

Changepoint detection methods as a practical tool for assessment of ecological restoration

Mengyi Gong¹, Simone Conte², Tom Jameson³, Rebecca Killick^{1,4}, James C Bull^{5,6*},
Luca Börger^{6,7}, Rachel McCrea¹

1. School of Mathematical Sciences, Lancaster University, Lancaster, LA1 4YF, United Kingdom

2. Mathematics-Driven Innovation Centre and The Maxwell Institute for Mathematical Sciences,
Heriot-Watt University, Edinburgh, EH14 4AS, UK

3. Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, United Kingdom

4. School of Mathematical and Statistical Sciences, Clemson University, South Carolina, USA

5. Centre for Nature-Based Solutions, Faculty of Science and Engineering, Swansea University,
Swansea, SA2 8PP, United Kingdom

6. Department of Biosciences, Swansea University, Swansea, SA2 8PP, Wales, UK

7. Centre for Biomathematics, Faculty of Science and Engineering, Swansea University, Swansea,
SA2 8PP, United Kingdom

*Corresponding author: J.C.Bull@swansea.ac.uk

Paper for submission to Ecological Solutions and Evidence: Practical Tools Article

Abstract

1. A key aim of many conservation initiatives is to induce a change, however detecting change from monitoring data can be challenging. Here, we introduce statistical methods for changepoint detection, demonstrating how this under-utilised analytical approach can provide practitioners and researchers insight and help in planning ecological restoration projects.
2. We show why it is important to consider the stage of an ecological intervention as well as the design of the study and how this impacts the power of detecting a change in observed data.
3. *Practical Implication:* We provide user-friendly software and guidance that can be used to explore and quantify the data needed to detect change, under different settings. This enables practitioners to understand data requirements before planning monitoring studies, and effectively manage ongoing projects through all stages.

Keywords: ecological monitoring; evidence-based decision making; experimental design

1 Introduction

Monitoring is a fundamental component of effective conservation and ecological restoration. In the context of managing a particular area, ecological monitoring can be defined as the regular and systematic collection of information to track ecological change over time. Site managers commonly use monitoring to understand the impacts of anthropogenic pressures, to identify priorities for management, and to evaluate the effectiveness of interventions designed to mitigate those pressures (Christie et al., 2019). Robust monitoring requires the ability to distinguish natural temporal variability from changes attributable to anthropogenic threats or management actions, to statistically assess the magnitude of such impacts, and to interpret their ecological significance (Christie et al., 2019; Hewitt et al., 2001). Assessing the effectiveness of management interventions is particularly important for determining whether projects are achieving their stated aims. Such assessment underpins accountability to funders and stakeholders, helps to maintain confidence and financial support, directly informs adaptive management through iterative learning and improvement (Allen et al., 2011), and contributes to the wider evidence base on what does and does not work in conservation

and restoration practice (Cardinale et al., 2012; Sutherland et al., 2019; Sutherland and Wordley, 2018).

Despite its importance, effective monitoring is often difficult to incorporate into conservation and restoration projects. Monitoring programmes can require substantial time, expertise, and financial investment, resources that are frequently limited or prioritised towards on-the-ground action rather than evaluation (Christie et al., 2019; Ockendon et al., 2021; Tinsley-Marshall et al., 2022). Furthermore, the design and implementation of monitoring programmes requires clear guidance, which has historically been lacking, although recent efforts have begun to address this gap (e.g. Beyers and Sinclair, 2022; Ockendon et al., 2021; Torres et al., 2018). Key challenges in monitoring design include selecting an appropriate experimental framework, identifying suitable variables to measure, and determining an adequate level of replication (Christie et al., 2019). These challenges are particularly acute in the context of open-ended restoration systems (Mutillod et al., 2024).

Restoration ecology is defined as the process of assisting the recovery of ecosystems that have been degraded, damaged, or destroyed (Gann et al., 2019). An increasingly influential set of restoration approaches adopt an open-ended perspective, emphasising ecological processes over fixed structural targets and accepting a range of potential trajectories rather than prescriptive end states (Hughes et al., 2011). This perspective underpins concepts such as Restoration v. 2.0 (Higgs et al., 2014) and rewilding (Perino et al., 2019), which seek to restore dynamic, self-regulating ecosystems in which nature has greater autonomy over outcomes (Carver et al., 2025; Mutillod et al., 2024; Torres et al., 2018). While conceptually appealing, the open-ended nature of these approaches complicates the monitoring and evaluation of management effectiveness. When end goals are deliberately flexible, determining whether actions are successful becomes less straightforward. As a result, projects often rely on broad measures of success, such as increases in biodiversity or ecological integrity.

Several frameworks have been proposed to support monitoring in rewilding and other open-ended restoration contexts. These include approaches that compare site-level ecological metrics to those of a reference ecosystem, such as the BARI framework (Beyers and Sinclair, 2022), conceptual models that define success along gradients of reduced human forcing and increased ecological integrity (Torres et al., 2018), and landscape-scale approaches that utilise short- and long-term indicators to assess restoration success (Ockendon et al., 2026). Nevertheless, important chal-

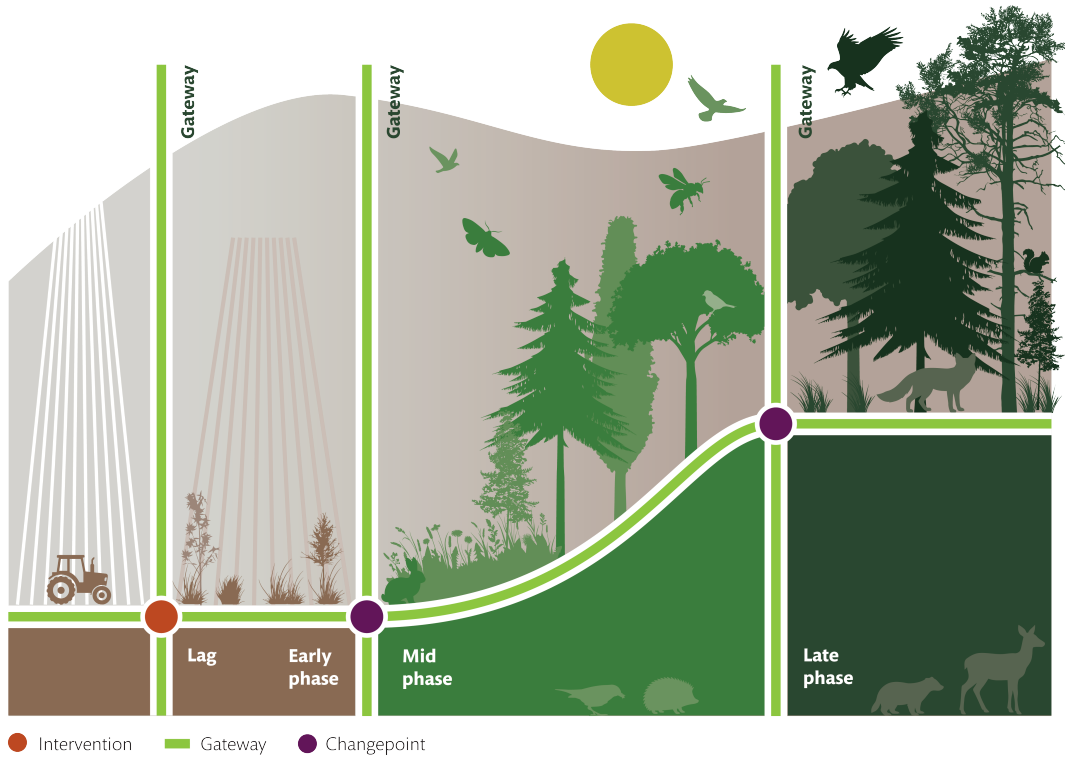


Figure 1: Visualisation of an abstract system changing across three phases and corresponding gateways: intervention, detectable ecological response, and stabilisation.

lenges remain in how monitoring data are interpreted and used to inform adaptive management. In open-ended systems, it is often unclear how much deviation from a desired condition should trigger management intervention, whether intervention should occur once a trajectory has been established, or how long managers should wait before acting when trajectories appear undesirable, particularly given the potential for time-lags in ecological responses (Watts et al., 2020).

These challenges highlight the need for analytical tools and conceptual frameworks to support both the design of monitoring programmes and the interpretation of monitoring data. There is a need for approaches that clarify what constitutes progress or success in the absence of fixed end points and that are sensitive to how ecological change unfolds through time.

