SPARSE BAYESIAN LEARNING APPROACH FOR DAMAGE DETECTION IN A POPULATION OF NOMINALLY IDENTICAL STRUCTURES

QIUHU ZHANG

PhD

The Hong Kong Polytechnic University

2020

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

Sparse Bayesian Learning Approach for Damage Detection in a Population of Nominally Identical Structures

ZHANG Qiuhu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

January 2020

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

ZHANG Qiuhu (Name of student)

To my family

for their love and support

Abstract

This thesis is dedicated to the development of a general damage detection framework for nominally identical structures (NISs) rather than only a particular single structure. The developed damage detection framework is formulated in an unsupervised learning scheme, which only makes use of response measurements from undamaged structures. It consists of two phases: the baseline and inspection phases.

In the baseline phase, historical response measurements from multiple nominally identically undamaged structures are utilized to establish a data-driven baseline model for representing healthy population features of all NISs. Three types of sparse Bayesian modelling approaches are proposed to deal with multiple sources of uncertainty in the measured responses, including measurement noise (intra-structure uncertainty) and structural variability in the materials and/or manufacturing processes (inter-structure uncertainty). The first modelling approach is introduced simply by pooling the inter-structure and intra-structure uncertainties such that standard sparse Bayesian learning (SSBL) can be implemented to model the population features of NISs. In the second modelling approach, an extension to SSBL termed heteroscedastic sparse Bayesian learning (HSBL) is proposed to address heteroscedastic training data, resulting from the pooling of multiple sources of uncertainty. In the third modelling approach, another extension to SSBL termed panel sparse Bayesian learning (PSBL) is proposed, in which

different sources of uncertainty can be modelled separately. Their performance is assessed in terms of three model quality indices, including the root mean square residual (RMSR), the mean standardized log loss (MSLL) and the sparsity ratio $\mathcal{K}$.

In the inspection phase, Bayesian residuals between new response measurements and population features predicted by the baseline model are examined for the identification of damage in NISs. Three categories of probabilistic diagnostic logics including frequentist null hypothesis significance testing (NHST), Bayesian point null hypothesis testing (PNHT), and the novel Bayesian NHST are compared in the capacities of the detection of damage, the quantification of damage extent, and the warning of diagnostic risk. The impact on structural damage diagnostics, of the three types of sparse Bayesian modelling approaches for constructing the baseline model in the baseline phase is investigated. The optimal baseline modelling approach and the optimal damage diagnostic logic are found. A case study of online condition assessment for railway wheels is conducted throughout this thesis to validate the feasibility and effectiveness of the proposed methods.

Publications

Journal Papers

Zhang Q. H. and Ni Y. Q. (2020). "An Improved Most Likely Heteroscedastic Gaussian Process Regression via Bayesian Residual Moment Estimator." *IEEE Transactions on Signal Processing*, vol. 68, pp. 3450-3460.

Ni Y. Q. and Zhang Q. H. (2020). "A Bayesian Machine Learning Approach for Online Detection of Railway Defects Using Track-side Monitoring." *Structural Health Monitoring*, doi.org/10.1177/1475921720921772.

Conference Papers

Zhang Q. H. and Ni Y. Q. (2018). "Bayesian Spectrum Approach for Wheel Flat Identification." *Proceedings of the 2018 Engineering Mechanics Institute Conference*, 29 May-1 June 2018, Boston, Massachusetts, USA.

Zhang Q. H. and Ni Y. Q. (2018). "Online Wheel Condition Detection Using Standard and Heteroscedastic Gaussian Process Learnings." *Proceedings of the 2nd International Workshop on Structural Health Monitoring for Railway System*, 17-19 October, Qingdao, Shandong, China.

Ni Y. Q. and Zhang Q. H. (2018). "A Bayesian Machine Learning Approach for Online Wheel Condition Detection Using Track-side Monitoring." *Proceedings of the 2018 International Conference on Intelligent Rail Transportation*, 12-14 December 2018,

Singapore.

Zhang Q. H. and Ni Y. Q. (2019). “A Bayesian probabilistic approach for damage detection for a population of nominally identical structures: a practical application in railway wheel condition assessment.” *Proceedings of the 12th International Workshop on Structural Health Monitoring*, San Francisco, California, USA.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Yi-Qing Ni, for his high sense of responsibility, valuable guidance, and constant support throughout the whole study. I am very lucky to share his vision, insight and passion in research. I thank him especially for his patient, constructive criticism and illuminating instruction. His conscientious attitude towards scientific research is the most valuable lesson I have learned from this PhD. It is my great honor to have this opportunity to have Prof. Ni as my supervisor.

