

RESEARCH ARTICLE OPEN ACCESS

Learning From Limited Temporal Data: Dynamically Sparse Historical Functional Linear Models With Applications to Earth Science

Joseph Janssen¹  | Shizhe Meng² | Asad Haris¹ | Stefan Schrunner³ | Jiguo Cao⁴  | William J. Welch⁵ | Nadja Kunz⁶ | Ali A. Ameli¹

¹Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, Canada | ²Department of Mathematics, University of British Columbia, Vancouver, Canada | ³Department of Data Science, Norwegian University of Life Sciences, Oslo, Norway | ⁴Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada | ⁵Department of Statistics, University of British Columbia, Vancouver, Canada | ⁶Norman B Keevil Institute of Mining Engineering, University of British Columbia, Vancouver, Canada

Correspondence: Joseph Janssen (joejanssen@eoas.ubc.ca)

Received: 24 July 2024 | **Revised:** 22 April 2025 | **Accepted:** 29 April 2025

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (<http://dx.doi.org/10.13039/501100000038>) and the Canadian Statistical Sciences Institute (<http://dx.doi.org/10.13039/100022210>).

Keywords: functional data analysis | hydrology | rainfall–runoff relationships | streamflow | unit hydrograph

ABSTRACT

Scientists and statisticians often seek to understand the complex relationships that connect two time-varying variables. Recent work on sparse functional historical linear models confirms that they are promising as a tool for obtaining complex and interpretable inferences, but several notable limitations exist. Most importantly, previous works have imposed sparsity on the historical coefficient function, but have not allowed the sparsity, hence lag, to vary with time. We simplify the framework of sparse functional historical linear models by using a rectangular coefficient structure along with Whittaker smoothing, then reduce the assumptions of the previous frameworks by estimating the dynamic time lag from a hierarchical coefficient structure. We motivate our study by aiming to extract the physical rainfall–runoff processes hidden within hydrological data. We show the promise and accuracy of our method using eight simulation studies, further justified by two real sets of hydrological data.

1 | Introduction

In many complex systems, a causal variable is filtered into some response of interest via complicated and unknown nonstationary processes. Exactly how the causal impact spreads and changes across time is of great interest to scientists and policy-makers in various disciplines. If a method for learning this filtering function in a data-driven manner existed, one would be able to extract useful knowledge about the mechanisms that generate variability from various time series data sources.

More specifically, this research is motivated by the desire to learn the data generating processes behind streamflow time series. Observational units in hydrology called catchments, filter precipitation into streamflow (Figure 1). This process may sound simple; however, slightly different geologic, climatic, or topographic features may produce vastly different filtering functions. Understanding the filtering function for a given catchment is vital for various reasons. For example, knowing that a catchment filters most rainfall slowly through subsurface and deep flow pathways would indicate that the stream is vulnerable to long-term soil and

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Environmetrics* published by John Wiley & Sons Ltd.

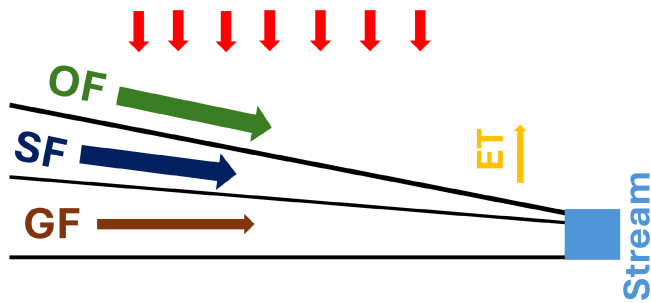


FIGURE 1 | A flow path diagram depicting how precipitation (red downward arrows) gets partitioned into evapotranspiration (ET), overlandflow (OF), shallow subsurface flow (SF), or deep groundwater flow (GF), with arrow width depicting illustrative relative effect size.

deep rock contamination from human activities such as mining (Kunz 2020). Furthermore, inferring the geological or climatological factors that cause process variation can enable better predictions in regions where streamflow is not measured. Indeed, accurately estimating the functions that filter precipitation into streamflow can help hydrologists understand the physical processes behind complex temporal hydrological data.

As indicated by Botter et al. (2010), for a given catchment, water input and streamflow can be causally related via a historical filtering function. This filtering function evolves through time due to changes in antecedent water input, evapotranspiration, temperature, and landscape variables (Mindham et al. 2023). Evaporation or transpiration can alter the effect sizes of the filtering function, since a catchment with high evaporation only has a small portion of water available for filtering into streamflow. On the other hand, the depth to which water flows or gets stored in/on the ground dictates the length of time between a precipitation event and a rise in streamflow (Jasechko et al. 2016; Somers and McKenzie 2020). Overland flow (OF) is the fastest and shallowest path that precipitation can follow, precipitation filtered into shallow subsurface flow (SF) travels down into the soil and towards the stream at a slower pace, and deep groundwater flow (GF) travels deeply into the ground and is the slowest responding flow path (Figure 1).

Streamflow and rainfall time series at the Koksilah River (British Columbia) and the Withlacoochee River (Florida), illuminate the fact that the relationship between rainfall and streamflow is indeed complex and highly nonstationary (Figure 2). At the Koksilah River, very little winter rainfall is evaporated (days 1–50 and days 320–365), rather it quickly filters through the catchment as streamflow. Conversely, in summer and early fall (days 100–310), any rainfall that may occur does not result in notable streamflow intensification. At the Withlacoochee River, streamflow falls less rapidly during periods of little-to-no rainfall compared with the Koksilah River (e.g., days 1–50). Evaporation is likely to be high in this catchment; therefore, streamflow may only increase in the event of a large rainfall event as seen on day 50.

The statistical model that most closely follows the causal filtering relation that we described in the preceding paragraphs comes from Xun et al. (2022), who considered the historical functional linear model,

$$y_i(t) = \int_{\max(0, t-\delta)}^t \beta(s, t) x_i(s) ds + \epsilon_i(t) \quad (1)$$

Here, $y_i(t)$ is the outcome of interest at time t of replicate series i , $x_i(s)$ is a temporal explanatory variable at the past time s for the same replicate, and $\epsilon_i(t)$ is the residual term. The bivariate coefficient function $\beta(s, t)$ represents the effect of the explanatory variable at past time s on the outcome at the current time t , where $\max(0, t - \delta) \leq s \leq t$. To estimate $\beta(s, t)$, Xun et al. (2022) minimized an objective function containing three terms: (1) a least-squares loss, (2) a nested group bridge penalty term which reveals the constant lag δ , after which $\beta(s, t)$ is zero, and (3) a smoothness penalty term. This builds upon many previous methods including Malfait and Ramsay (2003) who originally developed the functional historical linear model, and Harezlak et al. (2007) who imposed a discrete difference penalty on the coefficients similar to the P-spline framework (Eilers et al. 2015). While both Malfait and Ramsay (2003) and Harezlak et al. (2007) primarily focused on predictive capability, Xun et al. (2022) were more interested in accurate parameter estimation. Parameter estimation will also be the focus of this article.

The method introduced by Xun et al. (2022) is compelling, although it has several notable downsides. First, it is overly complex. Its original formulation is nonconvex, necessitating iterative lasso-equivalent optimization steps until convergence. The tent-like basis functions which parameterize $\beta(s, t)$ add computational cost such that the number of basis functions is usually restricted (about 230 basis functions were used in previous works). Second, the triangular finite element framework artificially imposes a constraint on the maximum time lag (i.e., the domain of integration cannot be negative in Equation (1)). For example, if one were to predict streamflow on day one for the Withlacoochee River in Figure 2, only one value of rainfall (0mm) could be used, leading to an inaccurate prediction of 1998s peak streamflow. In our application, however, except for the very first year, the model for streamflow on a day at the beginning of a year has rainfall data available from the last days of the previous year (see Section 2). Third, the previous frameworks do not allow for a time dependent time-lag parameter $\delta(t)$. Although it is common to assume an arbitrary fixed window of time for which precipitation can affect streamflow, this is a severe limitation of previous works (Janssen et al. 2021), so finding the true underlying $\delta(t)$ is of interest to the hydrological community (Tenant et al. 2020; Chiu and Bittler 1969). Other similar methods, with similar issues, can be found in the distributed lag literature (Almon 1965; Pesando 1972; Rushworth et al. 2013; Asencio et al. 2014; Gupta 1968; Liao et al. 2023).

This article has three major contributions. First, to the best of our knowledge, this is the first attempt to estimate a time-dependent time-lag parameter in historical functional linear models. Second, we simplify and increase the flexibility of previous historical functional linear model methods by replacing the tent-like basis functions with the tensor product Whittaker basis such that higher resolution coefficient and time-lag functions can be recovered. Finally, we successfully demonstrate that our method can lead to novel research directions within hydrology by accurately estimating the ground-truth filtering function $\beta(s, t)$ and time-dependent time-lag parameter $\delta(t)$.