1.1 Gateways

Ecological restoration is often characterised by long periods of apparently limited change followed by rapid shifts in composition and structure and, eventually, a slowing of change as the system

approaches a relatively stable state. Consequently, restoration progress can be conceptualised as a sigmoid (logistic) trajectory (Lanceman et al., 2025) through time (Figure 1). The vertical axis in this framework represents an abstract measure of conservation success, encompassing one or more indicators that reflect project objectives, which could include biodiversity recovery, carbon storage, habitat condition, ecosystem function, or other measures of ecological value. Individual indicators may follow different trajectories, but collectively they are expected to reflect increasing restoration success through time.

This conceptualisation suggests that restoration proceeds through *gateways*, defined to be points in time where a system change occurs, and they can be used to determine what monitoring can realistically achieve. The intervention gateway represents the time a management action was taken, intending to alter the state of the ecosystem. Examples include species reintroductions, woodland creation, invasive species removal, or changes in land management. The early phase is characterised by low levels of observable change in indicators of conservation success. Ecological processes may be occurring, but their effects remain difficult to detect with monitoring during this phase restricted to determining whether restoration has failed, for example through the loss of a reintroduced species or failure of planted individuals to establish. Challenges centre on low detection probability, false negatives, and limited information content within monitoring data.

The response gateway occurs when the effects of the intervention become detectable in monitored indicators and conservation success begins to increase measurably. In ecological terms, this gateway marks the transition from a system exhibiting little consistent observable change to one undergoing active restoration. Although the timing of this gateway is unknown, identifying when it has been crossed is a primary objective of monitoring because it provides evidence that restoration processes are progressing as intended.

The ecosystem subsequently enters a middle phase characterised by accelerating change. Populations increase, communities reorganise, ecosystem functions develop, and restoration outcomes become increasingly evident. Monitoring can move beyond simply assessing failure and towards understanding the processes that underpin ecosystem development and resilience. Although restoration trajectories can be evaluated during this phase, it is often still too early to assess the ultimate objectives of restoration.

The stabilisation gateway, representing the point at which rates of change slow and the ecosys-

tem approaches a relatively stable condition, may correspond to a new equilibrium, an alternative stable state, or a long-lived transient state. Crossing this gateway indicates that the major structural and functional changes associated with restoration have largely occurred and that the system is transitioning from development towards persistence. During the late phase, indicators of success fluctuate around a relatively stable level and restoration outcomes can be evaluated directly. Monitoring can now focus on long-term objectives such as biodiversity recovery, ecosystem service provision, carbon storage, and resilience, while continuing to assess the potential for future change.

For many restoration projects, particularly those with open-ended objectives and uncertain ecological trajectories, success cannot be adequately assessed using a single predefined endpoint. Instead, progress must be evaluated relative to the expected development of the system through time. We therefore suggest that monitoring programmes should be explicitly designed to identify when gateways have been crossed. The passage of gateways provides measurable milestones along the restoration trajectory, allowing practitioners to assess progress, evaluate success, and inform adaptive management even when ultimate restoration outcomes may take decades to emerge.

Within this framework, gateway transitions are expected to manifest as detectable shifts in ecological time series. Because each gateway reflects a change in the underlying processes governing ecosystem dynamics, crossing a gateway should alter the behaviour of monitored indicators, whether through changes in abundance, occupancy, rates of increase, variance, community composition, ecosystem function, or other measures of conservation success. Detecting these shifts provides evidence that the system has entered a new phase of recovery and within this paper we demonstrate how changepoint methodology provides a tool for identifying such transitions, supporting adaptive management throughout the restoration process.

2 Changepoint detection

A *changepoint* is the point at which there is a change in the structural pattern of a time-series. The segments of data before and after a changepoint are expected to be distinctively different in some way, whilst the observations within a segment are assumed to follow the same distribution. The changes in patterns can be explained by structural shifts in statistical properties, these include a mean shift, variance or covariance structure change, trend change and distribution

change. Changepoint analysis has been used within many different fields, including bioinformatics (Liehrmann et al., 2023), movement ecology (Edelhoff et al., 2016), space science (Xu et al., 2021), climate science (Tso et al., 2022), psychology (Mardanbegi et al., 2018), political science (Wang et al., 2024), and engineering (Hamed et al., 2026).

Depending on how data are collected and how the information from the analysis will be used, the methods for changepoint detection can be divided into two categories: offline (or retrospective) detection and online (or streaming) detection. For offline changepoint detection, the full time series is observed and the goal is to identify at which time point in history the structural change occurred. For online changepoint detection, the time series data are still being observed as time goes by and the goal is to flag when a change has occurred as quickly as possible whilst not creating too many false alarms. Depending on the problem, there may be more than one changepoint in the time series. Hence, the detection methods can be further grouped into those for detecting single or multiple changes. For the purpose of this paper, we focus on detecting a single changepoint, identifying a specific gateway and consider both offline and online detection as both scenarios are relevant to our setting.

Let y_t denote the observation at time t , $\mathbf{y}_{1:n}$ denote the full time series of length n , and $\mathbf{y}_{s:e}$ the segment from time s to e . We denote the changepoint by τ , $0 < \tau < n + 1$, where the time series $\mathbf{y}_{1:n}$ is divided into 2 segments. All observations within segment $\mathbf{y}_{1:\tau}$ are assumed to follow the same distribution, whilst the distribution of observations in segment $\mathbf{y}_{(\tau+1):n}$ will differ from that of segment $\mathbf{y}_{1:\tau}$. A typical changepoint model with a single changepoint at time τ can be written

$$y_t = \begin{cases} f_1(y_t; \boldsymbol{\theta}_1) & , \quad t \leq \tau \\ f_2(y_t; \boldsymbol{\theta}_2) & , \quad t > \tau \end{cases} \quad (1)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors of parameters of the models before and after the changepoint.

A standard approach for offline changepoint detection is to construct test statistics, based on all available data $\mathbf{y}_{1:n}$, that can reflect the difference between the segments before and after the changepoint. Under the null hypothesis of no change, the sample test statistic should be small and randomly fluctuating around zero. If the sample test statistic grows and passes a certain threshold, then there is a high probability that a change has occurred. The time point corresponding to the

172 maximum test statistic above threshold will be identified as the changepoint (Figure 2).

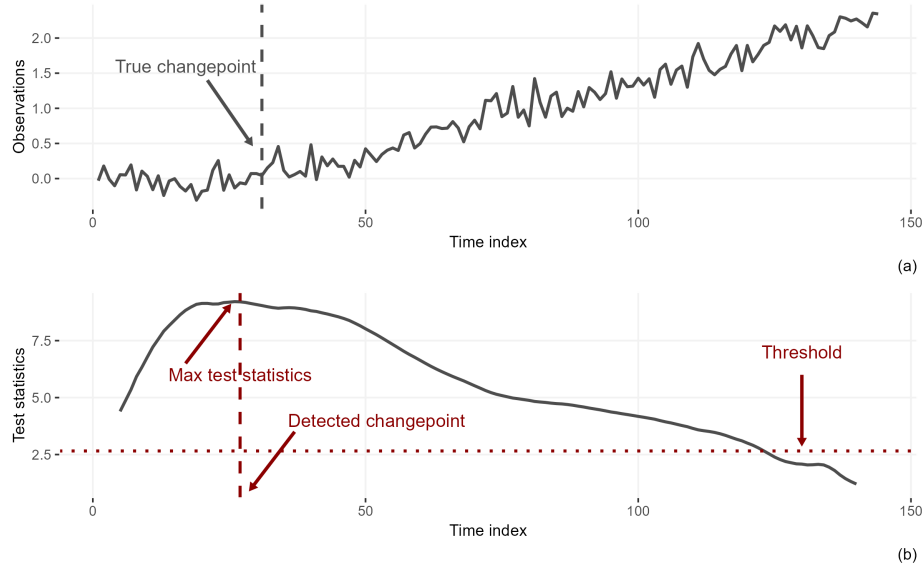


Figure 2: Illustration of offline detection of a single changepoint. The top panel shows the time series with a trend change. The bottom panel shows the test statistics reflecting the contrast before and after all potential changepoints. The changepoint is identified where the maximum of the test statistics occurred, providing the test statistic passes certain threshold.

173 Online changepoint detection methods make use of test statistics based on data collected up to
 174 current time $\mathbf{y}_{1:t}$, e.g., cumulative sum (CUSUM) type statistics (Page, 1954; Aue and Kirch, 2024),
 175 to evaluate the deviation of the distribution of the time series from the null hypothesis. As more
 176 observations after a change are collected, the evidence for a change grows. Once the test statistic
 177 passes the threshold, a claim can be made that a change has occurred at time $\tau < t$ (Figure 3).

