TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

STATISTICAL SIGNATURES FOR ADVERSE EVENTS

in molecular life sciences

Ville Laitinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. F OSA – TOM. 28 | TECHNICA – INFORMATICA | TURKU 2023

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

STATISTICAL SIGNATURES FOR ADVERSE EVENTS

in molecular life sciences

Ville Laitinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. F OSA – TOM. 28 | TECHNICA – INFORMATICA| TURKU 2023

University of Turku

Faculty of Technology
Department of Computing
Computer science
Doctoral Programme in Technology

Supervised by

Professor Leo Lahti
University of Turku

Professor Marko Mäkelä
University of Turku

Professor Jukka Heikkonen
University of Turku

Reviewed by

Associate Professor Steffen Leth Jørgensen
University of Bergen

Docent Tommi Vatanen
University of Helsinki

Opponent

Associate Professor Christopher Quince
University of Warwick

ISBN 978-951-29-9431-1 (PRINT)
ISBN 978-951-29-9432-8 (PDF)
ISSN 2736-9390 (PRINT)
ISSN 2736-9684 (ONLINE)
Painosalama, Turku, Finland, 2023

UNIVERSITY OF TURKU
Faculty of Technology
Department of Computing
Computer science
LAITINEN, VILLE: Statistical signatures for adverse events
Doctoral dissertation, 126 pp.
Doctoral Programme in Technology
October 2023

ABSTRACT

The ongoing evolution of computational sciences is helping to address the growing data analytical needs in applications. For instance, in biosciences, recent advances in measurement technologies have resulted in large amounts of data with domain-specific properties that are challenging to analyze with traditional statistical methods.

An example of such a domain is microbiomics, the study of microbial communities, which in humans, have been reported to be associated with health and diseases. Despite advances in the field, further research is needed, as there is still a lack of understanding of how microbiome data should be processed and of the universal ecological properties of these complex systems.

The objective of this thesis is to advance the field of microbiome data science by considering methods for predicting future outcomes based on current information. This is achieved through developing time series methods for complex systems and applying established statistical models in large population cohorts.

The thesis consists of two complementary parts. The first part consists of analyses of two prospective human gut microbiome data sets, and contains the first ever microbiome-based survival analysis. The second part is focused on the stability properties of dynamical systems. It shows that the Bayesian statistical framework can be used to improve accuracy in inferring stability features, such as systemic resilience and early warning signals for catastrophic state transitions.

The results of this thesis contribute to the best practices of human microbiome-related data science and demonstrate the advantages of the Bayesian framework in detecting adverse events in limited time series. Although the work was motivated by timely questions in microbiomics, the developed tools are generic and applicable in various contexts.

KEYWORDS: Early warning signals, time series analysis, probabilistic modeling, microbiome

TURUN YLIOPISTO
Teknillinen tiedekunta
Tietotekniikan laitos
Tietojenkäsittelytiede
LAITINEN, VILLE: Statistical signatures for adverse events
Väitöskirja, 126 s.
Teknologian tohtoriohjelma
Lokakuu 2023

TIIVISTELMÄ

Laskennallisten tieteiden jatkuva kehitys auttaa vastaamaan sovelluksissa ilmaantuviin uusiin kvantitatiivisiin tarpeisiin. Esimerkiksi biotieteissä mittausmenetelmien viimeaikaiset kehityssaskeleet ovat synnyttäneet huomattavia datamääriä, joiden ominaispiirteiden huomioiminen perinteisillä tilastollisilla menetelmillä on haastavaa.

Eräs tällainen tutkimusalue on mikrobiomiikka, jossa ihmisen kehossa ja ympärillä elävillä mikrobipopulaatioilla on todettu olevan yhteys terveyteen ja sairauksiin. Alan edistyksestä huolimatta tarvitaan lisää tutkimusta, koska on epäselvää, miten mikrobiomidataa tulisi käsitellä. Lisäksi mikrobiomien yleisiä ekologisia ominaisuuksia ymmärretään puutteellisesti.

Tämän väitöskirjan tavoite on edistää mikrobiomiin liittyviä kvantitatiivisia käytäntöjä tutkimalla menetelmiä, jotka ennustavat tulevia tapahtumia nykyisen tiedon valossa. Väitöskirjassa kehitetään aikasarjamenetelmiä kompleksisten systeemien tutkimukseen, ja sovitetaan vakiintuneita tilastollisia menetelmiä suuriin väestöaineistoihin.

Työ koostuu kahdesta toisiaan täydentävästä osasta, joista ensimmäisessä analysoidaan kahta prospektiivista suolistomikrobiomiaineistoa ja esitellään ensimmäinen mikrobiomidatan perusteella toteutettu elinaika-analyysin. Toinen osa keskittyy dynaamisten systeemien tasapaino-ominaisuuksiin. Tässä osassa näytämme, että systeemin palautuvuuden ja aikaisten varoitussignaalien mittaaminen onnistuu aiempaa tarkemmin Bayesiläistä viitekehystä hyödyntämällä.

Väitöskirjassa esitetyt menetelmät ja tulokset edistävät ihmisen mikrobiomiin liittyvän data-analytiikan parhaita käytäntöjä ja esittelevät Bayesiläisten menetelmien etuja tulevien tapahtumien ennakoimisessa puuttellisten aikasarjojen perusteella. Vaikka mikrobiomiikka on toiminut työn metodologisen kehitystyön motivaationa, ovat esitetyt menetelmät sovellettavissa myös muissa yhteyksissä.

ASIASANAT: Aikaiset varoitussignaalit, aikasarja-analyysi, probabilistinen mallinnus, mikrobiomi

Acknowledgements

I want to express my gratitude to my primary supervisor, Prof. Leo Lahti, for guiding me throughout this journey and offering me the chance to conduct research in a multidisciplinary setting. Your unwavering support, encouragement, and patience were crucial in completing this thesis.

I also wish to acknowledge my research director, Prof. Jukka Heikkonen, second supervisor, Prof. Marko Mäkelä, and thesis pre-examiners, Assoc. Prof. Steffen Leth Jørgensen and Doc. Tommi Mäkinen, for their valuable feedback on the manuscript.

I am grateful to the past and present members of the Turku Data Science and Algorithms group for their support and inspiration. Guilhem, Katariina, Matti, Aura, Pande, Moein, Chandler, Pyry, Chouaib, Shivang, Aaro, Wisam, and the shorter-term visitors, thank you for creating a warm and welcoming atmosphere. The journey toward obtaining a doctoral degree was sometimes hard, but you made the load lighter.

I also wish to acknowledge certain people I have crossed paths with during the past few years. I had the privilege of participating in the FINRISK and FinnBrain projects, where Prof. Teemu Niiranen and Dr. Anna Aatsinki were my primary collaborator, respectively. Thank you also to the FinnBrain cohort leaders, Prof. Hasse Karlsson and Assoc. Prof. Linnea Karlsson. My six-month visit at the lab of Prof. Jeroen Raes in KU Leuven, Belgium, was an unforgettable experience. I also want to thank Dr. Jorge F. Vázquez-Castellanos for his company during my stay in Leuven.

I extend my gratitude to those who assisted in maintaining a good work-life balance: my academic allies Mikko and Boris, my long-time friends Jarkko, Lauri, Juuso, Eero, Jukka, and the rest of the Paimio crew, my climbing buddies Joonas and Niko, guitar instructor Ilkka, and the Kaske's Hill Blues Band. Your company helped make acquiring a doctorate a manageable part of my life.

I am deeply grateful to my loved ones - my parents, Tuija and Antero, sisters Vilma and Tuuli, and grandmother Rauni, who always supported me with warmth, humor, and understanding. My mother-in-law and horticultural patron, Elina, allowed me to spend many enjoyable weekends gardening. Lastly, I want to express my love and gratitude to my beloved Anna and little Alma, who bring so much joy to my life.

September 4, 2023
Ville Laitinen

Table of Contents

Abbreviations

ABC	Approximate Bayesian computing
AIC	Akaike information criterion
CI	Confidence interval, credible interval
CLR	Centered log-ratio transformation
CSD	Critical slowing down
EWS	Early warning signal
EM	Euler-Maruyama approximation
FDR	False discovery rate
GLM	Generalized linear model
GP	Gaussian process
$GP(\boldsymbol{\mu}, \Sigma)$	A Gaussian process with mean-vector $\boldsymbol{\mu}$ and covariance function Σ
HHC	Hair cortisol concentration
HMC	Hamiltonian Monte Carlo algorithm
HMM	Hidden Markov model
HR	Cox regression hazard ratio
LV	Lotka-Volterra model
$MVN(\boldsymbol{\mu}, \Sigma)$	The multivariate normal distribution with mean-vector $\boldsymbol{\mu}$ and covariance function Σ
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings algorithm
MLE	Maximum likelihood estimate
MAP	Maximum a posteriori, mode of posterior distribution
$N(\mu, \sigma)$	The normal distribution with mean μ and variance σ
OOB	Out-of-bag error
OUP	Ornstein-Uhlenbeck process
PC	Principal component
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PPD	Postpartum Depression score
RBF	Radial basis function
TPR	True positive rate
VI	Variational inference

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Anna Aatsinki, Anniina Keskitalo, Ville Laitinen, Eveliina Munukka, Henna-Maria Uusitupa, Leo Lahti, Susanna Kortesluoma, Paula Mustonen, Ana João Rodrigues, Bárbara Coimbra, Pentti Huovinen, Hasse Karlsson, Linnea Karlsson. Maternal prenatal psychological distress and hair cortisol levels associate with infant fecal microbiota composition at 2.5 months of age. *Psychoneuroendocrinology*, 2020, 119: 104754.

II Aaro Salosensaari*, Ville Laitinen*, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfthan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti*, Teemu Niiranen*. Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nature Communication*, 2021, 12: 2671.

III Ville Laitinen, Leo Lahti. A Hierarchical Ornstein-Uhlenbeck Model for Stochastic Time Series Analysis. In: Duivesteijn, W., Siebes, A., Ukkonen, A. (eds) *Advances in Intelligent Data Analysis XVII. IDA 2018. Lecture notes in Computer Science*, 2018, vol 11191: 188 - 209.

IV Ville Laitinen, Vasilis Dakos, Leo Lahti. Probabilistic early warning signals. *Ecology and Evolution*, 2021, 11: 14101– 14114.

V Ville Laitinen, Leo Lahti. Probabilistic Multivariate Early Warning Signals. In: Petre, I., Păun, A. (eds) *Computational Methods in Systems Biology, 20th International Conference, CMSB2022, Lecture Notes in Computer Science*, 2022, vol 13447: 259 - 274.

The original publications have been reproduced with the permission of the copyright holders.

*Equal contribution

1 Introduction

The data analysis of complex systems is a significant branch of modern applied statistics. An important application domain of this discipline is life sciences, where, over the past decade, extensive research efforts have been focused on microbiomes - the communities of microbes inhabiting the environment and bodies of animals [1; 2]. The interest in microbiomes was ignited by advances in measuring technologies and bioinformatics that enabled sequencing of the genetic material in a given target and was further fueled by reports suggesting remarkable connections with the host's health. Most notably in humans, the microbiomes in our various body sites have been shown to have close ties to the health of the host [2]. Connections to various illnesses and conditions imply a massive potential for a more holistic understanding of human health and targeted clinical therapies.

At the same time, increasing computational resources are available for handling the collected data and answering the arising research questions with powerful algorithmic approaches [3; 4]. The world, however, is not ready. The new types of data generated by high-throughput technologies have characteristics that the standard statistical tools cannot adequately address [5]. In microbiome research, a limited sample size is the norm since collecting comprehensive data sets is costly and complicated by ethical hurdles. Studying microbiomes is further complicated since the underlying mechanistic processes are poorly understood. Moreover, the generated data is plagued with high levels of technical and biological variation, low signal-to-noise ratio, large dimensionality, zero-inflation, over-dispersion, and compositionality [5]. Considering such factors and ensuring that the relevant information is extracted as best as possible while ignoring noise requires customized data processing and modeling solutions. To realize the field's full potential, customized data processing and modeling solutions that can transform the raw data into useful information optimally are required [5; 6].

This thesis aims to contribute to the rapidly growing field of microbiome data science. While the majority of the microbiome experiments thus far have focused on cross-sectional population-level associations with health-related outcomes, the predictive power of microbiome features in prospective settings is largely unknown due to the scarcity of data sets with comprehensive follow-ups [7; 8]. Moreover, it has become evident that the dynamical properties of microbial communities are clinically relevant. For example, stability and variability of the microbiome are features

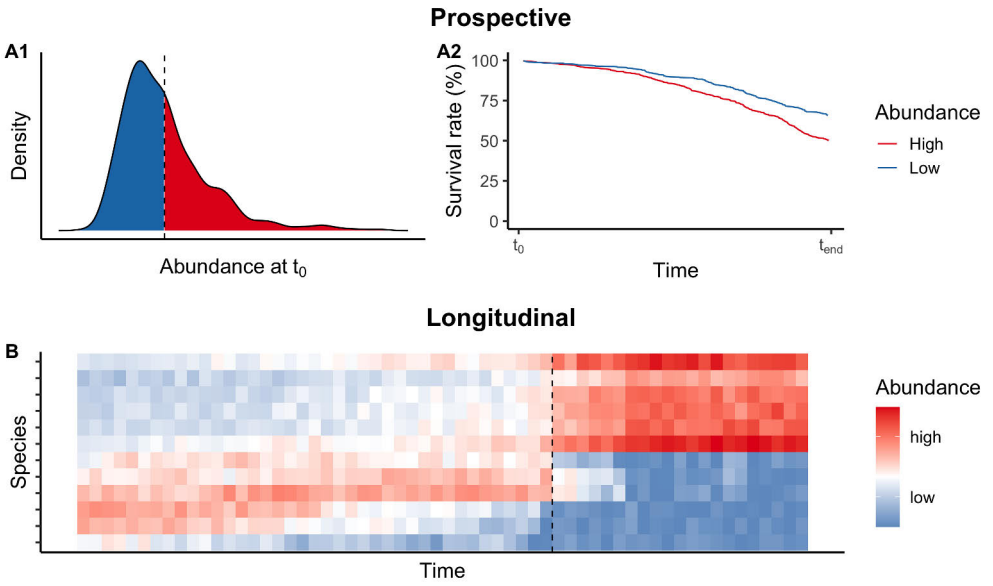

Figure 1. Depiction of the themes of the thesis using simulated microbiome data. The top and bottom rows illustrate the prospective and longitudinal perspectives, respectively. In panel **A1**, the study population is divided into two groups based on the abundance of a bacterial species, labeled as low (blue) and high (red) abundance groups. During the follow-up period, these groups exhibited different mortality rates, as shown in panel **A2**. Panel **B** displays the time series of the abundance of various taxa in a single subject. Notably, at approximately the vertical dashed line, the community undergoes a transition to an alternative state with no apparent warning signs beforehand.

that have been reported to reflect the health status of the host [9; 10]. As the number, length, and resolution of available prospective and time series data increases, data analysts need to be prepared with appropriate statistical tools that can adequately handle the temporal dimension.

The goal of this thesis is to study, apply and develop methods that can help address this gap in the toolkit of data scientists. More specifically, the focus is on predicting future events in prospective and longitudinal study designs (see Fig. 1 for illustration). The research objectives can be summarized in the following questions:

Q1: To what extent are the existing statistical tools suitable for analyzing prospective microbiome data, and what sort of model modifications and data preprocessing steps can improve the applicability?

Q2: Can the Bayesian statistical framework be used to improve the inference of characteristics of dynamical systems based on limited time series?

Publications I and II examine the first research question, and were made in collaboration with national cohort studies: FinnBrain [11] and FINRISK [12]. These studies analyze two prospective cohorts that include measurements of the human

gut microbiome and background variables. In Publication I, the relationship between stress levels of expectant mothers and gut microbiome composition of their infants at 2.5 months is explored. Publication II provides the first-ever instance of survival analysis using gut microbiome data in a large population cohort with all-cause mortality as the endpoint. This pioneering study has significant implications for understanding the relationship between the gut microbiome and health outcomes and suggests that the gut microbiome can be used as a general biomarker for overall health status. In terms of this thesis, the relevance lies in the methodology of these analyses. Both studies demonstrate the importance of customized modeling and data analysis techniques that take into account the characteristics of microbiome data. The temporal dimension is present as the gut microbiome profiles are separated from background information by a follow-up time. These publications complement the time series analysis found in the subsequent publications.

Publications III - V delve into the second research question and examine ways of enhancing the inference of the stability properties of dynamical systems. The presented methodological advances leverage the Bayesian modeling framework and focus on measuring stability, and the loss thereof, from time series data. The motivation for these works was the common challenge in real-world applications, especially in microbiomics - limited sample size.

Publication III introduces a hierarchical variant of a stochastic time series model that demonstrates improved accuracy in measuring stability in parallel data sets, compared to non-hierarchical models.

In publications IV and V, novel approaches for detecting early warning signals (EWS) for catastrophic transitions is presented. These data-driven methods are designed to detect ”critical slowing down,” a phenomenon some dynamic systems display as they approach a tipping point and possibly transition to an alternative stable state [13]. One of the benefits of these methods is their generic nature. They are agnostic to the application domain and make few assumptions about the data, which makes them versatile and applicable in a variety of settings. However, they are known to lack robustness at low sample sizes and other common data limitation [14]. Regardless, EWS have been detected in various natural and social systems [15]. However, it remains an open question if microbiomes display such signals before transitions between states of alternative abundance levels [16]. A more sensitive and robust methodology could help settle this question sooner. The publications approach the matter by formulating certain EWS indicators in the Bayesian framework, in contrast to previous EWS studies, which have relied solely on the frequentist framework, and show that the Bayesian formulation provides increased robustness and a more principled means of handling the model uncertainty.

