCO2 estimates across all rivers and timesteps for each velocity model. BIKER nearly perfectly captures FCO2 estimates above approximately the 60th percentile, while the Lauerwald, et al. (2015) model nearly perfectly captures FCO2 below the 60th percentile. Because the bulk efflux is more sensitive to peak evasion events, at the bulk properties level BIKER more accurately models carbon efflux (Figure 8a) than the rating curves do. The Raymond et al. (2012) model, which was mostly trained on small streams, uniformly underestimates FCO2 and thus bulk carbon efflux. The Raymond et al. (2013) model, which is an average of the other two rating curve estimates (Table S1), is in between.

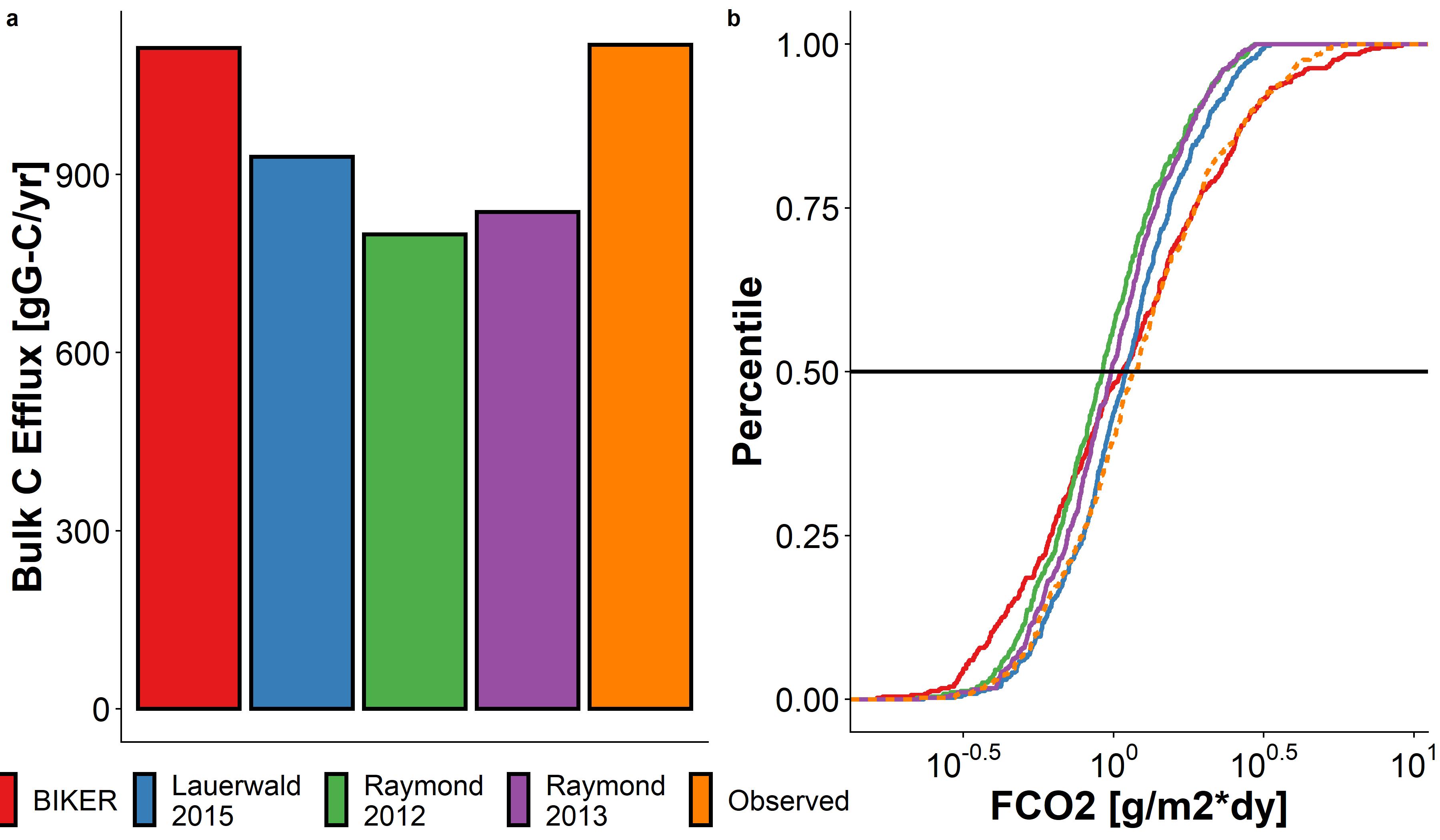


Figure 8. a: Bulk carbon efflux, per year, from the 22 SWOT rivers as calculated using five different average flow velocity estimates: 1) BIKER, 2) through 4) average flow velocity as calculated using rating curves from the literature, and 5) observed average flow velocity. b: Cumulative density functions (CDFs) of all FCO2 estimates across all timesteps and rivers for the same four velocity models. The median of the distributions is noted with a black line.

