



Evaluation of statistical methods for quantifying fractal scaling in water quality time series with irregular sampling

Qian Zhang¹, Ciaran J. Harman², James W. Kirchner^{3,4,5}

¹University of Maryland Center for Environmental Science at the US Environmental Protection Agency Chesapeake Bay Program Office, 410 Severn Avenue, Suite 112, Annapolis, MD 21403 (formerly, Department of Geography and Environmental Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218)

²Department of Environmental Health and Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

³Department of Environmental System Sciences, ETH Zurich, Universitätstrasse 16, CH-8092 Zurich, Switzerland

⁴Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

⁵Department of Earth and Planetary Science, University of California, Berkeley, California 94720

Correspondence to: Qian Zhang (qzhang@chesapeakebay.net)

1 **Abstract.** River water-quality time series often exhibit fractal scaling, which here refers to
2 autocorrelation that decays as a power law over some range of scales. Fractal scaling presents
3 challenges to the identification of deterministic trends, but traditional methods for estimating
4 spectral slope (β) or other equivalent scaling parameters (*e.g.*, Hurst exponent) are generally
5 inapplicable to irregularly sampled data. Here we consider two types of estimation approaches
6 for irregularly sampled data and evaluate their performance using synthetic time series. These
7 time series were generated such that (1) they exhibit a wide range of prescribed fractal scaling
8 behaviors, ranging from white noise ($\beta = 0$) to Brown noise ($\beta = 2$), and (2) their sampling gap
9 intervals mimic the sampling irregularity (as quantified by both the skewness and mean of gap-
10 interval lengths) in real water-quality data. The results suggest that none of the existing methods
11 fully account for the effects of sampling irregularity on β estimation. First, the results illustrate
12 the danger of using interpolation for gap filling when examining auto-correlation, as the
13 interpolation methods consistently under-estimate or over-estimate β under a wide range of
14 prescribed β values and gap distributions. Second, the long-established Lomb-Scargle spectral
15 method also consistently under-estimates β . A modified form, using only the lowest 5% of the



16 frequencies for spectral slope estimation, has very poor precision, although the overall bias is
17 small. Third, a recent wavelet-based method, coupled with an aliasing filter, generally has the
18 smallest bias and root-mean-squared error among all methods for a wide range of prescribed β
19 values and gap distributions. The aliasing method, however, does not itself account for sampling
20 irregularity, and this introduces some bias in the result. Nonetheless, the wavelet method is
21 recommended for estimating β in irregular time series until improved methods are developed.
22 Finally, all methods' performances depend strongly on the sampling irregularity, highlighting
23 that the accuracy and precision of each method are data-specific. Accurately quantifying the
24 strength of fractal scaling in irregular water-quality time series remains an unresolved challenge
25 for the hydrologic community and for other disciplines that must grapple with irregular sampling.

26 **Key Words**

27 Fractal scaling, autocorrelation, Hurst effect, river water-quality sampling, sampling irregularity,
28 trend analysis

29 **1. Introduction**

30 ***1.1. Autocorrelations in Time Series***

31 It is well known that time series from natural systems often exhibit auto-correlation, that is,
32 observations at each time step are correlated with observations one or more time steps in the past.
33 This property is usually characterized by the autocorrelation function (ACF), which is defined as
34 follows for a process X_t at lag k :

$$\gamma(k) = \text{cov}(X_t, X_{t+k}) \quad (1)$$

35 In practice, auto-correlation has been frequently modeled with classical techniques such as auto-
36 regressive (AR) or auto-regressive moving-average (ARMA) models (Darken *et al.*, 2002; Yue
37 *et al.*, 2002; Box *et al.*, 2008). These models assume that the underlying process has short-term
38 memory, *i.e.*, the ACF decays exponentially with lag k , which implies that the ACF is summable
39 (Box *et al.*, 2008).

40 Although the short-term memory assumption holds sometimes, it cannot adequately describe
41 many time series whose ACFs decay as a power law (thus much slower than exponentially) and
42 may not reach zero even for large lags, which implies that the ACF is non-summable. This



43 property is commonly referred to as long-term memory or fractal scaling, as opposed to short-
 44 term memory (Beran, 2010).

45 ***1.2. Overview of Approaches for Quantification of Fractal Scaling***

46 Several equivalent metrics can be used to quantify fractal scaling. Here we provide a review
 47 of the definitions of such processes and several typical modeling approaches, including both
 48 time-domain and frequency-domain techniques, with special attention to their reconciliation. For
 49 a more comprehensive review, readers are referred to Beran *et al.* (2013), Boutahar *et al.* (2007),
 50 and Witt and Malamud (2013).

51 Strictly speaking, X_t is called a stationary long-memory process if the condition

$$\lim_{k \rightarrow \infty} k^\alpha \gamma(k) = C_1 > 0 \quad (2)$$

52 where C_1 is a constant, is satisfied by some $\alpha \in (0,1)$ (Boutahar *et al.*, 2007; Beran *et al.*, 2013).

53 Equivalently, X_t is a long-memory process if, in the spectral domain, the condition

$$\lim_{\omega \rightarrow 0} |\omega|^\beta f(\omega) = C_2 > 0 \quad (3)$$

54 is satisfied by some $\beta \in (0,1)$, where C_2 is a constant and $f(\omega)$ is the spectral density function
 55 of X_t , which is related to ACF as follows (which is also known as the Wiener-Khinchin theorem):

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega} \quad (4)$$

56 where ω is angular frequency (Boutahar *et al.*, 2007).

57 One popular model for describing long-memory processes is the so-called fractional auto-
 58 regressive integrated moving-average model, or ARFIMA (p, q, d) , which is an extension of
 59 ARMA models and is defined as follows:

$$(1 - B)^d \varphi(B) X_t = \psi(B) \varepsilon_t \quad (5)$$

60 where ε_t is a series of independent, identically distributed Gaussian random numbers $\sim (0, \sigma_\varepsilon^2)$,
 61 B is the backshift operator (*i.e.*, $BX_t = X_{t-1}$), and functions $\varphi(\cdot)$ and $\psi(\cdot)$ are polynomials of order
 62 p and q , respectively. The fractional differencing parameter d is related to the parameter α in Eq.
 63 (2) as follows:

$$d = \frac{1 - \alpha}{2} \in (-0.5, 0.5) \quad (6)$$

64 (Beran *et al.*, 2013; Witt and Malamud, 2013).



65 In addition to a slowly decaying ACF, a long-memory process manifests itself in two other
66 equivalent fashions. One is the so-called “Hurst effect”, which states that, on a log-log scale, the
67 range of variability of a process changes linearly with the length of time period under
68 consideration. This power-law slope is often referred to as the “Hurst exponent” or “Hurst
69 coefficient” H (Hurst, 1951), which is related to d as follows:

$$H = d + 0.5 \quad (7)$$

70 (Beran *et al.*, 2013; Witt and Malamud, 2013). The second equivalent description of long-
71 memory processes, this time from a frequency-domain perspective, is “fractal scaling”, which
72 describes a power-law decrease in spectral power with increasing frequency, yielding power
73 spectra that are linear on log-log axes (Lomb, 1976; Scargle, 1982; Kirchner, 2005).
74 Mathematically, this inverse proportionality can be expressed as:

$$f(\omega) = C_3 |\omega|^{-\beta} \quad (8)$$

75 where C_3 is a constant and the scaling exponent β is termed the “spectral slope.” In particular, for
76 spectral slopes of zero, one, and two, the underlying processes are termed as “white”, “pink” (or
77 “flicker”), and “Brown” (or “red”) noises, respectively (Witt and Malamud, 2013). Illustrative
78 examples of these three noises are shown in **Figure 1a-1c**.

79 In addition, it can be shown that the spectral density function for ARFIMA (p, d, q) is

$$f(\omega) = \frac{\sigma_\varepsilon^2 |\psi(e^{-i\omega})|^2}{2\pi |\varphi(e^{-i\omega})|^2} |1 - e^{-i\omega}|^{-2d} \quad (9)$$

80 for $-\pi < \omega < \pi$ (Boutahar *et al.*, 2007; Beran *et al.*, 2013). For $|\omega| \ll 1$, Eq. (9) can be
81 approximated by:

$$f(\omega) = C_4 |\omega|^{-2d} \quad (10)$$

82 with

$$C_4 = \frac{\sigma_\varepsilon^2 |\psi(1)|^2}{2\pi |\varphi(1)|^2} \quad (11)$$

83 Eq. (10) thus exhibits the asymptotic behavior required for a long-memory process given by Eq.
84 (3). In addition, a comparison of Eq. (10) and (8) reveals that,

$$\beta = 2d \quad (12)$$

85 Overall, these derivations indicate that these different types of scaling parameters (*i.e.*, α , d , and
86 H and β) can be used equivalently to describe the strength of fractal scaling. Specifically, their
87 equivalency can be summarized as follows:



$$\beta = 2d = 1 - \alpha = 2H - 1 \quad (13)$$

88 It should be noted, however, that the parameters d , α , and H are only applicable over a fixed
89 range of fractal scaling, which is equivalent to $(-1, 1)$ in terms of β .

90 Fractal scaling has been increasingly recognized in studies of hydrological time series,
91 particularly for the common task of trend identification. Such hydrological series include
92 riverflow (Montanari *et al.*, 2000; Khaliq *et al.*, 2008; Khaliq *et al.*, 2009; Ehsanzadeh and
93 Adamowski, 2010), air and sea temperature (Fatichi *et al.*, 2009; Lennartz and Bunde, 2009;
94 Franzke, 2012b; Franzke, 2012a), conservative tracers (Kirchner *et al.*, 2000; Kirchner *et al.*,
95 2001; Godsey *et al.*, 2010), and non-conservative chemical constituents (Kirchner and Neal,
96 2013; Aubert *et al.*, 2014). Because for fractal scaling processes the variance of the sample mean
97 converges to zero much slower than the rate of n^{-1} (n : sample size), the fractal scaling property
98 must be taken into account to avoid "false positives" (Type I errors) when inferring the statistical
99 significance of trends (Cohn and Lins, 2005; Fatichi *et al.*, 2009; Ehsanzadeh and Adamowski,
100 2010; Franzke, 2012a). Unfortunately, as stressed by Cohn and Lins (2005), it is "surprising that
101 nearly every assessment of trend significance in geophysical variables published during the past
102 few decades has failed [to do so]", and a similar tendency is evident in the decade following that
103 statement as well.

104 ***1.3. Motivation and Objective of this Work***

105 To account for fractal scaling in trend analysis, one must be able to first quantify the strength
106 of fractal scaling for a given time series. Numerous estimation methods have been developed for
107 this purpose, including Hurst rescaled range analysis, Higuchi's method, Geweke and Porter-
108 Hudak's method, Whittle's maximum likelihood estimator, detrended fluctuation analysis, and
109 others (Taqqu *et al.*, 1995; Montanari *et al.*, 1997; Montanari *et al.*, 1999; Rea *et al.*, 2009;
110 Stroe-Kunold *et al.*, 2009). For brevity, these methods are not elaborated here; readers are
111 referred to Beran (2010) and Witt and Malamud (2013) for details. While these estimation
112 methods have been extensively adopted, they are unfortunately only applicable to regular (*i.e.*,
113 evenly spaced) data, *e.g.*, daily streamflow discharge, monthly temperature, *etc.* In practice,
114 many types of hydrological data, including river water-quality data, are often sampled irregularly
115 or have missing values, and hence their strengths of fractal scaling cannot be readily estimated
116 with the above traditional estimation methods.



117 Thus, estimation of fractal scaling in irregularly sampled data is an important challenge for
118 hydrologists and practitioners. Many data analysts may be tempted to interpolate the time series
119 to make it regular and hence analyzable (Graham, 2009). Although technically convenient,
120 interpolation can be problematic if it distorts the series' autocorrelation structure (Kirchner and
121 Weil, 1998). In this regard, it is important to evaluate various types of interpolation methods
122 using carefully designed benchmark tests and to identify the scenarios under which the
123 interpolated data can yield reliable (or, alternatively, biased) estimates of spectral slope.

124 Moreover, quantification of fractal scaling in real-world water-quality data is subject to
125 several common complexities. First, water-quality data are rarely normally distributed; instead,
126 they are typically characterized by log-normal or other skewed distributions (Hirsch *et al.*, 1991;
127 Helsel and Hirsch, 2002), with potential consequences for β estimation. Moreover, water-quality
128 data also tend to exhibit long-term trends, seasonality, and flow-dependence (Hirsch *et al.*, 1991;
129 Helsel and Hirsch, 2002), which can also affect the accuracy of β estimate. Thus, it may be more
130 plausible to quantify β in transformed time series after accounting for the seasonal patterns and
131 discharge-driven variations in the original time series, which is also the approach taken in this
132 work. For the trend aspect, however, it remains a puzzle whether the data set should be de-
133 trended before conducting β estimation. Such de-trending treatment can certainly affect the
134 estimated value of β and hence the validity of (or confidence in) any inference made regarding
135 the statistical significance of temporal trends in the time series. This somewhat circular issue is
136 beyond the scope of our current work -- it has been previously discussed in the context of short-
137 term memory (Zetterqvist, 1991; Darken *et al.*, 2002; Yue *et al.*, 2002; Noguchi *et al.*, 2011;
138 Clarke, 2013; Sang *et al.*, 2014), but it is not well understood in the context of fractal scaling (or
139 long-term memory) and hence presents an important area for future research.

