# Time-varying sediment yield models to support long-term fluvial sediment estimates and climate-smart planning in the Hudson River Watershed

Final Report to the Hudson River Foundation
Grant # 006/17A

Scott Steinschneider
Cornell University
Department of Biological and Environmental Engineering
111 Wing Drive
Ithaca, NY 14850

May 31, 2018

**Abstract**
The research conducted in this work developed dynamic (i.e., time-varying) sediment–discharge rating curves for freshwater tributaries of the tidal portion of the Hudson River. These dynamic rating curves were created in the Esopus, Mohawk, and Upper Hudson Rivers and were used to help quantify the long-term variability of fluvial sediment inputs into the Hudson River Estuary (HRE), with a focus on three primary questions 1) how has sediment yield per unit streamflow changed over time at multiple temporal scales (daily-decadal), 2) what aspects of climate and land use variability are responsible for these fluctuations, and 3) how could these fluctuations be modeled under future climate scenarios. Results of the analysis highlighted the lasting impact of major floods on the flow-sediment relationship in rivers originating in the Catskill Range (Esopus, Mohawk), but emphasized longer-term trends over short-term dynamics in in the Upper Hudson that may be related to land use change. Flood-induced hysteresis in the Esopus Creek watershed was effectively modeled using a latent-variable regression approach, enabling long-term climate change projections that could account for non-stationary sediment-discharge behavior. The reconstruction of Mohawk River suspended sediment loads was complicated by substantial differences in rating curve dynamics between the Mohawk and Upper Hudson Rivers. However, the analysis revealed important hysteretic behavior in the Mohawk that could be used in future work to develop improved suspended sediment reconstructions. Overall, the rating curves developed in this work advance the potential for more accurate retrospective and projected future sediment loading analyses throughout the Hudson River Watershed, supporting the generation of sediment loading scenarios needed to assess long-term sediment management strategies in the HRE.

**Five Summary Points of Interest**

- Dynamic Linear Regression Models applied to the Esopus Creek revealed interesting flood-induced positive hysteresis in sediment yield.
- This behavior can be modeled using a latent variable regression model to support long-term climate change impact assessments.
- Non-linear Dynamic Regression Models can be used to model time-varying sediment yield behavior in the Upper Hudson and Mohawk Basins.
- The behavior of sediment yield in the Upper Hudson and Mohawk basins differed significantly, precluding a reconstruction of Mohawk suspended sediment based on data in the Upper Hudson.
- Instead, the Mohawk River exhibits significant, long-term hysteresis, similar to the Esopus Creek and also likely induced by large flood events.

**Overview of Deliverables**

**Publications**

#1) Ahn, K.-H., B. Yellen, and S. Steinschneider (2017), Dynamic linear models to explore time-varying suspended sediment-discharge rating curves, Water Resour. Res., 53, 4802–4820, doi:10.1002/2017WR020381.

#2) Ahn, K-H, Steinschneider, S. (2017), Time-varying suspended sediment-discharge rating curves to estimate climate impacts on fluvial sediment transport. Hydrological Processes, 32, 102-117.

#3) Ahn, K-H, Steinschneider, S. (*under review*), A comparison of two methods to develop time-varying, non-linear suspended sediment-discharge rating curves: An application to the Upper Hudson River, NY. Journal of Water Resources Planning and Management.

**Presentations**

Steinschneider, S., and Ahn, K.-K. (2017), Dynamic linear models to explore time-varying sediment-discharge rating curves, *ASABE Annual International Meeting,* Presentation.

Ahn, K.-K., Yellen, B., and Steinschneider, S. (2017), Dynamic Rating Curves for Estimating Time-Varying Sediment Loads, *ASCE World Environmental and Water Resources Congress 2017*, Presentation.

**Training of Personnel**

- 1 postdoctoral research supported (Kuk-Hyun Ahn)
- 1 semester-long undergraduate research project (Mackenzie Scheerer)

**Primary Hypothesis and Objectives**

The primary hypothesis of this work was that sediment yield from freshwater fluvial systems in the Hudson River Watershed vary on intra- and inter-annual timescales in response to changes in watershed characteristics, such as vegetation cover and stream bank instability, which in turn may be initiated by anthropogenic and natural disturbances to the watershed. We postulated that time-varying sediment yields could be modeled using new classes of dynamic sediment-discharge rating curves and available data collected through the USGS/DEC Hudson River Watershed Suspended-Sediment Monitoring Network and NYCDEP's Watershed Water Quality Monitoring Program. We tested our hypothesis in three freshwater tributaries to the tidal Hudson: the Esopus Creek, Mohawk River, and Upper Hudson River. In addition to providing a diagnostic analysis of changes in sediment yield over time, we sought to demonstrate two different applications of the proposed modeling framework: 1) to develop extended estimates of sediment loading in the Mohawk River Basin and 2) to support climate change assessments in the Esopus Creek Basin. These applications were the basis for the three objectives proposed in this work:

***Objective 1.*** *Develop and test filtering models of the daily sediment-discharge relationships in the Esopus, Mohawk, and Upper Hudson Basins to diagnose the historical variability of sediment yield across daily-decadal time scales.*

***Objective 2.*** *Use the modeling framework to reconstruct suspended sediment loads in the Mohawk River Basin from 1976-2003 based on available discharge records and long-term (1976-2013) sediment loading data in the Upper Hudson.*

***Objective 3.*** *Develop and test a predictive model in the Esopus Basin that can project dynamic sediment yield-discharge behavior under alternative climate scenarios.*

In the reporting that follows, we first discuss the data used throughout our analyses, and then present the development of dynamic rating curve models for the Upper Hudson and Esopus Creek in the results for Objective 1. As the Mohawk River is the focus of Objective 2, we present the model for that river under our reporting for that objective. We were unsuccessful in reconstructing suspended sediment loads for the Mohawk during the 1976-2003 period, and use our reporting for Objective 2 to describe the challenges we faced and insights we gained through our analysis. Based on these insights, we provide possible directions for future research, which include extending the methods developed in Objective 3 to support the reconstruction of sediment loads in the Mohawk River. For all objectives, we first present details of the experimental design, and then selected results. Due to space limitations, we direct the reader to attached articles for a more in-depth reporting of methods and results. These articles are numbers above (see "Overview of Deliverables") and referenced throughout this report when appropriate.

## Data

Two long-term, daily records of suspended sediment loading in the Upper Hudson and Mohawk Basins were collected from the USGS/DEC Hudson River Watershed Suspended-Sediment Monitoring Network. A near continuous daily suspended sediment loading time series for the Upper Hudson was taken from the Hudson River at Waterford gage (#01335770) and extends from Oct. 1976 to Mar. 2014. Similar data for the Mohawk was taken from the Mohawk River at Cohoes gage (#01357500). These data are available across three separate periods: Jan. 1954 to Jun. 1959, Aug. 1976 to Sep. 1979, and Oct. 2004 to Sep. 2015. We curtail all data to between Oct. 1 1976 and Sep. 30 2013 to work with full water years. Continuous records of discharge were collected from these two gage sites for the period of suspended sediment data availability.

In the Esopus, long records of suspended sediment are not available. However, NYC DEP has collected an extensive set of turbidity grab sample measurements in the Esopus Creek near its inlet to Ashokan Reservoir between Oct. 1987 and Sep. 2015 (*NYC DEP*, 2016). These data have been collected between 1-7 times weekly with an average inter-sample separation of 2.6 days. Daily streamflow is available from the USGS Esopus Creek at Coldbrook gage (#01362500) over the same period. For Objective 1, we analyzed turbidity load, the product of discharge and turbidity, as a proxy for suspended sediment load, over the 1987-2015 period (see article #1). For Objective 3, we also estimated suspended sediment concentration from turbidity and used these data in the development of latent variable regression models for the Esopus Creek (see article #2).

## Objective 1

*Modeling Experimental Design*

We used dynamic regression models to explore how the relationship between daily discharge and sediment has changed over time in freshwater tributaries to the Hudson

River, with an initial focus on the Esopus Creek and Upper Hudson watersheds. Dynamic regression models are a class of state space model that represent a set of observations (i.e., suspended sediment concentration, load, or turbidity) as the sum of one or more latent (unobservable) state variables and observation noise, the latter often taken as the combination of measurement, sampling, and residual misspecification errors (Harrison and West, 1999). The latent states are designed to capture the underlying (i.e., process-level) changes to the system that drive variations in the data beyond the observational noise. In the rating-curve application forwarded here, the underlying processes considered are changes in the relationship between sediment and flow. This is modeled by letting the rating curve parameters be latent variables that vary in time and are inferred from the available observations.