The thesis structure is outlined as follows. The motivational subject of the thesis, microbiome, is presented in Chapter 2. The statistical methodology used in the publications is the focus of Chapter 3, and Chapter 4 provides an account of the stability

properties of dynamical systems. The theory and methods presented are mostly limited to what is necessary for understanding the publications, with references being provided to textbooks and articles where the reader can access more comprehensive information. The publications included in the thesis are summarized and discussed in Chapter 5, and the thesis is concluded in Chapter 6.

2 Microbiome

In this chapter, we first give a general-level overview of the microbiome, followed by a description of its statistical properties and significance to human health.

2.1 Basic concepts in microbiomics

The human microbiome refers to the aggregate of microorganisms residing in our tissues and biofluids of various body sites [1; 17]. The number of bacterial cells inhabiting a human is estimated to be at least of the same order of magnitude as the number of cells of the host, with microbiome-related genes significantly outnumbering those of the host [17]. While most microbiome studies to date have focused on bacteria, the definition of the microbiome encompasses archaea, fungi, and viruses as well.

Unlike the genetic makeups of humans, the microbiomes colonizing our bodies are remarkably diverse, and the taxonomic composition in different body sites and between individuals can be vary greatly [18]. It is estimated that some thousands of species of bacteria inhabit the human body at any given time, and the species overlap between individuals can be negligible [19]. Additionally, the taxonomic composition can change significantly even over short periods, while simultaneously displaying a degree of stability over long-term periods [18].

In recent years, it has become apparent that there is a close connection between the human microbiome and overall health [2; 20]. A multitude of conditions, such as autoimmune diseases [21], depression [22], cancer [23], and inflammatory bowel disease [24], have been found to correlate with the microbiome. However, the specific taxonomic features that can be regarded as pathological or health-promoting are largely unknown, and it is unclear to what extent the microbiome has a causal role in regulating health [1]. Causality is difficult to study, especially in human studies due to ethical and practical hurdles. The field is, however, starting to shed light on causal links, and animal experiments have provided evidence for the gut microbiome's regulating role in some conditions like obesity, for instance [25].

Initially, microbiome research was primarily focused on cross-sectional single-time-point data, which remains the case today, despite the known significance of the temporal dimension. As the field is still relatively young, prospective and longitudinal data sets with extensive follow-ups and substantial participants and samples

sizes are still scarce, although such data is becoming increasingly available. In one of the earliest prospective human microbiome studies [7], a link between the gut microbiome and the later onset of type I diabetes was discovered. Recently, more prospective and longitudinal studies have been published that explore the association between microbiome features with clinical variables later in time or vice versa. For instance, the microbiome has been linked to vaccine response in infants [26], post-acute COVID-19 syndrome [27], and gastric cancer [28] at distinct time points. Additionally, the dynamic properties of microbiomes inferred from time series have been shown to have clinical relevance. For example, a higher temporal variability has been found to be associated with inflammatory bowel disease [9].

While most biomedical microbiome research has concentrated on the relationship between the taxonomic composition and clinical outcomes, the abundance profile of a microbiome does not fully reflect the microbes' functional roles [29]. Closely related species may have vastly different physiological importance and, on the other hand, distantly related species can have similar roles, for example, in processing food molecules. This concept of functional redundancy highlights that, despite significant taxonomic variation, the functional differences may not be as pronounced.

While the clinical relevance of the human microbiome has become established, the specific mechanisms that govern the variation within and between individuals, as well as their effects on the host remain largely unknown [1]. The interplay between commensal, pathogenic, and symbiotic microbes, along with the host's physiology and their reaction to changes in the living environment and lifestyle, complicate our understanding of the microbiome [30]. The exact biochemical processes and the emergence of taxonomic composition and other properties from them are not fully understood [30], and the immense complexity, nonlinear effects, and personalized responses make the microbiome a challenging target to study.

2.2 Statistical properties of microbiome data

Despite the aforementioned challenges, factors affecting the microbiome and certain consistent statistical patterns have been revealed.

Most personal and population-level variation can be explained by environmental factors, while host genetics play only a minor role [31]. Factors such as diet [32], medication [2], lifestyle factors (such as having pets or traveling) [33; 34], mode of delivery and breastfeeding in early life [35], immunology [36], and stress [37] can all significantly affect the composition of microbiomes. Additionally, ecological factors like migration, evolution, and competition between microbes can also play a role [38].

The reaction to these factors can be gradual but nonlinear effects can also give rise to phenomena such as alternative stable states [39] that have secondary properties such as resilience [30]. In a pathological condition known as dysbiosis, the

microbiome remains persistently stuck in an unfavorable configuration that may be difficult to reverse [40]. The resilience of the dysbiotic state underscores the clinical significance of considering the microbiome from this perspective.

At the population level, the relative abundance densities of microbial groups at various levels of taxonomic resolution tend to exhibit characteristic distribution shapes [16] (see Figure 2). These abundances (on a log-scale) can be characterized as skewed, fat-tailed, or bimodal, while some bacteria are altogether absent in a proportion of the population causing a noticeable density mass at zero. At the community level, microbiome samples tend to cluster into regions of the species space. These clusters have been referred to as *enterotypes* [41; 42], although the discreteness of these groups has been a subject of debate [43].

Over time, microbiomes can vary significantly. Samples from a given target taken just a few days apart can show a vast difference in taxonomic composition. However, over more extended periods, the human adult microbiome is known to be relatively robust and tends to maintain its taxonomic configuration [8; 10; 44]. Despite this, a deeper understanding of more specific general properties of microbiome dynamics is lacking, although certain aspects have been uncovered. For example, it has been reported that gut microbiomes seem to follow universal dynamics that are determined by interactions between microbe species, whereas microbial communities of certain skin sites are more influenced by the environment [45]. In the future, gaining a better understanding, for instance, of the systemic stability properties [44] could aid in planning clinical interventions and recovering from therapies such as antibiotic treatments.

In terms of statistical analysis and modeling, microbiome data presents a considerable challenge, and a plethora of statistical and machine learning approaches have been explored to tackle this challenge [5; 6; 46; 47]. When selecting appropriate quantitative methods for microbiome analysis, the specific properties that require attention include: non-Gaussian distribution shapes, an abundance of zeros (both technical and biological), high dimensionality compared to sample size, heteroscedastic variations, and a low signal-to-noise ratio. Microbiome data is also count type, which means that observations from the target community consist of individual observations of individual bacteria. Often this count data is transformed into compositional type that represents the *proportions* of different taxonomic ground in the sample. Analysing compositional data as opposed to absolute abundances poses a challenge as the proportion of a single taxon is influenced by alterations in the proportions of other taxa [48]. Much of the literature on microbiome-related data analysis has focused on addressing these features by modifying established methodology [49] with, for example, latent variable models [50], probabilistic inference of network structure [51], or by data transformations [52].

In general, statistical and machine learning methods used in microbiome analysis can be categorized into five main classes [5]. *Dimensionality reduction methods*

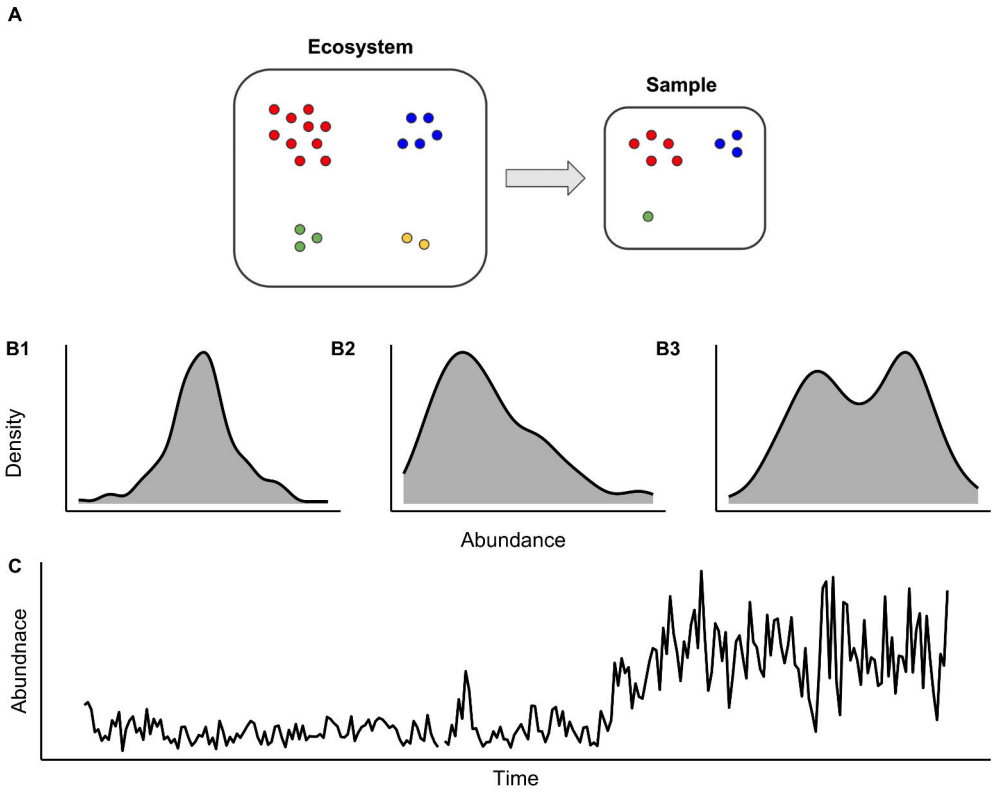

Figure 2. Illustration of selected statistical properties of the microbiome in simulated data. **A** Microbiome data is count data obtained from sampling the target ecosystem. The balls represent individual microbes, and the colors correspond to various species. Sampling preserves approximate proportions of the species but low abundance species have higher measurement error. The yellow species is present in the target community but not in the sample. On a population level, microbe abundances typically have fat-tailed (**B1**), skewed (**B2**), or bimodal (**B3**) distributions [16]. **C** The variance of individual species depends on their abundance (heteroscedasticity). Generally, higher abundance levels imply greater temporal variation.

such as PCA (principal component analysis) or PCoA (principal Coordinates Analysis) [53] are typically used to collapse a large number of dimensions into two or three that enable visualizing and exploring the population and community-wide characteristics. These methods can also be convenient for understanding dynamics. For example, switches between alternative stable states [54] and overall community variation [9] have been studied with dimensionality reduction methods. *Clustering methods* are used in annotating gene reads to microbe species using gene libraries and in identifying community types in collections of samples [41; 55]. *Classification methods* such as the random forest [56] are used to compare microbiomes of stratified populations in clinical studies. *Deep learning* can be used in a wide range of different tasks [57], although the black-box nature of these methods restricts recovering mechanistic

insights. Finally, *differential abundance analysis* methods such as generalized linear models [58] associate microbiome features with clinical variables. Several methods for differential abundance testing have been developed, which address the characteristics of microbiome data with different strategies. These methods can, however, give contradictory results [59] which implies that further development is required.

In addition, as more extended time series and follow-up data are being collected, methods designed specifically for time series and prospective data analysis are becoming increasingly relevant [60]. Ordinary and stochastic differential equations are a standard tool for time series modeling and, for example, the generalized Lotka-Volterra model has been used to infer interaction coefficients between species and to explain the emergence of alternative stable states [51; 54]. Non-parametric methods such as Gaussian processes are an alternative approach that have proven to be useful in microbiome time-series analysis [50].

In conclusion, despite being a relatively new area of research, microbiomics holds great promise as a component of a more holistic understanding of human disease and health. In order for the field to reach its full potential, quantitative methods that can address the microbiome-specific data characteristics need to be developed.

3 Methodological foundations

This chapter presents the methodological foundation of the quantitative methods that played a significant role in the thesis. We start by introducing the two main statistical paradigms, the frequentist and the Bayesian frameworks. An emphasis is given to the latter since it is the one most of the methodological developments of the thesis utilize. Then, we provide a cursory overview of model fitting with Markov chain Monte Carlo methods, as they are of great importance in practice. Then, we delve into more specific data analysis methods, starting with models for cross-sectional and prospective data. In the last subsection, we present stochastic models for time series analysis.

3.1 Applied statistics

Statistics is the discipline of gathering, analyzing, and presenting data [61]. A data set consists of samples, which in experimental studies represent measurements of specific characteristics of a population set. Applying statistics in practical data analysis typically involves analyzing data with descriptive statistics, such as the mean and higher moments, or with some statistical model, such as linear regression.

Inferential statistics is concerned with the latter, where samples are seen as realizations of a random process described by a model, M. The main objective is to fit this model to the available data, which means learning the model parameters that best describe the data. These parameters can then be used to make inferences about the population, such as differences in an outcome variable between groups or the impact of a continuous variable on another. The validity of any conclusions about the data and parameters is naturally conditional on the appropriateness of the chosen model and the quality and amount of available data [3]. Therefore, quantifying the level of statistical certainty in the learned parameter estimates is an integral part of the process.

The process of fitting models and handling and interpreting of the results can be approached in different ways. The two main approaches, both of which are used in this thesis, are the frequentist and the Bayesian probability interpretations [62]. Each framework interprets probability differently and has distinct implications for statistical learning. Next, we will explore these two interpretations and how they approach statistical modeling.

3.1.1 Frequentist interpretation

The frequentist interpretation of probability, as the name suggests, focuses on the frequency of events in a data set [61]. This interpretation defines the probability of an event as the hypothetical proportion of repeated idealized experiments in which the event occurs. In other words, the idea is that any experiment can be regarded as one in an infinite sequence of independent repetitions of the same experiment.

In the context of frequentist inference, model fitting amounts to maximizing the likelihood function $p(X|\theta)$ corresponding to the chosen model M, in terms of the model parameters θ. Likelihood is a real-valued function that quantifies the probability of observing the data X given the model and its parameters, and is used as the basis for model fitting. The values of θ mapping to higher likelihood values are considered to provide a better description of the data, and the value of θ that maximizes the likelihood function is referred to as the maximum likelihood estimate (MLE). The MLE is generally the primary target of frequentist model inference.

However, it is important to note that the MLE is a point estimate, meaning that it provides only a single value as a result of the inference process, without any information regarding the level of certainty of the estimate. To address this issue, the MLE is typically accompanied by a confidence interval (CI) and a p-value, which are used to quantify the reliability of the estimate [61]. These statistics are based on the philosophy of frequentism and play a critical role in providing a comprehensive understanding of the results [63].

The confidence interval (CI) is a range of values for the parameter θ that is calculated at a specified confidence level, often 95%. This estimate is based on the assumption that if the experiment was repeated multiple times, the computed 95% CI would contain the true parameter value 95% of the time. In other words, there is a 95% probability that the true value of the parameter falls within the 95% CI of a repeated experiment. There are different methods for determining the (approximate) confidence interval, including those based on the likelihood function or bootstrapping. However, the interpretation of the CI has often been found to be confusing and misunderstood [63]. One common misinterpretation, for instance, is that 95% of the estimates in future studies will fall within a computed CI.

The p-value is another critical component of frequentist inference and serves as a measure of compatibility between the data and the model used to generate the results [61]. It is defined as the probability of obtaining results that are as extreme or more extreme than the observed results, assuming that the null hypothesis is true. The null hypothesis, H_0, typically represents a model parameter or other quantity (such as the difference between group means) being equal to zero, indicating that any deviation from H_0 has arisen due to chance alone. A pre-set level, often set at $\alpha = 0.05$, is used as a threshold for statistical significance; if the p-value is less than α in an experiment, it implies that the observed results are incompatible with H_0. In such a

case, the result is considered to be statistically significant and a positive finding.

However, similar to the confidence interval, the p-value can be a source of confusion and even deliberate misuse. For example, obtaining a p-value that is less than α does not imply that H_0 is false with a probability of p. Another common misconception is that $p > \alpha$ is evidence for the absence of an effect [63]. Additionally, the practice of "p-hacking," where researchers manipulate the experimental design until they obtain results that give $p < \alpha$, is relatively common in research articles [64].

3.1.2 Bayesian interpretation

This subsection mostly follows the presentation in [65].

The Bayesian statistical paradigm represents a departure from the frequentist interpretation of probability. Rather than viewing probability as the relative frequency of events in hypothetical repeated trials, Bayesian statistics views probability as a degree of belief in an event, based on all available information. Additionally, in Bayesian statistics, model parameters are viewed as random variables, representing uncertain values, rather than fixed but unknown values as in the frequentist interpretation.

The foundation of the Bayesian framework is the Bayes' theorem formulated as

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}. \tag{1}$$

The formula relates the probability distribution of the model parameters θ conditional on the data X, the *posterior distribution* $p(\theta|X)$, as the product of the likelihood function $p(X|\theta)$ and the *prior distribution* $p(\theta)$ divided by evidence for the data $p(X)$. The posterior distribution is the combination of prior knowledge and the information in the data, and expresses the belief in different parameter values as probabilities.

The prior encodes probability information about the parameters before taking new data X into account, and can be based on previous experiments or on subjective beliefs about the studied phenomenon. This belief is then updated with X in a way defined by the likelihood function. Typically, $p(X)$ is difficult or practically impossible to compute, and analytical solutions for $p(\theta|X)$ are available only for special cases. However, the proportional form

$$p(\theta|X) \propto p(X|\theta)p(\theta) \tag{2}$$

can be used even when no analytical solution is available. Dropping the evidence function $p(X)$ can be addressed by normalizing the product of the prior and the likelihood. Normalization is not required when using posterior samples obtained from Markov chain Monte Carlo methods, which is the common practice in actual applications.