178 A large collection of changepoint detection methods are available for offline and online detection.
 179 Depending on the type of change, the inference problem, the length of the time series, and the
 180 dimension of the data (i.e., whether it is univariate or multivariate), different methods may be
 181 utilised for the specific modelling tasks. For example, offline methods such as PELT (Killick et al.,
 182 2012) and FPOP (Maidstone et al., 2017) aim to improve the computational efficiency for large
 183 n through the use of (functional) pruning; online methods include Bayesian online changepoint
 184 detection (Fearnhead and Liu, 2007), FOCuS (Romano et al., 2023), and the forecast error-based
 185 detection (Grundy et al., 2026). All these methods have the flexibility of handling a variety of
 186 structural changes in the time series, from the basic mean and variance change to the more complex

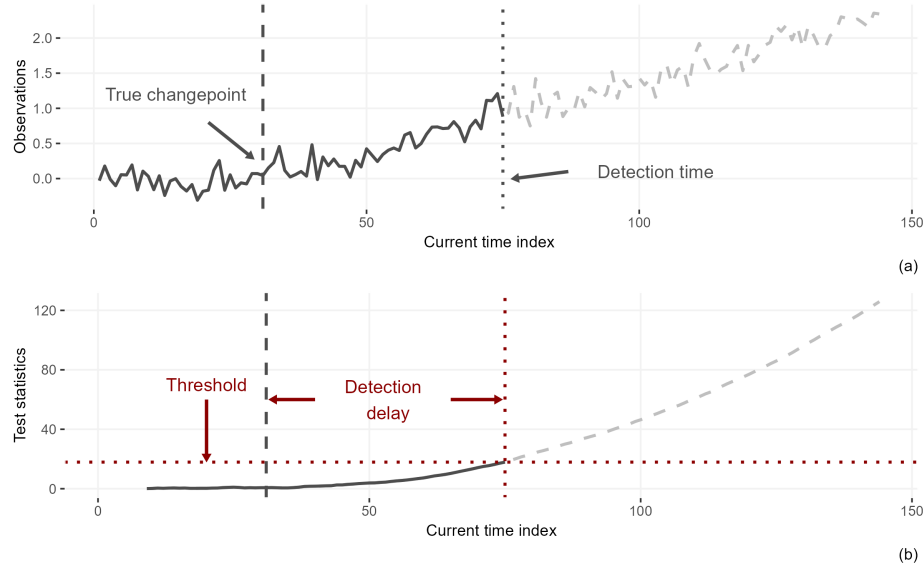


Figure 3: Illustration of online changepoint detection using CUSUM-type test statistics. The top panel shows the time series with a trend change. The bottom panel shows the CUSUM-type statistics. The light grey curves represent the “future” time series that are unobserved. A change is declared when the test statistics passes certain threshold.

change in model coefficients, orders, and correlation structure. In the simulation study below, we will focus on three methods which we feel are most relevant for the ecological restoration context we are considering.

3 Simulation study

To demonstrate the potential of the use of changepoint methodology within this setting, we present a detailed simulation study. We consider the experimental design with, (1) paired control and intervention time series, with observations from both before and after the management, referred to as BACI (Before-After-Control-Intervention), and (2) only the intervention time series, with observations from both before and after the management, referred to as BA (Before-After). We assume that the time series of the variable of interest are observed directly. After the intervention is applied, a change in trend is expected in the time series from the intervention plot, which would make the intervention time series deviate from the time series from the control plot (or from the pre-intervention period if no control data is available). There may be a delay before the onset of the change in trend, hence the true changepoint location is unknown in practice. For this scenario,

201 a series of increasing trend coefficients, reflecting different sizes of trend changes, are considered,
 202 alongside both independent and auto-correlated time series. An example of the simulated time
 203 series is given in Figure 4. An alternative scenario of a change in distribution is presented in the
 204 Supplementary material.

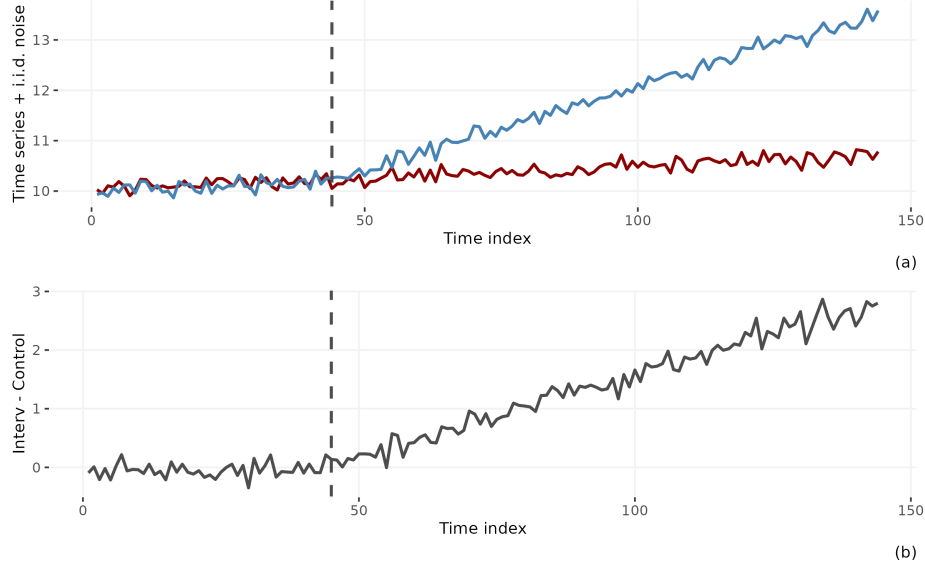


Figure 4: Change in trend after intervention (24 time points before intervention, 120 time points after intervention). (Top) Simulated time series from the control (red curve) and the intervention (blue curve) plots. (Bottom) The difference between the intervention and the control time series. The dotted vertical line indicates the (simulated) changepoint location.

205 Both BACI and BA scenarios require the detection of a change in trend. Therefore, we consider
 206 the generic change in regression coefficients model. With a changepoint at τ , the model can be
 207 written as

$$y_t = \begin{cases} \alpha_1 + \beta_1 t + \epsilon_t, & t \leq \tau \\ \alpha_1 + \beta_1 \tau + (\beta_1 + \Delta\beta)(t - \tau) + \epsilon_t, & t > \tau \end{cases} \quad (2)$$

208 where there is a step change in the linear trend coefficient of size $\Delta\beta$ after time τ . We denote
 209 the trend coefficient after the changepoint as $\beta_2 = \beta_1 + \Delta\beta$. Most changepoint methods assume
 210 the noise (error) series, $\epsilon_{1:n}$, to be independent but in reality it can be auto-correlated. The
 211 most common auto-correlation seen in practice is an AR(1) model, that is, $\epsilon_t = \phi\epsilon_{t-1} + u_t$, where
 212 u_t is white noise. For the BACI-type experiment, we detect changes by measuring the difference

between the intervention and control time series, as this avoids having to explicitly model dynamics we are not interested in. For the BA-type experiment, we investigate the difference between the time series before and after intervention and consider two sampling frequencies, mimicking the difference between a monthly (higher frequency) and annual (lower frequency) sampling rate (Table 1). Further detail is provided in Supplementary Material B.

Table 1: Summary of the simulation design and changepoint detection algorithms.

	Trend change			Distribution change	
	BACI + ind.	BA + ind.	BA + AR(1)	BACI	BA
Higher frequency					
Offline	AMOC	AMOC	AMOC	AMOC	AMOC
Online	Forecast	Forecast	Forecast	Forecast	Forecast
Lower frequency					
Offline	AMOC	AMOC	-	AMOC	AMOC
Online	BOCPD	BOCPD	-	BOCPD	BOCPD

3.1 Changepoint detection methods

Offline changepoint detection is used to illustrate how the size of trend change and the length of monitoring period after intervention affects the accuracy of changepoint detection whilst online changepoint detection is used to illustrate a more practical application, which addresses the key question of “how long after the intervention will one be able to detect a change in the time series?”.

Time series data from restoration experiments are often short and offline changepoint detection methods, such as the at most one change (AMOC) method (Beaulieu et al., 2024), are adequate to identify the changepoint location. For online changepoint detection, we consider the forecast error-based (Forecast) method (Grundy et al., 2026) and the Bayesian online changepoint detection (BOCPD) method (Fearnhead and Liu, 2007) for detecting the changes in the regression models. The Forecast method is used for detecting changes in high frequency (monthly) time series, and the BOCPD method is used for lower frequency (yearly) time series. This is due to the consideration that the Forecast method relies on an appropriate forecasting model constructed from pre-intervention data, which is impossible to obtain in the lower frequency setting where there are only two observations pre-intervention. Note, if a model was available from another setting or experiment, it could be used. Additional technical details about the AMOC, Forecast and BOCPD methods are provided in Supplementary Material C.

