Masterarbeit

Marina Siebold

Spatio-Temporal Shifts in Citizen Science Data: Detecting Disruptions in Bird Sightings with Change Point Analysis

Faculty of Engineering and Computer Science
Department Computer Science

Marina Siebold

Spatio-Temporal Shifts in Citizen Science Data: Detecting Disruptions in Bird Sightings with Change Point Analysis

Master thesis submitted for examination in Master´s degree
in the study course *Master of Science Informatik*
at the Department Computer Science
at the Faculty of Engineering and Computer Science
at University of Applied Science Hamburg

Supervisor: Prof. Dr. Thomas Clemen
Supervisor: Dr. André Ebert

Submitted on: 06. Mar 2025

Marina Siebold

Thema der Arbeit

Spatio-Temporal Shifts in Citizen Science Data: Detecting Disruptions in Bird Sightings with Change Point Analysis

Stichworte

Zeitreihenanalyse, Change-Point-Analyse, Citizen Science, Bayesischer Schätzer, Ornithologie

Kurzzusammenfassung

Spezies, einschließlich Vögel, können Verschiebungen in ihrer Verbreitung und Häufigkeit erfahren, was sich auf Ökosysteme und die Biodiversität auswirken kann. Change-Point-Detection (CPD)-Methoden sind wertvolle Werkzeuge zur Identifikation solcher Veränderungen. Citizen Science bietet hierfür großflächige Datensätze, bringt jedoch auch beobachterbedingte Verzerrungen mit sich. Dies wirft Fragen zur Verlässlichkeit etablierter CPD-Algorithmen bei der Anwendung auf solche Daten sowie zu ihrer Akzeptanz unter Fachexperten auf.

Diese Arbeit greift diese Problematik auf, indem sie einen CPD-Ansatz unter Verwendung des "Bayesian Estimation of Abrupt Change, Seasonality, and Trend" (BEAST)-Algorithmus auf einen Citizen-Science-Vogeldatensatz anwendet. Vor der BEAST-Analyse wird eine Preprocessing-Pipeline entwickelt, um Beobachterverzerrungen zu reduzieren. Die Evaluation untersucht die Genauigkeit von BEAST sowie dessen ökologische Relevanz im Kontext von Citizen Science. Detektierte Veränderungspunkte wurden quantitativ mit dokumentierten ökologischen Ereignissen validiert, während Ornithologen die ökologische Plausibilität und praktische Relevanz qualitativ bewerteten.

Die Ergebnisse zeigen, dass BEAST ökologisch bedeutsame Veränderungspunkte zuverlässig erkennt, wobei seine Sensitivität von der Datenaggregation und den gewählten Preprocessing-Strategien abhängt. Obwohl CPD manuelle Bewertungen nicht ersetzt, wird es als wertvolle Ergänzung angesehen, um subtile oder unerwartete Veränderungen aufzudecken, die Echtzeit-Überwachung ökologischer Prozesse zu unterstützen und das Retraining von Machine-Learning-Modellen zu informieren.

Trotz des spezifischen Anwendungsfalls, unterstreicht diese Studie das breitere Potenzial von CPD in der Citizen Science, indem sie zeigt, dass mit robustem Preprocessing und Expertenvalidierung zeitnahe ökologische Erkenntnisse gewonnen werden können.

Marina Siebold

Title of Thesis

Spatio-Temporal Shifts in Citizen Science Data: Detecting Disruptions in Bird Sightings with Change Point Analysis

Keywords

Time Series Analysis, Change Point Analysis, Citizen Science, Bayesian Estimator, Ornithology

Abstract

Species, including birds, can undergo sudden shifts in distribution and abundance due to environmental changes, human activities, or natural variability, which can impact ecosystems and biodiversity. Change Point Detection (CPD) methods are valuable for identifying these shifts. For this, citizen science offers large-scale datasets, but also introduces observer-related biases, raising questions about the reliability of established CPD algorithms when applied to such data, and their trustworthiness among domain experts. This thesis addresses this concern by implementing a CPD approach using the Bayesian Estimation of Abrupt Change, Seasonality, and Trend (BEAST) algorithm on a citizen science bird dataset. Prior to BEAST analysis, a tailored preprocessing pipeline is developed to mitigate user bias.

Evaluation examines BEAST's accuracy and ecological relevance in citizen science contexts. Detected change points were quantitatively validated against documented ecological events, while ornithologists qualitatively assessed ecological plausibility and practical relevance.

Findings indicate that BEAST reliably detects ecologically meaningful change points, though its sensitivity depends on data aggregation and preprocessing strategies. While not replacing manual assessments, CPD is seen as a valuable complement to uncover subtle or unexpected changes, supporting real-time ecological monitoring, and informing machine learning model retraining.

Though case-specific, this study underscores CPD's broader potential in citizen science, enabling timely ecological insights when paired with robust preprocessing and expert validation.

Contents

List of Figures

List of Tables

1 Introduction

Ecological systems are inherently dynamic, shaped by intricate interactions among species, environmental factors, and human activities. While gradual changes are common, ecosystems can also experience abrupt shifts due to external pressures such as environmental changes, human influence, and natural variability. These sudden transitions can fundamentally alter ecosystem structure, function, and biodiversity, often with long-term consequences (Storch & Day 2020).

Accurately detecting these shifts is critical in ecology and conservation biology, as it can inform effective environmental management, conservation strategies, and policy decisions. A key methodological approach for this task is Change Point Detection (CPD), a statistical technique used to identify moments when underlying processes experience sudden or sustained changes.

However, the reliability of such analyses depends on comprehensive and geographically dense data. In recent years, citizen science initiatives have become a valuable resource for large-scale ecological monitoring. Through online platforms and mobile applications, volunteers record their sightings, collectively producing datasets that can far exceed the spatial and temporal coverage of conventional survey methods.

Despite their strengths in data volume and geographic reach, citizen science data also pose unique challenges: sightings are typically "presence-only", meaning absence data are not systematically recorded, and volunteers vary widely in observation skill, reporting frequency, and geographic preferences. Such biases can obscure true ecological signals, making it difficult to differentiate between genuine ecological changes and artifacts of reporting behavior.

1.1 Research Gap

Regardless of these complexities, citizen science datasets continue to play a growing role in ecology, raising two key gaps regarding the adoption of Change Point Detection methods:

1. **Algorithmic Transferability to Citizen Science Contexts**: While Change Point Detection algorithms have proven reliable in various time-series applications, research is limited on how they perform with the distinct properties of citizen science data in general, or bird sighting data in particular. The biases introduced by volunteer-driven data may affect the detection of true shifts in species distributions and abundances, raising questions about an algorithm's trustworthiness when transferred to such data.

2. **Lack of Systematic Expert Involvement**: Currently, domain experts are seldom involved in evaluating the practical relevance of CPD algorithms in ecology. Yet, their systematic input is crucial for verifying whether detected change points align with meaningful ecological events and for assessing the overall utility of the method. Without expert involvement, any misalignment with established ecological understanding or issues with interpretability can undermine trust and hinder integration, regardless of the algorithm's technical sophistication.

1.2 Objective and Research Questions

This thesis aims to address these gaps in applying Change Point Detection to citizen science data by focusing on bird species distributions in Germany and Switzerland.

At the core of this investigation is the Bayesian Estimation of Abrupt Change, Seasonality, and Trend (BEAST) algorithm, applied to volunteer-collected observations. A structured framework is introduced to (1) mitigate biases through data preprocessing, (2) detect potential shifts with BEAST, and (3) present the results through an accessible user interface.

By centering on the transferability of CPD approaches to citizen science data and ensuring ornithological expertise is incorporated, this thesis pursues both practical and scientific outcomes. In particular, five key questions guide the research, aiming not only to provide use-case insights but also to extract broader lessons for applying CPD algorithms in citizen science contexts:

- **RQ1: Accuracy and Correspondence with Ecological Phenomena** How accurate is the BEAST algorithm in detecting significant change points in citizen science time series data, and to what extent do these change points reflect actual ecological shifts and align with known real-world events? *Motivation*: Assess whether BEAST can robustly identify true changes and minimize false positives, while determining the degree to which detected breakpoints match expert-confirmed bird population phenomena.
- **RQ2: Role of Data Preprocessing** How does the applied data preprocessing method affect BEAST's ability to detect true ecological change points? *Motivation*: Understanding how data aggregation strategies can either highlight or obscure certain types of change events, thus influencing both detection rates and precision.
- **RQ3: Influence of Citizen Science Bias** In what ways can observer bias lead BEAST to detect false change points? *Motivation*: Exploring how volunteer-collected data may introduce artifacts.
- **RQ4: Expert Perception and Usability** How do ornithological experts perceive the usability and trustworthiness of BEAST for monitoring bird populations, and what are their perspectives on integrating it into ongoing research and workflows? *Motivation*: Examining the practical value of BEAST, including ease of interpretation, confidence in its outputs, and potential roles in expert-driven ecological monitoring.
- **RQ5: Generalization and Best Practices** What best practices and methodological insights emerge from this study for applying Change Point Detection methods to citizen science data in ecology? *Motivation*: Consolidating findings on into actionable guidelines to inform future research and ecological applications utilizing Change Point Detection on citizen science datasets.

To answer these questions, a complementary evaluation strategy is employed:

- **Quantitative Testing** compares detected change points to documented ecological events, testing BEAST's accuracy in reflecting actual population shifts.
- **Qualitative Interviews** with ornithologists assess whether the algorithm's outputs align with expert knowledge and whether the framework is trustworthy, interpretable, and practically applicable.

Through this combined approach, the thesis offers two major contributions. First, it delivers a **case-specific assessment** of how effectively BEAST—supported by bias-mitigating preprocessing—deals with large-scale, volunteer-collected bird sightings. Second, it produces **broader guidelines** for practitioners seeking to implement CPD on presence-only data, informed by expert feedback and grounded in real-world ecological contexts. These findings aim to help researchers and conservation professionals design, validate, and deploy robust CPD strategies that meet technical standards while earning the confidence of domain experts.

1.3 Outline

The work is organized as follows. First, the **Background** chapter is presented. This chapter provides an introduction to citizen science and Change Point Detection (CPD). Further, the BEAST algorithm is detailed.

The **Related Work** chapter examines existing strategies for mitigating user bias in citizen science data, explores CPD algorithms and their applications in ecological research, and reviews common evaluation methods for CPD algorithms.

In the following **Implementation** chapter, the framework developed for this thesis is presented. It details the data preprocessing pipeline—spatial discretization, weekly aggregation, and user bias mitigation—followed by the application of BEAST to each species' time series. Finally, a web-based user interface for visualizing detected shifts is introduced.

The **Evaluation** chapter explains the overall design of the evaluation methods. It covers two complementary strategies—semi-structured interviews with ornithologists to gather expert feedback and a quantitative comparison of BEAST outputs with archival records of documented bird population changes.

The **Results** chapter compiles the outcomes of both the qualitative and quantitative evaluations. It presents ornithologists' perspectives on the accuracy and utility of BEAST, along with detection rates measured against the archival data.

Building on the evaluation results, the **Discussion** chapter systematically discusses each Research Question in light of the findings. It addresses the strengths and limitations of BEAST under the chosen preprocessing strategy, highlights the algorithm's real-world applicability, and presents broader implications for applying CPD methods to citizen science datasets.

The final **Outlook** chapter presents future directions for research and practical integration. It offers recommendations for improving the adopted methodology, addresses remaining challenges, and outlines potential avenues for future research, thereby offering a roadmap toward more agile, data-driven ecological monitoring.

1.4 Context of this work

This thesis was developed within the Ornitho project, supported by the Federation of German Avifaunists (DDA) and the Swiss Ornithological Institute. Their platforms collect millions of bird observations across multiple European countries, providing a major resource for biodiversity research but also posing challenges related to data reliability and volunteer biases.

To improve data quality, Ornitho is exploring AI-driven methods to flag implausible sightings. Motivated by this need for adaptive models, the concept of building a Change Point Detection system for each bird species was proposed. This thesis serves as a first step in this direction by assessing whether Change Point Detection is applicable to citizen science data in general and the Ornitho use case in particular.

This thesis represents a collaborative effort between the University of Applied Sciences Hamburg (HAW), the DDA, the Swiss Ornithological Institute, and inovex GmbH, bringing together academic, domain, and technical expertise to explore CPD's potential in large-scale citizen science contexts.

2 Background

2.1 Citizen Science

Citizen science refers to the involvement of non-professional scientists in the research process, encompassing activities such as data collection, analysis, and dissemination of scientific knowledge (Vohland et al. 2021). It represents a collaborative approach where members of the public contribute to scientific projects, often in partnership with professional scientists and institutions; this participation can take various forms, ranging from recording observations of natural phenomena to analyzing astronomical or medical data (Davis et al. 2023).

Advancements in technology and the proliferation of internet connectivity have significantly expanded the scope and scale of citizen science (Bonney 2021). Online platforms, mobile applications, and GPS-enabled devices have made it easier for volunteers to participate in projects worldwide, enabling real-time data collection and analysis (Johnston et al. 2023). This technological evolution has transformed citizen science into a global phenomenon, engaging participants in diverse research initiatives.

2.1.1 Relevance

Citizen science contributes substantially to the advancement of scientific knowledge by facilitating large-scale data collection that would be otherwise impractical for individual researchers or small teams (Berghen et al. 2025). Projects like *eBird* leverage the contributions of birdwatchers globally to monitor bird distributions and migration patterns, generating large datasets that inform conservation efforts and ecological studies (Zhu & Newman 2024).

Engaging the public in scientific research fosters a greater appreciation and understanding of science. According to Smith et al. (2024), citizen science projects not only advance scientific research but can also foster a more scientifically literate society and increase civic action.

Additionally, citizen science enhances the democratization of science by making it more accessible and participatory. It breaks down barriers between professional scientists and the public, encouraging collaboration and dialogue. This inclusive approach can lead to increased trust in scientific institutions (Hall et al. 2024).

2.1.2 Fields of Application

Citizen Science is applied to a vast amount of research fields, including Food Science (e.g. *PataFEST*), Astronomy (e.g. *Galaxy Zoo*), and Epidemiology (e.g. *Flu Near You*). However, one of the most prominent applications of citizen science is in environmental monitoring.

Figure 2.1 shows the proportion of ecological citizen science projects per taxonomic group, compared to their abundance on Earth and the number of professional scientists in that field. It highlights that birds, amphibians, reptiles, and mammals are significantly over-represented in citizen science projects compared to their relative abundance on Earth. This bias is likely due to the ease of observing and identifying these species, as well as the public's general interest in them; meanwhile, less charismatic species or taxonomic groups that are harder to access are underrepresented, leading to potential gaps in biodiversity monitoring (Theobald et al. 2015).

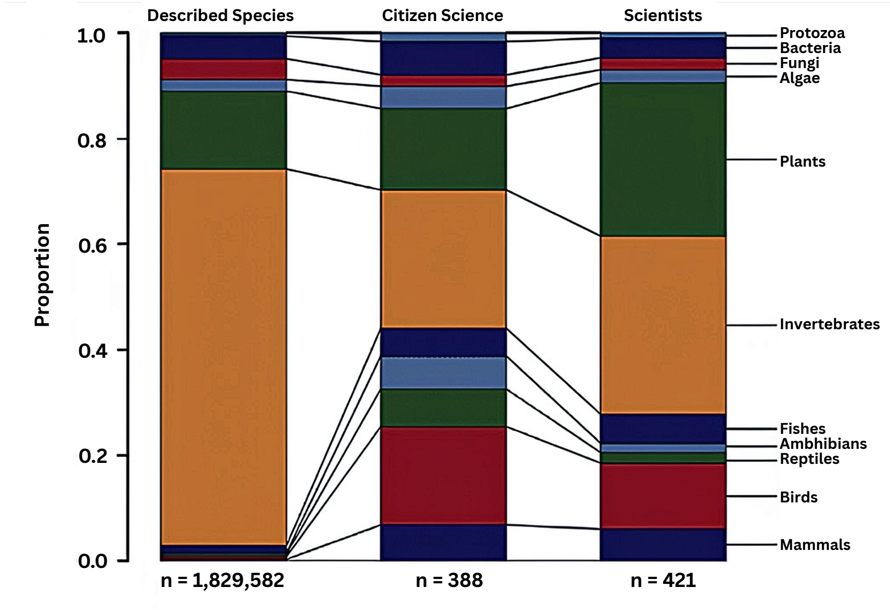


Figure 2.1: Proportion of sampled citizen science projects covering one or more taxonomic categories (middle stacked bar) compared to each group's relative abundance on Earth (left bar) and representation among professional scientists (right bar). Adopted from Theobald et al. (2015).

Further, Figure 2.2 shows that citizen science projects are predominantly focused on terrestrial and freshwater environments, while marine ecosystems are significantly underrepresented. This disparity presumably arises because marine environments, despite covering a vast portion of the Earth's surface, are less accessible to the public.

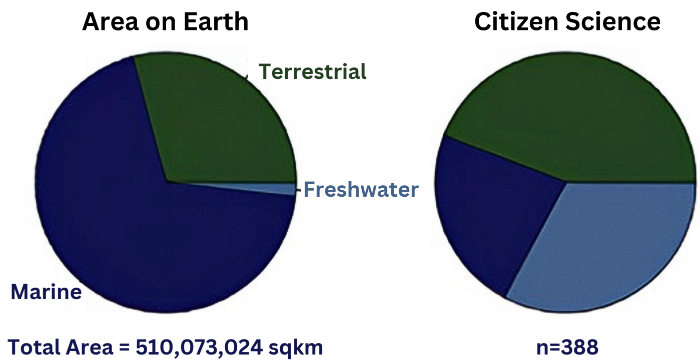
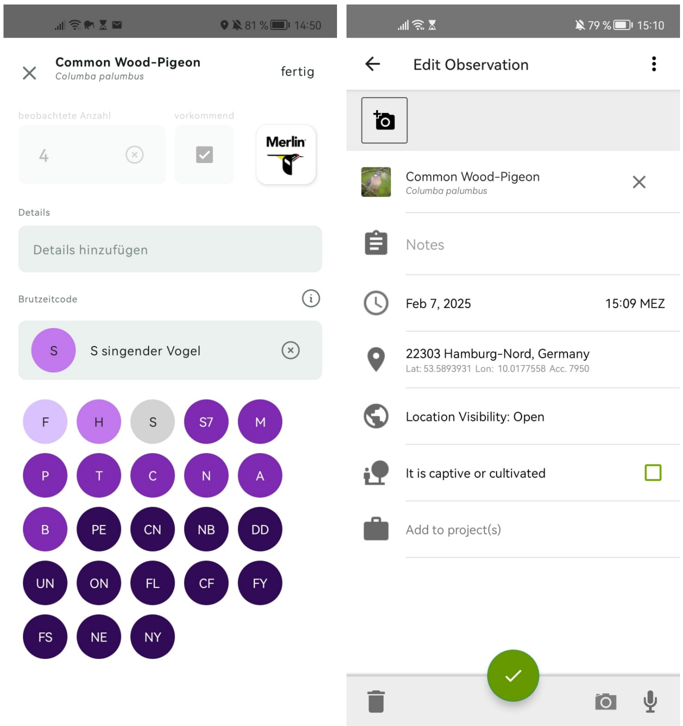


Figure 2.2: Proportion of sampled citizen science projects collecting data in terrestrial, freshwater, or marine environments relative to their global area coverage. Adopted from Theobald et al. (2015).



