

Affine-Invariant Inference for Local Trend Similarity in Nonstationary Time Series, with Application to Biogeochemical Decoupling in an Estuary

Trevor D. Ruiz^{*1}, Aléa J. Seifert¹, Emma N. Hamilton¹, Catherine M. Mispagel¹, Jose Garcia¹, and Emily E. Bockmon²

¹ Statistics Department, California Polytechnic State University, San Luis Obispo, CA

² Chemistry & Biochemistry Department, California Polytechnic State University, San Luis Obispo, CA

Abstract

Oceanographic models that incorporate biogeochemical processes are widely used to assess the potential trajectory and impact of phenomena associated with climate change, such as hypoxia and acidification. Models often assume coupling between the carbonate system and dissolved oxygen, and although this behavior is robust across environments, empirical observations demonstrate that it can be interrupted locally under a variety of conditions. Monitoring data show that coupling and decoupling are observable in the first-order structure of time series of pH and dissolved oxygen, and ocean observing systems provide an abundance of data that make coupling locally testable. However, both variables are subject to independent biological and environmental influences, so relative local levels drift gradually over time, which complicates inference using existing methods. We develop a local inference procedure for testing trend similarity in time series under an unknown affine transformation, and apply a multiple testing adjustment to provide simultaneous inference across overlapping windows under arbitrary dependence, allowing for detection and temporal localization of decoupling episodes. Our procedure is based on a local asymptotic approximation for the rolling correlation in which trend variation controls the limiting behavior. We develop the required results directly and construct a test with plug-in estimators for the mean and variance of the limiting distribution, which we implement in a companion R package. We validate the method empirically through simulation experiments, and present an application to a multi-year record from Morro Bay, CA that identifies a previously undescribed decoupling pattern that is structured seasonally and by location. We estimate the effect of the pattern on local predictive uncertainty in pH reconstructions and show that prediction error increases by 23–48% during decoupling episodes, depending on location.

Keywords: rolling correlation, multiple testing, locally stationary processes, ocean acidification, hypoxia, Morro Bay

^{*}Corresponding Author: truíz01@calpoly.edu

1 Introduction

1.1 Background and motivation

Ocean acidification poses a major risk to the health of global ecosystems and threatens biodiversity, food webs, fisheries, and aquaculture. Coastal waters often exhibit co-occurring acidification (low pH) and hypoxia (low dissolved oxygen) (Cai et al., 2011; Wallace et al., 2014), a pattern which persists broadly across ocean environments (Gobler and Baumann, 2016) and is driven by CO₂ exchange during photosynthesis and respiration according to fixed stoichiometric ratios (Baumann and Smith, 2018). As a result, metabolic coupling between the carbonate system (pH/CO₂) and dissolved oxygen (DO) is commonly embedded as an operating assumption in oceanographic models. Such models are used for interpolation and historical reconstruction (Alin et al., 2012; Carter et al., 2018, 2021), hindcasting (Andrews et al., 2017), and climate change projections (Kwiatkowski et al., 2020). However, recent work documenting isolated instances of decoupling demonstrates that this operating assumption does not hold universally (Su et al., 2021; Li et al., 2024). Where models rely on or reproduce coupling, such failures imply a risk of local model misspecification that current methods are not well-equipped to detect.

Ocean observing systems generate large volumes of monitoring data that can be leveraged to identify decoupling events, measure their frequency, explore patterns, and generate and test hypotheses about possible mechanisms. In California, the Central and Northern California Ocean Observing System (CeNCOOS) covers the coastal ocean near Humboldt, Bodega, San Francisco, Monterey, and Morro bays; the network provides meteorological, chemical, physical, and biological data products spanning 1977 to the present (Central and Northern California Ocean Observing System, 2026; U.S. Integrated Ocean Observing System, 2026).

Here we leverage CeNCOOS data to infer spatial and temporal patterns in pH/CO₂-DO decoupling over a multiyear period in Morro Bay, a shallow, seasonally low-inflow estuary on California’s central coast. Low-inflow estuaries are common but understudied systems, in which reduced and intermittent freshwater delivery can produce environments distinct from river-dominated systems (Largier, 2023). In Morro Bay, spatial variation in tidal flushing distinguishes the ocean-influenced mouth of the bay from the shallow and slowly exchanged interior (Taherkhani et al., 2023) and coupled with the freshwater inputs generate strong spatial and temporal gradients in environmental conditions (Walter et al., 2018; Bartoloni et al., 2023). The estuary supports ecologically important eelgrass habitat as well as commercial oyster aquaculture, making locally resolved carbonate-system

and oxygen conditions relevant to both habitat resilience and coastal resource use. Figure 1 shows the CeNCOOS water quality monitoring locations in Morro Bay and (standardized) example pH and DO data from each location over a 52-day period. The data shown illustrate an apparent decoupling event at the Bay Mouth location lasting approximately four weeks by visual inspection. In this paper we develop a method for detecting and localizing such events from long-term, high-frequency monitoring data, and then use this approach to identify a seasonal decoupling pattern in Morro Bay that has not previously been described.

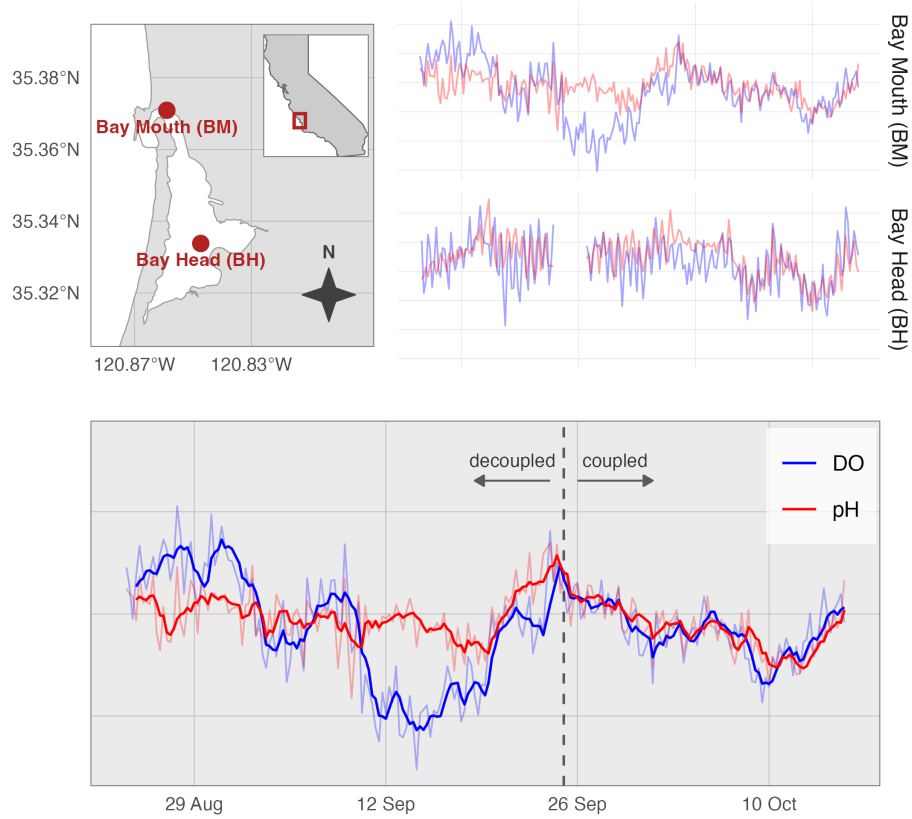


Figure 1: Top, Morro Bay water quality monitoring locations (left) and example dissolved oxygen (DO) and pH data over a 52-day period in 2022 (right) showing a decoupling event at the BM location only. Bottom, reproduction of BM location data with 24-hour moving average overlays and approximate timing of transition between decoupled and coupled states (identified by visual inspection). Data are standardized and passed through a band-stop filter to attenuate tidal frequencies.

1.2 Statistical perspectives

Considered in the abstract, coupling seems naturally expressed as a (cross) correlation, and decoupling as a change in correlation structure. Several existing methods can detect change points in

covariance and correlation under the assumption that the underlying processes are piecewise stationary (Aue et al., 2009; Wied et al., 2014; Dehling et al., 2017). Vogt and Dette (2015) consider the detection of change points under local stationarity (see Dahlhaus, 2012, for an overview), and Dette et al. (2019) develop a CUSUM-type test for correlation change points under piecewise local stationarity. Yet coupling in physical systems may not produce discrete transitions between fixed states as much as smoothly varying fluctuations between regimes. Zhao (2015) develops an estimator for continuous time-varying autocorrelation structure in nonstationary time series, and Zhang and Shao (2024) extend the method to dynamic cross-correlations under arbitrary unknown trend functions. In these approaches, time-varying trends enter the analysis as nuisance terms. However, the data in Figure 1 show time-varying coupling appears in the first-order structure: moving averages align closely at times and drift apart at others. In the application context, Baumann and Smith (2018) show empirically that coupling manifests at multiple time scales, but the decoupling mechanisms proposed in the scientific literature — such as seasonal transitions (Su et al., 2021) or gas exchange (Li et al., 2024) — are incremental processes that accrue gradually over time. Decoupling patterns are therefore more likely expressed in the smoothed series as co-movement between trends, rather than in the high-frequency variation. This is broadly consistent with the example in Figure 1, and implies that trend similarity, not cross-correlation, is the salient inferential target.

Statistical inference on time-varying trends is less standard than inference on second-order structure, but has been studied in a nonparametric regression framework (Wu and Zhao, 2007). Given data $\{X_{it}\}_{t=1}^T$ for series $i = 1, \dots, p$, these approaches consider the model

$$X_{it} = \mu_i\left(\frac{t}{T}\right) + Z_{it} \quad (1.1)$$

where $\mu_i : [0, 1] \rightarrow \mathbb{R}$ are unknown deterministic functions and Z_{it} are independent weakly stationary noise processes with $\mathbb{E}[Z_{it}] = 0$. Hall and Hart (1990) test identity of trends, *i.e.*, fixing $i \neq j$, $H_0 : \mu_i(u) = \mu_j(u)$ for all $u \in [0, 1]$ against $H_A : \mu_i(u) \neq \mu_j(u)$ for some $u \in [0, 1]$. This test can resolve global differences, but cannot localize where they occur. Park et al. (2009) propose a similar test together with a graphical diagnostic to identify the location of the differences, yet do not frame the inference in terms of local hypotheses. Recently, Khismatullina and Vogt (2023, 2026) formulate an explicitly localized inference procedure: for a local window $W \subset [0, 1]$, they test $H_0(W) : \mu_i(u) = \mu_j(u)$ for all $u \in W$ against $H_A(W) : \mu_i(u) \neq \mu_j(u)$ for some $u \in W$, and by performing simultaneous inference across a family generated by varying the location and width of

W , they obtain a test that identifies where trends diverge locally. Varying the widths of windows W in the underlying hypotheses allows for added sensitivity in localizing divergence and yields a “multiscale” procedure, and establishing valid simultaneous inference within this framework is the central theoretical difficulty in this approach. The window-level inferences use test statistics constructed from kernel-weighted averages of pointwise differences between precentered series, where precentering is performed once globally (*i.e.*, across all windows), and the multiscale procedure relies on the precentering estimation error being $O_p(T^{-1/2})$.

In the application context, the stoichiometry of O_2 and CO_2 exchange is fixed, but the pH response depends on buffer capacity, which is set by alkalinity and temperature (Sunda and Cai, 2012) and thus varies over time. Similarly, O_2 solubility varies with temperature, which also changes over time, so the same metabolic fluctuation in O_2 shifts DO concentration differently at different times. These mechanisms affect local levels and relative scales of pH and DO independently of metabolic coupling, and Figure 2 illustrates such locally shifting alignments. Co-movement between trends thus occurs under an unknown local transformation. In principle, one could correct for this feature by performing local realignments after estimating the transformation for each window. Härdle and Marron (1990) show that unknown transformations between regression curves can be estimated at the parametric $\sqrt{n}$ rate — here, $\sqrt{T}$ — for *iid* data, even when the curve estimates converge more slowly. Applied locally, however, the rate would be set by window size, and so for simultaneous inference across windows, the estimation error would have to be carried through the null distribution of every local test. This type of estimation error is not innocuous in general (see, *e.g.*, Zhang and Shao, 2024). A simple empirical experiment reported in Appendix B confirms that local realignment fails to provide an adequate preprocessing strategy within existing frameworks.

