Deciphering multiple changes in complex climate time series using Bayesian inference

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
in der Wissenschaftsdisziplin
Theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von

Nadine Berner

Potsdam, im November 2016

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-100065
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-100065

"If the universe is the answer, what is the question?"
L.M. Lederman

Abstract

Change points in time series are perceived as heterogeneities in the statistical or dynamical characteristics of the observations. Unraveling such transitions yields essential information for the understanding of the observed system's intrinsic evolution and potential external influences. A precise detection of multiple changes is therefore of great importance for various research disciplines, such as environmental sciences, bioinformatics and economics. The primary purpose of the detection approach introduced in this thesis is the investigation of transitions underlying direct or indirect climate observations. In order to develop a diagnostic approach capable to capture such a variety of natural processes, the generic statistical features in terms of central tendency and dispersion are employed in the light of Bayesian inversion. In contrast to established Bayesian approaches to multiple changes, the generic approach proposed in this thesis is not formulated in the framework of specialized partition models of high dimensionality requiring prior specification, but as a robust kernel-based approach of low dimensionality employing least informative prior distributions.

First of all, a local Bayesian inversion approach is developed to robustly infer on the location and the generic patterns of a single transition. The analysis of synthetic time series comprising changes of different observational evidence, data loss and outliers validates the performance, consistency and sensitivity of the inference algorithm. To systematically investigate time series for multiple changes, the Bayesian inversion is extended to a kernel-based inference approach. By introducing basic kernel measures, the weighted kernel inference results are composed into a proxy probability to a posterior distribution of multiple transitions. The detection approach is applied to environmental time series from the Nile river in Aswan and the weather station Tuscaloosa, Alabama comprising documented changes. The method's performance confirms the approach as a powerful diagnostic tool to decipher multiple changes underlying direct climate observations.

Finally, the kernel-based Bayesian inference approach is used to investigate a set of complex terrigenous dust records interpreted as climate indicators of the African region of the Plio-Pleistocene period. A detailed inference unravels multiple transitions underlying the indirect climate observations, that are interpreted as conjoint changes. The identified conjoint changes coincide with established global climate events. In particular, the two-step transition associated to the establishment of the modern Walker-Circulation contributes to the current discussion about the influence of paleoclimate changes on the environmental conditions in tropical and subtropical Africa at around two million years ago.

Zusammenfassung

Im Allgemeinen stellen punktuelle Veränderungen in Zeitreihen (*change points*) eine Heterogenität in den statistischen oder dynamischen Charakteristika der Observablen dar. Das Auffinden und die Beschreibung solcher Übergänge bietet grundlegende Informationen über das beobachtete System hinsichtlich seiner intrinsischen Entwicklung sowie potentieller externer Einflüsse. Eine präzise Detektion von Veränderungen ist daher für die verschiedensten Forschungsgebiete, wie den Umweltwissenschaften, der Bioinformatik und den Wirtschaftswissenschaften von großem Interesse. Die primäre Zielsetzung der in der vorliegenden Doktorarbeit vorgestellten Detektionsmethode ist die Untersuchung von direkten als auch indirekten Klimaobservablen auf Veränderungen. Um die damit verbundene Vielzahl an möglichen natürlichen Prozessen zu beschreiben, werden im Rahmen einer Bayes'schen Inversion die generischen statistischen Merkmale Zentraltendenz und Dispersion verwendet. Im Gegensatz zu etablierten Bayes'schen Methoden zur Analyse von multiplen Übergängen, die im Rahmen von Partitionsmodellen hoher Dimensionalität formuliert sind und die Spezifikation von Priorverteilungen erfordern, wird in dieser Doktorarbeit ein generischer, Kernel-basierter Ansatz niedriger Dimensionalität mit minimal informativen Priorverteilungen vorgestellt.

Zunächst wird ein lokaler Bayes'sche Inversionsansatz entwickelt, der robuste Rückschlüsse auf die Position und die generischen Charakteristika einer einzelnen Veränderung erlaubt. Durch die Analyse von synthetischen Zeitreihen die dem Einfluss von Veränderungen unterschiedlicher Signifikanz, Datenverlust und Ausreißern unterliegen wird die Leistungsfähigkeit, Konsistenz und Sensitivität der Inversionmethode begründet. Um Zeitreihen auch auf multiple Veränderungen systematisch untersuchen zu können, wird die Methode der Bayes'schen Inversion zu einem Kernel-basierten Ansatz erweitert. Durch die Einführung grundlegender Kernel-Maße können die Kernel-Resultate zu einer gewichteten Wahrscheinlichkeit kombiniert werden die als Proxy einer Posterior-Verteilung multipler Veränderungen dient. Der Detektionsalgorithmus wird auf reale Umweltmessreihen vom Nil-Fluss in Aswan und von der Wetterstation Tuscaloosa, Alabama, angewendet, die jeweils dokumentierte Veränderungen enthalten. Das Ergebnis dieser Analyse bestätigt den entwickelten Ansatz als eine leistungsstarke diagnostische Methode zur Detektion multipler Übergänge in Zeitreihen.

Abschließend wird der generische Kernel-basierte Bayes'sche Ansatz verwendet, um eine Reihe von komplexen terrigenen Staubdaten zu untersuchen, die als Klimaindikatoren der afrikanischen Region des Plio-Pleistozän interpretiert werden. Eine detaillierte Untersuchung deutet auf multiple Veränderungen in den indirekten Klimaobservablen hin, von denen einige als gemeinsame Übergänge interpretiert werden. Diese gemeinsam auftretenden Ereignisse stimmen mit etablierten globalen Klimaereignissen überein. Insbesondere der gefundene Zwei-Stufen-Übergang, der mit der Ausbildung der modernen Walker-Zirkulation assoziiert wird, liefert einen wichtigen Beitrag zur aktuellen Diskussion über den Einfluss von paläoklimatischen Veränderungen auf die Umweltbedingungen im tropischen und subtropischen Afrika vor circa zwei Millionen Jahren.

Contents

List of Figures

List of Tables

List of Symbols

x	scalar
$\boldsymbol{x}$	vector
X	matrix
$\mathbf{1}$	identity matrix
$(\cdot)^T$	transpose of a vector or a matrix
$(\cdot)^{-1}$	inverse of a matrix
$\lvert \cdot \rvert$	determinant of a matrix
$tr(\cdot)$	trace of a matrix
$rank(\cdot)$	rank of a matrix
$\lVert \cdot \rVert^2$	norm
$\langle \cdot \rangle$	average
$\mathbb{E}(\cdot)$	expected value
$STD(\cdot)$	standard deviation
$p(\cdot)$	probability distribution
$p(\cdot\lvert\cdot)$	conditional probability distribution
$\mathcal{L}(\cdot)$	likelihood function
$\hat{(\cdot)}$	maximum a posteriori (MAP) estimator
$(\cdot)^*$	other estimators such as the profiled likelihood function (PLH) or the best linear unbiased predictor (BLUP)
δ_{ij}	Kronecker delta
$\underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x})$	the value of $\boldsymbol{x}$ which minimizes $f(\boldsymbol{x})$
$\{\cdot\}$	set of observations (sequence) or models (family of distributions)
$\mathcal{U}(a,b)$	uniform distribution within the interval $[a,b]$
$\mathcal{N}(\boldsymbol{\mu},\Sigma)$	Gaussian or normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ

$t_\nu(\boldsymbol{\mu}, \Psi)$	t-distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$ and scale matrix Ψ
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^{n \times m}$	real $n \times m$-dimensional vector space
$\mathcal{O}(\cdot)$	Landau symbol indicating the order of algorithm complexity

1 Introduction

As summarized in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) the *"evidence has grown [...] that impacts of recent changes in climate on natural and human systems occur on all continents and across the oceans"* [1]. The scientific studies cited in the IPCC report suggest the attribution of changes in environmental and ecosystems and in economic and agricultural systems to climate changes. In consequence, a better understanding of the climate system and its changes provides a solid basis of decision-making to adapt to anticipated future changes. For that purpose the investigation of recent climate changes is realized by direct observation of climate components on decadal and millennial time scales. Different to recent changes past climate changes on time scales of thousands to millions of years are indicated by indirect observations in terms of proxy records sensitive to specific climate components. The past climate periods may not be analogues to modern climate conditions [2], but offer valuable study cases for the response of the climate system to various forcings and its potential influences on natural systems [3]. A prominent example of the importance of paleoclimate transitions on the biosphere is the study of the relationship between environmental change and human evolution in tropical and subtropical Africa of the last five million years [4–6]. The presented thesis aims to provide further probabilistic information about potential changes in the according paleoclimate observations to that debate. Nevertheless, before any causality explaining a change in the complex climate system [7] may be formulated or any influence on other natural systems may be attributed to this change, the potential change has to be unraveled from direct or indirect climate observations.

Climate change is generally defined in terms of variability, frequency, intensity, spatial extent, duration, and occurrence of extreme events [8]. According to the different domains of possible changes the analysis approaches are as well conceptually different. To infer on a change of a component within an observed time series advanced analysis techniques are used to investigate changes of (i) dynamical properties e.g. via recurrence network analysis [9], (ii) frequency patterns e.g. via wavelet [10] or singular value analysis [11] or (iii) regression parameters e.g. via estimation approaches [12, 13], to name only a few.

The general parametric approach to change point detection, such as regression, assumes a process describing the observed time series and comprising a change in one or multiple process properties. The process parameters are then estimated from the observations. By the reformulation of the estimation problem in the light of Bayesian statistics, each model parameter may be assigned with uncertainty. The performance of

the Bayesian inversion yields conditional probability densities of each model parameter given the observations. Besides the parameter estimation on its own, a degree of belief about the assumed model and about the uncertainties in the employed parameters can be obtained. In particular, the probability density of the parameter describing the change point intuitively indicates the degree of belief about a change occurring in the analyzed time series. However, the number of process parameters increases with the amount of change points. Thus, the models capable to describe complex paleoclimate observations potentially comprising multiple changes are of high dimensionality. The according probability densities can therefore only be derived by advanced sampling strategies [14] or dynamic programming algorithms [15]. In contrast to these Bayesian approaches, the presented thesis elaborates a Bayesian method to detect multiple changes without enlarging the change point model's dimensionality and combines the derived probabilistic information into a comprehensive credibility expression.

The incorporation of uncertainty in geophysical problems, including climate research, offers new perspectives for analysis approaches since advances in modeling strategies, computational performance and specific algorithms make Bayesian concepts more and more feasible [16, 17]. Despite ongoing debates about the realization of explicit Bayesian analysis strategies [18], Bayesian approaches have the potential to spark new exciting discussions to existing scientific problems. Notably, the Bayes principle offers new perspectives in the field of climate research, e.g. the inference on detection and attribution of recent climate changes in a single computational step [19, 20]. Since *"the use of statistics is pervasive in the climate sciences, not only for the extraction and quality control of data, but also for the synthesis of knowledge and information from that data"* [21], Bayesian approaches offers a powerful framework to contribute to that matter. In paleoclimate research the interest in and application of Bayesian methods is increasing. Some recent analysis examples illustrate the use of the Bayesian strategy, e.g., to calibrate age-depth models [22], to unravel uncertainty propagation from age calibration to proxy records [23], to investigate multiple climate time series up to millennial time scales [24] and single time series up to a time scale up to millions of years [15]. The essential aim of the Bayesian inference approach proposed in this thesis is the investigation of complex paleoclimate observations spanning the time scale of a few million years for multiple changes. The derived probabilistic information can then be used for detailed comparative studies between observations from different locations interpreted as indicators of the same climate aspect.

The presented thesis is organized as outlined in the following: In Chap. 2 the theoretical background of the Bayesian analysis approach to detect changes in climate time series is elucidated. In principle, climate signals are interpreted as complex observations assigned with uncertainties due to the variety of the involved generating processes. The basic formation mechanisms of climate observations are exemplified for the time series investigated in the course of this thesis, followed by a brief overview of advanced analysis approaches to climate time series. In contrast to such specialized approaches, the general change point problem is formulated and the main approaches to its solution -*inter*

alia Bayesian model selection- are presented, including an explanation of the Bayesian theorem to understand the mathematical concept of Bayesian inference. Finally, the basic components of Bayesian approaches to multi change point problems are described.

Chapter 3 begins with an introduction of a local Bayesian inference approach to a single transition. The explicit formulation of a generic transition model is motivated by the aim to at least locally approximate complex observations. Further, the Bayesian inversion is explained in its analytical and numerical form. In order to validate the approach as a detection method it is applied on synthetic time series comprising changes of different observational evidence. Moreover, the empirical convergence of the posterior distribution and the robustness of the employed estimators are assessed. By investigating environmental time series comprising documented changes the suitability of the generic transition model as a first order approach to direct climate observations is finally discussed.

To investigate multiple transitions, the local Bayesian inference approach is extended to a kernel-based inference approach as elaborated in Chap. 4. Basic kernel weights are proposed to allow the composition of a proxy probability to a hypothetical multi transition posterior distribution. The extended approach is employed to reanalyze the environmental time series of the previous chapter.

In Chap. 5 the multi transition inference is used to investigate a set of complex observations indicating a paleoclimate aspect of the subtropical and tropical African region of the last five million years. The inferred changes are discussed in the context of previous studies based on various climate records, simulations or previous advanced analysis methods. In Chap. 6 the main findings of this thesis are summarized and a further outlook with respect to the introduced Bayesian approach is provided.

2 Climate observations and change point problems

The aim of this thesis is to develop a diagnostic tool to investigate climate observations of different types for structural changes over different temporal scales. For that reason, the presented approach is based on a generic change point model transferable to various types of climate observations. To motivate the investigation of changes as generic features of climate time series, the main aspects of directly and indirectly measurable climate observations are outlined in the light of uncertainty. The term *climate changes* usually describes specific variations in the properties of climate variables. Thus, the according climate observations are commonly analyzed with methods adapted to the specific properties of interest as briefly summarized in this chapter.

In contrast to such highly specialized approaches, the theoretical background of the general change point problem in terms of statistical inhomogeneities is formulated. One way to analyze such generic changes is accomplished by model selection via Bayesian inversion. To provide an understanding of the concept of Bayesian inversion guiding the development of the approach presented in this thesis, the Bayesian theorem and the components used therein are explained. In this way the Bayesian perception of a probability as a degree of belief is emphasized. Finally, the basic strategies commonly employed in Bayesian change point detection approaches are described.

2.1 Uncertainty of climate time series

An important scope of climate research is the inference on the principles guiding the climate system by analyzing climate sensitive observables. The better understanding of the climate system, its properties and variations, enables scientists to approach related hypotheses, such as predictions of future climate conditions by increasingly precise climate modelling. The practical strategy to investigate the climate system despite its complexity consists of the decomposition into subsystems or climate components [25, 26]. These climate components are mutually interacting and exhibit various characteristic properties over a broad range of space [27, 28] and time scales [29, 30] over many orders of magnitude. Based on acquired climate observables assumed to reflect the evolution of specific climate components at a given spatio-temporal scale, aspects of the climate system can be studied (see Fig. 2.1).

Irrespective of the analysis approach, i.e. exploratory (inductive) or confirmatory

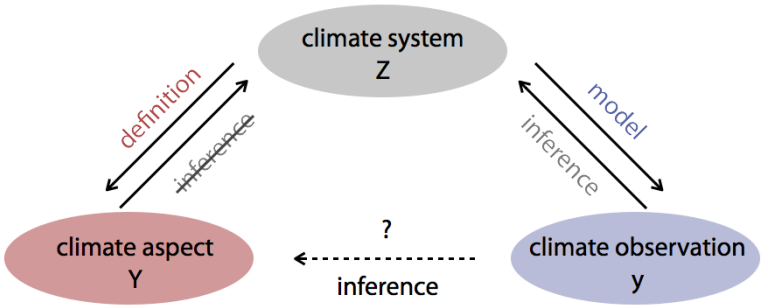


Figure 2.1: *The climate aspect may be interpreted in terms of a hypothesis related to climate system. In principle the inference from the aspect to the system is not possible due to identifiability. By integrating over the complete system the observations may be used to infer on the aspect.*

(deductive), one has to be aware of the fact, that climate observables are inherently uncertain. The sources of uncertainty are manifold, case-specific and only partly known for the different types of climate signals. In order to schematically illustrate the main sources of uncertainties commonly assumed for climate time series, a hierarchical model approach [31] serves as a basic mathematical framework. Thus the fundamental differences between the uncertainties assigned to direct and to indirect climate observations are elucidated. Clearly, a complete and detailed overview about incorporating uncertainty in climate signals is beyond the scope of this thesis. Instead, the following section aims to exhibit the challenges scientists face by interpreting observables as climate sensitive signals and to motivate the design of a generic change point model within this thesis.

A generic description of a climate component can be formulated in terms of a latent, continuous space-time climate process $\boldsymbol{Z}$

$$\boldsymbol{Z} = \{Z(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{S}, t \in \mathcal{T}\} \ , \tag{2.1}$$

within the spatial $\mathcal{S}$ and temporal $\mathcal{T}$ domain of interest. The attribute *latent* refers in a statistical sense to an unobserved quantity describing the underlying climate process as deduced from theoretical and empirical studies. The actually measurable climate variable can be defined as a further stochastic process $\boldsymbol{Y}$ conditional on the climate process under study $\boldsymbol{Z}$ [31]. For simplicity, the process is specified as a real-valued, time-discrete process located at a fixed point $\boldsymbol{s}'$ in space

$$\boldsymbol{Y} = \{Y(\boldsymbol{s}', t_i) : \boldsymbol{s}' \in \mathcal{S}, t_i \in \mathcal{T}, i = 1, ..., n\} : \Omega \to \mathbb{R}^n \tag{2.2}$$

describing n real *in situ* observations. The stochastic process $\boldsymbol{Y}$ defines a probability model of the climate variable on a probability space (Ω, Σ, P), determined by the sample space Ω, the σ-algebra Σ and the probability measure P. The sampling path of this process realizes the acquired climate observations $\boldsymbol{y}$ measured at a fixed location $\boldsymbol{s}'$ at

discrete times t_i for $i = 1, \ldots, n$:

$$\boldsymbol{y} = \{y(\boldsymbol{s}', t_i) : \boldsymbol{s}' \in \mathcal{S}, t_i \in \mathcal{T}, i = 1, ..., n\} \ , \tag{2.3}$$

that is a univariate time series. Note, that in this thesis only time series at fixed locations are considered and thus, the spatial variable $\boldsymbol{s}'$ is neglected for the sake of convenience. The common assumption of the equivalence between the stochastic process $\boldsymbol{Y}$ and the associated statistical model $\mathcal{M}$ [32], both defined on the same sampling space Ω and following the same σ-algebra Σ, allows to perform the investigation of the climate variable $\boldsymbol{Y}$ in the appropriate function space

$$\mathcal{M} = \{f(\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \Theta\} \tag{2.4}$$

of the family of suitable distributions $f(\boldsymbol{\vartheta})$ parametrized by $\boldsymbol{\vartheta}$. Hence, based on an assumed statistical model $\mathcal{M}$, the analysis of the climate signal $\boldsymbol{y}$ enables to infer on properties of the climate variable $\boldsymbol{Y}$ and provides information for the target climate process $\boldsymbol{Z}$ of interest [31, 33]. The investigation of climate variables naturally relies on the significance and precision of the derived climate observables. The inherent uncertainties influencing both aspects differ substantially for directly and indirectly observed climate signals as exemplified in the following.

2.1.1 Direct climate observations

Climate signals that can be directly accessed by physical measurements are referred to as direct climate observations and exist on a time scale up to 400 years before present [34]. Given the instrumentation equipment, the acquirement of the climate time series can be described by the generic scheme in Fig. 2.2. The annual mean temperature series from the Tuscaloosa weather station [35–37] investigated in Sec. 3.4.2 and 4.3.2 offers an example of direct climate observations.

- **climate variable** (e.g. local surface temperature)
  The climate variable $Y(t)$ (e.g. local surface temperature) represents an associated climate component of interest $Z(t, \boldsymbol{s})$ (e.g. surface temperature field).

- **direct observations** (e.g. ground-based daily temperature)
  The direct observations (e.g. temperature) are obtained by a physical measurement process given the instrumentation (e.g. ground-based thermometer) as time-discrete observations $y(t_i)$ at known time points t_i (e.g. daily). The measurement process is subject to uncertainty caused by a finite resolution, measurement errors etc.

Therefore, besides the random nature of the climate variable itself, further uncertainties rise from the measurement process. Moreover, any postprocessing of the observations (e.g. averaging the daily temperatures to an annual mean) influences the climate signal (e.g. reduces autocorrelation, see Sec. 3.4.2) as well as the uncertainty assigned to it.

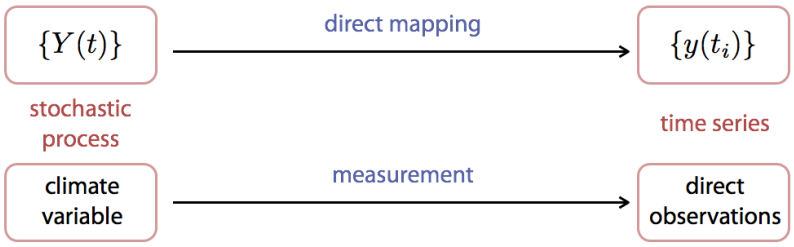


Figure 2.2: *Direct climate observations are accessible via direct physical measurements of the climate variable of interest. In mathematical terms, a time-continuous stochastic process* $\{Y(t)\}$ *is realized as a time series* $\{y(t_i)\}$ *measured at known discrete time points* t_i *for* $i = 1, ..., n$*. Besides the random nature of the stochastic process, uncertainty is attributed to the measurement process (e.g. resolution, errors) depending on the instrumentation.*

2.1.2 Indirect climate observations

On longer time scales, climate signals are not directly measureable and can be only indirectly accessed by proxy variables. A proxy variable describes an adequate substitute observable which is assumed to have systematically responded to a climate variable of interest. The proxy variable (e.g. treerings, pollen, ^{18}O oxygen isotopes) is assumed to be preserved within an archive (e.g. trees, sediment, ice) from which it can be obtained. Based on the characteristic archive stratification (e.g. growth layers, core depth) a time scale is commonly constructed by gauging the sequence at physically measured (e.g. radiometric) time reference points. The underlying processes leading to the final derived measurements are case-specific and usually incompletely understood [23]. Consequently, the interpretation of proxy variables as specific climate indicators is subject to an increased uncertainty compared to direct climate observations [38].

The scheme presented in Fig. 2.3 depicts a strong simplification of the set of processes generating the indirect observations and is adapted to the proxy variable terrigenous dust [5, 39], analyzed in Sec. 5. Even though exemplified for a specific proxy, the explanation of the scheme's components stress the fact, that uncertainties assigned to indirect observations emerge from various, vaguely known processes.

- **climate variable** (e.g. humidity)
  The climate variable $Y(t)$ (e.g. humidity) represents an associated climate component of interest $Z(t, \boldsymbol{s})$ (e.g. environmental conditions in tropical Africa).

- **proxy** (e.g. terrigenous dust)
  The qualitative dependency between the climate variable $Y(t)$ and proxy variable $Y'(t)$ (e.g. terrigenous dust) is described by a transfer hypothesis (e.g. more humidity, less available dust in the atmosphere). Note, that the proxy can be sensitive to multiple climate variables. Moreover, the transport mechanisms of the proxy from source to archive are an important aspect of the transfer hypothesis. The

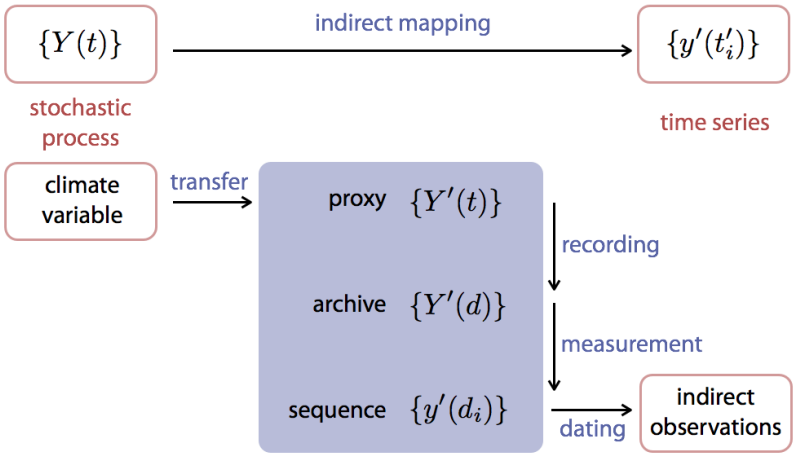


Figure 2.3: *Indirect climate observations are not directly accessible, but a result of a set of various processes. The scheme is simplified and adapted to the proxy terrigenous dust as explained in the text. By a transfer hypothesis the time-continuous climate variable $\{Y(t)\}$ of interest can be related to a proxy variable $\{Y'(t)\}$. A recording process describes the embedding of the proxy variable $\{Y'(t)\}$ into a continuously stratified archive $\{Y'(d)\}$. Given the instrumentation, the recorded proxy variable $\{y'(d_i)\}$ is physically measured at discrete stratification points d_i. A dating process models the mapping of the observed proxy $\{y'(t'_i)\}$ onto an estimated age scale t'_i. Naturally, all processes enclosed in the generation of the indirect observations are inherently uncertain and therefore decrease the significance and precision of the climate sensitive signal.*

assumptions leading to the estimated dependency are mainly assessed by analog empirical studies (e.g. dust transport given modern wind patterns).

- **archive** (e.g. oceanic sediment core)
  The embedding of the proxy variable $Y'(t)$ into the archive $Y'(d)$ (e.g. oceanic sediment) is considered as a recording process (e.g. sedimentation over depth d in the deep sea). Besides natural causes of perturbation within the recording process (e.g. bioperturbation), the acquirement of the archive itself can be disrupted (e.g. mistakes in the core drilling).

- **sequence** (e.g. terrigenous dust measured over sediment depth)
  The recorded proxy $y'(d_i)$ is physically measured (e.g. magnetic susceptibility) at discrete stratification points of the archive (e.g. depth points d_i). Note, that the applied measurement techniques can vary over the archive due to practical reasons.

- **indirect observations** (e.g. measured terrigenous dust related to age scale)
  The indirect climate observation $y'(t'_i)$ is the result of a complex dating process. Based on a synthesis of physically measured time reference points, tuning at presumably known cycles (e.g. oxygen isotopic stratigraphy) and correlation to presumably known events (e.g. magnetostratigraphic datums) the depth scale $\{d_i\}$ can be transformed into a derived age scale $\{t'_i\}$.

The scheme's categories are ambiguous and may be allocated differently. Indeed, the used categories are motivated by approaches that model aspects of the signal derivation in order to assess the underlying uncertainty structure. The subsets of the scheme's processes are investigated separately, e.g. the transfer from climate to proxy variable [40], aspects of the recording and age scaling process by using age-depth models [22, 41] and the propagation of uncertainties within a depth-age-proxy approach [23, 42]. However, a comprehensive framework enclosing the main aspects of the observation generating process is not yet accomplished. Generic hierarchical approaches to infer from observations on the climate component of interest are formulated, but applied so far on data rich settings, i.e. multiple sequences, spanning time scales smaller than one million years [31].

A prevalent approach in climate research is to favor specialized models to investigate climate observations of interest in order to make *"scientific progress"* [33]. Nevertheless, for many processes generating climate observations, in particular indirect climate observations, a specialized model framework is often not feasible. In order to investigate changes in few (direct or indirect) climate observation sequences at different time scales a generic non-hierarchical model is employed in this thesis. Clearly, a generic approach to climate time series can not substitute a detailed understanding of the specific processes leading to the measured climate signals. A generic model, however, offers an efficient diagnostic tool in order to capture a variety of different natural processes generating climate signals. Furthermore, the resulting generic probabilistic expression of changes occurring in the data is comparable between different types of climate observations and has the potential to supplement existing probabilistic analysis methods.

2.1.3 Analysis approaches to climate time series

As outlined in the previous section, climate observations are inherently uncertain but - for the sake of feasibility of the analysis - the uncertainties are commonly neglected. The climate time series may be as well irregularly sampled in time or correlated in between discrete observations. The analysis approaches to climate time series need to take into account these potential features of climate observations, in particular for the precise detection of changes. Changes in climate observations are considered as characteristic variations of associated climate components and are commonly interpreted as structural dynamical changes or structural statistical changes [34]. Therefore, the main analysis methods of changes in climate time series are based on one of these interpretations.

On the one hand, climate changes are considered as transitions between dynamical regimes of an assumed model describing an observed climate variable. The analysis methods employ descriptive state space parameters of interest in order to elucidate the domain of changes within the complex model. Based on the characteristics of the inferred transitions, this approach enables to explore dynamical correlation patterns between different climate variables. In this way, the dynamical investigation may provide auxiliar information to study the causality patterns of changes within the climate system. That

can be accomplished for example by recurrence network analysis approaches [6, 43].

On the other hand, climate changes are considered as structural changes in the statistical properties of the climate observations, generally referred to as statistical inhomogeneities. By formulating a specific statistical model describing the deterministic and stochastic component of an observed climate variable, the climate time series can be investigated. The inferred changes may be due to statistical variability of the climate variable but as well due to changes in the instrumentation process. For clarity, note that different to the use throughout this work the term *inhomogeneities* is in climatology primarily restricted to artificial changes due to technical reasons within the measurement of direct observations [12, 35, 44, 45]. Based on the estimated timing or patterns of the inferred changes, the approach has the potential to elucidate synchronous changes between various climate variables, but does not necessarily imply causality [5, 46, 47]. That can be accomplished for example by adapted regression techniques [13, 35], wavelet based concepts [10, 48], and Bayesian approaches [49–53].

In conclusion, most of the established analysis techniques rely on specific, and thus often complex, hypotheses adapted to each climate variable of interest. The information obtained by these approaches needs to be interpreted with respect to the specific model assumptions (e.g. explicit formulations) and application requirements (e.g. equidistance in time). Different to that, the analysis approach presented in this thesis is based on a generic model describing a statistical change. Consequently, the formulated change point problem is as well generic and may not answer complex hypotheses about specific climate variables. However, due to its generity, the approach is applicable to a variety of climate observations. The formulation as a linear model allows to realize processes as commonly assumed for environmental observations (normal and log-normal). The generic model design facilitates the explicit Bayesian inference. The method does not rely on temporal equidistant observations and so far, does not include correlation patterns. The resulting probabilistic expression indicates the degree of belief about multiple changes within the observations over different temporal scales and can thereby serve as an intuitive diagnostic tool for the investigation of climate time series.

2.2 Statistical approaches to change point problems

To understand the investigation of statistical structural changes in the context of change point problems, the framework of a general change point problem is explained in the following. Analysis techniques are generally based on the assumption of stochastic structural stability of the observations under study. Any structural change in the data therefore puts the application of techniques requiring stochastic homogeneity under question. Hence, the solution of change point problems is of fundamental importance for numerous research disciplines, such as climatology [5, 46], ecology [54, 55], bioinformatics [56–58], economics [59, 60] and many more. Commonly, each scientific discipline analyzes conceptually different processes characterized by different properties. Therefore,

the literature provides a variety of potential approaches [61–64]. The essential idea of all change point detection methods, is to validate the mere existence of changes in the data and, if present, to estimate the number of changes and the corresponding locations within the sequence of observations.

2.2.1 Hypothesis test

Traditionally, change point problems are formulated as hypothesis tests [64, 65]. In order to accept or reject the existence of change points within the observations, the assumed changes are stated as an alternative hypothesis H_1. The null hypothesis H_0 phrases that no changes of the assumed characteristics occur in the observations. As a generic example, let the observations consist of a multivariate time series $\{\boldsymbol{y}(t_i)\} = \{\boldsymbol{y}_i\}$ with $\boldsymbol{y}_i \in \mathbb{R}^m$ measured at discrete time points t_i with $i = 1, ..., n$. Each independent random vector $\boldsymbol{y}_i$ is assumed to be a sample drawn from a probability density function $f(\boldsymbol{y}_i) = f_i$ describing the underlying process. Structural stability or in other words stochastic homogeneity, as a consequence of the absence of any change is defined in the null hypothesis as

$$H_0 : f_1 = f_2 = ... = f_n \,. \tag{2.5}$$

An unknown number k of structural changes c_j located at the time points t_j for $j = 1, ..., k$ break the stochastic homogeneity of the time series. These stochastical inhomogeneities, are used to formulate the alternative hypothesis as

$$H_1 : f_1 = ... = f_{c_1-1} \neq f_{c_1} = ... = f_{c_2-1} \neq f_{c_2} = ... = f_{c_k-1} \neq f_{c_k} = ... = f_n \tag{2.6}$$

and divide the time series into $k + 1$ segments

$$\delta_j = [t_{c_j}, ..., t_{c_{j+1}}[\text{ for } j = 0, ..., k \tag{2.7}$$

with the convention $t_1 = t_{c_0}$ and $t_n = t_{(c_{k+1})-1}$.

As a reasonable assumption, in particular with respect to a common stochastic process $\{\boldsymbol{Y}_i\}$ generating the observations $\{\boldsymbol{y}_i\}$, the distribution functions f_i may be considered as members of a common parametric family $f(\boldsymbol{y}_i, \boldsymbol{\eta}_i) = f(\boldsymbol{\eta}_i)$, characterized by the population parameters $\boldsymbol{\eta}_i \in \mathbb{R}^p$. Thus, the change point problem can be reformulated in terms of structural stability in the population parameters via the null hypothesis as

$$H_0 : \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = ... = \boldsymbol{\eta}_n \,, \tag{2.8}$$

versus changes of the population parameters via the alternative hypothesis as

$$H_1 : \boldsymbol{\eta}_1 = ... = \boldsymbol{\eta}_{c_1-1} \neq \boldsymbol{\eta}_{c_1} = ... = \boldsymbol{\eta}_{c_2-1} \neq \boldsymbol{\eta}_{c_2} = ... = \boldsymbol{\eta}_{c_k-1} \neq \boldsymbol{\eta}_{c_k} = ... = \boldsymbol{\eta}_n \,. \tag{2.9}$$

Adjusted for the unknown change point locations c_j, the test statistics are considered as two-sample tests and are derived by maximum-type procedures. Based on the selected

significance level α, the critical values for the hypothesis test is obtained by evaluating the asymptotic properties of the employed test statistics [65]. By rejecting the null hypothesis H_0, the structural changes are accepted as the alternative hypothesis H_1 describing the observations. Next to the mere decision task, the change points c_j need then to be estimated from the observations. The standard methods of change point inference are usually based on the maximum likelihood ratio or asymptotic information criteria such as the Akaike information criterion (AIC) or Schwarz information criterion (SIC) [64]. Generally, the combination of hypothesis test and inference approach in the context of the change point problem is denoted by change point detection.

The detection methods are distinguished as *local* approaches, capable to describe the process in a subinterval of the time series at a local scale, or as *global* approaches capable to describe the process over the complete time series at a global scale. As a matter of feasibility, the hypotheses are in practice often formulated for a single change point, mainly because the number of inhomogeneities k in the observations are *a priori* not known. Therefore, a practicable strategy is necessary to analyze sequences for an arbitrary number of structural changes. The standard strategy is to employ the single change point analysis in a binary segmentation procedure as introduced in [66]. For that purpose, the single change point detection approach is applied on the whole sample and the sample is divided at the detected change point. The change point detection approach is then performed iteratively on the resulting two segments until no further change point is detected.

2.2.2 Model selection

By formulating a model describing the process generating the observations, the task of hypothesis test can be reformulated as a model test or model selection. The comparison of distributions from the competing hypotheses in terms of a test statistic can thus be generalized to a model comparison also taking into account the difference in parametric complexity [64, 67]. As an example, let the observations constitute a univariate time series $\boldsymbol{y} = \{y(t_i)\} = \{y_i\}$ with $y_i \in \mathbb{R}$ measured at discrete time points t_i with $i = 1, ..., n$. The random vector $\boldsymbol{y}$ is assumed to be generated by a stochastic process described by the parameter vector $\boldsymbol{\vartheta} \in \mathbb{R}^p$. Different to the population parameter $\boldsymbol{\eta}$ of the previous example, the $\boldsymbol{\vartheta}$ is a generic parameter vector that may contain parameters describing random as well as deterministic properties in the parameter space $\Theta \subseteq \mathbb{R}^p$ of finite dimension p. In this way the assumed underlying process can be defined as a parametric statistical model $\mathcal{M}_{\boldsymbol{\vartheta}} = \{f(\boldsymbol{y}, \boldsymbol{\vartheta})\}$ in terms of a family of distributions $f(\boldsymbol{y}, \boldsymbol{\vartheta})$, e.g. a linear mixed model. Structural stability, as a consequence of the absence of any change is defined in the null hypothesis as

$$H_0 : \mathcal{M}_{k=0} = \{f(\boldsymbol{y}, \boldsymbol{\vartheta}) : \quad \boldsymbol{\vartheta} \in \Theta, \quad \text{and} \quad \boldsymbol{y} = \{y_i\} \quad \text{for} \quad i = 1, ..., n\} \ . \tag{2.10}$$

By assuming k structural changes c_j within the process, the time series is divided in $k+1$ subseries $\boldsymbol{y}_j = \{y(\delta_j)\}$. Here the definition of the segmentation δ_j is employed. Since

a common statistical model is assumed, it can be referred to as $\mathcal{M}_{\boldsymbol{\vartheta},k} = \mathcal{M}_k$, solemnly with respect to the amount of the number of changes k. The alternative hypothesis can therefore be formulated as

$$H_1 : \mathcal{M}_k = \{f(\boldsymbol{y}_j, \boldsymbol{\vartheta}_j) : \quad \boldsymbol{\vartheta}_j \neq \boldsymbol{\vartheta}_{j+1}, \quad \boldsymbol{\vartheta}_j \in \Theta \quad \text{and} \quad \boldsymbol{y}_j = \{y(\delta_j)\} \quad \text{for} \quad j = 0, ..., k\}\,.$$

The resulting set of possible segments $\Lambda_k = \{\delta_j : \ j = 0, ..., k\}$ has the cardinality $\binom{n}{k+1}$. Based on Λ_k the set of all possible models can be formulated as

$$\Lambda = \{\mathcal{M}_k : \ k \in [0, 1, ..., n], \ \{\delta_j\} \in \Lambda_k\} \tag{2.11}$$

and leads to a high dimensional inference task [68]. The general approach to this change point problem is the successive model comparison in order to infer on the model $\mathcal{M}_k$ adequately fitting the observations. Consequently, the final decision rule is not based on a test statistic but on a model selection process with respect to the employed asymptotic information criteria such as AIC and SIC. Another approach to achieve the change point inference is based on Bayesian inversion providing *inter alia* an exact information criteria, the Bayes factor. With respect to the major aspects of Bayesian statistics employed in this thesis, the framework of Bayesian inference is introduced in the following.

2.3 Bayesian inference

The general concept of Bayesian statistics consists of inverting conditional probabilities in order to derive probabilistic information about certain quantities of interest. An important application of this concept depicts the inference from given observations on an underlying process generating these observations. In Bayesian statistics, the observations are considered as given realizations of a random process. The quantities parameterizing the assumed underlying process are considered as random variables, i.e. associated with uncertainty indicated by subjective beliefs about possible parameter values. Thus, a (high dimensional) joint probability function of the assumed process model given the observations can be formulated and used to infer on the (low dimensional) probability functions of each parameter of interest individually. Even though various mathematicians have studied the employed concept based on the product rule of probability theory [69], it is an established custom to label this principle *Bayesian* inversion after Thomas Bayes [70]. Given the observations, the Bayesian inference provides an intuitive quantification of credibility of the assumed underlying model, referred to as *degree of belief*. By using the derived probabilistic information, the subjective belief about the model's parameters *prior* to further observations of the same process can be revised [71]. Hence, the Bayesian approach realizes a *concept of learning from the data* [69].

2.3.1 Bayes theorem

Let a probabilistic model $\mathcal{M}_{\boldsymbol{\vartheta}}$ describe a process of interest defined by the generic parameter vector $\boldsymbol{\vartheta} = (\vartheta_1, ..., \vartheta_p) \in \mathbb{R}^p$. To infer on the model parameters with respect

to the observations $\boldsymbol{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ of the process, a joint probability distribution $p(\boldsymbol{\vartheta}, \boldsymbol{y})$ may be formulated to model the uncertainty. In general the explicit formulation of the joint probability distribution is not feasible, but the joint expression can be decomposed into a product of

- a *prior* distribution $p(\boldsymbol{\vartheta})$, that encodes the knowledge about the parameters prior to any observation, and
- a *sampling* distribution $p(\boldsymbol{y}|\boldsymbol{\vartheta})$ representing the probability of the observations given the model, respectively its parameters.

By conditioning the unknown parameters $\boldsymbol{\vartheta}$ on the known observations $\boldsymbol{y}$ and by using the product rule of probability theory, the probability of the parameters given the observations can be formulated as

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\vartheta}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{\vartheta}) \cdot p(\boldsymbol{y}|\boldsymbol{\vartheta})}{p(\boldsymbol{y})} \,. \tag{2.12}$$

This equation represents the fundamental Bayes theorem guiding all Bayesian inference approaches. The probability of the observations $p(\boldsymbol{y})$ is a constant containing no further information about the underlying process. The complete information about the parameters $\boldsymbol{\vartheta}$ is therefore encoded in the *posterior* distribution

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}) = C \cdot p(\boldsymbol{\vartheta}) \cdot p(\boldsymbol{y}|\boldsymbol{\vartheta}), \tag{2.13}$$

where C serves as a normalization constant. The probabilistic information with respect to each model parameter ϑ_i can be derived from the joint posterior distribution. By analytical or numerical integration of the posterior distribution over all parameters but the one of interest, the according probability can be obtained as

$$p(\vartheta_{i=1}|\boldsymbol{y}) = C' \cdot \int ... \int p(\vartheta_1, ..., \vartheta_p) \cdot p(\boldsymbol{y}|\vartheta_1, ..., \vartheta_p) \, d\vartheta_2 ... d\vartheta_p \,. \tag{2.14}$$

This procedure of *marginalization* is applicable in the Bayesian framework due to the law of total probability. In *frequentistic* approaches, the formulation of a similar probability is usually not feasible since it is not part of the classical concept. The interpretation of the components of the Bayesian theorem and their potential influence on the posterior distribution are further explained in detail.

Likelihood function

The sampling distribution $p(\boldsymbol{y}|\boldsymbol{\vartheta})$ is determined by the assumed probability model $\mathcal{M}_{\boldsymbol{\vartheta}}$. The term can be regarded as a function $f(\boldsymbol{y}; \boldsymbol{\vartheta})$ of the generic parameter vector $\boldsymbol{\vartheta}$ given the observations $\boldsymbol{y}$ and poses the only influence of the observations on the posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{y})$. In this way, the sampling distribution equates the *likelihood function*

$\mathcal{L}(\boldsymbol{y};\boldsymbol{\vartheta})$ [72]. As a consequence, the Bayesian inference satisfies the likelihood principle. This principle states that the information about the parameters $\boldsymbol{\vartheta}$ derived by the observations $\boldsymbol{y}$ is completely contained in the likelihood function $\mathcal{L}(\boldsymbol{y};\boldsymbol{\vartheta})$. Additionally, let $\boldsymbol{y}$ and $\boldsymbol{y}'$ be different observations with respect to the same assumed model $\mathcal{M}_{\boldsymbol{\vartheta}}$, such that there exists a constant C satisfying

$$\mathcal{L}(\boldsymbol{y};\boldsymbol{\vartheta}) = C \cdot \mathcal{L}'(\boldsymbol{y}';\boldsymbol{\vartheta}) \tag{2.15}$$

for every parameter vector $\boldsymbol{\vartheta}$. The observations then contain the same information and must yield identical inferences on $\boldsymbol{\vartheta}$. Different to Bayesian approaches which obey the likelihood principle per definition, *frequentistic* approaches may violate this principle due to the dependence of the estimator on the methodological procedure employed [73].

Prior distribution

The prior distribution $p(\boldsymbol{\vartheta}) = p(\vartheta_1, ..., \vartheta_p)$ encodes the subjective belief about the model's parameters prior to any observation. In general, prior distributions aim to describe either knowledge or ignorance about parameters and are therefore specified as either *informative* or *non-informative*. However, as soon as the analyst formulates a prior distribution it contains information of some kind and, thus, is not truly non-informative anymore [74, 75]. The main - historical and recent - critics about the Bayesian concept address the subjective choice of the prior distribution. This argument is justified to a certain extent since there is no unique way of choosing a prior distribution, thereby affecting the resulting inference [73]. Instead of distinguishing between those restrictive prior categories, a conceptual overview on common prior specification and the basic assumptions leading to these prior choices are discussed in the following. The principles employed to construct specific prior distributions are explained with respect to the commonly used prior types and prior types referred to in this thesis. For the sake of convenience, the model's parameters $\boldsymbol{\vartheta}$ are assumed as independent from each other, leading to the factorization of the joint prior distribution into independent prior distributions for each individual parameter $p(\boldsymbol{\vartheta}) = p(\vartheta_1) \cdot ... \cdot (\vartheta_p)$. Thus, in the following the prior distribution is mainly considered as an univariate distribution $p(\vartheta)$.

Informative prior distributions describe elicited information from experts' opinion or previous analyses. *The aim is to let the prior knowledge influence the posterior distribution.* The simplest approach to generate an informative prior $p(\vartheta)$ is to restrict considerations on a feasible set of reasonable parameter values $\{\vartheta_i^*\}$ with $i = 1, ..., m$. Each value ϑ_i^* is assigned with a probability mass such that the relative contributions reflect the prior belief and that the probability mass sum to a constant. Alternatively, the parameter ϑ can be assumed to follow a parametric distributional family $f(\vartheta, \eta^*)$. The population parameter η^* is chosen such that the resulting distribution adequately maps the prior beliefs [76].

A further approach to generate informative prior distributions is commonly accomplished via Bayesian *hierarchical models*. In hierarchical models, the observations $\boldsymbol{y}$ are modeled conditionally on the model parameter ϑ, which itself is given a probabilistic specification via a further hyperparameter ϕ, which is not directly observable. Thus, the joint prior distribution is defined by

$$p(\phi, \vartheta) = p(\phi) \cdot p(\vartheta|\phi) \tag{2.16}$$

enclosing the hyperprior distribution $p(\phi)$. Consequently, the posterior distribution is reformulated with respect to the hyperparameter as

$$\begin{aligned} p(\phi, \vartheta|\boldsymbol{y}) &= C \cdot p(\phi, \vartheta) \cdot p(\boldsymbol{y}|\phi, \vartheta) \\ &= C \cdot p(\phi) \cdot p(\vartheta|\phi) \cdot p(\boldsymbol{y}|\vartheta), \end{aligned} \tag{2.17}$$

where $p(\boldsymbol{y}|\phi, \vartheta)$ can be simplified to $p(\boldsymbol{y}|\vartheta)$ due to the independence of the observations from the hyperparameter. The hyperprior can be estimated from elicited information from previous analyses or from the observations of the same analysis. For example, by computing the probability of the hyperparameter given the observations as

$$p(\phi|\boldsymbol{y}) = C \cdot \int p(\phi) \cdot p(\vartheta|\phi) \cdot p(\boldsymbol{y}|\vartheta) d\vartheta \tag{2.18}$$

an estimate $\hat{\phi}$ of the hyperparameter can be obtained and used to construct an empirical prior distribution as $p(\vartheta|\hat{\phi})$. This *empirical Bayes* approach is a simplification to practically generate informative prior distributions. However, since the prior distribution is constructed given the observations and not prior to any observation, it is not a strict Bayesian approach anymore. Therefore, strict Bayesian analysts commonly prefer the hierarchical Bayes approach without this simplification to construct informative prior distributions [72].