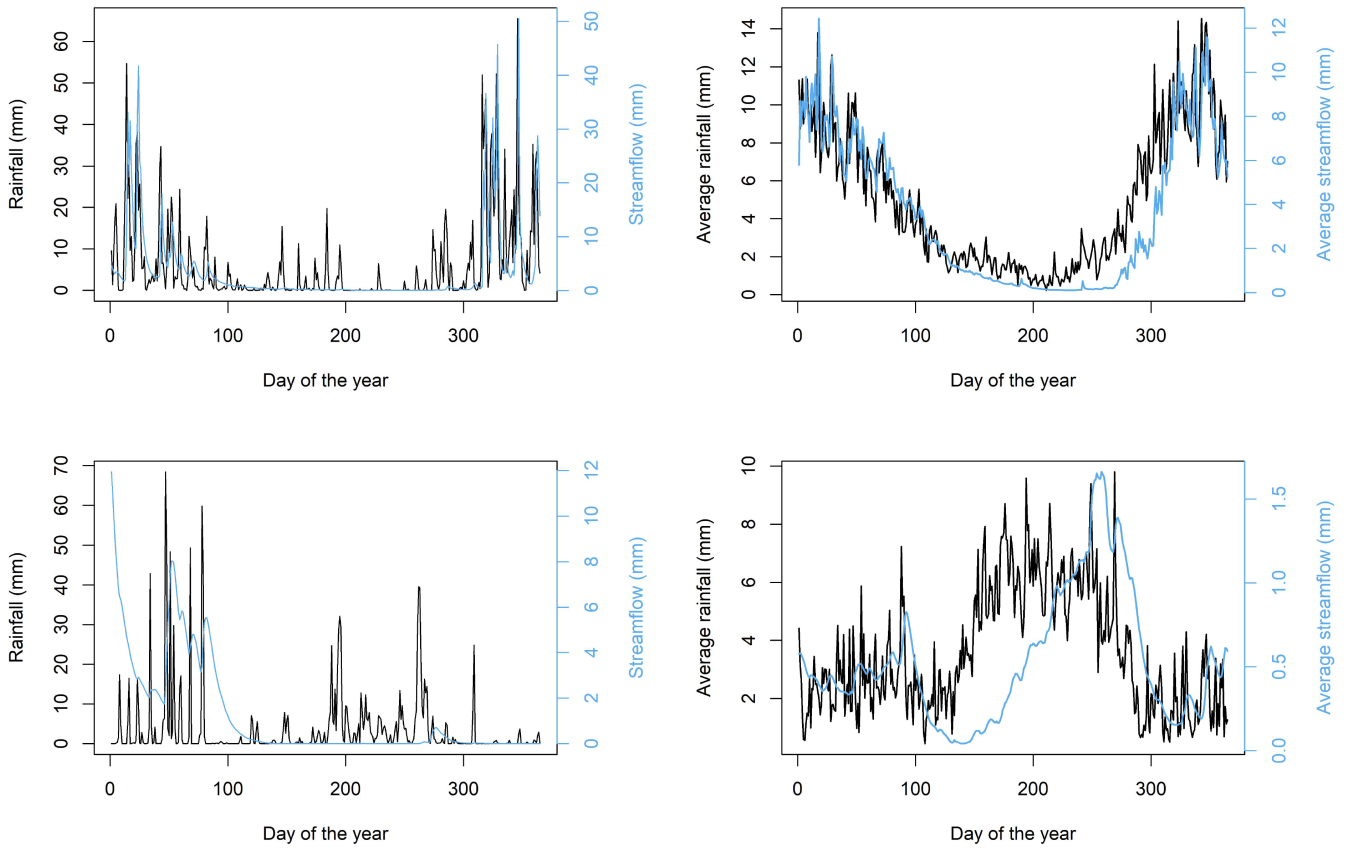


FIGURE 2 | Left: The daily streamflow (blue) and rainfall (black) for the Koksilah River (top) and the Withlacoochee River (bottom) for the year 1998. Right: The average daily streamflow (blue) and rainfall (black) for the Koksilah River (top) and the Withlacoochee River (bottom) for 1979–2018.

2 | Historical Functional Linear Model With Dynamic Sparsity (HFLM-DS)

Let $y_i(t)$ be the response at time $t \in [1, T]$ for replicate $i \in 1, \dots, n$, and let $x_i(t)$ be the corresponding covariate. In our application, $y_i(t)$ is streamflow for time t within year i , and $x_i(t)$ is rainfall. We begin by considering the exact causal streamflow relation, where $\mathbf{V}$ is the set of all (un)known and (un)measured antecedent variables which alter the way in which rainfall is filtered into streamflow. We then separate the true filtering function $h(s, \mathbf{V})$ into an identifiable component that is stable across years due to consistent seasonal variations of antecedent conditions ($\beta(s, t)$) and an unidentifiable component that changes from year to year ($h'(s, \mathbf{V})$):

$$\begin{aligned} y_i(t) &= \int_0^\infty h(s, \mathbf{V}) x_i(t-s) ds \\ &= \int_0^\infty [\beta(s, t) + h'(s, \mathbf{V})] x_i(t-s) ds \\ &= \int_0^\infty \beta(s, t) x_i(t-s) ds + \int_0^\infty h'(s, \mathbf{V}) x_i(t-s) ds \\ &= \int_0^\infty \beta(s, t) x_i(t-s) ds + \epsilon_i(t) \end{aligned}$$

After separating terms, letting the unidentifiable part become the autocorrelated error term $\epsilon_i(t)$, and letting $D-1$ be the maximum possible lag, our functional historical linear model is given by

$$y_i(t) = \int_0^{D-1} \beta(s, t) x_i(t-s) ds + \epsilon_i(t) \quad (2)$$

and the discretized version (e.g., assuming daily measurements) is visualized in Figure 3. One detail that distinguishes our application from most applications in functional data analysis is that in functional data, the replicates usually relate to independent units observed over a common time domain (e.g., experimental subjects). While, in a sense, each year can be treated as a replicate, in reality, we have one data series for each of x and y

Exact filtering function model

Expand h into subcomponents

Distribute and separate terms

Unparamaterized term as error

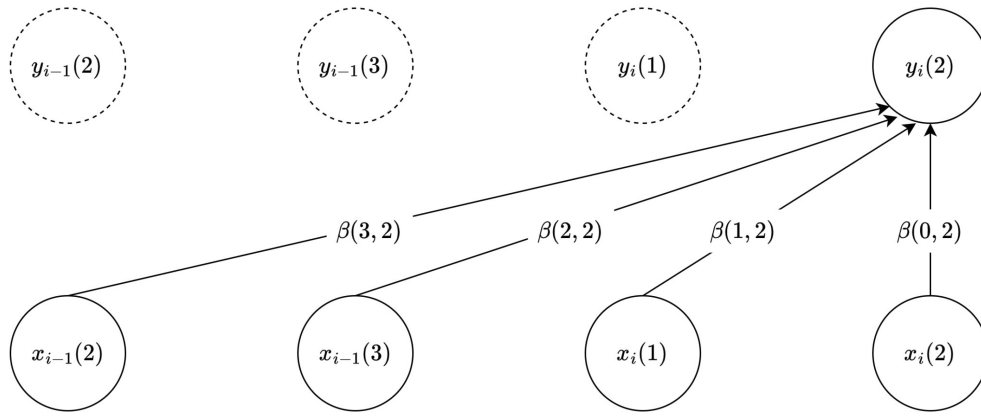


FIGURE 3 | The relationships between the response y and the cause x is shown for $D = 4$ and $T = 3$. Causal connections are shown for replicate i at $t = 2$.

over many years. Daily observations on x and y at the end of one year are followed by data on the first day of the next year. Due to this, in Equation (2), contrary to previous works, $t - s$ can be < 1 (i.e., outside the original range of $t \in [1, T]$). When this occurs, we draw from the end of the previous replicate $i - 1$. An example of this is illustrated in the final two lagged causes in Figure 3, where $y_i(2)$ is being predicted by the first two observations of replicate i ($x_i(1)$ and $x_i(2)$) as well as the last two observations of the previous replicate ($x_{i-1}(2)$ and $x_{i-1}(3)$).

As with Xun et al. (2022), we assume without loss of generality that the response and covariate have had their seasonal signals removed such that they have mean zero for each day of the year. Not only does the usage of rainfall and streamflow anomalies remove the need for a functional intercept term but also removes any potential periodic confounding effects present in both signals (Moges et al. 2022).

Let $\phi_1(s, t), \dots, \phi_K(s, t)$ denote a sequence of K known basis functions. Given these basis functions and their corresponding basis coefficients $\mathbf{b} = [b_1, \dots, b_K]^T$, the coefficient function $\beta(s, t)$ can be denoted as

$$\beta(s, t) = \sum_{k=1}^K b_k \phi_k(s, t).$$

Using this basis expansion, we may reformulate Equation (2) as

$$\begin{aligned} y_i(t) &= \int_0^{D-1} \sum_{k=1}^K b_k \phi_k(s, t) x_i(t-s) ds + \epsilon_i(t) \\ &= \sum_{k=1}^K b_k \int_0^{D-1} \phi_k(s, t) x_i(t-s) ds + \epsilon_i(t) \\ &= \sum_{k=1}^K b_k z_{ik}(t) + \epsilon_i(t), \end{aligned}$$

where

$$z_{ik}(t) = \int_0^{D-1} \phi_k(s, t) x_i(t-s) ds.$$

As stated in the introduction, one of our foremost goals is to infer the dynamic time lag $\delta(t)$ after which the cause has no discernible

effect on the response, that is,

$$\delta(t) = \max\{s\} \quad \text{subject to } \beta(s, t) \neq 0. \quad (3)$$

The accuracy in determining $\delta(t)$ is directly related to the resolution of the basis functions that parameterize $\beta(s, t)$, but increasing the number of bases can be computationally prohibitive if we continue the use of the tent-like basis functions utilized in previous works (Xun et al. 2022). Under these circumstances, using the Whittaker basis (or equivalently zero-degree P-splines) is an appropriate solution as it is a fast exact interpolator with few notable shortcomings (Eilers 2003; Whittaker 1922). In this work, we parameterize the coefficient function $\beta(s, t)$ with the finest resolution possible (one basis function per (lag, day of the year) pair) using tensor products of Whittaker basis functions $\phi_k(s, t) = w_m(s) \otimes w_\ell(t)$. If we have T observations in our time domain $[1, T]$, then $m \in \{0, \dots, D-1\}$, $\ell \in \{1, \dots, T\}$, and $k \in \{1, \dots, K\}$, where, for a specified k , $m = k - 1 - (\lceil k/D \rceil - 1) \times D$ and $\ell = \lceil k/D \rceil$. We note that m denotes where we are in lag space, ℓ denotes where we are in time space, and $\lceil x \rceil$ denotes the ceiling of x (x rounded up to nearest integer). Here $w_m(s)$ and $w_\ell(t)$ denote the Whittaker basis functions, which are equivalent to zero-degree B-spline basis functions with a knot spacing of one (Eilers 2003):

$$w_p(s) = \begin{cases} 1 & p \leq s < p+1 \\ 0 & \text{otherwise.} \end{cases}$$

Since the Whittaker basis either has a value of 1 or 0, the functions take on the same values as their discretized/observed versions. Due to this, in a slight abuse of notation, we continue to use functional notation (i.e., $\beta(s, t)$ or $y_i(t)$) even when we are discussing the computational/discrete values of these functions.