(a) eBird (b) iNaturalist

Figure 2.3: Input interface for a new observation in the two most commonly used bird citizen science project apps in Europe.

For bird species in particular, the most widespread citizen science projects include programs such as *eBird*, operated by the Cornell Lab of Ornithology, which allows birdwatchers worldwide to record and share their observations. In Europe, the most prevalent citizen science effort is the *Ornitho* project, used in multiple countries. The mobile app for this project is *iNaturalist*. These projects engage volunteers in large-scale data collection, helping researchers monitor bird populations, migration patterns, and long-term ecological changes. Figure 2.3 displays the two similar input interfaces used for logging new observations of both mobile apps.

2.1.3 Presence-Only and Presence-Absence Data Collection

Citizen science projects employ various data collection methods tailored to the project's objectives and the participants' capabilities. For citizen science projects in ecology, specifically, data can be collected in different ways. The most prevalent types are presence-only and presence-absence data.

For presence-only data, volunteers report sightings of species without noting their absences. This type of data collection is straightforward as it only involves recording observations when they occur, without the need for systematic surveys or additional tracking efforts. However, it requires careful analysis to address sampling bias and spatially uneven observation efforts (Di Febbraro et al. 2023), which can result in under-sampled regions. These gaps in data coverage may lead to inaccurate assumptions, such as falsely concluding the absence of species in poorly surveyed areas. Platforms like *eBird* and *Ornitho* collect presence-only data.

Another important type is presence-absence data, in which participants record both the presence and confirmed absence of species in surveyed areas. This comprehensive data provides deeper insights into species distributions and is crucial for monitoring changes over time. However, collecting reliable absence data necessitates rigorous protocols to ensure that non-detections represent true absences rather than overlooked presences (Cruickshank et al. 2019).

2.1.4 User Bias

A key challenge in analyzing ecological citizen science data lies in understanding that this data source does not provide direct estimates of abundance. In contrast to structured surveys, where individuals are systematically counted, citizen science data represent a continuous accumulation of sighting reports. This introduces several factors that distort the true number of individuals present in a given area.

First, spatial biases can arise when observations are concentrated in easily accessible or popular locations, leading to uneven data coverage (Backstrom et al. 2025). Consequently, remote or less accessible regions tend to be undersampled. Furthermore, users may select sites based on ecological factors, favoring biodiversity hotspots or areas where they anticipate finding a particular species of interest (Johnston et al. 2020).

Temporal biases can also be present in the data. Volunteers tend to be more active on weekends and during specific times of the year. In bird monitoring, for example, user activity peaks during spring and autumn due to bird migration, when birds are more visible and active (La Sorte & Somveille 2020).

Further, a bias exists due to user errors, which stem from inaccurate or incomplete submissions. These errors can take the form of False Positives, where an observer mistakenly records a species that was not present (Johnston et al. 2023). Likewise, False Negatives occur when a species is present but goes unreported due to lack of detection (Rempel et al. 2019).

Another critical source of bias arises from taxonomic preferences. Volunteers are often motivated by encounters with interesting wildlife. Observers are more likely to travel further to record rare or more charismatic species like the *White-tailed Eagle* while neglecting more common taxa, such as the *Wood Pigeon*. This introduces an additional layer of bias toward species that are considered more noteworthy (Johnston et al. 2023).

These factors, collectively, contribute to significant user bias in the data, making it difficult to directly link sighting records to the actual number of individuals present. If not addressed, this can significantly affect the outcomes of ecological studies (Backstrom et al. 2025). For Change Point Detection, this is particularly relevant as it aims to identify disruptions that reflect real changes in bird populations, rather than patterns influenced by user behavior. Therefore, it is critical to develop methodologies that account for these biases and ensure that the detected change points correspond to actual shifts.

2.2 Change Point Detection

Change Point Detection (CPD) is a fundamental statistical task used to detect shifts in a dataset over time. A change point marks an instance in time where the statistical properties of a time series experience a significant alteration. Detecting these points is essential for understanding underlying dynamics in data.

2.2.1 Definition

Formally, given a sequence of observations $\{X_t\}_{t=1}^{N}$, a change point exists if the statistical properties of the sequence before and after a certain time τ differ. That is, τ partitions the time series into segments with different probability distributions. Change Point Detection involves identifying the number and positions of these change points.

Considering a sequence of time series variables $\{x_m, x_{m+1}, \ldots, x_n\}$, Change Point Detection can be formulated as a hypothesis testing problem between two alternatives. According to Aminikhanghahi & Cook (2017), this can be expressed as:

- **Null hypothesis** H_0 (no change occurs):

$$H_0 : \mathbb{P}_{X_m} = \mathbb{P}_{X_{m+1}} = \cdots = \mathbb{P}_{X_n} \tag{2.1}$$

- **Alternative hypothesis** H_A (a change occurs at some point τ):

$$H_A : \exists\, m < \tau < n \text{ such that } \mathbb{P}_{X_m} = \cdots = \mathbb{P}_{X_\tau} \neq \mathbb{P}_{X_{\tau+1}} = \cdots = \mathbb{P}_{X_n} \tag{2.2}$$

where $\mathbb{P}_{X_i}$ denotes the probability distribution of X_i, and τ is the change point. The goal of CPD is to determine whether and where such a point τ exists within the interval (m, n) at which the distribution of the observations changes.

The types of changes that may constitute a change point include, but are not limited to, alterations in the mean, dispersion, count, or slope of the data (see Figure 2.4). A common type of change point is a sudden shift in the mean of the data, where the average value of observations changes abruptly at point τ. Similarly, a change in the variance or standard deviation refers to alterations in the data's dispersion, indicating that observations become more or less spread out after the change point. Further, a change in the slope denotes a shift in the trend component of the time series, where the rate of increase or decrease changes at τ. Lastly, changes in periodicity involve modifications in the cyclical patterns within the data, such as the emergence or disappearance of seasonal effects.

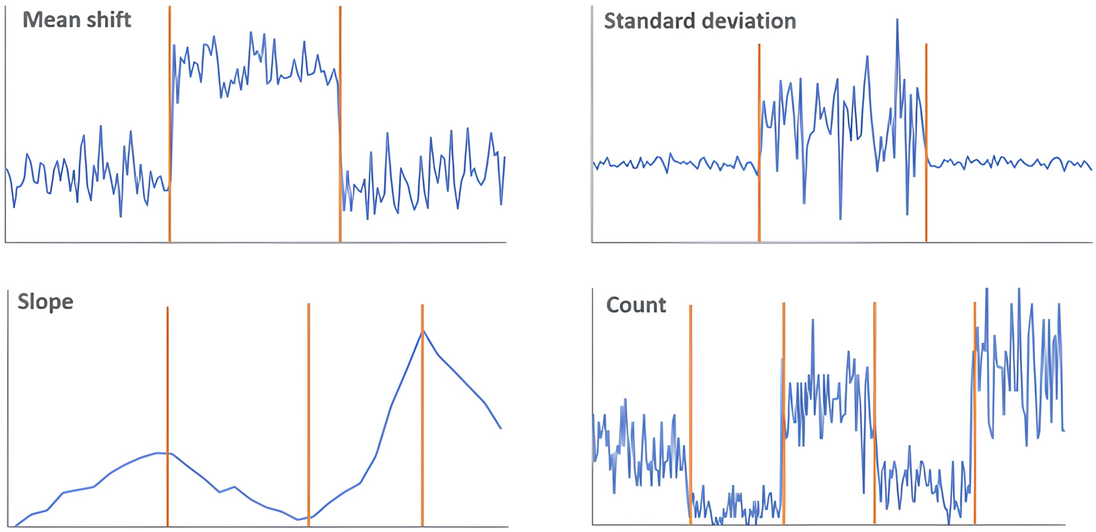


Figure 2.4: Types of change points in time series data, differentiated by statistical properties. Adopted from Arcgis (2024).

The exact nature of what constitutes a change point is not strictly defined regarding the suddenness or gradualness of the transition. Some researchers consider only abrupt shifts as change points, while others also include gradual transitions (Zhao et al. 2019). The ambiguity arises because real-world data often exhibit changes that are neither instantaneous nor entirely smooth. As such, the definition of a change point can be context-dependent, and detecting gradual changes may require different analytical approaches than detecting abrupt ones.

2.2.2 Relevance

Detecting change points in time series data is crucial across various scientific fields, including ecology, finance, and engineering (Aminikhanghahi & Cook 2017). It is vital for identifying events, which may be subtle yet significant. In ecological studies, for example, change point analysis can reveal shifts in species population, water quality, or vegetation development, indicating environmental changes, habitat loss, climate change effects, or anthropogenic impacts (Fan et al. 2024). Understanding these changes is essential for timely conservation efforts to mitigate adverse effects. Moreover, identifying periods of significant change can help scientists focus on investigating potential causes, leading to better-informed policy decisions.

Another relevant contribution of automatic Change Point Detection is the significant reduction in time required to identify predefined events within large-scale time series datasets. This automation minimizes the need for repetitive manual processing. For example, change point analysis can mark an organism's time of death based on a sudden decline in brain activity (Aqel et al. 2024).

2.2.3 Change Point Detection Methods

Methods for detecting change points changes can be broadly categorized into model-based and non-parametric approaches, as well as supervised and unsupervised techniques (Aminikhanghahi & Cook 2017). Additionally, they can be distinguished as offline methods, which analyze the entire time series retrospectively, and online methods, which detect changes as new data arrives (Van den Burg & Williams 2020).

Statistical methods form the foundation of Change Point Detection. Basic approaches, such as Cumulative Sum (CUSUM) and Likelihood Ratio Tests, identify structural breaks by measuring deviations from an expected pattern. CUSUM accumulates small deviations over time and signals a change when the cumulative sum exceeds a predefined threshold (Horváth et al. 2022). Likelihood Ratio Tests compare the probability of the data fitting two different models, one assuming no change and the other incorporating a possible break (Skrobotov 2023).

Segmentation-based methods take a different approach by dividing the time series into distinct segments, optimizing a predefined cost function. Binary Segmentation is a recursive technique that detects the most significant change point and then applies the same process to the resulting segments until no further substantial changes are found (Kovács et al. 2023). The Pruned Exact Linear Time (PELT) method improves upon this by applying a pruning strategy, which reduces computational complexity to $O(n)$ (Truong et al. 2020).

More advanced techniques include Bayesian approaches, which treat change points as random variables and use Bayesian inference to compute their distributions. This probabilistic framework allows the incorporation of prior knowledge and the quantification of uncertainty regarding the location and number of change points (Zhao et al. 2019). Bayesian methods can be particularly powerful in applications where prior information is available or where uncertainty quantification is crucial.

Machine learning methods, including Kernel-Based methods and Clustering algorithms, can capture complex, non-linear relationships in the data, detecting structural changes that traditional methods might miss (Aminikhanghahi & Cook 2017). While unsupervised clustering algorithms such as K-means can identify change points as transitions between clusters, supervised learning approaches such as Decision Trees can be trained to recognize patterns associated with change points (Aminikhanghahi & Cook 2017).

Online detection methods are essential in scenarios where immediate response to changes is required. Online CUSUM is an extension of the standard CUSUM approach that continuously updates its calculations as new data arrives, triggering an alert when deviations exceed a defined threshold (Wei & Xie 2022). Bayesian online Change Point Detection models estimate a probability distribution for the time that has passed since the most recent change point (Van den Burg & Williams 2020). Both supervised and unsupervised machine learning models can process streaming data within a sliding window with size n, allowing for near-real-time Change Point Detection (Aminikhanghahi & Cook 2017).

Ultimately, selecting the appropriate Change Point Detection method depends on multiple factors, including data complexity, computational efficiency, and the need for real-time analysis.

BEAST Algorithm

The Bayesian Estimator of Abrupt Change, Seasonal Change, and Trend (BEAST) by Zhao et al. (2019) is an advanced time series decomposition algorithm developed to address the challenges of analyzing complex, nonlinear dynamics in ecosystems. BEAST was initially developed using satellite time series data to track land-use changes, vegetation dynamics, and ecosystem disturbances. However, the authors emphasize that its ability to capture both subtle and prominent changes makes it applicable in fields such as disturbance ecology, climate science, and land resource management. The algorithm is suitable for a broad range of time series data types, including Normalized Difference Vegetation Index (NDVI), climate variables, and various ecological indicators. It has been successfully applied in various other research fields for Change Point Detection, including wildlife research, forestry, oceanography, geophysics, euthanasia, and food science (Smith & Pauli n.d., Mulverhill et al. 2024, Oehlert et al. 2023, Mu et al. 2023, Aqel et al. 2024, Zaytsev et al. 2024). Thise wide applicability and strong acceptance within the ecology research community were key factors in selecting BEAST for this thesis.

Unlike conventional single-model approaches, BEAST employs a Bayesian ensemble method, which integrates multiple model outcomes to produce a comprehensive and probabilistically reliable estimate of temporal patterns. The authors suggest that, rather than aiming to identify a single best-fit model, each model should be viewed as valuable in its own way, contributing unique insights that collectively enhance overall understanding (Zhao et al. 2019). This Ensemble approach helps mitigate issues stemming from trying to select one single model based on an arbitrary criterion, such as the Akaike's Information criterion (AIC) or the Bayesian information criterion (BIC), which might lead to misinterpretations and fortuitous conclusions of the data (Zhao et al. 2019). Such misconceptions may significantly influence our understanding of ecosystems and the policy decisions we implement.

Using BEAST, the input data can have multiple, overlapping sources of variation, including seasonality and trend. It is suited for cases where there is a need to detect both abrupt (high-magnitude) and subtle (low-magnitude) changes (Zhao et al. 2019). Further, in its latest version, BEAST accounts for outliers in the data, preventing these atypical data points from skewing the model's understanding of the underlying seasonal and trend patterns. The output of the BEAST analysis also includes additional valuable statistical diagnostics, such as change point probabilities and confidence intervals.

The BEAST algorithm decomposes a time series into four primary components: trend, seasonality, noise, and abrupt change (i.e., change points). Its goal is to accurately replicate the given, complex original signal by estimating these components. To achieve this, BEAST views a time series $y(t)$ as the sum of its trend, seasonal, and noise components. Mathematically, this can be expressed as:

$$y(t) = S(t; \Theta_s) + T(t; \Theta_T) + \epsilon \tag{2.3}$$

where $S(t; \Theta_s)$ represents the seasonal component parameterized by Θ_s, $T(t; \Theta_T)$ is the trend component parameterized by $T(\Theta_T)$, and ϵ is the noise component, accounting for the portion of the data not explained by the seasonal and trend components. The seasonality component $S(t; \Theta_s)$ is modeled as a piecewise harmonic function with specified frequencies, while the trend component $T(t; \Theta_T)$ is modeled as a piecewise linear function. Change points for seasonality and trend are introduced as breaks in the time series where either the seasonal frequency or the trend slope changes. Each change point in the trend or seasonal component allows the algorithm to capture abrupt shifts in the data's behavior.

The seasonal component of the time series is mathematically represented as a piecewise harmonic model, divided by p seasonal change points. These change points split the signal into $p+1$ segments, each with its own phase and amplitude. Thus, the seasonal signal is given by:

$$S(t) = \sum_{l=1}^{L_k} \left(a_{k,l} \sin \frac{2\pi l t}{P} + b_{k,l} \cos \frac{2\pi l t}{P} \right) \tag{2.4}$$

where L_k is the harmonic order of the k-th segment, P is the seasonal period (fundamental frequency), and $a_{k,l}$ and $b_{k,l}$ are the amplitudes for the sins and cosines term in the segment. Partitioning the harmonic component of the time series results in a non-continuous signal with unknown parameters, specifically the number and timing of change points, as well as the phase and amplitude of each harmonic oscillation.

The trend component of the time series is modeled as a piecewise linear function, where each segment is represented as

$$T(t) = a_j + b_j t \tag{2.5}$$

where a_j and b_j are the intercept and slope of the j-th segment, respectively. The points in time where segments transition are the change points of the trend component. In addition to the parameters of the linear segments, the trend component also has unknown parameters, similar to the seasonal component, namely the number and timing of change points, which need to be estimated.

Thus, in order to estimate the time series, the algorithm must estimate the following parameters, which collectively define the structure of the time series:

- **Model Structure** M
  - Number and timing of change points for trend and seasonality
  - Harmonic orders of the seasonal elements
- **Segment-Specific Coefficients** β_M

  - Trend: Slopes and intercepts of the linear trend segments
  - Seasonality: Sine and cosine coefficients of the harmonic functions

- **Noise Parameter σ^2**
  - Describes the noise component

The estimation process begins with an initial guess, after which the quality of this model is evaluated using the posterior probability. As formulated by Zhao et al. (2019), the posterior probability $p(\beta_M, \sigma^2, M|D)$ is calculated based on Bayes' Theorem:

$$p(\beta_M, \sigma^2, M|D) = \frac{p(D|\beta_M, \sigma^2, M) \cdot p(\beta_M, \sigma^2, M)}{p(D)} \tag{2.6}$$

where $p(\beta_M, \sigma^2, M|D)$ is the posterior probability, $p(D|\beta_M, \sigma^2, M)$ is the likelihood, $p(\beta_M, \sigma^2, M)$ is the prior probability, and $p(D)$ is the marginal likelihood, which normalizes the posterior distribution across all possible models, ensuring that the posterior probabilities sum to 1.

The prior $p(\beta_M, \sigma^2, M|D)$ reflects general assumptions set by the authors, including constraints on the change points' minimum and maximum numbers and their minimum separation distance. These settings are customizable, allowing for flexibility across different application contexts. The authors intentionally designed the prior to be generic and non-informative, ensuring the algorithm's applicability to a broad range of time series (Zhao et al. 2019).

The likelihood $p(D|\beta_M, \sigma^2, M)$ is distinct from the prior as it is purely data-driven, depending on the input time series. In BEAST, the likelihood quantifies how well the model M explains the observed data D. Given the model structure M, the coefficients β_M , and the noise parameters σ^2, the time series data is decomposed into trend and seasonal components, with the likelihood $p(D|\beta_M, \sigma^2, M)$ reflecting the fit of these components to the actual data points. The likelihood calculation involves estimating how closely the observed data y_i matches the predicted values based on the current model configuration, accounting for both the seasonal and trend components.

BEAST employs a Monte Carlo-based inference approach for exploring the vast model space of BEAST, as it allows efficient sampling of possible model configurations without requiring an exhaustive search. The MCMC sampling process iteratively generates samples of model parameters by alternately sampling the model structure M, the parameters β_M of the seasonal and trend components for the chosen model, and the noise variance σ^2. Each iteration updates the model parameters based on the current data and model, refining the posterior distribution with each step. This iterative process eventually converges, yielding a set of models and associated parameter estimates that reflect the data's underlying patterns.

After the MCMC sampling process, BEAST combines the results from all sampled models using Bayesian Model Averaging (BMA). This allows the algorithm to handle model uncertainty by combining multiple candidate models rather than selecting a single best-fit model. The posterior inference is performed on the aggregated set of sampled models, where each model M is weighted by its calculated posterior probability. This approach yields an overall estimate for the time series decomposition, combining the trend, seasonality, and change point estimates across models. The posterior distribution for each parameter is used to calculate uncertainty intervals, providing a measure of confidence in the results.

