Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Pure & Applied Mathematics (36^{th}cycle)

Statistical models for understanding biomedical data

By

Hiu Ching Yip

******

Supervisor(s):
Prof. Gianluca Mastrantonio, Supervisor
Prof. Enrico Bibbona, Co-Supervisor

Doctoral Examination Committee:

Politecnico di Torino
Tuesday 8th October, 2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Hiu Ching Yip
Tuesday 8th October, 2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Abstract

This work is a collection of statistical models that can be applied in biomedical research. It focuses on the development of statistical methods in the following research areas: cross-species comparison in bio-acoustic analysis, meta-analysis of randomized clinical trials and preferential attachment of an evolving network.

A statistical model is presented for each of the corresponding areas. In particular, these models and their implementation integrated novel techniques and methodologies in order to tackle challenges that are induced by the characteristics of real-world data sets. The first is a spatio-temporal model for bio-acoustic data that consists of periodic artifacts and is temporally non-stationary. The second is a modified proportional hazards model that can account for the impacts of the follow-up time of the trials in mixed treatment comparison. The third is a correction of a specific preferential attachment model that exhibit the power-law degree sequence which is empirically observed in real-world network data.

The models and the methodologies involved in the implementation have solid theoretical basis. More importantly, these models are useful quantitative tools for biomedical researchers to analyze and learn from data sets that possess the problematic characteristics encountered in this work. Extensive simulation studies and testings on real data sets were conducted.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Some history of mathematics in biology

The use of mathematics in research areas of other disciplines has a long history across the scientific communities. Mathematical biology and biostatistics are some of the most prominent fields that represent this type of inter-disciplinary research. However, the application of mathematical theories and theoretical tools for the purposes of understanding biological systems or decision-making in public healthcare management have been met with numerous challenges in terms of theoretical solidarity and technical practicality.

Topics that exemplify the use of mathematics in bio-medical research include the studies of evolutionary biology, statistically-based clinical trial studies and epidemic modeling. Firstly, the formal study of animal behaviors began with Darwin's evolutionary theory and the concept of natural selection. Subsequent empirical research on evolutionary paths of living organisms have since employed many mathematical tools with the most popular ones being the Poisson process for coalescent trees and the Wright-Fisher model for population genetics, among others. (Deonier et al., 20005; Durrett, 2008) Secondly, modern randomized controlled trials appear in the aftermath of World War II as a result of wartime scarcity, the development of new drugs for the war effort and the necessity of human trials. During the war, a series of human trials were conducted on over 170 patients to test the effectiveness of penicillin, which later became known as The Penicillin Project. (Lobanovska and Pilla, 2017) Following the success of the first randomized control trial that was conducted in 1946 to test antibiotics for tuberculosis, statistically-based study design gradually became universal in clinical trials and statistical methods such as random sampling, randomized allocations and hypothesis testings have been widely utilized in pharmaceutical research. (Council, 1948;

Hart, 1999) Finally, the mathematical modeling of the spread of infectious disease in human populations stretches back to the development of inoculation by Daniel Bernoulli in 1766 and the tracking of cholera outbreak by John Snow in 1857. (Brauer, 2017) Since the last century, the most popular tool for epidemic modeling is the compartmental model by the pioneering work of Kermack and McKendrick (1927) and most of the succeeding research thenceforth focused on the analytical solutions to the original or extended framework of this model.

Major issues remain in the application of theoretical tools from the mathematical sciences in bio-medical research. First and foremost, many technical limitations and assumptions of these tools have either inhibited or outright prevented researchers from obtaining accurate or even meaningful results. An example of this issue is the assumption of homogeneous mixing by the compartmental model in epidemic modeling. This assumption implies that the probability of disease transmission is identical for all members of the population, which was found to be unrealistic and non-factual by many disease and demographic case studies. (Anderson, 1991; Brauer, 2017)

Furthermore, the application of mathematical tools automatically raises new questions that need to be addressed. For example, the statistically-based study design of randomized control trial entails questions regarding the stopping points of different treatment phases and the minimum sample sizes, among others. The work by Bryant and Day (1995); Conaway and Petroni (1995), for instances, address these questions for the study designs of multiple-phase clinical trials.

Issues regarding the proper applicability of mathematical tools in bio-medical research topics have led to further investigations that continue to this date and have opened up new opportunities for inter-disciplinary studies. Specifically, the application of conventional mathematical tools often raise more questions than answers due to the discrepancy caused by the generality of mathematical tools and the particularity of biological questions. This has induced the need for the development of more suitable and specific quantitative tools for each respective bio-medical research topic.

1.2 Recent developments in biomedical data science

Recent developments in the disciplines of statistical sciences, theoretical computer science and mathematics have led to the emergence of new perspectives and novel approaches across many research topics in the fields of mathematical biology and biostatistics. As

previously elucidated, statistics has been practiced in biomedical research starting from the 20th century and earlier. Most of these statistical methods have their basis on frequentist statistics, which was systematically formalized by the work of Ronald Fisher, Jerzy Neyman and Egon Pearson. Recently, Bayesian statistics, which was first formulated by Thomas Bayes in the 16th century, has also gained popularity as an alternative to the conventional frequentist statistics due to the rise of modern computational power. Prior to the emergence of modern computational capacity, Bayesian statistical methods would have been deemed impossible and would have only been used when analytical solutions to the calculations are easily available. This increasing ubiquity of computational methods nowadays has enabled the statistical learning of enormous and highly complex biomedical data sets. Following the previously introduced topics in biomedical research, it can be seen that the these recent developments have indeed led to improvements in the understanding of the biomedical research topics under considerations here.

In recent years, empirical research in evolutionary biology began to focus on understanding how social structures and communicative systems might have impacted the evolution of different species. Empirical studies of this kind analyze interactive behaviors and vocal signals of social animals. This new perspective has arisen due to the prevalence of many computational methods which allow more complex bio-acoustic sequences to be studied. Signal processing techniques such as Fourier transform have allowed the bios-acoustic sequences to be analyzed in time-frequency format using statistical methods. (Kershenbaum et al., 2016) Quantitative tools that have been used to conduct such bio-acoustic analysis include feature engineering methods, time series as well as unsupervised clustering methods. (Sainburg et al., 2020) Results obtained by Gamba and Giacoma (2007) and Gamba et al. (2016), among other studies, have shown that bio-acoustics investigators are able to make use of statistical tools for identifying meanings in the vocal communication of a specific species. Bayesian cluster analysis method was also used in the bio-acoustic study by Valente et al. (2019).

Many randomized clinical trials have been independently conducted since their post-war success for understanding penicillin and antibiotics. As previously mentioned, the statistically-based study design is an issue that is left to be addressed by the researchers responsible for each trial. However, many of these trials might have aimed to evaluate the same treatment or drug even though they might have different study designs from each other. For this reason, historical data obtained from these trials could show conflicting statistical results on the very same treatment in question. As a result, a new approach called meta-analysis has appeared as a new statistical methodology to synthesize results from independent but related trials. (Normand, 1999; Schwarzer et al., 2015) The study by Gili et al. (2016) is

an example of meta-analysis that aims to understand the effectiveness of statin therapy on total cholesterol levels of HIV patients. Lately, the usage of Bayesian methods are also being investigated by clinical researchers such as the work by Hu et al. (2020). Many models and methodologies for meta-analysis are still being proposed or under development in order to further improve clinical research. (Schwarzer et al., 2015; van Houwelingen et al., 2002)

Since the introduction of random graphs and complex networks as a field of study in its own right, network models were shown to possess far more realistic assumptions in contact and transmission patterns than the classical compartmental model in epidemic modeling. (Newman, 2002) Subsequent research has confirmed the revolutionary success of the network approach in terms of applicability as well as computational efficiency. Therefore, epidemic network models have attracted considerable research interest in infectious disease modeling. A case study by Cauchemez et al. (2011) on H1N1 confirmed the impact of social network structure on the modeling of spread of disease and the recent study by Maheshwari and Albert (2020) also made use of the network model with edge-deletion for analyzing Covid-19. The computational efficiency of the network approach to epidemic modeling was demonstrated by the variational Bayesian approximation method in the work of Karrer and Newman (2010).

Altogether, these indicate that gradual improvement is being made in the understandings of the biomedical sciences due to the progress being made in the computational sciences, especially in computational statistics. It should be noted that most of the aforementioned computational methods for biomedical sciences are mostly based on the statistical sciences and therefore naturally require the availability of large data sets.

1.3 Challenges of biomedical data modeling

In spite of these recent breakthroughs, many challenges remain unresolved in the statistical learning of bio-medical data sets due to many methodological issues such as high dimensionality, statistical heterogeneity and inconsistency. However, it is beyond the scope of this introduction to explain the many methodological problems involved in the learning and analysis of biomedical data sets using statistical methods. Therefore, only the methodological issues that are relevant to the rest of this thesis are elucidated here. These relevant issues include high dimensionality, inapplicability of generic statistical tools or models as well as properties of some novel mathematical tools for modeling.

The issue of high dimensionality refers to the fact that the sizes of biomedical data sets are usually enormous. This often renders the direct application of many contemporary

statistical methods computationally infeasible. This is known as the "Big N Problem". Many approximation methods have been proposed for the different types of statistical models in both the frequentist and the Bayesian framework accordingly. For example, the "Big N Problem" in spatial statistics is explained in details in the work by Banerjee and Fuentes (2012). In biomedical data modeling, studies on biodiversity and wildlife populations such as the work by Bondo et al. (2023, 2024) often encounters this particular problem in their use of multivariate spatial process models and therefore requires suitable approximation methods to be applied on the statistical modeling of spatial data sets. For example, the work by Bondo et al. (2023, 2024) made use of the approximation method proposed by Rue et al. (2009) to model diseases in wildlife populations. The development and proper application of statistical models and approximation methods in biomedical data modeling require further research and investigation. In addition, the improvement of approximation methods for the purposes of easing computation and wider applications in different settings is also a research area of interest.

Another common issue in biomedical data modeling is the peculiar characteristics intrinsic to biological data which beget the inapplicability of many generic statistical tools. These peculiar characteristics of biomedical data sets almost always imply the need of distinctive statistical models or methodologies specialized to answer the biomedical research question of interest using the available data set. Examples of this include the study on microbiome by Denti et al. (2021) and the study on gene microarrays by Zou and Hastie (2005). Another example is clinical trial data sets that are obtained from different study designs, which require different statistical models and methodologies for trial data analysis. In light of the characteristics of different biomedical data sets, the role of statistician is required in such studies in order to build or apply specific statistical models or methodologies.

Another aspect that needs to be taken care of is the theoretical properties of the novel statistical tools. The theoretical properties and behaviors of many novel models and methodologies are not well established and is a topic of active research. This is particularly true for models that are based on random graphs and complex networks since the network science has only become a field of study on its own not so long ago. (Albert and Barabási, 2002; Newman, 2010)

There are many issues regarding the statistical analysis and modeling of large biomedical data sets that are beyond the scope of this thesis. In summary, the rise and challenges of biomedical data modeling have led to statisticians playing an ever more important role in biomedical research.

1.4 Content of this thesis

Many statistical methods, whether frequentist or Bayesian, are available and can be used for understanding the biomedical research topics that were discussed in this introduction. Nonetheless, as previously pointed out, many issues remain to be resolved and the previously introduced methodological challenges have led to the formation of the research projects that form the content of this thesis.

This thesis focuses on four research areas: the Bayesian learning of a set of bio-acoustic data (Chapter 2), the Bayesian meta-analysis of randomized clinical trials (Chapter 3) and the asymptotic behavior of the degree of a network model (Chapter 4). Each chapter of this thesis corresponds to a published or submitted article of a journal. The opening remark of each chapter indicates the corresponding published or submitted article.

Chapter 2 presents a novel Bayesian methodology that can learn from a set of bio-acoustic signals and can be used for cross-species comparison. The proposed model is a spatio-temporal model that obtains the latent spectral shape of the species-specific acoustic signals. This is achieved by accounting for periodic artifacts and synchronization in time. The implementation of the proposed model involves an approximation method due to the size of the data set.

Chapter 3 is the Bayesian network meta-analysis of a set of clinical trial data. The methodology utilized to perform the analysis is based on a modified proportional hazard model. A comparison of different approaches and the diagnostic statistics involved in the comparison are described.

Chapter 4 rigorously establishes a corrected statement regarding the behavior of the expected degree of a specific preferential attachment model, which is a type of network model that exhibits the power-law degree sequence and is highly suitable for modeling real-world network data. A simulation study is also conducted to numerically demonstrate the validity of the results.

Chapter 5 provides a concluding remark.

Chapter 2

Bayesian inference of latent spectral shapes

Background

This chapter was submitted as a journal article and was partially published in a conference proceedings:

Yip H.C., Mastrantonio G., Bibbona E., Gamba M. and Valente D. *Under review.* Bayesian inference of latent spectral shapes.

Yip H.C., Mastrantonio G., Bibbona E., Gamba M. and Valente D. Nearest neighbours Gaussian process model for time-frequency data: An application in bio-acoustic analysis. July 18-22, 2022. Conference Proceedings of the 36th International Workshop on Statistical Modelling. 603-607. ISBN: 978-88-5511-309-0.

The statistical model and results by Prof. Gianluca Mastrantonio, Prof. Enrico Bibbona and PhD candidate H.C. Yip in this work are based on the biological knowledge and bio-acoustic data set provided by Prof. Gamba and Dr. Valente. The candidate would like to express her deepest gratitude to all collaborators whose knowledge and expertise are the reasons for the success of this work.

2.1 Introduction

Bio-acoustic analysis is of significant research interests across different disciplines due to the potential insights into evolutionary biology that can be given by the information contained in animals' acoustic sequences. Exemplary areas of interests in the biological studies of bio-acoustics include cross-species animal behavioral studies, their vocal repertoire and anatomy. The work of Dunn and Smaers (2018) and McComb and Semple (2005), for example, made use of the repertoire size as an indicator for the level of communication complexity of primates.

Recent development in computational statistics has given rise to a form of bio-acoustic analysis that focuses on the spectral properties rather than the behavioral or biological contexts of the recorded analogue signals. In this form of analysis, the recorded analogue signals are usually processed and discretized by the Fourier Transform which results in a discretized signal being represented in a time-frequency format, called spectrogram, which consists of a time axis, a frequency axis and amplitude at each time-frequency coordinate. Bio-acoustic analysis therefore becomes a form of time-frequency analysis.

The recent survey by Sainburg et al. (2020) provided an overview of the contemporary computational methods for the spectrographic representations of bio-acoustic data. The most common practice generally involve the selection of a set of basis-features from the spectrograms for quantitative comparison between various sets of bio-acoustic data. The most common type of features are musical features such as pitch, timbre and harmony, whilst features in the signal domain such as amplitude, period and fundamental frequency are also very popular. Depending on the goal of the particular study, bio-acoustic investigators can make use of a combination of several types of features obtained from the spectrograms alongside with factors that are grounded in the biological contexts of the signals. Examples of these studies include the work of Gamba and Giacoma (2007); Gamba et al. (2016) and Valente et al. (2019).

The survey by Kershenbaum et al. (2016) provided additional background on several paradigms of basis-features and suggested a protocol to analyze them. However, it was also suggested that the identification of meaningful basis-features are often inaccurate by virtue of human subjectivity and difficult to generalize for cross-species comparison. Furthermore, these manually selected basis-features are almost always treated as independent features by bio-acoustic investigators such that the effects of time are customarily ignored. As pointed out by Sainburg et al. (2020), ignoring the relative relationships between the time-varying components and the spectral shape might result in the failure to capture the relevant

Table 2.1 Descriptions of the set of recordings of the "grunt" call-type.

	Species							
	EC	ER	FL	FU	II	MA	MO	PD
# recordings	1966	6594	1174	1041	1144	3615	1492	145
min. length	0.025	0.020	0.020	0.021	0.011	0.020	0.031	0.027
max. length	4.016	0.374	0.474	0.513	0.238	2.057	1.135	0.105
# length > 0.1	1311	4531	542	464	25	1599	1277	2
# length > 0.2	412	180	15	33	4	204	268	0
# length > 0.3	87	12	2	13	0	47	73	0

characteristics of the signal. A study conducted by Bregman et al. (2016) also challenges the conventional view that songbirds rely exclusively on the absolute pitch at a specific frequency for tone sequence pattern recognition. Instead, it was argued that songbirds are able to perceive and recognize tone sequences as long as the overall pattern of the spectral amplitudes are preserved.

This chapter presents a spatio-temporal model for bio-acoustic data in spectrogram representation for the purpose of facilitating cross-species comparison in bio-acoustic analysis. Specifically, this work deals with the problem of information extraction by parametric estimation of a Gaussian process model in the presence of non-stationary covariance structures in the time dimension. The aim of the proposed model is to obtain the latent spectral shape of the acoustic structure of a species. The latent spectral shape is then used to measure the dissimilarity between different species.

The subsequent sections are organized as follows. Section 2.2 describes and illustrates the available set of bio-acoustic data that motivates the work in this chapter. Section 2.3 thoroughly explains the proposed spatio-temporal model. Section 2.4 discusses the methodologies utilized in the implementation of the model. Finally, the analysis of the results obtained from applying the proposed model on the data set are provided in Section 2.5. A simulation study can be found in the Appendix A.

2.2 The data set

The available data set is a set of vocal signals of lemurs in Madagascar. Each signal was emitted by an individual lemur and was recorded as an audio file. A total of 39741 analogue audio recordings were obtained from 9 different species of lemurs. The format of the data

set resembles that of Pozzi et al. (2010). Each recording is categorized by a species label and a behavioral call-type label which respectively characterizes the biological species and behavioral context from which the signal came. The descriptions of some of the call-type labels can be found in the work of Maretti et al. (2010) and Pozzi et al. (2010), for example. The total number of behavioral call-types ranges from 9 to 20 for each species. The analysis of the entire data set is outside the scope of this work and thus the focus is put on a single call-type, called "grunt". The "grunt" call-type is of particular interest because it is a type of multi-contextual call that demonstrates species-specific characteristics and is used by each species to coordinate their collective movements. (Gamba et al., 2012; Sperber et al., 2017)

Table 2.1 summarizes the characteristics of the analogue signals that are labeled as "grunt". The "grunt" call-type is common to 8 out of 9 species. Note that each analogue audio recording lasts for a unique duration of time. The time lengths of the recordings are summarized in Table 2.1 from the second to the last row. Note that the maximum time length largely differs from the minimum time length for each species. It should also be noted that the number of recorded signals of longer time-lengths, e.g. # length > 0.3 in the last row, are especially few relative to that of shorter time-lengths, e.g. # length > 0.1 in the fourth row.

The analogue audio signals are discretized and converted into the time-frequency format, called the spectrogram, by means of the Short-Time Fourier Transform with a constant time-step of 0.01 seconds. The frequency axis of each spectrogram ranges from log 63 to log 20000 decibels and is divided into 26 evenly spaced frequency intervals of step size 0.221 approximately. On the other hand, the time axis of each spectrogram is discretized by a constant time-step of 0.01 seconds. As a consequence, all spectrograms have the same number of frequency coordinates, but each of them has a distinctive number of time coordinates of its own due to the varying time-lengths of the recordings as described in Table 2.1.

Two motivating examples of the spectrograms from the data set are shown in Figure 2.1 and Figure 2.2. The $x-$axis and $y-$axis are the discretized time and frequency domain, respectively. The sound points of the acoustic structure are measured at the time-frequency coordinates as amplitudes. Upon inspection of these figures, one can spot several noticeable attributes associated with the time domains of the spectrograms, which affect the acoustic structures in various ways.

Firstly, there exists misalignment between the starting time and ending time of different signals. The two spectrograms in Figure 2.1 depict the same "grunt" sound from the ER species. A closer look reveals the misalignment in the two acoustic structures: the higher amplitudes between the log-frequencies of 5.3 and 7.0 on the $y-$axis commence at time

Figure 2.1 Two discretized signals from species ER

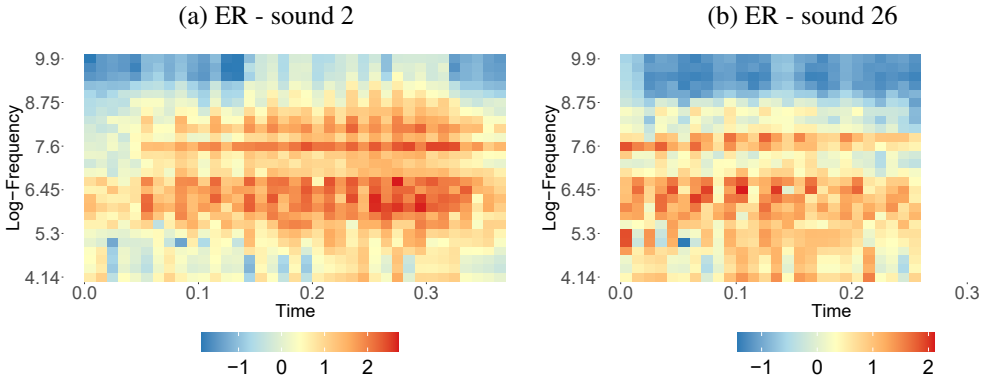
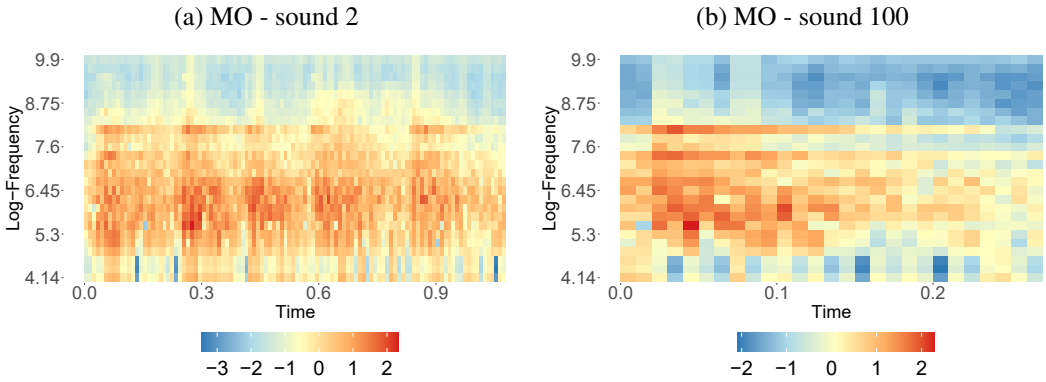


Figure 2.2 Two discretized signals from species MO



0 for Figure 2.1 (b) but approximately at 0.06 for Figure 2.1 (a). It is not unreasonable to contemplate the possibility that the actual beginning of Figure 2.1 (b) is 0.06 relative to the $x-$axis of Figure 2.1 (a).

Secondly, there is a possibility that the spectrograms are asynchronous in time. Figure 2.2 shows two spectrograms of the same sound from the MO species. Though it is not apparent to the naked eyes, Figure 2.2 (b) might be a portion of Figure 2.2 (a) being depicted at a different speed. This is because two individual animals may produce different parts of the same sound at different speeds, which lead to non-linear warping of the same sound and thus the dissimilarity between the two acoustic structures in Figure 2.2.

Lastly, it can be observed that there exists oscillations along the observed time domains of the spectrograms. For example, the temporal pattern of the spectrograms in Figure 2.1 exhibit oscillations along the time domains wherein a time bin of high sound intensity is immediately followed by 1 or 2 bins with lower intensity. However, the temporal pattern of Figure 2.2 (a) exhibits an oscillation that differs from those in Figure 2.1. In Figure 2.2 (a), it

seems that a time bin of the highest sound intensity appears approximately every interval of 0.15 seconds. These oscillating temporal patterns are induced by the presence of sampling artifacts that arise from the windowing effect of the Short-Time Fourier Transform (STFT), which is a well known phenomena that was explained in Tan and Jiang (2018).

In summary, the uniqueness of the recordings' time-lengths entails distortions in the acoustic structures, whilst the constant time-step of the STFT entails the presence of sampling artifacts. Regarding the distortions, the time axis of different signals might be misaligned or non-linearly warped with respect to one another. This distortion of each recording that arises from asynchronous time will have to be quantified. Regarding the sampling artifacts, they appear as cyclic components along the observed time domains in the spectrograms and so their periodicity will need to be accounted for.

In order to enable cross-species comparison between the bio-acoustic signals of the same call-type label, each recorded signal is assumed to be a noisy realization of a latent acoustic structure that is representative of the species to which the signal belong. This species-specific latent acoustic structure is envisioned to be the realization of a zero mean and stationary Gaussian process (GP) over a three-dimensional space. All recorded signals of the same species are assumed to be conditionally independent given this latent and stationary GP in three-dimensional space. Marginally, a recorded signal is the realizations of a temporally non-stationary GP over a two-dimensional space. The major novelties in this work lay in the features of the latent and stationary GP which represents a species' acoustic structure as well as its connection to the non-stationary GPs for the recorded signals. As will be discussed in the next Section 2.3, these major novelties are the resolutions to the issues illustrated by the above motivating examples.

The size of the available data set in Table 2.1 presents a computational challenge within a GP-based modeling framework, which arises from the inversion of the covariance matrix. This is commonly referred to as the Big N Problem and is usually addressed by replacing the original GP with an approximated but comparably close one that requires less computational power to be estimated. Many approaches have been proposed to obtain such an approximation , including the sparse nearest neighbor approach by Vecchia (1988), the low-rank approach by Banerjee et al. (2008), and the Gaussian Markov random field approach by Rue and Martino (2007), among many others.

In this work, the Nearest neighbor Gaussian Process (NNGP) method that was formulated in Datta et al. (2016a) is employed. This approach has been successfully applied to another data set with circular time that is similar to the periodic artifacts in this work. Datta et al. (2016b); Mastrantonio et al. (2017) A detailed description of the NNGP method and other

techniques involved in the implementation of the GP model will be provided in Section 2.4. The results are the posterior samples from the predictive distribution of the GP model. These samples can be interpreted as the smooth acoustic structure that is representative of a single species, which can then be used by bio-acoustic investigators for cross-species comparison. Moreover, the posterior values of the model parameters also allow investigators to explain and describe the acoustic structure of a specific species. The posterior samples are obtained by the Markov Chain Monte Carlo (MCMC) method.

2.3 The model

All the following notations are referred to the recorded signals of the same call-type from the same species. Signals of different species are assumed to be independent from each other and signals of the same species are assumed to follow the same model with the species-specific parameters. Let N be the total number of recorded signals of one call-type from the same species. Let T_i be the number of time coordinates on the observed time-axis of the $i-$th recorded signal where $i = 1,\ldots,N$ and let H be the number of log-frequency bins on the frequency-axis, respectively. Let

$$\mathcal{T}_i = \{0.01(k-1)|k = 1,\ldots,T_i\}, \quad \mathcal{H} = \{0.23k + \log 63|k = 1,\ldots,H\}, \tag{2.1}$$

such that the time-length of the $i-$th recorded signal is denoted by $l_i = \max(\mathcal{T}_i)$.

Let each $i-$th recorded signal be denoted by $\boldsymbol{y}_i = (y_{i,t,h})_{t\in\mathcal{T}_i,h\in\mathcal{H}}$, then $\boldsymbol{y}_i$ is assumed to be the discrete realizations of a two-dimensional process that is defined over time and log-frequency, denoted by $\mathcal{Y}_i(t,h) \in \mathbb{R}$ where $t \in \mathbb{R}_{\geq 0}$ is time and $h \in \mathbb{R}$ is log-frequency. The two dimensional process $\mathcal{Y}_i(t,h)$ is further assumed to be the noisy and non-stationary version of a noise-free latent process that is denoted by $\mathcal{A}_i(t,h)$. The model is

$$\begin{aligned} \mathcal{Y}_i(t,h) &= \mu_i + \mathcal{A}_i(t,h) + \varepsilon_i(t,h), \\ \varepsilon_i(t,h) &\overset{i.i.d.}{\sim} \mathrm{GP}(0,\tau_i), \end{aligned} \tag{2.2}$$

where $\mu_i \in \mathbb{R}$ is the mean sound intensity level and $\tau_i \in \mathbb{R}_{>0}$ is the variance term, otherwise known as the nugget effect. The latent process $\mathcal{A}_i(t,h)$ is defined to be the weighted average

of two stationary GPs that have two separate roles. The latent process $\mathcal{A}_i(t,h)$ is given by

$$\begin{aligned} \mathcal{A}_i(t,h) &= \sigma\left(\sqrt{\lambda}\mathcal{W}_1\big(\psi_i(t),h\big) + \sqrt{1-\lambda}\mathcal{W}_2(t,h)\right), \\ \mathcal{W}_1(d,h) &\sim \mathrm{GP}\left(0, \mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})\right), \\ \mathcal{W}_2(t,h) &\sim \mathrm{GP}\left(0, \mathcal{C}^c(\cdot,\cdot;\boldsymbol{\theta})\right), \end{aligned} \tag{2.3}$$

where $\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})$ and $\mathcal{C}^c(\cdot,\cdot;\boldsymbol{\theta})$ are the correlation functions that are dependent on the vector of parameters $\boldsymbol{\theta}$. The two components of $\mathcal{A}_i(t,h)$ respectively address the aforementioned distortion and periodic artifacts as shown by the previous Section 2.2. The first component $\mathcal{W}_1(d,h)$ is called the common-time component that models the distortion of recorded signals through a specific transformation encoded in the synchronization function $\psi_i(t) = d$. The common-time component and the construction of the synchronization function are discussed in details in Section 2.3.1. The second component $\mathcal{W}_2(t,h)$ is an additional component that is used solely for modeling the periodic artifacts and is called the cyclic component as such. The cyclic component is further discussed in Section 2.3.2.

