Functional regression influence measures for out-of-sample prediction

Ryan D. Pittman[1*] and David B. Hitchcock[2]

[1*]Department of Statistics, University of South Carolina, Columbia, 29208, SC, USA.
[2]Department of Statistics, University of South Carolina, Columbia, 29208, SC, USA.

*Corresponding author(s). E-mail(s): rpittman@email.sc.edu;
Contributing authors: hitchcock@stat.sc.edu;

Abstract

In linear regression, it is common to calculate measures of influence for each observation in the sample to understand the impact it has on the regression model. Typically, such measures are used to assess the effects of observations on predictions of the response for observations within the sample. However, in some scenarios, the primary interest is in predictions for out-of-sample observations. In this study, we present two measures of influence on out-of-sample predictions which apply to the concurrent functional regression model, in which both the predictor and response are functional observations. These two measures, which we call $\boldsymbol{\Delta}$ and Accumulated Influence Percentiles (AIP), describe how the prediction of an out-of-sample observation's response function changes when the model is fit with and without each of the functional observations in turn. We also describe a weighted bootstrapping method to assess formally whether an observation within the sample has significant influence on the out-of-sample prediction. We present a simulation study to illustrate the effectiveness of the method, and two relevant real data examples (on river stage functions and on air and water temperature functions) further illustrate the method.

Keywords: Functional data analysis, Out-of-sample influence measures, Concurrent model, Bootstrapping, River stage, Weather data

1 Introduction

In traditional regression, it is common to measure the influence an observation has on a fitted model. Traditional measures like DFFITS [1] and Cook's distance [2] to measure the influence an observation has on an in-sample prediction of that observation's response (or of all responses in the sample). The concurrent linear model for functional data involves using one set of (predictor) curves $X_i(t), i = 1, \ldots, N$ to predict or explain a corresponding set of (response) curves $Y_i(t), i = 1, \ldots, N$. Extensions of traditional measures of influence to the functional-data setting have been explored by Shen and Xu [3], Chiou and Möller [4], Febrero-Bande et al. [5], Chen et al. [6], and Pittman and Hitchcock [7]. In this article, we present two new measures of influence, which we call Δ and Accumulated Influence Percentiles (AIP), for use with the concurrent functional regression model, which calculate how the prediction of an out-of-sample response curve of interest, say $\hat{Y}^{new}(t)$, changes when the concurrent functional model is fit with and without each functional observation in turn. This is different than the traditional influence measures because it examines how the ith observation influences the prediction of a new target response curve $\hat{Y}^{new}(t)$, so that this out-of-sample observation is different from those used to fit the model. In practice, knowing which observations have significant influence on a specific new prediction is valuable to construct reliable predictions of a new response curve in the functional regression setting. We derive Δ_i and AIP_i for the ith functional observation $i = 1, \ldots, N$, where larger values indicate that the observation has higher influence when predicting an external observation's response curve. While Δ_i is a metric that indicates whether or not an entire observation is influential in the prediction by measuring the magnitude of the change, AIP_i is similar but takes into account the duration of the influence spanning throughout the functional observation. Section 2 describes both measures in detail, and Section 3 proposes a weighted bootstrap method to determine whether an observation's measure is significantly large. Section 4 discusses a simulation study that illustrates the effectiveness of our methods. We then apply our proposed measures to two real datasets in Sections 5 and 6.

2 Functional Regression Influence Measures

2.1 Influence Measure: $\boldsymbol{\Delta_i}$

The concurrent functional regression model has model equation $Y_i(t) = \beta_0(t) + \beta_1(t)X_i(t) + \epsilon_i(t),\ i = 1, \ldots, N$ where $\beta_0(t)$ is the functional intercept and $\beta_1(t)$ is the functional slope that relates predictor $X_i(t)$ and response $Y_i(t)$ at each t. Estimates of $\beta_0(t)$ and $\beta_1(t)$ are computed using all N pairs of functional data (which are represented via basis functions), using the `fRegress` function in the `fda` package [8] in `R` [9]. The resulting $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ are used in the fitted functional regression model, along with the external observation's known predictor curve $X^{new}(t)$, to calculate the prediction $\hat{Y}^{new}(t)$. The formula for

Δ_i is then

$$\Delta_i = \left[\int (\hat{Y}^{new}(t) - \hat{Y}^{new}_{(i)}(t))^p dt \right]^{1/p}. \quad (1)$$

$\hat{Y}^{new}(t)$ is the predicted value for a new response curve using all N sets of $X_i(t)$ and $Y_i(t)$ observations. $\hat{Y}^{new}_{(i)}(t)$ is that same prediction from a model fit with observation i withheld. In practice, we use $p = 2$ for an L_2 measure and we approximate this using trapezoidal approximation via the `trapz` function in the `pracma` package [10] in `R`. The observation with the highest Δ_i value has the most overall influence on the prediction of $\hat{Y}^{new}(t)$ and should be further investigated, especially if this value is much larger than the next highest Δ_i value. In section 4.2, we describe a weighted bootstrapping approach to approximate the null distribution of Δ for a particular out-of-sample target observation, which can be used to determine whether any observed Δ_i measures are significantly large.

2.2 Influence Measure: Accumulated Influence Percentiles (AIP_i)

The previously discussed measure calculates the overall influence each observation has on an out-of-sample prediction. Our second influence measure, which we call Accumulated Influence Percentiles (AIP), accounts for whether the difference (based on fits with and without each sample observation) in the prediction of an external observation is entirely at a single portion of the functional observation or spread throughout it, thus combining the magnitude of the difference with the duration of the difference. The formula for AIP_i is:

$$AIP_i = \int_0^1 \phi_p \left(\left| \hat{Y}^{new}(\mathbf{t}) - \hat{Y}^{new}_{(i)}(\mathbf{t}) \right| \right) dp, \quad (2)$$

where $\mathbf{t} = (t_1, \ldots, t_n)$ and $\phi_p(\mathbf{x})$ is the pth percentile of the values in the vector $\mathbf{x}$. If an observation has a large Δ, but it is caused by a large difference at a single portion of the prediction, then the resulting percentile curve of absolute differences will be relatively flat across $p \in [0, 1]$ and increase sharply at the higher percentiles. Taking the area under the curve minimizes the effect of a single large difference and balances the magnitude of the difference and the duration of the difference.

3 Approximate null distributions of Δ and AIP

To determine the significance of the observed influence measures, we employ a weighted bootstrapping method introduced by Pittman and Hitchcock [7] that approximates null distributions for each of Δ and AIP, i.e. a distribution for each metric under the condition that there is no especially influential curve. To accomplish this, when selecting our bootstrap sample we propose to sample the apparently less influential observations from our observed curves more often than the apparently most influential observations. We calculate the influence

measure Δ_i or AIP_i (generally denoted as r_i) for each observation, and then use the following equation to translate the metric value for observation i into a selection probability θ_i:

$$\theta_i = \frac{(1/r_i)^\alpha}{\sum_i [(1/r_i)^\alpha]}, \quad \alpha \geq 0. \tag{3}$$

Note that $\alpha = 0$ corresponds to equal selection probabilities for each observation. In practice, the tuning parameter α should not exceed 0.5 and is most crucial when N is small. The following weighted bootstrap method will approximate a null distribution of the measure:

1. Calculate r_i for each observation.
2. Select an appropriate value of α (or allow a range of choices) and calculate θ_i for $i = 1, \ldots, N$.
3. Sample N paired observations with replacement from the original set of data, where the ith observation has probability θ_i of being selected. Each bootstrap sample then consists of N functional pairs $\{(X_1^*(t), Y_1^*(t)), \ldots, (X_N^*(t), Y_N^*(t))\}$.
4. Using these new pairs of functional data, fit the concurrent functional regression model and calculate r_i for each observation $i = 1, \ldots, N$.
5. Repeat Steps 3-4 for the desired number of bootstrap iterations (B) to obtain NB values of the measure, which approximate draws from a null distribution for that influence measure.