1.3 Methodological framework and contributions

In what follows we develop a test of local trend similarity under an unknown local affine transformation based on rolling correlations between nonparametric trend estimates. Because correlation is invariant under affine transformation, this approach does not require local realignment of series and sidesteps the technical issue of controlling errors associated with re-estimations across a family of local tests. To formalize our approach, we rewrite model (1.1) with $p = 2$ as a bivariate time series $\{X_t \in \mathbb{R}^2\}_{t=1}^T$ following the model

$$X_t = \nu_t + Z_t \tag{1.2}$$

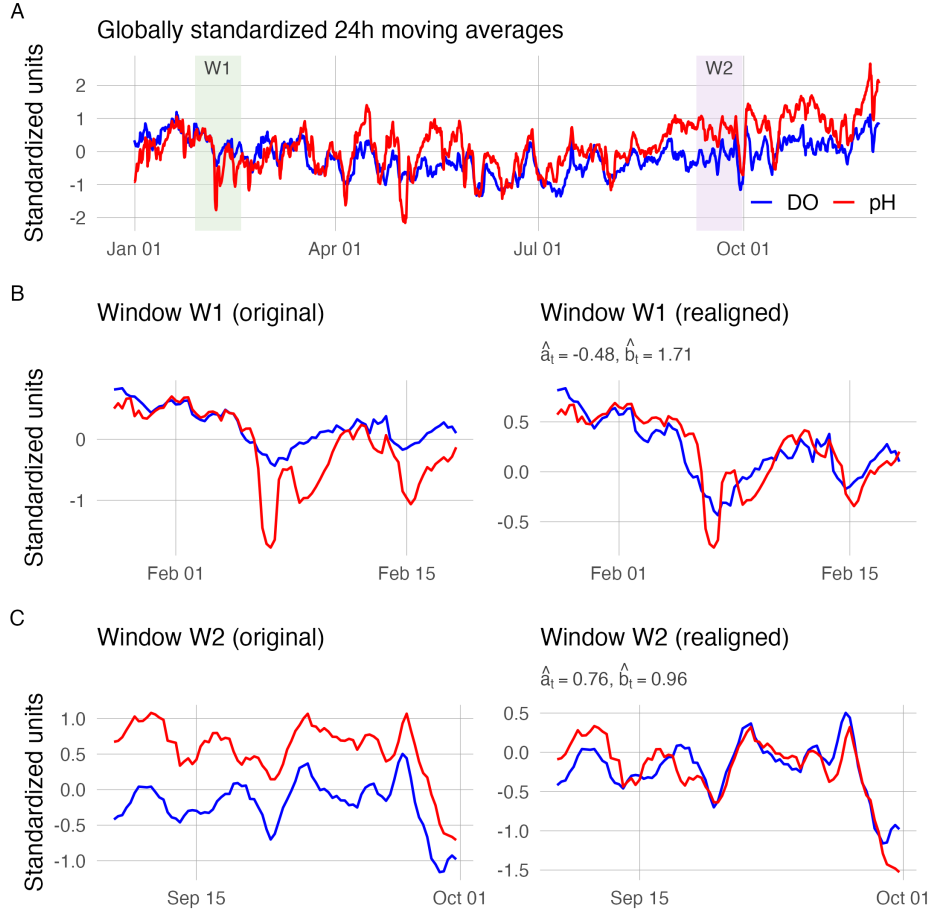


Figure 2: Trend similarity under time-varying affine transformations in the Morro Bay CeNCOOS data: (A) globally standardized trends show time-varying amplitude gain and location drift over a roughly one-year period (January to December, 2025) at the BH location; (B-C) specific three-week windows W1 and W2 show trends realign locally under affine transformation, but with different scale and location parameters.

where $\{Z_t \in \mathbb{R}^2\}_{t=1}^T$ is a weakly stationary noise process and $\nu : \{1, \dots, T\} \rightarrow \mathbb{R}^2$ is a deterministic trend (we suppress the time rescaling notation of model (1.1) for simplicity). We say the trends ν_{1t} and ν_{2t} exhibit *local affine similarity* if there exist $a_t \in \mathbb{R}$ and $b_t > 0$ such that $\nu_{1u} = a_t + b_t \nu_{2u}$ for all $u \in W_t$, where $W_t = \{t - s + 1, \dots, t\}$ is the local window of width s ending at time t . We measure the departure from local similarity by the smallest mean square separation achievable under affine transformation

$$\delta_t^2 = \frac{1}{\omega_{1t}^2} \min_{a_t \in \mathbb{R}, b_t \geq 0} \left\{ \frac{1}{s} \sum_{u \in W_t} [\nu_{1u} - (a_t + b_t \nu_{2u})]^2 \right\} \quad (1.3)$$

where $\omega_{1t}^2 = \text{Var}_{W_t}(\nu_1)$ denotes the (empirical) variance of the trend ν_1 on W_t . In (1.3), the scale constraint $b_t > 0$ is relaxed to its closure so that a minimizer always exists. We consider the local null hypothesis $H_{0t} : \delta_t = 0$, under which trends are similar on W_t up to an unknown affine transformation, against the local alternative hypothesis $H_{At} : \delta_t > 0$, under which the trends differ on W_t by more than an affine reparametrization.

The use of a rolling window correlation to test H_{0t} against H_{At} arises naturally from the inferential target δ_t : since the unconstrained minimizer of (1.3) is given by the least squares regression of ν_1 on ν_2 over W_t , the innermost term is the residual variance $\omega_{1t}^2(1 - r_t^2)$, where $r_t = \text{Corr}_{W_t}(\nu_1, \nu_2)$ denotes the local (empirical) correlation between trends on W_t . On substitution in (1.3), the ω_{1t}^2 term cancels and renders $\delta_t^2 = 1 - r_t^2$ a scale-invariant measure of local affine similarity. Under the constraint $b_t \geq 0$, the minimizer is at the boundary whenever $r_t \leq 0$, so (1.3) becomes

$$\delta_t^2 = 1 - (r_t^+)^2 \quad \text{where} \quad r_t^+ = r_t \vee 0 \quad (1.4)$$

Thus, the local null hypothesis holds exactly when $r_t = 1$. To test the hypothesis, we compare the rolling correlation between nonparametric trend estimates to its value under H_{0t} , which we characterize asymptotically and estimate from the data. Existing asymptotic results for local correlations target the correlation of the stochastic component and treat trend variation as a nuisance (*e.g.*, Choi and Shin, 2021; Zhang and Shao, 2024); here that relationship is inverted, so we develop the required results directly. We then provide simultaneous inference by applying the adjustment of Benjamini and Yekutieli (2001) across overlapping windows to control the false discovery rate under arbitrary dependence.

The remainder of the paper is organized as follows. In Section 2 we derive a local asymptotic approximation for the rolling correlation and construct a test of the hypotheses $H_{0t} : \delta_t = 0$ against $H_{At} : \delta_t > 0$. In Section 3 we measure the finite-sample accuracy of the approximation and provide empirical validation of the method in a simulation experiment designed to assess power relative to H_{At} . In Section 4 we return to the application of interest and use the method to identify seasonal and spatial decoupling patterns in Morro Bay. Lastly, in Section 5, we discuss advantages and known limitations of the methodology and indicate future directions for its development and application.

2 Methods

Consider estimating the trend in model (1.2) by the moving average over a window of width h :

$$Y_t := (K_h X)_t = \frac{1}{h} \sum_{j=0}^{h-1} X_{t-j}, \quad t \in \{h, h+1, \dots, T\} \quad (2.1)$$

We focus on moving averages throughout, but K_h may be any linear smoother with normalized weights (*e.g.*, Priestley and Chao, 1972; Fan and Gijbels, 1996). The moving average inherits the trend-plus-noise structure of the underlying model since $Y_t = (K_h \nu)_t + (K_h Z)_t$. Further, K_h preserves local affine similarity, so it is straightforward to reformulate the inferential problem on the smoothed scale. Extending the assumption of local affine similarity on W_t over the preceding $h-1$ observations, $\nu_{1u} = a_t + b_t \nu_{2u}$ implies $(K_h \nu_1)_u = a_t + b_t (K_h \nu_2)_u$, so under H_{0t} we have $\tilde{r}_t := \text{Corr}_{W_t}(K_h \nu_1, K_h \nu_2) = 1$. Departures from local affine similarity at timescales shorter than h are attenuated by smoothing, and under H_{At} we may have $r_t < \tilde{r}_t$, so smoothing may give up power. However, as noted in Section 1.2, decoupling is expressed on long time scales in the application of interest, and when departures from similarity vary slowly relative to h , we have $\tilde{r}_t \approx r_t$.

Now consider the empirical counterpart of $\tilde{r}_t$, the rolling window correlation between moving averages on W_t , which we express as

$$R_t := \frac{C_{12}}{\sqrt{C_{11}C_{22}}} \quad \text{where} \quad C = \frac{1}{s} \sum_{i=1}^s (Y_i - \bar{Y})(Y_i - \bar{Y})^T \quad \text{with} \quad Y_i := Y_{t-s+i} \quad (2.2)$$

Our approach to inference is to compare R_t with the value of its population counterpart ρ_t under H_{0t} . In Section 2.1 we characterize R_t asymptotically for any process with a locally stationary approximation. We then show in Section 2.2 that a parametric model for Z_t facilitates approximation of ρ_t and the limiting distribution of $\sqrt{s}(R_t - \rho_t)$ under H_{0t} , and use this approximation to characterize the relationship between ρ_t and $\tilde{r}_t$. Finally, in Section 2.3 we construct a test using plug-in estimates of ρ_t and the asymptotic variance.

2.1 Asymptotic approximation for the local correlation

We characterize the behavior of R_t asymptotically in two steps. First we establish a central limit theorem for the sample moments under the local asymptotic regime of Dahlhaus (1997), and then we apply the multivariate delta method to obtain a limiting distribution for R_t . We state these results here and defer proofs to the Supplementary Material.

Theorem 1. Consider a triangular array $\{Y_{t,T} \in \mathbb{R}^2 : t = 1, \dots, T\}_{T \in \mathbb{N}}$. Fix $u \in (0, 1)$ and define $W_T := \{t_0 - s_T + 1, \dots, t_0\}$ to be the window of width s_T with right endpoint $t_0 = \lfloor uT \rfloor$. For $y \in \mathbb{R}^2$ let $h(y) = (y_1, y_2, y_1^2, y_2^2, y_1 y_2)^T$, and define the moment process $H_{t,T} := h(Y_{t,T})$. Denote the sample moments within the window by:

$$\hat{\theta}_T(u) = \frac{1}{s_T} \sum_{t \in W_T} H_{t,T}$$

Assume that there exists a stationary approximation $\{Y_t(u)\}_{t \in \mathbb{Z}}$ of $\{Y_{t,T}\}_{t \in W_T}$ (in the sense of Dahlhaus, 2012) that satisfies the regularity conditions in the Supplementary Material and define $H_t(u) := h(Y_t(u))$ to be the local approximation to the moment process. Define $\theta(u) := \mathbb{E}[H_t(u)]$ and $\Gamma_H(l; u) := \text{Cov}(H_t(u), H_{t-l}(u))$. Then as $T \rightarrow \infty$, $s_T \rightarrow \infty$, and $\frac{s_T}{T} \rightarrow 0$,

$$\sqrt{s_T} \left(\hat{\theta}_T(u) - \theta(u) \right) \xrightarrow{d} N(0, \Sigma(u))$$

where $\Sigma(u) = \sum_{l \in \mathbb{Z}} \Gamma_H(l; u)$ is the long-run covariance matrix of the process $H_t(u)$.

The distribution of R_t is obtained by a straightforward application of the delta method to Theorem 1, which is summarized in the next result.

Theorem 2. Suppose the conditions of Theorem 1 hold. For $\theta \in \mathbb{R}^5$, let $C_{11}(\theta) = \theta_3 - \theta_1^2$ and $C_{22}(\theta) = \theta_4 - \theta_2^2$ and $C_{12}(\theta) = \theta_5 - \theta_1 \theta_2$, and define the map:

$$g(\theta) := \frac{C_{12}(\theta)}{\sqrt{C_{11}(\theta)C_{22}(\theta)}}$$

Define $R_T(u) := g(\hat{\theta}_T(u))$ to be the local sample correlation and define $\rho(u) := g(\theta(u))$ to be the local population correlation. Then as $T \rightarrow \infty$, $s_T \rightarrow \infty$, and $\frac{s_T}{T} \rightarrow 0$,

$$\sqrt{s_T} (R_T(u) - \rho(u)) \xrightarrow{d} N(0, V(u))$$

where $V(u) = [\nabla g(\theta(u))]^T \Sigma(u) [\nabla g(\theta(u))]$.

Results of this type have been obtained elsewhere; for example, Choi and Shin (2021) apply a similar moment central limit theorem and delta method argument to obtain a limiting distribution for a kernel-smoothed local correlation under a location-scale model. However, these results treat the trend as a nuisance and require that local trend variation vanish asymptotically, whereas here that variation carries the signal of interest. Theorems 1 and 2 provide an asymptotic characteriza-

tion of R_t under a locally stationary approximation and in the form needed for the application of interest.

2.2 Model-based approximation under stationary Gaussian noise

We specialize the results of Section 2.1 to the rolling correlation over the smoothed process $Y_t = (K_h X)_t$ under the trend-plus-noise model (1.2). Under the model, the noise Z_t is weakly stationary, so the locally stationary approximation to the smoothed process Y_t takes the form $Y_t(u) = m(u) + \eta_t$, where $m(u)$ is a deterministic term that depends only on $K_h \nu$ and $\eta_t = (K_h Z)_t$ is weakly stationary. Thus, the covariance structure of $Y_t(u)$ is entirely determined by the noise, and in particular, $\text{Cov}(Y_t(u), Y_{t+l}(u)) = \Gamma_\eta(l) = \frac{1}{h^2} \sum_{i,j=0}^{h-1} \Gamma_Z(l+i-j)$, where $\Gamma_Z(l) = \text{Cov}(Z_t, Z_{t+l})$. The entries of $\Gamma_H(l; u)$ include fourth-order moments of $Y_t(u)$, and thus of η_t , which do not in general admit closed-form expressions in terms of Γ_Z . However, if the noise Z_t is Gaussian, Γ_H is polynomial in Γ_Z , which allows for numerical approximation of the asymptotic covariance $\Sigma(u)$ from truncated sums involving Γ_Z . Below we use this structure to derive explicit approximations.

In the application of interest, we explicitly estimate locally nonconstant trends, but Theorems 1 and 2 assume a stationary approximation on W_T , which requires the smoothed trend to be roughly constant over the window. If this assumption were reflected in the underlying trends exactly, *i.e.*, if ν_t were constant, then there would be no meaningful trends to test for similarity, H_{0t} would hold trivially, and δ_t would be undefined since $\omega_{1t}^2 = 0$ in (1.3). Thus, for practical application of the theory, we evaluate the asymptotic expressions at window-averaged moments under the assumption that the trends are slowly varying. This approach provides model-based approximations for $\rho(u)$ and $V(u)$, which we later validate empirically in Section 3.1. For notational simplicity, we fix $u \in (0, 1)$ and suppress the dependence on u throughout.