The generic dynamic regression model framework can be expressed as follows:

$$\log S_t = f(\log Q_t|\boldsymbol{\beta}_t) + \varepsilon_t$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t$$

Here S is suspended sediment concentration, load, or turbidity, Q is discharge, $f(\log Q_t|\boldsymbol{\beta}_t)$ denotes some arbitrary function of discharge, and $\boldsymbol{\beta}$ is a vector of parameters that can vary through time. These parameters are updated from time step to time step as a multivariate random walk based on some state evolution error vector $\boldsymbol{w}$.

The goal of the updating procedure is to estimate the posterior distribution of the parameter vector given the data up to time $t$, $p(\boldsymbol{\beta}_t|S_{1:t}, Q_{1:t})$. For simplicity, we hereafter omit conditioning on discharge. The distribution is evaluated in a Bayesian recursive estimation that follows prediction and updating steps, given some distribution for initial values of $\boldsymbol{\beta}_{t=0}$:

Prediction: $$p\left(\boldsymbol{\beta}_t\middle|S_{1:(t-1)}\right) = \int h(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})\ p\left(\boldsymbol{\beta}_{t-1}\middle|S_{1:(t-1)}\right)d\boldsymbol{\beta}_{t-1}$$

Update: $$p(\boldsymbol{\beta}_t|S_{1:t}) = \frac{g(S_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|S_{1:(t-1)})}{p(S_t|S_{1:(t-1)})}$$

Here, $g(S_t|\boldsymbol{\beta}_t)$ is the conditional pdf of the data given the current parameter set, and is determined by the deterministic function $f(\log Q_t|\boldsymbol{\beta}_t)$ and the stochastic properties of $\varepsilon$. The pdf $h(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})$ describes the transition probabilities of moving from $\boldsymbol{\beta}_{t-1}$ to $\boldsymbol{\beta}_t$, and is determined by the random walk and behavior of $\boldsymbol{w}$. $p(S_t|S_{1:(t-1)})$ is a normalizing constant.

If $f(\log Q_t|\boldsymbol{\beta}_t)$ is a linear function and both $\boldsymbol{w}$ and $\varepsilon$ are assumed to be normally distributed, then the Bayesian recursive estimation can be conducted analytically. This simplifying case was found to be appropriate for the Esopus Creek watershed. Here, the dynamic regression model becomes a dynamic linear model (DLM), which can be expressed by the following observation and state equations:

$$\log S_t = \beta_{0,t} + \beta_1 \log Q_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_{\epsilon,t}^2\right)$$

$$\beta_{0,t} = \beta_{0,t-1} + w_t \qquad w_t \sim \mathcal{N}(0, \sigma_w^2)$$

In this application, we set S to turbidity load (turbidity×flow), and used this formulation to diagnose changes in S over the period between 1987-2015. We only allowed the intercept ($\beta_{0,t}$) of the regression to change over time, and examine its temporal variability to better understand dynamics in sediment yield.

For some arbitrary, nonlinear form of $f(\log Q_t|\boldsymbol{\beta}_t)$ and potential non-Gaussian behavior in the error terms, the model is a dynamic non-linear model (DNLM) and the solution is analytically intractable (Arulampalam et al., 2002). In this case, the conditional distributions are estimated sequentially using a particle filter, which is a sequential Monte Carlo approach to approximate the Bayesian recursive estimation (Schon et al., 2005). This application was used for the rating curve of the Upper Hudson River, since the curve was highly non-linear and therefore could not be modeled under the simpler linear framework.

For both the Esopus Creek and Upper Hudson River, we develop a baseline model, termed $M_{Baseline}$, against which to compare our dynamic model results. For the Esopus Creek, the baseline model is a simple log-linear regression between turbidity load and flow. For the Upper Hudson (and later the Mohawk), we use a piecewise linear model similar to that in Woodruff (1999) as a baseline:

$$\log S_t = \beta_0 + \beta_1 \log Q_t + \varepsilon_{1,t} \qquad \forall\, Q_t < \mathrm{Q}^*$$

$$\log S_t = \beta_0 + \beta_1 \log \mathrm{Q}^* + \beta_2(\log Q_t - \log \mathrm{Q}^*) + \varepsilon_{2,t} \qquad \forall\, Q_t \geq \mathrm{Q}^*$$

In this baseline model, there are four parameters, including an intercept ($\beta_0$), two slopes ($\beta_1, \beta_2$), and a threshold ($\mathrm{Q}^*$) that determines the cutoff between lower and upper slopes.

In addition, for both linear and nonlinear dynamic regression models above, we compare the results of the analysis to an existing time-varying rating curve approach called weighted regression on time, discharge, and season (WRTDS), which analyzes variations in S using time, discharge, and season as covariates in a linear regression (Hirsch et al., 2010):

$$\log S_t = \alpha_{0,t} + \alpha_{1,t} \log Q_t + \alpha_{2,t} i_t + \alpha_{3,t} sin(2\pi i_t) + \alpha_{4,t} cos(2\pi i_t) + \varepsilon_t$$

Here, $i_t$ is decimal year at time $t$, $\{\alpha_{0,t}, \alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t}, \alpha_{4,t}\}$ are fitted coefficients, and $\varepsilon_t$ is an error term. The coefficients are allowed to vary through time via a process of re-estimation for each observation in the record using localized regression. In this way, WRTDS produces a time series of parameter estimates than can be analyzed for trends in the underlying rating curve, similar to the dynamic regression frameworks. In the application of WRTDS to the Upper Hudson River, some alterations to the method were tested to improve its interpretability, including the removal of time (i.e., $\alpha_{2,t}$) from the equation above. This new version of the model was termed Weighted Regression on Normalized Discharge and Season (WRNDS), and is described further in article #3.

*Results*

*Esopus Creek*

Figure 1(a) shows the daily time series of the dynamic intercept from October 01, 1987 through September 30, 2015. We found that the dynamic intercept exhibits a complex array of variability across scales. For instance, the dynamic intercept exhibits a clear seasonality, as shown by the power spectrum for the 12-month period in Figure 1(b). However, the strength of this annual cycle varies over the period of record, suggesting that seasonal dynamics are not constant over long time scales. There is also significant and episodic oscillatory behavior in the 4-8 month band. Interestingly, these oscillations coincide somewhat with some of the largest flow events on record, depicted by vertical lines and tick marks in Figure 1. This indicates that large floods may initiate a longer-lived response in S yield beyond the peak in S associated with the large flow event. Overall, the large fluctuations in the intercept would suggest that, even after controlling for variability in discharge, there is a substantial amount of variability in S yield over the 29-year record.

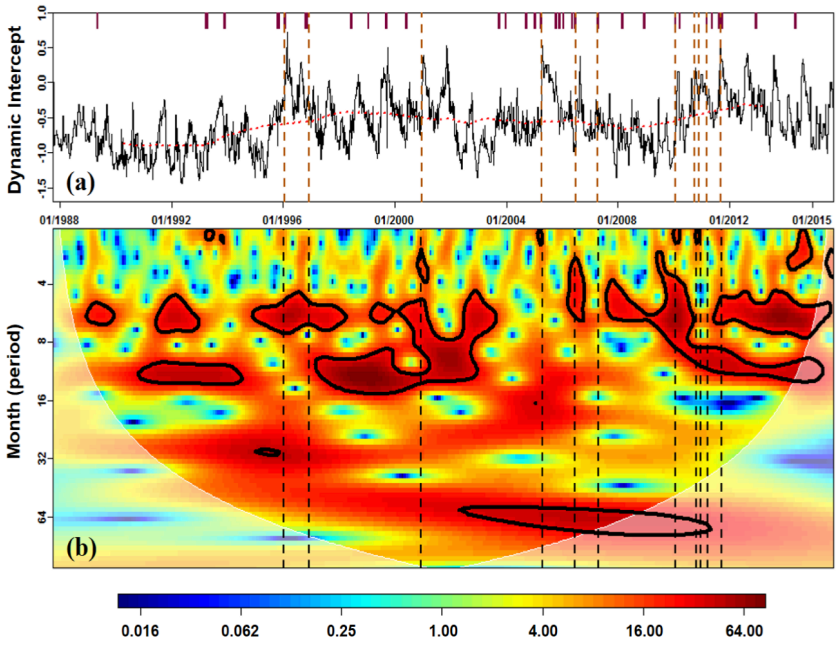


**Figure 1. Intercept parameter of the DLM, including (a) daily time series plot along with its 5 year moving average (red dotted line) and (b) power spectrum for monthly averaged $\beta_0$ values, with bolded regions delimiting significance at the 90 % level. Vertical dotted lines mark the date of large floods (> 350 $m^3/s$). Small ticks at the top of (a) mark medium floods (> 204 $m^3/s$, i.e., the 99.5$^{th}$ percentile of daily flows).**