Values of the metric from the originally observed dataset can then be compared to percentiles from the respective bootstrap distribution to determine whether the largest values identified in the original data analysis are significantly large relative to the null distribution. The ideal value of α in Equation (3) varies based on the observed measures from the initial dataset. In general, we recommend using $\alpha = 0.5$ when N is small or when one of the observed influence measures is noticeably larger than the rest. If values of the metric have little variability, the bootstrapped percentiles will be similar regardless of $\alpha \in (0, 0.5)$; however, when the observed influence measures are more spread out or one observation's influence measure is much larger than the rest, using $\alpha = 0.5$ dampens the effect that this observation has on the bootstrap sample and prevents the bootstrap distribution from being dominated by the values for the most influential observations, resulting a bootstrap sample that better resembles a null distribution. This allows truly significant influential observations to be flagged as extreme. For large sample sizes, an observation with a large influence measure has less impact on the approximate null distribution since it is less likely to be sampled in a given iteration (regardless of the value of α) compared to when the sample size is small; therefore, using $\alpha = 0$ in large sample scenarios is appropriate. If the sample size is moderate, or it is unclear whether the largest influence measure is much larger than the next highest, we recommend using both $\alpha = 0$ and $\alpha = 0.5$ separately and comparing the resulting percentiles to see the effect of the more influential observations.

After we perform this bootstrapping method, the significantly influential functional observations can be identified by comparing the observed measures from the original dataset to the upper percentiles of the null distribution.

4 Simulation Study

To elucidate the effectiveness of these new influence measures and the bootstrap method, we perform a simulation study in which we generate a sample of paired functional predictor and response observations and intentionally contaminate one of the response curves. After generating an additional predictor curve which corresponds to a hypothetical out-of-sample target observation, we calculate for each sample observation our measures of influence on the prediction of the response curve of the target observation. Once we have the influence measure values for each observation, we perform our bootstrap method to determine where these values fall within the null distribution.

This simulation study is constructed in a similar manner to the one presented by Pittman and Hitchcock [7], which investigated functional regression measures of influence on in-sample prediction and estimation of regression coefficients as opposed to the out-of-sample prediction that we focus on here. We generate as predictor functions N independent $X(t)$ curves over a grid of values $t \in \{1, 2, \ldots, 1000\}$ using the following formula:

$$X(t) = \frac{t}{12} + \left[a_s \sin\left[\frac{t - d_s}{k_s}\right] + c_s\right]\left[a_c \cos\left[\frac{t - d_c}{k_c}\right] + c_c\right]$$

where each of the N curves is generated by randomly selecting the parameters within the equation as follows: (1) a_s, a_c, c_s and c_c are independently sampled from the list $\{-3, -2, -1, 0, 1, 2, 3\}$; (2) k_s and k_c are sampled from the list $\{-300, -200, -100, 100, 200, 300\}$; and (3) d_s and d_c are sampled from the list $\{-100, -50, 0, 50, 100\}$. By varying the combinations of parameters used to generate the functional data, we produce curves that are similar, following the same underlying signal curve $m(t) = t/12$, but not exactly the same. An example of $N = 20$ such $X(t)$ curves is shown in Figure 1 (top). Note that the simulation results are not changed if the parameters' ranges are expanded as long as they are the same for all N curves. We set the functional slope and intercept functions to be $\beta_0(t) = \cos(t/200) + 2$ and $\beta_1(t) = \sin(t/200) + 2$.

We generate response signal curves $Y_i(t) = \beta_0(t) + \beta_1(t)X_i(t)$, $i = 1, \ldots, N$, and then we coarsen the relationship between the predictor and response curves by generating noise functions $\epsilon_i(t)$ to slightly distort the functional relationship between each pair of simulated $X(t)$ and $Y(t)$ curves. Specifically, we add to the signal curves realizations of the Ornstein-Uhlenbeck process [11], approximated using the Euler-Maruyama method. An example of a resulting set of 20 simulated response curves is shown in Figure 1 (center).

As an initial check that these generated data follow our concurrent functional linear model, before introducing any contamination, we generated 100 independent data sets, fit the model for each, and verified the estimates of

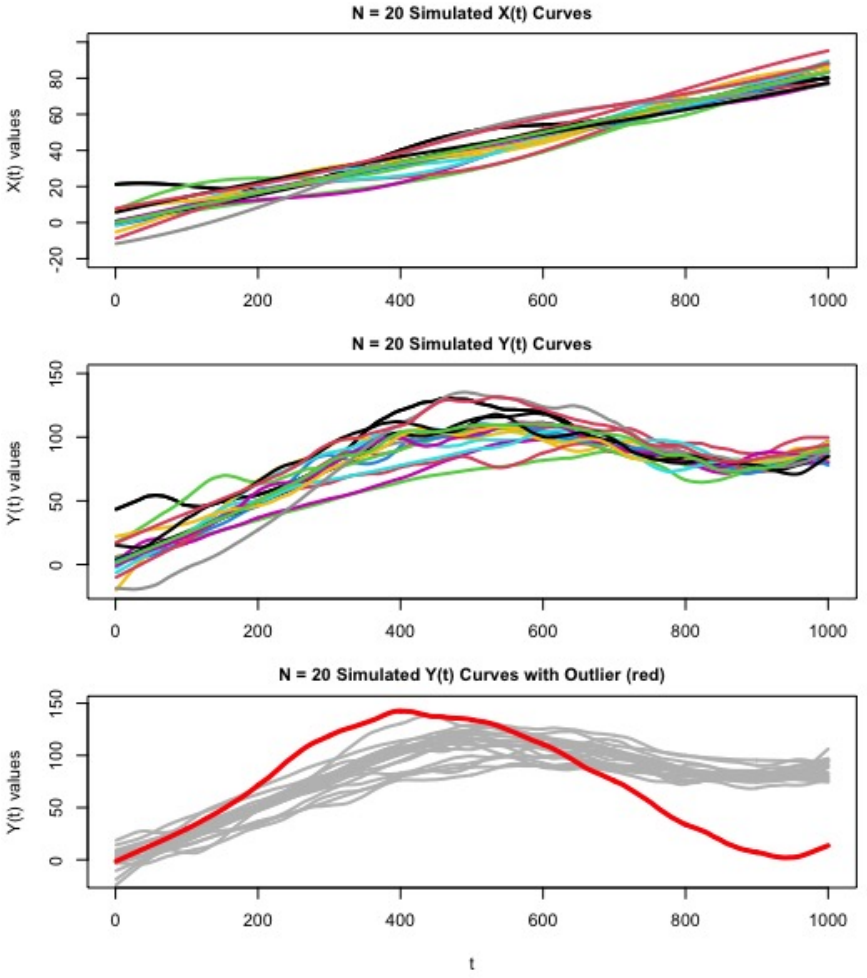

Fig. 1: Top: $N = 20$ generated $X(t)$ curves using the described functional data generation method. Center: $N = 20$ response $(Y(t))$ curves used in simulation with no contaminated observations ($\lambda = 1$). Bottom: $N = 20$ response $(Y(t))$ curves used in simulation with one outlier (red) using $\lambda = 2$.