3.2 Results from the simulation study

For offline detection scenarios, a changepoint is considered correctly identified if the estimated changepoint is within the ± 12 time points interval of the true changepoint for higher (monthly) frequency time series, and within the ± 1 time point interval for lower (yearly) frequency time series. The true positive rate (TPR) and false discovery rate (FDR) for different combinations of change sizes and post-intervention time series lengths are visualised as heat maps (Figures 5 and 6). For the online detection scenarios, summaries of the detection delay are given in Tables 2 and 3.

Note that the FDR is calculated as the proportion of simulation replicates that incorrectly identify the changepoint out of the simulation replicates that return a positive claim that there exists a changepoint. In the combinations with small change size and short post-intervention time series, the number of replicates with a positive claim of a changepoint can be very low. If these claims are mostly correct, then the FDR would be low. However, the reality is that only a few replicates successfully identified a changepoint, suggesting that the algorithms struggle to detect the true changepoint. Therefore, the heat maps of TPR and FDR need to be considered jointly.

From Figures 5 and 6 we observe that the size of the trend change in the intervention time series has a large impact on the accuracy of changepoint detection (reflected by the TPR). The length of the time series after intervention also impacts the changepoint detection. At very short post-intervention length, the TPR is visibly worse when compared to the TPR from a longer length. The impact of length becomes less substantial after the size of change goes beyond certain level (i.e., at around 20% for higher frequency independent time series and 40% for lower frequency data independent time series). The heat maps of FDR show a similar pattern, apart from the size and length combinations showing at the bottom left corner of the maps, which is a consequence of the definitions being used, explained earlier.

When the time series are auto-correlated, the detection of the changepoint becomes more challenging. Within Figure 6, we see that for time series of the same length, it requires a larger size of change for the algorithm to detect a change accurately in the auto-correlated time series. In addition, strong auto-correlation can create an artifact that looks like a trend, which can be mistaken as a changepoint by the detection algorithm. This explains the decrease of TPR with the increase of post-intervention length at smaller change sizes in the two heat maps for the scenarios

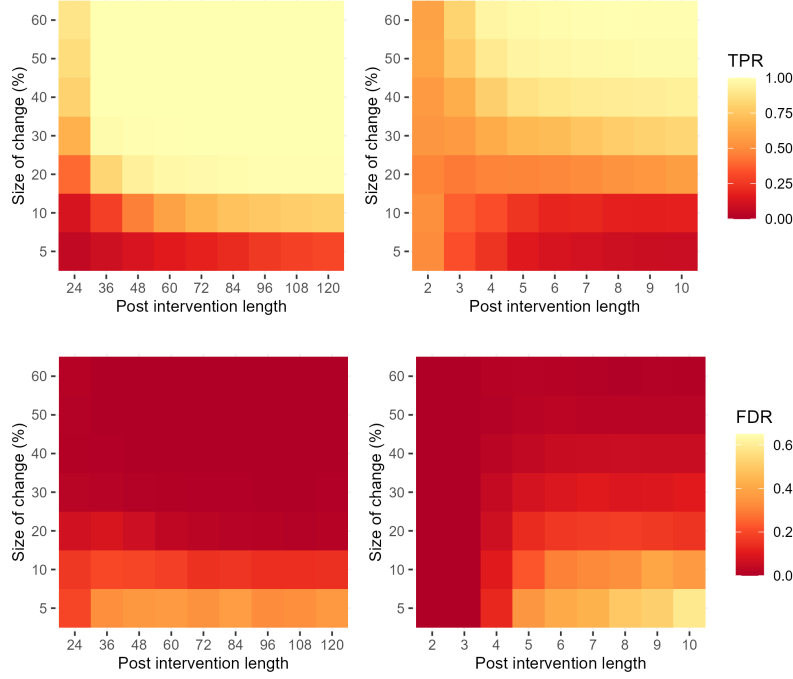


Figure 5: Heat maps showing true positive rates (TPR, top) and false discovery rate (FDR, bottom) of different combinations of post-intervention length and percentage change to pre-intervention level from the offline changepoint detection using AMOC, for trend change scenario BACI + independent noise. The left column shows results for time series with higher (monthly) sampling frequency and the right column shows results for time series with lower (yearly) sampling frequencies.

with AR(1) noise. This also echos the increasing FDR recorded for these combinations, due to the spurious changepoint detection.

From Tables 2 and 3 we see that for online changepoint detection using either the Forecast method or the BOCPD method, we observe similar impact from the size of changes on changepoint detection. Both the detection delay and the accuracy of changepoint detection improve as the size of change increases. The TPR from the simulation using the Forecast method on higher frequency data increases from around 50% to 95%; whereas the true positive rate from the simulation using BOCPD on lower frequency data show a more dramatic increase as the sizes of change increases, around 10% to 95%.

The Forecast method struggled with the online detection of changepoint in time series with AR(1) noise (Table 3). The TPRs are lower across almost all change sizes as compared to the independent scenario, despite the smaller detection delays. This can be explained by the short pre-intervention period, which does not provide enough data to build a reliable prediction model.

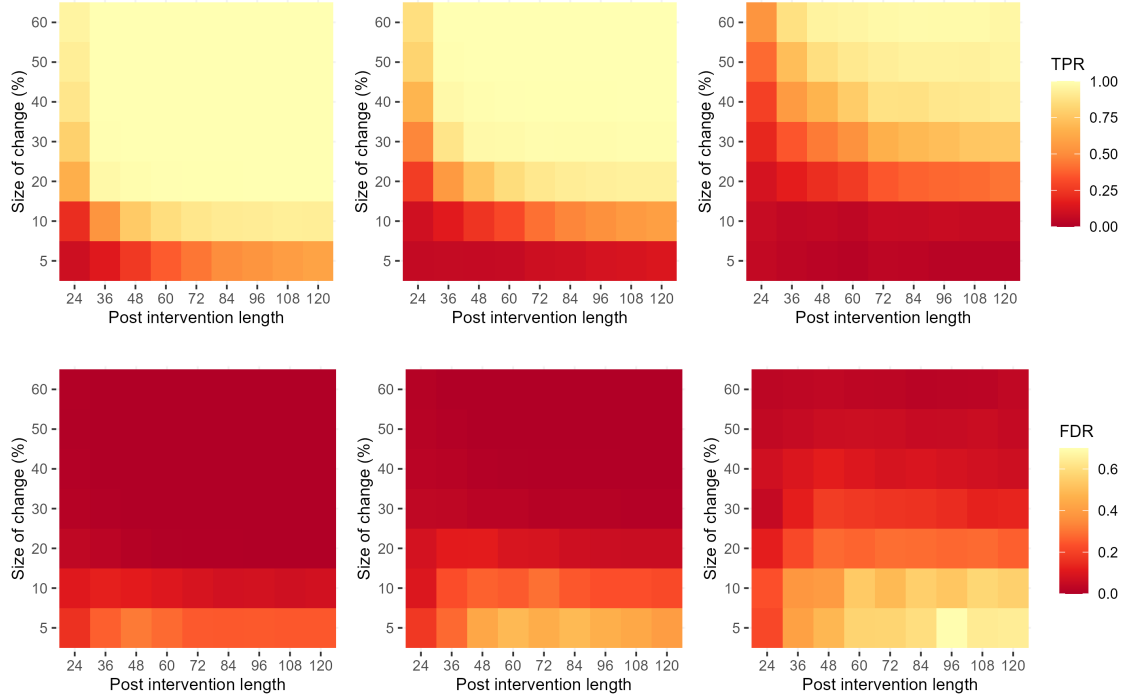


Figure 6: Heat maps showing true positive rates (TPR, top) and false discovery rate (FDR, bottom) of different combinations of post-intervention length and percentage change to pre-intervention level from offline changepoint detection using AMOC, for trend change scenario BA + independent noise (left column), BA + AR(1) noise with $\phi = 0.5$ (middle column) and BA + AR(1) noise with $\phi = 0.8$ (right column), all with higher (monthly) sampling frequencies.

277 Increasing the pre-intervention data improves results.

Table 2: True positive rate (TPR), mean detection delay and standard deviation of delay from online changepoint detection of the first change in trend under BACI with independent errors. The Forecast method was used for higher frequency time series, and the BOCPD method was used for lower frequency time series.