Weakly informative prior distributions are employed to prevent results that contradict the common knowledge or potential algorithm failure. *The aim is to use enough information to ensure regularization and stabilization of the computations, but no further information.* The common strategy to computationally benefit from the prior choice is the construction of a *conjugate prior* distribution [77]. The property of conjugacy is achieved in case the prior distribution $p(\vartheta)$ has the same parametric form as the posterior distribution $p(\vartheta|\boldsymbol{y})$. In this way the prior distribution is a conjugate family of probability distributions for the likelihood function $p(\boldsymbol{y}|\vartheta)$ and therefore closed under sampling [73]. By employing the corresponding likelihood function, the computation of the posterior distribution is basically a transparent update of the prior distribution and facilitates the computations considerably. For this reason, the prior distribution is often chosen to follow the same distributional family as the posterior distribution.

However, for all mentioned approaches to construct prior distributions encoding information the analyst always needs to ensure that the outcome does not contradict the prior belief. In case it does, the prior distribution needs to be revised accordingly.

Least informative prior distributions avoid the preference of any specific parameter value and are often referred to as non-informative. *The aim is "to let the data speak for themselves", with the intention that the prior has minimal influence on the posterior distribution.* A simple approach to encode as least information as possible realizes a *flat prior* distribution. By assigning every reasonable parameter value with equal probability no explicit value is preferred. This can be accomplished by a uniform distribution $\mathcal{U}(a, b)$, restricted to the interval $[a, b]$. The concrete restriction of the parameter space of prior distributions is mainly determined by the principal assumptions required to ensure the model's identifiability.

Another approach is the modification of the conjugate prior. By increasing the scale parameter of the conjugate prior distribution to a very high value, the distribution converges against a flat prior distribution and is called a *vague prior* distribution.

The general concern of constructing prior distributions is the dependency between the model parametrization and the representation of beliefs. In other words, a prior may be flat under one parametrization ϑ but this may not hold under a different parametrization ϕ. The invariance principle of Jeffreys [78] demands the equivalence of priors under parameter transformation and provides a further approach to derive least informative prior distributions. Note, in order to discuss the major aspects of the *Jeffreys prior*, the joint prior distribution $p(\boldsymbol{\vartheta}) = p(\vartheta_1) \cdot ... \cdot (\vartheta_p)$ is taken into account. Let a process be described by a model parameterized with the vector $\boldsymbol{\vartheta}$. The Fisher information matrix $\mathcal{J}(\boldsymbol{\vartheta})$ of the process, indicates the expected value of the observed information

$$\mathcal{J}_{ij}(\boldsymbol{\vartheta}) = -\mathbb{E}\left[\frac{\partial^2}{\partial\vartheta_i \partial\vartheta_j} ln\left(p(\boldsymbol{\vartheta}|\boldsymbol{y})\right)\right], \tag{2.19}$$

for the parameter vector $\boldsymbol{\vartheta}$ and the natural logarithm of the likelihood function $ln\left(p(\boldsymbol{\vartheta}|\boldsymbol{y})\right)$. The determinant of the Fisher information matrix is used to obtain a prior distribution of the model's parameters as

$$p(\boldsymbol{\vartheta}) \propto |\mathcal{J}(\boldsymbol{\vartheta})|^{\frac{1}{2}} . \tag{2.20}$$

It can be shown that the prior density yields an equivalent result,

$$p(\boldsymbol{\phi}) = p(\boldsymbol{\vartheta}) \left|\frac{d\boldsymbol{\vartheta}}{d\boldsymbol{\phi}}\right|, \tag{2.21}$$

if applied to a transformed representation of the model $\boldsymbol{\phi} = h(\boldsymbol{\vartheta})$ generated via a one-to-one transformation h as illustrated in Fig. 2.4. However, the computation of the Jeffreys' prior for models of multiple parameters is expensive. Furthermore, it is under debate whether it still holds the invariance [79]. Therefore, it is a common approach to assume the independence of the model's parameters and derive the Jeffreys' priors for each parameter individually, in particular for the scale and location parameter.

As a consequence of the flatness of the least informative prior distributions, the according prior distributions are often *improper*. That means that they do not integrate

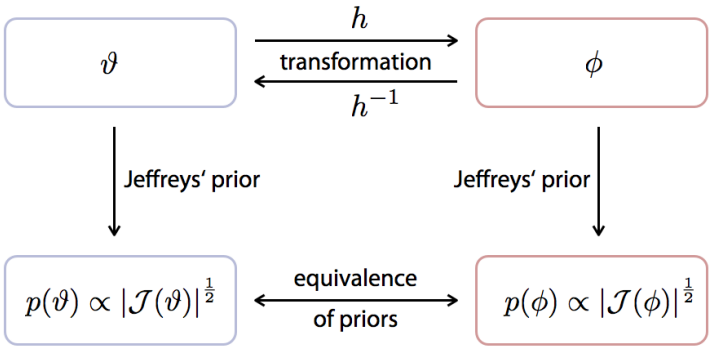


Figure 2.4: *A process may be modeled by the parameter ϑ. Under another parametrization the process may be represented by the transformed parameter $\phi = h(\vartheta)$. By computing the Jeffreys' prior distributions $p(\cdot)$ with the according Fisher information matrix $\mathcal{J}(\cdot)$, the resulting distributions represent the same beliefs, i.e. are equivalent under reparametrization (based on [67]).*

to a constant as required for every probability distribution. Nevertheless, in many settings the resulting posterior distribution remains proper due to the convergence of the likelihood function. Therefore, as soon as a least informative prior distribution is used the convergence of the posterior distribution has to be critically investigated.

2.3.2 Bayes factor

A fundamental feature of the Bayesian theorem is the formulation of an exact information criterion [80], in contrast to common asymptotic information criteria AIC and SIC (see Sec. 2.2). By using the Bayesian formula of Eq. 2.12, the probability of a scientific theory phrased as a statistical model $\mathcal{M}_i$ given the observations $\boldsymbol{y} \in \mathbb{R}^n$ can be derived. For this purpose, the sampling distribution $p(\boldsymbol{y}|\mathcal{M}_i)$ of the model is decomposed with respect to the model's parameters $\boldsymbol{\vartheta}_i \in \Theta_i$ as

$$\begin{aligned} p(\mathcal{M}_i|\boldsymbol{y}) &= C \cdot p(\boldsymbol{y}|\mathcal{M}_i) \cdot p(\mathcal{M}_i) \\ &= C \cdot \int_{\Theta_i} p(\boldsymbol{y}|\boldsymbol{\vartheta}_i, \mathcal{M}_i) \cdot p(\boldsymbol{\vartheta}_i|\mathcal{M}_i)\, d\boldsymbol{\vartheta}_i \cdot p(\mathcal{M}_i)\,, \end{aligned} \tag{2.22}$$

into an integral over the complete parameter space $\Theta_i \subseteq \mathbb{R}^{p_i}$. Under the model hypothesis $\mathcal{M}_i$ the parameters' likelihood function $p(\boldsymbol{y}|\boldsymbol{\vartheta}_i, \mathcal{M}_i)$ and prior distribution $p(\boldsymbol{\vartheta}_i|\mathcal{M}_i)$ are employed to compute the sampling distribution $p(\boldsymbol{y}|\mathcal{M}_i)$. This probability represents the evidence of the model provided by the observations irrespective of the assumed prior probability of the model $p(\mathcal{M}_i)$. By taking into account another statistical model $\mathcal{M}_j$ parametrized by $\boldsymbol{\vartheta}_j \in \Theta_j$ with $\Theta_j \subseteq \mathbb{R}^{p_j}$ and of different complexity p_j, the evidence is employed as a comparative measure of different model assumptions. Such a measure

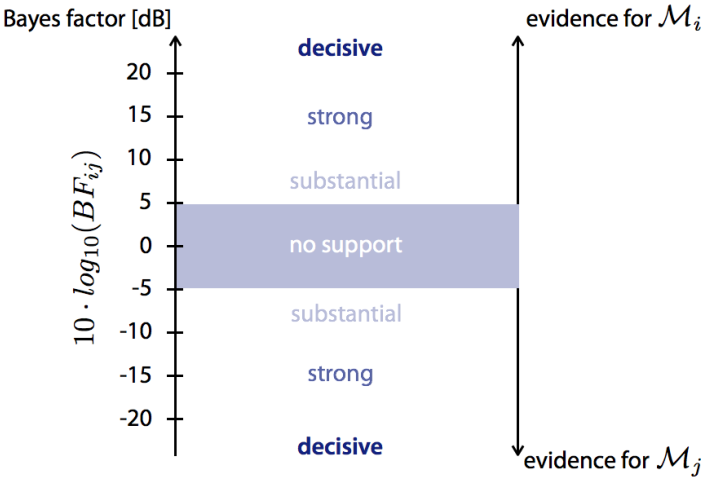


Figure 2.5: *The Bayes factor BF_{ij} is interpreted as a weight of evidence between two scientific theories in terms of stochastic models $\mathcal{M}_{i,j}$. Based on the value of the Bayes factor, the support for one of the assumed models is commonly categorized into substantial, strong and decisive evidence [67, 80]. Note, the Bayes factor is consistently presented in the unit deciban (dB) throughout this thesis.*

is accomplished by the ratio of the models evidences

$$BF_{ij} = \frac{p(\boldsymbol{y}|\mathcal{M}_i)}{p(\boldsymbol{y}|\mathcal{M}_j)} = \frac{\int_{\Theta_i} p(\boldsymbol{y}|\boldsymbol{\vartheta}_i, \mathcal{M}_i) \cdot p(\boldsymbol{\vartheta}_i|\mathcal{M}_i)\, d\boldsymbol{\vartheta}_i}{\int_{\Theta_j} p(\boldsymbol{y}|\boldsymbol{\vartheta}_j, \mathcal{M}_j) \cdot p(\boldsymbol{\vartheta}_j|\mathcal{M}_j)\, d\boldsymbol{\vartheta}_j}\,, \tag{2.23}$$

referred to as the Bayes factor BF_{ij}. As a convention, the Bayes factor is commonly defined with the more complex model in the denominator, i.e. $p_j > p_i$ [80]. The interpretation of the Bayes factor with respect to certain categories of evidence is based on [81] and is described in Fig. 2.5. An accomplishment of the Bayes factor with respect to ordinary hypothesis tests poses its ability to favor each assumed model, in contrast to common test capable to favor only the alternative hypothesis. In this way, the Bayes factor offers a descriptive measure for the evidence of competing statistical model assumptions.

2.3.3 Bayesian approaches to change point detection

As discussed in Sec. 2.2.2, the change point problem can be formulated as a model selection approach in the light of Bayesian inference. Under the assumption of a statistical model $\mathcal{M}$ or model family $\left\{\mathcal{M}^{(j)}\right\}$, the aim is to derive a degree of belief about the number k and locations $\boldsymbol{c} = (c_0, ..., c_k)$ of changes occurring in the observations. Note that all employed variables follow the definitions of Sec. 2.2.2. One of the first Bayesian change point problems was designed as a Markov process to detect random shifts in the model's parameters [82, 83]. Further Bayesian approaches were formulated for single or multiple change point models describing *inter alia* Markov models with time-varying

transition matrices [84, 85], Poisson processes with varying rate parameter [57, 86] and various linear regression models comprising parametric changes [12, 87].

With respect to multi change point problems, the inference task is high dimensional and requires sophisticated computational strategies. Even though the explicit formulation of the posterior distributions and applied computational algorithms vary, there are basic assumptions and techniques common to most approaches. These main concepts of Bayesian multi change point approaches are briefly outlined in the following.

Sampling distribution: product partition model

The product partition model yields a conditional independence property that facilitates the formulation of multi change point problems [88, 89]. For this purpose, let a sequence of observations $\boldsymbol{y}$ be partitioned into $k+1$ segments. The observations $\boldsymbol{y}_i$ of the ith segment δ_i are assumed as independent to the observations $\boldsymbol{y}_j$ with $j \neq i$ of the other segments. Within each segment δ_i the model is defined by the parameter vector $\boldsymbol{\vartheta}_i$. Thus, the segmentation $c_{0:k}$ determines the change point locations $\boldsymbol{c}$. Due to the independence between segments, the likelihood function can be formulated as the product

$$p(\boldsymbol{y}|k, c_{0:k}, \boldsymbol{\vartheta}_{0:k}) = \prod_{i=0}^{k} p(\boldsymbol{y}_i|\boldsymbol{\vartheta}_i)\,. \tag{2.24}$$

Consequently, the computation of the posterior distribution $p(k, c_{0:k}, \boldsymbol{\vartheta}_{0:k}|\boldsymbol{y})$ of the multi change point problem is considerably simplified. The product partition model is commonly employed as a multi change point model concept [68, 90].

Prior distributions: empirical Bayes estimator or intrinsic Bayes factor

Generally, in single change point problems least informative prior distribution $p(c)$ are used to describe the lack of knowledge about the change point location c. In multi change point problems, suitable prior distributions $p(k, c_{0:k})$ of the change specific variables are difficult to accomplish and often require the introduction of additional hyperparameters. Commonly, the joint prior distribution is considered to be of hierarchical nature, independently for each parameter [91] or as

$$p(k, c_{0:k}) = p(c_{0:k}|k) \cdot p(k)\,, \tag{2.25}$$

with the change point locations $c_{0:k}$ conditioned on the change point number k [87]. Based on the assumed model, the hierarchical approach employs different hyperparameters in order to derive adequate and consistent prior beliefs. By using subjective beliefs or empirical Bayes estimators [92], the prior distributions are constructed conditionally on the provided information or derived estimators, e.g. the number of changes $\hat{k}$. Alternatively, the SIC [68] or the intrinsic Bayes factor [93] can be derived from the posterior

distributions computed in the course of the model selection. Based on one of these information criteria, intrinsic prior distributions can be constructed [87, 94]. In general, however, no closed formulation of the intrinsic prior distributions can be accomplished and their computation is complicated, in particular with regard to complex statistical models.

Computation: reversible jump Markov chain Monte Carlo algorithm

In order to compute high dimensional probability distributions, as typical for Bayesian multi change point approaches, Markov chain Monte Carlo (MCMC) methods are applied. For a known number of change points k the inference algorithm is performed with MCMC techniques [95–97]. In case the number of changes k is unknown, a common technique offers the reversible jump MCMC (rjMCMC) algorithm [14]. Let $\{\mathcal{M}_k\}$ be a set of models where each member incorporates a different number of changes $k = 0, ..., m$. The rjMCMC algorithm is used to explore the joint space of models $\{\mathcal{M}_k\}$ and parameters $\{k, c_{0:k}, \boldsymbol{\vartheta}_{0:k}\}$. At each iteration step, the MCMC algorithm either

- deletes a change point, i.e. merges consecutive segments,
- adds a change point, i.e. splits consecutive segments,
- modifies the change point location,

whereby the decision criteria are derived from the posterior distributions obtained via Metropolis-Hasting algorithms. The main concern of the rjMCMC algorithm is the diagnosis of the convergence behavior. For example, a dynamic programming algorithm for exactly computing the posterior distribution $p(k, c_{1:k}|\boldsymbol{y})$ of the number k and positions $c_{1:k}$ of change points is proposed in Ref. [90].

In conclusion, most Bayesian multi change point approaches rely on the construction of complex prior distributions, to correctly formulate prior beliefs and to ensure stabilization and regularization of the algorithm. Another challenge is the appropriate formulation and efficient computation of high dimensional posterior distributions of the assumed models. The common computational strategies additionally require the diagnose of the numerical convergence of the algorithm against the formulated distributions. In contrast to these methods, the presented thesis introduces a Bayesian kernel-based approach employing a low dimensional, generic transition model to robustly infer on the change point location. By constructing a proxy expression to the multi change point posterior distribution a degree of belief about multiple changes is achieved, thereby avoiding complex prior distributions and expensive computational algorithms.

3 Detection of a single transition in time series

Unraveling changes in time series yields essential information about the system under observation. Such a transition may occur, and therefore be detected, in different domains such as statistical characteristics, frequency patterns or dynamical properties. For the purpose of developing an investigation approach capable to capture a variety of natural processes, generic features of time series are employed. The proposed formulation of a generic change point problem is thereby motivated from, but not restricted to complex climate observations. As a basic assumption the evolution of the statistical characteristics, in terms of central tendency and dispersion, are considered as generic features of time series comprising a transition. Both aspects are combined into a low-dimensional linear model with Gaussian noise, whereas heteroscedasticity gives rise to the non-linear part of the change point problem. Within the scope of climate research, the stochastic process generating the observations is not known with certainty and may lie outside the family of processes defined by the proposed generic change point model. In order to accept a generic approach appropriate to the investigation of a transition in climate observations, the robustness of the inference algorithm needs to be critically assessed in the presence of model errors and when applied to real world data.

As outlined in Chap. 2, the Bayesian inference offers a mathematical framework to derive an intuitive degree of belief about the transition location given an adequate statistical model. The beneficial model design and the choice of least informative prior distributions presented here, enable the separation of the Gaussian from the intrinsic non-linear part of the inference task. Besides clarifying the structure of the model, the approach speeds up the computational process considerably. As a result of the inference, the approach allows to derive the probability of the change point location over the time series. Moreover, by estimating the complete set of model parameters, the most likely shape of the underlying generic transition can be visualized. In practice, varying observational evidence, sparse data and outlier hamper the detection of transitions in time series. The performance and robustness of the proposed approach is therefore examined with respect to these difficulties. Finally, the introduced Bayesian method is applied to direct environmental observations comprising documented transition events: a hydrological time series of the Nile river in Egypt and a temperature series from the weather station in Tuscaloosa, Alabama. By investigating real world data, the suitability and limitation of the generic approach is critically discussed.

3.1 Formulation of the generic transition model

The design of the statistical model $\mathcal{M}$ introduced in the following aims to parametrize a generic transition and to facilitate the formulation of the model's sampling distribution, i.e. the likelihood function. Commonly, a transition is associated with a sudden change of local trend in the observations. This may be induced by a transition between regimes governed by different internal dynamics or external influences. Let a time series $\{y(t_i)\} = \{y_i\}$ be acquired at discrete time points t_i for $i = 1, ..., n$ and comprise a change at time θ. At least locally, the time series can be described by a first order approach $\mathcal{O}(1)$ prior and after the change point. The assumption essentially corresponds to the approximation of the signal with the Taylor series expansion of order $\mathcal{O}(2)$. In this way, such a model depicts a specific subclass of models locally parameterizing the time series. For this purpose, a set of basis functions is used to model a transition in terms of a continuous *break* or a discontinuous *shift* as illustrated in Fig. 3.1. The employed piecewise linear basis functions [98, 99], also referred to as ramp functions, simulating a break transition are defined as

$$\left(\zeta_-^\theta\right)_i = \zeta_-^\theta(t_i) = \begin{cases} \theta - t_i \text{ if } t_i \leq \theta, \\ 0 \quad \text{else,} \end{cases} \quad \text{and} \quad \left(\zeta_+^\theta\right)_i = \zeta_+^\theta(t_i) = \begin{cases} t_i - \theta \text{ if } t_i \geq \theta, \\ 0 \quad \text{else.} \end{cases} \tag{3.1}$$

By combining these ramp functions with piecewise constant basis functions, also referred to as Heaviside functions, defined as

$$\left(\varphi_-^\theta\right)_i = \varphi_-^\theta(t_i) = \begin{cases} 1 \text{ if } t_i \leq \theta, \\ 0 \quad \text{else,} \end{cases} \quad \text{and} \quad \left(\varphi_+^\theta\right)_i = \varphi_+^\theta(t_i) = \begin{cases} 1 \text{ if } t_i \geq \theta, \\ 0 \quad \text{else,} \end{cases} \tag{3.2}$$

a shift transition is modeled in form of a break and horizontal offset. Consequently, the basic types of signal $y(t_i)$ at time t_i undergoing a change at time θ can be expressed as

$$\mathcal{M}_{break}: \quad y(t_i) = \beta_0 + \beta_1 \cdot \zeta_-^\theta(t_i) + \beta_2 \cdot \zeta_+^\theta(t_i) + \varepsilon(t_i), \tag{3.3}$$

$$\mathcal{M}_{shift}: \quad y(t_i) = \beta_0 \cdot \varphi_-^\theta(t_i) + \beta_1 \cdot \zeta_-^\theta(t_i) + \beta_2 \cdot \varphi_+^\theta(t_i) + \beta_3 \cdot \zeta_+^\theta(t_i) + \varepsilon(t_i), \tag{3.4}$$

with the parameters β_j determining the signal's mean behavior.

However, natural observations can in general not be modeled by such simple mean behavior as given by these functions. Therefore, some random fluctuations $\varepsilon(t_i)$ are added around the mean signal. These random fluctuations may be due to measurement noise as well as to some intrinsic variability, which is not captured by the low dimensional mean dynamics on both sides of the change point θ. For this fluctuating part of the signal it is supposed that the intensity is constant around the transition. The intrinsic variability may - like the mean behavior of the system itself - undergo a sudden change in its amplitude. Hence, the intensity of the stochastic fluctuations $\varepsilon(t_i)$ at time t_i is considered to comprise a break transition at time θ according to

$$\mathrm{STD}\left[\varepsilon(t_i)\right] = \sigma \cdot \left(1 + s_1 \cdot \zeta_-^\theta(t_i) + s_2 \cdot \zeta_+^\theta(t_i)\right) . \tag{3.5}$$

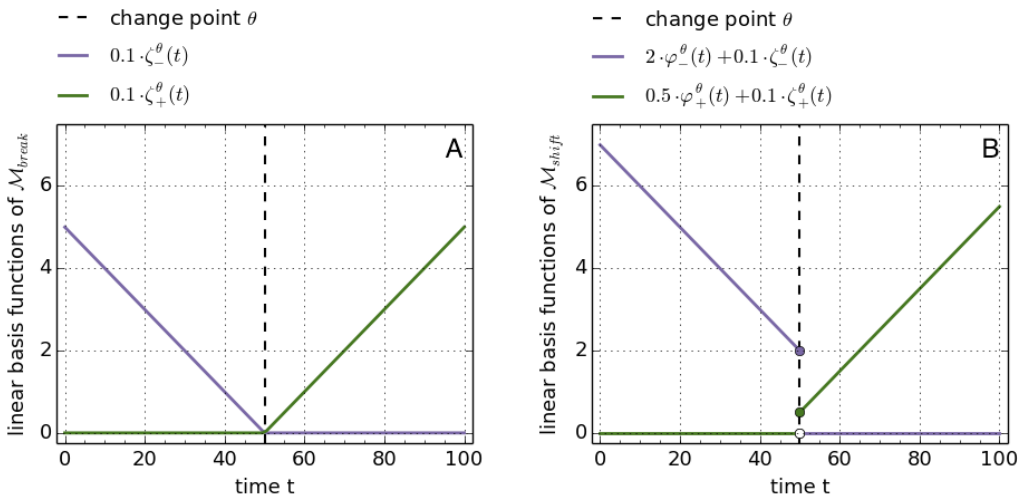


Figure 3.1: *The piecewise linear basis functions $\zeta^{\theta}_{\pm}$ and $\varphi^{\theta}_{\pm}$ are used to parametrize generic transition events around the change point θ in terms of a break defined in the transition model $\mathcal{M}_{break}$ (A) and in terms of a shift defined in the transition model $\mathcal{M}_{shift}$ (B).*

The scale factor σ may be the level of the measurement noise or some background level of the intrinsic fluctuations, whereas the parameters $s_{1,2}$ describe the systematic evolution of the model's intrinsic variability prior and after the change point measured in units of σ. Although the fluctuating component may naturally contain coherent parts, it is assumed that the fluctuations are Gaussian random variables, which are uncorrelated at different time points,

$$\mathbb{E}\left[\varepsilon(t_i)\varepsilon(t_j)\right] = 0\,, \quad \forall\, t_i \neq t_j\,. \tag{3.6}$$

Clearly, this is an approximation and its validity has to be questioned in concrete applications.[1] This assumption, however, allows the implementation of highly efficient algorithms for the inference on the model parameters. In order to minimize the number of parameters that can only be estimated by intensive numerical computations, the noise $\varepsilon(t_i)$ is assumed to undergo a transition solely in form of a break. Thus, the name of the statistical model $\mathcal{M}_j$ is induced by the specific approach $j = [\text{break}, \text{shift}]$ to the mean behavior, whereas the noise term may simultaneously undergo a break transition. Synthetic observations derived from simulations of the introduced models $\mathcal{M}_{break}$ and $\mathcal{M}_{shift}$ are presented in Fig. 3.2. The proposed models are by design capable to describe all functions locally differentiable up to the first order prior and after the change point and therefore, may be interpreted as generic transition models approximately describing a variety of natural signals.

[1] The assumption is particularly insufficient for climate time series with autocorrelation patterns [33]. The approximations used to design the generic transition models are discussed in the course of the application to real climate observations in Sec. 3.4.

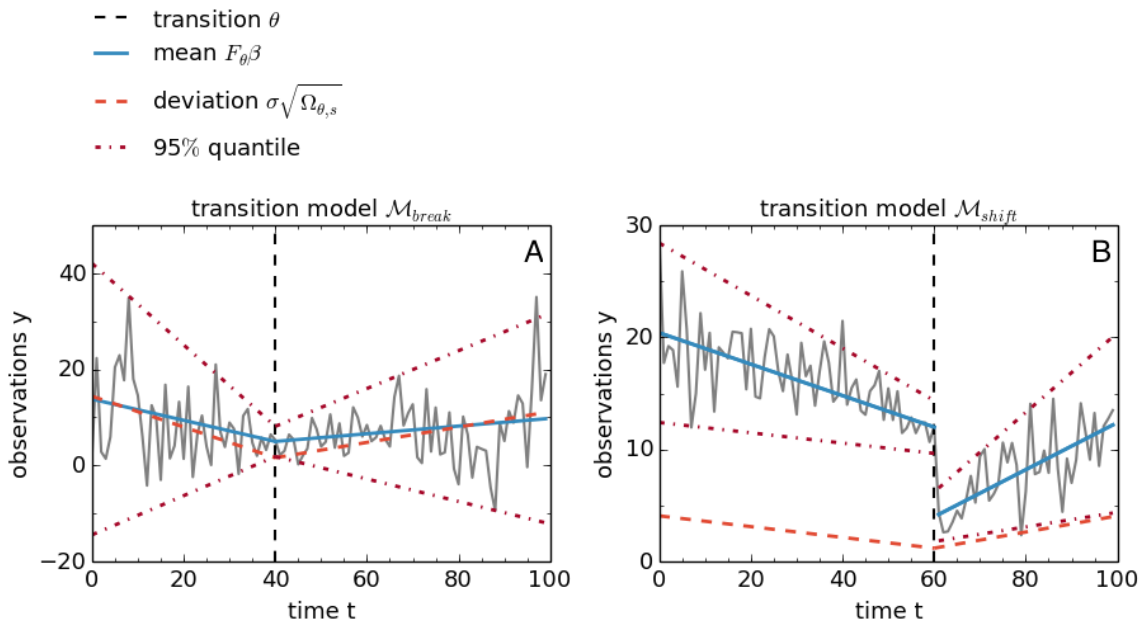


Figure 3.2: *Based on the generic transition models $\mathcal{M}_{break}$ (A) and $\mathcal{M}_{shift}$ (B) synthetic observations are generated that comprise a break or shift in mean and simultaneously a break in the deviation. The 95% quantiles of the random fluctuations illustrate the simulated variability and heteroscedasticity, respectively.*

Linear model approach

For convenience of the computations in the course of the Bayesian inversion, the generic transition models $\mathcal{M}_j$ with $j = [\text{break}, \text{shift}]$ are formulated in terms of a linear model [100]. At least locally, the models are supposed to be approximations to the observation vector $\boldsymbol{y} = [y(t_i)]^T \in \mathbb{R}^n$ obtained at the time points $t_i \in \mathcal{T} = [t_1, ..., t_n]$. Hence, the generic models may sufficiently describe the observations only within some sub-interval $\mathcal{T}' = [t'_1, ..., t'_{n'}]$ with $t_1 < t'_1$ and $t'_{n'} < t_n$ of the considered time series. The observations $\boldsymbol{y}_{|\mathcal{T}'} = [y(t'_i)]^T \in \mathbb{R}^{n'}$ enclosed by this sub-interval are referred to as the *local* observation vector and may be expressed as a linear model

$$\mathcal{M}_j : \ \boldsymbol{y}_{|\mathcal{T}'} = F_\theta^{(j)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\theta \quad \text{with} \ \ j = [\text{break}, \text{shift}] \tag{3.7}$$

as the sum of the *fixed effects* $F_\theta^{(j)}\boldsymbol{\beta}$ and the *random effect* $\boldsymbol{\varepsilon}_\theta$. Here, the index θ aims to highlight the model components dependent on the change point parameter. Due to the assumption of a local approach, the transition parameter $\theta \in \mathcal{T}'$ is explicitly defined within the sub-interval of the local observations $\boldsymbol{y}_{|\mathcal{T}'}$.

The distinction between the generic transition models $\mathcal{M}_j$ is given by the mean evolution, i.e. in the linear model formulation by the fixed effects. The fixed effect vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \in \mathbb{R}^p$ with $p = 3$ corresponds to the coefficients of the linear combination of the basis functions modeling the mean behavior of the transition model $\mathcal{M}_{break}$.

The according system matrix $F_\theta^{(break)} \in \mathbb{R}^{n' \times 3}$ is given by the sampling of the basis functions $\boldsymbol{\zeta}_\pm^\theta$ defined in Eq. 3.1 at the observation points t_i'

$$F_\theta^{(break)} = \begin{pmatrix} 1 & \left(\zeta_-^\theta\right)_1 & \left(\zeta_+^\theta\right)_1 \\ \vdots & \vdots & \vdots \\ 1 & \left(\zeta_-^\theta\right)_{n'} & \left(\zeta_+^\theta\right)_{n'} \end{pmatrix} . \tag{3.8}$$

By additionally using the piecewise constant basis functions $\boldsymbol{\varphi}_\pm^\theta$ from Eq. 3.2 at the observation points t_i' the system matrix $F_\theta^{(shift)} \in \mathbb{R}^{n' \times p}$ with $p = 4$ can be formulated as

$$F_\theta^{(shift)} = \begin{pmatrix} \left(\varphi_-^\theta\right)_1 & \left(\zeta_-^\theta\right)_1 & \left(\zeta_+^\theta\right)_1 & \left(\varphi_+^\theta\right)_1 \\ \vdots & \vdots & \vdots & \vdots \\ \left(\varphi_-^\theta\right)_{n'} & \left(\zeta_-^\theta\right)_{n'} & \left(\zeta_+^\theta\right)_{n'} & \left(\varphi_+^\theta\right)_{n'} \end{pmatrix} , \tag{3.9}$$

with the coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T \in \mathbb{R}^4$.

The random effect $\boldsymbol{\varepsilon}_\theta = [\varepsilon(t_i')]^T \in \mathbb{R}^{n'}$ is defined as a Gaussian random vector with zero mean, scale parameter $\sigma \in \mathbb{R}^+$ and covariance matrix $\Omega_\theta \in \mathbb{R}^{n' \times n'}$, that is

$$\boldsymbol{\varepsilon}_\theta \sim \mathcal{N}\left(0, \sigma^2 \Omega_\theta\right) . \tag{3.10}$$

The covariance itself is structured noise, which is parametrized by the deviation's slope parameters $\boldsymbol{s} = (s_1, s_2) \in \Theta_{s_1,s_2}$ prior and after the transition. Due to the assumption of uncorrelated noise, the resulting covariance matrix is of diagonal form and is given by

$$\left(\Omega_\theta\right)_{ij} = \left(\left[1 + s_1 \cdot \left(\zeta_-^\theta\right)_i + s_2 \cdot \left(\zeta_+^\theta\right)_j \right]^2 \right) \cdot \delta_{ij} \tag{3.11}$$

using the Kronecker delta δ_{ij}. The generic transition model is well defined and identifiable for $\boldsymbol{s} \in (\mathbb{R}^+)^2$. However, the parametrization becomes ambiguous for any $s_i < 0$, since different parameter values $\pm s_i$ may give rise to the same values of the likelihood function. Moreover, for $\boldsymbol{s} \in (\mathbb{R}^-)^2$ the deviation may become negative prior and after the transition and may be falsely interpreted as two additional changes in the variability of the signal. As a consequence, the transition parameter θ is not identifiable anymore. By adequately restricting the parameter space Θ_{s_1,s_2} of the deviation parameters the identifiability of the transition parameter can be accomplished. A negative deviation, i.e. a falsely interpreted change in the variability, is avoided in case each diagonal element of the covariance matrix holds

$$\sqrt{\left(\Omega_\theta\right)_{ij}} \cdot \delta_{ij} = 1 + s_1 \cdot \left(\zeta_-^\theta\right)_i + s_2 \cdot \left(\zeta_+^\theta\right)_i > 0 \quad \forall\ i = 1, ..., n' . \tag{3.12}$$

As an important side effect, the covariance remains non-singular and therefore allows the application of the QR decomposition for the practical computation of the likelihood

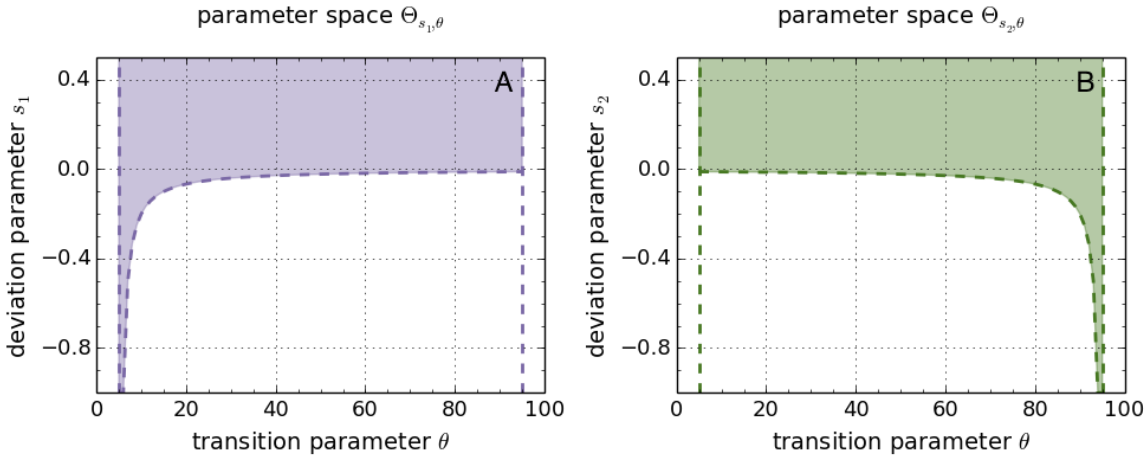


Figure 3.3: *To ensure the identifiability of the transition parameter θ the parameter spaces $\Theta_{s_1,\theta}$ (A) and $\Theta_{s_2,\theta}$ (B) of the slope parameters need to be restricted. The shaded areas indicate the parameter spaces as given by Eq. 3.13 and Eq. 3.14. The dashed lines are not part of the parameter space $\Theta_{s_1,s_2,\theta}$, such that the parameter space is the open inside of the polygon illustrated by the dashed lines.*

function, as explained in the next section. In this way, the parameter spaces of the deviation parameters s_1 and s_2 yield a lower limit dependent on the transition parameter θ and an open upper limit:

$$\begin{aligned} \Theta_{s_1} &= \left]\, -\frac{1}{\theta - t'_1}\,,\, \infty \right[\;=\; [\, s_{1,min}(\theta)\,,\, \infty[\,, \\ \Theta_{s_2} &= \left]\, -\frac{1}{t'_{n'} - \theta}\,,\, \infty \right[\;=\; [\, s_{2,min}(\theta)\,,\, \infty[\,. \end{aligned} \tag{3.13}$$

The parameter space of the change point parameter has to be restricted accordingly, such that $\theta_{min} > t'_1$ and $\theta_{max} < t'_{n'}$ is ensured by

$$\Theta_\theta = \,]\, t'_1\,,\, t'_{n'}\,[\;=\; [\, \theta_{min}\,,\, \theta_{max}\,]\,, \tag{3.14}$$

in order to avoid a division by zero in the lower limits of Θ_{s_1,s_2}. Hence, to reliably identify a generic transition, the joint parameter space $\Theta_{\theta,\boldsymbol{s}}$ of the change point and deviation parameters is defined as the open inside of the polygon given by the lower limits of $\Theta_{\boldsymbol{s}}$ and the limits of Θ_θ as illustrated in Fig. 3.3.

In conclusion, the generic transition model $\mathcal{M}_j$ with $j = [\text{break}, \text{shift}]$ is determined by the joint parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta)$ and defined within the according parameter spaces $\Theta_{\boldsymbol{\vartheta}}$ to provide an estimable model approach. Based on the formulation as a linear model, the probability distribution of the local observations $\boldsymbol{y}_{|\mathcal{T}'}$ for fixed parameters $\boldsymbol{\vartheta}$ is given by

$$\boldsymbol{y}_{|\mathcal{T}'} \sim \mathcal{N}\left(F_\theta^{(j)}\boldsymbol{\beta}, \sigma^2\Omega_\theta\right)\,, \tag{3.15}$$

and represents the statistical model of a generic transition event. The probability distribution of the local observations is defined with respect to the parametric model $\mathcal{M}_j$. Nevertheless, the formulation is independent of the explicit transition type $j = [\text{break}, \text{shift}]$. For the sake of clarity, the distinction between the transition types is only referred to if necessary in the following.

Likelihood function

The likelihood function can be formulated as the probability of the local observations $\boldsymbol{y}_{|\mathcal{T}'}$ under the generic transition model $\mathcal{M}$ in form of

$$\mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}\sqrt{|\Omega_\theta|}} \cdot e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta})^T\Omega_\theta^{-1}(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta})} . \tag{3.16}$$

Thereby, the functional dependency of the fixed effects $\boldsymbol{\beta} \in \mathbb{R}^{p_j}$ is a Gaussian density. The dimension of the fixed effects p_j is determined by the transition type j, i.e. $p_{break} = 3$ and $p_{shift} = 4$. Clearly, in the exponent $\boldsymbol{\beta}$ is of a quadratic form and since $\Xi = F_\theta^T\Omega_\theta^{-1}F_\theta$ is positive definite the exponent can be rewritten as

$$\mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}\sqrt{|\Omega_\theta|}} \cdot e^{-\frac{\mathcal{R}^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T\Xi(\boldsymbol{\beta} - \boldsymbol{\beta}^*)} , \tag{3.17}$$

where the mode of the Gaussian in $\boldsymbol{\beta}$ is the best linear unbiased predictor (BLUP) [101] of the fixed effects derived as

$$\begin{aligned} \boldsymbol{\beta}^* &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p_j}}{\operatorname{argmin}} (\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta})^T\Omega_\theta^{-1}(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta}) \\ &= (F_\theta^T\Omega_\theta^{-1}F_\theta)^{-1}F_\theta^T\Omega_\theta^{-1}\boldsymbol{y}_{|\mathcal{T}'} . \end{aligned} \tag{3.18}$$

In the course of the practical computation of the likelihood function, the QR decomposition is performed to obtain *inter alia* the BLUP numerically more efficient. To derive an unique QR factorization the system matrix F_θ is required to be of full rank $rank(F_\theta) = n'$. That is true in case the transition parameter lies within the sub-series of the local observations $\boldsymbol{y}_{|\mathcal{T}'}$, as already ensured by its parameter space Θ_θ. The application of the QR decomposition is explained in App. A.

The BLUP is used to reformulate the likelihood function and thus to facilitate the analytical integration over the fixed effects $\boldsymbol{\beta}$ in the course of marginalization of the joint posterior distribution. A further useful reformulation is achieved by defining the residuum $\mathcal{R}$ measured in the Mahalanobis distance [102, 103] and induced by the covariance matrix Ω_θ, as

$$\begin{aligned} \mathcal{R}^2 &= \min_{\boldsymbol{\beta}\in\mathbb{R}^{p_j}} (\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta})^T\Omega_\theta^{-1}(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta}) \\ &= (\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta}^*)^T\Omega_\theta^{-1}(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta\boldsymbol{\beta}^*) . \end{aligned} \tag{3.19}$$

Moreover, the BLUP is employed to obtain the profiled likelihood function $\mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}^*, \sigma, \boldsymbol{s}, \theta)$ and in this way the profiled likelihood estimator (PLH) of the scale parameter as

$$\frac{\partial \mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}^*, \sigma, \boldsymbol{s}, \theta)}{\partial \sigma} \stackrel{!}{=} 0 \quad \Rightarrow \quad \left(\sigma^2\right)^* = \frac{\mathcal{R}^2}{n - p_j} \,. \tag{3.20}$$

The estimator of the scale parameter σ^* is *inter alia* beneficial for the computation of the maximum of the likelihood function employed in the kernel-based extensions of the detection algorithm introduced in the Chap. 4.

3.2 Implementation of the Bayesian inversion

In the light of the Bayesian theorem (Eq. 2.13), the probability distribution of the transition model's parameters $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta)$ given the local observations $\boldsymbol{y}_{|\mathcal{T}'}$ can be formulated. For this purpose, the explicit likelihood function $\mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\vartheta})$ under the assumed transition model $\mathcal{M}$ as well as the prior distribution of the parameters $p(\boldsymbol{\vartheta})$ need to be specified to compute the posterior distribution of interest as

$$p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'}) \propto \mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) \cdot p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) \,. \tag{3.21}$$

The likelihood function and useful reformulations thereof are provided in the previous section. The prior distribution, however, needs to be specified to adequately encode the belief about each model parameter prior to any observation. Usually for change point problems, there is no *a priori* information about the model's parameters $\boldsymbol{\vartheta}$ available. The assumptions made so far in the design of the generic transition model $\mathcal{M}$ enclose the regression model (partly linear) and the noise distribution (Gaussian). The intention is to formulate the posterior distribution as general as possible by including no subjective assumptions on the model's parameter. But since the accurate use of the term *objective Bayesian analysis* is under debate [69, 75], it is reasonable to speak of an *approach transparent in all employed hypotheses*. By specification of the prior distributions, the joint posterior distribution can be factorized into a family of parametrized Gaussians. The factorization favorably mirrors the separation of the linear from the non-linear parts and thus, facilitates the explicit computation of the marginal distributions considerably.

Prior distributions

A common strategy to construct least informative prior distributions offers the Jeffreys prior defined in Eq. 2.20. Besides the transformation invariance, the Jeffreys prior is not affected by restrictions of the parameter space. Nevertheless, for high dimensional models the Jeffreys prior is cumbersome to compute and leads to complex prior expressions (see for example App. A). The performance of the Jeffreys prior for high dimensional models is even controversial [72, 104]. Alternatively, a practical approach is to assume

a priori no correlations between the parameters, in order to factorize the joint prior distribution into independent parts

$$p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) = p(\boldsymbol{\beta}) \cdot p(\sigma) \cdot p(\boldsymbol{s}, \theta)\,, \tag{3.22}$$

and to assign to each parameter a prior distribution individually [76]. The specification of a prior distributions needs to be considered always in tandem with dependencies between different parameters and with restrictions on the corresponding parameter spaces. In particular for the location and scale parameter of the Gaussian sampling distribution, i.e. $\boldsymbol{\beta}$ and σ, the according Jeffreys priors [78, 104] are commonly used and are derived as

$$p(\boldsymbol{\beta}) \;\sim\; 1 \;\text{ with }\; \boldsymbol{\beta} \in \mathbb{R}^{p_j}, \tag{3.23}$$

$$p(\sigma) \;\sim\; \frac{1}{\sigma} \;\text{ with }\; \sigma \in \mathbb{R}^{+}\,. \tag{3.24}$$

The computation of the Jeffreys priors of the transition parameter θ and deviation parameters $\boldsymbol{s}$ is not feasible. Another least informative prior function is a flat prior distribution such that every parameter may uniformly take every value within the joint parameter space,

$$p(\boldsymbol{s}, \theta) = p(\boldsymbol{s}|\theta) \cdot p(\theta) \sim 1 \;\text{ with }\; \begin{cases} \theta & \in \Theta_\theta = [\,\theta_{min}\,,\,\theta_{max}\,]\,, \\ \boldsymbol{s} & \in \Theta_{\boldsymbol{s}} = [\,\boldsymbol{s}_{min}(\theta)\,,\,\infty[\,. \end{cases} \tag{3.25}$$

The resulting prior is flat in the open inside of the polygon defined by the constraints of the parameter space $\Theta_{\boldsymbol{s},\theta}$ in Eq. 3.13 and 3.14. Note, that the decomposition of the prior distribution $p(\boldsymbol{s}, \theta)$ allows the kernel-based reformulation of the inference approach in Chap. 4. Even though these prior distributions are improper, the resulting posterior distribution can be numerically normalized and is convergent due to the shape of the likelihood function and the restriction of the prior distribution of the transition parameter.