$\beta_0(t)$ and $\beta_1(t)$ resembled the true functional slope and intercept on average. However, all further analysis was done on simulated data with a contaminated curve, as described next.

We intentionally contaminated the $\beta_1(t)$ function for one of the N observations, setting $\beta_1(t) = \lambda \times \sin(t/200) + 2$ for some $\lambda > 0$ for this curve. Clearly, $\lambda = 1$ represents the control case in which the contaminated observation is generated the same way as the others. In this simulation, we set $\lambda \in \{0.25,$

0.5, 0.75, 0.9, 1, 1.1, 1.25, 1.5, 1.75, 2}. Figure 1 (bottom) gives an example of $N = 20$ response curves with the contaminated curve generated using $\lambda = 2$.

We implemented the following algorithm for combinations of: $N = 100$, $N = 50$, $N = 20$, and $N = 10$; $\lambda \in \{0.25, 0.5, 0.75, 0.9, 1, 1.1, 1.25, 1.5, 1.75, 2\}$; and for $\alpha = 0, 0.5$.

1. Select λ.
2. Generate N sets of $\{X_i(t), Y_i(t)\}$ curves with one $Y_i(t)$ curve contaminated using λ.
3. Generate the predictor curve of an out-of-sample target observation.
4. For curve $i = 1, \ldots, N$, calculate the measure of influence (Δ_i or AIP_i) on the prediction of the response curve corresponding to the target observation.
5. For the selected α, calculate the selection probabilities θ_i for each observation using Equation (3).
6. Perform $B = 100$ bootstrap iterations, sampling the N observations with replacement, calculating the influence measure for each observation in each iteration (yielding NB values of the measure).
7. Determine the percentile relative to this bootstrap distribution of the originally contaminated observation's influence measure, checking whether it is above the 95th percentile.
8. Repeat 100 times for each combination of the influence measure, N, λ, and α.

Note that for each data generation, the bootstrapping process is executed using each choice of α on the same generated data.

With our simulated data, we tested whether the Δ measure successfully identifies the contaminated observation as influential, on average.

Figure 2 shows the average proportion of contaminated observations that were above the 95th percentile for Δ_i for $N \in \{10, 20, 50, 100\}$. This is analogous to the power of the procedure at a 0.05 significance level. When λ moved away from 1, the proportion of contaminated observations flagged increased. This correctly indicates that when an observation is more extreme, it is flagged as influential more often. Additionally, when N is small, this test had a substantially higher power when $\alpha = 0.5$, indicating that choosing $\alpha > 0$ is desirable for small N.

Furthermore, Figure 3 provides the average p-value, which is 1 minus the average percentile within the bootstrap distribution of the contaminated observation. When λ moved away from 1, the p-value decreased, indicating that the contaminated observation's influence measure was frequently significant. These results consistently indicate that our method successfully identifies the observation that is truly influential on an out-of-sample prediction and that with a small sample size, using $\alpha = 0.5$ improves the strength of the method.

Our simulation study also established the validity of AIP in identifying observations influential on out-of-sample prediction. Figure 4 and Figure 5 show the average estimated power and p-value of the AIP measure.

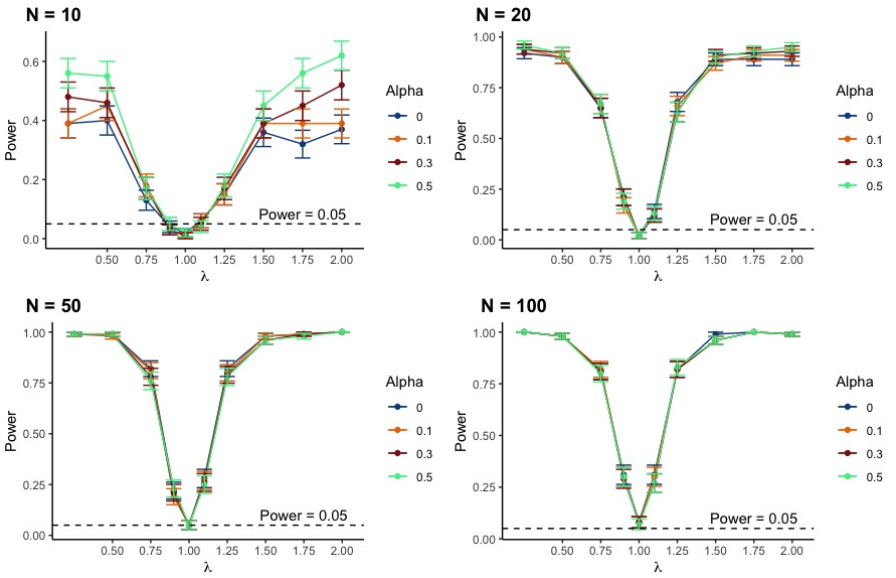

Fig. 2: Power functions displaying the average proportion of contaminated observations above the 95th percentile from the approximate null distribution of Δ for different values of α (with error bars representing one standard error) for different sample sizes N.

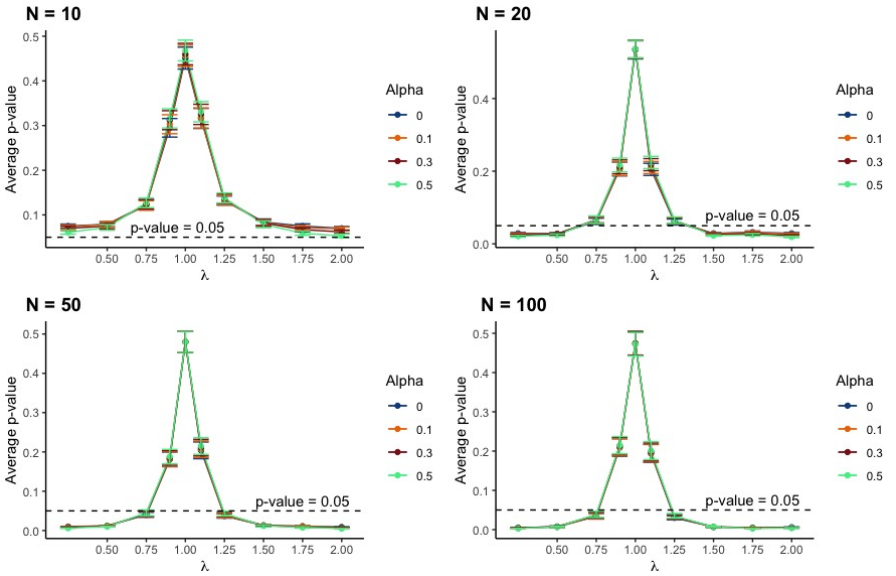

Fig. 3: Average p-value (1− percentile within bootstrap distribution) of contaminated observations for different values of α (with error bars representing one standard error) for different sample sizes N for Δ.

