**Supporting Information**

**Accompanying Manuscript:** A multi-scale approach to timescale analysis: Isolating diel signals from solute concentration time series

**Authors:**

Cathy Chamberlin
Duke University
9 Circuit Drive, Durham, NC 27709
catherine.chamberlin@duke.edu
(now at chamberlin.catherine@epa.gov)

Gaby Katul
Duke University
9 Circuit Drive, Durham, NC 27709
gaby@duke.edu

Jim Heffernan
Duke University
9 Circuit Drive, Durham, NC 27709
james.heffernan@duke.edu

**Contents:**

**Video Captions:**

*Video S1:* A demonstration of the IMFs derived from the Ichetucknee year-long timeseries through EEMD. The timeseries is shown in black, data gaps are highlighted in grey, and successive IMFs are displayed overlain on each other in blue, beginning from the highest frequency IMF and ending with the overall trend in the data.

*Video S2:* A demonstration of the IMFs derived from the Potomac year-long timeseries through EEMD. The timeseries is shown in black, data gaps are highlighted in grey, and successive IMFs are displayed overlain on each other in blue, beginning from the highest frequency IMF and ending with the overall trend in the data.

*Video S3:* A demonstration of the IMFs derived from the Connecticut year-long timeseries through EEMD. The timeseries is shown in black, data gaps are highlighted in grey, and successive IMFs are displayed overlain on each other in blue, beginning from the highest frequency IMF and ending with the overall trend in the data.

**Appendix S1:** *A model for the background variability - White versus colored noise*

To illustrate why the background concentration time series may follow a Lorentzian spectrum ('brown noise' instead of 'white noise'), the following argument and simplified model is offered. If the concentration dynamics underlying the chemical constituent C are given by a simplified balance

$$\frac{dC}{dt} = -\tau\, C + I_c(t),$$

where $\tau$ is the inverse reaction or transformation time scale, and $I_c(t)$ is a time series of sources or sinks of C (due to stochastic hydrological additions or biological activities or both), then it can be shown that the Fourier power spectrum of C, defined as $E_c(f)$, is given as

$$E_c(f) = \frac{E_I(f)}{\tau^2 + f^2},$$

where $E_I(f)$ is the Fourier spectrum of the sources and sinks $I_c(t)$, and $f$ is the frequency (i.e. inverse timescale)[1]. In this reduced dynamical system featuring a balance between changes in storage (d$C$/dt), a first-order reaction or transformation loss $(-\tau\, C)$, and time-dependent sources and sinks $I_c(t)$ for C, diel patterns can arise when $E_I(f)$ itself exhibits diurnal variations (e.g., due to photochemical activities driven by variability in incident shortwave radiation with stochasticity originating from clouds and water level/turbidity changes). At high frequencies ($f/\tau$>>1) and for an $I_c(t)$ exhibiting uncorrelated (white-noise) structure, the simplified model here predicts $E_c(f) \sim 1/f^2$ (brown noise)[2]. Likewise, at low frequencies ($f/\tau$<<1), $E_c(f) \approx \tau^{-2} E_I(f)$ and the concentration spectrum scales as the 'input' or $E_I(f)$ spectrum with no frequency modulations. White noise structure is characterized by $\alpha = 0$, where $\alpha$ is the slope of the power spectrum, meaning that $E_I(f)$ is a constant independent of $f$. Brown noise structure,

often produced by random walks, has $1/f^2$ scaling and shows a decreasing slope with frequency in the power spectra. When $E_I(f)$ maintains white noise properties ($\alpha = 0$) across all frequencies, $E_c(f)$ reduces to the so-called 'Lorentzian' spectrum (flat at low frequency and $1/f^2$at higher frequencies). In the analysis here, $E_I(f)$ is rarely described by 'white-noise' and at short times or high frequencies, it may exhibit its own statistical structure so that $E_I(f) \sim f^{-\beta}$. This noise structure for $E_I(f)$ leads to a concentration spectrum given by $E_c(f) \sim f^{-2-\beta}$ for $f/\tau$>>1. That is, concentration spectra here resemble 'black-noise' when $\beta > 0$.

The goal of the work here is to detect and isolate diel modes superimposed on a concentration spectrum given as $E_c(f) \sim f^{-2-\beta}$. In summary, the conjecture here is that at $f/\tau$>>1 (high frequency range), $E_I(f)$ can be decomposed into a background 'noise' characterized by $E_I(f) \sim f^{-\beta}$ (where $\beta \geq 0$) and other *deterministic modes* of variability (i.e. single spike at a fixed frequency) due to biological activity. It is these deterministic modes that we seek to detect from measured $E_c(f)$ using Empirical Mode Decomposition analysis.

In the Lorentzian synthetic timeseries models we produced, we introduced the diel, seasonal, and tidal signals as a timeseries of sources and sinks $I_c(t)$. We chose the decay rate $\tau$ to be one month so that the diel frequency $f_{diel}$ would be in the realm of $f/\tau$>>1 (high frequency range) where $E_I(f) \sim f^{-\beta}$. Choosing a $\tau$ of this magnitude produces timeseries power spectra similar to many signals reported in the literature[3, 4], and to signals of the measured timeseries in this work (Appendix S2, Figures S3, S4).

**Appendix S2:** *Fast Fourier Transform and Wavelet Transform*

*Fourier Analysis*

Fourier Transform in general and Fast Fourier Transforms (FFT) in particular are common methods for converting data from the time domain (e.g. a time-series of a water quality constituent) into the "power" of different frequencies[1]. This method can efficiently decompose a time series into a summation of sinusoidal waves, with each wave representing the 'energy' (also 'activity', 'variability' or 'variance') of a frequency in the series ranging from half the sampling frequency (or the "Nyquist Frequency") to the inverse of the record duration[5]. The energy of a certain frequency is determined by the squared amplitude of the associated wave. Fourier transforms are energy preserving in that the sum of the squared amplitudes of all frequencies recovers the energy of the original time series. Practically for the scientist, the precise definition of the sampling interval n and the normalization of the series of length N are incorporated differently in different software packages, necessitating caution when comparing activity across methods.

One major drawback to Fourier analysis is its inability to sense locality of events in time. For example, a time series characterized by a single 'spike' in time affects all the amplitudes describing the Fourier decomposition. All FFT methods assume linearity (i.e. all time-series are a linear combination of sine waves with different amplitudes, frequencies and periods) and stationarity (i.e. the median value does not change with time) of the time series.

To illustrate the use of Fast Fourier Transformation (FFT) compared to the method developed in this manuscript, we performed FFT on the synthetic and measured time series using the base R `spectrum()` and `fft()` functions. Gaps in the timeseries were filled with a linear spline. We

identified peaks in the power spectra using the pracma package in R, then filtering peaks by comparing them to $1/f^2$ ($\alpha = 2$) scaling. The $1/f^2$ scaling is representative of chemical constituents in channels with more catchment filtering, usually rivers of higher Strahler stream orders[4].

The resulting power spectra from the spectrum() R function show that the synthetic data sets initially all had scaling significantly steeper than 1/f or $1/f^2$, and converged on $1/f^2$ scaling when environmental variability was added to the series. The component frequencies for the linear timeseries were correctly identified (Figure S3a,d), however the non-linear series created by a pulse and exponential decay signal produced a large number of high harmonics in the power spectra (Figure S3g). This is consistent with spurious frequencies attributed to windowing and finite-size effects acting on low frequency trends[6]. The diel frequency had an energy in the range of 1.3-4.4 for the synthetic series, with the lower energies associated with the signals that had observational error (i.e. gaps) introduced, and signals resulting from the pulse-decay function.

The spectrum() function applies various normalization and smoothing functions, such that the total energy of the transformed signal differs from the total energy of the original timeseries, in violation of Parseval's theorem. To estimate the recovery of signal strength through FFT, we therefore used the fft() R function instead. The amplitudes of the recovered diel signals were closest to original (0.102 mg/l) for the clean S&D Sin and S,D&T Sin signals. FFT of the clean Pulse-Decay signal returned signal of amplitude 0.074 mg/l. Introduction of environmental variability and observational energy reduced the amplitude of the recovered diel signal to 0.067 mg/l for the S&D Sin and S,D&T Sin signals, and to 0.047 mg/l for the Pulse-Decay signal.

The concentration power spectra of the three measured data series using the spectrum() function showed an approximate $1/f^2$ scaling, consistent with catchment and hillslope filtering of

environmental variability[3, 4, 7] (Figure S4). Superimposed on this power-law behavior, the Ichetucknee showed a strong peak at daily frequencies, as well as a number of harmonics. Though only the largest of these harmonics was detected with our peak detection algorithm, 4 other higher harmonics can be seen visually. The Connecticut had tidal peaks strong enough to be detected by our peak detection criteria, a peak at the diel frequency could still be seen visually but was not strong enough to pass the peak detection criteria. The Potomac had no peaks strong enough for detection through our algorithm, however a small peak near the diel frequency can be seen. All three spectra flattened into white noise at frequencies commensurate to 5000/year or about a 2-hour period commensurate with the Nyquist frequency (defined as twice the sampling frequency). Diel signal amplitude extracted using the fft() function indicated a diel amplitude of 0.012 mg/l for the Ichetucknee, 0.001 mg/l for the Connecticut, and 0.004 mg/l for the Potomac.