140 In the above context, the main objective of this work was to use Monte Carlo simulation to
141 systematically evaluate and compare two broad types of approaches for estimating the strength
142 of fractal scaling (*i.e.*, spectral slope β) in irregularly sampled river water-quality time series.
143 Specific aims of this work include the following:

- 144 (1) To examine the sampling irregularity of typical river water-quality monitoring data and
145 to simulate time series that contain such irregularity; and
- 146 (2) To evaluate two broad types of approaches for estimating β in simulated irregularly
147 sampled time series.



148 The first type of approach includes several forms of interpolation techniques for gap filling, thus
149 making the data regular and analyzable by traditional estimation methods. The second type of
150 approach includes the well-known Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982) and
151 a recently developed wavelet method combined with a spectral aliasing filter (Kirchner and Neal,
152 2013). The latter two methods can be directly applied to irregularly spaced data; here we aim to
153 compare them with the interpolation techniques. Details of these various approaches are
154 provided in **Section 3.1**.

155 This work was designed to make several specific contributions. First, it uses benchmark tests
156 to quantify the performance of a wide range of methods for estimating fractal scaling in
157 irregularly sampled water-quality data. Second, it proposes an innovative and general approach
158 for modeling sampling irregularity in water-quality records. Third, while this work was not
159 intended to compare all published estimation methods for fractal scaling, it does provide and
160 demonstrate a generalizable framework for data simulation (with gaps) and β estimation, which
161 can be readily applied toward the evaluation of other methods that are not covered here. Last but
162 not least, while this work was intended to help hydrologists and practitioners understand the
163 performance of various approaches for water-quality time series, the findings and approaches
164 may be broadly applicable to irregularly sampled data in many other scientific disciplines.

165 The rest of the paper is organized as follows. We propose a general approach for modeling
166 sampling irregularity in typical river water-quality data and discuss our approach for simulating
167 irregularly sampled data (**Section 2**). We then introduce the various methods for estimating
168 fractal scaling in irregular time series and compare their estimation performance (**Section 3**). We
169 close with a discussion of the results and implications (**Section 4**).

170 **2. Quantification of Sampling Irregularity in River Water-Quality Data**

171 ***2.1. Modeling of Sampling Irregularity***

172 River water-quality data are often sampled irregularly. In some cases, samples are taken
173 more frequently during particular periods of interest, such as high flows or drought periods; here
174 we will address the implications of the irregularity, but not the (intentional) bias, inherent in such
175 a sampling strategy. In other cases, the sampling is planned with a fixed sampling interval (*e.g.*,
176 1 day) but samples are missed (or lost, or fail quality-control checks) at some time steps during
177 implementation. In still other cases, the sampling is intrinsically irregular because, for example,



178 one cannot measure the chemistry of rainfall on rainless days or the chemistry of a stream that
179 has dried up. Theoretically, any deviation from fixed-interval sampling can affect the subsequent
180 analysis of the time series.

181 To quantify the sampling irregularity, we propose a simple and general approach that can be
182 applied to any time series of monitoring data. Specifically, for a given time series with N points,
183 the time intervals between adjacent samples are calculated; these intervals themselves make up a
184 time series of $N-1$ points that we call Δt . In addition, the following parameters are calculated to
185 quantify its sampling irregularity:

- 186 • L = the length of the period of record,
- 187 • N = the number of samples in the record,
- 188 • $\Delta t_{nominal}$ = the nominal sampling interval under regular sampling (*e.g.*, $\Delta t_{nominal} = 1$ day
189 for daily samples),
- 190 • $\Delta t^* = \Delta t / \Delta t_{nominal}$, the sample intervals non-dimensionalized by the nominal sampling
191 interval,
- 192 • $\Delta t_{average} = L / (N - 1)$ the average of all the entries in Δt .

193 The quantification is illustrated with two simple examples. The first example contains data
194 sampled every hour from 1:00 am to 11:00 am on one day. In this case, $L = 10$ hours, $N = 11$
195 samples, $\Delta t = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$ hour, and $\Delta t_{nominal} = \Delta t_{average} = 1$ hour. The second
196 example contains data sampled at 1:00 am, 3:00 am, 4:00 am, 8:00 am, and 11:00 am. In this
197 case, $L = 10$ hours, $N = 5$ samples, $\Delta t = \{2, 1, 4, 3\}$ hours, $\Delta t_{nominal} = 1$ hour, and $\Delta t_{average} = 2.5$
198 hours. It is readily evident that the first case corresponds to fixed-interval (regular) sampling that
199 has the property of $\Delta t_{average} / \Delta t_{nominal} = 1$ (dimensionless), whereas the second case corresponds to
200 irregular sampling for which $\Delta t_{average} / \Delta t_{nominal} > 1$.

201 The dimensionless set Δt^* contains essential information for determining sampling
202 irregularity. This set is modeled as independent, identically distributed values drawn from a
203 negative binomial (NB) distribution. This distribution has two dimensionless parameters, the
204 shape parameter (λ) and the mean parameter (μ), which collectively represent the irregularity of
205 the samples. The NB distribution is a flexible distribution that provides a discrete analogue of a
206 gamma distribution. The geometric distribution, itself the discrete analogue of the exponential
207 distribution, is a special case of the NB distribution when $\lambda = 1$.



208 The parameters μ and λ represent different aspects of sampling irregularity, as illustrated by
209 the examples shown in **Figure 2**. The mean parameter μ represents the fractional increase in the
210 average interval between samples due to gaps: $\mu = \text{mean}(\Delta t^*) - 1 = (\Delta t_{\text{average}} - \Delta t_{\text{nominal}})/\Delta t_{\text{nominal}}$.
211 Thus the special case of $\mu = 0$ corresponds to regular sampling (*i.e.*, $\Delta t_{\text{average}} = \Delta t_{\text{nominal}}$), whereas
212 any larger value of μ corresponds to irregular sampling (*i.e.*, $\Delta t_{\text{average}} > \Delta t_{\text{nominal}}$) (**Figure 2c**). The
213 shape parameter λ characterizes the similarity of gaps to each other; that is, a small λ indicates
214 that the samples contain gaps of widely varying lengths, whereas a large λ indicates that the
215 samples contain many gaps of similar lengths (**Figure 2a-2b**).

216 To visually illustrate these gap distributions, representative samples of gappy time series are
217 presented in **Figure 1** for the three special processes described above (**Section 1.2**), *i.e.*, white
218 noise, pink noise, and Brown noise. Specifically, three different gap distributions, namely, NB(λ
219 = 1, $\mu = 1$), NB($\lambda = 1$, $\mu = 14$), and NB($\lambda = 0.01$, $\mu = 1$), were simulated and each was applied to
220 convert the three original (regular) time series (**Figure 1a-1c**) to gappy time series (**Figure 1d-**
221 **1f**). These simulations clearly illustrate the effects of the two parameters λ and μ . In particular,
222 compared with NB($\lambda = 1$, $\mu = 1$), NB($\lambda = 1$, $\mu = 14$) shows a similar level of sampling irregularity
223 (same λ) but a much longer average gap interval (larger μ). Again compared with NB($\lambda = 1$, $\mu =$
224 1), NB($\lambda = 0.01$, $\mu = 1$) shows the same average interval (same μ) but a much more irregular
225 (skewed) gap distribution that contains a few very large gaps (smaller λ).

226 2.2. Examination of Sampling Irregularity in Real River Water-Quality Data

227 The above modeling approach was applied to real water-quality data from two large river
228 monitoring networks in the United States to examine sampling irregularity. One such network is
229 the Chesapeake Bay River Input Monitoring program, which typically samples streams bi-
230 monthly to monthly, accompanied with additional sampling during stormflows (Langland *et al.*,
231 2012; Zhang *et al.*, 2015). These data were obtained from the U.S. Geological Survey National
232 Water Information System (<http://doi.org/10.5066/F7P55KJN>). The other network is the Lake
233 Erie and Ohio tributary monitoring program, which typically samples streams at a daily
234 resolution (National Center for Water Quality Research, 2015). For each site, we determined the
235 NB parameters to quantify sampling irregularity. The mean parameter μ can be estimated as
236 described above, and the shape parameter λ can be calculated directly from the mean and
237 variance of Δt^* as follows: $\lambda = \mu^2/[\text{var}(\Delta t^*) - \mu] = (\text{mean}(\Delta t^*) - 1)^2/[\text{var}(\Delta t^*) - \text{mean}(\Delta t^*) + 1]$.
238 Alternatively, a maximum likelihood approach can be used, which employs the “*fitdist*” function



239 in the “*fitdistrplus*” R package (Delignette-Muller and Dutang, 2015). In general, the two
240 approaches have produced similar results, which are summarized in **Table 1**, with two examples
241 of fitted NB distributions shown in **Figure 3**.

242 For the Chesapeake Bay River Input Monitoring program (9 sites), total nitrogen (TN) and
243 total phosphorus (TP) are taken as representatives of water-quality constituents. According to the
244 maximum likelihood approach, the shape parameter λ varies between 0.7 and 1.2 for TN and
245 between 0.8 and 1.1 for TP (**Table 1**). These λ values are around 1.0, reflecting the fact that
246 these sites have relatively even gap distributions (*i.e.*, relatively balanced counts of large and
247 small gaps). The mean parameter μ varies between 9.5 and 19.6 for TN and between 13.4 and
248 24.4 for TP in the Chesapeake monitoring network, corresponding to $\Delta t_{average}$ of 10.5–20.6 days
249 for TN and 14.4–25.4 days for TP, respectively. This is consistent with the fact that these sites
250 have typically been sampled bi-monthly to monthly, along with additional sampling during
251 stormflows (Langland *et al.*, 2012; Zhang *et al.*, 2015).

252 For the Lake Erie and Ohio tributary monitoring program (6 sites), the record of nitrate-plus-
253 nitrite (NO_x) and TP were examined. According to the maximum likelihood approach, the shape
254 parameter λ is approximately 0.01 for both constituents (**Table 1**). These very low λ values occur
255 because these time series contain a few very large gaps, ranging from 35 days to 1109 days (~3
256 years). The mean parameter μ varies between 0.06 and 0.22, corresponding to $\Delta t_{average}$ of 1.06
257 and 1.22 days, respectively. This is consistent with fact that these sites have been sampled at a
258 daily resolution with occasional missing values on some days (Zhang and Ball, 2017).

259 **2.3. Simulation of Time Series with Irregular Sampling**

260 To evaluate the various β estimation methods, our first step was to use Monte Carlo
261 simulation to produce time series that mimic the sampling irregularity observed in real water-
262 quality monitoring data. We began by simulating regular (gap free) time series using the
263 fractional noise simulation method of Witt and Malamud (2013), which is based on inverse
264 Fourier filtering of white noises. Our analysis showed this method performed well compared to
265 other simulation methods for β values between 0 and 1 (data not shown). In addition, this
266 method can also simulate β values beyond this range. The noises simulated by the Witt and
267 Malamud method, however, are band-limited to the Nyquist frequency (half of the sampling
268 frequency) of the underlying white noise time series, whereas true fractional noises would
269 contain spectral power at all frequencies, extending well above the Nyquist frequency for any



270 sampling. Thus these band-limited noises will be less susceptible to spectral aliasing than true
271 fractional noises would be; see Kirchner (2005) for detailed discussions of the aliasing issue.

272 Thirty replicates of regular (gap free) time series were produced for nine prescribed spectral
273 slopes, which vary from $\beta = 0$ (white noise) to $\beta = 2$ (Brownian motion or “random walk”) with
274 an increment of 0.25 (*i.e.*, 0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, and 2). These regular time series
275 each have a length (N) of 9125, which can be interpreted as 25 years of regular daily samples
276 (that is, $\Delta t_{nominal} = 1$ day).