Proposition 1. *Consider $Y_{kt} = m_{kt} + \eta_{kt}$ for $k = 1, 2$, where $m_{kt} \in \mathbb{R}$ are deterministic trends and η_{1t}, η_{2t} are independent mean-zero stationary Gaussian processes with autocovariance functions $\gamma_k(l) = \text{Cov}(\eta_{kt}, \eta_{k(t+l)})$, and fix a local window W . Denote the empirical trend variances over W by $\tau_k^2 := \frac{1}{|W|} \sum_{t \in W} (m_{kt} - \bar{m}_k)^2$, the empirical trend correlation by $r := \text{Corr}_W(m_1, m_2)$, and the noise variances by $\sigma_k^2 := \gamma_k(0)$. Then the local population correlation in Theorem 2 is*

$$\rho \approx \frac{r \tau_1 \tau_2}{\sqrt{(\tau_1^2 + \sigma_1^2)(\tau_2^2 + \sigma_2^2)}}$$

and, under H_0 (that is, when $r = 1$), the asymptotic variance is

$$V \approx \frac{Q_{12}}{BD} + \frac{\tau_2^2 \sigma_1^4 L_1}{B^3 D} + \frac{\tau_1^2 \sigma_2^4 L_2}{BD^3} + \frac{\tau_1^2 \tau_2^2 Q_1}{2B^3 D} + \frac{\tau_1^2 \tau_2^2 Q_2}{2BD^3}$$

where $B := \tau_1^2 + \sigma_1^2$ and $D := \tau_2^2 + \sigma_2^2$ denote the window population variances of Y_{1t}, Y_{2t} , respectively, and the remaining terms are the covariance sums $L_k := \sum_{l \in \mathbb{Z}} \gamma_k(l)$, $Q_k := \sum_{l \in \mathbb{Z}} \gamma_k(l)^2$, $Q_{12} := \sum_{l \in \mathbb{Z}} \gamma_1(l) \gamma_2(l)$.

Proposition 1 explicitly identifies the effect of departures from affine similarity on the local population correlation: ρ can be rewritten in terms of the signal-to-noise ratios $\lambda_1 := \frac{\tau_1^2}{\sigma_1^2}$ and $\lambda_2 := \frac{\tau_2^2}{\sigma_2^2}$ as

$$\rho \approx r \sqrt{\frac{\lambda_1 \lambda_2}{(1 + \lambda_1)(1 + \lambda_2)}}$$

Recalling that $r = 1$ under H_0 , the second factor is the local population correlation under local affine similarity.¹ That is, $\rho \approx r \rho_0$, where ρ_0 denotes the value of the local population correlation under H_0 . Since $r \leq 1$, we have $\rho \leq \rho_0$ with equality under H_0 . The relevant inference is therefore directional: H_0 implies $\rho = \rho_0$, and hence $\rho < \rho_0$ implies H_A .

2.3 Test for local similarity

The statistic $\sqrt{s}(R_t - \rho) / \sqrt{V}$ provides the basis for a test: under local affine similarity, ρ takes the value ρ_0 and V is given by Proposition 1. These quantities depend on τ_1^2, τ_2^2 , and the autocovariance functions γ_1, γ_2 , which are unknown and must be estimated from data. Although the noise is stationary under (1.2), trends vary across local windows, and therefore so do ρ and V ; we write ρ_t, V_t to denote the local population correlation and asymptotic variance from Proposition 1 on the window $W_t = \{t - s + 1, \dots, t\}$. Given estimates $\hat{\rho}_t, \hat{V}_t$, the test statistic at each time point t is

$$G_t := \frac{\sqrt{s}(R_t - \hat{\rho}_t)}{\sqrt{\hat{V}_t}} \quad (2.3)$$

Because trend separation attenuates local correlation, the relevant departures from H_{0t} are one-sided. We therefore test $H'_{0t} : \rho_t = \rho_{0t}$ against $H'_{At} : \rho_t < \rho_{0t}$ on the window W_t via the p-value $p_t = \Phi(G_t)$, construct a family of dependent tests of the null hypotheses $\{H'_{0t}\}_{t=h+s-1}^T$ against the

¹Here, r denotes trend correlation on the smoothed scale, *i.e.*, we use $r = \text{Corr}_W(K_h \nu_1, K_h \nu_2)$ and drop the previous distinction between $\tilde{r} = \text{Corr}_W(K_h \nu_1, K_h \nu_2)$ and $r = \text{Corr}_W(\nu_1, \nu_2)$ from Section 1.3. The abuse of notation is slight and limited in scope, as we do not make continued use of $r, r_t, \tilde{r}_t$ for the remainder of the paper.

corresponding alternatives H'_{At} , and apply the false discovery rate adjustment of Benjamini and Yekutieli (2001) to obtain simultaneous inference with false discovery rate control under arbitrary dependence between tests. Since H_{0t} implies H'_{0t} , the false discovery rate is controlled with respect to the family $\{H_{0t}\}$ as well. To localize departures from H_{0t} , we apply their step-up procedure: writing m for the number of tests in the family, $p_{(1)} \leq \dots \leq p_{(m)}$ for the ordered p-values, and $c_m = \sum_{i=1}^m i^{-1}$, we reject the hypotheses corresponding to the k smallest p-values, where $k = \max \{j : p_{(j)} \leq j\alpha / (m c_m)\}$.

Operationalizing the test requires estimates $\hat{\rho}_{0t}, \hat{V}_t$. The quantities $\sigma_1^2, \sigma_2^2, L_k, Q_k, Q_{12}$ are defined in terms of $\gamma_k(l)$, which is related to the noise autocovariance $\xi_k(l) = \text{Cov}(Z_{kt}, Z_{k(t+l)})$ via

$$\gamma_k(l) = \frac{1}{h^2} \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \xi_k(l+i-j) \quad (2.4)$$

Under the assumption that the noise components are independent, we estimate ξ_1, ξ_2 by direct application of the difference-based method of Hall and Van Keilegom (2003), and then filter to obtain estimates of γ_1, γ_2 following (2.4). Then, noting that $\mathbb{E}[\text{Var}_{W_t}(Y_1)] = \tau_1^2 + \sigma_1^2 - \text{Var}(\bar{\eta}_1)$, where $\text{Var}_{W_t}(Y_1)$ is the empirical variance of Y_{1t} on W_t and $\bar{\eta}_1$ is the window mean of the smoothed noise, we estimate τ_1^2 by the bias-adjusted local empirical variance

$$\hat{\tau}_1^2 = \left[\text{Var}_{W_t}(Y_1) - \frac{s-1}{s} \hat{\gamma}_1(0) + \frac{2}{s^2} \sum_{l=1}^{s-1} (s-l) \hat{\gamma}_1(l) \right]_+ \quad (2.5)$$

and estimate τ_2^2 similarly. Plug-in estimates $\hat{\rho}_{0t}, \hat{V}_t$ then follow from Proposition 1. In practice, we introduce a lag truncation to compute the covariance sums L_k, Q_k, Q_{12} for the plug-in estimate $\hat{V}_t$. The full estimation and inference procedure is implemented in the accompanying R package `lomad` (local moving average decoupling) (Ruiz et al., 2026).

3 Simulations

3.1 Finite-sample accuracy

We conducted a series of experiments to verify that the results of Section 2 provide adequate approximations at realistic sample sizes. The experiments first check whether the model-based expressions for ρ_t, V_t under window averaging in Proposition 1 remain faithful approximations of the empirical moments of R_t . Next, they isolate the asymptotic approximation from the effect of

estimation error by separately comparing both G_t , as defined in (2.3), and an oracle version of G_t to the limiting distribution. As window size was found to affect quality mainly through estimation error rather than the underlying approximation, we report only the results at a representative value $s = 100$.²

All datasets generated in the experiment share two fixed Fourier trends of length $T = 2000$ satisfying exact local affine similarity, as defined in Section 1.3, with $b_t \equiv 2$ and $a_t \equiv 0$. To simulate data, we add independent AR(1) noise to form $\{X_t\}_{t=1}^{2000}$ following model (1.2); we fix the innovation variance at 0.64 for both series and specify AR coefficients at $\phi = 0.3, 0.5$.

For estimation, we fix the moving average bandwidth at $h = 5$. For each simulated dataset, we compute $R_t, \rho_t, V_t, \hat{\rho}_t, \hat{V}_t$ on every fifth consecutive usable window (380 windows total). The true noise parameters ϕ_k, σ_k^2 and the true smoothed trend variances τ_{kt}^2 are used to compute the oracle mean and variance ρ_t and V_t ; plug-in estimates are computed as described in Section 2.3. We generate 1000 replicates. For the first 100 replicates, we compute the empirical mean and variance of R_t pointwise over time for comparison with theoretical counterparts. Across all replicates, we record pointwise rejection rates for each of the 380 windows using our method to estimate realized type I error. Finally, we inspect the empirical distribution of G_t (oracle and end-to-end) at four evenly-spaced evaluation points reflecting a range of local correlation regimes; Table 1 shows the relevant population quantities.

Table 1: Oracle quantities at evenly-spaced evaluation points in the validation experiment.

t	τ_{1t}^2	τ_{2t}^2	λ_{1t}	λ_{2t}	ρ_t	V_t
400	0.0030	0.0118	0.008	0.052	0.020	4.431
900	0.0189	0.0755	0.050	0.333	0.109	4.689
1400	0.0958	0.3834	0.252	1.690	0.356	3.718
1900	0.0011	0.0045	0.003	0.020	0.008	4.360

Results indicate that empirical moments of R_t closely track their population counterparts (Figure 3A) and the approximation produces a well-calibrated test under H_0 but one that becomes anti-conservative with error introduced by plug-in estimation (Figure 3B). In the experiment, empirical type I error realized by the end-to-end test more closely tracks its oracle counterpart (and the nominal level) when the signal-to-noise regime is favorable, but the two visibly separate when $\lambda_{kt} < 0.1$. In the worst-case scenario, the realized rejection rate reaches 0.103 (at $t = 1270$); here, $\tau_2^2 = 4\tau_1^2 = 0.01097$ produces $\lambda_1 = 0.0072, \lambda_2 = 0.0484$, and $\rho = 0.0182$. However, rejection rates

²For context, with hourly data this would correspond to a correlation window of 4.17 days.

remain elevated over nominal levels across windows in the experiment — the average rejection rate over all windows is 0.0714 for the end-to-end procedure, compared with 0.0541 for the oracle procedure. This is unsurprising, as plug-in estimation adds error not accounted for in the null distribution used in the test: the variance of $R_t - \hat{\rho}_t$ is inflated relative to $R_t - \rho_t$ and $\hat{V}_t$ estimates the latter, not the former. The magnitude of inflation varies with signal-to-noise ratio and the level of the test, with relative elevation above nominal increasing as α shrinks. Although downstream multiple testing adjustments have the potential to compound the issue, in practice we find that the adjustment used in our procedure controls the false discovery rate at nominal levels despite the pointwise anticonservatism of the test; we demonstrate this result in Section 3.2 below.

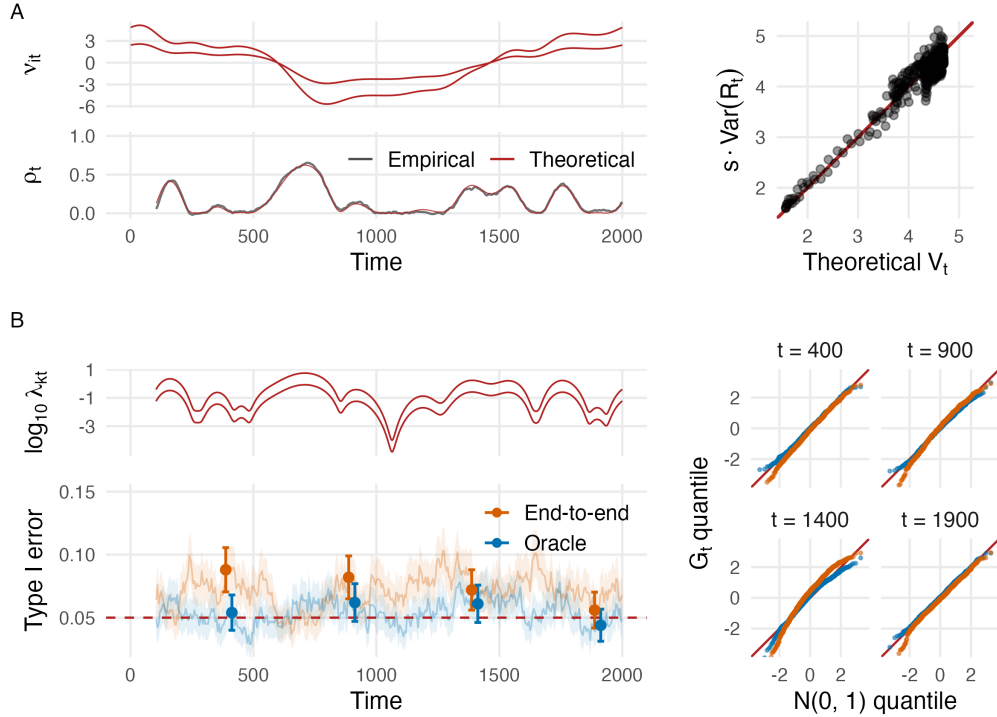


Figure 3: Results of validation experiment. (A) Trends used in the experiment; empirical mean and variance of R_t across the first 100 replicates against theoretical counterparts. (B) Underlying signal-to-noise ratios, type I error rates of pointwise nominal $\alpha = 0.05$ tests by window with 95% Wald intervals, and distributions of G_t at four evenly-spaced evaluation points.