Joint and marginal posterior distributions

Based on the specified likelihood function (Eq. 3.17) and the prior distributions (Eq. 3.23 - 3.25), the joint posterior distribution of the transition model's parameters given the local observations $\boldsymbol{y}_{|\mathcal{T}'}$ can be explicitly formulated as

$$p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}'}) = C \cdot \frac{1}{\sigma} \cdot \mathcal{L}(\boldsymbol{y}_{|\mathcal{T}'}; \boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta)\,. \tag{3.26}$$

The normalization constant C ensures that the right hand side actually defines a normalized probability distribution. The main purpose of the inference approach depicts the marginal posterior distribution of the transition parameter $p(\theta|\boldsymbol{y}_{|\mathcal{T}'})$ by integrating the

joint posterior distribution over all but the parameter of interest. Since each marginal distribution may be useful for the estimation of the underlying transition model, all distributions are provided in the following. It turns out, that only the integrals over the transition parameter θ and the deviation parameters $\boldsymbol{s}$ can not be carried out analytically. Due to the Gaussian nature of the $\boldsymbol{\beta}$ dependency the marginal distribution of the scale parameter σ,

$$p(\sigma, \boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'}) \propto \frac{\sigma^{p_j - n - 1}}{\sqrt{|\Omega_\theta||F_\theta^T \Omega_\theta^{-1} F_\theta|}} e^{-\frac{1}{2\sigma^2}\mathcal{R}^2}, \tag{3.27}$$

and the marginal distribution of the fixed effect parameter $\boldsymbol{\beta}$,

$$p(\boldsymbol{\beta}, \boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'}) \propto \frac{\left[(\boldsymbol{y}_{|\mathcal{T}'} - F_\theta \boldsymbol{\beta})^T \Omega_\theta^{-1} (\boldsymbol{y}_{|\mathcal{T}'} - F_\theta \boldsymbol{\beta})\right]^{-\frac{n}{2}}}{\sqrt{|\Omega_\theta|}}, \tag{3.28}$$

can be analytically computed. Further marginalization is performed to yield

$$p(\boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'}) \propto \frac{\mathcal{R}^{-(n-p_j)}}{\sqrt{|\Omega_\theta||F_\theta^T \Omega_\theta^{-1} F_\theta|}}\,. \tag{3.29}$$

Based on this marginal, the posterior distribution of the transition parameter θ can be computed by numerical evaluation of the integral

$$p(\theta | \boldsymbol{y}_{|\mathcal{T}'}) = C' \cdot \int_{\boldsymbol{s}_{min}(\theta)}^{\infty} d\boldsymbol{s}\; p(\boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'})\,, \tag{3.30}$$

over the parameter space $\Theta_{\boldsymbol{s}}$. In the same way, the posterior distribution of the deviation parameters $\boldsymbol{s}$ can be obtained by numerical evaluation of the integral

$$p(\boldsymbol{s} | \boldsymbol{y}_{|\mathcal{T}'}) = C'' \cdot \int_{\theta_{min}}^{\theta_{max}} d\theta\; p(\boldsymbol{s}, \theta | \boldsymbol{y}_{|\mathcal{T}'}) \tag{3.31}$$

over the parameter space Θ_θ. The normalization constants C' and C'' ensure that the derived expressions are normalized probability distributions. Each marginal distribution provides a degree of belief about the enclosed parameter given the local observations. Moreover, the distributions may be used to estimate the corresponding parameters and thus obtain an estimate of the complete generic transition with respect to the local observations as described in the following.

3.3 Performance of the detection approach

By assuming the generic transition model as an adequate approximation to a time series of interest, the Bayesian inversion may be used as a change point detection approach. To explain the application and the technical aspects of the algorithm, a set of synthetic time series is analyzed. For this purpose, temporal equidistant observations are generated for different change geometries of the transition model $\mathcal{M}_{break}$, that are interpreted as changes of different observational evidence. Besides the investigation of potential dependencies between the change geometry and the inference performance, different estimator approaches are compared to validate the balance between the accuracy and the computational costs of the inference. Given the set of optimal estimators, the sensitivity of the approach to data loss is studied in order to evaluate the minimal number of observations required for a reliable inference on the underlying transition. As a generic approach to complex time series, the method needs to be robust with respect to model errors. On the one hand, model errors may be due to an insufficient distribution assumption of the noise term, resulting in outliers. On the other hand, model errors may be due to process patterns that are too complex to be approximated by the low dimensional mean dynamics, resulting in divergent higher statistical moments. The robustness to outlier is evaluated for synthetic observations. However, the investigation of the higher statistical moments is done for real world observations in Sec. 3.4. The critical assessment of the inference performance aims to demonstrate the robustness of the proposed generic change point approach in order to justify its expansion to a multiple change point approach, as introduced in Chap. 4.

3.3.1 Transitions of different observational evidence

Depending on the change geometry, the transitions are more or less significant within the observations and thus, may be interpreted as changes of different observational evidence. For the exploratory analysis of real world time series, the influence of observational evidence on the efficiency of the applied inference approach depicts a crucial concern and needs to be understood in order to appropriately interpret the analysis results. Therefore, the main modifications of the generic transition model $\mathcal{M}_{break}$ are simulated to study the inference performance in the light of varying observational evidence. The introduced Bayesian inversion is applied to synthetic time series of $n = 100$ temporally equidistant observations $\boldsymbol{y}$ at the time points $\mathcal{T} = [0, ..., 99]$ generated for different geometries of the underlying change as presented in Fig. 3.4 and defined by the models' true parameter setting $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta)$ provided in Tab. C.1 (shaded columns). For clarity, the modifications may be grouped based on similar transition patterns in the properties subject to the change, i.e. mean and deviation:

- **uncorrelated or correlated transitions**, that are changes in one or simultaneously in both changing properties, and

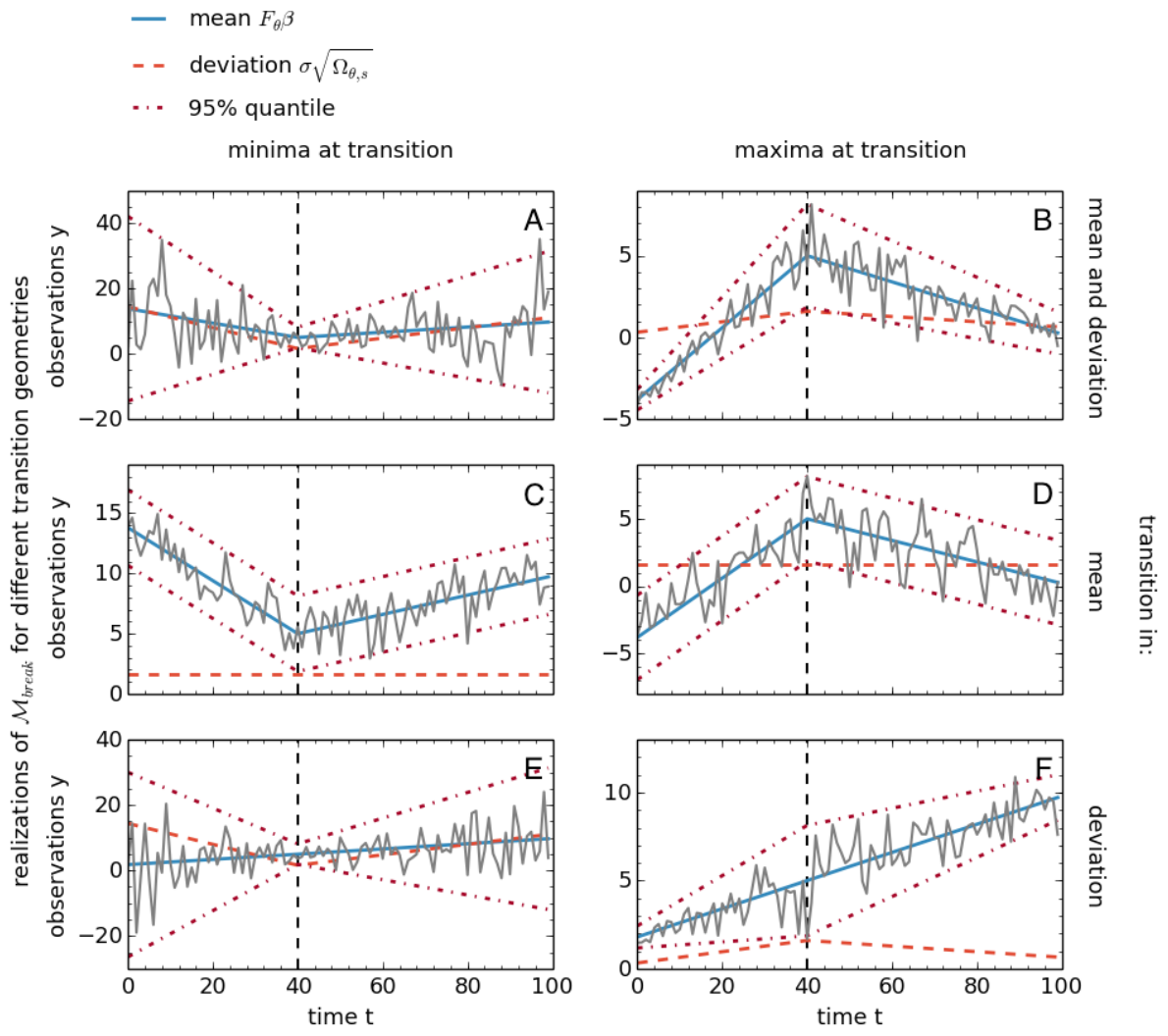


Figure 3.4: *To systematically investigate the inference performance with respect to transitions of different observational evidence, artificial observations with different change geometries are generated. Each true parameter setting of the underlying generic transition model $\mathcal{M}_{break}$ is listed in Tab. C.1. Thereby, the geometries are ordered based on the patterns of the changing properties mean and deviation. On the one hand, the changing properties may be correlated ((A), (B)) or uncorrelated ((C) - (F)). On the other hand, the evolution of the properties may undergo a minimum ((A), (C), (E)) or a maximum at the transition ((B), (D), (F)).*

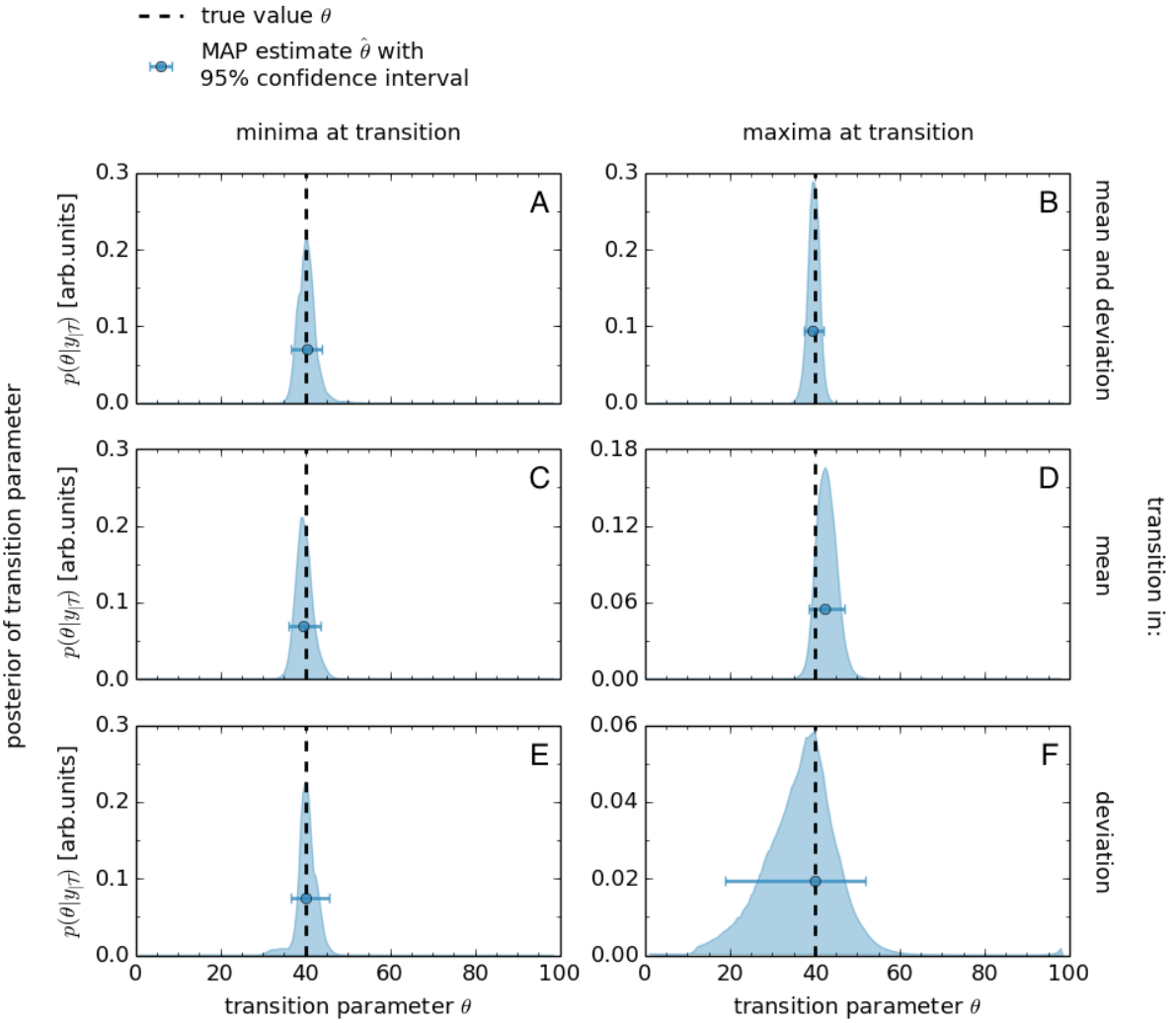


Figure 3.5: *Each marginal posterior distribution $p(\theta|\boldsymbol{y}_\mathcal{T})$ of the transition parameter θ indicates the degree of belief about at what times a generic change occurs in the corresponding synthetic observations $\boldsymbol{y}_\mathcal{T}$ presented in Fig. 3.4. The MAP estimate is considered as the point estimator $\hat{\theta}$ of the transition parameter while the confidence interval $CI_{0.95}$ marks the time period, in which the event occurs with 95% certainty. Even though the simulated transitions are of different observational evidence, the mode of the distributions is close to the true value θ and all but one indicate a change in approximately the same time interval $[35, 50]$ ((A) - (E)). Only for the transition realized as an uncorrelated deviation maximum (F), the approach yields a weak inference result in terms of a considerably broader posterior distribution.*

- **minimum or maximum** in the evolution of each changing property **at the transition**.

In other words, a transition is not only characterized by *what* property is changing, but also by *how* a changing property is evolving at the singularity. Naturally, the evolution of the deviation and the variability of the observations, respectively, is expected to have an influence on the observational evidence of the transition. Irrespective of a correlated or uncorrelated change, a minimum/maximum of the deviation results in a more/less distinct transition occurring in the data (see Fig. 3.4 (A), (E)). The influence of the evolution of mean is not expected to have a comparable influence. For the critical assessment of the inference performance, it is reasonable to employ the generic transition model $\mathcal{M}_{break}$ comprising a continuous change instead of the $\mathcal{M}_{shift}$ with a discontinuous and thus more distinct change.

Practical application of the inference

Due to the fact that the observations are generated by the generic transition model $\mathcal{M}_{break}$, the model approaches the time series over the complete interval $\mathcal{T}$. Hence, the inference on the transition model can be carried out with respect to the complete observation vector indicated by $\boldsymbol{y}_{|\mathcal{T}}$. The inference can be accomplished based on the joint posterior distribution as

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}_{|\mathcal{T}}) = p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}}) \text{ with } \boldsymbol{\vartheta} \in \Theta_{\boldsymbol{\vartheta}} = \Theta_{\boldsymbol{\beta}} \times \Theta_{\sigma} \times \Theta_{\boldsymbol{s}} \times \Theta_{\theta} \tag{3.32}$$

explicitly given in Eq. 3.26 and defined in the seven dimensional parameter space $\Theta_{\boldsymbol{\vartheta}}$. The computation and investigation of the full posterior distribution is not feasible.[2] Therefore, the inference on the model parameters $\boldsymbol{\vartheta}$ can be accomplished by the investigation of the marginal posterior distributions. In order to visualize the estimated transition models *point estimators* can be constructed.

The marginal posterior distribution $p(\vartheta_i|\boldsymbol{y}_{|\mathcal{T}})$ of each model parameter ϑ_i is obtained by integration of the full posterior distribution as implemented in Sec. 3.2. The mode of the marginal posterior distribution, referred to as maximum a posteriori estimate (MAP), can then be used as a point estimator,

$$\hat{\vartheta}_i = \underset{\vartheta_i \in \Theta_{\vartheta_i}}{\operatorname{argmax}}\, p(\vartheta_i|\boldsymbol{y}_{|\mathcal{T}}) = \left[p(\vartheta_i|\boldsymbol{y}_{|\mathcal{T}})\right]_{max}, \tag{3.33}$$

for the corresponding model parameter ϑ_i. The marginal posterior distribution of lowest dimensionality derived by analytical integration is defined in Eq. 3.29 as

$$p(\boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}}) \text{ with } \theta \in \Theta_{\theta},\ s_1 \in \Theta_{s_1},\ s_2 \in \Theta_{s_2} \tag{3.34}$$

[2] As explained in Sec. 2.3.3, there are MCMC approaches to approximate high dimensional posterior distributions. However, these techniques require further studies to ensure the convergence against the target distribution. With regard to the expansion of the Bayesian detection approach to multiple changes, the implementation of compute intensive techniques in the course of the basic inference task is deliberately avoided.

Table 3.1: *The sampling spaces for the numerical computation of the posterior distribution* $p(\boldsymbol{s}, \theta|\boldsymbol{y}_{\mathcal{T}})$.

$$
\begin{array}{rclcrcl}
\Theta_\theta & = & [\theta_{min},\, \theta_{max}] & \rightarrow & \Delta_\theta & = & [\theta_{min},\, \theta_{max}] = [5.0,\, 94.0]\ , \qquad \delta_\theta = 0.5 \\
\Theta_{s_1} & = & [s_{1,min}(\theta),\, \infty[& \rightarrow & \Delta_{s_1} & = & [s_{1,min}(\theta),\, s_{1,max}] = [s_{1,min}(\theta),\, 0.5]\,, \quad \delta_{s_1} = 0.01 \\
\Theta_{s_2} & = & [s_{2,min}(\theta),\, \infty[& \rightarrow & \Delta_{s_2} & = & [s_{2,min}(\theta),\, s_{2,max}] = [s_{2,min}(\theta),\, 0.5]\,, \quad \delta_{s_2} = 0.01
\end{array}
$$

of the deviation's slope parameters $\boldsymbol{s}$ and the change point parameter θ and is defined on the parameter space Θ_{θ,s_1,s_2} as given in Eq. 3.13 - 3.14. The explicit distribution can be computed on numerical grids, each defined within the range $\Delta_{\vartheta_i} = [\vartheta_{i,min}, \vartheta_{i,max}]$ and sampled by the step size δ_{ϑ_i}. The best choice of such a sampling space is the parameter space Θ_{ϑ_i} itself. However, for the generic transition models this is only applicable for the transition parameter θ. As a reasonable approach the sampling step of the transition parameter δ_θ is generally chosen as the half of the average distance between consecutive observations. In Tab. 3.1 the sampling spaces are listed as used in the course of this section.

For each sampling space Δ_{s_i} of the deviation parameters s_i for $i = 1, 2$ the lower limit $s_{i,min}(\theta)$ of the parameter spaces Θ_{s_i} can be used. By introducing a reasonable upper limit $s_{i,max}$ the sampling space can be bounded, and by choosing a sampling step $\delta_{s_i}$the sampling grid can be constructed. These initial sampling spaces Δ_{s_i} eventually need to be adjusted in case the marginal distribution $p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}})$ is influenced by artificial cutoffs or a too coarse sampling.

The marginal posterior distribution $p(s_1, s_2, \theta|\boldsymbol{y}_{|\mathcal{T}})$ can then be computed on the explicit sampling space $\Delta_\theta \times \Delta_{s_1} \times \Delta_{s_2}$. By numerical integration of the deviation parameters, the marginal posterior distribution of the transition parameter can be obtained as

$$
p(\theta|\boldsymbol{y}_{|\mathcal{T}}) = C \cdot \sum_{i=s_{1,min}(\theta)}^{s_{1,max}} \sum_{j=s_{2,min}(\theta)}^{s_{2,max}} p(s_i, s_j, \theta|\boldsymbol{y}_{|\mathcal{T}}) \cdot \delta_{s_1} \cdot \delta_{s_2}\ , \tag{3.35}
$$

and normalized afterwards by a numerical factor C. The posterior distributions $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ derived for each generated time series of Fig. 3.4 are presented in Fig. 3.5. Due to the random nature of the observations, the posterior distribution also depends randomly on the actual realization. It is therefore not surprising, that the mode's location $\hat{\theta} = \left[p(\theta|\boldsymbol{y}_{|\mathcal{T}})\right]_{max}$ does not exactly agree with the true parameter value θ. However, the MAP estimate $\hat{\theta}$ is within a certain quantile of the according probability distribution. By considering the shortest level interval, that encloses the highest posterior values covering at least $1 - \alpha$ percentage of the total probability weight, a confidence interval $\mathrm{CI}_{1-\alpha}$ for the MAP estimate is automatically obtained. This yields a natural way of uncertainty quantification.

In the same way, by numerical integration of the change point parameter θ, the

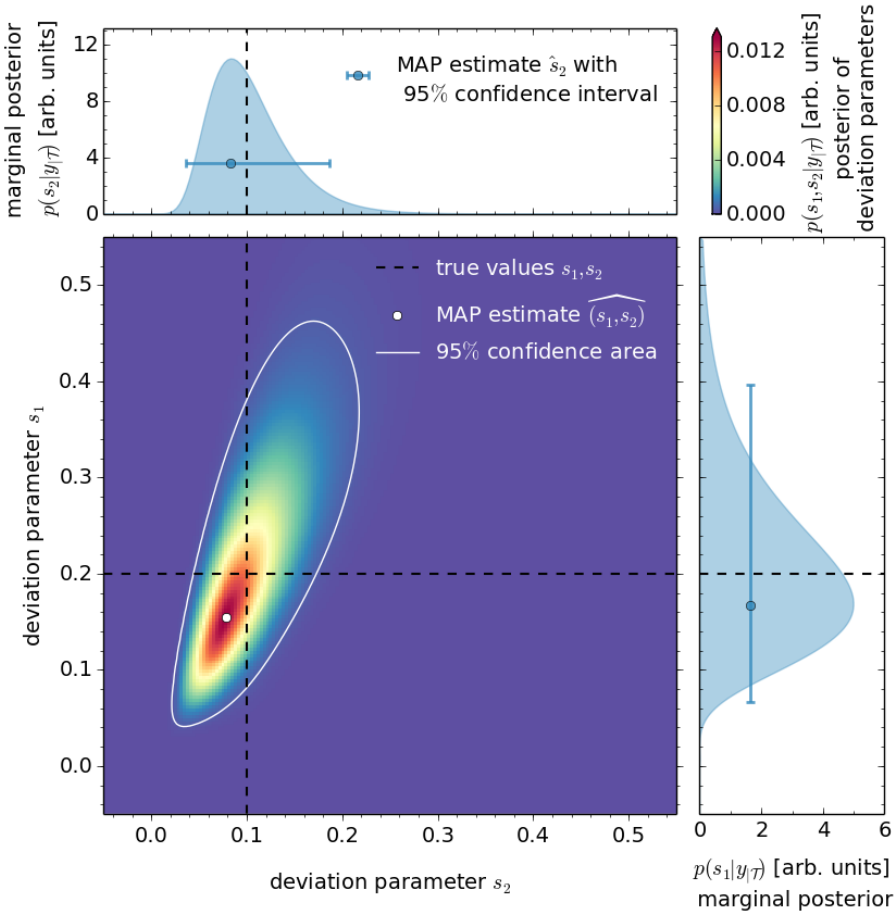


Figure 3.6: *The marginal posterior distribution $p(s_1, s_2|\boldsymbol{y}_\mathcal{T})$ of the slope parameters s_1, s_2 represents the degree of belief about the scedastic behavior of the observations $\boldsymbol{y}_\mathcal{T}$ from Fig. 3.4 (A). Non-zero parameter values (s_1, s_2) indicate a heteroscedastic behavior, whereas any deviation of the posterior distribution from the diagonal of the sampling space Δ_{s_1,s_2} indicates a change in the scedasticity of the observations. The confidence area $CA_{0.95}$ encloses the parameter combinations inferred with 95% certainty from the times series. Alongside are presented the further marginalized posterior distributions $p(s_i|\boldsymbol{y}_\mathcal{T})$ for $i = 1, 2$ providing the MAP estimates $\hat{s}_i$ and the according confidence intervals $CI_{0.95}$.*

marginal posterior distribution of the deviation parameters can be computed as

$$p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}}) = C' \cdot \sum_{i=\theta_{min}}^{\theta_{max}} p(s_1, s_2, \theta_i|\boldsymbol{y}_{|\mathcal{T}}) \cdot \delta_\theta \tag{3.36}$$

and normalized by the factor C'. As an example the resulting posterior distribution derived for the observations (A) in Fig. 3.4 is presented in Fig. 3.6. Based on the multivariate distribution the MAP estimate $\hat{\boldsymbol{s}} = \widehat{(s_1, s_2)} = \left[p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}})\right]_{max}$ can be assessed, whereas the isoline enclosing the highest posterior values covering at least $1-\alpha$ percentage of the total probability weight mark the confidence area $\mathrm{CA}_{1-\alpha}$. Further integration

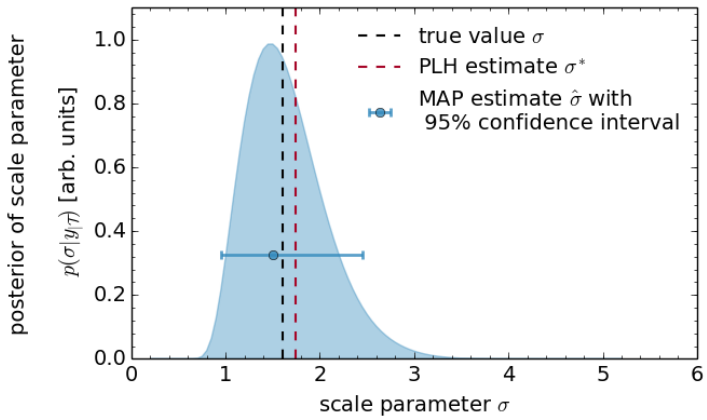


Figure 3.7: *Two optional point estimators for the scale parameter σ can be computed for the observations of Fig. 3.4 (A). Based on the marginal posterior distribution $p(\sigma|\boldsymbol{y}_{\mathcal{T}})$, the MAP estimate $\hat{\sigma}$ and the corresponding confidence interval $CI_{0.95}$ can be obtained. Another estimate provides the PLH estimator σ^*. By comparing the deviation of the estimates from the true value σ the accuracy of the estimation approaches can be assessed.*

is carried out to derive the univariate posterior distributions $p(s_i|\boldsymbol{y}_{|\mathcal{T}})$ for $i = 1, 2$ and the corresponding MAP estimates $\hat{s}_i = \left[p(s_i|\boldsymbol{y}_{|\mathcal{T}})\right]_{max}$ and confidence intervals $\mathrm{CI}_{1-\alpha}$.

The inference on the remaining model parameters, fixed effects $\boldsymbol{\beta}$ and scale σ, can be achieved by marginalization of the posterior distributions $p(\boldsymbol{\beta}, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}})$ given by Eq. 3.28 and $p(\sigma, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}})$ given by Eq. 3.27. A rational strategy to construct the parameter spaces $\Delta_{\boldsymbol{\beta}}$ and Δ_σ is shortly outlined in App. B.1. The estimated generic transition event can therefore be obtained as a set of MAP estimates

$$\hat{\boldsymbol{\vartheta}} = \left(\hat{\boldsymbol{\beta}}, \hat{\sigma}, \hat{\boldsymbol{s}}, \hat{\theta}\right) . \tag{3.37}$$

An alternative approach to construct point estimators offers the minimization of an adequate cost function, such as the likelihood function. For the designed transition model $\mathcal{M}_{break}$, the approach is only feasible for the fixed effects $\boldsymbol{\beta}$ in terms of the BLUP (Eq. 3.18) and for the scale parameter σ in terms of the PLH estimate (Eq. 3.20). The explicit estimators can be calculated numerically on the sampling space $\Delta_\theta \times \Delta_{s_1} \times \Delta_{s_2}$. The estimated generic transition event can therefore be also obtained as a set of MAP, BLUP and PLH estimates

$$\boldsymbol{\vartheta}^* = \left(\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta}\right) . \tag{3.38}$$

To compare the efficiency between the estimation approaches all introduced estimators are computed for all synthetic time series of varying observational evidence. The derived estimates are compared based on their accuracy, as exemplified in Fig. 3.7, and the required computational costs.

Interpretation of the marginal posterior distributions

Given the observations, the Bayesian inference provides a degree of belief about each parameter of the transition model over the investigated sampling spaces. Most importantly, the derived marginal posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ represents a degree of belief that a transition θ occurs at a time point $t_i \in \Delta_\theta$ (see Fig. 3.5). The unimodal shape of the distribution supports the model assumption of a single transition within the time series.[3] The position of the distribution's mode, the MAP estimator $\hat{\theta}$, lies in the smallest interval enclosing 95% of the complete probability weight. In this way, the location of a probable transition within the observations may be reduced with a confidence of 95% to the corresponding sub-interval $\mathrm{CI}_{0.95}$. The quality of the inference performance is assessed by the width of the posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$, that means the sharper the distribution the better the inference performance. A good performance enables a reliable inference on the transition patterns and is the essential requirement for the expansion of the generic transition model approach to a multi change point approach.

The marginal posterior distribution of the deviation parameters $p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}})$ represents a degree of belief about the deviation slope parameters s_1 and s_2 prior and after the transition (see Fig. 3.6). Thus, the distribution indicates the temporal dependency of the variability underlying the observations and the scedastic behavior, respectively. The continuous confidence area $\mathrm{CA}_{0.95}$ contains the mode of the unimodal distribution and 95% of the total probability weight. Based on the enclosed parameter combinations the posterior distribution reveals the underlying variability evolution across the potential transition. As a general pattern, non-zero parameter values, i.e. $s_i \neq 0$ for $i = 1, 2$, indicate a heteroscedastic behavior and zero parameter values indicate a homoscedastic behavior. Non-equal values $s_1 \neq s_2$, apparent due to their deviation from the diagonal of the sampling space $\Delta_{s_1} \times \Delta_{s_2}$, indicate a change in the scedasticity of the observations. In this way, the choice of a symmetric sampling space for the deviation parameters offers an intuitive visualization of the variability evolution of the investigated time series.

For concrete applications it might be of interest to study the marginal posterior distribution $p(\beta_1, \beta_2|\boldsymbol{y}_{|\mathcal{T}})$ of the mean slopes β_1 and β_2 on the joint sampling space $\Delta_{\beta_1} \times \Delta_{\beta_2}$ in a similar manner. Any change of the mean evolution across the transition becomes obvious within a symmetric sampling space.

Performance with respect to observational evidence

The numerical results derived from the Bayesian inference on transitions of different observational evidence are summarized in Tab. C.1. The generic approach reproduces the true underlying generic transition models convincingly as reasoned in the following.

[3] It is important to stress the fact that the reverse conclusion is in general not valid. More precisely, a multimodal posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ does not directly indicate multiple transitions. The only conclusion that can be drawn from a multimodal distribution is that the chosen generic model may not be correct, but not how an adequate model approach may look like.

On the one hand, the set of estimates parameterizing the underlying transition are all in good agreement with the true model settings, that is the relative errors of each estimator are of comparable order for the different realizations. On the other hand, the marginal posterior distributions correctly mirror the general behavior of the mean and deviation evolution in the corresponding sampling spaces. Additionally, the absolute errors of the BLUP $\boldsymbol{\beta}^*$ and the PLH σ^* estimates are of the same order as the absolute errors of the MAP estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$. Therefore, the estimation approaches are considered to have approximately the same estimation accuracy. The above findings hold for all change geometries and for time series comprising transitions of different observational evidence, respectively. In particular the occurrence of a transition θ is in average localized on a sub-interval of about 17% of the complete time series $\mathcal{T}$, as indicated by the confidence intervals $\mathrm{CI}_{0.95}$ of the posterior distributions $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$. However, a considerable loss of inference quality is evident for an uncorrelated change in the deviation in form of a maximum at the transition (see Fig. 3.5 (F)).

In conclusion, the performance of the inference validates the capability of the introduced Bayesian approach to convincingly localize a change point and to infer on the underlying transition patterns. Thus, the approach enables to detect and to describe a generic transition underlying the investigated time series. The findings justify that, without loss of estimation accuracy, the time expensive computation of the MAP estimates based on the high-dimensional posterior densities $p(\sigma, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}})$ and $p(\boldsymbol{\beta}, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{T}})$ can be avoided by employing instead the BLUP and PLH estimates. Consequently, the highlighted estimates in Tab. C.1 represent the optimal set of estimators, from now on referred to as the *estimated generic transition event* as

$$\hat{\mathcal{M}}_j : \ \boldsymbol{y}_{|\mathcal{T}} = F_{\hat{\theta}}^{(j)}\boldsymbol{\beta}^{*} + \boldsymbol{\varepsilon}_{\hat{\theta}} \ \text{ with } \ \boldsymbol{\varepsilon}_{\hat{\theta}} \sim \mathcal{N}\left(0, (\sigma^*)^2 \cdot \Omega_{\hat{\theta},\hat{s}_1,\hat{s}_2}\right) , \tag{3.39}$$

given the transition type indexed by $j = [\text{break}, \text{shift}]$. Nevertheless, the performance quality of the approach is not completely independent of the observational evidence. An effect on the performance can be observed for a change of the deviation, uncorrelated to the mean behavior and of maximal value at the singularity. The resulting broadening of the posterior distribution has to be taken into account, when interpreting the inference results of the transition detection approach.

3.3.2 Sensitivity to data loss

In real time series, analysts have to deal with sparse and irregularly sampled observations. The implemented Bayesian inference does not require uniformly sampled observations since from the beginning it employs only the available data. Consequently, no preprocessing of the time series, such as interpolation techniques, is required to apply the introduced detection approach. Nevertheless, the inference performance is affected by the amount of observations provided to the algorithm. The efficiency of the detection approach essentially depends on the posterior distribution of the transition parameter used to reveal the degree of belief about the transition location, and on the point estimators used to visualize the most likely underlying transition pattern. In order to evaluate the sensitivity of the inference to data loss, the obtained results are investigated with respect to the number of observations. Thereby, the sensitivity is deduced from the convergence of the posterior distribution $p(s_1, s_2, \theta|\boldsymbol{y}_{|\mathcal{T}})$ (Eq. 3.29) and from the bias of the employed estimators $\boldsymbol{\vartheta}^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})$ (Eq. 3.38) derived with respect to this distribution.

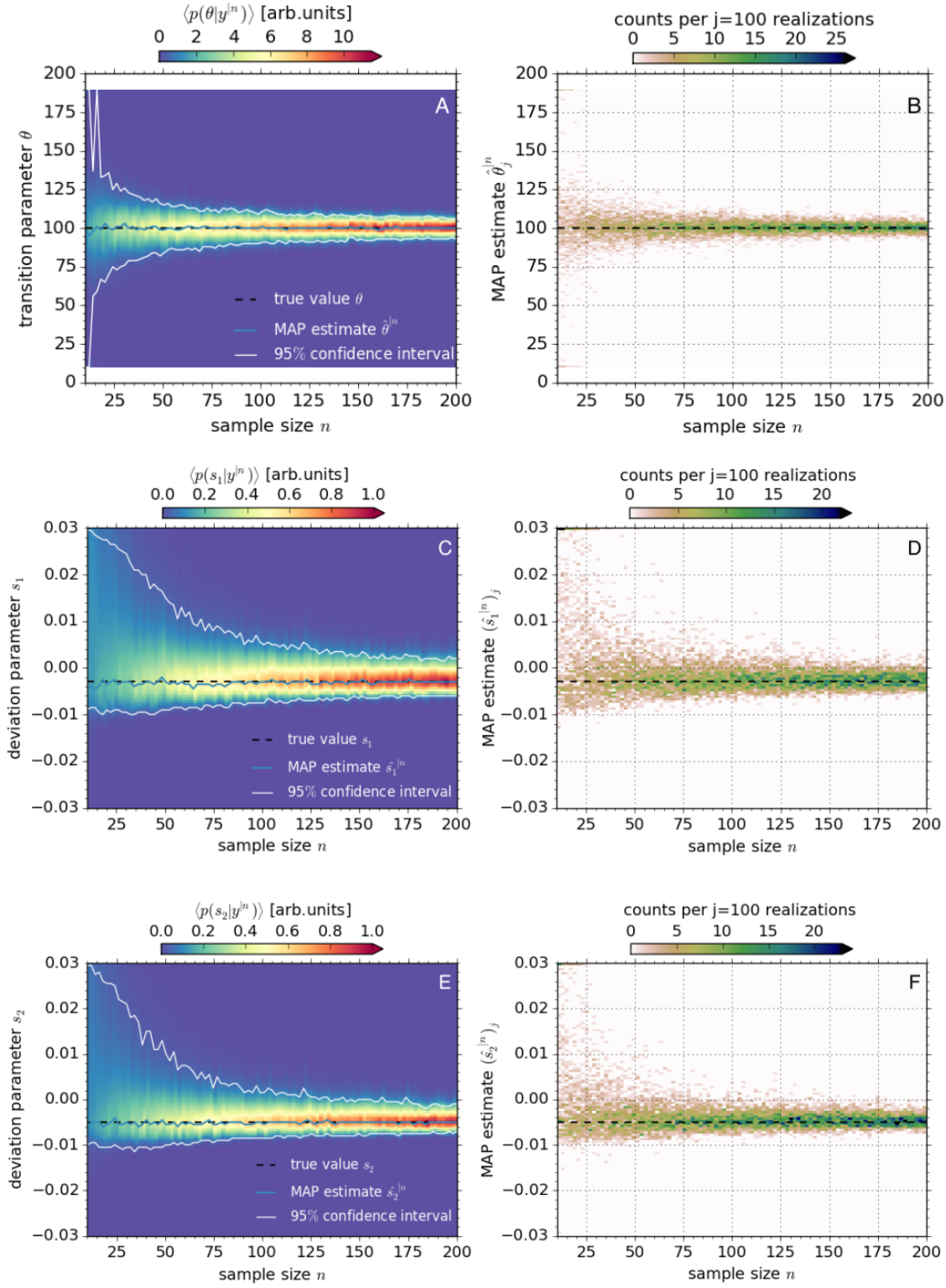
For this purpose, a generic transition model $\mathcal{M}_{break}$ is defined in Tab. B.2 and is used to generate sets of basic time series $\boldsymbol{y}_j \in \mathbb{R}^n$ of $n = 200$ temporally equidistant observations acquired at time points $t_i \in \mathcal{T}$ in the course of $j = 1, \ldots, 100$ random realizations. Based on a uniform distribution the basic time series are resampled by randomly ignoring a given portion of the data points in each set. In this way, time series with random gaps and irregular sampling steps are simulated. Note, each synthetic time series is analyzed over the complete interval of the basic time series $\mathcal{T}$. Therefore, the index indicating the local observations as $\boldsymbol{y}_{|\mathcal{T}}$ is avoided here for convenience. For each of the synthetic observations $\boldsymbol{y}_j^{|n}$ consisting of $n = 10, 12, \ldots, 200$ data points, the posterior distribution $p(s_1, s_2, \theta|\boldsymbol{y}_j^{|n})$ of the transition parameter θ and the deviation parameters s_i, and all point estimators $(\boldsymbol{\vartheta}_j^{|n})^*$ are computed given the realization j.

Convergence of the posterior distribution

To investigate the mean convergence of the marginal posterior distribution employed in the detection approach, the computed distributions are averaged over the realizations

$$\left\langle p(s_1, s_2, \theta|\boldsymbol{y}^{|n}) \right\rangle = \frac{1}{100} \sum_{j=1}^{100} p(s_1, s_2, \theta|\boldsymbol{y}_j^{|n}) \tag{3.40}$$

Figure 3.8 ***(facing page)*****:** *The sensitivity of the inference performance to the number of observations n is deduced from the convergence of the marginals of the mean posterior distribution $\langle p(\vartheta|\boldsymbol{y}^{|n}) \rangle$ of the parameter ϑ, i.e. the transition θ and the deviation parameters s_1, s_2, averaged over $j = 100$ realizations (left column). The confidence intervals $CI_{0.95}$ indicate a divergence for $n < 40$ for the transition and $n < 70$ for the deviation parameters. The individual estimates $\hat{\vartheta}_j^{|n}$ of each realization j support the assumption of consistency of the MAP estimators (right column).*


$\langle p(\theta|y^{|n}) \rangle$ [arb.units]
0 2 4 6 8 10
A
transition parameter θ
true value θ
MAP estimate $\hat{\theta}^{|n}$
95% confidence interval
sample size n
counts per j=100 realizations
0 5 10 15 20 25
B
MAP estimate $\hat{\theta}_j^{|n}$
sample size n
$\langle p(s_1|y^{|n}) \rangle$ [arb.units]
0.0 0.2 0.4 0.6 0.8 1.0
C
deviation parameter s_1
true value s_1
MAP estimate $\hat{s}_1^{|n}$
95% confidence interval
sample size n
counts per j=100 realizations
0 5 10 15 20
D
MAP estimate $(\hat{s}_1^{|n})_j$
sample size n
$\langle p(s_2|y^{|n}) \rangle$ [arb.units]
0.0 0.2 0.4 0.6 0.8 1.0
E
deviation parameter s_2
true value s_2
MAP estimate $\hat{s}_2^{|n}$
95% confidence interval
sample size n
counts per j=100 realizations
0 5 10 15 20
F
MAP estimate $(\hat{s}_2^{|n})_j$
sample size n

for constant sample sizes n of the observations $\boldsymbol{y}_j^{|n}$. By numerical integration, the marginal distributions $\langle p(\vartheta|\boldsymbol{y}^{|n})\rangle$ for each parameter ϑ, i.e. the transition parameter θ and the deviation parameters s_i for $i = 1, 2$, are derived over the sample size n as presented in Fig. 3.8 (left column). Apparently, the mode $\hat{\vartheta}^{|n} = \left[\langle p(\vartheta|\boldsymbol{y}^{|n})\rangle\right]_{max}$ of each mean distribution remains close to the corresponding true value ϑ irrespective of the sample size. The 95% credibility intervals $\mathrm{CI}_{0.95}$ indicate the width of each mean distribution and may be interpreted as the uncertainty of the inference on the corresponding parameter. A general assessment of the dependency between the inference performance and the sampling size can be conducted in the light of the asymptotic theory and under some regularity conditions [72]. Thus, it can be shown, that the posterior distribution of some parameter ϑ converges towards normality for increasing sample size n and can be approximated as

$$\lim_{n\to\infty} p(\vartheta|\boldsymbol{y}^{|n}) \to \mathcal{N}\left(\vartheta', (n \cdot \mathcal{J}(\vartheta))^{-1}\right) . \tag{3.41}$$

Here, the parameter value ϑ' minimizes the Kullback-Leibler information and the variance is formulated based on the Fisher information $\mathcal{J}(\vartheta)$ (Eq. 2.19). Hence, the width of the posterior distribution depends inversely proportional on the square root of the sample size n, that is

$$width\left[\langle p(\vartheta|\boldsymbol{y}^{|n})\rangle\right] \propto \frac{1}{\sqrt{n}} . \tag{3.42}$$

The broadening of the posterior distribution for $n < 200$ is therefore the natural consequence of the information loss induced by the decreased number of observations provided to the algorithm. For large numbers of observations the posterior distribution converges towards a delta distribution located at the true parameter value ϑ. In any case, even for sparse time series of about $n = 20$, the non-flatness of the transition posterior distribution $\langle p(\theta|\boldsymbol{y}^{|n})\rangle$ presented in Fig. 3.8 (A) clearly hints towards one distinct transition time within the time series. More precisely, the efficiency of the detection approach is determined by its ability to localize the credibility of a change within the time series. The degree of belief about a change is given by the transition posterior distribution. Hence, the smaller the distribution width is relative to the length $\mathcal{T}$ of the total time series, the less uncertain is the inferred transition location. The empirical divergence of the transition posterior $\langle p(\theta|\boldsymbol{y}^{|n})\rangle$ significantly increases for $n < 40$ such that the distribution encloses more than the half of the investigated time series $\mathcal{T}/2$. For this reason the sample size $n = 40$ may be considered as a practical minimum amount of observations required to obtain in average a transition posterior distribution of moderate width. This requirement is of particular importance for the resolution of multiple changes in the kernel-based expansion of the detection approach, as discussed in Chap. 4.

The posterior distributions of the deviation parameters $\langle p(s_i|\boldsymbol{y}^{|n})\rangle$ for $i = 1, 2$ presented in Fig. 3.8 (B) and (C) exhibit a considerable divergence for sample sizes of $n < 70$. In addition, the distributions are skewed for all numbers of observations as a consequence of the restrictions on the sample space $\Delta_{s_1} \times \Delta_{s_2}$ (see Tab. 3.1) required to ensure the

identifiability of the generic model approach. Therefore, the skewness of the deviation parameters posterior distributions is an inherent feature of the detection approach. However, except for the transition parameter θ, the model's parameters are primarily used as point estimators to infer on the estimated underlying transition pattern. For this reason the sensitivity of the inference to data loss is further evaluated based on the quality of the derived point estimators.

Bias of the estimators

The efficiency of the algorithm is also determined by its ability to reliably infer on the explicit location of a transition for every single time series and on the complete set of model parameters, respectively. Commonly, the dependency of an estimator $\hat{\vartheta}_j^{|n}$ on the sample size n may be assessed in terms of the consistency as

$$\lim_{n\to\infty} p(|\hat{\vartheta}_j^{|n} - \vartheta| > \varepsilon) = 0 \quad \forall\, \varepsilon > 0 \tag{3.43}$$

with respect to the true parameter value ϑ. Thus, a MAP estimator $\hat{\vartheta}_j^{|n} = \left[p(\vartheta|\boldsymbol{y}_j^{|n})\right]_{max}$ is convergent in case the weight of the posterior distribution concentrates in decreasing neighborhoods ε of the true value for an increasing amount of observations n. In other words, the distance between estimate and true value approaches zero. The individual MAP estimates $\hat{\vartheta}_j^{|n}$ for each realization j are presented in Fig. 3.8 (right column) and in general support the assumption of consistency. Nevertheless, even for sample sizes $n > 100$ the individual MAP estimates do occasionally not correctly reproduce the true underlying generic process patterns due to stochastic errors. In order to compare the quality of all employed estimators over the sample size n the systematic difference between the estimator $\hat{\vartheta}^{|n}$ and the true value ϑ offers a suitable measure and denotes the bias, defined as

$$bias(\hat{\vartheta}^{|n}) = \mathbb{E}\left[\hat{\vartheta}^{|n}\right] - \vartheta = \frac{1}{100} \cdot \sum_{j=1}^{100} \hat{\vartheta}_j^{|n} - \vartheta\,. \tag{3.44}$$

Here, the expected value of the estimator may be derived as the arithmetic mean of the estimates $\hat{\vartheta}_j^{|n}$ over the realizations j. To provide a comprehensive illustration of the bias of all estimators parameterizing the estimated generic transition model $\hat{\mathcal{M}}_{break}^{|n}$, each bias is computed relative to the corresponding true parameter value as

$$bias'(\hat{\vartheta}^{|n}) = \frac{bias(\hat{\vartheta}^{|n})}{\vartheta}\,, \tag{3.45}$$

and is presented over the sample size n in Fig. 3.9.