277 Each of the simulated regular time series was converted to irregular time series using gap
278 intervals that were simulated with NB distributions. To make these gap intervals mimic those in
279 typical river water-quality time series, representative NB parameters were chosen based on
280 results from **Section 2.2**. Specifically, μ was set at 1 and 14, corresponding to $\Delta t_{average}$ of 2 days
281 and 15 days respectively. For λ , we chose four values that span three orders of magnitude, *i.e.*,
282 0.001, 0.1, 1, and 10. Note that when $\lambda = 1$ the generated time series corresponds to a Bernoulli
283 process. With the chosen values of μ and λ , a total of eight scenarios were generated, which were
284 implemented using the “*rnbinom*” function in the “*stats*” R package (R Development Core Team,
285 2014):

- 286 1) $\mu = 1$ (*i.e.*, $\Delta t_{average} / \Delta t_{nominal} = 2$), $\lambda = 0.01$,
- 287 2) $\mu = 1$, $\lambda = 0.1$,
- 288 3) $\mu = 1$, $\lambda = 1$,
- 289 4) $\mu = 1$, $\lambda = 10$,
- 290 5) $\mu = 14$ (*i.e.*, $\Delta t_{average} / \Delta t_{nominal} = 15$), $\lambda = 0.01$,
- 291 6) $\mu = 14$, $\lambda = 0.1$,
- 292 7) $\mu = 14$, $\lambda = 1$,
- 293 8) $\mu = 14$, $\lambda = 10$.

294 Examples of these simulations are shown with boxplots in **Figure 2**.

295 **3. Evaluation of Proposed Estimation Methods for Irregular Time Series**

296 **3.1. Summary of Estimation Methods**

297 For the simulated irregular time series, β was estimated using the aforementioned two types
298 of approaches. The first type includes 11 different interpolation methods (designated as B1-B11
299 below) to fill the data gaps, thus making the data regular and analyzable by traditional methods:



- 300 B1) Global mean: all missing values replaced with the mean of all observations;
301 B2) Global median: all missing values replaced with the median of all observations;
302 B3) Random replacement: all missing values replaced with observations randomly drawn
303 (with replacement) from the time series;
304 B4) Next observation carried backward: each missing value replaced with the next available
305 observation;
306 B5) Last observation carried forward: each missing value replaced with the preceding
307 available observation;
308 B6) Average of the two nearest samples: it replaces each missing value with the mean of its
309 next and preceding available observations;
310 B7) Lowess (locally weighted scatterplot smoothing) with a smoothing span of 1: missing
311 values replaced using fitted values from a lowess model determined using all available
312 observations (Cleveland, 1981);
313 B8) Lowess with a smoothing span of 0.75: same as B7 except that the smoothing span is 75%
314 of the available data (similar distinction follows for B9-B11);
315 B9) Lowess with a smoothing span of 50%;
316 B10) Lowess with a smoothing span of 30%; and
317 B11) Lowess with a smoothing span of 10%.

318 B4 and B5 were implemented using the “*na.locf*” function in the “*zoo*” R package (Zeileis and
319 Grothendieck, 2005). B7-B11 were implemented using the “*loess*” function in the “*stats*” R
320 package (R Development Core Team, 2014). An illustration of these interpolation methods is
321 provided in **Figure 4**. The interpolated data, along with the original regular data (designated as
322 A1) were analyzed using the Whittle’s maximum likelihood method for β estimation, which was
323 implemented using the “*FDWhittle*” function in the “*fractal*” R package (Constantine and
324 Percival, 2014).

325 The second type of approaches estimates β in the irregularly sampled data directly, using
326 several variants of the Lomb-Scargle periodogram (designated as C1a-C1c below), and a
327 recently developed wavelet-based method (designated as C2 below). Specifically, these
328 approaches are:

- 329 C1a) Lomb-Scargle periodogram: the spectral density of the time series (with gaps) is
330 estimated and the spectral slope is fit using all frequencies (Lomb, 1976; Scargle, 1982).



331 This is a classic method for examining periodicity in irregularly sampled data, which is
332 analogous to the more familiar fast Fourier transform method often used for regularly
333 sampled data;

334 C1b) Lomb-Scargle periodogram with 5% data: same as C1a except that the fitting of the
335 spectral slope considers only the lowest 5% frequencies (Montanari *et al.*, 1999);

336 C1c) Lomb-Scargle periodogram with “binned” data: same as C1a except that the fitting of
337 the spectral slope is performed on binned data in three steps: (1) The entire range of
338 frequency is divided into 100 equal-interval bins on logarithmic scale. (2) The
339 respective medians of frequency and power spectral density are calculated for each of
340 the 100 bins. (3) The 100 pairs of median frequency and median spectral density are
341 used to estimate the spectral slope on a log-log scale.

342 C2) Kirchner and Neal (2013)’s wavelet method: uses a modified version of Foster’s
343 weighted wavelet spectrum (Foster, 1996) to suppress spectral leakage from low
344 frequencies and applies an aliasing filter (Kirchner, 2005) to remove spectral aliasing
345 artifacts at high frequencies.

346 C1a was implemented using the “*spec.ls*” function in the “*cts*” R package (Wang, 2013). C2 was
347 run in *C*, using codes modified from those in Kirchner and Neal (2013).

348 **3.2. Evaluation of Methods’ Performance**

349 Each estimation method listed above was applied to the simulated data (**Section 2.3**) to
350 estimate β , which were then compared with the prescribed (“true”) β to quantify the performance
351 of each method. Plots of method evaluation for all simulations are provided as **Figures S1-S10**
352 in the Supporting Information. Close inspections of these plots reveal some general patterns of
353 the methods’ performance. For brevity, these patterns are presented with a subset of the plots,
354 which correspond to the cases where true $\beta = 1$ and shape parameter $\lambda = 0.01, 0.1, 1, \text{ and } 10$
355 (**Figure 5**). In general, β values estimated using the regular data (A1) are very close to 1.0,
356 which indicates that the adopted fractional noise generation method and the Whittle’s maximum
357 likelihood estimator have small combined simulation and estimation bias. This is perhaps
358 unsurprising, since the estimator is based on the Fourier transform and the noise generator is
359 based on an inverse Fourier transform; thus, one method is essentially just the inverse of the
360 other. One should also note that when fractional noises are not arbitrarily band-limited at the
361 Nyquist frequency (as they inherently are with the noise generator that is used here), spectral



362 aliasing should lead to spectral slopes that are flatter than expected (Kirchner, 2005), and thus to
363 underestimates of LRD.

364 For the simulated irregular data, the estimation methods differ widely in their performance.
365 Specifically, three interpolation methods (*i.e.*, B4-B6) consistently over-estimate β , indicating
366 that they introduce additional correlations into the time series, reducing its short-timescale
367 variability. In contrast, the other eight interpolation methods (*i.e.*, B1-B3 and B7-B11) generally
368 under-estimate β , indicating that the interpolated points are less correlated than the original time
369 series, thus introducing additional variability on short timescales. As expected, results from the
370 lowess methods (B7-B11) depend strongly on the size of smoothing window, that is, more
371 severe under-estimation of β is produced as the smoothing window becomes wider. In fact, when
372 the smoothing window is 1.0 (*i.e.*, method B7), lowess performs the interpolation using all data
373 available and thus behaves similarly to interpolations based on global means (B1) or global
374 medians (B2), except that lowess fits a polynomial curve instead of constant values. However,
375 whenever a sampling gap is much shorter than the smoothing window, the infilled lowess value
376 will be close to the local mean or median, and the abrupt jumps produced by these infilled values
377 will artificially increase the variance in the time series at high frequencies, leading to an
378 artificially reduced spectral slope β and correspondingly, an underestimate of β . This mechanism
379 explains why lowess interpolation distorts β more when there are many small gaps (large λ), and
380 therefore more jumps to, and away from, the infilled values, than when there are only a few large
381 gaps (small λ).

382 Among the direct methods (*i.e.*, C1a, C1b, C1c, and C2), the Lomb-Scargle method, with
383 original data (C1a) or binned data (C1c) tends to under-estimate β , though the underestimation
384 by C1c is generally less severe. The modified Lomb-Scargle method (C1b), using only the
385 lowest 5% of frequencies, yields estimates that are centered around 1.0. However, C1b has the
386 highest variability (*i.e.*, least precision) in β estimates among all methods. Compared with all the
387 above methods, the wavelet method (C2) has much better performance in terms of both accuracy
388 and precision when λ is 1 or 10, a slightly better performance when λ is 0.1, but a worse
389 performance when λ is 0.01.

390 The shape parameter λ greatly affects the performance of the estimation methods. All the
391 interpolation methods that under-estimate β (*i.e.*, B1-B3 and B7-B11) perform worse as λ
392 increases from 0.01 to 10. This effect can be interpreted as follows: when the time series



393 contains a large number of relatively small gaps (*e.g.*, $\lambda = 1$ or 10), there are many jumps (which,
 394 as noted above, contain mostly high-frequency variance) between the original data and the
 395 infilled values, resulting in more severe under-estimation. In contrast, when the data contain only
 396 a small number of very large gaps (*e.g.*, $\lambda = 0.01$ or 0.1), there are fewer of these jumps, resulting
 397 in minimal under-estimation. Similar effects of λ are also observed with the interpolation
 398 methods that show over-estimation (*i.e.*, B4-B6) – that is, over-estimation is more severe when λ
 399 is larger. Similarly, the Lomb-Scargle method (C1a and C1c) performs worse (more serious
 400 underestimation) as λ increases. Finally, method C2 seems to perform the best when λ is large (1
 401 or 10), but not well when λ is very small (0.01), as noted above. This result highlights the
 402 sensitivity of the wavelet method to the presence of a few large gaps in the time series. For such
 403 cases, a potentially more feasible approach is to break the whole time series into several
 404 segments (each without long gaps) and then apply the wavelet method (C2) to analyze each
 405 segment separately. If this can yield more accurate estimates, then further simulation
 406 experiments should be designed to systematically determine how long the gap needs to be to
 407 invoke such an approach.

408 Next, the method evaluation is extended to all the simulated spectral slopes, that is, $\beta = 0$,
 409 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, and 2. For ease of discussion, three quantitative criteria were
 410 proposed for evaluating performance, namely, bias (B), standard deviation (SD), and root-mean-
 411 squared error (RMSE), as defined below:

$$B_i = \bar{\beta}_i - \beta_{true} \quad (14)$$

$$SD_i = \sqrt{\frac{1}{29} \sum_{j=1}^{30} (\beta_{i,j} - \bar{\beta}_i)^2} \quad (15)$$

$$RMSE_i = \sqrt{B_i^2 + SD_i^2} \quad (16)$$

412 where $\bar{\beta}_i$ is the mean of 30 β values estimated by method *i*, and β_{true} is the prescribed β value for
 413 simulation of the initial regular time series. In general, B and SD can be considered as the
 414 models' systematic error and random error, respectively, and RMSE serves as an integrated
 415 measure of both errors. For all evaluations, plots of bias and RMSE are provided in the main text.
 416 (Plots of SD are provided as **Figure S5** and **Figure S10** for simulations with $\mu = 1$ and $\mu = 14$,
 417 respectively.)



418 For simulations with $\mu = 1$, results of estimation bias and RMSE are summarized in **Figure 6**
419 and **Figure 7**, respectively. (More details are provided in **Figures S1-S4**.) For brevity, we focus
420 on three direct methods (C1a, C1b and C2) and three representative interpolation methods.
421 (Specifically, B1 represents B1-B3 and B7; B6 represents B4-B6, and B8 represents B8-B11.)
422 Overall, these six methods show mixed performances. In terms of bias (**Figure 6**), B1 (global
423 mean) and B8 (lowess with a smoothing span of 0.75) tend to have negative bias, particularly for
424 time series with (1) moderate-to-large β_{true} values and (2) large λ values (*i.e.*, less skewed gap
425 intervals). By contrast, B1 and B8 generally have minimal bias when (1) β_{true} is close to zero (*i.e.*,
426 when the simulated time series is close to white noise); and (2) λ is small (*e.g.*, 0.01), since
427 interpolating a few large gaps cannot significantly affect the overall correlation structure. In
428 addition, lowess interpolation with a larger smoothing window tends to yield more negatively
429 biased estimates (data not shown). The other interpolation method, B6 (mean of the two nearest
430 neighbors) tends to over-estimate β , particularly for time series with (1) small β_{true} values and (2)
431 large λ values. At large β_{true} values (*e.g.*, 2.0), the auto-correlation is already very strong such
432 that taking the mean of two neighbors for gap filling does not introduce much additional
433 correlation, as opposed to the case of small β_{true} values. The Lomb-Scargle methods (C1a and
434 C1b) generally have negative bias, particularly for time series with (1) moderate-to-large β_{true}
435 values (for both methods) and (2) large λ values (for C1a), which is similar to B1 and B8.
436 However, C1b overall shows less severe bias than C1a. Finally, the wavelet method (C2) shows
437 generally the smallest bias among all methods. However, its performance advantage is not as
438 great when the time series has small λ values (*i.e.*, very skewed gap intervals), as noted above,
439 which may be due to the fact that the aliasing filter was designed for regular time series. In terms
440 of SD (**Figure S5**), method C1b performs the worst among all methods (as noted above), method
441 B6 and B8 perform poorly for large β_{true} values, and method C2 performs poorly for $\beta_{true} = 0$. In
442 terms of RMSE (**Figure 7**), methods B1, B8, C1a, and C1b perform well for small β_{true} values
443 and small λ values, whereas method B6 performs well for large β_{true} values and small λ values. In
444 comparison, method C2 has the smallest RMSEs among all methods, and its RMSEs are
445 similarly small for the wide range of β_{true} and λ values. In general, the wavelet method can be
446 considered the best among all methods.