Note that dependence between all recorded signals of the same species is introduced through the latent process $\mathcal{A}_i(t,h)$ in equation (2.2). However, $\mathcal{A}_i(t,h)$ is defined over a three-dimensional space that consists of two different time dimensions and one frequency dimension such that

$$\begin{aligned} &Cov\big(\mathcal{A}_i(t,h), \mathcal{A}_{i'}(t',h')\big) \\ &\qquad := \sigma^2\big(\lambda\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta}) + (1-\lambda)\mathcal{C}^c(\cdot,\cdot;\boldsymbol{\theta})\big) : \mathbb{R}^2_{\geq 0} \times \mathbb{R} \to \mathbb{R}_{>0} \end{aligned}$$

where $(i,t,h) \not\equiv (i',t',h')$ and $(\sigma^2,\lambda) \in \boldsymbol{\theta}$ is the variance and weight, respectively. Note that $\mathcal{A}_i(t,h)$ and $\mathcal{A}_{i'}(t',h')$ are distinct from each other only through the data-specific synchronization functions, $\psi_i(t)$ and $\psi_{i'}(t')$. Meanwhile, the common-time component $\mathcal{W}_1(d,h)$ and the cyclic component $\mathcal{W}_2(t,h)$ are defined across all time and frequencies. This implies that $\mathcal{A}_i(t,h)$ and $\mathcal{A}_{i'}(t',h')$ where $(i,t,h) \not\equiv (i',t',h')$ are just different instances of the same latent process. Hence, the two-dimensional realizations of this three-dimensional latent process can be treated as though it is the representative acoustic structure of the species. The latent process $\mathcal{A}_i(t,h)$ specified by equation (2.3) is henceforth named the "Basic Acoustic Structure of Species" ("BASS") throughout this work.

2.3.1 Synchronization function & common time component

The synchronization function $\psi_i(\cdot) : \mathcal{T}_i \to \mathbb{R}_{\geq 0}$ in the definition of "BASS" in equation (2.3) is constructed to synchronize the observed time axis of the $i-$th recorded signal with a time frame that is common to all signals. This common time frame is one of the time dimensions for "BASS" and is henceforth called the common-time dimension. The coordinates within the common-time dimension, $d \in \mathbb{R}_{\geq 0}$, are henceforth termed the common-time coordinates. The issue of asynchronous time that was demonstrated by Figure 2.1 and Figure 2.2 in Section 2.2 can be addressed by mapping the real time coordinates $t \in \mathcal{T}_i$ for the non-stationary process $\mathcal{Y}_i(t,h)$ to the common-time dimension wherein temporal stationarity holds. Intuitively, the parameters of the synchronization function $\psi_i(t)$ should describe the portion of the common-time dimension to which the observed time axis $\mathcal{T}_i$ corresponds. The synchronization function is therefore defined as

$$\psi_i(t) = \alpha_i + \beta_i \Omega_i\left(t/l_i\right) l_i \tag{2.4}$$

where $\alpha_i \in \mathbb{R}_{\geq 0}$ is the *translation parameter*, $\beta_i \in \mathbb{R}_{>0}$ is the *scaling parameter* and $\Omega_i(q)$ is the non-linear *time-warping function* that depends on the vector of *warping parameters*. The time-warping function $\Omega_i(q)$ must be defined such that the boundary conditions $\Omega_i(0) = 0$ and $\Omega_i(1) = 1$ are always met. Furthermore, it must be continuous and strictly increasing in $[0,1]$ as a way to avoid the folding effect. Interpretation-wise, the boundary conditions imply that the beginning and the end of the $i-$th signal at real time $t = 0$ and $t = l_i$ always correspond to $\psi_i(0) = \alpha_i$ and $\psi_i(l_i) = \alpha_i + \beta_i l_i$, respectively. This means that the synchronization function maps the observed real-time coordinates $t \in \mathcal{T}_i \subset [0, l_i]$ to the the common-time coordinate $d \in [\alpha_i, \alpha_i + \beta_i l_i] \subset \mathbb{R}_{\geq 0}$ within the common-time dimension for $\mathcal{W}_1(d,h)$.

In principle, any function with the necessary features can be used as the time-warping function in theory. However, the choice of $\Omega_i(q)$ must retain an easy interpretation and flexibility with as few parameters as possible so as to avoid over-parametrization. The choice of $\Omega_i(q)$ is therefore the Beta cumulative distribution function (CDF) that is given by

$$\Omega_i\left(q\right) = \frac{\Gamma(\exp\zeta_i + \exp\delta_i)}{\Gamma(\exp\zeta_i)\Gamma(\exp\delta_i)} \int_0^q x^{\exp\zeta_i - 1}(1-x)^{\exp\delta_i - 1}\partial x, \qquad q \in [0,1], \tag{2.5}$$

where (ζ_i, δ_i) is the vector of warping parameters with $\zeta_i \in \mathbb{R}_{>0}$ and $\delta_i \in \mathbb{R}_{>0}$. Figure 2.3 shows several examples of the time-warping function under different parametrization. Indeed, a special case of equation (2.5) is when $\delta_i = \zeta_i = 0$, then $\Omega_i(q) = q$ and the Beta CDF is basically reduced to the Uniform CDF in which case non-linear time-warping is completely

absent such that $\psi_i(t) = \alpha_i + \beta_i t$. If further $\alpha_i = 0$ and $\beta_i = 1$, then there is no distortion in the signal whatsoever. Hence, the difference

$$\frac{\partial}{\partial q}\Omega_i(q) - 1$$

can be used to evaluate how the $i-$th recorded signal has been accelerated in time with respect to the common-time dimension for the latent process in equation (2.3). It is clear from Figure 2.3 that if $|\zeta_i| \gtrapprox 0.75$ or $|\delta_i| \gtrapprox 0.75$, then the warping becomes so severe that the derivative of $\Omega_i(q)$ evaluated at $q \approx 0$ or $q \approx 1$ is close to either 0 or ∞. This is unjustifiable from an application perspective and the parametric values need to be limited to a certain threshold. This argument will be reckoned with later in Section 2.4.3 when the distributions over the warping parameters are decided.

To simplify notations, let $\boldsymbol{\xi}_i = (\zeta_i, \delta_i)$ be the vector of warping parameters for the time-warping function in equation (2.5) and let $\boldsymbol{\chi}_i = (\alpha_i, \beta_i, \boldsymbol{\xi}_i)$ be the vector of all parameters of the synchronization function in equation (2.4). Let the distance between $\psi_i(t)$ and $\psi_{i'}(t')$ within the common-time dimension be denoted by

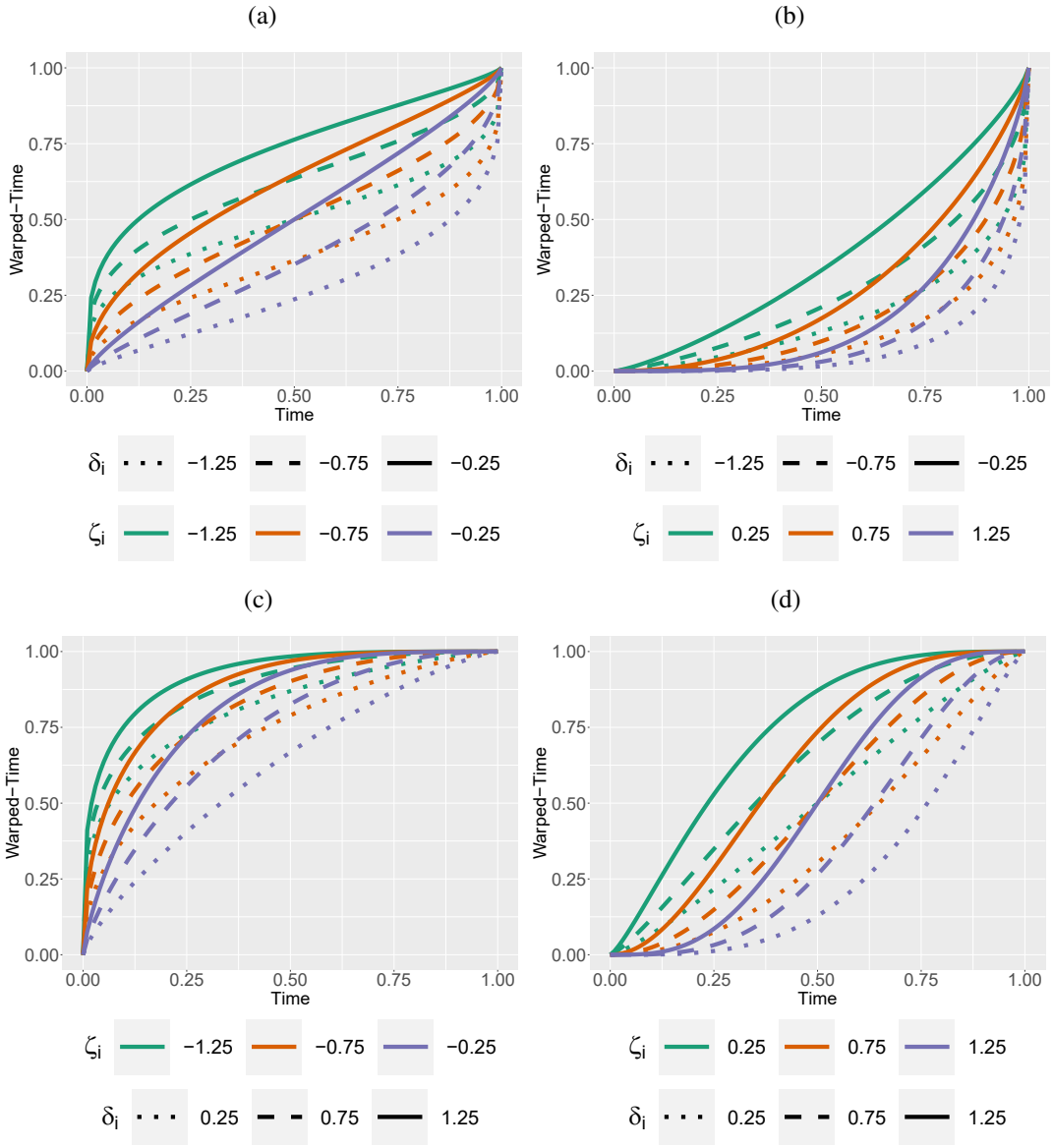
$$\begin{aligned} \Delta(t, t'; \boldsymbol{\chi}_i, \boldsymbol{\chi}_{i'}) &= |\psi_i(t) - \psi_{i'}(t')| \\ &= \left|\alpha_i + \beta_i \Omega_i\left(t/l_i\right) l_i - \alpha_{i'} + \beta_{i'} \Omega_i\left(t'/l_{i'}\right) l_{i'}\right| \end{aligned} \tag{2.6}$$

where $t \in \mathcal{T}_i, t' \in \mathcal{T}_{i'}$ and $i \neq i'$. The common-time component $\mathcal{W}_1(d,h)$ in equation (2.3) models the natural change in the spectral shape of "BASS" across the common-time and frequency dimensions. The correlation function $\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})$ for the common-time component is

$$\mathcal{C}^g(|h-h'|, |d-d'|; \boldsymbol{\theta}) = \frac{1}{\phi_d \Delta(t, t'; \boldsymbol{\chi}_i, \boldsymbol{\chi}_{i'}) + 1} \exp\left(-\frac{\phi_h |h-h'|}{\left(\phi_d \Delta(t, t'; \boldsymbol{\chi}_i, \boldsymbol{\chi}_{i'}) + 1\right)^{\rho/2}}\right) \tag{2.7}$$

where $(\phi_h, \phi_d, \rho) \in \boldsymbol{\theta}$. This is a parametrization of the Gneiting correlation function that was proposed in Gneiting (2002). The non-separable parameter $\rho \in [0,1]$ is the time-frequency interaction parameter, ϕ_d is the common-time decay and ϕ_h is the frequency decay. Since the temporal lags in $\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})$ must be measured for all real time coordinates in the common-time dimension, the data-specific synchronization function becomes an intrinsic part of the covariance function for "BASS". Note that there exists the issue of identifiability since only the relative difference between the translation parameters, α_i and $\alpha_{i'}$, affects the correlation function in equation (2.7). Similar issue arise for the scaling parameters, β_i and $\beta_{i'}$, as well as

Figure 2.3 Time-warping function $\Omega_i(q)$ under different parametric values.

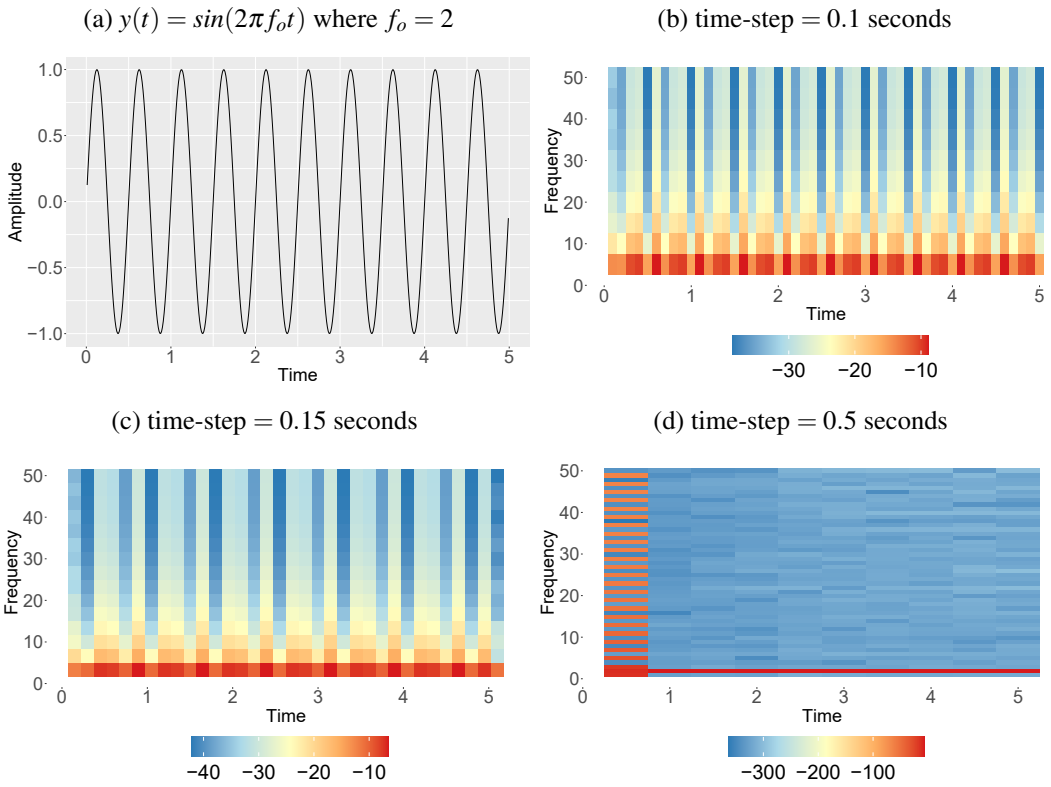


the time decay, ϕ_d. The issue of identifiability will be discussed in details in the subsequent Section 2.4.2.

2.3.2 Cyclic component

As illustrated by Figure 2.1 and Figure 2.2 (a) in Section 2.2, there exists oscillations in the temporal patterns in the sound intensities of the signals due to the presence of periodic sampling artifacts. Figure 2.4 is a toy example that illustrates how periodic sampling artifacts

Figure 2.4 Periodic sampling artifacts that arise from the STFT



arise from the STFT. When the time-step used by the STFT is less than the period of the true waveform, the sampled frequency does not capture the original periodicity of the true waveform and therefore becomes an artifact. In contrast to Figure 2.4 (d) in which the time-step of 0.5 seconds perfectly captures the fundamental period of Figure 2.4 (a), the periodic artifacts can be seen in both Figure 2.4 (b) and (c) when the time-steps are less than the fundamental period. Since real acoustics and sounds almost inevitably vary with time, their waveform are usually highly complex and non-periodic with respect to time. Thus, it is highly unlikely for the STFT to be able to capture the true waveform cycle. Furthermore, the size of the time-step of STFT cannot be easily changed or adapted to mitigate this problem as it is dictated by the Heisenberg uncertainty principle. (Graps, 1995; Kumar and Foufoula-Georgiou, 1997) Due to these reasons, the cyclic component $\mathcal{W}_2(t,h)$ is defined to model the relations of the periodic artifacts with the frequency dimension.

Let the circular distance between two real-time coordinates be defined as

$$\Delta_c(|t-t'|;\gamma) = \min(|t-t'| \bmod \gamma,\ \gamma - |t-t'| \bmod \gamma) \tag{2.8}$$

which measures the distance between two angle t and t' on a circle of circumference γ. This means that the circular distance between any two observed time coordinates is restricted to a circular scale with period $\gamma/2$ such that $\Delta_c(|t-t'|;\gamma) \in [0,\gamma/2]\ \forall\, t,t'$. Similar to the works of Shirota and Gelfand (2017) and Mastrantonio et al. (2017), the circular distance is then used to define the cyclic component which is

$$\mathcal{C}^c(|t-t'|,|h-h'|;\boldsymbol{\theta}) = \frac{1}{\phi_c\Delta_c(|t-t'|;\gamma)+1}\exp\left(-\frac{\phi_h|h-h'|}{\left(\phi_c\Delta_c(|t-t'|;\gamma)+1\right)^{\rho/2}}\right) \tag{2.9}$$

where $(\phi_c,\gamma) \in \boldsymbol{\theta}$ with ϕ_c being the circular-time decay and γ being the periodicity. The positive definiteness of this correlation function has been proven by Shirota and Gelfand (2017). The frequency decay parameter ϕ_h and the interaction parameter ρ are the same as those in the common-time component $\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})$ in equation (2.7). Note that the synchronization function is noticeably absent in $\mathcal{C}^c(\cdot,\cdot;\boldsymbol{\theta})$ owing to the fact that the cyclic component $\mathcal{W}_2(t,h)$ depends only on the periodic distances between the artifacts, which can only be given by the real time coordinates that are taken directly from the observed time axis $\mathcal{T}_i$.

The cyclic component is defined as such because the sampling artifacts that are induced by the discretization of the analogue signal almost always lead to the appearance of oscillations along the time domain as shown by Figure 2.1 and Figure 2.2 (a). There is no reason to assume that ϕ_h or ρ here in the cyclic component will be any different from that of the common-time component, especially with regard to the fact that the two correlation functions are just separate parts of the same covariance function. Ergo, for the purpose of avoiding over-parametrization, the frequency decay parameter ϕ_h and the non-separable interaction parameter ρ are the same for both equations (2.7) and (2.9).

2.3.3 Representative sound of species

The main objective of this work is to obtain a species-specific representative sound for cross-species comparison in bio-acoustic studies. This sound can be attained in the form of the finite realizations of the predictive "BASS", denoted by $\mathcal{A}_{N+1}(t,h)$. Care must be taken in the selection of the time-length and the corresponding synchronization function for the predicted sound. Since the realizations of $\mathcal{A}_{N+1}(t,h)$ serve as a comprehensive representative sound of a species, conditions must be imposed in order to ensure that it encompasses all available information from the recorded signals. Let the time-length and the synchronization function for the latent predictive "BASS" be denoted by l_{N+1} and $\psi_{N+1}(t)$, respectively.

Then, the constraints on l_{N+1} and $\psi_{N+1}(t)$ are defined as

$$\begin{aligned} &\psi_{N+1}(0) && = \min_{i\in\{1,\ldots,N\}} \psi_i(0), \\ &\psi_{N+1}(l_{N+1}) && = \max_{i\in\{1,\ldots,N\}} \psi_i(l_i), \\ &l_{N+1} > \gamma. \end{aligned} \tag{2.10}$$

These conditions ensure that all time-frequency coordinates of the observed sound points and at least one cycle of the cyclic component are included. The constraints on the synchronization function $\psi_{N+1}(t)$ specified in equation (2.10) implies a set of constraints for its vector of parameters $\boldsymbol{\chi}_{N+1} = (\alpha_{N+1}, \beta_{N+1}, \boldsymbol{\xi}_{N+1})$. Let D be a set that is defined as

$$D = \left\{ \chi_{N+1} \in \mathbb{R}^2_+ \times \mathbb{R}^2 \;\middle|\; \begin{aligned} &\alpha_{N+1} && = \min_{i\in\{1,\ldots,N\}} \alpha_i, \\ &\alpha_{N+1} + \beta_{N+1} l_{N+1} && = \max_{i\in\{1,\ldots,N\}} (\alpha_i + \beta_i l_i), \\ &l_{N+1} > \gamma \end{aligned} \right\}, \tag{2.11}$$

If $\boldsymbol{\chi}_{N+1} \in D$, then the constraints in equation (2.10) are satisfied. Note that there exists no constraints on the warping parameters $\boldsymbol{\xi}_{N+1}$ but only the translation parameter α_{N+1} and the scaling parameter β_{N+1} are constrained. Let $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_N^\top)^\top$ be the collection of all recorded signals in which the elements in $\boldsymbol{y}_i$ are sorted in the ascending order of time and log-frequencies. The representative sound of a species can be given by the two-dimensional realizations from the posterior distribution

$$\mathcal{A}_{N+1}(t,h) | \boldsymbol{y}, \boldsymbol{\chi}_{N+1} \in D$$

which is defined over "BASS", conditional on the data and subject to the constraints in equation (2.10) being satisfied. The details of the sampling procedure will be shown in Section 2.4.4.

2.3.4 The marginal model

Although the model is defined by the introduction of the "BASS", and is much easier to perceive conceptually, the sampling from "BASS" considerably impedes the direct MCMC implementation of the model since the latent process $\mathcal{A}_i(t,h)$ must be estimated at every time-frequency coordinate where a data point is observed. In particular, evaluating $\mathcal{W}_1(d,h)$ at different common-time coordinates requires the sampling of $\boldsymbol{\chi}_i$ of the synchronization function $\psi_i(t)$ for all $i = 1,\ldots,N$. Due to the size of the data set, it is more practical to marginalize the model with respect to the latent processes, $\mathcal{W}_1(d,h)$ and $\mathcal{W}_2(t,h)$, which

is a common practice in geo-spatial modeling. The covariance function that gives the cross-covariance, $Cov\big(\mathcal{Y}_i(t,h),\mathcal{Y}_{i'}(t',h')\big)$, is easily derived to be

$$\begin{aligned}\mathcal{C}^{y}_{i,i'}\big((t,h),(t',h');\boldsymbol{\chi}_i,\boldsymbol{\chi}_{i'},\boldsymbol{\theta}\big) = \\ \mathcal{C}^{A}_{i,i'}\big((t,h),(t',h');\boldsymbol{\chi}_i,\boldsymbol{\chi}_{i'},\boldsymbol{\theta}\big) + \tau_i^2\mathbb{1}_{\{(i,t,g)\equiv(i',t',g')\}}\end{aligned} \tag{2.12}$$

where $\mathbb{1}_E$ is the indicator function for the event E and

$$\begin{aligned}\mathcal{C}^{A}_{i,i'}\big((t,h),(t',h');\boldsymbol{\chi}_i,\boldsymbol{\chi}_{i'},\boldsymbol{\theta}\big) = \\ \frac{\sigma^2\lambda}{\phi_d\Delta(t,t';\boldsymbol{\chi}_i,\boldsymbol{\chi}_{i'})+1}\exp\left(-\frac{\phi_h|h-h'|}{\big(\phi_d\Delta(t,t';\boldsymbol{\chi}_i,\boldsymbol{\chi}_{i'})+1\big)^{\rho/2}}\right) + \\ \frac{\sigma^2(1-\lambda)}{\phi_c\Delta_c(|t-t'|;\gamma)+1}\exp\left(-\frac{\phi_h|h-h'|}{\big(\phi_c\Delta_c(|t-t'|;\gamma)+1\big)^{\rho/2}}\right).\end{aligned} \tag{2.13}$$

The observed processes, $\mathcal{Y}_i(t,h)$ and $\mathcal{Y}_{i'}(t',h')$ where $i\neq i'$, are conditionally independent given the "BASS" in equation (2.2). Clearly, this marginal covariance function is able to provide more insights into the actual relationship between the recorded signals than"BASS" alone since dependence between the observed processes are introduced through the marginalization with respect to the stationary latent processes, $\mathcal{W}_1(d,h)$ and $\mathcal{W}_2(t,h)$. Even though the stationary latent processes are essential to the formulation of the model, it is clear from equation (2.13) that the observed processes $\mathcal{Y}_i(t,h)$ is not assumed to be stationary because the marginal covariance function explicitly depends on t and t' instead of solely on $|t-t'|$. Another important issue regarding the warping parameters, $\boldsymbol{\xi}_i\in\boldsymbol{\chi}_i$ and $\boldsymbol{\xi}_{i'}\in\boldsymbol{\chi}_{i'}$, is the fact that they act exclusively as the non-stationary parameters of the marginal covariance function and are thus weakly identifiable in general. (Gelfand et al., 2010; Perrin and Meiring, 1999) This issue will be discussed later when the prior distributions for the warping parameters are specified in Section 2.4.2.

2.4 Implementation methodologies

This section concerns the methodologies that are utilized to resolve the aforementioned issues in the implementation of the model. In particular, the issues being addressed here are the non-identifiability of parameters in Section 2.4.1, the selection of prior distributions in Section 2.4.2, the approximation method NNGP in Section 2.4.3 and the posterior sampling from "BASS" in Section 2.4.4.

2.4.1 Identifiability

In general, the problem of non-identifiability in parametric models refers to the observational equivalence of multiple unique parametric estimates from whence the likelihood distributions given by the estimates are equal. Upon the inspection of the common-time component and the marginal covariance function, it is immediately revealed that the proposed model suffers from the problem of non-identifiability regarding the translation parameter, α_i, the scaling parameter, β_i, and the time decay parameter, ϕ_d. If a constant $c \in \mathbb{R}$ is added to all $(\alpha_i)_{i=1,\dots,N}$ or if all $(\alpha_i, \beta_i)_{i=1,\dots,N}$ are multiplied by a constant $c \in \mathbb{R}$ with time decay ϕ_d being divided by the same constant, then the values of $\mathcal{C}^g(\cdot,\cdot;\boldsymbol{\theta})$ remain the same despite the changes in parametric values, thereby leading to the same corresponding likelihood densities. To resolve this issue of non-identifiability, the MCMC algorithm explores the posterior space without constraints and proceeds to remap each retained posterior samples to an identifiable version using the constraints given by

$$\begin{cases} \min(\alpha_1, \dots, \alpha_N) = 0, \\ \max(\alpha_1 + \beta_1 l_1, \dots, \alpha_N + \beta_N l_N) = 1. \end{cases} \tag{2.14}$$

This means that the earliest common-time coordinate always starts at 0 and the longest recorded signal always has a maximum time-length of u in the common-time dimension. To explain this more thoroughly, let $(\alpha_i^{(b)}, \beta_i^{(b)}, \phi_d^{(b)})$ be the $b-$th posterior samples. Let $\alpha_{\min}^{(b)} = \min\left(\alpha_1^{(b)}, \dots, \alpha_N^{(b)}\right)$ and let $l_{\max}^{(b)} = \max\left(\alpha_1^{(b)} + \beta_1^{(b)} l_1, \dots, \alpha_N^{(b)} + \beta_N^{(b)} l_N\right) - \alpha_{\min}^{(b)}$. Then, each $b-$th posterior samples are remapped to

$$\begin{cases} \alpha_i^{(b)*} = \dfrac{(\alpha_i^{(b)} - \alpha_{min}^{(b)})}{l_{max}^{(b)}}, \\ \beta_i^{(b)*} = \dfrac{\beta_i^{(b)}}{l_{max}^{(b)}}, \\ \phi_d^{(b)*} = l_{max}^{(b)} \phi_d^{(b)} \end{cases} \tag{2.15}$$

such that the identification constraints given by equation (2.14) are obeyed. Consequently, the marginal covariance function in equation (2.13) evaluated at the remapped samples, $(\alpha_i^{(b)*}, \beta_i^{(b)*}, \phi_d^{(b)*})$, equals the one computed with $(\alpha_i^{(b)}, \beta_i^{(b)}, \phi_d^{(b)})$.