*Wavelet Transforms*

Wavelet transforms have some 'time-frequency' localization capabilities that circumvent the limitations of Fourier analysis in this regard. Full properties of wavelets are reviewed elsewhere[8] and are not repeated here. In wavelet analysis, a kernel of a given shape, amplitude and frequency is compared to the data at different time steps and the wavelet power of that frequency at that time is related to the goodness of fit. Unlike Fourier analysis, the wavelet kernel is not unique and numerous kernels or basis functions are allowed provided they satisfy a number of normalizing properties not discussed here. Like FFT, normalization methods differ amongst software packages, and the shapes of the kernels can impact the results under certain circumstances (e.g. symmetric versus asymmetric wavelets, degree of localization in frequency versus time, etc.) though wavelets under certain conditions (especially in detection of power-laws) are insensitive to kernel shape as demonstrated elsewhere[9]. A review article by Torrence

and Compo[10] presents a clear explanation of how to use wavelets and how to link frequencies in Fourier and wavelet analyses depending on choices made about the wavelet basis function.

Wavelet transforms are beneficial in decomposing signals in both space and time, however there are several circumstances under which empirical methods may be preferred. Tradeoffs exist, where better locality in one domain (say time) can be achieved if appropriate wavelet analyzing functions are selected, but this gain comes at a loss of resolution in the other domain (say frequency) and vice-a-versa. No wavelet transform can be fully accurate in both time and frequency domains simultaneously. Practically, this can mean a loss of timeseries properties such as phases of oscillations. Another drawback is redundancy – a time series of length $N$ results in $N^2$ wavelet coefficients (N coefficients required for all frequencies at every time instant for continuous wavelet transforms) – a huge redundancy in information content that is 'manufactured' by the transform, not the process under consideration[11, 12]. This manufacturing of information can produce bands of energy that are not descriptive of the underlying timeseries. Orthonormal wavelet transforms may be used to preserve the precise information content of a series without redundancy, but the arrangements of the wavelet coefficients require a courser resolution of the frequency-time wavelet half-plane making detection of some patterns more difficult [13, 14].

To demonstrate the use of wavelets, we used the WaveletComp package in R to compute continuous wavelets of our synthetic (Figure S5) and measured timeseries (Figure S6). Wavelet analysis of the synthetic timeseries were constant throughout time, and showed the same frequencies as the Fourier transform, including the first harmonics from the pulsed decay signal. The resolution of frequency in the Wavelet analysis is courser than Fourier, with wide bands appearing around the 24 hr periodicity rather than narrow and discrete horizontal bands (Figure

S5). The diel and tidal frequencies became more difficult to identify once environmental variability was added to the timeseries. The wavelet of the Potomac and Connecticut had very low power at most frequencies, with short durations (1-5 days) of slight power increases at the diel (Potomac & Connecticut) and tidal (Connecticut only) frequencies that are difficult to visually identify. The Ichetucknee showed consistent power at the diel frequency that was interrupted only by gaps in the data. All wavelet analyses showed significant distortion at the ends of the timeseries – an artifact of windowing and mathematical assumptions made about the form of the signal extrapolated past the ends of the measured or simulated timeseries.

**Appendix S3:** *Estimation of N-uptake and Gross Primary Productivity*

The N uptake predicted by the diel IMFs were calculated using a procedure described elsewhere[15], where the nighttime peaks of the timeseries were assumed to represent no autotrophic uptake, and the concentration differences between the diel trace and a linear spline that connected concentration peaks represented concentration deficits due to autotrophic biological uptake. We calculated this concentration deficit for each timestep and converted to a mass deficit though multiplication with the absolute value of simultaneous high-resolution sampled flowrate. Using the absolute value of discharge was necessary because of flow reversals in the Connecticut River on 160 of the days in the time series. After converting to mass deficits at each timestep, we summed the deficits at each timestep to a daily value. The conversion from concentration to mass allows for a consistent, biologically meaningful, interpretation of values across flow conditions. The same concentration deficit on a high flow day would indicate more N-uptake than the same concentration deficit on a low flow day, and by converting to mass uptake, this difference is more readily apparent.

The GPP estimates for the Potomac and Connecticut rivers were determined earlier using the StreamMetabolizer R package, which partitions changes in dissolved oxygen to GPP, ecosystem respiration and air-water gas exchange[16], and we accessed them through the StreamPulse data portal (data.streampulse.org). As the Connecticut River experiences flow reversals, and as these days were excluded from the published data set, we used the available oxygen data for the Connecticut River to remodel GPP using the StreamMetabolizer package[17]. Though flow reversals result in suboptimal metabolism estimates, given the high number of flow reversal days and the nature of our study, we decided that an increased number of less precise GPP estimates was preferable to fewer precise estimates. The GPP estimates derived here closely matched

published estimates for overlapping days. As these published estimates represent an ensemble of model fits, we used the mean and standard deviation of daily values in our calculations. GPP was estimated for the Ichetucknee using a 2 station approach, with the upstream $O_2$ concentration treated as a constant arising from the flow-weighted concentrations of the Ichetucknee's contributing springs[15], and gas evasion estimated from correlations among discharge, velocity, and gas exchange (as measured from floating domes).

All GPP estimates were provided as areal mass oxygen production, and were converted to estimates of N-uptake using a $C:O_2$ molar productivity ratio of 0.52 ± 0.15 for the Potomac and Connecticut which reflects a distribution of autotrophic respiration quotients[18-20], and a $C:O_2$ molar productivity ratio of 0.5 for the Ichetucknee to be consistent with previous work[15]. A C:N biomass molar ratio of 25:1[15, 21] was used for the Ichetucknee and 16.1 ± 5.46 :1[22] for the Potomac and Connecticut. Estimates of autotrophic N-uptake from these two methods were compared to assess how well the EEMD method isolated the diel signal from complex data series. These two metrics are nearly identical for previous work on the Ichetucknee River[15].

**Appendix S4:** *Solute footprint estimation*

We estimated the areal concentration footprint of a diel nitrate excursion using the width of a river and a one-dimensional advection-dispersion equation[23] rearranged to solve for longitudinal distance travelled by the excursion[24]:

$$x = \frac{u\left(\frac{1}{4D}\left(\frac{uT}{\mathrm{erfc}^{-1}\left(1 - 2\frac{C_\tau}{C_0}\right)}\right)^2 + T\right)}{1000}$$

where $x$ is the longitudinal distance travelled by the excursion (in km), $u$ is water velocity (in m $s^{-1}$), $T$ is the time duration of the excursion (in s), $D$ is the dispersion rate (in $m^2$ $s^{-1}$), erfc() is the complementary error function, $C_\tau$ is the concentration downstream at location $x$ and time τ (in mg $l^{-1}$), and $C_0$ is the magnitude of the concentration perturbation (in mg $l^{-1}$). We calculated this value for each river assuming a stream with no nitrate gets a 12-hour pulse ($T$ = 43,200 s) of magnitude equal to the diel nitrate range exhibited by the measured timeseries ($C_0$ = diel range on each day). The $C_\tau$ was defined as $C_0/2$, or complete mixing of the water. We parameterized $D = \alpha u^* h$ where $\alpha$ is specified as a distribution derived from published work[25], $h$ is the gage height (in m) and $u^*$ was parameterized as the slope of the relation between $u$ and (1/κ) log($h$) from USGS site visits where κ ≈ 0.41 is the van Karman's constant. For each site, we used data reported by the USGS from site visits to relate $u$ to discharge (in $m^3$ $s^{-1}$), and to relate depth of flow ($h$) and channel width ($w$) to $u$. For each day in the timeseries, average discharge was converted to $u$, $h$, and $w$ times using these relations, and footprints were computed as the product of distance $x$ (in km) and width $w$ (in km) of the river reach. This computation was performed 10,000 times for each day sampling the distributions of $u$, $h$, $w$, and $\alpha$. The mean and standard

error of the resulting concentration footprints were used in creating the vertical error bars in Figure S7 and are reported in the text. This approach may over-estimate the areal footprint because of 3 implicit assumptions: (i) lateral dispersion is ignored in the estimation of $x$, (ii) bulk estimates of $u$ and $D$ are used, which overestimate the local advection and dispersion in the vicinity of the biological sink (i.e. river bed and sides), and (iii) the flow field is assumed to be uniform along $x$ (channel section irregularities act as 'retardation' factors due to the presence of bends and dead-zones).