447 For simulations with $\mu = 14$, results of estimation bias and RMSE are summarized in
448 **Figure 8** and **Figure 9**, respectively. (More details are provided in **Figures S6-S9**.) Overall,



449 these methods show mixed performances that are generally similar to the cases when $\mu = 1$, as
 450 discussed above. These results highlight the generality of these methods' performances, which
 451 applies at least to the range of $\mu = [1, 14]$. In addition, all methods show generally larger RMSE
 452 for $\mu = 14$ than $\mu = 1$, indicating their dependence on the mean gap interval (**Figure 9**). Perhaps
 453 the most notable difference is observed with method C2, which in this case shows positive bias
 454 for small λ values (0.01 and 0.1) and negative bias for large λ values (1 and 10) (**Figure 8f**). It
 455 nonetheless generally shows the smallest RMSEs among all the tested methods.

456 3.3. Quantification of Spectral Slopes in Real Water-Quality Data

457 In this section, the proposed estimation approaches were applied to quantify β in real water-
 458 quality data from the two monitoring programs presented in **Section 2.2 (Table 1)**. As noted in
 459 **Section 1.3**, such real data are typically much more complex than our simulated time series,
 460 because of (1) strong deviations from normal distributions and (2) effects of flow-dependence,
 461 seasonality, and temporal trend (Hirsch *et al.*, 1991; Helsel and Hirsch, 2002). In this regard,
 462 future research may simulate time series with these important characteristics and evaluate the
 463 performance of various estimation approaches, perhaps following the modeling framework
 464 described herein. Alternatively, one may quantify β in transformed time series after accounting
 465 for the above aspects. In this work, we have taken the latter approach for a preliminary
 466 investigation. Specifically, we have used the published Weighted Regressions on Time,
 467 Discharge, and Season (WRTDS) method (Hirsch *et al.*, 2010) to transform the original time
 468 series. This widely accepted method estimates daily concentrations based on discretely collected
 469 concentration samples using time, season, and discharge as explanatory variables, *i.e.*,

$$\ln(C) = \beta_0 + \beta_1 t + \beta_2 \ln(Q) + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t) + \varepsilon \quad (17)$$

470 where C is concentration, Q is daily discharge, t is time in decimal years, β_i are fitted
 471 coefficients, and ε is the error term. The 2nd and 3rd terms on the right represent time and
 472 discharge effects, respectively, whereas the 4th and 5th terms collectively represent cyclical
 473 seasonal effects. For a full description of this method, see Hirsch and De Cicco (2015). In this
 474 work, WRTDS was applied to obtain the time series of estimated daily concentration for each
 475 constituent at each site. The difference between observed concentration (C_{obs}) and estimated
 476 concentration (C_{est}) was calculated in logarithmic space to obtain the concentration residuals,

$$residuals = \ln(C_{obs}) - \ln(C_{est}) \quad (18)$$



477 For our data sets, histograms of concentration residuals (expressed in natural log concentration
478 units) are shown in **Figures S11-S14**. Compared with the original concentration data, these
479 model residuals are much more nearly normal and homoscedastic. Moreover, the model residuals
480 are less susceptible to the issues of temporal, seasonal, and discharge-drive variations than the
481 original concentrations. Therefore, the model residuals are more appropriate than the original
482 concentrations for β estimation using the simulation framework adopted in this work.

483 The estimated β values for the concentration residuals are summarized in **Figure 10**. Clearly,
484 the estimated β varies considerably with the estimation method. In addition, the estimated β
485 varies with site and constituent (*i.e.*, TP, TN, or NO_x .) Our discussion below focuses on the
486 wavelet method (C2), because it is established above that this method performs better than the
487 other estimation methods under a wide range of gap conditions. We emphasize that it is beyond
488 our current scope to precisely quantify β in these water-quality data sets, but our simulation
489 results presented above (**Section 3.2**) can be used as references to qualitatively evaluate the
490 reliability of C2 and/or other methods for these data sets.

491 For TN and TP concentration data at the Chesapeake River Input Monitoring sites (**Table 1**),
492 μ varies between 9.5 and 24.4, whereas λ is ~ 1.0 . Thus, the simulated gap scenario of NB($\mu = 14$,
493 $\lambda = 1$) can be used as a reasonable reference to assess methods' reliability (**Figure 8**). Based on
494 method C2, the estimated β ranges between $\beta = 0.36$ and $\beta = 0.61$ for TN and between $\beta = 0.30$
495 and $\beta = 0.58$ for TP at these sites (**Figure 10**). For such ranges, the simulation results indicate
496 that method C2 tends to moderately under-estimate β under this gap scenario (**Figure 8**), and
497 hence spectral slopes for TN and TP at these Chesapeake sites are likely slightly higher than
498 those presented above.

499 For NO_x and TP concentration data at the Lake Erie and Ohio sites (**Table 1**), μ varies
500 between 0.06 and 0.22, whereas λ is ~ 0.01 . Thus, the simulated gap scenario of NB($\mu = 1$, $\lambda =$
501 0.01) can be used as a reasonable reference to assess the methods' reliability (**Figure 6**). For
502 such small λ (*i.e.*, a few gaps that are very dissimilar from others), C2 is not reliable for β
503 estimation, as reflected by the generally positive bias in the simulation results. By contrast,
504 methods B1 (interpolation with global mean) and B8 (lowess with span 0.75) both perform quite
505 well under this gap scenario (**Figure 6**). These two methods provide almost identical β estimates
506 for each site-constituent combination, ranging from $\beta = 1.0$ to $\beta = 1.5$ for NO_x and from $\beta = 1.0$
507 to $\beta = 1.4$ for TP (**Figure 10**).



508 Overall, the above analysis of real water-quality data has illustrated the wide variability in β
509 estimates, with different choices of estimation methods yielding very different results. To our
510 knowledge, these water-quality data have not heretofore been analyzed in this context. As
511 illustrated above, our simulation experiments (**Section 3.2**) can be used as references to coarsely
512 evaluate the reliability of each method under specific gap scenarios, thereby considerably
513 narrowing the likely range of the estimated spectral slopes. Nonetheless, our results demonstrate
514 that the analyzed water-quality time series can exhibit strong fractal scaling, particularly at the
515 Lake Erie and Ohio tributary sites. Thus, an important implication is that researchers and
516 analysts should be cautious when applying standard statistical methods to identify temporal
517 trends in such water-quality data sets (Kirchner and Neal, 2013). In future work, one may
518 consider applying Bayesian statistical analysis or other approaches to more accurately quantify
519 the spectral slope and associated uncertainty for real water-quality data analysis. In addition, the
520 modeling framework presented herein (including both gap simulation and β estimation) may be
521 extended to simulations of irregular time series that have prescribed spectral slopes and also
522 superimposed temporal trends, which can then be used to evaluate the validity of various
523 statistical methods for identifying trend and associated statistical significance.

524 **4. Conclusions**

525 River water-quality time series often exhibit fractal scaling behavior, which presents
526 challenges to the identification of deterministic trends. Because traditional estimation methods
527 are generally not applicable to irregularly sampled time series, we have examined two broad
528 types of estimation approaches and evaluated their performances against synthetic data with a
529 wide range of prescribed β values and gap intervals representative of the sampling irregularity of
530 real water-quality data.

531 The results of this work suggest several important messages. First, the results remind us of
532 the risks in using interpolation for gap filling when examining auto-correlation, as the
533 interpolation methods consistently under-estimate or over-estimate β under a wide range of
534 prescribed β values and gap distributions. Second, the long-established Lomb-Scargle spectral
535 method also consistently under-estimates β . Its modified form, using the 5% lowest frequencies
536 for spectral slope estimation, has very poor precision, although the overall bias is small. Third,
537 the wavelet method, coupled with an aliasing filter, has the smallest bias and root-mean-squared



538 error among all methods for a wide range of prescribed β values and gap distributions, except for
539 cases with small prescribed β values (*i.e.*, close to white noise) or small λ values (*i.e.*, very
540 skewed gap distributions). Thus, the wavelet method is recommended for estimating spectral
541 slope in irregular time series until improved methods are developed. In this regard, future
542 research should aim to develop an aliasing filter that is more applicable to irregular time series
543 with very skewed gap intervals. Finally, all methods' performances depend strongly on the
544 sampling irregularity in terms of both the skewness and mean of gap-interval lengths,
545 highlighting that the accuracy and precision of each method are data-specific.

546 Overall, these results provide new contributions in terms of better understanding and
547 quantification of the proposed methods' performances for estimating the strength of fractal
548 scaling in irregularly sampled water-quality data. In addition, the work has provided an
549 innovative and general approach for modeling sampling irregularity in water-quality records.
550 Moreover, this work has proposed and demonstrated a generalizable framework for data
551 simulation (with gaps) and β estimation, which can be readily applied toward the evaluation of
552 other methods that are not covered in this work. More generally, the findings and approaches
553 may also be broadly applicable to irregularly sampled data in other scientific disciplines. Last but
554 not least, we note that accurate quantification of fractal scaling in irregular water-quality time
555 series remains an unresolved challenge for the hydrologic community and for many other
556 disciplines that must grapple with irregular sampling.

557 **Data Availability**

558 River monitoring data used in this study are available through the U.S. Geological Survey
559 National Water Information System (<http://doi.org/10.5066/F7P55KJN>) and the Heidelberg
560 University's National Center for Water Quality Research.

561 **Supporting Information**

562 Supporting information to this article is available online.

563 **Competing Interests**

564 The authors declare that they have no conflict of interest.



565 **Acknowledgements**

566 Q. Zhang was supported by the Maryland Sea Grant through awards NA10OAR4170072 and
567 NA14OAR1470090 and by the Maryland Water Resources Research Center through a graduate
568 fellowship while he was a doctoral student at the Johns Hopkins University. C. Harman's contribution to
569 this work was supported by the National Science Foundation through grants CBET-1360415 and EAR-
570 1344664. We thank Bill Ball (Johns Hopkins University) and Bob Hirsch (U.S. Geological Survey) for
571 many useful discussions.

572 **References**

- 573 Aubert, A. H., J. W. Kirchner, C. Gascuel-Oudou, M. Faucheux, G. Gruau and P. Mérot, 2014. Fractal
574 water quality fluctuations spanning the periodic table in an intensively farmed watershed.
575 *Environ. Sci. Technol.* 48:930-937, DOI: 10.1021/es403723r.
- 576 Beran, J., 2010. Long-range dependence. *Wiley Interdisciplinary Reviews: Computational Statistics* 2:26-
577 35, DOI: 10.1002/wics.52.
- 578 Beran, J., Y. Feng, S. Ghosh and R. Kulik, 2013. *Long-Memory Processes: Probabilistic Properties and*
579 *Statistical Methods*. Berlin, Heidelberg, Springer Berlin Heidelberg, ISBN 978-3-642-35511-0
- 580 Boutahar, M., V. Marimoutou and L. Nourira, 2007. Estimation Methods of the Long Memory Parameter:
581 Monte Carlo Analysis and Application. *Journal of Applied Statistics* 34:261-301, DOI:
582 10.1080/02664760601004874.
- 583 Box, G. E. P., G. M. Jenkins and G. C. Reinsel, 2008. *Time Series Analysis, Fourth Edition*. Hoboken, NJ,
584 John Wiley & Sons, Inc., ISBN 9781118619193
- 585 Clarke, R. T., 2013. Calculating uncertainty in regional estimates of trend in streamflow with both serial
586 and spatial correlations. *Water Resour. Res.* 49:7120-7125, DOI: 10.1002/wrcr.20465.
- 587 Cleveland, W. S., 1981. LOWESS: A program for smoothing scatterplots by robust locally weighted
588 regression. *American Statistician* 35:54, DOI: 10.2307/2683591.
- 589 Cohn, T. A. and H. F. Lins, 2005. Nature's style: Naturally trendy. *Geophys. Res. Lett.* 32:L23402, DOI:
590 10.1029/2005GL024476.
- 591 Constantine, W. and D. Percival, 2014. *fractal: Fractal Time Series Modeling and Analysis*,
- 592 Darken, P. F., C. E. Zipper, G. I. Holtzman and E. P. Smith, 2002. Serial correlation in water quality
593 variables: Estimation and implications for trend analysis. *Water Resour. Res.* 38:1117, DOI:
594 10.1029/2001WR001065.
- 595 Delignette-Muller, M. L. and C. Dutang, 2015. fitdistrplus: An R Package for Fitting Distributions.
596 *Journal of Statistical Software* 64:1-34, DOI.



