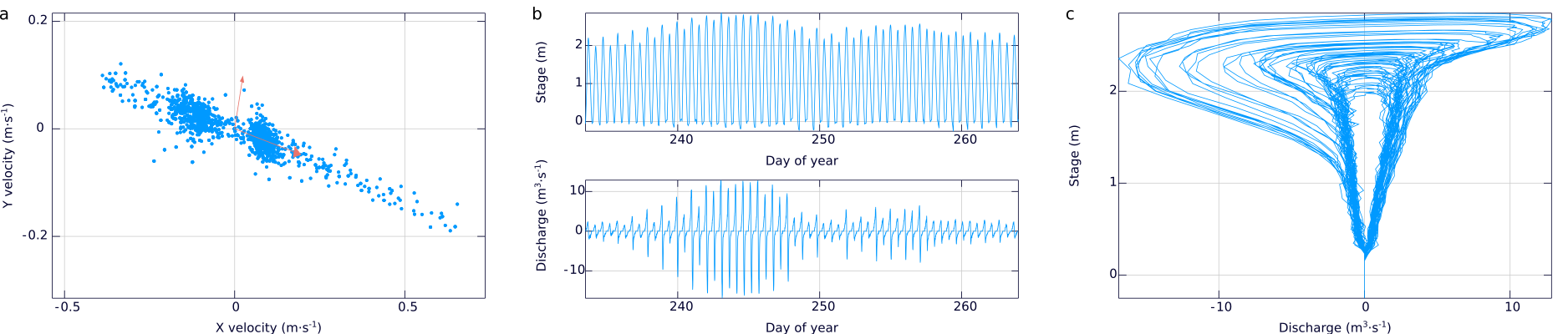
Stage-discharge relationship in tidal channels

Long-term records of the flow of water through tidal channels are essential to constrain the budgets of sediments and biogeochemical compounds in salt marshes. Statistical models which relate discharge to water level allow the estimation of such records from more easily obtained records of water stage in the channel. Here we compare four different types of stage-discharge models, each of which captures different characteristics of the stage-discharge relationship. We estimate and validate each of these models on a two-month long time series of stage and discharge obtained with an Acoustic Doppler Current Profiler in a salt marsh channel. We find that the best performance is obtained by models which account for the nonlinear and time-varying nature of the stage-discharge relationship. Good performance can also be obtained from a simplified version of these models which captures nonlinearity and nonstationarity without the complexity of the fully nonlinear or time-varying models.

# Introduction

The flow of water into and out of tidal channels carries with it nutrients, sediment and biota thus exerting a strong control on the biology and geomorphology of environments such as mudflats, mangroves and salt marshes (Morris et al. 2002; Chmura et al. 2003; Duarte et al. 2005; Cai 2011; Fagherazzi et al. 2013). Accurately estimating the volumetric flux of water, or discharge, through a channel is a crucial component of estimating the flux of materials transported through these systems. In fact, the flux of an advected material is equal to its concentration multiplied by the discharge. Precise estimates of discharge from water levels are important to quantify the exchange of biogeochemical compounds between marshes and nearby bays (Carey and Fulweiler 2014) and determine the stability of salt marshes from channel sediment fluxes (Ganju et al. 2013).

The development of rating curves which relate the easily measured water level, or stage, in a stream cross section to the flow through that cross section is routinely carried out in rivers (Kennedy 1984). Once a rating curve is calculated, discharge can be instantaneously estimated by measuring water level. In coastal streams influenced by tides, simple models for rating curves (such as power laws) fail because of the bidirectional and nonstationary nature of flow in these environments (Fig. c). Bidirectionality means that, in one tidal cycle, there are two discharges with opposite signs for a given stage. The asymmetry of the tides (Boon 1975; Pethick 1980; Healey et al. 1981; Fagherazzi et al. 2008) means these discharges are not simply negatives of each other, with a *hysteresis* between ebb and flood. Nonstationarity in tidal channel flow means that a single water level corresponds to many different discharges (Fig. c) over the course of a stage-discharge record. Nonstationarity comes from storm tides and from lower-frequency harmonics of the tide such as the spring-neap cycle, resulting in different discharges for the same water level. This multiplicity of discharge at a single stage means that we cannot estimate a single rating curve in which discharge at a point in time is a function of stage at that instant.



a) Velocities in the horizontal plane recorded by the Acoustic Doppler Current Profiler. The dominant direction of variability corresponds to the along-channel velocities, and that direction is extracted using principal components (the red arrows). b) The true discharge obtained with a handheld flow meter plotted against the index discharge derived from the ADCP. The red line represents the linear model used to calibrate the index discharge to the true discharge. c) Stage and discharge time series. The red dashed line in the stage time series represents the elevation of the channel bank. The spring-neap tidal cycle over the course of the month results in nonstationarity in the discharge time series. d) An example stage-discharge relationship from a one-month ADCP record in Sweeney Creek, Rowley, MA. Note the bidirectionality and hysteresis in ebb and flood.

The goal of this manuscript is to develop robust and interpretable models that predict discharge from stage in salt marsh channels. Given a record of stage and one of discharge, we aim to find some function of the stage which will return discharge so that, in the future, a record of stage from an independent water level logger can be used to estimate discharge. There are any number of approaches drawn from the statistical and machine learning disciplines, such as neural networks and support vector machines, which can perform fairly well at this problem (Tawfik et al. 1997; Sudheer and Jain 2003; Sahoo and Ray 2006; Ghorbani et al. 2016). However, demanding interpretability from any useful model of stage and discharge, constrains the tools which we can use. By interpretability, we mean that the structure of the stage-discharge model and its parameters are based in real properties of the marsh we are studying. If we possess an interpretable model, we can invert a model fit to stage-discharge data from a creek to learn about the structure of flow through the salt marsh, such as its channel topology, storage properties, and interactions between surface water and groundwater.

Here we examine a suite of models for estimating discharge from stage measurements. We explore the structure of each of these models and their relation to our physical understanding of flow in tidal systems and discuss the challenges to estimating the parameters of each of these models from stage and discharge data. We present a case study using stage-discharge records from a salt marsh creek along the Rowley River, Massachusetts, USA, to compare the performance of each of these methods. We conclude by discussing the advantages of each model and we make some recommendations for stage-discharge modelling in tidal creeks.

# Procedures

## Discharge measurements

We acquired a data set associating discharge with creek stage in August and September 2015 in a salt marsh creek (Sweeney Creek) along the Rowley River, MA. A Nortek Aquadopp acoustic Doppler current profiler (ADCP) operating at 2.0 MHz was programmed to record velocities in 20 cm bins at 10 minute intervals. The blanking distance of the ADCP was set to 10 cm, so that the center of the first bin is 20 cm above the ADCP. The ADCP parameters are given for reference in Table . We installed the ADCP looking upward in the creek thalweg.