As expected from the skewness of the marginal posterior distributions $\langle p(s_i|\boldsymbol{y}^{|n})\rangle$, the estimators of the deviation parameters s_i for $i = 1, 2$ are in average biased of about

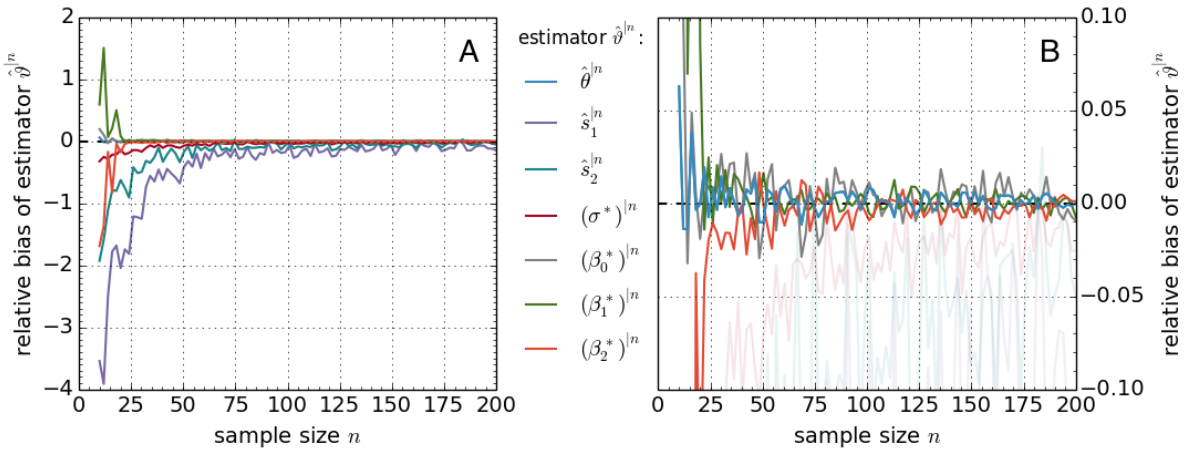


Figure 3.9: *The sensitivity of the inference performance to the number of observations n is further deduced from the relative bias of each estimator $(\boldsymbol{\vartheta}^*)^{|n} = ((\boldsymbol{\beta}^*)^{|n}, (\sigma^*)^{|n}, \hat{\boldsymbol{s}}^{|n}, \hat{\theta}^{|n})$ employed in the generic transition model. The relative biases of the estimators describing the deviation, that are deviation slopes s_i for $i = 1, 2$ and scale σ, exhibit a similar behavior due to the required restriction of the sampling space and increase for $n < 70$ (A). The relative biases of the estimators of the transition θ and the mean parameters β_i for $i = 0, 1, 2$ are much smaller and considered as negligible for $n > 40$ (B).*

-8% (see Fig. 3.9 (A)). The bias increases considerably for $n < 70$. Besides the deviation parameters, the scale parameter σ describes the variability of the generic transition as the non-linear part of the inference task. Thus, the relative bias of the estimator of the scale σ exhibits a synchronous increase for $n < 70$, but for considerably smaller values of around -3% in average. Clearly, the biases are resulting from the necessary restriction on the sample space $\Delta_{s_1} \times \Delta_{s_2}$ (see Tab. 3.1). In order to avoid further inaccuracy by the discrete sampling in the course of the numerical computations, the resolution within the sampling space has to be carefully chosen. For this reason, it is considered as essential to the performance of the detection approach to iteratively adapt the numerical sampling of the deviation parameters after a first inference based on an initial sampling grid and, if necessary, to perform the inference based on this refined sampling grid accordingly.

The relative biases of the estimator of the transition parameter θ and the estimators describing the mean of the generic transition, that are the coefficients β_i for $i = 0, 1, 2$ vary between $\pm 2\%$ and are therefore in general not considered as biased for sample sizes $n > 40$ (see Fig. 3.9 (B)). In summary, the detection approach infers in average convincingly on the transition parameter even for sample sizes $n < 50$. However, the biases of the remaining model parameters indicate that the underlying transition patterns are not reliably estimated from the observations anymore. Moreover, to a certain extent a bias of the deviation parameters can not be avoided due to the specific design of the generic transition model. As a pragmatic approach to the detection of transition events in real time series, the amount of required observations is suggested as $n \geq 50$ in order to reliably infer on all features of the generic transition model.

3.3.3 Robustness in the presence of outliers

Another challenge to the investigation of real time series are occasionally unusual observations. Such outliers may occur by chance in the random fluctuations of any stochastic process or may emerge from individual measurement errors in the course of data acquirement. The vulnerability of an analysis approach to outliers is a crucial concern and has to be studied in order to justify the application of the method on real observations. Commonly, outliers are described by heavy-tailed distributions and can in principle be incorporated in a robust parametric model of a stochastic process [105, 106]. The proposed generic transition model, however, is designed based on the assumption of normally distributed random fluctuations and the explicit implementation of the inference is carried out accordingly. Hence, the application of the detection approach to real observations may lead to model errors, since outliers may not be captured by the assumed random behavior of the generic transition model. In order to assess the influence of such model errors on the inference performance, the robustness of the approach to synthetic outliers are investigated. Thereby, the robustness is deduced from the comparison between the inference results obtained for observations without outliers and for observations with outliers.

To generate synthetic time series with approximately the same statistical characteristics as the assumed normal distribution but with additional outliers, a symmetric distribution with long tails is required. An adequate behavior offers the family of t-distributions $t_\nu\left(\boldsymbol{\mu}, \Psi\right)$ [93] characterized by the location vector $\boldsymbol{\mu} \in \mathbb{R}^n$, the scale matrix $\Psi \in \mathbb{R}^{n \times n}$ and the shape parameter $\nu \in \mathbb{R}^+$, referred to as degree of freedom. The corresponding multivariate probability distribution of the observation vector $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Psi, \nu) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{n}{2}}\sqrt{|\Psi|}} \cdot \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^T \Psi^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+n}{2}} , \tag{3.46}$$

with the Gamma function $\Gamma(\cdot)$ given by

$$\Gamma(k) = \int_0^\infty z^{k-1} \cdot e^{-z}\, dz \;\text{ with }\; k \in \mathbb{R}^+ . \tag{3.47}$$

The dependence of the distributional shape on the degree of freedom ν is illustrated for the standard normal t-distribution $t_\nu\left(0, \mathbf{1}\right)$ in Fig. 3.10. In the lower limit case $\nu = 1$ the t-distribution is equivalent to the Cauchy distribution, whereas in the upper limit case $\nu \to \infty$ the t-distribution approaches the Gaussian distribution $\mathcal{N}(0, \mathbf{1})$. For approximately $\nu < 30$ the tails of the t-distribution are longer than that of a normal distribution and result in the occurrence of outliers [72]. Thus, a stochastic process comprising a transition as well as outliers can be simulated by the generic transition model (Eq. 3.7) with a modified random term $\boldsymbol{\xi}$ as

$$\mathcal{M}_j' : \; \boldsymbol{y}_{|\mathcal{T}} = F^{(j)}\boldsymbol{\beta} + \boldsymbol{\xi} \;\text{ with }\; \boldsymbol{\xi} \sim t_\nu\left(0, \sigma^2\Omega\right) . \tag{3.48}$$

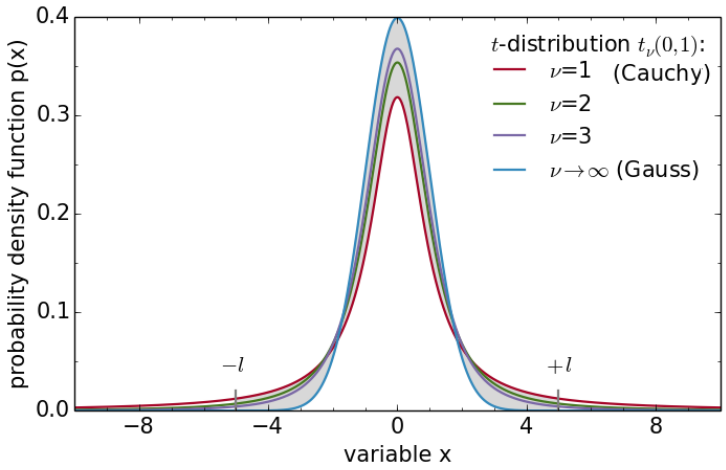


Figure 3.10: *The family of standard centralized t-distribution $t_\nu(0,\mathbf{1})$ with a degree of freedom $\nu < 30$ can be used to simulate outliers. For $\nu \to \infty$ the t-distribution converges into a Gaussian distribution $\mathcal{N}(0,\mathbf{1})$ and for $\nu = 1$ the t-distribution describes a Cauchy distribution. As a practical approach to avoid unrealistic outliers, the long tails of the Cauchy distribution can be truncated at some truncation limit l. The grey shaded area illustrates the redistribution of the probability weight of the Gaussian relative to the truncated Cauchy distribution $\mathcal{C}_{|5|}$ for the truncation limit $l = \pm 5$.*

Consequently, the synthetic observations $\boldsymbol{y}_{|\mathcal{T}}$ follow a multivariate t-distribution

$$\boldsymbol{y}_{|\mathcal{T}} \sim t_\nu \left(F^{(j)}\boldsymbol{\beta}, \sigma^2\Omega \right) \tag{3.49}$$

over the time interval $\mathcal{T}$. The degree of freedom ν is practically chosen with respect to the required amount and magnitude of simulated outliers. The Cauchy distribution contains the most probability weight of all t_ν-distributions in its tails and thus simulates the most outliers. But the observations resulting from the far tails of the distribution are usually not realistic for real time series. In order to avoid extreme outliers two strategies may be applied to reduce the weight in the long tails of the employed distribution:

- using a higher degree of freedom $(1 < \nu < 30)$,[4]
- symmetrically truncating the Cauchy distribution $(\nu = 1)$ using a truncation limit l, such that

$$\mathcal{C}_{|l|} = t_1\,(0,\mathbf{1}) \quad \text{with } -l < x_i < l \quad \forall\, i = 1,...,n\,. \tag{3.50}$$

Here, the truncated Cauchy distribution is employed to challenge the inference approach in the presence of multiple outliers. The corresponding random behavior of the generic transition $\boldsymbol{\xi}$ (Eq. 3.48) is derived by the linear transformation

$$\boldsymbol{\xi} = \sigma L_\Omega \boldsymbol{x}, \text{ with } \boldsymbol{x} \sim \mathcal{C}_{|l|}\,, \tag{3.51}$$

[4] A characteristic property of the Cauchy distribution is, that the mean and variance are infinite. By choosing a higher degree of freedom the derived distribution obtains a finite mean (for $\nu > 1$) and variance (for $\nu > 2$).

using the scale parameter σ and the Cholesky decomposition of the covariance matrix $\Omega = L_\Omega L_\Omega^T$. The difference in the probability weights between the Gaussian and the truncated Cauchy distribution is exemplified for $\mathcal{C}_{|5|}$ in Fig. 3.10.

For the purpose of evaluating the influence of outliers on the inference performance, time series $\boldsymbol{y}_j^{|\boldsymbol{\xi}} \in \mathbb{R}^n$ consisting of $n = 100$ temporally equidistant observations are generated for different random behavior $\boldsymbol{\xi}$ in the course of $j = 300$ random realizations. The synthetic observations are simulated based on a joint transition model $\mathcal{M}_{break}$ specified in Tab. B.3, but for different random terms $\boldsymbol{\xi}$ as listed in the following:

$\boldsymbol{y}_j^{|\mathcal{N}}$: reference data without outliers generated by a Gaussian distribution,
$\boldsymbol{y}_j^{|\mathcal{C}_{|5|}}$: data with outliers generated by a Cauchy distribution truncated at $l = \pm 5$,
$\boldsymbol{y}_j^{|\mathcal{C}_{|10|}}$: data with outliers generated by a Cauchy distribution truncated at $l = \pm 10$,
$\boldsymbol{y}_j^{|\mathcal{C}}$: data with (extreme) outliers generated by a Cauchy distribution .

The amount and magnitude of outliers increase with the truncation limit applied to the Cauchy distribution, i.e. $\mathcal{C}_{|5|}$ generates less outliers than $\mathcal{C}_{|10|}$, which generates less outliers than $\mathcal{C}_{|\infty|} = \mathcal{C}$. Even though the assumption is unrealistic, the Cauchy distribution is employed to identify model estimators that are most affected by the presence of outliers. The Bayesian approach, essentially designed for Gaussian random behavior, is then applied on the synthetic observations $\boldsymbol{y}_j^{|\boldsymbol{\xi}}$. By comparing the inference performance given the reference data with the performance given the data containing outliers, the relative robustness of the detection approach in the presence of outliers is deduced.

The mean posterior distributions $\left\langle p(\theta|\boldsymbol{y}^{|\boldsymbol{\xi}}) \right\rangle$ of the transition parameter θ is computed and presented for each simulated random behavior $\boldsymbol{\xi}$ in Fig. 3.11. The averaged distributions indicate the derived degree of belief about the transition location and indicate the efficiency of the approach to localize the underlying transition. In comparison to the reference posterior distribution $\left\langle p(\theta|\boldsymbol{y}^{|\mathcal{N}}) \right\rangle$ the width of the posterior distribution increases for an increasing amount and magnitude of outlier due to the increase of uncertainty in the observations. The posterior distributions for data generated by truncated Cauchy distributions are unimodal with approximately the same confidence intervals $\mathrm{CI}_{0.95}$ as the reference distribution, and yield correct MAP estimates $\hat{\theta}^{|\boldsymbol{\xi}}$ of the underlying transition location. The posterior distribution $\left\langle p(\theta|\boldsymbol{y}^{|\mathcal{C}}) \right\rangle$ is multimodal and fails to localize the transition, since its confidence interval $\mathrm{CI}_{0.95}$ nearly equates to the length of the time series.

Furthermore, the model estimates $(\boldsymbol{\vartheta}^*)_j^{|\boldsymbol{\xi}} = ((\boldsymbol{\beta}^*)^{|\boldsymbol{\xi}}, (\sigma^*)^{|\boldsymbol{\xi}}, \hat{\boldsymbol{s}}^{|\boldsymbol{\xi}}, \hat{\theta}^{|\boldsymbol{\xi}})_j$ are computed for each realization j of the simulated random behavior $\boldsymbol{\xi}$. The influence of outliers on the estimation accuracy of the underlying transition patterns is assessed by investigating

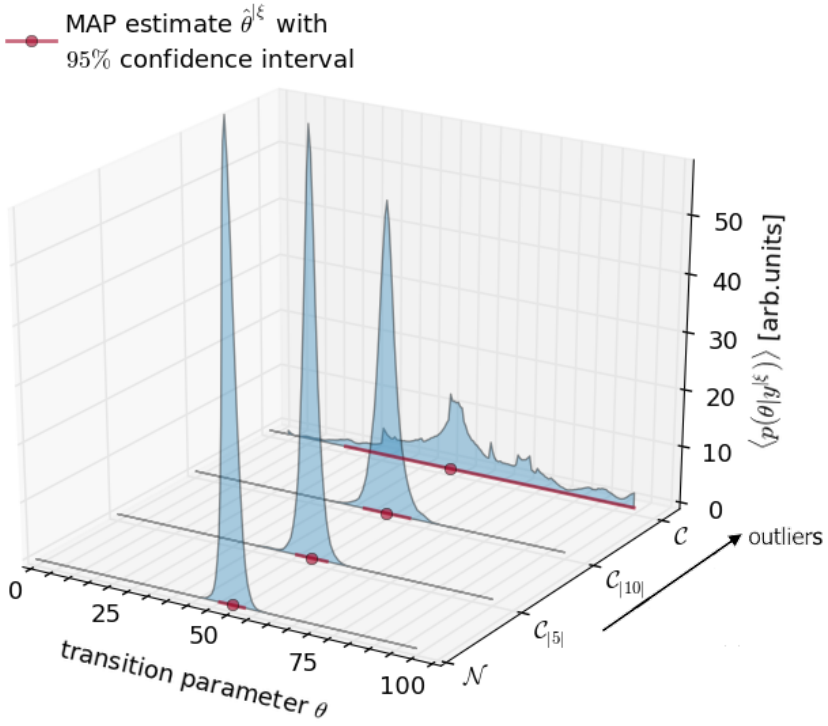


Figure 3.11: *The robustness of the inference performance in the presence of outliers is deduced from the convergence of the mean posterior distribution $\langle p(\theta|\boldsymbol{y}^{|\boldsymbol{\xi}})\rangle$ of the transition parameter θ for different random behavior $\boldsymbol{\xi}$. Similar to the reference posterior distribution $\langle p(\theta|\boldsymbol{y}^{|\mathcal{N}})\rangle$, the underlying transition can be convincingly localized by the posterior distributions given observations generated by truncated Cauchy distributions $\mathcal{C}_{|5|}$ and $\mathcal{C}_{|10|}$. The posterior distribution derived from observations generated by the Cauchy distribution $\mathcal{C}$ is multimodal and is not able to localize the underlying transition.*

the relative error for each estimate $\hat{\vartheta}_j^{|\boldsymbol{\xi}}$ as

$$\text{error}'(\hat{\vartheta}_j^{|\boldsymbol{\xi}}) = \frac{\hat{\vartheta}_j^{|\boldsymbol{\xi}} - \vartheta}{\vartheta}\,. \tag{3.52}$$

The comprehensive histograms of the relative errors are presented in Fig. 3.12. The estimates given the Cauchy random behavior exhibit extreme errors and are not discussed here for the sake of convenience. In general, there is no significant difference between the relative errors for the estimates $\hat{\vartheta}_j^{|\mathcal{N}}$ of the reference observations and the estimates $\hat{\vartheta}_j^{|\mathcal{C}_{|5|}}$ of the observations generated from the truncated Cauchy distribution. Apparently, the scedasticity estimators scale $(\sigma^*)^{|\boldsymbol{\xi}}$ and deviation slopes $\hat{s}_1^{|\boldsymbol{\xi}}$ and $\hat{s}_2^{|\boldsymbol{\xi}}$ are most affected by the presence of outliers, to a small extent for the observations $\boldsymbol{y}_j^{\mathcal{C}_{|5|}}$ and more distinct for the observations $\boldsymbol{y}_j^{\mathcal{C}_{|10|}}$. Since the scale parameter σ directly depends on the dispersion of the observations, that is basically the width of the employed distribution, its estimator is most affected by the presence of outliers. The systematic overestimation of the scale mirrors the high uncertainty in the observations. As a consequence the deviation slope

Table 3.3: *The relative bias of the model estimates $\hat{\vartheta}_j^{|\boldsymbol{\xi}}$ is derived from $j = 300$ realizations of the synthetic observations $\boldsymbol{y}_j^{|\boldsymbol{\xi}}$ for the random behavior $\boldsymbol{\xi}$ simulated by a Gaussian (without outliers) or by Cauchy distributions (with outliers). The inference performance is robust against moderate outliers as generated by $\mathcal{C}_{|5|}$. The model estimators most affected by unusual observations are the scedasticity estimators scale σ^* and deviation slopes $\hat{s}_1$ and $\hat{s}_2$.*

estimator	$bias'(\hat{\vartheta}^{\|\boldsymbol{\xi}})$ for modeled random term			
	$\boldsymbol{\xi} \sim$ Gauss	$\boldsymbol{\xi} \sim$ Cauchy		
$\hat{\vartheta}^{\|\boldsymbol{\xi}}$	$\mathcal{N}$	$\mathcal{C}_{\|5\|}$	$\mathcal{C}_{\|10\|}$	$\mathcal{C}$
$\hat{\theta}^{\|\boldsymbol{\xi}}$	$-4.0 \cdot 10^{-4}$	$-8.7 \cdot 10^{-4}$	$3.1 \cdot 10^{-3}$	$5.6 \cdot 10^{-2}$
$\hat{s}_1^{\|\boldsymbol{\xi}}$	$2.4 \cdot 10^{-1}$	$3.3 \cdot 10^{-1}$	$5.5 \cdot 10^{-1}$	1.0
$\hat{s}_2^{\|\boldsymbol{\xi}}$	$2.8 \cdot 10^{-1}$	$3.9 \cdot 10^{-1}$	$6.5 \cdot 10^{-1}$	2.1
$(\sigma^*)^{\|\boldsymbol{\xi}}$	$-8.0 \cdot 10^{-2}$	$8.3 \cdot 10^{-3}$	$4.1 \cdot 10^{-1}$	31.5
$(\beta_0^*)^{\|\boldsymbol{\xi}}$	$-1.7 \cdot 10^{-3}$	$-6.0 \cdot 10^{-3}$	$-4.8 \cdot 10^{-3}$	$2.8 \cdot 10^{-1}$
$(\beta_1^*)^{\|\boldsymbol{\xi}}$	$7.7 \cdot 10^{-3}$	$2.2 \cdot 10^{-2}$	$7.0 \cdot 10^{-3}$	$2.5 \cdot 10^{-1}$
$(\beta_2^*)^{\|\boldsymbol{\xi}}$	$6.2 \cdot 10^{-3}$	$1.7 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	-20.9

parameters are also systematically overestimated, though of a smaller order.

In order to derive a measure suitable to compare the effect of the simulated random behavior $\boldsymbol{\xi}$ on the estimators the relative bias as defined in Eq. 3.45 is computed. The results are listed in Tab. 3.3. The relative biases support the finding, that the estimators $\hat{\vartheta}^{|\mathcal{C}_{|5|}}$ are of approximately the same accuracy as the reference estimators $\hat{\vartheta}^{|\mathcal{N}}$. The relative biases of the estimators $\hat{\vartheta}^{|\mathcal{C}_{|10|}}$ is of comparable order, but differs particularly for the scale estimator $(\sigma^*)^{|\boldsymbol{\xi}}$. The relative biases of the estimators $\hat{\vartheta}^{|\mathcal{C}}$ substantiate the conclusion, that outliers primarily affect the estimators of parameters describing the scedasticity of the observations, that are the scale σ and deviation parameters s_1 and s_2.

Therefore, the inference approach is considered as robust for outliers, albeit only for moderate outliers as simulated by the truncated Cauchy distribution $\mathcal{C}_{|5|}$. So far, the robustness of the inference performance to outliers is assessed by taking into account symmetrical distributions. Indeed, outliers in real time series may be asymmetric or singular, and may not be characterized by any distribution at all. To facilitate a robust inference, two common approaches may be employed to better meet the basic assumption of a Gaussian random behavior. In some cases of asymmetric outliers, the observations can be log-transformed to better satisfy the requirement of a symmetric distribution [107]. The aspect of distributional asymmetry is discussed for real observations in the course of measurement processes in Sec. 5.3. In case of singular outliers, an effect analysis can be performed to identify potential extreme observations. By computing the Cook's distance [108, 109] from the standardized time series a measure of influence of each

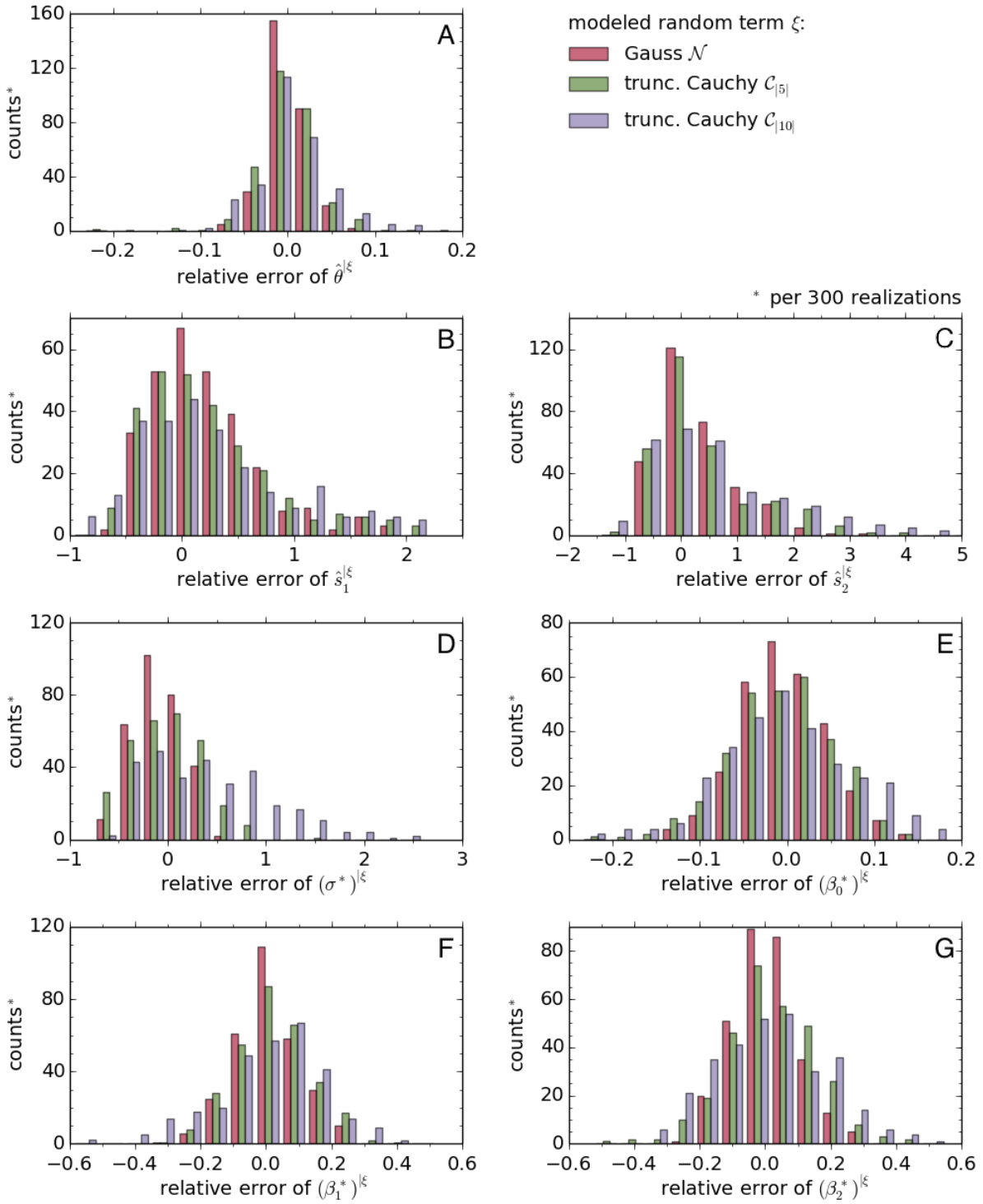


Figure 3.12: *The robustness of the inference performance in the presence of outliers is further deduced from the relative error of the transition model estimates* $(\boldsymbol{\vartheta}^*)_j^{|\boldsymbol{\xi}} = ((\boldsymbol{\beta}^*)^{|\boldsymbol{\xi}}, (\sigma^*)^{|\boldsymbol{\xi}}, \hat{\boldsymbol{s}}^{|\boldsymbol{\xi}}, \hat{\theta}^{|\boldsymbol{\xi}})_j$ *for different random behavior* $\boldsymbol{\xi}$*. The relative errors of the reference estimates* $\hat{\vartheta}^{|\mathcal{N}}$ *and the estimates* $\hat{\vartheta}^{|\mathcal{C}_{|5|}}$ *are of the same order. The relative errors increase for the estimates* $\hat{\vartheta}^{|\mathcal{C}_{|10|}}$ *and exhibit a significant influence of the outliers particularly on the scedasticity estimators scale* $(\sigma^*)^{|\boldsymbol{\xi}}$ *(D) and deviation slopes* $\hat{s}_1^{|\boldsymbol{\xi}}$ *(B) and* $\hat{s}_2^{|\boldsymbol{\xi}}$ *(C).*

data point on the employed estimators is derived. The observations that show extreme distances relative to the majority of the remaining observations may be interpreted as outliers and excluded from the original time series for the further inference process. Clearly, since the introduced Bayesian approach is based on a generic transition model it may not capture the complexity of the real underlying process. Thus, model errors are expected but shall not be solved by manipulating individual data points, such as simply ignoring unusual observations. For the sake of transparency, any local time series $\boldsymbol{y}_{\mathcal{T}'}$ that comprises outliers is deliberately not taken into account within the proposed inference process, since under this conditions the generic transition model is considered as invalid. This aspect is systematically incorporated in the expansion of the detection approach to a kernel-based inference approach explained in Sec. 4.2 by introducing a normality indicator function. The assessment whether the assumed stochastic process hold for a time series is discussed for the application of the approach to real observations in the following section.

3.4 Application on direct climate observations

The introduced Bayesian inference approach is applied to direct environmental observations comprising documented transition events: a hydrological time series of the Nile river in Egypt and a temperature series from the weather station in Tuscaloosa, Alabama. The meta data of the measurement stations report on specific events directly influencing the measurand. Each time series consists of 100 observations and since the transitions are known to be rather abrupt than gradual, the generic shift transition model $\mathcal{M}_{shift}$ is assumed as defined in Eq. 3.7. Given the information about the true transition time, the efficiency of the inference approach to detect the underlying change can be tested. More importantly, in contrast to the performance studies on synthetic observations the validity of the generic transition model as an adequate approximation to a real world process can be investigated.

Consideration of the generic model assumptions

The Bayesian inference approach is carried out with respect to the assumptions of the generic transition model. The derived estimates $\boldsymbol{\vartheta}^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})$ of the model parameters can be used to transform the investigated observations $\boldsymbol{y}_{|\mathcal{T}'}$ into

$$\boldsymbol{y}'_{|\mathcal{T}'} = \sqrt{\left(\sigma^* \, \Omega_{\hat{\theta},\hat{\boldsymbol{s}}}\right)^{-1}} \left(\boldsymbol{y}_{|\mathcal{T}'} - F_{\hat{\theta}} \boldsymbol{\beta}^*\right) \;\sim\; \mathcal{N}(0, \mathbf{1}) \,, \tag{3.53}$$

such that the obtained time series $\boldsymbol{y}'_{|\mathcal{T}'}$ is standard normally distributed. However, this holds only if the real random process is sufficiently close to a normal distribution. Due to the generic formulation of the model as a robust first order approach, any departure from normality is assumed to be primarily attributed to a differing random behavior of the observed process. Therefore, the suitability of the generic transition model as an adequate approximation to the underlying process can be assessed by considering the validity of the normality assumption for $\boldsymbol{y}'_{|\mathcal{T}'}$.

A common approach to validate the assumption of normality, offers the Shapiro-Wilk hypothesis test (SWT) [110].[5] Compared to other normality tests, the SWT performs with constant quality over a wide range of sample sizes [111] which is a crucial concern for the kernel-based expansion of the inference approach (see Sec. 4). For large sample sizes, however, the SWT becomes very sensitive to outliers. An alternative strategy to assess departures from normality yields the direct investigation of the higher statistical moments m_i of order i as defined in App. A.3. Instead of using test statistics that generally employ two statistical moments, the aim is to use the moments m_i for $i = 1, \ldots, 4$ as qualitative measures for departures from normality. The theoretical moments for a standard, central normal distribution and the interpretation thereof are summarized

[5] The general advantage of the SWT is, that due to the general formulation of the null hypothesis the approach does not require the observations to be of an explicitly parametrized distribution. Nevertheless, the mean and deviation are required to be constant over time.

Table 3.4: *The higher statistical moments can be used to deduce the actual characteristics of a given distribution. By comparing the theoretical statistical moments of the standard normal distribution with the empirical moments derived from the standardized observations of the Nile (Fig. 3.13 (C)) and the Tuscaloosa (Fig. 3.14 (C)) time series, the validity of the normality assumption can be assessed and relative differences described.*

statistical moment	theoretical value	empirical value Nile	empirical value Tusc.	descriptive characteristics	
m_1	0	$8 \cdot 10^{-16}$	$-6 \cdot 10^{-15}$	**mean**	
m_2	1	0.97	0.97	**variance**	
m_3	0	0.02	-0.31	**skewness**	< 0: left-skewed > 0: right-skewed
m_4	3	3.32	4.12	**kurtosis**	< 3: round peak, thin tails > 3: sharp peak, fat tails

in Tab. 3.4. Any deviation of the empirical from the theoretical statistical moments indicates a departure from normality. Furthermore, the values of the empirical moments provide valuable information about the kind of difference of the observations' random behavior relative to the standard normal distribution. In this way, the set of empirical statistical moments serve as a diagnostic tool to justify the application of the generic transition model to a specific real world times series.

3.4.1 Annual Nile river flow (1871-1970) from Aswan, Egypt

A prominent time series to study changes within environmental data depicts the annual water discharge of the Nile river measured at the Aswan dam in Egypt. Based on Nile river observations, Harold Edwin Hurst for example formulated an environmental memory process parametrized by the Hurst coefficient in order to predict maximal river levels. Here, the Nile river flow between the years 1871 to 1970 is investigated (see Fig. 3.13 (B)). Historical records provide the fact, that in the year 1899 a shift in the flow levels is attributed partly to weather changes and partly to the start of construction work for a new dam at Aswan. Several investigation methods have verified a change underlying the observed flow levels in 1899 [55, 112–114]. Thus, the observations offer an established study case for the proposed detection approach.

The Bayesian inference is applied by computing the posterior distribution $p(\theta, \boldsymbol{s}|\boldsymbol{y}_{|\mathcal{T}})$ in the sampling space defined in Tab. 3.5 given the complete time series $\boldsymbol{y}_{|\mathcal{T}}$ for $\mathcal{T} = [1871\,,\, 1970]$. The marginal posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ of the transition parameter θ is presented in Fig. 3.13 (A). The MAP estimate $\hat{\theta}$ indicates the most probable transition year as $\hat{\theta} = 1898$. The confidence interval $\mathrm{CI}_{0.95}$ of the posterior distribution localize the occurrence of the transition into a time interval of about four years between 1896 and

Table 3.5: *The parameters of the generic transition model $\mathcal{M}_{shift}$ are estimated from the annual Nile river flow (Fig. 3.13) by performing the inference approach in the sampling space $\Delta_\theta \times \Delta_{s_1} \times \Delta_{s_2}$.*

parameter	MAP		PLH/BLUP	sampling space	
ϑ_i	estimate	$\mathrm{CI}_{0.95}$	estimate	range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	1898.0 a	[1896.0 , 1899.5] a	-	[1875.0, 1965.0] a	0.5 a
s_1	0.007	[-0.014 , 0.042]	-	[-0.030, 0.070]	0.001
s_2	-0.001	[-0.006 , 0.007]	-	[-0.030, 0.070]	0.001
σ	-	-	0.128	-	-
β_0	-	-	1.119	-	-
β_1	-	-	-0.002	-	-
β_2	-	-	0.001	-	-
β_3	-	-	0.825	-	-

1900. Based on the derived point estimates $\boldsymbol{\vartheta}^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})$ of the model parameters (see Tab. 3.5) the estimated transition model $\hat{\mathcal{M}}_{shift}$ is visualized in Fig. 3.13 (B). The inferred model reveals a correlated transition in mean and variability, meanwhile the confidence area $\mathrm{CA}_{0.95}$ of the posterior distribution $p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}})$ of the deviation slopes also encloses homoscedastic behavior, i.e. the combination $s_1 = s_2 = 0$.

The assumption of a generic transition approach needs to be justified. To study the investigated time series for potential departures from normality, the Nile observations are standardized and centralized by the linear transformation given in Eq. 3.53 (see Fig. 3.13 (C)). At a significance level of $\alpha = 0.05$, the SWT yields the p-value = 0.82 such that the assumption of a normal distribution can not be rejected for the transformed observations $\boldsymbol{y}'_{|\mathcal{T}}$. Thus, the generic transition model apparently captures the underlying evolution of the stochastic process sufficiently. In addition, the empirical higher statistical moments of the transformed observations are compared to the theoretical values of a standard, central normal distribution (see Tab. 3.4). The deviation in mean m_1 is practically zero whereas the variance m_2 apparently reflects the inherent bias of the scedasticity parameters discussed in Sec. 3.3.2. Due to the finding of the SWT, the deviations in skewness m_3 and kurtosis m_4 is considered as insignificant.

In conclusion, the Bayesian approach correctly infers on the true transition year at 1899 and localizes the transition occurrence between the years 1896 and 1900. The considerations of normality confirm the generic transition model as an adequate approximation to the underlying process patterns of the Nile river flow.

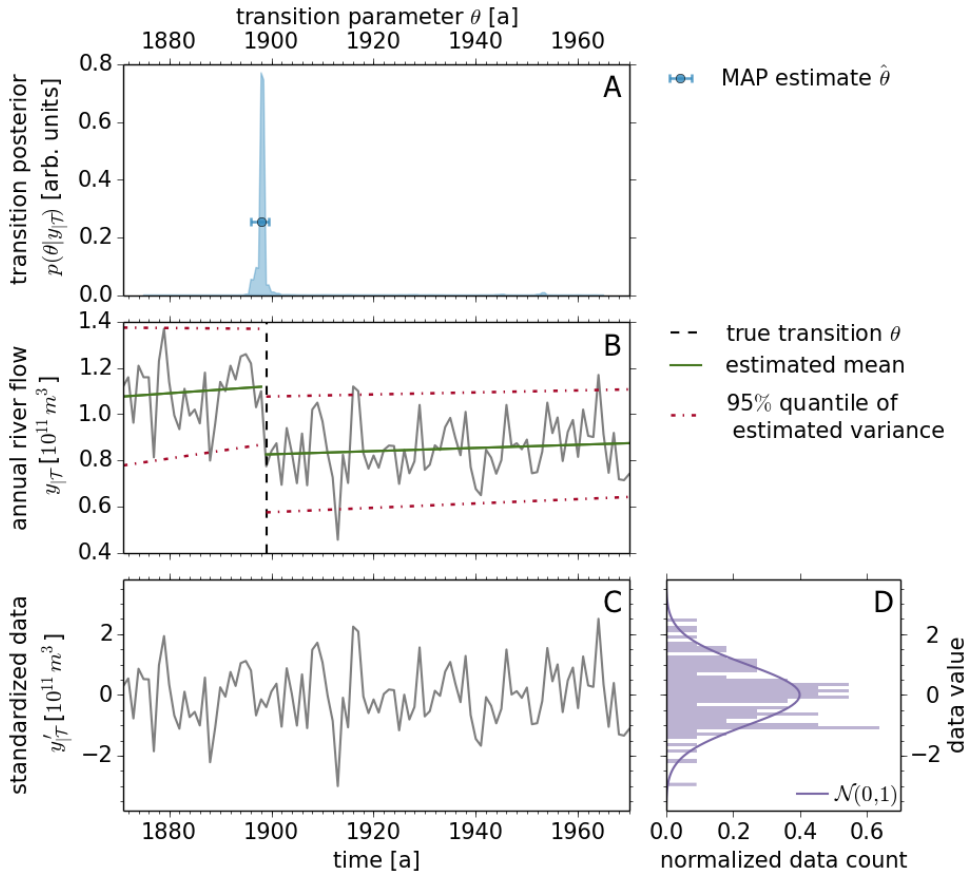


Figure 3.13: *The annual Nile river flow (B) comprises a documented flow shift in the year 1899. (A) The posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ of the transition parameter θ given the observations $\boldsymbol{y}_{|\mathcal{T}}$ localizes the change with 95% uncertainty within the period* $[1896\,,\,1900]$*. (B) The estimated underlying model reveals a significant shift in the mean and a minor change in the scedastic behavior of the data. (C) The standardized time series $\boldsymbol{y}'_{|\mathcal{T}}$ is used to investigate the validity of the assumed generic transition model. (D) The departure from normality is qualitatively compared to a standard normal distribution $\mathcal{N}(0,1)$ and quantitatively assessed by the empirical statistical moments in Tab. 3.4. Both approaches support the assumption of normality.*

3.4.2 Annual average temperature (1901-2000) from Tuscaloosa, Alabama

The investigation of trends in direct climate observations often requires a preliminary homogenization of the time series to assess discontinuities potentially affecting the analysis [115]. On this account, change point detection techniques are employed to locate inhomogeneities. Thereby, a suitable study case offers the time series of the annual mean temperature measured at the weather station in Tuscaloosa, Alabama [35–37] (see Fig. 3.14 (B)). Based on the meta data of the measurement station the time series comprises multiple transitions as a consequence of station relocations or modifications of the measurement instrumentation documented for the years 1921, 1939, 1956 and 1987. Despite the fact that more events underlie the observations than assumed by the generic transition model, the proposed detection approach is applied to the time series in order to investigate its performance.

The Bayesian inference is applied by computing the posterior distribution $p(\theta, \boldsymbol{s}|\boldsymbol{y}_{|\mathcal{T}})$ in the sampling space defined in Tab. 3.6 given the complete time series $\boldsymbol{y}_{|\mathcal{T}}$ for $\mathcal{T} = [1901\,,\,2000]$. The derived marginal posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ of the transition parameter θ (see Fig. 3.14 (A)) is of unimodal shape and localizes the change within a narrow interval of about one year. The MAP estimate $\hat{\theta}$ =1957 confirms previous studies, all inferring on a significant change in the year 1957 and interpreting the outcome as the documented transition in the year 1956 [35–37]. The estimated transition model $\hat{\mathcal{M}}_{shift}$ (see Fig. 3.13 (B)) is obtained from the point estimates $\boldsymbol{\vartheta}^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})$ of the model parameters (see Tab. 3.6) and reveals a distinct shift in the mean as well as in the variability of the time series. The confidence area $\mathrm{CA}_{0.95}$ of the posterior

Table 3.6: *The parameters of the generic transition model $\mathcal{M}_{shift}$ are estimated from the annual temperature in Tuscaloosa (Fig. 3.14) by performing the inference approach in the sampling space $\Delta_\theta \times \Delta_{s_1} \times \Delta_{s_2}$.*

parameter	MAP		PLH/BLUP	sampling space	
ϑ_i	estimate	$\mathrm{CI}_{0.95}$	estimate	range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	1957.5 a	[1957.0 , 1957.5] a	-	[1905.0, 1995.0] a	0.5 a
s_1	0.032	[0.008 , 0.082]	-	[-0.020, 0,180]	0.002
s_2	0.026	[0.002 , 0.072]	-	[-0.020, 0,180]	0.002
σ	-	-	0.331	-	-
β_0	-	-	18.022	-	-
β_1	-	-	-0.006	-	-
β_2	-	-	0.031	-	-
β_3	-	-	16.629	-	-

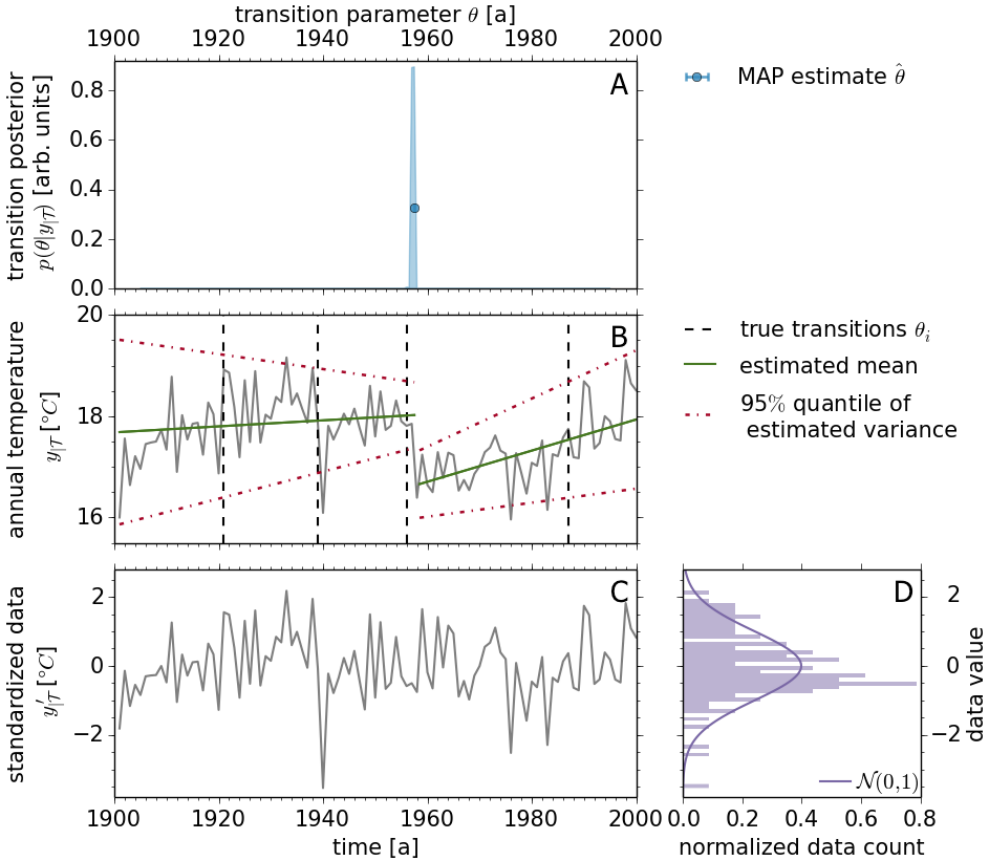


Figure 3.14: *The annual temperature measured in Tuscaloosa (B) comprises multiple documented changes in the years 1921, 1939, 1956 and 1987, but is investigated here with respect to a single transition model. (A) The posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{T}})$ of the transition parameter θ given the observations $\boldsymbol{y}_{|\mathcal{T}}$ localizes the change with 95% uncertainty within the period* [1956 , 1957]. *(B) The estimated underlying model reveals a significant transition in the mean as well as in the scedastic behavior of the data. (C) The standardized time series $\boldsymbol{y}'_{|\mathcal{T}}$ is used to investigate the validity of the assumed generic transition model. (D) The departure from normality is qualitatively compared to a standard normal distribution $\mathcal{N}(0,1)$ and quantitatively assessed by the empirical statistical moments in Tab. 3.4. Both approaches indicate a left-skewed distribution with a long tail and the applied SWT rejects the assumption of normality.*

distribution $p(s_1, s_2|\boldsymbol{y}_{|\mathcal{T}})$ of the deviation slopes only includes combinations that lead to a heteroscedastic behavior, i.e. $s_i \neq 0$ for $i = 1, 2$.

The standardized observations $\boldsymbol{y}'_{|\mathcal{T}}$ are presented in Fig. 3.14 (C). The SWT yields the p-value $= 0.02$ such that the assumption of normality is rejected at a significance level of $\alpha = 0.05$. Therefore, the assumption of normality and consequently the assumed generic transition model, does not sufficiently capture the underlying process patterns of the annual temperature.

Furthermore, the empirical higher statistical moments are computed for the transformed observations (see Tab. 3.4). The deviation in mean m_1 is practically zero, whereas

the variance m_2 has the exact same value as for the Nile observations. As discussed in Sec. 3.3.2 the extent of the inherent bias of the scedasticity parameters depends on the sample size of the data, which are similar for both investigated time series. The deviations in skewness m_3 and kurtosis m_4 reflect the shape of the count data, i.e. left-skewed with a long tail. In this way, the explicit deviation from the theoretical moments may help to characterize the random behavior of the observations relative to the standard normal distribution.

In conclusion, the approach infers correctly on one of the four underlying transitions in the year 1957 interpreted similar to previous studies as the documented change in the year 1956. As expected and as indicated by the departure from normality, the generic transition model does not capture the real underlying process of multiple changes as documented in the meta data of the Tuscaloosa weather station.

3.5 Discussion and Summary

In this chapter, a generic transition model is designed to locally approximate a change in complex natural observations. Based on the introduced first order approach a Bayesian inference is implemented to derive the degree of belief about the transition location and to estimate the underlying transition patterns. In order to investigate the performance of the approach, synthetic time series of different change geometries are analyzed. Irrespective of the underlying geometry, that is irrespective of the observational evidence of the change, the approach convincingly reproduces the transition patterns. However, for an increasing variability at the transition, the efficiency of the approach to localize the change decreases. In general, the approximation of real complex observations by a generic transition model is only reasonable, if the inference performance is robust for sparse sampling and in the presence of model errors. By studying the convergence of the posterior distribution and the sensitivity of the estimators for varying sample sizes a minimal required amount of 50 observations is assessed. Moreover, the inference performance is robust against symmetric outliers as simulated by truncated Cauchy distributions. Common to all performed inference studies, the estimator of the transition parameter shows to be most robust, whereas the scedasticity estimators of scale and deviation show to be most sensitive of all employed model estimators.