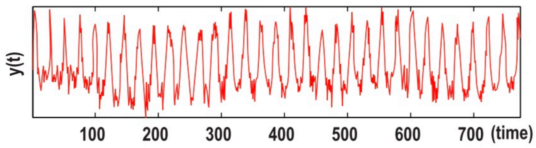
The final estimate $\hat{y}(t)$ of the time series is derived as follows:

$$\hat{y}(t) = \frac{1}{N} \sum_{i=1}^{N} y_M^{(i)}(t) \tag{2.7}$$

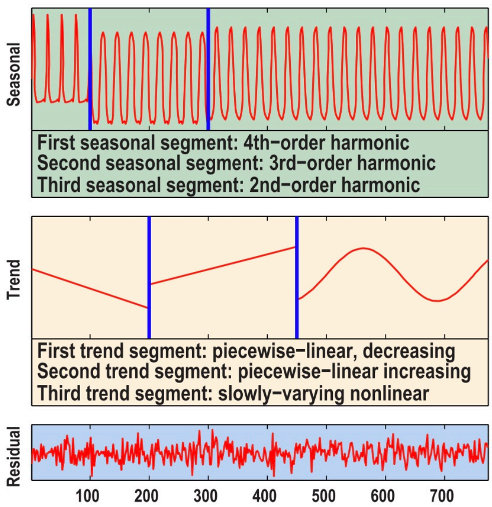
where N is the total number of sampled models, and $y_M^{(i)}(t)$ represents the time series decomposition from the i-th sampled model.

Through posterior inference, BEAST produces a robust time series decomposition with confidence intervals around trend and seasonal estimates, capturing the variability and uncertainties in the model space. This approach allows for the reliable detection of change points and seasonal patterns, which are essential for interpreting ecological and environmental dynamics accurately.

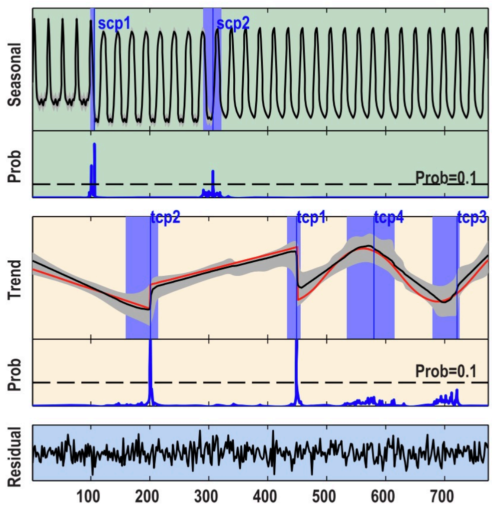
After performing the change point analysis, the BEAST package provides a comprehensive array of output plots, compiled in Figure 2.5. In Figure 2.5a, an exemplary input time series $y(t)$ is shown in red, illustrating the observed data that is analyzed. Its true underlying dynamics are decomposed in Figure 2.5b, with distinct seasonal and trend components shown as separate curves.



(a) A simulated time series with seasonal and trend components.



(b) True dynamics of the time series in (a).



(c) Dynamics extracted by BEAST.

Figure 2.5: Illustration of BEAST analysis on a simulated time series. The figure shows the input time series (a), the true dynamics (b), and the outputs of BEAST (c), including detected seasonal and trend signals, change points (*scp*/*tcp*), uncertainty envelopes, and the probability of observing a *scp* or *tcp* at any given time. Adopted from Zhao et al. (2019).

Figure 2.5c depicts the dynamics extracted by BEAST. Highlighted in green, the seasonal component detected by BEAST is shown, along with the identified seasonal change points, labeled as *scp*1 and *scp*2, and marked with blue vertical bars. Bayesian Model Averaging enables the a confidence estimation at each time step, providing probabilities for the presence of a change point, which are shown below the extracted seasonal signal.

The subsequent plot, highlighted in yellow, illustrates the trend component, which is modeled as piecewise linear segments. This plot also includes the probabilities for a trend change point, shown as probability curves below the trend. Credible intervals of the estimated trend signal are displayed as gray envelopes, offering a more comprehensive uncertainty range compared to single-best-model approaches. The final plot shows the residuals in blue, which represent the portions of the signal not classified as either seasonal or trend components by BEAST.

Additionally, BEAST provides further insights (see Figure 2.6), such as the probability distribution of the estimated total number of change points for both seasonal and trend components, the order of seasonal harmonics per seasonal segment, and the probability of a positive trend change at any given time.

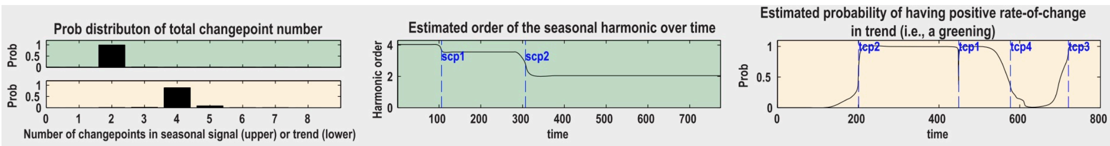


Figure 2.6: Additional output of BEAST, showing further estimation analytics. Adopted from Zhao et al. (2019).

3 Related Work

In the development and evaluation of the proposed framework, three main components are considered: a preprocessing scheme to reduce user bias in the applied citizen science data, the application of a Change Point Detection algorithm, and the evaluation of its performance. To provide context for these components, this chapter presents an overview of Related Work in these areas.

3.1 User Bias Reduction of Citizen Science Data

While citizen science has democratized data collection and greatly expanded geographical coverage, it also introduces challenges pertaining to data quality and reliability. As discussed in Chapter 2.1.4, a major concern is user bias, which arises from variations in observer effort, expertise, and reporting practices.

To mitigate the spatial unevenness of sampling effort, Matutini et al. (2021) and Steen et al. (2021) propose subsampling procedures. Such approaches have, among others, been tested on *eBird* data (Johnston et al. 2021). Typically, density-based sampling is employed, whereby a lower proportion of records is retained from densely sampled areas, thus achieving more balanced dataset coverage.

However, subsampling leads to a loss of valuable data. As an alternative, artificial imputation can replace missing values with plausible estimates (Bowler et al. 2025). For instance, Grattarola et al. (2023) employ species distribution models (SDMs) to estimate species presences, thereby reducing spatial imbalances. Similarly, Dakki et al. (2021) developed an imputation method to address spatial and temporal data gaps simultaneously.

Zbinden et al. (2014) applied the SOPM abundance index to address the issue of uneven sampling efforts in opportunistic biological records. Although published some time ago, this method remains relevant (Bowler et al. 2025). SOPM (*Summe der Ortspentadenmaxima*) is a German acronym describing its computation: for a target species, it represents the sum of the maximum number of records within a selected area and over a five-day period. This index has already been successfully applied to the same dataset used in this thesis by ornithologists from the Swiss Ornithological Institute, and was therefore chosen to be applied in this thesis.

In contrast to post-processing solutions, Callaghan et al. (2023) suggest reducing biases at the data collection stage. They recommend motivating participants to record observations during underrepresented periods and in underrepresented locations.

3.2 Change Point Detection for Ecological Studies

3.2.1 Algorithms

Performing Change Point Detection in ecology is inherently complex due to the intricate and dynamic nature of ecological systems. These complexities necessitate sophisticated CPD methods that can accurately discern genuine shifts from natural fluctuations. This section reviews prominent Change Point Detection algorithms applied for ecological studies, emphasizing traditional methods, their limitations, and recent advancements that have enhanced robustness and flexibility.

Traditional methods, such as global linear models applied by Myneni et al. (1997), provided foundational insights but often oversimplified nonlinear ecological processes (Zhao et al. 2019). Piecewise linear models improved flexibility by segmenting time series, enabling the detection of abrupt shifts, but remained sensitive to noise and required user-defined parameters (Banesh et al. 2019).

To better capture ecological dynamics, additive models like DBEST and bfast were introduced. DBEST integrates a level-shift detection mechanism with thresholds to identify discontinuities in trends, followed by decomposition into trend and seasonal components using STL (Seasonal and Trend decomposition using Loess) (Li et al. 2022). The Breaks For Additive Seasonal and Trend (bfast) algorithm developed by Verbesselt et al. (2010) decomposes time series data into trend, seasonal, and remainder components, allowing for the detection of structural changes in both trend and seasonality. However, both models rely on fixed parameter choices, which can lead to inconsistencies. For example, conflicting interpretations of Amazon forest responses to droughts highlight the limitations of single-model approaches (Zhao et al. 2019).

Probabilistic methods, such as BEAST, address these challenges by estimating change point probabilities rather than making binary decisions. Unlike bfast, BEAST does not require pre-specified change point numbers and can model complex trends and seasonal variations more flexibly. By averaging multiple models, it provides a nuanced perspective on ecosystem shifts, improving reliability in ecological decision-making (Zhao et al. 2019). BEAST is further detailed in Chapter 2.2.3.

Li et al. (2022) conducted a comparative evaluation of BEAST, DBEST, and bfast for land surface temperature time series analysis, specifically. Their study demonstrated that BEAST outperformed the other methods in terms of accuracy and robustness to noise, particularly in capturing abrupt changes and non-linear dynamics. Further, they proposed an improved version of BEAST, which addressed its susceptibility to false breakpoints by introducing thresholds based on trend magnitude, slope changes, and breakpoint probabilities.

Beyond its measured superiority, BEAST is also widely applied in recent research to detect regime shifts, disturbances, and ecological changes. It has been used across various domains, including wildlife research, forestry, oceanography, and geophysics (Smith & Pauli n.d., Mulverhill et al. 2024, Oehlert et al. 2023, Mu et al. 2023), making it a valid choice for the analyses conducted in this thesis.

3.2.2 Applications using Citizen Science Data

While citizen science data is increasingly recognized as a valuable resource for large-scale ecological monitoring, the application of CPD algorithms to such datasets remains limited. Several studies have explored methods for detecting ecological changes using citizen science data, demonstrating both the potential and the challenges associated with these approaches. In this section, some influential papers will be reviewed, with no claim to completeness.

Gouraguine et al. (2019) investigated the effectiveness of marine citizen science programs in detecting long-term ecosystem changes, particularly in resource-limited areas where no alternative data sources exist. Using an 11-year dataset collected by volunteers in

Southeast Sulawesi, Indonesia, the study analyzed changes in coral reef ecosystems by examining benthic cover and fish communities. The authors employed generalized linear models (GLMs) to identify trends and assess whether changes over time were statistically significant. Their findings demonstrate that citizen science data can effectively track long-term ecosystem changes, though data quality control remains a significant challenge.

Although relatively dated, the work of Walker & Taylor (2017) remains highly relevant, as it closely aligns with the use case examined in this thesis. The study evaluates the reliability of citizen science data from the eBird platform in modeling long-term population trends of migratory bird species. By comparing eBird data with the North American Breeding Bird Survey (BBS), the study assessed whether volunteer-collected data could provide accurate estimates of population changes. The results showed that eBird data successfully detected similar population trends as the BBS for many bird species. However, for species with low detection rates in eBird, trend estimates were more uncertain.

Habel et al. (2022) integrated presence-only data from multiple sources, including volunteer-collected contributions, and employed Linear and segmented regression to identify change points in species traits and habitat associations over time. These techniques enabled researchers to pinpoint critical changes in butterfly populations.

3.2.3 Methods for CPD algorithm evaluation

Due to the wide use of CPD, extensive effort has gone into evaluating algorithm performance, often by assessing accuracy in retrieving known change points. While straightforward with ground-truth labels, this becomes challenging in fields like ecology, where true changes are often uncertain (Zhao et al. 2019). Researchers address this with diverse validation approaches, offering a broader view of algorithm reliability.

A common approach is using synthetic data with known parameters, providing a baseline for algorithm performance. If an algorithm fails to detect embedded change points in such data, it is unlikely to succeed in real-world scenarios. Li et al. (2022) compared BEAST and bfast using this method.

Beyond controlled settings, Zhao et al. (2019) recommend qualitative and quantitative validation against general or expected patterns to assess whether detected changes align with expected patterns. For instance, a failure to identify deforestation in a documented region may indicate algorithmic shortcomings.

Van den Burg & Williams (2020) demonstrated that expert annotations offer a validation method by comparing algorithm-detected changes to human-labeled events using metrics such as accuracy and F1-score. This be particularly useful for evaluating whether algorithms capture the transitions that humans interpret as significant.

In addition, Browning et al. (2017) applied proxy data to provide indirect confirmation of change. They demonstrated how a detected major shift in vegetation greenness can be cross-referenced with climate variables, fire severity indices, and changes in moisture levels.

Each validation method has limitations: synthetic data oversimplifies real-world complexities, qualitative assessments and expert annotations can be subjective, and proxy data may not establish causality. Consequently, Zhao et al. (2019) recommend combining multiple strategies to build a more holistic view of performance.

4 Implementation

To answer the research questions posed in this thesis, a framework capable of identifying significant shifts within bird sighting data is developed. This framework is designed to analyze spatio-temporal changes in bird populations, specifically targeting abrupt alterations in the trend trajectories of various bird species over the defined observation period and within specified spatial grids. This approach includes a preprocessing scheme to mitigate biases, Change Point Detection with BEAST, and result presentation via a user interface.

The chapter starts with a requirement analysis, outlining the functional and non-functional requirements. It then introduces the modular software design and details key modules: data acquisition, preprocessing, time series construction, Change Point Detection, post-processing, and result presentation.

4.1 Requirement Analysis

The following requirement analysis outlines the core objectives and constraints that the solution must satisfy to effectively detect spatio-temporal shifts in bird sighting data. These requirements ensure that the system is not only functionally robust—capable of accurate data preprocessing, Change Point Detection, and result presentation—but also adheres to stringent performance, reproducibility, and maintainability standards.

4.1.1 Functional Requirements

1. **Data Preprocessing:**
    - The system shall aggregate individual sighting records into defined spatial units and organize them into time series.
    - The system shall implement mechanisms to reduce known user-induced biases.
2. **Change Point Detection:**
    - The system shall perform statistical Change Point Detection on the aggregated time series to identify significant shifts in the temporal patterns of species observations.
    - The system shall provide interpretable outputs, including the identification of abrupt trend changes and corresponding confidence measures for each detected change.
3. **Result Presentation:**
    - The system shall offer a clear, interpretable User Interface for evaluation purposes.

4.1.2 Non-Functional Requirements

1. **Performance:**
    - The system shall handle millions of sighting records efficiently.
    - The system shall perform Change Point Detection on numerous species without prohibitive runtimes.
2. **Reproducability & Maintainability:**
    - The system shall ensure that data preprocessing, merging, and Change Point Detection are reproducible, documented, and transparent for future validation and extension.

4.2 Software Design

To address the above requirements effectively, a modular software architecture was chosen. This design facilitates a clear separation of the functional responsibilities into discrete modules. Each module can then be implemented, tested, and extended independently, which is particularly advantageous in notebook-based development environments. The system is structured into the following components:

1. **Data Acquisition**
2. **Data Preprocessing**
3. **Time Series Construction**
4. **User Bias Correction**
5. **Change Point Detection (BEAST Analysis)**
6. **Postprocessing & Result Consolidation**
7. **User Interface**

Figure 4.1 conceptually illustrates how these components interact, starting with raw data ingestion through to final result presentation.

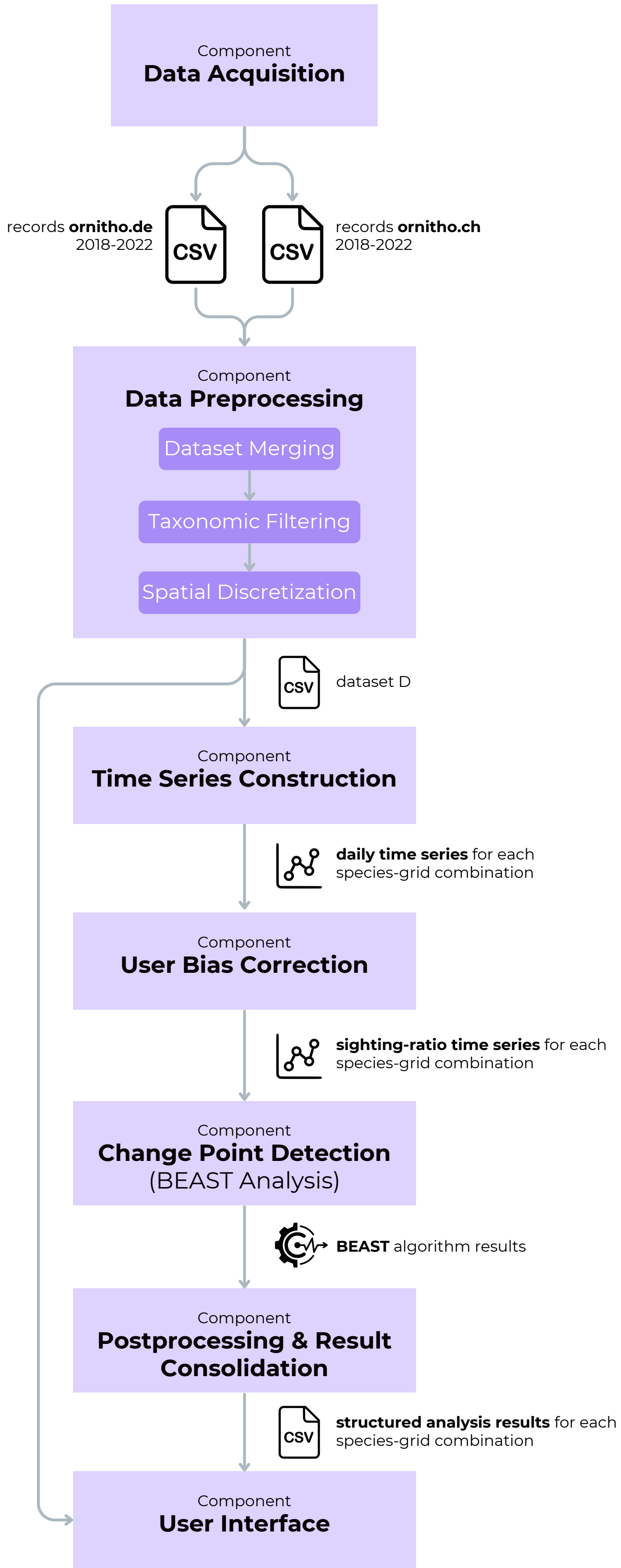


Figure 4.1: Component Diagram of the Implemented Architecture.

4.3 Component Descriptions

4.3.1 Data Acquisition

The core input data consist of bird sighting records collected via two national citizen-science platforms where users can log observations of various taxa, including birds, via websites or the mobile application *iNaturalist*. The Ornitho network operates across multiple countries; however, this study exclusively employs data from the platforms of Germany (`ornitho.de`) and Switzerland (`ornitho.ch`), available as dataframes. The time period is limited to 2018–2022.

Each record contains the following essential fields:

$$(\text{species name and ID}, \text{date}, \text{latitude}, \text{longitude}, \text{observer ID}, \text{optional metadata})$$

For the purposes of this study, only the date, location, and species are relevant for Change Point Detection; therefore, observer ID and optional features are not considered.

4.3.2 Data Preprocessing

Dataset Merging

While both datasets fundamentally have the same information and structure, they differ in some aspects. To address this, each dataset was preprocessed to standardize data structure, format, and terminology. The difference and resulting standardized format are detailed in the Table 4.1 below.