The velocity data retrieved from the ADCP come in the form of three matrices where is the number of bins and is the number of points recorded in time. Each of the three matrices represents velocity in one of three directions (east, north and up, ENU). In addition, the water pressure recorded by the ADCP is retrieved. This pressure is converted to a height of water above the ADCP by dividing by the specific weight of water. The velocity data are filtered to remove velocities recorded in bins above the water level and then the valid velocities are averaged to provide a trivariate time series of average velocity above the ADCP in each of the three directions.

Because the velocity variations in the creek channels which we examine are driven entirely by tides, the variability in velocity is dominated by the along-channel component of velocity. This motivates the use of principal components analysis to rotate the trivariate ENU velocity time series into another trivariate time series with components in the along-channel, across-channel and vertical directions. The first principal component (that with the largest eigenvalue) represents the along-channel velocity, the second, the across-channel, and the third, the vertical (Fig. a). The along-channel velocity is the index velocity in the cross section, and it is this component that we use to calculate the discharge.

We use the stage measurements from a pressure transducer in the ADCP along with a channel cross section measured by an RTK-GPS to calculate the flooded cross-sectional area. The index discharge is calculated by multiplying this area by the index velocity. This overestimates the true discharge through the channel because the velocity in the center of the channel tends to be higher than the velocity near the creek banks. Calibration of the index discharge to the true discharge is essential for any consistent estimate of material flux in the channel (Ruhl and Simpson 2005). We calibrate the index discharge to a true discharge with measurements taken with a handheld flow meter (Marsh-McBirney Flo-Mate 2000). We fit a linear regression from the index discharge to the true discharge (Ruhl and Simpson 2005) and then apply this regression model to the entire index discharge time series to obtain a true discharge time series. The final output of these discharge measurements are two time series: one of true discharge and one of stage. We deployed the ADCP for two thirty-day deployments.

## Modelling of the discharge

We examine four different classes of model: an instantaneous, geometric model of flow proposed Boon (1975), a linear, time-invariant model inspired by the unit hydrograph formulation of flow in rivers (the TIGER model presented in Fagherazzi et al. 2008), a nonlinear, time-invariant model based on the Volterra series (Rugh 1981), and a new linear, time-variant model inspired by the recent interest in time-variable travel time distributions (Fagherazzi et al. 2008; Botter et al. 2010; Harman 2015; Beven and Davies 2015). Below we briefly describe the models we estimate on our stage-discharge time series. More detail on each model and on the procedures used to estimate the parameters of these models can be found in the supplemental information.

Throughout, we use the notation to represent the time-varying discharge in a cross section and the time-varying stage in that cross section, and are the discrete stage-discharge time series of length taken at a sampling interval of (i.e. and likewise for the stage).

### The Boon model

Boon (1975) proposed a stage-discharge model as follows

where represents the hypsometric curve, the distribution of area within the salt marsh as a function of height. This model can be derived from the continuity of mass under the assumption that water surface slopes are negligible throughout the marsh. We assume here that we do not possess a description of the hypsometric curve as might be derived from LiDAR or aerial photography surveys. Instead, we estimate a representation of the hypsometric curve from the stage-discharge data. Specifically, we use a power law form for the hypsometric curve, . We approximate by the backward difference operator: . That is, we obtain a nonlinear system of equations in the parameters and of the form

for which we solve for the optimal values of and using nonlinear least squares with the Nelder-Mead method (Kelley 1999).

Extensions of Boon's model have been studied by Pethick (1980), who proposed theoretical forms of based on simple models of channel geometry which are encompassed by the power law model we use here.

### Linear, time-invariant models

The assumption in the Boon model that water surface slopes are negligible has been pointed out as unrealistic in many situations (Healey et al. 1981; Fagherazzi et al. 2008), and the model requires an asymmetric tide to generate asymmetric discharges (Pethick 1980). More fundamentally, the Boon model assumes that the tide propagates instantaneously into the marsh. Fagherazzi et al. (2008) put forward a model based on the instantaneous unit hydrograph developed for river runoff which relaxes this assumption, assuming that the tidal propagation can be described by a travel time distribution which determines how much of the flow at time is due to the increase in stage at time . The tidal discharge is obtained by convolving this travel time distribution with the Boon model.

Because of the dependence of the hypsometric curve and the travel time distribution on water stage, this formulation is naturally time-variant. We first consider a time-invariant version of this model ( for all ) which is both very simple to estimate and able to draw on the rich literature on system identification in linear, time-invariant systems

where we note that we have also incorporated the hypsometric curve into the time-invariant travel time distribution, averaging out its temporal variation to preserve the time-invariance of the model. In other words, we do not estimate a hypsometric curve explicitly in this or any of our later models. This integral equation can be discretized at our sampling frequency, which results in an overdetermined system of linear equations in the parameters, .

Since we ultimately approximate the derivative by a backward difference, the linear model is equivalent to one with replaced by in Eq. and the backward difference incorporated into the kernel coefficients, .

is the system order which determines how far back in time the discharge depends on stage. The system order is a hyperparameter of the problem which needs to be selected before estimating the model parameters . We perform hyperparameter optimization for this and all models using cross-validation, explained below.

### Nonlinear, time-invariant models

Frictional interactions between water, the banks of the channel and the marsh surface introduce nonlinearities into the continuity formulation (Speer and Aubrey 1985). Heterodyning of the stage signal by the nonlinear friction terms introduces higher frequency harmonics of the tide into the discharge, which helps explain the tidal discharge asymmetry (Speer and Aubrey 1985; Blanton et al. 2002). A linear model such as the system above is unable to account for this behavior: it cannot generate frequencies in the output signal which are not present in the input signal; it can only attenuate or amplify the strength of the signal at certain frequencies. We therefore investigate a nonlinear (but still time-invariant) model which is capable of generating these harmonics.

The canonical nonlinear equivalent to the linear, time-invariant system is the Volterra series, also seen in its orthogonalized version, the Wiener series. The Volterra series bears the same relationship to a linear, time-invariant system as a Taylor series does to the evaluation of a function at a point: it can be thought of as a Taylor series with memory. The Volterra series expands the system as a series of integrals of products of the stage signal at different lags

so that the first few terms look like

Note that the first convolution in this series is simply the linear time-invariant system, and the n-th term in the series involve n-th degree monomials of the stage at n different times in the past. We can likewise discretize the Volterra series, giving us a set of nonlinear equations in the coefficients (the discrete versions of the functions ). To estimate the coefficients effectively, we exploit the duality between the Volterra series and polynomial kernel regression (Franz and Schölkopf 2006).