2.4.2 Priors specifications

Under the Bayesian framework, a prior distribution must be chosen for the data-specific parameters as well as the general parameters. For each $i-$th recorded record, the vector of data specific parameters are denoted by $\boldsymbol{\eta}_i = (\mu_i, \tau_i, \alpha_i, \beta_i, \boldsymbol{\xi}_i)$ as per equations (2.2), (2.4) and (2.5).The general parameters of the correlation functions are $\boldsymbol{\theta} = (\sigma^2, \lambda, \phi_d, \phi_c, \phi_h, \rho, \gamma)$ as per the definitions in equations (2.7) and (2.9).

The priors for the scalar mean μ_i and variance term τ_i are $\mathrm{N}(m_\mu, v_\mu)$ and $\mathrm{IG}(a_{\tau^2}, b_{\tau^2})$, respectively, for all $i = 1, \ldots, N$. Let the variations between the non-linear time-warping of different recorded signals be random effects such that both $(\zeta_i)_{i=1,\ldots,N}$ and $(\delta_i)_{i=1,\ldots,N}$ are treated as a vector of independent samples from a common distribution. Although each of the warping parameters, ζ_i and δ_i, are supposed to assume values in domain $\mathbb{R}$ theoretically, it is hardly justifiable due to the case of severe time-warping that was shown in Figure 2.3. The severe warping means that the spectral structure of a recorded signal is explained only by an extremely small portion of its finite continuous time-length. Application-wise, this means that one observed vocalization could be, for example, 10 times faster than that of another within the very same portion of time-length. Taking these into account, the parameters are assumed to have a finite domain with $\zeta_i \in (-b_\zeta, b_\zeta)$ and $\delta_i \in (-b_\delta, b_\delta)$ for all $i = 1, \ldots, N$. Given the hyper-parameters, the prior distributions for ζ_i and δ_i are defined to be

$$\begin{aligned} \log\left(\frac{\zeta_i - b_\zeta}{b_\zeta - \zeta_i}\right) &\sim \mathrm{N}(m_\zeta, v_\zeta) \ , \\ \log\left(\frac{\delta_i - b_\delta}{b_\delta - \delta_i}\right) &\sim \mathrm{N}(m_\delta, v_\delta) \end{aligned} \tag{2.16}$$

for all $i = 1, \ldots, N$. The means, m_ζ and m_δ, and the variances, v_ζ and v_δ, are themselves random variables. The selection of the distributions for these hyper-parameters must be conducted with care on account of the restrictions imposed on the domains of the two warping parameters. For example, a large variance v_ζ coupled with $m_\zeta = 0$ induces a prior distribution over ζ_i which is bi-modal with modes close to the boundaries of the domain but with almost zero density everywhere else and so it becomes very informative. This is apparently the opposite of random effect since a large variance is often used to define weakly informative priors even in the non-constrained setting. For this reason, the distributions for

the hyper-parameters are defined to be

$$\begin{aligned} m_\zeta &\sim \mathrm{N}_{(-b_m,b_m)}(m_{0,\zeta}, v_{0,\zeta}), \\ m_\delta &\sim \mathrm{N}_{(-b_m,b_m)}(m_{0,\delta}, v_{0,\delta}), \\ v_\zeta &\sim \mathrm{IG}_{<b_v}(a_{0,\zeta}, b_{0,\zeta}), \\ v_\delta &\sim \mathrm{IG}_{<b_v}(a_{0,\delta}, b_{0,\delta}). \end{aligned} \tag{2.17}$$

The advantage of random effects on the warping parameters is two-fold. A random effect on each set of the warping parameters facilitates the estimation since they are non-stationary parameters, which are generally hard to be estimated. Secondly, as will be shown in the next Section 2.4.3, the random effect facilitates the sampling of from "BASS". The prior distributions for the remaining data-specific parameters of the synchronization function are

$$\begin{aligned} \alpha_i &\sim \mathrm{Uni}(a_\alpha, b_\alpha), \\ \tilde{\beta}_i &= \frac{\beta_i l_i}{1-\alpha_i} \sim \mathrm{Uni}(a_\beta, b_\beta) \end{aligned} \tag{2.18}$$

where $0 \leq a_\beta < b_\beta \leq 1$. The assumption, $0 \leq a_\beta < b_\beta \leq 1$, ensures that the last common-time coordinate will not exceed 1 and that the total possible time-length of the $i-$th recorded signal always assumes values in the interval $[0,1]$ within the common-time dimension.

The priors for the weight λ and for the interaction parameter ρ are both $\mathrm{Uni}(0,1)$. The prior for the variance σ^2 is $\mathrm{IG}(a_\sigma, b_\sigma)$. For the decay parameters and the periodicity γ, the uniform distribution is employed but some care must be taken during the selection of their hyper-parameters. If γ is too small such that the periodic distance is smaller than the minimum distance between two observed time coordinates, then there is no cyclic dependence on the artifacts at all. On the other hand, if it is too large such that $\gamma > 2\max\{l_1, \ldots, l_N\}$, then the circular distance is equivalent to the distance between any two observed time coordinates such that $\Delta_c(|t-t'|;\gamma) = |t-t'|$ as per the definition in equation (2.8), thereby losing its interpretation as a circular distance. In this case, $\mathcal{C}^c(|t-t'|, |h-h'|; \boldsymbol{\theta})$ is no longer a cyclic component and γ is non-identifiable. In fact, even if $\gamma < 2\max\{l_1, \ldots, l_N\}$, the periodicity γ can still be weakly identifiable if $\gamma < 2l_i$ only for a few $i-$th recorded signals. Hence, the prior for γ is set as $\mathrm{Uni}(a_\gamma, b_\gamma)$ with $a_\gamma = 0.02$ which is twice the minimum temporal distance of 0.01 seconds and with b_γ set to be two times the median of the observed time-lengths of the recordings.

The notion of practical range is invoked in order to ensure the identifiability of the decay parameters. In general, the correlation is required to be greater than 0.05 at the minimum

observed distance and less than 0.05 at the maximum observed distance. (Gelfand et al., 2010) Here the so-called practical range is defined to be the observed time-frequency distances at which the correlations in equations (2.7) and (2.9) equal 0.05 in the separable case of $\rho = 0$. Let pr_h, pr_c and pr_d be the practical ranges of the decays, ϕ_h, ϕ_c and ϕ_d, respectively. The practical ranges are given by

$$\begin{aligned} \text{pr}_h &= -\frac{\log(0.05)}{\phi_h} \ , \\ \text{pr}_c &= \frac{1.0-0.05}{0.05\phi_c} \ , \\ \text{pr}_d &= \frac{1.0-0.05}{0.05\phi_d}. \end{aligned} \tag{2.19}$$

Since the minimum and maximum distances in the frequency domain given by data are $|h-h'| = 0.23$ and $(H-1)0.23$, respectively, the prior for the frequency decay is simply defined as $\phi_h \sim \text{Uni}(0.521, 13.025)$ using the above equation (2.19). On the other hand, the minimum and maximum distances between the common-time coordinates and circular-time coordinates are themselves random variables because both equations (2.6) and (2.8) are dependent on the model parameters. This implies that the prior distributions for the decay parameters ϕ_c and ϕ_d must be conditional on the appropriate model parameters. The prior for the circular-time decay ϕ_c is thus defined to be conditional on γ such that

$$\phi_c|\gamma \sim \text{Uni}\left(\frac{1.0-0.05}{0.05(0.5\gamma)}, \frac{1.0-0.05}{0.05\times 0.01}\right) \tag{2.20}$$

where 0.01 and 0.5γ are the minimum and maximum circular distance given by equation (2.7), respectively, for any value of γ. As for the common-time decay ϕ_d, the prior is defined to be conditional on $(\beta_1, \beta_2, \ldots, \beta_N)$ such that

$$\phi_d|\beta_1, \beta_2, \ldots, \beta_N \sim \text{Uni}\left(\frac{1.0-0.05}{0.05\text{max}(\{\beta_i l_i\}_{i=1}^N)}, \frac{1.0-0.05}{0.05 \text{ min } (\{\beta_i l_i/(T_i-1)\}_{i=1}^N)}\right) \tag{2.21}$$

where $\beta_i l_i/(T_i-1)$ and $\beta_i l_i$ are the minimum and maximum common-time distance of the $i-$th recorded signal, respectively. Note that the minimum distance is computed in the absence of time-warping. The rationale for the absence of time-warping is similar to the rationale for the distances being used to define the range of the circular-time decay. For instance, if only one recorded signal is severely warped in time, then only very few time coordinates are going to give a distance that is close to the minimal common-time distance, which makes the common-time decay ϕ_d weakly identifiable.

2.4.3 Sampling from "BASS"

The process $\mathcal{A}_i(t,h)$ is defined on a tri-dimensional space that consists of the common-time dimension, the real time dimension and the frequency dimension. As previously mentioned in Section 2.3.3, the representative sound of a species are the two-dimensional realizations sampled from the posterior distribution

$$\mathcal{A}_{N+1}(t,h)|\boldsymbol{y},\boldsymbol{\chi}_{N+1} \in D. \tag{2.22}$$

Let $\boldsymbol{A}_{N+1}$ be the discrete realizations of $\mathcal{A}_{N+1}(t,h)$ at a given set of time-frequency coordinates To simplify notations, let $\boldsymbol{\eta} = (\boldsymbol{\eta}_i)_{i=1}^N$ be the collection of all data-specific parameters where $\boldsymbol{\eta}_i = (\mu_i, \tau_i, \alpha_i, \beta_i, \boldsymbol{\xi}_i)$ for the $i-$th recorded signal and let $\boldsymbol{\mu} = (\mu_i \mathbf{1}_{T_i H})_{i=1,\dots,N}$ be the collection of scalar means for the collection of signals $\boldsymbol{y}$ where $\mu_i \mathbf{1}_{T_i H}$ is the vector of scalar means for the $i-$th recorded signal $\boldsymbol{y}_i$. The joint distribution is

$$\begin{pmatrix} \boldsymbol{A}_{N+1} \\ \boldsymbol{y} \end{pmatrix} \Bigg| \boldsymbol{\eta}, \boldsymbol{\theta} \sim \mathrm{GP}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{A,y} \\ \boldsymbol{\Sigma}_{A,y}^\top & \boldsymbol{\Sigma}_y \end{pmatrix} \right) \tag{2.23}$$

where $\boldsymbol{\Sigma}_y$ is the cross-covariance matrix between the observed data given by equation (2.12), while $\boldsymbol{\Sigma}_{A,y}$ and $\boldsymbol{\Sigma}_A$ are given by equation (2.13). The conditional distriubtion of $\boldsymbol{A}_{N+1}$ can then be derived from equation (2.23) using standard results from multivariate Gaussian distribution which results in

$$\boldsymbol{A}_{N+1}|\boldsymbol{y},\boldsymbol{\eta},\boldsymbol{\theta} \sim \mathrm{GP}\left(\boldsymbol{\Sigma}_{A,y}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\Sigma}_y - \boldsymbol{\mu}), \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,y}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{A,y}^\top\right). \tag{2.24}$$

Equation (2.24) can then be used to sample from equation (2.22) since the latter is conditional on the former over the constraints $\boldsymbol{\chi}_{N+1} \in D$ in equation (2.11). Taking into consideration the identifiability constraints specified in equation (2.14) and the finite domains of the periodicity γ, the conditioning implies that $\alpha_{N+1} = 0, \beta_{N+1} l_{N+1} = 1$ and $l_{N+1} > b_\gamma$ where b_γ is the right end limit of the uniform prior for γ. This means the posterior distribution of interest is

$$\mathcal{A}_{N+1}(t,h)|\boldsymbol{y}, \alpha_{N+1} = 0, \beta_{N+1} l_{N+1} = 1.$$

which can be marginalized with respect to the warping parameters. The random effects on the warping parameters specified in equation (2.16) are thus exploited for the sampling from

the posterior of interest by a standard Monte Carlo procedure with

$$f\left(\boldsymbol{A}_{N+1}|\boldsymbol{y},\alpha_{N+1}=0,\beta_{N+1}l_{N+1}=1\right)=$$
$$\int\int\int f\left(\boldsymbol{A}_{N+1}|\boldsymbol{y},\alpha_{N+1}=0,\beta_{N+1}l_{N+1}=1,\boldsymbol{\xi}_{N+1}\right)f\left(\boldsymbol{\xi}_{N+1}|\boldsymbol{\eta},\boldsymbol{\theta}\right)f\left(\boldsymbol{\eta},\boldsymbol{\theta}|\boldsymbol{y}\right)\partial\boldsymbol{\eta}\partial\boldsymbol{\theta}\partial\boldsymbol{\xi}_{N+1}$$

2.4.4 Nearest neighbors Gaussian process

The computation of the multivariate Gaussian likelihood of the marginal model requires the inversion of the covariance matrix of dimension $\prod_{i=1}^{N}T_iH$, which is computationally infeasible and is known as the Big N Problem. The NNGP method from the work of Datta et al. (2016a) is therefore adopted in order to counter the Big N Problem.

Let $n_i=T_i\times H$ be the number of sound points in the $i-$th recorded signal $\boldsymbol{y}_i$. Let $y_{i,j}\in\boldsymbol{y}_i$ be the $j-$th element of the $i-$th recorded signal $\boldsymbol{y}_i$ and let $\boldsymbol{y}_i^{1:j}$ be the vector composed of the first $j-$th elements in $\boldsymbol{y}_i$. Define $\mathcal{P}$ as an arbitrary permutation of the integers in $\{1,2,\ldots,N\}$ and let p_i be the $i-$th element of $\mathcal{P}$. Let $\boldsymbol{y}_{\{p_1,\ldots,p_k\}}=\{\boldsymbol{y}_{p_1},\ldots,\boldsymbol{y}_{p_k}\}$ for any k. Then, the joint density of all recorded signals can be decomposed as

$$f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})=\prod_{j=1}^{n_{p_1}}f(y_{p_1,j}|\boldsymbol{y}_{p_1}^{1:j-1},\boldsymbol{\theta},\boldsymbol{\eta})\prod_{i=2}^{N}\prod_{j=1}^{n_{p_i}}f(y_{p_i,j}|\boldsymbol{y}_{p_i}^{1:j-1},\boldsymbol{y}_{\{p_1,\ldots,p_{i-1}\}},\boldsymbol{\theta},\boldsymbol{\eta}) \tag{2.25}$$

which is valid for any permutation. The idea of the NNGP is that, if the covariance function is monotonic with respect to the distances, then only the immediate neighborhoods rather than the entire conditional sets are necessary to approximate the likelihoods as the closest neighbors are strongly correlated with the observations. Let $\mathcal{N}_{p_i,j}$ be a subset of variables in the conditional set of $y_{p_i,j}$. The joint density in equation (2.25) can then be approximated by

$$f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})\approx\prod_{i=1}^{N}\prod_{j=1}^{n_{p_i}}f(y_{p_i,j}|\mathcal{N}_{p_i,j},\boldsymbol{\theta},\boldsymbol{\eta}) \tag{2.26}$$

where $\mathcal{N}_{p_1,1}=\emptyset$. The elements in $\mathcal{N}_{p_i,j}$ are called the neighbors of $y_{p_i,j}$ and the set $\mathcal{N}_{p_i,j}$ is called the neighbor set. To have a good approximation, the elements in the neighbor set must have a high correlation with $y_{p_i,j}$.

The maximum number of neighbors in the neighbor set is set as $4k$ and the neighbor set is divided into four distinct groups with each group having maximum k elements. The elements of the four distinct groups are then selected using the correlation functions, $C^g(\cdot,\cdot;\boldsymbol{\theta})$ and $C^c(\cdot,\cdot;\boldsymbol{\theta})$, in equations (2.7) and (2.9), as well as the dependencies between sound points

within the very same and previous spectrogram. Using $C^g(\cdot,\cdot;\boldsymbol{\theta})$, define $\mathcal{N}_{p_i,j}^{p_i,g}$ as the group that contains k elements from the p_i-th signal that are most strongly correlated with each other and define $\mathcal{N}_{p_i,j}^{p_{i-1},g}$ as the group that contains k elements from the $p_{i-1}-$th signal that are most strongly correlated with each other. Similarly using $C^c(\cdot,\cdot;\boldsymbol{\theta})$, let $\mathcal{N}_{p_i,j}^{p_i,c}$ and $\mathcal{N}_{p_i,j}^{p_{i-1},c}$ be the groups in which the elements are selected from the p_i-th signal and $p_{i-1}-$th signal, respectively. The full neighbor set is then defined as

$$\mathcal{N}_{p_i,j} = \mathcal{N}_{p_i,j}^{p_i,g} \cup \mathcal{N}_{p_i,j}^{p_{i-1},g} \cup \mathcal{N}_{p_i,j}^{p_i,c} \cup \mathcal{N}_{p_i,j}^{p_{i-1},c}. \tag{2.27}$$

If $i = 1$, then $\mathcal{N}_{p_i,j}^{p_{i-1},g} = \mathcal{N}_{p_i,j}^{p_{i-1},c} = \emptyset$. Since the correlation functions are dependent on unknown parameters, the neighbor sets must be adapted in accordance with the changing values of the parameters during model fitting. Additionally, the arbitrary permutation $\mathcal{P}$ is treated as an additional parameter with a uniform prior over the space of permutations and is updated by a Metropolis step.

2.5 Results of application on real bio-acoustics

The results obtained by the application of the model on the real data set are shown in this section. The maximum number of neighbors in the NNGP approximation is set as $4k = 40$. Since the recorded signals that are shorter in time convey much less information and are much more difficult to analyze, the bio-acoustic analysis in this paper focuses exclusively on the 100 longest recorded signals of each species. The model is applied on a set of recorded signals from each of 8 different species. Each of the 8 models is estimated with 60000 iterations. The burning is set at 4800 and the thinning is set at 6. A total of 2000 posterior samples are obtained from the inference.

The hyperparameters of the prior distributions for the warping parameters in equation (2.17) are set as $m_{0,\zeta} = m_{0,\delta} = 0$ and $v_{0,\zeta} = v_{0,\delta} = 0.75$ such that

$$\begin{aligned} m_\zeta &\sim \mathrm{N}_{(-5,5)}(0, 0.75), \\ m_\delta &\sim \mathrm{N}_{(-5,5)}(0, 0.75), \\ v_\zeta &\sim \mathrm{IG}_{<0.75}(0.01, 0.01), \\ v_\delta &\sim \mathrm{IG}_{<0.75}(0.01, 0.01). \end{aligned}$$

For the translation and scaling paraters, the prior distributions in equation (2.18) are set as $\alpha_i \sim \mathrm{Uni}(0, 0.2)$ and $\tilde{\beta}_i \sim \mathrm{U}(0.75, 1)$. For the remaining data-specific parameters, the prior

distributions are set to be $\tau_i^2 \sim \text{IG}(1.0, 1.0)$, and $\mu_i \sim \text{N}(0, 100000)$. As for the variance, the prior is set as $\sigma^2 \sim \text{IG}(1.0, 1.0)$.

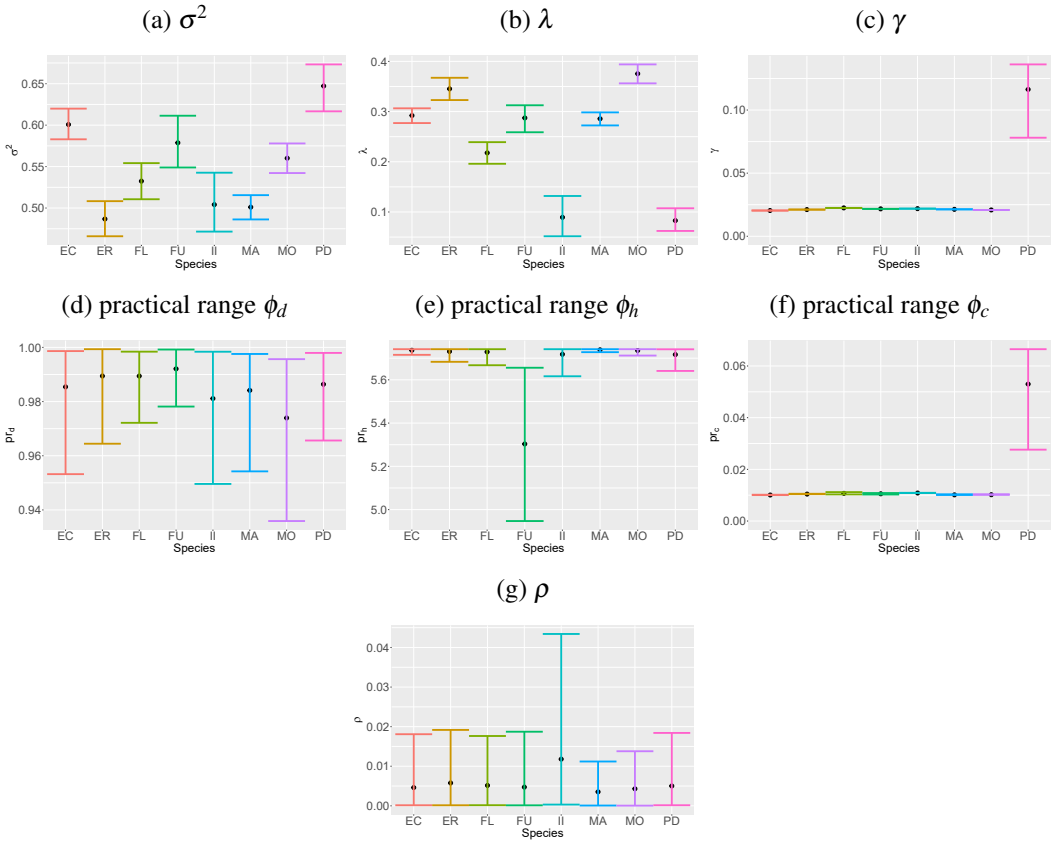
Figure 2.5 displays the 95% credible intervals (CIs) of the general parameters in $\boldsymbol{\theta}$ of each species. Note that the CIs of the decay parameters, ϕ_d, ϕ_h and ϕ_c in Figure 2.5 (d), (e) and (f) are depicted in terms of the practical ranges defined in equation (2.19). In Figure 2.5 (b), the CIs of the weights λ are all consistently below 0.5 which mean that the contribution of the cyclic component is smaller comparing to the common-time component for all species. Figure 2.5 (c) shows that, with the exception of PD, the cyclic components for all species exhibit a period of $\gamma \approx 0.025$. This is unsurprising due to the fact that the spectrograms are obtained from STFT with the same time-step of 0.01 seconds and the fact that the cyclic component is interpreted as an artifact, as explained in Section 2.2 and Section 2.3.2. The deviance of PD from the rest of the species might be caused by the extremely short time lengths of their signals, which might have hindered the learning of these parameters. Indeed, going back to Table 2.1, no signals of PD exceed 0.2 seconds and only two signals exceed 0.1 seconds in time length. In Figure 2.5 (d) and (e), the large values of the practical ranges of ϕ_d and ϕ_h for all species indicate a strong dependence on both the common-time and log-frequency distances. Figure 2.5 (g) indicates that the non-separability of the covariance is almost negligible.

Figure 2.6 displays the distributions of the posterior means of the data-specific parameters, α_i, $\tilde{\beta}_i$, μ_i, and τ_i^2, of all $i-$th recorded signals for each species. The distributions are represented by box-plots, which show that the misalignment, means and variability are rather consistent across species.

The warping function $\Omega_i(q)$ of each $i-$th recorded signal is computed using the posterior means of $\boldsymbol{\xi}_i = \{\zeta_i, \delta_i\}$ for all $i = 1, \ldots, N$ of each species. The posterior predictive distributions of the warping functions are then plotted in Figure 2.7 with the solid lines depicting the posterior means, the dashed lines depicting the posterior medians, and the two shaded areas depicting the 50% and 95% CIs, respectively. As explained in Section 2.3.1, the absence of non-linear warping will result in a 45 degree linear line. Notably, the CIs in Figure 2.7 (b), (c), (e) and (h) reveal that there exists distinctive non-linear warping in the sounds of species ER, FL, II, and PD. However, the results for PD might not be reliable due to the aforementioned limitations in the availability of signals that are longer in time. For the other species in Figure 2.7 (a), (d), (f) and (g), it appears that the warping functions are primarily asymmetric fluctuations around the linear line.

The posterior means and variances of the posterior distribution of the representative sound $\mathcal{A}_{N+1}(t,h)$ in equation (2.23) are derived under the assumption that the time length

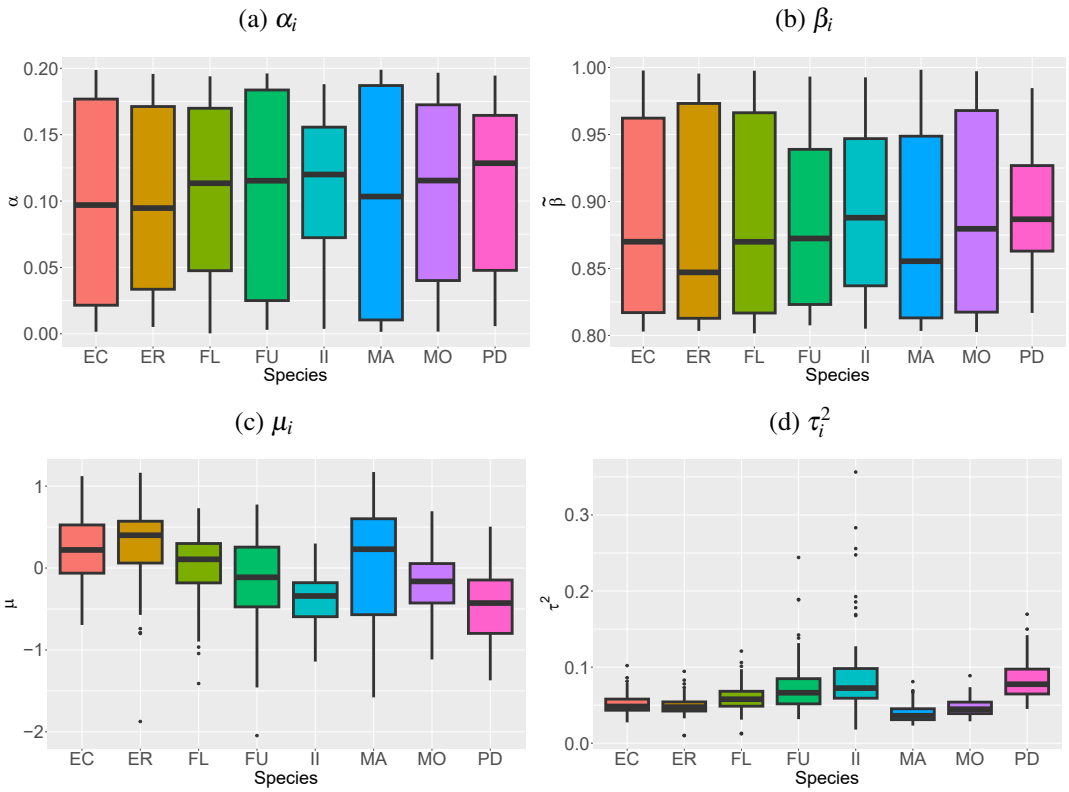
Figure 2.5 The 95% credible intervals of the general parameters $\boldsymbol{\theta}$.



l_{N+1} equals the median of the observed time lengths. The time coordinates for $\mathcal{A}_{N+1}(t,h)$ are equally spaced by a constant time step of 0.01 seconds in the exact same way that the data is observed. The log-frequency coordinates also have the same number and positions as observed. The posterior means and variances are plotted in Figure 2.8 and Figure 2.9, respectively. Looking at Figure 2.8, significant differences between the species can be spotted at 4.14 and at the range of $6.9-8.75$ in log-frequencies. All species possess strong sound intensities at the range of $5.06-6.65$ in log-frequencies, which generally persist across time for all species except EC, MA and PD. Instead, the strong intensities across time gradually decrease for EC and MA, but gradually increase for PD. Meanwhile, the posterior variances in Figure 2.9 are approximately constant across the time-frequency grid for all species. The occasional large values at the first time coordinates can be attributed to the synchronization function which leads to fewer observations for the estimation at these coordinates. The apparent cyclic pattern can be attributed to the cyclic component of the process.