Figure 2 shows observed S prior to, during, and after the six largest flood events recorded in the Esopus Creek at the Coldbrook gage for the sampled period of record, along with loading predictions under the DLM ($M_{\beta 0}$) and a static rating curve model ($M_{Static}$), the time series of discharge, and the time series of the dynamic intercept, $\beta_0$. After each event, $\beta_0$ drifts higher and remains elevated for a prolonged period of time. Following certain floods (Jan. 1996, Apr. 2005, Aug. 2011), $\beta_0$ remained elevated above pre-flood values for over four months. Correspondingly, $M_{\beta 0}$ tends to capture S variability following these floods better than $M_{Static}$. These results suggest a degree of flood event-driven positive hysteresis in turbidity loads. This is thought to be linked to stream bank erosion and mass wasting that persists for prolonged periods after the flood (Yellen et al.,

2014; Dethier et al., 2016). These effects are simulated by the dynamic intercept. From the viewpoint of short- and medium-term load predictions, it is critical to capture such behavior, particularly if large floods occur in close succession. Otherwise, loadings may be severely under-predicted for the later storms.

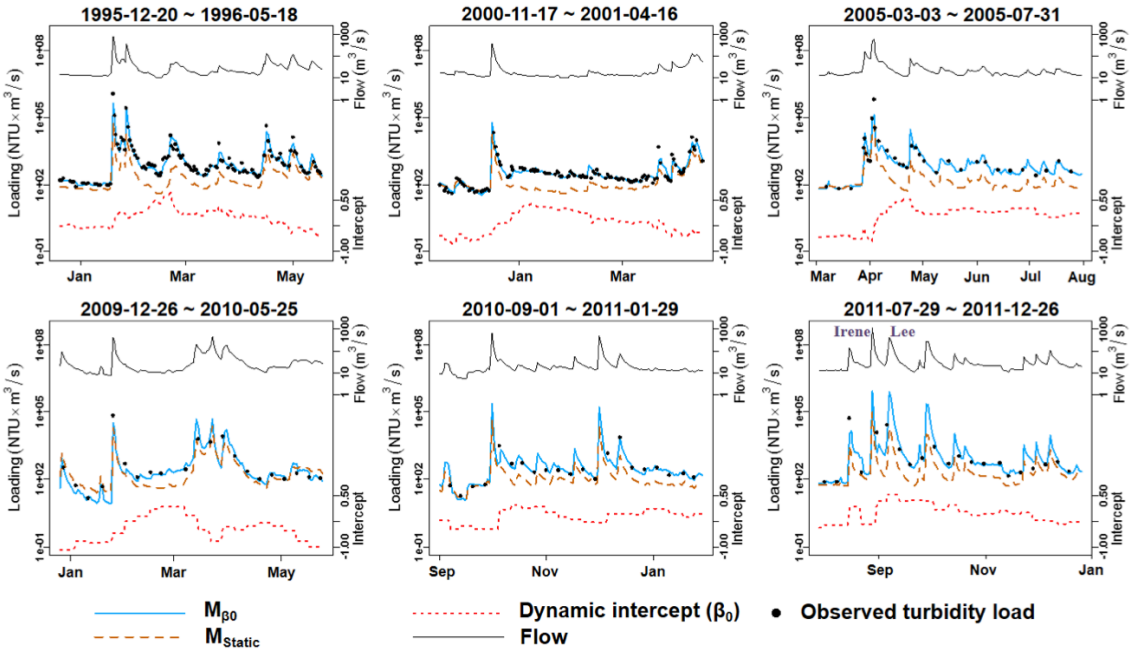


**Figure 2. Observed turbidity loading with model predictions from $M_{\beta 0}$ and $M_{Static}$ during six of the largest flood events on record. Here, hurricanes Irene and Lee are also denoted in the bottom-right pane. Time series of daily streamflow and $\beta_0$ are also shown.**

*Upper Hudson*

Figure 3 presents observed versus estimated sediment loads for the entire record calculated by $M_{Baseline}$, $M_{WRNDS}$ and $M_{DNLM}$. The models were calibrated using all of the data in the record and also compared under three different training and testing data sets. When compared to $M_{Baseline}$, both $M_{WRNDS}$ and $M_{DNLM}$ exhibit a tighter fit to the observations and achieve lower RMSE and higher NSE values in log-space in all three testing experiments, suggesting that both models capture important aspects of dynamic behavior in sediment load. The large outperformance in log space indicates significant improvements in representing dynamics in the data that occur in the body of the distribution for sediment load. $M_{DNLM}$ does over-estimate a few extreme loading values, producing large RMSE values in the original data scale, even though this model exhibits the best NSE values for all testing periods in both log and original scales. Overall, the performance of both $M_{WRNDS}$ and $M_{DNLM}$ is sufficiently higher than $M_{Baseline}$, particularly under cross-validation, to suggest they are capturing important changes in the underlying rating curve relationship for the Upper Hudson River.

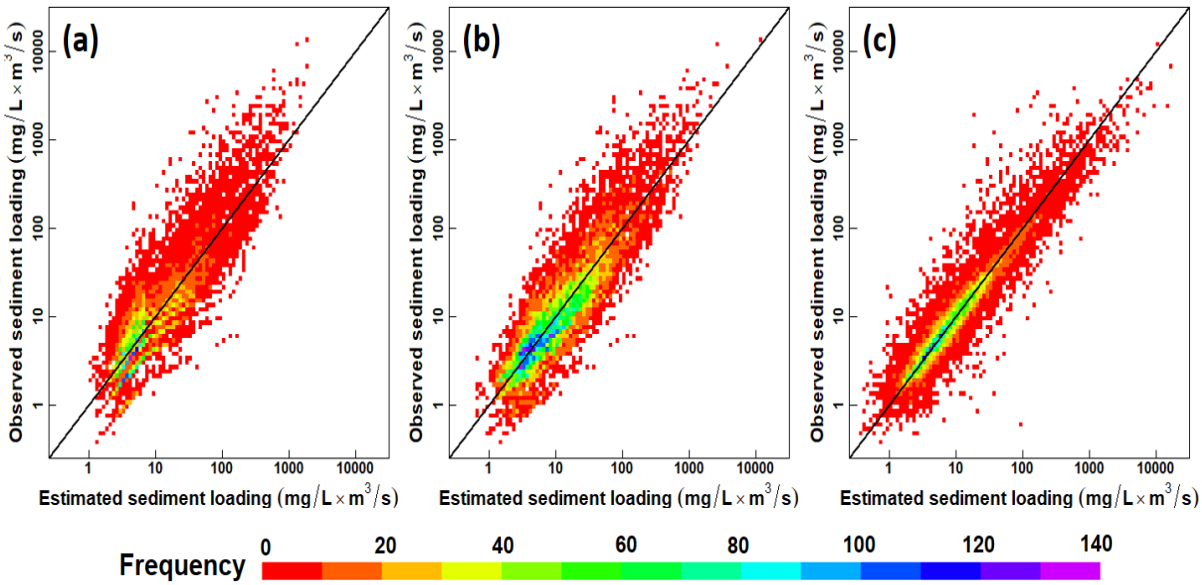


**Figure 3. Two-dimensional histograms between observed and modeled sediment load estimated by (a) $M_{Baseline}$ (b) $M_{WRNDS}$ and (c) $M_{DNLM}$**

Figure 4 shows the daily time series of the time-varying parameters for $M_{WRNDS}$ and $M_{DNLM}$. It is helpful to examine this variability through a physical interpretation of the rating curve parameters. Asselman (2000) proposed that the intercept of a rating curve represents the erodability of the watershed, while the slope represents the erosive power of the river. Therefore, changes in the intercept could reflect changes in land use/cover or river morphology that alter the exposure and ease with which materials are transported, while changes in slope reflect changing erosive power, perhaps related to shifts in streamflow generation from snowmelt or rainfall processes. The goal here is to determine which parameter trends reflect true, underlying changes in the erodability of the basin or erosive power of the river, and which are artifacts of sampling variability. A comparison between the two methods provides some insight towards this end.