**Supporting Figures:**

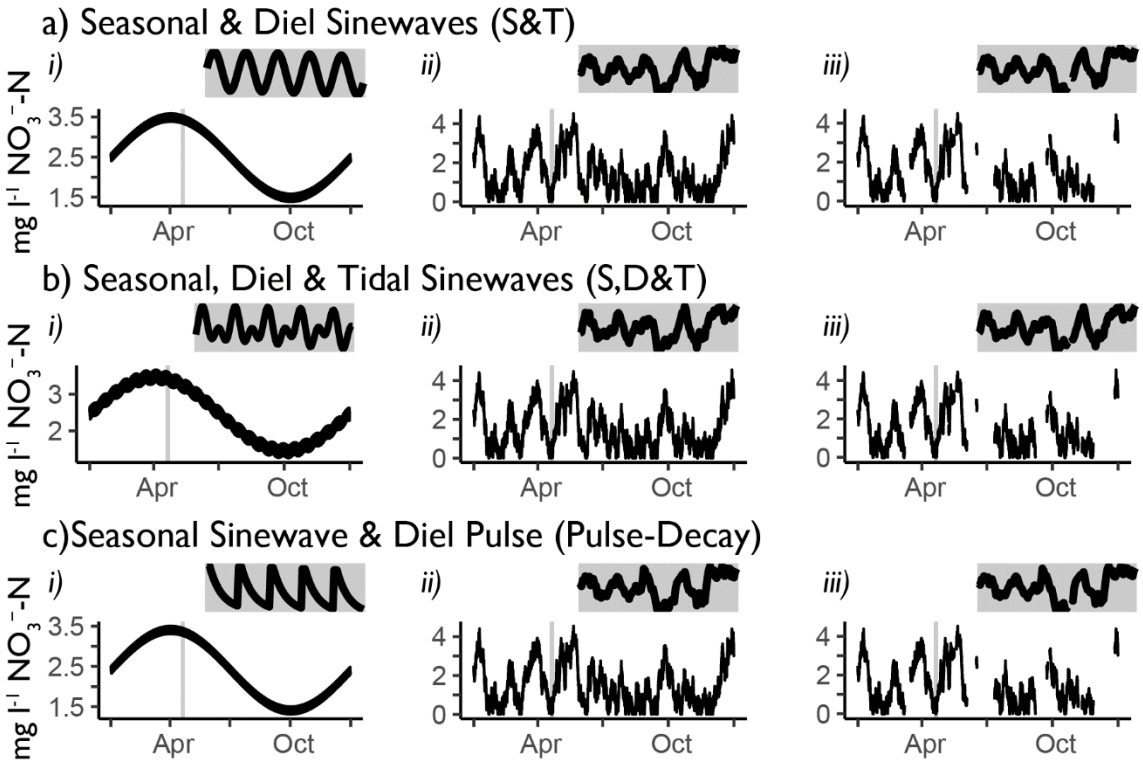


**Figure S1:** Synthetic time series used for illustrating detection and separation of diel patterns from other modes of variability. a) seasonal and diel sinewaves superimposed on each other b) seasonal, diel and tidal sinewaves superimposed on each other and c) seasonal sinewave and diel pulse-decay patterns superimposed on each other. Subsections i) show the simple time series, ii) show the time series with environmental variability included, and iii) show the time series with environmental and sensor induced variability and gaps included. In all plots, time periods highlighted with shaded boxes are amplified in the insets to illustrate fine-scale variability.

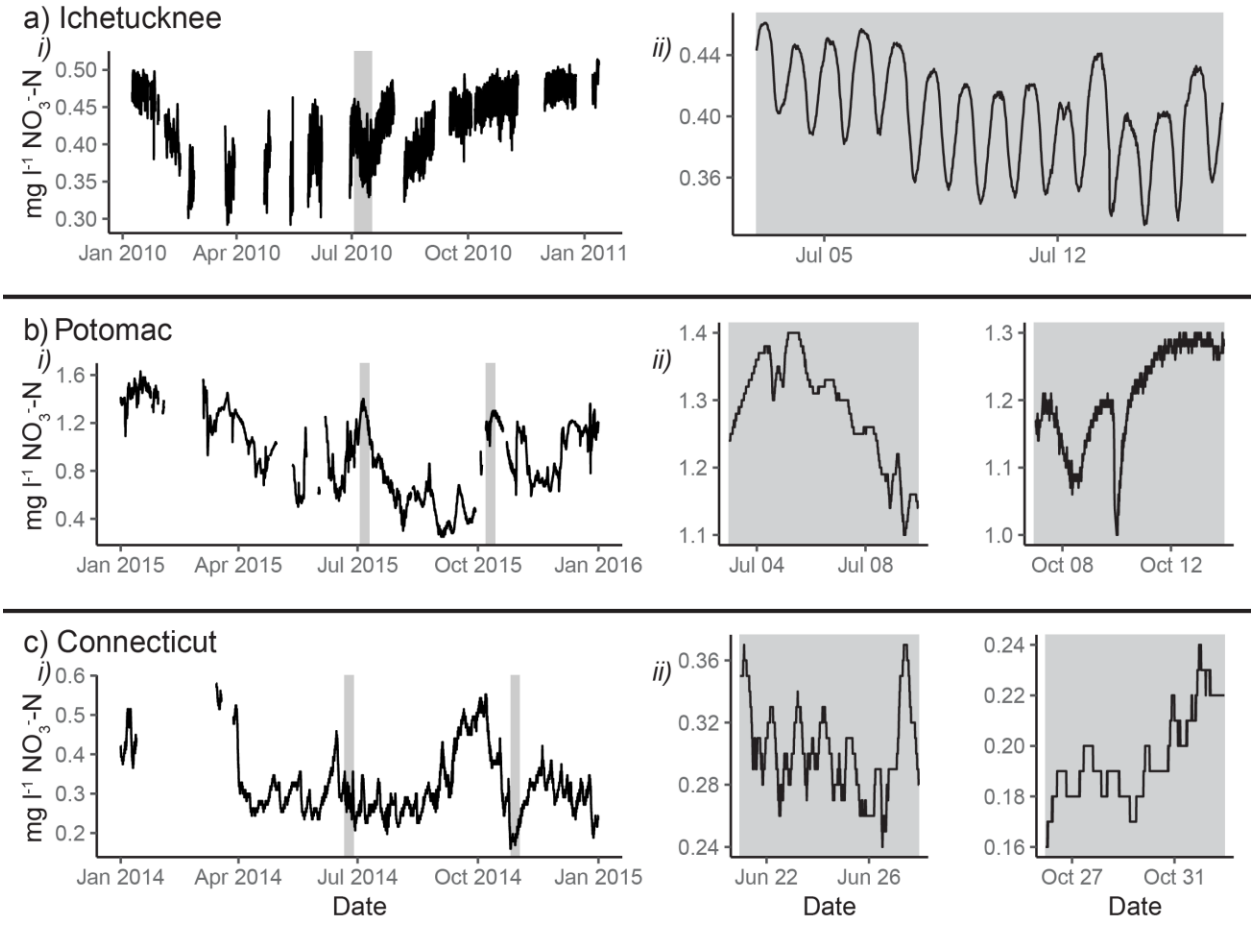


**Figure S2:** Measured nitrate concentration time series sampled at 15-minute intervals from the a) Ichetucknee River in 2010 b) Potomac River in 2015 and c) Connecticut River in 2014. Subsections i) show the whole series, and time periods highlighted with shaded boxes are amplified in subsections ii) for visualization of diel features.