I would also like to thank Prof. Sheng-Guo Wang at the University of North Carolina at Charlotte and Prof. Wen-Hua Wu at the Yunlin University of Science and Technology for their valuable suggestions on my research.

I am grateful to Dr. Song-Ye Zhu, Dr. Siu-Kai Lai, Dr. You Dong, Dr. Hua-Ping Wan, and Dr. Lu Zhou for the kindness in sharing their research experience which broadens my horizons about the academic world.

In addition, I am grateful to all of my friends and colleagues at the Hong Kong Polytechnic University, particularly Mr. Tai-Tung Wai, Mr. Wing-Hong Kwan, Dr. Jun-Fang Wang, Dr. Xiao-Zhou Liu, Dr. Xiao Wang, Dr. You-Wu Wang, Dr. Ran Chen, Mr. Chao Zhang, Mr. Si-Qi Ding, Mr. Chi Xu, Mr. Si-Xin Chen, Mr. Bei-Yang Zhang, Mr. Yang Lu, Ms. Si-Yi Chen, and Mr. Yun-Hao Wei for their very precious helps and

suggestions.

Finally, I dedicate this thesis to my beloved parents and elder brothers with gratitude for their everlasting love, encouragement and support.

The financial supports from the PhD research studentship of the Hong Kong Polytechnic University, the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant No. PolyU 152014/18E), and the Innovation and Technology Commission of Hong Kong Branch of Chinese National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1) are sincerely acknowledged.

Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike's information criterion
ANN	Artificial neural network
AR	Autoregressive
ARMA	Autoregressive moving average
ARMAX	Autoregressive moving average with exogenous
ARX	Autoregressive with exogenous
BF	Bayes factor
BIC	Bayesian information criterion
BNN	Bayesian neural network
BWNN	Bayesian wavelet neural network
CDF	Cumulative distribution function
CNN	Convolutional neural network
DFT	Discrete Fourier transform
EM	Expectation-Maximization
FAS	Fourier amplitude spectrum
FBG	Fiber Bragg grating
FRF	Frequency response function
GP	Gaussian process
GPR	Gaussian process regression
HGP	Heteroscedastic Gaussian process
HSBL	Heteroscedastic sparse Bayesian learning
IBF	Intrinsic Bayes factor
MCMC	Markov chain Monte Carlo

MM	Multiple model
MSLL	Mean standardized log loss
NARMAX	Nonlinear autoregressive moving average with exogenous
NARX	Nonlinear autoregressive with exogenous
NFAS	Normalized Fourier amplitude spectrum
NHST	Null hypothesis significant testing
NISs	Nominally identical structures
PCA	Principle component analysis
PDF	Probability density function
PSBL	Panel sparse Bayesian learning
PNHT	Point null hypothesis testing
RMSR	Root mean square residual
SBL	Sparse Bayesian Learning
SHM	Structural health monitoring
SHT	Simple hypothesis testing
SSBL	Standard sparse Bayesian learning
SVM	Support vector machine

Chapter 1

Introduction

1.1 Research Motivation

Modern society relies heavily on critical infrastructures such as airplanes, high-speed railways, high-rise buildings and long-span bridges. However, these infrastructures may deteriorate inevitably over time due to aging, and natural or man-made disasters such that damage accumulates throughout their service life, giving rise to increasing risk to people. Given the crucial importance of these infrastructures to the national economy and the people's life, proper measures should be taken to ensure their operational safety while considering the budget constraint.

In general, there exist three types of maintenance strategies to ensure infrastructure safety. In the reactive maintenance strategy, the concerned structure continues to operate until it fails and then is replaced. This maintenance strategy is obviously unacceptable when the life safety is of a big concern. The preventive maintenance strategy could be the most common practice to ensure infrastructure safety that is usually performed on a time or mileage basis. Yet, the preventive maintenance strategy depends heavily on expert knowledge and its drawbacks include: (1) the expert knowledge is subjective such that different engineers may make different judgments on the same structure; (2) potential

structural damage cannot be revealed in time before arriving at a scheduled maintenance cycle; (3) the preventive maintenance strategy combined with regular inspection is often time-consuming. The predictive maintenance strategy with the help of advanced sensing technologies allows for real-time monitoring of the concerned structure and thus it is advocated in recent years. However, the monitoring cost can be very expensive. The best choice of maintenance strategies is essentially a trade-off between life safety, technology feasibility and maintenance cost.