The posterior can be summarized using various summary statistics such as mean, mode, median, and variance that give information about the location and spread of posterior mass. *Credible intervals* are a way to quantify posterior uncertainty and represent the range where an unknown parameter falls with a specified certainty. Credible intervals can be defined in different ways. *The central interval* is formed using posterior quantiles, so for example, a 95% credible interval is defined by the 2.5% and 97.5% quantiles. *The highest density region*, on the other hand, is defined as the smallest set containing 95% of the posterior mass. The distribution of the posterior mass can conversely be used to test hypotheses about the parameters. For instance, the probability that a parameter lies between 0 and 1 can be calculated as the posterior mass over that set, $p(0 < \theta < 1|X)$.

Whereas the likelihood function is directly determined by the chosen statistical model, the prior distribution is chosen based on subjective beliefs, unless recovered from a previous experiment. Prior distributions can be categorized based on the level of information they possess. An *informative prior* reflects specific, precise knowledge of the event leading to strict restrictions on posterior values. A *weakly informative prior* expresses general information about a variable and guides the inference away from implausible values. An *uninformative prior* holds only vague information about the parameters and may have mass at values clearly unrealistic for the application.

For example, if one were to estimate the average temperature for the month of July, an informative prior might be the normal distribution based on the mean and variance of June temperatures from previous years, while a weakly informative prior might be the mean and variance of all temperature measurements on record. An uninformative prior, on the other hand, might be a normal distribution with mean 0 and variance 100.

This example illustrates the subjectivity of prior selection, although, in general, full objectivity is unattainable in data analysis. Data collection, modeling, presentation of results, and conclusions made thereof are all subjective to some extent. However, when there is sufficient data, the impact of the prior on the posterior is minimal but in cases where data is limited or noisy, the prior can have a major impact on the inference. In such cases, choosing a well-justified prior becomes a critical task. Prior selection plays a crucial role in Publications IV and V, as they focus on time series with low sample sizes.

The parameters of the priors, such as the mean and variance for normal prior, are referred to as *hyperparameters*. These hyperparameters can also be treated as random variables with their own priors, known as *hyperpriors*. This type of model is called a *hierarchical model* and is an appropriate choice for situations where subgroups of a population are comparable or connected by the structure of the problem. For instance, if different experiments are conducted to determine the effectiveness of a drug, the results of these experiments may vary due to factors such as sample size

or random variation in the test groups. Such experiments would evidently be related and a hierarchical model would consider the experiment-specific parameters, θ_i, as sampled from a common prior distribution with unknown hyperparameters. These hyperparameters can then be learned during the inference process, providing insight into the population-level variation of the drug's efficacy, in addition to the individual experiments.

In microbiome modeling, hierarchical models could be an appropriate choice due to reported universal dynamical characteristics between microbiomes of different individuals [45]. This was the motivating premise for Publication III where we employed hierarchical time series method to modeling time series of limited sample size.

3.1.3 Inference in probabilistic models

Formulating a statistical model can be relatively straight-forward, and optimization algorithms make it easy to compute the MLE or the posterior mode. In some cases, the posterior can be computed analytically but, in most cases, accessing the full or marginal posterior distributions can be a considerably more challenging task. For simple models with only a few parameters, it is possible to approximate the posterior by computing it point-wise in a grid of parameter values and normalizing the resulting function. But as the number of model parameters increases, this approach becomes impractical because the number of grid points grows exponentially.

Other approximate solutions include approximate Bayesian computation (ABC) [66] and variational inference (VI) [67]. ABC is employed in cases where the likelihood is computationally infeasible, or even impossible to evaluate. It is based on simulating data from an approximation of the likelihood using various samples form the prior distribution. The simulated data are then compared with the actual data and any simulations that are too dissimilar (as determined by some metrics and tolerance level) are discarded. The values corresponding to the remaining simulations make up the approximation of the posterior.

In VI, on the other hand, the posterior is approximated with a variational distribution that belongs to a simpler and more manageable family, such as the Gaussian distributions. The dissimilarity between the posterior and variational distributions is minimized in terms of the model parameters, with the Kullback-Leibner divergence being the most common dissimilarity metric.

Both ABC and VI can be challenging to use, as they require making distributional assumptions and mathematical derivations. In practice, the posterior distribution is usually inferred by drawing samples from it. However, sampling from a high-dimensional posterior is a complex task, as the curse of dimensionality causes the posterior mass to concentrate in increasingly small areas of the parameter space [68]. In the following, we focus on Markov chain Monte Carlo algorithms, which are

the most commonly used method for accessing posterior distributions, and that was extensively utilized in Publications III-V.

Markov chain Monte Carlo

Markov chain Monte Carlo algorithms (MCMC) are a class of methods that can approximate probability distributions and are the standard solution for drawing posterior samples efficiently [65; 68]. The methods are based on constructing a Markov chain, a sequence in which each value depends on the previous value, whose stationary distribution is the targeted distribution, in our context, the posterior distribution. Each sequence element is a sample from the posterior and as the chain advances, it converges towards the areas of the posterior with the highest density. The simulation starts by specifying some, often random, initial value θ^0, and then drawing the subsequent values iteratively from a *transition distribution* $T_t(\theta^t|\theta^{t-1})$, which is often dependent on the iteration number t.

Efficiency

Although convergence to the target distribution is guaranteed in theory, a poorly specified MCMC may work inefficiently from a practical perspective if the samples generated from the transition distribution do not explore the parameter space efficiently.

In order to improve efficiency of the parameter space exploration, an ensemble of separate chains is typically run. Using long enough chains with dispersed initial values helps with efficiency and discarding early iteration as warm-up ensures that chains started at low posterior density areas do not bias the estimate.

However, MCMC convergence still needs to be monitored. High sample autocorrelation and discrepancies between the chains are common signs of inefficient sampling and not reaching convergence, respectively. A standard convergence metric is the $\hat{R}$ which compares sample variances between and within chains [69] and measures agreement between different chains and whether convergence has been reached. Ideally, the value should be close to 1 and values above 1.1 have been regarded as a sign of poor convergence [69]. Effective sample size measures the chains' autocorrelation and large values after the post warm-up iterations signal inefficient sampling. The autocorrelation can be reduced by *thinning* the samples, which refers to discarding all but every kth sample [70]. However, thinning considerably reduces the algorithm's efficiency as a large number of samples are dismissed.

Using alternative parameterizations is another technique for improving posterior sampling efficiency. For example, in the normal model, using precision instead of variance can lead to more accurate results in some situations. The prior distributions can also play a crucial role in the exploration of the parameter space. By providing

more information, informative priors can limit the areas that are explored, leading to more efficient sampling.

Algorithms

The Metropolis-Hastings (MH) algorithm [71] forms a subclass of MCMC methods that use an acceptance/rejection rule for propositions from the transition distribution that ensure convergence to the target distribution. The algorithm is stated as follows:

1. Set an initial value θ^0, such that $p(\theta^0|X) > 0$. These can be drawn from some *starting distribution* or set manually.

2. Sample a proposal θ^* from a transition distribution $T_t(\theta^*|\theta^{t-1})$.

3. Compute $r = \frac{p(\theta^*|y)/T_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/T_t(\theta^{t-1}|\theta^*)}$.

4. Set $\theta^t = \begin{cases} \theta^*, & \text{with probability } \min(r, 1) \\ \theta^{t-1}, & \text{otherwise.} \end{cases}$

A common choice for the transition distribution in MH is a multivariate normal centered on the current value, which results in a random walk exploration of the parameter space [65]. However, specifying the variance of the transitions can be challenging and poor choices may lead to poor algorithm performance, even in low-dimensional problems [68]. Too short jumps lead to a slow exploration of the parameter space, whereas with oversized jumps the transitions are rejected too often, resulting in the sampler standing still for much of the time. With multidimensional posteriors, it is possible that no fixed transition variance will give satisfactory results. This is because different regions within the distribution would need transition variances of varying sizes to ensure efficient exploration [68].

The Gibbs sampler [72] is a popular, and a relatively simple example of an MCMC algorithm. It is a variant of the MH and is based on sampling the components of θ separately, conditional on the other components. Each iteration consists of d steps specified by the number of components in θ, with each step i updating the ith component of θ, conditional on the current values of the other components. Thus, each component of θ is updated separately and the algorithm proceeds componentwise. Given its simple structure, the Gibbs sampler is easy to program, as long as the conditional probabilities can be formulated. However, the conditional densities can be difficult to sample from if they are not members of any of the standard distributions. Moreover, Gibbs can be inefficient, especially with higher dimensional posterior distributions [73].

Hamiltonian Monte Carlo (HMC) is a more computationally efficient variant of MCMC [68; 74]. It modifies the standard MH algorithm by generating the transition

proposals in a more intricate and efficient manner. It adds an auxiliary momentum variable ϕ_i for each component of the parameter vector θ_i. HMC generates transition proposals by simulating the dynamics of a particle moving in a potential landscape defined by the posterior distribution. At each step, a random value is sampled for the momentum variable, the particle's dynamics are simulated, and a transition proposal is generated at the end of the simulated path. The trajectory is computed with a discretization, consisting of so-called "leapfrog steps," that use the log-posterior density gradient to approximate small advancing jumps of both θ and ϕ. HMC convergence can be monitored with the convergence of individual MCMC transitions, as well as the previously mentioned convergence statistics. In areas of the posterior with significant curvature, the simulated approximate particle trajectory can substantially propel away from the true one resulting in poor exploration of that particular area of the posterior [68]. Such divergent transitions are an essential method for monitoring HMC convergence.

While HMC is more involved both theoretically and in terms of implementation compared to the random walk MH algorithm, it has the advantage of being able to efficiently explore complex posterior distributions, if the sampler parameter are well specified. Choosing the appropriate momentum proposals, the number of leapfrog steps and step size is challenging, but algorithms that tune these parameters automatically have been introduced [68]. Moreover, automated software, such as Stan [4], are available for running HMC, requiring little manual effort beyond specifying the statistical model. The probabilistic models presented in Publications III-V were implemented in Stan.

3.2 Frameworks for prospective data

Cross-sectional data consists of samples taken from a population taken at a single time point and can be used to investigate correlative relationships between observed variables using statistical models. In microbiome research, cross-sectional inference is a common setup and involves comparing relative abundances of taxonomic groups or ecosystem-summarizing variables across population strata or against continuous background variables [5]. As more microbiome data is constantly beings collected, prospective data sets are also becoming increasingly available. The distinction to cross-sectional data is that a follow-up time is included; microbiome samples are separated in time from background variable measurements.

Two popular frameworks in both cross-sectional and prospective data analysis are generative models like generalized linear models (GLMs) [58] and discriminative models such as random forests [56]. Generative models provide a distribution of the data itself and can be used to generate new samples from the learned distribution after model fitting. Discriminative models, on the other hand, model decision boundaries between classes. In this section, we will delve into these approaches by presenting

GLMs and random forests and give an overview of the methods used in the cross-sectional analyses of Publications I and II.

3.2.1 Generative models

Generalized linear models (GLM) are a broad class of models that enable regression in a variety of contexts [58; 75]. The idea is to model the dependent variable Y as generated by a probability distribution in the exponential family. The members of this family can be written in the form

$$p(y|\theta) = h(y) \exp\left(\eta(\theta)T(y) - A(\theta)\right), \tag{3}$$

where the shape of the functions h, η, T and A are known and θ are the model parameters. The shape of these functions determines the particular probability distribution. For example, using $h(y) = \frac{1}{y!}, \eta(\theta) = \log\lambda, T(y) = y$ and $A(\theta) = \lambda$ produces the Poisson distribution:

$$\begin{aligned} p(y|\theta) &= \frac{1}{y!} \exp\left(\log(\lambda)y - \lambda\right) \\ &= \frac{\lambda^y e^{-\lambda}}{y!}, \end{aligned} \tag{4}$$

where y is the number of events in a unit of time. After specifying the distribution defining functions, the mean of the particular probability distribution is written as the linear predictor transformed with a link function $X\beta = g(\mu)$. For Poisson regression, the link function is $g(\mu) = \ln\mu$ and using the fact that the mean of the Poisson distribution is $\mu = \lambda$, the model likelihood can be recovered by substituting $\lambda = e^{X\beta}$ in equation (4). Similar reasoning can be used to formulate many of the usual regression models, such as linear, logistic or multinomial regression [58].

In publication I, we utilized DESeq2, a GLM specifically designed for high-throughput sequencing data [76]. DESeq2 takes into account the unique characteristics of this type of data and has been applied to microbiome data sets, although it was originally developed for gene expression data. The method assumes that the read counts follow a negative binomial distribution and uses an internal normalization process to adjust for bias introduced by varying total read counts among samples. Additionally, the method utilizes shrinkage in dispersion and effect size estimates, which enhance the reliability of the results when only a limited number of samples are available.

Cox regression

The Cox proportional hazards model is another example of a GLM [77; 78; 79] and was utilized in Publication II to associate microbiome features with risk for all-cause mortality. The model is the most commonly used tool for survival analysis, and is based on estimating the likelihood of a subject to experience an event at a given time based on explanatory covariate variables. The model is composed of two parts: a baseline hazard $\lambda_0(t)$ that represents the hazard at a specific time when the covariates are at their established baseline levels, and a hazard function that quantifies the multiplicative effect of the covariates on the hazard.

More specifically, the model can be specified as follows. Let $X_i = (X_{i1}, \ldots, X_{ip})$ be a p dimensional covariate vector for subject i. The hazard function, which quantifies the hazard for subject i at time t has the form

$$\lambda(t|X_i) = \lambda_0(t)\exp(X_i\beta), \tag{5}$$

where β is the vector of coefficients and λ_0 the baseline hazard function. The likelihood for observing an event for subject i at time t_i can then be written as

$$L_i(\beta) = \frac{\lambda(t_i|X_i)}{\sum_{j:Y_j \geq Y_i} \lambda(t_j|X_j)} = \frac{\exp(X_i\beta)}{\sum_{j:t_j \geq t_i} \exp(X_j\beta)}, \tag{6}$$

where the latter equation follows directly from the definition of the hazard function. The summation in the formula only considers subjects who are still participating in the study at time t_i. Subjects who have either experienced the event being studied or have been censored are excluded from the calculation at this time point. Censoring refers to a situation where only partial information is available about a subject. For example, in a clinical trial, if a subject drops out before the end of the follow-up period, their information about the event after the removal is not known, making them a censored case.

This partial likelihood is the ratio of a hazard for the individual i to the sum of the hazards for those who have not experienced an event at Y_i. It should be noticed that there is no need to specify the form of $\lambda_0(t)$ as it is cancelled from the likelihood function. Assuming the subjects are statistically independent from each other, the likelihood for all realized events can be written as:

$$L(\beta) = \prod_i L_i(\beta), \tag{7}$$

where the product is taken over the subjects for which the event has not occurred and that are not censored.

The exponentiated parameters e^{β_i} are called hazard ratios (HR) and quantify the impact of covariates on survival while holding other covariates at their baseline

levels. A HR of 1 ($\beta = 0$) means that the covariate has no effect on survival, while HR values less than 1 or greater than 1 indicate lower or higher hazards, respectively.

The standard Cox regression described above models the covariates linearly, which may be a too restrictive assumption in some cases. Nonlinear effects can be estimated with splines, which are piece-wise defined polynomial (often cubic) functions [80]. In spline regression, the parameters are estimated in disjoint intervals separated by automatically chosen points called knots, and the separate functions are optimized so that the first and second degree derivatives are equal at these knots. Another variation of the Cox model is to allow for time-varying covariates, allowing for modeling scenarios where the X_i change during the follow-up time. [81].

3.2.2 Predictive models

Random forest is a predictive ensemble learning algorithm that combines the predictions of multiple decision trees fitted to random sub-samples of the data [56]. The method can be used both in classification and regression tasks.

The algorithm is based on the idea of bootstrap aggregating, also known as bagging. To formulate the model, let $X = \{x_1, \ldots, x_n\}$ be a set of covariate data points, where each x_i may be a vector, and $Y = \{y_1, \ldots, y_n\}$ the corresponding responses. In bagging we first construct a collection of training sets $(X_b, Y_b), b = 1, \ldots, B$ by taking a sample of size n, with replacement, from (X, Y). Then, for each b, a decision tree f_b is trained, and predictions for unseen data x' are computed by averaging the predictions of individual trees $f_b(x')$ (in regression tasks) or by using the class chosen by most trees (in classification). The tree fitting in random forests differs from ordinary decision trees in that is also randomly selects the features (components of x_i, $i = 1, \ldots n$) used at each split. The purpose of this randomization is to reduce the correlation between trees. Uncertainty for the prediction y' can be computed as the standard deviation of predictions of individual trees

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B}(f_b(x') - y')^2}{B-1}}.$$

The algorithm is non-parametric, so it does not provide an effect size for model covariates like GLMs do. However, an importance score that reflects the significance of each covariate in the regression or classification task can be computed using the out-of-bag (OOB) error. This score is based on the average prediction error for a training sample x_i over the trees that don't include that sample. To calculate the importance score, the values of each feature are permuted and the difference in sample-wise OOB errors before and after the permutation is computed. The final importance score is the average of these errors normalized with their standard deviation.