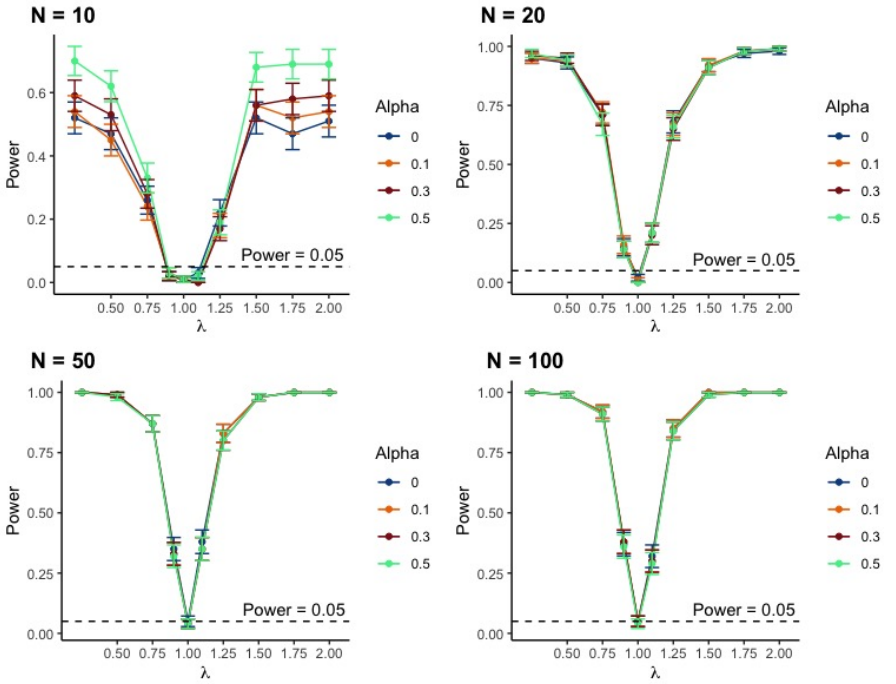

Fig. 4: Power functions displaying the average proportion of contaminated observations above the 95th percentile from the approximate null distribution of AIP for different values of α (with error bars representing one standard error) for different sample size N.

Figure 4 shows the average proportion (i.e., power) of contaminated observations that were above the 95th percentile of the approximate null distribution of AIP for $N \in \{10, 20, 50, 100\}$. When λ moved away from 1, the proportion of contaminated observations flagged increased. This indicates that when an observation is more extreme, it is appropriately flagged as influential more often. Additionally, when N is small, this test had a substantially higher power when $\alpha = 0.5$, again indicating that setting $\alpha > 0$ was useful in this case.

Figure 5 provides plots of the average p-value. When λ moved away from 1, the p-value decreased, indicating that the contaminated observation's AIP was often significantly large. These results consistently indicate that our new influence measure AIP successfully identifies the observation that is truly influential on an out-of-sample prediction.

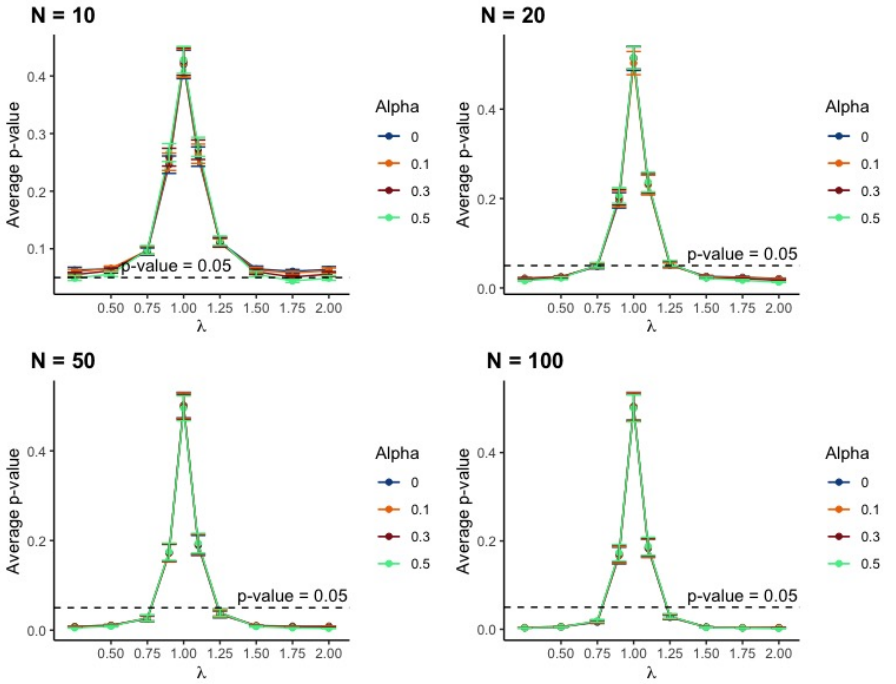

Fig. 5: Average p-value (1− percentile within bootstrap distribution) of contaminated observations for different values of α (with error bars representing one standard error) for different sample sizes N for AIP.

5 Application: River Stage Data During Flood Events

We applied these measures to river stage data studied by Pittman et al. [12], who analyzed river stages (heights) from two related gage locations at Congaree National Park near Columbia, South Carolina. They used a landmark alignment technique to objectively determine the optimal start and end points of ten flood events in which the Congaree River [13] flowed over-bank, through the floodplains, and into Cedar Creek [14]. This resulted in 10 historic flood events that could be used as paired observations in the concurrent functional model. The purpose of the functional regression was to relate the Congaree River stage to the Cedar Creek stage during flood events, in order to reconstruct the Cedar Creek stage during a major flood event in October 2015 when the Cedar Creek gage went offline but the Congaree River gage remained functioning. The Δ influence measure will determine the 10 historic flood events' influence on the eventual reconstruction of the October 2015 Cedar Creek stage and whether any of these events' influence measures are significantly larger than expected.

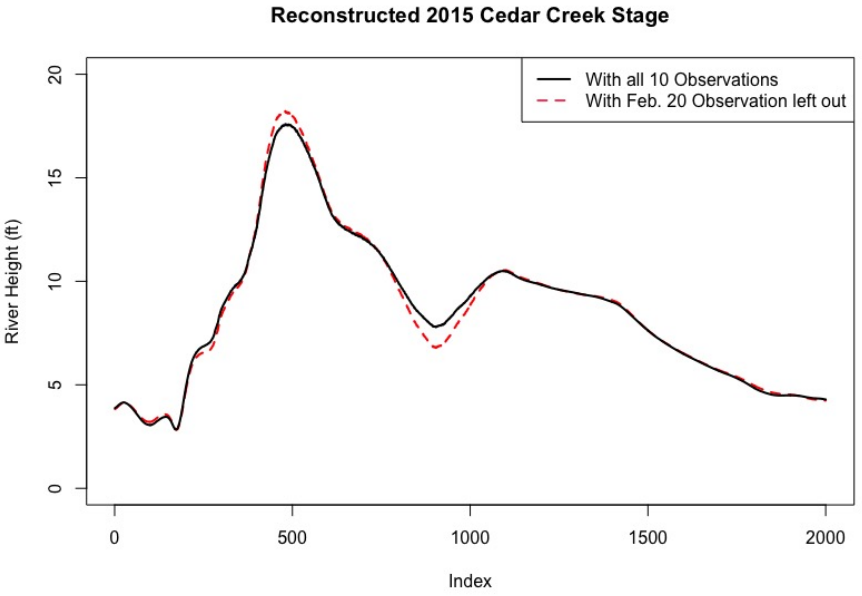

Fig. 6: The reconstructed October 2015 Cedar Creek stage with all 10 observations (solid black) and with the February 2020 observation withheld (dashed red).