With the fast development of cost-competitive sensing and data acquisition technologies, structural health monitoring enables a novel and viable means for real-time inspection and condition-based maintenance for crucial infrastructures. Structural health monitoring (SHM) refers to the process of implementing a damage detection and characterization strategy for these infrastructures (Farrar and Worden 2006, Worden et al. 2007, Farrar et al. 2012). Damage is physically defined as changes to the material and/or geometric properties of a structure, including changes to the boundary conditions and structural connectivity, which adversely affect the structural performance. The SHM process involves observation of a structure over time using periodically or continuously sampled structural response measurements from an array of sensors, extraction of damage-sensitive features from these measurements and statistical analysis of these features to identify the current structural state. The most important benefit it brings about

is to allow the current regular inspection or the time-based maintenance plan to evolve into a condition-based maintenance philosophy and thus maintenance activities are only required to be performed when certain indicators show signs of decreasing performance or upcoming failure of the structure in concern.

Damage detection is an essential step for a successful SHM strategy via either physics-based approaches or data-driven approaches (Sohn et al. 2002, Worden et al. 2007, Farrar et al. 2012). Most current damage detection approaches in place are assumed to operate on a single structure of interest. Nevertheless, from the point of view of engineering needs, one may be often faced in monitoring and performing damage detection not on a single, but rather on a population of nominally identical structures that are made of the same materials, manufactured by the same specifications and assembled in the same factory. Typical examples of such structures are railway vehicles, aircraft fleets, wind turbines and many other more. If these nominally identical structures are possible to be monitored using the same device and assessed using the same damage detection method, the cost and time required will be significantly reduced.

Damage detection in nominally identical structures is more challenging compared to the counterpart in a single structure due to multiple sources of uncertainty present in SHM data (Papatheou et al. 2014 and 2015, Vamvoudakis et al. 2014, 2016 and 2018). For any structural unit, SHM data are inevitably polluted by measurement noise, yielding

observation error in the measured structural responses (uncertainty within structural units). Besides, even nominally identical, these structures are not truly identical due to variabilities in material, manufacturing, boundary conditions and so on, which cause another type of uncertainty in the measured structural responses (uncertainty between structural units). Obviously, damage detection cannot be effectively achieved among nominally identical structures by using a single physical or data-based model that characterizes structural condition of one particular structural unit. The alternative of performing damage detection separately on each structure is not more advisable due to the increased computational cost. More importantly, structural condition information extracted from SHM data cannot be exchanged among nominally identical structures by using a single physical or data-based model. The most desirable solution to this problem might be to establish a single population model for representing structural healthy state of all nominally identical structures in the same group using all SHM data measured from a number of healthy structures such that the damage states of other structural units can be identified by checking the differences between new SHM data and the corresponding structural responses estimated by the population model.

Stimulated by many engineering needs in the field of SHM applications, the research effort of this thesis is devoted to the exploration of damage detection in a population of nominally identical structures via sparse Bayesian learning (SBL) (Tipping 2001, Tipping

and Faul 2003). This work starts with the introduction of a model-free and response-only population model which enables the quantification of uncertainty in SHM data in nominally identical structures with the aid of SBL. Then, standard SBL is extended to the time-varying measurement noise case, which allows for heteroscedasticity in SHM data to be considered. After that, the panel SBL is proposed to allow different types of uncertainties resulting from measurement error within and structural variability between nominally identical structures to be considered separately in order to avoid the information overestimation in SHM data. Finally, a novel damage detection approach for the nominally identical structures is proposed based on the aforementioned three SBL population models and Bayesian hypothesis testing. It should be mentioned that the newly built structures are typically subjected to various load tests before they are put into service. These load tests ensure that initial defects or flaws, if any, in the newly built structures can be detected in time. Therefore, the group of nominally identical structures considered in this study are presumed to be initially undamaged.

1.2 Research Objectives

This study is intended to develop model-free and response-only damage detection approaches in nominally identical structures via sparse Bayesian learnings:

1. To make use of standard SBL to establish a data-driven population model to quantify

uncertainty in SHM data due to measurement noise as well as structural unit-to-unit variability between nominally identical structures.

2. To extend standard SBL to the heteroscedastic case to allow heteroscedasticity in the population dynamics of nominally identical structures to be modelled and quantified.
3. To extend standard SBL to panel counterpart to avoid information overestimation in the modelling of population features of nominally identical structure.
4. To develop an unsupervised damage detection approach for nominally identical structures based on Bayesian residual analysis and Bayesian hypothesis testing.

1.3 Thesis Outline

The thesis consists of the following seven chapters:

Chapter 1 gives the research motivation, the research objectives and the structure of this thesis.

Chapter 2 begins with a comprehensive review of research efforts on damage detection in individual structures, followed by the latest damage detection investigations achieved in a population of nominally identical structures. The merits and weaknesses of the existing damage detection approaches are discussed.