For $M_{WRNDS}$, the primary pattern that emerges is a gradual, decreasing trend for the local intercept and, to a lesser extent, a gradual increasing trend in the local slope. In addition, the slope appears to be increasing more in its lower range, and less so for the upper range of values. However, this difference may be obscured by the fact that WRNDS smoothes slope estimates across months and between the two braches of the piecewise linear curve near the median flow value. Therefore, changes in the upper slope could end up propagating into estimated changes in the lower slope, or changes in certain months could be propagating into parameter estimates for other months. This hypothesis is somewhat supported by examining the DNLM results. The DNLM also predicts a decrease in the intercept, but shows little support for a change in the lower slope value. Rather, the DNLM estimates an increase in the upper slope. However, it is important to note that the 95% credible intervals for the DNLM are highly variable over the record, particularly for the slopes, making it different to assess the significance of any trends. The variability in the credible interval widths is due to runs of dry and wet flows that are below or above the median value, leading to stretches of time when there is no information to update

either the upper or lower slope. This leads to periodic expansions of the posteriors for these parameters as uncertainty around their true value increases. Still, one can focus on times when the credible intervals are tighter for either slope to try and infer trend significance. If this is done for the upper slope, the upward trend appears to be mildly significant when comparing the posterior modes and range of credible intervals between the first and second half of the record. Overall, the WRNDS and DNLM models suggest that the erodability (intercept) of the watershed appears to be decreasing over time, while there may be some evidence that the erosive power of the river (slope) is increasing.

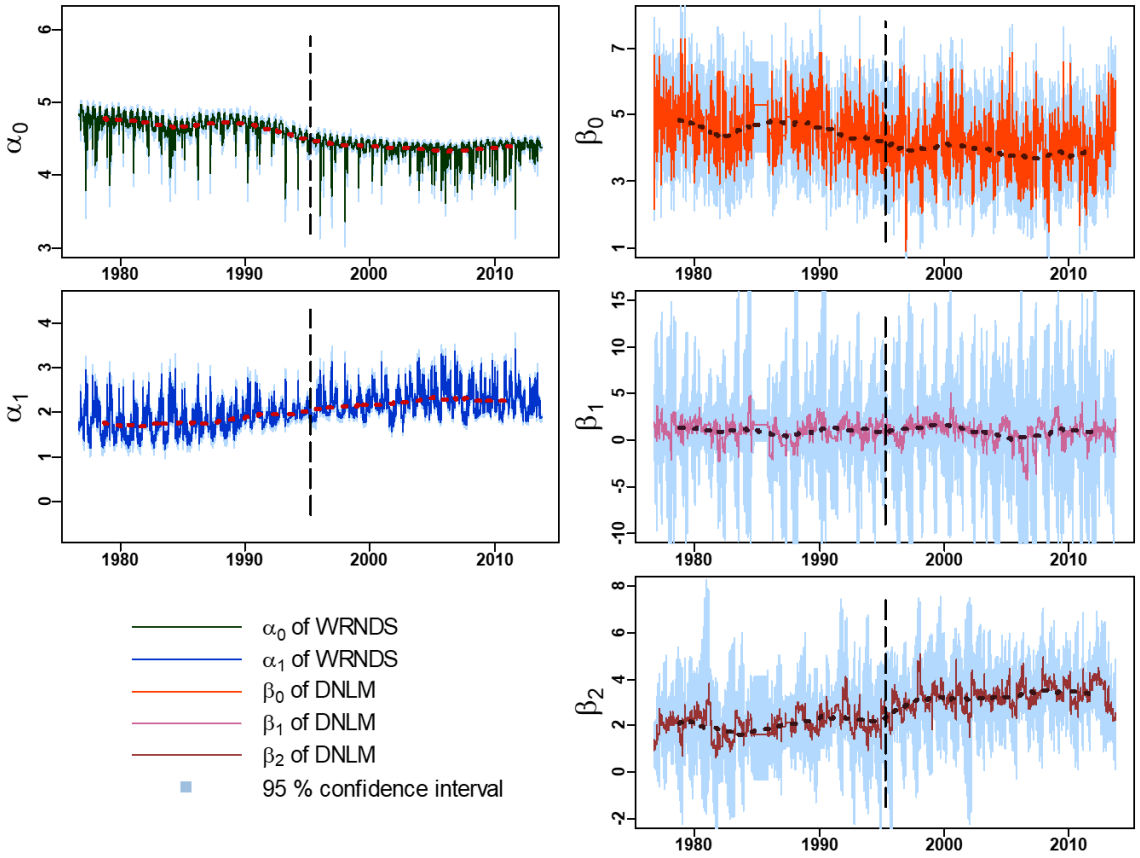


**Figure 4. Time-varying parameters of the $M_{WRNDS}$ and $M_{DNLM}$ models applied to observed data set from the Upper Hudson River. Daily time series are shown with 95% confidence intervals and a 4 year moving average of the median (dotted line). A dashed vertical line is also shown at 1995 where the data are separated for a composite analysis (see Figure 5).**

To further validate these results, we also estimated $M_{Baseline}$ separately by month for two periods of the record (1977-1995 and 1995-2014) and compare the estimated parameter values (Figure 5). This approach is helpful to visualize the changes in the rating curve by month, although we emphasize that it would not have been clear when to define the two periods *a priori* without first examining the DNLM and WRNDS results. In fact, this highlights a major benefit of both approaches – they can be used to determine change points in the rating curves that can then inform composite analyses, supporting a staged approach to the examination of changes in complex rating curves. Figure 5 generally confirms the results of both the WRNDS and DNLM models. It also provides some additional detail, for instance showing a decline in the intercept in most non-summer months, but not in the summer. There also appears to be some steepening in the slope in

most months, with increases in the lower slope being most prominent in the summer and increases in the upper slope more evenly distributed throughout the year. Perhaps most importantly, this analysis shows substantial differences in the rating curve not only across years, but also by month, which we later show to be statistically significant (see Obj. 2).

What is driving the rating curve trends for the Upper Hudson River, particularly a decline in the intercept? A complete diagnosis of the causes of these trends is beyond the scope of this work, but an initial analysis of land use/cover change over the study period can provide some clues as to the possible mechanisms involved. Between the periods of 1977 and 2006, the Upper Hudson River saw a general decline in agricultural land use (15.10% to 9.2%), similar to most of New York State (New York State Department of Environmental Conservation, 2005; Stanton & Bills, 1996). In conjunction with this change, wetland cover also increased (1.9% to 10.4%), somewhat reversing the trend of draining wetlands for agricultural land development that was prevalent earlier in the 20$^{th}$ century (Miller, 2013). The decline in agricultural lands, coupled with the increase in wetlands, may have reduced the erodability of the watershed by stabilizing soils and generally reducing the ease with which material is transported away from watershed. This could explain the declining intercept observed for most months.