### Linear time-variant models

Both the linear, time-invariant model and the Volterra series model are unable to capture the activation of different flow paths, such as when the water overtops channel banks and moves as sheet flow over the marsh platform. By estimating a single travel time distribution for the entire stage-discharge record, we tend to underestimate the high magnitude discharges just before and after the high slack water and to overestimate the discharge at relatively low flows which are dominated by residual drainage from the low-order creeks and ditches in the system and from seepage out of channel banks. Thus the TIGER model of Fagherazzi et al. (2008) and similar models developed for river basins (Botter et al. 2010; Harman 2015) explicitly account for time-varying travel time distributions. Estimating these travel time distributions is not a well-posed problem because one needs to estimate both the distribution itself and the dynamics of the distribution as it changes change continuously in time.

We therefore have to approximate the dynamics of travel time distributions so that they can be estimated with the finite amount of data that we have. We assume that there are a finite number of states which the flow can be in. We partition the time series into these states and estimate a linear, time-invariant travel time distribution for each state with only those data points representing these states. To predict discharge from a new stage trajectory, we assign the new trajectory to the appropriate state and use the linear model associated with that state to estimate the discharge.

We need to devise a principled way to partition the training data set into states and to assign a new, unobserved stage trajectory to a state. Here, for simplicity, we use an unsupervised clustering method (k-means; Xu and Wunsch 2009) to partition the -dimensional training stage trajectories into clusters such that each trajectory belongs to the cluster with the closest mean in the Euclidean distance. Upon receiving a new trajectory, we compute the distance from the new trajectory to each of the cluster centers and assign it to the cluster with the smallest distance.

This unsupervised method uses only the information in the stage trajectories to form the clusters. It does not take into account the predictive performance of each cluster; this is not necessarily the optimal clustering for discharge estimation. One could, in principle, construct a clustering to optimize the estimation performance, but one would then need to model separately the process which assigns new stage trajectories to these clusters using a supervised classification technique. In practice, the unsupervised clustering performs well without this additional complication.

### Regularization

The individual stage measurements at each ten minute interval are highly correlated with each other, so that each stage data point does not provide independent information for the discharge prediction. This is the collinearity problem familiar to users of any multiple regression (Hocking 1976; Wold et al. 1984). When performing a straightforward regression with this collinear data, we will tend to overfit our model to the training data, reducing its ability to generalize to new data. We will also obtain unphysical estimates of the parameters that oscillate rapidly and are sensitive to noise. Regularization trades off fitting the training data set and constraining the parameters in some way. Variable selection by a stepwise procedure or model selection with the Akaike information criterion (Burnham and Anderson 2002) is one form of regularization. Here, we use Tikhonov regularization (also called ridge regression) which adds a penalty term to the least-squares objective function

where is some positive semi-definite matrix. The penalty term enforces some constraints on the structure of the coefficients, , constraints chosen by the regularizing matrix . For a multiple of the identity matrix, , we obtain the common regularization which penalizes solutions with higher Euclidean norms, leading to smooth parameter estimates with the degree of smoothness controlled by the hyperparameter . Other choices of impose different constraints on the system which may enhance the interpretability of the model. For example, stable spline kernels (Pillonetto and De Nicolao 2010) enforce stability of the linear, time-invariant system, leading to an appropriately decaying impulse response, while in kernel regression methods such as that used to implement the Volterra series model, the matrix corresponds to the measurement error covariance. However, we use regularization in our assessment below, as it offers reasonable performance without much additional complexity.

### Cross-validation

To estimate the hyperparameters of each model, such as the system order or the regularization parameter, we use a cross-validation approach. We divide our training data set evenly into two blocks, and, for each value of each hyperparameter on a grid, we estimate the model using the data from the first block. We apply the estimated model to the second half of the training data set and measure the mean squared error between the estimated discharge and the observed discharge in that block. We choose the values of the hyperparameters which minimize this prediction mean squared error and reestimate the model on the entire training data set using these optimal hyperparameters before applying it to any further stage records from the same creek.

### Discharge estimation with a fitted model

To apply these models to develop a stage-discharge relationship for a particular channel, one must first collect a training data set with an ADCP and fit the model as described above. From then on, discharge can be estimated with only an independent water level logger instrumenting the channel. One records water level in the same cross section at the same sampling rate as the training data in the same cross section. Different cross sections will exhibit different stage-discharge relationships, and a model estimated on one cross-section is not valid at another cross section even in the same channel. The sampling rate must be identical because the each of the parameters in all of the models take the form of a coefficient that is applied to stage a certain amount of time in the past. To estimate discharge at the present time, one collects the stage time series from the present stretching back into the past a certain amount of time. We call this short record a "stage trajectory." In our measurements, at time steps of 10 minutes each, a 25 hour long stage trajectory is a vector of length 150. Each model takes a stage trajectory and applies some transformation to it -- a linear combination of the stages in the linear, time-invariant model, for instance -- and returns an estimate of discharge. If estimates of uncertainty are required for the estimated discharge value, bootstrap methods adapted for time series (Bühlmann 2002) can be easily applied to each of the models, though we will not specifically address methods for uncertainty quantification here.

# Assessment

To compare the performance of each of these models, we estimate each model on our ADCP stage-discharge records from the Rowley marshes. We follow the cross-validation procedure outlined above to estimate the parameters for each model on the first of the two stage-discharge records and apply the model to the second ADCP record. We examine, in turn, the parameters estimated for each model, the estimation performance of each model on the second stage-discharge record, the behavior of the residuals, and the impact that regularization has on both the estimated parameters and the estimation performance.

## Model structure

Each of the four classes of model uses a slightly different type of parameter set, and we show each of the resulting parameters in Fig. . The Boon model produces an estimated hypsometric curve in a power law form (Fig. a). The linear time-invariant model produces a single impulse response, representing the contribution of stage in the past to flow in the present (Fig. b). The Volterra series model generates a set of multidimensional impulse response functions. For simplicity, we show just the first order Volterra operator, which is just a linear, time-invariant impulse response, and the second order Volterra operator, which is a two-dimensional set of coefficients (Fig. c). The k-means model produces impulse responses, one for each of the clusters, and also assigns each point in the time series to one of these clusters (Fig. d).

a) The hypsometric curve estimated in the Boon model. b) The impulse response estimated in the linear, time-invariant model. c) The first-order Volterra kernel is equivalent to a linear, time-invariant impulse response (top). The second-order kernel is a two-dimensional analogue of the impulse response. The distance along the x- and y-axes are the lags backwards in time for each of the directions of the impulse response. The color is the amplitude of the impulse response. d) The k-means model estimates k impulse responses (left). Each impulse response is used to estimate from the correspondingly colored point in the stage time series (right) 

a) The hypsometric curve estimated in the Boon model. b) The impulse response estimated in the linear, time-invariant model. c) The first-order Volterra kernel is equivalent to a linear, time-invariant impulse response (top). The second-order kernel is a two-dimensional analogue of the impulse response. The distance along the x- and y-axes are the lags backwards in time for each of the directions of the impulse response. The color is the amplitude of the impulse response. d) The k-means model estimates impulse responses (left). Each impulse response is used to estimate from the correspondingly colored point in the stage time series (right)

The ideal system order in the linear models and the Volterra series describes how much memory is needed to estimate discharge effectively. Using cross-validation to select the system order ensures that we do not choose an order too large, in which case the model would overfit the data and have poor prediction performance on the validation data set. We find that, for the linear models, the optimal system order corresponds to approximately 25 hours or two full tidal cycles.