In addition to standard regression and classification tasks, variations of the random forest have been developed for different settings. For example, the random

survival forest [82] was created for survival data and differs from the regular algorithm by taking survival times and right-censoring into account in the tree splitting rules. Another application is missing value imputation, as demonstrated in the R package *missForest* [83]. These variants were utilized in Publications I and II of the thesis.

3.3 Stochastic processes for longitudinal data

Time series refers to data that consist of repeated observations of a systems taken at different time points [84]. The main distinction to the cross-sectional data is the temporal ordering of the observations. Typically, the time interval between consecutive data points is constant, but this is not always the case, as in many practical scenarios it is impossible to observe a system at will. Time series analysis methods try to extract information about the characteristic of the data, such as serial dependence and seasonality. In this section, we will present some of the basics of time series analysis methods that are utilized in the thesis Publications III-V. The focus is on stochastic methods, that take into account random data variations as opposed to deterministic models like ordinary differential equations.

3.3.1 Discrete time

A natural starting point for time series modeling is the Gaussian white noise ϵ_t where each element is independently sampled from a zero-mean normal distribution $N(0, \sigma^2)$, where σ^2 is the variance. This white noise process is commonly assumed to be the source of random variations in time series [84].

The next level in complexity in time series modeling is the random walk. In discrete time, it can be generated as the cumulative sum of a white noise process: $x_t = \sum_{i=1}^{t} \epsilon_i$, or, equivalently, with the recursion $x_{t+1} = x_t + \epsilon_t$.

The random walk and Gaussian noise can both be seen as special cases of the AR(p) process, where each element is generated as a linear combination of the previous p elements with added noise. This means that the current value depends on a set of past values, leading to a more complex and nuanced time series model. The AR(p) process is specified with the following recursion:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_{t-1}. \tag{8}$$

The AR(p) process can be used to model scenarios where the data oscillates around a long-term mean level. For instance, when $p = 1$, the equation becomes $x_t = \phi x_{t-1} + \epsilon_{t-1}$, and the values tend to revert back to the process mean of 0 at a rate of ϕ.

However, in many cases, a simple AR process may not capture the complexity of the data, but it may be used as a component in a more sophisticated model [84]. If the data exhibits an apparent linear, cyclical, or random walk trend, for example, this feature can be incorporated into the AR process by adding a corresponding term to the right-hand side of Eq. (8).

Likelihood function for the AR(1) process, which enables parameter estimation is

$$\mathcal{L}(\theta) = \prod_{i=2}^{T} N(\phi x_{t-1}, \sigma^2).$$

It should be noted that the first observation x_1 is not included in the calculation of the likelihood as there is no previous observation from which it could be generated.

In certain cases, it is reasonable to assume that the process parameters can vary in time [85]. In natural processes, such changes may occur due to intrinsic evolution or changes in external conditions, and using time-varying models may provide a more accurate and justified description of system. A time-varying variant of the AR(p) process can be formulated by adding time-dependence in the parameters of Eq. (8) [86]:

$$x_t = \sum_{i=1}^{p} \phi_{i,t} x_{t-i} + \epsilon_{t-1}. \tag{9}$$

The time-varying model, however, requires additional assumptions since otherwise each $\phi_{i,t}$ would need to be estimated solely based on the time points x_{i-1} and x_i. One approach is to assume that ϕ_i evolve as a random walk processes or as smooth functions [86]. In the thesis, we used Gaussian processes as priors for the time-dependent autoregressive parameter, as described in Publications IV and V.

The parameters ϕ_i in Eq. (8) define the deterministic characteristics of the system [84]. Stationarity is a particularly important aspect of the AR(p) process, as it has close ties to early warning signal indicators [13] and stability metrics [87], which will be discussed in Section 4. Intuitively, stationary means that a system behaves in a predictable manner and its properties do not change over time, whereas non-stationarity suggests that the system is becoming chaotic and difficult to predict. With the AR(p) process, stationarity can be studied by examining the characteristic polynomial of the AR(p) process $\Phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$. The magnitude of the roots of Φ determines the stationarity of the process, in that the process is stationary if and only if all of its root (which may be complex) have absolute values greater than 1 [84].

3.3.2 Continuous time

The stochastic processes discussed above are formulated for integer time points, meaning that applying them to data with uneven observation intervals would require data pre-processing with imputation, interpolation, or some approximation methods [88]. Such additional steps can lead to loss of information, and they should ideally be avoided. A more appropriate approach for handling data with uneven time points is to use continuous time stochastic processes.

Gaussian processes

The generalization of the random walk to continuous domain is called *Brownian motion* or *Wiener process* [89], and is defined as the collection of random variables $W_t, t \geq 0$ with the properties that each increment $W_{t+u} - W_t$, $u \geq 0$ is normally distributed with mean 0 and variance u, independently of any past value W_s, $s \leq t$. A property relevant for simulating data from the Wiener process is that $W_t - W_s \sim N(0, s-t)$ for $0 \leq s \leq t$, which directly implies that for any $\Delta t > 0$, $W(t+\Delta t) - W(t) \sim N(0, \Delta t)$.

The Wiener process belongs to a larger class of stochastic processes known as Gaussian processes (GP) [90], which are widely used for non-parametric regression, both in time series and cross-sectional data. A Gaussian process $GP(\mu, \Sigma)$ is defined as a set of random variables X_t with the property that each finite collection of these variables is multivariate normally distributed with mean μ and covariance function Σ. The process is completely specified by μ and Σ. While μ determines the average level of the process, the covariance function Σ has a more defining impact on the process behaviour, as it determines the relationship between different points in the process [90].

A commonly used class of covariance functions is the Matérn covariance functions [91]. The general definition, which we omit, depends on a parameter ν that determines how "wrinkled" the generated functions are. More specifically, ν defines the level of differentiability of the functions generated from the process. When the value of ν is restricted to $\nu = \frac{1}{2} + p$, $p \in N_+$, the definition of the Matérn covariance functions reduces to the form

$$C_{p+1/2}(d) = \sigma^2 \exp\Big(-\frac{d \cdot \sqrt{2p+1}}{\rho}\Big) \frac{p!}{(2p)!} \sum_{i=0}^{p} \frac{(p+1)!}{i!(p-1)!} \Big(\frac{2d\sqrt{2p+1}}{\rho}\Big)^{p-i}.$$

For specific choices of p the formula reduces to even more manageable forms. For $p = 0$ we get the so-called Ornstein-Uhlenbeck (OU) kernel:

$$C_{1/2}(d) = \sigma^2 \exp\Big(-\frac{d}{\rho}\Big) \tag{10}$$

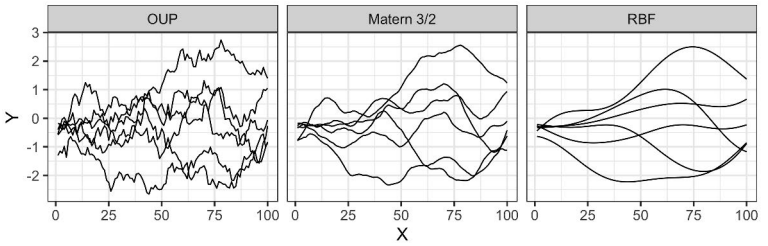

Figure 3. Sample functions from a Gaussian process with the Ornstein-Uhlenbeck (OUP), Matern 3/2, and radial basis (RBF) covariance functions. Length-scale was set to $\rho = 25$ and variance $\sigma^2 = 1$. Differences in the level of smoothness are apparent with the OUP producing the most jagged functions, while RBF produces the smoothest. The same random number generator seed was used for each panel, which makes the functions comparable.

Realizations from an OU process are functions that are continuous but nowhere differentiable (see Figure 3). By choosing $p = 1$ we get

$$C_{3/2}(d) = \sigma^2\Big(1 + \frac{\sqrt{3}d}{\rho}\Big)\exp\Big(-\frac{\sqrt{3}d}{\rho}\Big), \tag{11}$$

which produces once differentiable functions. Usually the difference between, say, 5 and 6 times differentiable functions is negligible and there is typically no need to consider Matérn kernels beyond $p = 2$ [90]. However, by taking the limit $\nu \to \infty$ we recover the important case of the squared exponential, or radial basis kernel:

$$\lim_{\nu\to\infty} C_\nu = \sigma^2 \exp\Big(-\frac{d^2}{2\rho^2}\Big). \tag{12}$$

This kernel produces infinitely differentiable (smooth) functions and is often the first choice for Gaussian process regression. Despite it being the most utilized member of the kernel universe, it has been argued that smooth functions may be unrealistic for most applications and that other Matérn members might be a more appropriate choice [91].

In practice, the choice of the kernel and its hyperparameters depends on the nature of the data and the desired level of smoothness or complexity in the resulting model. The hyperparameters of the Matérn kernel determine the degree to which the produced functions vary [90]. The length scale ρ affects the variations in the horizontal directions, while the variance parameter σ is responsible for the amplitude. As usual, hyperparameter choices should reflect prior beliefs about the system, although it is also possible to determine suitable hyperparameter values based on the data by treating them as unknown model parameters.

Several other types of covariance functions exist that are more suitable for different scenarios [92]. For instance, the periodic kernel is useful when considering cyclical phenomena, while the polynomial kernels, in fact, provide a Bayesian for-

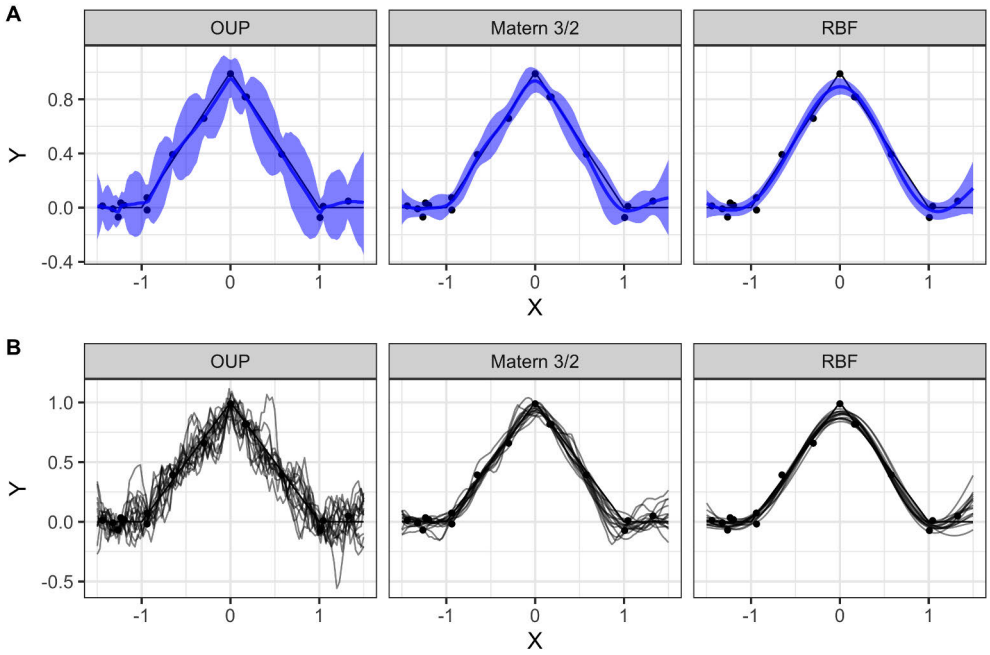

Figure 4. Non-parametric regression with a Gaussian process using the Ornstein-Uhlenbeck (OUP), Matern 3/2, and radial basis (RBF) covariance functions. The points represent noisy observations from the the triangular function (black line). Posterior mean (blue line) and 95% credible interval (blue ribbon) capture the triangle function with different smoothness characteristics.

mulation for polynomial regression. Non-stationary kernels are a relatively recent development [93] and more flexible, allowing the hyperparameters to vary as a function of the input variable (time in time series). However, their increased complexity makes them challenging to fit in terms of MCMC convergence and computation time.

Conditioning a GP on data, that is, performing regression is also relatively straightforward [90]. The likelihood of data X is a multivariate normal distribution $MVN(X|\mu, \Sigma)$ and predictions for values Y^* at test locations X^* can be computed by including the test locations in X and then marginalizing over the training data, which results in

$$p(Y^*|X^*, X, Y) = N(\Sigma(X^*, X)\Sigma(X, X)^{-1}Y, \\ \Sigma(X^*, X^*) - \Sigma(X^*, X)\Sigma(X, X)^{-1}\Sigma(X, X^*))$$

See Figure 4 for an illustration of GP regression.

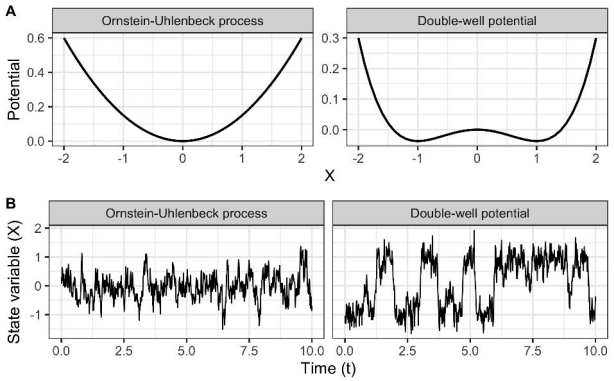

Figure 5. **A** Potential functions for single and double well potentials. **B** Example trajectories from the solutions to the corresponding SDEs.

Stochastic differential equations

Stochastic differential equations (SDEs) are another approach to modeling stochastic dynamics [89]. They are similar to ordinary differential equations with the distinction that the equation contains a term for a stochastic process. This distinction makes the solution of an SDE also a stochastic processes. While SDEs can be defined on a more general level, for many applications, the following equation form is adequate [89; 94]:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t. \tag{13}$$

The deterministic part of the equation, f, is called the drift and g is the diffusion that scales the differential of the Wiener process.

The equation (13) is a formal notation and should be interpreted as the integral equation [89]:

$$X_{t+s} - X_t = \int_t^{t+s} f(X_u, u)du + \int_t^{t+s} g(X_u, u)dB_u. \tag{14}$$

The second term on the right side of the equation is a stochastic integral, which is a central concept a field of mathematics called Itô calculus [94]. Here, we omit the precise definition and computing of stochastic integrals as it would require extensive theoretical developments that are not necessary for the purpose of this thesis.

One of the simplest examples of SDEs is the generalization of the AR(1) process to continuous domain, defined as

$$dX_t = \phi(\mu - X_t)dt + \sigma dW_t$$

[89]. This equation describes the random movements of a single attractor system, determined by the mean parameter μ. The process has the tendency to revert towards

its mean at a rate determined by the mean-reversion rate ϕ. The solution to this equation is the Ornstein-Uhlenbeck process presented in Section 3.3.2 [89; 90]. Such processes are relatively common in practice and, for example, abundance levels of microbiome groups have been modelled with the OUP [95].

A convenient property of the OU process is that its analytical transition density is known:

$$p(X_t|X_0 = x_0) = N\left(\mu + (x_0 - \mu)e^{-\phi t}, \frac{\sigma^2}{2\phi}(1 - e^{-2\phi t})\right).$$

Knowing the analytical density enables exact simulations from the process and parameter inference without numerical methods. However, for most SDEs, no explicit transition density can be written. An example of such an equation is the extension of the OUP to the case of two attractors [89]. This equation describes the dynamics in a double-well potential and can be formulated with a 3rd degree polynomial drift function: $f = X_t - X_t^3$. The transition density for such intractable processes can be handled with approximation methods [89; 96]. The Euler-Maruyama (EM) approximation, which is an extension of the Euler method used with ordinary (non-stochastic) differential equations, is one commonly used method. Assuming the process X_t is a solution to the SDE of the form in Eq. (13), the following iterative scheme gives the EM approximation of X_t:

$$p(Y_{i+1}|Y_i) = N\left(Y_i + f(t_i, Y_i)\Delta t, g(t_i, Y_i)^2\Delta t\right), \tag{15}$$

where $\Delta t = t_{i+1} - t_i$ is the time difference between Y_{i+1} and Y_i.

Given an initial value Y_0, the EM approximation can be used to generate an approximate realization of the process X_t. By choosing a small enough step size, typically 0.01 as commonly found in literature, the simulation will approximate the true solution X_t. The choice of step size is important in ensuring the simulation converges to the true solution, and a smaller step size will result in a more accurate approximation.

In addition to the simulation of the process, the transition density provides a means for parameter estimation. This is because it defines a generative process for the data, making it possible to calculate the likelihood for a time series X_t. The likelihood can be expressed as a product of the transition densities between consecutive time points:

$$\mathcal{L} = \prod_{t=2}^{N} p(X_t|X_{t-1}). \tag{16}$$

In cases where the transition density cannot be written analytically, it is possible to use the EM approximation to formulate an approximate likelihood function

[89]. Approximate Bayesian computations are another option means for handling intractable SDEs [66].

An intuitive way to understand SDEs is through potential analysis. This approach views the deterministic forces of the SDE, which are specified by the drift, as a potential landscape where a ball is moving [89]. The idea is to write the drift function in the form $f = -U(z)'$, where U is the potential function. The local extrema of U, which are also roots of f, correspond to locations where the deterministic forces vanish. At these locations, the qualitative behavior of the system is determined by sign of the second derivative $U(z)''$, and the location is either an attractor, a repeller, or a saddle point. By plotting the potential landscape, it is possible to get a visual understanding of the qualitative properties of the SDE, as shown in Figure 5.

4 Stability in dynamical systems

In this chapter, we present dynamical systems with a focus on the stability properties, which are the topic of publications III-V.

