



An algorithm for detecting and quantifying disturbance and recovery in high-frequency time series

Jonathan A. Walter ^{1*}, Cal D. Buelo,¹ Alice F. Besterman,^{1,2,3} Spencer J. Tassone,¹ Jeff W. Atkins,^{4,5} Michael L. Pace ¹

¹Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia

²Woodwell Climate Research Center, Falmouth, Massachusetts

³Buzzards Bay Coalition, New Bedford, Massachusetts

⁴Department of Biology, Virginia Commonwealth University, Richmond, Virginia

⁵Center for Forest Watershed Science, Southern Research Station, USDA Forest Service, New Ellenton, South Carolina

Abstract

Determining when a disturbance has occurred, its severity, and when the system recovered, is important to numerous questions in the aquatic sciences. This problem can be conceptualized as the timing and degree of perturbation from a typical state, and when the system returns to that typical state. We present an algorithm for detecting disturbance and recovery designed for high-frequency time series, e.g., data produced by automated sampling devices in instrumented buoys and flux towers. The algorithm quantifies differences in the empirical cumulative distribution functions of moving windows over reference and evaluation periods, and is sensitive to changes in the mean, variance, and higher statistical moments. Tests on simulated data show it accurately identifies disturbance and recovery. Three case studies illustrate the application of our algorithm in different empirical settings. A case study on dissolved oxygen in a Florida, USA estuary following a hurricane identified the disturbance and recovery 73 d later. A case study on air temperature and net ecosystem exchange in the Florida everglades identified cold snaps coinciding with periods of reduced carbon uptake. A case study on rotifer abundance following zebra mussel invasion in the Hudson River, NY showed rotifer collapse following invasion and recovery over a decade later. Methods such as ours can improve understanding response to disturbance and facilitate comparative and synthetic study of disturbance impacts across ecosystems.

Understanding patterns of disturbance and recovery, which can be thought of in terms of a system being perturbed from, and returning to, a typical state, is central to basic and applied analysis of aquatic systems (Rykiel 1985; Pickett et al. 1989; Lake 2000; Yannarell et al. 2007). Fundamental to this research is determining when a disturbance occurred, its severity, and when the system recovered (Weathers et al. 2016). Major disturbances, such as severe storms like hurricanes and major ocean–atmosphere oscillations like the El Niño Southern Oscillation, can be easy to identify because they are discrete, drive extreme changes in ecosystem state variables, and have obvious manifestations in the appearance of the system. Yet moderate disturbances can still exert substantial controls

on ecosystems. For example, the erosion of coastal saltmarshes is driven mainly by wave action outside of major storm events (Wiberg et al. 2019). Quantification of changes in ecosystem state variables can improve detection of subtler effects that may not be readily apparent as a qualitative shift in ecosystem state. However, moderate disturbances are more difficult to characterize (Atkins et al. 2020), requiring adequate data and statistical procedures to detect.

Even for major disturbances, determining when a system has recovered is difficult. Aquatic systems are complex, with components that may respond to perturbation in different ways or over different timescales (Engstrom et al. 1985; Pickett et al. 1989; Wolin and Stoermer 2005; Yannarell et al. 2007). Furthermore, it can sometimes be difficult to distinguish natural modes of variability, including seasonal and diurnal patterns, from disturbance. In addition, disturbance can have multifaceted effects on ecosystem state variables, not only impacting the mean but also the variance and higher statistical moments (Scheffer and Carpenter 2003). To take one hypothetical example, following disturbance the mean productivity of a system might return to predisturbance values,

*Correspondence: jaw3es@virginia.edu

Additional Supporting Information may be found in the online version of this article.

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

but if the system has lost resilience, the variance in productivity might remain elevated because the system responds more strongly to minor perturbations than before the disturbance occurred (Scheffer and Carpenter 2003; Scheffer et al. 2015).

Modern sampling technologies have reduced barriers to detecting disturbance and recovery in ecological systems. Most germane to this study, high-frequency data, such as from automated sampling buoys (Hanson et al. 2016; Meinson et al. 2016; Pace et al. 2017), aquatic autonomous vehicles (de Lima et al. 2020), or eddy covariance towers (Novick et al. 2018), which capture detailed records of disturbance/recovery and natural variation, are increasing in availability and time series length. However, statistical methods that leverage these detailed data and apply systematic quantitative criteria to detect disturbance and recovery are needed. Systematic quantitative definitions of disturbance and recovery strengthen individual investigations by reducing ambiguity and subjectivity in the timing of events and in metrics of disturbance severity, which is especially important for low- and moderate-intensity disturbances that do not produce unambiguous qualitative changes in ecosystem state. Systematic quantitative definitions of disturbance and recovery also strengthen synthesis and meta-analysis, e.g., to compare how different ecosystems or ecosystem components respond to disturbance.

A variety of statistical approaches have been used to quantify ecological disturbance and/or recovery, or the related more general problem of detecting anomalous periods in time series (Pace et al. 2010; Jones and Lennon 2015; Hobday et al. 2016; Zwart 2017; Thayne et al. 2021). However, those of which we are aware tend to have one or more of the following shortcomings when applied to the problem of detecting disturbance and recovery in high-frequency time series of ecosystem variables. One inferential strategy not specific to a particular statistical method uses statistics post hoc to support that an ecosystem state variable was affected by a known event and possibly later recovered (e.g., Pace et al. 2010). Although useful, different approaches are needed when disturbance and recovery may be ambiguous, as can be case for weak or moderate disturbances, and when the causal event is not known a priori. Tools like changepoint detection (Killick and Eckley 2014), autoencoders (Li et al. 2020), and intervention analysis (Carpenter et al. 1989) can be used, without a priori knowledge of an event, to detect disturbances but may not be well-suited to identifying recovery, especially when recovery is gradual, and may be confounded by time series with complex patterns of natural variability. The latter is particularly important to our case since the short sampling interval of high frequency data means that the time series will reflect diurnal and seasonal patterns, and possibly other influences such as tides and periodic climate oscillations. In data with somewhat coarser temporal resolution (e.g., daily to weekly), events can be identified as periods of anomalously low or high values outside the typical range of variability

(Verbesselt et al. 2012; Hobday et al. 2016), but a more holistic picture of disturbance could also include periods, e.g., with suppressed variance and these would be undetected.

Here, we present an algorithm for quantifying disturbance and recovery in high-frequency time series exhibiting strong natural variation. We evaluate its performance on simulated time series, focusing on (1) its accuracy at detecting disturbances and how accuracy depends on disturbance severity and duration; (2) the frequency of false positives; and (3) the error in estimated recovery dates. In addition, we illustrate the application of the algorithm in empirical settings by presenting three case studies in which we apply the algorithm to time series from different ecosystems, focal variables, and sampling technologies.