The introduced approach is finally applied to real time series comprising documented transitions. To justify the approximation of the actual observations with the generic model, the empirical higher statistical moments are used to characterize the random behavior of the observations relative to the assumed normal behavior. For the annual Nile river flow, the approach correctly infers on the underlying change and the statistical moments indicate minor departures from the assumed normality. For the annual temperature measured at Tuscaloosa, the approach infers on one out of four underlying changes. The observations reveal a departure from the model assumptions and hence, the approximation by the generic transition model is correctly considered as not adequate

to the temperature series.

Based on these findings, the introduced Bayesian inference is proposed as a practical and robust transition detection approach to real time series comprising a single change. Despite its simplicity the generic transition model yields valuable information on the underlying transition patterns and random behavior. Nevertheless, to investigate complex observations the local inference approach needs to be generalized in order to capture multiple transitions in a time series. Therefore, in the following chapter a kernel-based extension of the proposed local detection approach is introduced in principle capable to investigate a variety of natural processes.

4 Detection of multiple transitions in time series

The major challenge of change point detection is, that for real world observations neither number nor location of the underlying transitions are *a priori* known. A stochastic process describing multiple changes is naturally high dimensional and based on specific assumptions on the number and scale of the transitions. Instead of defining a specified and complex process model to infer on multiple events, the introduced detection approach based on a generic transition model is extended. Applying the inference approach globally on the complete time series may not be justified. However, locally the generic model assumption may still be valid as a first order approach. For this reason, the generic inference is systematically applied to investigate local sub series of the observations in terms of a kernel-based approach. By employing specific information criteria, the derived *local* posterior distributions may be combined into a *global* proxy distribution expressing the credibility of a transition for each investigated time point. Clearly, the kernel-based approach does not yield an estimate of a closed global stochastic model of arbitrary transitions, respectively arbitrary complexity. Nevertheless, it offers a powerful tool to reveal the degree of belief about multiple changes and to visualize the locally inferred generic transition patterns. As a beneficial side effect a kernel-based approach proves to be numerically more efficient compared to a global inference approach.

In order to localize the transition detection on a general time grid, a kernel-based reformulation of the inference approach introduced in Chap. 3 is elaborated. To adequately assemble the individual posterior distributions, an information criterion is proposed that indicates the existence of a transition within each kernel. By using the criterion as a kernel weight a credibility expression is derived that is considered as a proxy probability of transitions occurring at the employed kernel scale. The scale dependent investigation of synthetic time series demonstrates, that a further kernel weight indicating the local suitability of the generic model facilitates the inference on multiple transitions. Hence, the extended approach provides a proxy probability indicating multiple transitions and their inferred patterns for observations investigated at a given kernel scale. Finally, the introduced kernel-based Bayesian method is applied to direct environmental observations comprising documented transition events: a hydrological time series of the Nile river in Egypt and a temperature series from the weather station in Tuscaloosa, Alabama. The inference results are compared to the previous analysis in Sec. 3.4 and the applicability and limitation of the kernel-based approach is critically discussed.

4.1 Kernel-based inference approach

The previously introduced detection approach locally infers on a generic transition model $\mathcal{M}_{trans}$ within a given central sub interval $\mathcal{T}'$ of the complete time series of length $\mathcal{T}$.[1] For the purpose of exploring real observations for an unknown transition, the approach needs to be generalized to flexibly investigate different sub series of the observations. Therefore, a global time grid Θ_t spanning the whole time series $\boldsymbol{y}$ is defined. Around each of these time grid points t a time window $\mathcal{I}_t = [t - \frac{\lambda}{2}, t + \frac{\lambda}{2}]$ of length λ is constructed. Inside the corresponding data window $\boldsymbol{y}_{|\mathcal{I}_t}$, the prior distribution of the transition parameter θ is taken as a flat prior distribution

$$p_t(\theta) = \begin{cases} \frac{1}{\lambda'} & \text{for } \; t - \frac{\lambda'}{2} \leq \theta \leq t + \frac{\lambda'}{2} \\ 0 & \text{else} \end{cases} \quad \text{with } \; 0 < \lambda' < \lambda \,, \tag{4.1}$$

inside some sub interval of length λ' centered around the query time t. Note, that the ratio λ/λ' is constant throughout the investigation of a time series. The resulting local observations $\boldsymbol{y}_{|\mathcal{I}_t}$ may be interpreted as kernels of neighborhood $\mathcal{I}_t$ with the kernel support $p_t(\theta) \neq 0$ (based on Ref. [98]). Hence, the joint posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{y}_{|\mathcal{I}_t})$ of the generic transition model given by Eq. 3.26 is localized with respect to a query time t at a temporal scale λ as

$$p(\boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{I}_t}) \propto \mathcal{L}(\boldsymbol{y}_{|\mathcal{I}_t}; \boldsymbol{\beta}, \sigma, \boldsymbol{s}, \theta) \cdot p(\boldsymbol{\beta}) \cdot p(\sigma) \cdot p(\boldsymbol{s}|\theta) \cdot p_t(\theta) \,. \tag{4.2}$$

The kernel posterior distribution yields the kernel marginal distribution $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$ of the transition parameter as given by Eq. 3.30 and the estimated kernel transition model given by the set of estimators $\boldsymbol{\vartheta}_t^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})_t$ as summarized in Eq. 3.38.

However, for real world observations neither the kernel location t nor the scale λ, at which the time series may be locally approximated by the generic transition model, are known. As a pragmatic approach, the inference is systematically applied within kernels sampled across the complete time grid and with respect to multiple scales. For each scale, the individual kernel results are then combined to comprehensively deduce on multiple transitions and their local generic transition patterns as explained in the following.

Synthesis of the kernel posterior distributions $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$

In the kernel-based approach the Bayesian inference is applied in an automated way on each kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$. To combine the individual kernel results, a kernel weight $f(t)$ is required to quantify the credibility of the existence of a transition θ within

[1] Here, the generic transition models $\mathcal{M}_j$ with $j = [\text{break}, \text{shift}]$ defined in Sec. 3.1 are generally referred to as $\mathcal{M}_{trans}$.

the kernel support $p_t(\theta) \neq 0$. The kernel posterior distributions $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$ can then be integrated over the complete time grid Θ_t as a weighted superposition

$$p(\{\theta\}|\boldsymbol{y}^{|\lambda}) := C \cdot \int_{\Theta_t} f(t') \cdot p(\theta|\boldsymbol{y}_{|\mathcal{I}_{t'}}) dt' \,, \tag{4.3}$$

whereas the constant C ensures the normalization to a probability density. The derived probability expression is interpreted as a proxy probability of the transitions $\{\theta\} = (\theta_1, \theta_2, ...)$ underlying the time series $\boldsymbol{y}^{|\lambda}$ at the scale of investigation λ. The proxy probability does not provide an estimator for the number k of inferred transitions but assigns a degree of belief about a generic transition to each investigated time point t.

To elucidate the interpretation of Eq. 4.3 as a proxy probability, let a time series $\boldsymbol{y}$ be described by a stochastic model $\mathcal{M}_{k-trans}$ containing k transitions and be locally differentiable up to the first order prior and after each transition θ_i. Moreover, for a local scale $\tilde{\lambda}$ let maximal one transition θ_i be in the inner prior support $p_t(\theta)$ of each kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$. In the limit of the kernel scale λ against the local scale $\tilde{\lambda}$

$$\lim_{\lambda \to \tilde{\lambda}} p(\{\theta\}|\boldsymbol{y}^{|\lambda}, \mathcal{M}_{trans}) \to \sum_{i=1}^{k} p(\theta_i|\boldsymbol{y}, \mathcal{M}_{k-trans}) \,, \tag{4.4}$$

the proxy probability based on the generic transition model $\mathcal{M}_{trans}$ is assumed to converge towards the sum of posterior distributions of the individual transitions θ_i based on the high dimensional k-transition model $\mathcal{M}_{k-trans}$. In this way, the proxy probability $p(\{\theta\}|\boldsymbol{y}^{|\lambda})$ may be used as a diagnostic tool to investigate complex signals $\boldsymbol{y}$, that may be locally approximated by the generic transition model at the scale λ, for multiple changes $\{\theta\}$.

Nevertheless, the assumed proxy convergence is only reasonable, if an efficient kernel measure $f(t)$ is employed. In order to illustrate the need of a kernel weight, a schematic example of the kernel-based approach is presented in Fig. 4.1 for a synthetic time series comprising a single transition. The explicit setting is provided in Tab. B.4. The naïve integration over the kernel posterior distributions $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$ yields a multimodal probability expression. Even though the scale λ encloses maximal one transition per kernel support, the underlying transition can not be efficiently detected. To precisely infer on the underlying transition a kernel weight is required that quantifies the credibility that a transition actually exists within a kernel support.

The maximum likelihood function $\mathcal{L}_{max}(t)$ indicates the goodness of fit of the generic transition model $\mathcal{M}_{trans}$ to the kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$. For kernels that do not enclose a transition in their support $p_t(\theta) \neq 0$, the goodness of fit is very low and the maximum of the likelihood function is of very small value as shown in Fig. 4.2 (A). Thus, $\mathcal{L}_{max}(t)$ may be interpreted as an approximate measure for the existence of a generic

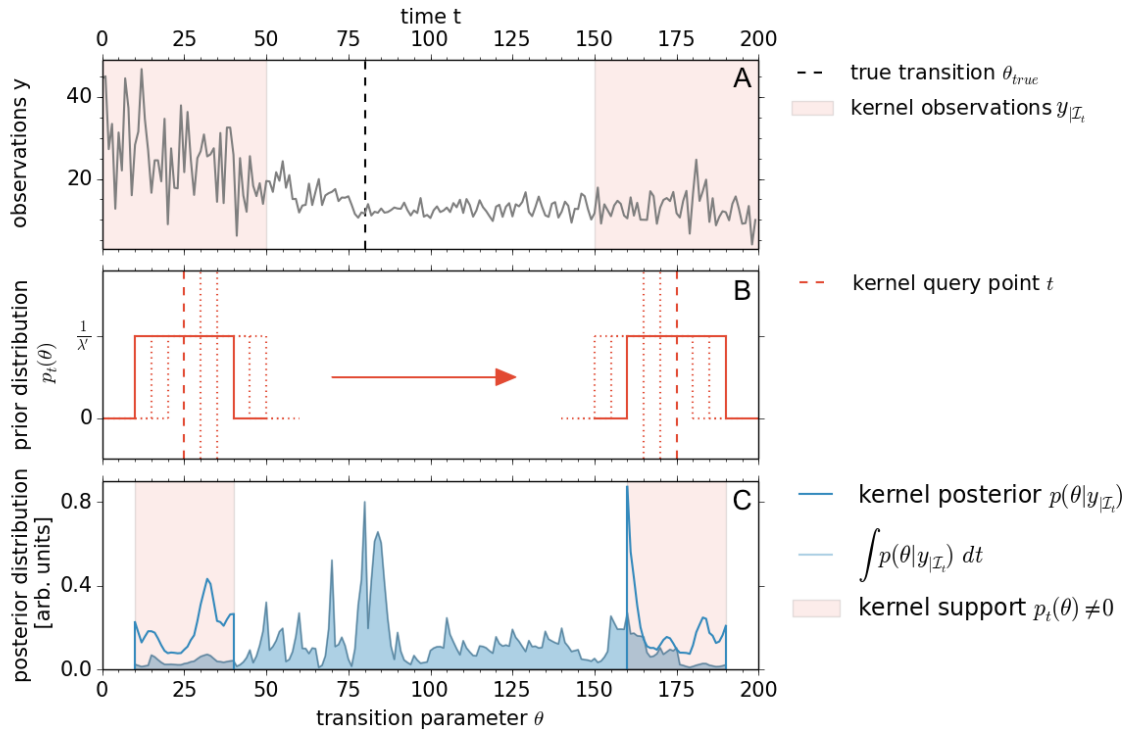


Figure 4.1: *In the kernel-based approach the transition inference is systematically applied on sub series of the complete time series to locally investigate the observations for an unknown transition. (A) The synthetic time series* $\boldsymbol{y}$ *comprises a transition at* $\theta = 80$ *as defined in Tab. B.4. (B) The kernels are of length* $\lambda = 50$ *and centered at a query time* t*. The kernel support of length* $\lambda' = 30$ *is given by a time dependent flat prior distribution* $p_t(\theta)$ *of the transition parameter* θ *as defined in Eq. 4.1. (C) Based on the kernel data* $\boldsymbol{y}_{|\mathcal{I}_t}$*, the kernel posterior distributions* $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$ *are computed. The plain sum over all kernel posterior distributions fails to localize the underlying transition* θ_{true}*, since the basic assumption that a transition actually exists is not true for every kernel* $\boldsymbol{y}_{|\mathcal{I}_t}$*.*

transition θ within the kernel $\boldsymbol{y}_{|\mathcal{I}_t}$. Based on Eq. 3.17, the maximum of the likelihood function can be obtained as

$$f(t) = \max_{\theta \in [t-\frac{\lambda'}{2}, t+\frac{\lambda'}{2}], \boldsymbol{s} \in [\boldsymbol{s}_{min}(\theta), \infty]} \mathcal{L}(\boldsymbol{y}_{|\mathcal{I}_t}; \boldsymbol{\beta}^*, \sigma^*, \boldsymbol{s}, \theta) = \mathcal{L}_{max}(t)\,, \tag{4.5}$$

by using the BLUP $\boldsymbol{\beta}^*$ (Eq. 3.20) and the PLH σ^* estimators (Eq. 3.18). In this way, the derived proxy probability in Eq. 4.3 may be interpreted as

$$p(\{\theta\}|\boldsymbol{y}^{|\lambda}) \mathrel{\widehat{=}} \sum_{\text{kernels}} (\text{goodness of } \mathcal{M}_{trans} \text{ fit in kernel}) \cdot p(\theta|\text{trans. in kernel})\,, \tag{4.6}$$

that is the product of the goodness of transition fit on the kernel observations $\boldsymbol{y}_{\mathcal{I}_t}$ and the probability of the transition location $p(\theta|\boldsymbol{y}_{\mathcal{I}_t})$ under the assumption that a transition occurs in the kernel support $p_t(\theta) \neq 0$. The efficiency of $\mathcal{L}_{max}(t)$ as a kernel weight is demonstrated in Fig. 4.2 (C). In general, each continuous interval, in which the proxy

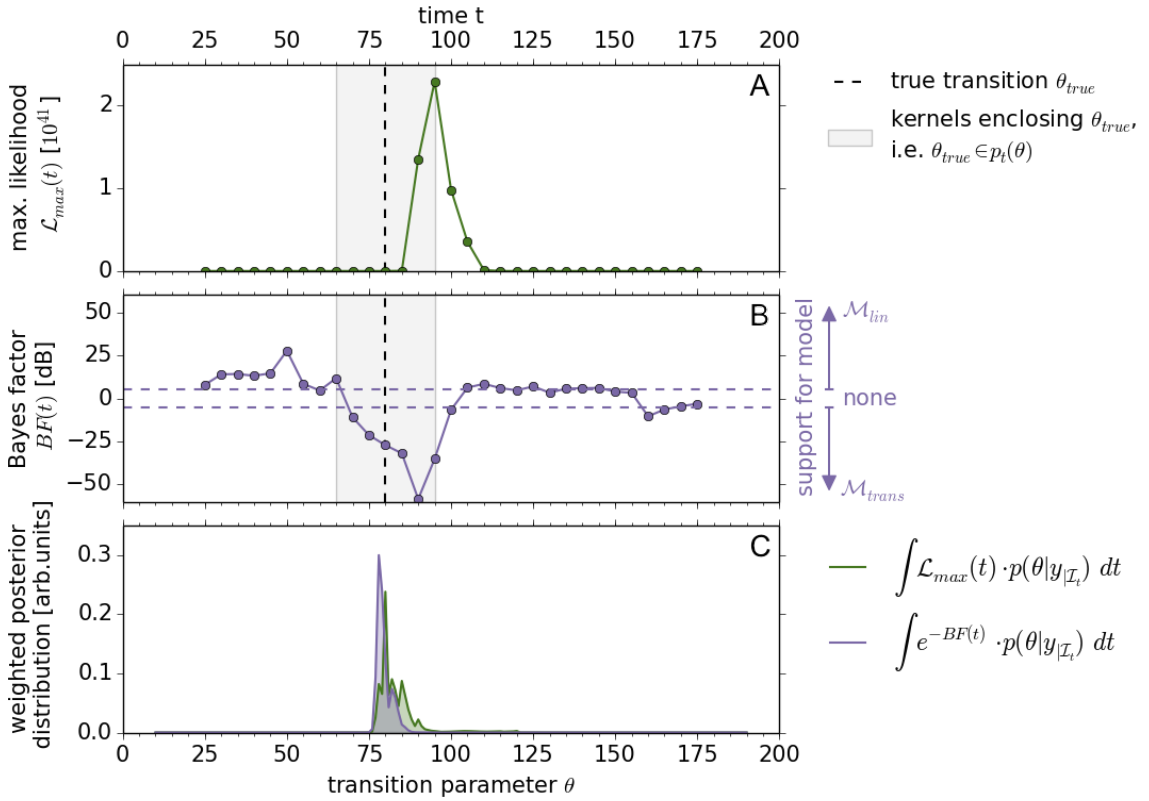


Figure 4.2: *To precisely infer on the underlying transition in Fig.4.1 (A) kernel weights are proposed in order to account for the credibility that a transition occurs within a kernel support $p_t(\theta) \neq 0$. The kernel weights $f(t)$ are presented over the kernel query points t of the employed time grid Θ_t. (A) The maximum of the kernel likelihood function $\mathcal{L}_{max}(t)$ indicates the goodness of fit of the generic transition model $\mathcal{M}_{trans}$ to the kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$. (B) The kernel Bayes factor $BF(t)$ indicates the support for either a generic transition model $\mathcal{M}_{trans}$ or a model $\mathcal{M}_{lin}$ without any transition. (C) By weighting each kernel posterior distribution $p(\theta|\boldsymbol{y}_{|\mathcal{I}_t})$ with the proposed quantities $f(t) = [\mathcal{L}_{max}(t), e^{-BF(t)}]$, the derived proxy probabilities accomplish to localize the underlying transition θ_{true}.*

probability $p(\{\theta\}|\boldsymbol{y}^{|\lambda})$ is non-zero, indicates the occurrence of an individual transition event. The resulting proxy probability therefore localizes a transition within the interval $[78, 121]$ enclosing the true underlying event. As a consequence of the artificial synthesis of the kernel posterior distributions, the assessment of confidence intervals $\mathrm{CI}_{1-\alpha}$ for the proxy probability is not feasible anymore.

The Bayes factor $BF(t)$ indicates the goodness of fit and the complexity of the generic transition model $\mathcal{M}_{trans}$ with respect to the kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$ and a competing model $\mathcal{M}_{lin}$ without any transition. Based on Eq. 2.23, the Bayes factor realizes a model test between $\mathcal{M}_{lin}$ and $\mathcal{M}_{trans}$ as

$$BF(t) = 10 \cdot log_{10} \left(\frac{p(\mathcal{M}_{lin}|\boldsymbol{y}_{|\mathcal{I}_t})}{p(\mathcal{M}_{trans}|\boldsymbol{y}_{|\mathcal{I}_t})} \right) , \tag{4.7}$$

here formulated in deciban. Both model assumptions are considered *a priori* as equally probable with $p(\boldsymbol{\vartheta}|\mathcal{M}_{trans}) = p(\boldsymbol{\vartheta}|\mathcal{M}_{lin})$. Thus, the kernel Bayes factor $BF(t)$ indicates the credibility of the generic transition model $\mathcal{M}_{trans}$ as a suitable local approach to the kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$ by its actual value [80]:

$$BF(t) \begin{cases} > 5 & \text{supports } \mathcal{M}_{lin}, \\ < -5 & \text{supports } \mathcal{M}_{trans}, \\ \text{else} & \text{no substantial support.} \end{cases} \tag{4.8}$$

The dependency of the Bayes factor from the kernel location t is presented in Fig. 4.2 (B). In all kernels, for which the true transition θ_{true} is in the support of the inner prior distribution $p_t(\theta) \neq 0$, the Bayes factor $BF(t)$ favors the transition model $\mathcal{M}_{trans}$ over the linear model $\mathcal{M}_{lin}$. Hence, the kernel Bayes factor itself can be used as a diagnostic tool like the likelihood weighted proxy probability, but the techniques may be also combined by using the Bayes factor as a kernel weighting function in Eq. (4.3) by setting

$$f(t) = e^{-BF(t)} \,. \tag{4.9}$$

In this form the proxy probability of transitions corresponds essentially to the total probability decomposition of the transition by

$$p(\{\theta\}|\boldsymbol{y}^{|\lambda}) \mathrel{\widehat{=}} \sum_{\text{kernels}} p(\text{trans. exists in kernel}) \cdot p(\theta|\text{trans. in kernel}) \tag{4.10}$$

into the probability of the existence of a transition in the kernel support $p_t(\theta) \neq 0$ and the probability of the transition location $p(\theta|\boldsymbol{y}_{\mathcal{I}_t})$ under the assumption that a transition occurs in the kernel support. The efficiency of the kernel weight based on $BF(t)$ is demonstrated in Fig. 4.2 (C). The resulting proxy probability indicates a transition within the interval $[78, 89]$, an apparently more precise localization than for the likelihood weighted probability. A detailed comparison of the influence of the proposed kernel weights on the localization of an underlying transition is discussed in the next section for multiple changes.

Computational costs

A major advantage of the introduced kernel-based approach, even in a single transition context, is the considerable speed-up of the computational process. The generic approach allows to treat long time series numerically more efficient since the computation scales with the complexity of the algorithm. A qualitative order of the complexity may be estimated by the most time consuming computational step incorporated and thus dominating the performance. In the introduced algorithm a matrix inversion is the most expensive step scaling approximately with the third power $\mathcal{O}(n_t^3)$ of the number of observations n_t within a kernel $\boldsymbol{y}_t^{|\lambda}$. To further speed up the computations, the individual

kernel inferences may be executed as independent processes and thus easily distributed on a computer cluster.

For instance, for a time series with $n = 2000$ observations the *global* computation of the posterior distribution $p(\boldsymbol{s}, \theta|\boldsymbol{y}_{\mathcal{T}})$ in 3 h 41 min 40 s pass to *local* computations of $p(\boldsymbol{s}, \theta|\boldsymbol{y}_{|\mathcal{I}_t})$ in 40 overlapping kernels with $n_{|\mathcal{I}_t} = 100$ kernel observations in 7 min 44 s. In average, this leads to a reduction in computational time of about 95% [116]. The example is performed using Python 2.6.5 on a Supermicro Intel(R) Core(TM)i7 CPU 920 @ 2.68 GHz with 12 GB RAM. In the context of complex multiple transition scenarios, as real time series mostly are, the proxy probability $p(\{\theta\}|\boldsymbol{y}^{|\lambda})$ realizes a powerful tool to investigate observations for local generic transition events $\{\theta\}$ at different temporal scales λ, as implemented in the following.

4.2 Extension of the kernel-based approach to infer on multiple transitions

The assumed convergence of the proxy probability in Eq. 4.4 elucidates the condition on the kernel-based detection approach required to infer on multiple transitions: Each kernel should enclose maximal one transition event. In principle, consecutive transitions θ_i and θ_{i+1} may be resolved, if separated by at least one kernel size λ. The kernel scale λ in turn determines the kernel observations $\boldsymbol{y}_{|\mathcal{I}_t}$, i.e. the number of observations n_t per kernel, to which the generic transition model $\mathcal{M}_{trans}$ is assumed an approximation. As discussed in Sec. 3.3.2, a minimal amount of observations $n_t \geq 50$ is necessary to reliably infer on the generic transition, thus inducing a minimal kernel scale λ. Since for real observations neither number nor location nor distance between the transitions are known, the kernel-based inference approach essentially needs to be applied at multiple scales. To adequately interpret the scale-resolved inference, the dependency between the performance of the kernel-based approach and the employed kernel scale needs to be understood. Thereby, the performance of the kernel-based approach is deduced from the method's ability to localize individual and resolve multiple underlying transitions.

The general influence of the kernel scale on the kernel-based inference performance substantiates the demand of a further kernel weight. Let a local scale $\tilde{\lambda}$ enclose maximal one transition event that may be approached by a generic transition model $\mathcal{M}_{trans}$. And let the employed scale λ of investigation result in kernels that contain at least $n_t = 50$ observations. For $\lambda \gg \tilde{\lambda}$ and in comparison to the underlying event, the estimated transition patterns $\boldsymbol{\vartheta}_t^* = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})_t$ exhibit a small variance due to averaged random behavior but a high bias since the linear approximation may not be suitable. For $\lambda \ll \tilde{\lambda}$, the estimated transition patterns $\boldsymbol{\vartheta}_t^*$ may exhibit a high variance due to random fluctuations but a small bias since the generic model offers an adequate approximation. For $\lambda \to \tilde{\lambda}$, the estimated transition patterns $\boldsymbol{\vartheta}_t^*$ are expected to have the best trade-off between bias and variance from the underlying event (based on Ref. [98]). The

introduced kernel weights $f(t) = [\mathcal{L}_{max}(t), e^{-BF(t)}]$ are used to take into account the existence of a transition. The resulting proxy probability is expected to be unimodal and hence, to clearly localize the underlying transition. However, to indicate the trade-off between bias and variance at each scale of investigation, a further kernel weight $f'(t)$ is required. By additionally penalizing the kernel inference in this way, the derived proxy probabilities $p(\{\theta\}|\boldsymbol{y}^{|\lambda})$ indicate the suitability of the generic transition assumption $\mathcal{M}_{trans}$ with respect to the scale of investigation λ.

For the purpose of elaborating efficient kernel weights, the inference performance of the kernel-based approach is applied in a multiple transition scenario for varying kernel scales. A synthetic time series $\boldsymbol{y}$ of $n = 200$ observations is generated as defined in Tab. B.5 comprising three transition events $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ as presented in Fig. 4.3 (A). The time points $t_1, ..., t_n$ are used as the global time grid Θ_t, respectively as the kernel query points. The employed kernel scales λ are defined within the range Δ_λ and sampled by the step size δ_λ. A constant ratio between the kernel size λ and kernel support λ', i.e. the actual applied sampling range Δ_θ of the transition parameter θ, allows to reveal potential scale influences on the inference results. The complete computational setting of the kernel-based inference approach is given by

$$\begin{array}{lcllcl} \Delta_\lambda & = & [20,\, 200]\,, & \delta_\lambda & = & 10, \\ \Delta_{\lambda'} = \Delta_\theta & = & \frac{3}{5}\lambda\,, & \delta_t = \delta_\theta & = & 1, \\ \Delta_{\boldsymbol{s}} & = & [-0.08,\, 0.44]\,, & \delta_{\boldsymbol{s}} & = & 0.005\,. \end{array}$$

Note, that the sampling step of the transition parameter δ_θ is taken as the time step δ_t of the kernel sampling. The sampling range $\Delta_{\boldsymbol{s}}$ of the deviation parameters $\boldsymbol{s}$ is used as defined in Eq. 3.13. Based on Eq. 3.30, the proxy probability of transitions $\{\theta\}$ is computed at different kernel scales λ and for different kernel weights f as

$$p_f(\{\theta\}|\boldsymbol{y}^{|\lambda}) := C \cdot \sum_{i=t_1+\frac{\lambda}{2}}^{t_n-\frac{\lambda}{2}} f_i^{|\lambda} \cdot p(\theta|\boldsymbol{y}_i^{|\lambda})\,, \tag{4.11}$$

In this way, the optimal kernel weight may be deduced from its influence on the weighted sum of kernel posterior distributions $p(\theta|\boldsymbol{y}_i^{|\lambda})$ of the transition parameter θ.

Kernel weight with respect to transition existence

As proposed in Sec. 4.1, the existence of a transition in a kernel support $p_t(\theta) \neq 0$ may be indicated by the maximum of the kernel likelihood function $\mathcal{L}_{max,t}^{|\lambda}$ or by the kernel Bayes factor $BF_t^{|\lambda}$. The kernel quantities obtained for the synthetic observations are presented with respect to the employed kernel scales λ in Fig. 4.3 (B) and (C). The desired effect of these quantities is the minimization of the proxy probability in between consecutive transitions such that individual events are clearly resolved. This may be

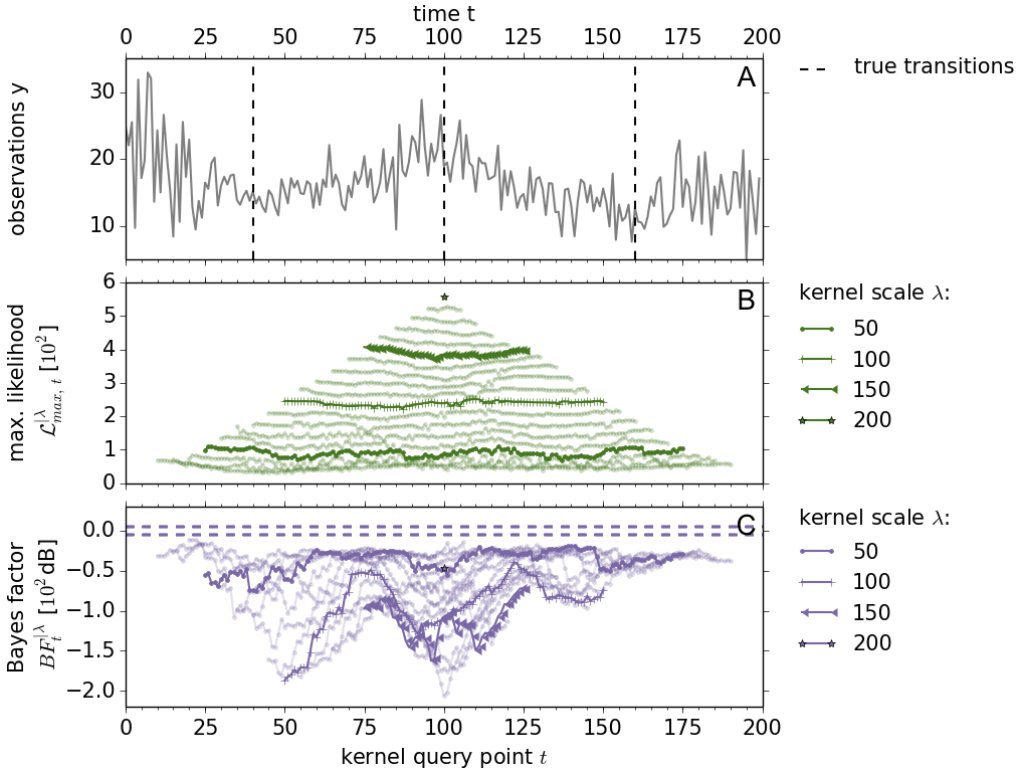


Figure 4.3: *The kernel-based inference approach is applied to a multiple transitions scenario to investigate its performance with respect to the kernel scale. (A) The synthetic time series comprises three transitions at* $\boldsymbol{\theta} = (40, 100, 160)$ *as defined in Tab. B.5. (B) For each employed kernel scale* $\lambda \in \Delta_\lambda$*, the proposed kernel weight maximum likelihood* $\mathcal{L}^{|\lambda}_{max,t}$ *is presented over the kernel query points* t*. (C) Another kernel weight is based on the Bayes factor* $BF^{|\lambda}_t$ *as a measure for the credibility of a transition. In comparison to* $\mathcal{L}^{|\lambda}_{max}$ *and for a constant scale, the* $BF^{|\lambda}$ *varies significantly across the time series as desired for an effective resolution of consecutive transitions.*

achieved by strong variations of the kernel weights between kernels containing and kernels not containing a transition. At a constant kernel scale λ, the maximum likelihood function does not vary significantly across the times series, whereas the Bayes factors considerably increases at the underlying transitions. Even though the Bayes factor favors in every kernel a transition $\mathcal{M}_{trans}$ over a linear model $\mathcal{M}_{lin}$, its variation across the time series clearly mirrors the underlying events. Therefore, the quantity may still be used as a kernel weight, but not in the exponential form $e^{-BF^{|\lambda}_t}$, since the kernel credibility for all transitions except the most dominant one is eliminated due to the strong difference in magnitude. In order to obtain a moderate kernel weight $bf^{|\lambda}_t$, the Bayes factor may be used as

$$bf^{|\lambda}_t = \begin{cases} -BF^{|\lambda}_t & \text{if } \mathcal{M}_{trans} \text{ supported in } \boldsymbol{y}^{|\lambda}_t, \\ 0 & \text{else.} \end{cases} \tag{4.12}$$

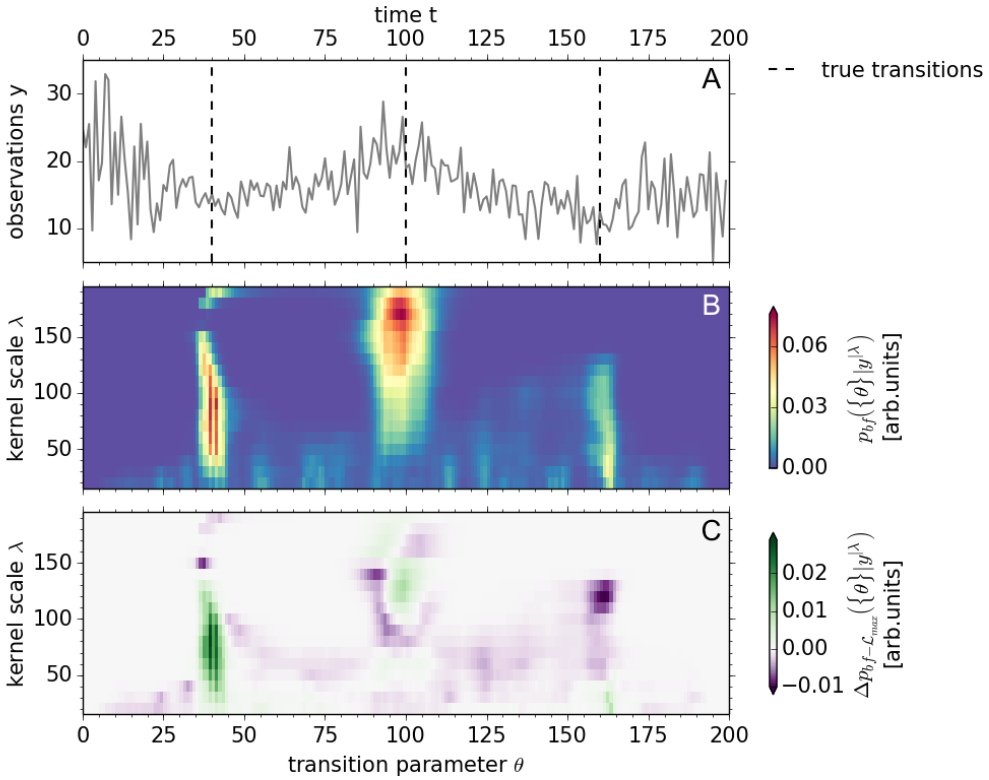


Figure 4.4: *Based on the kernel weights* $f = [bf^{|\lambda}, \mathcal{L}_{max}^{|\lambda}]$ *indicating the existence of a transition, the proxy probability* $p_f(\{\theta\}|\boldsymbol{y}^{|\lambda})$ *of transitions* $\{\theta\}$ *is computed for the synthetic time series* $\boldsymbol{y}$ *(A) at multiple kernel scales* λ*. (B) The modes of the proxy probability derived with the kernel weight* $bf^{|\lambda}$ *indicate the three underlying transitions correctly over most kernel scales. (C) The proxy difference* $\Delta p_{bf-\mathcal{L}_{max}}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ *represents the relative effect of the kernel weights* $bf^{|\lambda}$ *to* $\mathcal{L}_{max}^{|\lambda}$ *on the resulting proxy probability. As desired for the resolution of multiple transitions, the Bayes factor kernel weight favorably assigns more credibility to the modes of the proxy probability and less credibility to the intervals between consecutive transitions.*

By using one of the kernel weights $f^{|\lambda} = [bf^{|\lambda}, \mathcal{L}_{max}^{|\lambda}]$, the proxy probability $p_f(\{\theta\}|\boldsymbol{y}^{|\lambda})$ of transitions $\{\theta\}$ is derived.

In Fig. 4.4 (B), the proxy probability $p_{bf}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ is presented over the kernel scales λ. For most scales the underlying transitions are indicated correctly by the three dominant modes of the proxy probability. At small scales $\lambda < 50$, the approach infers on multiple events causing a diffuse broadening of the proxy probability. In the presented example, the number of kernel observations n_t equates the employed kernel scale λ. Thus, random fluctuations may falsify the inference since the kernel observations $\boldsymbol{y}_t^{|\lambda}$ do not comply with the minimal required number of observations $n_t \geq 50$ (see Sec. 3.3.2). In order to compare the relative influence of the kernel weights $f^{|\lambda} = [bf^{|\lambda}, \mathcal{L}_{max}^{|\lambda}]$ on the proxy probability $p_f(\{\theta\}|\boldsymbol{y}^{|\lambda})$ the difference in terms of

$$\Delta p_{bf-\mathcal{L}_{max}}(\{\theta\}|\boldsymbol{y}^{|\lambda}) = p_{bf}(\{\theta\}|\boldsymbol{y}^{|\lambda}) - p_{\mathcal{L}_{max}}(\{\theta\}|\boldsymbol{y}^{|\lambda}) \tag{4.13}$$

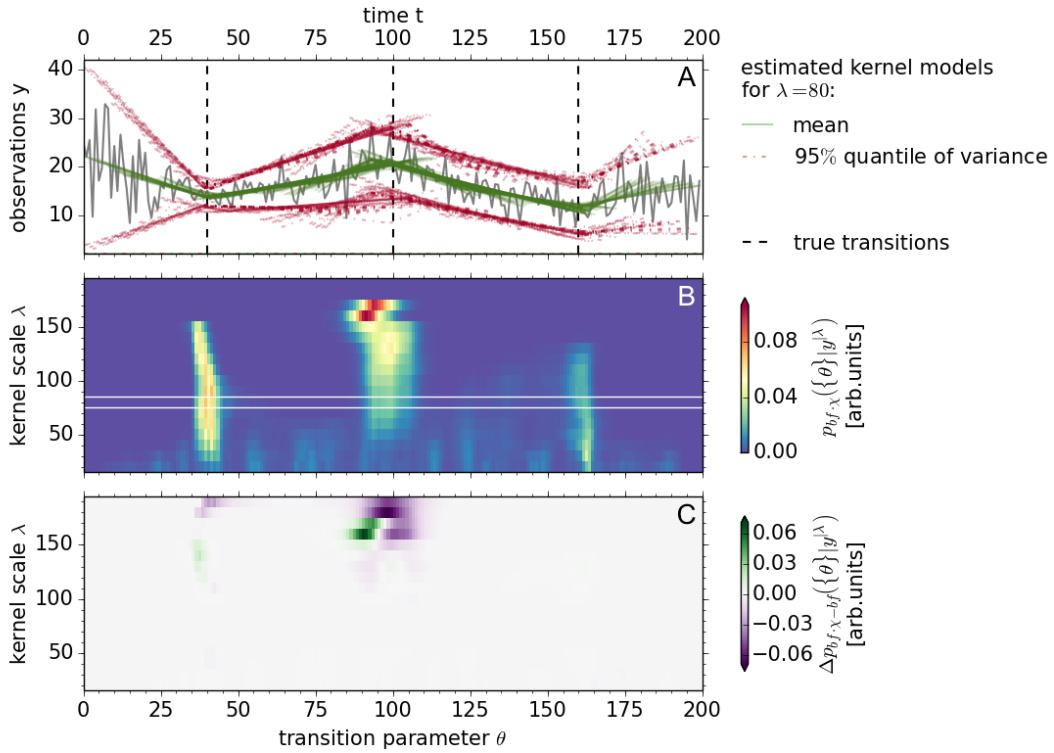


Figure 4.5: *Based on the binary kernel weight $\chi^{|\lambda}$ indicating the normality assumption in a kernel as valid or not, the proxy probability $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ of transitions $\{\theta\}$ is computed for the synthetic time series $\boldsymbol{y}$ (A) over multiple kernel scales λ. (B) For $\lambda < 170$, the derived proxy probability indicates the three underlying transitions correctly. At larger scales the proxy probability does not significantly indicate a transition. (C) The difference between the proxy probability here and with respect the simple bf weighting of Fig. 4.4 (B) illustrate the ignores kernels at larger scales. Moreover, to visualize the inferred transition patterns at a particular scale $\lambda = 80$ (marked white in (B)), the kernel estimates $(\boldsymbol{\vartheta}^*)_t^{|80}$ for kernels with $\chi_t^{|80} = 1$ are shown in (A).*

is shown over the kernel scales λ in Fig. 4.4 (C). The weight of the proxy probability at the transitions $(\theta_1, \theta_2) = (40, 100)$ is bigger for the quantity $bf^{|\lambda}$, whereas the transition $\theta_3 = 160$ gains more weight for the quantity $\mathcal{L}_{max}^{|\lambda}$. The decisive advantage provides the kernel weight $bf^{|\lambda}$ by decreasing the proxy probability at the borders of the modes faster against zero and in between the modes closer to zero, particularly at scales $\lambda < 100$. In consequence, the kernel Bayes factor weight $bf_t^{|\lambda}$ is proposed as the more efficient kernel weight to convincingly resolve multiple transitions.

Kernel weight with respect to generic model assumptions

However, an important issue concerns the visualization of the inferred transition patterns $(\boldsymbol{\vartheta}^*)_t^{|\lambda}$. In principle, the estimated transition models are only presented for kernels, in which the generic transition model $\mathcal{M}_{trans}$ is supported by $BF_t^{|\lambda}$. Here, the approach

supports the transition model in every kernel, respectively the inferred transition patterns are presented for each kernel. Moreover, even at large scales $\lambda > 170$ the proxy probability $p_{bf}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ indicates two out of three transitions correctly, though per definition the generic transition model $\mathcal{M}_{trans}$ can not capture multiple events. These findings indicate, that the suitability of the assumed model as an adequate approximation to the kernel observations $\boldsymbol{y}_t^{|\lambda}$ might not be sufficiently considered in the approach. Based on Sec. 3.4, the kernel observations can be standardized with respect to the estimated kernel transition patterns $(\boldsymbol{\vartheta}^*)_t^{|\lambda}$ as defined in Eq. 3.53. Any deviation of the observations to the generic model is therefore mapped on the deviation of the transformed kernel observations $(\boldsymbol{y}')_t^{|\lambda}$ from normality. This strategy may be used to define a further kernel measure $(f')_t^{|\lambda}$.

In the presented example, the validation of normality is approached via the Shapiro-Wilk test (SWT) [110]. The assumption of normality is considered to be true for p-values greater than an α-level of 0.05. Instead of employing the p-value itself as a kernel weight the consideration of normality is expressed in terms of a binary indicator function $\chi_t^{|\lambda}$ defined as

$$\chi_t^{|\lambda} = \begin{cases} 1 & \text{if } (\boldsymbol{y}')_t^{|\lambda} \sim \mathcal{N}(0, \mathbf{1}) \text{ considered as valid,} \\ 0 & \text{else.} \end{cases} \tag{4.14}$$

Hence, the normality indicator function either accepts the kernel posterior distribution $p(\boldsymbol{s}, \theta|\boldsymbol{y}_t^{|\lambda})$ and all corresponding estimates $(\boldsymbol{\vartheta}^*)_t^{|\lambda}$ or rejects the complete kernel outcome. In this way, the kernel weight $\chi_t^{|\lambda}$ essentially realizes a removal strategy of kernels enclosing outliers. The cumulative acceptance $X^{|\lambda}$ per kernel scale λ defined as

$$X^{|\lambda} = \sum_{i=t_1+\frac{\lambda}{2}}^{t_n-\frac{\lambda}{2}} \chi_i^{|\lambda}\,, \tag{4.15}$$

that is the amount of kernels for which the normality assumption is considered to be valid, is summarized in Tab. 4.1. The maximal cumulative acceptance occurs for $60 \leq \lambda \leq 90$ and indicates the scales at which the underlying transition patterns may be best approximated by the generic transition model $\mathcal{M}_{trans}$.[2] Larger kernel scales $\lambda > 90$ successively contain more than one transition and the generic transition model becomes less suitable as mirrored by the decreasing cumulative acceptance $X^{|\lambda}$. At smaller kernel scales $\lambda < 60$ the cumulative acceptance decreases only slightly and does not achieve a more accurate resolution of consecutive transitions. As already stated before, for $n_t < 50$ the robustness of the inference approach can not be guaranteed anymore.

Additionally, the amount of observations n_t generally affects the performance of the normality validation. In particular for the SWT, the performance is considered to be

[2]Similar to real observations there actually is no *true* scale to the synthetic time series, since the observations are generated by a different model approach (Tab. B.5). Nevertheless, the kernel weight $\chi_t^{|\lambda}$ correctly indicates the kernel scale λ, at which only one transition is in the kernel support $p_t(\theta)$.

weak for $n_t < 30$, since not enough observations are available for the test statistics [111]. Up to $n_t < 5000$, the SWT infers relatively robust but the test is extremely sensitive to outliers, such that a single outlier may lead to a false decision. To better understand the real observations, the kernel-based approach may eventually use another strategy to validate the normality assumption in a more transparent way, e.g., by higher statistical moments as suggested in Sec. 2.1.1.

In conclusion, the complete proxy probability of transitions $\{\theta\}$ analyzed at the kernel scale λ may be formulated as

$$p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda}) := \sum_{i=t_1+\frac{\lambda}{2}}^{t_n-\frac{\lambda}{2}} bf_i^{|\lambda} \cdot \chi_i^{|\lambda} \cdot p(\theta|\boldsymbol{y}_i^{|\lambda}) \tag{4.16}$$

and is presented for the analyzed synthetic time series in Fig. 4.5 (B). To illustrate the impact of the additional kernel weight $\chi_t^{|\lambda}$ on the proxy probability the difference

$$\Delta p_{bf\cdot\chi-bf}(\{\theta\}|\boldsymbol{y}^{|\lambda}) = p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda}) - p_{bf}(\{\theta\}|\boldsymbol{y}^{|\lambda}) \tag{4.17}$$

is shown in Fig. 4.5 (C). The most significant effect occurs for kernel scales $\lambda \geq 150$ enclosing more than one transition per kernel. At each scale λ, the inferred transition patterns may be visualized by the derived kernel estimates $(\boldsymbol{\vartheta}_t^{|\lambda})^*$ for kernels $\boldsymbol{y}_t^{|\lambda}$ with $\chi_t^{|\lambda} = 1$ as exemplified for $\lambda = 80$ in Fig. 4.5 (A). In this way, the proposed proxy probability $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ offers a powerful diagnostic tool to investigate observations for multiple generic transition events $\{\theta\}$ at different temporal scales λ. Although none of the estimated kernel models is ignored ($X^{|80} = 100\%$), the underlying transition patterns, respectively the observations, are convincingly approached. This is possible since in the presented synthetic example the transition events $(\theta_1, \theta_2, \theta_3) = (40, 100, 160)$ are equidistant. For real observations, the transitions are most likely unevenly distributed over the time series and consequently the cumulative acceptance might never reach 100% for a specific scale λ. It is more likely, that the transitions occur at a different scales and hence, the locations of $\chi_t^{|\lambda} = 1$ might be more useful to address. To demonstrate the use of the extended kernel-based approach, it is applied on real time series in the following.