We first used the concurrent model fitted on the full data set to reconstruct the missing Cedar Creek stage during the October 2015 flood. We then repeated the fits and reconstructions, leaving out each of the functional observations in turn. Figure 6 provides an example of the difference between the reconstruction $\hat{Y}^{new}(t)$ based on the full data and the reconstruction $\hat{Y}^{new}_{(10)}(t)$ based on data with the February 2020 event withheld. Δ was the L_2 distance between the two curves based on Equation (1). The resulting Δ_i for each flood event i is given in Table 1 and plots of the October 2015 Cedar Creek reconstruction based on fits with and without each event can be found in the Supplementary Material.

The March 2003 and February 2020 flood events stand out with Δ_i values about twice as large as the next highest values, indicating that these two events have the greatest impact on the target event's reconstruction.

Given our small sample size ($N = 10$), we used $\alpha = 0.5$ when we approximated the null distribution. The March 2003 event had the highest Δ (14.165), but this was not as large as the 90th percentile (15.070). This indicates that none of the ten prior flood events have a significantly large impact on the reconstructed October 2015 Cedar Creek stage.

We can also identify potentially influential functional observations in the river stage data using AIP. This measure involves the absolute difference

Table 1: The L_2 distance (Δ_i) between the 2015 Cedar Creek reconstructions with all 10 observations included and with each observation individually withheld.

Event	Δ
August 1995	8.706
February 1998	2.559
March 2003	*14.165*
May 2003	6.684
September 2004	6.285
March 2007	8.628
February 2010	2.118
May 2013	3.204
November 2018	5.580
February 2020	13.377

Table 2: Δ percentiles of each influential measurement from 5000 bootstrapped river stage observations along with the observed maximum of each metric in the river stage data context. Note that $\alpha = 0.5$ is most appropriate to use given the small sample size $N = 10$.

Percentile	$\alpha = 0$	$\alpha = 0.5$
90%	13.253	15.070
95%	18.078	19.730
99%	27.942	27.586
Max Obs. 14.165 (March 2003)		

between $\hat{Y}^{new}(t)$ and $\hat{Y}^{new}_{(i)}(t)$ (see, e.g., the difference between curves in Figure 6). All 10 observed absolute difference functions are shown in the left panel of Figure 7.

The percentiles of each vector of absolute differences for each curve are plotted the right panel of Figure 7, showing the observations whose influence persisted over a long duration. At the lower percentiles, this right panel shows no pronounced differences across curves; however, the red curve representing the February 2020 flood event begins a noticeable increase around its 80th percentile, and the green curve depicting the March 2003 event begins a sharp increase shortly thereafter. The area under each curve indicates which of these two observations has the most overall impact on the October 2015 Cedar Creek reconstruction based on the magnitude and duration of the difference. Table 3 provides these AIP values for each flood event.

These results show that the February 2020 flood event was the most influential when we account for the duration of the influence. However, the bootstrapped percentiles of our approximate null distribution of AIP, given in Table 4, indicate that in terms of AIP, none of the prior flood events were significantly influential on the reconstruction of the October 2015 Cedar Creek stage. These results have several explanations. The first reason could be that Pittman et al. [12] intentionally aligned the predictor curves to best resemble the out-of-sample October 2015 Congaree River curve, specifically to ensure

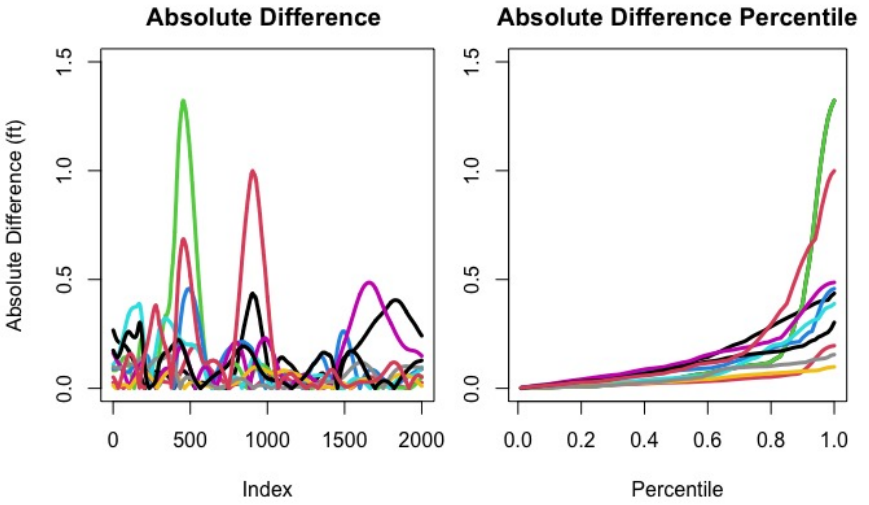

Fig. 7: All 10 absolute difference curves (left) and percentiles of absolute differences (right) between the October 2015 reconstruction using all 10 observations and when each observation was removed. In each panel, the green curve is from the March 2003 event and the pronounced red curve is from the February 2020 event.

Table 3: The area (AIP) under each curve in Figure 7 (right).

Event	AIP
August 1995	14.615
February 1998	3.909
March 2003	14.668
May 2003	11.146
September 2004	9.475
March 2007	14.689
February 2010	3.868
May 2013	5.495
November 2018	10.201
February 2020	*18.200*

that none of the 10 events carried too much weight and that the reconstruction was accurate. These results suggest that the landmark alignment method of Pittman et al. [12] was successful. Additionally, the simulation study in the previous section indicated that when $N = 10$, it takes a very influential observation to surpass the 95th percentile, since the average power was around 0.60 when λ was far from 1, so it is possible that the lack of significantly influential curves was also a result of the small sample size.

Table 4: *AIP* percentiles from each 5000 bootstrapped river stage observations along with the observed maximum of each metric in the river stage data context. Note that $\alpha = 0.5$ is most appropriate to use given the small sample size $N = 10$.

Percentile	$\alpha = 0$	$\alpha = 0.5$
90%	18.168	19.668
95%	22.949	25.282
99%	33.374	36.747
Max Obs. 18.2 (Feb. 2020)		

6 Application: Air and Water Temperature

We calculated our influence measures and applied our method on an air and water temperature dataset coming from 35 weather stations along the US coastline in 2020. This dataset was studied by Pittman and Hitchcock [7] in the context of functional regression influence for in-sample prediction and estimation of regression coefficients. We obtained the data from the National Data Buoy Center [15]. These 35 stations are located all around the United States coastline, including the East Coast, West Coast, Gulf of Mexico, Alaskan coastline, and Hawaii (map of specific locations provided in the Supplementary Material).