Procedures

To give an overview, our algorithm identifies disturbance and recovery by quantifying the difference between the empirical distribution of a time series variable observed within moving windows and a reference distribution representing the typical state of the system. It then quantifies how atypical these differences are by benchmarking them against the distribution of differences observed in moving windows of the same width during the reference period. An increase in the difference from the reference state past a specified threshold defines the onset of disturbance, as detected in our algorithm, and a return toward the reference state, past another (possibly different) threshold, defines recovery. In addition, we optionally filter detections by specifying a minimum duration for an excursion from (return toward) the reference state to be considered a disturbance (recovery). A general outline of our algorithm is given in Table 1. Software implementing this algorithm is available in the “disturbhf” package for the R environment for statistical computing (R Core Team 2019) on GitHub <<https://github.com/jonathan-walter/disturbhf>>. Installation instructions are provided in Supporting Information Material S1.

More specifically, we define an evaluation time series, x_{eval} over which we test for disturbance through comparison to a reference time series, x_{ref} . In temporal moving windows of

Table 1. Outline of procedures for implementing our algorithm.

Step
1. Define reference period, test period, and test window width
2. Get ECDF for reference period
3. Find mean and standard deviation of d_w for test-width moving windows in reference period
4. Quantify d_w for moving windows in test period
5. Express d_w during test period as a Z score benchmarked against reference period
6. Assign disturbance and recovery based on thresholds of Z score
7. (optional) filter to remove, e.g., transient disturbances

specified width, we compute the empirical cumulative distribution function (ECDF) of observations within each moving window, and compare it to the ECDF of the reference period by evaluating it over a range spanning the joint minima and maxima of x_{eval} and x_{ref} and computing the difference statistic d_w . We provide two ways to compute d_w . In the main text, we consider d_w as the maximum difference between the ECDFs: $d_w = \max(X_w - X_{\text{ref}})$. Here, X_w and X_{ref} are the ECDFs for the test window and the reference period, respectively, evaluated over some common sequence of values spanning the joint minima and maxima of x_w and x_{ref} , with interval small relative to the range. This is similar to the Kolmogorov–Smirnov test (Kolmogorov 1933; Smirnov 1948). Figure 1 illustrates how d_w differs between a disturbed and undisturbed portion of a synthetic time series. In Supporting Information Material S2 we additionally consider d_w as the integral of the absolute difference between X_w and X_{ref} and provide parallel results showing that both versions of d_w perform similarly.

We provide two options for defining the reference period: (1) using the reference period in its entirety, or (2) using a seasonally adaptive reference period. In both cases, the user specifies a time series reflecting the typical state of the system over

some period of time. This might include, e.g., data from the site in one or more prior years, or contemporary data from a similar nearby undisturbed location. The evaluation and reference time series can also be the same; note that this will identify anomalous periods in the time series, but will have reduced sensitivity to identify disturbances. The simplest way to use the reference data is to consider its entirety, but this can be ineffective for disturbances that are overlain on baselines with strong natural variation (e.g., diurnal or seasonal cycles). Consequently, we also implemented an adaptive reference period, in which the reference time series X_{ref} is determined by indexing the provided reference time series by day-of-year and subsetting a window (of user-defined width) centered on the corresponding day-of-year of X_{test} . The reference window will typically be wider than the test window by several times; see Supporting Information Material S3 for a quantitative investigation of the effect of reference window width on algorithm performance. We refer to this as the “seasonally adaptive” reference period given the cases considered by this study, but the approach could generalize to coarser or finer timescales given appropriate data.

We next quantify the magnitude of departure for moving windows during the evaluation period from typical conditions

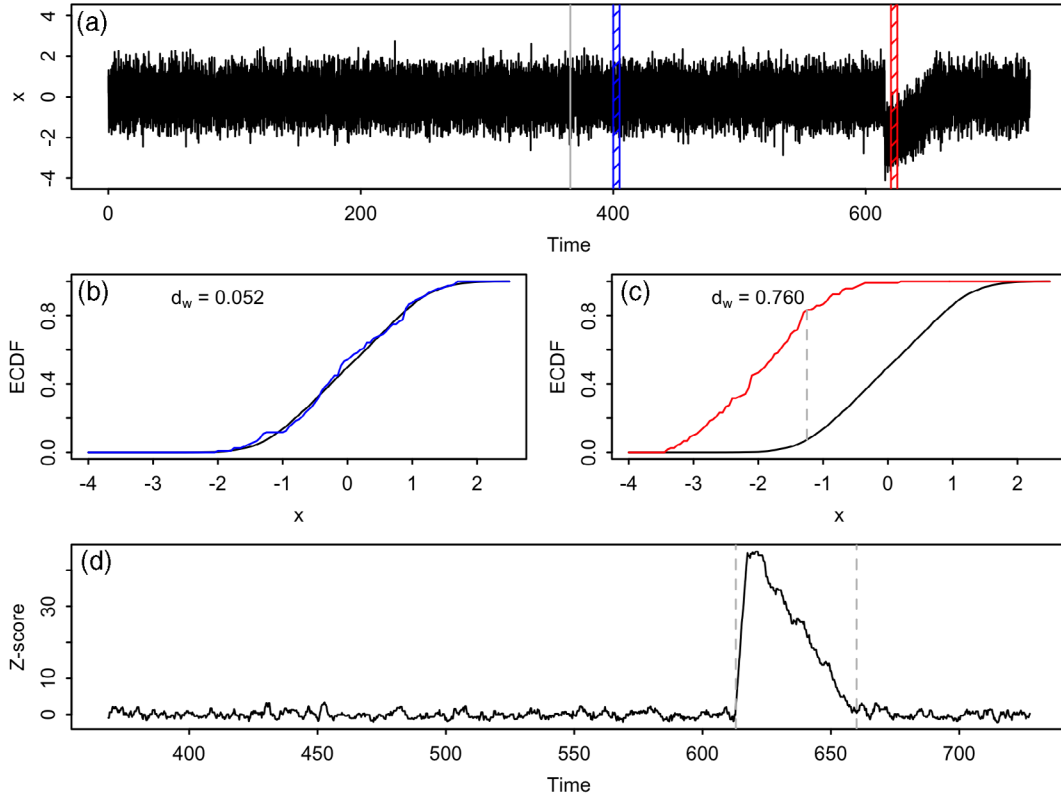


Fig. 1. Illustration of disturbance statistic using a synthetic time series. In (a) the reference period is shown to the left of the vertical gray line; the evaluation period is to the right. The blue hatched area shows an undisturbed 5-d evaluation window and the red shows a 5-d evaluation window following disturbance. Panels (b) and (c) illustrate how the difference statistic d_w is computed from empirical cumulative distribution functions for the reference period and evaluation window. As above, blue and red show windows without and with disturbance, respectively. Panel (d) shows the time series of Z scores for d_w over the test period, with dashed gray lines indicating the detected onset of disturbance and recovery.

using a Z score. To do so, we apply the moving window analysis over the reference time series to characterize the distribution of d_w while the system is “normal” via its sample mean, $\hat{\mu}_{d_{\text{ref}}}$, and standard deviation, $\hat{\sigma}_{d_{\text{ref}}}$. We compute Z scores for all d_w during the test period as $Z_w = (d_w - \hat{\mu}_{d_{\text{ref}}}) / \hat{\sigma}_{d_{\text{ref}}}$. Recall that the Z score for an observation quantifies in units of standard deviations how far that observation is from the mean. Our statistic Z_w is modified from a standard Z score in that the mean and standard deviation are taken from the reference period, not the full time series.