3.2 Power and localization

We carried out a separate simulation study to evaluate the efficacy and reliability of simultaneous local inference across a variety of conditions. The study measures power, as a function of the effect size δ_t , across factorial combinations of window size, autocorrelation, and average signal-to-noise

ratio. Combinations are specified to produce inferential regimes of varying relative difficulty while remaining broadly plausible in the context of the motivating application. In each factorial cell, we evaluate the procedure on three distinct “structures” that differ in how local trend separation is distributed in time, under the hypothesis that irregular distributions may produce more challenging inferential problems. In addition to power curves, we also examine effect sizes conditional on test outcomes to measure how informative the test is about the presence or absence of an effect within each factorial cell.

Data for the study are simulated under model (1.2). Trend construction proceeds in stages, starting with a pair μ_1, μ_2 generated from an orthonormal Fourier basis. The coefficients for μ_1 are drawn randomly with spectral decay and normalized to fixed total variance; the coefficients for μ_2 are the displacement by a unit-length vector in a random direction orthogonal to μ_1 and weighted toward lower-frequency harmonics; we rescale so that $\|\mu_2\| = \|\mu_1\|$. To concentrate separation within local windows, we mix the base trends μ_1, μ_2 using a time-varying coupling weight w_t to obtain $\tilde{\nu}_{it} = \bar{\mu}_t + d(1 - w_t)(\mu_{it} - \bar{\mu}_t)$, where $\bar{\mu}_t := (\mu_{1t} + \mu_{2t})/2$ and d is a scaling factor. When $w_t = 1$, the trends are identical; separation increases at a rate set by d as $|1 - w_t|$ grows. We consider three separation “structures”, illustrated in Figure 4 and described in detail in Appendix B, and for any given structure, we use the scale factor d as a device to generate effect sizes of different magnitudes. Finally, the trend mixtures are passed through a slowly time-varying affine transformation to obtain the pair $\nu_{1t} := \tilde{\nu}_{1t}$ and $\nu_{2t} := a_t + b_t\tilde{\nu}_{2t}$, where the coefficients a_t, b_t are generated by smoothed Gaussian random walks constrained to remain near-constant on local windows. To form a pair of simulated series, we add independent AR(1) noise with shared coefficient ϕ and distinct innovation variances calibrated to the trends so that the signal-to-noise ratios λ_{it} average to a target $\bar{\lambda}$ within each series. The 18 factorial cells in the study are the combinations of $\bar{\lambda} \in \{0.5, 1.5\}$, $\phi \in \{0.3, 0.5, 0.7\}$, $s \in \{50, 100, 150\}$, and in each cell, series length is fixed in proportion to window size at $T = 25s$, and 500 replicate datasets are generated per trend structure at each of three values of the effect size generator d . Thus, for each cell and structure, we obtain 1500 replicates. For estimation, we fix the smoothing bandwidth at $h = 5$ throughout. This design produces 81,000 unique datasets and 194,157,000 usable windows.

The study endpoints are defined in terms of the uniform probability measure $\mathbb{P}$ over simulated local windows from a given structure within a factorial cell. We compute power as the probability $\mathbb{P}(G_t \in \mathcal{R}_t(\alpha) \mid \delta_t)$, where $\mathcal{R}_t(\alpha)$ denotes the rejection region on W_t under the step-up procedure at level α . We characterize localization through precision, computed as $\mathbb{P}(\delta_t > c \mid G_t \in \mathcal{R}_t(\alpha))$,

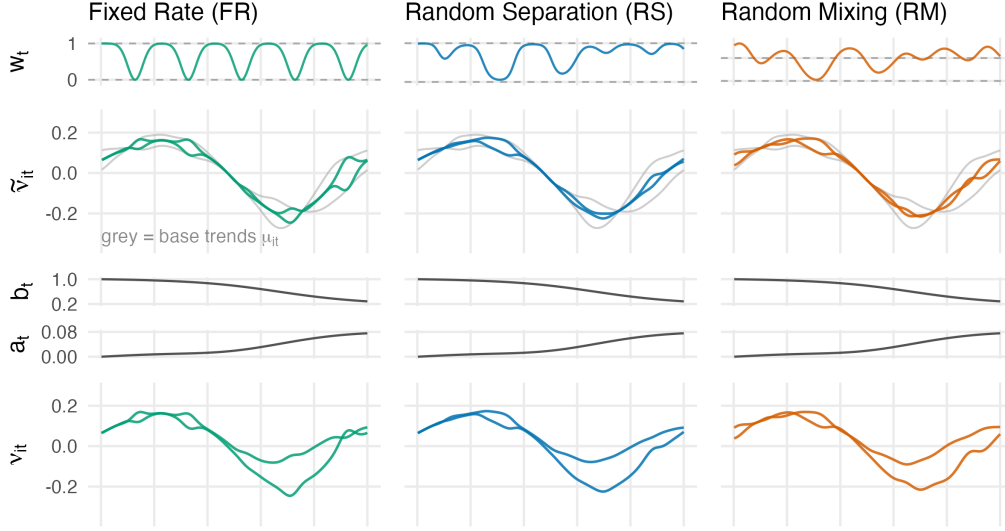


Figure 4: Examples illustrating the trend construction process for each of the three structured separation methods; all three examples are generated from the same base trends and apply the same affine transformation. Panels within each column are horizontally aligned by time t ; from top to bottom, panels show the coupling weight w_t , the mixtures $\tilde{\nu}_{it}$ with the base trends shown in the background in grey, the affine transformation coefficients a_t, b_t , and the final trends ν_{it} .

together with negative predictive value (NPV), computed as $\mathbb{P}(\delta_t \leq c \mid G_t \notin \mathcal{R}_t(\alpha))$. Profiles over $c \in [0, 1]$ provide a characterization independent of any particular threshold for what counts as a meaningful effect; for a summary metric, we compute the area under the precision/(1 – NPV) curve (AUC), which can be shown to be formally equivalent to the probability that effect sizes are well-ordered by the test, *i.e.*, $\mathbb{P}(\delta_t > \delta_{t'} \mid G_t \in \mathcal{R}_t(\alpha), G_{t'} \notin \mathcal{R}_{t'}(\alpha))$ for t, t' drawn independently under $\mathbb{P}$. Because the trends are known under the simulation design, the effect size δ_t can be computed exactly on every window, but varies widely across any given series. We compute power by binning realized effect sizes into a grid of 100 cells of equal width 0.01 over the interval $[0, 1]$; because windows with uncorrelated or anti-correlated trends resolve to an effect size $\delta_t = 1$, we treat this value separately as a point mass. For localization, precision and negative predictive value are computed along a grid of 1000 cells of equal width 0.001 over the interval $(0, 1]$.

Figure 5 shows the results of the study. After disaggregating usable windows by factorial cell and trend structure, the design specifications and our specific run³ yield minimum Monte Carlo sample sizes of 2,080 windows for estimating power (pointwise), 29,533 rejections for estimating precision, and 1,414,039 non-rejections for estimating NPV. Considering the substantial overlap between windows, effective sample sizes are much lower: simple estimates (see, *e.g.*, Rao and Scott,

³Realized effect sizes are subject to randomness from the trend generation procedure.

1992) are 566, 715, and 1,899, respectively. In the easiest regimes, defined by high signal-to-noise, large window size, and low autocorrelation, the method detects small effects with high probability. In the hardest regimes, defined by low signal-to-noise, small window size, and high autocorrelation, the method detects only the largest effects and does so infrequently. The saturation rate varies predictably between regimes, increasing with signal strength and window size and decreasing with autocorrelation. Power is nearly identical between trend separation structures within each cell, except on small windows, in which case fixed rate separation is detected at a higher rate than the others. Localization accuracy is uniformly high, indicating that the test orders effect sizes consistently, but this result must be interpreted relative to available power. For example, the combination of very low power with very high localization accuracy on small windows indicates that the test is reliable in the few instances where it identifies a local difference, but the absence of a detection carries much less information. Across all cells, maxima for precision and NPV are stable around 0.865, suggesting adequate localization is available at some effect size threshold, but the maximizing threshold varies by window size with medians of $c = 0.45, 0.32, 0.25$ for $s = 50, 100, 150$, respectively.

Lastly, Table 2 shows average realized false discovery rates across all simulated datasets. Owing to the method of data generation δ_t is never exactly zero on any given window; null effects correspond instead to small values of δ_t . We compute the realized false discovery rate for a given simulated dataset as the share of rejections for which $\delta_t \leq \varepsilon$; the table reports the average and standard error of these rates across all 81,000 simulated datasets. Considering the opportunity for error is constrained when ε is small, and realized FDR could be artificially controlled, we vary the value of ε between 0.001 and 0.050 and report the average per-dataset share of windows with $\delta_t < \varepsilon$. At the largest threshold value, roughly 40% of windows exhibit no effect. Realized false discoveries increase slightly as ε is relaxed, but the rate remains under 2% throughout, well below the nominal 5%. Thus, despite the pointwise anticonservatism of the test shown in Section 3.1, the overall procedure controls false discoveries at an acceptable level for reliable inference.

4 Application

Water quality data from Morro Bay CeNCOOS stations date back to 2016 (Central and Northern California Ocean Observing System, 2026); here, we analyze data from March 2020 through December 2025 after automated and manual quality control and filtering. Figure 1 shows the lo-

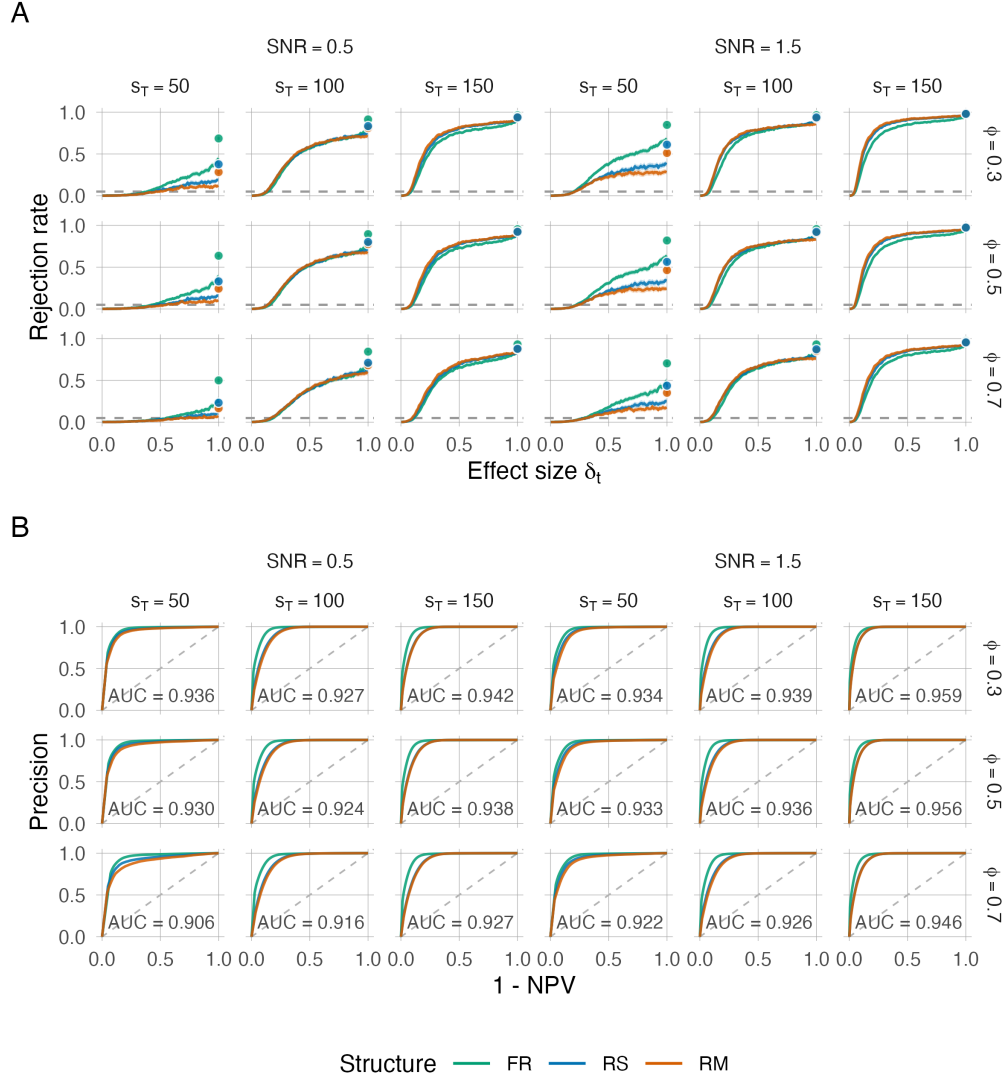


Figure 5: Results of the simulation study, arranged in panels corresponding to the factorial cells $(\bar{\lambda} \times \phi \times s)$. (A) Power curves for each structure as functions of the effect size δ_t and with power at $\delta_t = 1$ shown separately as a point mass; the horizontal dashed line indicates the nominal $\alpha = 0.05$ level. (B) Localization accuracy across windows shown by profiling precision and negative predictive value across $c \in (0, 1]$, with annotations indicating the average AUC across profiles; the diagonal dashed line indicates no discrimination.

Table 2: Average realized false discovery rates across all simulated datasets.

Null effect threshold ε	True null share	Realized FDR (SE)
0.001	0.040	0.0013 (0.0001)
0.005	0.217	0.0084 (0.0003)
0.010	0.282	0.0113 (0.0003)
0.020	0.331	0.0129 (0.0004)
0.050	0.403	0.0156 (0.0004)

cations of the two CeNCOOS monitoring stations: Bay Mouth (BM) and Bay Head (BH). At the Bay Mouth station, water quality data were recorded at 15-minute intervals through May 2023 and at 10-minute intervals thereafter; at the Bay Head station, data were recorded at 15-minute intervals through June 2023 and at 10-minute intervals thereafter. We prepared data for analysis in three steps: quality control removals at the native resolution, aggregation to hourly resolution and pre-smoothing to remove tidal and diurnal periodicity, and downsampling to 6-hour resolution (*i.e.*, four observations per day). We describe this procedure in detail below before reporting the results of our analysis of decoupling patterns over the study period.