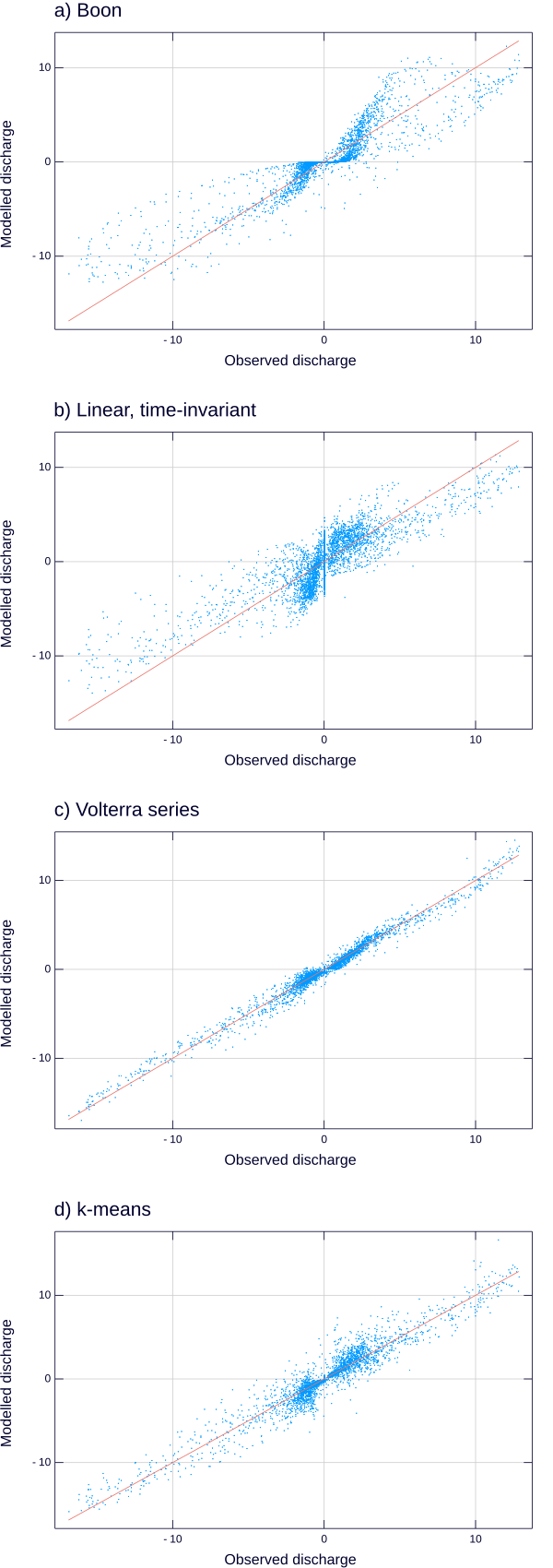
For the Volterra series, however, fewer lagged measurements of stage are required to predict the discharge, with an optimal system order around three hours. In estimating the Volterra series by a polynomial kernel regression, we exchange memory for degrees of nonlinearity as the number of parameters for each order of the Volterra operator scales as for a system order of and a Volterra operator order of . Given our finite data set, we will be able to estimate only a finite total number of these parameters, so using a higher system order -- a longer memory -- forces the order of the Volterra series down. And indeed the optimal Volterra order for a three-hour system order is 5, corresponding to polynomials up to quintics, while that for a 25-hour system is 3, corresponding to cubic polynomials.

The k-means model uses an unsupervised method to determine which cluster a new stage trajectory belongs to, so that the clustering is determined entirely by the shape of the stage signal. Two given stage trajectories will be closest in the Euclidean metric when they are perfectly in phase and farthest apart when they are perfectly out of phase, so any unsupervised clustering method using the Euclidean metric will naturally cluster based on the phase of the tidal signal, as we find in Fig. d. For a system order of 25 hours, the optimal number of clusters is around four, corresponding roughly to a low flood tide, a high flood tide, a high ebb tide and a low ebb tide.

We have found in practice, that the k-means clustering approach can be easily replaced by an ad hoc procedure which extracts the four clusters mentioned above simply using local information on the stage and stage derivative, making this approximation useful for real-time discharge estimation. The distinction between flood and ebb tides can be found where the time derivative of stage (approximated with the backward difference) changes sign. It is positive on the flood tides and negative on the ebb tides. The distinction between high and low stages can be based on a threshold, which we choose by cross-validation. The threshold selected by cross-validation for the Sweeney Creek stage-discharge time series is approximately 1.4 m which is just below the top of the right bank (Fig. b). This threshold indicates a fundamental difference between flow when the water level is above the bank, flooding the vegetated marsh surface and flow when the water is below the bank.

## Model performance

For each of the models (Boon, LTI, Volterra, k-means), we use cross validation to estimate the model with good choices for hyperparameters. We reestimate the model on the entire first time series using the good hyperparameters and apply each estimated model to our second stage-discharge time series and plot the modeled discharge values against the observed values in Fig. . The ideal modelled discharge values would lie on the red one-to-one line in Fig. . We report the Nash-Sutcliffe efficiency and the mean squared error of each model in Table to compare the prediction performance of the four models.