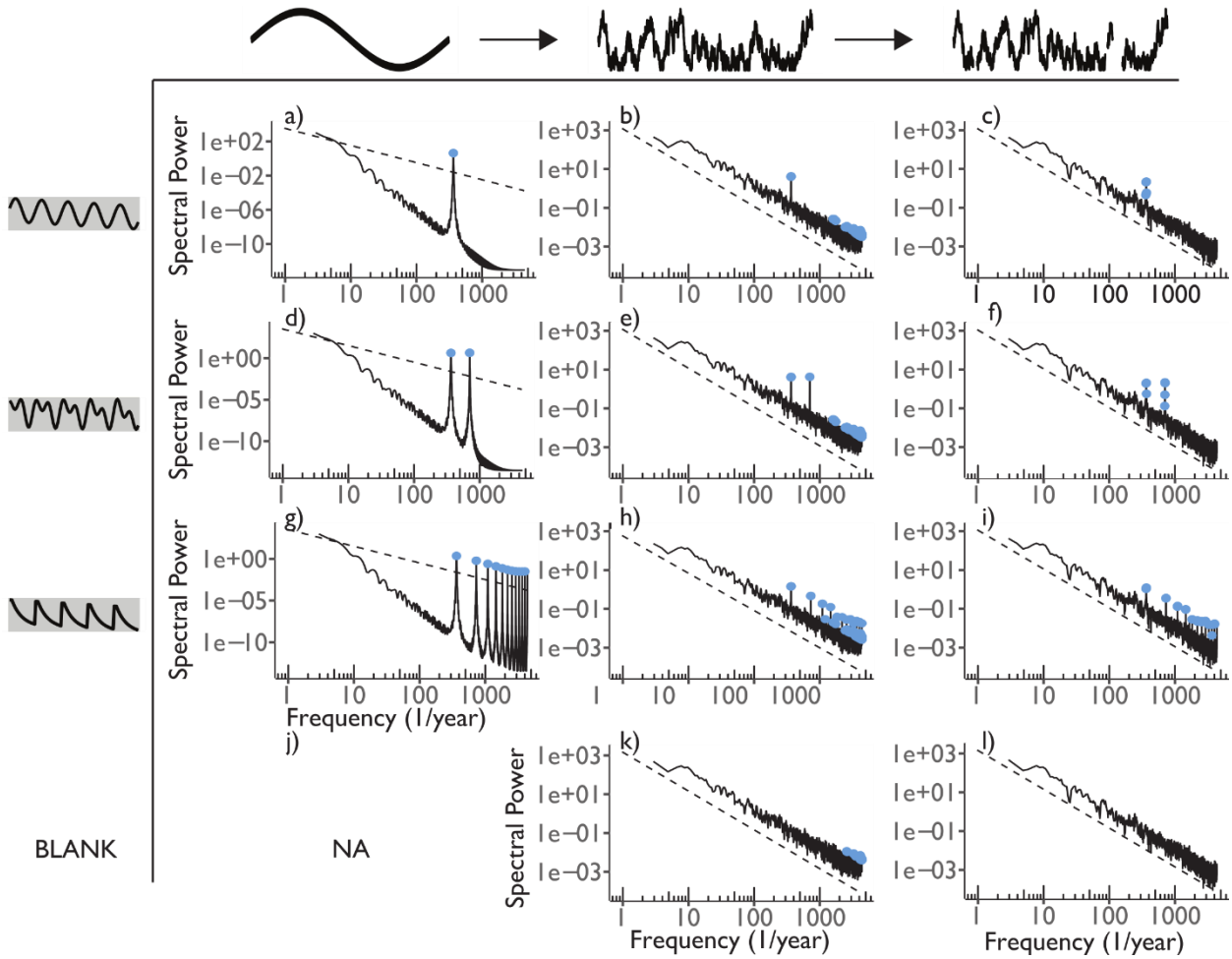


**Figure S3:** Power spectra of the three synthetic data series; Seasonal & Diel Sine: a-c; Seasonal, Diel & Tidal Sine: d-f; Pulse-Decay g-i; Blank with just the environmental structure j-l). Power spectra are shown of the simple signal (a, d, g & j), the signals with environmental variability (b, e, h & k) and the signals with variability from environmental and detection processes (c, f, I & l). The dashed line shows $1/f^2$ scaling ($\alpha = 2$), and blue dots show peaks with power at least 50x $1/f^2$ scaling.

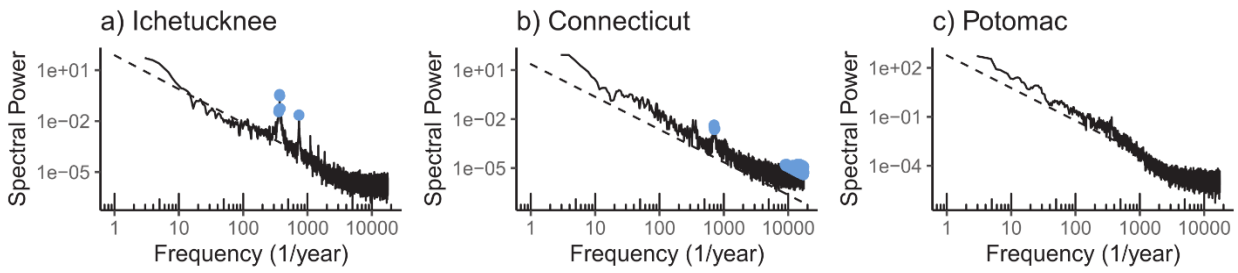


**Figure S4:** Power spectra of the example three measured data series Power spectra are shown for the a) Ichetucknee River, b) Connecticut River, and c) Potomac River. The dashed line shows $1/f^2$ scaling ($\alpha = 2$), and blue dots show peaks with power at least 50x $1/f^2$ scaling.

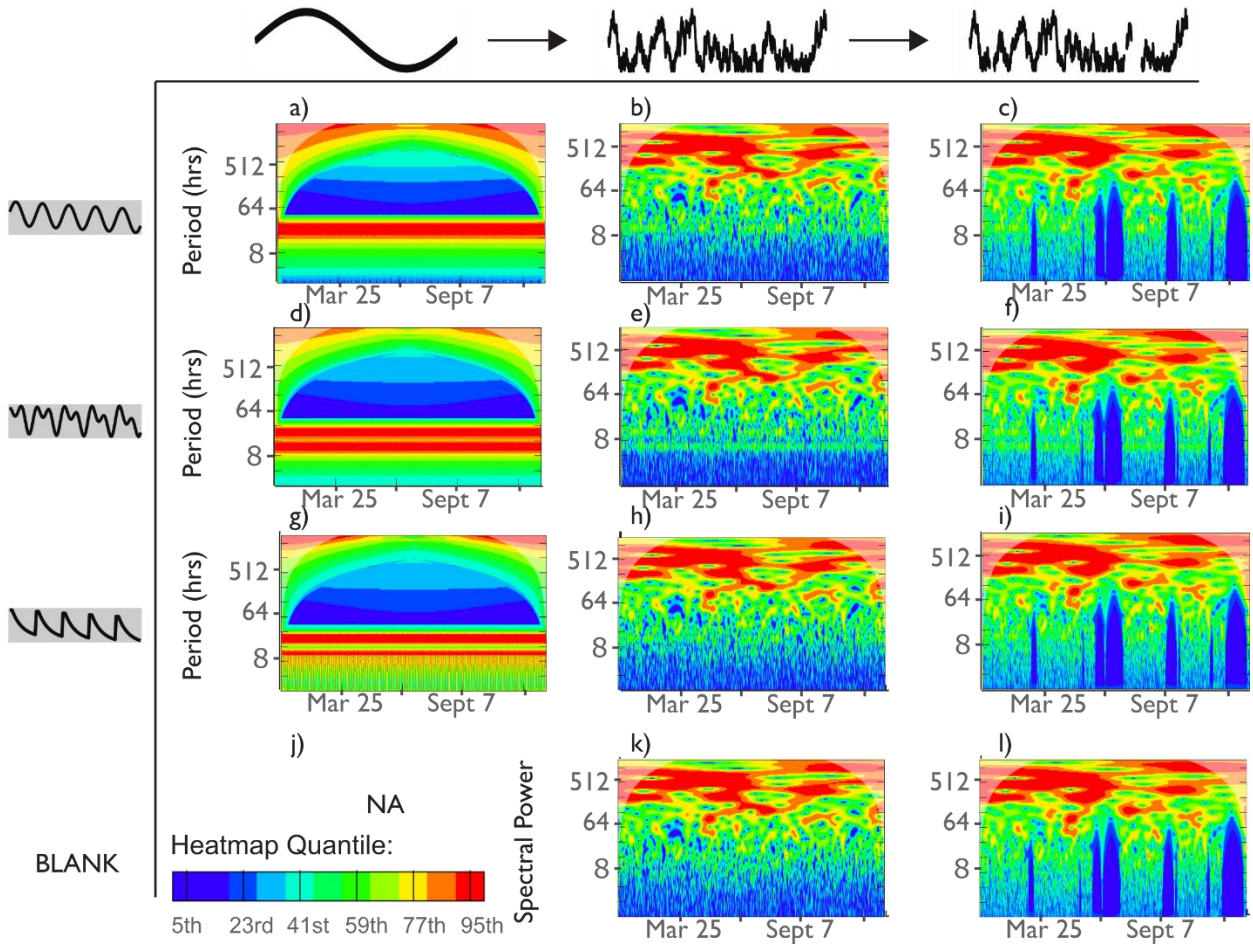


**Figure S5:** Wavelet analysis of the three synthetic data series; Seasonal & Diel Sine: a-c; Seasonal, Diel & Tidal Sine: d-f; Pulse-Decay g-i; Blank with just the environmental structure j-l). Wavelet heatmaps are shown of the simple signal (a, d, g & j), the signals with environmental variability (b, e, h & k) and the signals with variability from environmental and detection processes (c, f, I & l). The color maps on all panels correspond to the quantiles of energy within each measured timeseries. Thus, colors within panels are comparable, but colors are not comparable across panels.