4.1 Dynamical systems

The branch of mathematics called dynamical systems theory is concerned with models that describe how state variables change over time [97]. Formally, a dynamical system can be defined as a triple $(\mathcal{T}, \mathcal{S}, \Phi)$, where $\mathcal{T}$ is the parameter space, $\mathcal{S}$ the state space and Φ a function that maps from the trajectory space $(\mathcal{S} \times \mathcal{T})$ into the state space $\Phi : (\mathcal{S} \times \mathcal{T}) \rightarrow \mathcal{S}$. In applied contexts, $\mathcal{T}$ is typically the time parameter, and the dimensions of $\mathcal{S}$ correspond to some observable features, such as location, velocity, or magnitude of, for example, animal species in a population.

In order to specify the dynamics governing the system, ordinary and stochastic differential and difference equations, such as the Lotka-Volterra (LV) model [98] and its variants, are utilized. The LV is a population dynamics model consisting of dimension-specific growth rates and interactions between its components, such as species in population ecology. While the LV and its extensions have successfully been applied in various contexts, including microbial communities [51; 99], it is a relatively simplistic model, as it only contains linear pairwise interactions that are constant in time [30; 54]. However, it is often used as the starting point when considering more sophisticated models for population dynamics. In Publication V, we employ a generalized LV model variant as the basis for the simulation experiments.

The study of systemic stability properties is an essential aspect in gaining comprehensive understanding a dynamical systems [97]. Features in the state space are, by definition, constantly evolving, and understanding the qualitative nature of the dynamics and how they react to both internal and external disturbances is often essential in applications. Stability analysis can, for instance, identify fixed point attractors, or stable states, that indicate where the state variable tends to drift over time and after perturbations [100]. Understanding the properties of these attractors, such as local attraction strength [87] or the risk of transitioning into another state [101], can be crucial for systems management, especially when there are multiple stable states. These features can be important in defining what ”normal” and, perhaps more importantly, ”abnormal” behavior looks like, what type and what strength

of disturbances the system can withstand, and when systems management actions are necessary. Such attributes can be derived from the differential equations governing the dynamics. However, these equations are often not available and difficult to determine from data [102], making it necessary to find alternative ways of inferring stability characteristics. In this chapter, we will present some ways in which systemic stability characteristics can be inferred without relying on specific mechanistic models.

4.2 Stability and resilience

Stability is an intuitive and seemingly simple concept but, in practice, often more multifaceted than it may seem. The concept and its related terms are often used without clear definition, which can lead to confusion and disconnection between theoretical and empirical studies [103]. Moreover, disturbances that natural real-world systems experience vary in terms of magnitude, duration, frequency, and how they change over time and space. Similarly, the way a system reacts to the these perturbations can be multidimensional [103; 104].

To better understand stability, it has been characterized through five different components that describe the reaction of a system to perturbations [105]. *Asymptotic stability*, a binary variable, indicates whether or not a system will ultimately recover its equilibrium state following perturbations a small distance away from it. *Variability* is measured as the coefficient of variation of a variable over time or across space. *Persistence* quantifies the time a system is expected to maintain its current state before undergoing some fundamental change. *Resistance* is the ratio of a system variable before and after a perturbation event, while *resilience* describes the ability of a system to recover from perturbations and the rate at which it does so. In ecological literature, two qualitatively distinct definitions for resilience are used [102]. *Ecological resilience* refers to the magnitude of perturbations a system can absorb while maintaining its current state, without transitioning to an alternative functional or structural configuration. *Engineering resilience*, on the other hand, describes the ability of a system to recovery near a stable state. The defining difference between these concepts is that the former focuses on global properties, while the latter is concerned with local properties.

Stability and resilience can be quantified and measured in several ways, many of which are explicable from, or related to, the properties of the potential landscape in which the state of the system evolves (see subsection 3.3.2). These properties include potential valley depth, width, and curvature (second derivative of the potential) at the valley bottom, for example, and quantify the strength of an attractor. Many of these metrics are, however, correlated and function as alternative ways of measuring the same systemic aspects [102]. In [104], for instance, the authors compared 27 different stability metrics for multivariate systems and found them to group into

3 relatively independent components based on the responses to different types of perturbations.

If the system is well-understood and a mechanistic model exists, these metrics can be directly computed from the model. In practice, however, these mechanisms are typically poorly understood, and may be indeterminable based on the available data. Furthermore, in multidimensional systems the potential may even not exist [102]. In such cases, the dynamics may be approximated with simpler models. For instance, the Ornstein-Uhlenbeck (OU) process can be used to approximate dynamics of a single stable state, while bistable dynamics can be emulated with SDEs having a higher degree polynomial drift or a non-parametric function.

4.3 Early warning signals for critical transitions

In nature, systems are constantly in a state of change as a results of both internal and external factors. Such developments may affect the system in a way that causes the forces maintaining the current stable state to deteriorate, leading to a loss in resilience [13]. In certain systems, if the conditions change drastically, the system may cross a so-called tipping point and transition into a new, alternative stable state, which may have significant implications for the overall functioning of the system. As transitions between states may be undesired, it is crucial that such events can be anticipated [106]. The theory of early warning signals (EWS) suggests that there are specific statistical signatures that can be observed as a systems approaches a tipping point.

A key aspect of EWS is critical slowing down (CSD), which refers to the dynamics of the system becoming slower. Intuitively, as a tipping point is approaching, the walls of the current potential well tend to lose steepness, which allows the state to wander farther from the potential minimum [13; 107]. In other words, the effect of random variations compared to the deterministic forces increases, and the drift towards the stable state loses strength [13; 108]. This slowing down is reflected in measurable properties of the state variable. Correlation between consecutive time points (lag-1 autocorrelation) and the variance of the state variable and, in particular, *increase* in these statistics are the most commonly used signals [107].

CSD and EWSs have been detected in laboratory experiments and in a range of natural and social systems from animal populations to financial systems (see Table 1 in [109] for a list of recent research). In terms of the motivational theme of this thesis, alternative stable states and transitions between these states have been observed in microbiome time series as well [16], but whether or not the concepts of CSD and EWS apply to microbiomes remains an open questions.

Let us illustrate resilience loss in a well-studied simulation model [110] which consists of logistic growth limited by harvesting of the state variable X :

$$dX = \left(rX \left(1 - \frac{X}{K} \right) - c \frac{X^2}{X^2 + h^2} \right) dt + \sigma X dW_t. \tag{17}$$

Above, r is the growth rate, K the carrying capacity that determines the maximum population size, c the harvest rate, h the half-saturation constant and σ the level of instantaneous stochastic variations. The parameters c and h determine the rate of removal of biomass (as represented by X) from the systems. By changing the value of the harvest rate c, it is possible induce a bifurcation in the parameter space which leads to the emergence of an alternative stable state [107]. Within a certain range of parameters the system exhibits bistability: two stable states coexist and random variations may drive the system over the potential barrier between the states. If the stochastic noise is sufficiently large or the observation time long enough, the state may switch back and forth between the alternative states. If the switches occur frequently relative to the observation time the system displays a phenomenon called *flickering* [13]. However, if the value of c is further increased, the original stable state will eventually disappear altogether and the system will collapse into the new potential minimum (see Figure 6 for illustration). The harvest model and the induction of EWS as described above is a common strategy for generating data for simulation experiments in literature. This approach was also used in Publication IV of the thesis.

A more holistic understanding of the stability properties of the harvest model can be obtained with be bifurcation diagram [107] displayed in Figure 7. The bifurcation diagram presents a graph of the potential extrema for a range of values of c, and provides a visual representation of the number and locations of the stable states and the emergence of tipping points. The bifurcation diagram also illustrates the concept of *hysteresis*, which refers to the phenomenon where if the system is driven beyond a tipping point, recovering the original state may require restoring the conditions beyond another tipping point leading to the original state [111].

EWS detection

The process of EWS detection as typically performed in literature can be outlined in the following steps [107]:

1. Remove long-term mean-level variations.
2. Compute the chosen EWS indicator in sliding windows over the data.
3. Compute Kendall's rank correlation τ between time and the inferred indicator trend.
4. Assess statistical significance for $\tau > 0$ with surrogate data analysis methods.

The purpose of the first step is to pre-process the data so that mean-level variations unrelated to the stability of the system are removed [112]. These variations could arise, for example, because of seasonality or other processes that are unrelated

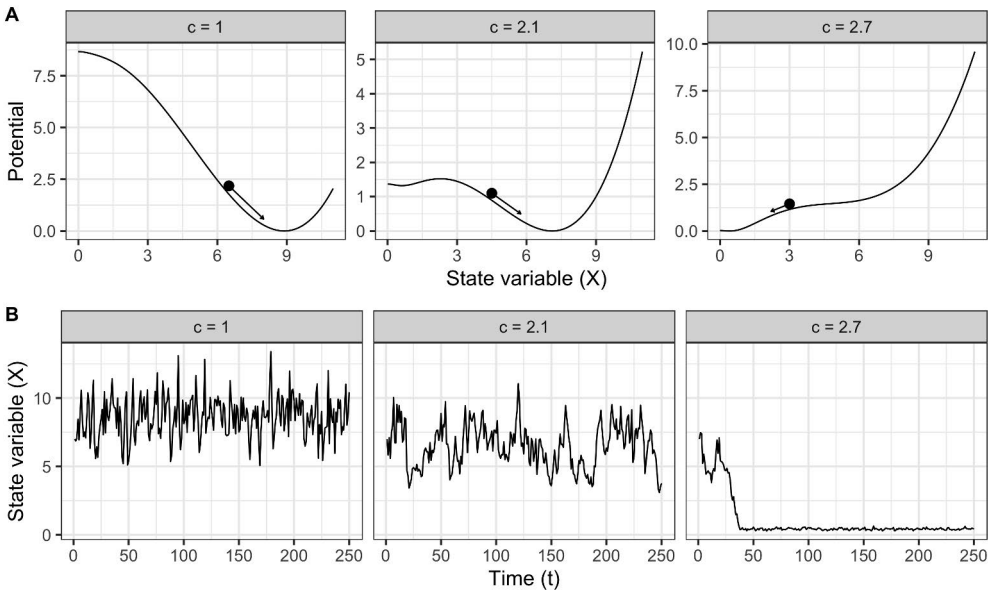

Figure 6. **A** Potential landscapes for the harvest model at different values of the harvest parameter c. The system state, represented by the ball, reverts towards the potential minima (valley bottoms) at a rate denoted by the arrow length. At $c = 1$ there is a single stable state, at $c = 2.1$ another stable state has emerged, and at $c = 2.7$ the original stable state has vanished. **B** Example trajectories of the solutions to the system defined by Eq. (17). Increasing the parameter from $c = 1$ to $c = 2.1$ has caused a loss of resilience, which is reflected in the increase of lag-1 autocorrelation (0.24, 0.76) and variance (2.3, 2.5, for $c = 1$ and $c = 2.1$, respectively). At $c = 2.7$ the system quickly collapses to the lower state. The minimum potential levels are set to 0 in each panel. Parameter values used in the simulation are $K = 10, h = 1, r = 1$ and $\sigma = 0.15$.

to short-term oscillations around the stable state that reflect the properties of the equilibrium state. The common methods of detrending include Gaussian smoothing or first-differencing, and omitting this step can cause spurious conclusions in the next steps [107]. Detrending with Gaussian smoothing requires setting a bandwidth parameter that determines the range of time points affecting the trend estimate at each time point. The bandwidth has a significant influence on the resulting residuals (difference between data and estimated trend), and the effect propagates trough the subsequent steps [107; 112]. The first step could also include interpolation if the time points are not equidistantly distributed or if there are missing observations [113].

In the second step, the EWS indicator is computed from the residuals using sliding windows along the time series. Similarly to the bandwidth selection in the previous step, the length of the window can have a significant impact on the results and conclusions of the process [114]. An alternative to sliding windows is time-varying models, which we utilized in publications IV and V. However, formulating these models for many EWS indicators can be challenging because the generative models are not apparent, and sliding windows are still generally used in the literature.

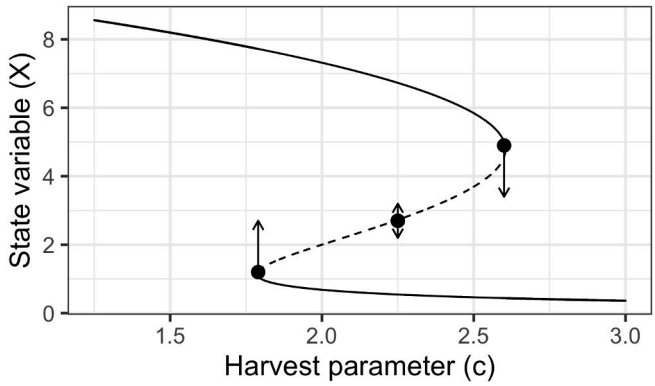

Figure 7. Bifurcation diagram for the harvest model. The black line denotes the potential extrema as a function of the harvest parameter c and the balls illustrate what happens to the system at these critical points. The system has two stable states when the control parameter is between $c_1 \approx 1.8$ and $c_2 \approx 2.6$ and a single one outside this range. Increasing c from a value below c_1 to one above c_2 inevitably induces a state transition. In a deterministic systems, recovering the original state requires decreasing c below c_1 (hysteresis) but in a stochastic one the transition may occur in the bi-stable range due to random fluctuations.

The second step results in an indicator trajectory, which is then evaluated in step three. For most indicators, an increasing trajectory trend implies critical slowing down, and the standard way to quantify the trend in EWS context is to compute Kendall's rank correlation τ between the trajectory and time [107]. It is defined as

$$\tau = (N_{\text{concordant pairs}} - N_{\text{disconcordant pairs}})/N_{\text{all pairs}},$$

where N refers to the number of elements in the subscript set, and a pair $(t_i, \phi_i), (t_j, \phi_j)$ is said to be concordant if $t_i \leq t_j$ implies $\phi_i \leq \phi_j$ and otherwise disconcordant, where ϕ_i is the indicator at time t_i. A value of τ close to 1 indicates a clearly increasing trend and an EWS, while $\tau \approx 0$ implies a negative finding.

However, it is possible that a positive trend has arisen by chance, so assessing statistical evidence for a true finding is performed in the final step [115]. An issues with this is that the probability distribution of τ needed for hypothesis testing is not readily available. To overcome this, an often-used solution is to use the null hypothesis that the indicator trend has arisen randomly and generate an approximate probability distribution based on this assumption with surrogate data analysis methods [107]. This is achieved by first performing model selection for an autoregressive-moving-average ARMA(p, q) model, in terms of p, q with some information criterion, such as the Akaike information criterion (AIC). Surrogate time series are then generated from the selected model, and the EWS indicator trend is evaluated for each surrogate series. The distribution of these surrogate Kendall's τ's forms the approximate sampling distribution, against which the actual data estimate can be compared [116].

This procedure can be applied for both one-dimensional and multivariate time series data. However, in multivariate systems, the phenomenology is more complex as the degrees of freedom increase dramatically. Nevertheless, many multivariate indicators have been developed, many of which are generalizations of the univariate indicators [117].

Challenges in EWS detection

Generic data-based EWS are attractive in their simplicity and because they can be used regardless of the application domain and in the absence of mechanistic models [106]. However, detecting them is acknowledged to be a challenging task [14; 109; 114; 118].

Firstly, EWS methods are validated based on simulations data, which poses several problems. Typically, data is simulated so that the tipping point is approached gradually and linearly over the available time series, leading to a state transition at the end. However, collecting such data in real-world scenarios would happen only by chance, apart from controlled experimental settings, as the observations might begin at a considerable distance from the tipping point and the system may maintain stable conditions for an arbitrarily long period of time. Moreover, developments leading to a state transition can occur rapidly compared to the observed time period [118], and the sampling interval needs to be set appropriately to capture the characteristics of the dynamics [14].

Even in ideal data collection scenarios, other challenges remain. EWS indicators are data-hungry relative to the amounts of data practical to collect in most cases, especially in the field of biomedicine [14]. Low signal-to-noise ratio, low resolution, uneven observation intervals, partially observed systems in multivariate settings, and measuring a proxy variable instead of the actual target are other prevalent data-related issues that hinder signal detection [109; 114; 117]. These issues are particularly pronounced in fields such as ecology and biomedical studies, where systems are multivariate, difficult to isolate from their environment, and challenging to monitor for extended time periods.

While several promising EWS indicator studies have been published, their applicability to real-world data is not self-evident. The model used for data generation affects the conclusions in simulations studies [119; 120], which poses a challenge for generalizability to actual data. Additionally, EWS indicators may not be robust to different types of bifurcations, and this information may not be available for natural systems [119]. Furthermore, in some cases, transitions can even occur without detectable dynamical changes [121].

Finally, using Kendall's τ as the test statistic for the chosen EWS indicator can present challenges. The statistic measures *changes* in conditions, not the absolute risk at a given time point, meaning that if the observation period starts close to a

tipping point, the analysis may produce a misleading understanding of the situation: τ may be close to zero even if there is a substantial risk for transition. Moreover, the use of surrogate data analysis to test hypotheses about τ also requires additional assumptions about the data (that an ARMA process serves as a baseline model), which complicates the analysis.

It should be noted, however, that lag-1 autocorrelation could be used to measure absolute risk as values above 1 imply non-stationarity regardless of prior developments [122]. Using the metric in this way does not require computing or statistical testing for τ, although applying this approach to continuous systems is problematic since the definition of a unit observation interval is arbitrary.

In conclusion, while EWS indicator studies have been promising, there are a number of limitations and potential issues that need to be taken into account when applying these methods to real-world data. In publications IV and V we utilize, for the first time, the Bayesian statistical framework in this context and show how some of these issues can be addressed with probabilistic time-varying methods.