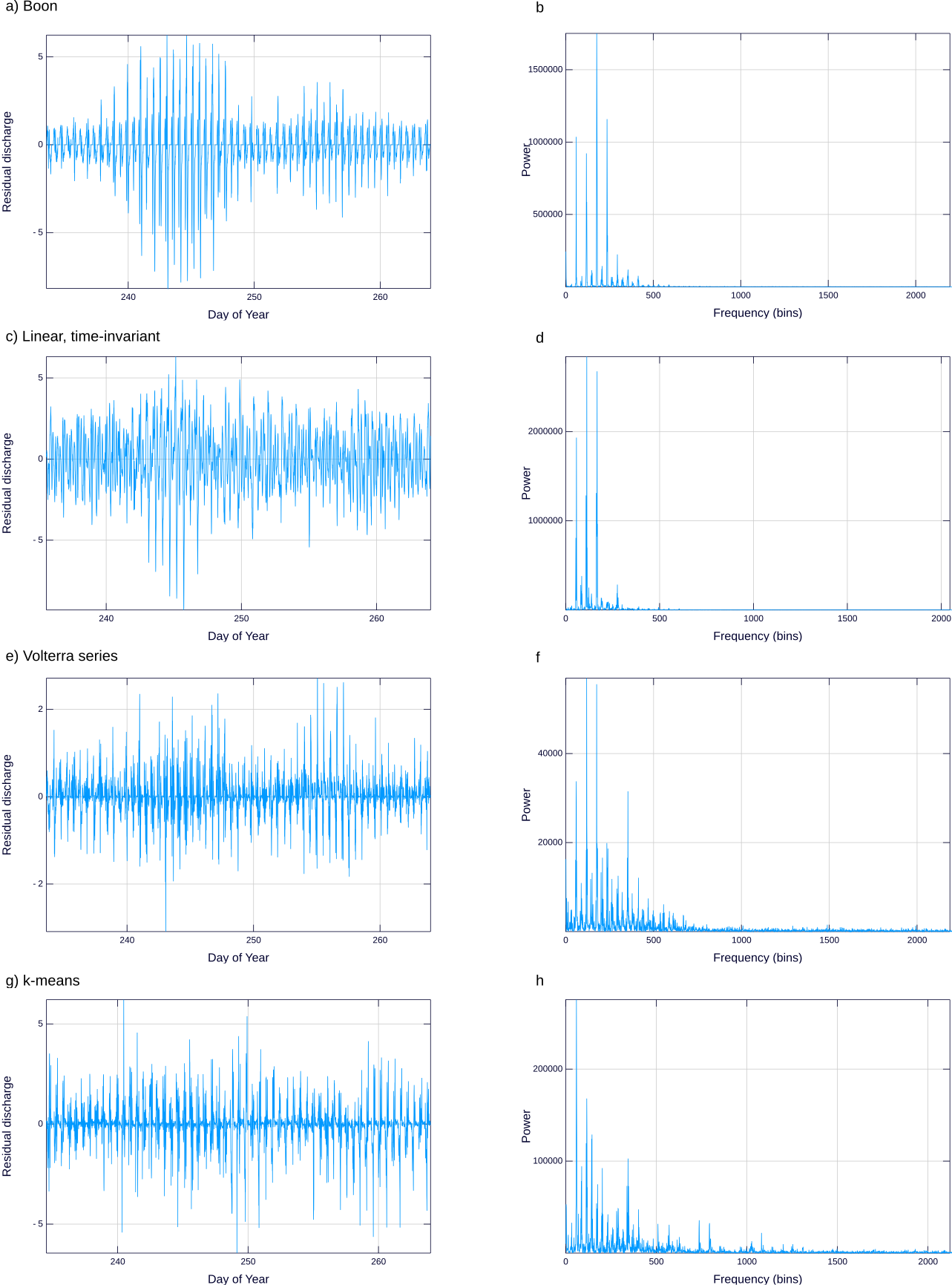
The modeled discharge plotted against the observed discharge. The red line in each plot is the one-to-one line.

The Volterra series model is the best performing (has the highest Nash-Sutcliffe efficiency and lowest mean squared error), followed by the k-means model, the Boon model and the linear, time-invariant model, a ranking which is supported by the visual representation of model fit, Fig. . Each of the four models tends to underestimate the high discharges and to overestimate the low discharges. At high magnitudes of the discharge, both positive and negative, points in Fig. tend to lie on the side of the one-to-one line closer to the x-axis, while at smaller discharges, they tend to lie on the side further from the x-axis. This effect is more pronounced in the more poorly performing models (Boon and linear, time-invariant).

## Residual structure

If our model completely captured the discharge-generating behavior of our salt marsh system, we would expect the residuals to be roughly independently distributed, in other words the error in the model comes not from systematically misestimating discharge at certain points of the time series but from random fluctuations in the velocity or from instrument noise. In addition to examining the fit of each model, we therefore also want to examine the structure present in the residuals. The predictive capability of two models being equal, we prefer the one with the least correlation in the residuals, or, in the frequency domain, the model with the flattest spectrum. We plot the residual time series and power spectra for each of the four models in Fig. . While we observe some structure in the residuals, it is hard to determine visually which of the models whitens the residuals the best. We would like a quantitative measure of the residual structure. The Ljung-Box test (Ljung and Box 1978) provides a statistical test of the the autocorrelation of the residual time series, but as we expect, the test rejects the null hypothesis of no autocorrelation for all of the models here, so the test itself does not adequately discriminate between the models. Instead, we use the spectral flatness (the ratio of the geometric mean of the power spectrum to the arithmetic mean) to measure how close to a white spectrum the residuals are. Flatness ranges from zero, at a signal with a single frequency, to one, at a purely white spectrum, so higher values of the spectral flatness indicate a better-specified model.

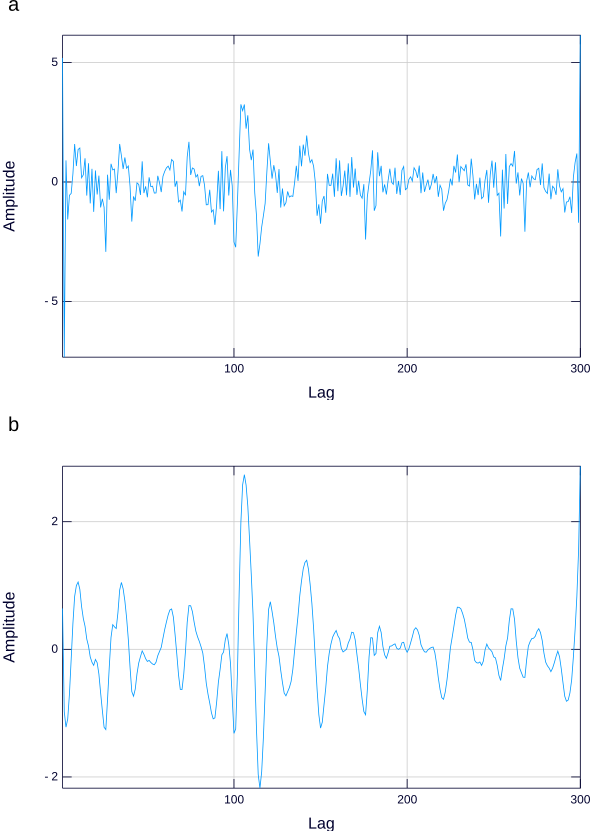
The estimated flatness of the residuals range from 0.022 for the linear, time-invariant model to 0.27 for the Volterra series model (Table ). These values suggest that the Volterra series model is the best specified model of the four.