The distance between the posterior means of $\mathcal{A}_{N+1}(t,h)$ are computed as a mean to compare the representative sounds of each species. All $\mathcal{A}_{N+1}(t,h)$ are evaluated on the

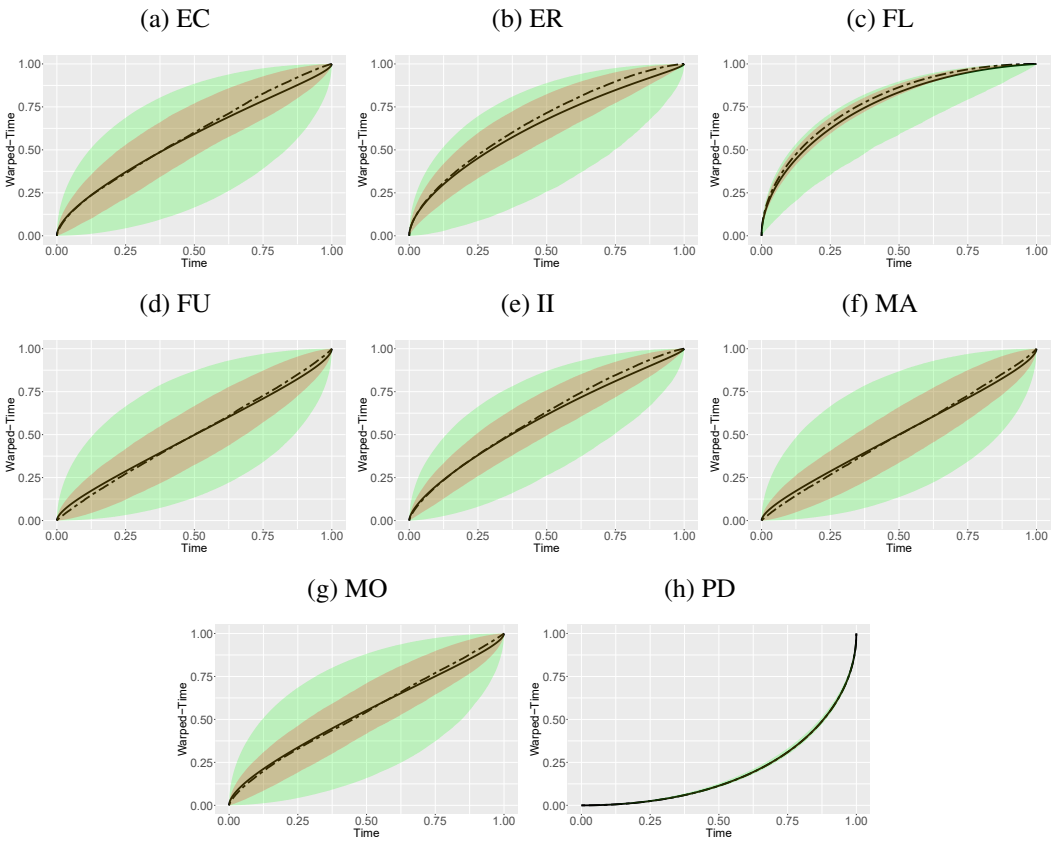
Figure 2.6 Posterior means of α_i, β_i, μ_i and τ_i^2 of all $i-$th recorded signals.



same set of time and frequency coordinates. The distance between two species is then given by the mean square difference between the two sets of sound intensity points over the same set of time and frequency coordinates. Figure 2.10 (a) is the resulted distance matrix and Figure 2.10 (b) is the phylogenetic tree that is associated with the distance matrix. The phylogenetic tree is computed using the R package, phylogram, given by the work of Wilkinson and Davy (2018).

Cross-validation is implemented in order to determine whether all components of the proposed model are necessary. Specifically, 5% of the time-frequency coordinates from each spectrogram of each species is removed. The model is then fit to the remaining 95% of the data under different settings: (i.) the proposed model specified by equations (2.2) and (2.3); (ii.) the model in the absence of non-linear warping where $\Omega_i(q) = 1 \; \forall \; i$ (NoWarp); (iii.) the model in the absence of the cyclic component (NoCirc); and (iv.) the model without temporal alignment such that $\alpha_i = 0$ and $\beta_i = 1$ (NoAl). The Continuous Ranked Probability Score (CRPS) are computed using the holdout sample data and the results are presented in Table 2.2. The CRPS of the best model is highlighted in bold. The proposed model exhibits the lowest CRPS in 5 out of the 8 species. In general, the temporal alignment and

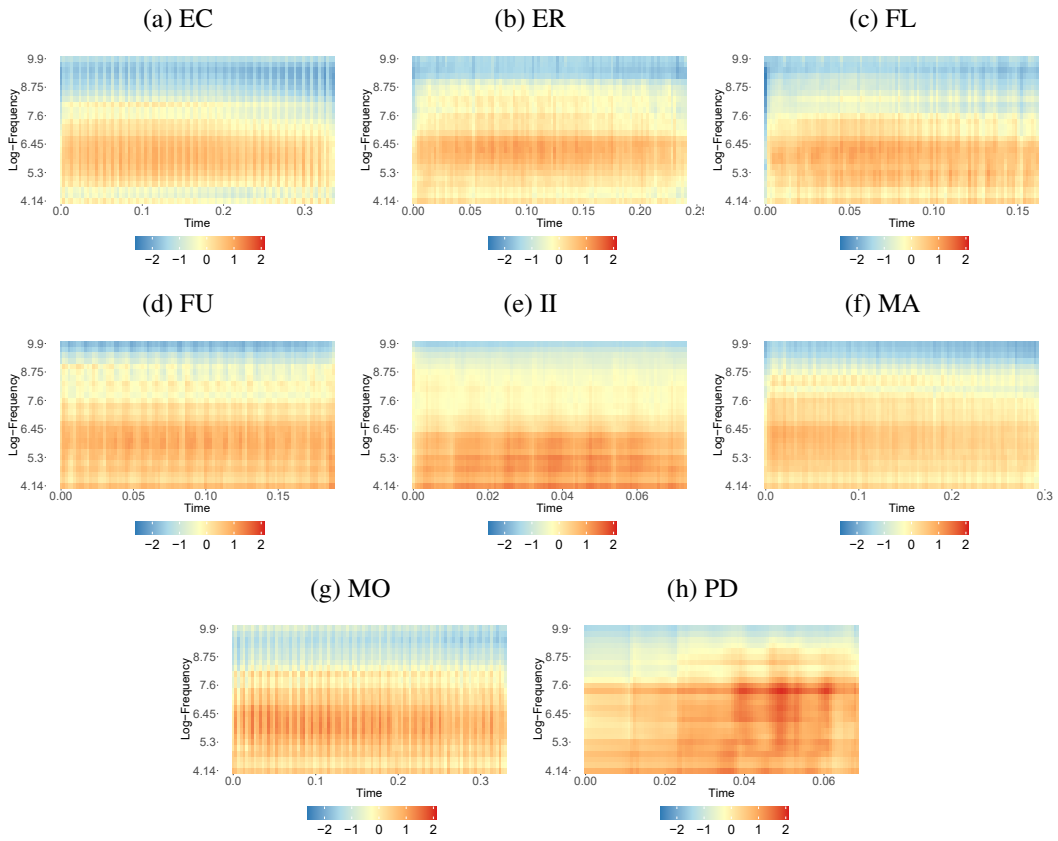
Figure 2.7 $\Omega_i(q)$ given by posterior values of $\boldsymbol{\xi}_i = \{\zeta_i, \delta_i\}$ of all $i-$th recorded signals.



non-linear warping produce smaller CRPS values. However, in future it might be a better option to estimate the model without these components, especially when the data is not highly informative, since the differences are not substantial. Note that for the species PD, not only is the proposed model not preferable, but its CRPS is very small for NoAl which yields the best index. This can be attributed to the short time lengths of the recorded signals which do not allow for temporal alignment. Interestingly, the removal of the the cyclic component results in a worsened CRPS for all species with a significant increase in the margin. This might testify the effects of the periodic artifacts from STFT on the sound intensities over the spectrograms.

The results in this work generally confirm previous findings on the lemur species. The variation in log-frequency distributions across the species reflects previous findings on the Eulemur species. According to Gamba et al. (2012, 2016), distinctive species-specific traits were found in low-pitched "grunt" vocalizations. In here, the log-frequency distribution reflects the formant distribution of the respective species, with the frequency variation in ER, FU, and MA being higher than those in the other species at higher frequencies. This

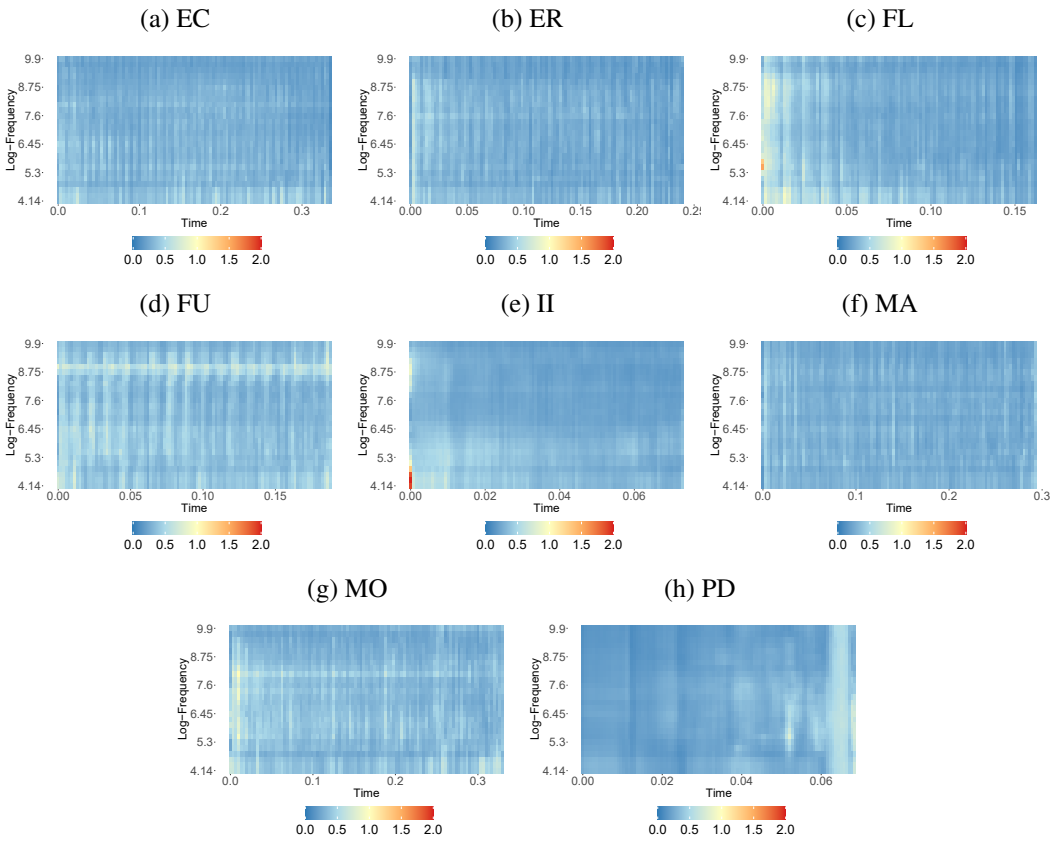
Figure 2.8 Posterior means of $\mathcal{A}_{N+1}(t,h)$.



agrees with the findings on formant distribution and variation across lemur species in the paper by Gamba et al. (2012), in which the fourth formant in FU, MA and ER was estimated and measured at frequency 6.

The results in this work also demonstrate the critical role played by communicative signals in species-specific interactions. The phylogenetic tree generated in this study show a relatively different picture from previous investigations. According to the distance matrix in Figure 2.8 (a), MA and FL are more dissimilar than in the other taxa, which is surprising considering that they are the two former subspecies of the Eulemur macaco. However, a study by Gamba and Giacoma (2008) has shown differences in the "grunt" vocalizations among individuals of what were then the two subspecies of the Eulemur macaco. This suggests that divergence across vocalizations may only partially map the differences at the phylogenetic level, which is in agreement with the findings by Macedonia and Stanger (1994). Vocalizations in MA and FL may have diverged to prevent hybridization, which has been observed in the wild. Gamba and Giacoma (2008) interpreted this divergence as informative differences for morphological differences that can be mainly attributed to

Figure 2.9 Posterior variances of $\mathcal{A}_{N+1}(t,h)$.



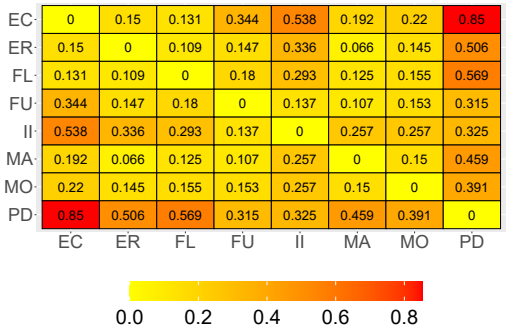
formants and fundamental frequency. It should be noted that phylogenetic reconstructions from the analyses of the lemuri are often based on very different data sets and that analyses based on communication data are rare. Research on DNA data must still establish a clear connection in some clades. (DelPero et al., 2006)

2.6 Discussion

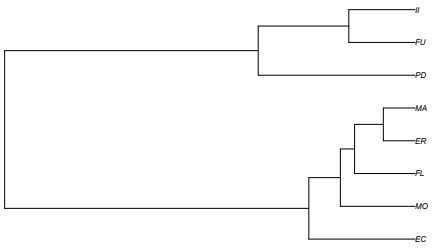
This work has presented a spatio-temporal model for bio-acoustic data with non-stationary temporal patterns and periodic artifacts. The model combines the novel idea of time synchronization with several techniques, including NNGP, in order to describe the complex structures that are intrinsic to bio-acoustic data as well as to handle the sheer size of the available data set. The sources of temporal non-stationarity are the unique time length of each recorded signal and the ensuing non-linear distortions of the signals' sound intensities with respect to each other. This distortion constitutes the most problematic issue that is resolved by the proposed model. The major novelty of this work is the construction of the synchronization

Figure 2.10 Quadratic distance between the posterior means of $\mathcal{A}_{N+1}(t,h)$ and associated phylogenetic tree.

(a) Quadratic distance



(b) Phylogenetic tree



function that is used to project the real time coordinates of the recorded signals into the latent common-time dimension wherein the latent process for the species-specific representative acoustic structure, named "BASS", is stationary. The synchronization function provides an easy interpretation on the relationships between the observed time dimension of each recorded signal and the latent common-time dimension. This easy interpretation is attained by means of the translation, scaling and warping parameters that are specific to each individual recording. That is, the major contribution of this work is the construction of a more interpretable transformation that permits the quantification of temporal non-stationarity in bio-acoustic data. The presence of periodic artifacts that arise from STFT is resolved by a circular representation of the observed time coordinates. The computational efficiency of NNGP allows the model to be estimated by a very large data set.

The results of the application on real bio-acoustic data are very promising. The model is able to obtain the species-specific representative acoustic structure over a time-frequency

Table 2.2 CRPS index of each model.

	Prop	NoWarp	NoCirc	NoAl
EC	**0.95**	0.95	1.22	0.97
ER	**0.91**	0.92	1.17	0.98
FL	**0.92**	0.96	1.31	0.94
FU	0.91	1.02	1.33	**0.90**
II	**0.93**	1.06	1.38	0.94
MA	0.82	0.87	1.18	**0.81**
MO	**0.92**	0.91	1.29	0.93
PD	1.17	1.35	1.43	**0.71**

grid, namely the "BASS", which is described by features that are used for cross-species comparison in terms of quadratic distance and the phylogenetic tree. Outputs that are specific to the individual recordings are also successfully captured by the model and temporal non-stationarity is quantified through the warping parameters. Cross-validation has shown that the proposed model is almost always the best for describing the available data set, though not always true for all species. For bio-acoustic investigators, this work offers a tool to look at all the information offered by the full recording, rather than just a few hand-picked pitches after signal decomposition. Information specific to each individual recording such as the scalar mean and the warping parameters might also prove useful for learning about characteristics of individual animals of each species.

The future will find the candidate looking for a proper way to extend the proposed model to become a predictive model for the behavioral call-type labels. Indeed, the proposed model is limited to a specific call-type of a species. The application of this methodology to a larger data set that comprises of different call-type labels will provide a better description of the distances between different species in relation to a phylogenetic tree.

Chapter 3

Anatomic vs ischemia-driven strategies for percutaneous coronary revascularization in chronic coronary syndrome: a network meta-analysis

Background

This chapter is submitted as a journal article:

Giacobbe F., Valente E., Giannino G., Yip H.C. and et al. Anatomic versus ischemia-driven strategies for percutaneous coronary revascularization in chronic coronary syndrome: A network meta-analysis. *Under review.*

The statistical contributions made by Prof. Mauro Gasparini and PhD candidate H.C. Yip is only possible with the medical expertise of Dr. Federico Giacobbe, Dr. Eduardo Valente, Dr. Giuseppe Giannino and the other collaborators. The PhD candidate would like to thank all the collaborators for their invaluable contributions.

3.1 Introduction

Meta-analysis is a statistical methodology that is frequently employed by the medical research communities in order to pool related results from independent sources. Factors intrinsic

to the study design such as sample size of patient groups, features of enrolled patients and hospital management might have substantial impacts on the treatment effects in an individual clinical trial. Meta-analysis allows conflicting conclusions from various studies on medical treatments or interventions to be synthesized through estimating the average treatment effect, thereby producing a more comprehensive picture of the treatment. The tutorial by Normand (1999) provides a protocol for conducting meta-analysis which includes an overview of the relevant analytical and statistical methods. Another overview of the more advanced statistical methods for meta-analysis is available in the tutorial by van Houwelingen et al. (2002).

The network approach to meta-analysis, known as network meta-analysis (NMA), has recently emerged as a popular tool for the integration of results from several randomized controlled trials (RCTs) in which multiple treatments are compared. (Ahn and Kang, 2021; Schwarzer et al., 2015) NMA makes use of both direct and indirect evidence from multi-arms RCTs to achieve simultaneous mixed treatment comparison. The Bayesian framework of NMA is explained in the tutorial by Hu et al. (2020).

This chapter presents a Bayesian NMA in order to compare multiple strategies for treating patients that suffer from chronic coronary syndromes. In contrast to the non-invasive, drug-based treatment called optimal medical therapy, the overall effects of treatment strategies that involve the invasive surgery known as percutaneous coronary intervention (PCI) remain unclear to date. (Boden et al., 2007; Pavasini et al., 2020) In fact, previous RCTs that were conceived to investigate the benefits of different PCI strategies have demonstrated conflicting results. (Bruyne et al., 2012; Maron et al., 2020) This might be due to the fact that the different guidance strategies which PCI depends on have not been thoroughly compared.

A Bayesian NMA is conducted on a set of 18 RCTs data that encompasses a total of 17512 patients. A total of 5 treatments are compared: optimal medical therapy (OMT), PCI guided by angiography (PCI angio), PCI guided by non-invasive ischemia assessment (PCI ischemia), PCI guided by fractional flow reserve (PCI FFR) and PCI guided by instantaneous wave-free ratio (PCI IFR). The primary clinical endpoint concerned by this study is the major adverse clinical events (MACE) as defined by each included RCT in the data set. The secondary clinical endpoints of concern are all cause deaths (All deaths), cardiovascular deaths (CV deaths), myocardial infarction (MI) and unplanned revascularization (Revas).

The sections of this chapter are organized as follow. Section 3.2 provides a detailed descriptions of the available data set. Section 3.3 is an overview of the models and methodologies involved in the Bayesian network meta-analysis. Section 3.4 summarizes the statistical results of the comparison on the angina treatments that concerns the primary clinical endpoint, MACE. Section 3.5 is a discussion about the recent developments on the methodologies

Table 3.1 Labels of treatments and clinical endpoints

treatments	labels	clinical endpoints	labels
OMT	1	MACE	1
PCI angio	2	All cause deaths	2
PCI ischemia	3	CV deaths	3
PCI FFR	4	MI	4
PCI IFR	5	Revas	5

in meta-analysis for mixed treatment comparison. Results that concerns the remaining secondary clinical endpoints are provided in the Appendix B.

3.2 Randomized controlled trial data

The available data set consists a total of 18 RCTs. Each RCT has a control arm and a treatment arm in order to perform pairwise comparison. The follow-up time of the trials range from 0.5 years to 10 years. A total of 5 treatments and 5 clinical endpoints are recorded in this data set. All clinical endpoints are hazardous events. Table 3.1 lists the discrete labels of the treatments and clinical endpoints. A Bayesian NMA is conducted for each clinical endpoint because the number of trials available for each endpoint is different. The number of trials recorded for each of the 5 clinical endpoints is $18, 15, 10, 15$ and 14, respectively.

Table 3.2 provides the proper format of the trial data for clinical endpoint MACE, which is labeled as endpoint 1. The first column of Table 3.2 lists the trials and their corresponding follow-up time in years. Each trial has a control arm or placebo group (arm 1) and a treatment arm (arm 2) for comparison between two treatments. The second column of the table records the results of the control group, whilst the third column of the table records that of the patients group which received the treatment of interest. The last column of the table is the treatment-contrast label indicating which two treatments are being compared in each trial. For example, the first row of Table 3.2 means that trial 1 recorded the comparison between OMT (treatment 1) and PCI angio (treatment 2) with a follow-up period of 1.5 years. By the end of the follow-up period, 25 out of the 164 patients who received OMT ended up with MACE, whilst 36 out of the 177 patients who received PCI angio ended up with MACE. The total number of observed data points available for clinical endpoint 1, MACE, alone is 36 because there is a total of 18 trials with 2 treatment arms each. Subsequently, the total number of observed data points for each of the 5 clinical endpoints is $36, 30, 20, 30$ and 28,

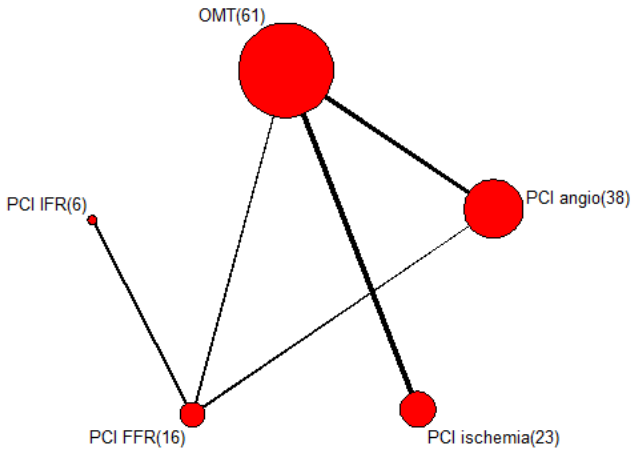
Table 3.2 Trial data available for clinical endpoint 1 (MACE)

trial	follow-up	arm 1	occurrence	sample	arm	occurrence	sample	contrast
1	1.5	1	25	164	2	36	177	1,2
2	1.6	1	5	22	3	3	19	1,3
3	6	1	93	162	3	84	166	1,3
4	3	1	131	389	3	133	388	1,3
5	5	1	471	2591	3	424	2588	1,3
6	2	1	86	441	4	36	447	1,4
7	4.1	1	52	148	2	62	153	1,2
8	4.6	1	213	1138	2	222	1149	1,2
9	10	1	67	105	3	27	96	1,3
10	2.7	1	17	514	2	32	504	1,2
11	10	1	120	203	2	87	205	1,2
12	1	1	6	51	3	21	50	1,3
13	2	1	146	366	2	44	192	1,2
14	1	1	15	101	2	13	104	1,2
15	1	4	83	1250	5	78	1242	4,5
16	0.5	1	4	107	2	5	105	1,2
17	5	2	154	496	4	143	509	2,4
18	1	4	35	1018	5	37	1019	4,5

respectively. Tables similar to Table 3.2 for the remaining secondary clinical endpoints can be found in the Appendix B.

Figure 3.1 is the evidence network plot that shows the treatment-contrasts from the pairwise treatment comparisons of all trials for all 5 clinical endpoints. Each node of the network represents a treatment. The existence of an edge between two nodes indicates that there exists direct pairwise comparison between the two connected treatments in the data set. If there exists direct pairwise comparison between two treatments, then there is direct evidence available for the treatment effects. The thickness of the edge represents the number of existing direct comparisons in the data. The thicker the edge, the higher the number. If direct pairwise comparison is not available between two treatments at all, then there exists indirect evidence based on the direct evidence. The number in the bracket of each node is the total number of the specified treatment in the full data set. For example, the total number of OMT (treatment 1) available for each of the 5 clinical endpoints is $15, 13, 8, 13$ and 12, respectively. Thus, the number in the bracket of OMT is the sum total of 61. This network representation of the RCTs data will prove to be useful in the next section when the model for this data set and the Bayesian NMA are explained.

Figure 3.1 Evidence network plot of the treatment-contrasts from the full data set.



3.3 Methods

In the classical generalized linear model framework for meta-analysis, the number of occurrence of the clinical endpoint in each arm is usually assumed to be Binomial with the unknown probability of occurrence as a latent parameter. Essentially, this is the logistic regression with the probability of occurrence in the logit scale as the dependent variable. However, this is not suitable for the RCTs data set here in this chapter as it is noticeable that the follow-up periods are vastly different from each other. In fact, the shortest follow-up period is 0.5 years while the longest is 10 years for in Table 3.2. This implies that the lengths of the follow-up period have statistical impact on the survival outcomes of the patients and thus the recorded number of occurrence of the clinical endpoint. A detailed descriptions of the suitable statistical methodologies involved in the case study on the available RCTs data set is provided in this section.

3.3.1 Cox proportional hazards model

The rate of occurrence of an event is almost always treated as a Poisson variable in the regression setting. In this specific case, the number of occurrence can be treated as though it is a Poisson variable by means of a link function that can account for the effects of the follow-up period. This method was applied on the medical studies of death rates by Berry (1983) and Frome (1983). The application of this method in the context of meta-analysis of

clinical trial data with multiple follow-up times can be found in the work of Lu et al. (2007). Following the work of Lu et al. (2007), a similar model is implemented in this work.