4.1 Quality control

Raw sensor data are subject to measurement error, instrument drift, and periodic maintenance interruptions. Sensors at each monitoring station are physically cleaned approximately once per month, but even so, sensor drift can occur for extended periods between cleanings. Data are reviewed for retrospective quality control on an approximately annual basis; this process consists of visual identification and removal of periods subject to obvious drift or malfunction, followed by removal of data collected shortly before, during, and shortly after equipment cleanings. Next, we implemented a series of automated tests from the National Oceanic and Atmospheric Administration’s Quality Assurance/Quality Control of Real-Time Oceanographic Data (QARTOD) protocol (Bushnell et al., 2019). Briefly, automated tests detect measurements exceeding reasonable limits (“range” test), changing at implausible rates (“rate of change” test), remaining constant for extended periods (“flatline” test), or deviating from adjacent measurements by an unrealistic magnitude (“spike” test). Thresholds for each test were set at parameter-specific values, and observations failing one or more tests were removed from the analysis. After quality control, both pH and DO were available simultaneously for roughly 70-81% of the study period at each station and spread over a total of 63 contiguous blocks between the two stations. We removed blocks less than five

days in length, leaving 56 blocks of usable data.

4.2 Tidal presmoothing and tuning parameters

Both pH and DO exhibit strong periodicity at tidal and diurnal frequencies that are not of direct interest in our analysis. Modeling periodicity explicitly would have required accounting for estimation error in the downstream inference, which poses challenges similar to those discussed in Section 1.2. We instead used a band-stop filter designed to remove periodicity prior to analysis. In this approach, we first aggregated data to hourly resolution and applied a discrete Fourier transformation; hourly resolution yields a regular grid needed for the transform and places the Fourier frequencies in cycles per hour so that tidal constituents are easily identifiable. We then replaced coefficients within bands spanning periods of roughly 11.5-13.2 and 22-28 hours — covering the principal semidiurnal and diurnal constituents (M2, S2, N2, K1, O1) — by zeroes and inverted the transform.

For the decoupling analysis, we sought to correlate moving averages taken over at most a few days over windows of at most a few weeks. Although results in Section 3.2 indicate that inference becomes conservative on shorter windows, detections based on longer periods may fail to match the expected resolution of decoupling patterns in real data. Exploratory analysis of standardized data, such as the visualization in Figure 1, suggests that decoupling occurs on the order of weeks. We then sought to keep h small, considering the rolling correlation test degrades in an obvious way as h grows: if the moving averages become too smooth, rolling correlations vacillate between -1 and 1 and reflect only whether trends agree in direction over any given window. This produces competing pressures: modeling the data at high temporal resolution effectively increases s , and thus power, over a fixed time interval; but this also requires increasing h to capture moving averages at the same scale, and thus risks degrading the test. To balance these considerations, we downsampled the data to 6-hour resolution and fixed the smoothing bandwidth and local window size at $h = 4$ and $s = 60$, which correspond to 24-hour moving averages over a 15-day period.

4.3 Decoupling patterns

We first return to the motivating example from Section 1 — the 52-day period from the Bay Mouth location with a visually apparent decoupling event shown in Figure 1 — to illustrate the inference procedure. On this data, the method rejects the local affine similarity hypothesis on 13 consecutive windows covering a total of 18.75 days; Figure 6 highlights both the pointwise rejections

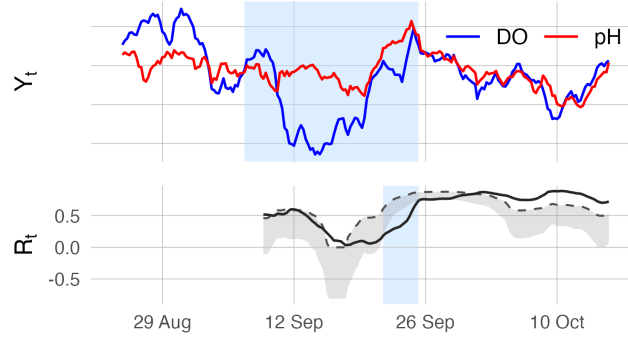


Figure 6: Illustration of inferred local decoupling on the example data from Section 1. Top, moving averages with interval underlying flagged windows highlighted; bottom, rolling correlation series, with estimated local population correlation (dashed line) and one-sided rejection region after BY adjustment (ribbon) shown in background.

in the rolling correlation series and the underlying interval in the moving average series. Visually, the right endpoint of the inferred interval aligns almost exactly with the apparent transition time between decoupled and coupled states. The left endpoint excludes a period showing some visual separation, but during an underdetermined regime where signal is low while pH is relatively flat, so the inference does not rule out the possibility that the decoupling period began earlier.

Extending the inference procedure to the full study period reveals that decouplings concentrate in spring with a peak in May at both locations and secondary peaks in autumn at Bay Mouth (September and October) and in autumn and winter at Bay Head (October through January). Marginally, decouplings appear more frequent at the Bay Mouth location; the detection rate is 2.8 times higher at the Bay Mouth location compared with the Bay Head location. These results are shown in Figure 7. We observe that low-power regimes appear throughout the study period: by Proposition 1 and under the operating rejection threshold in the application, the inference procedure cannot identify an effect when $\hat{\rho} < 3.32\sqrt{\hat{V}_t/s}$, which occurs in over half of the pooled record. Reported detection rates therefore likely understate the frequency of decoupling.

It is worth noting that quality control produced removals of extended periods at both locations, and that these removals are not uniformly distributed in time; we summarize temporal coverage in detail in Appendix C. Nonetheless, neither pattern can be explained away by uneven temporal coverage. At the Bay Head location, data availability during the spring peak is near average compared with other months, and at the Bay Mouth location, the months with the highest detection rates are in fact the least represented (May and June contribute only 3.8% of available data but 19.6% of detections). Overall, data are available from the Bay Mouth location during 36.8% of the

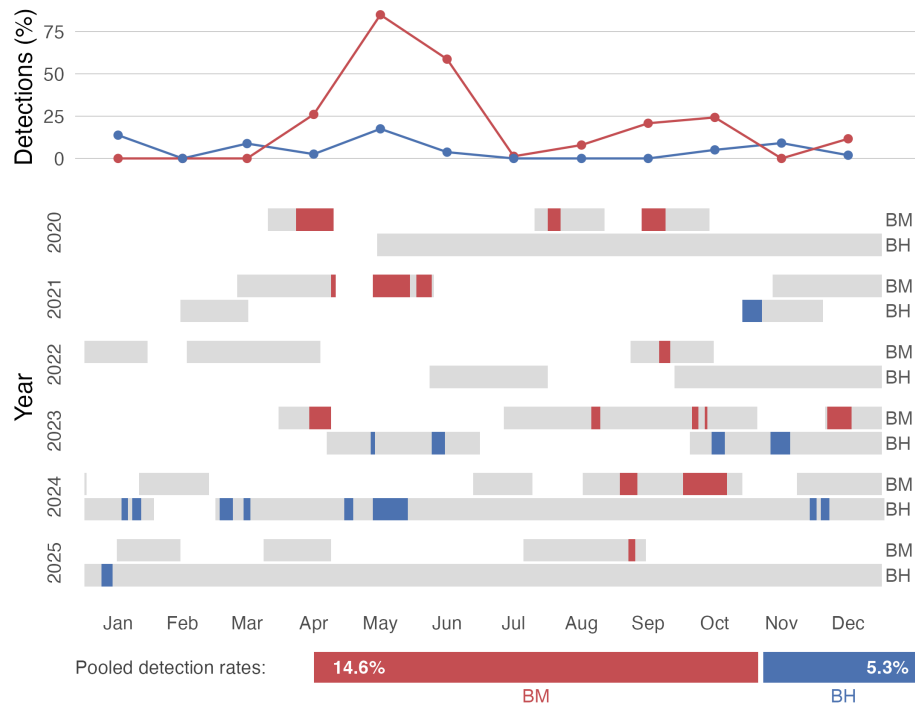


Figure 7: Decoupling patterns in Morro Bay during 2020-2025 across two monitoring locations. The center panel shows available data in each year (*i.e.*, contiguous blocks after quality control) aligned by day of year, with “lanes” within each year corresponding to monitoring location, and intervals flagged as decoupled shaded in color (blue, BH; red, BM). The top panel shows, by month and location, the percentage of available windows across all years flagged as decoupled; a peak occurs in May at both locations. Bottom, pooled detection rates as percentages of available windows across all years by location.

study period, compared with 61.6% at the Bay Head location, and the gap falls disproportionately in months with higher detection rates. Potential bias from missing data would therefore attenuate, rather than inflate, the reported contrast between locations.

The presence of inferred decouplings suggests that noise in the pH-DO relationship is not homogeneous in time, but aggregation conceals the variation. The relationship between pH and DO is practically identical between coupling regimes; it is only when temporal structure is accounted for that visible differences emerge on local scales. Conditional on coupling regime and local structure, underlying data generally exhibit lower correlation on windows flagged as decoupled, but because the test procedure evaluates correlations against local expectations, those window correlations do not distinguish coupling regimes uniformly. Nonetheless, local prediction error based on the pH-DO relationship is elevated by 23-48% in the decoupled regime, depending on location. On the pH scale, this translates to an increase of roughly 0.02-0.03 units. Figure 8 provides an illustration contrasting the difference in the aggregate relationship with that between local relationships, and the corresponding increase in local prediction error estimated by cross-validation. Taken together, this result and the seasonal and spatial patterns suggest that, for example, a pH reconstruction reporting a single record-wide error understates uncertainty locally, and data recorded at the Bay Mouth location in spring are relatively more likely to be affected.

5 Discussion

This work provides a new procedure for evaluating trend divergence in time series data. The method is motivated by an oceanographic application in which common metabolic processes of production and respiration often produce strong coupling between pH and DO, but temperature, buffering, mixing, and gas exchange can alter local levels and amplitudes differently. This coupling is observable in first-order structure but intermittent over extended periods, and the relative location/scale alignment of trends shifts over time. To detect periods of decoupling under unknown local realignment, we develop an affine-invariant test of local trend similarity and apply a simultaneous inference adjustment controlling false discoveries under arbitrary dependence across overlapping windows to identify local periods of decoupling.

Applied to CeNCOOS records from Morro Bay during 2020-2025, the procedure reveals that pH-DO decoupling is structured in both season and location. The observed timing and location of decoupling provide useful targets for process-based investigation, but do not identify a mechanism.

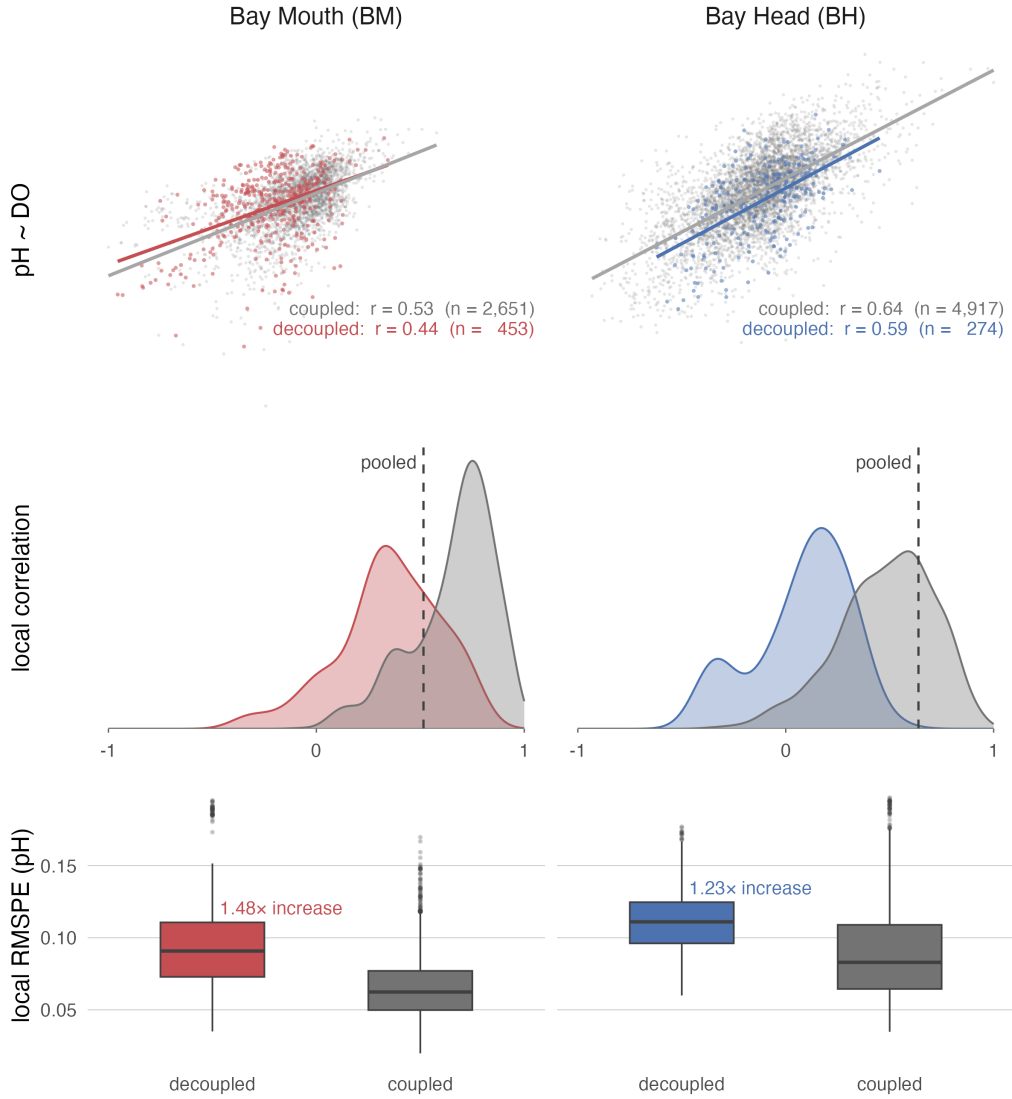


Figure 8: pH-DO relationship, conditional on coupling regime, for Bay Mouth (left column) and Bay Head (right column) locations. Within each column: top, global DO-pH relationships by coupling regime, with regression lines and annotations showing global correlations and sample sizes in each regime; middle, distributions of local between-series correlations within windows used in the decoupling analysis, by regime; bottom, cross-validated root mean square error (RMSPE) for predicting pH from DO, on the scale of pH units, for each regime.