The residual time series for each of the four classes of models: a) Boon, c) Linear, time-invariant, e) Volterra series g) k-means. The power spectrum of the residual time series for each of the four models b) Boon, d) Linear time-invariant, f) Volterra series, b) k-means.

## Effect of regularization

The unregularized linear, time-invariant impulse response is compared to that estimated with regularization in Fig. . We see that the effect of regularization is to smooth out the estimated coefficients. The main features of the response such as the high peak just after 100 lags (approximately 17 hours) are preserved in the regularized impulse response, but the finer scale oscillations are damped by the regularization. As the regularization parameter increases, lower and lower frequency oscillations are filtered out, and the resulting impulse response is smoother. Regularization improves the predictive ability of the linear, time-invariant model very slightly as measured by a larger out-of-sample Nash-Sutcliffe efficiency (from 0.40 to 0.645) and a smaller mean squared error (from 0.976 to 0.962).



a) The unregularized impulse response for the linear, time-invariant model. b) The regularized impulse response

The impact of regularization is much greater on the Volterra series model. The unregularized Volterra series parameters are a set of coefficients each corresponding to one of the data points in the training data set. The estimation procedure, as a result, is extremely sensitive to noise in the data -- the Gram matrix of the polynomial kernel is ill-conditioned -- and regularized is necessary to achieve any predictive ability with the model. When the fifth-order Volterra series model with 19 lags, the optimal model shown above, is estimated with no regularization (), the model is flatly unable to predict the discharge. The Nash-Sutcliffe efficiency is (note that negative Nash-Sutcliffe efficiencies correspond to models which predict the discharge worse than a constant model) while the mean squared error is (the respective values for the regularized model are 0.980 and 0.212). Also notable is the stark increase in the variance of the parameters, from to , and the corresponding inflated discharges, reaching as high as . For such a high-dimensional regression problem, regularization is absolutely essential. With regularization, however, the Volterra series performs the best of the four models examined here.

# Discussion

## Physical realism and stage-discharge models

The physical realism of each model roughly corresponds to its success in estimating the discharge. The Boon and linear, time-invariant models both perform fairly poorly in all of the measures examined (Table ). The Boon model is derived from a continuity law and is both nonlinear and nonstationary because of its dependence on the hypsometric curve. However, it has long been recognized as incapable of matching the asymmetry and hysteresis between flood and ebb tides because of its lack of memory. Only the slight asymmetry of the stage on the ebb and flood tides enables a discharge asymmetry. The linear, time-invariant model can generate asymmetry because it estimates discharge from the history of the stage over the course of two full tidal cycles. It is therefore aware of whether it is on a flood or an ebb tide and whether it is the higher or lower high tide of the day. The linearity and, more importantly, the stationarity of this model are nonphysical, and this lack of physical realism shows up in the performance of the model. The linear, time-invariant model systematically underpredicts very high discharges and overpredicts the low discharges because a single linear model is trained on the entire data set. It essentially aims to interpolate between the high and the low discharges which causes poor predictive performance on both, the worst performance of the four models.

The k-means model attempts to overcome this unphysical assumption of stationarity by estimating several different models and switching between the models throughout the tidal cycle. In doing so, it accounts somewhat for the nonlinearity problem as well. It segments the high-dimensional space of the stage trajectories into Voronoi cells and constructs a piecewise linear approximation to the nonlinear function which predicts discharge from stage trajectories. The piecewise linear approximation should converge to the true nonlinear function as the number of partitions increases, and the number of partitions is here limited mostly by the amount of data available for training. As a result of this ability, it performs significantly better than the first two models. The Volterra series, while time-invariant and, like the linear, time-invariant model, unable to account for nonstationarity, captures naturally the nonlinearity present in the shallow water equations, which ultimately govern the system. The spectral flatness results show that this model is the best specified of the four. The Volterra series model is a parametric nonlinear system, but the duality between the Volterra series and polynomial kernel regression means we estimate the series with the latter, a nonparametric estimator of the system response. Because the kernel regression is nonparametric, it is not restricted by our misspecification and, with infinite training data and appropriate regularization to reduce the effect of noise, we should be able to converge on as close an approximation to the true system as is possible with a time-invariant model.

regularization is straightforward to implement, and for the discharge estimation problem, it is sufficient for estimating effective parameters. However, it does not necessarily lead to straightforwardly interpretable model coefficients. The impulse response of the linear, time-invariant model, for example, is a combination of the travel-time distribution, the hypsometric curve and the action of the time derivative, all of which are approximations because of the assumptions of linearity and time-invariance. A more sophisticated regularization scheme would take into account knowledge of the behavior of these parameters -- such as the non-negativity and decaying tail of the travel-time distribution. If formulated carefully, these prior assumptions can be easily incorporated into the present regularization scheme by choosing an appropriate Tikhonov matrix (as in stable spline kernels (Pillonetto and De Nicolao 2010)). More complex prior assumptions such as sparsity of the impulse response coefficients can not be handled with the quadratic penalty term of Tikhonov regularization, but other frameworks exist for these alternative forms of regularization (Tibshirani 1996; Zou and Hastie 2005; Aravkin et al. 2013) and in a Bayesian formulation of the estimation problem, characterizing our physical assumptions on the models by an arbitrary prior distribution is a type of regularization.

## Limitations of these models

We have tested our models on stage-discharge records from a channel in a macrotidal salt marsh where the channel flow is almost entirely driven by regular tidal forcing. The models almost certainly do not work as well in environments with multiple drivers of flow such as microtidal channels with strong effects of wind on flow, tidally influenced streams with significant upland freshwater inputs, or loops in a channel network where the inputs and outputs do not flow through the same cross section. Future work will quantify which properties of our suite of models remain useful and what additional data might be necessary to extend this modeling framework to these other environments.

## Low flows and missing data

When the stage in the creek is below the first cell of the ADCP profile, no valid velocities are recorded by the instrument. While the velocities at these stages can be fast, the flooded cross-sectional area of the channel is very small, so the true discharges are also small. We fill these missing discharges with zeros, and we estimate all of the models on these zero-filled discharge time series. This imputation is likely to bias our discharge estimates (Little and Rubin 2002), and it certainly prevents us from consistently estimating the discharge during these low-flow periods. Volumes exchanged during these periods are small relative to the entire tidal prism, so the imputation with zeros has little impact on the estimated water balance of the marsh. If one is not particularly interested in the exact discharge during these periods, the Boon, Volterra series and k-means model are all able to estimate zero discharges during these periods. These low flows during ebb tides, however, represent slow drainage out of the marsh and creek system and so have the potential to transport significant amounts of nutrients from the marsh (Gardner 1975; Fagherazzi et al. 2013). If it is important to capture these effects or to quantify the uncertainty which results from imputation, more sophisticated modeling of the discharge at low stages is possible (Hopke et al. 2001).