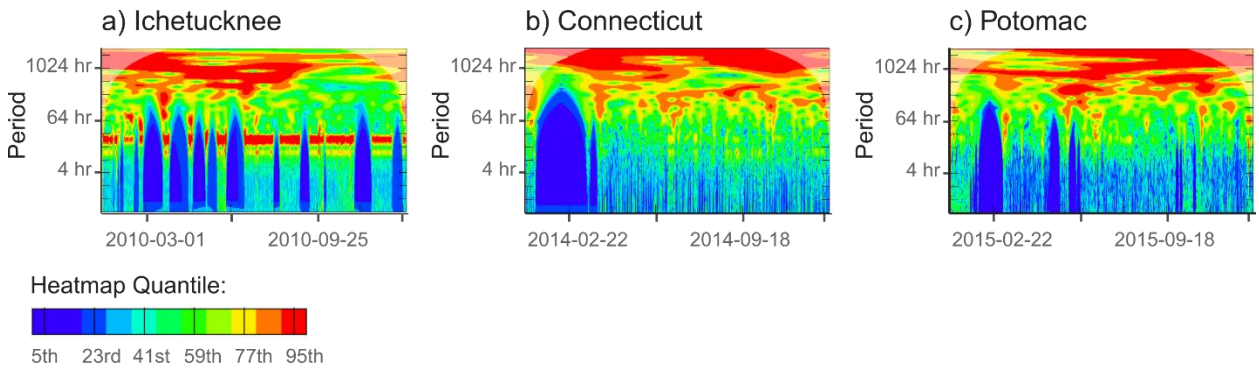


**Figure S6:** Wavelet heatmaps of the example three measured data series. Wavelets are shown for the a) Ichetucknee River, b) Connecticut River, and c) Potomac River. The color maps on all panels correspond to the quantiles of energy within each measured timeseries. Thus, colors within panels are comparable, but colors are not comparable across panels.

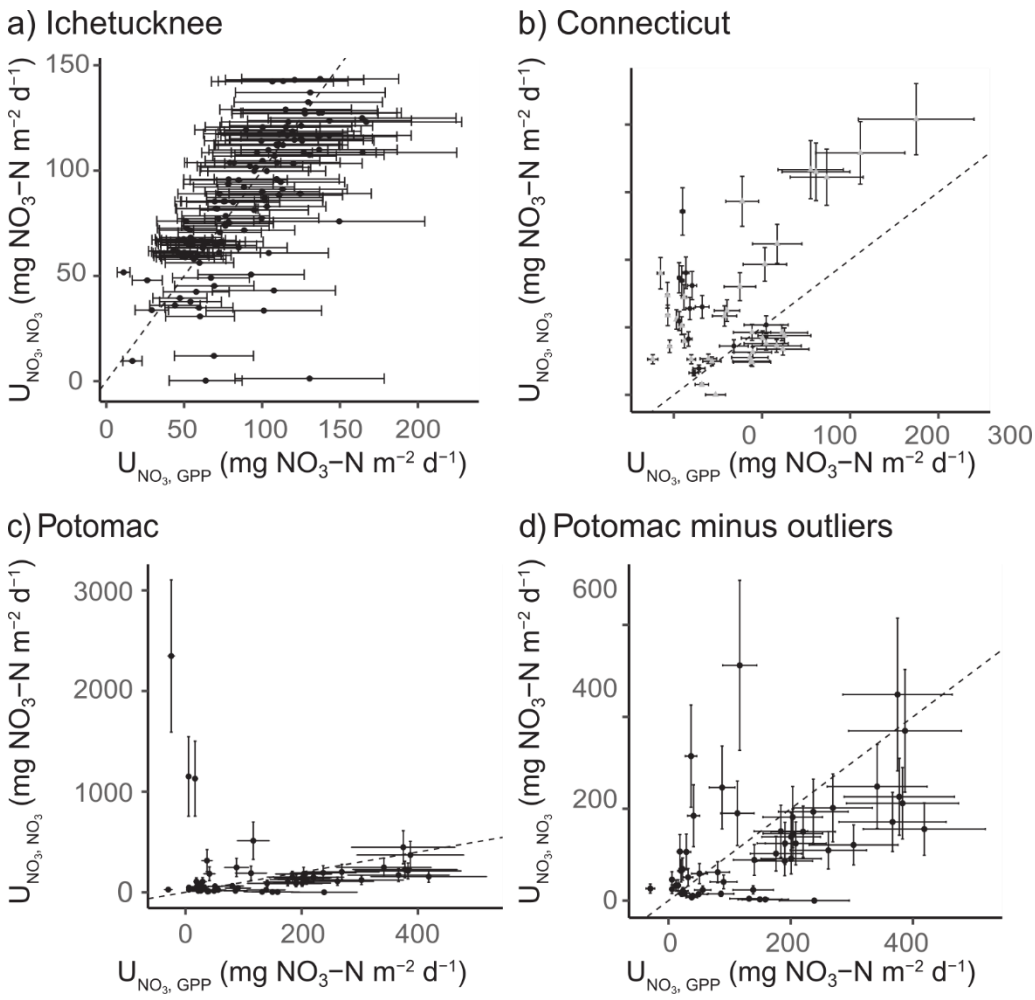


**Figure S7**: Areal daily N-uptake estimated from concentration deficit in the diel IMF (y-axis; $U_{NO3, NO3}$) plotted against areal daily N-uptake estimated from GPP (x-axis; $U_{NO3, GPP}$) with accompanying error in a) the Ichetucknee River, b) the Connecticut River, c) the Potomac River, and d) the Potomac River minus 3 outliers that correlated with storm events. Error bars represent the 95% confidence interval (± 2 standard error). Error for $U_{NO3, GPP}$ stems from uncertainty around estimates of GPP, autotrophic respiration quotients, and C:N stoichiometry of photoautotrophs. Error for $U_{NO3, NO3}$ derives from uncertainty around the size of the chemical signature footprint. The black dotted line shows the 1:1 line. Grey triangles in a) denote estimates from days that had flow reversals and are therefore less reliable than days shown in black circles. They are maintained because of their prevalence in the time series. In all panels,

only days on which signal was extracted from the timeseries are shown. Though theoretically, negative uptake values are impossible, published model results of GPP are occasionally negative[16], and are kept for consistency in describing uncertainty. Error bars in the Ichetucknee River do not account for error associated with GPP, as it was computed arithmetically and not statistically.

# MetadataS1

---

## For: Data S1

### All Compiled Data and Code

---

## Author(s) [of the material provided in DataS1.zip]

Catherine Chamberlin
Duke University
9 Circuit Drive, Durham, NC 27709
catherine.chamberlin@duke.edu

Jim Heffernan
Duke University
9 Circuit Drive, Durham, NC 27709
james.heffernan@duke.edu

Gabriel Katul
Duke University
9 Circuit Drive, Durham, NC 27709
gaby@duke.edu

---

## File list (files found within DataS1.zip)

```
README.txt
/Code/
./AnalyzeConnecticutIMFs.Rmd
./AnalyzeIchetuckneeIMFs.Rmd
./AnalyzeLorentzianIMFs.Rmd
./AnalyzePotomacIMFs.Rmd
./CreateSyntheticSeries.Rmd
./Download+Save_MetabolismEstimates.Rmd
./EEMDComputation.Rmd
./FFTAnalysis.R
./HelperFunctions.R
./MakeFigS2.Rmd
./MakeFig1.Rmd
./MakeFig2.Rmd
```

```
./MakeFigS1.Rmd
./MakeVideo1.Rmd
./MakeVideo2.Rmd
./MakeVideo3.Rmd
./ModelConnecticutMetabolism.R
./Nuptake_Connecticut.Rmd
./Nuptake_Ichetucknee.Rmd
./Nuptake_Potomac.Rmd
./ProcessIchetuckneeData.Rmd
./USGSDataDownload.Rmd
./WaveletAnalysis.R
/Data/
./Ichetucknee_2010_parsed_jbh.csv
./Ichetucknee_metabolism.csv
/Output/
./Conn_ratingcurve.RData
./Conn_USGSdata.RData
./CT_nwis-0113050_metab.RData
./Ich_data.RData
./MD_nwis-0164500_metab.RData
./Pot_ratingcurve.RData
./Pot_USGSdata.RData
./SyntheticSeries.RData
./Figures/
./../FigS2.pdf
./../Fig1a.pdf
./../Fig1b.pdf
./../Fig1c.pdf
./../Fig1d.pdf
./../Fig2.pdf
./../Fig3a.pdf
./../Fig3b.pdf
./../Fig3c.pdf
./../Fig3d.pdf
./../FigS1.pdf
./../FigS1insets.pdf
./../FigS3.pdf
./../FigS4.pdf
./../FigS5a.tif
./../FigS5b.tif
./../FigS5c.tif
./../FigS5d.tif
./../FigS5e.tif
./../FigS5f.tif
```

```
././FigS5g.tif
././FigS5h.tif
././FigS5i.tif
././FigS5k.tif
././FigS5l.tif
././FigS6a.tif
././FigS6b.tif
././FigS6c.tif
././FigS7a.pdf
././FigS7b.pdf
././FigS7c.pdf
././FigS7d.pdf
./IMFs_Connecticut/
././Connecticut_compiled.RData
./IMFs_Ichetucknee/
././Ichetucknee_compiled.RData
./IMFs_Lorentzian/
././ts.blank.c_compiled.RData
././ts.blank_compiled.RData
././ts1.b_compiled.RData
././ts1.c_compiled.RData
././ts1_compiled.RData
././ts2.b_compiled.RData
././ts2.c_compiled.RData
././ts2_compiled.RData
././ts3.b_compiled.RData
././ts3.c_compiled.RData
././ts3_compiled.RData
./IMFs_Potomac/
././Potomac_compiled.RData
./VideoFrames/
././EEMDConn_p1.png … EEMDConn_p17.png
././EEMDIch_p1.png … EEMDIch_p16.png
././EEMDPot_p1.png … EEMDPot_p16.png
```

---

## Description

This compilation should run on most systems provided the folder structure and file organization is left intact.