The stronger signal at Bay Mouth is compatible with greater exposure to coastal water masses. The spring peak of decoupling is consistent with seasonal upwelling-associated inputs and water mixing, and the fall decoupling is associated with warmer waters and potentially increased biological production. The secondary pattern at Bay Head likely reflects enclosed estuarine processes, as well as high levels of atmospheric gas exchange in this very shallow estuarine system. Surprisingly, decoupling that might have been expected from changes to buffering with freshwater input was not observed during the intermittently wet-season in Central California, typically between January and April. Ultimately, coupling and decoupling could result from several processes happening simultaneously, and to different degrees.

There are both practical and methodological limitations to our work that deserve mention. Practically, data gaps are common in environmental records, mostly due to bio-fouling and maintenance, and reduce opportunity to assess interannual variability in decoupling patterns. Additionally, decoupling is not captured on time intervals shorter than the scale chosen for analysis. The smoothing bandwidth and local window size constrain the temporal scale of inference; appropriate choices should be informed by both application context and native sampling resolution. Concerning the method itself, our empirical experiments demonstrate its central vulnerability: plug-in estimators introduce additional error that renders the local test pointwise anticonservative. However, the reliability of the test is only a concern for rejections falling near the significance threshold, and our power experiments showed that after simultaneous inference adjustment the false discovery rate is controlled well below nominal levels. In principle, correcting the pointwise anticonservatism through an adjusted variance estimator would improve the method, but the truncation of local trend variance estimates at zero makes this extension less straightforward than it initially appears and is left for future work.

Considered broadly, the application demonstrates why a single pH-DO relationship can be inadequate for estimation and uncertainty quantification. While the local prediction error for pH from DO is modest in absolute terms (0.02-0.03 pH units) and should not be overstated as a biological or management effect, the error is nevertheless appreciably larger than the short-term repeatability of the pH sensor (< 0.001), indicating a robust predictive uncertainty. Thus, reconstructions based on DO or a single record-wide prediction error should account for greater uncertainty in the seasons and locations where decoupling is more likely. While direct observation can be locally informative in this respect, our approach provides an avenue for identifying patterns systematically and could, in principle, be applied across broad geographical extents. Ocean observing systems provide ex-

tensive historical records at regional, continental, and global scales; analyses of such data offer a promising direction for future work.

CRediT authorship contributions

Trevor D. Ruiz: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. **Aléa J. Seifert:** Data curation, Formal analysis, Software, Validation, Writing – review & editing. **Emma Hamilton:** Software, Validation. **Catherine M. Mispagel:** Software, Validation. **Jose Garcia:** Data curation. **Emily E. Bockmon:** Conceptualization, Data curation, Investigation, Writing – review & editing.

Acknowledgments

The authors thank Andres Rocha Jayasinha for assistance testing software implementations, Otis Hunt for assistance deploying simulation experiments on HPC infrastructure, and Julia Schedler for early input on the project.

Funding

This work was supported by the Santa Rosa Creek Foundation under a mini-grant administered by the Center for Coastal Marine Sciences at California Polytechnic State University, and by the Central and Northern California Ocean Observing System (CeNCOOS) via NOAA Integrated Ocean Observing System (IOOS) grant number NA21NOS0120090.

Data and code availability

CeNCOOS data are publicly available at <https://erddap.cencoos.org/erddap>. The QARTOD quality-control procedures applied to the CeNCOOS data are implemented at <https://github.com/ruizt/mb-qartod> (currently private). The lomad R package implementing the method is available at <https://github.com/ruizt/lomad-package> and archived at <https://doi.org/10.5281/zenodo.21695840>, which resolves to the most recent release; version 0.1.0, used for the results reported here, is archived at <https://doi.org/10.5281/zenodo.23323394>. Code reproducing the simulation studies and the Morro Bay analysis is available at <https://github.com/ruizt/lomad-analysis>, and the raw data, quality-controlled data supporting the analysis, and simulation study results are archived at <https://doi.org/10.5281/zenodo.20388153>.

References

- S. R. Alin, R. A. Feely, A. G. Dickson, J. M. Hernández-Ayón, L. W. Juranek, M. D. Ohman, and R. Goericke. Robust empirical relationships for estimating the carbonate system in the southern California Current System and application to CalCOFI hydrographic cruise data (2005–2011). *Journal of Geophysical Research: Oceans*, 117:C05033, 2012.
- O. Andrews, E. Buitenhuis, C. Le Quéré, and P. Suntharalingam. Biogeochemical modelling of dissolved oxygen in a changing ocean. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2102):20160328, 2017.
- A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087, 2009.
- S. E. Bartoloni, R. K. Walter, S. N. Wewerka, J. Higgins, J. K. O’Leary, and E. E. Bockmon. Spatial distribution of seawater carbonate chemistry and hydrodynamic controls in a low-inflow estuary. *Estuarine, Coastal and Shelf Science*, 281:108195, 2023.
- H. Baumann and E. M. Smith. Quantifying metabolically driven pH and oxygen fluctuations in US nearshore habitats at diel to interannual time scales. *Estuaries and Coasts*, 41(4):1102–1117, 2018.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.
- P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 2009.
- M. Bushnell, C. Waldmann, S. Seitz, E. Buckley, M. Tamburri, J. Hermes, E. Heslop, and A. Lara-Lopez. Quality assurance of oceanographic observations: Standards and guidance adopted by an international partnership. *Frontiers in Marine Science*, 6:706, 2019.
- W.-J. Cai, X. Hu, W.-J. Huang, M. C. Murrell, J. C. Lehrter, S. E. Lohrenz, W.-C. Chou, W. Zhai, J. T. Hollibaugh, Y. Wang, P. Zhao, X. Guo, K. Gundersen, M. Dai, and G.-C. Gong. Acidification of subsurface coastal waters enhanced by eutrophication. *Nature Geoscience*, 4(11):766–770, 2011.
- B. R. Carter, R. A. Feely, N. L. Williams, A. G. Dickson, M. B. Fong, and Y. Takeshita. Updated

- methods for global locally interpolated estimation of alkalinity, pH, and nitrate. *Limnology and Oceanography: Methods*, 16(2):119–131, 2018.
- B. R. Carter, H. C. Bittig, A. J. Fassbender, J. D. Sharp, Y. Takeshita, Y.-Y. Xu, M. Álvarez, R. Wanninkhof, R. A. Feely, and L. Barbero. New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, 19(12):785–809, 2021.
- Central and Northern California Ocean Observing System. CeNCOOS, 2026. URL <https://www.cencoos.org>. Accessed 2026-08-01.
- J.-E. Choi and D. W. Shin. Nonparametric estimation of time varying correlation coefficient. *Journal of the Korean Statistical Society*, 50:333–353, 2021.
- R. Dahlhaus. Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37, 1997.
- R. Dahlhaus. Locally stationary processes. In *Handbook of Statistics*, volume 30, pages 351–413. Elsevier, 2012.
- H. Dehling, D. Vogel, M. Wendler, and D. Wied. Testing for changes in Kendall’s tau. *Econometric Theory*, 33(6):1352–1386, 2017.
- H. Dette, W. Wu, and Z. Zhou. Change point analysis of correlation in non-stationary time series. *Statistica Sinica*, 29(2):611–643, 2019.
- J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1996.
- C. J. Gobler and H. Baumann. Hypoxia and acidification in ocean ecosystems: coupled dynamics and effects on marine life. *Biology Letters*, 12(5):20150976, 2016.
- P. Hall and J. D. Hart. Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, 85(412):1039–1049, 1990.
- P. Hall and I. Van Keilegom. Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 65(2):443–456, 2003.

- W. Härdle and J. S. Marron. Semiparametric comparison of regression curves. *The Annals of Statistics*, 18(1):63–89, 1990.
- M. Khismatullina and M. Vogt. Nonparametric comparison of epidemic time trends: The case of covid-19. *Journal of econometrics*, 232(1):87–108, 2023.
- M. Khismatullina and M. Vogt. Multiscale comparison of nonparametric trend curves. *Journal of Business & Economic Statistics*, 2026. Forthcoming.
- L. Kwiatkowski, O. Torres, L. Bopp, O. Aumont, M. Chamberlain, J. R. Christian, J. P. Dunne, M. Gehlen, T. Ilyina, J. G. John, A. Lenton, H. Li, N. S. Lovenduski, J. C. Orr, J. Palmieri, Y. Santana-Falcón, J. Schwinger, R. Séférian, C. A. Stock, A. Tagliabue, Y. Takano, J. Tjiputra, K. Toyama, H. Tsujino, M. Watanabe, A. Yamamoto, A. Yool, and T. Ziehn. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, 17(13):3439–3470, 2020.
- J. L. Largier. Recognizing low-inflow estuaries as a common estuary paradigm. *Estuaries and Coasts*, 46(8):1949–1970, 2023.
- D. Li, B. Wang, H. Jin, Y. Miao, Q. Sun, H. Lin, H. Li, Q. Liu, F. Zhou, and J. Chen. Decoupling of high-resolution surface pH and DO reveals temporal algal bloom dynamics on the East China Sea. *Water Research*, 261:122030, 2024.
- C. Park, A. Vaughan, J. Hannig, and K.-H. Kang. SiZer analysis for the comparison of time series. *Journal of Statistical Planning and Inference*, 139(12):3974–3988, 2009.
- M. B. Priestley and M. T. Chao. Non-parametric function fitting. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 34(3):385–392, 1972.
- J. N. K. Rao and A. J. Scott. A simple method for the analysis of clustered binary data. *Biometrics*, 48(2):577–585, 1992.
- T. D. Ruiz, A. J. Seifert, E. Hamilton, C. M. Mispagel, and A. E. R. Jayasinha. lomad: Local moving average decoupling, 2026. R package version 0.1.0.
- J. Su, W.-J. Cai, J. Brodeur, B. Chen, N. Hussain, Y. Yao, C. Ni, J. M. Testa, M. Li, X. Xie, W. Ni, K. M. Scaboo, Y.-Y. Xu, Y. Cai, W. Zhang, and E. W. Boyer. Supply-controlled calcium

- carbonate dissolution decouples the seasonal dissolved oxygen and pH minima in Chesapeake Bay. *Limnology and Oceanography*, 66(10):3796–3810, 2021.
- W. G. Sunda and W.-J. Cai. Eutrophication induced CO₂-acidification of subsurface coastal waters: Interactive effects of temperature, salinity, and atmospheric pCO₂. *Environmental Science & Technology*, 46(19):10651–10659, 2012.
- M. Taherkhani, S. Vitousek, R. K. Walter, J. O’Leary, and A. P. Khodadoust. Flushing time variability in a short, low-inflow estuary. *Estuarine, Coastal and Shelf Science*, 284:108277, 2023.
- U.S. Integrated Ocean Observing System. IOOS catalog, 2026. URL <https://data.ioos.us>. Accessed 2026-08-01.
- M. Vogt and H. Dette. Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713–740, 2015.
- R. B. Wallace, H. Baumann, J. S. Grear, R. C. Aller, and C. J. Gobler. Coastal ocean acidification: The other eutrophication problem. *Estuarine, Coastal and Shelf Science*, 148:1–13, 2014.
- R. K. Walter, E. J. Rainville, and J. K. O’Leary. Hydrodynamics in a shallow seasonally low-inflow estuary following eelgrass collapse. *Estuarine, Coastal and Shelf Science*, 213:160–175, 2018.
- D. Wied, H. Dehling, M. van Kampen, and D. Vogel. A fluctuation test for constant Spearman’s rho with nuisance-free limit distribution. *Computational Statistics & Data Analysis*, 76:723–736, 2014.
- W. B. Wu and Z. Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 69(3):391–410, 2007.
- T. Zhang and Y. Shao. Time-varying correlation for noncentered nonstationary time series. *Statistica Sinica*, 34(4):2141–2160, 2024.
- Z. Zhao. Inference for local autocorrelations in locally stationary models. *Journal of Business & Economic Statistics*, 33(2):296–306, 2015.

A Proofs

A.1 Asymptotic framework and notation

For reference, we restate the asymptotic framework and key notation introduced in Theorem 1. Consider a triangular array $\{Y_{t,T} \in \mathbb{R}^2 : t = 1, \dots, T\}_{T \in \mathbb{N}}$. Fix $u \in (0, 1)$ and define $W_T := \{t_0 - s_T + 1, \dots, t_0\}$ to be the window of width s_T with right endpoint $t_0 = \lfloor uT \rfloor$. For $y \in \mathbb{R}^2$ let $h(y) = (y_1, y_2, y_1^2, y_2^2, y_1 y_2)^T$, and define the moment process $H_{t,T} := h(Y_{t,T})$. Denote the sample moments within the window by:

$$\hat{\theta}_T(u) = \frac{1}{s_T} \sum_{t \in W_T} H_{t,T}$$

We write the local sample correlation as $R_T(u) = g(\hat{\theta}_T(u))$ where g is defined by

$$g(\theta) := \frac{C_{12}(\theta)}{\sqrt{C_{11}(\theta)C_{22}(\theta)}}$$

with $C_{11}(\theta) = \theta_3 - \theta_1^2$ and $C_{22}(\theta) = \theta_4 - \theta_2^2$ and $C_{12}(\theta) = \theta_5 - \theta_1 \theta_2$.