5 Summary of publications and discussion

In this chapter, we provide an overview of the articles included in this thesis. Each section will be dedicated to a separate publication and will present an overview of the motivation, methods and results, along with discussion and specification of the contributions of the thesis author. The discussions will outline some of the challenges and impasses encountered during the projects, along with ideas for further research that sprouted during the projects. While the publications included in this thesis have biomedical significance, this chapter will primarily focus on the statistical and modeling aspects of the works. The reader who is interested in the biomedical implications of the work can find more information in the original publications and supplementary materials.

5.1 Publication I: Microbiome-based prospective analysis

Motivation: Maternal prenatal stress is known to be associated with infant developmental outcomes but the specific mechanisms of this link are not fully understood [123]. It has been speculated that the infant gut microbiome may have a mediating role, and this hypothesis is supported by animal models [124] and a small previous study that examined the association between maternal stress and gut microbiome features [125]. The aim of publication I was to shed further light on this potential link.

Methods: In publication I, we compared prenatal psychological distress (PPD) and hair cortisol concentration (HCC) of mothers to samples of the infant's gut microbiome at 2.5 months. The study was carried out as part of the FinnBrain Birth Cohort Study [11].

We used the DESeq2 model [76], a type of generalized linear model, to determine the association between the different stress scores and gut microbiome composition at the genus level. In order to control for confounding effects, we included breastfeeding status, mode of delivery, infant age at sample collection, and infant sex as covariates in the model. The Benjamini-Hochberg procedure was applied to account for multiple comparisons [126].

Results: We observed statistically significant associations between maternal

stress and gut microbiome features. More specifically, the bacterial phylum *Proteobacteria* and genera *Akkermansia* and *Lactobacillus* were associated with stress.

Discussion: The results of the analysis shed light on the potential role of the infant gut microbiome in mediating the link between maternal prenatal stress and infant developmental outcomes. DESeq2 was chosen for modeling the association between microbiome features and background variables as it has been developed for high-troughput sequencing assays, and can take into account some of the critical aspects of such data: count type, mean-variance dependence and presence of outliers [76].

Author's contribution: The author was responsible for the data analysis evaluating the association between PPD and infant gut microbiome composition. The task included selecting appropriate statistical tools, designing, optimizing, and implementing the analysis workflow, interpreting the results, and participating in writing the manuscript.

5.2 Publication II: Survival modeling in a population cohort

Motivation: A wealth of evidence from extensive cross-sectional studies suggests an association between human microbiome composition and various disease and lifestyle factors [1; 2; 20]. However, prospective links between the microbiome and health are largely unknown due to the lack of cohorts with extensive follow-ups. Moreover, it is unknown how well the existing statistical methodology is suitable for analyzing prospective time-to-event microbiome data sets.

In publication II, we aimed to investigate the predictive power of gut microbiome features on all-cause mortality in a representative random sample of the Finnish adults population. The data set is unique in that data of comparable sample size and follow-up time does not exist. In this sense, the study offered a unique opportunity to investigate applicability of survival analysis methods on microbiome data.

Methods: The stool microbiome samples were collected in 2002 as part of the FINRISK health examination survey [12]. Comprehensive health records were available for the following 15 years and included information on several background variables, including the time of death for the individuals who deceased during the follow-up.

We performed survival analysis based on the Cox regression [77] and survival random forest [82], with taxonomic features as explanatory variables and the time until death as the response. We used the centered log-ratio (CLR) transformed genus abundances, alpha and beta diversity, taxonomic co-occurrence networks [127], and functional Kegg Orthology groups as predictors for all-cause mortality. To account for confounding effects, we included covariates in the model that are known to affect both the microbiome and mortality risk: baseline age, body mass index, sex, smoking

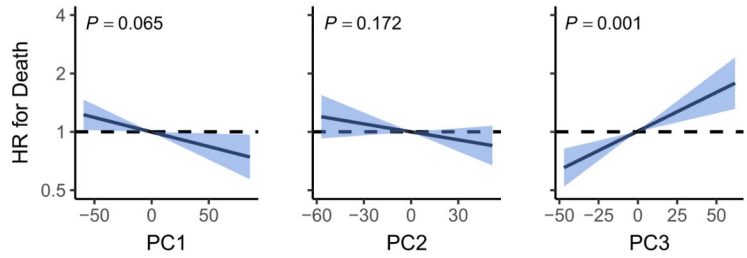

Figure 8. Association between mortality risk and the first three principal components of beta diversity (PC). The black line indicates the estimated hazard ratio compared to median PC value and blue area the 95% confidence interval (CI). Unit variance increase in the PCs were related to hazard ratios of 0.92 (95% CI, 0.85–0.99; FDR-adjusted P=0.065; two-tailed Wald test), 0.95 (95% CI, 0.87–1.02; FDR-adjusted P=0.17; two-tailed Wald test) and 1.14 (95% CI, 1.07–1.23; FDR-adjusted P=0.001; two-tailed Wald test) for PC1–PC3, respectively. Analyses are adjusted for age, body mass index, sex, smoking, diabetes, use of antineoplastic and immunomodulating agents, systolic blood pressure and self-reported antihypertensive medication. The dashed line represents a hazard ratio of 1 set at a median PC value. HR hazard ratio. The figure has been reproduced under the CC BY 4.0 licence.

status, systolic blood pressure, prevalent diabetes, antihypertensive medication, and use of antineoplastic or immunomodulating agents.

Results: While the often-used alpha diversity was not a statistically significant predictor for mortality, beta diversity measured with the principal component loadings produced a robust signal. The third principal component, PC3, was strongly associated with all-cause mortality risk, with a hazard ratio (HR) of 1.14, confidence interval (CI) 1.07-1.23 and *P*-value = 0.001 (see Figure 8). Investigating the drivers of PC3, we found that the axis was largely driven by genera of the Enterobacteriaceae family, many of which are known pathogens [128]. The association could also be observed individually in Eastern and Western Finnish populations, groups that differ genetically and in lifestyle factors. Enterobacteriaceae abundance and PC3 were also related to cause-specific mortality, with particularly strong predictive signals for gastrointestinal and respiratory causes.

Moreover, we found statistically significant associations when using individual genera and co-occurrence networks as predictors for mortality in the Cox regression and for the whole community in the random survival forest analysis. In each case, Enterobacteriaceae were strongly represented in the findings. Finally, we assessed the predictive power of microbial functions represented by Kegg Orthology groups. Here we found both positive and negative associations with mortality, for instance, in functions related to nutrient metabolism.

Discussion: In publication II, we present one of the largest prospective human gut microbiome data sets to date and provide the first instance of survival modeling based on microbiome features. Comparable data sets that include human microbiome samples with an extensive follow-up time and detailed health records have not been

published. Our work demonstrates, for the first time, that human gut microbiome features can function as biomarkers for general health status.

The study follows standard survival analysis procedures, and the main contributions are biomedical. From a methodological perspective, the study indicates that standard survival analysis tools are applicable in microbiome context. The supplementary of the publication provides a recommended workflow for microbiome-data-based survival analysis. However, methods further customized to the characteristics of microbiome data would likely give more robust results, and provide a direction for future research.

In the preliminary stages of the analysis, we also experimented with different methodological approaches. We attempted using the whole community composition in a Cox model equipped with ridge, elastic net, and lasso regularizations [129]. The goal was to perform variable selection that would reveal microbiome features important for mortality. However, we faced unresolved model convergence issues and opted to use the survival random forest algorithm instead for feature selection.

Author's contribution: The author participated in designing and executing the analysis workflow as a shared first author. This included method selection, implementation, and performing of the survival analyses, analyzing, interpreting, and reporting the results in the manuscript.

5.3 Publication III: Hierarchical stability analysis

Motivation: In microbiomics, several taxonomic units have been reported to maintain relatively stable long-term average abundance levels, despite considerable stochastic fluctuations [10; 18; 34]. This stability and variability have been linked to certain health outcomes [1], and the ability to measure these attributes reliably from microbiome time series could have material implications in clinical settings [9]. However, the complexity and limitations of microbiome data make it a challenging task to infer stability properties from taxon abundance time series, especially at commonly encountered sample sizes and traditional time series analysis methods. [5]

One potential solution would be to use data aggregation methods, where information from several related taxa can be combined to improve detection sensitivity [65]. This approach is based on the assumption that taxa may share certain behavioral characteristics [45], and by aggregating the parallel information, we could obtain improved detection sensitivity compared to processing the individual time series separately. With this in mind, the aim of the Publication III was to explore data aggregation methods within the Bayesian statistical framework that could be used to infer stability properties in related and parallel time series.

Methods: The methodology of the publication was based on the Student-t type Ornstein-Uhlenbeck process [130] (OUP, see Chapter 3), which we used to mea-

sure the mean-reversion rate and variance in collections of a time series data. We compared three levels of information aggregation: complete, partial, and no pooling. These methods refer to qualitatively different ways of assimilating related information with prior distributions [65]. In the case of complete pooling, the different time series are assumed to be generated by a single process, while no pooling assumes complete independence between the series. Partial pooling (a Bayesian hierarchical model), is an intermediate approach that assumes the series to be generated by distinct but related OU processes.

We compared the performance of the model variants in a simulation study based on limited time series. The performance of each model was assessed by investigating their ability to accurately recover the true simulation parameters.

Results The results of the simulation study showed that the partially pooled model was more effective in recovering the simulation parameters compared to the other model variants.

Discussion: The results demonstrate that data aggregation with partially pooled parameters can be an effective strategy for inferring properties of related time series. In addition to offering more sensitive inference on properties of individual series, the partially pooled model provides a general, summarizing overview of the system by estimating the population distributions of the parameters.

This sub-project of the thesis provided numerous ideas for further research.

While the publication focused on single-stable-state systems, an apparent extension would be to examine bistable systems, which involve systems that evolve in a potential landscape with potential wells. This type of systems can be approached in different ways. SDEs with a third-degree polynomial drift function can generate bistable dynamics, but they are more challenging to fit, and interpreting the parameters is not as straightforward [89].

Hidden Markov models (HMM), on the other hand, are based on a discrete latent process with two or more possible states that determine the properties of the system at a given time [131]. The properties of each state are separate and can consist of, for example, OUPs or a white noise processes. In fitting an HMM both the latent transition process and the properties of the states are learned. An interesting feature is that an HMM provides probabilities for state transitions, which could be used as an alternative for a stability metric called exit time, which is a measure of the average time a system spends in a given state before transitioning to another state. [101].

Another option would be to use SDEs without assuming a parametric form for the drift or the diffusion functions and to learn these from the data [132]. The drift and diffusion provide important information in and of itself, but they can also be used to compute the stationary distribution of the system. In clinical microbiome studies, for example, this could allow for the estimation of an individual's microbiome profile from the time series and a comparison with population distributions.

Finally, the project led to early warning signals (see Chapter 4), which are the

topic of publications IV and V.

Author's contribution: The author was responsible for the original idea, planning and conducting the experiments, interpreting the results, and writing the manuscript.

5.4 Publication IV: Early warning signals

Motivation: The objective of this sub-project of the thesis was to explore the potential benefits of formulating EWS in the Bayesian statistical framework. All previous EWS research had been conducted using the classical frequentist framework, and here our goal was to see if shifting away from this framework would help in improving signal detection accuracy in limited data. We hypothesized that imposing prior distributions could regularize the inference process by steering it away from unrealistic areas of the parameter space. Furthermore, we also hypothesized that the Bayesian framework's treatment of uncertainties would prove helpful in EWS detection.

Methods: We measured lag-1 autocorrelation trajectories in time series using three different autoregressive-1 process variants [84]. We did this by using regular non-probabilistic AR(1) process and its time-varying variant TVAR(1), in addition to a novel probabilistic TVAR(1), denoted pTVAR(1). In the pTVAR(1) model, we used Gaussian process (GP) priors with a Matérn-3/2 covariance kernel on the time-varying parameters μ_t and ϕ_t.

To measure the magnitude of the indicator trend, we used Kendall's rank correlation τ [115]. To test the hypothesis that $\tau > 0$ (positive EWS), we utilized surrogate data analysis methods for the non-probabilistic methods [116] and based the hypothesis testing on the posterior distribution in the probabilistic case. More specifically, we computed τ for each individual posterior sample for ϕ, and this way obtained the posterior of τ [65].

Results: We conducted a simulation benchmark study based on the stochastic population model presented in Section 4.3, in which we compared the proposed TVAR(1) model against the non-probabilistic AR(1) and TVAR(1) models. First, we tested the effect of the smoothing hyperparameter on inference and found that the pTVAR(1) model was more robust to this choice than the other models. Then, we used a large collection of simulated time series, both with induced state transitions and with stable conditions, and varying levels of observation noise, to assess the ability of the models to detect true and false signals. Our results showed that pTVAR(1) achieved the best overall performance, as measured with the F1 score that takes into account both precision and recall.

Finally, we applied the models to actual data from three previously published data sets where transitions between alternative stable states had occurred. In this demonstration, we could replicate previously reported positive EWS in a paleocli-

matic [113] and an experimental cyanobacteria population time series [133]. Furthermore, we found the first preliminary evidence of EWS in a human gut microbiome time series [34].

Discussion: The publication introduces (to our knowledge) the first Bayesian formulation of EWS detection. Additionally, using EWS methods to classify collections of time series and measuring the performance with standard classification metrics is a novel approach. The motivation arose from the discrepancy between what the available EWS detection techniques require in terms of data quantity and quality and what is typically available in real-world scenarios [14]. This is the case, especially in microbiomics, where time series data is limited and the existence of EWS has been challenging to establish [60].

The presented pTVAR(1) model showed robustness to hyperparameter choices and superior overall performance in the simulation experiments implying that the Bayesian model is helpful in practice. One key advantage of the pTVAR(1) model is the ability to perform hypothesis testing without the need for surrogate data methods, simplifying the analysis pipeline and reducing the number of modeling assumptions. Additionally, the time-varying formulation eliminates the need for the sliding window approach used in most other EWS methods.

While the results of the experiments were encouraging, further analysis is necessary to fully establish the utility of the proposed model. This includes more comprehensive simulation experiments using different models and a more extensive evaluation of the robustness, for instance, to missing observations, different noise structures, and count type data [134]. While the results of the experiments were encouraging, further analysis is necessary to fully establish the utility of the proposed model. This includes more comprehensive simulation experiments using different models and a more extensive evaluation of the robustness, for instance, to missing observations, different noise structures, and count type data [108; 134].

The proposed pTVAR(1) model could also be developed further. Including observation error modeling in the pTVAR(1) process could improve the performance of the model. Theoretically, this is a simple addition but results in a state-space model, which can be challenging to fit even in simple cases [135]. The GP hyperparameter selection process could be another target for improvement [90]. Although we show that the choices have less impact on the analysis conclusions compared to the non-probabilistic models, making the selection process automatic would further reduce subjectivity in the analysis.

The extension to the continuous time domain could be formulated with non-stationary GPs with time-varying parameters [93]. The length-scale parameter of a non-stationary OU process would work as an EWS indicator, as it is analogous with the autoregressive parameter in the pTVAR(1) process. We implemented this model during the project but encountered unresolved convergence issues in MCMC sampling.

Finally, another line of study could be to formulate other than autocorrelation-based indicators in the probabilistic framework, although the generative processes may not be as apparent.

Author's contribution: The author was responsible for the conceptualization of the work, design and implementation of the experiments, interpretation of results, and writing of the manuscript.

5.5 Publication V: Multivariate early warning signals

Motivation: In natural systems, features such as the abundance of an animal species are intricately connected to a multitude of environment and other variables, such as competing species or availability of resources. Simply reducing such a complex system to a single variable causes a loss of information, and a more holistic analysis would therefore be preferable. With this in mind, we investigated probabilistic EWS in the multivariate context [136]. Encouraged by the results of Publication IV, we decided to study extensions of the pTVAR(1) model into the multivariate domain.

Methods: In Publication V, we developed a multivariate extension of the time-varying autoregressive-1 process and compared it against previously presented autocorrelation-based multivariate early warning indicators in a simulation benchmark. The time-varying probabilistic vector AR(1) process, or tvPVAR(1) is defined as $\mathbf{X}_{t+1} = \Phi_t \cdot \mathbf{X}_t + \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t$ is a multivariate Gaussian random variable with diagonal covariance matrix [84]. For simplicity, we assumed the target variable, Φ_t, to be of the form $\phi_t I$, where I is the identity matrix and ϕ_t a real-valued function. In line with Publication IV, we used a GP prior with a Matérn-3/2 covariance kernel for ϕ_t.

The previously published autocorrelation-based indicators included in the comparison were maximum node autocorrelation, average node autocorrelation [137], degenerate fingerprinting [138], and eigenvalues of min/max autocorrelation factor analysis [139]. The simulations were based on a stochastic generalized LV model [140] that can describe the dynamics of a community with competition and mutualism.

Results: In the first part of the experiments, we evaluated the ability of the models to classify data based on EWS detected in time series data. To test the performance under different conditions, we generated collections of time series where we altered dimensionality, observation error, and sample size. These data included instances both with and without induced state transitions. The results indicated that the proposed tvPVAR(1) model consistently outperformed the other models in terms of true positive rate (see Figure 9). However, we found no statistically significant differences in the true negative rate among the models.