Table 4.1: *The binary kernel weight $\chi_t^{|\lambda}$ indicates whether the normality assumption of the transformed kernel observations $(\boldsymbol{y}_t^{|\lambda})'$ is considered as valid or not. Based in the SWT, the cumulative acceptance per kernel scale λ are listed for the synthetic time series, the Nile and the Tuscaloosa data. Note, that the kernel scale λ equates the number of kernel observations n_t. The highlighted values mark the scale, at which the observations may be best approximated by the generic transition model $\mathcal{M}_{trans}$.*

kernel scale λ	cumulative acceptance $X^{\|\lambda}$ [%]		
	synthetic data Fig. 4.4	Nile data Fig. 4.6	Tuscaloosa data Fig. 4.7
20	98.9	92.5	83.8
30	98.8	87.1	77.1
40	99.4	96.7	**88.3**
50	99.3	96.0	80.0
60	**100**	**100**	67.5
70	**100**	**100**	43.3
80	**100**	**100**	35.0
90	**100**	**100**	10.0
100	98.0	-	-
110	93.4	-	-
120	83.5	-	-
130	70.4	-	-
140	68.9	-	-
150	51.0	-	-
160	14.6	-	-
170	6.5	-	-
180	0.0	-	-
190	0.0	-	-
200	0.0	-	-

4.3 Application on direct climate observations

The introduced kernel-based inference approach is applied to direct environmental observations comprising documented transition events. In Sec. 3.4 the observations have already been investigated by using the single transition approach. The kernel-based approach realizes a consistent extension thereof, designed to detect multiple changes within a given time series. As previously assumed, the shift model $\mathcal{M}_{shift}$ is used as the generic transition model. The time series are sampled equidistant in time, thus directly providing a reasonable global time grid Θ_t, respectively the kernel query points t. The numerical setting is similarly taken for both time series and defines the kernel scale λ in years $[a]$, and the sampling range of the transition parameter θ in years $[a]$ and of the deviation parameters $\boldsymbol{s}$ as

$$\begin{array}{lcllcl} \Delta_\lambda & = & [20,\,80]\,a\,, & \delta_\lambda & = & 10\,a, \\ \Delta_{\lambda'} = \Delta_\theta & = & \frac{3}{5}\lambda\,, & \delta_t = \delta_\theta & = & 1\,a, \\ \Delta_{\boldsymbol{s}} & = & [-0.25,\,0,25]\,, & \delta_{\boldsymbol{s}} & = & 0.005\,. \end{array} \tag{4.18}$$

Due to the fact, that both time series consist of temporally equidistant observations, the investigated kernel scales $\lambda \in \Delta_\lambda$ equate the number n_t of the corresponding kernel observations $\boldsymbol{y}_t^{|\lambda}$. By analyzing real time series, the inference performance of the kernel-based approach may be deduced from the localization of individual and the resolution of multiple underlying transitions documented in the meta data.

4.3.1 Annual Nile river flow (1871-1970) from Aswan, Egypt

Given the numerical setting, the kernel-based inference approach is applied on the annual Nile river flow measured at a dam in Aswan. The employed kernel observations $\boldsymbol{y}_t^{|\lambda}$ are centered at the query time t and are of kernel width $\lambda \in \Delta_\lambda$. For each kernel at each scale, the kernel posterior distribution $p(\theta|\boldsymbol{y}_t^{|\lambda})$ (Eq. 3.30) of the transition θ, the kernel estimates $(\boldsymbol{\vartheta}^*)_t^{|\lambda} = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})_t^{|\lambda}$ (Eq. 3.38) of the transition patterns and the kernel weight $bf_t^{|\lambda}$ (Eq. 4.12) based on the Bayes factor $BF_t^{|\lambda}$ are computed.

The derived Bayes factors only support the transition model $\mathcal{M}_{shift}$ over the non-transition model $\mathcal{M}_{lin}$. Therefore, the kernel estimators $(\boldsymbol{\vartheta}^*)_t^{|\lambda}$ of the transition model are used to standardize the kernel observations into $(\boldsymbol{y}')_t^{|\lambda}$ as defined in Eq. 3.53. By applying the SWT, the binary kernel weight $\chi_t^{|\lambda}$ (Eq. 4.14) indicating the validity of the generic transition model $\mathcal{M}_{shift}$ within each kernel is obtained. The cumulative acceptance $X^{|\lambda}$ (Eq.4.15) is summarized in Tab. 4.1 and is maximal, i.e. $X^{|\lambda} = 100\%$, for $\lambda \geq 60\,a$. Hence, the generic transition model convincingly captures the time series at scales enclosing at least 60% of the observations. Thus, the findings point towards a single underlying transition event within the Nile river observations.

The proxy probability $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ of transitions $\{\theta\}$ defined in Eq. 4.16 is presented over the kernel scales λ in Fig. 4.6 (A). The major mode of the proxy probability remains

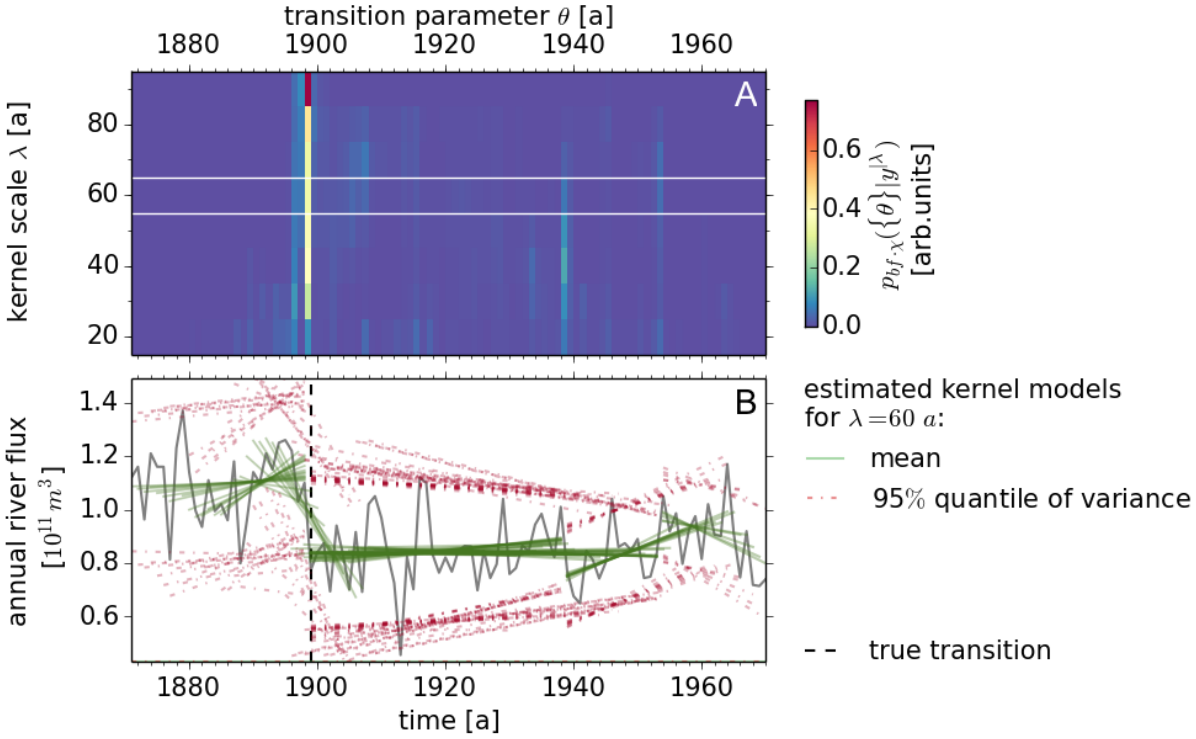


Figure 4.6: *The annual Nile river flow (B) comprises a documented flow shift in the year* 1899. *(A) By applying the kernel-based inference approach, the proxy probability* $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ *of transitions* $\{\theta\}$ *is derived over the kernel scale* λ. *The major mode indicates a transition at* $\hat{\theta}_1 = 1898$, *in good agreement with the documented event. (B) To visualize the inferred transition patterns, the estimated kernel models are presented for* $\lambda = 60\,a$ *(marked white in (A)), the smallest kernel scale, at which the cumulative acceptance is maximal, i.e.* $X^{|\lambda} = 100\%$ *(see Tab 4.1). The major transition* $\hat{\theta}_1 = 1898$ *is characterized by a sharp shift in mean and a slightly declining variability after the change.*

dominant and within a small time interval around the year $\hat{\theta}_1 = 1898$ for all kernel scales. A minor mode appears on scales $\lambda \leq 60\,a$ around the year $\hat{\theta}_2 = 1939$. At scales $\kappa < 60\,a$ the proxy probability broadens mainly around $\hat{\theta}_1$ and $\hat{\theta}_2$. The broadening for $\lambda < 50\,a$ is essentially attributed to the fact, that the kernel observations $\boldsymbol{y}_t^{|\lambda}$ fall below the minimal recommended number of observations $n_t = 50$ required for a robust inference (see Sec. 3.3.2).

To comprehensively visualize the inferred transition patterns, the estimated kernel transition models are presented at the kernel scale $\lambda = 60\,a$ in Fig. 4.6 (B). Obviously, the major event $\hat{\theta}_1$ is characterized by a sharp shift in the mean, similar to the estimated characteristics of the previous analysis in Sec. 3.4.1. Different to the previously estimated minor change of the observational variability, the kernel-based approach infers on a more significant decline of the variability. Moreover, a second minor event $\hat{\theta}_2$ is induced by an apparent shift in mean and an approximately constant variability after the event. Therefore, the kernel-based approach correctly detects the documented un-

derlying change taking place in the year 1899 and the major characteristics induced by the net effect of the associated changes, i.e. an abrupt water level shift [55, 112–114].

4.3.2 Annual average temperature (1901-2000) from Tuscaloosa, Alabama

Based on the numerical setting of Eq. 4.18, the kernel-based inference approach is applied on the annual temperature measured at the weather station in Tuscaloosa. The computational steps are performed in accordance to the investigation of the Nile river observations in the previous section. The obtained kernel Bayes factors $BF_t^{|\lambda}$ favor a non-transition model $\mathcal{M}_{lin}$ over a transition model $\mathcal{M}_{shift}$ for only few kernels. For the remaining kernels with $bf_t^{|\lambda} > 0$, the kernel estimators $(\boldsymbol{\vartheta}^*)_t^{|\lambda} = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})_t^{|\lambda}$ are used to standardize the corresponding kernel observations in order to apply the SWT. In comparison to the Nile observations, the derived cumulative acceptance $X^{|\lambda}$ is listed in Tab. 4.1 and does for none of the employed kernel scales $\lambda \in \Delta_\lambda$ reach 100%. The highest cumulative acceptances occur for scales $\lambda \leq 50\,a$, with the maximal value of $X^{|\lambda} = 88.3\%$ for $\lambda = 40\,a$. The proxy probability $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ of the transitions $\{\theta\}$ is presented over the kernel scale λ in Fig. 4.7 (A) and confirms a dominant change at $\hat{\theta}_2 = 1957$ for all scales of investigation. Moreover, the kernel-based approach reveals two minor modes at scales $\lambda \leq 50\,a$ at $\hat{\theta}_1 = 1939$ and at $\hat{\theta}_3 = 1975$. However, these kernel scales contain less than the minimal recommended number of observations $n_t = 50$ required for a robust inference.

Besides the documented change in the year 1956 also detected with the previous single transition detection (see Sec. 3.4.2), the kernel-based approach accomplishes to additionally infer on the documented change in the year 1939. Given the meta data of the measurements, the modes of the proxy probability at the smallest scale of investigation $\lambda = 20\,a$ may each be associated to one of the documented changes. But clearly, the average distance between consecutive changes is about $22\,a$ and the inference performance of the approach is generally not considered to be reliable for such sparse observations. Noteworthy is the investigation of the Tuscaloosa temperature data at a monthly scale [37], hence providing sufficient observations to the introduced kernel-based approach. Even in this data rich situation, the applied minimum description length approach does not achieve to detect more changes than the presented Bayesian approach. On the contrary, the monthly data raises even more difficulties to the analysis strategy. The analysis approach has to explicitly incorporate autocorrelation in order to appropriately describe the underlying process. Furthermore, a reference series based on the measurements from neighboring weather stations is required to adjust the monthly observations to the seasonal impact [37].

To comprehensively visualize the inferred transition patterns, the estimated kernel transition models are presented at the kernel scale $\lambda = 40\,a$ in Fig. 4.7 (B). The dominant event $\hat{\theta}_2 = 1957$ is characterized by a sharp shift in the mean and an increase in

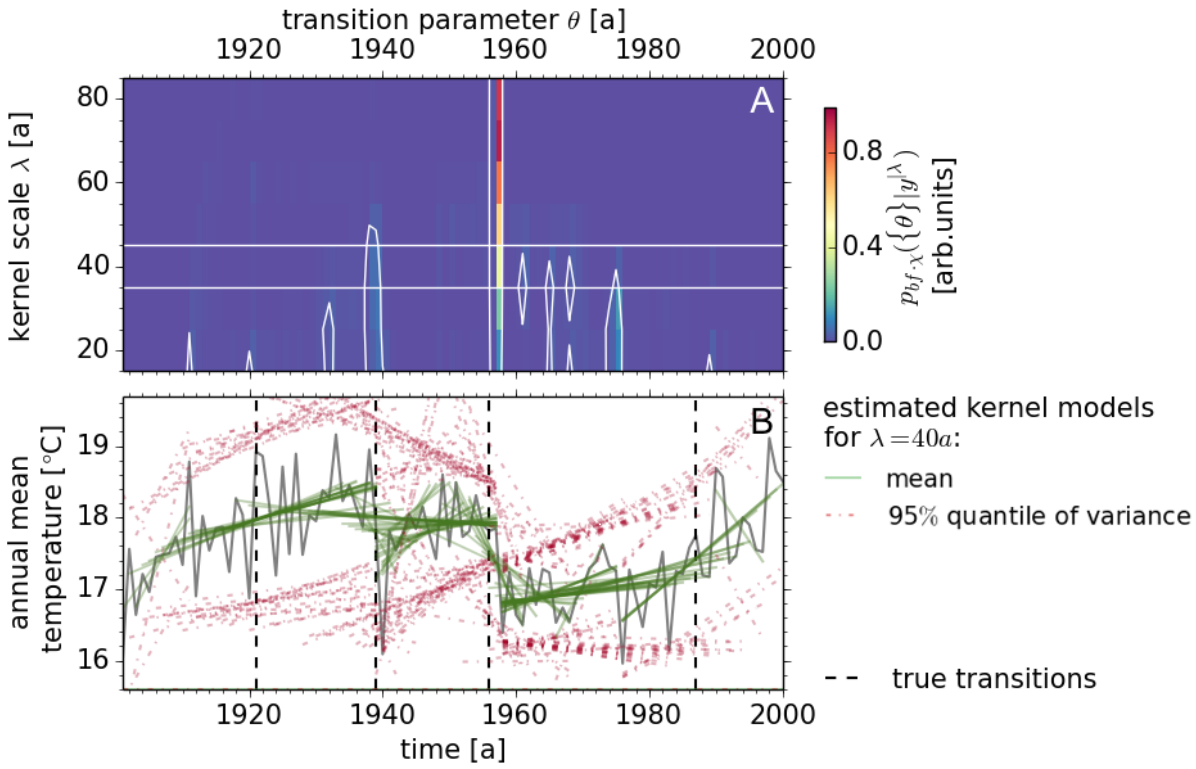


Figure 4.7: *The annual temperature measured in Tuscaloosa (B) comprises multiple documented changes in the years* 1921, 1939, 1956 *and* 1987. *(A) By applying the kernel-based inference approach, the proxy probability* $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ *of transitions* $\{\theta\}$ *is derived over the kernel scale* λ. *The dominating transition in 1957 confirms the global event already inferred in the global detection approach. At smaller scales further events are indicated (marked by the isoline* $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda}) = 0.04$*) inter alia at the known transition in 1939 and one additional transition in 1975. (B) To visualize the characteristics of the underlying change, the estimated kernel models are displayed for the scale* $\lambda = 40\,a$*, the scale with the maximum amount of kernels in which normality is not rejected.*

the variability prior and after the change. Moreover, the event $\hat{\theta}_1 = 1939$ describes as a sharp shift in the mean, whereas the variability is difficult to interpret, since one extreme observations apparently affects the inference due to small number of data points. The event $\hat{\theta}_3 = 1975$ describes a rather continuous change in the mean but no considerable change in the increasing variability. In conclusion, the kernel-based approach accomplishes to infer on two transitions in the years out of four transitions convincingly, albeit at kernel scales containing less observations than recommended for a robust inference. Although the investigated observations do not comply with all requirements, the kernel-based approach serves as a useful diagnostic tool and unravels valuable information from the time series.

4.4 Discussion and Summary

In this chapter, the Bayesian transition detection introduced in Chap. 3 is extended to a kernel-based inference approach in order to investigate time series for multiple changes. By defining a time dependent prior distribution for the transition parameter, the inference approach can be systematically applied on kernel observations sampled across the complete time series. Thus, the kernel-based approach achieves a considerable computational speed-up, since the inference task is sub-divided into a set of smaller kernel computations. As demonstrated for a synthetic time series comprising a single change, the derived kernel posterior distributions of the transition parameter can be combined into a comprehensive credibility expression, interpreted as a proxy probability of transitions underlying the time series at a particular scale. Thereby, the kernel Bayes factor is used to construct a kernel weight taking into account the existence of a transition in order to precisely localize individual and resolve multiple transitions. Moreover, to take into account the suitability of the generic transition model depending on the kernel scale of investigation an additional binary kernel weight is proposed. The inference performance of the proxy probability over the kernel scales confirms the desired influence of each kernel weight for a synthetic time series comprising multiple changes. The inferred transition patterns are visualized at a particular scale by presenting all local estimates for kernels in which the product of both kernel weights is non-zero.

The kernel-based approach is finally applied to the real time series comprising documented transitions and investigated in Sec. 3.4. For the annual Nile river flow, the approach correctly infers on one underlying change dominating over all kernel scales and thus confirming the result of the previous analysis. For the annual temperature measured at Tuscaloosa, the approach infers on two out of four underlying changes correctly for kernel scales enclosing enough observations for a robust inference. Despite the sparse data situation at small kernel scales, the inferred transition patterns reveal multiple of the underlying changes.

In conclusion, the introduced kernel-based Bayesian inference is proposed as a powerful diagnostic tool to investigate real time series for multiple transitions. The assumed generic transition model is thereby designed to approximate signals that are locally differentiable up to the first order prior and after each transition. Hence, the model captures a variety of natural processes. In case complex kernel observations do not comply with the generic assumptions the corresponding kernel results are ignored for the composition of the proxy probability and the visualization of the inferred transition patterns. Therefore, the kernel-based inference approach is considered as an appropriate diagnostic tool to investigate complex indirect climate observations for multiple transitions as intended in the following chapter.

5 Bayesian inference about Plio-Pleistocene climate transitions in Africa

During the last 5 Ma the Earth's ocean-atmosphere system passed through several major transitions, many of which were tectonically induced, such as the intensification of the Northern Hemisphere Glaciations through the closure of the Panama seaway and the onset of the monsoon system through the uplift of the Tibetian plateau. Many of these transitions may have had an impact on the evolution of humans and animals. A popular explanation for the first appearance of the genus *Homo* is (or was) the aridification of Africa in the course of the Northern Hemisphere glaciation. Apart from the fact that the correlation does not necessarily imply causality, many attempts to establish a relationship between climate and evolution fail because of the complexity of the climate records. The kernel-based Bayesian inference approach used here helps to detect multiple transitions in established records of Plio-Pleistocene African climate. The generic changes in central tendency and dispersion identified in marine records of terrestrial dust coincide with established global climate events: (i) the closure of the Central American Seaway around (3.95-3.70) Ma, (ii) the change of the Indonesian throughflow at (3.25-3.15) Ma, (iii) the intensification of the Northern Hemisphere Glaciation between (2.90-2.80) Ma, (iv) the two-step establishment of the zonal sea-surface temperature gradient in the tropics at (2.35-2.10) Ma and (1.70-1.50) Ma, associated with the reorganization of the Hadley-Walker-Circulation, and (v) the Middle Pleistocene transition at (0.95-0.85) Ma. These transitions towards a more arid and more variable climate in Africa coincide with important steps in the evolution of humans and other mammals, and therefore provide an important basis for the current discussion on climate-evolution linkages.

5.1 Introduction

Abrupt climate transitions provide an important insight into the dynamics of the climate system of the Earth. Since such transitions possibly have dramatic consequences for the biosphere, their investigation is of particular importance. On time scales of thousands or millions of years, detecting and deciphering the causes of such climate transitions is essentially difficult. The available observations are generally interpreted as the result of multiple stochastic processes each describing a specific aspect of the Earth's climate

system or the data acquirement. Hence, their investigation requires the use of powerful statistical methods in order to detect and understand underlying climate transitions.

A prominent example of the importance of climate transitions on the biosphere is the study of the relationship between environmental change and human evolution in the Plio-Pleistocene Africa. A popular hypothesis of such a relationship that has been proposed in the past was about the intensification of the Northern Hemisphere Glaciations near 2.5 Ma causing a rapid burst of speciation in Africa (e.g., [117, 118]). Elisabeth Vrba used the temporal coincidence of the intensification of the Northern Hemisphere Glaciations, a shift towards a drier climate in Africa, and the replacement of forest antilopes by species that graze only in dry, open savannas. Others have made this climate transition responsible for the first appearance of the genus *Homo* [4, 39], an interpretation that has recently been superseded by the discovery of the correlation of major speciation events and the establishment of the Hadley-Walker-Circulation at around 2 Ma ago [5, 119].

Apart from the fact that the correlation does not necessarily imply causality, many attempts to establish a relationship between climate and evolution fail due to the complexity of the climate records. Therefore, an objective criteria is needed to study conjoint events potentially imprinted in a set of paleoclimate records. We propose here the use of a kernel-based Bayesian inference approach to accomplish a precise detection, generic characterization and mutual comparison of multiple changes in Plio-Pleistocene African climate [116]. The method is applied to records of terrigenous dust fluxes to the oceans around the African continent from the Ocean Drilling Programme (ODP) sites 659 [120], 721/722 [4, 39, 121] and 967 [122] presented in Fig. 5.1. The amount of dust eroded and transported from the continent to the oceans is an established measure for wind strength and aridity of the tropical and subtropical African region [4, 5].

The method used to analyze these proxy records is based on the stochastic modeling of a generic transition, such as a break or a shift of the statistical moments of the data. The assumption of a generic transition model allows to at least locally approximate complex changes for a variety of natural processes. In our case, the generic model is parametrized by the mean value (as a measure of central tendency) and the variance (as a measure of variability). To investigate complex time series for an unknown number of transitions, we define a window (also called the *kernel*) of a certain size (the *scale* of the analysis), which we shift along the complete time series with a certain delay. For each of these kernels, we perform a Bayesian inference to compute the probability of a transition for every time point within the kernel. The mode of the resulting probability indicates the transition location within the individual kernel and its width enables us to derive an associated confidence interval at a given significance level (e.g. 95%).

The individual kernel probabilities are integrated over the complete time series in order to derive a proxy probability of multiple transitions[1]. Thereby, the kernel probabilities

[1]For the sake of clarity, the term *proxy probability* of transitions denotes a practical substitute to a multi-transition posterior distribution (cf. Sec. 4.1). The term *proxy record* denotes a practical substitute to a direct climate observation (cf. Sec 2.1.2).

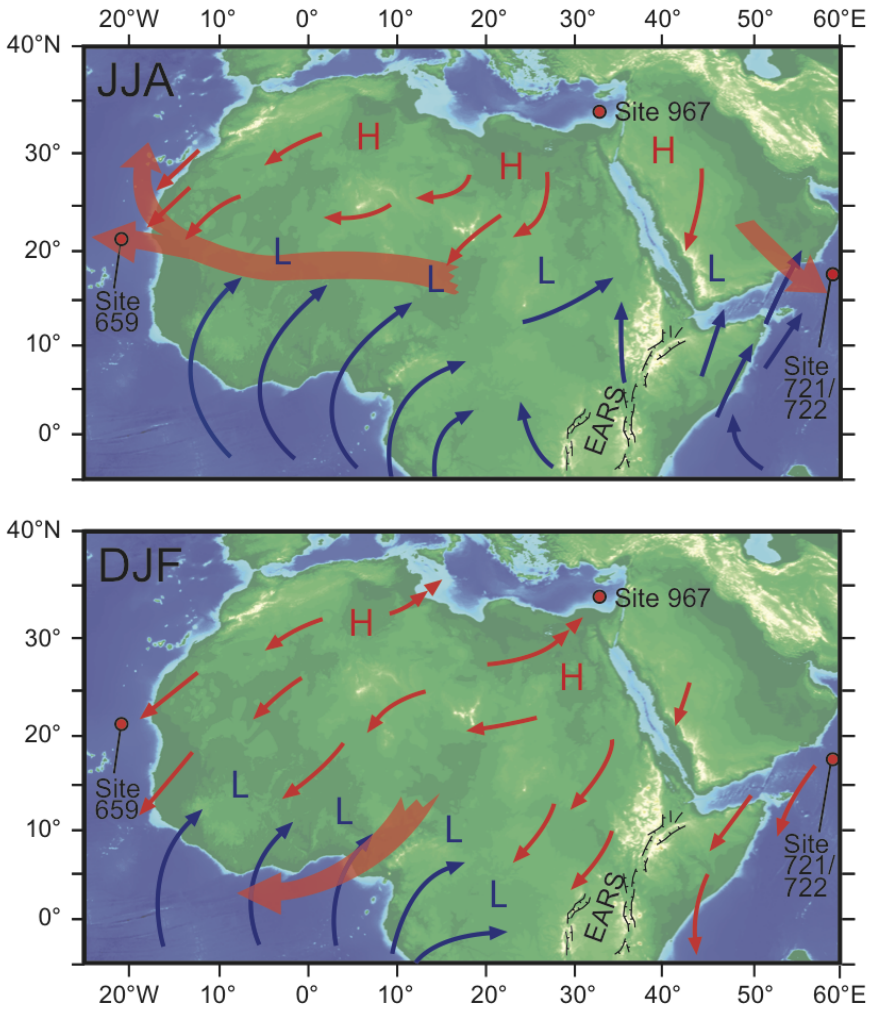


Figure 5.1: *The map shows the locations of the studied ODP sites 659 in the Atlantic Ocean [120], 721/722 in the Arabian Sea [4, 39] and 967 in the Mediterranean Sea [122] used as archives of Plio-Pleistocene African climate and the East African Rift System (EARS). The wind patterns between high (H) and low (L) pressure regions, presented for summer (JJA) and winter (DJF) conditions, are considered as the major transport mechansims of the investigated terrigenous dust from the continent to the deposit sites.*

are weighted by kernel measures in order to precisely localize individual and distinctly resolve multiple transitions. By comparing the proxy probability of transitions derived for each proxy record, we search for conjoint changes characterized by the synchronicity of the estimated transition locations and the congruence of the estimated transition patterns. Furthermore, we distinguish between short-term and long-range changes by computing the proxy probability of transitions for various kernel scales to study the dependency of the detected transition events on the size of the window.

5.2 Terrigenous dust as climate indicator

We investigate a representative set of marine records of terrigenous dust flux from the African continent spanning the Plio-Pleistocene period. Dust flux in marine sediments is an established proxy for aridity on the adjacent continent but also for the strength of the wind transporting dust particles from the source on land to the sink in the ocean [5]. The records are derived from the ODP sites shown in Fig. 5.1 off subtropical West Africa (ODP site 659 [120]), the eastern Mediterranean Sea (ODP site 956 [122]) and the Arabian Sea (ODP site 721/722 [4, 39]), and enable us to investigate multiple time series enclosing an integrated climate signal of Northern Africa over the last 5 Ma. Detecting and quantifying dominant changes in the trends and the variability of these proxy records yields potential correlations with global tectonic and climate shifts and supports the identification of important linkages to the African climate.

Even though the measurement methods may not be similar between different ODP records nor within one ODP record ([4, 39, 120–122], discussion in [5]) and the age calibration may not be optimal due to orbital tuning or relatively sparse age control points [123], we consider the time scale as given. The measurement errors of the terrigenous dust flux are indirectly incorporated in our Bayesian approach by the assumption of additive Gaussian noise as the sum of intrinsic and external random fluctuations.

5.3 Bayesian transition detection

Paleoclimatic observations may be considered as a convolution of the actual climate signal (e.g. aridity on the African continent), the recording process (e.g. weathering, mobilization, transportation and deposition of terrestrial dust in the ocean) and the data reconstruction (e.g. whole-core measurements of magnetic susceptibility of sediments). A schematic illustration of the basic processes underlying the generation of indirect climate observations is provided in Fig. 2.3. The deconvolution of the climate signal requires a solid knowledge of these processes including their particular noise structure induced by different factors such as uncertain age-depth relationship of paleoclimate archives (e.g. [41, 124, 125]), misinterpretation or miscalibration of climate proxies (e.g. ignoring wind strength as an important factor controlling the flux of terrestrial dust to the ocean in addition to aridity [122]), sediment mixing or bioturbation (e.g. [126–129]) and analytical errors [28, 130, 131], that all together complicate the detection of important climate transitions. In our study, we consider the climate signal as the primary component of the time series. Since the noise may also contain important information on climate variability, we will include it in our investigation of climate transitions. As a general approach we assume a Gaussian model for the noise with a linearly varying mean and standard deviation (i.e. heteroscedastic behavior) through time. Consequently, the transition events are generically defined as a mutual break in the linear evolution of both mean and variance.

Table 5.1: *To derive the kernel weight $\chi_t^{|\lambda}$, thresholds for the statistical kernel moments $m_{1-4,t}^{|\lambda}$ are assumed. Thus, the normality assumption in a kernel is considered to be valid in case the empirical moments do not exceed these limits.*

statistical moment	theoretical value	threshold	descriptive characteristics
m_1	0	± 0.01	mean
m_2	1	± 0.06	deviation
m_3	0	± 1	skewness
m_4	3	± 2	kurtosis

Established methods of transition detection in climate time series address adapted regression techniques [13, 35], wavelet based concepts [10, 48], recurrence network analysis [6] and Bayesian approaches [51, 90]. Different to these techniques our method, introduced in [116] and further developed in this thesis, provides a probabilistic quantification and a diagnostic visualization of multiple underlying generic transitions. Based on a locally linear, heteroscedastic transition model, we compute a proxy probability of generic transitions given the proxy record, by employing a kernel-based (i.e. basically a sliding window approach) Bayesian inference approach. In this way we are able to infer on multiple transitions within complex time series while avoiding high computational costs. Moreover, we are able to provide associated credibility measures about the location and the existence of such events, as objective criteria to evaluate the detected transitions. Additionally, we obtain in a natural way local model estimates to supportingly visualize the evolution of mean and variability over the record and support our understanding of the inferred transition characteristics.

5.3.1 Formulation of the proxy probability of transitions

The terrigenous dust records are sparsely sampled relative to the long time range they span. For instance, from all records the ODP 659 signal exhibits the minimal average time step of about 4.1 ka over a time scale of 5 Ma. Therefore, we assume that the underlying events may be adequately approximated by the continuous generic transition model $\mathcal{M}_{break}$ as defined in Eqs. 3.7, 3.8 and illustrated in Fig. 3.2 (A). As elaborated in Chap. 4, the kernel-based approach is constructed on a global equidistant time grid $\Theta_t = [t_1, ..., t_n]$, thus providing a common grid of investigation to the three diversely sampled ODP records. Given the kernel observations $\boldsymbol{y}_t^{|\lambda}$ of length λ centered at time point t of the global time grid, the kernel posterior distribution $p(\theta|\boldsymbol{y}_t^{|\lambda})$ of the transition parameter θ (Eq. 3.30) and the kernel estimates $(\boldsymbol{\vartheta}^*)_t^{|\lambda} = (\boldsymbol{\beta}^*, \sigma^*, \hat{\boldsymbol{s}}, \hat{\theta})_t^{|\lambda}$ of the transition

patterns (Eq. 3.38) are computed. Moreover, the kernel Bayes factor $BF_t^{|\lambda}$ as defined in Eq. 4.7 is computed to construct the kernel weight $bf_t^{|\lambda}$ formulated in Eq. 4.12. The binary kernel weight $\chi_t^{|\lambda}$ is given by Eq. 4.14, whereas the normality assumption used therein is verified by the higher statistical moments $m_{1-4,t}^{|\lambda}$ as discussed in Sec. 3.4. The normality assumption is considered to be valid in case all kernel moments are within the corresponding thresholds given in Tab. 5.1. Based on these kernel quantities, the proxy probability of transitions $\{\theta\}$ can be calculated at different kernel scales λ as

$$p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda}) := \sum_{i=t_1+\frac{\lambda}{2}}^{t_n-\frac{\lambda}{2}} bf_i^{|\lambda} \cdot \chi_i^{|\lambda} \cdot p(\theta|\boldsymbol{y}_i^{|\lambda}) \,. \tag{5.1}$$

Hence, for every terrigenous dust record a degree of belief about a generic transition is assigned to each investigated time point t over the conjoint time grid Θ_t. A schematic illustration of the computational steps of the kernel-based inference approach is presented in Fig. 5.2.

5.3.2 Specification of the kernel-based approach

In order to systematically investigate the terrigenous dust flux from the ODP sites 659, 721/722 and 967 for multiple transitions, we perform the inference approach at kernel scales λ between 0.25 and 1.25 Ma with a sampling step δ_λ of 0.25 Ma. The average number of kernel observations $\langle n_t^{|\lambda} \rangle$ vary with respect to the kernel scale λ as well as to the ODP records as summarized in Tab. 5.2. The lowest considered kernel scale is determined by the minimal number of kernel observations $n_t = 50$ required to ensure a robust inference performance (cf. Sec. 3.3.2). The upper kernel scale is particularly chosen to resolve at least two consecutive transitions in the shortest, 3 Ma-long record of ODP site 967. For this reason, the scale of investigation should not exceed the half of the smallest considered record. In order to reveal potential scale dependencies of the changes, we consider the ratio between the kernel scale λ and the sampling range Δ_θ of the transition parameter θ as constant. The numerical resolution of the transition

Figure 5.2 *(facing page)*: *The schematic flow diagram describes the computational steps of the kernel-based inference approach. (A) Here, the terrigenous dust flux of ODP site 721/722 is investigated at the kernel scale λ =1.0 Ma. (B) Based on a model without and with a transition, we infer via Bayesian inversion on the local mean (green), variance (red, represented as 95% quantile) and transition θ (purple) of the kernel centered at t =1.50 Ma. (C) The kernel probability of transition $p(\theta|\boldsymbol{y}_t^{|\lambda})$ (blue) yield the MAP estimator $\hat{\theta}_t$ (purple) within its 90% confidence interval. To generate the kernel weighting functions, the Bayes factor $BF_t^{|\lambda}$ and the statistical moments $m_{1-4,t}^{|\lambda}$ are computed. (D) By repeating the computations for kernels $t \in [1.50, 2.15]$ Ma, the sum of the weighted kernel probabilities of transition (blue shaded) provide the proxy probability of transitions $p_{bf\cdot\chi}(\{\theta\}|\boldsymbol{y}^{|\lambda})$ (blue line). Thus, two transitions (purple shaded) at around $\hat{\theta}_1$ =(1.50-1.65) Ma and $\hat{\theta}_2$ =(2.10-2.25) Ma are indicated by the intervals enclosing at least 90% of the proxy credibility weight associated to each inferred event.*

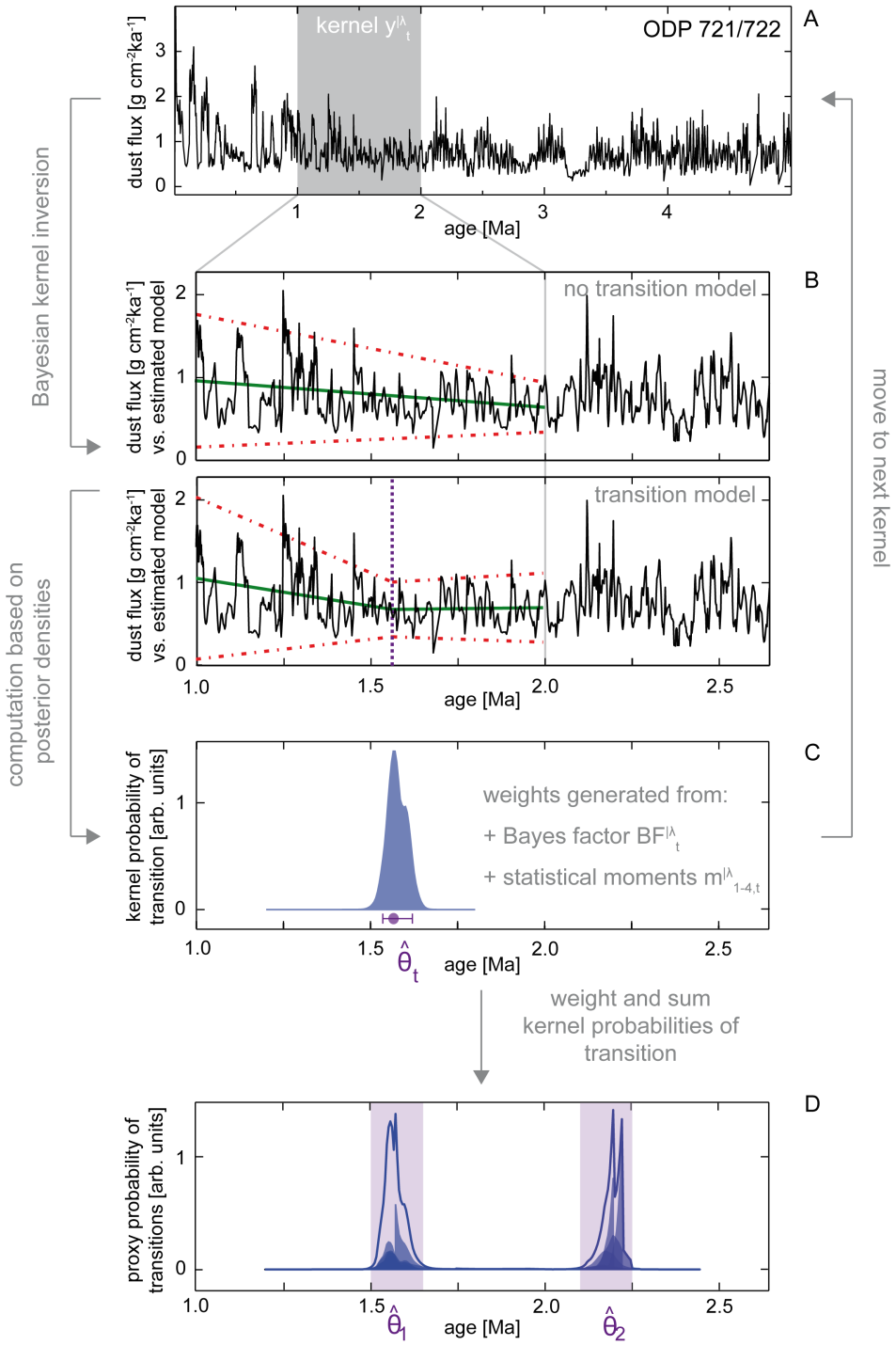

A
ODP 721/722
kernel y^λ_t
dust flux [g cm⁻²ka⁻¹]
age [Ma]
B
no transition model
dust flux [g cm⁻²ka⁻¹] vs. estimated model
transition model
C
weights generated from:
+ Bayes factor BF^λ_t
+ statistical moments m^λ_1-4,t
kernel probability of transition [arb. units]
θ̂_t
weight and sum kernel probabilities of transition
D
proxy probability of transitions [arb. units]
θ̂_1
θ̂_2
Bayesian kernel inversion
computation based on posterior densities
move to next kernel

Table 5.2: *Depending on the considered kernel scale λ, the terrigenous dust records from the ODP sites 659, 721/722 and 967 have in average a different number $\langle n_t^{|\lambda} \rangle$ of kernel observations.*

kernel scale λ [Ma]	average number of kernel observations $\langle n_t^{|\lambda} \rangle$:		
	ODP 659	ODP 721/722	ODP 967
0.25	61	139	679
0.50	121	279	1345
0.75	180	417	2020
1.00	240	553	2704
1.25	299	686	3356
total signal	1221	2757	8417
time range [Ma]:	5.0	5.0	3.0

parameter is given by the step size of the global time grid as $\delta_\theta = \delta_t = 5\,\text{ka}$. The complete numerical sampling of the kernel-based approach can be summarized as

$$\begin{array}{lcllcl} \Delta_\lambda & = & [0.25,\, 1.25]\,\text{Ma}\,, & \delta_\lambda & = & 0.25\,\text{Ma} \\ \Delta_{\lambda'} = \Delta_\theta & = & \frac{3}{5}\lambda\,, & \delta_t = \delta_\theta & = & 5\,\text{ka} \\ \Delta_{\boldsymbol{s}} & = & [-0.5,\, 1.0]\,, & \delta_{\boldsymbol{s}} & = & 0.125\,. \end{array} \tag{5.2}$$

Thereby, the sampling space $\Delta_{\boldsymbol{s}}$ of the deviation parameters $\boldsymbol{s}$ is dynamically adapted to the location specific restrictions defined in Eq. 3.13 during the computational process.

Since measurement techniques of climate proxies influence the distributional shape of the resulting record $\boldsymbol{y}$ (e.g. the dust flux is naturally bounded at zero), it is a common approach to additionally investigate the log-transformed signal $ln(\boldsymbol{y})$ to better meet the model's assumption of normality [107]. The kernel weight $\chi_t^{|\lambda}$ ensures that departures from normality for both, the original and the log-transformed signal, are of comparable magnitude. Moreover, variation patterns of statistical kernel moments $m_{1-4,t}^{|\lambda}$ provide valuable information about the local signal noise and may be used to develop models with more specialized distributional assumptions. Note, even though the log-transformation weakens the influence of outliers, it also reduces the sensitivity of the method to detect changes in the variability domain. A comparative study of both, the original and the log-transformed record, is therefore not negligible but essential for a comprehensive investigation.

5.4 Results of the Bayesian inference

Based on the numerical sampling spaces defined in Sec. 5.3.2 the kernel quantities summarized in Sec. 5.3.1 are computed for each original and for each log-transformed ODP record. The derived kernel quantities are interpreted and compared for all investigated signals before composed to the final proxy probabilities of transitions. The proxy probabilities are investigated for changes occurring in all ODP records within a common time interval. In order to substantiate the interpretation of these events as conjoint transitions, the corresponding inferred transition patterns are compared. The analysis results of the ODP 659 signals for the smallest and largest kernel scales λ=[0.25, 1.25] Ma are exemplarily shown in Fig. 5.3. The results of all ODP records are provided for all investigated kernel scales in Fig. C.2-C.4.

5.4.1 Higher statistical moments - departures from normality

To derive the indicator function of normality $\chi_t^{|\lambda}$, we use the kernel estimators $(\boldsymbol{\vartheta}^*)_t^{|\lambda}$ to standardize the kernel observations (Eq. 3.53) in order to compute the statistical kernel moments $m_{1-4,t}^{|\lambda}$ (Eq. A.9). In this way, the generic transition model is considered a suitable local approximation to the kernel observations if none of the empirical kernel moments exceeds the predefined thresholds from Tab. 5.1. The kernel indicator function is presented over the global time grid Θ_t at the smallest and largest kernel scales λ=[0.25, 1.25] Ma for all investigated ODP signals in Fig. C.1. Apart from some isolated departures, the kernel mean $m_{1,t}^{|\lambda}$ and variance $m_{2,t}^{|\lambda}$ are relatively close to the theoretical values of a standard normal distribution, indicating that the normality assumption is considered to be true for most signals. General deviations from the theoretical values are obvious in the values of the kernel skewness $m_{3,t}^{|\lambda}$ and kurtosis $m_{4,t}^{|\lambda}$, indicating more asymmetric, long-tailed kernel distributions. On the one hand, however, these effects are compensated by the log-transformation of the record, e.g. in the ODP 659 record since around 2.5 Ma. On the other hand, these effects are locally enhanced by the log-transformation, e.g. in the ODP 721/722 record at around 1.7 Ma. Irrespective of the specific scale or the deviating moment, the maximal attained amount of kernels not considered to be suitably approximated by the generic transition model ($\chi_t^{|\lambda} = 0$) is listed in Tab. 5.3 for each ODP signal.

An explanation for gradually increasing deviations from normality, which appear on all scales λ, may be due to possible changes of measurement techniques of the terrigenous dust flux within an individual ODP record. Moreover, in particular singular deviations from normality may be attributed to complex patterns, which are not sufficiently approximated by the assumed generic transition model. These two effects presumably explain the deviation from normality in the log-transformed ODP 967 record, where the kernel $m_{1,t}^{|\lambda}$ and variance $m_{2,t}^{|\lambda}$ differ significantly from the theoretical values of a standard normal distribution (cf. Tab. 5.3). Also the original ODP 721/722 record is better ap-

Table 5.3: *The maximal attained number of kernels for which the assumption of normality ($\chi_t^{|\lambda} = 0$), the existence of a transition ($bf_t^{|\lambda} = 0$) or both are considered as not to be true for the investigated signals of the ODP records 659, 721/722 and 967, irrespective of the specific kernel scale λ.*

signal	maximal attained number of kernels with:		
	$\chi^{\|\lambda} = 0$ [%]	$bf^{\|\lambda} = 0$ [%]	$\chi^{\|\lambda} \cdot bf^{\|\lambda} = 0$ [%]
$\boldsymbol{y}_{659}$	52	20	57
$ln(\boldsymbol{y}_{659})$	2	19	19
$\boldsymbol{y}_{721/722}$	4	2	5
$ln(\boldsymbol{y}_{721/722})$	18	4	19
$\boldsymbol{y}_{967}$	10	2	12
$ln(\boldsymbol{y}_{967})$	61	0	61

proximated by the generic transition model than the log-transformed signal. In contrast to that, the normality indicator function accepts more kernels for the log-transformed ODP 659 record than for the original signal.