- 597 Ehsanzadeh, E. and K. Adamowski, 2010. Trends in timing of low stream flows in Canada: impact of
598 autocorrelation and long-term persistence. *Hydrol. Process.* 24:970-980, DOI: 10.1002/hyp.7533.
- 599 Faticchi, S., S. M. Barbosa, E. Caporali and M. E. Silva, 2009. Deterministic versus stochastic trends:
600 Detection and challenges. *Journal of Geophysical Research* 114:D18121, DOI:
601 10.1029/2009JD011960.
- 602 Foster, G., 1996. Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal*
603 112:1709-1729, DOI, <http://articles.adsabs.harvard.edu/full/1996AJ....112.1709F>.
- 604 Franzke, C., 2012a. Nonlinear Trends, Long-Range Dependence, and Climate Noise Properties of Surface
605 Temperature. *J. Clim.* 25:4172-4183, DOI: 10.1175/JCLI-D-11-00293.1.
- 606 Franzke, C., 2012b. On the statistical significance of surface air temperature trends in the Eurasian Arctic
607 region. *Geophys. Res. Lett.* 39:L23705, DOI: 10.1029/2012GL054244.
- 608 Godsey, S. E., W. Aas, T. A. Clair, H. A. de Wit, I. J. Fernandez, J. S. Kahl, I. A. Malcolm, C. Neal, M.
609 Neal, S. J. Nelson, S. A. Norton, M. C. Palucis, B. L. Skjelkvåle, C. Soulsby, D. Tetzlaff and J. W.
610 Kirchner, 2010. Generality of fractal 1/f scaling in catchment tracer time series, and its
611 implications for catchment travel time distributions. *Hydrol. Process.* 24:1660-1671, DOI:
612 10.1002/hyp.7677.
- 613 Graham, J., 2009. Missing Data Analysis: Making It Work in the Real World. *Annu. Rev. Psychol.* 60:549-
614 576, DOI: 10.1146/annurev.psych.58.110405.085530.
- 615 Helsel, D. R. and R. M. Hirsch, 2002. Statistical Methods in Water Resources. *U.S. Geological Survey*
616 *Techniques of Water-Resources Investigations Book 4, Chapter A3*. U.S. Geological Survey,
617 Reston, VA, p. 522. <http://pubs.usgs.gov/twri/twri4a3/>.
- 618 Hirsch, R. M., R. B. Alexander and R. A. Smith, 1991. Selection of methods for the detection and
619 estimation of trends in water quality. *Water Resour. Res.* 27:803-813, DOI: 10.1029/91WR00259.
- 620 Hirsch, R. M. and L. De Cicco, 2015. User guide to Exploration and Graphics for RivEr Trends (EGRET)
621 and dataRetrieval: R packages for hydrologic data (version 2.0, February 2015). U.S. Geological
622 Survey Techniques and Methods Book 4, Chapter A10, Reston, VA, p. 93.
623 <http://dx.doi.org/10.3133/tm4A10>.
- 624 Hurst, H. E., 1951. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* 116:770-808,
625 DOI.
- 626 Khaliq, M. N., T. B. M. J. Ouarda and P. Gachon, 2009. Identification of temporal trends in annual and
627 seasonal low flows occurring in Canadian rivers: The effect of short- and long-term persistence.
628 *Journal of Hydrology* 369:183-197, DOI: 10.1016/j.jhydrol.2009.02.045.
- 629 Khaliq, M. N., T. B. M. J. Ouarda, P. Gachon and L. Sushama, 2008. Temporal evolution of low-flow
630 regimes in Canadian rivers. *Water Resour. Res.* 44:W08436, DOI: 10.1029/2007WR006132.



- 631 Kirchner, J., 2005. Aliasing in $1/f^\alpha$ noise spectra: Origins, consequences, and remedies. *Physical*
632 *Review E* 71:066110-066110, DOI: 10.1103/PhysRevE.71.066110.
- 633 Kirchner, J. W., X. Feng and C. Neal, 2000. Fractal stream chemistry and its implications for contaminant
634 transport in catchments. *Nature* 403:524-527, DOI: 10.1038/35000537.
- 635 Kirchner, J. W., X. Feng and C. Neal, 2001. Catchment-scale advection and dispersion as a mechanism
636 for fractal scaling in stream tracer concentrations. *Journal of Hydrology* 254:82-101, DOI:
637 10.1016/S0022-1694(01)00487-5.
- 638 Kirchner, J. W. and C. Neal, 2013. Universal fractal scaling in stream chemistry and its implications for
639 solute transport and water quality trend detection. *Proc. Natl. Acad. Sci. U. S. A.* 110:12213-
640 12218, DOI: 10.1073/pnas.1304328110.
- 641 Kirchner, J. W. and A. Weil, 1998. No fractals in fossil extinction statistics. *Nature* 395:337-338, DOI:
642 10.1038/26384.
- 643 Langland, M. J., J. D. Blomquist, D. L. Moyer and K. E. Hyer, 2012. Nutrient and suspended-sediment
644 trends, loads, and yields and development of an indicator of streamwater quality at nontidal sites
645 in the Chesapeake Bay watershed, 1985-2010. U.S. Geological Survey Scientific Investigations
646 Report 2012-5093, Reston, VA, p. 26. <http://pubs.usgs.gov/sir/2012/5093/pdf/sir2012-5093.pdf>.
- 647 Lennartz, S. and A. Bunde, 2009. Trend evaluation in records with long-term memory: Application to
648 global warming. *Geophys. Res. Lett.* 36:L16706, DOI: 10.1029/2009GL039516.
- 649 Lomb, N. R., 1976. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space*
650 *Science* 39:447-462, DOI: 10.1007/BF00648343.
- 651 Montanari, A., R. Rosso and M. S. Taqqu, 1997. Fractionally differenced ARIMA models applied to
652 hydrologic time series: Identification, estimation, and simulation. *Water Resour. Res.* 33:1035-
653 1044, DOI: 10.1029/97WR00043.
- 654 Montanari, A., R. Rosso and M. S. Taqqu, 2000. A seasonal fractional ARIMA Model applied to the Nile
655 River monthly flows at Aswan. *Water Resour. Res.* 36:1249-1259, DOI:
656 10.1029/2000WR900012.
- 657 Montanari, A., M. S. Taqqu and V. Teverovsky, 1999. Estimating long-range dependence in the presence
658 of periodicity: An empirical study. *Mathematical and Computer Modelling* 29:217-228, DOI:
659 10.1016/S0895-7177(99)00104-1.
- 660 National Center for Water Quality Research, 2015. Tributary Data Download. Accessed July 23, 2015,
661 [https://www.heidelberg.edu/academics/research-and-centers/national-center-for-water-quality-](https://www.heidelberg.edu/academics/research-and-centers/national-center-for-water-quality-research/tributary-data-download)
662 [research/tributary-data-download](https://www.heidelberg.edu/academics/research-and-centers/national-center-for-water-quality-research/tributary-data-download).
- 663 Noguchi, K., Y. R. Gel and C. R. Duguay, 2011. Bootstrap-based tests for trends in hydrological time
664 series, with application to ice phenology data. *Journal of Hydrology* 410:150-161, DOI:



- 665 10.1016/j.jhydrol.2011.09.008.
- 666 R Development Core Team, 2014. R: A language and environment for statistical computing. R Foundation
667 for Statistical Computing, Vienna, Austria. ISBN 3900051070. <http://www.r-project.org>.
- 668 Rea, W., L. Oxley, M. Reale and J. Brown, 2009. Estimators for Long Range Dependence: An Empirical
669 Study. *Electronic Journal of Statistics*, <http://arxiv.org/abs/0901.0762>, DOI.
- 670 Sang, Y.-F., Z. Wang and C. Liu, 2014. Comparison of the MK test and EMD method for trend
671 identification in hydrological time series. *Journal of Hydrology* 510:293-298, DOI:
672 10.1016/j.jhydrol.2013.12.039.
- 673 Scargle, J. D., 1982. Studies in Astronomical Time-Series Analysis. II. Statistical Aspects of Spectral-
674 Analysis of Unevenly Spaced Data. *Astrophysical Journal* 263:835-853, DOI: 10.1086/160554.
- 675 Stroe-Kunold, E., T. Stadnytska, J. Werner and S. Braun, 2009. Estimating long-range dependence in time
676 series: an evaluation of estimators implemented in R. *Behav. Res. Methods* 41:909-923, DOI:
677 10.3758/BRM.41.3.909.
- 678 Taqqu, M. S., V. Teverovsky and W. Willinger, 1995. Estimators for long-range dependence: an empirical
679 study. *Fractals*, <http://www.worldscientific.com/doi/abs/10.1142/S0218348X95000692>, DOI.
- 680 Wang, Z., 2013. cts: An R Package for Continuous Time Autoregressive Models via Kalman Filter.
681 *Journal of Statistical Software* 53:1-19, DOI.
- 682 Witt, A. and B. D. Malamud, 2013. Quantification of Long-Range Persistence in Geophysical Time
683 Series: Conventional and Benchmark-Based Improvement Techniques. *Surv. Geophys.* 34:541-
684 651, DOI: 10.1007/s10712-012-9217-8.
- 685 Yue, S., P. Pilon, B. Phinney and G. Cavadias, 2002. The influence of autocorrelation on the ability to
686 detect trend in hydrological series. *Hydrol. Process.* 16:1807-1829, DOI: 10.1002/hyp.1095.
- 687 Zeileis, A. and G. Grothendieck, 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series.
688 *Journal of Statistical Software* 14:1-27, DOI.
- 689 Zetterqvist, L., 1991. Statistical Estimation and Interpretation of Trends in Water Quality Time Series.
690 *Water Resour. Res.* 27:1637-1648, DOI: 10.1029/91wr00478.
- 691 Zhang, Q. and W. P. Ball, 2017. Improving Riverine Constituent Concentration and Flux Estimation by
692 Accounting for Antecedent Discharge Conditions. *Journal of Hydrology* 547:387-402, DOI:
693 10.1016/j.jhydrol.2016.12.052.
- 694 Zhang, Q., D. C. Brady, W. R. Boynton and W. P. Ball, 2015. Long-Term Trends of Nutrients and
695 Sediment from the Nontidal Chesapeake Watershed: An Assessment of Progress by River and
696 Season. *JAWRA Journal of the American Water Resources Association* 51:1534-1555, DOI:
697 10.1111/1752-1688.12327.

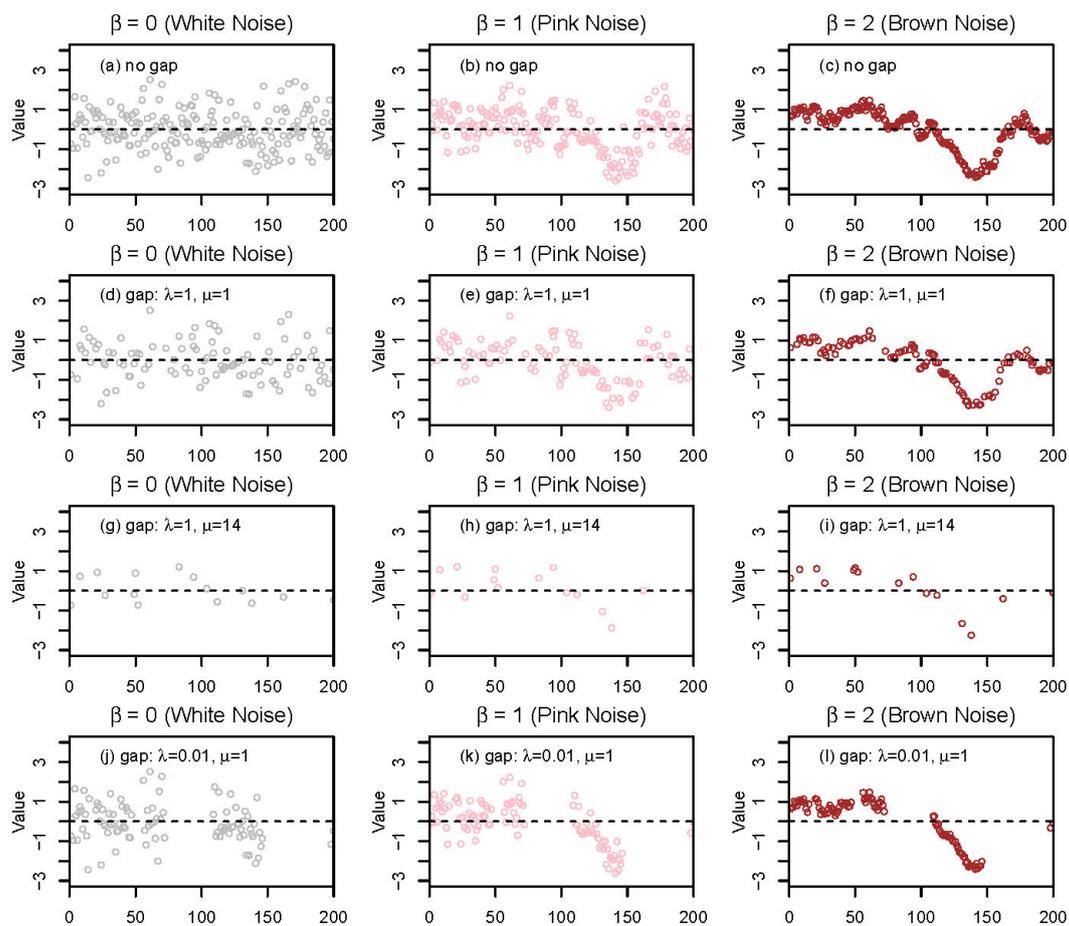


Table 1. Quantification of sampling irregularity for selected water-quality constituents at nine sites of the Chesapeake Bay River Input Monitoring program and six sites of the Lake Erie and Ohio tributary monitoring program. (μ : mean parameter; λ : shape parameter estimated using maximum likelihood; λ' : shape parameter estimated using the direct approach (see **Section 2.2**). $\Delta t_{average}$: average gap interval; N : total number of samples.)