Let N be the total number of trials available for a clinical endpoint and let the trial number be denoted by $i \in \{1,2,...,N\}$. Let T be the total number of treatments such that $\{1,\ldots,T\}$ is the set of discrete labels for the treatments. Let k denote the label of a treatment such that $k \in \{1,\ldots,T\}$. Let $n_{i,k}$ be the total number of patients that were assigned to treatment k in the $i-$th trial and let $y_{i,k}$ be the number of patients that ended up with the clinical endpoint at the end of the follow-up period. Let $p_{i,k}$ be the probability of the clinical endpoint's occurrence for the patients after receiving the treatment k in the $i-$th trial. The classical Binomial model assumes that

$$y_{i,k}|n_{i,k},p_{i,k} \sim \text{Bin}(n_{i,k},p_{i,k}), \tag{3.1}$$

which is not directly applicable in the regression setting of this case since the follow-up periods of the RCTs need to be accounted for. Let f_i be the follow-up period of the $i-$th trial. It is assumed that the repeated occurrences of the clinical endpoint for each patient is a Poisson process such that

$$y_{i,k}|\lambda_{i,k} \sim \text{Poi}(\lambda_{i,k}f_i), \tag{3.2}$$

where $\lambda_{i,k}$ is the rate at which a clinical endpoint occurred. Let $t_{i,k}$ be the first arrival time in the Poisson process specified above such that

$$t_{i,k} \sim \text{Exp}(\lambda_{i,k}) \tag{3.3}$$

and that the probability $p_{i,k}$ can be written in terms of the rate and the follow-up period by

$$p_{i,k} = \mathbb{P}(t_{i,k} < f_i) = 1 - \exp(-\lambda_{i,k}f_i). \tag{3.4}$$

The treatment of interest that is under investigation in each trial is in contrast to a baseline treatment which was received by patients in the control group. Note that the baseline treatment differs across different trials as per Table 3.2. Let the baseline treatment of each $i-$th trial be denoted by $b(i) = b$ for simplification of notation. The probability $p_{i,k}$ of the clinical endpoint's occurrence after having received treatment k in the $i-$th trial is modeled by

$$\begin{aligned}\theta_{i,k} &= g(p_{i,k}) \\ &= \log(f_i) + \log(\lambda_{i,k}) \\ &= \log(f_i) + \mu_i + \delta_{i,(b,x)}\mathbb{1}(x=k),\end{aligned} \tag{3.5}$$

where $\theta_{i,k}$ is the log cumulative hazard, μ_i is the trial-specific baseline effect, $\delta_{i,(b,x)}$ is the trial-specific log hazard ratio of treatment y to baseline treatment b and $g(p) = \log(-\log(1-p))$ is the clog-log link function that can be given by the inverse of equation (3.4). This is also known as the Cox proportional hazards model. In meta-analysis, the trial-specific log hazard ratio $\delta_{i,(b,x)}$ is called the relative effect of a treatment and the expected log hazard ratio is called the expected relative effect. In Bayesian NMA, the trial-specific relative effects, denoted by $\{\delta_{i,(b,x)}\}_{i=1,\ldots,N}$, are assumed to be exchangeable with the expected relative effects, denoted by $d_{b,x}$. These expected relative effects are the parameters of interests.

3.3.2 Evidence consistency

As explained by Lu and Ades (2006), direct evidence and indirect evidence is combined for mixed treatment comparison through the key assumption of evidence consistency. Let r be the fixed label of the reference treatment such that $r \in \{1,\ldots,T\}$. Let the expected relative effect of treatment k to the fixed reference r be denoted by $d_{r,k}$ where $r \neq k$. Theses are called the basic parameters and their representation in an evidence network is a spanning tree. Thus, there is a total of $T-1$ basic parameters. The basic parameters are the main parameters of interest. By convention, the label of the reference treatment is usually set as $r = 1$ such that $r < k$ for all $k \in \{2,\ldots,T\}$. The remaining expected relative effects are called the functional parameters and can be represented as linear functions of the basic parameters by

$$d_{k,k'} = d_{r,k} - d_{r,k'}, \tag{3.6}$$

where $r < k < k'$. This is known as the consistency equation. Models that follow the consistency equation are said to be under the assumption that all available evidence about the basic and functional parameters is consistent. The assumption of evidence consistency can be illustrated by the contrasts of the pairwise treatment comparisons in the evidence network in Figure 3.1. If the information given by the direct evidence from treatment 1 (OMT) versus treatment 2 (PCI angio) is inconsistent with the information given by the indirect evidence from treatment 1 (OMT) versus treatment 5 (PCI IFR), then $d_{2,5} \neq d_{1,5} - d_{1,2}$ and it is said that the evidence is inconsistent.

The presence of evidence inconsistency in two-arms trial data can be detected by means of the Bucher's method for single-loop evidence that is defined in the work of Dias et al. (2011b). Let $\widehat{d}^{dir}_{k,k'}$ be the direct estimate of $d_{k,k'}$ that is obtained from direct evidence that exists in the trial data. Define $\widehat{d}^{ind}_{k,k'} = \widehat{d}^{dir}_{r,k'} - \widehat{d}^{dir}_{r,k}$ as the indirect estimate using the direct estimates as well as the consistency equation (3.6). The Bucher's method states that a hypothesis test for the

non-existence of inconsistency can be conducted using the z-score test statistic given by

$$z_{k,k'} = \frac{\widehat{d}^{dir}_{k,k'} - \widehat{d}^{ind}_{k,k'}}{\sqrt{\mathrm{Var}(\widehat{d}^{dir}_{k,k'}) + \mathrm{Var}(\widehat{d}^{ind}_{k,k'})}}. \tag{3.7}$$

The null hypothesis of the Bucher's method states that there is no difference between the estimates obtained from direct evidence, $\widehat{d}^{dir}_{k,k'}$, and the estimates obtained through indirect evidence, $\widehat{d}^{ind}_{k,k'}$. The issue of evidence inconsistency will be discussed later in Section 3.5.

3.3.3 Fixed effects, homogeneous random effects & heterogenous random effects

The model in this case study abides by the same assumption of evidence consistency as per equation (3.6). However, in order to obtain the statistically best-fit results, the model was implemented under three different assumptions on the variation: the fixed effect, the homogeneous random effect and the heterogeneous random effect. As explained by Dias et al. (2011a) and Hu et al. (2020), these three versions of the Cox proportional hazards model differ in the assumed sources of variation within the Bayesian framework.

In the fixed effect (FE) model, the only source of variation of the trial-specific relative effects is the random effects of the basic parameters. Thus, the trial-specific relative effects are directly specified by the basic parameters and the consistency equation (3.6) in the FE model, which is

$$\begin{aligned} \theta_{i,k} &= \log(f_i) + \mu_i + \delta_{i,(b,x)}\mathbb{1}(x = k), \\ \delta_{i,(b,x)} &= d_{b,x}, \\ d_{b,x} &= d_{r,x} - d_{r,b}, \end{aligned} \tag{3.8}$$

for all $i = 1, \ldots, N$. This is equivalent to saying that there is no variation across the independent trials and that the underlying true treatment effects can be directly obtained from all trials regardless of their intrinsic differences in study design, sample sizes or other factors. The trial-specific relative effects therefore equal their corresponding expected relative effects and the only variation comes solely from the priors for the basic parameters.

The homogeneous random effect (RE-Hom) model admits that the trial-specific relative effects should vary across the trials. The source of variation to the relative effects is no longer just the Bayesian prior for the basic parameters in this case. Instead, it assumes that the

random effects for all treatment-contrasts is homogeneous. The RE-Hom model is

$$\begin{aligned} \theta_{i,k} &= \log(f_i) + \mu_i + \delta_{i,(b,x)} \mathbb{1}(x = k), \\ \delta_{i,(b,x)} &\sim \mathrm{N}(d_{b,x}, \sigma^2), \\ d_{b,x} &= d_{r,x} - d_{r,b}, \end{aligned} \tag{3.9}$$

for all $i = 1, \ldots, N$. The source of variation for the trial-specific relative effects $\delta_{i,(b,x)}$ here in the RE-Hom model comes from across the independent trials rather than the different treatment-contrasts of the pairwise comparison themselves.

The heterogeneous random effect (RE-Het) model assumes both random effects for the trial-specific relative effects as well as statistical heterogeneity. This statistical heterogeneity refers to the variation that arises from the different pairwise contrasts of treatment comparison such that the trial-specific relative treatment effect $\delta_{i,(k,k')}$ is specified by

$$\delta_{i,(k,k')} \sim \mathrm{N}(d_{k,k'}, \sigma^2_{k,k'}),$$

for all $k \neq k'$ and $k, k' \in \{1, \ldots, T\}$. This formulation implies that there exists a covariance structure associated with the vector of basic and functional parameters. Let $\rho^{(r)}_{k,k'}$ where $r < k < k'$ be the correlation coefficient that indicates how the two basic parameters $d_{r,k}$ and $d_{r,k'}$ are related to each other with respect to the reference treatment. In a mixed treatment comparison with only 2 arms present in all trials, the RE-Het model is

$$\begin{aligned} \theta_{i,k} &= \log(f_i) + \mu_i + \delta_{i,(b,x)} \mathbb{1}(x = k), \\ \delta_{i,(b,x)} &\sim \mathrm{N}(d_{b,x}, \sigma^2_{b,x}), \\ \sigma^2_{b,x} &= \sigma^2_{r,x} + \sigma^2_{r,b} - 2\rho^{(r)}_{b,x}\sigma_{r,b}\sigma_{r,x}, \\ d_{b,x} &= d_{r,x} - d_{r,b}, \end{aligned} \tag{3.10}$$

for all $i = 1, \ldots, N$. This means that the source of variation for the trial-specific relative effects $\delta_{i,(b,x)}$ here in the RE-Het model comes from across the independent trials as well as the different treatment-contrasts of the pairwise comparison between the reference treatment and other treatments of interests. (Lu and Ades, 2006)

In the Bayesian framework, the variation between the basic parameters and their associated parameters are assumed to be random effects. The priors for each basic parameter is set as $d_{r,k} \sim \mathrm{N}(m_d, v_d)$ in all three models. The prior for the standard deviation σ of the RE-Hom model in equation (3.9) is set as $\mathrm{Uni}(a_\sigma, b_\sigma)$. As for the RE-Het model in equation (3.10),

the prior for the standard deviations $\sigma_{r,x}$ and the correlations $\rho^{(r)}_{x,x'}$ are set to be Uni(a_σ, b_σ) and Uni(a_ρ, b_ρ), respectively.

3.3.4 Diagnostic deviance statistics

Since all the models specified in Section 3.3.3 will be implemented by Gibbs sampling later in the case study on angina treatments in Section 3.4, model comparison needs to be conducted. The best model for the data is chosen according to model fit and complexity, which are measured by the diagnostic deviance statistics that are specified in the work of Dias et al. (2011a).

The goodness-of-fit of a model is measured by the overall residual deviance. The overall residual deviance is the summation of the residual deviance of all observed data points since the observed data points are assumed to be independent and identically distributed. Let $\widehat{p}_{i,k}$ be the estimate of $p_{i,k}$ at each iteration of Gibbs sampling. Let $\widehat{y}_{i,k} = \widehat{p}_{i,k} n_{i,k}$ be the number of occurrence of the clinical outcomes calculated at each iteration. Let $dev_{i,k}$ be the residual deviance of each observed data point. Let D be the overall residual deviance at each iteration, then

$$\begin{aligned} D &= \sum_{i=1}^{n} \sum_{k=1}^{m} dev_{i,k} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{m} 2 \left[y_{i,k} \log\left(\frac{y_{i,k}}{\widehat{y}_{i,k}}\right) + (n_{i,k} - y_{i,k}) \log\left(\frac{n_{i,k} - y_{i,k}}{n_{i,k} - \widehat{y}_{i,k}}\right) \right], \end{aligned} \tag{3.11}$$

for the Binomial likelihood. This can be obtained by the estimates of the Cox proportional model in equation (3.5) using $p_{i,k} = 1 - \exp(-\exp(\theta_{i,k}))$. The overall residual deviance can be interpreted as the distance between the model and the perfect fit with the data. The optimal value of D is the total number of observed data points such that each $dev_{i,k}$ should contribute 1 to D at each iteration. The posterior mean of the overall residual deviance, denoted by $\overline{D}$, summarizes the goodness-of-fit of the model.

Model complexity is measured by the effective number of parameters. Let $\overline{dev}_{i,k}$ be the posterior mean of of the residual deviance of each observed data point. The effective number of parameters is given by the summed distances between $\overline{dev}_{i,k}$ and the estimate of $dev_{i,k}$ based on the posterior means of the parameters. Let $\tilde{p}_{i,k}$ be the posterior mean of $p_{i,k}$ and let $\tilde{y}_{i,k} = \tilde{p}_{i,k} n_{i,k}$ be the estimated number of occurrence calculated by the posterior mean. Let $\widetilde{dev}_{i,k}$ be the estimate given by replacing $\widehat{y}_{i,k}$ with $\tilde{y}_{i,k}$ in equation (3.12). The effective

number of parameters is defined to be

$$pD = \sum_{i=1}^{n} \sum_{k=1} (\overline{dev}_{i,k} - \widetilde{dev}_{i,k}). \tag{3.12}$$

Another deviance statistic is the global Deviance Information Criteria (*DIC*), which is used to measure the goodness-of-fit with penalized model complexity. It is thus useful for comparing model of different parametrization with the same likelihood. The *DIC* is defined to be

$$\begin{aligned} DIC &= \overline{D} + pD \\ &= \overline{D} + \sum_{i=1}^{n} \sum_{k=1} (\overline{dev}_{i,k} - \widetilde{dev}_{i,k}). \end{aligned} \tag{3.13}$$

The diagnostic statistics specified by equations (3.11) and (3.13) here give an adequate picture on the overall fit of the model and its complexity. These will be used in the next Section 3.4 in the case study on angina treatments in order to determine which models in Section 3.2 is the best fit for the available data set.

3.4 Case study on angina treatments

There are a total of 5 clinical endpoints of interest in this case study on the comparison of 5 angina treatments. Table 3.1 provides the labels of the clinical endpoints and angina treatments. The primary clinical endpoint is clinical endpoint 1 (MACE). The clinical endpoints labeled by $2, 3, 4$ and 5 in Table 3.1 are the secondary clinical endpoints. The trial data available for MACE is described by Table 3.2 and its network plot is Figure 3.1.

A Bayesian NMA is performed for each clinical endpoint using the methodologies specified in the previous Section 3.3. All three models, fixed effect (FE), homogeneous random effect (RE-Hom) and heterogeneous random effect (RE-Het), that are specified in Section 3.3.3 are implemented in each Bayesian NMA. As explained in Section 3.3.1 and Section 3.3.2, the main parameters of interests are the expected log hazard ratios of the treatment of interest k to the reference treatment r, which are defined to be the basic parameters, $d_{r,k}$ where $r \neq k$. The hyper-parameters of the priors for the model implementation are set to be as uninformative as possible. The mean and standard deviation of the Normal priors for the basic parameters are therefore set as $m_d = 0.000$ and $v_d = 100$, respectively, in all three models. For the RE-Hom and RE-Het model, the hyper-parameters

of the Uniform priors for σ and $\sigma_{r,k}$ are both set as $a_\sigma = 0$ and $b_\sigma = 5$, respectively. The hyper-parameters of the Uniform prior for the correlations $\rho^{(r)}_{k,k'}$ are set as $a_\rho = 0$ and $b_\rho = 1$ for all $r < k < k'$.

3.4.1 Results

In this case study, the non-invasive and drug-based treatment 1 (OMT) is designated as the reference treatment $r = 1$, whilst the invasive and PCI-guided treatments are designated as the treatments of interests $k = 2,3,4,5$. Table 3.3 gives the posterior means and the standard deviations in brackets of the model parameters for the primary clinical endpoint 1 (MACE). Similar tables for the secondary clinical endpoints can be found in Appendix B. The numeric results of the model parameters are then used to compute the $z-$score test statistic in equation (3.7) and their $p-$values. Hypothesis testings at the significance level of 0.05 are then conducted using the numerical values in Table 3.4 and it was concluded that the available data do not suffer from the problem of evidence inconsistency.

Model comparison is then performed using equations (3.11) and (3.12) in order to determine which model is the best for describing the data. Table 3.5 summarizes the diagnostic statistics of all models for each clinical endpoint and the lowest deviance statistics are highlighted in bold. The residual deviance D reveal that the FE model fits very poorly with the data, whilst the RE-Hom and RE-Het models both have very good fit. In comparison to the RE-Hom model, the RE-Het model has a slightly better fit, but it comes at a higher cost of model complexity as well as a much more difficult interpretation. The standard deviations of the RE-Het model parameters in Table 3.3 also indicate that the RE-Het model might be suffering from over-fitting. Since the RE-Hom model is very close to the RE-Het model in terms of model fit, it is thus chosen to be the best model with the optimal trade-off between model fit and model complexity.

Figure 3.2 is a forest plot that shows the results of the expected log hazard ratios of all treatments that are given by the RE-Hom model. The posterior means and 95% credible intervals of the expected log hazard ratios are plotted for all clinical endpoints. It should be noted that the 95% credible intervals of the expected log hazard ratios for clinical endpoints 1 (MACE) and 4 (MI) are much narrower than that of the other clinical endpoints. This makes the use of the expected log hazard ratios for decision-making more difficult. For example, the 95% credible interval of $d_{1,3}$ completely encompasses that of $d_{1,2}$ for clinical endpoint 1 (MACE) whilst their posterior means are very close to each other. Meanwhile, the 95% credible interval of $d_{1,5}$ completely encompasses that of $d_{1,2}$ for clinical endpoint 4 (MI)

Table 3.3 Results of different models for clinical endpoint 1 (MACE).

model parameters	FE	RE-Hom	RE-Het
$d_{1,2}$	−0.160 (0.060)	−0.0737 (0.223)	−0.061 (0.217)
$d_{1,3}$	−0.133 (0.053)	−0.102 (0.268)	−0.077 (0.510)
$d_{1,4}$	−0.464 (0.108)	−0.559 (0.447)	−0.561 (1.250)
$d_{1,5}$	−0.487 (0.170)	−0.562 (0.627)	−0.564 (2.009)
$d_{2,3}$	0.027 (0.080)	−0.028 (0.348)	−0.016 (0.554)
$d_{2,4}$	−0.304 (0.101)	−0.485 (0.445)	−0.500 (1.249)
$d_{2,5}$	−0.327 (0.166)	−0.489 (0.626)	−0.503 (2.009)
$d_{3,4}$	−0.331 (0.120)	−0.457 (0.521)	−0.484 (1.350)
$d_{3,5}$	−0.354 (0.178)	−0.460 (0.681)	−0.487 (2.072)
$d_{4,5}$	−0.023 (0.131)	−0.003 (0.440)	−0.003 (1.568)
σ		0.567 (0.159)	
$\sigma_{1,2}$			0.513 (0.221)
$\sigma_{1,3}$			1.065 (0.560)
$\sigma_{1,4}$			1.392 (1.166)
$\sigma_{1,5}$			1.801 (1.373)
$\sigma_{2,3}$			1.215 (1.732)
$\sigma_{2,4}$			2.883 (4.468)
$\sigma_{2,5}$			4.520 (5.755)
$\sigma_{3,4}$			3.265 (4.278)
$\sigma_{3,5}$			4.661 (5.387)
$\sigma_{4,5}$			4.909 (5.995)

whilst their posterior means are both extremely close to zero. It should be noted that the expected log hazard ratios $d_{1,5}$ are consistently more varied with a much higher range of 95% credible intervals than the other treatments. This might be caused by the fact that there is no direct evidence available for treatment 1 versus treatment 5 as per Figure 3.1. The exponential transformation of the expected log hazard ratios gives the expected hazard ratios and might be easier for numerical comparison.

Table 3.6 summarizes the expected hazard ratios of the treatments of interests k relative to the reference treatment $r = 1$ and the statistical heterogeneity for each clinical endpoint. If the expected hazard ratio is less than 1.0, then it can be said that the treatment of interest is more effective than the reference treatment against the hazardous clinical endpoint and vice versa, if the hazard ratio is larger than 1.0, then the treatment of interest can be said as less effective. The lower the expected hazard ratio, the more effective the treatment of interest. The statistical heterogeneity across the trials is quantified in the last row by the standard deviation of the cumulative log-hazard ratio of the treatment arm to control arm of each trial. The smaller the standard deviation, the less variation there is across the trials. The

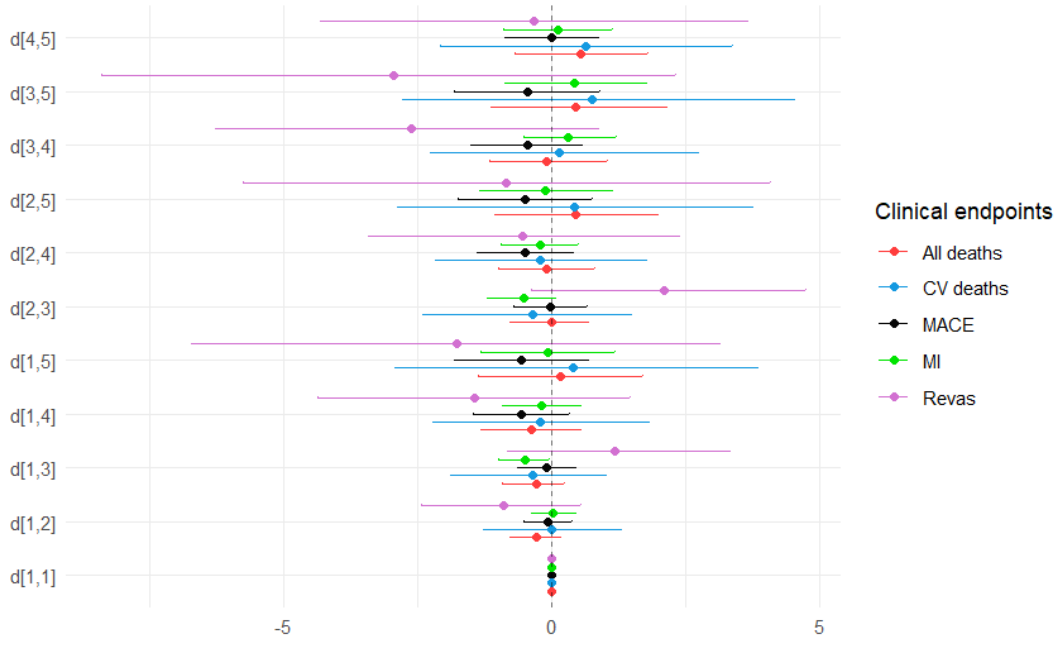
Table 3.4 Bucher's method: $p-$values of the $z-$score in equation (3.7).

Clinical endpoints	$p-$value of				
	$z_{1,2}$	$z_{1,3}$	$z_{1,4}$	$z_{2,4}$	$z_{4,5}$
MACE	0.545	0.567	0.827	0.500	0.687
All deaths	0.731	0.720	0.740	0.500	0.736
CV deaths	0.614	0.501	0.537	0.500	0.582
MI	0.930	0.529	0.755	0.500	0.591
Revas	0.979	0.708	0.966	0.500	0.754

Table 3.5 Diagnostic statistics of the models for all clinical endpoints.

Clinical endpoints	*D*			*DIC*		
	FE	RE-Hom	RE-Het	FE	RE-Hom	RE-Het
MACE	101.91	36.81	**36.01**	124.01	70.11	**70.01**
All deaths	44.73	31.51	**31.44**	63.71	**57.28**	58.46
CV deaths	48.32	19.79	**19.10**	62.31	38.52	**37.30**
MI	45.56	**29.58**	29.74	64.56	**55.23**	56.10
Revas	324.47	**27.29**	27.37	342.58	**52.64**	52.85

Figure 3.2 Log hazard ratios of treatments by RE-Hom for all clinical endpoints.



posterior means and the 50% credible intervals of the expected hazard ratios are reported since they suffice for a practical evaluation on the effects of the treatments. The lowest

Table 3.6 Hazard ratios of treatments for all clinical endpoints.

	1 MACE	2 All deaths	3 CV deaths	4 MI	5 Revas
$\exp(d_{1,2})$	0.929 (0.806, 1.068)	0.761 (0.668, 0.875)	0.996 (0.686, 1.430)	1.028 (0.903, 1.164)	0.404 (0.261, 0.636)
$\exp(d_{1,3})$	0.903 (0.761, 1.068)	0.757 (0.642, 0.914)	**0.706** (0.484, 1.064)	**0.609** (0.529, 0.708)	3.273 (1.713, 6.142)
$\exp(d_{1,4})$	0.572 (0.432, 0.758)	**0.684** (0.523, 0.900)	0.806 (0.447, 1.433)	0.823 (0.660, 1.024)	0.237 (0.099, 0.572)
$\exp(d_{1,5})$	**0.570** (0.385, 0.846)	1.190 (0.750, 1.895)	1.510 (0.557, 4.021)	0.923 (0.632, 1.346)	**0.171** (0.039, 0.763)
σ	0.567 (0.456, 0.652)	0.448 (0.292, 0.571)	1.076 (0.721, 1.290)	0.394 (0.280, 0.483)	1.884 (1.474, 2.168)

posterior means of the expected hazard ratios are highlighted in bold. Using the posterior samples of the expected hazard ratios, the effectiveness of the treatments can be ranked and the probability of a treatment being the best or the worst against each clinical endpoint can be obtained. Table 3.7 summarizes the probability of each treatment being the best against each clinical endpoint. The highest probability of a treatment being the best are highlighted in bold. Clearly, the highest probability of a treatment being the best coincides with the lowest posterior mean of a treatment's expected hazard ratio. Looking at Table 3.7 together Table 3.6 helps to reliably determine which treatment is indeed the most effective against each clinical endpoint.

It can be easily seen that all four invasive PCI-guided treatments of interests are more effective than the non-invasive, drug-based reference treatment 1 against the primary clinical endpoint 1 (MACE), with treatments 4 and 5 being significantly more effective than treatments 2 and 3. At least three invasive PCI-guided treatments are more effective than OMT against each of the secondary clinical endpoints. However, it appears that only treatment 4, PCI FFR, is consistently more effective than OMT against all clinical endpoints although it is not consistently the most effective. Treatment 4 is the most effective treatment that is closely followed by treatment 3 being the second most effective treatment against clinical endpoint 2 (All cause deaths). Treatment 3 is certainly the most effective against both clinical endpoints 3 (CV deaths) and 4 (MI). Treatment 5 is the most effective against clinical endpoint 5 (Revas).

Table 3.7 The probability of treatment k being the best against each clinical endpoint.

probability	treatment k	1 MACE	2 All deaths	3 CV deaths	4 MI	5 Revas
$\mathbb{P}(\text{best treatment} = k)$	1	0.013	0.005	0.062	0.003	0.014
	2	0.044	0.188	0.117	0.011	0.190
	3	0.100	0.269	**0.363**	**0.651**	0.010
	4	0.386	**0.424**	0.265	0.131	0.276
	5	**0.457**	0.114	0.193	0.205	**0.509**

3.5 Discussion

This chapter focused on the Bayesian network meta-analysis of randomized control trials for mixed treatment comparison and addressed the issue of heterogeneity that arises from the wide range of follow-up time of different trials as well as evidence consistency. The methodologies employed in this chapter can be found in the literature by Dias et al. (2011b); Hu et al. (2020); Lu and Ades (2006) and Lu et al. (2007). These methodologies are then applied on the available trial data of the case study on angina treatments for coronary diseases.

The methodologies employed in this case study have several limitations that should be acknowledged. The Bucher's method for detecting evidence inconsistency is essentially a very basic and simplistic hypothesis test on each pairwise treatment contrast, though it might be sufficient for this case study due to the presence of only two treatment arms for each trial in the data set. Many alternative methods for detecting evidence inconsistency have been proposed and should be further studied using a larger data set with multiple arms trials. The random inconsistency factor method by Lu and Ades (2006), node-splitting method proposed by Dias et al. (2010) and the two-stage linear inference approach formulated by Lu et al. (2011), for example, could potentially provide significant improvements on the detection of evidence inconsistency as well as the quantification of the discrepancy between direct and indirect evidence. Furthermore, extending the current Cox proportional hazards model for network meta-regression might be of future research interests since it might help to evaluate the bias of a specific study on the effects as well as to determine the specific study characteristics that cause such bias.

The case study in this work is motivated by the unclear benefits of the invasive surgery known as PCI, which depends on different guidance strategies. The results of the case study demonstrated overwhelming evidence for the effectiveness of the invasive PCI-guided treatments against the conventional drug-based OMT. However, there exists variations between different PCI-guided strategies in regards to their effectiveness against different hazardous

clinical outcomes. In this work, the effectiveness of the treatments is measured against each specific hazardous clinical outcome. Further analysis needs to be performed in order to determine the reasons for such variations across different types of hazardous clinical outcomes and to produce a more accurate picture of PCI-guided strategies. Nevertheless, the statistical findings of this study have offered insights into the effectiveness of the PCI-guided strategies in spite of the limitations of the methodologies.

Chapter 4

Degree evolution in a general growing network

Background

This chapter presents the published paper by De Ambroggio and Yip (2024):

De Ambroggio U. and Yip H.C. Degree evolution in a general growing network. Statistics & Probability Letters, 211:110151, 2024. ISSN 0167-7152. doi:https://doi.org/10.1016/j.spl.2024.110151.