Finally, we apply a disturbance and recovery “alarm system” that uses user-specified thresholds on Z_w for disturbance and recovery to assign whether excursions in d_w during the test period are sufficiently large and returned to normal to qualify, respectively, as disturbances and recoveries. Additional requirements can be imposed, such as requiring the system to remain under the recovery threshold for a period of time, to add confidence that the system has recovered, not merely passed quickly and transiently through a distribution similar to the predisturbance state. We provide an example of such an algorithm. Taking the alarm output, we filter it by first merging together disturbances separated by short recoveries, where short recoveries are those less than a user-specified minimum threshold. We then ignore disturbances shorter than a second, possibly different, user-specified threshold.

The primary assumption made by this algorithm is that the reference period effectively describes all important processes driving variation in the focal variable, except for the disturbance. By benchmarking changes in the evaluation period against data in the reference period, the algorithm avoids making explicit assumptions about attributes of the data, such as its distribution and autocorrelation structure, that are common to statistical tests for difference. The cost of avoiding such assumptions is that relatively large amounts of data that are temporally dense relative to the duration of the disturbance and relative to other important sources of variability, such as diurnal, tidal, or seasonal cycles, are needed.

We also caution that changes in time series behavior that violate this algorithm’s primary assumption, such as may be due to long-term trends or that are unrelated to the disturbance of interest, can result in inaccurately defined disturbances or misattribution of disturbance effects. Environmental time series often feature trends and nonstationarity, e.g., due to climate change. Statistical pretreatment of the data time series, such as removing long-term trends, should improve disturbance detection where such phenomena are present. Since many approaches to disturbance detection used in the environmental sciences identify observations or patterns of observations that are outside norms for that system, uncertainty in attributing specific mechanisms to anomalous observations is a challenge common to many different approaches to this problem, including the algorithm presented here. In some cases, specific mechanisms of disturbance could have characteristic patterns that machine learning algorithms could be

trained to detect given adequate training data (Kennedy et al. 2015, Kong et al. 2018). Alternatively, more process-based models of variation in an undisturbed system could better represent changes in a nonstationary system that are driven by nondisturbance processes. Disturbance detection based on a process-based model can plausibly be implemented on lower-volume data time series but requires greater a priori knowledge of the system that must generally come from previous studies and independent data.

Assessment

Simulation tests: Procedures

To test the performance of our methods for detecting disturbance and recovery, we applied them to simulated time series representing two idealized patterns of disturbance overlain on two idealized patterns of background natural variation. We considered “wedge-shaped” disturbances, in which the disturbance manifest as a sudden drop in the time series, followed by a linear return to its normal state (Fig. S1). We also considered “S-shaped” disturbances, in which the disturbance manifest as a continuous decline in the time series, followed by an increase through the mean of the normal state, before peaking and returning to its normal state (Fig. S1). An example of a qualitatively S-shaped disturbance patterns is estuary salinity during hurricanes: storm surge initially causes salinity to increase above background levels, but later an influx of freshwater from precipitation and flooding causes salinity to decrease below background levels. We expected S-shaped disturbances to be more challenging for the algorithm to detect. The patterns of background variation we considered were “flat,” in which observations fluctuated around a single mean; and “seasonal” in which a sine wave with a 365-day period was used to generally approximate seasonal variation (Fig. S1). All cases used hourly time steps and included diurnal variation, simulated as a sine wave with a 24-h cycle, and Gaussian noise. We considered disturbances in both the positive and negative direction.

For our simulation tests, we considered time series where the seasonal component (if present) had an amplitude of 0.5, the diurnal component had an amplitude of 1, and the Gaussian noise had mean = 0 and standard deviation = 0.5. We varied the severity and duration of disturbances by drawing them randomly and independently from uniform distributions. Severity ranged from 0 to 5 times the standard deviation of the time series before the disturbance was applied. Duration ranged from 3 to 60 d. For each of 16 scenarios [2 background variation types (flat and seasonal) \times 4 disturbance types (negative wedge, negative S, positive wedge, positive S) \times 2 reference period types (regular and seasonally adaptive)] we conducted 5000 simulations in which severity and duration were selected independently, randomly, and uniformly on the intervals stated above. Simulated time series for algorithm tests contained 2 years of hourly data, with the disturbance

occurring in the second year. The first simulated year was used as the reference time series. The width of the evaluation window was 5 d and the width of the adaptive reference window was 60 d. The disturbance and recovery thresholds were $Z_w = 2$ and $Z_w = 0.5$, respectively. These values were chosen because they ensure that detected disturbances are rare events with substantial changes from typical conditions and that detected recoveries represent a substantial shift back toward typical conditions. Exploration of the method on simulated and empirical data suggest that these values generally balanced errors associated with false positives and false negatives. In some cases, increasing or decreasing these values will be appropriate.

To evaluate the performance of our method at detecting disturbances in simulated data, we focused on the true detection rate and the false detection rate, and how these depend on the severity and intensity of disturbance. We considered the true detection rate to be the proportion of simulations in which the algorithm produced an alarm during the true disturbance, with a tolerance equal to the width of the moving window for the test period. We considered the false detection rate to be the number of alarms per simulation that were not during a true disturbance, inclusive of the tolerance window, divided by the number of moving windows that were evaluated. In addition, we considered the degree of error in reported disturbance and recovery dates.

In addition to the tests shown in the main text of the manuscript, we conducted simulation experiments to compare two ways of computing d_w (Supporting Information Material S2), to evaluate the importance of the seasonally adaptive reference window width (Supporting Information Material S3), and to compare the performance of our algorithm to a disturbance and recovery detection procedure based on the CUSUM method (Page 1954; Supporting Information Material S4). These supplements describe methodological details and results.

Simulation tests: Results

The rate of true disturbance detections was high, overall, averaging 80–98%, but depended on disturbance severity and duration (Fig. 2). Across disturbance types, the lowest true detection rates were experienced with short-duration and low-severity disturbances, and were as low as 7–71% depending on simulation scenario and whether the adaptive reference window was used. When simulated time series featured sinusoidal seasonal variation, using the adaptive reference window improved performance by as much as 84%, particularly for detection of short, low-severity disturbances (Fig. 2). However, increases in the true detection rate using the adaptive reference window came at the cost of a higher rate of false positives. Although the false positive rate did not depend on disturbance severity or duration (Fig. S2), using the adaptive reference window the number of false positives was as much as 20 times greater. The recovery date was generally

well-estimated, as across simulation sets the great majority of estimated recovery dates were within 5 d (the width of the test window) of the true recovery date (Fig. 3). However, recovery for long, low-severity disturbances was often estimated before true recovery (Fig. 3). Filtering the disturbance detections by combining detected disturbances separated by short transient detected recoveries (<3 d), and by omitting very short detected disturbances (<3 d), tended to reduce the false positive rate (Fig. S3) and to modestly reduce error in the estimated recovery date (Fig. S5) at the expense of modest decreases in the true detection rate (Fig. S4).

As expected, results were identical (to within sampling variation) regardless of the direction of change associated with disturbance, so only results for disturbances in the negative direction are shown in the main text; parallel results for positive disturbances are shown in Supporting Information Material Figs. S6, S7.