Suppose the maximum reasonable lagged influence between $x_i(t)$ and $y_i(t)$ for all t occurs at lag $s = D - 1$, then we must estimate $K = DT$ parameters $\{b_k : k = 1, \dots, K\}$. In general, the amount of available data is dense with few missing values but limited in temporal extent (i.e., we can have T observations in $[1, T]$), thus we expect that $nT = N \ll K$, leaving us with an underdetermined system. For example, in our real data study shown in Section 3.1, $n \leq 40$, $T = 365$, and $D = 150$. This implies that $N \leq 40 \cdot 365 = 14600 \ll K = 150 \cdot 365 = 54750$. This problem cannot

be overcome without some assumptions about the structure of the coefficient function $\beta(s, t)$, or equivalently the coefficients b_k , and we therefore stipulate three major assumptions. First, for the same lag, the coefficients at consecutive time points should have similar coefficient values (e.g., $\beta(s, t) \approx \beta(s, t + 1)$). Second, for the same time point, coefficients at consecutive lags should have similar values (e.g., $\beta(s, t) \approx \beta(s + 1, t)$). Finally, we expect $\beta(s, t)$ to be fairly sparse. Before fitting a model, the extent of the sparsity is unknown, although we expect the sparsity will be in regions with larger lags (i.e., for each day there will exist a lag $\delta(t)$ after which the predictor stops contributing to the response). As with the coefficients themselves, we expect that $\delta(t)$ will smoothly vary with day of the year. We limit the number of assumptions so that our methods are flexible and widely applicable (Clark et al. 2011), although we could have further assumed that the coefficient function should be non-negative for our application since additional precipitation cannot reduce streamflow. Instead of adding this constraint, we keep our model flexible to allow an additional evaluation of the scientific validity of our method.

With these three assumptions in mind, we now impose regularization and sparsity constraints on the coefficients of $\beta(s, t)$. With the first assumption, we regularize the squared differences of the coefficient function at consecutive points in time. This can be done with the horizontal first difference penalty matrix (D_H), which was first applied to functional historical linear models in Harezlak et al. (2007) and further used by Xun et al. (2022). We further add periodic boundary conditions to D_H in a similar fashion as Garcia (2010), such that $\beta(s, T) \approx \beta(s, 1)$. Likewise, given the second assumption, we penalize the squared differences between coefficients with the same time but with consecutive lag values by introducing the vertical first difference penalty matrix (D_V). We additionally consider a zero top boundary condition on the vertical penalties since lags past $D - 1$ are assumed to have zero influence. The structure of the coefficients and the penalties are visualized in Figure 4. As an illustrative example, if we take $D = 2$, $T = 3$, and $K = 6$, then D_H , D_V , and $\beta(s, t)$ are given by:

$$D_H = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \quad D_V = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\beta(s, t) = \begin{bmatrix} \beta(0, 1) \\ \beta(1, 1) \\ \beta(0, 2) \\ \beta(1, 2) \\ \beta(0, 3) \\ \beta(1, 3) \end{bmatrix}$$

In statistics, sparsity is usually induced via the lasso penalty (Tibshirani 1996), however, it does not have the oracle and sign consistency properties, meaning it cannot distinguish which coefficients should be exactly zero according to the data-generating process. This was shown experimentally in the

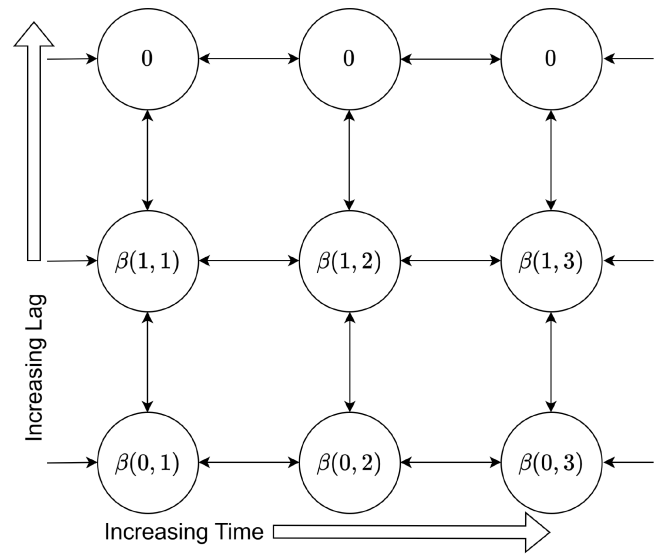


FIGURE 4 | The general structure of the discretized $\beta(s, t)$ is shown for $D = 2$ and $T = 3$, where s is lag and t is time. The bidirectional lines between nodes indicate that the nodes should have similar values. The unidirectional lines indicate that there is periodic behavior with a period of $T + 1$ (i.e., $\beta(0, 3)$ and $\beta(0, 1)$ are connected). The bottom row of nodes represents the coefficients relating the current time's cause to the current time's response while the next rows represent the coefficients for consecutive past lags. The top row shows the zero boundary conditions.

original lasso article (Tibshirani 1996), and it was shown theoretically in Meinshausen and Yu (2009), Zhao and Yu (2006), Fan and Li (2001), and Leng et al. (2006). The oracle and consistency properties of bridge penalties are better than lasso, although they still fail when the number of parameters exceeds the number of observations and when there is extensive multicollinearity (Huang et al. 2008). Both of these troublesome properties are expected in problems surrounding our intended application. As an alternative, Zhao and Yu (2006) suggests using the L0 penalty, although this poses a new problem of nonconvexity and NP-hard computational complexity (Huo and Ni 2007). The L0-norm has the oracle property and is selection consistent, meaning it will select the correct features which are non-zero in the ground-truth model with a probability converging to 1 under weaker conditions compared with the lasso (Zhang and Zhang 2012; Staerk et al. 2018). Generally, the L0-norm problem is $O(2^K)$, but due to the strongly hierarchical nature of our model in which we assume we know the ordering of how the coefficients should be dropped to zero, optimizing the sparsity problem in our case changes to $O(K)$. Thus, for our application it is computationally tractable and theoretically preferable for variable selection.

Most previous works that have developed historical functional linear models have imposed sparsity in their model (Harezlak et al. 2007; Malfait and Ramsay 2003; Xun et al. 2022), however, they all assume that the lag after which the explanatory feature has no effect on the response is static in time. This assumption is indeed too strong for many applications, especially in Earth science. Instead, we impose sparsity with a more flexible approach via nested group 2-norm thresholding such that a dynamic time lag $\delta(t)$ can be reliably estimated. The nested group framework allows us to place each coefficient in a series

of groups according to a hierarchical structure (Zhao et al. 2009). Let $A_{0,t}, A_{1,t}, \dots, A_{D-1,t}$ be a series of groups defined for each time point $t \in \{1, \dots, T\}$. The group $A_{s,t}$ contains the indices of the coefficients corresponding to $\beta(s, t), \dots, \beta(D-1, t)$. Notice that the groups are nested such that $A_{D-1,t} \subset A_{D-2,t} \subset \dots \subset A_{0,t}$, meaning a specific group contains all coefficients with lags greater or equal to its s . Since $A_{s,t} \subset A_{s-r,t}$ for $r \in \{1, \dots, s\}$, we know that $\|\mathbf{b}_{A_{s,t}}\|_2^2 \leq \|\mathbf{b}_{A_{s-r,t}}\|_2^2$. In contrast to this, if we look at two different times $t \neq t'$, then $A_{s,t} \cap A_{s,t'} = \emptyset$. These properties of our groups and their norms allow us to ensure that (1) coefficients at later lags must be removed and set to zero before coefficients at earlier lags when they represent the same time and (2) groups of coefficients at different times are formed independently such that the sparsity can remain dynamic and $\delta(t)$ can be estimated accurately. We further note that coefficients only act locally, so if $\mathbf{b}_3 = 0$ (corresponding to $\beta(0, 2)$), then the surrounding coefficients may not be zero (i.e., $\beta(0, 1) \neq 0$ and $\beta(0, 3) \neq 0$) unless the time is the same and the lag is greater ($\beta(1, 2) = 0$). These groups will be further discussed when our method and algorithm are detailed in Section 2.2.

2.1 | Autocorrelated Errors

As we illustrated in the derivation at the beginning of this section, our model is somewhat misspecified due to replacing unmeasured dynamic components with the day of the year t (Granger and Newbold 1974). While t serves as a good proxy for these dynamic antecedent conditions, it is imperfect due to climatic variations across various years (Wu et al. 2024). This imperfection, as well as inherent day-to-day autocorrelation in streamflow and possible systematic measurement errors (Horner et al. 2018; Sorooshian and Dracup 1980; Schoups and Vrugt 2010; Kim et al. 2023), can lead to strongly autocorrelated errors (Sun et al. 2021; Thursby 1987). For example, suppose that for a given catchment, May is a dry month across most years, but for a particular year, it is extremely wet. Our model will likely choose $\beta(s, t)$ to be fairly small in May to fit the data across most years, however, for the particularly wet May, our model will systematically underpredict streamflow (i.e., autocorrelated errors) since streamflow has persistent magnitudes and our $\hat{\beta}(s, t)$ is too low for May in that particular year. If we ignore this property of the errors, we violate the assumptions of linear regression (independent errors). Naturally, one could consider previous lags of the response to improve predictive capabilities and reduce autocorrelation, as is done in the classical time series framework ARIMAX (Box et al. 2015), however, prioritizing forecasting accuracy leads to several issues. First, previous streamflow does not cause future streamflow, thus incorporating it into the model leads us further from the true data-generating processes thereby instigating incorrect causal effect estimates (Dafoe 2018; Barnett et al. 2017; Runge 2021; Achen 2000). Furthermore, while data for precipitation is globally available and accurate, other data such as streamflow or evaporation is usually inaccurate or unavailable. Since hydrologists often require a model that is easily transferable to all ungauged or data-limited regions (Janssen and Ameli 2021), explicit use of previous streamflow values must be excluded from the model.

The most common way to account for autocorrelated errors while not explicitly using lagged response values in the model, has

been the Cochrane-Orcutt or Prais-Winsten correction method (Cochrane and Orcutt 1949; Prais and Winsten 1954). Unfortunately, both Cochrane-Orcutt and Prais-Winsten corrections failed to outperform our method without autocorrelation correction for parameter estimation in our initial simulation studies. It is unclear exactly why these methods failed, but we have several suspicions. First, we noticed that artificially reducing the smoothness penalty term found from our validation set led to better results, but there was no consistent reduction we could apply to produce optimal results. Second, we noticed that if we only consider a few lags (i.e., $D < 5$), the corrective methods far outperformed default regression, however if $D > 100$, default regression far outperformed the corrective methods. Reviewing the causes of these and other issues could be an interesting direction for future work (Dagenais 1994; Sims 1972; McGuirk and Spanos 2009; Mizon 1995; Dafoe 2018; Thursby 1987).