Feature	**German Data**	**Swiss Data**	**Merged Data**
Date Format	`dd.mm.yyyy`	`yyyy-mm-dd`	`yyyy-mm-dd`
Feature Names	lowercase	UPPERCASE	lowercase
Species IDs	German-specific IDs	International IDs	International IDs
Species Names	German	Swiss German	German

Table 4.1: Comparison of key feature differences between the German and Swiss bird sighting datasets and the resulting standardized format in the merged dataset.

Since the two datasets are gathered independently and cover disjoint geographical areas (i.e., no overlap in exact coordinates), the merged dataset D is their set union:

$$D = D_{\text{de}} \cup D_{\text{ch}}.$$

where D_{de} and D_{ch} are the german and swiss sightings, respectively. As each record is uniquely identified, there is no intersection. Hence, the datasets can be merged by simply concatenating their records without requiring deduplication or conflict resolution. Therefore, the total number of records in the merged dataset D is simply the sum of the records from both original datasets:

$$|D| = |D_{\text{de}}| + |D_{\text{ch}}|.$$

Taxonomic Filtering

The merged dataset comprises observations of 821 unique bird species, of which 497 and 708 species have been recorded in Switzerland and Germany, respectively. Collectively, the dataset encompasses approximately 50 million individual sightings, equating to an average of about 27,000 sightings per day. While this extensive dataset offers a wealth of information, the volume significantly exceeds the processing capacity of available resources.

Therefore, only a relevant species subset

$$S' \subseteq S$$

is chosen for deeper analysis, based on domain-expert recommendations. They identified 27 bird species as especially valuable for investigation based on their ecological significance, population trends, or other relevant criteria. Additionally, species essential for the quantitative analysis (see Chapter 5.2.2) were included based on ornithologists' evaluations indicating that they had undergone significant changes in recent years. In total, S' encompasses 184 out of 821 bird species available. They provide a focused yet representative sample for the initial evaluation of the methodology. All subsequent steps apply only to species $s \in S'$.

Spatial Discretization

For change point analysis, multiple sightings must be aggregated within a defined spatial area to construct a cohesive time series from a collection of individual sightings. To achieve this, each sighting is assigned to a standardized grid cell based on geographic location (ϕ, λ) (latitude ϕ and longitude λ). To discretize the sighting locations, observations are aggregated into standardized grid cells defined by the European Environment Agency (EEA)[1] in EPSG:4326, which references the WGS84 coordinate system. As a balance between grid size, species density per grid cell, and computational complexity, a 50x50 km grid was selected.

To perform grid assignment, let $\mathcal{G}$ be the set of all $50 \times 50\,\text{km}$ EEA grid cells covering Germany and Switzerland. Each grid cell

$$g \in \mathcal{G}$$

can be represented as a polygon $\text{poly}(g)$ in WGS84 coordinates. Figure 4.2 shows all selected grid cells of $\mathcal{G}$.

[1] https://www.eea.europa.eu/en. Accessed on 26th Oct 2024

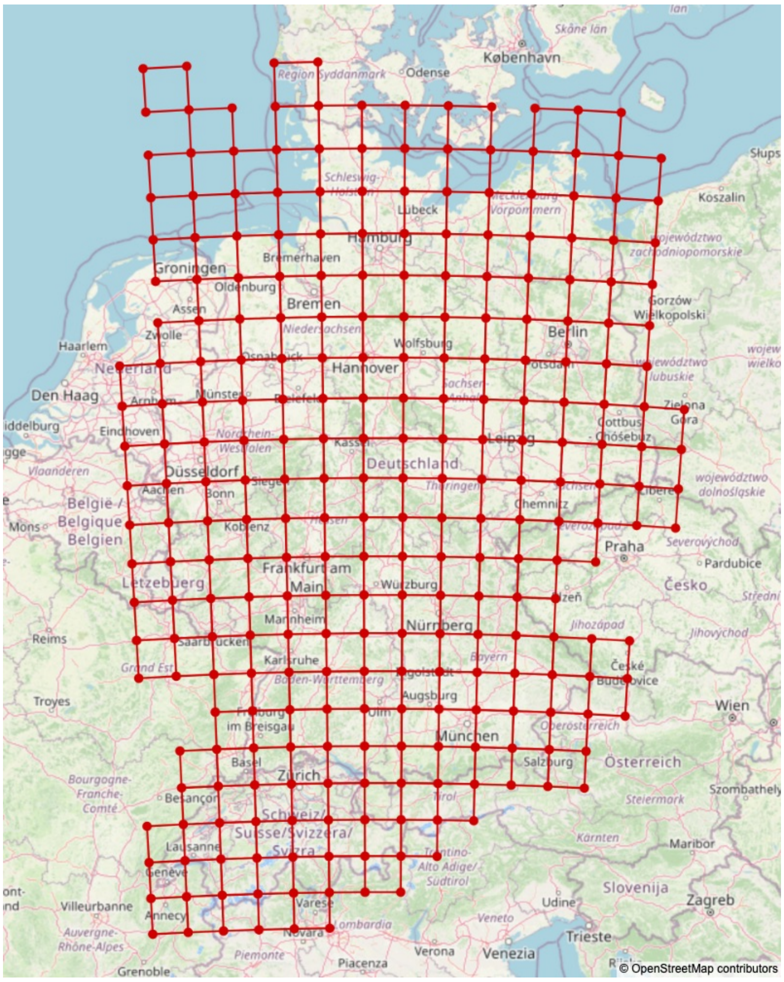


Figure 4.2: Standardized 50x50 km grid cells covering Germany and Switzerland, provided by the European Environment Agency (EEA). Each cell serves as a spatial unit for aggregating sightings per species, enabling the construction of species- and grid-specific time series for detecting change points.

Each sighting record $(s_i, t_i, \phi_i, \lambda_i)$ with species s_i, timestamp t_i, and location ϕ_i, λ_i is mapped to exactly one grid cell g_i for which:

$$(\phi_i, \lambda_i) \in \text{poly}(g_i).$$

Since each location belongs to exactly one $50 \times 50\,\text{km}$ cell, the assignment is unique.

4.3.3 Time Series Construction

The primary goal is to detect temporal changes in bird sightings. Hence, the merged and filtered records are transformed into time series defined per species–grid combination. To achieve this, the number of sightings is counted for each combination of species, grid cell, and date. Mathematically, the time series for each species-grid combination is defined as:

$$S(s, g, t) = \sum_{i} \mathbf{1}(s_i = s, g_i = g, t_i = t)$$

where:

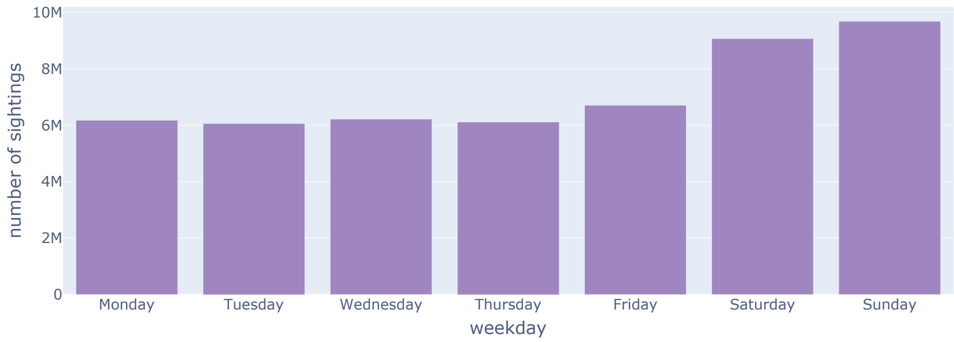
- $S(s, g, t)$ represents the number of sightings for species s in grid cell g on date t,
- i indexes individual sightings,
- s_i, g_i, t_i are the species, grid cell, and date of sighting i, respectively,
- $\mathbf{1}(\cdot)$ is the indicator function, which equals 1 if the condition inside holds (i.e., if the sighting matches the species, grid, and date), and 0 otherwise.

This results in continuous time series spanning the entire observation period for each species-grid combination, providing a temporal framework for tracking occurrences and detecting shifts in species presence over time.

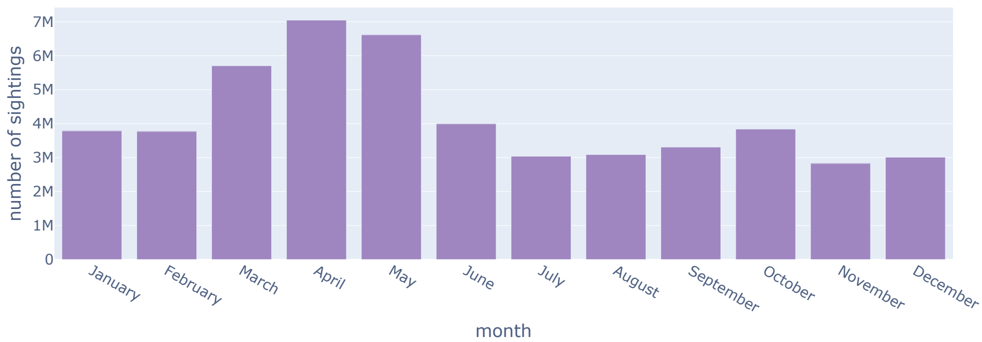
An example time series for the *Pygmy Owl* sightings in a grid cell near Basel, Switzerland, is shown in Figure 4.4a, illustrating this methodology for a single species within a defined spatial area.

4.3.4 User bias correction

The factors described in Chapter 2.1.4 highlight that the true number of individuals present cannot be accurately inferred from the number of daily sightings, as the data is influenced by user bias. While most of these biases cannot be easily quantified, temporal biases can be observed in the data itself. Figure 4.3a demonstrates that both Swiss and German users tend to report sightings predominantly on weekends, leading to an overrepresentation of sightings during these periods. Additionally, it is evident from Figure 4.3b that users primarily record bird sightings in spring.



(a) weekday distribution



(b) monthly distribution

Figure 4.3: Temporal distributions of submitted sightings in Germany and Switzerland

As detailed in Chapter 3.1, the SOPM abundance index is used to address user bias. While effective, adjustments were made for this thesis. First, data is aggregated weekly instead of using pentads to account for temporal bias, as Ornitho users are more active on weekends. Pentads would mix or omit weekends, causing inconsistencies. Additionally, instead of using the maximum count, a sighting ratio—the percentage of days within a week with at least one sighting—is applied. This prevents overestimation from isolated large counts and ensures Change Point Detection is driven by consistent patterns rather than outliers.

To achieve this, the overall study period (i.e., 2018–2022) is divided into weekly intervals $W_1, W_2, \ldots, W_N$. For a given species s and grid g, define

$$x_{s,g}(W_n) \;=\; \sum_{\tau \in W_n} \mathbf{1}\big(\text{at least one sighting of } s \text{ in } g \text{ on day } \tau\big),$$

where $\mathbf{1}(\cdot)$ is the indicator function. Typically, $|W_n| = 7$ days, so $x_{s,g}(W_n) \in \{0, 1, \ldots, 7\}$.

To mitigate user bias, a *sighting ratio* is defined:

$$r_{s,g}(W_n) \;=\; \frac{x_{s,g}(W_n)}{|W_n|}$$

where:

- $r_{s,g}(W_n)$ represents the fraction of days in week W_n on which species s was observed in grid g,
- $x_{s,g}(W_n)$ is the number of days within W_n where at least one sighting occurred,
- $|W_n|$ is the total number of days in the week (typically 7).

For a standard 7-day week, $r_{s,g}(W_n)$ is the fraction of days in W_n on which species s was observed in grid g. By definition, the sighting ratio is bounded within the interval:

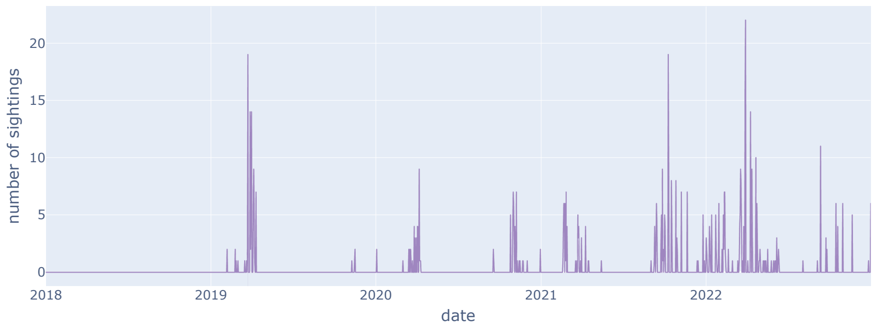
$$0 \leq r_{s,g}(W_n) \leq 1$$

with

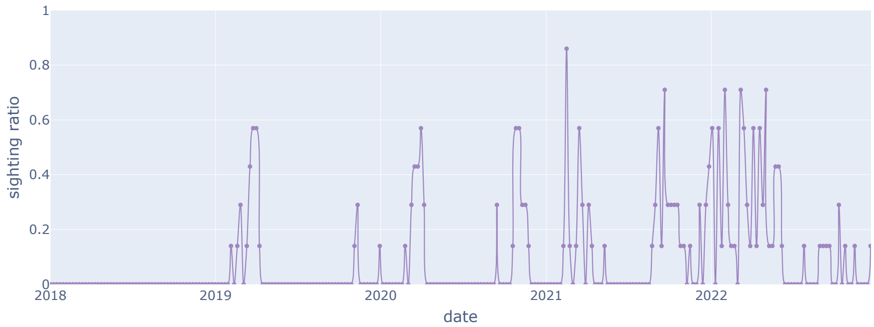
- $r_{s,g}(W_n) = 0$: no sightings in that week,
- $r_{s,g}(W_n) = 1$: sightings occurred every day of that week.

This ensures the sighting ratio remains a normalized metric, preventing overestimation from sporadic high counts while capturing consistent observation patterns. The impact of this procedure on the time series is illustrated in Figure 4.4, using an example time series of *Pygmy Owl* sightings. As shown, the time series after user bias reduction, depicted in Figure 4.4b exhibits reduced fluctuations and fewer outliers when compared to its raw state, as seen in Figure 4.4a. The value of each data point is consistently between 0 and 1, reflecting the proportion of days within each week where sightings were recorded.

Unlike the number of sightings, which does not allow for inferences about the actual abundance of the bird, the overall presence or absence and the regularity of sightings are now emphasized. However, a notable drawback of this approach is the significant reduction in the number of data points, which decreases by a factor of seven.



(a) Raw time series, depicting the number of daily sightings over the observation period from January 2018 to December 2022.



(b) Time series after user bias reduction, where data has been aggregated into weekly intervals and converted into sighting ratios (proportion of days per week where the bird was sighted).

Figure 4.4: Example time series of Pygmy Owl sightings in a 50x50 km grid near Basel, Switzerland before and after user bias reduction.

Summary

To detect and analyze temporal changes in bird sighting patterns across different regions, a total of 184 bird species were considered, with each species analyzed individually to capture species-specific trends and changes. The study area was divided into 50x50km spatial grids. For each species, all grids where the species had been observed were extracted, ensuring that the analysis was geographically comprehensive for that species.

The time series were constructed using data from January 1, 2018, to December 31, 2022. Weekly intervals were used to provide a consistent temporal resolution and smoothing out daily fluctuations. For each species-grid combination, a sighting ratio was calculated, defined as the ratio of days per week where the respective species was sighted at least once.

4.3.5 Change Point Detection (BEAST Analysis)

After preprocessing the data, aggregating sightings spatially into grid cells, and constructing weekly time series, a univariate weekly time series

$$\mathcal{T}_{s,g} = \{(W_1, r_{s,g}(W_1)), \ldots, (W_N, r_{s,g}(W_N))\}$$

can be obtained for each species–grid pair (s, g). For each time series, the Bayesian Estimation of Abrupt Change, Seasonality, and Trend (BEAST) (Zhao et al. 2019) algorithm is applied to detect statistically significant shifts. It is available as open-source in the corresponding package *Rbeast*[2] For mathematical details on the algorithm, please refer to Chapter 2.2.3.

[2] https://github.com/zhaokg/, Accessed on 9th Feb 2025.

Implementation Parameters

The following parameters were specified for the algorithm:

- **Time Parameters**:
  - Start date: January 1, 2018
  - End date: December 31, 2022
  - Delta t: 7 days (weekly intervals)
- **Priors**:
  - Seasonality: A period of one year is specified to account for annual seasonal effects in bird sightings. This is used as the lowest harmonic order, i.e., the fundamental frequency
  - Accepted range of number of changepoints: A minimum of 0 and a maximum of 10 change points for both the seasonal and trend components are allowed.
  - Accepted minimal space between neighbouring changepoints: A minimum spacing of one data point between neighboring change points is allowed for both seasonal and trend components.
- **Outlier Detection**: Outlier detection is enabled to allow the model to identify and account for anomalous data points.
- **Number of MCMC samples to collect**: A total of 8,000 MCMC samples are collected during the analysis. This number is recommended to balance computational efficiency with the need for sufficient samples to accurately estimate the posterior distributions of the model parameters (*Rbeast Documentation* 2019).
- **Reproducibility**: A fixed random seed is used to ensure consistent results across runs.

BEAST Results

For each time series $\mathcal{T}_{s,g}$, BEAST outputs:

- The posterior median of trend change points, $\hat{k}$.
- The estimated times $\{\hat{t}_1, \ldots, \hat{t}_{\hat{k}}\}$ of these change points.
- Posterior probabilities $P(\hat{t}_i)$ for each $\hat{t}_i$.
- Trend slopes m_t at each time point t in $\mathcal{T}_{s,g}$.

These results are stored in a dataframe for subsequent postprocessing.

4.3.6 Postprocessing & Result Consolidation

For each time series $\mathcal{T}_{s,g}$, the median of the estimated number of change points $\hat{k}$ is used to determine how many change points to consider. If $\hat{k} = 0$ or $\hat{k} = NaN$ (e.g., due to insufficient data), no trend changes are recorded.

Otherwise, each trend change point $\hat{t}_i$ is further labeled as:

$$\text{label}(\hat{t}_i) = \begin{cases} \text{positive}, & \text{if } P(\text{slope} > 0) > P(\text{slope} < 0), \\ \text{negative}, & \text{otherwise.} \end{cases}$$

Based on this classification, the number of positive trend change points, indicating sudden trend increases, is denoted as k_+, with estimated times $\{\hat{t}_1^+, \ldots, \hat{t}_{k_+}^+\}$, and the number of negative trend change points, indicating sudden trend decreases, is denoted as k_-, with estimated times $\{\hat{t}_1^-, \ldots, \hat{t}_{k_-}^-\}$.

These change points are associated with their posterior probabilities $P(\hat{t}_i^+)$ and $P(\hat{t}_i^-)$, indicating the confidence in each detected shift.

BEAST further provides the trend slope at each time point in the time series. Let m_t denote the estimated slope at time t. The mean trend $\bar{m}$ is then computed as the average of all individual slopes across the time series:

$$\bar{m} = \frac{1}{T} \sum_{t=1}^{T} m_t$$

where m_t is the trend slope at time t, and T is the total number of time points in the series. This metric provides an additional summary of the species' temporal development over the observation period.