Case studies: Procedures

We present three case studies illustrating how our algorithm can be applied to empirical data. These case studies were selected because they represent different kinds of focal variables, ecosystems, and data types.

Our first case study examines the impact of a major hurricane on patterns of dissolved oxygen (DO) dynamics in Apalachicola Bay, Florida, USA. Water quality in Apalachicola Bay is monitored by the National Estuarine Research System's System-Wide Monitoring Program (NOAA National Estuarine Research Reserve System [NERRS] 2020), which collects high-frequency (15 min) data on several variables, including DO saturation. We focused on DO because it is a broad indicator of biological activity that had low missingness during a >2-year period from December 2016 to February 2019, including in the days following a major hurricane that knocked out some other sensors. Hurricane Michael made landfall on the Gulf coast of Florida as a Category 5 hurricane on October 10, 2018. We applied our algorithm using January 1, 2018 to separate the time series into reference and test periods, a test window width of 5 d, and an adaptive reference window width of 60 d. The threshold for disturbance was a $Z_w = 3$ and the threshold for recovery was a $Z_w = 0.5$. Here, we used a higher disturbance threshold, $Z_w = 3$, because we were a priori interested in a very large disturbance event and wanted to limit the number of detected disturbances not associated with Hurricane Michael.

Our second case study investigates the co-occurrence of cold snaps with net ecosystem exchange (NEE) anomalies using AmeriFlux eddy covariance data from a mangrove forest at Shark River Slough in the everglades of Florida, USA (Barr et al. 2010; Fuentes 2016). The site is exposed to semi-diurnal tidal inundation, hence aquatic productivity and respiration along with forest metabolism, contribute to observed NEE patterns. This case study also supports our objective of evaluating our method on a variety of data types. Previously, Kominoski

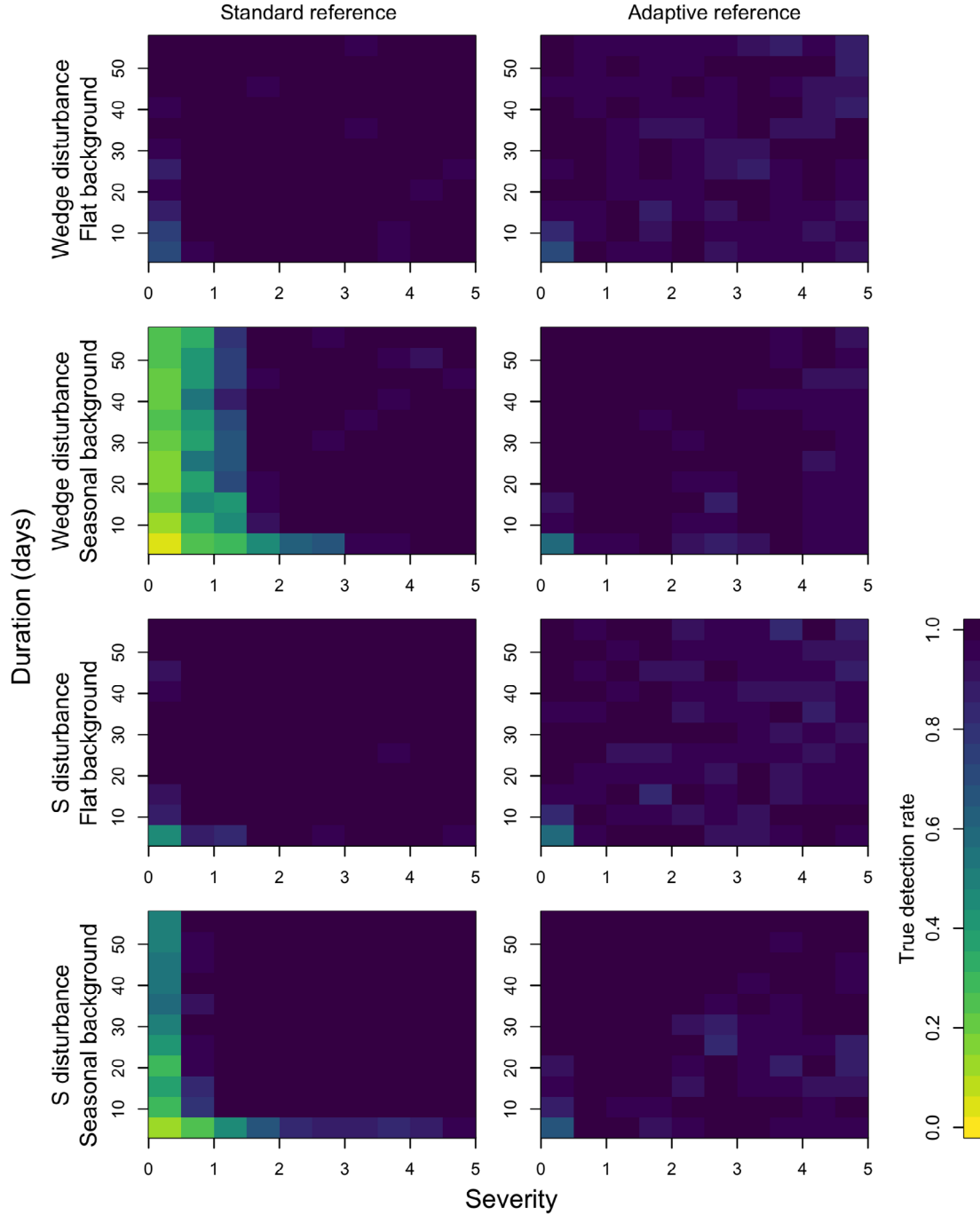


Fig. 2. Rates of true positives for “negative” disturbance simulations, in which disturbance reduced the value of the variable in question.

et al. (2020) showed effects of extreme cold snaps in winter 2009–2010 had substantial impacts on this ecosystem. We used half-hourly air temperature and NEE data from January 1, 2006 to June 30, 2010, using observations prior to July 1, 2009 as the reference data and tested for disturbances in the subsequent observations. We applied the disturbance algorithm independently to each variable, asking whether periods of particularly low temperature coincided with periods with

atypical NEE. For both variables, the threshold for disturbance was $Z_w = 2$ and for recovery, $Z_w = 0.5$.