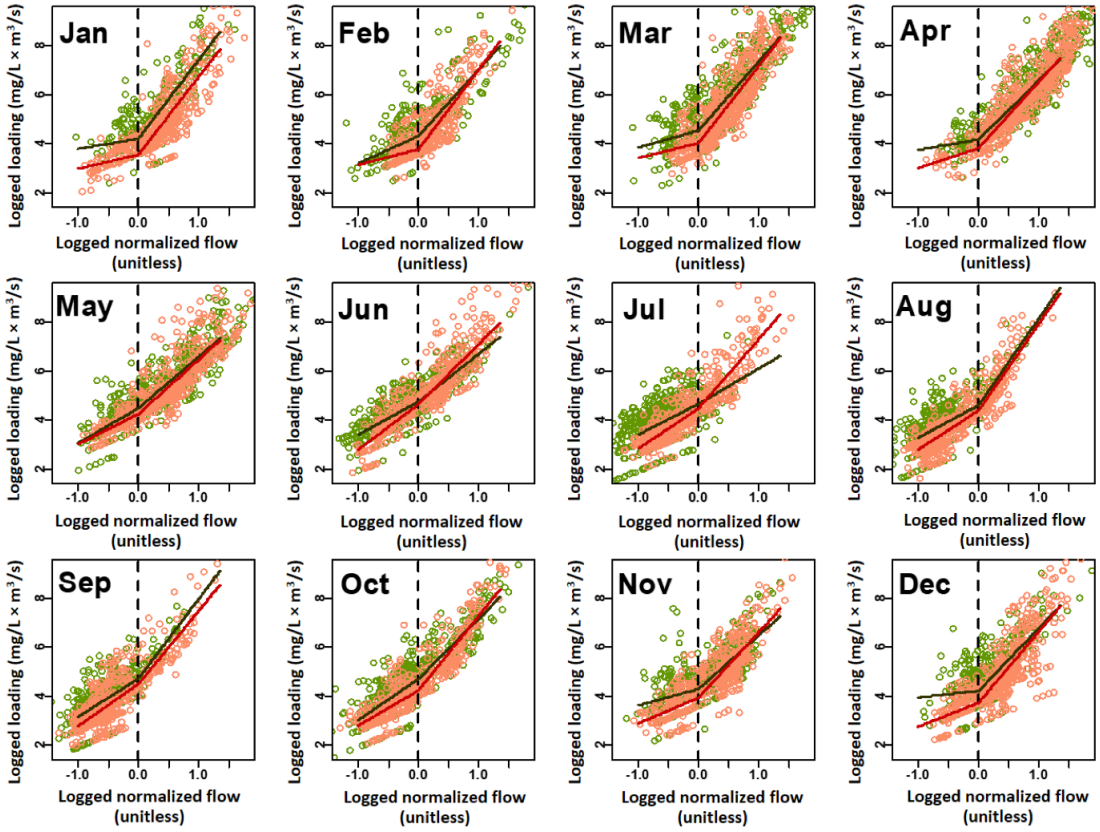


**Figure 5. Monthly nonlinear regressions ($M_{Baseline}$) in the logged S-Q relationship for two periods. Green indicates data between Oct. 01 1976 and Mar. 31 1995, whereas orange represents data between Apr. 01 1995 and Sep. 30 2012.**

The change in slope is somewhat harder to interpret. Such a change generally indicates an increase in erosive power, or alternatively reflects the belief that new sediment sources become available for transport under relatively weathered conditions (Asselman, 2000). One hypothesis for the increased slope follows this latter explanation, where at high flow values, the material available for erosion is relatively unchanged across the record. It is only at low and moderate flows that sediment yield declines due to land use changes. Under this hypothesis, the decline in sediment yield in the lower and mid flow range (i.e., the declining intercept) must be accompanied by an increase in the upper slope to explain the data in the upper flow range. Still, one cannot entirely discount the possibility that the trend seen for the upper slope may be an artifact of sampling variability, particularly given the large uncertainty bounds for this parameter provided by the DNLM.

## Objective 2

*Modeling Experimental Design*

The goal in Objective 2 was to reconstruct suspended sediment loading in the Mohawk basin for the period prior to 2004 using available data for that period in the Upper Hudson. This effort was thought to hold promise because these sub-basins have experienced similar climate conditions and have undergone similar landscape change over the last several decades (Swaney et al., 2006). We intended to use the dynamic regression filtering models to help extract these common signals, which could then be transferred from the Upper Hudson to the Mohawk basins during the pre-2004 time frame. Specifically, we intended to parameterize a common trend into the filtering models for both the Mohawk and Upper Hudson basins, and then use continuous shifts in the rating curve for the Upper Hudson to inform changes in the rating curve for the Mohawk during years (1976-2003) in which the Mohawk has flow but either sparse or no suspended sediment data.

However, upon further data analysis, we determined that this strategy could not correct for the substantial differences in rating curve dynamics between the two basins. To best demonstrate the issue, we highlight here two fundamental differences in rating curve dynamics between the Upper Hudson and the Mohawk basins: flood-induced, long-term hysteresis and seasonality.

First, we develop baseline piecewise linear rating curve models for both basins (described above). We then examine the residuals of these models through time and juxtapose large runs of positive and negative residuals (i.e., long-term hysteresis) against the timing of major floods in each of the basins, which we define as the $99.5^{th}$ percentile of discharge. We also examine the concurrent and lagged relationships between residuals from the two basins, since correlations between residuals would indicate that information regarding suspended sediment load in the Upper Hudson that is not accounted for by discharge could improve the estimation of suspended sediment load in the Mohawk over a rating curve in the Mohawk that only uses discharge.

We also compare the seasonality of piecewise linear rating curves between the two basins. Here, we re-fit the piecewise linear model to both basins by month and examine seasonal cycles in the parameter estimates. We compare the monthly variations against a 95% confidence interval developed through bootstrapping. Specifically, for each basin, we calculate the number of suspended sediment observations $n_m$ available for each month $m$ across the entire record. We then resample with replacement $n_m$ observations from the record without distinguishing by month, and re-fit the piecewise linear rating curve. We repeat this resampling and estimation procedure 1,000 times, and then create confidence intervals for each of the four parameters (intercept, two slopes, threshold). These confidence intervals indicate the range of parameter values one might expect by chance if a random subset of data of length $n_m$ is used for estimation. If the monthly parameter estimates lie outside the 95% confidence interval, this indicates that those monthly parameters differ significantly from a rating curve fit to all the data regardless of month.

*Results*

Figure 6 shows the time series of residuals from piecewise linear rating curves fit to both the Mohawk and the Upper Hudson basins. For the Upper Hudson, residuals are shown for both the overlapping period with the Mohawk (2004-2013) and for the entire 1976-2013 period. When comparing the time series of residuals between the Mohawk and the Upper Hudson basins, there is a clear distinction between the degree of persistence in the time series. For the Mohawk basin, rating curve residuals exhibit substantial persistence, on the order of

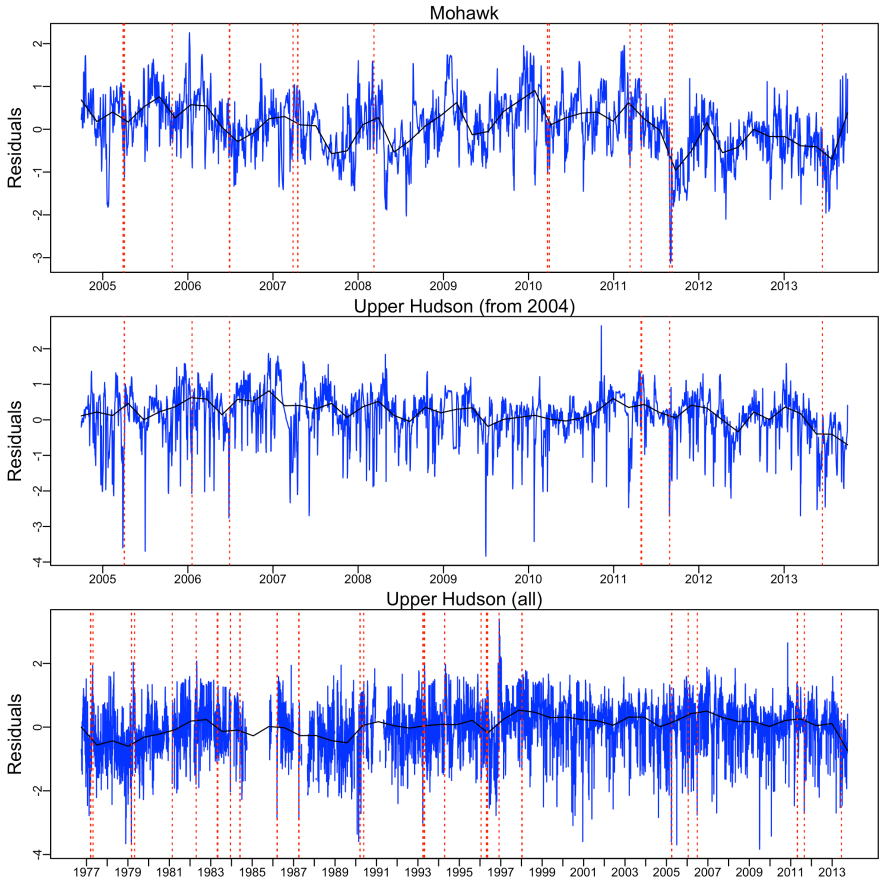


**Figure 6. Time series of residuals from a piecewise linear rating curve model between discharge and sediment load for the a) Mohawk, b) Upper Hudson with data from 2004 onward, and c) Upper Hudson will all available data. A loess smoothing curve is also shown, along with the timing of major floods in the historic record (vertical dashed lines).**

several months. Often, residuals become negative (i.e., the rating curve underestimates sediment load) after large flood events. This is similar to the behavior seen in the Esopus Creek, (see Figure 1), suggesting that the Mohawk River experiences flood-induced long-term hysteresis. We speculate that this behavior is emerging from the Schoharie basin that originates in the Catskills and drains into the Mohawk, but further work is needed to verify this conclusion.