# Comments and recommendations

## A simplified method to compute tidal discharges from water levels

Based on the results presented herein, we suggest the following simplified method to estimate discharge in tidal channels from water stage using the threshold-based approximation to the k-means model. Choose a threshold stage which corresponds to the elevation of the bank. If the left and right banks are asymmetric or there are multiple steps up to the marsh platform, choose the lowest bank elevation. Segment the time series into four groups: flood tide below the threshold, flood tide above the threshold, ebb tide below the threshold and ebb tide above the threshold. The flood/ebb distinction can be made quantitatively by taking differences between the current stage and the stage at the previous time step. These differences will be positive on the flood tide and negative on the ebb tide.

For each of the four groups of data, form a design matrix where each row represents a data point and each column contains the stage data from the previous time steps. That is, for row , the first column contains the stage at time step , , the second column contains , the third column and so on. The number of columns, , should cover two whole tides. At the 10 minute sampling interval of the time series presented here, this is approximately time steps, resulting in a design matrix with 150 columns. If the time series is at a different sampling interval, change the width of the design matrix accordingly.

One should now have a design matrix for each of the four time series segments , , and , and four vectors of discharge values , , and each of which contains the corresponding discharge values for each of the data points. The coefficients of the model are the four vectors which can be obtained with standard routines for linear regression. Once the four vectors of coefficients are obtained, prediction of discharge at a new point proceeds by first deciding to which of the four groups (high flood, low flood, high ebb, low ebb) the water stage belongs. Each of the previous time steps of the stage is then multiplied by each of the model coefficients of the corresponding group and added together to provide an estimate of discharge.

## Model recommendations

The complexity of estimating each of these models tracks closely their performance. The linear, time-invariant model is a straightforward linear regression, but it performs the worst (as measured by any of our error measures presented in Table ). The Boon model (as formulated here) requires a nonlinear least squares algorithm but does significantly better. The k-means model has a mean squared error half that of the Boon model, but requires some clustering either through k-means or the simplified threshold model presented above. The Volterra model performs the best of all four models but requires a computationally-intensive kernel regression. Choosing between the models is an exercise in trading off complexity for predictive ability and requires a rigorously defined selection criterion adapted to the particular application. We have used the mean squared error, Nash-Sutcliffe efficiency and spectral flatness of residuals to argue that the cubic Volterra series model with 25 hours of lagged stage observations performs the best of the four models. However, each of these measures simply reflects the discrepancy between modeled and observed instantaneous discharges which may not be appropriate for all applications. One could envision the integrated volume of water over a tide being more important than the instantaneous discharge, in which case it might be worth selecting model which slightly misestimates the discharge to get a more accurate estimate of the tidal prism. They also do not take into account the complexity of each model.

To help quantify the tradeoff between complexity and performance for applications, we have calculated the mean absolute percent error for each model as a function of stage (Fig. ). We bin the stage into 50 cm bins and calculate the mean of the absolute value of the percent error between the modeled and estimated discharge within each bin. This gives some estimate of how far off one might expect to be when using each model to predict discharge over a certain range of stages. The general pattern follows our conclusions from the other measures of the model error with the Volterra series model performing the best, followed by k-means, Boon and the linear, time-invariant model. The Volterra series percent error is around 10-15% at all stages, while the k-means percent error ranges from around 20-30%. While the Boon model has a percent error around 50% at high and low stages, it is within one percent at stages just above the bankfull stage for our channel. If one is interested in estimating only the bankfull discharge in a channel, the Boon model performs just as well as the significantly more complex k-means model.

Mean absolute percent error for each of the four models as a function of stage. 

Mean absolute percent error for each of the four models as a function of stage.

The k-means model, and especially the threshold version represents, we believe, the best model for applications which need to estimate discharge from long-term records of stage such as biogeochemical and ecological investigations. It offers good estimation performance throughout a long time series, because its estimation complexity resides in the selection of clusters, which, as we have shown, can be well-approximated with a heuristic, and because of the appealing interpretation of the clusters in terms of flow regimes.

# Acknowledgements

This research was supported by the National Science Foundation (awards OCE1354251, OCE1354494, and OCE1238212). Data are publicly available through the Long Term Ecological Research (LTER) Network Data Portal at http://portal.lternet.edu. We would like to thank Neil Ganju for helpful comments on a draft of this manuscript.

# Tables

ADCP parameters

|  |  |
| --- | --- |
| Parameter | Value |
| Acoustic frequency | 2.0 MHz |
| Blanking distance | 10 cm |
| Cell size | 20 cm |
| Sampling interval | 10 min |

Performance of the models

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Mean squared error | Nash-Sutcliffe efficiency | Spectral flatness |
| Boon | 0.483 | 0.816 | 0.041 |
| Linear, time-invariant | 0.962 | 0.645 | 0.022 |
| Volterra series | 0.052 | 0.980 | 0.272 |
| k-means | 0.245 | 0.910 | 0.257 |

# References

Aravkin, A. Y., J. V. Burke, and G. Pillonetto. 2013. Linear system identification using stable spline kernels and PLQ penalties. 52nd IEEE Conference on Decision and Control. doi:[10.1109/cdc.2013.6760701](https://doi.org/10.1109/cdc.2013.6760701)

Beven, K., and J. Davies. 2015. Velocities, celerities and the basin of attraction in catchment response. Hydrological Processes **29**: 5214–5226. doi:[10.1002/Hyp.10699](https://doi.org/10.1002/Hyp.10699)

Blanton, J. O., G. Lin, and S. A. Elston. 2002. Tidal current asymmetry in shallow estuaries and tidal creeks. Continental Shelf Research **22**: 1731–1743. doi:[10.1016/S0278-4343(02)00035-3](https://doi.org/10.1016/S0278-4343(02)00035-3)

Boon, J. D. 1975. Tidal discharge asymmetry in a salt marsh drainage system. Limnology and Oceanography **20**: 71–80. doi:[10.4319/lo.1975.20.1.0071](https://doi.org/10.4319/lo.1975.20.1.0071)

Botter, G., E. Bertuzzo, and A. Rinaldo. 2010. Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox. Water Resources Research **46**: W03514. doi:[10.1029/2009wr008371](https://doi.org/10.1029/2009wr008371)

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference, Springer.