Our third case study investigates changes in the abundance of rotifers in the Hudson River, New York, USA following invasion by zebra mussels. The first two case studies illustrate typical applications of our method, while this third case study illustrates one way it can be extended for application to data that are not conventionally high frequency, but for which the

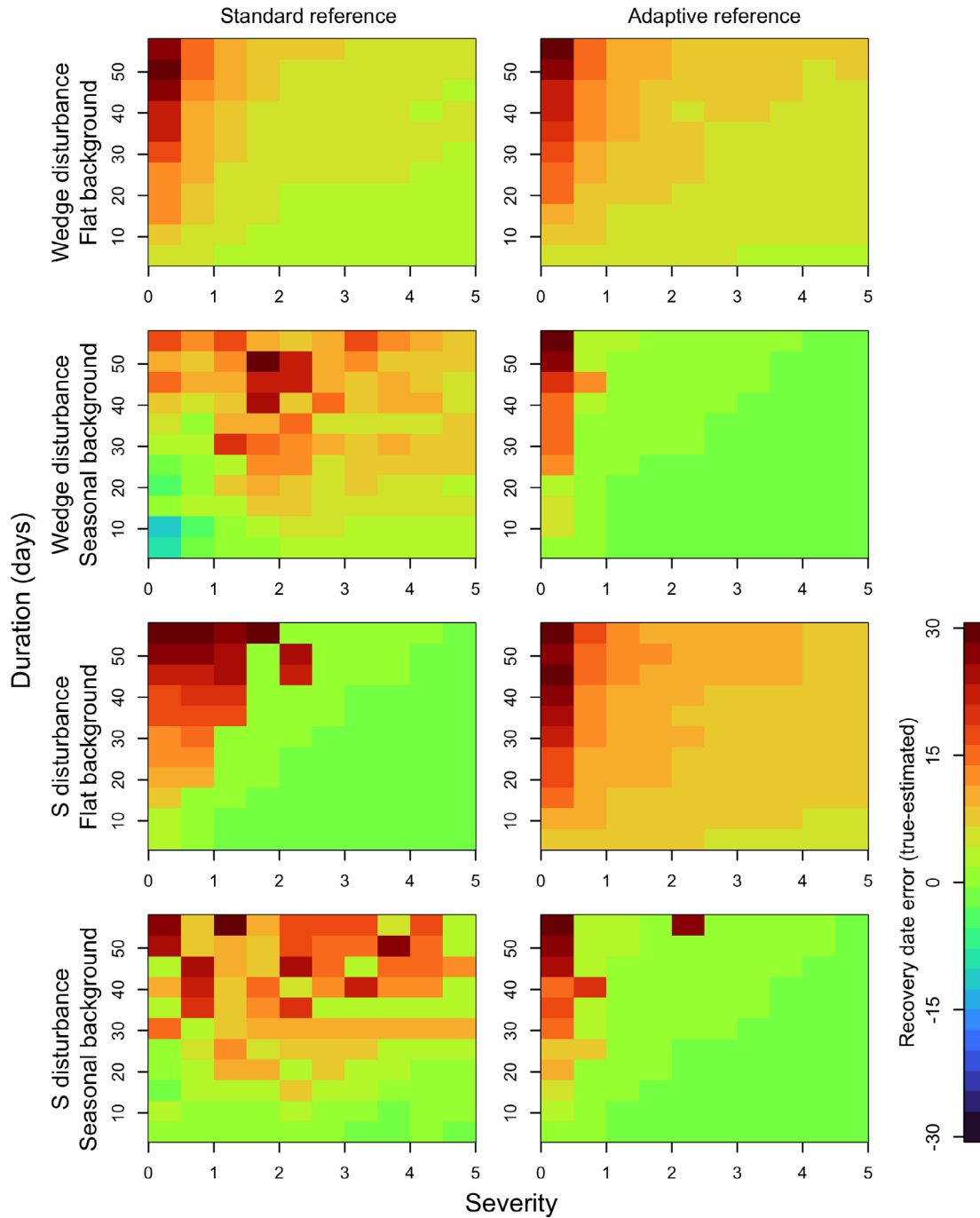


Fig. 3. Error (number of days) in estimated recovery date in simulated disturbance time series, in which disturbance reduced the value of the variable in question. Positive values indicate that recovery was detected by the algorithm before the true date.

time series is long and the sampling interval is dense relative to the temporal scale of a significant change in the system. Zebra mussels were first observed in the Hudson River in 1991 and became dominant by fall 1992 (Strayer et al. 2020), causing a crash in microzooplankton abundance; rotifer abundance later recovered following changes in zebra mussel size structure (Pace et al. 2010). We applied our algorithm to a

1987–2015 time series of rotifer abundance (number/L) collected using methods described in Pace et al. (2010). Samples were taken approximately every 2 weeks during the ice-free season, typically April–November. The data are publicly available (Cary Institute of Ecosystem Studies and Strayer 2016). Observations spanning 1987–1991 were used as the reference period, the width of test window was 18 months, and no

adaptive reference window was used in this case. The threshold for disturbance was $Z_w = 2$ and for recovery, $Z_w = 0.5$.

Case studies: Results

Our case study examples illustrate that the method can be applied to a range of data types, systems, and study questions, including some beyond the ideal use case of high frequency sensor data, provided that the timescale of a phenomenon of interest is long relative to the sampling frequency.

In the case study of hurricane Michael's impact on DO in Apalachicola Bay, FL, the algorithm signaled a disturbance immediately after landfall (landfall on October 10, 2018; disturbance detected on October 11, 2018), an interval much smaller than the test window width of 5 d (Fig. 4). Recovery was detected approximately 73 d later, on December 23, and visual inspection suggests that at this time both the mean and variance had returned to qualitatively similar values to the same time the previous year (Fig. 4). Our algorithm also detected four other periods of anomalous DO values, three of which were periods of low DO. The causes of these drops in DO are not known, but could include other storm events, as named tropical storms Alberto and Gordon also impacted the Gulf Coast of Florida earlier that season. The maximum Z_w (4.42) during the disturbance corresponding to Hurricane Michael was the greatest of these events, corroborating that this was the most severe disturbance during the evaluation period.

In the case study concerning coincidence of cold snaps and anomalous NEE in Shark River Slough, FL, notable cold periods in January and February–March 2010 coincided with periods of distinctly low carbon uptake (Fig. 5). We also detected a third, less extreme cold snap in late March 2010, that was not associated with a change in NEE. There were also

two detected periods of elevated carbon uptake in August and September 2009. Although the drivers of these changes are not known, detecting disturbances associated with both increases and decreases in NEE is consistent with our simulation studies that showed no difference in detections between positive and negative disturbances. In addition, this example highlights how our method can detect disturbances manifesting as distributional changes that do not move the mean outside the system's historical range of variability (Fig. 5).

In the case study of rotifer abundance in the Hudson River following zebra mussel invasion, we found that shortly after zebra mussels became dominant in fall 1992, a disturbance was detected associated with a decline in rotifer abundance that persisted through the mid-2000s (Fig. 6). In August 2006, we detected a recovery of rotifer abundance, similar to an earlier finding by Pace et al. (2010). Despite the lower frequency of these data, our algorithm recovered a disturbance–recovery pattern consistent with known changes in the system.