All detected change points and relevant metadata are compiled into a unified dataset. For each species-grid combination (s, g), this dataset includes:

- The number of positive trend change points, k_+, their estimated times $\{\hat{t}_1^+, \ldots, \hat{t}_{k_+}^+\}$, and their posterior probabilities $P(\hat{t}_i^+)$.
- The number of negative trend change points, k_-, their estimated times $\{\hat{t}_1^-, \ldots, \hat{t}_{k_-}^-\}$, and their posterior probabilities $P(\hat{t}_i^-)$.
- The mean trend slope $\bar{m}$
- A data availability flag indicating whether the time series $\mathcal{T}_{s,g}$ contained sufficient data for analysis.

By consolidating these results into a comprehensible dataset, further analyses can be performed without re-running the entire pipeline. This was particularly important for the implemented User Interface (see Chapter 4.3.7) where rapid access to results is necessary.

4.3.7 User Interface

To facilitate the systematic exploration and analysis of species abundance, distribution, temporal dynamics, and the identification of significant shifts (i.e., change points), an interactive, web-based user interface was developed. This interface serves both as a tool for subsequent evaluations and a potential resource for ornithologists and researchers in related fields. It integrates data visualizations and statistical BEAST analysis, enabling the detection and examination of change points and trends in bird sightings.

This chapter delineates the key components of the user interface, detailing the functionalities and methodologies employed to analyze species data, filter change points, visualize spatial and temporal trends, and conduct BEAST analyses. The user interface was built with *gradio*[3] (Python). Figure 4.5 provides a screenshot of the entire application. The individual components are presented and described in detail below.

[3] https://www.gradio.app/. Accessed on 9th Feb 2025.

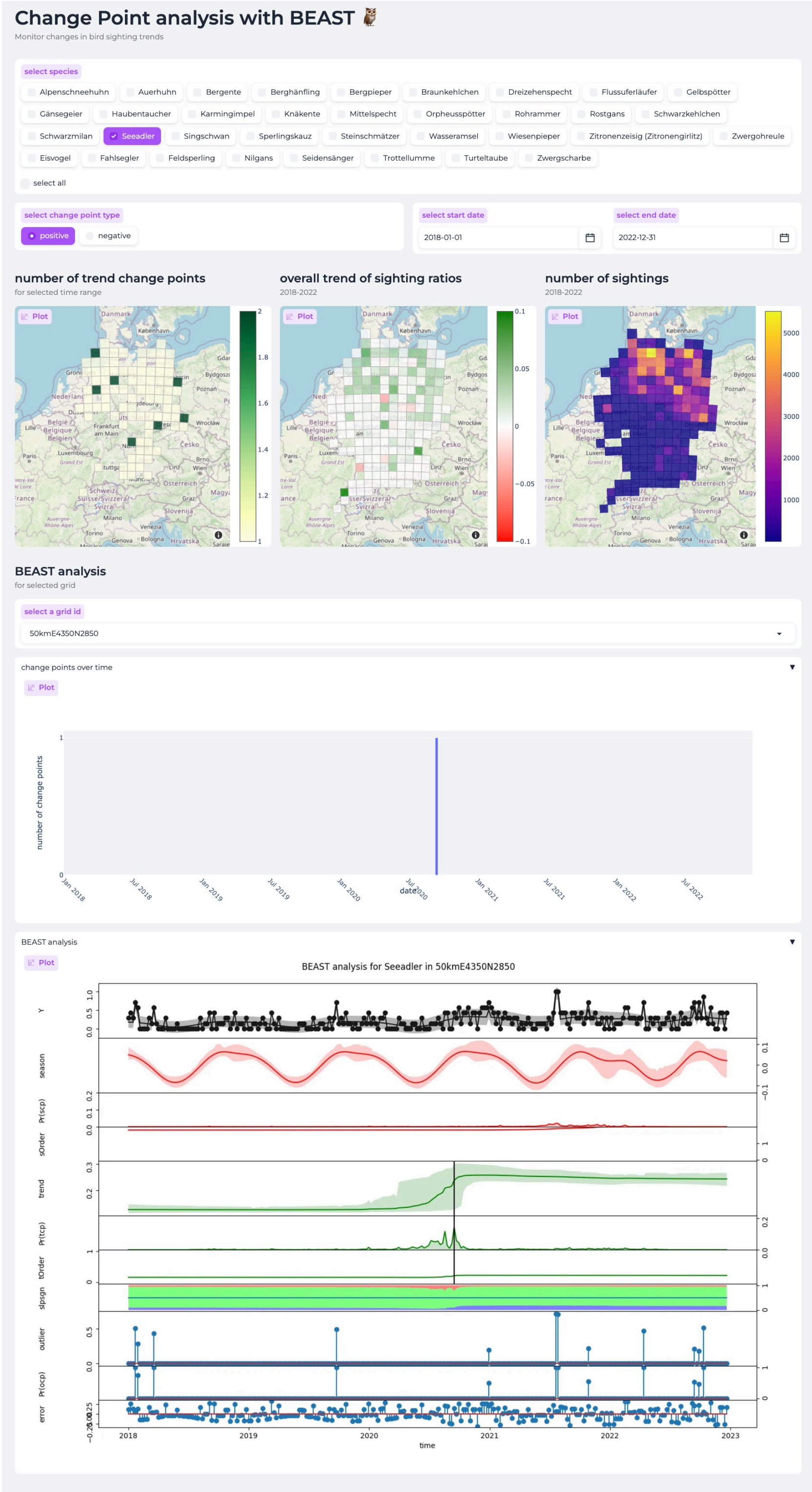


Figure 4.5: Overview of the user interface built with *gradio*. For a detailed view and description of each component, refer to the screenshots provided below.

Species selection

Figure 4.6 shows the Species Selection field. It allows users to select the species to be analyzed. Multiple selections are supported, enabling comparisons of change points and trends across different species or identifying regions with a high concentration of change points.

Figure 4.6: Species selection field. This component enables users to select one or multiple species for analysis.

Change Point filters

Below the species selection, change points can be filtered (see Figure 4.7). These filters directly impact the change point map, as described in Chapter 4.3.7.

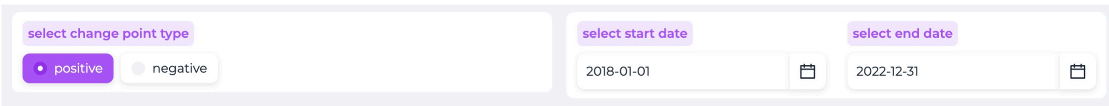


Figure 4.7: Change point filters field. This allows to select the type and date range of the change points to be shown in the change point heatmap.

Change Point Type With the first filter, users can specify whether to display positive or negative change points. This distinction allows for targeted analysis of trends, enabling researchers to focus on either growth or decline patterns.

Temporal Filtering Temporal filtering enables the restriction of change points to a specific time interval within the overall study period of 2018-2022. Users can adjust the timeframe with day-level granularity, allowing for precise temporal analysis. By default, the entire period from 2018 to 2022 is selected.

Map visualizations

This section encompasses all map-based visualizations, each presented as a heatmap with data points corresponding to individual grid cells (see Figure 4.8). These visualizations provide insights into the spatial distribution of change points, trends in sighting ratios, and the number of sightings across the entire study period.

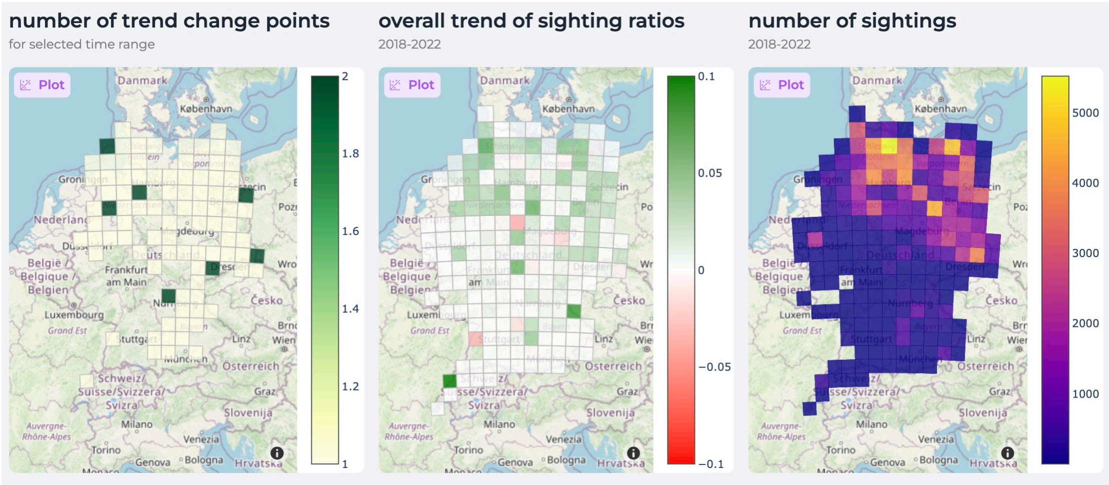


Figure 4.8: Map visualizations. This component shows the distribution of change points, trends, and number of sightings for the selected species. As an example, maps of the *White-tailed Eagle* are presented for the entire study period.

Number of change points This map shows the distribution of change points across Germany and Switzerland. Depending on the selection in the change point filters (see Section 4.3.7), either positive or negative change points are displayed for the chosen species. Each colored grid cell indicates the presence of change points within the selected timeframe, with color intensity representing the number of detected change points—the darker the color, the higher the count. Hovering over a tile reveals detailed information, including the grid ID, the number of change points, and the corresponding dates for each change point.

Overall trend of sighting ratios This map presents the average trend slope of sighting ratios over the entire period as a heatmap. Green tiles denote a positive trend, indicating an overall increase in sighting ratios, while red tiles signify a negative trend, representing a decrease. The intensity of the color reflects the magnitude of the trend—the darker the shade, the more pronounced the trend. White tiles indicate stability in sighting ratios for the respective species within that grid cell over the study period. The exact trend slope can be obtained by hovering over a tile.

Number of sightings The third map provides an overview of the absolute number of sightings for the selected species from 2018 to 2022. Unlike sighting ratios, this plot displays the actual count of sightings. Each grid cell indicates that the species was sighted at least once, with lighter colors representing a higher number of sightings. Detailed counts are accessible via the hover text on each tile.

BEAST analysis

The BEAST analysis component allows for a detailed examination of the time series and its change points for an individual grid cell. It consists of three subcomponents: grid selection, a change point timeline, and the BEAST results visualization.

Grid selection BEAST is designed to perform analyses on univariate time series; it can process data from only one grid cell at a time. Consequently, users must select a specific grid ID for analysis via the dropdown menu shown in the Figure 4.9 below. The selection pool includes all grid cells where the chosen species have been sighted at least once, ensuring that the analysis is relevant and data-driven.

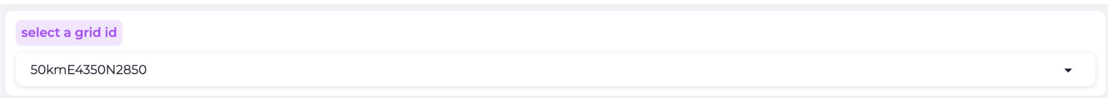


Figure 4.9: Grid selection field. This allows to select the grid ID for which the BEAST results are shown.

Change point timeline The change point timeline visualization displays the absolute number of change points over the study period. If a single species is selected, this allows for quick identification of when change points occurred. An example for this is shown in Figure 4.10, where a single changepoint is shown in Summer 2020.

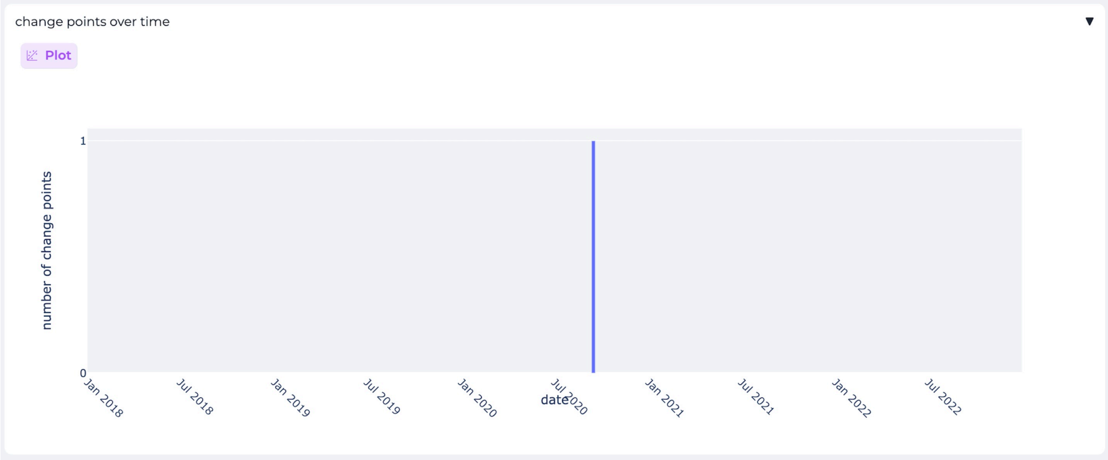


Figure 4.10: Changepoint timeline. This shows identified change points of the selected type over time.

However, this temporal representation is particularly valuable when multiple species are selected for simultaneous examination, as shown in Figure 4.11. It enables the identification of coinciding change points across different species within the same region and timeframe. Such patterns may suggest that the observed change points are not species-specific but rather attributable to local environmental factors, such as landscape modifications or weather variations. Identifying these non-species-specific change points can provide deeper insights into broader ecological or climatic influences affecting multiple species concurrently.

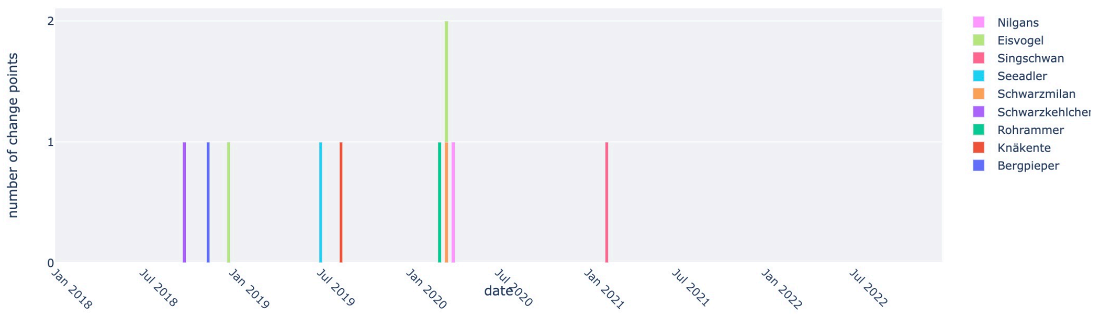


Figure 4.11: Change point timeline when multiple species are selected. Exemplarily, the timeline for a grid near Magdeburg, Germany is depicted.

Visualization of BEAST results In the last component of the interface, the BEAST analysis results for the specified species and grid are presented, providing detailed insights into the temporal dynamics of the species' sighting ratios. These results include the sighting ratios over time, the decomposition of this time series into seasonal and trend components, and the identification of change points within both components.

Figure 4.12 below shows an exemplary visual output of the BEAST algorithm. This output can be directly generated using the *Rbeast* package.

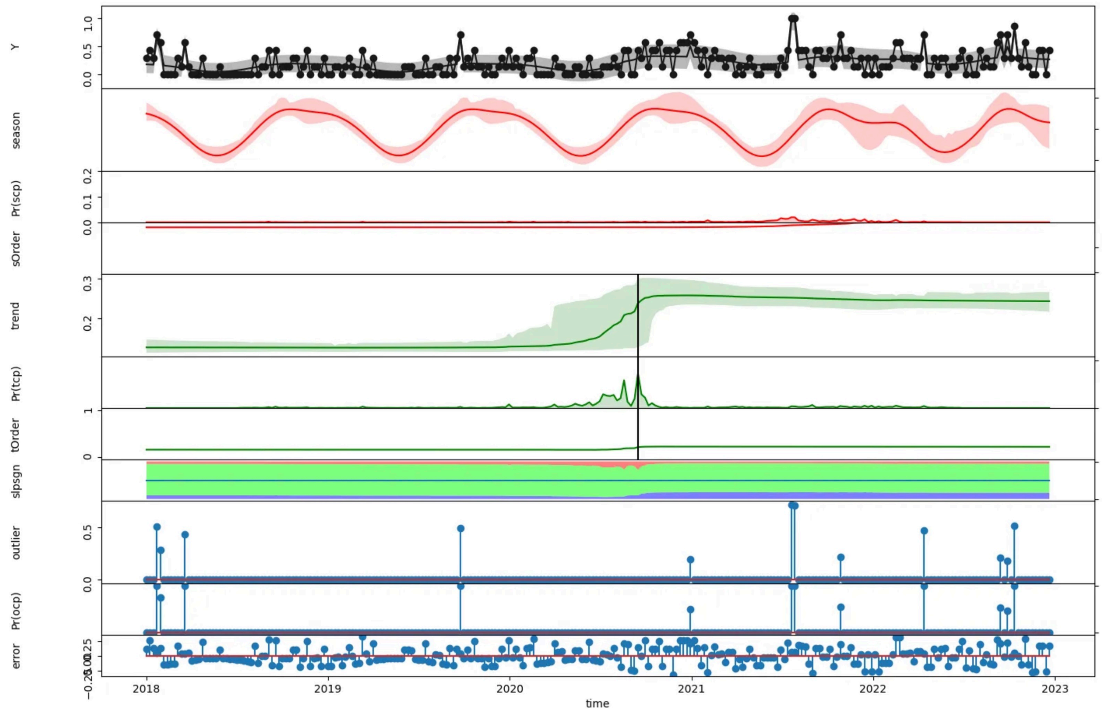


Figure 4.12: An exemplary output from BEAST. This analysis pertains to the *White-tailed Eagle* within a 50x50km grid near Ingolstadt, Germany.

At the top, depicted in black, is the input time series, where the sighting ratio is plotted over the observation period. In this instance, the time series represents the sighting ratios of *White-tailed Eagles* within a grid near Ingolstadt, Germany. The data cover the observation period from 2018 to 2022. Each entry in the time series indicates the ratio of days per week during which at least one eagle sighting occurred.

In red below is the seasonal component identified by BEAST, along with the estimated probability of an existing seasonal change point, denoted as $Pr(scp)$. The seasonal component demonstrates regular fluctuation between summer and winter months, accompanied by a credible interval surrounding the assumed seasonal curve.

In green is the estimated trend component, as well as the estimated probability of an existing trend change point, denoted as $Pr(tcp)$. A notable increase is observed in autumn 2020. At the peak of the probability curve, which results from the averaging of multiple models through Bayesian Model Averaging (BMA), BEAST identifies a change point, marked by a vertical black line. Since the trend does not exhibit an abrupt increase but rather extends over several months, BEAST indicates that the precise location of the change point may not be exact but is likely probabilistically distributed within the range of the increase.

Beneath these main components, additional information is provided regarding the slope sign, outlier analysis, and the residual or error component, which comprises data points that BEAST could neither assign to the seasonal nor to the trend component and did not classify as outliers. The slope sign is visually represented as $slpsgn$. In this panel, the upper red portion indicates the probability of a positive trend slope, the middle green portion the probability of a zero slope, and the lower blue portion the probability of a negative trend slope (*Rbeast Documentation* 2019). The outlier analysis highlights data points that BEAST has detected as outliers.