Site ID	River and station name	Drainage area (mi ²)	Total nitrogen (TN)				Total phosphorus (TP)					
			λ	λ'	μ	$\Delta t_{average}$ (days)	N	λ	λ'	μ	$\Delta t_{average}$ (days)	N
01578310	Susquehanna River at Conowingo, MD	27100	0.8	1.1	13.5	14.5	876	0.8	1.0	13.4	14.4	881
01646580	Potomac River at Chain Bridge, Washington D.C.	11600	0.9	0.6	9.5	10.5	1385	1.1	1.0	24.4	25.4	579
02035000	James River at Cartersville, VA	6260	0.8	1.0	13.9	14.9	960	0.8	1.1	13.7	14.7	974
01668000	Rappahannock River near Fredericksburg, VA	1600	0.8	0.6	15.6	16.6	776	0.8	0.6	15.2	16.2	796
02041650	Appomattox River at Matoaca, VA	1340	0.8	0.8	15.1	16.1	798	0.8	0.8	14.9	15.9	810
01673000	Pamunkey River near Hanover, VA	1071	0.8	0.9	15.1	16.1	873	0.8	1.0	14.7	15.7	894
01674500	Mattaponi River near Beulahville, VA	601	0.7	0.9	14.3	15.3	810	0.8	0.9	14.2	15.2	820
01594440	Patuxent River at Bowie, MD	348	0.9	1.1	15.3	16.3	787	0.8	0.8	14.0	15.0	861
01491000	Choptank River near Greensboro, MD	113	1.2	1.5	19.6	20.6	680	1.1	1.0	20.5	21.5	690

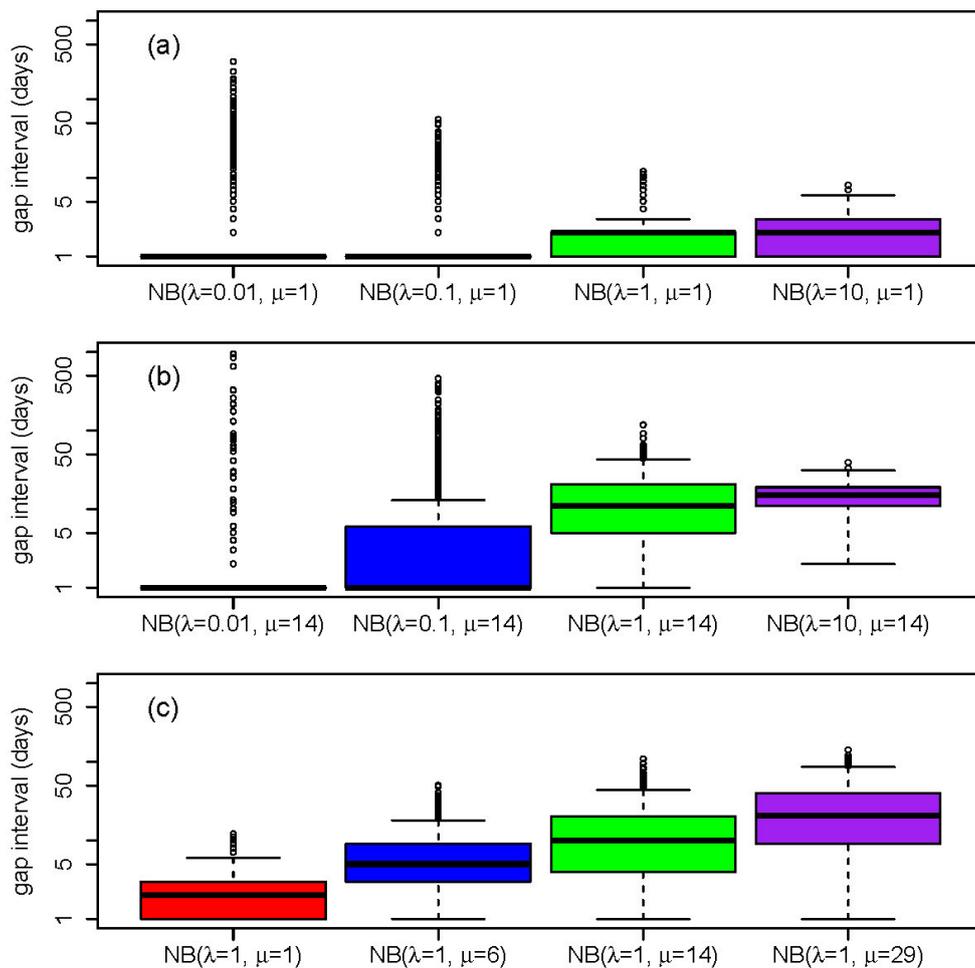
Site ID	River and station name	Drainage area (mi ²)	Nitrate-plus-nitrite (NO _x)				Total phosphorus (TP)					
			λ	λ'	μ	$\Delta t_{average}$ (days)	N	λ	λ'	μ	$\Delta t_{average}$ (days)	N
04193500	Maumee River at Waterville, OH	6330	0.005	0.0003	0.19	1.19	9101	0.005	0.0003	0.19	1.19	9101
04198000	Sandusky River near Fremont, OH	1253	0.01	0.003	0.22	1.22	9641	0.01	0.003	0.22	1.22	9655
04208000	Cuyahoga River at Independence, OH	708	0.007	0.006	0.13	1.13	7421	0.007	0.006	0.13	1.13	7426
04212100	Grand River near Painesville, OH	686	0.01	0.005	0.21	1.21	5023	0.01	0.005	0.22	1.22	4994
04197100	Honey Creek at Melmore, OH	149	0.007	0.005	0.06	1.06	9914	0.007	0.005	0.06	1.06	9914
04197170	Rock Creek at Tiffin, OH	34.6	0.007	0.008	0.06	1.06	8422	0.007	0.008	0.06	1.06	8440

II. Lake Erie and Ohio tributary monitoring program



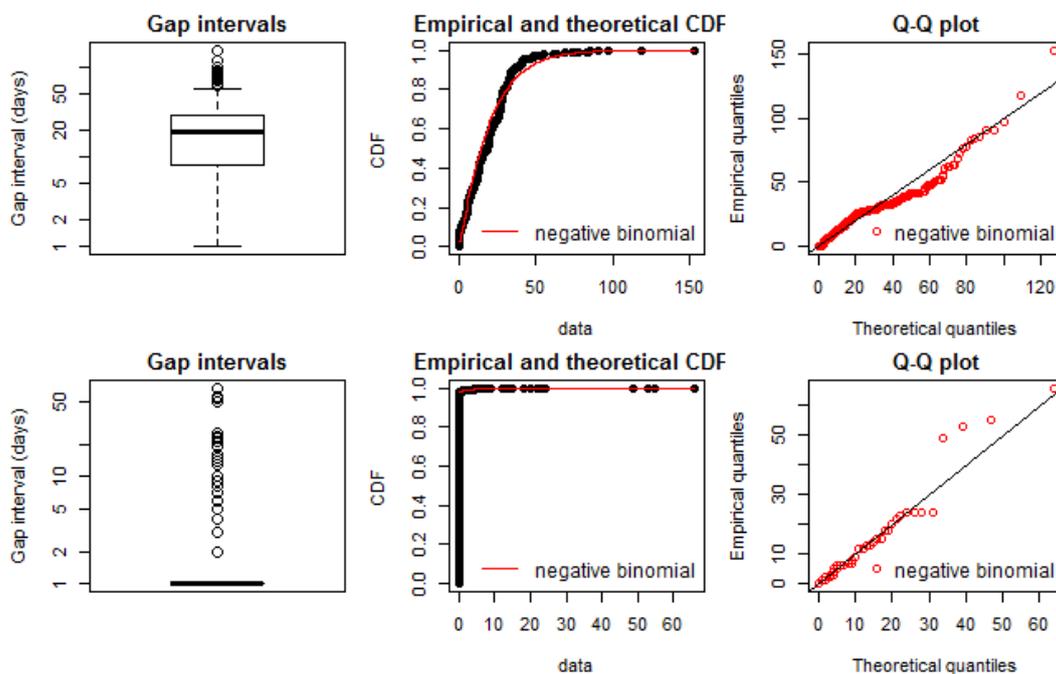
698

699 **Figure 1.** Synthetic time series with 200 time steps for three representative fractal scaling
 700 processes that correspond to white noise ($\beta = 0$), pink noise ($\beta = 1$), and Brown noise ($\beta = 2$).
 701 The 1st row shows the simulated time series without any gap. The three rows below show the
 702 same time series as in the 1st row but with data gaps that were simulated using three different
 703 negative binomial (NB) distributions, that is, 2nd row: NB($\lambda = 1, \mu = 1$); 3rd row: NB($\lambda = 1, \mu =$
 704 14); 4th row: NB($\lambda = 0.01, \mu = 1$).

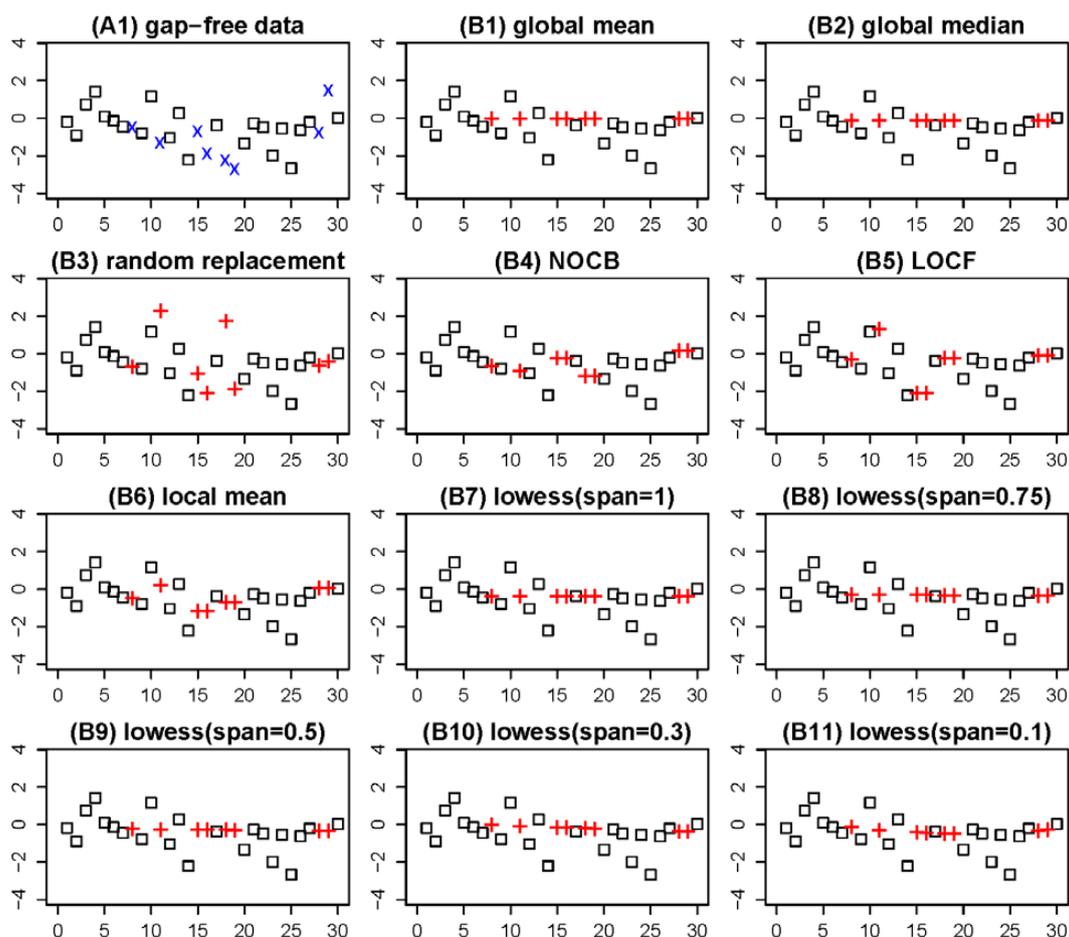


705

706 **Figure 2.** Examples of gap interval simulation using binomial distributions, NB (shape λ , mean
 707 μ). Simulation parameters: $L = 9125$ days, $\Delta t_{nominal} = 1$ day. The three panels show simulation
 708 with fixed (a) $\mu = 1$, (b) $\mu = 14$, and (c) $\lambda = 1$. Note that $\Delta t_{average}/\Delta t_{nominal} = \mu + 1$.