Size of change	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Higher frequency										
TPR	0.55	0.73	0.87	0.92	0.95	0.94	0.95	0.95	0.95	0.96
mean delay	29.90	28.24	17.33	13.42	11.50	9.60	9.08	8.01	7.59	7.09
std. dev. delay	25.50	22.87	12.03	8.04	6.34	5.61	4.63	4.33	3.83	3.47
no detection	121	25	0	0	0	0	0	0	0	0
Lower frequency										
TPR	0.12	0.19	0.49	0.74	0.86	0.92	0.93	0.95	0.94	0.95
mean delay	5.03	5.19	4.79	3.92	3.25	2.75	2.44	2.28	2.12	2.06
std. dev. delay	2.74	2.58	2.21	1.80	1.28	0.88	0.65	0.53	0.41	0.32
no detection	721	612	245	40	3	3	0	0	0	0

278 When time series are short, using AMOC for auto-correlated data may be more appropriate.

Table 3: True positive rate (TPR), mean detection delay and standard deviation of delay from online changepoint detection of the first change in trend under BA with independent or AR(1) errors. The Forecast method was used for both scenarios which have higher sampling frequencies.

Size of change	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Independent										
TPR	0.64	0.81	0.92	0.93	0.95	0.95	0.94	0.96	0.95	0.97
mean delay	30.17	23.28	14.00	10.55	9.21	7.91	7.13	6.69	6.10	5.73
std. dev. delay	26.39	17.07	8.26	6.14	4.86	4.17	4.01	3.57	3.16	2.97
no detection	67	2	0	0	0	0	0	0	0	0
AR(1), $\phi = 0.5$										
TPR	0.63	0.71	0.81	0.85	0.89	0.89	0.89	0.88	0.89	0.88
mean delay	15.71	14.77	9.10	7.43	6.29	5.22	4.78	4.21	3.72	3.47
std. dev. delay	21.94	19.83	11.81	9.35	7.37	6.81	6.12	5.89	5.55	5.33
no detection	32	6	0	0	0	0	0	0	0	0

Each time a new observation is obtained, the AMOC procedure is rerun to see if any time point in history can be a changepoint. This recursion may be stopped after a changepoint is detected for the first time. Across all scenarios, it is clear that a lower sampling frequency makes the changepoint detection harder for both offline and online methods. A larger size of change is required to achieve a similar true positive rate or detection delay, however the presence of a control time series does not appear to have a large impact on the results.

3.3 Software

We provide R code to help practitioners explore scenarios that more closely match the parameter values expected in their own studies¹, with R scripts provided for each of the simulation scenarios listed in Table 1. The code leverages the `changepoint.forecast` package² for implementing the forecast error-based method and the `transport` package (Schuhmacher et al., 2024) for evaluating the Wasserstein distance.

For users who prefer an interactive interface, we additionally provide a Python Streamlit web application, **TRACE** (Temporal Restoration Analysis for Changepoint Estimation), that reproduces these scenarios in a browser³. The application brings together all the detection methods considered in this paper, allowing practitioners to compare their behaviour under a common set of assumptions. Users can adjust the simulation inputs directly through interface controls (for example the effect

¹The R scripts are available on GitHub at: <https://github.com/GMY2018/SimRewilding>

²<https://github.com/grundy95/changepoint.forecast>

³The source code for the web application is available on GitHub at: <https://github.com/M-DICE/trace>

size, the timing of the intervention, the length of the monitoring window, or the noise level), allowing exploration of relevant empirical findings for specific studies.

4 Discussion

Changepoint detection is a well-established statistical approach that has been applied across a wide range of disciplines to identify significant shifts in temporal data. Despite its broad use elsewhere, its potential for ecological restoration monitoring remains largely unexplored. Many contemporary ecological restoration approaches, particularly those aiming to restore ecological processes and promote self-regulating ecosystems, do not have clearly defined targets, trajectories, or end points against which success can be measured. Consequently, assessing ecosystem responses to management interventions can be challenging.

For ecological restoration, progress may be more usefully evaluated in relation to the restoration trajectory itself. Under the conceptual framework developed here, restoration can be viewed as a progression through a series of ecological gateways, beginning with intervention, followed by a detectable ecological response, and ultimately the emergence of a relatively stable restored state. The passage of these gateways provides a set of measurable milestones against which restoration progress can be assessed, even when the ultimate endpoint remains uncertain. From this perspective, evidence that a system has passed through successive gateways represents increasing confidence that restoration is progressing in the desired direction (or not). Monitoring therefore shifts from asking whether a predefined target has been achieved to asking whether the ecosystem has undergone meaningful changes consistent with recovery and development.

This trajectory-based perspective is consistent with existing frameworks for monitoring ecological restoration (e.g., Ockenden et al. 2025). Changepoint approaches offer a complementary quantitative tool for operationalising long-term indicators of recovery and assessing whether systems are undergoing directional, rather than transient, change and can be used to link empirical findings to conceptual restoration trajectories. The gateways framework recognises delayed ecological responses, highlighting the patience required to distinguish meaningful change from short-term variability. It also identifies how monitoring approaches and measures of restoration success should change as ecosystems progress through recovery stages. The tools provided can be used to identify

how much data is required before you can expect to detect a change, and can be used to inform management and monitoring decision-making. We highlight that detection of a change does not imply causality and any detected changes should be considered in conjunction with causal inference to ensure erroneous conclusions are not reached.

The changepoint detection methods within this paper are all structured within a linear modelling framework and at short temporal scale this is usually a good approximation (see Figure 1). It is however, worth highlighting we have considered a simple model structure here, and higher order, non-linear terms or more complex correlation structures are often included when detecting changepoints in environmental data, see e.g., Gallagher et al. (2012); Shi et al. (2022).

Extensions of changepoint detection methods for multivariate data exist, e.g., Healy (1987); Matteson and James (2014); Pishchagina et al. (2026), which can be useful if it is anticipated that variables are going to exhibit similar changes in mean, rate or covariance structure. However, if variables respond differently to a given intervention, this can create additional noise in the data and it would be better to consider univariate methods on individual variables. Similarly, spatiotemporal changepoint detection (Agarwal et al., 2025; Holloway and Killick, 2024) may also be of interest, but consideration would need to be given to the structure of experimental design in an ecological restoration context to evaluate the potential for such methods to contribute in this area.

The time-scale and magnitude of expected changes will vary dramatically for different studied systems, as will the feasibility for collecting data across variables. The analytical tools we have presented allow exploration of data needs and taking such an analytical approach will ensure that monitoring can help contribute to the broad evidence-base for ecological restoration.

Acknowledgements

We would like to acknowledge the International Centre for Mathematical Sciences that funded a *Rewilding Mathematics* group and further Knowledge Exchange event which enabled the discussions that led to the writing of this paper. JB was funded by NERC Exploring the Frontiers grant: NE/X012697/1. We thank Nicky Regan at Forest Research for assistance with the graphic design of Figure 1.

Author Contributions

Following the CRediT contribution structure the authors have contributed in the following ways:

- Mengyi Gong - Methodology; Software; Validation; Formal Analysis; Writing - Original Draft; Visualization
- Simone Conte - Software; Visualization
- Tom Jameson - Conceptualization; Writing - Original Draft
- Rebecca Killick - Conceptualization; Methodology; Writing - Original Draft; Supervision
- James Bull - Conceptualization; Writing - Original Draft
- Luca Börger - Conceptualization; Writing - Original Draft
- Rachel McCrea - Conceptualization; Writing - Original Draft; Supervision; Project administration

Conflict of Interest

No declared conflict of interest.

Data Availability Statement

Simulated data have been used to demonstrate the methods of the paper and the code to replicate these data and findings have been provided on GitHub. See Section 3.3 for links.

References

- Agarwal, G., Eckley, I. A., and Fearnhead, P. (2025). Efficient likelihood-based temporal change-point detection in spatio-temporal processes. *Statistics and Computing*, 35:213.
- Allen, C. R., Fontaine, J. J., Pope, K. L., and Garmestani, A. S. (2011). Adaptive management for a turbulent future. *Journal of Environmental Management*, 92(5):1339–1345.

- Aue, A. and Kirch, C. (2024). The state of cumulative sum sequential changepoint testing 70 years after page. *Biometrika*, 111(2):367–391.
- Beaulieu, C., Gallagher, C., Killick, R., Lund, R., and Shi, X. (2024). A recent surge in global warming is not detectable yet. *Communications Earth & Environment*, 5(576).
- Beyers, R. and Sinclair, A. R. E. (2022). Measuring success in rewilding: Ecological overview. In *Routledge Handbook of Rewilding*. Routledge.
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., and Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486:59–67.
- Carver, S., Convery, I., Hawkins, S., Hertel, S., Fallon, J., Lyons, K., Beyers, R., Locquet, A., Derham, T., and Kun, Z. (2025). Guidelines for rewilding. *Switzerland: IUCN*.
- Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., and Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, 56(12):2742–2754.
- Eckley, I. A., Fearnhead, P., and Killick, R. (2011). Analysis of changepoint models.
- Edelhoff, H., Signer, J., and Balkenhol, N. (2016). Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*, 4:21.
- Fearnhead, P. and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):589–605.
- Gallagher, C., Lund, R., and Robbins, M. (2012). Changepoint detection in daily precipitation data. *Environmetrics*, 23(5):407–419.
- Gann, G. D., McDonald, T., Walder, B., Aronson, J., Nelson, C. R., Jonson, J., Hallett, J. G., Eisenberg, C., Guariguata, M. R., Liu, J., Hua, F., Echeverría, C., Gonzales, E., Shaw, N., Decler, K., and Dixon, K. W. (2019). International principles and standards for the practice of ecological restoration. second edition. *Restoration Ecology*, 27(S1):S1–S46.