4.3.8 Summary

In summary, the software framework follows a modular design that separates data acquisition, aggregation, and preprocessing from the core BEAST analysis. Postprocessing and an interactive user interface complete the workflow, enabling transparent and reproducible detection of spatio-temporal shifts in citizen-science bird sighting data. The chosen approach satisfies both functional and non-functional requirements, offering scalability, maintainability, and user-friendliness.

5 Evaluation

The BEAST algorithm has been previously evaluated on diverse datasets by its developers and independent researchers. These studies have demonstrated satisfactory accuracy and have shown that BEAST outperforms similar algorithms employing best-model approaches (Li et al. 2022).

However, two gaps in the application of change point detection methods to citizen science data were identified (see Chapter 1). First, no dedicated study currently investigates the performance of CPD algorithm—such as BEAST—on citizen science data. This gap is significant because citizen science data may contain inconsistencies or biases, which could affect the algorithm's ability to capture genuine real-world phenomena. The second challenge is the insufficient involvement of domain experts in most existing studies. Consequently, the perceived trustworthiness and practical value of BEAST from a domain-expert standpoint have not been thoroughly explored, despite being critical factors in determining whether the algorithm will be adopted and relied upon in real-world applications.

To address these challenges, this chapter evaluates the BEAST algorithm using two complementary approaches. First, qualitative interviews with ornithological experts capture in-depth, domain-specific perspectives on algorithm outputs, perceived trustworthiness, and usability. Second, a quantitative comparison of detected change points with archival records from the Swiss Ornithological Institute assesses how closely BEAST aligns with documented real-world changes.

The remainder of this chapter presents a detailed overview of the methodology for both the qualitative and quantitative evaluations. For overview purposes, the evaluation methodology is visualized at the end of the chapter.

5.1 Evaluation Design

To investigate the five Research Questions posed in Chapter 1, this study employs two complementary strategies:

- **Qualitative Expert Interviews:** Semi-structured interviews with ornithological experts uncover in-depth, context-rich feedback on BEAST's outputs, revealing how well the algorithm's detected changes align with real-world ecological shifts and whether experts find the method trustworthy and useful (RQ1, RQ2, RQ3, RQ4). The interviews also yield broader insights on best practices for CPD in citizen science (RQ5), helping to generalize the outcomes beyond the specific use case at hand.
- **Quantitative Archival Comparison:** Documented changes in bird populations from the Swiss Ornithological Institute serve as a partial ground truth. By matching BEAST-detected change points to these archival records, detection rates across different types of events are measured, providing a systematic, data-driven perspective on algorithm performance (RQ1, RQ2). These findings also inform best practices for CPD in citizen science (RQ5).

5.2 Methodology

This chapter details the methodology for each evaluation component, emphasizing how each contributes to the Research Questions. As an overview, Figure 5.1 summarizes the Evaluation procedure.

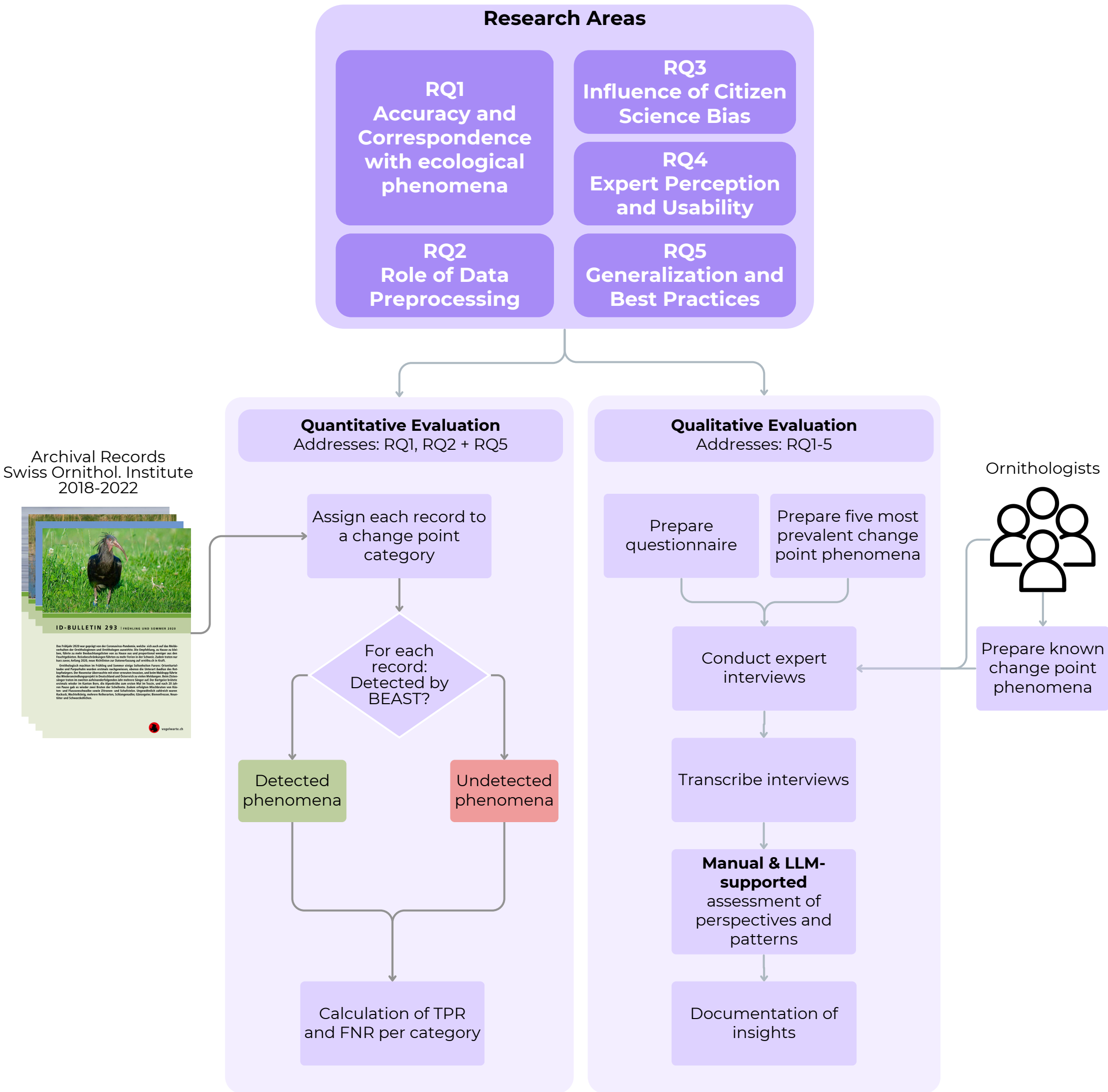


Figure 5.1: Evaluation Methodology overview. The evaluation assesses BEAST's accuracy, ecological validity, and practical applicability by combining a quantitative approach to measure performance metrics (left) with a qualitative expert feedback (right) to ensure both data-driven and domain-specific relevance.

5.2.1 Qualitative Evaluation

This section describes the design of the semi-structured interviews conducted with four ornithologists of varying statistical and ecological expertise. The primary purpose is to gather domain-specific insights related to RQ1, RQ2, RQ3, and RQ4, while also informing broader methodological considerations (RQ5). The methodology is visualized in Figure 5.1.

Interview Design and Procedures

A semi-structured interview format was chosen due to its flexibility and depth of inquiry, aligning with guidelines from Kallio et al. (2016), Castillo-Montoya (2016), and Turner (2010). An initial set of open-ended questions was developed to guide discussions around:

1. **Algorithm Accuracy:** Experts were shown 10 BEAST outputs (plots indicating change points and trends) for a range of bird sighting time series. They were prompted to evaluate the perceived correctness of the detected change points, the adequacy of the extracted trends, and the overall placement of these change points within the time series data.
2. **Correspondence to Real-World Phenomena:** Experts were presented with five change points phenomena detected by BEAST, and asked to evaluate whether the detected change points align with known real-world events (detailed below). Further, experts were asked to prepare known real-world phenomena, and results produced by BEAST were compared against this ground truth. The questionnaire seeked to identify reasons for any discrepancies between the algorithm's detections and actual ecological changes, such as algorithm limitations, data quality issues, or user biases. Lastly, experts were prompted to quantify the perceived trustworthiness of the algorithm to detect real-world phenomena.
3. **Utility, Usability, and possible application fields of BEAST:** The questionnaire gathered feedback on the clarity and comprehensibility of the algorithm's outputs, evaluated the perceived benefits it brings to their work, and seeked to identify areas for improvement.

Following best practices in semi-structured interviewing (Turner 2010, Castillo-Montoya 2016), the questionnaire was pilot-tested with a colleague to refine clarity and flow. The final protocol include both predefined open-ended questions and space for probing emergent topics. The questionnaire is provided in the Appendix.

To identify the most relevant change points for part (2.) of the interviews, each detected change point $\hat{k}_{i,s,g}$ with timestamp $\hat{t}_{i,s,g}$ for species s in spatial grid g was assigned to its respective species-grid combination (s, g). These change points were then grouped into clusters based on temporal proximity, such that any two change points timestamps $\hat{t}_{i,s,g}$ and $\hat{t}_{j,s,g}$ belong to the same cluster C_k if:

$$|\hat{t}_{i,s,g} - \hat{t}_{j,s,g}| \leq 30 \text{ days}$$

for all $\hat{t}_{i,s,g}, \hat{t}_{j,s,g} \in C_k$.

Each resulting cluster C_k thus represents a cohesive spatio-temporal event rather than isolated outliers, ensuring that change points occurring within a 30-day window are treated as a single event rather than independent detections.

Two species, the *Kingfisher* and the *European Stonechat*, collectively accounted for approximately 21% of C_k (13% and 8%, respectively). To avoid skewing the assessment toward these species, only one representative cluster per species was selected, prioritizing the largest and most spatially widespread. Additionally, a single phenomenon indicating a synchronous decrease in sightings during summer 2021 was formed by grouping phenomena observed in multiple species. A distinct example involving the *Pygmy Cormorant* was deliberately included to illustrate a striking transition from zero to numerous sightings, despite it being less geographically widespread. Figure 5.2 summarizes the phenomena chosen for expert review.

Figure 5.2: Overview of the five most prevalent change point phenomena for expert review. Grids refer to 50×50 km cells, and clusters represent change points within a 30-day interval. The map shows the distribution of change points.

Data Collection and Analysis

Each of the four interviews lasted approximately 90 minutes and was audio-recorded with participant consent. The recordings were transcribed automatically using *f4*[1], then manually verified for accuracy. Transcripts were thematically analyzed, comparing recurring insights across participants. To enhance consistency and uncover patterns, a Large Language Model (*ChatGPT 4o*) was employed as a supportive tool for grouping statements with shared themes and highlighting differences. This application of *ChatGPT* followed the guidelines laid out by Zhang et al. (2023), ensuring transparent and structured prompts, traceability of model outputs, and thorough manual cross-checking to avoid hallucinations or misinterpretations.

Mapping to the Research Questions

The qualitative analysis is designed to provide insights into all proposed research questions in the following ways:

- **RQ1 (Accuracy & Correspondence)**: Experts' perceptions of BEAST's outputs and alignment with known ecological phenomena uncover the algorithm's accuracy in real-world contexts.
- **RQ2 (Data Preprocessing)**: Discussions on how the proposed preprocessing scheme, grid size, and outlier handling influenced detected change points illuminate the role of preprocessing.
- **RQ3 (Citizen Science Bias)**: Experts' reflections on volunteer observer behavior provide evidence of how BEAST might be influenced by user-induced biases.
- **RQ4 (Expert Perception & Usability)**: The interview data capture whether experts found BEAST's outputs and results trustworthy, comprehensible, and valuable for their work, and informed possible areas for integration and improvement.
- **RQ5 (Generalization & Best Practices)**: Emerging strengths and limitations contribute to broader guidelines for CPD in citizen science settings.

[1] https://www.audiotranskription.de/f4/automatische-transkription/. Accessed 30th Jan 2025

5.2.2 Quantitative Analysis Using Archival Records

The second part of this evaluation quantifies how well BEAST detects documented real-world changes in bird populations, responding especially to RQ1, RQ2, and RQ5. The methodology is visualized in Figure 5.1.

Data Collection

A ground truth was assembled from the *ID-Bulletin* journals curated by the Swiss Ornithological Institute. These records, referred to as archival data, included 748 documented changes in bird species distribution and abundance from January 1, 2018, through December 31, 2022. These records are sourced from the same data as was used for BEAST analysis. However, the archives themselves employed a different preprocessing workflow (e.g., calculating occurrence indices while correcting for general increases in reporting effort).

Data Categorization

To determine the strengths and weaknesses of the BEAST algorithm in different types of changes, the records were divided into categories. These categories were formed by feeding all archival records into a Large Language Model (*ChatGPT o1*), which extracted patterns by clustering similar descriptions of change events based on contextual similarities. The following six categories were identified: (1) *rare observations and individual cases*, (2) *population dynamics*, (3) *discrete day events*, (4) *phenological shifts*, (5) *spatial expansion and decline*, and (6) *invasions and influxes*. Additionally, population dynamics events were further subdivided into (1) *real change points*, denoting clear, sudden shifts in population trends, and (2) *trends* to account for more gradual changes or ambiguous records. This subdivision aimed to clarify whether BEAST's results aligned better with distinct population breakpoints rather than long-term trend shifts.

Comparison Procedure

In line with how BEAST was configured (e.g., ignoring outliers and change points in the seasonal component), changes classified as *rare observations and individual cases*, *discrete day events*, and *phenological shifts* were deemed out of scope for detection under current parameters. For the remaining three categories—*population dynamics*, *spatial expansion and decline*, and *invasions and influxes*—an evaluation was conducted by comparing the archival record for each documented change with the algorithm's detected change point(s), in the relevant time period and spatial grid(s).

Limitations In many archival records, change events were documented only at the semiannual level and often lacked precise geographical information at the 50×50 km grid resolution. This uncertainty required a broader matching strategy, whereby a detection was considered valid if at least two grids in the relevant six-month window showed a change. Consequently, the evaluation results should be interpreted with an awareness of both temporal and spatial imprecision in the ground truth.

Mapping to the Research Questions

Although this quantitative analysis is most directly aligned with RQ1 (accuracy and ecological correspondence), it also informs RQ2 and RQ5:

- **RQ1 (Accuracy & Correspondence)**: By matching BEAST detections against documented changes in the archival records, it is quantified how often BEAST identifies true ecological shifts. True Positive Rates (TPR) and False Negative Rates (FNR) indicate whether the algorithm is capturing relevant phenomena or missing key events.
- **RQ2 (Data Preprocessing)**: The evaluation highlights how weekly presence data and 50×50km grids affect detection capabilities. If certain archival changes—especially local or short-term events—are consistently missed, it may underscore limitations in the current preprocessing strategy.
- **RQ5 (Generalization & Best Practices)**: Identifying systematic mismatches between BEAST and the archives helps pinpoint limitations and guide recommendations for adapting CPD methods to citizen science data.

5.3 Summary

The evaluation strategy in this thesis draws on two complementary methods—expert interviews and archival data comparisons—to address five central Research Questions. By combining qualitative insights from domain experts with a quantitative performance analysis, a comprehensive view of BEAST's capabilities and limitations in citizen science contexts are realized.

6 Results

This chapter presents the outcomes of both the qualitative and quantitative evaluations of the BEAST algorithm. The first part details the outcomes of the expert interviews conducted with ornithologists. The second part offers a quantitative comparison of BEAST's detections against documented archival records of bird population changes, exploring its performance across various categories such as population dynamics, spatial expansions, and invasions.

6.1 Qualitative Study with Ornithological Experts

This chapter presents the results of the qualitative evaluation conducted with ornithological experts, focusing on the accuracy, reliability, and usability of the BEAST algorithm for detecting change points in bird sighting time series data. The results are structured according to the main discussion points (1) algorithm accuracy, (2) correspondence to real-world phenomena, and (3) perceived usability, utility, and possible applications fields.

6.1.1 Accuracy of the BEAST Algorithm in Identifying Change Points

Overall, the interviewed ornithologists expressed that BEAST generally performs well in detecting change points that are visually evident in the data. In instances where a bird population trend changed abruptly or in a sustained fashion, BEAST reliably flagged these points. In the 10 examples presented, there were no instances where an expert indicated a desire for BEAST to have detected a change point where it had not. Hence, all interviewees rated their confidence in BEAST's ability to detect a visually apparent change point in the data as 5 out of 5.

Despite this generally positive assessment, all experts recognized a tendency of BEAST to produce False Positives. In other words, BEAST sometimes flagged minor, subtle changes as significant change points. The interviews revealed that while BEAST is sensitive and rarely misses a clearly visible abrupt change, it can also be too sensitive in some cases, setting change points at relatively subtle or inconsequential shifts. Consequently, all interviewees rated their confidence that a change point detected by BEAST would align with a change point they would identify themselves as 3 out of 5, indicating a moderate level of trust.

6.1.2 Correspondence of Identified Change Points with Real-World Changes

Phenomena prepared by interviewer

When experts were asked about concrete phenomena and whether the change points identified by BEAST aligned with known real-world changes, the results were generally promising. To confirm or reject the detected phenomena, the ornithologists used various tools, such as biannual expert reports or direct examination of sighting data through their current visualization tools. Five main phenomena were discussed (visualized in Figure 5.2):

Kingfisher: Crash in Winter 2021 Three out of four experts confirmed that the winter 2020/21 event was a well-known real-world occurrence. A severe winter with frozen waterways caused substantial *Kingfisher* mortality. Consequently, their numbers dropped abruptly, which was detected by BEAST. While BEAST detected this phenomenon across numerous grids throughout Germany and Switzerland, it was primarily concentrated in northern regions. The interviewees indicated that this geographic distribution of the identified changes aligns with observed real-world patterns. Further, they praised BEAST for capturing not only the sharp decline but also, in some instances, the subsequent recovery period.

Stonechat: Increase in Spring 2020 For the *Stonechat*, BEAST identified a positive shift in spring 2020. Experts confirmed this event and attributed it to mild winter conditions leading to higher survival and possibly earlier breeding. Although they found the initial increase plausible and correlating with warmer weather, subsequent declines or plateaus were not always as clearly explained. Nonetheless, experts acknowledged the initial positive trend as realistic.

Meadow Pipit: Increase in Autumn/Winter 2020 BEAST detected a rise in sightings in late 2020. Experts explained that unusually mild conditions delayed migration or caused an extended stay of these birds, resulting in higher sighting probabilities. This phenomenon was seen as well-founded and reflective of actual ecological conditions. Thus, the alignment with a real-world event was again positively confirmed.

Pygmy Cormorant: Influx in 2021 There was a noted influx of *Pygmy Cormorants* in 2021, which significantly changed their status from a rare vagrant to a more regularly encountered species. Experts confirmed that BEAST's detection of these changes closely matched a known influx and establishment event, which was a well-documented shift in the species' presence.

Multiple Species (e.g., Tree Sparrow, Egyptian Goose, Great Crested Grebe): Decrease in Summer 2021 BEAST identified concurrent declines for multiple species in summer 2021. For the *Tree Sparrow* in particular, an interviewee confirmed this change point; there is an ongoing mystery about steep declines potentially linked to agricultural chemicals, which BEAST's shown trend and detected change points highlighted. However, experts struggled to pinpoint a common ecological factor that explains why multiple species experienced a simultaneous shift during summer 2021. They proposed several theories: a particularly rainy summer may have reduced detectability due to less field activity by observers, lowered breeding success due to unfavorable weather conditions, or other external factors like changes in observer behavior post-COVID lockdowns.