All code was developed using R 4.0.2. The following packages must be installed to run the code:

streamMetabolizer_0.11.4
StreamPULSE_0.0.0.9043
Cairo_1.5-12.2
shiny_1.5.0
padr_0.5.2
dataRetrieval_2.7.6
xts_0.12.1
dygraphs_1.1.1.6
pracma_2.2.9
gridExtra_2.3
lubridate_1.7.9
hht_2.1.3
zoo_1.8-8
doParallel_1.0.15
iterators_1.0.12
forcats_0.5.0
stringr_1.4.0
dplyr_1.0.2
purr_0.3.4
readr_1.3.1
tidyr_1.1.2
tibble_3.0.3
ggplot2_3.3.2
tidyverse_1.3.0
foreach_1.5.0
WaveletComp_1.1

`README.txt` – This file includes a brief description of the purpose of the zip folder DataS1.zip.

`AnalyzeConnecticutIMFs.Rmd` – This file accesses files `Conn_USGSdata.RData` and `Connecticut_compiled.RData`. This file analyzes the compiled EEMD results of the decomposition of the measured timeseries from the Connecticut River. This R Markdown file can be knitted or run chunk by chunk. Code provided first views all of the IMFs, examines their periods, isolates the Diel IMF using functions sourced from the `HelperFunctions.R` file, and calculates statistics on the extent, magnitude and seasonality of diel signal.

`AnalyzeIchetuckneeIMFs.Rmd` – This file accesses files `Ich_data.RData` and `Ichetucknee_compiled.RData`. This file analyzes the compiled EEMD results of the decomposition of the measured timeseries from the Ichetucknee River. This R Markdown file can be knitted or run chunk by chunk. Code provided first views all of the IMFs, examines their periods, isolates the Diel IMF using functions sourced from the `HelperFunctions.R` file, and calculates statistics on the extent, magnitude and seasonality of diel signal.

`AnalyzeLorentzianIMFs.Rmd` – This file accesses files `SyntheticSeries.RData`, `ts.blank.c_compiled.RData`, `ts.blank_compiled.RData`, `ts1.b_compiled.RData`, `ts1.c_compiled.RData`, `ts1_compiled.RData`, `ts2.b_compiled.RData`, `ts2.c_compiled.RData`, `ts2_compiled.RData`, `ts3.b_compiled.RData`, `ts3.c_compiled.RData` and `ts3_compiled.RData`. This file analyzes the compiled EEMD results of the decomposition of the Lorentzian timeseries. This R Markdown file can be knitted or run chunk by chunk for one timeseries at a time. Code provided first views displays all of names of the Lorentzian timeseries, and requires user input to choose which series to view for the remainder of the file. Once a timeseries is selected, the code displays all IMFs, examines their periods, isolates the Diel IMF using functions sourced from the `HelperFunctions.R` file, and calculates statistics on the extent, magnitude, and percent signal recovery.

`AnalyzePotomacIMFs.Rmd` – This file accesses files `Pot_USGSdata.RData` and `Potomac_compiled.RData`. This file analyzes the compiled EEMD results of the decomposition of the measured timeseries from the Potomac River. This R Markdown file can be knitted or run chunk by chunk. Code provided first views all of the IMFs, examines their periods, isolates the Diel IMF using functions sourced from the `HelperFunctions.R` file, and calculates statistics on the extent, magnitude and seasonality of diel signal.

`CreateSyntheticSeries.Rmd` – This file creates the Lorentzian timeseries following the mathematical description in the Supplemental Information. A base timeseries of either sinusoidal or exponential decay shape is successively contaminated with brown noise (i.e. a random walk) of a certain magnitude representing environmental variability, and then with white noise and gaps representing measurement error or observational variability. The power spectra for each timeseries is available for viewing. "Blank" timeseries of just background variability are also produced that do not have any structure as their base. Produces output `SyntheticSeries.RData`.

`Download+Save_MetabolismEstimates.Rmd` – This file accesses the metabolism estimates for the Potomac and Connecticut rivers that were available in the literature[16], and accessible through the StreamPulse data portal (data.streampulse.org). As the metabolism of the Connecticut River was also estimated locally, several graphical comparisons of the two estimations are provided in this file. Produces output `CT_nwis-01193050_metab.RData` and `MD_nwis-01646500_metab.RData`.

`EEMDComputation.Rmd` – This file accesses files `Conn_USGSdata.RData`, `Pot_USGSdata.RData`, `Ich_data.RData` and `SyntheticSeries.RData`. It uses the hht package to compute 200 trial decompositions of each timeseries, then compiles them and resifts them according to the method described by the authors of the package[26]. Outputs of this code file include `Connecticut_compiled.RData` ,

Ichetucknee_compiled.RData, Potomac_compiled.RData, ts.blank.c_compiled.RData, ts.blank_compiled.RData, ts1.b_compiled.RData, ts1.c_compiled.RData, ts1_compiled.RData, ts2.b_compiled.RData, ts2.c_compiled.RData, ts2_compiled.RData, ts3.b_compiled.RData, ts3.c_compiled.RData, and ts3_compiled.RData. Running the code also creates the folders Connecticut_eemd/, Ichetucknee_eemd/, Potomac_eemd/, ts.blank.c_eemd/, ts.blank_eemd/, ts1.b_eemd/, ts1.c_eemd/, ts1_eemd/, ts2.b_eemd/, ts2.c_eemd/, ts2_eemd/, ts3.b_eemd/, ts3.c_eemd/, and ts3_eemd/, and 200 trials in each folder. For space reasons, these trial files are not included in the zipped folder. Please contact the authors if they are needed.

FFTAnalysis.R contains code to perform Fourier analysis of the measured and synthetic timeseries and produce the power spectra plots in Fig S3 and Fig S4.

HelperFunctions.R contains 5 functions that are called in many of the other code files. They are loaded when this script is sourced by the other RMarkdown files. The first function is auto.installer() which combines installation and loading libraries. Alternatively, experienced R users can install packages manually. The second is my_PlotIMFs(), which is a very slight modification of PlotIMFs() from the hht package to allow for more descriptive plot titles. Investigate.IMF() creates an interactive graph of all of the selected IMFs overlain on each other. This is a helpful function for inspecting regions of a timeseries more closely. format_EEMD_result() and filter_imfs() are meant to be used in concert, where the input of filter_imfs() is the output of format_EEMD_result(). filter_imfs() will isolate the signal of a given period and phase from all of the IMFs. Phase is only supported for 24 hour periodicity at this point.

MakeFigS1.Rmd contains code to create Fig S1 of the Lorentzian data series. This file loads SyntheticSeries.RData and writes FigS1.pdf and FigS1insets.pdf. The presented Fig S1 was created using these pdf files in Adobe Illustrator.

MakeFigS2.Rmd contains code to create Fig S2 of the Measured data series. This file loads Ich_data.RData, Conn_USGSdata.RData, and Pot_USGSdata.RData and writes FigS2.pdf. The presented Fig S2 was created using this pdf file in Adobe Illustrator.

MakeFig1.Rmd contains code to create Fig 1. This file loads Conn_USGSdata.RData and Connecticut_compiled.RData and writes Fig1a.pdf, Fig1b.pdf, Fig1c.pdf and Fig1d.pdf. The presented Fig 1 was created using these pdf files in Adobe Illustrator.

MakeFig2.Rmd contains code to create Fig 2. This file loads Ich_data.RData, Ichetuknee_compiled.RData, Conn_USGSdata.RData,

`Connecticut_compiled.RData`, `Pot_USGSdata` and `Potomac_compiled.RData` and writes `Fig2.pdf`.