In the next phase, we conducted a sensitivity analysis to assess the impact of hyperparameter choices on the model performance. This involved performing a grid

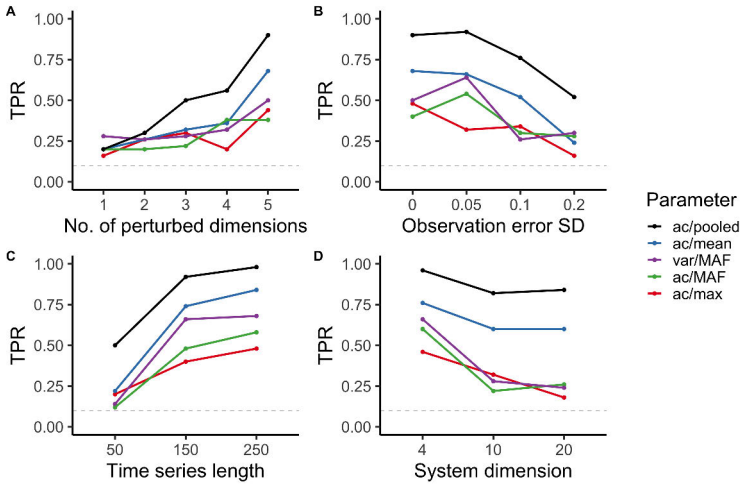

Figure 9. EWS classification accuracy measured in true positive rate (TPR). Changes in the number of perturbed dimensions (out of 10) **A**, level of Gaussian observation error **B**, time series length **C** and total system dimension **D** affect TPR. The tvPVAR(1) based indicator was named ac/pooled (black lines). The horizontal grey line marks the level of TPR for a random guess. The figure has been reproduced under the Springer Nature licence 5507031502591.

search for the smoothing bandwidth and length scale/sliding window length. The results revealed that, as in Publication IV, the probabilistic model was more robust to hyperparameter choices than the non-probabilistic variants. This indicates that the presented probabilistic model performed more consistently and reliably that the previously presented approaches.

Discussion: The final publication of the thesis introduces a probabilistic EWS indicator for multivariate systems. This approach, to the best of our knowledge, has never been presented before and the results from our experiments demonstrate its practical value.

Initially, we attempted to implement a more general probabilistic vector AR(1) model that would not impose such strong restrictions on the entries of Φ_t. In the most general case, we assigned GP priors to all the entries. However, while this variant behaved well when with low-dimensional systems $(D \leq 3)$, it scaled poorly as the dimension increased.

Next, we restricted the off-diagonal elements to zero and experimented with a different diagonal structures, including independent entries and a composite of a shared and individual dimension-specific process. However, while these attempts were showed some degree of success, the final model used in the publication demonstrated superior performance. This could be due to the relatively simple model structure which avoids overfitting in low-sample-size data.

The project sparked ideas for future research on probabilistic multivariate EWS indicators. In particular, probabilistic dimension reduction techniques could be in-

teresting to study in this context. For instance, principal component analysis (PCA) that finds the direction of the largest variance in the state space has been used in this context [141], and the probabilistic PCA might be more sensitive, because it can decompose variation into actual and technical variation [142].

Author's contribution: The author conceived the idea for the work, and was responsible for the design, and implementation of the experiments, the interpretation of the results, and writing of the manuscript.

6 Conclusion

The main objective of this thesis has been to explore the use of statistical models in the analysis of the temporal dimension from both the prospective and longitudinal perspectives. Throughout the research, the central focus was the prediction of future events based on current information, specifically regarding the risk of extreme events in the future, such as all-cause mortality in humans and critical state transitions in time series.

In the first part, the goal was to assess the applicability of established models in the context of prospective microbiome data. As data with longer follow-up times are becoming more abundant in microbiome research, the ability to analyze such data is becoming increasingly important. To tackle this challenge, we introduced the use of survival analysis methodology in this context and demonstrated its effectiveness in providing reliable and robust results. Our work showcased, for the first time, the potential of this approach in the field.

In the second part of the thesis, we took a complementary approach to temporal data analysis by shifting the focus to time series. The attention was centered on the stability properties of dynamical systems, which we investigated using probabilistic methods. To shed light on this topic, three methods were developed for measuring mean-reversion rate, a statistic that can be used as a metric for systemic resilience. These methods provide the first probabilistic treatment of measuring changes in autocorrelation and use the recovered findings to detect early warning signals for critical transitions in complex systems. By utilizing key aspects of the probabilistic framework, such as hierarchical model structure, regularization with prior distributions, and principled uncertainty management, more sensitive and robust results were obtained in data with common limitations: low sample size and high levels of noise. Although the work was motivated by questions in microbiomics, the methodology is generic and can be applied in a variety of contexts.

In conclusion, this thesis has demonstrated the application and further development of quantitative methods for prospective microbiome data and time series analysis with limited data. The results provide evidence for the potential of probabilistic modeling in the stability analysis of complex systems and early warning signals and contribute to the practices of microbiome data science. Building on the findings of this thesis offers ample opportunities for future research.

List of References

[1] Gilbert Jack, Martin Blaser, J. Gregory Caporaso, Janet Jansson, Susan Lynch, and Rob Knight. Current understanding of the human microbiome. *Nature Medicine*, 24:392–400, 2018.

[2] Matthew Jackson, Serena Verdi, Maria-Emanuela Maxan, Cheol Shin, Jonas Zierer, Ruth Bowyer, Tiphaine Martin, Frances Williams, Cristina Menni, Jordana Bell, Tim Spector, and Claire Steves. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, 9, 2018.

[3] Susan Holmes and Wolfgang Huber. *Modern Statistics for Modern Biology*. Cambridge University Press, 2019. ISBN 9781108705295.

[4] Stan Development Team. RStan: the R interface to Stan, 2020. R package version 2.21.2.

[5] Isabel Moreno-Indias, Leo Lahti, Miroslava Nedyalkova, Ilze Elbere, Gennady Roshchupkin, Muhamed Adilovic, Onder Aydemir, Burcu Bakir-Gungor, Enrique Carrillo-de Santa Pau, Domenica D'Elia, Mahesh S. Desai, Laurent Falquet, Aycan Gundogdu, Karel Hron, Thomas Klammsteiner, Marta B. Lopes, Laura Judith Marcos-Zambrano, Cáudia Marques, Michael Mason, Patrick May, Lejla Pačić, Gianvito Pio, Sándor Pongor, Vasilis J. Promponas, Piotr Przymus, Julio Saez-Rodriguez, Alexia Sampri, Rajesh Shigdel, Blaz Stres, Ramona Suharoschi, Jaak Truu, Ciprian-Octavian Truică, Baiba Vilne, Dimitrios Vlachakis, Ercument Yilmaz, Georg Zeller, Aldert L. Zomer, David Gómez-Cabrero, and Marcus J. Claesson. Statistical and machine learning techniques in human microbiome studies: Contemporary challenges and solutions. *Frontiers in Microbiology*, 12, 2021.

[6] Junghyun Namkung. Machine learning methods for microbiome studies. *Journal of Microbiology*, 58:206–216, 2020.

[7] Tommi Vatanen, Eric A. Franzosa, Randall Schwager, Surya Tripathi, Timothy D. Arthur, Kendra Vehik, Åke Lernmark, William A. Hagopian, Marian J. Rewers, Jin-Xiong She, et al. The human gut microbiome in early-onset type 1 diabetes from the teddy study. *Nature*, 562 (7728):589–594, 2018.

[8] Mirjana Rajilić-Stojanović, Hans GHJ Heilig, Sebastian Tims, Erwin G. Zoetendal, and Willem M. de Vos. Long-term monitoring of the human intestinal microbiota composition. *Environmental microbiology*, 15(4):1146–1159, 2013.

[9] Jonas Halfvarson, Colin J. Brislawn, Regina Lamendella, Yoshiki Vázquez-Baeza, William A. Walters, Lisa M. Bramer, Mauro D'Amato, Ferdinando Bonfiglio, Daniel McDonald, Antonio Gonzalez, Erin E. McClure, Mitchell F. Dunklebarger, Rob Knight, and Janet K. Jansson. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology*, 2 (5):17004, 2017.

[10] Lianmin Chen, Daoming Wang, Sanzhima Garmaeva, Alexander Kurilshikov, Arnau Vich Vila, Ranko Gacesa, Trishla Sinha, Eran Segal, Rinse Weersma, Cisca Wijmenga, Alexandra Zhernakova, and Jingyuan Fu. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*, 184:2302–2315.e12, 2020.

[11] Linnea Karlsson, Mimmi Tolvanen, Noora Scheinin, Henna-Maria Uusitupa, Riikka Korja, Eeva Ekholm, Jetro Tuulari, Marjukka Pajulo, Minna Huotilainen, Tiina Paunio, and Hasse Karlsson. Cohort profile: The finnbrain birth cohort study (FinnBrain). *International Journal of Epidemiology*, 47, 2017.

[12] Katja Borodulin, Erkki Vartiainen, Markku Peltonen, Pekka Jousilahti, Anne Juolevi, Tiina Laatikainen, Satu Männistö, Veikko Salomaa, Jouko Sundvall, and Pekka Puska. Forty-year trends in cardiovascular risk factors in Finland. *The European Journal of Public Health*, 25(3): 539–546, 2015.

[13] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.

[14] Alex Arkilanian, Christopher F. Clements, Arpat Ozgul, and Gaurav Baruah. Effect of time series length and resolution on abundance- and trait-based early warning signals of population declines. *Ecology*, 101(7):e03040, 2020.

[15] Marten Scheffer. *Critical Transitions in Nature and Society*. Princeton University Press, New Jersey, United States, 2009.

[16] Leo Lahti, Jarkko Salojärvi, Anne Salonen, Marten Scheffer, and Willem M. de Vos. Tipping elements in the human intestinal ecosystem. *Nature Communications*, 5:4344, 2014.

[17] Gabriele Berg, Daria Rybakova, Doreen Fischer, Tomislav Cernava, Marie-Christine Vergés, Trevor Charles, Xiaoyulong Chen, Luca Cocolin, Kellye Eversole, Gema Corral, Maria Kazou, Linda Kinkel, Lene Lange, Nelson Lima, Alexander Loy, James Macklin, Emmanuelle Maguin, Tim Mauchline, Ryan McClure, and Michael Schloter. Microbiome definition re-visited: Old concepts and new challenges. *Microbiome*, 8, 2020.

[18] J. Gregory Caporaso, Christian Lauber, Elizabeth Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, Noah Fierer, Jeffrey Gordon, and Rob Knight. Moving pictures of the human microbiome. *Genome Biology*, 12:R50, 2011.

[19] Sigal Leviatan, Saar Shoer, Daphna Rotshchild, Maria Gorodetski, and Eran Segal. An expanded reference map of the human gut microbiome reveals hundreds of previously unknown species. *Nature Communications*, 13:3863, 2022.

[20] Gwen Falony, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, Marc Jan Bonder, Mireia Valles-Colomer, Doris Vandeputte, Raul Tito Tadeo, Samuel Chaffron, Leen Rymenans, Chloë Verspecht, Lise Sutter, Gipsi Lima-Mendez, Kevin D'hoe, Karl Jonckheere, Daniel Homola, and Jeroen Raes. Population-level analysis of gut microbiome variation. *Science*, 352:560–564, 2016.

[21] Fabrizio De Luca and Yehuda Shoenfeld. The microbiome in autoimmune diseases. *Clinical Experimental Immunology*, 195, 2018.

[22] Therese Limbana, Farah Khan, and Noha Eskander. Gut microbiome and depression: How microbes affect the way we think. *Cureus*, 12, 2020.

[23] Gregory D. Sepich-Poore, Laurence Zitvogel, Ravid Straussman, Jeff Hasty, Jennifer A. Wargo, and Rob Knight. The microbiome and human cancer. *Science*, 371(6536):eabc4552, 2021.

[24] Kerri Glassner, Bincy Abraham, and Eamonn Quigley. The microbiome and inflammatory bowel disease. *Journal of Allergy and Clinical Immunology*, 145:16–27, 2020.

[25] Julia Goodrich, Jillian Waters, Angela Poole, Jessica Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, Willian Treuren, Rob Knight, Jordana Bell, Tim Spector, Andrew Clark, and Ruth Ley. Human genetics shape the gut microbiome. *Cell*, 159:789–799, 2014.

[26] Yuka Moroishi, Jiang Gui, Kari C. Nadeau, Hilary G. Morrison, Juliette Madan, and Margaret R. Karagas. A prospective study of the infant gut microbiome in relation to vaccine response. *Pediatric Research*, 26:1–7, 2022.

[27] Qin Liu, Joyce Wing Yan Mak, Qi Su, Yun Kit Yeoh, Grace Chung-Yan Lui, Susanna So Shan Ng, Fen Zhang, Amy YL Li, Wenqi Lu, David Shu-Cheong Hui, et al. Gut microbiota dynamics in a prospective cohort of patients with post-acute covid-19 syndrome. *Gut*, 71(3):544–552, 2022.

[28] Yaohua Yang, Jinshun Long, Cong Wang, William J. Blot, Zhiheng Pei, Xiang Shu, Fen Wu, Nathaniel Rothman, Jie Wu, Qing Lan, Qiuyin Cai, Wei Zheng, Yu Chen, and Xiao-Ou Shu. Prospective study of oral microbiome and gastric cancer risk among Asian, African American and European American populations. *International Journal of cancer*, 150(6):916–927, 2022.

[29] Sara Vieira-Silva, Gwen Falony, Youssef Darzi, Gipsi Lima-Mendez, Roberto Yunta, Shujiro Okuda, Doris Vandeputte, Mireia Valles-Colomer, Falk Hildebrand, Samuel Chaffron, and Jeroen Raes. Species-function relationships shape ecological properties of the human gut microbiome. *Nature Microbiology*, 1:16088, 2016.

[30] Naomi Iris van den Berg, Daniel Machado, Sophia Santos, Isabel Rocha, Jeremy Chacón, William Harcombe, Sara Mitri, and Kiran R. Patil. Ecological modelling approaches for predicting emergent properties in microbial communities. *Nature Ecology & Evolution*, 6:1–11, 2022.

[31] Julia Goodrich, Jillian Waters, Angela Poole, Jessica Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, Willian Treuren, Rob Knight, Jordana Bell, Tim Spector, Andrew Clark, and Ruth Ley. Human genetics shape the gut microbiome. *Cell*, 159:789–799, 2014.

[32] Owen Qin, Aki Havulinna, Yang Liu, Pekka Jousilahti, Scott Ritchie, Alex Tokolyi, Jon Sanders, Liisa Valsta, Marta Brożyńska, Qiyun Zhu, Anupriya Tripathi, Yoshiki Vázquez-Baeza, Rohit Loomba, Susan Cheng, Mohit Jain, Teemu Niiranen, Leo Lahti, Rob Knight, Veikko Salomaa, and Guillaume Méric. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nature Genetics*, 54:1–9, 2022.

[33] Ashley Kates, Omar Jarrett, Joseph Skarlupka, Ajay Sethi, Megan Duster-Matz, Lauren Watson, Garret Suen, Keith Poulsen, and Nasia Safdar. Household pet ownership and the microbial diversity of the human gut microbiota. *Frontiers in Cellular and Infection Microbiology*, 10, 2020.

[34] Lawrence A. David, Arne C. Materna, Jonathan Friedman, Maria I. Campos-Baptista, Matthew C. Blackburn, Allison Perrotta, Susan E. Erdman, and Eric J. Alm. Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, 15:R89, 2014. ISSN 1474-760X.

[35] Shohei Akagawa, Shoji Tsuji, Chikushi Onuma, Yuko Akagawa, Tadashi Yamaguchi, Mitsuru Yamagishi, Sohsaku Yamanouchi, Takahisa Kimata, Shin-ichiro Sekiya, Atsushi Ohashi, Masaki Hashiyada, Atsushi Akane, and Kazunari Kaneko. Effect of delivery mode and nutrition on gut microbiota in neonates. *Annals of Nutrition and Metabolism*, 74:132–139, 2019.

[36] Danping Zheng, Timur Liwinski, and Eran Elinav. Interaction between microbiota and immunity in health and disease. *Cell Research*, 30:492–506, 2020.

[37] Rachel Moloney, Lieve Desbonnet, Gerard Clarke, Timothy Dinan, and John Cryan. The microbiome: Stress, health and disease. *Mammalian genome : official journal of the International Mammalian Genome Society*, 25, 2013.

[38] Maria Bauer, Katharina Kainz, Didac Carmona-Gutierrez, and Frank Madeo. Microbial wars: Competition in ecological niches and within the microbiome. *Microbial Cell*, 5:215–219, 2018.

[39] Didier Gonze, Leo Lahti, Jeroen Raes, and Karoline Faust. Multi-stability and the origin of microbial community types. *The ISME Journal*, 11(10):2159, 2017. ISSN 1751-7370.

[40] José E. Belizário and Joel Faintuch. *Microbiome and Gut Dysbiosis*, pages 459–476. Springer International Publishing, Cham, 2018. ISBN 978-3-319-74932-7.

[41] Manimozhiyan Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel Mende, Gabriel Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, Marcelo Bertalan, Natalia Borruel, Francesc Casellas, Leyden Fernandez, Laurent Gautier, Torben Hansen, Masahira Hattori, Tetsuya Hayashi, Michiel Kleerebezem, and Peer Bork. Enterotypes of the human gut microbiome. *Nature*, 473:174–80, 2011.

[42] Paul Costea, Falk Hildebrand, Manimozhiyan Arumugam, Fredrik Bäckhed, Martin Blaser, Frederic Bushman, Willem M. De Vos, S. Dusko Ehrlich, Claire Fraser, Masahira Hattori, Curtis Huttenhower, Ian Jeffery, Dan Knights, James Lewis, Ruth Ley, Howard Ochman, Paul O'Toole, Christopher Quince, David Relman, and Peer Bork. Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, 3, 2017.