Overall, based on the shown examples, all experts rated their trust that a phenomenon detected by BEAST aligns with real-world events as 4 out of 5. While this suggests a high trust level, it remains below their trust in domain expert assessments.

Phenomena prepared by interviewees

To evaluate whether BEAST identified known change points, two of four experts presented several known changes, such as unusual mortality events of the *Common Murre*, *Eurasian Woodcock* influxes, or an increase in population of the *Whooper Swan* at Lake

Constance. In most cases, these events were not clearly detected by BEAST. Experts concluded that the reason often lay in the chosen data preprocessing steps. In the given examples, the aggregated weekly presence data did not produce a strong enough signal for BEAST to detect a clear change point. Further, using presence in a week rather than absolute counts or maximum counts per time unit made it harder for BEAST to detect abundance-driven events. Moreover, the chosen grid size of 50x50km was found to be too broad to detect some smaller-scaled changes. In such scenarios, the lack of detection does not mean no change occurred; it simply was not apparent in the processed data available to BEAST.

Additionally, for species that appear sporadically or for singular events (like rare migrants), BEAST tended to mark these occurrences as outliers rather than genuine change points. The experts noted that this can be appropriate methodologically but also means that known short-term or sporadic changes may go unnoticed as change points. Hence, genuine changes that do not produce a strong or sustained alteration in the weekly presence ratio may remain undetected.

When asked whether they would recommend an alternative preprocessing method or a smaller grid size, the experts noted that it would be challenging to define a preprocessing strategy or grid size that ensures BEAST can detect all possible change points. This difficulty arises from the inherent variability and diverse characteristics of change points.

The experts indicated that the chosen presence-focused methodology is particularly suitable for detecting large-scale changes and general spatial expansions or declines of a species. A method focused on abundance instead of presence could reveal change points associated with population size fluctuations but might fail to detect broader spatial patterns. Similarly, reducing the grid size could capture more localized changes while potentially overlooking larger-scale dynamics.

Given these trade-offs, the experts expressed confidence that BEAST can detect various types of changes but acknowledged that its ability is inherently limited to the characteristics of the data it processes—thus constrained by the chosen preprocessing method. Therefore, they rated their confidence that BEAST would detect any real-world change as moderate (ranging from 2 to 4 out of 5). Instead of pursuing a one-size-fits-all approach, one expert suggested applying BEAST across multiple preprocessing methods and spatial scales to capture a wider range of change types.

6.1.3 Usefulness and Possible Application Fields

Regarding BEAST's usefulness and user-friendliness, experts agreed that the tool provides a valuable addition to their analytical methods. After being introduced to the plots, the experts found the visual outputs, underlying logic, and interpretation process easy to follow and relatively clear. Some suggested small improvements, such as increasing the size of the input data plot for better visibility, refining axis labeling for greater intuitiveness, and using terminology other than "seasonality" to prevent biological misunderstandings. Others mentioned that normalizing or refining the scale could help highlight more subtle changes.

When asked about fields in which they see value in integrating BEAST, the experts highlighted two main areas. First, they noted that BEAST could support retraining machine learning models in an adaptive manner, initiating updates when the data indicate actual shifts in species presence rather than relying on fixed schedules. One expert suggested that, for this specific use case, it could be beneficial to delay retraining after a detected change to avoid premature retraining triggered by very short-lived events.

Secondly, they emphasized that BEAST could offer valuable insights into when and where a species undergoes a drastic change, enabling experts to recognize shifts they might otherwise miss. Currently, identifying sudden changes often depends on retrospective annual or biannual data checks, personal expertise, networking, or pre-formed expectations based on external events (e.g., harsh winter periods). Experts noted that while they have exceptionally high trust in these outcomes, two limitations remain: First, feedback is delayed as reports are only published biannually; incorporating BEAST could introduce a more proactive, data-driven approach, where shifts are flagged as they occur rather than being detected long after the fact. Second, relying primarily on networking and focusing on species where changes are anticipated may lead to confirmation bias, potentially overlooking more subtle or unexpected shifts; BEAST, which systematically analyzes all species and locations, could detect such changes.

Further, one experts considered BEAST's capacity to operate as an automated alert system a key advantage. Instead of relying on human observers to detect anomalies over months or years, BEAST could rapidly highlight unexpected changes, allowing ornithologists to respond more promptly.

While experts note that they do not expect BEAST to detect every change point—particularly those not well-represented by the data's preprocessing methods—they view it as an addition for their current range of analytical tools. Rather than replacing existing workflows or human expertise, BEAST could serve as a complementary analytical tool, highlighting detected shifts that domain experts can further evaluate, confirm, or reject.

6.1.4 Summary

In summary, experts believe that BEAST can deliver valuable insights, revealing previously unnoticed trends and allowing more real-time responses to shifts in bird populations. This potential for more immediate and data-driven monitoring stands in contrast to current, more reactive methods. While they highlighted that BEAST may not capture every type of change, and they would always recheck the correctness of a detected change point with a human expert, its capacity to uncover subtle or unexpected shifts underscored their perceived usefulness to the field. The interviewees believe that with thoughtful adjustments to data preprocessing (e.g., exploring appropriate spatial and temporal resolutions), BEAST could become an even more powerful instrument for ornithologists.

6.2 Quantitative Analysis Using Archival Records

This chapter presents the results of the quantitative evaluation conducted by comparing BEAST results with archival records, as described in the Evaluation methodology in chapter 5.2.2.

Out of the total 748 documented changes, 409 were classified as *rare observations and individual cases*, 33 were *discrete day events*, and 27 involved *phenological shifts*—categories that, by design, BEAST was not set up to detect under the current parameters, as highlighted in 5.2.2. As shown in Figure 6.1, these categories account for 62.7% of the observed changes. In contrast, the changes corresponding to the categories *population dynamics*, *spatial expansions and declines*, and *invasions and influxes*—which are expected to be captured by BEAST—comprise 37.3% of the total recorded changes. Here, the absolute changes amount to 238, 21, and 20 listed changes, respectively. Within the context of *population dynamics*, further subdivision revealed an identifiable group of 40 records that explicitly correspond to abrupt or significant population shifts, designated as *real change points*. Consequently, these represent 16.8% of the total *population dynamics* subset.

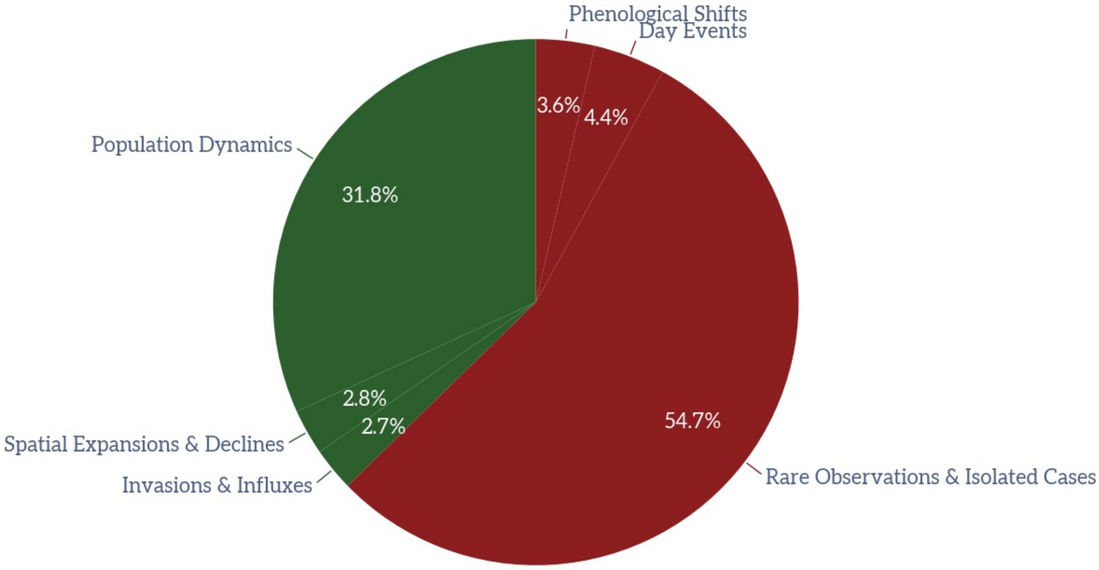


Figure 6.1: Distribution of change points in the archival records, by category. Categories highlighted in red denote those for which the BEAST algorithm is not configured to detect, while those highlighted in green denote categories for which change points are expected to be detected by BEAST.

Table 6.1 summarizes the percentages of archival changes that were recognized by BEAST for each category where the algorithm is expected to detect changes.

Category	**Detected (%)** (TPR)	**Not Detected (%)** (FNR)
Population Dynamics (all)	47.48	52.52
Population Dynamics (real change points)	62.86	37.14
Invasions and Influxes	50.00	50.00
Spatial Expansions and Declines	66.67	33.33

Table 6.1: Performance of the BEAST algorithm in detecting change points across categories where it is expected to be effective. Percentages represent the proportion of archival changes successfully identified (True Positive Rate, TPR) and those missed (False Negative Rate, FNR) for each category.

For *population dynamics*, BEAST identified 47.48% of changes, whereas for the *real change point* subset, the detection rate increased to 62.86%. In the *invasions and influxes* category, half of the documented events were captured by the algorithm. BEAST performed best with *spatial expansions and declines*, detecting two-thirds of the changes reported by the archives in that category.

7 Discussion

This chapter discusses the implications of the evaluation results for each of the five Research Questions (RQ1–RQ5). Drawing on both the qualitative interviews with ornithologists and the quantitative comparison with archival records, the discussion combines two distince vantage points and reflects on the strengths and limitations of BEAST in detecting real-world phenomena in citizen science data. Further, it identifies methodological considerations and best practices for applying Change Point Detection in citizen science contexts.

RQ1: Accuracy and Correspondence with Ecological Phenomena

The first question (RQ1) concerns how accurately the BEAST algorithm detects significant change points in citizen science time series data and whether these detected change points align with known real-world phenomena.

When evaluated solely based on the provided time series during the interviews, ornithological experts gave high marks to BEAST's ability to detect abrupt shifts in bird populations that are visually evident in the data. Experts never felt that BEAST missed an obvious change. However, the interviewees also identified False Positives. These observations align with the findings of Li et al. (2022). This emphasizes an inherent trade-off: while high sensitivity helps minimize the risk of missing genuine changes, it can also lead to extraneous detections that require domain experts to distinguish truly significant shifts from background noise.

However, examining whether BEAST detections correspond to actual ecological phenomena goes beyond straightforward time series analysis. Both qualitative and quantitative evaluations offer useful perspectives.

From a qualitative standpoint, ornithologists demonstrated high confidence in the ecological validity of identified phenomena: out of five phenomena presented, four were confirmed as genuine. While they do not expect BEAST to detect every true change, they do anticipate the tool to be most effective at uncovering spatial expansions, declines, and pronounced influxes.

These observations align with the quantitative results. Although a 67% detection rate indicates room for improvement, spatial expansions and declines produced the highest detection rates—presumably because the shift from near-zero to consistent presence (or vice versa) is relatively pronounced. Similarly, invasions and influxes were detected reliably when the species remained present over multiple weekly intervals. However, short-term influxes lasting only a few days were generally classified as outliers rather than true changes, thereby reducing the detection rate to only 50%. More gradual population dynamics or short-lived influxes fell below the detection threshold in many cases. Conversely, the subset of population dynamics labeled as real change points exhibited more abrupt changes, aligning better with how BEAST detects breakpoints in the time series. In these instances, the algorithm successfully captured the transition points in approximately 63% of cases.

It is important to note that the sample sizes for these categories were highly imbalanced, ranging from $n = 20$ for influxes to $n = 238$ for population dynamics. Consequently, any conclusions drawn from these results should be interpreted with caution, as the uneven distribution of samples may influence the observed detection rates and limit the generalizability of the findings.

Nevertheless, the findings from both evaluations underscore that BEAST, in its current parameterization and given the preprocessed input data, is more attuned to distinct breakpoints than to subtle or quickly vanishing anomalies.

RQ2: Role of Data Preprocessing

RQ2 examines how data preprocessing decisions influence BEAST's ability to detect true ecological breakpoints. Both the interviews and the archival comparison show that the choice of metrics (e.g., presence versus abundance) and spatial resolution (e.g., 50×50 km grids) plays a central role in shaping BEAST's detection capabilities. When species presence is aggregated weekly over relatively large geographic areas, the resulting time series effectively captures large-scale trends and abrupt declines or expansions that persist over multiple weeks. However, such a setup can overlook local or short-lived spikes and dips, even if those shifts represent ecologically significant events. Several experts illustrated this point with examples of localized or transient phenomena—such as rare influxes or short mortality events—that did not produce a clear enough change signal at the chosen level of aggregation.

Significantly, neither the quantitative nor the qualitative findings suggest a fundamental flaw in BEAST itself; rather, they point to a mismatch between how certain ecological changes manifest in nature and how the current preprocessing pipeline translates them into time series data.

Thus, the success of BEAST is partly determined by whether a target phenomenon is prominent under the data's chosen temporal and spatial breakdown. As experts suggested, no single data-processing approach can reliably capture every type of change because different types of ecological change points vary considerably, and any single method risks discarding information relevant to certain phenomena. Instead, applying BEAST across multiple data representations may be necessary to detect the broadest possible range of changes. This approach increases the likelihood of detecting a broader spectrum of important changes, particularly those that might otherwise remain hidden under a single preprocessing scheme.

RQ3: Influence of Citizen Science Bias

The third question (RQ3) considers how observer bias in volunteer-collected data might lead BEAST to detect spurious change points. The expert interviews, in particular, shed light on how fluctuations in observer behavior may create artificial signals.

The findings partially confirm this concern. While experts deemed most of the algorithmically identified change points to be consistent with plausible ecological events, the simultaneously detected declines in summer 2021 across multiple species illustrate how observer-related factors, such as altered monitoring efforts, pandemic-induced changes in field activities, or simply shifts in reporting behavior, could explain the apparent 'change' rather than any real ecological phenomenon. This underscores that citizen science data can introduce extraneous signals, which in turn can yield False Positives if not carefully interpreted. Hence, although BEAST remains robust in highlighting distinct shifts within the time series, and despite efforts to minimize user biases, some detected changes still likely stem from user bias rather than genuine population dynamics.

In light of these findings, it is apparent that citizen science data require careful handling if they are to be used reliably in automated change detection. Temporally changing participation rates, uneven sampling intensities, and geographical biases can all create data artifacts. Although BEAST itself does not differentiate between an ecological shift and a shift caused by observer behavior, expert review can help distinguish real from artificial breakpoints. This underscores the need for domain expertise and methodological safeguards—such as standardizing by observation effort—in any final interpretation of detected change points.

RQ4: Expert Perception and Usability

RQ4 addresses how ornithologists perceive the usability and trustworthiness of BEAST and whether they envision integrating it into their ongoing workflows.

Across the interviews, experts unanimously stated that they found the BEAST outputs intuitive, once they became familiar with the method's underlying logic. They expressed substantial enthusiasm for using BEAST in a complementary capacity.

An especially promising application identified by the experts was to use BEAST in near-real-time alert systems, notifying them of emerging shifts rather than relying on biannual reporting and ad hoc communications. They viewed such an automated approach as a useful complement to current analytical practices, able to capture unexpected or subtle changes they might otherwise overlook. Moreover, this data-driven method could help reduce the risk of confirmation bias by highlighting phenomena that observers might not anticipate. Further, experts highlighted the possible integration of BEAST to realize an empirically driven model retraining schedule.

Equally important is the realization that BEAST does not replace expert judgment. Experts viewed the tool as a complement to, rather than a replacement for, existing analytical workflows. While BEAST's capacity to detect unexpected or subtle changes is considered an advantage, the experts underscored that final validation requires human interpretation and reference to additional evidence, such as biannual species reports or specialized knowledge. This notion resonates with the dual requirements of high sensitivity and strategic expert oversight, especially in contexts where erroneous or unverified change point detection could lead to misallocated conservation efforts. Nevertheless, the experts were positive regarding BEAST's utility, believing that it can streamline the process of monitoring bird populations by rapidly highlighting anomalies that merit further inquiry.

RQ5: Generalization and Best Practices for Applying CPD to Citizen Science Data

Beyond the specific findings concerning BEAST, this study highlights several broader implications for the use of change point detection methods on citizen science data in ecological research.

First, while the large spatial and temporal coverage provided by volunteer-collected observations is a clear strength, the potential for observer biases and inconsistencies means that automated analytics must always be interpreted with caution. Despite efforts to mitigate user bias, certain change points indicative of observer behavior—such as changes associated with Covid-related restrictions or inclement weather—persisted in the data. Small shifts in reporting habits can appear as abrupt changes in a time series, underscoring the importance of transparent preprocessing strategies designed to mitigate or flag these effects.

Second, although advanced methods like BEAST can identify meaningful patterns even in noisy data, domain expertise remains central for validating whether these patterns reflect real ecological phenomena or simply artifacts of sampling variability. Involving experts early and often fosters a reciprocal learning process: algorithmic outputs gain ecological credibility while domain practitioners gain exposure to new analytical insights they might not otherwise detect.

Finally, the importance of selecting or combining complementary preprocessing methods becomes especially pronounced in citizen science contexts, where data quality and granularity can vary dramatically across regions, seasons, and species. As illustrated in this study, detecting a specific ecological shift often depends on the interplay between how data are aggregated and the nature of the phenomenon itself. A multi-scale or multi-metric approach broadens the scope of detectability and reduces the risk of missing crucial events.

In summary, change point detection can be effectively applied to citizen science data when approached with care. This involves thoughtfully addressing potential observer biases, selecting appropriate preprocessing methods, and conducting thorough validation with the help of domain experts. These steps ensure that the identified shifts accurately reflect genuine ecological processes rather than being mere artifacts of data collection.

Conclusion

Taken together, the results from both the qualitative and quantitative analyses demonstrate that BEAST can indeed detect a wide array of ecologically relevant changes, but is constrained by how input data are aggregated.

Both the qualitative interviews and the quantitative evaluation revealed in which cases detected breakpoints corresponded with real phenomena, and highlighted how easily smaller-scale or short-lived phenomena can be overlooked. Both perspectives emphasize employing multiple, complementary preprocessing methods can expand the range of events that BEAST can detect while still benefiting from its strengths in identifying large-scale trend changes.

In sum, this study shows that BEAST can deliver considerable value in ecological change detection. While it does not replace existing workflows or human expertise, it can operate as a complementary analytical layer, providing timely alerts, uncovering subtle or unexpected shifts, and informing Machine Learning related operational applications. The experts' proposals for a proactive alert system and adaptive model retraining demonstrate avenues for practical deployment that could enhance ecological monitoring.

At the same time, the findings confirm that neither the algorithm nor its results can stand alone; effective change point detection in citizen science data requires an iterative process, combining algorithmic sensitivity, expert verification, and flexible data representations to capture the full spectrum of ecological shifts.

Ultimately, by combining methodical preprocessing, bias correction, and continuous expert oversight, the BEAST algorithm can serve as a powerful complement to existing approaches, helping researchers, conservationists, and citizen scientists alike respond more effectively to shifting bird population dynamics.