Discussion

Our algorithm for detecting disturbance and recovery in high-frequency time series data had substantial skill, identifying disturbances and recoveries across a range of simulated and empirical cases. Although admittedly lacking some complexities of empirical time series data, our simulation tests evaluated algorithm performance while directly manipulating patterns of disturbance and background “natural” variability. Disturbances and recoveries were identified with high accuracies, with errors increasing when disturbances had very low severity or were shorter than the width of the evaluation window. In addition, uncertainty in recovery dates was generally within the width of the moving window, but could increase

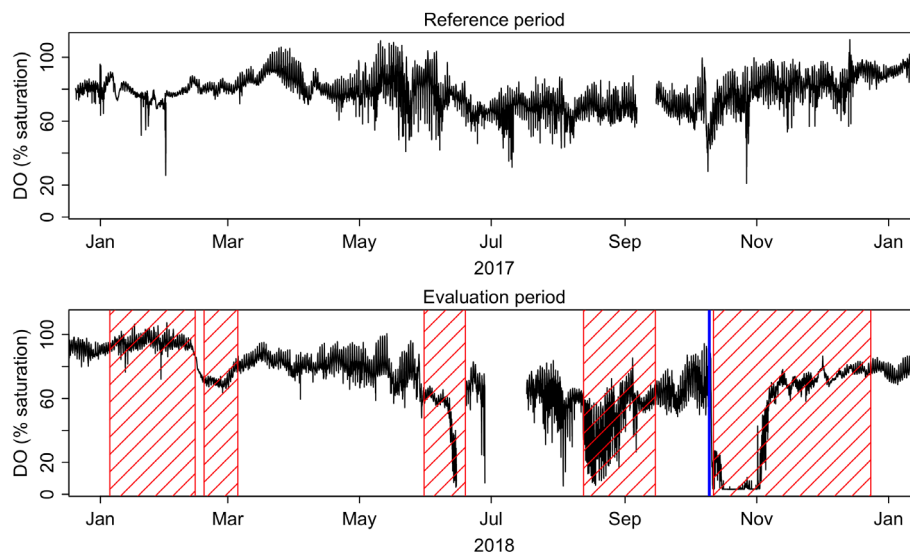


Fig. 4. Empirical case study of dissolved oxygen saturation in Apalachicola Bay, FL. Areas hatched in red indicate disturbances detected using our algorithm; the blue vertical line indicates landfall of Hurricane Michael on October 10, 2018.

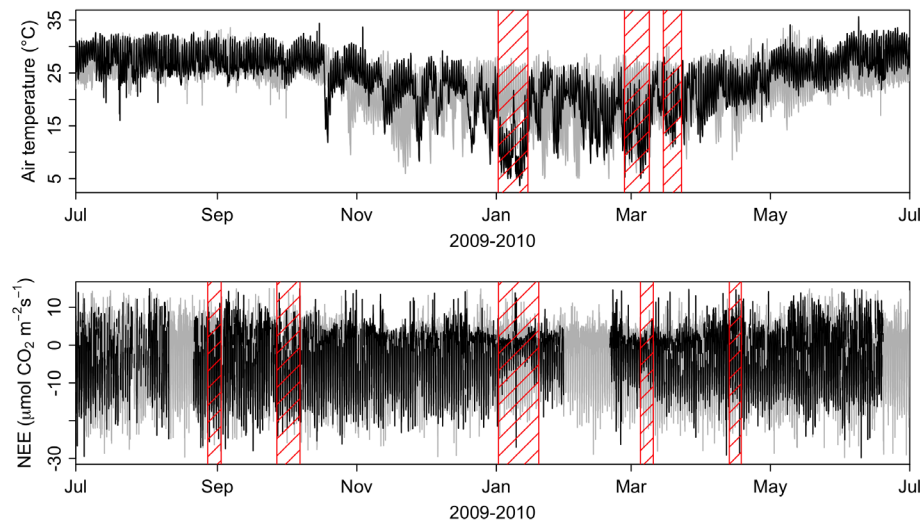


Fig. 5. Empirical case study of coincidence of cold snaps and NEE anomalies at Shark River Slough, FL. The black time series gives values of air temperature and NEE during the test period; gray shading indicates the 0.25–0.75 quantiles of hourly temperatures across the reference time series. Areas hatched in red indicate disturbances detected by our algorithm. January and March 2010 cold snaps coincide with reduced carbon uptake. Negative values of NEE represent carbon uptake.

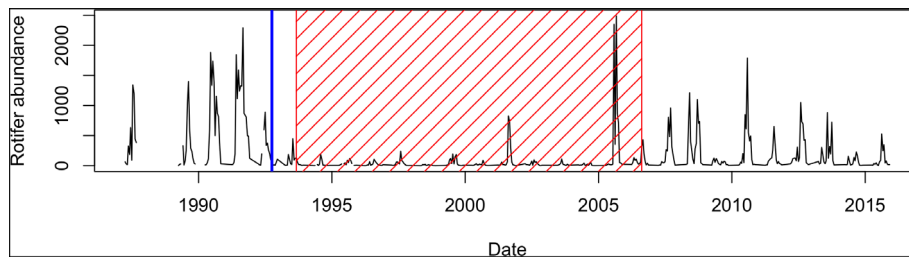


Fig. 6. Empirical case study of effects of zebra mussel invasion on rotifer abundance in the Hudson River, NY. The red hatched area indicates the disturbance detected by our algorithm; the blue vertical line is placed at October 1, 1992 as zebra mussels became dominant in the Hudson river by fall 1992.

for long disturbances in which the system slowly approached baseline conditions. The case studies illustrate applications of this method to different types of empirical data spanning ecosystem types. While data from high frequency sampling buoys (e.g., the Apalachicola Bay case study) and flux towers (e.g., the Shark River Slough case study) are envisioned as major use cases for this method, we also demonstrated how the method could be applied to data with lower sampling frequency when the length of disturbance is also long relative to the sampling frequency.

The approach has strengths and weaknesses. A notable strength is that our method is simultaneously sensitive to any aspect of statistical distributions, including changes in the mean, variance, skewness, and kurtosis. Other change detection approaches often focus on only one, typically the mean, but our method can detect subtler changes as illustrated in the Shark River Slough case study (Fig. 5). However, this advantage comes at the cost of high data requirements, which in many cases can only be achieved with high-frequency data from automated sensors. Although some applications in

particularly long time series are possible, as shown in the Hudson River rotifer case study, the density of observations relative to the temporal scale of the change we observed was still quite high and the changes to the system were unambiguous. There are also tradeoffs to benchmarking differences between test and references periods in terms of ECDFs and standard deviations. We used this approach to create a generally applicable method that could be applied when exactly what values or patterns in the focal variable constitute disturbance is unclear, as is often this case. However, there will tend to be false positives, simply because any distribution contains values that are extremes relative to the rest of that distribution. We observed that these false positives tended to be short-lived excursions away from reference conditions, and these can be reduced by increasing the threshold for disturbance, defining a minimum duration for disturbances of interest, or identified through post hoc interpretation.

Our investigations of this disturbance detection method yield some practical recommendations for investigators. The method performed poorly when the duration of disturbance

was less than the width of the test window, but there must be enough observations in the test window to effectively define the ECDF, and the test window should generally be wide enough to encompass repeated iterations of natural cycles with short periods, such as days or tides. Exactly how wide the test window should be depends on the sampling frequency, the complexity of natural underlying variation in the data, and of the complexity of ecosystem response to disturbance the investigator seeks to detect. In the examples presented here, the number of observations in the test window ranged from 24 (Hudson River rotifers) to 480 (NERRS data). The simulation studies featured 120 observations per test window, and were wide enough to contain multiples of the major short-timescale (diurnal) “natural” variations that were unrelated to disturbance, and so not of interest, but represented an important pattern of variation in these simulated cases. Making the test window wider than diurnal variation ensured that the algorithm did not see diurnal patterns as a source of difference in the time series. Since more observations are required to accurately characterize higher statistical moments than the mean, we infer that the Hudson River rotifer case study, while recovering a result that makes sense in context, was based largely on changes in the mean and is at the lower limit of data necessary to use our method. However, this suggests promise for applying our approach, or an adaptation thereof, to aquatic ecosystem time series derived from satellite remote sensing (e.g., Khandelwal et al. 2020).