The mathematical proof by Dr. Umberto De Ambroggio and the simulation study contributed by PhD candidate H.C. Yip constitute this work. The PhD candidate is forever indebted to Dr. Umberto De Ambroggio.

4.1 Introduction

The study of complex networks has attracted considerable research interests in recent years due to the emergence of real-world network data and the broad applicability of random graph models. The book by Hofstad (2016) provides a comprehensive introduction to random graph models for complex networks. Many studies on random graph models were also conducted in order to investigate the theoretical and empirical properties of complex networks. It has been observed empirically that many real world networks exhibit power-law degree sequences such that the nodes of degree k decays as $k^{-\tau}$ for some $\tau > 0$. Cooper and Frieze (2003) and

Frieze et al. (2006) studied the degree structure of such networks in a more applied setting of the World Wide Web graph process, for example.

One of the most studied models for networks with power-law degree distribution is the preferential attachment model that was introduced by Barabási and Albert (1999) and rigorously studied by Bollobás et al. (2001). In the basic version of preferential attachment, it is always more likely for the newly added nodes to connect to high-degree nodes instead of low-degree ones. However, this is not always a reasonable assumption from a modeling perspective and many alternative models have been proposed to generalize preferential attachment for the purpose of incorporating a wider range of behaviors that are empirically observed in real-world networks. One of these alternatives is preferential attachment with node or edge deletion, which was studied by Cai et al. (2011) and Cooper et al. (2003). In these models, a random graph is generated by node addition and deletion such that the power-law degree distribution is preserved, even though the overall growth of the degree is slowed down when a high-degree node is selected to be deleted. Another variant of preferential attachment was studied by Rudas et al. (2007) and further investigated by Betken et al. (2019); Dereich and Mörters (2009). In this variant, newly added nodes connect to old nodes with a probability that is proportional to a sub-linear function of their degree. It was established that the limiting degree distribution does not have a power-law decay, but instead exhibits a stretched exponential tail. It was also established that the limiting behavior of the vertex degree is dictated by the asymptotic of the sub-linear function characterizing the connection probabilities. Other versions of preferential attachment were discussed and analyzed in the work by Janssen and Prałat (2010); Oliveira and Spencer (2005); Prałat and Wormald (2007). Readers can refer to the work by Durrett (2006); Hofstad (2016); Mitzenmacher (2003) for a more comprehensive survey on random graphs and complex networks.

This chapter focuses on a particular preferential attachment model that was introduced by Deijfen and Lindholm (2009) and was further investigated by Lindholm and Vallier (2011). At every discrete time step, either a new node is added and connected to another existing node in the random graph, or a new edge is added between two nodes, or an edge is deleted. The work of Deijfen and Lindholm (2009) showed that the degree sequence of the network undergoes a phase transition such that the degree distribution loses the power-law behavior and decays exponentially if the probability of edge deletion is strictly larger than the critical case of $1/3$. In the work of Lindholm and Vallier (2011), it was stated that if the probability of edge deletion equals $1/3$, then the expected degree of any given node is of constant order.

This work, however, shows that the statement by Lindholm and Vallier (2011) is wrong. Instead, it is demonstrated that the expected degree of a node grows logarithmically under the critical case of $1/3$. Furthermore, it is also demonstrated that when the probability of edge deletion is strictly less than $1/3$, the degree process of a given node converges almost surely as the number of nodes goes to infinity.

This chapter presents a correction on the statement made by Lindholm and Vallier (2011) with regard to the expected degree of nodes of a particular preferential attachment model that was introduced by Deijfen and Lindholm (2009). Section 4.2 summarizes the precise formulation of the preferential attachment model of interest and presents the main results. Section 4.3 demonstrates a simulation study that numerically validates the main results. Section 4.4 is the complete proof of the main results. Section 4.5 is a discussion about the future developments and potential applications of preferential attachment.

4.2 Model & main results

The notations for the rest of this chapter are defined here. Given a random variable X and a distribution function F, let $X \sim F$ if X has law F. Let $(x_n)_n, (y_n)_n$ be two sequences of non-negative real numbers. If $x_n/y_n \to 0$ as $n \to \infty$, then $x_n = o(y_n)$ or $x_n \ll y_n$. If there exist constants C_1, C_2 such that $C_1 x_n \leq y_n \leq C_2 x_n$, then $x_n \asymp y_n$. If $x_n/y_n \to 1$ for all large enough n, then $x_n \sim y_n$. Let a, b be real numbers, then $a \vee b := \max\{a,b\}$ and $a \wedge b := \min\{a,b\}$.

Let the graph process produced by the model be denoted by $(G_t)_{t\in\mathbb{N}}$. Let the number of nodes and the number of edges at time $t \in \mathbb{N}$ be denoted by $V_t := |V(G_t)|$ and $E_t := |E(G_t)|$, respectively. Note that both V_t and E_t are random. Given a node $u \in V(G_t)$ that was born at time $s \in \{1,\ldots,t\}$, the degree of the node u at time t is denoted by $d_s(u,t)$. The graph process G_1 is initiated at time $t = 1$ with an isolated node with a self-loop. At $t \in \mathbb{N}$, the graph process G_{t+1} is constructed iteratively from G_t by:

1. A new node u is added with probability $p_1 > 0$ and connected to an existing node $v \in G_t$ selected with probability $d_s(v,t)/2E_t$.

2. An edge is added with probability p_2 between two existing nodes, $u, v \in G_t$ selected with probabilities $d_s(u,t)/2E_t$ and $1/V_t$, respectively.

3. An edge is selected with probability $1/E_t$ and deleted with probability $p_3 := 1 - p_1 - p_2$.

If $E_t = 0$, then either a new node with a self-loop is introduced with probability p_1, or an edge is added between two distinct nodes selected uniformly at random with probability $1 - p_1$. Note that if $E_t = 0$, then an edge is always added at time $t+1$ such that $E_{t+1} = 1$ with probability 1. As shown by Lindholm and Vallier (2011), this value of the edge deletion probability p_3 is critical for the behavior of the limiting expected degree of a given node. Set

$$G_s(t) := (t/s)^{\frac{1-3p_3}{2(1-2p_3)}} \tag{4.1}$$

for $t \geq s \geq 1$. It was shown that if $p_3 < 1/3$, then $\mathbb{E}[d_s(u,t)] \asymp G_s(t)$ independently of p_2 as $t \to \infty$. On the other hand, if $p_3 > 1/3$, then $\lim_{t\to\infty} \mathbb{E}[d_s(u,t)] = \gamma$ where γ is a finite positive constant that depends on p_1, p_2 and p_3 such that $\gamma = 0$ if $p_2 = 0$. The following are the main results of this chapter.

Proposition 4.2.1. Let $p_3 = 1/3$, $p_2 > 0$. Then, as $t \to \infty$,

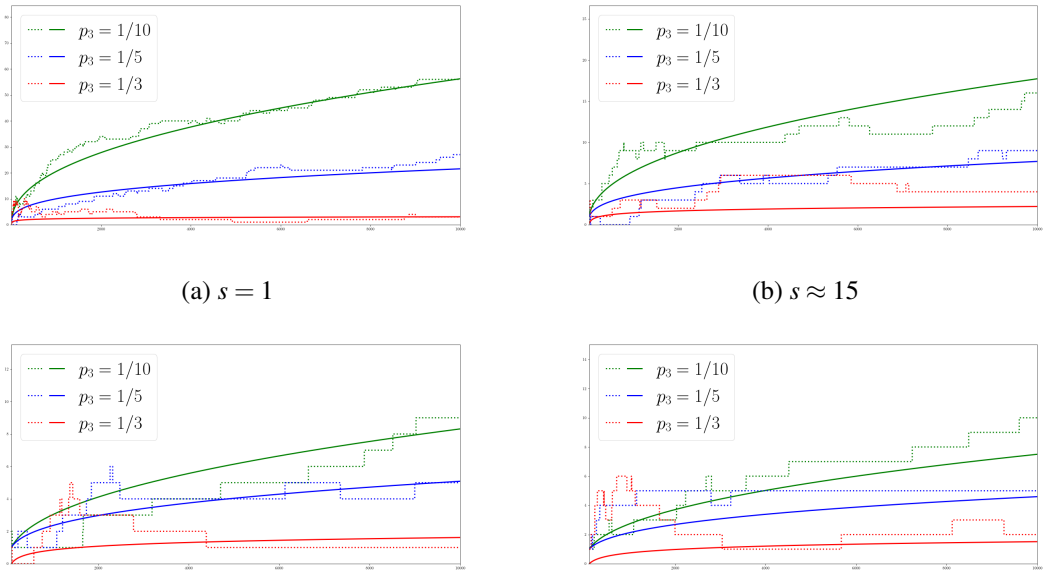
$$\mathbb{E}[d_s(u,t)] \sim \sum_{k=s}^{t-1} \frac{p_2}{kp_1} \prod_{r=k+1}^{t-1} \left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right) \asymp \log(t/s), \tag{4.2}$$

where the constants in the $\asymp$ notation depend on p_1 and p_2.

Proposition 4.2.2. Let $p_3 \in [0, 1/3)$. Then, $d_s(u,t)/G_s(t)$ converges almost surely as $t \to \infty$ to a non-negative random variable with finite mean independently of the value of p_2.

The main result in Proposition 4.2.1 corrects a statement that was made in the work of Lindholm and Vallier (2011), which stated that the expected degree of a given node is of constant order if $p_3 = 1/3$. Note that $\sum_{k=s}^{t-1} 1/k \sim \log(t/s)$ and $\prod_{r=k+1}^{t-1} \left[1 - p_2(\frac{2}{3}r^2 p_1)^{-1}\right] \asymp 1$ with constants depending on p_1, p_2. The asymptotic $\asymp$ then follows. The second main result in Proposition 4.2.2 shows the almost sure convergence of the degree process of a given node if $p_3 < 1/3$.

The asymptotics stated in Propositions 4.2.1 and 4.2.2 are valid whenever $s \ll t$, which is arguably the more relevant range for s. Furthermore, Proposition 4.2.2 does not exclude the case of a degenerate limit. Proving that the limiting random variable is strictly positive with probability one is not a trivial task. However, a simulation study in Section 4.3 below shows that the functions displayed in equations (4.1) and (4.2) indeed describe the asymptotic behavior of a node degree in the corresponding regimes, $p_3 = 1/3$ and $p_3 < 1/3$.



(a) $s = 1$ (b) $s \approx 15$

(c) $s \approx 70$ (d) $s \approx 105$

Figure 4.1 Node degree against time $t = 10^4$.

4.3 Simulation study

Here the main results in Propositions 4.2.1 and 4.2.2 are illustrated by simulation. Three independent graph processes $(G_t)_{t\in\mathbb{N}}$ were simulated under three different edge deletion probabilities p_3. More precisely, p_3 is the probability of an edge being deleted and each graph process was generated under a different p_3. One of the three graph processes was generated using the critical value of $p_3 = 1/3$. The other two graph processes were generated under $p_3 = 1/5$ and $p_3 = 1/10$, respectively. The probabilities of node addition and edge addition at each time step were set as $p_1 = 1/2$ and $p_2 = 1 - p_1 - p_3$, respectively, for all three graph processes. The number of time steps for each graph process is set as $t = 10^4$.

Figure 4.1 illustrates the results of the simulation. Nodes that were approximately born at the same time s are plotted in the same figure. The solid lines represent the asymptotic node degree given by Propositions 4.2.1 and 4.2.2, whilst the dotted lines represent the simulated degree of the nodes. The nodes of each graph process are represented by a color.

Due to the relatively small range of the asymptotic degree in the critical case of $p_3 = 1/3$, the convergence for the critical case $p_3 = 1/3$ is not as apparent in comparison to the cases $p_3 < 1/3$. For example, the asymptotic degrees at time $t = 10^4$ in Figure 4.1 (a) with $s = 1$ under the cases of $p_3 = 1/3, p_3 = 1/5$ and $p_3 = 1/10$ are $\approx 3, \approx 22$ and ≈ 56, respectively.

In contrast, the asymptotic degrees at time $t = 10^4$ in Figure 4.1 (c) with $s \approx 70$ under the cases of $p_3 = 1/3, p_3 = 1/5$ and $p_3 = 1/10$ are $\approx 2, \approx 5$ and ≈ 8, respectively. This uneven scale of the $y-$axis results in the lack of visible growth of the red curves.

4.4 Proofs

A few preliminary results are used to prove the main results, Propositions 4.2.1 and 4.2.2, in the subsections below. For the convenience of readers, the preliminary results are stated here as Lemma 4.4.1, Theorem 4.4.1 and Lemma 4.4.2. Lemma 4.4.1 provides some classical concentration inequalities for the binomial distribution. Theorem 4.4.1 is the Robbins-Siegmund Theorem, which guarantees the almost sure convergence for sequences of non-negative random variables that behave almost like supermartingales. Lemma 4.4.2 provides a concentration inequality for E_t, the number of edges in the graph process G_t.

Lemma 4.4.1. Let X be distributed by $Bin(n, p)$. Then,

$$\begin{cases} \mathbb{P}(X \geq \mathbb{E}(X) + t) & \leq \exp\left(-\frac{t^2}{2(\mathbb{E}(X)+t/3)}\right), \\ \mathbb{P}(X \leq \mathbb{E}(X) - t) & \leq \exp\left(-\frac{t^2}{2\mathbb{E}(X)}\right), \end{cases}$$

for all $t \geq 0$. It follows that

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 2\exp\left(-\frac{t^2}{2(\mathbb{E}(X)+t/3)}\right).$$

Theorem 4.4.1 (Robbins-Siegmund Theorem)**.** *Let $(V_n)_{n\geq 1}, (\beta_n)_{n\geq 1}, (\gamma_n)_{n\geq 1}$ and $(\delta_n)_{n\geq 1}$ be sequences of random variables that take values in $[0, \infty)$ and are adapted to some filtration $(\mathcal{F}_n)_{n\geq 1}$. If*

$$\mathbb{E}(V_{n+1}|\mathcal{F}_n) \leq V_n(1 + \beta_n) + \gamma_n - \delta_n$$

for all $n \geq 1$, then $(V_n)_{n\geq 1}$ converges almost surely to a finite random variable and $\sum_{n\geq 1} \delta_n < \infty$ almost surely on the event that both $\sum_{n\geq 1} \beta_n$ and $\sum_{n\geq 1} \gamma_n$ are finite.

Lemma 4.4.2. Let $p_3 \in (0, 1/2)$. Let $\lambda = \lambda_t$ such that $C \leq \lambda = o(t)$ as $t \to \infty$ for some finite constant $C > 1$. Then, there exists $t_0 \in \mathbb{N}$ such that for all $t \geq t_0$,

$$\mathbb{P}(|E_t - t(1 - 2p_3)| > \lambda) \leq 3\exp\left(-\frac{\lambda^2(1 - 1/C)^2}{8t}\right). \tag{4.3}$$

Note that the edge deletion probability p_3 in Lemma 4.4.2 is at most $1/2$. This is sufficient for the proofs of the main results in Section 4.4.2 and Section 4.4.3 because only the regimes, $p_3 = 1/3$ and $p_3 < 1/3$, are considered.

The proof of Lemma 4.4.2 is provided in Section 4.4.1 below. For the proof of Lemma 4.4.1 and Theorem 4.4.1, readers are referred to the works of Hofstad (2016) and Brémaud (2020), respectively.

4.4.1 Proof of Lemma 4.4.2

A result of this type was in fact derived in Section 2.2 of (Lindholm and Vallier, 2011). However, the argument provided here is self-contained as no classical results about one-dimensional random walks was used. The proof makes use of Lemma 4.4.3, the Hoeffding's inequality, which is provided here for the convenience of the readers.

Lemma 4.4.3 (Hoeffding's inequality)**.** Let $X_1, \ldots, X_n$ be independent random variables. Let there exists constants a_i, b_i with $a_i < b_i$ such that $a_i \leq X_i \leq b_i$ almost surely for all $i \in \{1, \ldots, n\}$. Define $S_n \coloneqq \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))$ and $D_n \coloneqq \sum_{i=1}^{n} (b_i - a_i)^2$. Then,

$$\begin{cases} \mathbb{P}(S_n \geq x) & \leq \exp\left(-2x^2/D_n\right) \\ \mathbb{P}(S_n \leq -x) & \leq \exp\left(-2x^2/D_n\right) \end{cases}$$

for all $x > 0$.

Let $(U_t)_{t \geq 1}$ be a sequence of independent and identically distributed random variables taking values in $\{-1, 1\}$ with $p_3 = \mathbb{P}(U_1 = 1) = 1 - \mathbb{P}(U_1 = -1)$. Define $\widehat{E}_0 \coloneqq 1$ and let

$$\widehat{E}_t = \widehat{E}_{t-1} + 1 - 2\mathbb{1}_{\{U_t = 1\}} \mathbb{1}_{\{\widehat{E}_{t-1} \geq 1\}} \tag{4.4}$$

for all $t \geq 1$. Then, $\widehat{E}_t \overset{d}{=} E_t$ for all $t \geq 0$ by induction. Define $X_i \coloneqq \mathbb{1}_{\{U_i = 1\}} \mathbb{1}_{\{\widehat{E}_{i-1} \geq 1\}}$ for all $i \geq 1$, then $\widehat{E}_t = 1 + \sum_{i=1}^{t} (1 - 2X_i)$ by equation (4.4). Since $X_i \leq \mathbb{1}_{\{U_i = 1\}}$, therefore $\widehat{E}_t \geq \sum_{i=1}^{t} (1 - 2\mathbb{1}_{\{U_i = 1\}})$ for all t. Then, it follows that

$$\begin{aligned} \mathbb{P}(E_t \leq t(1 - 2p_3) - \lambda) &\leq \mathbb{P}\left(\sum_{i=1}^{t} (1 - 2\mathbb{1}_{\{U_i = 1\}}) \leq t(1 - 2p_3) - \lambda \right) \\ &= \mathbb{P}\left(\sum_{i=1}^{t} (\mathbb{1}_{\{U_i = 1\}} - p_3) \geq \lambda/2 \right) \leq \exp\left(-\frac{\lambda^2}{2t} \right). \end{aligned} \tag{4.5}$$

by Lemma 4.4.3. Let $H = H_t \in \mathbb{N}$ with $H \ll t$ to be specified later. Define the event $\mathcal{E}_H = \{\widehat{E}_k \geq 1 \; \forall \, H \leq k \leq t-1\}$. Since $E_t \overset{d}{=} \widehat{E}_t$,

$$\mathbb{P}(E_t \geq t(1-2p_3)+\lambda) \leq \mathbb{P}\left(\{\widehat{E}_t \geq t(1-2p_3)+\lambda\} \cap \mathcal{E}_H\right) + \mathbb{P}(\mathcal{E}_H^c).$$

Note that $\widehat{E}_H = 1 + \sum_{i=1}^{H}(1-2X_i) \leq H+1$ since $X_i \geq 0$. Since $1-2X_i = 1-2\mathbb{1}_{\{U_i=1\}}$ for $H+1 \leq i \leq t$ on the event $\mathcal{E}_H$, therefore

$$\begin{aligned}
&\mathbb{P}\left(\{\widehat{E}_t \geq t(1-2p_3)+\lambda\} \cap \mathcal{E}_H\right) \\
&\qquad = \mathbb{P}\left(\left\{\widehat{E}_H + \sum_{i=H+1}^{t}(1-2X_i) \geq t(1-2p_3)+\lambda\right\} \cap \mathcal{E}_H\right) \\
&\qquad \leq \mathbb{P}\left(\sum_{i=H+1}^{t}(1-2\mathbb{1}_{\{U_i=1\}}) \geq t(1-2p_3)+\lambda-H-1\right) \\
&\qquad = \mathbb{P}\left(\sum_{i=H+1}^{t}(\mathbb{1}_{\{U_i=1\}}-p_3) \leq Hp_3 - \frac{\lambda}{2}\left(1-\frac{1}{\lambda}\right)\right).
\end{aligned} \tag{4.6}$$

Taking $H \coloneqq \lfloor \frac{\lambda}{4p_3}(1-1/\lambda) \rfloor$, the probability in equation (4.6) is at most

$$\mathbb{P}\left(\sum_{i=H+1}^{t}(\mathbb{1}_{\{U_i=1\}}-p_3) \leq -\frac{\lambda}{4}(1-1/\lambda)\right) \leq \exp\left(-2\frac{\lambda^2(1-1/\lambda)^2}{16(t-H)}\right) \leq \exp\left(-\frac{\lambda^2(1-1/C)^2}{8t}\right),$$

where the first inequality follows again by Lemma 4.4.3. Now, to bound $\mathbb{P}(\mathcal{E}_H^c)$, note that

$$\mathbb{P}(\mathcal{E}_H^c) = \mathbb{P}\left(\exists\, k \in [H, t-1] : \widehat{E}_k = 0\right) \leq \sum_{k=H}^{t-1}\mathbb{P}\left(1+\sum_{i=1}^{k}(1-2X_i) = 0\right)$$

by a union bound. Since $1-2X_i \geq 1-2\mathbb{1}_{\{U_i=1\}}$ such that $\sum_{i=1}^{k}(1-2X_i) \geq \sum_{i=1}^{k}(1-2\mathbb{1}_{\{U_i=1\}})$, thus

$$\begin{aligned}
\sum_{k=H}^{t-1}\mathbb{P}\left(\sum_{i=1}^{k}\left(1-2\mathbb{1}_{\{U_i=1\}}\right) \leq 0\right) &\leq \sum_{k=H}^{t-1}\mathbb{P}\left(\sum_{i=1}^{k}(\mathbb{1}_{\{U_i=1\}}-p_3) \geq \frac{k}{2}(1-2p_3)\right) \\
&\leq \sum_{k=H}^{t-1}\exp\left(-\frac{k(1-2p_3)^2}{2}\right),
\end{aligned} \tag{4.7}$$

where the last inequality follows once more by Lemma 4.4.3. Using the formula for the geometric sum,

$$\sum_{k=H}^{t-1} \exp\left(-\frac{k(1-2p_3)^2}{2}\right) = \frac{1-\exp\left(-t(1-2p_3)^2/2\right)}{1-\exp\left(-(1-2p_3)^2/2\right)} - \frac{1-\exp\left(-H(1-2p_3)^2/2\right)}{1-\exp\left(-(1-2p_3)^2/2\right)} \leq \frac{\exp\left(-H(1-2p_3)^2/2\right)}{1-\exp\left(-(1-2p_3)^2/2\right)}.$$

Now, recall that $H = \lfloor \frac{\lambda}{4p_3}(1-1/\lambda) \rfloor$ such that

$$\frac{H(1-2p_3)^2}{2} \geq \frac{\lambda}{8}\left(1-\frac{1}{C}\right)^2 \frac{(1-2p_3)^2}{p_3}.$$

Thus, setting $c_1 \coloneqq \left(1-e^{-\frac{(1-2p_3)^2}{2}}\right)^{-1}$ and $c_2 \coloneqq \left(1-\frac{1}{C}\right)^2 \frac{(1-2p_3)^2}{8p_3}$, the bound $\mathbb{P}(\mathcal{E}_H^c) \leq c_1 \exp(-c_2\lambda)$ is obtained. Summarizing,

$$\mathbb{P}(E_t \geq t(1-2p_3)+\lambda) \leq \exp\left(-\frac{\lambda^2(1-1/C)^2}{8t}\right) + c_1 \exp\left(-c_2\lambda\right), \tag{4.8}$$

and since $\lambda \ll t$, the desired result follows.

4.4.2 Proof of Proposition 4.2.1

Let $p_3 = 1/3$ and $p_2 > 0$. Define $\mathcal{F}_t \coloneqq \sigma(\{G_s : 1 \leq s \leq t\})$ where $t \in \mathbb{N}$. It is not difficult to see that

$$\begin{aligned}\mathbb{E}\left[d_s(u,t+1)|\mathcal{F}_t\right] &= d_s(u,t)\left[1+\frac{1-3p_3-p_2/(V_t-1)}{2E_t}\right] + \frac{p_2}{V_t-1} \\ &= d_s(u,t)\left[1-\frac{p_2}{2E_t(V_t-1)}\right] + \frac{p_2}{V_t-1}.\end{aligned} \tag{4.9}$$

Let $a_t > 0$ be such that $a_t = o(t)$. Let $C > 1$ and let $\lambda_t \geq C$ such that $\lambda_t = o(t)$. Define the events $D_t \coloneqq \{|V_t - tp_1| \leq a_t\}$ and $H_t \coloneqq \{E_t \leq t/3 + \lambda_t\}$ where $t/3 = t(1-2p_3)$ since $p_3 = 1/3$. Then, using the fact that $d_s(u,t+1) \leq t$,

$$\mathbb{E}\left[d_s(u,t+1)|\mathcal{F}_t\right] \leq \mathbb{E}\left[d_s(u,t+1)|\mathcal{F}_t\right]\mathbb{1}_{D_t\cap H_t} + t\,\mathbb{1}_{(D_t\cap H_t)^c}. \tag{4.10}$$

Observe that $\frac{p_2}{V_t-1} \leq \frac{p_2}{tp_1-a_t-1}$ on the event $D_t \cap H_t$ and for sufficiently large t,

$$1-\frac{p_2}{2E_t(V_t-1)} \leq 1-\frac{p_2}{2[(t/3+\lambda_t)(tp_1+a_t)]} \leq 1-\frac{p_2}{\frac{2}{3}t^2p_1+h_t},$$

where $h_t := 4t(a_t/3 \vee 2\lambda_t p_1)$. Therefore,

$$\mathbb{E}\left[d_s(u,t+1)|\mathcal{F}_t\right]\mathbb{1}_{D_t\cap H_t} \leq d_s(u,t)\left(1-\frac{p_2}{\frac{2}{3}t^2p_1+h_t}\right)+\frac{p_2}{tp_1-a_t-1}. \tag{4.11}$$

Using the inequality, $1/(1+y) \geq 1-y$ which is valid for all $y>-1$, it is clear that

$$\frac{p_2}{\frac{2}{3}t^2p_1+h_t} \geq \frac{p_2}{\frac{2}{3}t^2p_1}\left(1-\frac{h_t}{\frac{2}{3}t^2p_1}\right) = \frac{p_2}{\frac{2}{3}t^2p_1}-\frac{p_2h_t}{\frac{4}{9}t^4p_1^2}.$$

Thus, using again the inequality $d_s(u,t) \leq t$, equation (4.11) is at most

$$d_s(u,t)\left(1-\frac{p_2}{\frac{2}{3}t^2p_1}\right)+\frac{p_2}{tp_1-a_t-1}+\frac{p_2h_t}{\frac{4}{9}t^3p_1^2}.$$

Summarizing,

$$\mathbb{E}\left[d_s(u,t+1)|\mathcal{F}_t\right] \leq d_s(u,t)\left(1-\frac{p_2}{\frac{2}{3}t^2p_1}\right)+\frac{p_2}{tp_1-a_t-1}+\frac{p_2h_t}{\frac{4}{9}t^3p_1^2}+t\mathbb{1}_{(D_t\cap H_t)^c}.$$

Taking expectations on both sides of the above inequality,

$$\mathbb{E}\left[d_s(u,t+1)\right] \leq \mathbb{E}\left[d_s(u,t)\right]\left(1-\frac{p_2}{\frac{2}{3}t^2p_1}\right)+\frac{p_2}{tp_1-1-a_t}+\frac{p_2h_t}{\frac{4}{9}t^3p_1^2}+t\mathbb{P}\left((D_t\cap H_t)^c\right).$$

Thus, the expected value of $d_s(u,t)$ is at most

$$\prod_{k=s}^{t-1}\left(1-\frac{p_2}{\frac{2}{3}k^2p_1}\right)+\sum_{k=s}^{t-1}\frac{p_2}{kp_1-1-a_k}\prod_{r=k+1}^{t-1}\left(1-\frac{p_2}{\frac{2}{3}r^2p_1}\right)+\sum_{k=s}^{t-1}\frac{p_2h_k}{\frac{4}{9}k^3p_1^2}+\sum_{k=s}^{t-1}k\mathbb{P}\left((D_k\cap H_k)^c\right).$$

Using Lemma 4.4.1, it can be shown that $\mathbb{P}(D_k^c) \leq 2\exp(-c_1a_k^2/k)$ for some finite constant $c_1>0$. Using Lemma 4.4.2, it can be shown that $\mathbb{P}(H_k^c) \leq 3\exp(-c_2\lambda_k^2/k)$ for some finite

constant $c_2 > 0$ if $k \geq t_0$. Therefore,

$$\sum_{k=s}^{t-1} k\mathbb{P}\left((D_k \cap H_k)^c\right) \leq C \sum_{k=t_0}^{t-1} k e^{-(c_1 \frac{a_k^2}{k} \wedge c_2 \frac{\lambda_k^2}{k})}$$

for all $t > t_0 \vee s$ and some finite positive constant C. Take $a_k := \sqrt{k\log(k^\alpha)}$ and $\lambda_k := \sqrt{k\log(k^\beta)}$, where $\alpha, \beta > 0$ are constants such that $c_1\alpha \wedge c_2\beta > 2$. Note that both $a_k, \lambda_k \ll k$. Recall that $h_k := 4k(a_k/3 \vee 2\lambda_k p_1)$ as defined prior to equation (4.11). Then,

$$\sum_{k=s}^{t-1} \frac{p_2 h_k}{\frac{4}{9}k^3 p_1^2} = \sum_{k=s}^{t-1} O\left(\frac{a_k \vee \lambda_k}{k^2}\right) = \sum_{k=s}^{t-1} O\left(\frac{\log^{1/2}(k)}{k^{3/2}}\right),$$
$$\sum_{k=t_0}^{t-1} k e^{-(c_1 \frac{a_k^2}{k} \wedge c_2 \frac{\lambda_k^2}{k})} = \sum_{k=t_0}^{t-1} \frac{1}{k^C},$$

where $C := (c_1\alpha \wedge c_2\beta) - 1 > 1$. Note that since

$$\frac{p_2}{kp_1 - 1 - a_k} = \frac{p_2}{kp_1}\left(1 - \frac{a_k + 1}{kp_1}\right)^{-1} = \frac{p_2}{kp_1}\left(1 + O(a_k/k)\right),$$

thus leading to

$$\sum_{k=s}^{t-1} \frac{p_2}{kp_1 - 1 - a_k} \prod_{r=k+1}^{t-1}\left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right) = O(1) + \sum_{k=s}^{t-1} \frac{p_2}{kp_1} \prod_{r=k+1}^{t-1}\left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right) + \sum_{k=s}^{t-1} O\left(\frac{\log(k)}{k^{3/2}}\right).$$

Since $\lim_{t\to\infty} \prod_{r=s}^{t-1}\left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right) \in (0, \infty)$, it can be concluded that

$$\limsup_{t\to\infty} \frac{\mathbb{E}[d_s(u,t)]}{\sum_{k=s}^{t-1} \frac{p_2}{kp_1} \prod_{r=k+1}^{t-1}\left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right)} \leq 1.$$

Similar calculations yield

$$\liminf_{t\to\infty} \frac{\mathbb{E}[d_s(u,t)]}{\sum_{k=s}^{t-1} \frac{p_2}{kp_1} \prod_{r=k+1}^{t-1}\left(1 - \frac{p_2}{\frac{2}{3}r^2 p_1}\right)} \geq 1$$

and the proof is concluded.