By-river performance scores for (Figure S3) show the Lauerwald et al. (2015) model as the best performing in KGE, NRMSE and RRMSE, while BIKER arguably has the least prediction bias (rBIAS) of the models and is approximately the same in KGE as the Raymond et al. (2013) model. Again, the high performance of the Lauerwald et al. (2015) and Raymond et al. (2013) models makes sense as the rating curves rely on the in-situ discharge record. Across all four metrics, the Raymond et al. (2012) model is the worst, performing marginally worse than BIKER.

## 4 Discussion

### 4.1 Ungauged estimation of gas exchange velocity

BIKER is therefore successful at 1) reproducing without any in situ river information for (Figure 7) and 2) nearly perfectly reproducing the bulk carbon efflux without any in situ information for (Figure 5.8). It is important to stress that, unlike BIKER, the rating curves in Figure 8 rely on the observed discharge record. This means that Figure 8 represents the best performance that the rating curves could ever have; if ran using modeled discharge their accuracy would necessarily decrease. However, this decrease would not be much due to the high accuracy of those data. Thus, these results suggest that BIKER will be useful in two settings: upscaling in ungauged rivers as hypothesized, but also in potentially improving our bulk efflux understandings at gauged sites too. Future work should systematically quantify prediction error from using both global-scope rating curves and modeled discharge, like is the default approach used to date in upscaling studies.

However, Figures 8 and S3 still confirm that rating curves with a streamgauge are preferred when possible, particularly when interested in properties finer than the bulk efflux (Figure S3). Further, methods like the Raymond et al. (2013) and Lauerwald et al. (2015) models are implementable in rivers of any size, while BIKER is limited to the largest rivers on Earth due to SWOT’s coarse spatial resolution. However, our results suggest that, upon SWOT's launch, a BIKER-style approach to estimating gas exchange could be coupled with existing upscaling workflows to improve gas flux predictions where gauges are unavailable.

### 4.2 Exploring spatiotemporal dynamics of

To date, most field-scale studies of riverine gas exchange have focused on 1) its relationship with wind speed (e.g. Beaulieu et al. 2012; Borges et al. 2004; Zappa et al. 2007), 2) average flow velocity (e.g. Alin et al. 2011; Beaulieu et al 2012; Schelker et al. 2016), 3) discharge (e.g. Roberts et al. 2007; Uehlinger & Naegeli, 1998), or 4) oxygen diel curves (where gas exchange is still assumed constant in time- e.g. Appling et al. 2018; Demars et al. 2015). This leaves the spatiotemporal dynamics of gas exchange weakly constrained. A few studies have investigated these dynamics, but these studies have been limited to individual rivers and/or limited field seasons (e.g. Hall et al. 2012; Sand-Jensen & Staher, 2012). Wallin et al. 2011 performed a preliminary analysis in northern Sweden relating river specific (‘at-a-station’) temporal variability in gas exchange with channel slope, but they were limited to an average of only 8 measurements per river in a single watershed. Thus, there is a fundamental knowledge gap in understandings the geomorphic and hydraulic drivers of riverine gas exchange. This limitation is due both to a lack of process-level understanding (Hall & Ulseth, 2020) but also due to a lack of measurements. Authors have argued that the key to explaining the large residual variation in upscaling models is to explore at-a-station temporal variability in (Hall & Ulseth, 2020).

Therefore, to explore the spatiotemporal dynamics of at fine temporal resolution and but also at the global-scale, SWOT data is an attractive option for upscaling (providing daily hydraulic measurements for a 3 month fast sampling period for calibration and validation and sampling thereafter between 1 and 7 days per 21 day repeat cycle thereafter- Biancarma, et al. 2016). BIKER's success on simulated SWOT rivers bodes well for its implementation on real SWOT data, and functionally opens the door for daily estimation of riverine gas exchange globally once SWOT launches.

### 4.3 Biases in upscaling k600

Systematic sampling biases in the data used to develop upscaling models (Ulseth et al. 2019; Raymond et al. 2012).

Finally, in reasonably wide rivers, wind begins to exert a non-trivial influence on gas exchange. It is well established that in lakes and the ocean, wind controls near-surface turbulence, and thus gas exchange (Beaulieu et al. 2012; Read et al. 2012). Authors have argued that large rivers are a hybrid of the hydraulics-driven turbulence in small rivers and the wind-driven turbulence in lakes (Beaulieu et al. 2012). As SWOT will measure only rivers wider than 50m, it follows that wind is likely exerting some influence on gas exchange in SWOT-observable reaches. However, we opted to ignore wind effects in our upscaling model and in BIKER to favor global scalability and implementability for two reasons: 1) current upscaling efforts do not account for wind in their estimation of and 2) it is infeasible to parameterize every set of SWOT data with local wind data, and relying on an in situ understanding of wind defeats the purpose of BIKER for ungauged settings. Rather, we argue that

## 5 Conclusions

## 6 Acknowledgements

BIKER is available at <https://github.com/craigbrinkerhoff/BIKER>. All results and code to generate results, figures, and the manuscript are archived at zenodod something something. Data used in this study is available in Frasson et al. (2019) and Ulseth et al. (2019). This project benefitted from early comments from Tamlin Pavelsky, as well as the large body of existing knowledge from the SWOT River Discharge Working Group.