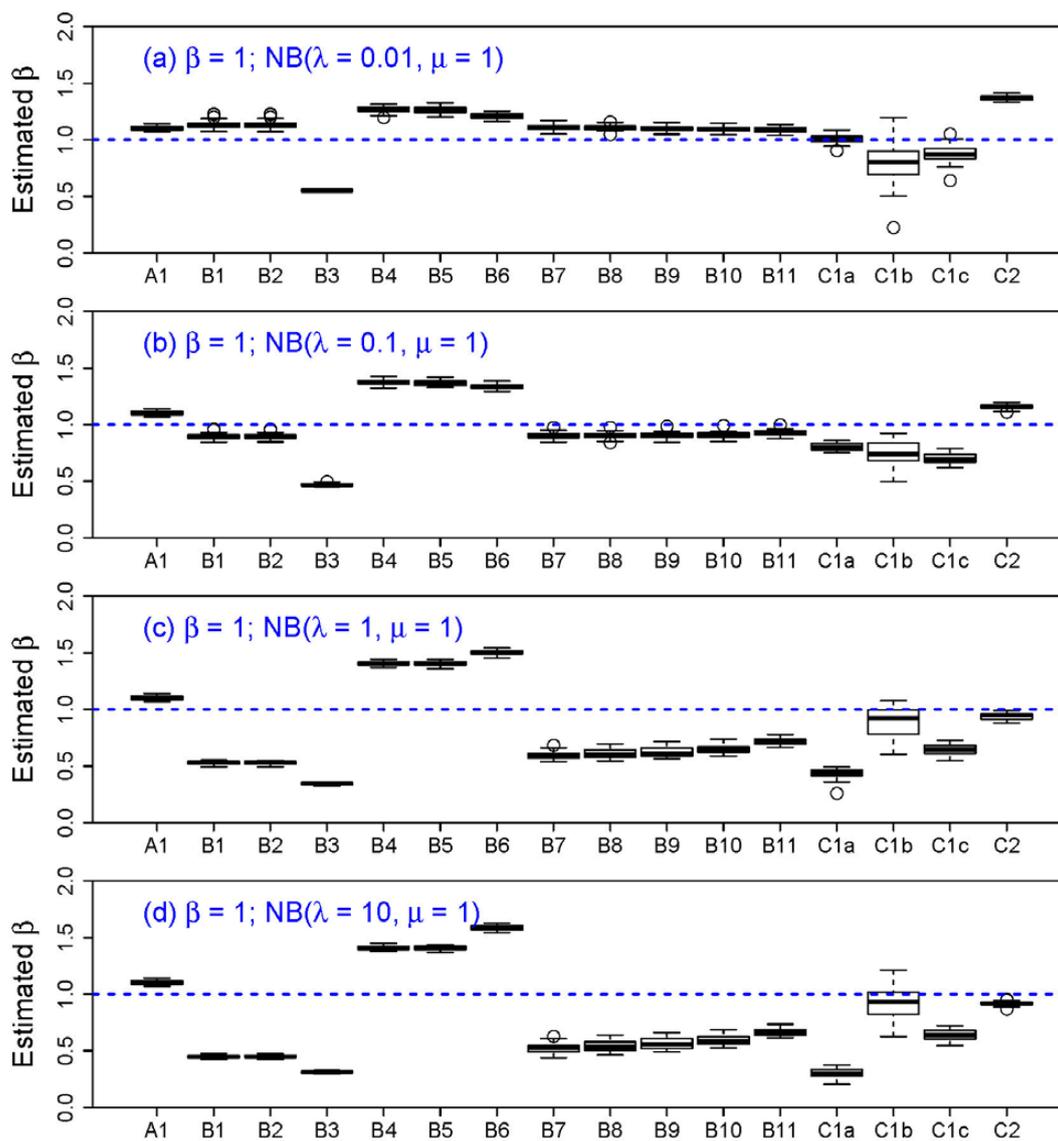


711 **Figure 3.** Examples of quantified sampling irregularity with negative binomial (NB)
712 distributions: total nitrogen in Choptank River (top) and total phosphorus in Cuyahoga River
713 (bottom). Theoretical CDF and quantiles are based on the fitted NB distributions. See **Table 1**
714 for estimated mean and shape parameters.



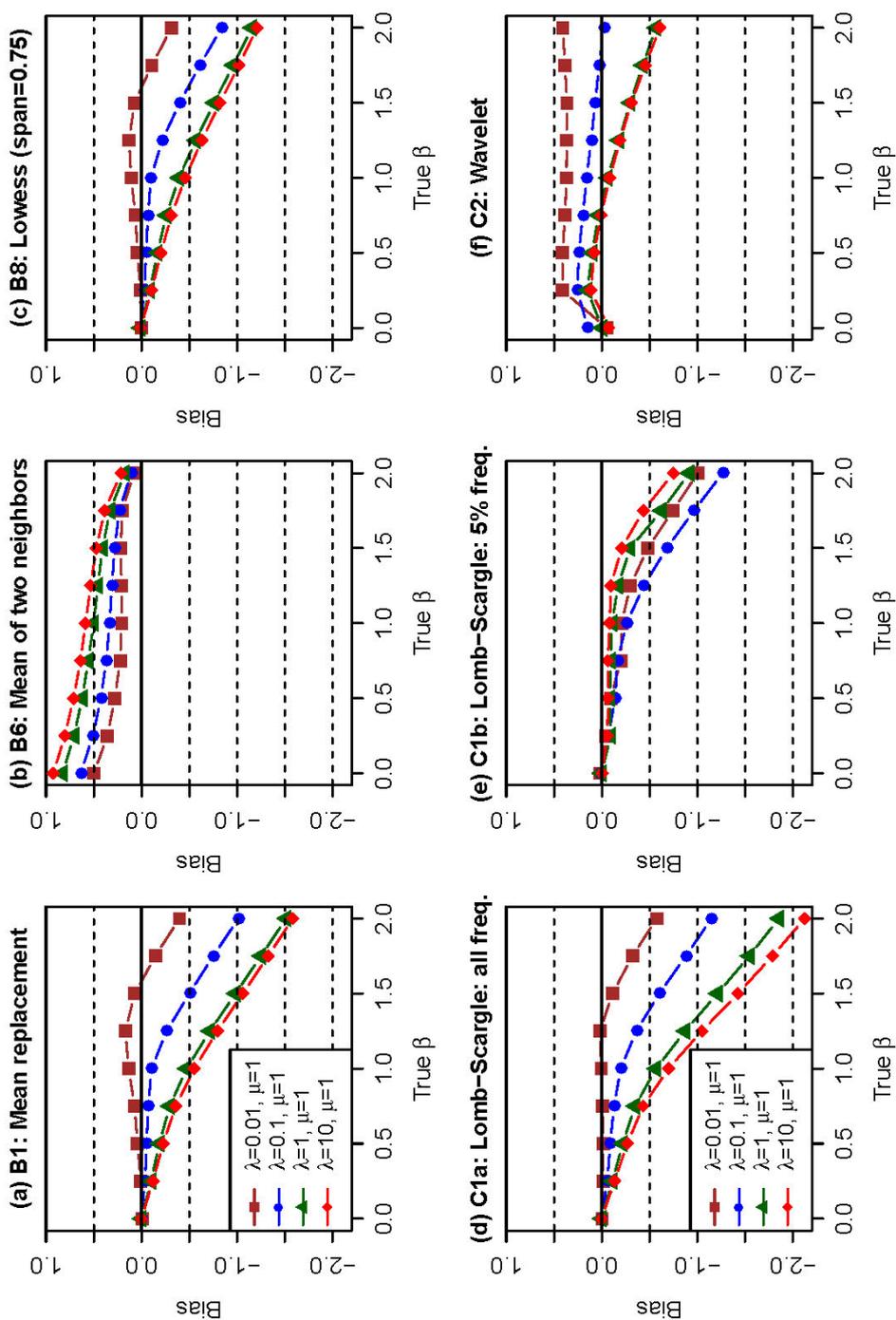
715

716 **Figure 4.** Illustration of the interpolation methods for gap filling. The gap-free data (A1) was
717 simulated with a series length of 500, with the first 30 data shown. (x: omitted data for gap filling;
718 +: interpolated data; NOCB: next observation carried backward; LOCF: last observation carried
719 forward; lowess: locally weighted scatterplot smoothing.)