The long memory in the Mohawk residuals contrasts starkly with residuals of the Upper Hudson basin, which exhibit behavior more akin to white noise. Because of this key difference, it is unlikely that residuals from the Upper Hudson basin, which are available for most of the 1976-2003 period, could be used to inform the reconstruction of Mohawk suspended sediment loads during that time period. We confirm this by examining cross-correlations between residuals from the two basins at different lead and lag times (Figure 7). There is a peak in the correlation when residuals for the Upper Hudson basin lead the Mohawk by two days. However, this correlation peaks at 0.23, indicating there is a relatively weak relationship between the two series.

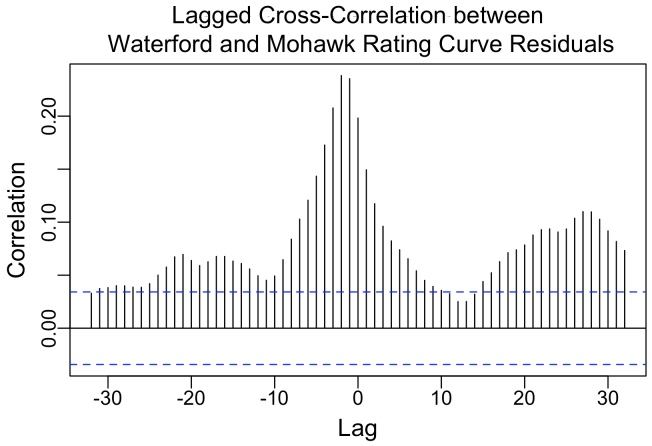


**Figure 7. Lagged cross-correlations between rating curve residuals for the Upper Hudson and Mohawk basins. Negative lags indicate that the Upper Hudson leads the Mohawk basin.**

In addition, we compare the seasonality of rating curve parameters between the Upper Hudson and Mohawk basins in Figure 8, which shows monthly estimates for each of the four piecewise linear rating curve parameters. These estimates are compared against a 95% confidence interval assuming no seasonality in the parameter values. The results suggest that most of the monthly parameters in the Mohawk basin do not vary significantly from the values that would be expected if no seasonality

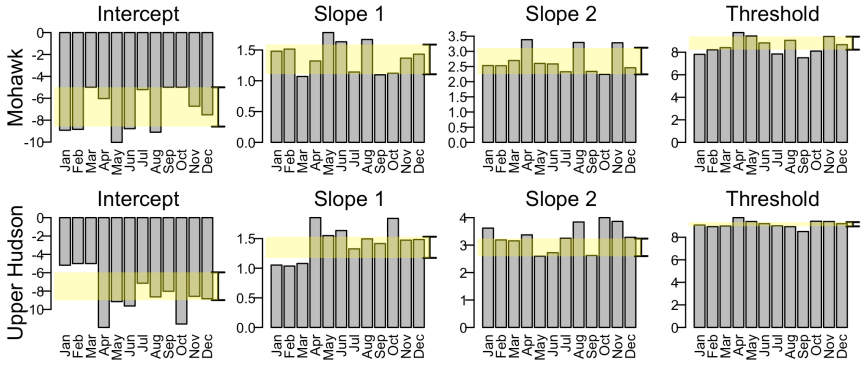


**Figure 8. Monthly values for piecewise linear rating curve parameters for the Mohawk and Upper Hudson River basins. The whiskers and yellow shaded area indicate a 95% confidence interval for parameter value with no seasonality, as determined through bootstrapping.**

were present in the rating curve. There are some slight monthly variations for some parameters that extend outside of the 95% confidence interval, but the magnitude of these deviations in small. In contrast, the Upper Hudson basin shows substantially more seasonality in parameter values. In particular, the parameters in the winter appear to have a lower absolute intercept and lower slope value, while the upper slope is significantly higher in late summer and early fall months. These results support some of the monthly rating curve differences seem in Figure 5 above. The difference in seasonality again complicates an attempt to use data from the Upper Hudson basin to reconstruct sediment loads in the Mohawk.

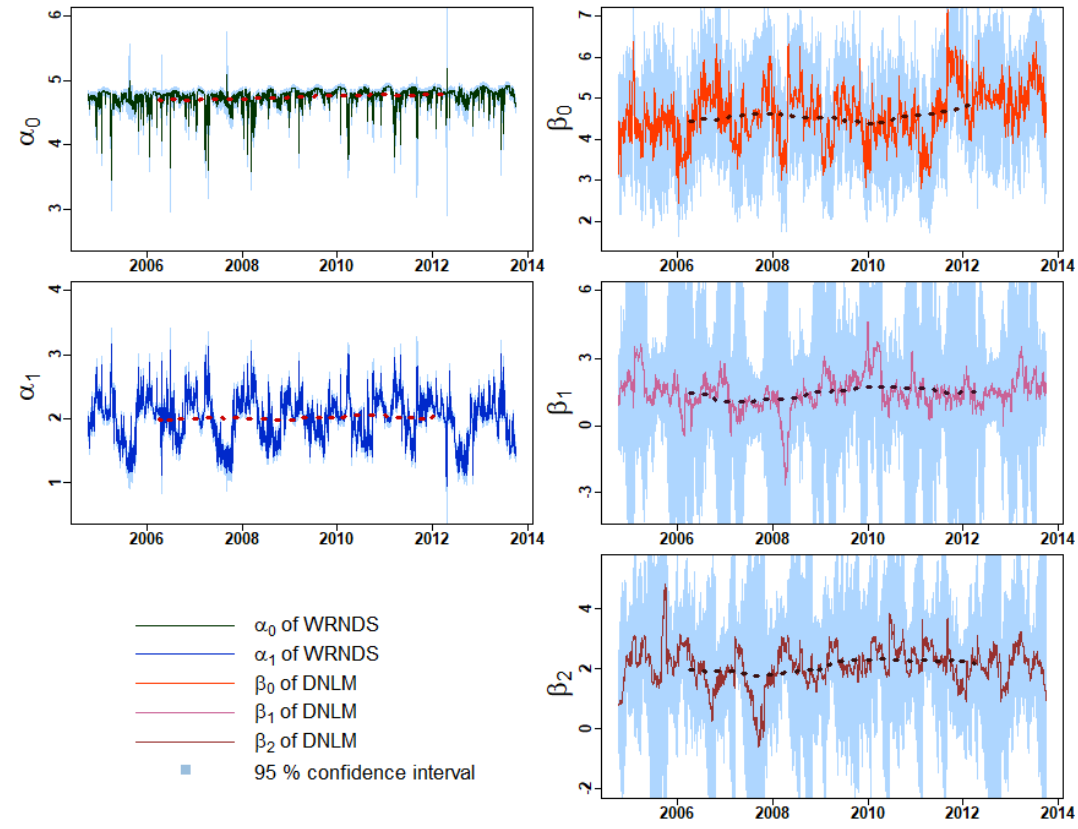


**Figure 9. Same as Figure 4, but for the Mohawk River.**

While variability in the Mohawk rating curve may not be related to that in the Upper Hudson, one can examine changes in the Mohawk curve to better assess sediment-discharge dynamics in that basin. This is shown in Figure 9, which presents the daily time series of the time-varying parameters for $M_{WRNDS}$ and $M_{DNLM}$. For $M_{WRNDS}$, there is no substantial variability in the intercept of the model through time, but the slope for discharge does exhibit quasi-oscillatory variations on a seasonal timescale. This seasonal variability likely relates to the change in slope across low- and high- flow conditions (i.e., the basic nonlinearity of the rating curve), and does not imply any additional seasonal dynamics (as discussed for Figure 8). Overall, there are no significant trends over the 2004-2013 period in the WRNDS parameters, which contrasts the situation in the Upper Hudson. In the DNLM results, the most prominent feature is the step change in the

dynamic intercept after Hurricane Irene in 2011, which persists until the end of the record. This behavior reflects the type of long-term flood-induced hysteresis seen for the Esopus Creek (see Figures 1 and 2), rather than behavior seen in the Upper Hudson. As suggested before, we speculate that this behavior is originating from the Schoharie Creek inputs to the Mohawk, which flows from the Catskills, but further analysis is needed to verify this claim. The DNLM also predicts some annual variations in both the lower and upper slopes of the piecewise linear rating curve (e.g., 2007, 2008, 2010), but the error bounds around these estimates are quite large, indicating that these variations are highly uncertain. Again, the variability in the credible interval widths is due to runs of dry and wet flows that are below or above the median value, leading to stretches of time when there is no information to update either the upper or lower slope. Overall, the DNLM confirms the results from WRNDS and suggests little change in rating curve slope, but does identify an important step-shift in the intercept that was not identified by WRNDS.