- Grundy, T., Killick, R., and Svetunkov, I. (2026). Online detection of forecast model inadequacies using forecast errors. *Journal of Time Series Analysis*, 47(3):715–726.
- Hamed, H., Reis, A., Choobar, B., Pang, Q., Killick, R., and M., S. (2026). Changepoint detection as a light data-driven approach to battery state-of-health prediction. *Cell Reports Physical Science*, 7(3):103157.
- Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics*, 29(4):409–412.
- Hewitt, J. E., Thrush, S. E., and Cummings, V. J. (2001). Assessing environmental impacts: effects of spatial and temporal variability at likely impact scales. *Ecological Applications*, 11(5):1502–1516.
- Higgs, E., Falk, D. A., Guerrini, A., Hall, M., Harris, J., Hobbs, R. J., Jackson, S. T., Rhemtulla, J. M., and Throop, W. (2014). The changing role of history in restoration ecology. *Frontiers in Ecology and the Environment*, 12(9):499–506.
- Holloway, M. J. and Killick, R. (2024). Detection of spatiotemporal changepoints: a generalised additive model approach. *Statistics and Computing*, 34:162.
- Hughes, F. M., Stroh, P. A., Adams, W. M., Kirby, K. J., Mountford, J. O., and Warrington, S. (2011). Monitoring and evaluating large-scale ‘open-ended’ habitat creation projects: A journey rather than a destination. *Journal for Nature Conservation*, 19(4):245–253.
- Hughes, F. M. R., Stroh, P. A., Mountford, J. O., Warrington, S., Gerrard, C., and Jose, P. (2008). Monitoring large-scale wetland restoration projects: Is there an end in sight? In *Landscape ecology and conservation. Proceedings of the 15th Annual Conference of the International Association for Landscape Ecology*. International Association for Landscape Ecology.
- Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.
- Lanceman, D., Mayer-Pinto, M., and Glamore, W. (2025). Ecosystem service trajectories in restored coastal habitats. *Global Change Biology*, 31(3):e70151.

423 Liehrmann, A., Delannoy, E., Launay-Avon, A., Gilbault, E., Loudet, O., Castandet, B., and
 424 Rigaill, G. (2023). Diffsegr: an rna-seq data driven method for differential expression analysis
 425 using changepoint detection. *NAR Genomics and Bioinformatics*, 5(4):lqad098.

426 Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017). On optimal multiple changepoint
 427 algorithms for large data. *Statistics and Computing*, 27:519–533.

428 Mardanbegi, D., Killick, R., Wilcockson, T., Xia, B., Gellersen, H.-W., Sawyer, P., and Crawford,
 429 T. (2018). Effect of aging on post-saccadic oscillations. *Vision Research*, 143:1–8.

430 Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point
 431 analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345.

432 Mutillod, C., Buisson, , Mahy, G., Jaunatre, R., Bullock, J. M., Tatin, L., and Dutoit, T. (2024).
 433 Ecological restoration and rewilding: two approaches with complementary goals? *Biological*
 434 *Reviews*, 99(3):820–836.

435 Ockendon, N., Amano, T., Cadotte, M., Downey, H., Hancock, M. H., Thornton, A., Tinsley-
 436 Marshall, P., and Sutherland, W. J. (2021). Effectively integrating experiments into conservation
 437 practice. *Ecological Solutions and Evidence*, 2(2):e12069.

438 Ockendon, N., Gullett, P. R., Hancock, M. H., Noble, D., Shaw, T., and Thomas, D. H. L. (2026).
 439 A framework for monitoring ecosystem restoration at landscape scale. *Conservation Science and*
 440 *Practice*, 8(2):e70214.

441 Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

442 Perino, A., Pereira, H. M., Navarro, L. M., Fernández, N., Bullock, J. M., Ceașu, S., Cortés-
 443 Avizanda, A., van Klink, R., Kuemmerle, T., Lomba, A., Pe’er, G., Plieninger, T., Benayas, J.
 444 M. R., Sandom, C. J., Svenning, J.-C., and Wheeler, H. C. (2019). Rewilding complex ecosystems.
 445 *Science*, 364(6438):eaav5570.

446 Pishchagina, L., Romano, G., Fearnhead, P., Runge, V., and Rigaill, G. (2026). Online multivariate
 447 changepoint detection: leveraging links with computational geometry. *Journal of the Royal*
 448 *Statistical Society Series B: Statistical Methodology*, 88(1):171–194.

449 Romano, G., Eckley, I. A., Fearnhead, P., and Rigai, G. (2023). Fast online changepoint detection
450 via functional pruning cusum statistics. *Journal of Machine Learning Research*, 24(81):1–36.

451 Schuhmacher, D., Bähre, B., Bonneel, N., Gottschlich, C., Hartmann, V., Heinemann, F.,
452 Schmitzer, B., and Schrieber, J. (2024). *transport: Computation of Optimal Transport Plans
453 and Wasserstein Distances*. R package version 0.15-4.

454 Shi, X., Beaulieu, C., Killick, R., and Lund, R. (2022). Changepoint detection: An analysis of the
455 central england temperature series. *Journal of Climate*, 35(19):6329 – 6342.

456 Sutherland, W. J., Taylor, N. G., MacFarlane, D., Amano, T., Christie, A. P., Dicks, L. V., Lemas-
457 son, A. J., Littlewood, N. A., Martin, P. A., Ockendon, N., Petrovan, S. O., Robertson, R. J.,
458 Rocha, R., Shackelford, G. E., Smith, R. K., Tyler, E. H., and Wordley, C. F. (2019). Build-
459 ing a tool to overcome barriers in research-implementation spaces: The conservation evidence
460 database. *Biological Conservation*, 238:108199.

461 Sutherland, W. J. and Wordley, C. F. R. (2018). A fresh approach to evidence synthesis. *Nature*,
462 558:364–366.

463 Tinsley-Marshall, P., Downey, H., Adum, G., Al-Fulaij, N., Bourn, N. A., Brotherton, P. N.,
464 Frick, W. F., Hancock, M. H., Hellon, J., Hudson, M. A., Kortland, K., Mastro, K., McNicol,
465 C. M., McPherson, T., Mickleburgh, S., Moss, J. F., Nichols, C. P., O’Brien, D., Ockendon, N.,
466 Paterson, S., Parks, D., Pimm, S. L., Schofield, H., Simkins, A. T., Watuwa, J., Wormald, K.,
467 Wilkinson, J., Wilson, J. D., and Sutherland, W. J. (2022). Funding and delivering the routine
468 testing of management interventions to improve conservation effectiveness. *Journal for Nature
469 Conservation*, 67:126184.

470 Torres, A., Fernandez, N., zu Ermgassen, S., Helmer, W., Revilla, E., Saavedra, D., Perino, A.,
471 Mimet, A., Rey-Benayas, J. M., Selva, N., Schepers, F., Svenning, J.-C., and Pereira, H. M.
472 (2018). Measuring rewilding progress. *Philosophical Transactions of the Royal Society B: Bio-
473 logical Sciences*, 373(1761):20170433.

474 Tso, C.-H. M., Monteith, D., Scott, T., Watson, H., Dodd, B., Pereira, M. G., Henrys, P., Hollaway,
475 M., Rennie, S., Lowther, A., Watkins, J., Killick, R., and Blair, G. (2022). The evolving role of

476 weather types on rainfall chemistry under large reductions in pollutant emissions. *Environmental*
477 *Pollution*, 299:118905.

478 Walters, C. (1986). *Adaptive Management of Renewable Resources*. McMillan.

479 Wang, Z., Cheng, H., Zhong, W., Ma, P., and Murdie, A. (2024). Shifting sands: How change-point
480 and community detection can enrich our understanding of international politics. *International*
481 *Interactions*, 50(2):349–369.