Real-world applications sometimes have missing data, e.g., Figs. 4–6. The algorithm is designed to bypass time windows with insufficient nonmissing observations and can detect disturbance immediately following a period of missing data. For relatively short periods with missing data, it may be appropriate to widen the test window so that the missing period is smaller relative to the width of the test window, but we caution that the probability of a false negative increases when the width of the test window is greater than the period between disturbance and recovery.

We also found that use of the adaptive reference window was important for detecting shorter and lower-severity disturbances in time series with a seasonal pattern, but its use came with an increased false positive rate. Many ecosystem variables feature at least subtle seasonal patterns, and particularly where seasonality is strong this trade-off is likely worth it. Seasons are expressed differently in different climates (e.g., temperate vs. Mediterranean climates), and the width of the adaptive reference window can be tailored to this. The width of the reference window should be greater than that of the test window, typically by several times, up to the length of a season (Supporting Information Material S3).

Systematic quantitative tools for detecting disturbance and recovery are needed to better understand responses, and to facilitate quantitative syntheses across datasets or studies of marine and inland waters. Even though aspects of disturbance regime are well-quantified for major discrete events like inland

floods, reef bleaching, and coastal hurricanes (Paerl et al. 2006; Baker et al. 2008), other perturbations are often more subtle, and even when the disturbance event is unambiguous, responses can be complex. Quantitative synthesis has proven powerful for answering questions about how pattern and process differ across space, time, and system or variable of interest, thereby leading to new understandings (Carpenter et al. 2009). Tools such as the one presented here have potential to facilitate a synthetic analysis of disturbance and recovery, of potentially great importance in the context of increasing severe weather and natural disasters related to climate change.

Data and availability statement

All data used in this study are already publicly available via the referenced sources. Data files, simulation, and analysis code used in this study are publicly available on Zenodo (doi: 10.5281/zenodo.6472554) <https://github.com/jonathan-walter/disturbhf-test-ms>, and an R package implementing our algorithm is available at <https://github.com/jonathan-walter/disturbhf>. The version of the R package used to produce this manuscript is archived on Zenodo (doi: 10.5281/zenodo.6472546). <https://github.com/jonathan-walter/disturbhf>

References

- Atkins, J. W., and others. 2020. Application of multi-dimensional structural characterization to detect and describe moderate forest disturbance. *Ecosphere* **11**: e03156. doi:10.1002/ecs2.3156
- Baker, A. C., P. W. Glynn, and B. Riegl. 2008. Climate change and coral reef bleaching: An ecological assessment of long-term impacts, recovery trends and future outlook. *Estuar. Coast. Shelf Sci.* **80**: 435–471. doi:10.1016/j.ecss.2008.09.003
- Barr, J. G., V. Engel, J. D. Fuentes, J. C. Zieman, T. L. O'Halloran, T. J. Smith, and G. H. Anderson. 2010. Controls on mangrove forest-atmosphere carbon dioxide exchange in western everglades national park. *J. Geophys. Res.: Biogeo.* **115**: G2. doi:10.1029/2009JG001186
- Carpenter, S. R., T. M. Frost, D. Heisey, and T. K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology* **70**: 1142–1152. doi:10.2307/1941382
- Carpenter, S. R., and others. 2009. Accelerate synthesis in ecology and environmental sciences. *Bioscience* **59**: 699–701. doi:10.1525/bio.2009.59.8.11
- Cary Institute of Ecosystem Studies, and D. L. Strayer. 2016. Hudson River data (Kingston, biweekly). Knowledge Network for Biocomplexity knb.766.2.
- Engstrom, D. R., E. B. Swain, and J. C. Kingston. 1985. A palaeolimnological record of human disturbance from Harvey's Lake, Vermont: Geochemistry, pigments and