`MakeVideo1.Rmd` contains code to create the frames for Video S1. This file loads `Ich_data.RData`, and `Ichetucknee_compiled.RData` and writes `EEMDIch_p[1:16].png`. The video was created using these images through Windows Video Editor.

`MakeVideo2.Rmd` contains code to create the frames for Video S2. This file loads `Pot_USGSdata.RData`, and `Potomac_compiled.RData` and writes `EEMDPot_p[1:16].png`. The video was created using these images through Windows Video Editor.

`MakeVideo3.Rmd` contains code to create the frames for Video S3. This file loads `Conn_USGSdata.RData`, and `Connecticut_compiled.RData` and writes `EEMDConn_p[1:17].png`. The video was created using these images through Windows Video Editor.

`ModelConnecticutMetabolism.R` contains code to model metabolism using the streamMetabolizer package. Settings match those used in published literature[16] except for the decisions on which days to include in the model. This code produces the file `CT_Modelled_metab.RData`. As the model output file is large, the output is not included in this zipped folder. Please contact authors if the file is needed.

`Nuptake_Connecticut.Rmd` contains code to calculate the N-uptake for the Connecticut River predicted from the diel nitrate variability according to methods developed in the literature[15]. Independent estimates of N-uptake are also produced from GPP estimates. This file also calculates the theoretical concentration footprints based on literature methods[24]. All estimates are compared with each other, and error surrounding these estimates are provided. This file loads `Conn_USGSdata.RData`, `Connecticut_compiled.RData`, `CT_modelled_metabolism.RData` and `Conn_ratingcurve.RData`, and writes `Fig3b.pdf` and `FigS7b.pdf`. As `CT_modelled_metabolism.RData` is large, it is not included in this zipped folder. Please contact authors if the file is needed.

`Nuptake_Ichetucknee.Rmd` contains code to calculate the N-uptake for the Ichetucknee River predicted from the diel nitrate variability according to methods developed in the literature[15]. Independent estimates of N-uptake are also produced from GPP estimates. All estimates are compared with each other. This file loads `Ich_data.RData`, `Ichetucknee_compiled.RData` and `Ichetucknee_metabolism.csv`, and writes `Fig3a.pdf` and `FigS7b.pdf`.

`Nuptake_Potomac.Rmd` contains code to calculate the N-uptake for the Potomac River predicted from the diel nitrate variability according to methods developed in the literature[15]. Independent estimates of N-uptake are also produced from GPP estimates. This file also calculates the theoretical concentration footprints based on literature methods[24]. All estimates are compared with each other, and error surrounding these estimates are provided. This file loads `Pot_USGSdata.RData, Potomac_compiled.RData, MD_nwis-01646500_metab.RData` and `Pot_ratingcurve.RData,` and writes `Fig3c.pdf, Fig3d.pdf, FigS7b.pdf` and `FigS7c.pdf.`

`ProcessIchetuckneeData.Rmd` contains code that formats the Ichetucknee nitrate data to be consistent with data retrieved from the USGS NWIS site. The code loads `Ichetucknee_2010_parsed_jbh.csv` and writes `Ich_data.RData.`

`USGSDataDownload.Rmd` contains code to access relevant data from the USGS NWIS site using the dataRetrieval R package. Continuous nitrate and discharge data are downloaded from the Connecticut River (gage 01193050) and the Potomac River (gage 016456500). The data are lightly processed to ensure a consistent sampling frequency, no duplicate measurements, and units of m3s for discharge. Rating curves for both sites are also downloaded and saved. This script exports files `Conn_USGSdata.RData, Pot_USGSdata.RData, Conn_ratingcurve.RData` and `Pot_ratingcurve.RData.`

`WaveletAnalysis.R` contains code to compute and print the continuous wavelet analyses of the measured and synthetic timeseries. The code saves the wavelet analyses in ./Output, and due to their size, these output files are not included in the initial DataS1 zip file. The code also creates .tif files to create Fig S5 and Fig S6. Fig S5 and S6 were compiled in Inkscape.

`Ichetucknee_2010_parsed_jbh.csv` contains direct observations of $NO_3^-$ and other water quality parameters as well as derived estimates of metabolism from O2 and nitrogen uptake from diel $NO_3^-$ variability. $NO_3^-$ data were collected from Satlantic SUNA V1 sensors and other water quality parameters collected by YSI sonde. Sensors were deployed ~5km downstream of the Ichetucknee Headspring at US Highway 27 near USGS gage 02322700, and cleaned weekly concurrent with data download. Diel N uptake was estimated based on diel $NO_3^-$ deviation from two different baselines: the prior nighttime maxima and the mean of the prior and subsequent nighttime maxima[15].

`Ichetucknee_metabolism.csv` contains estimates of metabolism (Gross Primary Production and Ecosystem Respiration) derived from diel variation in $O_2$ concentrations and other physical and chemical parameters. Because the Ichetucknee's flow is derived from 5 distinct large springs, assumptions of longitudinal $O_2$ equilibrium required by single-station methods were not applicable. Instead, metabolism was calculated using a 2-station approach with the upstream boundary $O_2$ concentration estimated from the flow-weighted contributions of the hydrologically- and chemically-stable springs. Gas exchange was estimated from discharge

based on prior studies of the relationship between dome-gas exchange and flow velocity and applied to the river surface area upstream of the sensors.

`Conn_ratingcurve.RData` is data accessed from the USGS describing their site visits to gage 01193050. The file is written by `USGSDataDownload.Rmd`.

`Conn_USGSdata.RData` is continuous monitoring data from USGS NWIS for site 01193050. The file is written by `USGSDataDownload.Rmd`.

`CT_nwis-01193050_metab.RData` contains metabolism estimates for site 01193050 accessed through StreamPulse. The file is written by `Download+Save_MetabolismEstimates.Rmd`.

`Ich_data.RData` contains the reformatted data from `Ichetucknee_2010_parsed_jbh.csv`. The file is written by `ProcessIchetuckneeData.Rmd`.

`MD_nwis-01646500_metab.RData` contains metabolism estimates for site 016456500 accessed through StreamPulse. The file is written by `Download+Save_MetabolismEstimates.Rmd`.

`Pot_ratingcurve.RData` is data accessed from the USGS describing their site visits to gage 016456500. The file is written by `USGSDataDownload.Rmd`.

`Pot_USGSdata.RData` is continuous monitoring data from USGS NWIS for site 016456500. The file is written by `USGSDataDownload.Rmd`.

`SyntheticSeries.RData` contains the Lorentzian timeseries created by `CreateSyntheticSeries.Rmd`.

`FigS1.pdf` is written by `MakeFigS1.Rmd` and contains segments of Figure S1. Figure S1 was created in Adobe Illustrator.

`FigS1insets.pdf` is written by `MakeFigS1.Rmd` and contains segments of Figure S1. Figure S1 was created in Adobe Illustrator.

`FigS2.pdf` is written by `MakeFigS2.Rmd` and contains segments of Figure S2. Figure S2 was created in Adobe Illustrator.

`Fig1a.pdf` is written by `MakeFig1.Rmd` and contains segments of Figure 1. Figure 1 was created in Adobe Illustrator.

`Fig1b.pdf` is written by `MakeFig1.Rmd` and contains segments of Figure 1. Figure 1 was created in Adobe Illustrator.

`Fig1c.pdf` is written by `MakeFig1.Rmd` and contains segments of Figure 1. Figure 1 was created in Adobe Illustrator.

`Fig1d.pdf` is written by `MakeFig1.Rmd` and contains segments of Figure 1. Figure 1 was created in Adobe Illustrator.

`Fig2.pdf` is written by `MakeFig2.Rmd` and is the same as published Figure 2.

`Fig3a.pdf` is written by `Nuptake_Ichetucknee.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate. Fig 3 in the manuscript was created in Adobe Illustrator and Inkscape.

`Fig3b.pdf` is written by `Nuptake_Connecticut.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate. Fig 3 in the manuscript was created in Adobe Illustrator and Inkscape.

`Fig3c.pdf` is written by `Nuptake_Potomac.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate. Fig 3 in the manuscript was created in Adobe Illustrator and Inkscape.

`Fig3d.pdf` is written by `Nuptake_Potomac.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate. Fig 3 in the manuscript was created in Adobe Illustrator and Inkscape.

`FigS3.pdf` is written by `FFTAnalysis.R` Fig S3 in the manuscript was created in Inkscape.

`FigS4.pdf` is written by `FFTAnalysis.R` Fig S4 in the manuscript was created in Inkscape.

`FigS5a.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5b.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5c.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5d.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5e.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5f.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5g.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5h.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5i.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5k.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS5l.tif` is written by `WaveletAnalysis.R` Fig S5 in the manuscript was created in Inkscape.

`FigS6a.tif` is written by `WaveletAnalysis.R` Fig S6 in the manuscript was created in Inkscape.

`FigS6b.tif` is written by `WaveletAnalysis.R` Fig S6 in the manuscript was created in Inkscape.