5.4.2 Bayes factor - evidence for the existence of a transition

In order to derive the kernel weight $bf_t^{|\lambda}$ indicating the existence of a transition, we compute the kernel Bayes factor $BF_t^{|\lambda}$ (see Fig. 5.3 (C,D)). The obtained $BF_t^{|\lambda}$ values are marked as empty circles in case the normality assumption is not considered as to be true in the same kernel ($\chi_t^{|\lambda} = 0$). The shaded area $-5 \leq BF_t^{|\lambda} \leq 5$ includes the values that support none of the assumed models, while the values underneath that area favor the transition model over the no-transition model, and *vice versa* (cf. Sec. 4.1). In all ODP records, we observe a change in the evidence of the transitions with varying kernel scale λ. The most dominant transitions occur after 3.0 Ma, including the most obvious two-step transition at around 2.2 Ma and 1.6 Ma. The kernel Bayes factors $BF_t^{|\lambda}$ that are discarded due to the normality indicator function $\chi_t^{|\lambda} = 0$, are often captured by the other transformation of the signal, for instance the two-step event in the ODP 967 record as well as the entire ODP 659 record since 3.0 Ma. Hence, the investigation of the original and log-transformed signals prove to be supportive by providing complementary information to the transition detection analysis. The kernels in which a no-transition or none of the proposed models is favored by the Bayes factor help us to resolve consecutive transition events by ignoring the complete kernel result via the weight $bf_t^{|\lambda} = 0$ (cf. Eq. 4.12). In our analysis, however, only a small number of kernels suggest a no-transition model as indicated in Tab. 5.3.

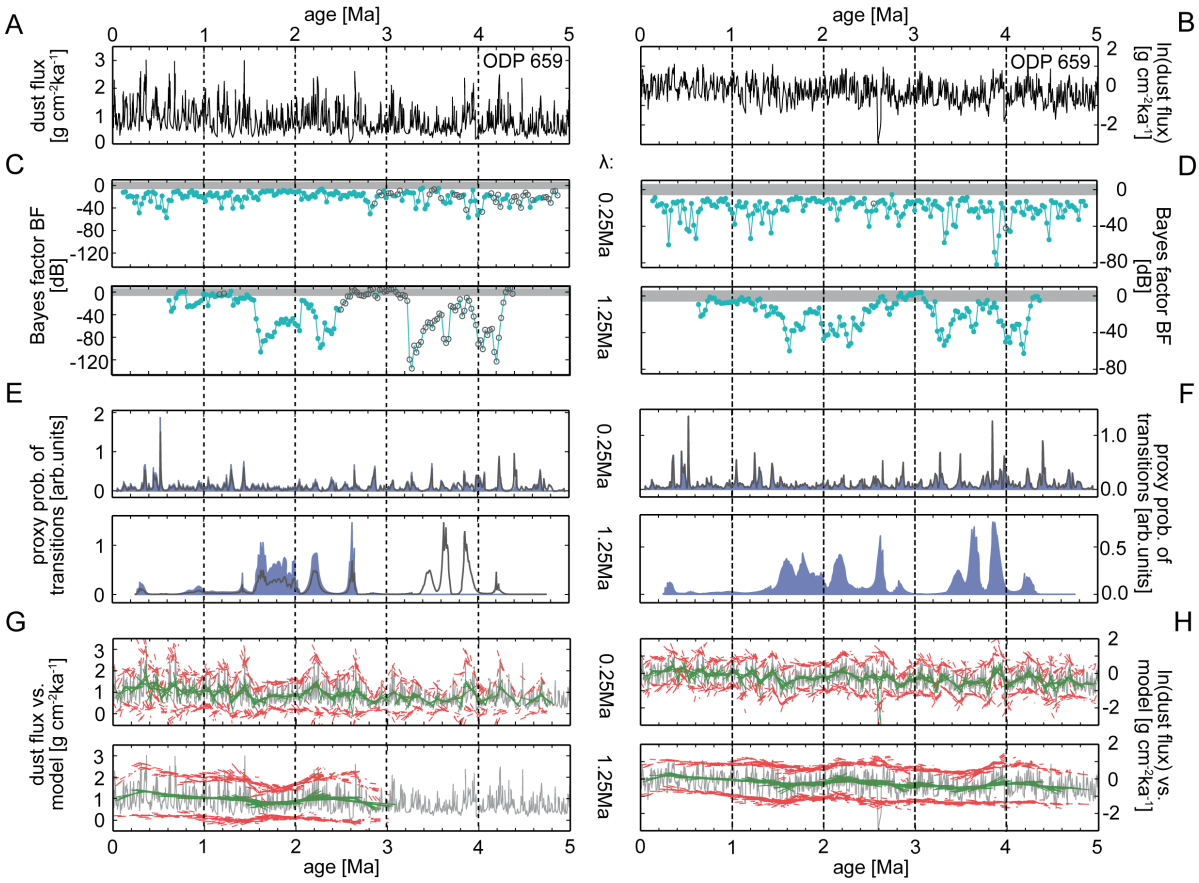


Figure 5.3: *The inference results are exemplarily presented at the smallest and largest kernel scale λ =[0.25, 1.25] Ma for (A) the ODP 659 record and (B) the log-transformed signal. (C,D) The kernel Bayes factors $BF_t^{|\lambda}$ indicate the existence of a transition for values underneath the shaded area. In case the generic transition model is considered not to adequately approximate the kernel observations ($\chi_t^{|\lambda} = 0$), the corresponding $BF_t^{|\lambda}$ is marked as an empty circle. (E,F) The derived proxy probability of transitions (blue) indicates the final degree of belief about underlying transitions. The proxy probability only weighted with respect to the Bayes factor (grey) illustrates the concept of our kernel-based inference approach: inference results are removed from the analysis in case the generic transition model does not sufficiently describe the complex observations. (G,H) The estimated kernel mean (green) and variability in terms of the 95% quantiles (red) visualize the transition patterns for kernels complying with the basic assumptions of the approach, i.e. $\chi_t^{|\lambda} \cdot bf_t^{|\lambda} \neq 0$.*

5.4.3 Proxy probability of transitions - location and credibility of transitions

The primary purpose of our analysis is the composition of the proxy probability of transitions $p(\{\theta\}|\boldsymbol{y}^{|\lambda})$ as defined in Eq. 5.1 and elaborated in Sec. 4.1. This probability represents our degree of belief about the estimated transition events actually existing in the terrigenous dust records. The analysis results of the ODP 659 record for the smallest

and largest scale λ are exemplarily shown in Fig. 5.3. The results for all investigated ODP signals at all employed kernel scales are provided in Fig. C.2-C.4. In general, consecutive transitions are successfully resolved if the proxy probability of transitions approaches zero between adjacent modes, provided the number of data points between those modes is sufficiently large. This numerical fact explains the occurrence of relatively broad local modes in the proxy probability of the ODP 659 record, whereas the ODP 967 record exhibits well-defined local modes. In order to discard spurious transitions caused by random fluctuations rather than by a deterministic cause, e.g. a climatic event, we only consider transition events that occur on multiple scales. And more precisely, the term transition event refers the time interval enclosing at least 90% of the credibility weight of the corresponding probability mode. Hereby, the relative amplitudes of the proxy probability of transitions may be interpreted as indicators for the relative dominance of the detected events at a constant scale. Using this criteria, multiple dominant transition events occur in the dust records at around (3.95-3.70) Ma, (3.35-3.15) Ma, (3.10-3.00) Ma, (2.90-2.80) Ma, (2.35-1.50) Ma, (1.00-0.80) Ma and (0.70-0.60) Ma.

To better understand the consequences of ignoring kernels that do not satisfy the normality criterion $\chi_t^{|\lambda}$, the probabilities of transitions weighted only with the transition existence criterion $bf_t^{|\lambda}$ are additionally presented in Fig. 5.3 (E,F). Even though the proxy probability weighted only by $bf_t^{|\lambda}$ clearly resolves multiple events reoccurring in the fully weighted proxy probability of the log-transformed signal, these findings are ignored within our inference approach. This is due to the fact that the generic transition model does not adequately capture the corresponding changes for this signal type.

Since not all indicated events may be properly resolved, in particular for the ODP 659 record, we survey the inferred transition patterns. In this way we aim to better understand the performance of the kernel-based inference approach, to further substantiate the resolution of pronounced events and to reasonably interpret synchronous events as conjoint transitions.

5.4.4 Estimated transition patterns - structures of transitions

In addition to the synchronicity of inferred changes in the terrigenous dust records, we compare the structure of the synchronous events as an additional criterion for the significance of conjoint transitions. As an example, the temporal variation of the mean (green) and variance (red dashed) of the ODP 659 record is show in Fig. 5.3 (G,H). The estimated models are only displayed for kernels satisfying the normality assumption and the existence of a transition, i.e. $\chi_t^{|\lambda} \cdot bf_t^{|\lambda} > 0$. The estimated kernel models reveal significant changes in mean or the variance at the detected transitions in all ODP records as presented in Fig. C.2-C.4. Therefore, the structures of the transition events in the dust records exhibit patterns that we can use to identify distinct events within a signal, to compare the distinct events between different signals and to understand the shape of the corresponding proxy probability modes. Note, that each comparative

conclusion between the ODP signals has logically to be drawn at the same kernel scale of investigation. For instance, we find events comprising a local minimum in the variance, e.g. at around 2.8 Ma in ODP 659 and 721/722 signals causing a relatively narrow proxy probability mode. In contrast, a rather gradual transition in the mean, such as at around 1.6 Ma in the ODP 659 record, causes a rather broad proxy probability mode.

Correlating the estimated transition patterns within the interval (2.35-1.50) Ma across the ODP records 659 and 721/722 enables us to resolve two distinct transitions around (2.35-2.10) Ma and (1.70-1.50) Ma. Comparing the estimated structure of the transitions provide evidence to correlate events which are slightly different in their incidence. Hence, we correlate the events at around (2.10-2.00) Ma and (1.85-1.70) Ma in the ODP 967 record with the events at around (2.35-2.10) Ma and (1.70-1.50) Ma of the ODP 659 and 721/722 records. By employing these and other inferred transitions, we are able to identify conjoint events occurring in the terrigenous dust records and discuss their significance in Plio-Pleistocene African climate based on the comprehensive overview presented in Fig .5.4.

5.5 Discussion of the identified conjoint transitions

Our Bayesian algorithm identifies multiple significant transition events in the terrigenous dust records of the OPD sites 659, 721/722 and 967 documenting tropical and subtropical African Plio-Pleistocene climate. Based on the synchronicity of these events, the similarity of their scale dependencies and the congruence of their estimated transition patterns, we conclude that these transitions are continent-wide climate events.

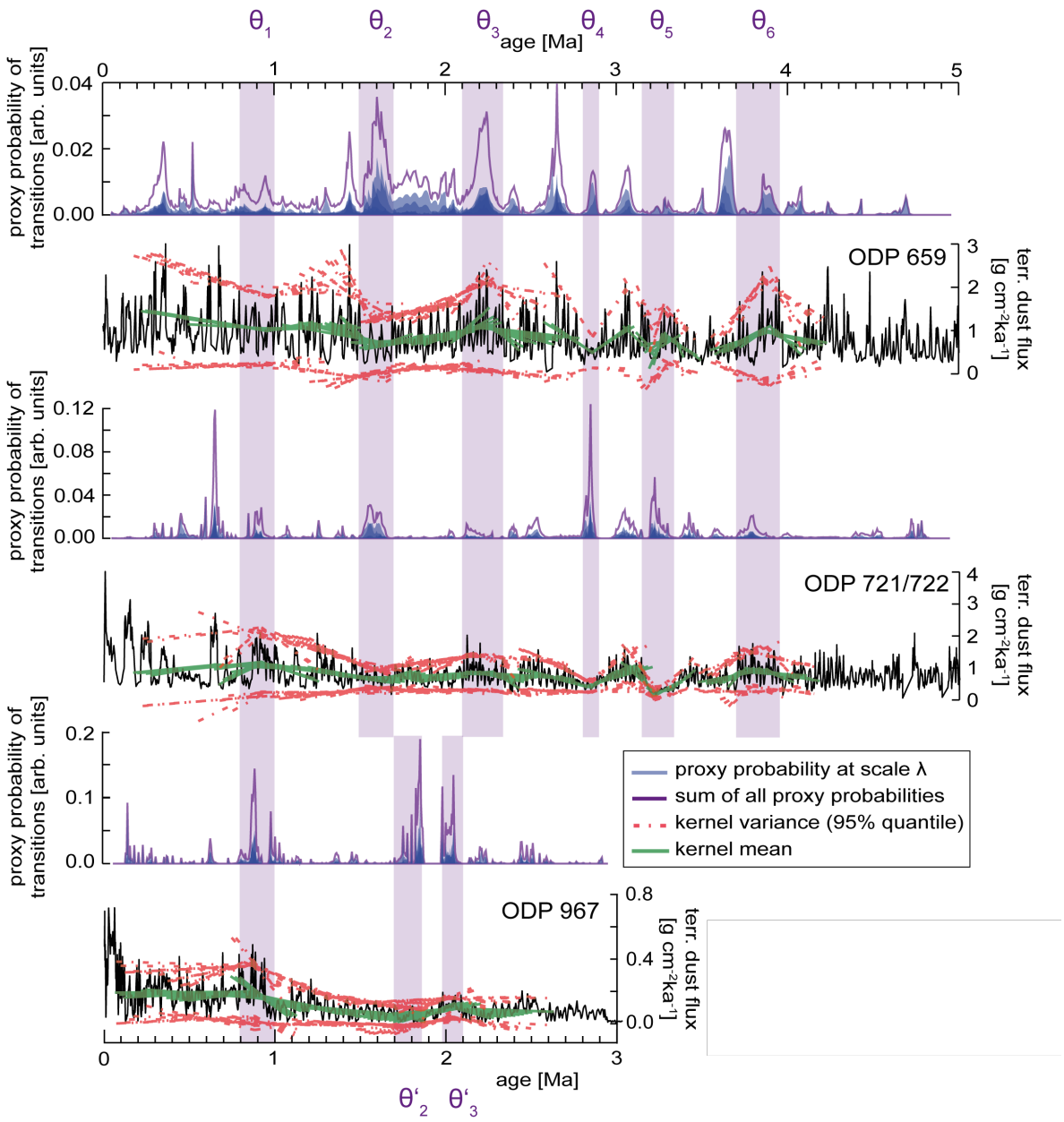
The first criterion to classify these transitions, the synchronicity of their occurrence, helps to recognize events which are representative for the climate of the entire African continent. We regard a transition to be synchronous if the transition emerges in all records independently from the scale of analysis within a constant time interval based on a given significance. According to this criterion, we find synchronous transition events at around (3.95-3.70)Ma, (3.35-3.15) Ma, (3.10-3.00) Ma, (2.90-2.80) Ma, (2.35-2.10) Ma, (1.70-1.50) Ma, (1.00-0.80) Ma and (0.70-0.60) Ma. For the second criterion to characterize these transitions, we survey the scale dependency of these events. In climatological terms, we distinguish between short-term events from those which occur at long time scales. For example, the transition at around (2.90-2.80) Ma looses its relative importance on larger scales, indicating a short-term effect of the event. In contrast, the event at around (2.35-2.10) Ma becomes more important on longer time scales. Using the third criterion, the congruence of the estimated transition patterns, we compare the temporal variation of the statistical parameters mean and variance through the synchronous events of all records. In climatological terms, we search for those events characterized by similar structures of transitions indicating a similar effect on the African climate. As an example, at around (3.35-3.15) Ma both the mean and the variance in the dust records of ODP 659 and 721/722 are increasing, remembering the dust record of

ODP 967 has no data before 3 Ma. A comprehensive summary of all important findings of the kernel-based inference approach is presented in Fig. 5.4.

While comparing the congruence and the scale dependencies of the events for the different records, we claim synchronous transitions even if they are slightly out of phase. We believe that this approach is feasible in the light of the uncertainties of the dust records, both in their age models and the interpretation of the dust flux as a climate proxy itself (e.g. [5]). We therefore correlate the transitions at (2.35-2.10) Ma and (1.70-1.50) Ma documented in the dust records of ODP 659 and 721/722 with the congruent transitions at (2.10-2.00) Ma and (1.85-1.70) Ma identified in the dust record of ODP 967. The transitions initiate a decline of mean and variance followed by a pronounced long-term increase of both properties persisting until about 1.0 Ma.

The multiple transitions identified in the dust records of the OPD sites 659, 721/722 and 967 coincide with established transitions in proxy records of Plio-Pleistocene African climate. The transition θ_6 at around (3.95-3.70) Ma, which is dominant in the Atlantic record of ODP 659, and the transition θ_5 around (3.35-3.15) Ma, which is dominant in the Arabian Sea record of ODP 721/722, correlate with important paleo-environmental shifts between (4.0-3.0) Ma in Eastern Africa towards more arid climate, accompanied by a transition from predominant C_3 (woodland) to C_4 (grassland) vegetation [132, 133]. Furthermore, the pronounced transition θ_6 in the Atlantic Ocean coincides with a cooling event punctuating the upwelling at the South African margin associated with the gradual closure of the Central American Seaway (CAS) [134]. Even though the importance of the CAS for continental African climate is under debate [135], simulation studies support a CAS induced large-scale thermal redistribution in the Atlantic Ocean leading to modified upwelling patterns [136] and a shallowing of the tropical thermocline [137]. The resulting change of tropical, coastal Atlantic SSTs are considered as a determining factor of African environmental conditions [4, 138] due to modified local climate conditions. From 3.50 to 2.95 Ma, including the event θ_5, planktonic foraminifera records reveal a gradual decrease of Indian Ocean SSTs [139], potentially influencing the coastal Atlantic

Figure 5.4 *(facing page)*: *For the terrigenous dust fluxes of the ODP sites 659, 721/722 and 967 (black) the kernel-based inference approach yields the proxy probabilities of transitions for multiple kernel scales (blue shaded). By integrating these probabilities over the kernel scales a cumulative credibility expression (purple line) is derived enhancing the modes of the identified conjoint transitions θ_{1-6} (purple shaded). For each conjoint event θ_i the inferred transition patterns, i.e. mean (green) and variability (red), are presented at a particular scale λ_i. These changes coincide with established climate events and are associated to: θ_6 at (3.95-3.70) Ma to the closure of the Central American Seaway (λ_6 =0.50 Ma), θ_5 at (3.35-3.15) Ma to the change of the Indonesian throughflow (λ_5 =0.25 Ma), θ_4 at (2.90-2.80) Ma to the intensification of the Northern Hemisphere Glaciation (λ_4 =0.50 Ma), $\theta_{3,2}$ at (2.35-2.10) Ma and (1.70-1.50) Ma ($\lambda_{3,2}$ =0.75 Ma) to the step-wise establishment of the modern state tropical SST gradients, associated with the reorganization of the Hadley-Walker-Circulation, and θ_1 at (1.00-0.80) Ma to the Middle Pleistocene transition (λ_1 =1.00 Ma). Based on the congruence of the transition patterns, we correlate the ODP 967 events $\theta'_{3,2}$ at (2.10-2.00) Ma and (1.85-1.70) Ma with the double transition $\theta_{3,2}$.*


θ1
θ2
θ3
θ4
θ5
θ6
age [Ma]
proxy probability of transitions [arb. units]
ODP 659
terr. dust flux [g cm-2ka-1]
ODP 721/722
ODP 967
proxy probability at scale λ
sum of all proxy probabilities
kernel variance (95% quantile)
kernel mean
θ'2
θ'3

SST patterns [140, 141]. Simulation studies support this shift in the thermal balance via a shallowing of the Indonesian seaway and hence the reduction of the Indonesian throughflow (ITF) as a major cause for the East African aridification [132]. Our findings of an important climate transition is corroborated by a recurrence network study on the same dust records, suggesting a dynamical transition at around (3.35-3.15) Ma, also interpreted as a consequence of the reduction of the ITF [6]. The authors of this study argued that the reduced ITF had an effect on the convective precipitation patterns over subtropical and tropical Africa.

In the early Pliocene, we detect a synchronous small-scale (short-term) event θ_4 at around (2.90-2.80) Ma. This event occurred within the time period of the gradual initiation of the Northern Hemisphere Glaciations between (3.15 -2.85) Ma as inferred from ice-rafted debris [142, 143] and the global ice volume [144]. The impact of the Intensification of the Northern Hemisphere Glaciation (INHG) at around 2.75 Ma on African climate is debated [4, 5, 39], with a weak tendency towards a more arid climate in West Africa rather than in East Africa at around 2.75 Ma, as pollen data suggest [145, 146]. In contrast to the small-scale event θ_4 at around (2.90-2.80) Ma, our analysis reveals two prominent events θ_3 and θ_2 at around (2.35-2.10) Ma and (1.70-1.50) Ma in the Atlantic Ocean (ODP 659) and the Arabian Sea (ODP 721/722), respectively, and at around (2.10-2.00) Ma and (1.85-1.70) Ma in the Eastern Mediterranean Sea (ODP 967). Previous studies have already identified an important climate transition between (2.25-1.60) Ma [4–6] and associated these with the establishment of the Walker-Circulation (WC) increasing environmental variability in the tropics [3]. In contrast with these earlier studies, however, our results clearly suggest a two-step transition rather than a single event.

While trying to understand the effect of the establishment of the WC on Plio-Pleistocene African climate, the relative importance of meridional and zonal SST gradients are the most dominant factors. During the Pliocene, tropical climate was characterized by prevailing El Niño-like conditions indicated by relatively symmetric zonal and significantly reduced meridional SST gradients [147–150]. At around 3.6 Ma, the increase of low-latitudinal SST gradients begins [151] and intensifies between (2.2-2.0) Ma [152]. The modern meridional SST patterns established around (2.2-2.0) Ma [150, 152, 153]. The increase of the meriodional gradient strengthened the Hadley-Circulation causing stronger trade winds [150, 152, 154, 155]. Moreover, the zonal SST contrast over the Indian Ocean may have contributed to the gradual strengthening of the Asian monsoon [150], supported by a dynamic transition in the Indian monsoon patterns around (1.9-1.5) Ma [43].

Simulation studies of atmospheric general circulation models demonstrate the effect of the zonal and meridional SST gradient changes on aridification of Plio-Pleistocene African climate [153, 154]. On shorter (Milankovitch) time scales, the two-step reorganization of the ocean-atmosphere system causes rapid fluctuations between wet and dry conditions, as suggested by lake-level records in the East African Rift System (EARS) with pronounced humid periods between (2.1-1.6) Ma [156], and complex

paleo-vegetation patterns linked *inter alia* to oscillating precipitation [157].

The most recent conjoint event θ_1 at around (1.00-0.80) Ma coincides with the Middle Pleistocene transition (MPT) around (1.2 - 0.7) Ma. The MPT marks a profound shift of glacial-interglacial sequences from low-amplitude 41 ka to large-amplitude 100 ka cycles [4, 158]. At this time, the tropical SST patterns experienced a resembling variability shift [158, 159], which is reflected in both the time and frequency domain of terrestrial dust records [4, 39], again accompanied by another episode of high but fluctuating lakes in the EARS between (1.1-0.9) Ma [156, 160].

5.6 Conclusion

We used a kernel-based Bayesian inference approach to detect generic transitions in the mean and variability of terrigenous dust records from the ODP sites 659, 721/722 and 967. The estimated conjoint transition events are interpreted as major shifts towards a more arid and more variable climate in Plio-Pleistocene Africa. The most remarkable finding of this study is the establishment of the Walker-Circulation and its influence on African climate as a two-step event centered at (2.35-2.10) Ma and (1.70-1.50) Ma, which has not yet been resolved by other studies. The well-constrained chronology of the two-step increase in aridity and variability contributes to the current discussion on climate-evolution linkages at around 2 Ma ago, which has recently shifted away from the outdated correlation of the first appearance of the genus *Homo* near the onset of the high-latitude glaciations [161].

6 Summary

In this thesis a kernel-based Bayesian inference approach is developed to decipher changes in complex climate observations. For the purpose of investigating direct as well as indirect climate observations a generic transition model is proposed. Moreover, a degree of belief about multiple transitions is formulated in order to facilitate the comparison of the inference results between different time series.

The first step towards the intended approach is the design of a generic transition model as a first order approximation to the statistical moments of complex observations. In principle every time series that is at least locally differentiable up to the first order prior and after the change may be approximated by the generic model. The beneficial model design and the choice of least informative prior distributions enables the separation of the Gaussian from the intrinsic non-linear part of the inference task. Besides clarifying the structure of the model, the formulation speeds up the computations considerably. The basic inference approach proves to convincingly infer on an underlying change, independent of the specific transition geometry. In particular the transition parameter proves to be robust even for sparse data situations and in the presence of outliers. Based on the analysis of synthetic time series a minimal empirical threshold of 50 observations is recommended to ensure a reliable inference on the complete transition patterns.

The suitability of the generic transition model is investigated for direct climate observations of the annual Nile river level in Aswan, Egypt, and the annual mean temperature measured at a weather station in Tuscaloosa, Alabama, both comprising documented changes. For the Nile observations, the approach correctly infers on a single underlying transition. For the temperature observations, the approach correctly infers on one out of four underlying transitions. The consideration of the normality assumption clearly indicates that the generic transition model does not sufficiently capture the time series comprising more than one change and motivates the generalization of the approach in order to investigate time series for an unknown number of changes.

By introducing a time dependent prior distribution for the transition parameter the basic Bayesian inference is extended to a flexible kernel-based approach. For the purpose of adequately combining the kernel posterior distributions of the transition parameter the kernel Bayes factor is used to formulate a kernel weight indicating the existence of a transition within the kernel. Hence, the composition of the weighted kernel posterior distributions yields a proxy probability to a multiple transition posterior distribution given that each kernel comprises maximal one underlying transition. Since for climate observations neither number nor location of the changes are *a priori* known the inference approach essentially needs to be applied at multiple kernel scales. Thereby, a further

kernel weight is introduced indicating the suitability of the generic transition model as an approximation to the data at a particular scale.

The previously investigated direct observations are reanalyzed with the extended inference approach. For the Nile observations, the approach confirms one underlying change. For the temperature observations, the approach infers on two out of four changes correctly. Here, the limitation of the approach with respect to the number of available observations per kernel is revealed. To resolve the underlying changes kernel scales are required that contain less observations than the recommended minimal amount for a robust inference. The proposed kernel weights achieve to remove the kernel distributions from the analysis for kernels observations that do not comply to the basic model assumptions. On the one hand, the kernel measures thus offer an essential outlier removal and on the other hand, provide further insight into the specific deviations, in terms of statistical moments, from the basic model assumptions.

Based on the acquired understanding of the performance and requirements of the introduced kernel-based Bayesian inference, the approach is proposed as a powerful diagnostic tool to investigate real time series for multiple transitions. For this reason, the approach is used to investigate a set of terrigenous dust fluxes obtained from the ODP sites 659, 721/722 and 967 of the last 5 Ma. The terrigenous dust is considered as an indicator of the tropical and subtropical African climate and thus offers indirect climate observations of the Plio-Pleistocene period. By investigating the synchronicity of the inferred transitions, the congruence of the corresponding transition patterns and the scale dependency of these events, conjoint changes are identified that are interpreted as continent-wide climate events.

Based on findings from previous analysis approaches on various proxy records and simulation studies the conjoint changes are associated to established climate events: (i) the closure of the Central American Seaway around (3.95-3.70) Ma, (ii) the change of the Indonesian throughflow at (3.25-3.15) Ma, (iii) the intensification of the Northern Hemisphere Glaciation between (2.90-2.80) Ma, (iv) the two-step establishment of the zonal sea-surface temperature gradient in the tropics at (2.35-2.10) Ma and (1.70-1.50) Ma, associated with the reorganization of the Hadley-Walker-Circulation, and (v) the Middle Pleistocene transition at (0.95-0.85) Ma. The most remarkable finding is the well-constrained two step increase in aridity and variability associated to the establishment of the Walker-Circulation. Thus, the investigation contributes to the current discussion on climate-evolution linkages in tropical and subtropical Africa at around 2 Ma ago.

In conclusion, the introduced kernel-based Bayesian inference approach realizes a powerful diagnostic tool to investigate direct and indirect climate observations for multiple generic transitions. In principle a variety of natural processes can be investigated, however, it might be reasonable to provide a set of generic transition models in order to take as well into account specific process properties, e.g. generic correlation patterns. An appropriate way to select the most suitable generic process model to an individual transition may then be accomplished by a kernel measure based on the Bayes factor.

By unraveling the locations, patterns and scale dependencies of generic transition events the introduced Bayesian approach offers an important insight into multiple changes potentially underlying direct and indirect climate observations.

A Computations and definitions

A.1 Computational steps used within the program

Cholesky decomposition of the covariance matrix

Every symmetric and positive-definite matrix may be factorized into the product of a unique lower triangular matrix and its transpose. Thus we may decompose the inverse covariance matrix Ω^{-1} as

$$\Omega^{-1} = L_{\Omega^{-1}} L_{\Omega^{-1}}^T \tag{A.1}$$

to speed up our numerical computations. The exponent of the likelihood function $\mathcal{L}$ defined in Eq.(3.16) may be rewritten as the quadratic norm

$$(\boldsymbol{y} - F\boldsymbol{\beta})^T \Omega^{-1} (\boldsymbol{y} - F\boldsymbol{\beta}) = \| L_{\Omega^{-1}}^T (\boldsymbol{y} - F\boldsymbol{\beta}) \|^2 , \tag{A.2}$$

from which we benefit in the computation of the BLUP.

QR decomposition to obtain the BLUP

To solve the linear least square problem in Eq.(3.18) we use the common strategy of the QR decomposition. Any matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and $rank(A) = n$ may be factorized into a product of an orthogonal matrix $Q \in \mathbb{R}^{m \times m}$ and an upper triangular matrix $R \in \mathbb{R}^{m \times n}$

$$A = QR = [Q_1, Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} . \tag{A.3}$$

The matrices may be further partitioned into the orthogonal submatrices $Q_1 \in \mathbb{R}^{m \times n}$ and $Q_2 \in \mathbb{R}^{m \times (m-n)}$ with orthogonal columns and the upper triangular matrix $R_1 \in \mathbb{R}^{n \times n}$ and the zero matrix $O \in \mathbb{R}^{(m-n) \times n}$. Since the orthogonal matrix does not change a norm, we may rewrite Eq.(A.2) as

$$\begin{aligned} \| Q^T L_{\Omega^{-1}}^T (\boldsymbol{y} - F\boldsymbol{\beta}) \|^2 &= \| Q^T L_{\Omega^{-1}}^T \boldsymbol{y} - Q^T L_{\Omega^{-1}}^T F\boldsymbol{\beta} \|^2 = \\ &= \| Q^T L_{\Omega^{-1}}^T \boldsymbol{y} - Q^T QR\boldsymbol{\beta} \|^2 = \end{aligned}$$

$$
\begin{aligned}
&= \left\| \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} L_{\Omega^{-1}}^T \boldsymbol{y} - \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} [Q_1, Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \boldsymbol{\beta} \right\|^2 = \\
&= \left\| Q_1^T L_{\Omega^{-1}}^T \boldsymbol{y} - R_1 \boldsymbol{\beta} \right\|^2 + \left\| Q_2^T L_{\Omega^{-1}}^T \boldsymbol{y} \right\|^2 \stackrel{!}{=} \min \qquad \text{(A.4)}
\end{aligned}
$$

by using the QR decomposition of

$$
\left\| L_{\Omega^{-1}}^T F \right\|^2 = \left\| Q_1 R_1 \right\|^2 = R_1^T R_1 = F^T \Omega^{-1} F \,.
$$

The minimum condition in Eq.(A.4) may be used to determine an alternative formulation of the BLUP as defined Eq.(3.18) within the framework of the employed decompositions,

$$
\boldsymbol{\beta}^* = R_1^{-1} Q_1^T L_{\Omega^{-1}}^T \boldsymbol{y} \,. \qquad \text{(A.5)}
$$

A.2 Supplementary computations

Jeffreys prior of the liner model

The Fisher information matrix for the parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \sigma, \theta, \boldsymbol{s})$ and the likelihood function $\mathcal{L}$ defined in Eq.(3.16) can be written as:

$$
\mathcal{J}(\boldsymbol{\vartheta}) = -\mathbb{E}\left[\begin{pmatrix} \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial^2 \boldsymbol{\beta}} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \boldsymbol{\beta} \partial \sigma} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \boldsymbol{\beta} \partial \theta} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \boldsymbol{\beta} \partial \mathbf{s}} \\ \vdots & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial^2 \sigma} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \sigma \partial \theta} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \sigma \partial \mathbf{s}} \\ \vdots & \cdots & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial^2 \theta} & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial \theta \partial \mathbf{s}} \\ \vdots & \cdots & \cdots & \frac{\partial^2 ln(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y}))}{\partial^2 \mathbf{s}} \end{pmatrix} \right] \qquad \text{(A.6)}
$$

The derived diagonal elements are

$$
\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial^2 \boldsymbol{\beta}} \right] &= \frac{1}{\sigma^2} F^T \Omega^{-1} F = j_{11} \\
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial^2 \sigma} \right] &= -\frac{n}{\sigma^2} = j_{22} \\
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial^2 \theta} \right] &= \frac{1}{2} tr\left(\Omega^{-1} \frac{\partial^2 \Omega}{\partial^2 \theta} - \Omega^{-1} \frac{\partial \Omega}{\partial \theta} \Omega^{-1} \frac{\partial \Omega}{\partial \theta} \right) + \frac{1}{\sigma^2} \left(\frac{\partial F}{\partial \boldsymbol{\theta}} \boldsymbol{\beta} \right)^T \Omega^{-1} \frac{\partial F}{\partial \boldsymbol{\theta}} \boldsymbol{\beta} = j_{33} \\
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial^2 \mathbf{s}} \right] &= \frac{1}{2} tr\left(\Omega^{-1} \frac{\partial^2 \Omega}{\partial^2 \mathbf{s}} - \Omega^{-1} \frac{\partial \Omega}{\partial \mathbf{s}} \Omega^{-1} \frac{\partial \Omega}{\partial \mathbf{s}} \right) = j_{44} \,,
\end{aligned}
$$

and the non-zero, symmetric off-diagonal elements are

$$
\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial\boldsymbol{\beta}\partial\theta}\right] = -\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial\boldsymbol{\theta}\partial\beta}\right] &= \frac{1}{2\sigma^2}\left(\left(\frac{\partial F}{\partial\boldsymbol{\theta}}\boldsymbol{\beta}\right)^T\Omega^{-1}F + F^T\Omega^{-1}\frac{\partial F}{\partial\boldsymbol{\theta}}\boldsymbol{\beta}\right) \\
&= j_{13} = j_{31} \\
-\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial\theta\partial\mathbf{s}}\right] = -\mathbb{E}\left[\frac{\partial^2 ln\left(\mathcal{L}(\boldsymbol{\vartheta}|\boldsymbol{y})\right)}{\partial\mathbf{s}\partial\theta}\right] &= \frac{1}{2}tr\left(\Omega^{-1}\frac{\partial^2\Omega}{\partial\mathbf{s}\partial\theta} - \Omega^{-1}\frac{\partial\Omega}{\partial\mathbf{s}}\Omega^{-1}\frac{\partial\Omega}{\partial\theta}\right) \\
&= j_{34} = j_{43}\,.
\end{aligned}
$$

The determinant of the Fisher information matrix may be written as

$$
|\mathcal{J}(\boldsymbol{\vartheta})|^{\frac{1}{2}} = \left(j_{22}\cdot\left(j_{11}j_{33}j_{44} - j_{11}j_{34}j_{43} - j_{13}j_{31}j_{44}\right)\right)^{\frac{1}{2}} \tag{A.7}
$$

and therefore leading to a rather complex expression dependent on all parameters of the model.

A.3 Supplementary definitions

Higher statistical moments

The statistical moment m_i of order i about a value c and of the probability density function $p(x)$ is generally defined as

$$
m_i(c) = \int_{-\infty}^{+\infty}(x-c)^i\cdot p(x)dx = \mathbb{E}\left[(x-c)^i\right]\,. \tag{A.8}
$$

Based on these formula we derive the theoretical m_i and empirical m_i' higher centralized and standardized statistical moments of a normal distribution:

$$
\begin{aligned}
m_1(0) &= \mathbb{E}\left[(x-0)\right] \;\rightarrow\; m_1'(0) &&= \tfrac{1}{n}\sum_{j=0}^{n}(x_j-0) \\
m_2(0) &= \mathbb{E}\left[(x-0)^2\right] \;\rightarrow\; m_2'(0) &&= \tfrac{1}{n}\sum_{j=0}^{n}\left(x_j - m_1'(0)\right)^2 \\
m_3(0) &= \mathbb{E}\left[(x-0)^3\right] \;\rightarrow\; m_3'(0) &&= \tfrac{1}{n}\sum_{j=0}^{n}\left(\tfrac{x_j-m_1'(0)}{m_2'(0)}\right)^3 \\
m_4(0) &= \mathbb{E}\left[(x-0)^4\right] \;\rightarrow\; m_4'(0) &&= \tfrac{1}{n}\sum_{j=0}^{n}\left(\tfrac{x_j-m_1'(0)}{m_2'(0)}\right)^4
\end{aligned}
\tag{A.9}
$$

.

B Numerical settings

To achieve a convenient text flow all employed numerical settings for generating artificial observations (model setting) and, if performed, for inferring on the underlying model (computational setting) are listed in the following. Besides the model's parameters $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \sigma, \theta, \boldsymbol{s})$, we summarize the number of observation points n_{obs}, the numerical sampling ranges $\Delta\boldsymbol{\vartheta}$ and sampling step size $\delta\boldsymbol{\vartheta}$.

B.1 Strategy to obtain the numerical sampling spaces

The marginal posterior density $p(\theta, \boldsymbol{s}|\boldsymbol{y})$ of Eq. (3.29) can be derived by pure analytical integration. This probability distribution depicts the most objective marginalization of the joint posterior density since so far no numerical approaches need to be used to evaluate the integration. Consequently, we begin the inference by estimating the transition parameter θ and the deviation parameters $\boldsymbol{s}$ of the deviation. The numerical sampling range Δ_θ of the transition θ is chosen as maximal, namely we assume all time points t_i of the series as possible locations of the transition, except the first and last one. To avoid side effects we additionally ignore the first and last ten data points. The chosen sampling step $\delta\theta$ defines the resolution of our Bayesian inference. For the deviation parameters the common sampling range $\Delta_{s_1} = \Delta_{s_2} = \Delta_{\boldsymbol{s}}$ and reasonable sampling steps $\delta s_1 = \delta s_2 = \delta \boldsymbol{s}$ may be roughly estimated from the obvious variability of the observations. It is noteworthy, that the choice of a common sampling range of the deviation slope parameters prior s_1 and after s_2 the transition event is motivated by its intuitive interpretation. Any deviation from the diagonal of the parameter space $\Delta_{\boldsymbol{s}}$ points to a variability change at the transition. To ensure the subjective choice of the deviation sampling as reasonable, the derived marginal posterior $p(\boldsymbol{s}|\boldsymbol{y})$ has to be checked for artificial cutoffs and the sampling grid needs to be adjusted if necessary.

In order to minimize the computational costs for the higher-dimensional marginal posterior densities $p(\theta, \boldsymbol{s}, \sigma|\boldsymbol{y})$ and $p(\theta, \boldsymbol{s}, \boldsymbol{\beta}|\boldsymbol{y})$ we use the gained information about the transition and deviation parameters. The sampling range Δ_θ is reduced to the 95% confidence interval $\mathrm{CI}_{0.95}$ of the MAP estimate $\hat{\theta}$. The sampling range $\Delta_{\boldsymbol{s}}$ is reduced to the minimal lower and maximal upper limit of the $\mathrm{CI}_{0.95}$ of the MAP estimates $\hat{s}_1$ and $\hat{s}_2$. The numerical sampling ranges of the scale Δ_σ and the mean parameters Δ_{β_j} with $j = 0, 1, 2$ are centered at the PLH and BLUP estimates σ^* and $\boldsymbol{\beta}^*$, derived from $p(\theta, \boldsymbol{s}|\boldsymbol{y})$. The remaining information of the numerical setting, that is the extension of

Table B.1: *General strategy to evaluate the numerical sampling grids, i.e. the range Δ and step δ, necessary for the computation of the marginal posterior densities. Based on the confidence intervals $CI_{0.95}$ of the MAP estimates $\hat{\theta}$, $\hat{s}_1$ and $\hat{s}_2$ and the PLH σ^* and BLUP $\boldsymbol{\beta}^*$ estimates derived from $p(\theta, \boldsymbol{s}|\boldsymbol{y})$ the adjusted sampling grids may be constructed. The variables $x_{(\cdot)}$ and $\delta_{(\cdot)}$ may be chosen with respect to the obvious behavior of the time series.*

step 1: initial numerical setting to compute $p(\theta, \boldsymbol{s}|\boldsymbol{y})$:

parameter	θ	(s_1, s_2)	σ	β_0	(β_1, β_2)
sampling range Δ	$[10.0, 90.0]$	$[0.00, 0.50]$	-	-	-
sampling step δ	0.5	0.01	-	-	-
derived estimates	$\hat{\theta}$	$(\hat{s}_1, \hat{s}_2)$	σ^*	β_0^*	β_1^*, β_2^*

step 2: adapted setting to compute $p(\theta, \boldsymbol{s}, \sigma|\boldsymbol{y})$ and $p(\theta, \boldsymbol{s}, \boldsymbol{\beta}|\boldsymbol{y})$:

parameter	θ	(s_1, s_2)	σ	β_0	(β_1, β_2)
adapted range Δ'	$\mathrm{CI}_{0.95}$ of $\hat{\theta}$	$[\min(^\dagger), \max(^\dagger)]$ † $\mathrm{CI}_{0.95}$ of $\hat{s}_1$ and $\hat{s}_2$	$[0.0, \sigma^* + x_\sigma]$	$\beta_0^* \pm x_{\beta_0}$	$\beta_{1,2}^* \pm x_{\beta_{1,2}}$
adapted step δ'	0.5	0.01	δ_σ	δ_{β_0}	δ_{β_1,β_2}
derived estimates	$\hat{\theta}'$	$(\hat{s}_1', \hat{s}_2')$	$\hat{\sigma}'$	$\hat{\beta}_0'$	$\hat{\beta}_1', \hat{\beta}_2'$

the sampling ranges and steps, may be chosen with respect to the obvious behavior of the time series. Similar to the common sampling of the deviation slopes we choose a common sampling grid for the mean slopes $\Delta_{\beta_1} = \Delta_{\beta_2} = \Delta_{\beta_{1,2}}$. Thus, any deviation from the diagonal of the parameter space $\Delta_{\beta_{1,2}}$ points towards a trend change at the transition. Again, the marginal posteriors have to be checked for artificial cutoffs and the sampling grids need to be adjusted if necessary. The numerical setting to compute the marginal posterior density $p(\theta, \boldsymbol{s}|\boldsymbol{y})$ given the observations $\boldsymbol{y}$ of Fig 3.2 and the strategy to obtain the adapted setting to sample the high-dimensional marginal posterior densities $p(\theta, \boldsymbol{s}, \sigma|\boldsymbol{y})$ and $p(\theta, \boldsymbol{s}, \boldsymbol{\beta}|\boldsymbol{y})$ are exemplified in Tab. B.1.