## Objective 3

*Modeling Experimental Design*

The DLM results for the Esopus Creek in Objective 1 indicated important variations in the rating curve between discharge and turbidity load. However, because the DLM is effectively a time series filter, it cannot be used to model this behavior outside of the historical record of sediment/turbidity data, as is commonly required in climate change impact studies. Therefore, we developed six rating curves of increasing complexity ($M_1$ through $M_6$) to provide a basis for evaluating whether the behavior captured by the DLM can be parameterized into a rating curve that would be appropriate for modeling flood-induced hysteresis under future climate change scenarios (Table 1). All models utilize two covariates including log-transformed discharge and seasonality, and are used to predict estimated suspended sediment concentration (SSC) instead of turbidity load (see data section and article #2). The first model, $M_1$, presents a baseline rating curve with just these two variables. $M_2$ employs a first order autoregressive (AR) term, which can flexibly capture short-term memory in the system but is not necessarily associated with hysteresis dynamics during high flow events. $M_3$ represents short-term hysteresis using the first difference of log(Q). The fourth model, $M_4$, utilizes a latent, time-varying intercept term $\beta_0^*$ to represent long-term hysteresis following major flood events. The behavior of this latent intercept is designed to capture changes to SSC driven by mass wasting of steep cut banks that persists for prolonged periods after a flood. This model abstracts the flood-induced hysteretic behavior inferred from the DLM above. $M_5$ considers both short and long-term hysteresis simultaneously, combing the first order autoregressive term and the time-varying intercept of models $M_2$ and $M_4$, respectively. Similarly, $M_6$ adds the first time derivative of discharge used in $M_3$ to the structure of $M_4$.

**TABLE 1** Model formulations of six rating curve models

| Model | Model equation | Time-varying intercept |
|---|---|---|
| $M_1$ | $\log rSSC_t = \beta_0 + \beta_1 logQ_{wt} + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \varepsilon_t$ | No |
| $M_2$ | $\log rSSC_t = \beta_0 + \beta_1 logQ_w + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \beta_4 \log\widehat{Q_{st-1}} + \varepsilon_t$ | No |
| $M_3$ | $\log rSSC_t = \beta_0 + \beta_1 logQ_w + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \beta_5 \frac{dlogQ_{wt}}{dt} + \varepsilon_t$ | No |
| $M_4$ | $\log rSSC_t = \beta_0 + \beta^*_{0,t-1} + \beta_1 logQ_{wt} + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \varepsilon_t$ | Yes |
| $M_5$ | $\log rSSC_t = \beta_0 + \beta^*_{0,t-1} + \beta_1 logQ_{wt} + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \beta_4 \log\widehat{Q_{st-1}} + \varepsilon_t$ | Yes |
| $M_6$ | $\log rSSC_t = \beta_0 + \beta^*_{0,t-1} + \beta_1 logQ_{wt} + \beta_2 \sin(2\pi i_t) + \beta_3 \cos(2\pi i_t) + \beta_5 \frac{dlogQ_{wt}}{dt} + \varepsilon_t$ | Yes |

$$\beta^*_{0,t} = \varphi_1 \beta^*_{0,t-1} + \varphi_{2,t}(\log Q_{wt} - \log Q_{threshold})$$

$$\varphi_{2,t} = \begin{cases} 0 & \forall Q_{wt} < Q_{threshold} \\ \gamma & \forall Q_{wt} \geq Q_{threshold} \end{cases}$$

A comparison between $M_1$, $M_2$, and $M_3$ enables us to test the benefit of modeling different aspects of short-term memory in the system, including structured memory targeted specifically for event-based hysteresis. By comparing $M_4$ to the previous models, we can determine the utility of our formulation for long-term hysteresis, while a comparison of $M_4$ with $M_5$ and $M_6$ will highlight the benefits that emerge when considering both short- and long-term hysteresis effects simultaneously.

Finally, a climate risk assessment is presented to explore how these different formulations influence the analysis of water quality risk under plausible scenarios of alternative climate and hydrology. As part of this assessment, we utilized a partial duration model of peak flow events to compare SSC simulations between rating curves under alternative hydrologic scenarios. We first identify historical flow events greater than the value of $Q_{Threshold}$ calibrated for $M_6$. Inter-event timing between these events is modeled using an exponential distribution, and the magnitude of exceedances over the threshold is modeled using a Generalized Pareto distribution. The frequency and magnitude of new sequences of peak flows are then simulated through random draws from these distributions, and SSC for a subset of rating curve models is estimated for each event. Importantly, we consider two versions of $M_6$ in this experiment: 1) the original version calibrated from historical data, and 2) an altered version that uses the fitted parameters from the original version but suppresses any time-varying behavior in the latent intercept ($\beta^*_0 = 0$ for all time steps). This approach enables a direct assessment of how the latent intercept impacts SSC during consecutive flood events.

*Results*

Figure 10 shows 2D histograms between observed and estimated SSC for the entire record using the six nested models. $M_1$ clearly underestimates the most extreme SSC values, while $M_2$ and $M_3$ show some improvement but still underestimate the largest values. Accordingly, both $M_2$ and $M_3$ generally have lower AIC values than the $M_1$ values, with $M_3$ modestly outperforming $M_2$. This suggests that an accounting of memory effects, particularly short-term hysteresis, improves predictions overall and to an extent also helps in the reproduction of extremes.

Compared to $M_1$-$M_3$, $M_4$ generally performs better on the log-scale data, but this is not the case in the original scale. This suggests that $M_4$ is able to capture some post-flood behavior in the data better than the other models, but high SSC predictions are not as robust under the formulation. This can be seen in Figure 10d, which shows more scatter in high SSC predictions under $M_4$ compared to the previous models. Extreme SSC values are better represented when the long-term hysteresis in $M_4$ is coupled with additional terms to account for shorter-term hysteresis and additional system memory ($M_5$, $M_6$). We note that the fitted values of $\varphi_1$ (all $\varphi_1 = 0.98$) in $M_4$-$M_6$ are consistently quite high, indicating substantial long-term memory following large floods. Both $M_5$ and $M_6$ improve performance over all other models when compared in logarithmic space. In the original scale, the results are more mixed, but overall $M_6$ tends to outperform the other models. $M_6$ also appears to show the least systematic bias for large SSC values (Figure 10f). We therefore conclude that $M_6$ provides the most robust predictions of the six models tested.

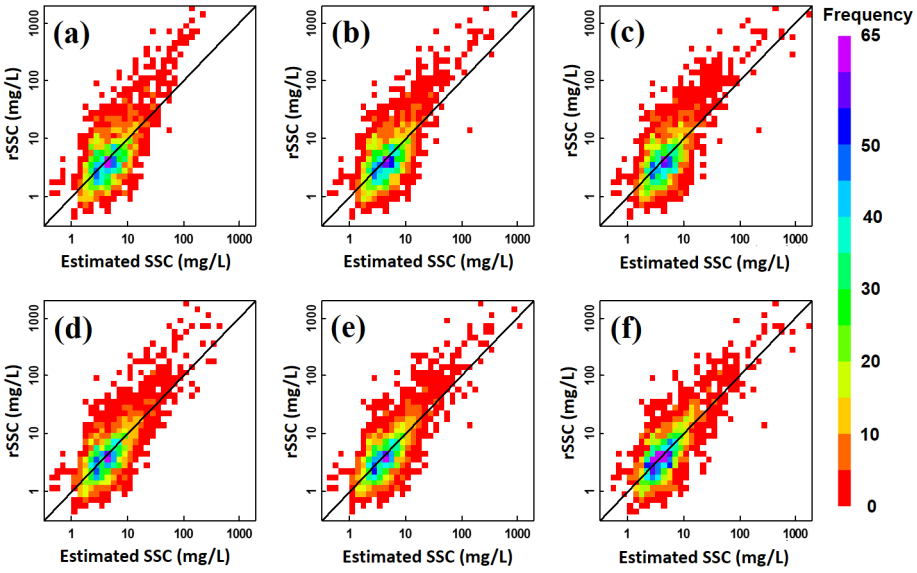


**Figure 10. 2D histograms between observed SSC (rSSC) and modeled SSC estimated by (a) $M_1$ (b) $M_2$ (c) $M_3$ (d) $M_4$ (e) $M_5$ and (f) $M_6$.**

We also examine the impact of the latent intercept and long-term hysteresis on extreme SSC estimates using flood simulations under a partial duration series approach. Figure 11 shows random samples of flood exceedances over $Q_{threshold}$ (fitted for $M_6$) for a future five-year period, given exponential and GPD distributions fitted to historical inter-event timing and magnitudes. Note that daily flow between each flood is not generated in this example. We consider four rating curve formulations including $M_1$, $M_3$, $M_6$ and $M_6^*$. Here, $M_6^*$ uses the fitted parameters from $M_6$ but suppresses any time-varying behavior in the latent intercept. A comparison between $M_6$ and $M_6^*$ allows for a direct assessment of how long-term hysteresis impacts extreme SSC estimates.