Bühlmann, P. 2002. Bootstraps for time series. Statist. Sci. **17**: 52–72. doi:[10.1214/ss/1023798998](https://doi.org/10.1214/ss/1023798998)

Cai, W.-J. 2011. Estuarine and coastal ocean carbon paradox: CO2 sinks or sites of terrestrial carbon incineration? Ann. Rev. Marine. Sci. **3**: 123–145. doi:[10.1146/Annurev-Marine-120709-142723](https://doi.org/10.1146/Annurev-Marine-120709-142723)

Carey, J. C., and R. W. Fulweiler. 2014. Salt marsh tidal exchange increases residence time of silica in estuaries. Limnology and Oceanography **59**: 1203–1212. doi:[10.4319/lo.2014.59.4.1203](https://doi.org/10.4319/lo.2014.59.4.1203)

Chmura, G. L., S. C. Anisfeld, D. R. Cahoon, and J. C. Lynch. 2003. Global carbon sequestration in tidal, saline wetland soils. Global Biogeochemical Cycles **17**: 1111. doi:[10.1029/2002GB001917](https://doi.org/10.1029/2002GB001917)

Duarte, C. M., J. J. Middelburg, and N. Caraco. 2005. Major role of marine vegetation on the oceanic carbon cycle. Biogeosciences **2**: 1–8. doi:[10.5194/Bg-2-1-2005](https://doi.org/10.5194/Bg-2-1-2005)

Fagherazzi, S., M. Hannion, and P. D’Odorico. 2008. Geomorphic structure of tidal hydrodynamics in salt marsh creeks. Water Resources Research **44**: W02419. doi:[10.1029/2007wr006289](https://doi.org/10.1029/2007wr006289)

Fagherazzi, S., P. L. Wiberg, S. Temmerman, E. Struyf, Y. Zhao, and P. A. Raymond. 2013. Fluxes of water, sediments, and biogeochemical compounds in salt marshes. Ecological Processes **2**: 3. doi:[10.1186/2192-1709-2-3](https://doi.org/10.1186/2192-1709-2-3)

Franz, M. O., and B. Schölkopf. 2006. A unifying view of Wiener and Volterra theory and polynomial kernel regression. Neural Computation **18**: 3097–3118. doi:[10.1162/Neco.2006.18.12.3097](https://doi.org/10.1162/Neco.2006.18.12.3097)

Ganju, N. K., N. J. Nidzieko, and M. L. Kirwan. 2013. Inferring tidal wetland stability from channel sediment fluxes: Observations and a conceptual model. Journal of Geophysical Research: Earth Surface **118**: 2045–2058. doi:[10.1002/jgrf.20143](https://doi.org/10.1002/jgrf.20143)

Gardner, L. R. 1975. Runoff from an intertidal marsh during tidal exposure-recession curves and chemical characteristics. Limnology and Oceanography **20**: 81–89. doi:[10.4319/Lo.1975.20.1.0081](https://doi.org/10.4319/Lo.1975.20.1.0081)

Ghorbani, M. A., R. Khatibi, A. Goel, M. H. Fazelifard, and A. Azani. 2016. Modeling river discharge time series using support vector machine and artificial neural networks. Environmental Earth Sciences **75**. doi:[10.1007/S12665-016-5435-6](https://doi.org/10.1007/S12665-016-5435-6)

Harman, C. J. 2015. Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. Water Resources Research **51**: 1–30. doi:[10.1002/2014wr015707](https://doi.org/10.1002/2014wr015707)

Healey, R., K. Pye, D. Stoddart, and T. Bayliss-Smith. 1981. Velocity variations in salt marsh creeks, norfolk, england. Estuarine, Coastal and Shelf Science **13**: 535–545. doi:[10.1016/S0302-3524(81)80056-4](https://doi.org/10.1016/S0302-3524(81)80056-4)

Hocking, R. R. 1976. The analysis and selection of variables in linear regression. Biometrics **32**: 1–49.

Hopke, P. K., C. Liu, and D. B. Rubin. 2001. Multiple imputation for multivariate data with missing and below-threshold measurements: Time-series concentrations of pollutants in the arctic. Biometrics **57**: 22–33. doi:[10.1111/j.0006-341x.2001.00022.x](https://doi.org/10.1111/j.0006-341x.2001.00022.x)

Kelley, C. T. 1999. Iterative methods for optimization, Society for Industrial & Applied Mathematics (SIAM).

Kennedy, E. J. 1984. Discharge ratings and gaging stations. U.S. Geological Survey Techniques of Water-Resources Investigations.

Little, R. J. A., and D. B. Rubin. 2002. Statistical analysis with missing data, 2nd ed. Wiley-Blackwell.

Ljung, G. M., and G. E. P. Box. 1978. On a measure of lack of fit in time series models. Biometrika **65**: 297–303. doi:[10.1093/biomet/65.2.297](https://doi.org/10.1093/biomet/65.2.297)

Morris, J. T., P. V. Sundareshwar, C. T. Nietch, B. Kjerfve, and D. R. Cahoon. 2002. Responses of coastal wetlands to rising sea level. Ecology **83**: 2869–2877. doi:[10.1890/0012-9658(2002)083[2869:ROCWTR]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2869:ROCWTR]2.0.CO;2)

Pethick, J. 1980. Velocity surges and asymmetry in tidal channels. Estuarine and Coastal Marine Science **11**: 331–345. doi:[10.1016/S0302-3524(80)80087-9](https://doi.org/10.1016/S0302-3524(80)80087-9)

Pillonetto, G., and G. De Nicolao. 2010. A new kernel-based approach for linear system identification. Automatica **46**: 81–93. doi:[10.1016/J.Automatica.2009.10.031](https://doi.org/10.1016/J.Automatica.2009.10.031)

Rugh, W. J. 1981. Nonlinear system theory, Johns Hopkins University Press Baltimore.

Ruhl, C. A., and M. R. Simpson. 2005. Computation of discharge using the index-velocity method in tidally affected areas. US Geological Survey.

Sahoo, G., and C. Ray. 2006. Flow forecasting for a Hawaii stream using rating curves and neural networks. Journal of Hydrology **317**: 63–80. doi:[10.1016/j.jhydrol.2005.05.008](https://doi.org/10.1016/j.jhydrol.2005.05.008)

Speer, P., and D. Aubrey. 1985. A study of non-linear tidal propagation in shallow inlet/estuarine systems. Part II: Theory. Estuarine, Coastal and Shelf Science **21**: 207–224. doi:[10.1016/0272-7714(85)90097-6](https://doi.org/10.1016/0272-7714(85)90097-6)

Sudheer, K., and S. Jain. 2003. Radial basis function neural network for modeling rating curves. Journal of Hydrologic Engineering **8**: 161–164.

Tawfik, M., A. Ibrahim, and H. Fahmy. 1997. Hysteresis sensitive neural network for modeling rating curves. Journal of Computing in Civil Engineering **11**: 206–211.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 267–288.

Wold, S., A. Ruhe, H. Wold, and W. Dunn III. 1984. The collinearity problem in linear regression. the partial least squares (pLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing **5**: 735–743.

Xu, R., and D. Wunsch. 2009. Clustering, John Wiley & Sons.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**: 301–320.