4.4.3 Proof of Proposition 4.2.2

Let $p_3 < 1/3$. Let $a_t > 0$ and let λ_t be defined as the same as the previous section. Recall that $D_t := \{|V_t - tp_1| \leq a_t\}$. Define $H_t := \{E_t \geq t(1-2p_3) - \lambda_t\}$. Proceeding in the exact same way as the previous section, it can be shown that $\mathbb{E}[d_s(u,t+1)|\mathcal{F}_t]$ is bounded from above by

$$\mathbb{1}_{D_t \cap H_t} \mathbb{E}[d_s(u,t+1)|\mathcal{F}_t] + t\mathbb{1}_{D_t^c} + t\mathbb{1}_{H_t^c}. \tag{4.12}$$

The first term from the left in equation (4.12) is at most

$$\mathbb{1}_{H_t} \left\{ d_s(u,t) \left[1 + \frac{1 - 3p_3 - p_2/(tp_1 - 1 + a)}{2E_t} \right] + \frac{p_2}{tp_1 - 1 - a} \right\}. \tag{4.13}$$

Let $t_0 = \min\{t \geq 1 | p_3 + p_2/3(tp_1 - 1 + a) \leq 1/3\}$ where $p_3 < 1/3$. Then, $1 - 3p_3 - p_2/(tp_1 - 1 + a) \geq 0$ for $t \geq t_0$ and hence, by the definition of H_t, the expression in equation (4.13) here is bounded from above by

$$d_s(u,t) \left[1 + \frac{1 - 3p_3 - p_2/(tp_1 - 1 + a)}{2t(1-2p_3) - 2\lambda} \right] + \frac{p_2}{tp_1 - 1 - a}.$$

Thus, setting $L_t := \prod_{h=s}^{t} \left(1 + \frac{1-3p_3 - p_2/(hp_1 - 1 + a)}{2h(1-2p_3) - 2\lambda} \right)$ leads to

$$\mathbb{E}\left[\frac{d_s(u,t+1)}{L_t} \middle| \mathcal{F}_t \right] \leq \frac{d_s(u,t)}{L_{t-1}} + \gamma_t \text{ where } \gamma_t := \frac{2p_2}{tp_1 L_t} + \frac{t}{L_t}\mathbb{1}_{D_t^c} + \frac{t}{L_t}\mathbb{1}_{H_t^c}$$

for $t \geq t_0$ where t_0 is so large such that the condition $1 - (a+1)/tp_1 \geq 1/2$ is satisfied. The upper and lower bounds on L_t can be given using the following:

Lemma 4.4.4. There exists $f : \mathbb{N} \times \mathbb{N} \mapsto \mathbb{R}_+$ with $f(s,t) \to 0$ as $s,t \to \infty$ such that $(1 - f(s,t))G_s(t) \leq L_t \leq (1 + f(s,t))G_s(t)$.

Lemma 4.4.4 can be easily proved using the Taylor series expansion. Now, in order to apply Theorem 4.4.1 to the sequence $(d_s(u,t)/L_{t-1})_{t \geq t_0}$, it must first be shown that

$$\sum_{t \geq t_0} \frac{2p_2}{tp_1 L_t} + \sum_{t \geq t_0} \frac{t}{L_t}\mathbb{1}_{D_t^c} + \sum_{t \geq t_0} \frac{t}{L_t}\mathbb{1}_{H_t^c} < \infty \tag{4.14}$$

almost surely, which can be easily done. Indeed, it follows from Lemmas 4.4.1 and 4.4.2 that

$$\begin{aligned}\mathbb{E}\left(\sum_{t\geq t_0}\frac{t}{L_t}\mathbb{1}_{D_t^c}+\sum_{t\geq t_0}\frac{t}{L_t}\mathbb{1}_{H_t^c}\right)\\ \leq\sum_{t\geq t_0}\frac{t}{L_t}2\exp\left(-\frac{a^2}{2tp_1+\frac{2}{3}a}\right)+3\sum_{t\geq t_0}\frac{t}{L_t}\exp\left(-\frac{\lambda_t^2(1-1/C)^2}{8t}\right).\end{aligned} \tag{4.15}$$

Finally, since $a_t = o(t)$, the right hand side of equation (4.15) for all large enough t is at most

$$C^* s^{\frac{1-p_3}{2(1-2p_3)}}\left(\sum_{t\geq t_0} t^{\frac{1-p_3}{2(1-2p_3)}}\exp\left(-\frac{a_t^2}{3tp_1}\right)+\sum_{t\geq t_0} t^{\frac{1-p_3}{2(1-2p_3)}}\exp\left(-\frac{\lambda_t^2(1-1/C)^2}{8t}\right)\right) \tag{4.16}$$

where $C^* = C_s^*$ is a constant which depends on s. Let $a = \sqrt{t\log(t^\alpha)}$ and $\lambda = \sqrt{t\log(t^\beta)}$ where $\alpha,\beta > 0$ are appropriately chosen, then the two series in equation (4.16) converge. Therefore,

$$\sum_{t\geq t_0}\frac{t}{L_t}\mathbb{1}_{D_t^c}+\sum_{t\geq t_0}\frac{t}{L_t}\mathbb{1}_{H_t^c}<\infty$$

almost surely.

By Lemma 4.4.4, the series $\sum_{t\geq t_0}\frac{2p_2}{tp_1L_t}$ converges since $L_t \asymp (t/s)^\eta$ for some $\eta > 0$ and so $tL_t \asymp t^{1+\eta}s^{-\eta}$. It can be concluded that equation (4.14) holds true. By Theorem 4.4.1, $d_s(u,t)/L_{t-1}$ converges almost surely to a non-negative random variable with finite mean as $t\to\infty$. Using Lemma 4.4.4 once again, the proof is concluded.

4.5 Discussion

The asymptotic behavior of the expected vertex degree of the preferential attachment model proposed by Deijfen and Lindholm (2009) is rigorously studied in this work, thereby providing a correction on a statement made by Lindholm and Vallier (2011). The focus of this work is put on the regime where the edge deletion probability is $p_3 \leq 1/3$.

The type of computation that was used in this work might be followed by other studies on preferential attachment models with similar attachment rules. In particular, the result of this work might motivate further research on the edge deletion properties of other variants of preferential attachment. For example, it might be of interest to make studies on the relation between edge deletion probability and degree dynamics, such as the work of Brot et al. (2013), more robust.

In addition, the results of this work may be of interest to future studies on real complex networks in various application settings. Indeed, there are many real-life scenarios where preferential attachment models with edge deletion is applicable. The work by Shaw et al. (2011), for example, studies the effects of gossips on social network structure by weakening the weight of an edge through a quadratic rule, which clearly could have been studied by preferential attachment with edge deletion. In particular, preferential attachment with edge deletion would be extremely applicable in mathematical epidemiology due to its suitability in network-based epidemic models. (Durrett, 2006). The study on Covid-19 by Maheshwari and Albert (2020), for example, made predictions on the second wave under social distancing using social networks with edge deletion. This is remarkably useful for biomedical researchers that aim to understand the propagation and patterns of infectious diseases in real-life. An alternative topic of interest is to measure the relation between edge deletion probability and topological properties of real complex networks, which is similar to the work by Vázquez (2003) in which the relations between the connection probabilities of vertices, the vertex degree and the topological properties of several scale-free networks are studied. A discussion on different applications of preferential attachment models can be found in Chapter 1 of the textbook by Hofstad (2016).

Chapter 5

Conclusion & discussion

Several statistical learning methods for understanding complex biomedical data sets have been proposed, thoroughly examined and discussed in this thesis. Each chapter presents a detailed description and analysis of the involved statistical methodologies. The ultimate goal is to provide useful statistical tools for biomedical researchers to conduct statistical analysis by learning from biomedical data sets that possess specific characteristics. This thesis focuses on the statistical aspect of three research topics: a spatio-temporal model to learn the latent spectral shape from bio-acoustic data for the purpose of cross-species comparison, a thorough Bayesian network meta-analysis on a set of randomized control trials data and the behavior of a particular preferential attachment model with edge deletion that is suitable for epidemic modeling.

It has been shown that the proposed spatio-temporal model for bio-acoustic data is able to obtain the latent spectral shape of the acoustic structure of a species. The proposed model is able to perform time synchronization and thereby quantifying temporal non-stationarity in the bio-acoustic signals. This is achieved by the construction of a synchronization function. More importantly, it has also been shown that the results obtained by the proposed model are useful for cross-species bio-acoustic analysis, which has not been possible using conventional statistical tools. The proposed model has the potential to be widely applied in the bio-acoustic analysis of other animals and facilitate the understanding on the relation between communication systems and evolutionary paths of different species. Statistically, the construction of the synchronization function also provides a novel methodology for modeling data that is non-stationary in time.

For the Bayesian network meta-analysis of randomized clinical trials, it is clear that the application of the methodologies has obtained valuable insights in the case study of

angina treatments comparison for chronic coronary syndromes. It is shown that the specific proportional hazard model used in the case study is able to account for the effects of varying follow up times. The comparison of the model under different assumptions has revealed future research directions for further development on the matter of evidence consistency in the context of network meta-analysis.

The study on the asymptotic behavior of the expected degree of a particular preferential attachment model has encouraged the further development of preferential attachment with edge deletion. The proof of the established results does not make use of classical results about one-dimensional random walks and similar computations of this type can be applied on other preferential attachment models with similar attachment rules. Furthermore, this study also motivated the wider application of preferential attachment with edge deletion in more real-life settings, especially for understanding the spread of disease under social distancing.

In conclusion, the statistical methods presented and studied in this thesis are shown to possess great potentials for the application in biomedical sciences for learning from complex and large data sets. The performances of these methods were demonstrated by simulation studies or application on real data sets. Finally, the results obtained by the application of these methods and the in-depth examination of these methods have shown many directions for future research.

References

EunJin Ahn and Hyun Kang. Concepts and emerging issues of network meta-analysis. *Korean Journal of Anesthesiology*, 74:371–382, 2021. doi: https://doi.org/10.4097/kja.21358.

Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002. doi: 10.1103/RevModPhys.74.47. URL https://link.aps.org/doi/10.1103/RevModPhys.74.47.

Roy M. Anderson. Discussion: The kermack-mckendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53(1):3–32, 1991. ISSN 0092-8240. doi: https://doi.org/10.1016/S0092-8240(05)80039-4. URL https://www.sciencedirect.com/science/article/pii/S0092824005800394.

Sudipto Banerjee and Montserrat Fuentes. Bayesian modeling for large spatial datasets. *WIREs Computational Statistics*, 4(1):59–66, 2012. URL https://doi.org/10.1002/wics.187.

Sudipto Banerjee, Alan E. Gelfand, Andrew O. Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society*, 70(4):825–848, 2008. doi: 10.1111/j.1467-9868.2008.00663.x. URL https://doi.org/10.1111/j.1467-9868.2008.00663.x. PMID: 19750209.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. doi: 10.1126/science.286.5439.509. URL https://www.science.org/doi/abs/10.1126/science.286.5439.509.

G. Berry. The analysis of mortality by the subject-years method. *Biometrics*, 39(1):173–184, 1983.

Carina Betken, Hanna Döring, and Marcel Ortgiese. Fluctuations in a general preferential attachment model via Stein's method. *Random Structures & Algorithms*, 55(4):808–830, 2019. doi: https://doi.org/10.1002/rsa.20852. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20852.

William E. Boden, Robert A. O'Rourke, Koon K. Teo, et al. Optimal medical therapy with or without pci for stable coronary disease. *The New England journal of medicine*, 356(15):1503–1516, 2007. doi: https://doi.org/10.1056/NEJMoa070829.

Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001. doi: https://doi.org/10.1002/rsa.1009. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.1009.

Kristin J. Bondo, Diego Montecino-Latorre, Lisa Williams, Matt Helwig, Kenneth Duren, Michael L. Hutchinson, and W. David Walter. Spatial modeling of two mosquito vectors of west nile virus using integrated nested laplace approximations. *Ecosphere*, 14(1):e4346, 2023. doi: https://doi.org/10.1002/ecs2.4346. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecs2.4346.

Kristin J. Bondo, Christopher S. Rosenberry, David Stainbrook, and W. David Walter. Comparing risk of chronic wasting disease occurrence using bayesian hierarchical spatial models and different surveillance types. *Ecological Modelling*, 493:110756, 2024. ISSN 0304-3800. doi: https://doi.org/10.1016/j.ecolmodel.2024.110756. URL https://www.sciencedirect.com/science/article/pii/S0304380024001443.

Fred Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, 2(2):113–127, 2017. ISSN 2468-0427. doi: https://doi.org/10.1016/j.idm.2017.02.001.

Micah R. Bregman, Aniruddh D. Patel, and Timothy Q. Gentner. Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences*, 113(6):1666–1671, 2016. doi: 10.1073/pnas.1515380113. URL https://www.pnas.org/doi/abs/10.1073/pnas.1515380113.

Pierre Brémaud. *Probability Theory and Stochastic Processes*. Universitext. Springer Cham, 2020. doi: https://doi.org/10.1007/978-3-030-40183-2.

Hilla Brot, Michal Honig, Lev Muchnik, Jacob Goldenberg, and Yoram Louzoun. Edge removal balances preferential attachment and triad closing. *Physical Review E*, 88:042815, 2013. doi: 10.1103/PhysRevE.88.042815.

Bernard De Bruyne, Nico H.J. Pijls, Bindu Kalesan, Emanuele Barbato, Pim A.L. Tonino, Zsolt Piroth, Nikola Jagic, Sven Möbius-Winkler, Gilles Rioufol, Nils Witt, Petr Kala, Philip MacCarthy, Thomas Engström, Keith G. Oldroyd, Kreton Mavromatis, Ganesh Manoharan, Peter Verlee, Ole Frobert, Nick Curzen, Jane B. Johnson, Peter Jüni, William F. Fearon, and FAME 2 Trial Investigators. Fractional flow reserve-guided pci versus medical therapy in stable coronary disease. *The New England journal of medicine*, 367(11):991–1001, 2012. doi: https://doi.org/10.1056/NEJMoa1205361.

J Bryant and R Day. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics*, 51(4):1372–1378, 1995.

Kai-Yuan Cai, Zhao Dong, Ke Liu, and Xian-Yuan Wu. Phase transition on the degree sequence of a random graph process with vertex copying and deletion. *Stochastic Processes and their Applications*, 121(4):885–895, 2011. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2010.12.008. URL https://www.sciencedirect.com/science/article/pii/S0304414910002930.

Simon Cauchemez, Achuyt Bhattarai, Tiffany L. Marchbanks, Ryan P. Fagan, Stephen Ostroff, Neil M. Ferguson, David Swerdlow, and Pennsylvania H1N1 working group. Role of social networks in shaping disease transmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences of the United States of America*, 108:2825–2830, 2011. URL https://doi.org/10.1073/pnas.1008895108.

M.R. Conaway and G.R. Petroni. Bivariate sequential designs for phase II trials. *Biometrics*, 51(2):656–664, 1995.

Colin Cooper and Alan Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003. doi: https://doi.org/10.1002/rsa.10084. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.10084.

Colin Cooper, Alan Frieze, and Juan Vera. Random deletion in a scale-free random graph process. *Internet Mathematics*, 1(4):463 – 483, 2003.

Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal*, 2(4582):769–782, 1948. ISSN 0007-1447. doi: 10.1136/bmj.2.4582.769. URL https://www.bmj.com/content/2/4582/769.

Abhirup Datta, Sudipto Banerjee, Andrew O. Finley, and Alan E. Gelfand. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016a. doi: 10.1080/01621459.2015.1044091. URL https://doi.org/10.1080/01621459.2015.1044091. PMID: 29720777.

Abhirup Datta, Sudipto Banerjee, Andrew O. Finley, Nicholas A. S. Hamm, and Martijn Schaap. Nonseparable dynamic nearest neighbor gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *The Annals of Applied Statistics*, 10(3):1286 – 1316, 2016b. doi: 10.1214/16-AOAS931. URL https://doi.org/10.1214/16-AOAS931.

Umberto De Ambroggio and Hiu Ching Yip. Degree evolution in a general growing network. *Statistics & Probability Letters*, 211:110151, 2024. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2024.110151. URL https://www.sciencedirect.com/science/article/pii/S0167715224001202.

Maria Deijfen and Mathias Lindholm. Growing networks with preferential deletion and addition of edges. *Physica A: Statistical Mechanics and its Applications*, 388(19):4297–4303, 2009. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2009.06.032. URL https://www.sciencedirect.com/science/article/pii/S0378437109004890.

Massimiliano DelPero, Luca Pozzi, and Judith C. Masters. A composite molecular phylogeny of living lemuroid primates. *Folia Primatologica*, 77(6):434–445, 2006. doi: 10.1159/000095390.

Francesco Denti, Michele Guindani, Fabrizio Leisen, Antonio Lijoi, William Duncan Wadsworth, and Marina Vannucci. Two-group poisson-dirichlet mixtures for multiple testing. *Biometrics*, 77(2):622–633, 2021. doi: 10.1111/biom.13314.

Richard C. Deonier, Simon Tavaré, and Michael S. Waterman. *Computational Genome Analysis: An Introduction*. Springer Science and Business Media, 20005.

Steffen Dereich and Peter Mörters. Random networks with sublinear preferential attachment: Degree evolutions. *Electronic Journal of Probability*, 14:1222 – 1267, 2009. doi: 10.1214/EJP.v14-647. URL https://doi.org/10.1214/EJP.v14-647.

S. Dias, N.J. Welton, D.M. Caldwell, and A.E. Ades. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine*, 29:932–944, 2010. doi: 10.1002/sim.3767.

Sofia Dias, Nicky J. Welton, Alex J. Sutton, and A.E. Ades. Nice dsu technical support document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011a. URL https://www.ncbi.nlm.nih.gov/books/NBK310366/.

Sofia Dias, Nicky J. Welton, Alex J. Sutton, and A.E. Ades. Nice dsu technical support document 4: Inconsistency in networks of evidence based on randomised controlled trials. 2011b. URL https://www.ncbi.nlm.nih.gov/books/NBK310372/.

Jacob C. Dunn and Jeroen B. Smaers. Neural correlates of vocal repertoire in primates. *Frontiers in Neuroscience*, 9(12), 2018. doi: 10.3389/fnins.2018.00534. URL https://doi.org/10.3389/fnins.2018.00534.

Richard Durrett. *Probability Models for DNA Sequence Evolution*. Springer New York, NY, 2008. doi: https://doi.org/10.1007/978-0-387-78168-6.

Rick Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2006. doi: 10.1017/CBO9780511546594.

Alan Frieze, Juan Vera, and Soumen Chakrabarti. The influence of search engines on preferential attachment. *Internet Mathematics*, 3(3):361–381, 2006. doi: 10.1080/15427951.2006.10129129. URL https://doi.org/10.1080/15427951.2006.10129129.

E.L. Frome. The analysis of rates using poisson regression models. *Biometrics*, 39(3): 665–674, 1983. doi: https://doi.org/10.2307/2531094.

M. Gamba and C. Giacoma. Subspecific divergence in the black lemur's low-pitched vocalizations. *The Open Acoustics Journal*, 7(1):49–53, 2008.

Marco Gamba and Cristina Giacoma. Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution*, 19(4):323–343, 2007. doi: 10.1080/08927014.2007.9522555. URL https://doi.org/10.1080/08927014.2007.9522555.

Marco Gamba, Olivier Friard, and Cristina Giacoma. Vocal tract morphology determines species-specific features in vocal signals of lemurs (eulemur). *International Journal of Primatology*, 33:1453–1466, 2012. doi: 10.1007/s10764-012-9635-y.

Marco Gamba, Valeria Torti, Vittoria Estienne, Rose M. Randrianarison, Daria Valente, Paolo Rovara, Giovanna Bonadonna, Olivier Friard, and Cristina Giacoma. The indris have got rhythm! timing and pitch variation of a primate song examined between sexes and age classes. *Frontiers in Neuroscience*, 10(249), 2016. doi: 10.3389/fnins.2016.00249. URL https://doi.org/10.3389/fnins.2016.00249.

A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp. *Handbook of Spatial Statistics*. Chapman and Hall, 2010.

Sebastiano Gili, Walter Grosso Marra, Fabrizio D'Ascenzo, Enrica Lonni, Andrea Calcagno, Margherita Cannillo, Flavia Ballocca, Enrico Cerrato, Martina Pianelli, Umberto Barbero, Massimo Mancone, James J. DiNicolantonio, Carl J. Lavie, Pierluigi Omedè, Antonio Montefusco, Stefano Bonora, Mauro Gasparini, Giuseppe Biondi-Zoccai, Claudio Moretti, and Fiorenzo Gaita. Comparative safety and efficacy of statins for primary prevention in human immunodeficiency virus-positive patients: a systematic review and meta-analysis. *European Heart Journal*, 37:3600–3609, 2016. doi: 10.1093/eurheartj/ehv734.

Tilmann Gneiting. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600, 2002. URL https://doi.org/10.1198/016214502760047113.

A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2): 50–61, 1995. doi: 10.1109/99.388960.

P D'Arcy Hart. A change in scientific approach: from alternation to randomised allocation in clinical trials in the 1940s. 319(7209):572–573, 1999. doi: 10.1136/bmj.319.7221.1372.

Remco van der Hofstad. *Random Graphs and Complex Networks Volume 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016. doi: 10.1017/9781316779422.

Dapeng Hu, Annette M. O'Connor, Chong Wang, Jan M. Sargeant, and Charlotte B. Winder. How to conduct a bayesian network meta-analysis. *Frontiers in Veterinary Science*, 7(271), 2020. doi: 10.3389/fvets.2020.00271.

Jeannette Janssen and Paweł Prałat. Rank-based attachment leads to power law graphs. *SIAM Journal on Discrete Mathematics*, 24(2):420–440, 2010. doi: 10.1137/080716967. URL https://doi.org/10.1137/080716967.

Brian Karrer and M. E. J. Newman. Message passing approach for general epidemic models. *Phys. Rev. E*, 82:016101, Jul 2010. doi: 10.1103/PhysRevE.82.016101. URL https://link.aps.org/doi/10.1103/PhysRevE.82.016101.

W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927. ISSN 09501207. URL http://www.jstor.org/stable/94815.

Arik Kershenbaum, Daniel T. Blumstein, Marie A. Roch, Caglar Akcay, and et al. Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Review*, 91(1):13–52, 2016. URL https://doi.org/10.1111/brv.12160.

Praveen Kumar and Efi Foufoula-Georgiou. Wavelet analysis for geophysical applications. *Reviews of geophysics*, 35(4):385–412, 1997.

Mathias Lindholm and Thomas Vallier. On the degree evolution of a fixed vertex in some growing networks. *Statistics & Probability Letters*, 81(6):673–677, 2011. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2011.02.015. URL https://www.sciencedirect.com/science/article/pii/S0167715211000617.

Mariya Lobanovska and Giulia Pilla. Penicillin's discovery and antibiotic resistance: Lessons for the future? *The Yale journal of biology and medicine*, 90(1):135–145, 2017. doi: https://doi.org/10.1016/j.idm.2017.02.001.

G. Lu and A.E. Ades. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101:447–459, 2006. doi: 10.1198/016214505000001302.

G. Lu, A.E. Ades, A.J. Sutton, N.J. Cooper, A.H. Briggs, and D.M. Caldwell. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in medcine*, 26: 3681–3699, 2007. doi: 10.1002/sim.2831.

Guobing Lu, Nicky J. Welton, Julia P.T. Higgins, Ian R. White, and A.E. Ades. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research synthesis methods*, 2:43–60, 2011. doi: 10.1002/jrsm.34.

Joseph M. Macedonia and Kathrin F. Stanger. Phylogeny of the lemuridae revisited: Evidence from communication signals. *Folia Primatologica*, 63(1):1 – 43, 1994. doi: 10.1159/000156787. URL https://brill.com/view/journals/ijfp/63/1/article-p1_1.xml.

Paul Maheshwari and Réka Albert. Network model and analysis of the spread of covid-19 with social distancing. *Applied network science*, 5:100, 2020. doi: 10.1007/s41109-020-00344-5.

Giovanna Maretti, Viviana Sorrentino, Andriamasitoly Finomana, Marco Gamba, and Cristina Giacoma. Not just a pretty song: an overview of the vocal repertoire of indri indri. *Journal of Anthropological Sciences*, 88:151–165, 2010. PMID: 20834055.

David J. Maron, Judith S. Hochman, Harmony R. Reynolds, Sripal Bangalore, Sean M O'Brien, William E. Boden, Bernard R. Chaitman, Roxy Senior, Jose López-Sendón, Karen P. Alexander, Renato D. Lopes, Leslee J. Shaw, Jeffrey S. Berger, Jonathan D. Newman, Mandeep S. Sidhu, Shaun G. Goodman, Witold Ruzyllo, Gilbert Gosselin, Aldo P. Maggioni, Harvey D. White, Balram Bhargava, James K. Min, G.B. John Mancini, Daniel S. Berman, Michael H. Picard, Raymond Y. Kwong, Ziad A. Ali, Daniel B. Mark, John A. Spertus, Mangalath N. Krishnan, Ahmed Elghamaz, Nagaraja Moorthy, Whady A. Hueb, Marcin Demkow, Kreton Mavromatis, Olga Bockeria, Jesus Peteiro, Todd D. Miller, Hanna Szwed, Rolf Doerr, Matyas Keltai, Joseph B. Selvanayagam, P. Gabriel Steg, Claes Held, Shun Kohsaka, Stavroula Mavromichalis, Ruth Kirby, Neal O. Jeffries, Frank E. Harrell Jr., Frank W. Rockhold, Samuel Broderick, T. Bruce Ferguson Jr., David O. Williams, Robert A. Harrington, Gregg W. Stone, Yves Rosenberg, and ISCHEMIA Research Group. Initial invasive or conservative strategy for stable coronary disease. *The New England journal of medicine*, 382(15):1395–1407, 2020. doi: https://doi.org/10.1056/NEJMoa1915922.