482 Watts, K., Whytock, R. C., Park, K. J., Fuentes-Montemayor, E., Macgregor, N. A., Duffield,
483 S., and McGowan, P. J. K. (2020). Ecological time lags and the journey towards conservation
484 success. *Nature Ecology and Evolution*, 4:304–311.

485 Xu, C., Günther, H. M., Kashyap, V. L., Lee, T. C. M., and Zezas, A. (2021). Change-point
486 detection and image segmentation for time series of astrophysical images. *The Astronomical*
487 *Journal*, 161(4):184.

Supplementary Material

A. Glossary of Terms

Within this section we define some of the general terms used within the paper.

- **Adaptive Management:** An approach to natural resource management that emphasizes learning through management based on the philosophy that knowledge is incomplete and much of what we think we know is actually wrong, but despite uncertainty managers and policy makers must act (Walters, 1986).
- **Changepoint:** the point at which there is a change in the structural pattern of a time-series. Mathematically, this corresponds to the time point which divides a data set into distinct homogeneous segments (Eckley et al., 2011).
- **Ecological Monitoring:** The regular and systematic collection of information about an ecosystem to track changes over time.
- **Ecological restoration:** The process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed (Gann et al., 2019).
- **Ecological Restoration v 2.0:** An approach to ecological restoration that uses historical knowledge as a guide and not as a template, accepts multiple potential trajectories for ecosystems, emphasizes process over structure, and embraces pragmatic goals for human well-being (Higgs et al., 2014).
- **Gateway:** Points in time where a system change occurs.
- **Milestone:** A significant point or event in a project that marks the completion of a major phase, deliverable, or decision point.
- **Open-ended restoration:** An approach to ecological restoration which views restoration as a journey whose destination is uncertain rather than necessarily producing specific habitats or species populations (Hughes et al., 2008).
- **Rewilding:** The process of rebuilding, following major human disturbance, a natural ecosystem by restoring natural processes and the complete or near complete food web at all trophic

levels as a self-sustaining and resilient ecosystem with biota that would have been present had the disturbance not occurred. This will involve a paradigm shift in the relationship between humans and nature (Carver et al., 2025).

- **Time-series:** A sequence of data points recorded in chronological order, typically at regular intervals.

B. Technical details of simulation study settings

- (1) Size of changes: For both change types, a series of different change sizes are considered. In particular, the sizes of changes (in trend coefficient for change type A or in the location changes for change type B) are specified as percentage change to the pre-intervention level. For example, if the pre-intervention mean is $\mu = E[y_t]$, then the post-intervention trend would be $\delta\%\mu$, for a series of increasing $\delta\%$. In this simulation study, we will consider $\delta\% = 5\%, 10\%, 20\%, \dots, 100\%$.
- (2) Sampling frequency: For both change types, for the same rate of change, a higher and a lower sampling frequencies will considered, mimicking the monthly and annual sampling frequencies. That is, the sampling frequency of the higher frequency time series is 12 times that of the lower frequency time series. For a fair comparison of the impact of sampling frequencies, we further assume that the same amount of change will be observed in a given period of time, regardless of sampling frequencies. Therefore, if the higher frequency time series has trend coefficient β , then the lower frequency time series will have trend coefficient $\beta^* = 12 \times \beta$.
- (3) Length of time series: Based on a typical experimental design, the length of the pre-intervention period is set to 24 for higher frequency time series and 2 for lower frequency time series, and the maximum length of the post-intervention period is set to 120 for higher frequency time series and 10 for lower frequency time series. For the “BA + AR(1)” scenario, only the higher frequency time series are simulated due to data requirements for the increased model complexity.
- (4) Time delay: To mimic the real-life situation, we simulate the time delay before the intervention starts to take effect from a uniform distribution. In particular, we use $U(1, 20)$ for higher

frequency data and $U(1, 2)$ for lower frequency data. The time delay can have some influence on the result, as it results in a longer period of “pre-change” data.

(5) Auto-correlation: Two levels of auto-correlation are considered for the AR(1) noise in the simulated time series. They are $\phi = 0.5$ for a moderate auto-correlation and $\phi = 0.8$ for a strong auto-correlation. This can help us better investigate the impact of auto-correlation on changepoint detection.

(6) Simulation runs: Each scenario is replicated 1000 times with distinct random seeds.

C. Technical details about AMOC, Forecast and BOCPD methods

Here we provide some details on the offline and online detection methods used in this simulation study. For offline changepoint detection, we used the at most one change (AMOC) method to determine the location of the changepoint provided it exists. For online changepoint detection, the forecast error-based (Forecast) method and the Bayesian online changepoint detection (BOCPD) method are used. Both methods process the data sequentially and can flag a changepoint once certain detection criterion is met.

(1) For offline changepoint detection using AMOC, the test statistic

$$T_{max} = \max_{\tau=s}^e |T_{\tau}|, \quad (3)$$

is computed for each simulated time series, where

$$T_{\tau} = \frac{\hat{\beta}_2(\tau) - \hat{\beta}_1(\tau)}{\sqrt{\text{Var} [\hat{\beta}_2(\tau) - \hat{\beta}_1(\tau)]}},$$

and s is the beginning of the intervention, e is the end of the observing period, and $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ are estimated assuming that there is a changepoint at τ . If T_{max} is greater than some estimated threshold, then the τ that corresponds to T_{max} is identified as the changepoint. The threshold used in the simulation study are obtained via Monte Carlo simulation. As time series of different lengths are considered, different thresholds are simulated for different length

sizes (i.e., at $\log(n)$ sizes). A similar approach is used in Beaulieu et al. (2024) to investigate whether a global warming surge is detectable.

- (2) For online changepoint detection using the Forecast method, we stop the algorithm once the CUSUM-based test statistic goes beyond the estimated threshold. In particular, the test statistic is computed as

$$Q(m, k) = \sum_{t=m+1}^{m+k} \epsilon_t - \frac{k}{m} \sum_{t=1}^m \epsilon_t$$

$$T(m, k) = \max_{0 \leq i \leq k} |Q(m, k) - Q(m, i)|$$

where m is the length of the training time series (which is used to build the forecast model) and k is the length of the post training period. In this simulation study, m is taken to be the length of the pre-intervention time series. A changepoint is declared the first time $T(m, k)$ passes certain threshold, which can be derived based on asymptotics or obtained via simulation. Detail of the Forecast method can be found in Grundy et al. (2026). In this study, the Forecast is applied to the higher frequency scenarios. This is because the Forecast method relies on an adequate forecast model, which may be difficult to obtain on a small number of observations in a lower frequency scenario.

- (3) For the BOCPD method, we stop the algorithm once the first changepoint τ ($\tau > 0$) is identified from the maximum a posterior (MAP) estimation. In a nutshell, the BOCPD method assumes a hidden process z_t , which represents the last changepoint before the current time t (i.e., $z_t = s$ means the last changepoint before t is at s). Given the hidden process, the likelihood of all segments in the time series up to current time can be evaluated, and the posterior distribution of z_t given observed time series $y_{1:t}$ can be obtained iteratively under the Bayesian framework. From the posterior probabilities, one can get the MAP estimation of the first changepoint. Details of the BOCPD method can be found in Fearnhead and Liu (2007). In this study, BOCPD is applied to the lower frequency scenarios as an alternative to the Forecast method. This is because the BOCPD method does not require such models and hence can be applied to lower frequency time series. However, BOCPD requires the user to specify the prior distributions and some essential global parameters in order to initialise

the algorithm. Therefore, its implementation is not as straightforward as the forecast error method. In the simulation study, the Normal prior is given to the trend coefficients and the inverse-Gamma prior is given to the residual variance. The geometric distribution is used as the run-length distribution, giving a fixed transition probability at each time step.

D. Change in Distribution Scenario

This scenario considers that the distributions of the variable of interest (e.g., distribution of tree sizes) over time are observed. After intervention, it is expected that the distribution of the variable from the intervention plot start to diverge from the distribution of the variable from the control plot (or from the pre-intervention period if no control data is collected). Similar to the trend change, there may be some delay before the change happens. For this case, time series of distributions with location change will be simulated. An example of the simulated data is given in the top panel of Figure 7.