Instead, we found that using an autoregressive distributed lag model followed by deconvolution achieved better results in the presence of autocorrelated errors. This method is comprehensively explained in Kirchner (2022) but also appears in many earlier works (Baltagi and Baltagi 2011; Pagano and Hartley 1981; Tsay 1985; Young 2002; Wilkins 2018). As a simple illustration, consider the equations

$$y_t = \alpha_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \epsilon_t \quad (4)$$

and

$$y_t = \beta_{0,t} x_t + \beta_{1,t} x_{t-1} + u_t, \quad (5)$$

where the ϵ_t is a white noise process from $N(0, \sigma^2)$ and $u(t) = \alpha_1 u_{t-1} + \epsilon_t$ is an AR(1) process. The two models are equivalent: we can subtract from (5) its lagged self-multiplied by α_1 to obtain

$$y_t = \alpha_1 y_{t-1} + \beta_{0,t} x_t + (\beta_{1,t} - \alpha_1 \beta_{0,t-1}) x_{t-1} - \alpha_1 \beta_{1,t-1} x_{t-2} + \epsilon_t.$$

We can therefore readily estimate the parameters of (4) for the simpler case of independent errors, then use deconvolution to obtain the β coefficients of interest in (5) from the modified b coefficients (see Kirchner (2022) for further details). The above simple example is easily applied to our more complex methodology and implemented in the algorithm below. Fortunately, this also reduces the computational cost compared with the Prais-Winsten procedure due to only requiring a single model fit, with a diagnostic test of no autocorrelation, followed by a simple deconvolution, instead of two model fits. Note that all autocorrelated error correction approaches implicitly assume the common factor restriction implying that the response does not cause the lagged explanatory variables (McGuirk and Spanos 2009; Cook and Webb 2021), a safe assumption for our application.

To summarize, our problem setting incurs autocorrelated errors (which violates one of the core assumptions of linear regression) due to slight model misspecification and the inherent properties of our problem setting. The more traditional methods for autocorrelation correction from Cochrane-Orcutt and Prais-Winsten work poorly in our setting for reasons that should be explored in future work. In the end, we use the deconvolution method suggested by Kirchner (2022), as it removes the autocorrelation, however, from our extensive testing, it does not significantly improve parameter estimation capabilities.

2.2 | Algorithm

Let the discretized response $\mathbf{Y}$ be defined as a column vector of length $N - (D - 1)$. We then define the sparse matrix $\mathbf{Z} \in \mathbb{R}^{(N-(D-1)) \times K}$ which has D non-zero entries in each row corresponding to the valid entries of $\beta(s, t)$ given the current time of $y_i(t)$ and the Whittaker basis. Using the Whittaker basis saves significant computational effort when computing the design matrix $\mathbf{Z}$. Instead of needing to evaluate NK integrals, no integration is needed as each entry of $\mathbf{Z}$ is either 0 or equal to the data point itself. Quantitatively, given similar amounts of data/parameters (about 10000 data points and about 50000 parameters), in MATLAB using the code and triangular basis functions of Xun et al. (2022), computing $\mathbf{Z}$ takes about 10 min while in R with Whittaker basis functions, it takes less than half a second. This is the case even while R is known to be slower compared with MATLAB according to the latest benchmarks done by the Julia team (Bezanson et al. 2017). Here, $\mathbf{b}$ is the vector of $b_k \forall k \in 1, \dots, K$. For example, suppose $n = 2$, $D = 2$, $T = 3$, and $K = 6$, then

$$\mathbf{Y} = \begin{bmatrix} y_1(2) \\ y_1(3) \\ y_2(1) \\ y_2(2) \\ y_2(3) \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} z_{1,1}(2) & z_{1,2}(2) & z_{1,3}(2) & z_{1,4}(2) & z_{1,5}(2) & z_{1,6}(2) \\ z_{1,1}(3) & z_{1,2}(3) & z_{1,3}(3) & z_{1,4}(3) & z_{1,5}(3) & z_{1,6}(3) \\ z_{2,1}(1) & z_{2,2}(1) & z_{2,3}(1) & z_{2,4}(1) & z_{2,5}(1) & z_{2,6}(1) \\ z_{2,1}(2) & z_{2,2}(2) & z_{2,3}(2) & z_{2,4}(2) & z_{2,5}(2) & z_{2,6}(2) \\ z_{2,1}(3) & z_{2,2}(3) & z_{2,3}(3) & z_{2,4}(3) & z_{2,5}(3) & z_{2,6}(3) \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix}$$

Our method first estimates a smooth coefficient function via:

$$\min_{\mathbf{b}} \|\mathbf{Y} - \mathbf{Z}\mathbf{b}\|_2^2 + w_h \|D_H \mathbf{b}\|_2^2 + w_v \|D_V \mathbf{b}\|_2^2 \quad (6)$$

and the group 2-norms are computed using this initial coefficient function estimation (see the end of Section 2). Then, to optimize for our time-varying maximum lag parameter $\delta(t)$, a sequence of models are fit using coefficients with corresponding group norms greater than a sequence of q values in

$$\min_{\mathbf{b}} \|\mathbf{Y} - \mathbf{Z}\mathbf{b}\|_2^2 + w_h \|D_H \mathbf{b}\|_2^2 + w_v \|D_V \mathbf{b}\|_2^2 \quad \text{s.t. } b_{A_{ij}} = 0 \text{ when } \|b_{A_{ij}}\|_2^2 < q.$$

Once the optimal q is chosen (selected via a user's choice of methods), to account for autocorrelated errors, previous response values are added as covariates and modified coefficients are computed. Finally, the coefficients are deconvolved back into their intended form. A detailed description of our algorithm is provided in Algorithm 1.

3 | Hydrology Data Application

We apply our methodology to data from two diverse catchments to show its versatility for hydrology. For the real data experiments, the relationship between actual rainfall and streamflow is modeled. The simulation study also uses real rainfall data, to mimic actual patterns in the time series, but streamflow is simulated according to known coefficient functions with known lag structures, so we can assess estimation performance.

The first catchment we explore is the Koksilah River which is located in Cowichan, British Columbia, Canada. The area of this

ALGORITHM 1 |

Require: One response time series $\mathbf{Y} = [y_1, \dots, y_N]$ and one explanatory time series $\mathbf{X} = [x_1, \dots, x_N]$, split into training and validation sets (and test set if needed).

- 1: Optimize $\mathbf{b}$ on the training set given a set of hyperparameter pairs (w_h, w_v) , and find the optimal hyperparameters such that some objective function (i.e., R^2) is maximized on a validation set.
- 2: Find optimal $\mathbf{b}$ using the optimal weights w_h and w_v and the least squares criterion in Equation (6).
- 3: Compute the group norms $\|\mathbf{b}_{A_{ij}}\|_2^2$ for all (s, t) .
- 4: Find optimal sparsity threshold q to obtain $\delta(t)$.
- 5: Given q and using Step 1, reoptimize the smoothing weights, w_h and w_v , and refit the model by computing $\mathbf{b}$ using all data and $y_{t-1} \dots y_{t-c}$ as additional covariates.
- 6: Deconvolve the optimized coefficients to obtain $\beta(s, t)$.
- 7: **return** $\beta(s, t)$ and $\delta(t)$

catchment is 236 km² with an average elevation of 461 meters. About 11% of precipitation falls as snow and it is an extremely wet catchment with an aridity index (fraction of potential evapotranspiration over precipitation) of just 0.37. The second catchment we explore is the Withlacoochee River which is located in Dade City, Florida, United States. The area of the catchment is 650 km², and it is low lying with an average elevation of only 42 meters above sea level. It does not experience any snow and it is a fairly dry catchment with an aridity index of 1.07. To obtain rainfall for both these areas, we use EMDNA (Tang et al. 2021), a high-quality climate dataset that has complete daily precipitation and temperature values for 1979–2018 at 10 km grid squares across North America. From this data, daily rainfall values are generated for the Koksilah and Withlacoochee rivers by averaging precipitation and temperature values across the catchments, then using a temperature threshold of 0°C, we compute daily rainfall (Figure 2). Leap days are removed, leaving us with $365 \times 40 = 14,600$ observations. We decided that the maximum reasonable lag should be $D = 150$ days. Indeed, regardless of the catchment location, all precipitation in rain-dominated catchments should drain, become captured by deep groundwater systems, or be evaporated from the catchment within 150 days (5 months) (Jasechko et al. 2016; Tennant et al. 2020), leaving a total of $365 \times 150 = 54,750$ parameters to be estimated to construct $\beta(s, t)$ (Equation (6)). The choice of maximum lag also alters the total number of available observations since to predict the response we require the previous 150 days of explanatory data. Therefore, the final number of observations is 14,451.

Hyperparameters w_h and w_v are optimized with Bayesian optimization using the `GPfit` R package (MacDonald et al. 2015). We start by evaluating 30 uniformly random initial hyperparameter sets on the validation set and fitting a Gaussian process to said points. The following three steps are then repeated 35 times to further optimize hyperparameters: (1) we choose the next point based on the Gaussian process model by maximizing the expected increase in validation set R^2 , (2) we evaluate the new point, and (3) we refit the Gaussian process model with all points that have thus far been evaluated. After experimenting on several catchments, we noticed that the optimal w_h is often much larger than

the optimal w_v , thus we set our search range as $e^{[8,24]}$ for w_h and $e^{[-5,15]}$ for w_v .

Choosing the optimal threshold q is often the most difficult step in Algorithm 1. We do not specifically define how this should be done in the methods section as we leave it up to the user and their goals, although we will give some guidance in the context of our hydrological application. Minimizing errors on the validation set seemed to give high test set R^2 scores, however, the final coefficient function indicated too little sparsity. The phenomena was also found in Rushworth et al. (2013) who observed optimizing the AIC leads to too little smoothness, pushing the authors to use the AIC-optimal smoothness parameter as a lower bound and artificially increasing this value. In all experiments below, we find the “knee-point” as outlined in Satopaa et al. (2011) to choose q since it gave consistently strong results (see Supplement).