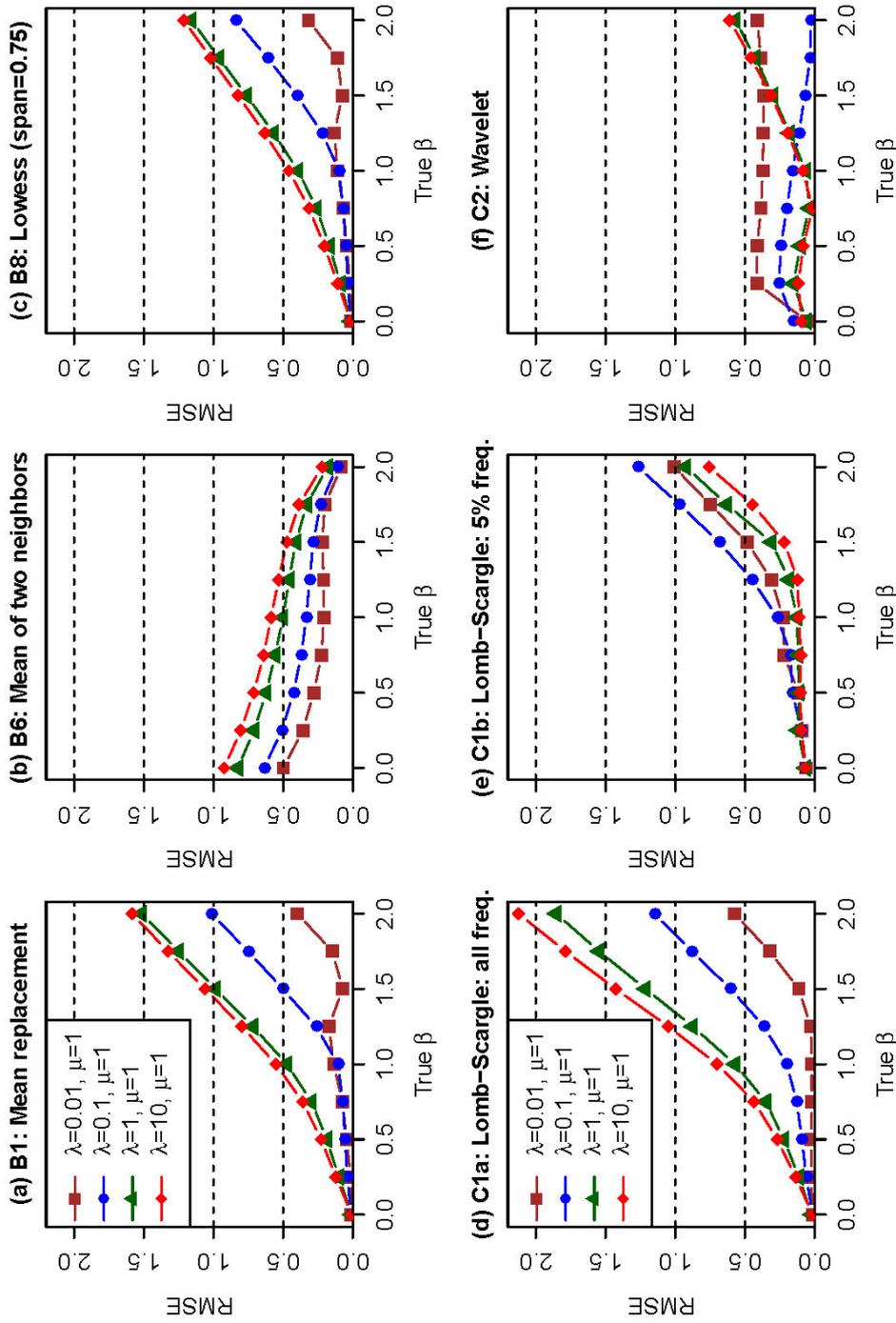


720

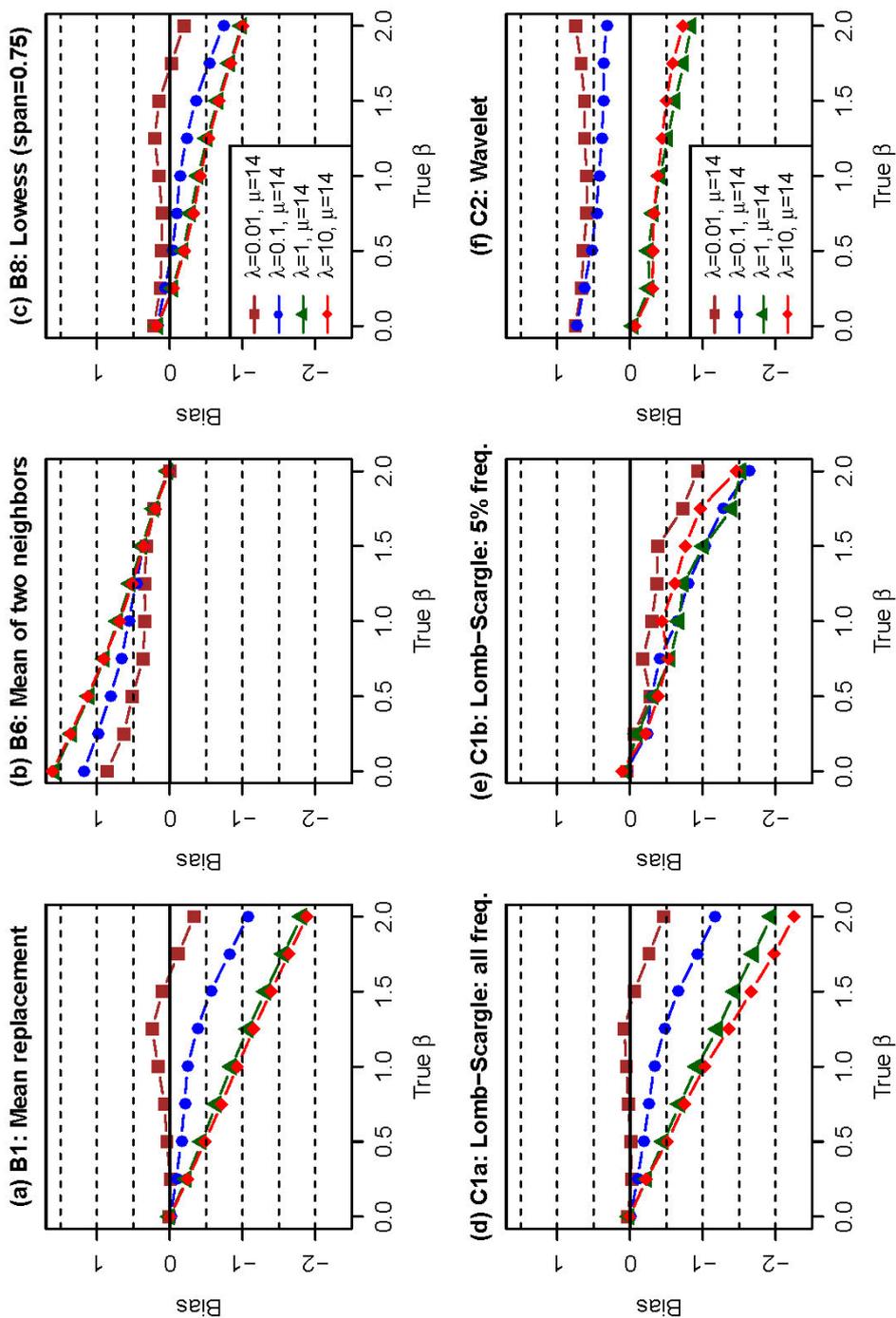
721 **Figure 5.** Comparison of bias in estimated spectral slope in irregular data that are simulated with
 722 prescribed $\beta = 1$ (30 replicates), series length of 9125, and gap intervals simulated with (a) NB ($\lambda = 0.01, \mu = 1$),
 723 (b) NB ($\lambda = 0.1, \mu = 1$), (c) NB ($\lambda = 1, \mu = 1$), and (d) NB ($\lambda = 10, \mu = 1$). The blue
 724 dashed lines indicate the true β value.



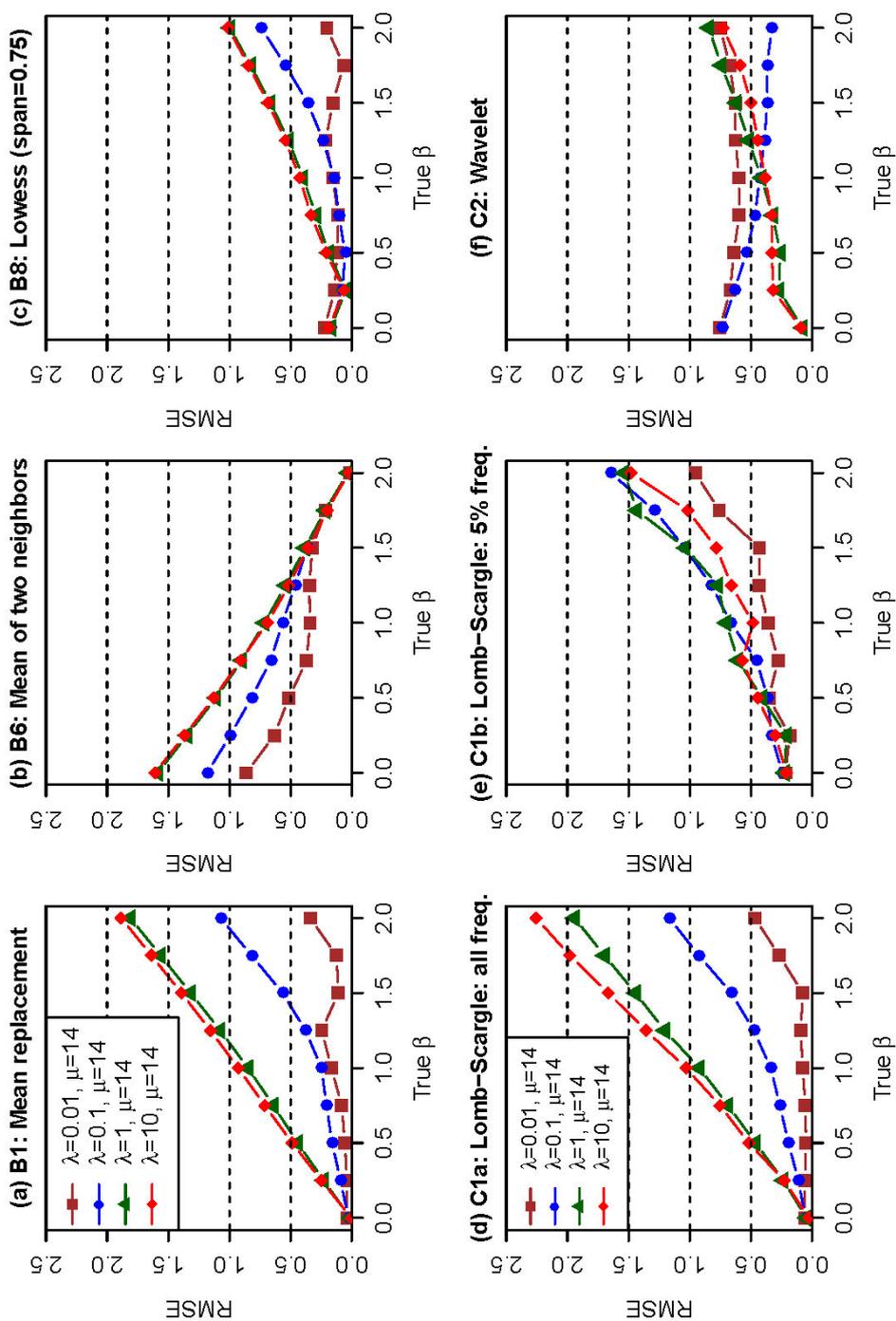
725
 726 **Figure 6.** Comparison of bias in estimated spectral slope in irregular data that are simulated with varying prescribed β values (30
 727 replicates), series length of 9125, and mean gap interval of 2 (*i.e.*, $\mu = 1$).



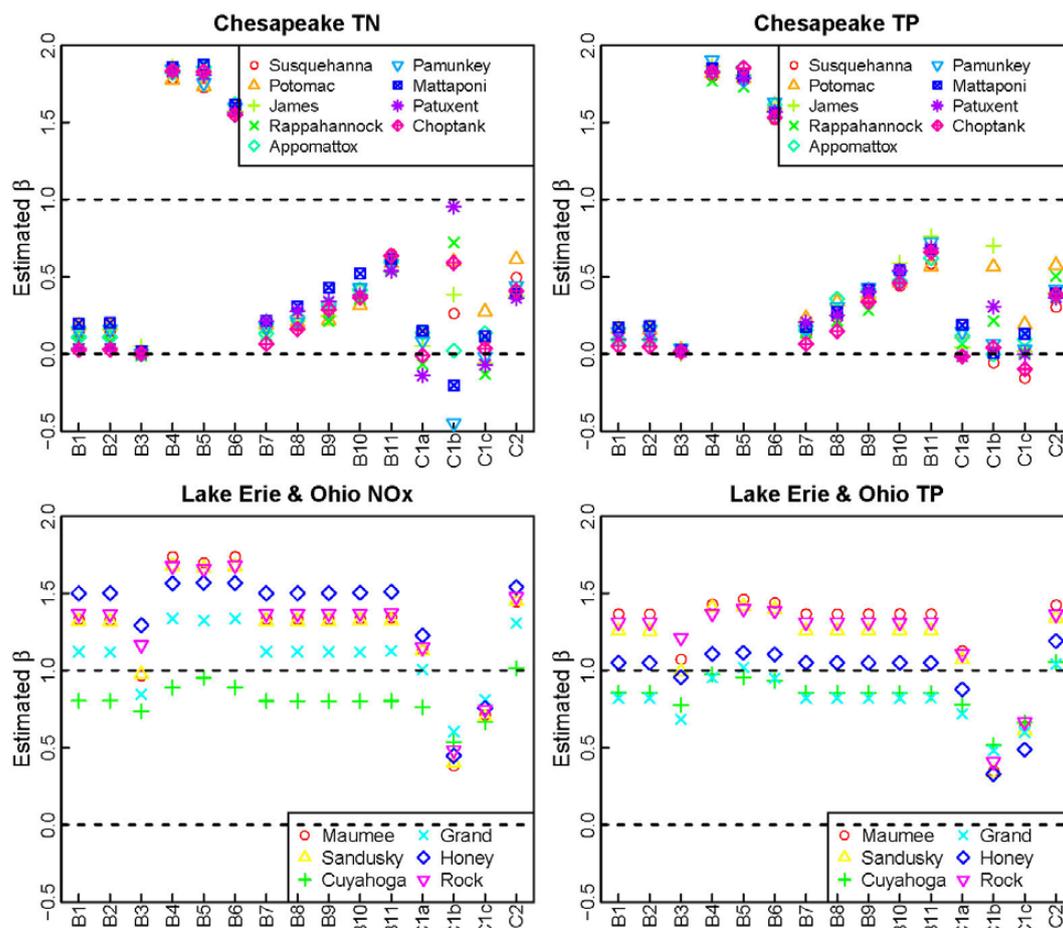
728
 729 **Figure 7.** Comparison of root-mean-squared error (RMSE) in estimated spectral slope in irregular data that are simulated with varying
 730 prescribed β values (30 replicates), series length of 9125, and mean gap interval of 2 (*i.e.*, $\mu = 1$).



731
 732 **Figure 8.** Comparison of bias in estimated spectral slope in irregular data that are simulated with varying prescribed β values (30
 733 replicates), series length of 9125, and mean gap interval of 15 (*i.e.*, $\mu = 14$).



734
 735 **Figure 9.** Comparison of root-mean-squared error (RMSE) in estimated spectral slope in irregular data that are simulated with varying
 736 prescribed β values (30 replicates), series length of 9125, and mean gap interval of 15 (*i.e.*, $\mu = 14$).



737

738 **Figure 10.** Quantification of spectral slope in real water-quality data from the two regional
 739 monitoring networks, as estimated using the set of examined methods. All estimations were
 740 performed on concentration residuals (in natural log concentration units) after accounting for
 741 effects of time, discharge, and season. The two dashed lines in each panel indicate white noise (β
 742 = 0) and pink (flicker) noise ($\beta = 1$), respectively. See **Table 1** for site and data details.