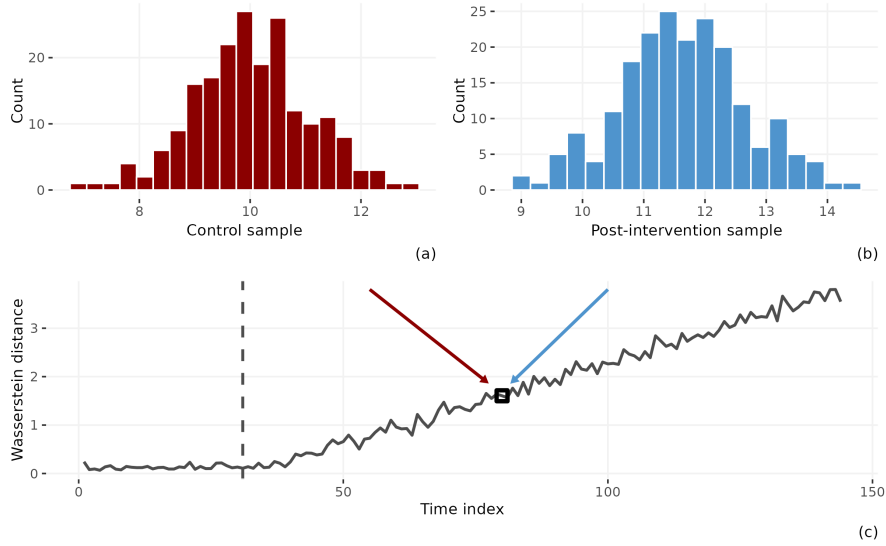


Figure 7: Change in distributions after intervention (24 time points before intervention, 120 time points after intervention). (Top) The histograms are the simulated distributions, assuming increasing mean after intervention. (Bottom) The time series of Wasserstein distance metric. The dotted vertical line indicates the changepoint location.

For the time series of distributions, we first create certain type of distance metric that can reflect the changes in the distributions. Note that although there are methods (e.g., non-parametric changepoint detection methods) for detecting a change in distribution, they are designed for a

different setting from the one considered in this study. Here, a sample of m observations $\{x_{ti}\}_{i=1}^m$ is collected at each time point, and we are interested in the change of distributions of $\{x_{ti}\}_{i=1}^m$ at each time t . Whereas the non-parametric detection methods typically consider the problem of whether the distribution of $x_{1:\tau}$ is different from that of $x_{(\tau+1):n}$. Therefore, in this case, we choose to convert the difference between two distributions to the Wasserstein distance metric and apply changepoint detection methods to the time series of distance metrics. Specifically, for the BACI scenario, the Wasserstein distance is taken to be the distance between the distributions from the control and the intervention plots at time t ; whereas for the BA only scenario, the Wasserstein distance is taken to be the distance between the distribution at a post-intervention time t and the distribution from the pre-intervention (or baseline) period. The bottom panel of Figure 7 shows an example of the time series of the Wasserstein distance from the BACI scenario.

The results from the simulation are shown in Figures 8, 9 and Table 4. Similar to the change in trend scenario, the size of the parameter change plays an important role in changepoint detection using the AMOC method. The length of the post-intervention period also affects changepoint detection substantially. At very short post-intervention length, the TPR is visibly worse. Although the impact of lengths becomes much smaller when the size of change goes beyond certain level (about 20% of the mean level in this study). For online changepoint detection using either the Forecast method or the BOCPD method, we observe similar impact from the size of changes on changepoint detection. Both the detection delay and the accuracy of changepoint detection improve as the size of change increases. The TPR from the simulation using the Forecast method on higher frequency data increases from around 40% to 95%; whereas the TPR from the simulation using BOCPD on lower frequency data increases from 0% to nearly 100%. The sampling frequency can affect both the accuracy in offline detection and the detection delay in online detection. Although the difference becomes smaller as the size of change gets larger. Whether there is a control time series or not does not appear to have a large impact on the result.

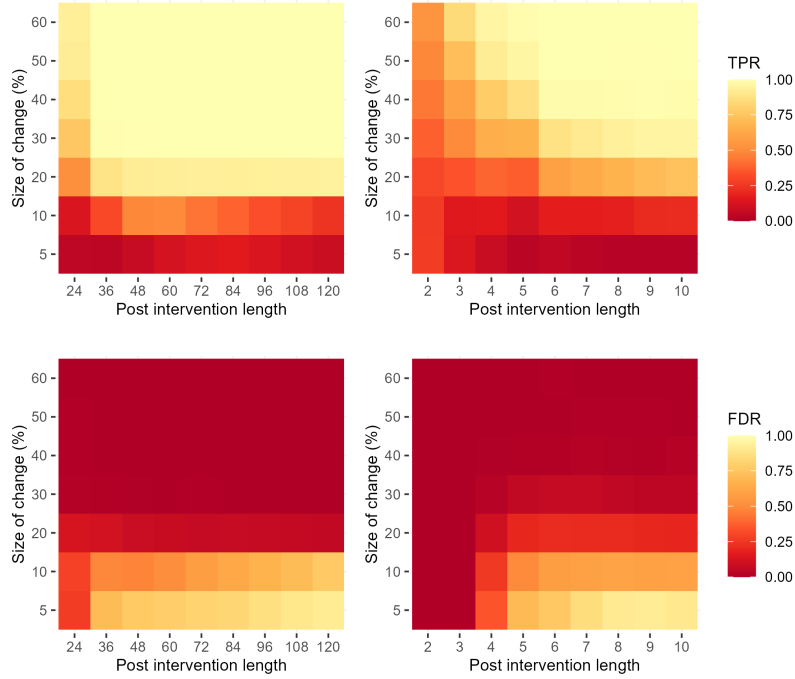


Figure 8: Heatmaps showing true positive rates (TPR) and false discovery rate (FDR) of different combinations of post-intervention length and percentage change to pre-intervention level from offline changepoint detection using AMOC, for time series of distribution change under BACI. The left column shows results for time series with higher sampling frequency and the right column shows results for time series with lower sampling frequencies.

Table 4: True positive rate (TPR), mean detection delay and standard deviation of delay from online changepoint detection of the first change in distribution under BACI with independent errors. The Forecast method was used for higher frequency time series and the BOCPD method was used for lower frequency time series.

Size of change	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Higher frequency										
TPR	0.40	0.56	0.85	0.95	0.95	0.95	0.95	0.94	0.95	0.96
mean delay	32.97	21.32	13.45	10.21	8.45	7.15	6.35	5.62	5.33	4.97
std. dev. delay	24.02	12.62	7.10	5.28	4.23	3.77	3.46	3.20	2.74	2.75
no detection	0	0	0	0	0	0	0	0	0	0
Lower frequency										
TPR	-	0.00	0.41	0.89	0.99	1.00	1.00	1.00	1.00	1.00
mean delay	-	8.33	7.49	5.69	4.22	3.45	3.06	2.83	2.52	2.21
std. dev. delay	-	1.30	1.46	1.42	0.84	0.58	0.40	0.40	0.50	0.41
no detection	1000	988	411	11	0	0	0	0	0	0

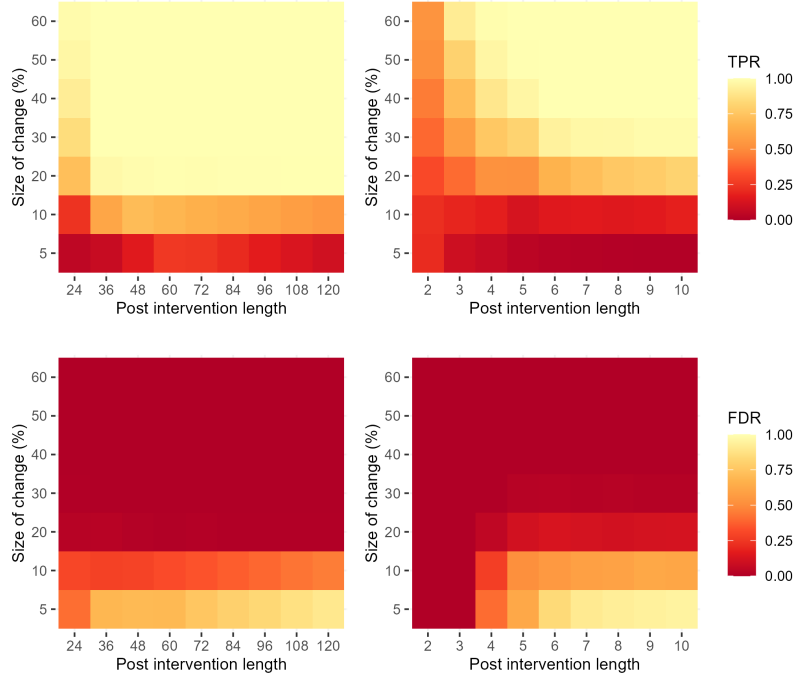


Figure 9: Heatmaps showing true positive rates (TPR) and false discovery rate (FDR) of different combinations of post-intervention length and percentage change to pre-intervention level from retrospective changepoint detection using AMOC, for time series of distribution change under BA. The left column shows results for time series with higher sampling frequency and the right column shows results for time series with lower sampling frequencies.