A.2 Regularity conditions

Assume that for each $u \in (0, 1)$ there exists a strongly stationary process $\{Y_t(u) \in \mathbb{R}^2\}_{t \in \mathbb{Z}}$ with autocovariance function $\Gamma(l; u)$ and associated moments $H_t(u) := h(Y_t(u))$ satisfying:

$$(A1) \quad \Gamma(0; u) \succ 0$$

$$(A2) \quad s_T \sup_{t \in W_T} \mathbb{E} \|H_{t,T} - H_t(u)\|^2 \rightarrow 0$$

$$(A3) \quad \text{the process } \{H_t(u)\}_{t \in \mathbb{Z}} \text{ is strongly mixing with coefficients } \alpha_H(k; u) \text{ satisfying, for some } \delta > 0,$$

$$\sup_u \sum_{k=1}^{\infty} \alpha_H(k; u)^{\delta/(2+\delta)} < \infty \quad \text{and} \quad \sup_u \mathbb{E} [\|H_t(u)\|^{2+\delta}] < \infty$$

A.3 Proof of Theorem 1

Proof. Suppose the conditions of Theorem 1 hold. We write

$$\sqrt{s_T} (\hat{\theta}_T(u) - \theta(u)) = \frac{1}{\sqrt{s_T}} \sum_{t \in W_T} (H_{t,T} - \theta(u)) = A_T + B_T$$

where

$$A_T = \frac{1}{\sqrt{s_T}} \sum_{t \in W_T} (H_{t,T} - H_t(u)) \quad \text{and} \quad B_T = \frac{1}{\sqrt{s_T}} \sum_{t \in W_T} (H_t(u) - \theta(u))$$

It suffices to show that as $T, s_T \rightarrow \infty$,

$$A_T \xrightarrow{p} 0 \quad \text{and} \quad B_T \xrightarrow{d} N(0, \Sigma(u))$$

For the A_T term, let $D_{t,T} = H_{t,T} - H_t(u)$ and write $A_T = \frac{1}{\sqrt{s_T}} \sum_{t \in W_T} D_{t,T}$. Using the bound

$$\left\| \sum_{t \in W_T} D_{t,T} \right\|^2 \leq s_T \sum_{t \in W_T} \|D_{t,T}\|^2$$

we obtain

$$\mathbb{E}\|A_T\|^2 = \frac{1}{s_T} \mathbb{E} \left\| \sum_{t \in W_T} D_{t,T} \right\|^2 \leq \sum_{t \in W_T} \mathbb{E}\|D_{t,T}\|^2 \leq s_T \sup_{t \in W_T} \mathbb{E}\|H_{t,T} - H_t(u)\|^2.$$

By condition (A2), the right-hand side converges to zero. Hence $\mathbb{E}\|A_T\|^2 \rightarrow 0$ and therefore

$$A_T \xrightarrow{p} 0$$

For the B_T term, note that $\{H_t(u)\}_{t \in \mathbb{Z}}$ is strongly stationary by construction and, by condition (A3), strongly mixing with finite $(2 + \delta)$ moments. Hence, by a standard multivariate central limit theorem for strongly mixing sequences (*e.g.*, Brockwell and Davis, 2009, Theorem 7.1.1), together with Cramér–Wold, we obtain

$$\frac{1}{\sqrt{s_T}} \sum_{t=1}^{s_T} (H_t(u) - \theta(u)) \xrightarrow{d} N(0, \Sigma(u))$$

where $\Sigma(u) = \sum_{l \in \mathbb{Z}} \Gamma_H(l; u)$ and $\Gamma_H(l; u) = \text{Cov}(H_t(u), H_{t-l}(u))$. Since $\{H_t(u)\}$ is stationary, the same limit holds for the sum over the shifted window W_T , and therefore

$$B_T \xrightarrow{d} N(0, \Sigma(u))$$

Finally, combining the limits and applying Slutsky's theorem yields

$$\sqrt{s_T} \left(\hat{\theta}_T(u) - \theta(u) \right) = A_T + B_T \xrightarrow{d} N(0, \Sigma(u)),$$

as claimed. $\square$

Remark. The condition $s_T/T \rightarrow 0$ is not used directly in the argument; rather, it underlies the stationary local approximation on which assumption (A2) is based by ensuring that the entire window W_T remains asymptotically concentrated around the rescaled location u .

A.4 Proof of Theorem 2

Proof. Suppose the conditions of Theorem 1 hold. We have

$$\sqrt{s_T} \left(\hat{\theta}_T(u) - \theta(u) \right) \xrightarrow{d} N(0, \Sigma(u)).$$

Since $\theta(u) = \mathbb{E}[H_t(u)]$ by definition,

$$C_{11}(\theta(u)) = \theta_3(u) - \theta_1(u)^2 = \mathbb{E}[Y_{1t}(u)^2] - (\mathbb{E}[Y_{1t}(u)])^2 = \text{Var}[Y_{1t}(u)] = \Gamma_{11}(0; u)$$

and similarly $C_{22}(\theta(u)) = \Gamma_{22}(0; u)$. By condition (A1), $\Gamma(0; u)$ is positive definite, so in particular $C_{11}(\theta(u)) > 0$ and $C_{22}(\theta(u)) > 0$. Therefore g is well-defined and continuously differentiable in a neighborhood of $\theta(u)$, and by the multivariate delta method we obtain

$$\sqrt{s_T} \left(g \left(\hat{\theta}_T(u) \right) - g(\theta(u)) \right) \xrightarrow{d} N \left(0, [\nabla g(\theta(u))]^T \Sigma(u) [\nabla g(\theta(u))] \right)$$

Substituting $R_T(u) = g \left(\hat{\theta}_T(u) \right)$, $\rho(u) = g(\theta(u))$, and $V(u) = [\nabla g(\theta(u))]^T \Sigma(u) [\nabla g(\theta(u))]$ yields

$$\sqrt{s_T} (R_T(u) - \rho(u)) \xrightarrow{d} N(0, V(u))$$

as claimed. $\square$

Remark. The gradient in Theorem 2 may be written explicitly. For ease of notation define $A := C_{12}(\theta)$, $B := C_{11}(\theta)$, $D := C_{22}(\theta)$ so that $g(\theta) = A(BD)^{-1/2}$. Then

$$\nabla A = (-\theta_2, -\theta_1, 0, 0, 1)^T, \quad \nabla B = (-2\theta_1, 0, 1, 0, 0)^T, \quad \nabla D = (0, -2\theta_2, 0, 1, 0)^T,$$

and we have by the product rule and chain rule

$$\nabla g(\theta) = (BD)^{-1/2} \nabla A - \frac{1}{2} A(BD)^{-3/2} (D \nabla B + B \nabla D)$$

Equivalently,

$$\nabla g(\theta) = \begin{pmatrix} -\theta_2(BD)^{-1/2} + \theta_1 AD(BD)^{-3/2} \\ -\theta_1(BD)^{-1/2} + \theta_2 AB(BD)^{-3/2} \\ -\frac{1}{2} AD(BD)^{-3/2} \\ -\frac{1}{2} AB(BD)^{-3/2} \\ (BD)^{-1/2} \end{pmatrix}$$

A.5 Proof of Proposition 1

Proof. For the proof we reintroduce the local asymptotic notation and write $Y_{kt}(u) = m_{kt}(u) + \eta_{kt}$ for $k = 1, 2$. Assume the conditions of the proposition hold and define the local signal means $m_k(u) := \frac{1}{s_T} \sum_{t \in W_T} m_{kt}(u)$. Since the sample correlation $g(\theta)$ is invariant to location shifts applied separately to each component, we may without loss of generality work with the centered process $\tilde{Y}_{kt}(u) := Y_{kt}(u) - m_k(u)$, which has signal components $\tilde{s}_{kt}(u) = m_{kt}(u) - m_k(u)$ satisfying $\frac{1}{s_T} \sum_{t \in W_T} \tilde{s}_{kt}(u) = 0$, $\frac{1}{s_T} \sum_{t \in W_T} \tilde{s}_{kt}(u)^2 = \tau_k^2(u)$, and $\frac{1}{s_T} \sum_{t \in W_T} \tilde{s}_{1t}(u) \tilde{s}_{2t}(u) = r(u) \tau_1(u) \tau_2(u)$. The population moment vector of the centered process is

$$\theta(u) \approx \frac{1}{s_T} \sum_{t \in W_T} \mathbb{E}[h(\tilde{Y}_t(u))] = \begin{pmatrix} 0 \\ 0 \\ \tau_1^2(u) + \sigma_1^2 \\ \tau_2^2(u) + \sigma_2^2 \\ r(u) \tau_1(u) \tau_2(u) \end{pmatrix}$$

which yields

$$C_{11}(\theta(u)) = \tau_1^2(u) + \sigma_1^2, \quad C_{22}(\theta(u)) = \tau_2^2(u) + \sigma_2^2, \quad C_{12}(\theta(u)) = r(u) \tau_1(u) \tau_2(u)$$

and therefore

$$\rho(u) \approx \frac{r(u) \tau_1(u) \tau_2(u)}{\sqrt{(\tau_1^2(u) + \sigma_1^2) (\tau_2^2(u) + \sigma_2^2)}}$$

Now write

$$A := r(u)\tau_1(u)\tau_2(u), \quad B := \tau_1^2(u) + \sigma_1^2, \quad D := \tau_2^2(u) + \sigma_2^2$$

Since $\theta_1(u) = \theta_2(u) = 0$ in the centered parameterization, the first two coordinates of $\nabla g(\theta(u))$ vanish and the gradient in Theorem 2 reduces to

$$\nabla g(\theta(u)) = \begin{pmatrix} 0 \\ 0 \\ -\frac{1}{2}AD(BD)^{-3/2} \\ -\frac{1}{2}AB(BD)^{-3/2} \\ (BD)^{-1/2} \end{pmatrix}$$

In order to write the entries of $\Sigma(u)$, define

$$L_k := \sum_{l \in \mathbb{Z}} \gamma_k(l), \quad Q_k := \sum_{l \in \mathbb{Z}} \gamma_k(l)^2, \quad Q_{12} := \sum_{l \in \mathbb{Z}} \gamma_1(l)\gamma_2(l).$$

The entries of $\Sigma(u)$ relevant for the delta method variance calculation are the lower-right 3×3 block. Using the window-averaged approximation

$$\Sigma_{ij}(u) \approx \sum_{\ell \in \mathbb{Z}} \Gamma_{H,ij}^W(\ell; u) \quad \text{where} \quad \Gamma_{H,ij}^W(\ell; u) := \frac{1}{s_T} \sum_{t \in W_T} \text{Cov}(\tilde{H}_{i,t}(u), \tilde{H}_{j,t-\ell}(u))$$

and $\tilde{H}_t(u) := h(\tilde{Y}_t(u))$, we obtain, using Wick's theorem for the fourth-order Gaussian moments and the independence of η_1, η_2 ,

$$\Sigma_{(3,4,5)}(u) \approx \begin{pmatrix} 2Q_1 + 4\tau_1^2(u)L_1 & 0 & 2AL_1 \\ 0 & 2Q_2 + 4\tau_2^2(u)L_2 & 2AL_2 \\ 2AL_1 & 2AL_2 & Q_{12} + \tau_1^2(u)L_2 + \tau_2^2(u)L_1 \end{pmatrix}$$

Then substitution into $V(u) = \nabla g(\theta(u))^T \Sigma(u) \nabla g(\theta(u))$ gives

$$V(u) \approx \frac{Q_{12}}{BD} + \frac{\tau_2^2(u)L_1}{BD} [1 - r^2(u)u_1(2 - u_1)] + \frac{\tau_1^2(u)L_2}{BD} [1 - r^2(u)u_2(2 - u_2)] + \frac{A^2 Q_1}{2B^3 D} + \frac{A^2 Q_2}{2BD^3}$$

where $u_1 := \tau_1^2(u)/B$ and $u_2 := \tau_2^2(u)/D$ denote the local signal fractions. Under H_0 we have $r(u) = 1$, so that $u_k(2 - u_k) = 1 - (1 - u_k)^2$ with $1 - u_1 = \sigma_1^2/B$ and $1 - u_2 = \sigma_2^2/D$, and $A = \tau_1(u)\tau_2(u)$; the expression reduces to

$$V(u) \approx \frac{Q_{12}}{BD} + \frac{\tau_2^2(u)\sigma_1^4 L_1}{B^3 D} + \frac{\tau_1^2(u)\sigma_2^4 L_2}{BD^3} + \frac{\tau_1^2(u)\tau_2^2(u) Q_1}{2B^3 D} + \frac{\tau_1^2(u)\tau_2^2(u) Q_2}{2BD^3}$$

which establishes the result. □

B Simulations

B.1 Trend separation structures

Here we describe in detail our approach to generating structured trend separation. Given a base pair μ_1, μ_2 generated as described in the main text, we compute $\tilde{\nu}_{it} = \bar{\mu}_t + d(1 - w_t)(\mu_{it} - \bar{\mu}_t)$, where $\bar{\mu}_t := (\mu_{1t} + \mu_{2t})/2$ and d is a scaling factor. We consider three methods of generating the coupling weight w_t ; these correspond to the three “structures” of trend separation described in the paper.

FR The first method distributes separation at evenly spaced times based on a specified separation rate r (local separation events per unit time): we determine $m = \lfloor rT \rfloor$ evenly spaced times $\tau_1, \dots, \tau_m$ and compute

$$w_t = \left[1 - \sum_{j=1}^m f(t - \tau_j) \right]_+$$

where $f(u)$ is a Gaussian density normalized to unit peak.