8 Outlook

The findings of this study highlight that the BEAST algorithm can detect meaningful shifts within ornithological time series data, especially when such shifts manifest as spatial expansions, declines, or noteworthy influxes of particular bird species. However, as underscored by both quantitative and qualitative evaluations, the algorithm's performance ultimately depends on data preprocessing choices, parameter settings, and subsequent interpretation. This final chapter therefore outlines potential directions for further research and development, focusing on algorithmic refinements, integration with ornithological workflows, and more comprehensive assessments of BEAST's detection capabilities.

A primary avenue for algorithmic refinement lies in reducing the number of False Positives. The expert assessments demonstrated that BEAST tends to identify certain minor fluctuations as significant change points. A straightforward yet effective approach to addressing this issue would be to impose a threshold on the probability of a trend change point, $Pr(tcp)$, to filter out less certain breakpoints. Moving beyond this probability-based approach, researchers could adopt strategies proposed by Li et al. (2022), who introduced four additional features for identifying unjustified breakpoints: the magnitude of trend change, the change in slope, the probability of the breakpoint itself, and the fraction of anomalous residuals. By defining suitable thresholds for these features, false breaks can be identified and removed. Such refined filtering can further ensure that only change points with strong supporting evidence are retained, thus making the detection process more robust and reliable.

In addition, the built-in features provided by BEAST can be evaluated for their overall utility. For instance, BEAST's outlier detection can be deactivated. In contexts where short-lived yet intense events are ecologically significant, designating such outliers as meaningful change points—rather than excluding them—can yield more informative results. Moreover, BEAST is capable of detecting change points in the seasonal component, which may offer further insights and potentially reveal shifts in phenology.

However, possibly the most significant avenue for further investigation is the systematic experimentation with alternative preprocessing strategies. As the interviews showed, under the current preprocessing scheme, user biases remained in the data. Further, different spatial resolutions, temporal aggregations, and measures of abundance or presence can capture different types of ecological change. Comparative studies of multiple preprocessing pipelines would help determine which strategy is most suitable for user bias reduction and for each class of change (e.g., local population expansions vs. short-lived influx events), enabling a more targeted design of analytical workflows. Possible alternative preprocessing approaches are discussed in the *Related Work* chapter. For example, species distribution models (SDMs) could be utilized to artificially increase sample density in undersampled regions, as suggested by Grattarola et al. (2023).

In parallel to algorithmic developments, there is considerable scope for the integration of BEAST into active ornithological workflows. Domain experts from both DDA and the Swiss Ornithological Institute are developing models that automatically flag implausible bird sightings contributed by citizen scientists. These models aim to reduce the manual effort required to validate untrustworthy reports. Currently, these models are trained on historical data. The findings of this thesis demonstrate that bird populations are constantly in flux—through invasions, expansions, contractions, and other events—so that existing outlier-detection models must be periodically retrained to avert model drift. A reactive approach to retraining, triggered by the detection of genuine shifts, would be particularly valuable. For example, whenever BEAST finds a plausible change point for a given species, an adaptive mechanism could schedule a model update. This dynamic strategy addresses model drift more precisely than a rigid, calendar-based retraining. As noted by one expert, however, a brief delay in retraining may be necessary to ensure that short-lived fluctuations do not prematurely trigger full model updates. Formalizing and testing such a reactive training strategy would be an essential step in operationalizing automated model updates.

Beyond adaptive training, many experts saw promise in using BEAST as a near-real-time alert mechanism, enabling practitioners to receive timely notifications of emergent changes. By offering an early-warning system, BEAST would complement current observer-based monitoring and biannual reporting cycles, making it easier for ornithologists to detect unexpected shifts at an earlier stage. Technical and organizational details surrounding how BEAST could be triggered, how often it would run, and how results would be fed back to domain experts would require further planning. Nonetheless, adopting an online or continually updated approach would help ensure that the algorithm's outputs remain as current as possible.

In practice, an incremental approach—running BEAST weekly or monthly on newly added data—would be required to operationalize both an alert system and reactive model update schedules. The user interface developed in this thesis, alongside the available software code, provides a foundation for integrating BEAST into existing analytical platforms, thereby making continual or near-real-time change point monitoring feasible.

Lastly, a deeper, more structured evaluation of BEAST's performance would further reinforce its utility for long-term ecological monitoring. Although the present evaluation employed archival records and expert interviews, these sources are not free from spatial and temporal uncertainties, thereby constraining the precise validation of change points. Future studies could undertake a more fine-grained labeling effort by narrowing the geographical scope or by assigning higher temporal resolutions to documented ecological events. To complement these expert-based approaches, one might also incorporate simulation techniques and species distribution modeling (SDM). Such methods could generate synthetic or semi-synthetic datasets that incorporate known population shifts, providing a robust ground truth for detecting not only the strengths of BEAST but also its vulnerabilities. More accurate benchmarks would sharpen the algorithm's calibration and further clarify which categories of change points are captured consistently and which remain elusive.

In summary, while this thesis demonstrates the potential of Bayesian-based change point detection for ornithological time series, the algorithm's successful application relies on additional refinements in post-processing, data preparation, model integration, and evaluation design. Implementing advanced filtering methods, expanding or customizing the scope of preprocessing, and aligning algorithmic workflows more closely with expert practices could collectively enhance both the scientific and practical value of automated change point analysis in bird population studies. Integrating BEAST into regular monitoring processes—whether as a trigger for adaptive model retraining, a near-real-time alert system, or a tool for retrospective discovery of hidden shifts—promises more agile, data-driven responses to the ongoing changes in bird distributions and abundances.

Bibliography

Aminikhanghahi, S. & Cook, D. J. (2017), 'A survey of methods for time series change point detection', *Knowledge and information systems* **51**(2), 339–367.

Aqel, K., Wang, Z., Peng, Y. B. & Maia, P. D. (2024), 'Reconstructing rodent brain signals during euthanasia with eigensystem realization algorithm (era)', *Scientific Reports* **14**(1), 12261.

Arcgis (2024), 'Change point detection (space time pattern mining)'. Accessed: 2024-10-28.
URL: *https://doc.arcgis.com/en/allsource/1.0/analysis/geoprocessing-tools/space-time-pattern-mining/change-point-detection.htmL*

Backstrom, L. J., Callaghan, C. T., Worthington, H., Fuller, R. A. & Johnston, A. (2025), 'Estimating sampling biases in citizen science datasets', *Ibis* **167**(1), 73–87.

Banesh, D., Petersen, M., Wendelberger, J., Ahrens, J. & Hamann, B. (2019), 'Comparison of piecewise linear change point detection with traditional analytical methods for ocean and climate data', *Environmental Earth Sciences* **78**, 1–16.

Berghen, B. V., Vanermen, I. & Vranken, L. (2025), 'Testing four design interventions for sustained participation in an agricultural citizen science project', *Environmental Science & Policy* **163**, 103981.

Bonney, R. (2021), 'Expanding the impact of citizen science', *BioScience* **71**(5), 448–451.

Bowler, D. E., Boyd, R. J., Callaghan, C. T., Robinson, R. A., Isaac, N. J. & Pocock, M. J. (2025), 'Treating gaps and biases in biodiversity data as a missing data problem', *Biological Reviews* **100**(1), 50–67.

Browning, D. M., Maynard, J. J., Karl, J. W. & Peters, D. C. (2017), 'Breaks in modis time series portend vegetation change: verification using long-term data in an arid grassland ecosystem', *Ecological Applications* **27**(5), 1677–1693.

Callaghan, C. T., Thompson, M., Woods, A., Poore, A. G., Bowler, D. E., Samonte, F., Rowley, J. J., Roslan, N., Kingsford, R. T., Cornwell, W. K. et al. (2023), 'Experimental evidence that behavioral nudges in citizen science projects can improve biodiversity data', *BioScience* **73**(4), 302–313.

Castillo-Montoya, M. (2016), 'Preparing for interview research: The interview protocol refinement framework.', *Qualitative report* **21**(5).

Cruickshank, S. S., Bühler, C. & Schmidt, B. R. (2019), 'Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends', *Conservation Science and Practice* **1**(7), e54.

Dakki, M., Robin, G., Suet, M., Qninba, A., El Agbani, M. A., Ouassou, A., El Hamoumi, R., Azafzaf, H., Rebah, S., Feltrup-Azafzaf, C. et al. (2021), 'Imputation of incomplete large-scale monitoring count data via penalized estimation', *Methods in Ecology and Evolution* **12**(6), 1031–1039.

Davis, L. S., Zhu, L. & Finkler, W. (2023), 'Citizen science: Is it good science?', *Sustainability* **15**(5), 4577.

Di Febbraro, M., Bosso, L., Fasola, M., Santicchia, F., Aloise, G., Lioy, S., Tricarico, E., Ruggieri, L., Bovero, S., Mori, E. et al. (2023), 'Different facets of the same niche: Integrating citizen science and scientific survey data to predict biological invasion risk under multiple global change drivers', *Global Change Biology* **29**(19), 5509–5523.

Fan, X., Qu, Y., Zhang, J. & Bai, E. (2024), 'China's vegetation restoration programs accelerated vegetation greening on the loess plateau', *Agricultural and Forest Meteorology* **350**, 109994.

Gouraguine, A., Moranta, J., Ruiz-Frau, A., Hinz, H., Reñones, O., Ferse, S. C., Jompa, J. & Smith, D. J. (2019), 'Citizen science in data and resource-limited areas: A tool to detect long-term ecosystem changes', *PLoS One* **14**(1), e0210007.

Grattarola, F., Bowler, D. E. & Keil, P. (2023), 'Integrating presence-only and presence–absence data to model changes in species geographic ranges: An example in the neotropics', *Journal of Biogeography* **50**(9), 1561–1575.

Habel, J. C., Schmitt, T., Gros, P. & Ulrich, W. (2022), 'Breakpoints in butterfly decline in central europe over the last century', *Science of the Total Environment* **851**, 158315.

Hall, D. M., Avellaneda-Lopez, P. M., Ficklin, D. L., Knouft, J. H. & Lowry, C. (2024), 'Citizen silence: Missed opportunities in citizen science', *BioScience* **74**(5), 319–321.

Horváth, L., Rice, G. & Zhao, Y. (2022), 'Change point analysis of covariance functions: A weighted cumulative sum approach', *Journal of Multivariate Analysis* **189**, 104877.

Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T. & Fink, D. (2021), 'Analytical guidelines to increase the value of community science data: An example using ebird data to estimate species distributions', *Diversity and Distributions* **27**(7), 1265–1277.

Johnston, A., Matechou, E. & Dennis, E. B. (2023), 'Outstanding challenges and future directions for biodiversity monitoring using citizen science data', *Methods in Ecology and Evolution* **14**(1), 103–116.

Johnston, A., Moran, N., Musgrove, A., Fink, D. & Baillie, S. R. (2020), 'Estimating species distributions from spatially biased citizen science data', *Ecological Modelling* **422**, 108927.

Kallio, H., Pietilä, A.-M., Johnson, M. & Kangasniemi, M. (2016), 'Systematic methodological review: developing a framework for a qualitative semi-structured interview guide', *Journal of advanced nursing* **72**(12), 2954–2965.

Kovács, S., Bühlmann, P., Li, H. & Munk, A. (2023), 'Seeded binary segmentation: a general methodology for fast and optimal changepoint detection', *Biometrika* **110**(1), 249–256.

La Sorte, F. A. & Somveille, M. (2020), 'Survey completeness of a global citizen-science database of bird occurrence', *Ecography* **43**(1), 34–43.

Li, J., Li, Z. L., Wu, H. & You, N. (2022), 'Trend, seasonality, and abrupt change detection method for land surface temperature time-series analysis: Evaluation and improvement', *Remote Sensing of Environment* **280**.

Matutini, F., Baudry, J., Pain, G., Sineau, M. & Pithon, J. (2021), 'How citizen science could improve species distribution models and their independent assessment', *Ecology and Evolution* **11**(7), 3028–3039.

Mu, Y., Biggs, T. W. & Jones, C. (2023), 'Importance in shifting circulation patterns for dry season moisture sources in the brazilian amazon', *Geophysical Research Letters* **50**(9), e2023GL103167.

Mulverhill, C., Coops, N. C., White, J. C., Tompalski, P. & Achim, A. (2024), 'Evaluating the potential for continuous update of enhanced forest inventory attributes using optical satellite data', *Forestry: An International Journal of Forest Research* p. cpae029.

Myneni, R. B., Keeling, C., Tucker, C. J., Asrar, G. & Nemani, R. R. (1997), 'Increased plant growth in the northern high latitudes from 1981 to 1991', *Nature* **386**(6626), 698–702.

Oehlert, A. M., Hunter, H., Riopelle, C. & Purkis, S. J. (2023), 'Perturbation to north atlantic ocean-climate dynamics tripled whitings mud production in the bahamas', *Journal of Geophysical Research: Oceans* **128**(11), e2023JC020021.

Rbeast Documentation (2019), *Remote Sensing of Environment* **232**.

Rempel, R. S., Jackson, J. M., Van Wilgenburg, S. L. & Rodgers, J. A. (2019), 'A multiple detection state occupancy model using autonomous recordings facilitates correction of false positive and false negative observation errors.', *Avian Conservation & Ecology* **14**(2).

Skrobotov, A. (2023), 'Likelihood ratio test for change in persistence', *Communications in Statistics-Theory and Methods* **52**(17), 5952–5965.

Smith, H. E., Cooper, C. B., Busch, K., Harper, S., Muslim, A., McKenna, K. & Cavalier, D. (2024), 'Facilitator organizations enhance learning and action through citizen science: a case study of girl scouts' think like a citizen scientist journey on scistarter', *Environmental Education Research* **30**(2), 190–213.

Smith, M. M. & Pauli, J. N. (n.d.), 'Connectivity maintains genetic diversity and population persistence within an archipelagic refugia even under declining lake ice', *Mechanisms of species recovery for a forest carnivore in a changing landscape* p. 173.
URL: *https://www.proquest.com/openview/1b5708a82ec7871b6115ba7a09aa2d10/1.pdf?pq-origsite=gscholarcbl=18750diss=ypage=181*

Steen, V. A., Tingley, M. W., Paton, P. W. & Elphick, C. S. (2021), 'Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data', *Methods in Ecology and Evolution* **12**(2), 216–226.

Storch, L. S. & Day, S. L. (2020), 'Computational topology and early warning signals of population collapse'.

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A. et al. (2015), 'Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research', *Biological Conservation* **181**, 236–244.

Truong, C., Oudre, L. & Vayatis, N. (2020), 'Selective review of offline change point detection methods', *Signal Processing* **167**, 107299.

Turner, D. W. (2010), 'Qualitative interview design: A practical guide for novice investigators', *The qualitative report* **15**(3), 754.

Van den Burg, G. J. & Williams, C. K. (2020), 'An evaluation of change point detection algorithms', *arXiv preprint arXiv:2003.06222* .

Verbesselt, J., Hyndman, R., Zeileis, A. & Culvenor, D. (2010), 'Phenological change detection while accounting for abrupt and gradual trends in satellite image time series', *Remote Sensing of Environment* **114**(12), 2970–2980.

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R. & Wagenknecht, K. (2021), *The science of citizen science*, Springer Nature.

Walker, J. & Taylor, P. D. (2017), 'Using ebird data to model population change of migratory bird species.', *Avian Conservation & Ecology* **12**(1).

Wei, S. & Xie, Y. (2022), 'Online kernel cusum for change-point detection', *arXiv preprint arXiv:2211.15070* .

Zaytsev, V., Tutukina, M. N., Chetyrkina, M. R., Shelyakin, P. V., Ovchinnikov, G., Satybaldina, D., Kondrashov, V. A., Bandurist, M. S., Seilov, S., Gorin, D. A. et al. (2024), 'Monitoring of meat quality and change-point detection by a sensor array and profiling of bacterial communities', *Analytica Chimica Acta* **1320**, 343022.

Zbinden, N., Kéry, M., Häfliger, G., Schmid, H. & Keller, V. (2014), 'A resampling-based method for effort correction in abundance trend analyses from opportunistic biological records', *Bird Study* **61**, 506–517.

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J. & Carroll, J. M. (2023), 'Redefining qualitative analysis in the ai era: Utilizing chatgpt for efficient thematic analysis'.
URL: *http://arxiv.org/abs/2309.10771*

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X. & Brown, M. (2019), 'Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A bayesian ensemble algorithm', *Remote Sensing of Environment* **232**.

Zhu, R. & Newman, G. (2024), 'Methodological overview for ebird, inaturalist, gap data, and wildlife richness', *Contemporary Landscape Performance Methods and Techniques* pp. 77–83.

A Anhang

A.1 Applied Tools

Table A.1 lists the tools and resources used for this Master thesis.

Table A.1: Used Tools and Resources

Tool	Usage
LaTeX	Typesetting and layout tool used for the creation of this document
gradio	Python package for building the presented user interface
Rbeast	Python package for performing the BEAST analysis
EEA Reference grid	Geographical grid system for spatially discretizing bird sightings
f4	Automatic transcription of expert interview recordings for the qualitative analysis
ChatGPT 4o and o1	Extraction of categories for the quantitative analysis; Extraction of patterns for the qualitative analysis; Translation tasks for the creation of this document

A.2 Questionnaire for Expert Interviews

A.2.1 Accuracy of BEAST

- Per displayed time series:
  - Do you find the extracted trend reasonable?
  - Do you find the placed change points meaningful?
  - Would you place an additional change point at another location?
  - How would you rate the placement of the change points overall? (1 = very inappropriate, 5 = very appropriate)
- To what extent do you agree with the following statement: "I trust that when I see a change point in the data, BEAST will recognize it as well." (1 = strongly disagree, 5 = strongly agree)
- To what extent do you agree with the following statement: "I trust that when BEAST displays a change point, there is indeed something visible there." (1 = strongly disagree, 5 = strongly agree)

A.2.2 Agreement of Change Points with Reality

Prepared Ecological Events

- Per bird species: Based on your expertise, do the change points match real phenomena?
  - In cases where BEAST identifies a change point that does not correspond to real events: What could be the reasons?
  - How do you assess your knowledge about this bird species? (1 = little knowledge, 5 = expert)
- To what extent do you agree: "I trust that a change point reported by BEAST is associated with a real change in the bird species." (1 = strongly disagree, 5 = strongly agree)

Bird Species from Ornithologists

- Per bird species: Based on your expertise, do the change points correspond to actual changes?
  - In cases where BEAST failed to detect a known change: What could be the reason? (e.g., gradual change, insufficient data, inappropriate preprocessing)
  - How do you assess your knowledge about this bird species? (1 = little knowledge, 5 = expert)
- To what extent do you agree: "I am confident that if a change point is occurring in reality, then BEAST detects it." (1 = strongly disagree, 5 = strongly agree)

A.2.3 Usefulness

- How easy to understand do you find BEAST's plots? (1 = difficult to understand, 5 = easy to understand)
- Does BEAST contribute to new insights for you, or do you already have a tool for that?
- In which application areas do you see a benefit for BEAST, whether for trend analysis or change point analysis?
- How likely is it that you would use BEAST? (1 = very unlikely, 5 = very likely)
- Where do you see possible improvements?
- Do you have any additional comments?

Erklärung zur selbständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort	Datum	Unterschrift im Original