[43] Dan Knights, Tonya Ward, Christopher McKinlay, Hannah Miller, Antonio Gonzalez, Daniel Mcdonald, and Rob Knight. Rethinking "enterotypes". *Cell Host & Microbe*, 16:433–437, 2014.

[44] Ashley Shade, Hannes Peter, Steven Allison, Didier Baho, Mercé Berga, Helmut Bürgmann, David Huber, Silke Langenheder, Jay Lennon, Jennifer Martiny, Kristin Matulich, Thomas Schmidt, and Jo Handelsman. Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, 3:417, 2012.

[45] Amir Bashan, Travis E. Gibson, Jonathan Friedman, Vincent J. Carey, Scott T. Weiss, Elizabeth L. Hohmann, and Yang-Yu Liu. Universality of human microbial dynamics. *Nature*, 534 (7606):259–262, 2016. ISSN 1476-4687.

[46] Rob Knight, Alison Vrbanac, Bryn C. Taylor, Alexander Aksenov, Chris Callewaert, Justine Debelius, Antonio Gonzàlez, Tomasz Kosciolek, Laura-Isobel Mccall, Daniel Mcdonald, Alexey Melnik, James Morton, Jose Navas, Robert Quinn, Jon Sanders, Austin Swafford, Luke Thompson, Anupriya Tripathi, Zhenjiang Xu, and Pieter Dorrestein. Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7):410–422, 2018.

[47] Laura Marcos-Zambrano, Kanita Karaduzovic-Hadziabdic, Tatjana Tutukalo, Piotr Przymus, Vladimir Trajkovik, Oliver Aasmets, Magali Berland, Aleksandra Gruca, Jasminka Hasic Telalovic, Karel Hron, Thomas Klammsteiner, Mikhail Kolev, Leo Lahti, Marta Lopes, Victor Moreno, Irina Naskinova, Elin Org, Inês Paciência, Georgios Papoutsoglou, and Jaak Truu. Applications of machine learning in human microbiome studies: A review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in Microbiology*, 12:313, 2021.

[48] Michael Greenacre. Compositional data analysis. *Annual Review of Statistics and Its Application*, 8(1):271–299, 2021.

[49] Pratheepa Jeganathan and Susan P. Holmes. A statistical perspective on the challenges in molecular microbial biology. *Journal of Agricultural, Biological and Environmental Statistics*, 26(2): 131–160, 2021.

[50] Tarmo Äijö, Christian L. Müller, and Richard Bonneau. Temporal probabilistic modeling of bacterial compositions derived from 16s rrna sequencing. *Bioinformatics*, 34(3):372–380, 2018.

[51] Travis Gibson, Younhun Kim, Sawal Acharya, David Kaplan, Nicholas DiBenedetto, Richard Lavin, Bonnie Berger, Jessica Allegretti, Lynn Bry, and Georg Gerber. Intrinsic instability of the dysbiotic microbiome revealed through dynamical systems inference at scale. *bioRxiv 2021.12.14.469105*, 2021.

[52] Paul J. McMurdie and Susan P. Holmes. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLOS Computational Biology*, 10(4):1–12, 2014.

[53] George Armstrong, Gibraan Rahman, Cameron Martino, Daniel Mcdonald, Antonio Gonzalez, Gal Mishne, and Rob Knight. Applications and comparison of dimensionality reduction methods for microbiome data. *Frontiers in Bioinformatics*, 2, 2022.

[54] Didier Gonze, Katharine Z. Coyte, Leo Lahti, and Karoline Faust. Microbial communities as dynamical systems. *Current Opinion in Microbiology*, 44:41–49, 2018.

[55] Ian Holmes, Keith Harris, and Christopher Quince. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLOS ONE*, 7:e30126, 2012.

[56] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[57] Ricardo Medina, Svetlana Kutuzova, Knud Nor Nielsen, Joachim Johansen, Lars Hansen, Mads Nielsen, and Simon Rasmussen. Machine learning and deep learning applications in microbiome research. *ISME Communications*, 2:98, 2022. doi: 10.1038/s43705-022-00182-9.

[58] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. 2019. ISBN 9780203753736.

[59] Jacob Nearing, Gavin Douglas, Molly Hayes, Jocelyn MacDonald, Dhwani Desai, Nicole Allward, Casey Jones, Robyn Wright, Akhilesh Dhanani, André Comeau, and Morgan Langille. Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, 13, 2022.

[60] Ashley Coenen, Sarah Hu, Elaine Luo, Daniel Muratore, and Joshua Weitz. A primer for microbiome time-series analysis. *Frontiers in Genetics*, 11, 2020.

[61] David A. Freedman, Robert Pisani, and Roger Purves. *Statistics*. W.W. Norton, 1998. ISBN 9780393970838.

[62] Tony O'Hagan. Dicing with the unknown. *Significance*, 1(3):132–133, 2004.

[63] Sander Greenland, Stephen Senn, Kenneth Rothman, John Carlin, Charles Poole, Steven Goodman, and Douglas Altman. Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31, 2016.

[64] Megan L. Head, Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. The extent and consequences of p-hacking in science. *PLOS Biology*, 13(3):1–15, 2015.

[65] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013. ISBN 9781439840955.

[66] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate Bayesian computation. *PLOS Computational Biology*, 9:e1002803, 2013.

[67] David Blei, Alp Kucukelbir, and Jon McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112, 2016.

[68] Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434*, 2017.

[69] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992. ISSN 08834237.

[70] William Link and Mitchell Eaton. On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3:112–115, 2011.

[71] Wilfred K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. ISSN 00063444.

[72] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[73] George Casella and Edward I. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992. ISSN 00031305.

[74] Simon Duane, Anthony D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987. ISSN 0370-2693.

[75] John Ashworth Nelder and Robert W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[76] Michael Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, page 550, 2014.

[77] David R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972. ISSN 00359246.

[78] Ping Wang, Yan Li, and Chandan K. Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 51(6), 2019. ISSN 0360-0300.

[79] Frank E. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer Series in Statistics. Springer International Publishing, 2015. ISBN 9783319194240.

[80] Lynn A. Sleeper and David P. Harrington. Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412): 941–949, 1990.

[81] Zhongheng Zhang, Jaakko Reinikainen, Kazeem Adedayo Adeleke, Marcel E. Pieterse, and Catharina G. M. Groothuis-Oudshoorn. Time-varying covariates and coefficients in Cox regression models. *Annals of translational medicine*, 6(7), 2018.

[82] Hemant Ishwaran, Udaya Kogalur, Eugene Blackstone, and Michael Lauer. Random survival forests. *The Annals of Applied Statistics*, 2, 2008.

[83] Daniel J. Stekhoven and Peter Bühlmann. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2011. ISSN 1367-4803.

[84] Robert Shumway and David Stoffer. *Time Series Analysis and Its Applications With R Examples*, volume 9. 2011. ISBN 978-1-4419-7864-6.

[85] Jianqing Fan and Wenyang Zhang. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1:179–195, 2008.

[86] Jonas Haslbeck, Laura Bringmann, and Lourens Waldorp. A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, 56:1–30, 2020.

[87] Anthony Ives and Stephen R. Carpenter. Stability and diversity of ecosystems. *Science*, 317: 58–62, 2007.

[88] Irfan Pratama, Adhistya Erna Permanasari, Igi Ardiyanto, and Rini Indrayani. A review of missing values handling methods on time-series data. In *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*, pages 1–6. IEEE, 2016.

[89] Stefano M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples (Springer Series in Statistics)*. Springer Publishing Company, Incorporated, 1 edition, 2008. ISBN 0387758380, 9780387758381.

[90] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. ISBN 026218253X.

[91] Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer-Verlag, New York, 1999. ISBN 0-387-98629-4.

[92] Marc Genton. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2:299–312, 2001.

[93] Markus Heinonen, Henrik Mannerström, Juho Rousu, Samuel Kaski, and Harri Lähdesmäki. Non-Stationary Gaussian Process Regression with Hamiltonian Monte Carlo. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, PLMR*, 51:732–740, 2016.

[94] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*, volume 82. 2000. ISBN 978-3-540-60243-9.

[95] Toby Kenney, Junqiu Gao, and Hong Gu. Application of OU processes to modelling temporal dynamics of the human microbiome, and calculating optimal sampling schemes. *BMC Bioinformatics*, 21:450, 2020.

[96] Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

[97] Stephen Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, volume 4. 2003. ISBN 0-387-00177-8.

[98] Alfred J. Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3):271–274, 1910.

[99] Richard R. Stein, Vanni Bucci, Nora C. Toussaint, Charlie G. Buffie, Gunnar Rätsch, Eric G. Pamer, Chris Sander, and João B. Xavier. Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLOS Computational Biology*, 9(12): 1–11, 2013.

[100] Beatrix Beisner, Daniel Haydon, and Kim Cuddington. Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, 1(7):376–382, 2003.

[101] Babak M. S. Arani, Stephen R. Carpenter, Leo Lahti, Egbert H. van Nes, and Marten Scheffer. Exit time as a measure of ecological resilience. *Science*, 372(6547):eaay4895, 2021. ISSN 0036-8075.

[102] Vasilis Dakos and Sonia Kéfi. Ecological resilience: What to measure and how. *Environmental Research Letters*, 17, 2022.

[103] Ian Donohue, Helmut Hillebrand, José M. Montoya, Owen L. Petchey, Stuart L. Pimm, Mike S. Fowler, Kevin Healy, Andrew L. Jackson, Miguel Lurgi, Deirdre McClean, Nessa E. O'Connor, Eoin J. O'Gorman, and Qiang Yang. Navigating the complexity of ecological stability. *Ecology Letters*, 19(9):1172–1185, 2016.

[104] Virginia Domínguez-García, Vasilis Dakos, and Sonia Kéfi. Unveiling dimensions of stability in complex ecological networks. *Proceedings of the National Academy of Sciences*, 116: 201904470, 2019.

[105] Stuart Pimm. The complexity and stability of ecosystems. *Nature*, 307:321–326, 1984.

[106] Marten Scheffer, Stephen R. Carpenter, Timothy M. Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan van de Koppel, Ingrid A. van de Leemput, Simon A. Levin, Egbert H. van Nes, Mercedes Pascual, and John Vandermeer. Anticipating critical transitions. *Science*, 338 (6105):344–348, 2012.

[107] Vasilis Dakos, Stephen R. Carpenter, William A. Brock, Aaron M. Ellison, Guttal Vishwesha, Anthony R. Ives, Sonia Kéfi, Valerina Livina, David A. Seekell, Egbert H. van Nes, and Scheffer Marten. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLOS ONE*, 7(7):1–20, 2012.

[108] Christopher F. Clements and Arpat Ozgul. Indicators of transitions in biological systems. *Ecology Letters*, 21(6):905–919, 2018.

[109] Haoyu Wen, Massimo Pica Ciamarra, and Siew Ann Cheong. How one might miss early warning signals of critical transitions in time series data: A systematic study of two major currency pairs. *PLOS ONE*, 13(3):1–22, 2018.

[110] Robert M. May. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, 269(5628):471–477, 1977.

[111] Marten Scheffer, Steve Carpenter, Jonathan A. Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596, 2001.

[112] Timothy M. Lenton, Valerie N. Livina, Vasilis Dakos, Egbert H. van Nes, and Marten Scheffer. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962):1185–1204, 2012.

[113] Vasilis Dakos, Marten Scheffer, Egbert H. van Nes, Victor Brovkin, Vladimir Petoukhov, and Hermann Held. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, 105(38):14308–14312, 2008. ISSN 0027-8424.

[114] Christopher F. Clements, John M. Drake, Jason I. Griffiths, and Arpat Ozgul. Factors influencing the detectability of early warning signals of population collapse. *The American Naturalist*, 186 (1):50–58, 2015. ISSN 00030147, 15375323.

[115] Shiyang Chen, Amin Ghadami, and Bogdan I. Epureanu. Practical guide to using Kendall's tau in the context of forecasting critical transitions. *Royal Society Open Science*, 9(7):211346, 2022.

[116] James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992.

[117] Els Weinans, Rick Quax, Egbert H. van Nes, and Ingrid A. van de Leemput. Evaluating the performance of multivariate indicators of resilience loss. *Scientific Reports*, 11(1):9148, 2021.

[118] Christopher F. Clements and Arpat Ozgul. Rate of forcing and the forecastability of critical transitions. *Ecology and Evolution*, 6(21):7787–7793, 2016.

[119] Partha Sharathi Dutta, Yogita Sharma, and Karen C. Abbott. Robustness of early warning signals for catastrophic and non-catastrophic transitions. *Oikos*, 127(9):1251–1263, 2018.

[120] Vasilis Dakos and Fernando Soler-Toscano. Measuring complexity to infer changes in the dynamics of ecological systems under stress. *Ecological Complexity*, 32, 2016.

[121] Alan Hastings and Derin B. Wysham. Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, 13(4):464–472, 2010.

[122] Anthony R. Ives and Vasilis Dakos. Detecting dynamical changes in nonlinear time series using locally linear state-space models. *Ecosphere*, 3(6):58, 2012.

[123] Catherine Monk, Claudia Lugo-Candelas, and Caroline Trumpff. Prenatal developmental origins of future psychopathology: Mechanisms and pathways. *Annual review of clinical psychology*, 15:317, 2019.

[124] Eldin Jašarević, Christopher D. Howard, Ana M. Misic, Daniel P. Beiting, and Tracy L. Bale. Stress during pregnancy alters temporal and spatial dynamics of the maternal and offspring microbiome in a sex-specific manner. *Scientific Reports*, 7(1):1–13, 2017.

[125] Christine Offergeld Née Hechler, Klaudyna Borewicz, Roseriet Beijers, Edoardo Saccenti, M. Riksen-Walraven, Hauke Smidt, and Carolina Weerth. Association between psychosocial stress and fecal microbiota in pregnant women. *Scientific Reports*, 9(1):1–10, 2019.

[126] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289 – 300, 1995.

[127] Zachary D. Kurtz, Christian L. Müller, Emily R. Miraldi, Dan R. Littman, Martin J. Blaser, and Richard A. Bonneau. Sparse and compositionally robust inference of microbial ecological networks. *PLOS Computational Biology*, 11(5):e1004226, 2015.

[128] Claudia Lupp, Marilyn L. Robertson, Mark E. Wickham, Inna Sekirov, Olivia L. Champion, Erin C. Gaynor, and B. Brett Finlay. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of enterobacteriaceae. *Cell host & microbe*, 2(2):119–129, 2007.

[129] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

[130] Amar Shah, Andrew Gordon Wilson, and Zoubin Ghahramani. Student-t processes as alternatives to Gaussian processes. *The Seventeenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.

[131] Bhavya Mor, Sunita Garhwal, and Ajay Kumar. A systematic review of hidden Markov models and their applications. *Archives of computational methods in engineering*, 28(3):1429–1448, 2021.

[132] Cagatay Yildiz, Markus Heinonen, Jukka Intosalmi, Henrik Mannerström, and Harri Lähdesmäki. Learning stochastic differential equations with Gaussian processes without gradient matching. *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*.

[133] Annelies J. Veraart, Elisabeth J. Faassen, Vasilis Dakos, Egbert H. van Nes, Miquel Lürling, and Marten Scheffer. Recovery rates reflect distance to a tipping point in a living system. *Nature*, 481(7381):357–359, 2012.

[134] Vasilis Dakos, Egbert H. van Nes, Paolo D'Odorico, and Marten Scheffer. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, 93(2):264–271, 2012.

[135] Marie Auger-Méthé, Chris Field, Christoffer M. Albertsen, Andrew E. Derocher, Mark A. Lewis, Ian D. Jonsen, and Joanna Mills Flemming. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. *Scientific Reports*, 6(1):26677, 2016.

[136] Caleb P. Roberts, Dirac Twidwell, Jessica L. Burnett, Victoria M. Donovan, Carissa L. Wonkka, Christine L. Bielski, Ahjond S. Garmestani, David G. Angeler, Tarsha Eason, Brady W. Allred, Matthew O. Jones, David E. Naugle, Shana M. Sundstrom, and Craig R. Allen. Early warnings for state transitions. *Rangeland Ecology Management*, 71(6):659–670, 2018. ISSN 1550-7424.

[137] Vasilis Dakos. Identifying best-indicator species for abrupt transitions in multispecies communities. *Ecological Indicators*, 94:494–502, 2018.

[138] Hermann Held and Thomas Kleinen. Detection of climate system bifurcations by degenerate fingerprinting. *Geophysical Research Letters*, 312:L23207, 2004.

[139] Els Weinans, J. Jelle Lever, Sebastian Bathiany, Rick Quax, Jordi Bascompte, Egbert H. Nes, Marten Scheffer, and Ingrid van de Leemput. Finding the direction of lowest resilience in multivariate complex systems. *Journal of The Royal Society Interface*, 16:20190629, 2019.

[140] J. Jelle Lever, Egbert H. van Nes, Marten Scheffer, and Jordi Bascompte. The sudden collapse of pollinator communities. *Ecology letters*, 17:350–359, 2014.

[141] Shiyang Chen, Eamon O'Dea, John Drake, and Bogdan Epureanu. Eigenvalues of the covariance matrix as early warning signals for critical transitions in ecological systems. *Scientific Reports*, 9:2572, 2019.

[142] Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2023

ISBN 978-951-29-9431-1 (PRINT)
ISBN 978-951-29-9432-8 (Online)
ISSN 2736-9390 (PRINT)
ISSN 2736-9684 (Online)