We let $c = 2$ for all experiments, meaning in step 5 of Algorithm 1, we add the previous two values of the response as additional covariates in an AR(2) distributed lag model. This is a conservative decision since we will also be able to accurately model AR(1) errors (Schmidt 1971). All experiments were run in R version 4.2.1.

3.1 | Applications to Real Streamflow

Daily streamflow data from the Koksilah River and the Withlacoochee River were gathered from the Environment Canada HYDAT database and from CAMELS (Addor et al. 2017), respectively. The average daily streamflow for both rivers is visualized in Figure 2. To improve predictive performance and remove patterns and non-Gaussianity from the residual plots, we transform streamflow y via $y_{new} = \log(y + 1)$. The estimates of our hyperparameters w_h and w_v are obtained after splitting the data with 80% of the data for training and 20% of the data for validation. The

final estimates of $\beta(s, t)$, $\delta(t)$, and the whole dataset R^2 are then obtained by training on all available data. To quantify the overall confidence, certainty, and predictive performance of our model, we evaluate our model on a test set. To compute test set R^2 values, we split the data with 60% for training, 20% for optimizing the hyperparameters w_h and w_v , and 20% for testing.

The estimates $\hat{\beta}(s, t)$ for the Koksilah River after training on the entire dataset are shown in Figure 5a. In the final model, the autocorrelated errors were modeled via an AR(2) process with fitted coefficients (0.704, 0.122), showing substantial autocorrelation. In winter, rainfall frequency and magnitude peaks after a rapid ramp up from September to October to November (days 260–320) (Figure 2), therefore, by December (day 335), the soil in the catchment is fully saturated and groundwater is fully connected. When this happens, any additional rain either immediately transmits to flood waters heading towards the stream or the additional rainfall quickly pushes the water already in the soil towards the stream. Furthermore, this behavior can be seen in Figure 2, where December, January, and February (days 330–70) streamflow correlates strongly with rain falling the previous day. Therefore, observing a $\delta(t)$ of around 3 days from days 350–60 empirically confirms and supports existing expert knowledge. Two distinct peaks of maximum lag appear in summer and fall in Figure 5a. The peak in summer at around day 190, with a $\delta(t)$ of about 18 days, displays long term recession behavior in the catchment. Starting in March, this catchment appears to enter a drying phase where evapotranspiration and runoff exceed water input (Figure 2). During the drying phase (see Supplement), rainfall further back in the past begins to make a relatively larger impact on current streamflow, since the relative amount of antecedent dryness is a significant driver for streamflow production. During August (around day 230), the soil is completely dry and temperatures are high, therefore most rainfall contributes to soil wetness and evaporation instead of streamflow. Starting in mid-September (day 260), the Koksilah River catchment enters

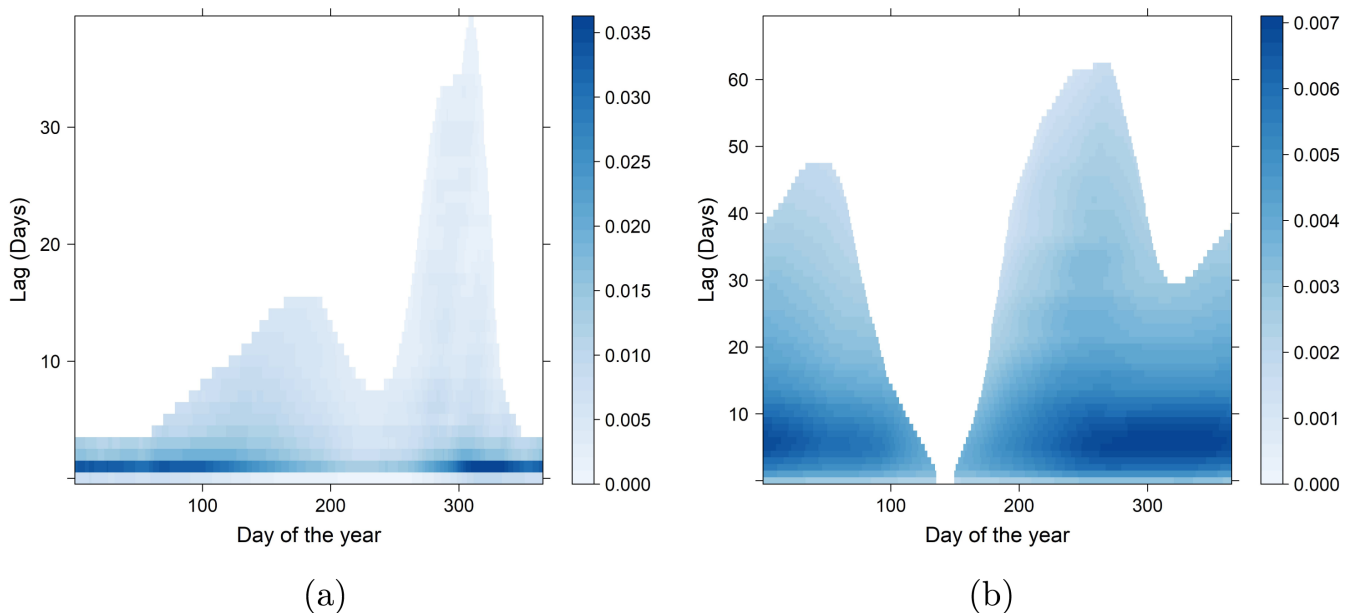


FIGURE 5 | The estimated function $\hat{\beta}(s, t)$ for each lag s and time t for the Koksilah River located in British Columbia (left) and Withlacoochee River, located in Florida (right) from the whole data period (1979–2018). (a) Koksilah River and (b) Withlacoochee River.

a rapid wetting phase as rainfall quickly becomes more frequent (see Supplement). Rainfall occurring up to around 40 days in the past begins to have a significant impact on streamflow, since during this time if the previous days were dry any additional precipitation will contribute to soil wetness, but if previous days were wet, additional precipitation will contribute to streamflow. The wetting phase ends in December (about day 350) when the soil becomes fully saturated regardless of the between-year variations of precipitation in the wettest month, November. On the test set, $R^2 = 0.82$ (in the transformed space), indicating that we can have a high level of confidence in estimates of $\beta(s, t)$ and $\delta(t)$ in Figure 5a. Clearly, the Koksilah River shows substantial nonstationarity in both the magnitude of the impulse response ($\beta(s, t)$) and the maximum lag ($\delta(t)$), thus our methodological contributions allow hydrologists to draw unprecedented and important inferences about this catchment.

For the second catchment, the Withlachoochee River in Florida, the R^2 was 0.55 on the test dataset (in the transformed space), indicating that we can have a moderate level of confidence in the coefficients plotted in Figure 5b. In the final model, the autocorrelated errors were modeled via an AR(2) process with fitted coefficients (1.389, -0.408), showing extreme autocorrelation. Although the Withlachoochee River results presented in Figure 5b look very different compared with the results from the Koksilah River (Figure 5a), they can be interpreted in a similar fashion. We first note that this catchment may not experience overland flow where rainfall bypasses the storage and strongly affects streamflow in a short period of time. This is reasonable since this is a dry catchment with an aridity index of greater than one, meaning that the potential for evaporation is greater than total precipitation. Starting in April (days 100–175), the catchment enters a dry phase. By this time of the year, the catchment has experienced about 6 months of relatively dry weather and temperatures have begun to rise, therefore in May, most rainfall evaporates or contributes to soil moisture instead of travelling as streamflow. During this dry period, both $\beta(s, t)$ and $\delta(t)$ are near zero indicating very little streamflow response. By mid to late June (about day 180), a consistent amount of rainfall usually falls, outpacing evaporation (Figure 2), thus the catchment soils become wet enough to produce streamflow. During this wetting period, streamflow responds slowly to water input ($\delta(t)$) rises to over 60 days), with strong and persistent recession behavior (fairly constant $\beta(s, t)$ across lags). As the catchment becomes more saturated, in the fall and early winter (day 275–365), we begin to see the largest magnitudes of $\beta(s, t)$ and the response times $\delta(t)$ decrease from 60 to 30 days, indicating a wet phase. From the beginning of the calendar year until day 100, rainfall falls below evaporation and streamflow, leading to decreasing impulse response magnitudes (decreasing $\beta(s, t)$) and larger lags becoming significant again ($\delta(t)$ rises to almost 50) (see Supplement Figure S1 for further comparisons).

Both real-world studies lead to interesting and meaningful results with good R^2 scores on the test set. In both cases, wetting, wet, drying, and dry periods could be delineated. Furthermore, even although the positivity of the coefficient function was not imposed as a constraint for our models, in both cases, the model correctly identified that there is a strong positive relationship between rainfall and streamflow at all times and all lags. This is a further indication that our method is reliable and can correctly

identify dominant processes. We also note that the estimated $\delta(t)$ for each catchment varies smoothly with day of the year as expected. The test set R^2 results for both catchments are similar to those seen from neural networks which are more complex and less interpretable (Hoedt et al. 2021; Anderson and Radić 2022; Kratzert et al. 2018). Importantly, both catchments show the importance of considering a time-varying maximum lag $\delta(t)$, which was our primary methodological contribution. Without inferring the changes in these lags over the year, learning how to divide the year into wet, drying, dry, and wetting periods for each catchment would have been substantially more difficult.

3.2 | Benchmark Against Previous Hydrological Inference Methods

In recent years, hydrologists and statisticians have developed several methods that are specifically made for understanding the temporal dynamics of rainfall–runoff relationships (Schranner et al. 2025; Kirchner 2022). Each method parameterizes their models differently, therefore the correctness of their inferences about the ground-truth filtering function is difficult to compare; however, we can compare their ability to fit and predict streamflow.