- diatoms. *Freshw. Biol.* **15**: 261–288. doi:[10.1111/j.1365-2427.1985.tb00200.x](https://doi.org/10.1111/j.1365-2427.1985.tb00200.x)
- Fuentes, J. D. 2016. AmeriFlux US-Skr Shark River Slough (tower SRS-6) Everglades, Ver. 1-1, AmeriFlux AMP (dataset). doi:[10.17190/AMF/1246105](https://doi.org/10.17190/AMF/1246105)
- Hanson, P. C., K. C. Weathers, and T. K. Kratz. 2016. Networked lake science: How the Global Lake Ecological Observatory Network (GLEON) works to understand, predict, and communicate lake ecosystem response to global change. *Inland Waters* **6**: 543–554. doi:[10.1080/IW-6.4.904](https://doi.org/10.1080/IW-6.4.904)
- Hobday, A. J., L. V. Alexander, S. E. Perkins, and others. 2016. A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.* **141**: 227–238. doi:[10.1016/j.pocean.2015.12.014](https://doi.org/10.1016/j.pocean.2015.12.014)
- Jones, S. E., and J. T. Lennon. 2015. A test of the subsidy–stability hypothesis: The effects of terrestrial carbon in aquatic ecosystems. *Ecology* **96**: 1550–1560. doi:[10.1890/14-1783.1](https://doi.org/10.1890/14-1783.1)
- Kennedy, R. E., Z. Yang, J. Braaten, C. Copass, N. Antonova, C. Jordan, and P. Nelson. 2015. Attribution of disturbance agent change from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sens. Environ.* **166**: 271–285. doi:[10.1016/j.rse.2015.05.005](https://doi.org/10.1016/j.rse.2015.05.005)
- Khandelwal, A., R. Ghosh, Z. Wei, H. Kuang, H. Dugan, P. Hanson, A. Karpatsne, and V. Kumar. (2020). RealSAT: A new reservoir and lake surface area timeseries dataset created using machine learning and satellite imagery. Available from <https://hdl.handle.net/11299/216044>.
- Killick, R., and I. A. Eckley. 2014. changepoint: An R package for changepoint analysis. *J. Stat. Soft.* **58**: 1–19. doi:[10.18637/jss.v058.i03](https://doi.org/10.18637/jss.v058.i03)
- Kolmogorov, A. 1933. Sulla determinazione empirica di una legge di distribuzione. *Giorn. Inst. Ital. Attuari.* **4**: 83–91.
- Kominoski, J. S., and others. 2020. Disturbance legacies increase and synchronize nutrient concentrations and bacterial productivity in coastal ecosystems. *Ecology* **101**: e02988. doi:[10.1002/ecy.2988](https://doi.org/10.1002/ecy.2988)
- Kong, Y.-L., Q. Huang, C. Wang, J. Chen, J. Chen, and D. He. 2018. Long short-term memory neural networks for online disturbance detection in satellite image time series. *Remote Sens.* **10**: 452. doi:[10.3390/rs10030452](https://doi.org/10.3390/rs10030452)
- Lake, P. S. 2000. Disturbance, patchiness, and diversity in streams. *J. North Am. Benthol. Soc.* **19**: 573–592. doi:[10.2307/1468118](https://doi.org/10.2307/1468118)
- Li, H., C. Wang, J. T. Ellis, Y. Cui, G. Miller, and J. T. Morris. 2020. Identifying marsh dieback events from Landsat image series (1998–2018) with an autoencoder in the NIWB estuary, South Carolina. *Int. J. Digit. Earth* **13**: 1467–1483. doi:[10.1080/17538947.2020.1729263](https://doi.org/10.1080/17538947.2020.1729263)
- de Lima, R. L. P., F. C. Boogaard, and R. E. de Graaf-van Dinther. 2020. Innovative water quality and ecology monitoring using underwater unmanned vehicles: Field applications, challenges and feedback from water managers. *Water* **12**: 1196. doi:[10.3390/w12041196](https://doi.org/10.3390/w12041196)
- Meinson, P., A. Idrizaj, P. Nöges, T. Nöges, and A. Laas. 2016. Continuous and high-frequency measurements in limnology: History, applications, and future challenges. *Environ. Rev.* **24**: 52–62. doi:[10.1139/er-2015-0030](https://doi.org/10.1139/er-2015-0030)
- NOAA National Estuarine Research Reserve System (NERRS). 2020. System-wide monitoring program.
- Novick, K. A., J. A. Biederman, A. R. Desai, M. E. Litvak, D. J. P. Moore, R. L. Scott, and M. S. Torn. 2018. The AmeriFlux network: A coalition of the willing. *Agric. For. Meteorol.* **249**: 444–456. doi:[10.1016/j.agrformet.2017.10.009](https://doi.org/10.1016/j.agrformet.2017.10.009)
- Pace, M. L., D. L. Strayer, D. Fischer, and H. M. Malcom. 2010. Recovery of native zooplankton associated with increased mortality of an invasive mussel. *Ecosphere* **1**: 1–10. doi:[10.1890/ES10-00002.1](https://doi.org/10.1890/ES10-00002.1)
- Pace, M. L., R. D. Batt, C. D. Buelo, S. R. Carpenter, J. J. Cole, J. T. Kurtzweil, and G. M. Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Natl. Acad. Sci. U. S. A.* **114**: 352–357. doi:[10.1073/pnas.1612424114](https://doi.org/10.1073/pnas.1612424114)
- Paerl, H. W., and others. 2006. Ecological response to hurricane events in the Pamlico sound system, North Carolina, and implications for assessment and management in a regime of increased frequency. *Estuaries Coasts* **29**: 1033–1045. doi:[10.1007/BF02798666](https://doi.org/10.1007/BF02798666)
- Page, E. S. 1954. Continuous inspection schemes. *Biometrika* **14**: 100–115. doi:[10.2307/2333009](https://doi.org/10.2307/2333009)
- Pickett, S. T. A., J. Kolasa, J. J. Armesto, and S. L. Collins. 1989. The ecological concept of disturbance and its expression at various hierarchical levels. *Oikos* **54**: 129–136. doi:[10.2307/3565258](https://doi.org/10.2307/3565258)
- R Core Team. 2019. R: A language and environment for statistical computing. Available from <https://www.r-project.org>
- Rykiel, E. J. 1985. Towards a definition of ecological disturbance. *Austral. Ecol.* **10**: 361–365. doi:[10.1111/j.1442-9993.1985.tb00897.x](https://doi.org/10.1111/j.1442-9993.1985.tb00897.x)
- Scheffer, M., and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends Ecol. Evol.* **18**: 648–656. doi:[10.1016/j.tree.2003.09.002](https://doi.org/10.1016/j.tree.2003.09.002)
- Scheffer, M., S. R. Carpenter, V. Dakos, and E. H. van Nes. 2015. Generic indicators of ecological resilience: Inferring the chance of a critical transition. *Annu. Rev. Ecol. Evol. Syst.* **46**: 145–167. doi:[10.1146/annurev-ecolsys-112414-054242](https://doi.org/10.1146/annurev-ecolsys-112414-054242)
- Smirnov, N. 1948. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* **19**: 279–281. doi:[10.1214/aoms/1177730256](https://doi.org/10.1214/aoms/1177730256)
- Strayer, D. L., D. T. Fischer, S. K. Hamilton, H. M. Malcom, M. L. Pace, and C. T. Solomon. 2020. Long-term variability and density dependence in Hudson River *Dreissena* populations. *Freshw. Biol.* **65**: 474–489. doi:[10.1111/fwbi.13444](https://doi.org/10.1111/fwbi.13444)
- Thayne, M. W., B. M. Kraemer, J. P. Mesman, B. W. Ibelings, and R. Adrian. 2021. Antecedent lake conditions shape resistance and resilience of a shallow lake ecosystem following extreme wind storms. *Limnol. Oceanogr.* **67**: S101–S120. doi:[10.1002/lno.11859](https://doi.org/10.1002/lno.11859)

- Verbesselt, J., A. Zeileis, and M. Herold. 2012. Near real-time disturbance detection using satellite image time series. *Remote Sens. Environ.* **123**: 98–108. doi:[10.1016/j.rse.2012.02.022](https://doi.org/10.1016/j.rse.2012.02.022)
- Weathers, K. C., and others. 2016. Frontiers in ecosystem ecology from a community perspective: The future is boundless and bright. *Ecosystems* **19**: 753–770. doi:[10.1007/s10021-016-9967-0](https://doi.org/10.1007/s10021-016-9967-0)
- Wiberg, P. L., S. R. Taube, A. E. Ferguson, M. R. Kremer, and M. A. Reidenbach. 2019. Wave attenuation by oyster reefs in shallow coastal bays. *Estuaries Coasts* **42**: 331–347. doi:[10.1007/s12237-018-0463-y](https://doi.org/10.1007/s12237-018-0463-y)
- Wolin, J. A., and E. F. Stoermer. 2005. Response of a Lake Michigan coastal lake to anthropogenic catchment disturbance. *J. Paleolimnol.* **33**: 73–94. doi:[10.1007/s10933-004-1688-2](https://doi.org/10.1007/s10933-004-1688-2)
- Yannarell, A. C., T. F. Steppe, and H. W. Paerl. 2007. Disturbance and recovery of microbial community structure and function following Hurricane Frances. *Environ. Microbiol.* **9**: 576–583. doi:[10.1111/j.1462-2920.2006.01173.x](https://doi.org/10.1111/j.1462-2920.2006.01173.x)
- Zwart, J. A. 2017. The influence of hydrologic residence time on Lake carbon cycling dynamics following extreme precipitation events. *Ecosystems* **20**: 1000–1014. doi:[10.1007/s10021-016-0088-6](https://doi.org/10.1007/s10021-016-0088-6)

Acknowledgments

This research was supported by a research grant from the Environmental Resilience Institute at the University of Virginia and NSF DEB-1832221. J.A.W. was also supported by NSF grants OAC-1839024 and OCE-2023555. Dat Ha provided helpful comments on the manuscript. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science.

Submitted 06 July 2021

Revised 29 March 2022

Accepted 11 April 2022

Associate editor: Ben Surridge