Gianluca Mastrantonio, Giovanna Jona Lasinio, Aleesio Pollice, Giulia Capotorti, Lorenzo Teodonio, Giulio Genova, and Carlo Blasi. A hierarchical multivariate spatio-temporal model for large clustered climate data with annual cycles. *Annals of Applied Statistics*, 13(2):797–823, 2017. doi: 10.1214/18-AOAS1212.

Karen McComb and Stuart Semple. Coevolution of vocal communication and sociality in primates. *Biology Letters*, 1:381–385, 2005. doi: 10.1098/rsbl.2005.0366. URL https://doi.org/10.3389/fnins.2018.00534. PMCID: PMC1626386.

Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226 – 251, 2003.

M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002. doi: 10.1103/PhysRevE.66.016128. URL https://link.aps.org/doi/10.1103/PhysRevE.66.016128.

M. E. J. Newman. *Networks: an introduction*. Oxford University Press, Oxford; New York, 2010. ISBN 9780199206650 0199206651.

Sharon-Lise T. Normand. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in medicine*, 18(3):321–359, 1999. doi: https://doi.org/10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P.

Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Mathematics*, 2(2):121–163, 2005. doi: 10.1080/15427951.2005.10129101. URL https://doi.org/10.1080/15427951.2005.10129101.

R. Pavasini, S. Biscaglia, E. Barbato, M. Tebaldi, D. Dudek, J. Escaned, G. Casella, A. Santarelli, V. Guiducci, E. Gutierrez-Ibanes, G. Di Pasquale, L. Politi, A. Saglietto, F. D'Ascenzo, and G. Campo. Complete revascularization reduces cardiovascular death in patients with st-segment elevation myocardial infarction and multivessel disease: systematic review and meta-analysis of randomized clinical trials. *European Heart Journal*, 41 (42):4103–4110, 2020. doi: 10.1093/eurheartj/ehz896.

Olivier Perrin and Wendy Meiring. Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36(4):1244–1250, 1999. URL https://www.jstor.org/stable/3215594.

Luca Pozzi, Marco Gamba, and Cristina Giacoma. The use of artificial neural networks to classify primate vocalizations: a pilot study on black lemurs. *American Journal of Primatology*, 72(4):337–348, 2010. doi: https://doi.org/10.1002/ajp.20786. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ajp.20786.

Paweł Prałat and Nicholas Wormald. Growing protean graphs. *Internet Mathematics*, 4 (1):1–16, 2007. doi: 10.1080/15427951.2007.10129135. URL https://doi.org/10.1080/15427951.2007.10129135.

Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007. doi: https://doi.org/10.1002/rsa.20137. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20137.

Havard Rue and Sara Martino. Approximate bayesian inference for hierarchical gaussian markov random fields models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192, 2007. ISSN 0378-3758. URL https://doi.org/10.1016/j.jspi.2006.07.016.

Havard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society*, 71(2):319–392, 2009. URL https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Tim Sainburg, Marvin Thielk, and Timothy Q. Gentner. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, 16(10), 2020. URL https://doi.org/10.1371/journal.pcbi.1008228.

Guido Schwarzer, James R. Carpenter, and Gerta Rücker. *Meta-Analysis with R*. Springer Cham, 2015. doi: https://doi.org/10.1007/978-3-319-21416-0.

Allison K. Shaw, Milena Tsvetkova, and Roozbeh Daneshvar. The effect of gossip on social networks. *Complexity*, 16(4):39–47, 2011. doi: https://doi.org/10.1002/cplx.20334. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cplx.20334.

Shinichiro Shirota and Alan E. Gelfand. Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, 11(2):481 – 503, 2017. doi: 10.1214/16-AOAS960. URL https://doi.org/10.1214/16-AOAS960.

Anna Lucia Sperber, Lynne M. Werner, Peter M. Kappeler, and Claudia Fichtel. Grunt to go - vocal coordinate of group movements in redfronted lemurs. *Ethology*, 123(12):894–905, 2017. URL https://doi.org/10.1111/eth.12663.

L. Tan and J. Jiang. *Digital Signal Processing: Fundamentals and Applications*. Elsevier Science, 2018. ISBN 9780128150726. URL https://books.google.de/books?id=MxlxDwAAQBAJ.

Daria Valente, Chiara De Greogorio, Valeria Torti, Longondraza Miaretsoa, Olivier Friard, Rose Marie Randrianarison, Cristina Giacoma, and Marco Gamba. Finding meanings in low dimensional structures: stochastic neighbor embedding applied to the analysis of indri indri vocal repertoire. *Frontiers in Neuroscience*, 9(12), 2019. doi: 10.3389/fnins.2018.00534. URL https://doi.org/10.3389/fnins.2018.00534.

Hans C van Houwelingen, Lidia R Arends, and Theo Stijnen. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, 21(4):589–624, 2002. doi: https://doi.org/10.1002/sim.1040.

Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67:056104, 2003. doi: 10.1103/PhysRevE.67.056104.

A.V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2):297 – 312, 1988. doi: 10.1080/01621459.2015.1123632.

Shaun P. Wilkinson and Simon K. Davy. phylogram: an r package for phylogenetic analysis with nested lists. *Journal of Open Source Software*, 3(26):790, 2018. URL https://doi.org/10.21105/joss.00790.

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, 2005. URL http://www.jstor.org/stable/3647580.

Appendix A

Supplementary material for A hierarchical spatio-temporal model for time-frequency data: An application in bio-acoustic analysis

The model is estimated using a set of synthetic data in this section in order to see if the model is able to retrieve the correct parameters that were used to simualte the data. A total of $N = 15$ synthetic sounds are generated over predefined time-frequency grids that are designed to closely resemble the real data. The number of log-frequency bins for all synthetic sounds is set as $H = 26$ which is identical to the available real data, whilst the number of time coordinates for each synthetic sound T_i is simulated uniformly from an interval of integers in the range $[10, 25]$. The temporal distance between two consecutive time coordinates is set as 0.01 seconds, which precisely matches that of the real data.

The general parameters, $\theta = (\phi_h, \phi_d, \phi_c, \gamma, \rho, \sigma, \lambda)$, are deliberately set to make the estimation challenging for the purpose of providing a level of assurance regarding the model's performance in real data applications. Specifically, the decays are set as $\phi_h = 0.69$, $\phi_d = 206$, and $\phi_c = 766$, which result in the practical ranges being 4.3, 0.092, and 0.024, respectively. The periodicity is set to be $\gamma = 0.06$. Notably, the practical range for the circular decay implies that the circular correlation between any two time coordinates separated by three time steps is nearly zero, thereby increasing the difficulty of its estimation. The non-separable parameter, variance, and weight between the two latent processes are set as $\rho = 0.85$, $\sigma^2 = 10$, and $\lambda = 0.5$, respectively.

Table A.1 The true simulated values, posterior means and 95% credible intervals (from top to bottom) of the general parameters.

σ^2	λ	γ	ρ	ϕ_d	ϕ_h	ϕ_c
10	0.5	0.06	0.85	206	0.69	766
9.702	0.476	0.06	0.805	167.821	0.038	1021.255
(8.637 10.763)	(0.434 0.519)	(0.06 0.06)	(0.394 0.992)	(123.967 228.794)	(0.034 0.044)	(730.176 1246.129)

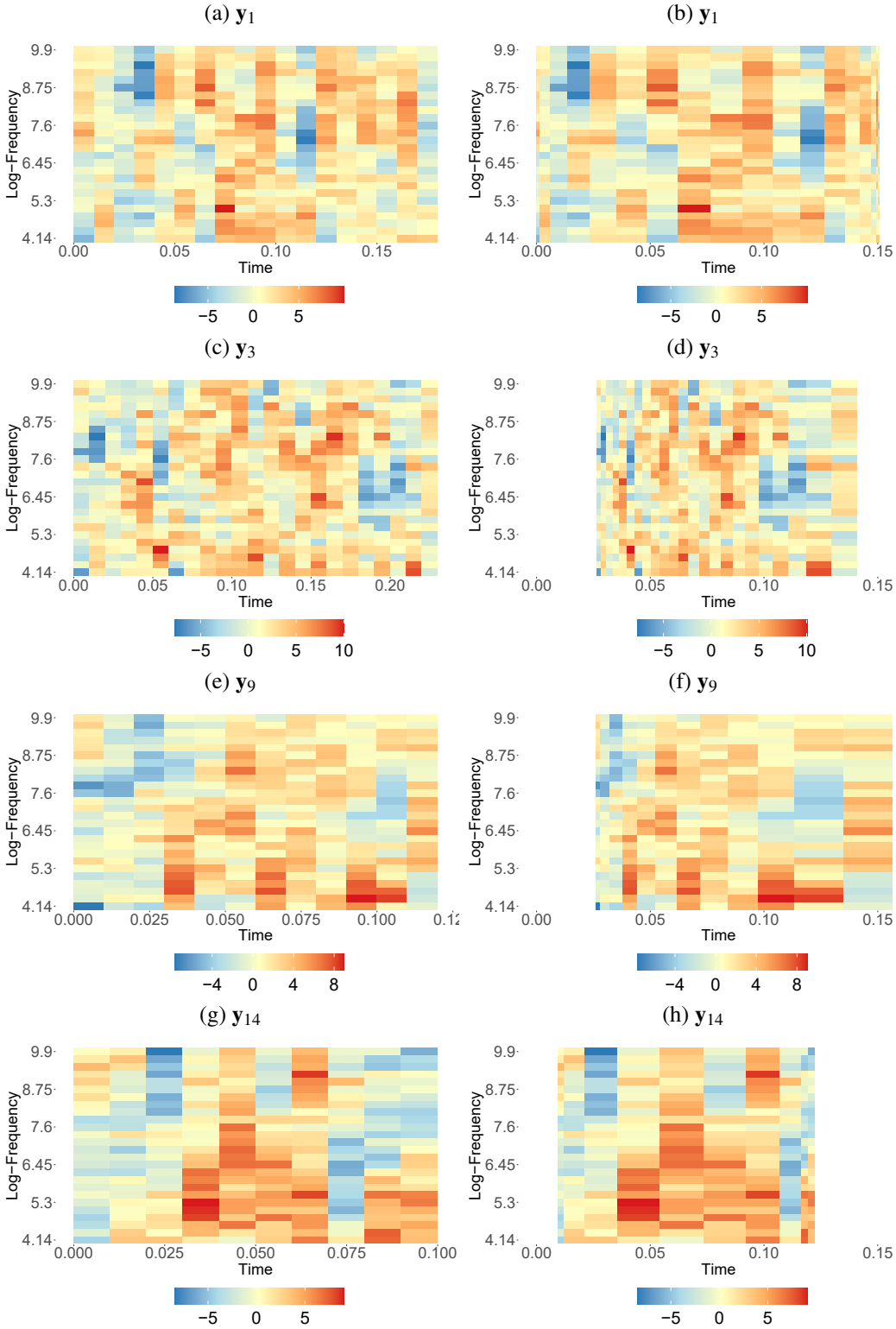
The values of the data-specific parameters, $\chi_i = (\alpha_i, \beta_i, \xi_i)$, used for the simulation can be found in Table A.2. With the exception of $\tilde{\alpha}_1$ and $\tilde{\beta}_1$, which are respectively set to 0 and 1, the parameters α_i and $\tilde{\beta}_i$ are simulated from $\alpha_i \sim \text{Uni}(0.05, 0.2)$ and $\tilde{\beta}_i \sim \text{Uni}(0.75, 0.95)$, respectively. The warping parameters $\xi_i = (\zeta_i, \delta_i)$ are simulated from their random effect distributions under the assumption that $a_\zeta = a_\delta = -1.5$, $b_\zeta = b_\delta = 1.5$, $\mu_\zeta = \mu_\delta = 0.7$, and $v_\zeta = v_\delta = 0.3$. A selected few of the synthetic sounds are depicted in Figure A.1. Each row of Figure A.1 contains two different plots of the same synthetic sound. The 4 plots on the left column are the spectrogram representation with the x-axis being the real-time axis given by the simulated T_i and the constant 0.01 time-step. The 4 plots on the right column are the same spectrogram representations with the x-axis being given by $\psi_i(t)$ under the simulated parametrizations.

The model is implemented using the same number of iterations, thinning, burn-in, and priors as the real data. Table A.2 and Table A.1 present the estimation summary for the data-specific parameters and the general parameters, respectively. The estimation of the frequency decay ϕ_h, and the warping parameters δ_2 and ζ_9 yields posterior means with relatively small 95% CIs that do not include the true values. However, the inference has largely demonstrated good recovery of the other 94 of the 97 simulated parameters.

Table A.2 The true simulated values, posterior means and 95% credible intervals (from top to bottom) of the data-specific parameters.

	α_i	$\tilde{\beta}_i$	ζ_i	δ_i	μ_i	τ_i^2
	0.00	1.00	0.75	0.88	-1.12	1.16
1	0.083	0.932	0.507	0.696	-0.217	0.764
	(0.00 0.218)	(0.759 1.00)	(0.112 0.858)	(0.232 1.062)	(-1.689 1.207)	(0.516 1.066)
	0.07	0.78	0.34	0.42	-0.08	1.27
2	0.113	0.695	0.596	0.911	1.024	1.41
	(0.00 0.228)	(0.561 0.808)	(0.109 1.034)	(0.56 1.198)	(-0.482 2.503)	(1.025 1.826)
	0.18	0.84	0.19	-0.26	3.86	1.39
3	0.174	0.873	-0.001	-0.214	3.843	1.529
	(0.104 0.226)	(0.767 0.964)	(-0.427 0.372)	(-0.568 0.126)	(2.144 5.737)	(1.233 1.846)
	0.08	0.92	0.94	0.42	7.13	1.53
4	0.082	0.761	0.879	0.482	7.777	1.935
	(0.033 0.156)	(0.57 1)	(0.531 1.163)	(-0.016 0.831)	(6.165 9.284)	(1.471 2.509)
	0.16	0.82	-0.14	0.12	2.00	0.96
5	0.151	0.96	-0.014	0.368	2.643	0.866
	(0.098 0.234)	(0.878 1)	(-0.278 0.25)	(0.015 0.707)	(1.125 4.196)	(0.572 1.201)
	0.12	0.88	0.85	0.60	-1.95	0.87
6	0.069	0.913	0.753	0.64	-0.735	1.127
	(0.01 0.143)	(0.815 1)	(0.419 0.942)	(0.287 0.884)	(-2.157 0.623)	(0.871 1.415)
	0.09	0.91	0.13	0.78	-6.12	0.85
7	0.089	0.964	0.202	0.886	-5.289	0.952
	(0.04 0.171)	(0.905 1)	(0.018 0.383)	(0.621 1.117)	(-6.716 -3.862)	(0.731 1.215)
	0.10	0.77	0.60	0.50	-3.76	0.64
8	0.034	0.853	0.401	0.436	-2.861	0.641
	(0 0.108)	(0.781 0.908)	(0.068 0.625)	(0.198 0.646)	(-4.206 -1.569)	(0.467 0.855)
	0.17	0.88	0.44	-0.19	-0.38	0.70
9	0.185	0.962	0.721	0.242	0.131	0.578
	(0.14 0.26)	(0.895 1)	(0.508 0.979)	(-0.02 0.547)	(-1.492 1.761)	(0.326 0.894)
	0.07	0.90	0.83	0.68	5.34	0.60
10	0.03	0.959	0.777	0.69	6.014	0.657
	(0 0.122)	(0.907 0.995)	(0.489 1.022)	(0.418 0.935)	(4.467 7.618)	(0.446 0.917)
	0.07	0.76	0.66	0.90	-1.46	1.31
11	0.072	0.893	0.327	0.568	-1.379	1.06
	(0 0.21)	(0.657 1)	(-0.104 0.87)	(0.093 1.07)	(-2.896 0.098)	(0.695 1.498)
	0.08	0.79	0.61	-0.06	-6.44	0.82
12	0.067	0.855	0.798	0.258	-6.604	1.074
	(0 0.172)	(0.735 0.958)	(0.504 1.13)	(-0.053 0.658)	(-8.022 -4.992)	(0.807 1.344)
	0.19	0.85	0.15	0.11	-2.81	1.20
13	0.196	0.946	0.197	0.293	-2.575	1.329
	(0.15 0.262)	(0.816 1)	(-0.038 0.501)	(0.009 0.696)	(-4.026 -0.83)	(1.055 1.629)
	0.06	0.78	0.82	0.81	-0.32	0.67
14	0.027	0.835	0.574	0.69	0.548	0.637
	(0 0.109)	(0.75 0.984)	(0.126 0.89)	(0.201 1.005)	(-0.897 2.041)	(0.372 0.956)
	0.07	0.85	0.13	0.82	-2.14	1.06
15	0.07	0.875	0.06	0.918	-2.476	1.276
	(0 0.173)	(0.78 0.966)	(-0.172 0.298)	(0.58 1.187)	(-4.388 -0.411)	(1.009 1.603)

Figure A.1 Plots of 4 synthetic sounds.



Appendix B

Supplementary materials for Anatomic versus ischemia-driven strategies for percutaneous coronary revascularization in chronic coronary syndrome: A network meta-analysis

Table B.1 Trial data available for clinical endpoint 2 (All cause deaths).

trial	follow-up	arm 1	occurrence	sample	arm	occurrence	sample	contrast
1	1.5	1	1	164	2	1	177	1,2
2	1.6	1	1	22	3	1	19	1,3
3	6	1	15	162	3	9	166	1,3
4	3	1	98	389	3	94	388	1,3
5	5	1	215	2591	3	233	2588	1,3
6	2	1	8	441	4	6	447	1,4
7	4.1	1	40	148	2	45	153	1,2
8	7	1	25	69	2	13	69	1,2
9	4.6	1	95	1138	2	85	1149	1,2
10	10	1	22	105	3	6	96	1,3
11	2.7	1	7	514	2	11	504	1,2
12	10	1	63	203	2	49	205	1,2
13	2	1	20	366	2	2	192	1,2
14	1	4	13	1250	5	22	1242	4,5
15	5	2	49	496	4	44	509	2,4

Table B.2 Results of different models for clinical endpoint 2 (All cause deaths).

model parameters	**FE**	**RE-Hom**	**RE-Het**
$d_{1,2}$	−0.201 (0.093)	−0.273 (0.231)	−0.274 (0.282)
$d_{1,3}$	−0.022 (0.076)	−0.278 (0.290)	−0.327 (0.519)
$d_{1,4}$	−0.337 (0.211)	−0.379 (0.460)	−0.375 (1.432)
$d_{1,5}$	0.217 (0.414)	0.174 (0.769)	0.168 (3.291)
$d_{2,3}$	0.179 (0.120)	−0.006 (0.364)	−0.053 (0.590)
$d_{2,4}$	−0.136 (0.195)	−0.107 (0.443)	−0.101 (1.429)
$d_{2,5}$	0.418 (0.407)	0.447 (0.760)	0.442 (3.290)
$d_{3,4}$	−0.315 (0.223)	−0.101 (0.544)	0.048 (1.522)
$d_{3,5}$	0.239 (0.421)	0.452 (0.821)	0.495 (3.332)
$d_{4,5}$	0.554 (0.357)	0.553 (0.619)	0.543 (2.962)
σ		0.448 (0.229)	
$\sigma_{1,2}$			0.522 (0.389)
$\sigma_{1,3}$			0.829 (0.634)
$\sigma_{1,4}$			1.563 (1.337)
$\sigma_{1,5}$			2.499 (1.441)
$\sigma_{2,3}$			1.080 (1.912)
$\sigma_{2,4}$			3.767 5.390)
$\sigma_{2,5}$			7.445 (6.836)
$\sigma_{3,4}$			4.024 (5.307
$\sigma_{3,5}$			7.343 (6.641)
$\sigma_{4,5}$			8.672 (7.225)

Table B.3 Trial data available for clinical endpoint 3 (CV deaths).

trial	follow-up	arm 1	occurrence	sample	arm	occurrence	sample	contrast
1	1.5	1	1	164	2	1	177	1,2
2	3.0	1	82	389	3	76	388	1,3
3	5.0	1	52	2591	3	101	2588	1,3
4	2.0	1	3	441	4	3	447	1,4
5	4.1	1	32	148	2	32	153	1,2
6	10.0	1	22	105	3	3	96	1,3
7	2.7	1	3	514	2	5	504	1,2
8	10.0	1	42	203	2	29	205	1,2
9	1.0	4	4	1250	5	7	1242	4,5
10	5.0	2	28	496	4	21	509	2,4

Table B.4 Results of different models for clinical endpoint 3 (CV deaths).

model parameters	**FE**	**RE-Hom**	**RE-Het**
$d_{1,2}$	−0.176 (0.164)	−0.004 (0.643)	−0.088 (0.509)
$d_{1,3}$	0.1458 (0.110)	−0.349 (0.720)	−0.453 (1.481)
$d_{1,4}$	−0.450 (0.311)	−0.216 (1.014)	−0.262 (1.597)
$d_{1,5}$	0.174 (0.730)	0.412 (1.696)	0.354 (3.447)
$d_{2,3}$	0.322 (0.197)	−0.345 (0.972)	−0.366 (1.567)
$d_{2,4}$	−0.274 (0.275)	−0.212 (0.977)	−0.174 (1.593)
$d_{2,5}$	0.350 (0.715)	0.416 (1.674)	0.442 3.445)
$d_{3,4}$	−0.596 (0.330)	0.133 (1.247)	0.191 (2.180)
$d_{3,5}$	0.028 (0.738)	0.761 (1.839)	0.807 (3.752)
$d_{4,5}$	0.624 (0.660)	0.628 (1.365)	0.616 (3.061)
σ		1.076 (0.529)	
$\sigma_{1,2}$			0.644 (0.671)
$\sigma_{1,3}$			2.256 (1.124)
$\sigma_{1,4}$			1.748 (1.363)
$\sigma_{1,5}$			2.506 (1.448)
$\sigma_{2,3}$			5.773 (5.598)
$\sigma_{2,4}$			4.538 (5.772)
$\sigma_{2,5}$			7.629 (6.905)
$\sigma_{3,4}$			7.313 (6.409)
$\sigma_{3,5}$			9.085 (7.127)
$\sigma_{4,5}$			8.900 (7.281)

Table B.5 Trial data available for clinical endpoint 4 (MI).

trial	follow-up	arm 1	occurrence	sample	arm	occurrence	sample	contrast
1	1.5	1	4	164	2	5	177	1,2
2	1.6	1	2	22	3	2	19	1,3
3	6.0	1	42	162	3	24	166	1,3
4	3.0	1	56	389	3	46	388	1,3
5	5.0	1	308	2591	3	266	2588	1,3
6	2.0	1	30	441	4	26	447	1,4
7	4.1	1	18	148	2	18	153	1,2
8	4.6	1	128	1138	2	134	1149	1,2
9	10.0	1	40	105	3	11	96	1,3
10	2.7	1	10	514	2	21	504	1,2
11	10.0	1	42	203	2	27	205	1,2
12	2.0	1	18	366	2	7	192	1,2
13	1.0	4	28	1250	5	31	1242	4,5
14	0.5	1	3	107	2	5	105	1,2
15	5.0	2	60	496	4	49	509	2,4

Table B.6 Results of different models for clinical endpoint 4 (MI).

model parameters	**FE**	**RE-Hom**	**RE-Het**
$d_{1,2}$	−0.000 (0.092)	0.027 (0.205)	0.021 (0.219)
$d_{1,3}$	−0.271 (0.071)	−0.497 (0.236)	−0.532 (0.450)
$d_{1,4}$	−0.216 (0.168)	−0.195 (0.362)	−0.198 (1.320)
$d_{1,5}$	−0.1048 (0.312)	−0.081 (0.622)	−0.092 (3.201)
$d_{2,3}$	−0.271 (0.116)	−0.524 (0.315)	−0.554 (0.501)
$d_{2,4}$	−0.216 (0.160)	−0.222 (0.358)	−0.219 (1.320)
$d_{2,5}$	−0.105 (0.308)	−0.108 (0.620)	−0.114 (3.201)
$d_{3,4}$	0.055 (0.183)	0.302 (0.434)	0.334 (1.395)
$d_{3,5}$	0.166 (0.320)	0.416 (0.665)	0.440 (3.231)
$d_{4,5}$	0.111 (0.264)	0.114 (0.503)	0.106 (2.923)
σ		0.394 (0.169)	
$\sigma_{1,2}$			0.381 (0.280)
$\sigma_{1,3}$			0.762 (0.544)
$\sigma_{1,4}$			1.383 (1.305)
$\sigma_{1,5}$			2.500 (1.442)
$\sigma_{2,3}$			0.809 (1.575)
$\sigma_{2,4}$			3.268 (5.210)
$\sigma_{2,5}$			7.603 (6.983)
$\sigma_{3,4}$			3.442 (5.007)
$\sigma_{3,5}$			7.297 (6.659)
$\sigma_{4,5}$			8.486 (7.202)

Table B.7 Trial data available for clinical endpoint 5 (Revas).

trial	follow-up	arm 1	occurrence	sample	arm	occurrence	sample	contrast
1	1.5	1	20	164	2	29	177	1,2
2	6.0	1	2	162	3	38	166	1,3
3	5.0	1	287	2591	3	396	2588	1,3
4	2.0	1	179	441	4	36	447	1,4
5	4.1	1	148	148	2	18	153	1,2
6	4.6	1	348	1138	2	228	1149	1,2
7	10.0	1	46	105	3	26	96	1,3
8	2.7	1	101	514	2	62	504	1,2
9	10.0	1	80	203	2	85	205	1,2
10	1.0	1	1	51	3	7	50	1,3
11	2.0	1	105	366	2	21	192	1,2
12	1.0	4	63	1250	5	46	1242	4,5
13	0.5	1	11	107	2	16	105	1,2
14	5.0	2	101	496	4	92	509	2,4

Table B.8 Results of different models for clinical endpoint 5 (Revas).

model parameters	**FE**	**RE-Hom**	**RE-Het**
$d_{1,2}$	−0.759 (0.059)	−0.905 (0.738)	−0.952 (0.854)
$d_{1,3}$	0.355 (0.072)	1.186 (1.038)	1.225 (1.351)
$d_{1,4}$	−1.294 (0.112)	−1.439 1.449)	−1.5189 (1.670)
$d_{1,5}$	−1.612 (0.225)	−1.764 (2.457)	−1.825 (3.482)
$d_{2,3}$	1.115 (0.0932)	2.0912 (1.276)	2.176 (1.599)
$d_{2,4}$	−0.535 (0.111)	−0.534 (1.443)	−0.567 (1.715)
$d_{2,5}$	−0.853 (0.225)	−0.859 (2.453)	−0.873 (3.504)
$d_{3,4}$	−1.649 (0.133)	−2.625 (1.781)	−2.744 (2.150)
$d_{3,5}$	−1.967 (0.236)	−2.950 (2.669)	−3.050 (3.734)
$d_{4,5}$	−0.318 (0.195)	−0.325 (1.984)	−0.306 (3.049)
σ		1.884 (0.580)	
$\sigma_{1,2}$			2.156 (0.817)
$\sigma_{1,3}$			2.413 (1.022)
$\sigma_{1,4}$			2.015 (1.371)
$\sigma_{1,5}$			2.495 (1.444)
$\sigma_{2,3}$			6.997 (5.590)
$\sigma_{2,4}$			6.545 (5.786)
$\sigma_{2,5}$			8.243 (6.486)
$\sigma_{3,4}$			7.956 (6.459)
$\sigma_{3,5}$			9.171 (7.017)
$\sigma_{4,5}$			9.216 (7.339)

Figure B.1 The probability of treatment k being the best against each clinical endpoint.