In this subsection, we benchmark our model (HFLM-DS) against the most recent hydrological inference methods by comparing their predictive performances on the test sets of our two catchments. To the best of our knowledge, there are currently two primary methods that aim to infer the filtering function between rainfall and streamflow. The first method, developed by Kirchner (2022), extends the popular unit hydrograph model by allowing users to develop a series of unit hydrograph models that change with precipitation intensity and antecedent conditions (Kirchner 2024). This method is similar to the method developed in this article; however, instead of using day of the year to automatically group data, the user must prespecify how to break the full time series into groups based on antecedent conditions and rainfall intensity. In particular, to implement ensemble rainfall–runoff analysis (ERRA) models, we follow the modeling advice of Gao et al. (2025) by using the previous 5 days of precipitation as an indication of antecedent wetness while using 4 thresholds of precipitation intensity. The second method (Schranner et al. 2025), called Gaussian sliding window regression (GSWR), uses a series of Gaussian kernels to infer the timing and magnitudes of the different flow paths as shown in Figure 1 (i.e., overland flow, shallow subsurface flow, and groundwater flow). We follow the implementation details described in Schranner et al. (2025) during our experiment.

From the results summarized in Table 1, it is clear that HFLM-DS (average $R^2 = 0.74$) represents an important improvement over ERRA (average $R^2 = 0.64$) since it is more interpretable (due to its ability to infer the dynamics of the largest important lag) while having significantly better predictive performance. Furthermore, HFLM-DS (average $R^2 = 0.74$) vastly improves over GSWR (average $R^2 = 0.55$) in terms of the ability of each model to represent the complexities in streamflow data. This difference in predictive capability is likely due to the ability of HFLM-DS to change its filtering function with time (day of the year), whereas

TABLE 1 | Summary of the test set R^2 values (on raw observed streamflow) across three different methods (columns) and two different catchments (rows). The best result in each row is bolded.

Location	GSWR	ERRA	HFLM-DS (ours)
Koksilah river	0.70	0.78	0.81
Withlacoochee river	0.39	0.50	0.67

GSWR assumes that the rainfall–runoff relationship is stationary in time.

4 | Simulation Studies

The aim of our simulation study is to quantify and understand our ability to estimate $\beta(s, t)$ and $\delta(t)$ in various noisy environments.

4.1 | Evaluation Criteria

Xun et al. (2022) compared methods by computing the root mean squared error of δ , the percent bias of δ , and the mean integrated squared error of $\beta(s, t)$. The percent bias of δ does not fully capture the accuracy of δ estimates when δ varies over time, and the other metrics, such as the root integrated squared error, are less interpretable compared with R^2 , thus we introduce three evaluation criteria for our simulation studies. The first criterion, $R^2(\beta, \hat{\beta})$, aims to evaluate the accuracy in estimating the true coefficient function in an interpretable fashion:

$$R^2(\beta, \hat{\beta}) = 1 - \frac{\int \int \{\beta(s, t) - \hat{\beta}(s, t)\}^2 ds dt}{\int \int \{\beta(s, t) - \bar{\beta}\}^2 ds dt} \quad (7)$$

The second criterion, $\delta(t)$ -bias, is the average of $\hat{\delta}(t)$ across time minus the average of $\delta(t)$ across time. Our third criterion, $\delta(t)$ -correlation evaluates how well we can estimate the inter-period variability in $\delta(t)$ across time. This metric is simply calculated as the correlation between the ground-truth $\delta(t)$ and the estimated $\hat{\delta}(t)$. While biases could cancel over time, and strong correlation does not imply that $\hat{\delta}(t)$ has the correct scale, together these evaluation metrics account for each other's shortcomings while independently providing information about the accuracy in which we are estimating the scale and temporal dynamics of $\delta(t)$. Ideally, β - R^2 and $\delta(t)$ -correlation are close to one, while $\delta(t)$ -bias is close to zero.

4.2 | Simulation Scenarios

In this section, we run eight simulation studies that aim to quantify our method's ability to estimate $\beta(s, t)$ and $\delta(t)$ in noisy environments. We use the same real rainfall data used in the previous section to form the covariate matrix $\mathbf{Z}$, then we define eight scenarios with different values for $\beta(s, t)$, $\delta(t)$, noise levels, and noise autocorrelation coefficients, such that known ground-truth values for the response $y_i(t)$ can be simulated. Streamflow is notoriously difficult to predict. Depending on the location of the catchment and its dynamics, even when using powerful black-box

methods such as long short-term memory (LSTM) neural networks (Hochreiter and Schmidhuber 1997), combined with more predictor variables, the test set R^2 can range all the way from zero to one (Ayzel and Heistermann 2021; Hoedt et al. 2021; Kratzert et al. 2018). In future scenarios for which our method may be applied, we hypothesize that inferences from models with $R^2 < 0.4$ will be difficult. Furthermore, from previous works we know that interpretable models with $R^2 > 0.8$ will be rare in hydrology. Thus, we simulate the response vector with additional noise such that if we calculate the true response $y_{\text{true}}(t)$ from the ground-truth $\beta(s, t)$ we would produce $R^2(y, y_{\text{true}}) = 0.8$ or $R^2(y, y_{\text{true}}) = 0.4$, where

$$R^2(y, y_{\text{true}}) = 1 - \frac{\int \{y(t) - y_{\text{true}}(t)\}^2 dt}{\int \{y(t) - \bar{y}\}^2 dt}, \quad (8)$$

$y(t)$ is the simulated data after adding noise to $y_{\text{true}}(t)$, and $\bar{y}$ is the average of $y(t)$ over t . Furthermore, we specify two possible levels of autocorrelation for our simulated noise. From our results in the previous section, we know that the noise can be highly autocorrelated, thus we specify medium AR(0.6, 0.1) and high AR(1.5, −0.52) autocorrelation scenarios.

The eight scenarios emulate the patterns obtained in the real hydrology study in Section 3.1. The estimated $\beta(s, t)$ values from Figure 5a and 5b for the two catchments, along with their respective rainfall values, allow us to simulate two sets of noiseless ground-truth response time series. With two noise levels and two autocorrelation levels for the additive error, there are a total of eight simulation scenarios. For each scenario, we repeat data generation 100 times such that stable results and uncertainty levels can be obtained. Each simulation iteration takes less than 10 min on an Intel i9-9980HK 2.4 GHz processor with 32 GB of RAM, so our methods are computationally feasible for any modern hardware setup.

4.3 | Simulation Results

Table 2 shows that the introduced method is quite promising and robust. We consistently observe that regardless of location, the data with less noise allowed for more accurate and stable inferences of $\beta(s, t)$ and $\delta(t)$. We also consistently observed that higher autocorrelation leads to poorer $\delta(t)$ -corr results. The average results displayed in Table 2 as well as the individual simulation runs consistently resulted in negative $\delta(t)$ -biases, revealing systematic errors. This indicates that on average $\delta(t) > \hat{\delta}(t)$, meaning that the estimated coefficient function is often too sparse and the chosen q should decrease, suggesting that our method for choosing the threshold q is not optimal. Although our chosen sparsity levels are too large due to wrongfully removing some small but non-zero coefficients, we note that the biases seen in Table 2 are relatively small compared with the magnitudes of $\delta(t)$ shown in Figure 5a and 5b, thus we believe our results are satisfactory (see Figure S3 for a visualization).

For the Koksilah River, $R^2(\beta, \hat{\beta})$ was consistently above 0.9 for the high noise scenarios and above 0.96 for the low noise scenarios. The high noise scenarios also gave consistently worse estimates of $\delta(t)$, with higher autocorrelation also having an impact.

TABLE 2 | Summary of the simulation study results.

Location	$R^2(y, y_{\text{true}})$	AR(2) coeffs	$R^2(\beta, \hat{\beta})$	$\delta(t)$ -bias	$\delta(t)$ -corr
Koksilah river	0.4	0.6, 0.1	0.926 (0.016)	−2.699 (1.806)	0.772 (0.126)
Koksilah river	0.8	0.6, 0.1	0.971 (0.005)	−1.773 (0.769)	0.925 (0.043)
Koksilah river	0.4	1.5, −0.52	0.904 (0.030)	−4.133 (2.60)	0.534 (0.280)
Koksilah river	0.8	1.5, −0.52	0.961 (0.009)	−2.42 (0.944)	0.850 (0.145)
Withlacoochee river	0.4	0.6, 0.1	0.927 (0.014)	−2.773 (1.684)	0.943 (0.039)
Withlacoochee river	0.8	0.6, 0.1	0.968 (0.005)	−2.391 (0.571)	0.992 (0.003)
Withlacoochee river	0.4	1.5, −0.52	0.888 (0.037)	−4.912 (4.155)	0.848 (0.097)
Withlacoochee river	0.8	1.5, −0.52	0.934 (0.015)	−3.736 (1.045)	0.982 (0.009)

Note: The average (standard deviation) across the 100 replications is shown for each of the evaluation criteria defined in Section 4.1.

For the Withlacoochee River, the estimates for $\beta(s, t)$ are consistently accurate with R^2 ranging from 0.888 to 0.927 in the high noise scenario and from 0.934 to 0.968 in the low noise scenario. For both noise scenarios, the bias was about −4 days for the high correlation scenarios days while it was about −2.5 days for the low correlation scenarios. Considering $\delta(t)$ can range from 2 to over 60 (Figure 5b), these bias results are quite promising. Furthermore, our method could recover $\delta(t)$ with a correlation above 0.98 in both low noise scenarios. The $\delta(t)$ correlation was slightly lower for the high noise low autocorrelation scenario (0.943) and much lower for the high noise high autocorrelation scenario (0.848).

5 | Discussion and Conclusions

In this work, we build off the iterations of the functional historical linear model introduced by Malfait and Ramsay (2003), Harezlak et al. (2007), Rushworth et al. (2013), and Xun et al. (2022). We greatly simplify their formulations while allowing for dynamic sparsity and high resolution estimates of the coefficient function using the Whittaker basis. In Section 2, we illustrate our algorithm with several diagrams and introduce the three minor assumptions we make to reduce the highly overparameterized problem to one that is tractable. We only assume horizontal smoothness, vertical smoothness, and that sparsity is concentrated at larger lags to keep our methods flexible and applicable to various applications outside of our main hydrologically centered goal. Throughout the work, we are strongly driven by the goal of accurately estimating rainfall–runoff relationships. After gathering rainfall and streamflow data from Vancouver Island and Florida, we estimated the coefficient functions that filter rainfall into streamflow at both locations and compared our conclusions with expert knowledge. We found that the functions have an interesting hydrological interpretation where catchments go through four distinct phases: (1) wetting, (2) wet, (3) drying, and (4) dry. Because the Koksilah River (Vancouver Island) has high seasonal variability in temperature and rainfall, the four phases can be distinctly parsed. On the other hand, we found that the Withlacoochee River (Florida) is an extremely dry catchment, therefore it never enters the wet phase, but all other phases can be inferred. Indeed, none of these inferences could have been made without the methodological improvements pertaining to the dynamic sparsity. Finally, in an extensive simulation study, we found that our simple methodology can successfully

recover the ground-truth coefficient function $\beta(s, t)$ as well as the dynamic time lag $\delta(t)$ with high accuracy. Although the accuracy for which we can recover the coefficient function and time-lag clearly depends on the level of noise and autocorrelation, we found that even when the R^2 is less than or equal to 0.4 and the errors are extremely autocorrelated, accurate conclusions can be made from our inferences.