`FigS6c.tif` is written by `WaveletAnalysis.R` Fig S6 in the manuscript was created in Inkscape.

`FigS7a.pdf` is written by `Nuptake_Ichetucknee.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate with error bars for both methods. Fig S7 in the manuscript was created in Inkscape.

`FigS7b.pdf` is written by `Nuptake_Connecticut.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate with error bars for both methods. Fig S7 in the manuscript was created in Inkscape.

`FigS7c.pdf` is written by `Nuptake_Potomac.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate with error bars for both methods. Fig S7 in the manuscript was created in Inkscape.

`FigS7d.pdf` is written by `Nuptake_Potomac.Rmd` and shows the correlation between N-uptake estimates from GPP and from isolated diel oscillations of nitrate with error bars for both methods. Fig S7 in the manuscript was created in Inkscape.

`Connecticut_compiled.RData` contains the compiled EEMD results from all 200 trials for the Connecticut River. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`Ichetucknee_compiled.RData` contains the compiled EEMD results from all 200 trials for the Ichetucknee River. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts.blank.c_compiled.RData` contains the compiled EEMD results from all 200 trials for the blank Lorentzian timeseries with observational variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts.blank_compiled.RData` contains the compiled EEMD results from all 200 trials for the blank Lorentzian timeseries with environmental variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts1.b_compiled.RData` contains the compiled EEMD results from all 200 trials for the S&D Lorentzian timeseries with environmental variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts1.c_compiled.RData` contains the compiled EEMD results from all 200 trials for the S&D Lorentzian timeseries with observational variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts1_compiled.RData` contains the compiled EEMD results from all 200 trials for the S&D Lorentzian timeseries. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts2.b_compiled.RData` contains the compiled EEMD results from all 200 trials for the S,D&T Lorentzian timeseries with environmental variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts2.c_compiled.RData` contains the compiled EEMD results from all 200 trials for the S,D&T Lorentzian timeseries with observational variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts2_compiled.RData` contains the compiled EEMD results from all 200 trials for the S,D&T Lorentzian timeseries. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts3.b_compiled.RData` contains the compiled EEMD results from all 200 trials for the Pulse-Decay Lorentzian timeseries with environmental variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts3.c_compiled.RData` contains the compiled EEMD results from all 200 trials for the Pulse-Decay Lorentzian timeseries with observational variation. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`ts3_compiled.RData` contains the compiled EEMD results from all 200 trials for the Pulse-Decay Lorentzian timeseries. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`Potomac_compiled.RData` contains the compiled EEMD results from all 200 trials for the Potomac River. The results contain timeseries for each IMF. The file was produced by `EEMDComputation.Rmd.`

`EEMDConn_p1.png` … `EEMDConn_p17.png` contain image frames to make Video S2. Images are written by `MakeVideo2.Rmd`

`EEMDIch_p1.png` … `EEMDIch_p16.png` contain image frames to make Video S1. Images are written by `MakeVideo1.Rmd`

`EEMDPot_p1.png` … `EEMDPot_p16.png` contain image frames to make Video S3. Images are written by `MakeVideo3.Rmd`

---

## Supporting References


1. Priestley, M. B., *Spectral analysis and time series*. Academic Press: London ; New York, 1981.
2. Halley, J. M., Ecology, evolution and 1/f-noise. *Trends in Ecology & Evolution* **1996,** *11*, (1), 33-37.
3. Kirchner, J. W.; Neal, C., Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, (30), 12213-8.
4. Hensley, R. T.; Cohen, M. J.; Jawitz, J. W., Channel Filtering Generates Multifractal Solute Signals. *Geophys Res Lett* **2018,** *45*, (21), 11722-11731.
5. Lloyd, C. E. M.; Freer, J. E.; Collins, A. L.; Johnes, P. J.; Jones, J. I., Methods for detecting change in hydrochemical time series in response to targeted pollutant mitigation in river catchments. *J Hydrol* **2014,** *514*, 297-312.
6. Press, W. H., *Numerical recipes : the art of scientific computing*. Cambridge University Press: Cambridge ; New York, 1989; p xx, 702 p.
7. Thompson, S. E.; Katul, G. G., Multiple mechanisms generate Lorentzian and 1/f(alpha) power spectra in daily stream-flow time series. *Adv Water Resour* **2012,** *37*, 94-103.
8. Mallat, S. G., *A wavelet tour of signal processing*. 2nd ed.; Academic Press: San Diego, 1999; p xxiv, 637 p.
9. Katul, G. G.; Parlange, M. B., On the Active-Role of Temperature in Surface-Layer Turbulence. *J Atmos Sci* **1994,** *51*, (15), 2181-2195.
10. Torrence, C.; Compo, G. P., A practical guide to wavelet analysis. *B Am Meteorol Soc* **1998,** *79*, (1), 61-78.
11. Yamada, M.; Ohkitani, K., An Identification of Energy Cascade in Turbulence by Orthonormal Wavelet Analysis. *Prog Theor Phys* **1991,** *86*, (4), 799-815.
12. Katul, G. G.; Parlange, M. B.; Chu, C. R., Intermittency, Local Isotropy, and Non-Gaussian Statistics in Atmospheric Surface-Layer Turbulence. *Phys Fluids* **1994,** *6*, (7), 2480-2492.
13. Katul, G.; Lai, C. T.; Schafer, K.; Vidakovic, B.; Albertson, J.; Ellsworth, D.; Oren, R., Multiscale analysis of vegetation surface fluxes: from seconds to years. *Adv Water Resour* **2001,** *24*, (9-10), 1119-1132.
14. Katul, G.; Vidakovic, B.; Albertson, J., Estimating global and local scaling exponents in turbulent flows using discrete wavelet transformations. *Phys Fluids* **2001,** *13*, (1), 241-250.
15. Heffernan, J. B.; Cohen, M. J., Direct and indirect coupling of primary production and diel nitrate dynamics in a subtropical spring-fed river. *Limnol Oceanogr* **2010,** *55*, (2), 677-688.
16. Appling, A. P.; Read, J. S.; Winslow, L. A.; Arroita, M.; Bernhardt, E. S.; Griffiths, N. A.; Hall, R. O., Jr.; Harvey, J. W.; Heffernan, J. B.; Stanley, E. H.; Stets, E. G.; Yackulic, C. B., The metabolic regimes of 356 rivers in the United States. *Sci Data* **2018,** *5*, 180292.
17. Appling, A. P.; Hall, R. O.; Yackulic, C. B.; Arroita, M., Overcoming Equifinality: Leveraging Long Time Series for Stream Metabolism Estimation. *J Geophys Res-Biogeo* **2018,** *123*, (2), 624-645.
18. Webster, J. R.; Meyer, J. L., Stream organic matter budgets - Introduction. *J N Am Benthol Soc* **1997,** *16*, (1), 5-13.
19. Hall, R. O.; Beaulieu, J. J., Estimating autotrophic respiration in streams using daily metabolism data. *Freshwater Science* **2013,** *32*, (2), 507-516.
20. Hall, R. O.; Tank, J. L., Ecosystem metabolism controls nitrogen uptake in streams in Grand Teton National Park, Wyoming. *Limnol Oceanogr* **2003,** *48*, (3), 1120-1128.
21. Canfield, D. E.; Hoyer, M. V., Influence of Nutrient Enrichment and Light Availability on the Abundance of Aquatic Macrophytes in Florida Streams. *Can J Fish Aquat Sci* **1988,** *45*, (8), 1467-1472.

22. Webster, J. R.; Mulholland, P. J.; Tank, J. L.; Valett, H. M.; Dodds, W. K.; Peterson, B. J.; Bowden, W. B.; Dahm, C. N.; Findlay, S.; Gregory, S. V.; Grimm, N. B.; Hamilton, S. K.; Johnson, S. L.; Marti, E.; McDowell, W. H.; Meyer, J. L.; Morrall, D. D.; Thomas, S. A.; Wollheim, W. M., Factors affecting ammonium uptake in streams - an inter-biome perspective. *Freshwater Biology* **2003,** *48*, (8), 1329-1352.

23. Runkel, R. L., Solution of the advection-dispersion equation: Continuous load of finite duration. *J Environ Eng-Asce* **1996,** *122*, (9), 830-832.

24. Hensley, R. T.; Cohen, M. J., On the emergence of diel solute signals in flowing waters. *Water Resour Res* **2016,** *52*, (2), 759-772.

25. Fischer, H. B., Longitudinal dispersion and turbulent mixing in openchannel flow. *Annu. Rev. Fluid Mech.* **1973,** *5*, 59-78.

26. Bowman, D. C.; Lees, J. M., The Hilbert-Huang Transform: A High Resolution Spectral Method for Nonlinear and Nonstationary Time Series. *Seismol Res Lett* **2013,** *84*, (6), 1074-1080.