Each station's data contained roughly 87,600 temperature measurements in 6-minute intervals across 2020. To be eligible for inclusion, the station's air and water temperatures had to be at least 90% non-missing. Then we preprocessed the data, which included linear interpolation to fill in any missing record and presmoothing each observation to remove the day-to-day variability and focus on the yearly trends. Lastly, without disrupting the underlying relationship between the air and water temperature curves, we resized the length of each smoothed discretized curve to 1000 equally-spaced observations across the year to facilitate the functional calculations. Figure 8 provides all 35 air and water temperature curves used in the model, which we represented using 21 B-spline basis functions. Note that the low gold curve in Figure 8 is from Red Dog Dock, Alaska, which is the most northern station used in the sample.

We then used air temperature to predict concurrent water temperature throughout the calendar year. In this study, we had 35 complete air and water temperature functional observations, and we introduced 5 out-of-sample observations that had available air temperature data but incomplete or completely missing water temperatures. Using the 35 complete air and water temperature observations, we found estimates $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ for the concurrent model from Section 2.1 and predicted the five missing stations' water temperature functions using each the known air temperature functions at those stations. Using $\hat{Y}^{new}(t)$ and $\hat{Y}^{new}_{(i)}(t)$, we calculated each of the 35 observations' $\Delta_i, i = 1, \ldots, 35$ for each of the five out-of-sample stations. Note that the predicted water temperature of a held-out in-sample station using the other 34 stations is generally very close to the true water temperature of the omitted 35th station. The complete results are given in Table A1. Red Dog Dock

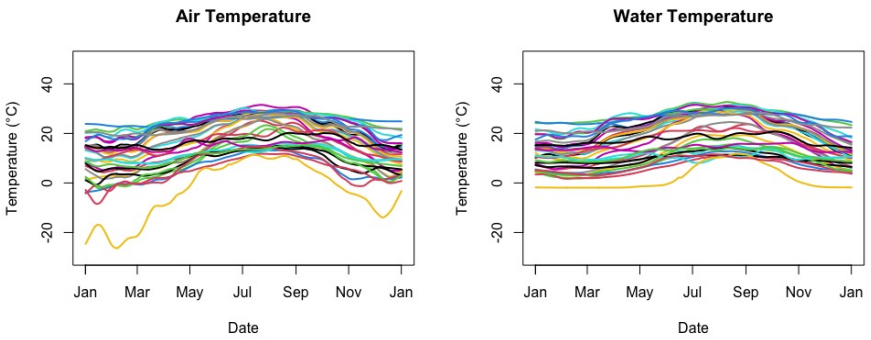

Fig. 8: All 35 smoothed air (left) and water (right) temperatures used in the model.

has the highest Δ for each of the five out-of-sample observations; it was so much larger than the rest that there is little doubt that this observation has massive influence on the predicted values. This station corresponds to the gold curve in Figure 8 which has a clearly low air temperature during the winter months compared to the other 34 locations, whereas Red Dog Dock's water temperature is only slightly lower than the rest. Note that which specific out-of-sample observation that was considered greatly affected the magnitude of Δ, indicating that using a universal threshold to judge Δ values is not appropriate. Like Red Dog Dock, Prudhoe Bay in Alaska had a low air temperature with a moderate water temperature, and the values of the Δ_i were large for all 35 locations corresponding to the Prudhoe Bay prediction. Using the approximate null distribution from the bootstrap is necessary to judge significant influence on prediction of some particular response curve.

With a moderate sample size $N = 35$, using either $\alpha = 0$ or $\alpha = 0.5$ in our weighted bootstrap method may be appropriate; however, given the large magnitude of Δ for the Red Dog Dock station for each out-of-sample observation, we suggest $\alpha = 0.5$ since it dampened the effect of that observation on the approximate null distribution, facilitating testing whether any other stations significantly influenced the prediction.

We applied the weighted bootstrap method (with $B = 100$) independently for each of the five target external observations and approximated a null distribution of Δ. Table 5 gives the resulting 90th, 95th, and 99th percentiles. The clearest conclusion from Table 5 is that regardless of the target observation, Red Dog Dock (observation 31) had a significantly large Δ, falling above the 99th percentile for all five out-of-sample observations. Note that different functional observations in the sample provide the second largest observed Δ depending on the target observation, but in some cases, that second-most-influential observation is also highly influential on the predicted water temperature curve of the out-of-sample location. For predicting the Ship John

Table 5: Percentiles from an approximate null distribution of Δ for each of the five target observations and the highest and 2nd highest observed Δ for each observation.

		Adak Island	Kahului	Prudhoe Bay	Rockport	Ship John Shoal
$\alpha = 0$	90%	5.22	4.91	9.63	3.94	2.52
	95%	7.02	6.17	12.43	4.91	3.77
	99%	22.41	15.24	88.64	8.07	8.50
$\alpha = 0.5$	90%	7.36	4.80	12.64	3.74	2.52
	95%	10.43	6.87	16.68	4.99	3.56
	99%	20.90	10.23	30.64	9.66	7.08
	Max Obs.	**26.88 (31)**	**24.66 (31)**	**101.58 (31)**	**9.82 (31)**	**16.90 (31)**
	2nd Highest	7.15 (29)	6.81 (19)	11.7 (34)	5.7 (14)	3.87 (7)

Shoal water temperature, the Boston observation's Δ was above the 95th percentile, indicating a significant amount of influence on the Ship John Shoal prediction. For the Kahului prediction, the Lake Worth Pier observation's Δ was slightly below the 95th percentile, indicating a moderate influence on the Kahului prediction. For the Rockport prediction, the Fernandina Beach observation's Δ was well above the 95th percentile, showing significant influence on the water temperature prediction at the Rockport station.

Lastly, we calculated AIP_i of each of the 35 sampled observations for each of the five out-of-sample stations. We first calculated the predicted water temperature using all 35 observations for each of the five target locations independently, and then did likewise with each of the 35 observations in turn sequentially removed. We found the percentiles of the absolute differences as described in Section 2. The plot of the percentiles of absolute differences for one out-of-sample location, Adak Island, is given in Figure 9. The other four target observations' absolute difference percentiles plots are very similar. The area under each curve yields the AIP of each of the 35 observations for each target observation (provided in Table A2).

As it did with Δ, Red Dog Dock had the largest AIP for all five target observations. This suggests that the observation has a large impact on the water temperature prediction at each of the five target locations. To confirm the formal significance of its influence and investigate the potential influence of other observations, we independently carried out the weighted bootstrap method to approximate the null distribution of AIP for each target observation. Given the large magnitude of Red Dog Dock's AIP compared to the others', we suggest using $\alpha = 0.5$. Table 6 provides the approximate null distribution's percentiles using $\alpha = 0$ and $\alpha = 0.5$, respectively, along with the largest and second largest observed AIP.