Several shortcomings of our methods and experiments can be identified. First, although our algorithm can potentially include multiple features, these additional features may not follow our sparsity and smoothness assumptions. For example, if we included snowfall as a predictor of streamflow in a snow-rich catchment, we may expect snowfall to take days or weeks or even months to melt and begin to contribute to streamflow. This would contradict our sparsity assumption since very low and very high lags would be expected to have zero coefficients. Second, we assume throughout the work that the behavior of the catchment does not change across different years. Although our method does not capture year-to-year relationship variability, it can be used to extract multiple coefficient functions before and after some known change. For example, one could estimate the rainfall–runoff relationship before and after a large change in land use or before and after a period of severe climate change. Focusing now on the limitations of our experiments, we note that our experiments are only limited to a specific application in hydrology, although we hypothesize that our methods can work well in other domains. Furthermore, our simulation study is limited since we provided ground-truths that we know can be reached from the algorithm. This limitation points back to the second methodological limitation since our method may only be able to recover certain types of smoothly varying coefficient functions.

In future work, one could solve one or more of the above shortcomings. Perhaps the flexibility of our methods could also be improved by adding adaptive smoothness in a similar fashion as the adaptive lasso, since different days or lags could have different smoothness (Zou 2006; Centofanti et al. 2022; Yang and Hong 2017; Martinez and Carroll 2010; Ballout et al. 2023). Adding time-varying AR terms to the autoregressive distributed lag model or accounting for heteroscedasticity may also improve the extraction of filtering functions (Kim et al. 2023; Rao 2004). Furthermore, although for our purposes we adequately quantify our certainty about the estimated coefficients through a test set

evaluation, this could be extended by creating confidence intervals (Chatterjee and Lahiri 2011; Lahiri 1999; Paparoditis and Politis 2003). Specifically, we suggest using the block residual bootstrap as described in Asencio et al. (2014) or Paparoditis and Politis (2003). Extending our method to be appropriate for spatial analysis could also be interesting. With smooth spatial attribution, one could understand how impactful a specific grid cell of a cause is on the response. Furthermore, with sparsity, one could delineate the spatial area that is impactful on the response. Our current work takes the perspective of penalized least squares and sparsity thresholding, but another interesting perspective leading to different methods could come from a causal inference and conditional independence framework (Laumann et al. 2023). Better methods for choosing our hyperparameters could also be explored (Bach 2008; Hall et al. 2009; Chatterjee and Lahiri 2011).

We hope our method can serve as a reliable tool for learning from time series data across multiple areas of science or serve as an inspiration for further methodological development. This area of research is certainly rich with potential scientific discoveries.

Acknowledgments

The supplementary document contains some additional simulation results. The computing codes for replicating the application studies can be downloaded at <https://github.com/HydroML/HFLM-DS>. This research was funded by the Collaborative Research Team Project of Canadian Statistical Sciences Institute (CANSSI) awarded to Ali A. Ameli, William J. Welch, and Jiguo Cao. Joseph Janssen was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) PhD Scholarship. Asad Haris was supported by postdoctoral funding from UBC's Data Science Institute. Jiguo Cao's and William J. Welch's research was partially supported by NSERC Discovery grants (RGPIN-2018-06008 and RGPIN-2019-05019, respectively). Stefan Schrunner gratefully acknowledges financial support from the Norwegian University of Life Sciences (project number 1211130114) for an international stay at the University of British Columbia, Canada.

Data Availability Statement

The data that support the findings of this study are available in HFLM-DS at <https://github.com/HydroML/HFLM-DS>. These data were derived from the following resources available in the public domain: – CAMELS, <https://ral.ucar.edu/solutions/products/camels> – HYDAT, <https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html> – EMDNA, <https://essd.copernicus.org/articles/13/3337/2021/>.

References

Achen, C. H. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." In *Annual Meeting of the Political Methodology Section of the American Political Science Association*, vol. 20, 7–2000. UCLA.

Addor, N., A. J. Newman, N. Mizukami, and M. P. Clark. 2017. "The Camels Data Set: Catchment Attributes and Meteorology for Large-Sample Studies." *Hydrology and Earth System Sciences* 21, no. 10: 5293–5313.

Almon, S. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica: Journal of the Econometric Society* 33: 178–196.

Anderson, S., and V. Radić. 2022. "Evaluation and Interpretation of Convolutional Long Short-Term Memory Networks for Regional

Hydrological Modelling." *Hydrology and Earth System Sciences* 26, no. 3: 795–825.

Asencio, M., G. Hooker, and H. O. Gao. 2014. "Functional convolution models." *Statistical Modelling* 14, no. 4: 315–335.

Ayzel, G., and M. Heistermann. 2021. "The Effect of Calibration Data Length on the Performance of a Conceptual Hydrological Model Versus Lstm and Gru: A Case Study for Six Basins From the Camels Dataset." *Computers & Geosciences* 149: 104708.

Bach, F. R. 2008. "Bolasso: Model Consistent Lasso Estimation Through the Bootstrap." In *Proceedings of the 25th International Conference on Machine Learning*, 33–40. ACM Digital Library.

Ballout, N., L. Etievant, and V. Viallon. 2023. "On the Use of Cross-Validation for the Calibration of the Adaptive Lasso." *Biometrical Journal* 65, no. 5: 2200047.

Baltagi, B. H., and B. H. Baltagi. 2011. "Distributed Lags and Dynamic Models." in *Econometrics*, 4th ed., 131–147. Springer.

Barnett, A. G., D. Stephen, C. Huang, and M. Wolkewitz. 2017. "Time Series Models of Environmental Exposures: Good Predictions or Good Understanding." *Environmental Research* 154: 222–225.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. "Julia: A Fresh Approach to Numerical Computing." *SIAM Review* 59, no. 1: 65–98.

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, no. 3: 1–18.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Centofanti, F., A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini. 2022. "Adaptive Smoothing Spline Estimator for the Function-on-Function Linear Regression Model." *Computational Statistics* 38: 1–26.

Chatterjee, A., and S. N. Lahiri. 2011. "Bootstrapping Lasso Estimators." *Journal of the American Statistical Association* 106, no. 494: 608–625.

Chiu, C.-L., and R. P. Bittler. 1969. "Linear Time-Varying Model of Rainfall-Runoff Relation." *Water Resources Research* 5, no. 2: 426–437.

Clark, M. P., D. Kavetski, and F. Fenicia. 2011. "Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling." *Water Resources Research* 47, no. 9: 1–16.

Cochrane, D., and G. H. Orcutt. 1949. "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms." *Journal of the American Statistical Association* 44, no. 245: 32–61. <https://doi.org/10.1080/01621459.1949.10483290>.

Cook, S. J., and C. Webb. 2021. "Lagged Outcomes, Lagged Predictors, and Lagged Errors: A Clarification on Common Factors." *Political Analysis* 29, no. 4: 561–569.

Dafoe, A. 2018. "Nonparametric Identification of Causal Effects Under Temporal Dependence." *Sociological Methods & Research* 47, no. 2: 136–168.

Dagenais, M. G. 1994. "Parameter Estimation in Regression Models With Errors in the Variables and Autocorrelated Disturbances." *Journal of Econometrics* 64, no. 1–2: 145–163.

Eilers, P. H. 2003. "A Perfect Smoother." *Analytical Chemistry* 75, no. 14: 3631–3636. <https://doi.org/10.1021/ac034173t>.

Eilers, P. H., B. D. Marx, and M. Durbán. 2015. "Twenty Years of p-Splines." *SORT: Statistics and Operations Research Transactions* 39, no. 2: 149–186.

Fan, J., and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association* 96, no. 456: 1348–1360.