RS The second method concentrates separation locally but distributes separation events irregularly: we generate a standardized kernel-smoothed Gaussian process G_t with bandwidth b and compute

$$w_t = \Phi(\Phi^{-1}(c) - G_t)$$

where $c \in (0, 1)$ controls the fraction of time the trends spend near the midpoint, *i.e.*, with $w_t > 1/2$.

RM The third method relaxes the constraint $w_t \in [0, 1]$, which forces trends to move symmetrically away from the base trend midpoint during local separation events, and allows trends to separate freely in either direction. We generate G_t as before but instead compute

$$w_t = 1 - \sigma_w G_t$$

where $\sigma_w = 0.5/\Phi^{-1}((1+c)/2)$ is chosen so that $P(|w_t - 1| < 1/2) = c$ and $c \in (0, 1)$ again controls the fraction of time the trends spend near the midpoint.

B.2 Affine transformations

Here we briefly describe our method of generating the coefficients a_t, b_t for the transformation $\nu_2 = a_t + b_t \tilde{\nu}_2$ described in the main text. Both coefficient paths are generated from a common construction. Let $Z_t = \sum_{u \leq t} \varepsilon_u$ with $\varepsilon_u \stackrel{iid}{\sim} N(0, 1)$ denote a Gaussian random walk, and let $\tilde{Z}$ be its kernel smoothing with a normal kernel of bandwidth βn , recentered so that $\tilde{Z}_1 = 0$. We rescale $\tilde{Z}$ so that its displacement over any window of width s meets a fixed budget κ :

$$W_t = \kappa \tilde{Z}_t / \max_{u \leq n-s} |\tilde{Z}_{u+s} - \tilde{Z}_u|$$

The coefficients are then obtained from two independent realizations $W^{(a)}, W^{(b)}$ by setting $a_t = \tilde{\omega}_1 W_t^{(a)}$ and $b_t = 1 + W_t^{(b)}$, where $\tilde{\omega}_1$ denotes the standard deviation of $\tilde{\nu}_1$ over the series, so that the location shift is expressed on the scale of the first trend and the scale factor fluctuates about unity. For our simulation studies, we fix $\beta = 0.5$ and $\kappa = 0.015$ throughout.

The budget κ is what keeps the construction inside H_{0t} : local affine similarity constrains a_t and b_t to be nearly constant within a window but places no restriction across windows, allowing distant windows to realize substantially different transformations. The bandwidth β governs accumulated displacement: larger β yields greater total drift across the series at the same κ .

B.3 Local realignment study

If trends are related by an unknown time-varying affine map, a natural strategy is to estimate that map on each local window, remove it, and then test for local differences between the realigned trends directly. We conducted a small simulation experiment to assess empirically whether such local realignment provides for adequate inference using existing methods; the experiment takes the method of Khismatullina and Vogt (2026) as representative.

For the experiment we construct a series pair to satisfy the local null, so that the trends are affinely similar up to the within-window variation the drift cap permits (median $\delta_t = 0.004$), and essentially every rejection reported is a type I error. We test for trend similarity using the method in the main text at three window sizes ensuring our method is assessed under varying power regimes, and again using the method of Khismatullina and Vogt (2026) as implemented in the **MSinference** package in R with three different preprocessing approaches: global standardization, local realignment of raw series, and local realignment of smoothed series. We generate two series pairs, one in which the trends are identical, and one in which they are related by a drifting affine

map. These series pairs are shown in Figure S1, along with the affine transformation used, and estimation errors associated with local realignments using estimates of the affine map at the “raw” scale (*i.e.*, directly from the series) and at the smoothed scale. Estimates of the map are computed by local regression of one series (or smoothed series) on the other.

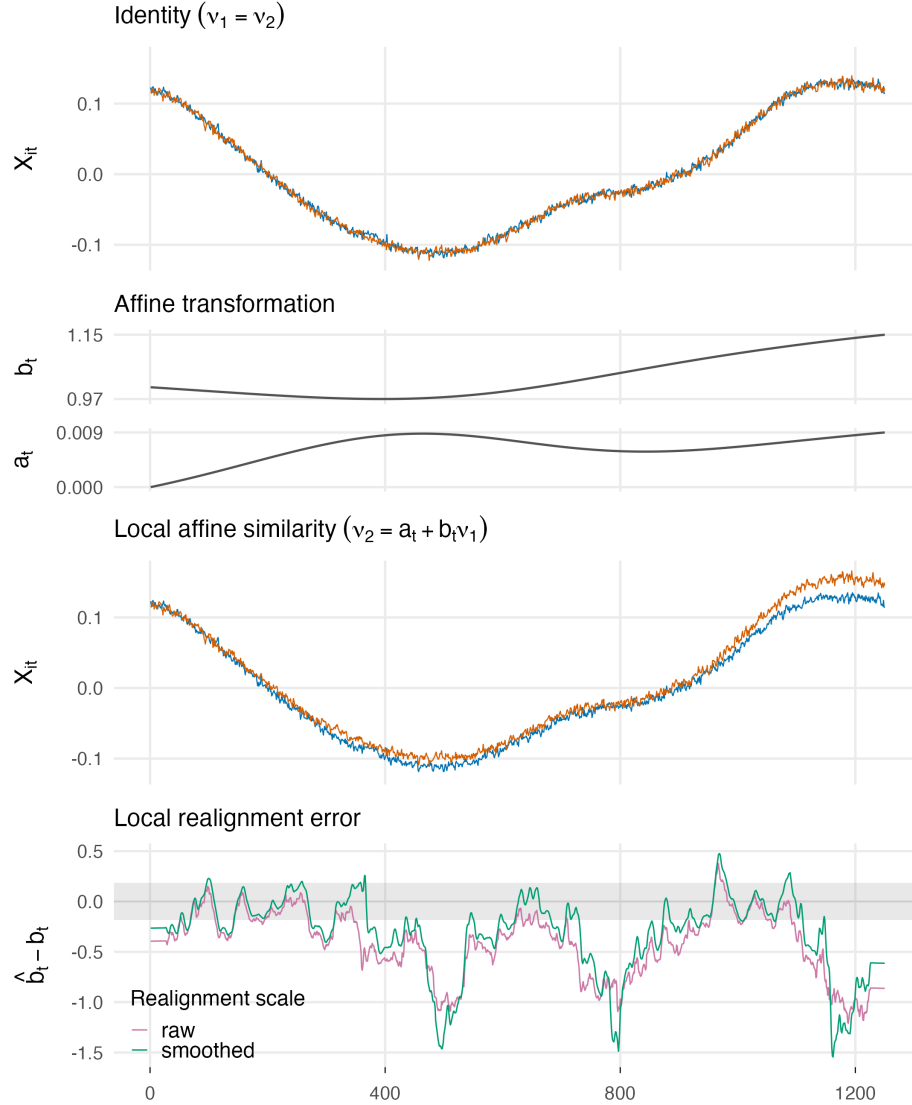


Figure S1: Series used in the realignment study. From top to bottom: series with identical trends; time-varying affine transformation coefficients used to induce local affine similarity with 19% scale drift; series with local affine similarity; estimation errors for the scale coefficient b_t from each realignment strategy, with perfect estimation shown by the solid horizontal line and a band corresponding to the range of b_t over time.

We carried out the comparison using a fixed pair of trends — shown exactly in Figure S1 — but repeated the comparison with a secondary trend pair that exhibited milder scale drift in the affine transformation as a robustness check. The results, shown in Table S1, demonstrate that

global standardization behaves as expected in the absence of a locally affine transformation and correctly flags no trend differences, but that when the transformation is present, local realignment produces falsely inferred differences for 10-26% of intervals. Moreover, the intervals flagged after local realignment span 35-78% of the series. The results under typical scale drift show that the failure of local realignment persists independently of whether global standardization leads to correct inference, *i.e.*, even when a single global map suffices. Consistent with Figure S1, this suggests that unaccounted-for estimation errors in local maps severely compromise simultaneous inference when local realignment is used as a preprocessing strategy.

Table S1: Results of realignment study, repeated at two levels of drift.

<i>(A) Typical 19% drift</i>				
Scenario	Method	Framing	Reject	Flagged
Equality	lomad	$s = 50$	No	0 / 1197
Affine similarity	lomad	$s = 50$	No	0 / 1197
Affine similarity	lomad	$s = 80$	No	0 / 1167
Affine similarity	lomad	$s = 100$	No	0 / 1147
Equality	MSinference	Global standardization	No	0 / 15250
Affine similarity	MSinference	Global standardization	Yes	227 / 15250
Affine similarity	MSinference	Local realignment (raw)	Yes	2670 / 15250
Affine similarity	MSinference	Local realignment (smoothed)	Yes	1453 / 15250
<i>(B) Mild 14% drift</i>				
Scenario	Method	Framing	Reject	Flagged
Equality	lomad	$s = 50$	No	0 / 1197
Affine similarity	lomad	$s = 50$	No	0 / 1197
Affine similarity	lomad	$s = 80$	No	0 / 1167
Affine similarity	lomad	$s = 100$	No	0 / 1147
Equality	MSinference	Global standardization	No	0 / 15250
Affine similarity	MSinference	Global standardization	No	0 / 15250
Affine similarity	MSinference	Local realignment (raw)	Yes	2121 / 15250
Affine similarity	MSinference	Local realignment (smoothed)	Yes	4017 / 15250

C Data analysis

C.1 Coverage and detections

The analysis covers 8,295 windows drawn from 26 contiguous blocks across the two locations between 2020 and 2025. Accounting for both smoothing and rolling correlation window size, underlying intervals span $s + h - 1 = 63$ observations at 6-hour spacing, or 2.25 weeks; consecutive windows share all but one of their observations. Table S2 shows how these windows are distributed in time via per-year counts of available windows and temporal coverage of those windows measured as the number of months with at least one window available for analysis. The table also shows interannual variability in detection rates and the number of episodes (contiguous runs of consecutive detected windows). The analysis described in the main text identifies 31 episodes in total, 16 at the Bay Mouth and 15 at the Bay Head.

Table S2: Temporal coverage, available data, and detections by location and year.

Location	Year	Months covered	Windows	Detection rate	Episodes
Bay Mouth (BM)	2020	6	358	35.8%	3
	2021	6	478	20.1%	3
	2022	6	496	2.6%	1
	2023	8	657	15.7%	5
	2024	9	669	15.2%	3
	2025	7	446	2.5%	1
	All		3104	14.6%	16
Bay Head (BH)	2020	8	913	0.0%	0
	2021	5	261	12.6%	1
	2022	6	584	0.0%	0
	2023	6	628	12.6%	4
	2024	12	1346	10.9%	9
	2025	12	1459	1.0%	1
	All		5191	5.3%	15

There is notable interannual variability in both coverage and detections, although available windows span five months or more in all years for both locations; typical coverage is six to eight months of the year. Table S3 shows data coverage by month, aggregated across year.

Interannual effects may therefore contribute alongside the seasonal and spatial variation described in the main paper.

Table S3: Temporal coverage, available data, and detections by location and month

Month	Bay Mouth (BM)			Bay Head (BH)		
	Windows	Rate	Episodes	Windows	Rate	Episodes
Jan	196	0.0%	0	247	13.8%	3
Feb	199	0.0%	0	170	0.0%	0
Mar	258	0.0%	0	307	8.8%	3
Apr	438	26.0%	3	275	2.5%	1
May	73	84.9%	1	440	17.5%	3
Jun	46	58.7%	1	564	3.7%	1
Jul	236	1.3%	1	494	0.0%	0
Aug	404	7.9%	2	371	0.0%	0
Sep	452	20.8%	4	369	0.0%	0
Oct	334	24.3%	4	612	5.1%	2
Nov	124	0.0%	0	715	9.1%	3
Dec	344	11.6%	1	627	1.9%	1
All	3104	14.6%	16	5191	5.3%	15

C.2 Sensitivity of results to bandwidth and window size

We conducted a brief sensitivity check to explore robustness of analysis conclusions — specifically the seasonal pattern, spatial difference in detection rate, and local deterioration of marginal predictive relationship — to choices of smoothing bandwidth h and window size s . The actual analysis uses $h = 4$ and $s = 60$; we compare these against combinations of $h = 4, 5$ with $s = 50, 60, 70$; bandwidths $h = 4, 5$ correspond to smoothing over 24 and 30 hours; widths $s = 50, 60, 70$ to windows of 12.5, 15 and 17.5 days. At smoothing scales under 24h, residual diurnal periodicity contaminates the trend estimates and the procedure degrades, so we do not consider $h < 4$ at this resolution. For each combination, we compute detection rates by site, number and typical length of decoupling events, Pearson and Spearman correlation (r, ρ respectively) between monthly detection rates at the specified combination with the “reference” cell $h = 4, s = 60$, and the cross-validated RMSPE ratio between coupled and decoupled windows (pooled across locations). The block set is held fixed at the length used in the reference cell. Results are shown in Table S4. While detection rates at each location increase with window size — expected based on the power experiments — the detection ratio between locations remains between 1.8 and 3.0 across combinations. All combinations produce comparable seasonal patterns and local RMSPE inflation.

Table S4: Sensitivity of analysis results to choices of smoothing bandwidth h and window size s .

h	s	Windows	Detection rate			Episodes		Seasonal profile		RMSPE ratio
			BM	BH	BM/BH	Runs	Median	r	ρ	
4	50	8555	10.66%	3.66%	2.91	32	3.25	0.964	0.958	1.44
4	60	8295	14.59%	5.28%	2.76	31	5.00	–	–	1.39
4	70	8035	18.26%	6.10%	2.99	28	6.88	0.966	0.930	1.27
5	50	8529	16.04%	7.43%	2.16	43	5.00	0.929	0.958	1.36
5	60	8269	21.16%	10.80%	1.96	40	7.38	0.901	0.874	1.31
5	70	8009	25.77%	13.97%	1.84	37	9.25	0.891	0.916	1.29