B.2 Numerical settings of investigated synthetic time series

Table B.2: *Setting of the transition model $\mathcal{M}_{break}$ analyzed in Sec. 3.3.2 over the numbers of observations $n = 10, 12, ..., 200$ within the sampling space given below.*

parameter ϑ_i	value	sampling space	
		range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	100	$[\,10.0, 190.0\,]$	1.0
s_1	-0.003	$[\,$-0.03, 0.03$\,]$	0.005
s_2	-0.005	$[\,$-0.03, 0.03$\,]$	0.005
σ	1.4	-	-
β_0	4.00	-	-
β_1	-0.14	-	-
β_2	0.10	-	-

Table B.3: *Setting of the transition model $\mathcal{M}_{break}$ analyzed in Sec. 3.3.3 within the sampling space given below.*

parameter ϑ_i	value	sampling space	
		range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	50	$[\,5.0, 95.0\,]$	0.5
s_1	0.15	$[\,$-0.10, 0.40$\,]$	0.05
s_2	0.05	$[\,$-0.10, 0.40$\,]$	0.05
σ	1.5	-	-
β_0	10.0	-	-
β_1	0.6	-	-
β_2	0.3	-	-

Table B.4: *Setting of the transition model $\mathcal{M}_{break}$ analyzed in Sec. 4.1 within the sampling space given below.*

parameter ϑ_i	value	sampling space range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	80	given by $p_t(\theta)$	1
s_1	0.18	$[$-0.20, 0.50$]$	0.05
s_2	0.04	$[$-0.20, 0.50$]$	0.05
σ	1.2	-	-
β_0	12.0	-	-
β_1	0.24	-	-
β_2	0.02	-	-

Table B.5: *Setting of the transition model with the mean is defined as* $F\boldsymbol{\beta} = 14 + 0.2 \cdot \boldsymbol{\zeta}_{-}^{40} + 0.1 \cdot \boldsymbol{\zeta}_{+}^{40} - 0.25 \cdot \boldsymbol{\zeta}_{+}^{100} + 0.3 \cdot \boldsymbol{\zeta}_{+}^{160}$ *and the variance as* $\sigma^2\Omega = \left[1.6(1 + 0.2 \cdot \boldsymbol{\zeta}_{-}^{40} + 0.03 \cdot \boldsymbol{\zeta}_{+}^{40} - 0.05 \cdot \boldsymbol{\zeta}_{+}^{100} + 0.1 \cdot \boldsymbol{\zeta}_{+}^{160})\right]^2$, *analyzed in Sec. 4.2 within the sampling space given below.*

parameter ϑ_i	value	sampling space range Δ_{ϑ_i}	step δ_{ϑ_i}
θ	-	given by $p_t(\theta)$	1
s_1	-	$[$-0.20, 0.50$]$	0.05
s_2	-	$[$-0.20, 0.50$]$	0.05

C Results and Figures

Figure C.1 ***(facing page)*****:** *The empirical kernel moments* $m^{|\lambda}_{1-4,t}$ *computed for the dust signals from the ODP sites 659, 721/722 and 967 exemplarily presented for each kernel center* t *at the smallest and largest kernel scales* $\lambda = [0.25, 1.25]$ *Ma. The black dashed lines mark the theoretical statistical moments of normally distributed observations (cf. Tab. 3.4). The shaded areas indicate the predefined thresholds for which departures from normality are considered as negligible (cf. Tab 5.1).*

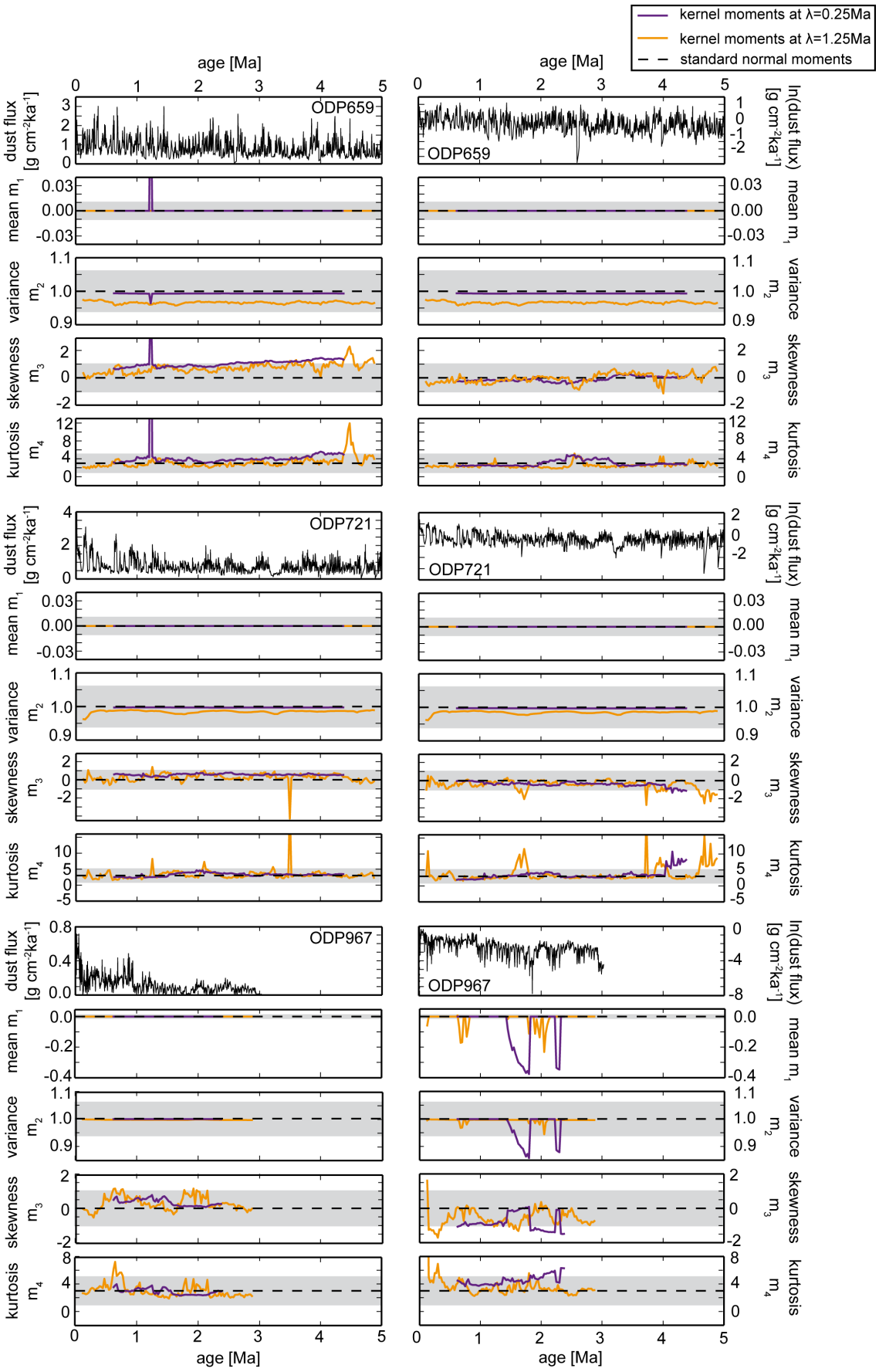

kernel moments at λ=0.25Ma
kernel moments at λ=1.25Ma
standard normal moments
age [Ma]
ODP659
ODP721
ODP967
dust flux [g cm-2ka-1]
ln(dust flux) [g cm-2ka-1]
mean m1
variance m2
skewness m3
kurtosis m4

Figure C.2 ***(facing page)*****:** *The complete analysis results for (A) the ODP 659 record and (B) the log-transformed signal for all investigated kernel scales λ. (C,D) The kernel Bayes factors $BF_t^{|\lambda}$ indicate the existence of a transition for values underneath the shaded area. In case the generic transition model is considered not to adequately approximate the kernel observations ($\chi_t^{|\lambda} = 0$), the corresponding $BF_t^{|\lambda}$ is marked as an empty circle. (E,F) The derived proxy probability of transitions (blue) indicates the degree of belief about underlying transitions. (G,H) The estimated kernel mean (green) and variability in terms of the 95% quantiles (red) visualize the transition patterns for kernels complying with the basic assumptions of the approach, i.e. $\chi_t^{|\lambda} \cdot bf_t^{|\lambda} \neq 0$.*

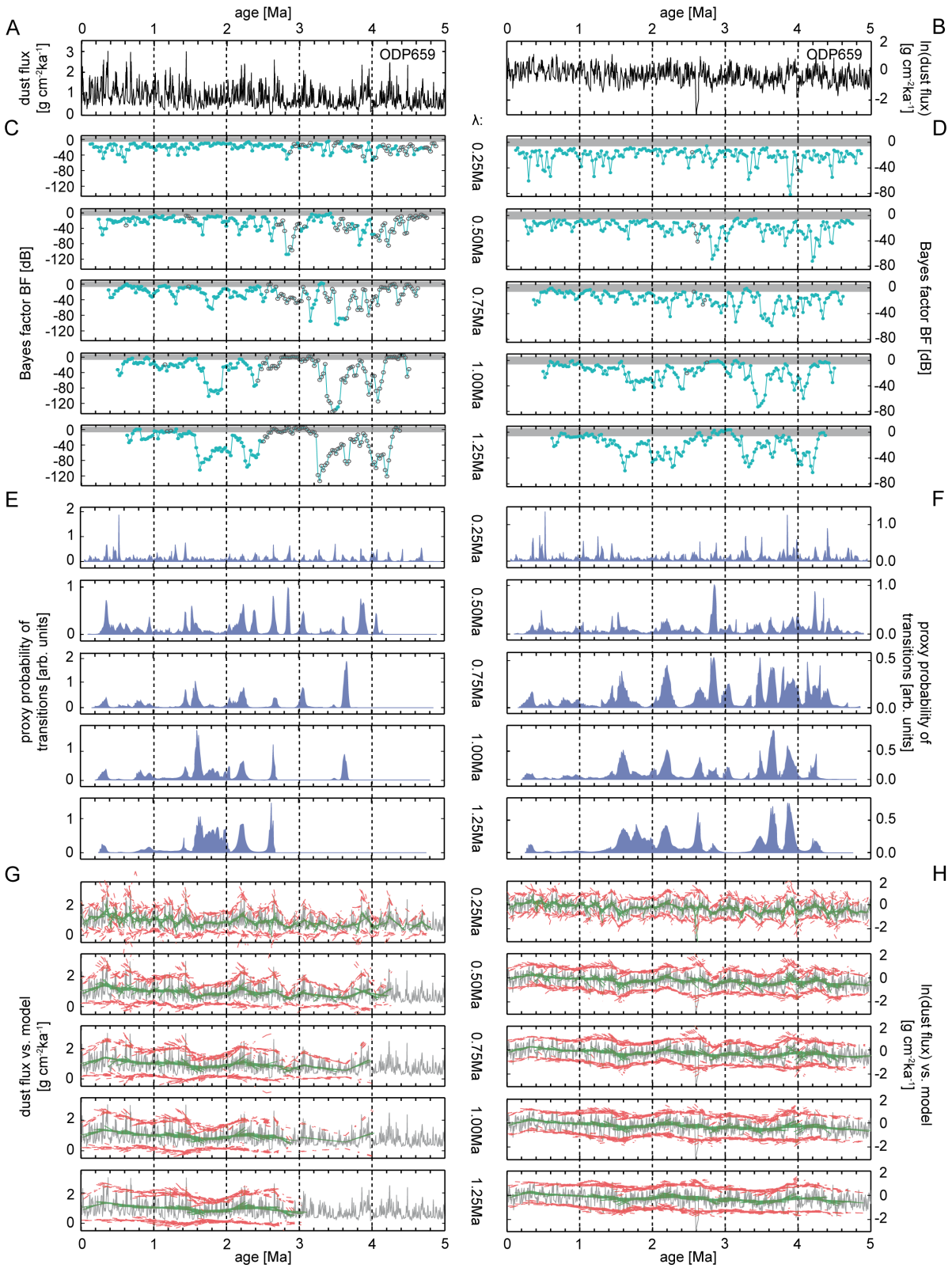

A
B
C
D
E
F
G
H
age [Ma]
dust flux [g cm-2ka-1]
ln(dust flux) [g cm-2ka-1]
ODP659
λ:
0.25Ma
0.50Ma
0.75Ma
1.00Ma
1.25Ma
Bayes factor BF [dB]
proxy probability of transitions [arb. units]
dust flux vs. model [g cm-2ka-1]
ln(dust flux) vs. model [g cm-2ka-1]

Figure C.3 *(facing page)*: *The complete analysis results for (A) the ODP 721/722 record and (B) the log-transformed signal for all investigated kernel scales λ. (C,D) The kernel Bayes factors $BF_t^{|\lambda}$ indicate the existence of a transition for values underneath the shaded area. In case the generic transition model is considered not to adequately approximate the kernel observations ($\chi_t^{|\lambda} = 0$), the corresponding $BF_t^{|\lambda}$ is marked as an empty circle. (E,F) The derived proxy probability of transitions (blue) indicates the degree of belief about underlying transitions. (G,H) The estimated kernel mean (green) and variability in terms of the 95% quantiles (red) visualize the transition patterns for kernels complying with the basic assumptions of the approach, i.e. $\chi_t^{|\lambda} \cdot bf_t^{|\lambda} \neq 0$.*

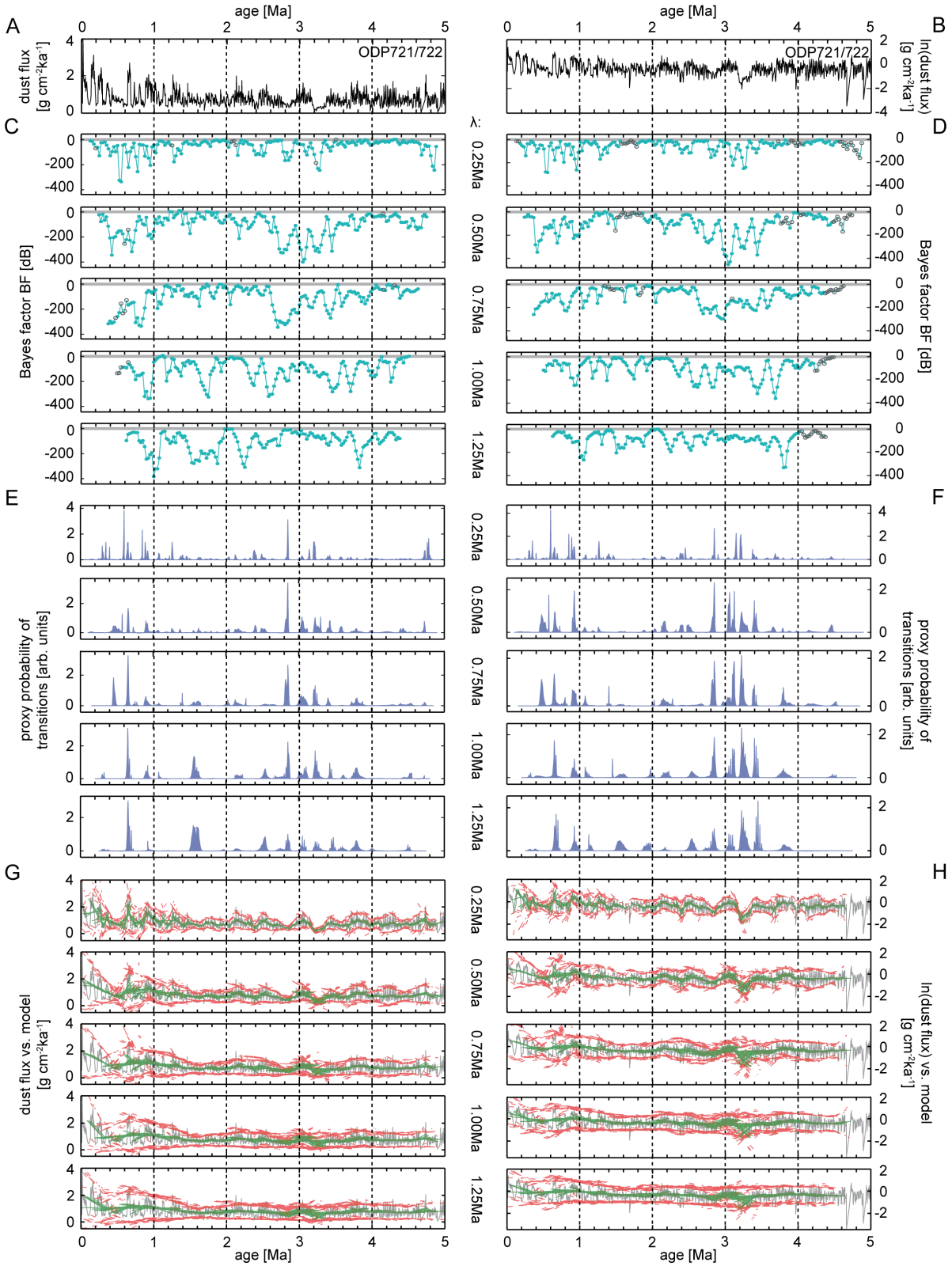

A
B
C
D
E
F
G
H
age [Ma]
ODP721/722
dust flux [g cm-2ka-1]
ln(dust flux) [g cm-2ka-1]
Bayes factor BF [dB]
proxy probability of transitions [arb. units]
dust flux vs. model [g cm-2ka-1]
ln(dust flux) vs. model [g cm-2ka-1]
λ:
0.25Ma
0.50Ma
0.75Ma
1.00Ma
1.25Ma

Figure C.4 ***(facing page)*****:** *The complete analysis results for (A) the ODP 967 record and (B) the log-transformed signal for all investigated kernel scales λ. (C,D) The kernel Bayes factors $BF_t^{|\lambda}$ indicate the existence of a transition for values underneath the shaded area. In case the generic transition model is considered not to adequately approximate the kernel observations ($\chi_t^{|\lambda} = 0$), the corresponding $BF_t^{|\lambda}$ is marked as an empty circle. (E,F) The derived proxy probability of transitions (blue) indicates the degree of belief about underlying transitions. (G,H) The estimated kernel mean (green) and variability in terms of the 95% quantiles (red) visualize the transition patterns for kernels complying with the basic assumptions of the approach, i.e. $\chi_t^{|\lambda} \cdot bf_t^{|\lambda} \neq 0$.*

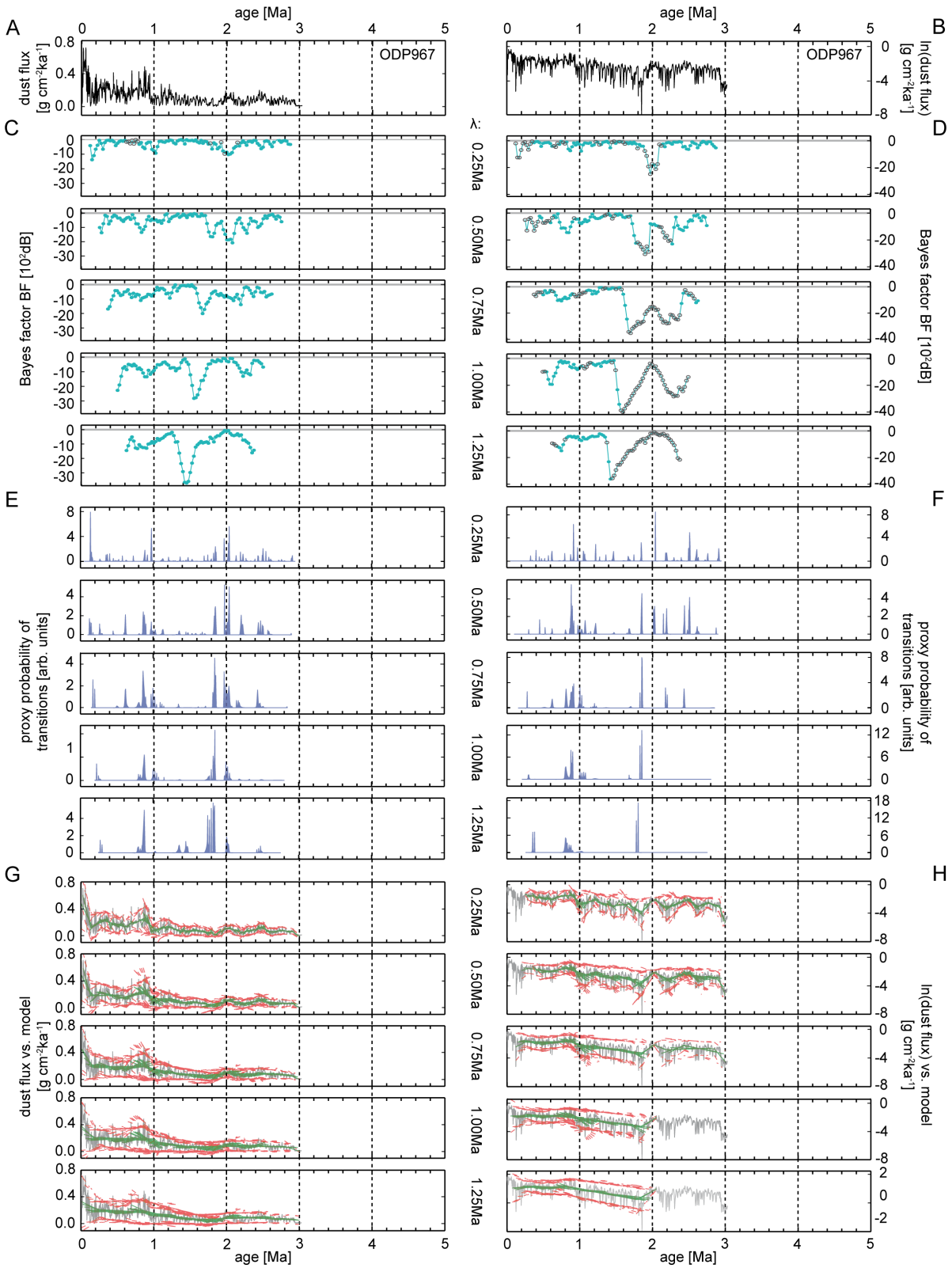

A
B
C
D
E
F
G
H
ODP967
ODP967
age [Ma]
age [Ma]
dust flux [g cm-2ka-1]
ln(dust flux) [g cm-2ka-1]
Bayes factor BF [10²dB]
Bayes factor BF [10²dB]
proxy probability of transitions [arb. units]
proxy probability of transitions [arb. units]
dust flux vs. model [g cm-2ka-1]
ln(dust flux) vs. model [g cm-2ka-1]
λ:
0.25Ma
0.50Ma
0.75Ma
1.00Ma
1.25Ma
age [Ma]
age [Ma]

Table C.1: *Numerical results of the inference on the parameters of the transition model $\mathcal{M}_{break}$ given the synthetic time series of different observational evidence presented in Fig. 3.4. The true parameter settings generating the different change geometries are presented in the shaded columns. For each model parameter the maximum a posteriori (MAP) estimate with according confidence interval of 95% significance ($CI_{0.95}$) is computed. To compare the estimation performance alternative estimators in terms of the BLUP for the mean $\boldsymbol{\beta}$ and in terms of the PLH for the scale σ are calculated. The accuracy is comparable for both approaches. The computational cost, however, are considerably smaller for the BLUP and the PLH estimators. Therefore, the highlighted estimates represent the optimal set of estimators to infer on the underlying transition.*

parameter	**minima at transition**				**maxima at transition**				transition in
		MAP		PLH/BLUP		MAP		PLH/BLUP	
	true	estimate	$CI_{0.95}$	estimate	true	estimate	$CI_{0.95}$	estimate	
θ	40.0	**40.5**	[36.5 , 44.0]	-	40.0	**39.5**	[37.5 , 42.0]	-	**mean and deviation**
s_1	0.20	**0.17**	[0.07 , 0.40]	-	-0.02	**-0.021**	[-0.024 , -0.017]	-	
s_2	0.10	**0.08**	[0.04 , 0.19]	-	-0.01	**-0.011**	[-0.014 , -0.006]	-	
σ	1.60	1.50	[0.95 , 2.45]	**1.74**	1.60	1.50	[1.25 , 1.80]	**1.52**	
β_0	5.00	4.25	[3.25 , 5.45]	**4.30**	5.00	5.10	[4.80 , 5.60]	**5.17**	
β_1	0.22	0.20	[0.08 , 0.36]	**0.21**	-0.22	-0.23	[-0.22 , -0.24]	**-0.23**	
β_2	0.08	0.08	[0.04 , 0.16]	**0.09**	-0.08	-0.08	[-0.07 , -0.09]	**-0.08**	
θ	40.0	**39.5**	[36.0 , 43.5]	-	40.0	**42.5**	[38.5 , 47.0]	-	**mean**
s_1	0.00	**-0.01**	[-0.02 , 0.01]	-	0.00	**-0.01**	[-0.02 , -0.00]	-	
s_2	0.00	**0.00**	[-0.01 , 0.01]	-	0.00	**-0.01**	[-0.01 , 0.00]	-	
σ	1.60	1.50	[1.20 , 1.90]	**1.58**	1.60	1.95	[1.55 , 2.50]	**2.03**	
β_0	5.00	4.85	[4.40 , 5.45]	**4.91**	5.00	5.15	[4.55 , 5.75]	**5.17**	
β_1	0.22	0.22	[0.19 , 0.25]	**0.23**	-0.22	-0.19	[-0.21 , -0.16]	**-0.19**	
β_2	0.08	0.08	[0.07 , 0.10]	**0.08**	-0.08	-0.10	[-0.11 , -0.08]	**-0.10**	
θ	40.0	**40.0**	[36.5 , 45.5]	-	40.0	**40.0**	[19.0 , 52.0]	-	**deviation**
s_1	0.20	**0.17**	[0.07 , 0.40]	-	-0.02	**-0.022**	[-0.043 , -0.012]	-	
s_2	0.10	**0.09**	[0.03 , 0.20]	-	-0.01	**-0.007**	[-0.012 , 0.001]	-	
σ	1.60	1.20	[0.75 , 1.95]	**1.40**	1.60	1.40	[1.15 , 1.70]	**1.40**	
β_0	5.00	4.85	[3.95 , 5.75]	**4.78**	5.00	5.00	[3.65 , 5.90]	**5.18**	
β_1	-0.08	-0.06	[-0.18 , 0.05]	**-0.06**	-0.08	-0.09	[-0.11 , -0.08]	**-0.09**	
β_2	0.08	0.08	[0.02 , 0.12]	**0.07**	0.08	0.07	[0.07 , 0.08]	**0.07**	

Bibliography

[1] W. Cramer, G. W. Yohe, M. Auffhammer, C. Huggel, U. Molau, M. A. F. S. Dias, A. Solow, D. A. Stone, and L. Tibig, "Detection and attribution of observed impacts," in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*, edited by C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, E. S. Kissel, A. N. Levy, S. MacCracken, P. R. Mastrandrea, and L. L. White (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014) pp. 979–1037.

[2] E. Jansen, J. Overpeck, K. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, D. Rind, O. Solomina, R. Villalba, and D. Zhang, "Palaeoclimate," in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, Q. D., M. M., Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007).

[3] A. C. Ravelo, D. H. Andreasen, M. Lyle, A. O. Lyle, and M. W. Wara, Nature **429**, 263 (2004).

[4] P. B. deMenocal, Earth Planet. Sci. Lett. **220**, 3 (2004).

[5] M. H. Trauth, J. C. Larrasoaña, and M. Mudelsee, Quat. Sci. Rev. **28**, 399 (2009).

[6] J. F. Donges, R. V. Donner, M. H. Trauth, N. Marwan, H.-J. Schellnhuber, and J. Kurths, Proc. Natl. Acad. Sci. U. S. A. **108**, 20422 (2011).

[7] R. Donner, S. Barbosa, J. Kurths, and N. Marwan, Eur. Phys. J.-Spec. Top. **174**, 1 (2009).

[8] P. K. Thornton, P. J. Ericksen, M. Herrero, and A. J. Challinor, Glob. Chang. Biol. **20**, 3313 (2014).

[9] J. Donges, R. Donner, K. Rehfeld, N. Marwan, M. Trauth, and J. Kurths, Nonlin. Processes Geophys. **18**, 545 (2011).

[10] L. R. Olsen, P. Chaudhuri, and F. Godtliebsen, Comput. Stat. Data. An. **52**, 3310 (2008).

[11] N. Itoh and N. Marwan, Nonlin. Processes Geophys. **20**, 467 (2013).

[12] R. Lund and J. Reeves, J. Climate **15**, 25472554 (2002).

[13] M. Mudelsee, Eur. Phys. J.-Spec. Top. **174**, 49 (2009).

[14] P. J. Green, Biometrika **82**, 711 (1995).

[15] E. Ruggieri and C. E. Lawrence, J. Comput. Graph. Stat. **23**, 87 (2014).

[16] J. O. Berger, J. Am. Stat. Assoc. **95**, 1269 (2000).

[17] L. M. Berliner, J. Geophys. Res.: Atmos. **108** (2003), 8776.

[18] B. Efron, Science **340**, 1177 (2013).

[19] L. M. Berliner, R. A. Levine, and D. J. Shea, J. Climate **13**, 3805 (2000).

[20] R. Schnur and K. Hasselmann, Climate Dyn. **24**, 45 (2005).

[21] F. W. Zwiers and H. Von Storch, Int. J. Climatol. **24**, 665 (2004).

[22] M. Blaauw and J. A. Christen, Bayesian Anal. **6**, 457 (2011).

[23] B. Goswami, J. Heitzig, K. Rehfeld, N. Marwan, A. Ambili, S. Prasad, and J. Kurths, Nonlin. Processes Geophys. Discuss. **1**, 1023 (2014).

[24] M. P. Tingley and P. Huybers, J. Climate **23**, 2759 (2009).

[25] D. S. E. Ahlonsou, Y. Ding, "Chapter 1 The Climate System: an Overview," in *Climate change 2001: The Scientific Basis*, edited by J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. Johnson (The Press Syndicate of the University of Cambridge, 2001) pp. 85 – 98.

[26] T. R. Karl and K. E. Trenberth, Science **302**, 1719 (2003).

[27] H. Von Storch and F. Zwiers, *Statistical analysis in Climate Research* (Cambridge University Press, Cambridge, 1999).

[28] W. F. Ruddiman, *Earth's Climate: past and future*, 2nd ed. (Macmillan, 2007).

[29] J. Peixoto and A. H. Oort, *Physics of climate* (Springer, New York, 1992).

[30] B. Saltzman, *Dynamical Paleoclimatology - A generalized theory of global climate change* (Academic Press, San Diegos, 2002).

[31] M. P. Tingley, P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam, Quat. Sci. Rev. **35**, 1 (2012).

[32] M. M. Rao, *Stochastic processes: Inference Theory*, 2nd ed. (Springer International Publishing, 2014).

[33] M. Mudelsee, *Climate time series analysis - Classical Statistical and Bootstrap Methods*, 2nd ed., Vol. 51 (Springer International Publishing, 2014).

[34] C.-D. Schönwiese, *Klimatologie* (UTB, 2013).

[35] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, J. Appl. Meteorol. Climatol. **46**, 900 (2007).

[36] Q. Lu and R. B. Lund, Can. J. Stat. **35**, 447 (2007).

[37] Q. Lu, R. Lund, and T. C. M. Lee, Ann. Appl. Stat. **4**, 299 (2010).

[38] M. Latif, *Klimawandel und Klimadynamik* (UTB Ulmer, 2009).

[39] P. B. deMenocal, Science **270**, 53 (1995).

[40] J. Emile-Geay and M. P. Tingley, Clim. Past Discuss. **11**, 2763 (2015).

[41] M. H. Trauth, Quat. Geochronol. **22**, 65 (2014).

[42] S. F. M. Breitenbach, K. Rehfeld, B. Goswami, J. U. L. Baldini, H. E. Ridley, D. J. Kennett, K. M. Prufer, V. V. Aquino, Y. Asmerom, V. J. Polyak, H. Cheng, J. Kurths, and N. Marwan, Clim. Past **8**, 1765 (2012).

[43] N. Malik, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **97**, 40009 (2012).

[44] T. C. Peterson, D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker, International Journal of Climatology **18**, 1493 (1998).

[45] C. Gallagher, R. Lund, and M. Robbins, J. Climate **26**, 4994 (2013).

[46] M. Mudelsee and M. E. Raymo, Paleoceanography **20**, PA4022 (2005).

[47] E. Ruggieri, T. Herbert, K. T. Lawrence, and C. E. Lawrence, Paleoceanography **24**, 87 (2009).

[48] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, Remote Sens. Environ. **114**, 2970 (2010).

[49] L. Perreault, J. Bernier, B. Bobée, and E. Parent, J. Hydrol. **235**, 221 (2000).

[50] P.-S. Chu and X. Zhao, *Journal of Climate*, J. Climate **17**, 4893 (2004).

[51] K. Gallagher, T. Bodin, M. Sambridge, D. Weiss, M. Kylander, and D. Large, Earth Planet. Sci. Lett. **311**, 182 (2011).

[52] C. Beaulieu, J. Chen, and J. L. Sarmiento, Philos. Trans. R. Soc. London, Ser. A **370**, 1228 (2012).

[53] S. Li and R. Lund, J. Climate **25**, 674 (2012).

[54] M. P. Girardin, A. A. Ali, C. Carcaillet, M. Mudelsee, I. Drobyshev, C. Hély, and Y. Bergeron, Glob. Chang. Biol. **15**, 2751 (2009).

[55] P. de Jong and J. Penzer, J. Am. Stat. Assoc. **93**, 796 (1998).

[56] V. N. Minin, K. S. Dorman, F. Fang, and M. A. Suchard, Bioinformatics **21**, 3034 (2005).

[57] Y. Ritov, A. Raz, and H. Bergman, J. Neurosci. Methods **122**, 25 (2002).

[58] J. S. Liu and C. E. Lawrence, Bioinformatics **15**, 38 (1999).

[59] P. Li and B.-H. Wang, Physica A **378**, 519 (2007).

[60] D. W. K. Andrews, Econometrica **61**, pp. 821 (1993).

[61] M. Basseville and N. Nikiforov, *The Detection of Abrupt Changes: Theory and Applications*, Vol. 104 (Prentice Hall Englewood Cliffs, 1993).

[62] E. Brodsky and B. Darkhovsky, *Nonparametric Methods in Change Point Problems*, Mathematics and its Applications (Springer Netherlands, 1993).

[63] M. Csörgö and L. Horváth, *Limit theorems in change-point analysis*, Wiley series in probability and statistics (Wiley, Chichester, 1997).

[64] J. Chen and A. K. Gupta, *Parametric statistical change point analysis: With applications to genetics, medicine, and finance* (Birkhäuser Boston, 2012).

[65] A. Aue and L. Horváth, J. Time Ser. Anal. **34**, 1 (2013).

[66] L. Y. Vostrikova, Theor. Probab. Appl. **26**, 362 (1981).

[67] L. Held and D. Sabanés Bové, *Applied statistical inference, Likelihood and Bayes* (Springer Berlin Heidelberg, 2014).

[68] A. Hannart and P. Naveau, Technometrics **54**, 256 (2012).

[69] E. T. Jaynes, *Probability theory: The logic of science* (Camebridge University Press, 2003).

[70] M. Bayes and M. Price, Philos. Trans. R. Soc. Lond. **53**, 370 (1763).

[71] W. M. Bolstad, *Introduction to Bayesian statistics* (John Wiley & Sons, 2007).

[72] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis* (CRC press, 2013).

[73] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. (Springer-Verlag New York, 2007).

[74] T. Z. Irony and N. D. Singpurwalla, J. Stat. Plan. Inf. **65**, 159 (1997).

[75] J. Berger, Bayesian Anal. **1**, 385 (2006).

[76] B. P. Carlin and T. A. Louis, *Bayesian Methods for Data Analysis* (CRC Press, 2009).

[77] R. Schlaifer and H. Raiffa, *Applied statistical decision theory* (Wiley-Interscience, 1961).

[78] H. Jeffreys, Proc. R. Soc. Lon. Ser.-A **186**, 453 (1946).

[79] J. M. Bernardo and A. F. Smith, *Bayesian theory*, Vol. 405 (John Wiley & Sons, 2009).

[80] R. E. Kass and A. E. Raftery, J. Am. Stat. Assoc. **90**, 773 (1995).

[81] H. Jeffreys, *Theory of probability* (Oxford University Press, Oxford UK, 1961).

[82] A. Shiryaev, Theor. Probab. Appl. **8**, 22 (1963).

[83] H. Chernoff and S. Zacks, Ann. Math. Statist. **35**, 999 (1964).

[84] J. V. Braun and H.-G. Müller, Statist. Sci. **13**, 142 (1998).

[85] J. V. Braun, R. Braun, and H.-G. Müller, Biometrika **87**, 301 (2000).

[86] A. E. Raftery and V. E. Akman, Biometrika **73**, 85 (1986).

[87] F. J. Girón, E. Moreno, and G. Casella, "Objective bayesian analysis of multiple changepoints for linear models," in *Bayesian statistics 8: Proceedings of the* 8^{th} *Valencia International Meeting, 2006*, Vol. 8, edited by J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West (Oxford University Press, USA, 2007) pp. 1–27.

[88] J. Hartigan, Commun. Stat. A-Theor. **19**, 2745 (1990).

[89] D. Barry and J. A. Hartigan, J. Am. Stat. Assoc. **88**, 309 (1993).

[90] P. Fearnhead, Stat. Comput. **16**, 203 (2006).

[91] B. P. Carlin, A. E. Gelfand, and A. F. M. Smith, J. Roy. Stat. Soc. C-App. **41**, 389 (1992).

[92] H. Lian, Commun. Stat. A-Theor. **38**, 419 (2008).

[93] J. O. Berger, E. Moreno, L. R. Pericchi, M. J. Bayarri, J. M. Bernardo, J. A. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, *et al.*, Test **3**, 5 (1994).

[94] E. Moreno, G. Casella, and A. Garcia-Ferrer, Stoch. Env. Res. Risk. A. **19**, 191 (2005).

[95] D. A. Stephens, J. Roy. Stat. Soc. C-App. **43**, 159 (1994).

[96] S. Chib, J. Econometrics **75**, 79 (1996).

[97] S. Chib, J. Econometrics **86**, 221 (1998).

[98] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. (Springer-Verlag New York, 2009) pp. pp.117–120 and pp.165–183.

[99] J. H. Friedman, Ann. Stat. **19**, 1 (1991).

[100] M. McCulloch, S. Searle, and J. Neuhaus, *Generalized, Linear, and Mixed Models*, 2nd ed. (Wiley, New York, 2008).

[101] G. K. Robinson, Statist. Sci. **6**, 15 (1991).

[102] P. C. Mahalanobis, Proc. Nat. Instit. Sci. India. **2**, 49 (1936).

[103] G. McLachlan, Resonance **4**, 20 (1999).

[104] J. O. Berger, *Statistical decision theory and Bayesian analysis*, 2nd ed. (Springer New York, 1985).

[105] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, J. Am. Stat. Assoc. **84**, 881 (1989).

[106] S. Kotz and S. Nadarajah, *Multivariate t-distributions and their applications* (Cambridge University Press, 2004).

[107] E. Limpert, W. A. Stahel, and M. Abbt, Bioscience **51**, 341 (2001).

[108] R. D. Cook, Technometrics **19**, 15 (1977).

[109] R. D. Cook, J. Am. Stat. Assoc. **74**, 169 (1979).

[110] S. S. Shapiro and M. B. Wilk, Biometrika **52**, 591 (1965).

[111] N. M. Razali and Y. B. Wah, Journal of Statistical Modeling and Analytics **2**, 21 (2011).

[112] G. W. Cobb, Biometrika **65**, 243 (1978).

[113] A.-E. S. Abd-Rabou and A. M. Gad, Journal of Data Science **5**, 379 (2007).

[114] A. B. Downey, ArXiv e-prints **0812.1237** (2008).

[115] E. Aguilar, I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa, World Meteorological Organization Techdoc. **1186** (2003).

[116] N. Schütz and M. Holschneider, Phys. Rev. E **84**, 021120 (2011).

[117] E. S. Vrba, "Ecological and adaptive changes associated with early hominid evolution," in *Ancestors: The Hard Evidence : Proceedings of the Symposium Held at the American Museum of Natural History April 6-10, 1984 to Mark the Opening of the Exhibition "Ancestors, Four Million Years of Humanity"* (Alan R. Liss, New York, 1985) pp. 63–71.

[118] E. S. Vrba, Am. J. Sci. **293**, 418 (1993).

[119] P. B. deMenocal and J. Tierney, Nature Education Knowledge **3** (2012).

[120] R. Tiedemann, M. Sarnthein, and N. J. Shackleton, Paleoceanography **9**, 619 (1994).

[121] P. deMenocal, J. Bloemendal, and J. King, Proceedings of the Ocean Drilling Program, Scientific Results **177**, 389 (1991).

[122] J. Larrasoaña, A. Roberts, E. Rohling, M. Winklhofer, and R. Wehausen, Climate Dyn. **21**, 689 (2003).

[123] M. Blaauw, Quat. Sci. Rev. **36**, 38 (2012).

[124] M. Blaauw, Quat. Geochronol. **5**, 512 (2010).

[125] A. Ojala, P. Francus, B. Zolitschka, M. Besonen, and S. Lamoureux, Quat. Sci. Rev. **43**, 45 (2012).

[126] W. Berger and R. Heath, J. Mar. Res. **26**, 134 (1968).

[127] P. Schiffelbein, **311**, 651 (1984).

[128] L. Löwemark and P. M. Grootes, Paleoceanography **19**, PA2001 (2004).

[129] M. H. Trauth, Comput. Geosci. **61**, 1 (2013).

[130] R. S. Bradley, *Paleoclimatology: reconstructing climates of the Quaternary*, 2nd ed., Vol. 68 (Academic Press, 1999).

[131] G. Fischer and G. Wefer, *Use of Proxies in Paleoceanography: Examples from the South Atlantic* (Springer-Verlag Berlin Heidelberg, 1999).

[132] M. A. Cane and P. Molnar, Nature **411**, 157 (2001).

[133] S. J. Feakins, P. B. deMenocal, and T. I. Eglinton, Geology **33**, 977 (2005).

[134] J. R. Marlow, C. B. Lange, G. Wefer, and A. Rosell-Mel, Science **290**, 2288 (2000).

[135] P. Molnar, Paleoceanography **23**, PA2201 (2008).

[136] M. Prange and M. Schulz, Geophys. Res. Lett. **31**, L17207 (2004).

[137] S. Steph, R. Tiedemann, M. Prange, J. Groeneveld, M. Schulz, A. Timmermann, D. Nürnberg, C. Rühlemann, C. Saukel, and G. H. Haug, Paleoceanography **25**, PA2202 (2010).

[138] L. M. Dupont, B. Donner, L. Vidal, E. M. Pérez, and G. Wefer, Geology **33**, 461 (2005).

[139] C. Karas, D. Nürnberg, A. K. Gupta, R. Tiedemann, K. Mohan, and T. Bickert, Nature Geosci. **2**, 434 (2009).

[140] C. Karas, D. Nürnberg, R. Tiedemann, and D. Garbe-Schönberg, Paleoceanography **26**, PA2217 (2011).

[141] A. Rosell-Melé, A. Martínez-Garcia, and E. L. McClymont, Earth Planet. Sci. Lett. **386**, 10 (2014).

[142] G. H. Haug, A. Ganopolski, D. M. Sigman, A. Rosell-Mele, G. E. Swann, R. Tiedemann, S. L. Jaccard, J. Bollmann, M. A. Maslin, M. J. Leng, *et al.*, Nature **433**, 821 (2005).

[143] E. Jansen, M. Raymo, P. Blum, *et al.*, eds., "Shipboard. scientific party, 1996. site 982," in *Proc. ODP, Init. Repts. 162: College Station, TX (Ocean Drilling Program)*, Vol. 162 (Texas A & M Univ., 1996) pp. 91–138.

[144] L. E. Lisiecki and M. E. Raymo, Paleoceanography **20**, PA1003 (2005).

[145] R. Bonnefille, R. Potts, F. Chalié, D. Jolly, and O. Peyron, Proc. Natl. Acad. Sci. U. S. A. **101**, 12125 (2004).

[146] R. Bonnefille, Global and Planetary Change **72**, 390 (2010).

[147] M. W. Wara, A. C. Ravelo, and M. L. Delaney, Science **309**, 758 (2005).

[148] A. V. Fedorov, P. S. Dekens, M. McCarthy, A. C. Ravelo, P. B. deMenocal, M. Barreiro, R. C. Pacanowski, and S. G. Philander, Science **312**, 1485 (2006).

[149] A. C. Ravelo, P. S. Dekens, and M. McCarthy, GSA TODAY **16**, 4 (2006).

[150] A. V. Fedorov, C. M. Brierley, K. T. Lawrence, Z. Liu, P. S. Dekens, and A. C. Ravelo, Nature **496**, 43 (2013).

[151] A. C. Ravelo, Nature Geosci. **3**, 672 (2010).

[152] J. Etourneau, R. Schneider, T. Blanz, and P. Martinez, Earth Planet. Sci. Lett. **297**, 103 (2010).

[153] C. M. Brierley and A. V. Fedorov, Paleoceanography **25**, PA2214 (2010).

[154] C. M. Brierley, A. V. Fedorov, Z. Liu, T. D. Herbert, K. T. Lawrence, and J. P. LaRiviere, Science **323**, 1714 (2009).

[155] A. V. Fedorov, C. M. Brierley, and K. Emanuel, Nature **463**, 1066 (2010).

[156] M. H. Trauth, M. A. Maslin, A. Deino, and M. R. Strecker, Science **309**, 2051 (2005).

[157] D. Barboni, Quat. Int. **322–323**, 264 (2014).

[158] P. U. Clark, D. Archer, D. Pollard, J. D. Blum, J. A. Rial, V. Brovkin, A. C. Mix, N. G. Pisias, and M. Roy, Quat. Sci. Rev. **25**, 3150 (2006).

[159] M. Medina-Elizalde and D. W. Lea, Science **310**, 1009 (2005).

[160] R. B. Owen, R. Potts, A. K. Behrensmeyer, and P. Ditchfield, Palaeogeogr. Palaeocl. **269**, 17 (2008).

[161] P. B. deMenocal, "The Ocean's Role in the Early Pleistocene Aridification of East Africa," (2012), aAAS Annual Meeting.

List of Publications

Publications

- **N. Berner**, M.H. Trauth, M. Holschneider, *Bayesian inference about Plio-Pleistocene climate transitions in Africa* (in preparation)

- **N. Schütz**, M. Holschneider, *Detection of trend changes in time series using Bayesian inference*, Physical Review E **84**, 021120 (2011)

Conference Contributions

- **N. Berner**, M.H. Trauth, M. Holschneider, *Deciphering transitions in climate time series using Bayesian inference* (talk), Frühjahrstagung der Deutschen Physikalischen Gesellschaft 2014, Berlin

- **N. Schütz**, M. Hayn, M.C. Brown, M. Holschneider, *Quantifying conjoint patterns of multiple paleo signals based on Bayesian inversion* (poster), European Geoscience Union 2012, Vienna, Austria

- **N. Schütz**, N. Makarava, M. Holschneider, *Deciphering scale-dependencies and trend changes in hydrological time series using Bayesian inversion* (talk), European Geoscience Union 2012, Vienna, Austria

- N. Makarava, **N. Schütz**, M. Holschneider, *Bayesian algorithm to quantify scale-dependencies in climatic time series* (poster), American Geoscience Union 2011, San Francisco, California

- **N. Schütz**, M.H. Trauth, M. Holschneider, *Deciphering trend changes in climate proxies using Bayesian inference* (poster), American Geoscience Union 2011, San Francisco, California

- **N. Schütz**, M.H. Trauth, M. Holschneider, *Detection of trend changes in the Plio-Pleistocene climate of Northern Africa using linear mixed modeling* (talk), European Geoscience Union 2011, Vienna, Austria

- **N. Schütz**, M.H. Trauth, M. Holschneider, *Detection of change points in palaeo-climate time series using Bayesian inference on linear mixed models* (poster), EXTREMES 2010: International Workshop in Recent Achievements on the Study of Extreme Events, Potsdam

- **N. Schütz**, M.H. Trauth, M. Holschneider, *Analysis of trend changes in Northern African palaeo-climate by using Bayesian inference* (talk), European Geoscience Union 2010, Vienna, Austria

Danksagung

Ich möchte die Gelegenheit nutzen, mich bei all denjenigen zu bedanken, die zum Gelingen meiner Arbeit beigetragen haben.
Allen voran bedanke ich mich bei Prof. Dr. Matthias Holschneider für die wissenschaftliche Zusammenarbeit, für sein Vertrauen in meine Fähigkeiten und seine andauernde Unterstützung.
Außerdem danke ich den Gutachtern dieser Arbeit, Prof. Dr. Jürgen Kurths und Prof. Dr. Charles-Antoine Guérin für die Zeit und die Aufmerksamkeit, mit der sie sich meiner Arbeit gewidmet haben. (Moreover, I greatly appreciate the time and effort the reviewers Prof. Dr. Jürgen Kurths and Prof. Dr. Charles-Antoine Guérin have spent on my thesis.)
Des weiteren möchte ich dem GRK 1364 danken, der es mir ermöglichte, eine transdisziplinäre Arbeit zu verfolgen. Insbesondere bedanke ich mich bei apl. Prof. Dr. Martin Trauth, der mich stets unterstützt und gefördert hat, und mir so ermöglichte, mich mit geowissenschaftliche Fragestellungen zu beschäftigen.
Ich möchte mich auch herzlich bei meinen (ehemaligen) Kollegen bedanken: Bei Dr. habil. Aneta Koseska, von der ich viel lernen durfte und die stets bemüht ist, alle meine Fragen zu beantworten, sei es über Bifurkationen oder über Bären. Von der wissenschaftlichen Zusammenarbeit mit Dr. Michael Hayn habe ich sehr profitiert, da er ein präziser und ausdauernder Erklärer ist. Vielen Dank für die Zeit, die Du in unsere spannenden Diskussionen gesteckt hast.
Und auch Jun.-Prof. Dr. Annett Junginger, die eine mindestens ebenso passonierte wie routinierte Erklärerin paläoklimatischer Zusammenhänge ist, verdanke ich viel. Gerne möchte ich auch Dr. Norbert Marwan für die fachlichen Anregungen und das ehrliche Feedback danken, dass ich in den Gesprächen mit ihm erhalten habe.
Ich danke PD Dr. Gert Zöller und Dr. Julien Bärenzug für alle hilfreichen Kommentare beim Korrekturlesen meiner Doktorarbeit. Tatkräftige Unterstützung hatte ich auch stets von Frau Sonja Neisse und Herrn Volker Gustavs genossen, die mir in vielen Belangen des Wissenschaftsalltags weitergeholfen haben.
Der wohl wichtigste Dank geht aber an meine Familie, die mich in jeder Minute meiner Doktorarbeit immer unterstützt haben und mir selbst in den schwierigen Momenten den entscheidenden Rückhalt gaben.