- Gao, H., Q. Ju, D. Zhang, Z. Wang, Z. Hao, and J. W. Kirchner. 2025. "Quantifying Dynamic Linkages Between Precipitation, Groundwater Recharge, and Streamflow Using Ensemble Rainfall-Runoff Analysis." *Water Resources Research* 61, no. 1: e2024WR037821.
- Garcia, D. 2010. "Robust Smoothing of Gridded Data in One and Higher Dimensions With Missing Values." *Computational Statistics & Data Analysis* 54, no. 4: 1167–1178.
- Granger, C. W., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2, no. 2: 111–120.
- Gupta, Y. 1968. "An Efficient Method of Estimating a Distributed Lag Model. Technical Report."
- Hall, P., E. R. Lee, and B. U. Park. 2009. "Bootstrap-Based Penalty Choice for the Lasso, Achieving Oracle Performance." *Statistica Sinica* 19: 449–471.
- Harezlak, J., B. A. Coull, N. M. Laird, S. R. Magari, and D. C. Christiani. 2007. "Penalized Solutions to Functional Regression Problems." *Computational Statistics & Data Analysis* 51, no. 10: 4911–4925.
- Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9, no. 8: 1735–1780.
- Hoedt, P.-J., F. Kratzert, D. Klotz, et al. 2021. "Mc-Lstm: Mass-Conserving LSTM." In *International Conference on Machine Learning*, 4275–4286. PMLR.
- Horner, I., B. Renard, J. Le Coz, F. Branger, H. McMillan, and G. Pierrefeu. 2018. "Impact of Stage Measurement Errors on Streamflow Uncertainty." *Water Resources Research* 54, no. 3: 1952–1976.
- Huang, J., J. L. Horowitz, and S. Ma. 2008. "Asymptotic Properties of Bridge Estimators in Sparse High-Dimensional Regression Models." *Annals of Statistics* 36, no. 2: 587–613.
- Huo, X., and X. Ni. 2007. "When Do Stepwise Algorithms Meet Subset Selection Criteria?" *Annals of Statistics* 35: 870–887.
- Janssen, J., and A. A. Ameli. 2021. "A Hydrologic Functional Approach for Improving Large-Sample Hydrology Performance in Poorly Gauged Regions." *Water Resources Research* 57, no. 9: e2021WR030263.
- Janssen, J., V. Radić, and A. Ameli. 2021. "Assessment of Future Risks of Seasonal Municipal Water Shortages Across North America." *Frontiers in Earth Science* 9: 730631.
- Jasechko, S., J. W. Kirchner, J. M. Welker, and J. J. McDonnell. 2016. "Substantial Proportion of Global Streamflow Less Than Three Months Old." *Nature Geoscience* 9, no. 2: 126–129.
- Kim, M., H. H. Bauser, K. Beven, and P. A. Troch. 2023. "Time-Variability of Flow Recession Dynamics: Application of Machine Learning and Learning From the Machine." *Water Resources Research* 59, no. 5: e2022WR032690.
- Kirchner, J. W. 2022. "Impulse Response Functions for Nonlinear, Nonstationary, and Heterogeneous Systems, Estimated by Deconvolution and Demixing of Noisy Time Series." *Sensors* 22, no. 9: 3291.
- Kirchner, J. W. 2024. "Characterizing Nonlinear, Nonstationary, and Heterogeneous Hydrologic Behavior Using Ensemble Rainfall-Runoff Analysis (Erra): Proof of Concept." *Hydrology and Earth System Sciences Discussions* 2024: 1–42.
- Kratzert, F., D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger. 2018. "Rainfall-Runoff Modelling Using Long Short-Term Memory (Lstm) Networks." *Hydrology and Earth System Sciences* 22, no. 11: 6005–6022.
- Kunz, N. C. 2020. "Towards a Broadened View of Water Security in Mining Regions." *Water Security* 11: 100079.
- Lahiri, S. N. 1999. "Theoretical Comparisons of Block Bootstrap Methods." *Annals of Statistics* 27, no. 1: 386–404. <https://doi.org/10.1214/aos/1018031117>.
- Laumann, F., J. von Kügelgen, J. Park, B. Schölkopf, and M. Barahona. 2023. "Kernel-Based Independence Tests for Causal Structure Learning on Functional Data." *Entropy* 25, no. 12: 1597.
- Leng, C., Y. Lin, and G. Wahba. 2006. "A Note on the Lasso and Related Procedures in Model Selection." *Statistica Sinica* 16: 1273–1284.
- Liao, Z., M. Qian, I. M. Kronish, and Y. K. Cheung. 2023. "Analysis of n-Of-1 Trials Using Bayesian Distributed Lag Model With Autocorrelated Errors." *Statistics in Medicine* 42: 2044–2060.
- MacDonald, B., P. Ranjan, and H. Chipman. 2015. "Gpfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs." *Journal of Statistical Software* 64: 1–23.
- Malfait, N., and J. O. Ramsay. 2003. "The Historical Functional Linear Model." *Canadian Journal of Statistics* 31, no. 2: 115–128.
- Martinez, J. G., and R. J. Carroll. 2010. "Oracle Is Not Optimal: Adapting the Adaptive Lasso." *Biostatistics* 11, no. 1: 1–27.
- McGuirk, A., and A. Spanos. 2009. "Revisiting Error-Autocorrelation Correction: Common Factor Restrictions and Granger Non-Causality." *Oxford Bulletin of Economics and Statistics* 71, no. 2: 273–294.
- Meinshausen, N., and B. Yu. 2009. "Lasso-Type Recovery of Sparse Representations for High-Dimensional Data." *Annals of Statistics* 37, no. 1: 246–270.
- Mindham, D., K. Beven, and N. Chappell. 2023. "Rainfall-Streamflow Response Times for Diverse Upland UK Micro-Basins: Quantifying Hydrographs to Identify the Nonlinearity of Storm Response." *Hydrology Research* 54, no. 2: 233–244.
- Mizon, G. E. 1995. "A Simple Message for Autocorrelation Correctors: Don't." *Journal of Econometrics* 69, no. 1: 267–288.
- Moges, E., B. L. Ruddell, L. Zhang, J. M. Driscoll, and L. G. Larsen. 2022. "Strength and Memory of Precipitation's Control Over Streamflow Across the Conterminous United States." *Water Resources Research* 58, no. 3: e2021WR030186.
- Pagano, M., and M. J. Hartley. 1981. "On Fitting Distributed Lag Models Subject to Polynomial Restrictions." *Journal of Econometrics* 16, no. 2: 171–198.
- Paparoditis, E., and D. N. Politis. 2003. "Residual-Based Block Bootstrap for Unit Root Testing." *Econometrica* 71, no. 3: 813–855.
- Pesando, J. E. 1972. "Seasonal Variability in Distributed Lag Models." *Journal of the American Statistical Association* 67, no. 338: 311–312.
- Prais, S. J., and C. B. Winsten. 1954. "Trend Estimators and Serial Correlation. Technical Report, Cowles Commission Discussion Paper Chicago."
- Rao, S. S. 2004. "On Multiple Regression Models With Nonstationary Correlated Errors." *Biometrika* 91, no. 3: 645–659.
- Runge, J. 2021. "Necessary and Sufficient Graphical Conditions for Optimal Adjustment Sets in Causal Graphical Models With Hidden Variables." *Advances in Neural Information Processing Systems* 34: 15762–15773.
- Rushworth, A. M., A. W. Bowman, M. J. Brewer, and S. J. Langan. 2013. "Distributed Lag Models for Hydrological Data." *Biometrics* 69, no. 2: 537–544.
- Satopaa, V., J. Albrecht, D. Irwin, and B. Raghavan. 2011. "Finding a Kneedle in a Haystack: Detecting Knee Points in System Behavior." In *Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171. IEEE.
- Schmidt, P. 1971. "Estimation of a Distributed Lag Model With Second Order Autoregressive Disturbances: A Monte Carlo Experiment." *International Economic Review* 12: 372–380.
- Schoups, G., and J. A. Vrugt. 2010. "A Formal Likelihood Function for Parameter and Predictive Inference of Hydrologic Models With

Correlated, Heteroscedastic, and Non-Gaussian Errors.” *Water Resources Research* 46, no. 10: 1–17.

Schranner, S., P. Pishrobat, J. Janssen, et al. 2025. “A Gaussian Sliding Windows Regression Model for Hydrological Inference.” *Journal of the Royal Statistical Society Series C: Applied Statistics*: qlaf009.

Sims, C. A. 1972. “The Role of Approximate Prior Restrictions in Distributed Lag Estimation.” *Journal of the American Statistical Association* 67: 169–175.

Somers, L. D., and J. M. McKenzie. 2020. “A Review of Groundwater in High Mountain Environments.” *Wiley Interdisciplinary Reviews: Water* 7, no. 6: e1475.

Sorooshian, S., and J. A. Dracup. 1980. “Stochastic Parameter Estimation Procedures for Hydrologic Rainfall-Runoff Models: Correlated and Heteroscedastic Error Cases.” *Water Resources Research* 16, no. 2: 430–442.

Staerk, C., M. Kateri, and I. Ntzoufras. 2018. “Adaptive Subspace Methods for High-Dimensional Variable Selection. PhD thesis, Universitätsbibliothek der RWTH Aachen.”

Sun, F.-K., C. Lang, and D. Boning. 2021. “Adjusting for Autocorrelated Errors in Neural Networks for Time Series.” *Advances in Neural Information Processing Systems* 34: 29806–29819.

Tang, G., M. P. Clark, S. M. Papalexiou, et al. 2021. “Emdna: An Ensemble Meteorological Dataset for North America.” *Earth System Science Data* 13, no. 7: 3337–3362.

Tennant, C., L. Larsen, D. Bellugi, E. Moges, L. Zhang, and H. Ma. 2020. “The Utility of Information Flow in Formulating Discharge Forecast Models: A Case Study From an Arid Snow-Dominated Catchment.” *Water Resources Research* 56, no. 8: e2019WR024908.

Thursby, J. G. 1987. “Ols or Glis in the Presence of Specification Error?: An Expected Loss Approach.” *Journal of Econometrics* 35, no. 2–3: 359–374.

Tibshirani, R. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society: Series B: Methodological* 58, no. 1: 267–288.

Tsay, R. S. 1985. “Model Identification in Dynamic Regression (Distributed Lag) Models.” *Journal of Business & Economic Statistics* 3: 228–237.

Whittaker, E. T. 1922. “On a New Method of Graduation.” *Proceedings of the Edinburgh Mathematical Society* 41: 63–75.

Wilkins, A. S. 2018. “To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis.” *Political Science Research and Methods* 6, no. 2: 393–411.

Wu, J., P. An, C. Zhao, et al. 2024. “Effects of Multi-Year Droughts on the Precipitation-Runoff Relationship: An Integrated Analysis of Meteorological, Hydrological, and Compound Droughts.” *Journal of Hydrology* 634: 131064.

Xun, X., T. Guan, and J. Cao. 2022. “Sparse Estimation of Historical Functional Linear Models With a Nested Group Bridge Approach.” *Canadian Journal of Statistics* 50, no. 4: 1254–1269.

Yang, L., and Y. Hong. 2017. “Adaptive Penalized Splines for Data Smoothing.” *Computational Statistics & Data Analysis* 108: 70–83.

Young, P. C. 2002. “Advances in Real-Time Flood Forecasting.” *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 360, no. 1796: 1433–1450.

Zhang, C.-H., and T. Zhang. 2012. “A General Theory of Concave Regularization for High-Dimensional Sparse Estimation Problems.” *Statistical Science* 27, no. 4: 576–593.

Zhao, P., G. Rocha, and B. Yu. 2009. “The Composite Absolute Penalties Family for Grouped and Hierarchical Variable Selection.” *Annals of Statistics* 37, no. 6A: 3468–3497.

Zhao, P., and B. Yu. 2006. “On Model Selection Consistency of Lasso.” *Journal of Machine Learning Research* 7: 2541–2563.

Zou, H. 2006. “The Adaptive Lasso and Its Oracle Properties.” *Journal of the American Statistical Association* 101, no. 476: 1418–1429.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1.** Supporting Information.