When using $\alpha = 0.5$, for each target observation, Red Dog Dock's AIP was well above the 95th percentile, and for every location except Rockport it was well above the 99th percentile. This indicates that the Red Dog Dock observation was significantly influential on water temperature prediction regardless of the target observation. Moreover, since it was substantially above the 99th percentile for most of the target locations, this observation should be investigated further for possible removal, as it may have incorrectly distorted the model's

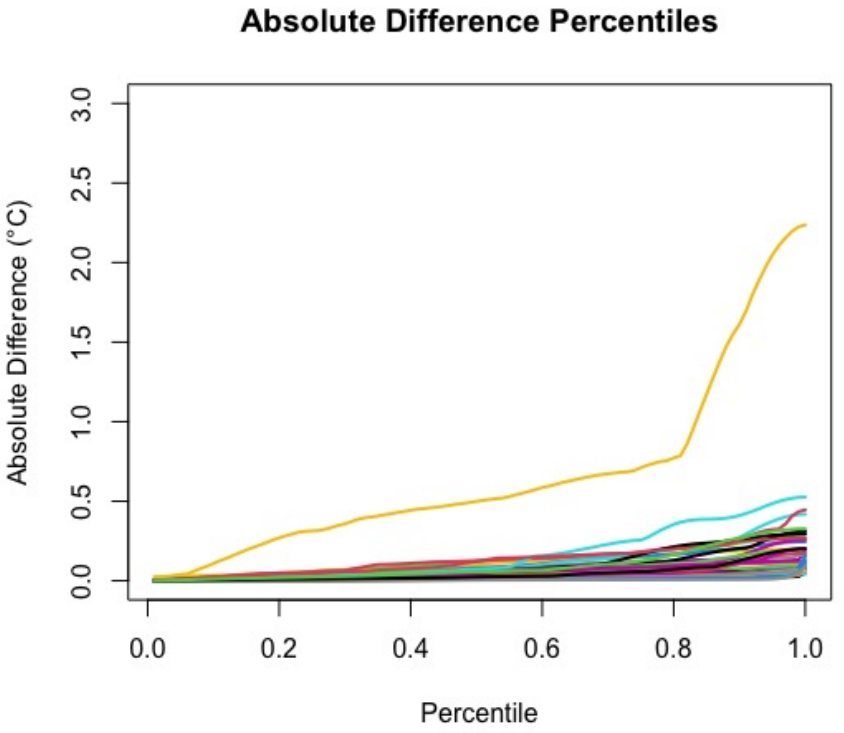

Fig. 9: All 35 observations' percentiles of absolute differences between the Adak Island water temperature prediction using all 35 observations and with each observation removed in turn. The gold curve is the Red Dog Dock observation's results.

prediction of water temperatures. Once we dampened the effect of Red Dog Dock on the null distribution by using $\alpha = 0.5$, the Fernandina Beach AIP was above the 90th percentile for predicting Kahului and the 95th percentile for predicting Rockport, indicating that it also had significant influence on the prediction of water temperatures at these locations. The Boston station's AIP was above the 95th percentile for predicting the Ship John Shoal water temperature, suggesting that it (along with Red Dog Dock) was influential.

7 Conclusion

One application of functional data analysis is to use a concurrent functional relationship between sets of observations to predict an out-of-sample (target) observation's response. Our new measures of influence, Δ and AIP, offer a pragmatic way to detect functional observations that have a large impact on

Table 6: Approximate null distribution percentiles of AIP for each target observation for $\alpha = 0$ and $\alpha = 0.5$ along with the largest (Red Dog Dock) and 2nd largest observed AIP given.

		Adak Island	**Kahului**	**Prudhoe Bay**	**Rockport**	**Ship John Shoal**
$\alpha = 0$	90%	12.87	11.987	24.21	10.33	6.38
	95%	17.62	15.35	31.02	13.004	9.36
	99%	50.21	24.65	109.71	19.90	25.68
$\alpha = 0.5$	90%	18.41	11.31	30.35	9.92	6.29
	95%	25.98	16.06	40.19	14.57	8.40
	99%	44.65	26.52	80.29	24.42	16.18
	Max Obs.	**64.81**	**43.23**	**190.87**	**22.51**	**32.39**
	2nd Highest	15.89 (29)	14.33 (14)	25.83 (34)	15.42 (14)	10.57 (7)

specific predictions of target response curves. Additionally, simulation shows that our weighted bootstrapping approach performs well in identifying whether the most influential observations truly have a significant impact on the prediction. In both the river stage and air and water temperature examples, we sensibly identify certain observations as more influential than the rest, and then the bootstrap method confirms whether their influence is significantly large, further illustrating that our method satisfactorily identifies functional observations in the concurrent model that influence the prediction of an out-of-sample response curve.

Supplementary Information. The Supplementary Material includes plots of the reconstructed October 2015 Cedar Creek stage function based on fits with and without each flood event and a table and a map of all 35 locations in the air and water temperature dataset.

Declarations

7.1 Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

7.2 Conflict of Interest

All authors declare that they have no conflicts of interest.

References

[1] Belsley, D.A., Kuh, E., Welsch, R.E.: Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New York (2005)

[2] Cook, R.D.: Detection of influential observation in linear regression. Technometrics **19**(1), 15–18 (1977). https://doi.org/10.1080/00401706.1977.10489493

[3] Shen, Q., Xu, H.: Diagnostics for linear models with functional responses. Technometrics **49**(1), 26–33 (2007). https://doi.org/10.1198/004017006000000444

[4] Chiou, J., Müller, H.: Diagnostics for functional regression via residual processes. Computational Statistics and Data Analysis **51**(10), 4849–4863 (2007). https://doi.org/10.1016/j.csda.2006.07.042

[5] Febrero-Bande, M., Galeano, P., González-Manteiga, W.: Measures of influence for the functional linear model with scalar response. Journal of Multivariate Analysis **101**(2), 327–339 (2010). https://doi.org/10.1016/j.jmva.2008.12.011

[6] Chen, G., Huang, C., Lin, J.: Statistical diagnostics for functional linear regression models with Gaussian process errors. Communication on Applied Mathematics and Computation **28**(1), 118–126 (2014). https://doi.org/10.3969/j.issn.1006-6330.2014.01.015

[7] Pittman, R.D., Hitchcock, D.B.: Identifying Influential Observations in Concurrent Functional Regression with a Weighted Bootstrap. Under revision. Preprint at https://people.stat.sc.edu/hitchcock/PreprintFuncInfluence.pdf (2022)

[8] Ramsay, J.O., Graves, S., Hooker, G.: Fda: Functional Data Analysis. (2020). R package version 5.1.5.1. https://CRAN.R-project.org/package=fda

[9] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020). R Foundation for Statistical Computing. https://www.R-project.org/

[10] Borchers, H.W.: Pracma: Practical Numerical Math Functions. (2019). R package version 2.2.9. https://CRAN.R-project.org/package=pracma

[11] Uhlenbeck, G.E., Ornstein, L.S.: On the theory of the Brownian motion. Phys. Rev. **36**, 823–841 (1930). https://doi.org/10.1103/PhysRev.36.823

[12] Pittman, R.D., Hitchcock, D.B., Grego, J.M.: Concurrent functional regression to reconstruct river stage data during flood events. Environmental and Ecological Statistics **28**(1), 219–237 (2021)

[13] United States Geological Survey: USGS 02169625 Congaree River at Congaree NP Near Gadsden, SC. https://waterdata.usgs.gov/sc/nwis/uv?site_no=02169625

[14] United States Geological Survey: USGS 02169672 Cedar Creek at Congaree NP Near Gadsden, SC. https://waterdata.usgs.gov/sc/nwis/uv?site_no=02169672

[15] National Oceanic and Atmospheric Administration: National Data Buoy Center. https://www.ndbc.noaa.gov/obs.shtml

Appendix A Tables for Air and Water Temperature Example

Table A1: Δ_i influence measure for all 35 observations for all five locations with missing water temperatures along with the average for each of them.