In Figure 11a, $M_1$ consistently produces the lowest SSC under all peak events. $M_3$ and $M_6$ produces higher estimates although the predictions are somewhat lower in $M_3$ than in $M_6$. $M_6$ also produces higher estimates than $M_6^*$ after some events, like in 2020 and 2023, but not others. The difference emerges because of the dynamic intercept $\beta_0^*$, whose time-varying behavior is shown in Figure 11b. $\beta_0^*$ remains elevated after the first flood in 2020 and 2023, increasing the estimate for $M_6$ during the second flood in each year. This

highlights how the risk profile of extreme SSC fundamentally changes for a relatively brief (several month) window after major floods, which should be considered when considering how to manage water systems following these events.

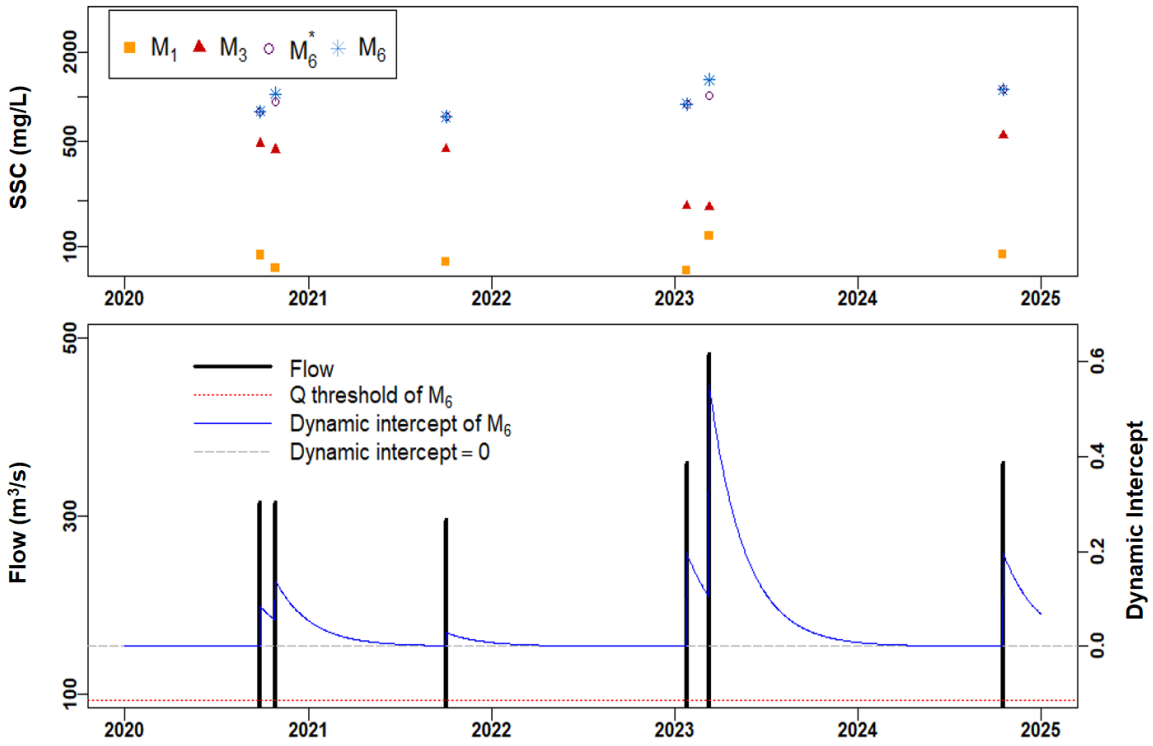


**Figure 11. (top) Simulated SSC under $M_1$, $M_3$, $M_6$ and $M_6^*$ for floods generated using the partial duration model. (bottom) Simulated floods (black) and corresponding changes in $\beta_0^*$ (blue).**

## Conclusion

This work provided two primary contributions to the field of sustainable water resource management in the Hudson River Watershed. First, it elucidated the capacity of multi-scale (daily to decadal) climate processes to shift discharge-sediment load rating curves used to characterize and predict suspended sediment in fluvial systems. Because basin sediment yield information is generally considered a static datum, documenting the mechanisms by which yields vary at different time scales comprises a major addition to our understanding of denudation and sediment delivery to estuary systems like the HRE. Second, this work developed an innovative modeling framework for the Esopus basin to simulate how sedimentation varies across time scales in response to different aspects of climate variability and change. The results of this work also found significant differences in the rating curve dynamics of the Upper Hudson and Mohawk Rivers, with the Upper Hudson exhibiting trends potentially linked to land use change, and the Mohawk River exhibiting significant inter-annual variability following large flood events (particularly Hurricane Irene). These differences precluded a reconstruction of Mohawk sediment loadings based on data from the Upper Hudson. Future work should consider extending the latent regression models developed for the Esopus to the Mohawk River, with the possibility of developing improved reconstructions of sediment load for that basin.

**References**


Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Asselman, N. (2000). Fitting and interpretation of sediment rating curves. *Journal of Hydrology*, *234*(3), 228–248.

Dethier, E., F. J. Magilligan, C. E. Renshaw, and K. H. Nislow (2016), The role of chronic and episodic disturbances on channel–hillslope coupling: the persistence and legacy of extreme floods, Earth Surf. Process. Landf., 41(10), 1437–1447, doi:10.1002/esp.3958.

Gray, A.B., et al., (2015a), Effects of antecedent hydrologic conditions, time dependence, and climate cycles on the suspended sediment load of the Salinas River, California, J. Hydrol., 525, 632-649.

Gray, A.B., et al., (2015b), The effect of El Niño Southern Oscillation cycles on the decadal scale suspended sediment behavior of a coastal dry-summer subtropical catchment, Earth Surface Processes and Landforms, 40, 272-284.

Harrison, J., and M. West (1999), *Bayesian Forecasting & Dynamic Models*, Springer.

Hirsch, R. M., Moyer, D. L., & Archfield, S. A. (2010). Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs. *JAWRA Journal of the American Water Resources Association*, *46*(5), 857–880.

Miller, D. E. (2013). *The Hudson River Estuary Habitat Restoration Plan*. New York State Department of Environmental Conservation, Hudson River Estuary Program.

New York State Department of Environmental Conservation. (2005). New York State Comprehensive Wildlife Conservation Strategy. Retrieved from http://www.dec.ny.gov/docs/wildlife_pdf/cwcs2005.pdf

NYCDEP (2016), Watershed protection program summary and assessment, 2016 Filtration Avoidance Determination Assessment Report, 393 pp.

Schon, T., Gustafsson, F., & Nordlund, P.-J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, *53*(7), 2279–2289.

Stanton, B. F., & Bills, N. L. (1996). *The Return of Agricultural Lands to Forest*. EB.

Swaney, D.P., Sherman, D., Howarth, R.W. (1996), Modeling water, sediment, and organic carbon discharges in the Hudson-Mohawk Basin: coupling to terrestrial sources, Estuaries, 19(4), 833-847.

Warrick, J.A., et al., (2013), Trends in the suspended-sediment yields of coastal rivers of northern California, 1955-2010, J. Hydrol., 489, 108-123.

Woodruff, J. D. (1999). *Sediment deposition in the lower Hudson River estuary*. Massachusetts Institute of Technology.

Yellen, B., et al., (2014), Source, conveyance and fate of suspended sediments following Hurricane Irene. New England , Geomorphology, 226, 124-134.