	Location	Adak Island AK	Kahului HI	Prudhoe Bay AK	Rockport TX	Ship John Shoal NJ	Average
1	Amerada Pass, LA	1.12	5.07	5.25	4.42	1.64	3.50
2	Atlantic City, NJ	1.85	1.05	3.58	1.29	1.74	1.90
3	Bar Harbor, ME	4.38	1.54	7.66	0.99	2.26	3.37
4	Bay Waveland Yacht Club, MS	0.92	2.40	2.49	2.28	0.94	1.81
5	Beaufort, NC	0.45	1.09	0.80	1.20	0.85	0.88
6	Bishops Head, MD	2.02	1.03	3.78	1.37	1.89	2.02
7	Boston, MA	3.79	2.83	7.60	3.32	3.87	4.28
8	Bridgeport, CT	2.84	1.08	5.78	1.21	2.24	2.63
9	Calcasieu Pass, LA	0.45	1.33	1.48	1.28	0.72	1.05
10	Charleston Cooper River Entrance, SC	0.36	1.02	0.92	1.04	0.63	0.79
11	Clearwater Beach, FL	1.49	3.88	4.87	3.96	1.96	3.23
12	Cordova, AK	3.80	0.82	6.44	0.69	1.36	2.62
13	Crescent City, CA	4.93	0.60	6.37	0.82	1.38	2.82
14	Fernandina Beach, FL	3.10	5.16	7.71	5.74	2.06	4.76
15	Fort Pulaski, GA	0.39	1.10	0.96	1.09	0.64	0.84
16	Johnny Mercer Pier, Wrightsville Beach, NC	0.69	1.29	0.87	1.21	0.91	0.99
17	Ketchikan, AK	3.62	0.38	5.20	0.56	1.14	2.18
18	King Cove, AK	4.75	0.80	8.51	1.01	1.90	3.39
19	Lake Worth Pier, FL	0.90	6.81	10.99	4.57	0.84	4.82
20	Mokuoloe, HI	0.84	5.22	8.97	3.56	0.78	3.87
21	Naples, FL	0.83	3.30	5.02	2.68	1.00	2.57
22	Old Port Tampa, FL	0.97	3.31	4.30	3.06	1.42	2.62
23	Oregon Inlet Marina, NC	0.71	0.88	0.82	0.96	0.86	0.85
24	Panama City Beach, FL	0.74	2.79	1.98	2.38	1.27	1.83
25	Port Angeles, WA	4.24	0.44	5.92	0.46	1.44	2.50
26	Port Chicago, CA	1.92	1.38	2.20	1.22	1.21	1.59
27	Portland, ME	2.80	1.39	4.74	1.23	1.81	2.39
28	Port Isabel, TX	0.42	1.87	2.60	1.40	0.56	1.37
29	Port Orford, OR	7.15	1.22	9.26	0.99	2.09	4.14
30	Port San Luis, CA	2.37	2.70	2.69	2.42	1.86	2.41
31	Red Dog Dock, AK	*26.88*	*24.66*	*101.58*	*9.82*	*16.90*	*35.97*
32	Sand Island, Midway Islands	0.31	2.76	4.08	1.95	0.56	1.93
33	Santa Monica Pier, CA	2.14	1.41	2.66	1.08	1.02	1.66
34	Skagway, AK	5.12	2.31	11.71	0.92	2.69	4.55
35	Westport, WA	4.18	0.45	5.65	0.63	1.41	2.47

Table A2: *AIP* influence measure for all 35 observations for all five locations with missing water temperatures along with the average for each of them.

	Location	Adak Island AK	Kahului HI	Prudhoe Bay AK	Rockport TX	Ship John Shoal NJ	Average
1	Amerada Pass, LA	2.69	12.51	12.39	11.20	4.34	8.63
2	Atlantic City, NJ	4.58	2.36	7.96	2.96	4.39	4.45
3	Bar Harbor, ME	11.57	3.33	20.56	1.70	6.03	8.64
4	Bay Waveland Yacht Club, MS	2.01	6.07	6.49	5.96	2.47	4.60
5	Beaufort, NC	1.22	3.11	1.78	3.34	2.39	2.36
6	Bishops Head, MD	4.84	2.84	8.43	3.80	5.02	4.99
7	Boston, MA	10.58	5.40	19.07	6.76	10.58	10.48
8	Bridgeport, CT	7.78	2.72	14.44	3.24	6.39	6.91
9	Calcasieu Pass, LA	1.15	3.39	3.39	3.38	1.71	2.60
10	Charleston, Cooper River Entrance, SC	0.86	2.69	2.13	2.73	1.58	2.00
11	Clearwater Beach, FL	3.76	11.89	14.43	11.93	5.03	9.41
12	Cordova, AK	9.75	1.78	16.94	1.63	3.36	6.69
13	Crescent City, CA	11.09	1.53	14.97	2.08	3.52	6.64
14	Fernandina Beach, FL	7.10	14.34	20.35	15.42	5.30	12.50
15	Fort Pulaski, GA	0.89	2.66	2.01	2.64	1.57	1.95
16	Johnny Mercer Pier, Wrightsville Beach, NC	1.34	2.97	1.35	2.88	2.10	2.13
17	Ketchikan, AK	8.49	0.88	13.35	1.19	2.92	5.37
18	King Cove, AK	12.83	2.07	23.30	2.42	4.69	9.06
19	Lake Worth Pier, FL	2.16	13.92	21.45	9.94	1.96	9.89
20	Mokuoloe, HI	2.12	11.67	17.87	8.85	2.05	8.51
21	Naples, FL	2.27	8.63	12.35	7.65	2.50	6.68
22	Old Port Tampa, FL	2.57	9.70	11.88	9.23	3.74	7.42
23	Oregon Inlet Marina, NC	1.57	2.23	1.93	2.42	2.18	2.07
24	Panama City Beach, FL	1.77	7.40	5.32	6.71	3.52	4.94
25	Port Angeles, WA	9.65	0.94	14.17	1.06	3.47	5.86
26	Port Chicago, CA	4.39	3.78	4.59	3.38	3.35	3.90
27	Portland, ME	7.22	2.68	12.47	1.94	4.57	5.77
28	Port Isabel, TX	0.97	4.40	5.52	3.55	1.44	3.18
29	Port Orford, OR	15.89	3.10	21.15	2.50	5.37	9.60
30	Port San Luis, CA	5.80	5.84	6.07	5.12	4.29	5.42
31	*Red Dog Dock, AK*	*64.81*	*43.23*	*190.87*	*22.51*	*32.39*	*70.76*
32	Sand Is., Midway Islands	0.55	5.88	7.66	4.57	1.46	4.02
33	Santa Monica Pier	4.43	3.70	5.67	2.82	2.70	3.87
34	Skagway, AK	12.94	3.83	25.83	1.95	5.66	10.04
35	Westport, WA	9.24	1.09	13.13	1.46	3.39	5.66