Chapter 3 develops a data-driven population model using standard SBL to quantify uncertainties in damage-sensitive features due to measurement noise and structural

variability in nominally identical structural characteristics. In this population model, no physical model is needed and only response data are required. The uncertainties due to reducible and irreducible between nominally identical structures are pooled together such that standard SBL can be used to model the population features. The hyperparameters associated with standard SBL are estimated by type-II maximum likelihood estimation. This population model can serve as a reference model later to evaluate the condition of the rest of nominally identical structures when new monitoring data are available. A practical application to model SHM data for assessing railway wheel quality is presented by using a track-side monitoring system based on a distributed fiber-optic strain sensor array, which enables the monitoring of multiple wheels and multiple trains.

Chapter 4 extends standard SBL to the heteroscedastic counterpart to account for heteroscedasticity in SHM data due to the combined uncertainty from measurement noise and inter-structure variability between nominally identical structures. In the heteroscedastic SBL, two standard sparse Bayesian models are employed, one to estimate the mean population features and the other to update the associated uncertainty. Such extension is critical for uncertainty quantification in SHM data due to various uncertainties and the subsequent damage identification based on statistical hypothesis tests. The effectiveness of the heteroscedastic SBL in uncertainty quantification is verified by comparison with the developed population model in the previous chapter.

Chapter 5 further extends standard SBL to panel SBL, allowing measurement error and inter-structure variability between nominally identical structures to be considered separately. It avoids information overestimation in SHM data. In addition, various uncertainties in the population features can be tracked and compared through analysis of variance due to measurement error and inter-structure variability between nominally identical structures. The superiority of the panel SBL in uncertainty quantification is verified by comparing the results given in Chapters 3 and 4.

Chapter 6 develops a novel Bayesian damage detection approach for a population of nominally identical structures based on Bayesian residual analysis and Bayesian hypothesis testing. The novel damage detection approach is realized by examining the discrepancies between the estimated population characteristics from sparse Bayesian population models proposed in the previous three chapters and new observations from the rest of nominally identical structures. If the residuals are as small as expected, the new structural units are considered healthy; otherwise, they are considered damaged. The threshold of the expected residuals is related to the diagnostic risks in the procedure of structural damage detection. If a smaller threshold is used, it may result in a larger rate of false positive errors (the healthy structures are falsely diagnosed as damaged); if a larger threshold is used, it could give rise to a larger rate of false negative errors (the damaged structures are falsely diagnosed as healthy). Therefore, the choice of the residual threshold

is essentially a balance between the false positive and false negative error rates. Bayesian residuals are decorrelated and standardized by the Mahalanobis linear transformation to meet the fundamental assumption of statistical hypothesis tests: the residuals are independent and identically distributed from the same population. To overcome the Jeffreys-Lindley paradox in the current Bayesian point null hypothesis testing that has been commonly used in structural damage detection for individual structures, a novel Bayesian point-to-point hypothesis testing is proposed to identify and quantify damage in a population of nominally identical new structures, in terms of intrinsic Bayes factor. The effectiveness of the proposed damage detection approach in a population of nominally identical structures are verified using online monitoring data of high-speed railway wheels.

Chapter 7 provides a summary of this thesis, the major findings and the potential future work.

Chapter 2

Literature Review

2.1 Introduction

The vibration-based damage detection, as part of the broader structural health monitoring, has been investigated across civil, mechanical and aerospace engineering fields over the past 40 years. The most important benefits it brings out include the fact that the vibration-based monitoring technique is relatively mature, the acquisition device is kept at a reasonable cost, the vibrational data can be available without interrupting the normal operation of the monitored structure.

The basic premise of the vibration-based damage detection lies in that damage will introduce changes into the material properties, geometric properties, boundary conditions or energy dissipation characteristics of the structure, which, in turn cause changes into the dynamic response characteristics and adversely affect the current or future performance of the structure. The vibration-based damage detection allows for damage, even invisible in the monitored structure to be identified and located at the earliest time, thus giving rise to the increased safety and serviceability.

In general, the vibration-based damage detection can be classified into two main categories, the physics-based approach and the data-driven approach. The physics-based

approach is usually implemented by building a physical model of the structure of interest such as a finite element model. The data-driven approach also needs to establish a model, but rather a statistical one to characterize structural conditions from all the possible healthy and damage states for the interested structure with the use of regression approaches or classification techniques such as neural networks, support/relevance vector machines and Gaussian processes.

We start with literature review of the physics-based and data-driven damage detection approaches for individual structures in Section 2.2, followed by the review of the damage detection studies for a population of nominally identical structures in Section 2.3. Some concluding remarks are made in Section 2.4.

2.2 Damage Detection for Individual Structures

2.2.1 Physics-based damage detection methods

Most of the current physics-based damage detection methods for individual structures examine changes in conventional basic modal properties such as natural frequencies, mode shapes and damping ratios (Cawley and Adams 1979, Idichandy and Ganapathy 1990, Farrar and Cone 1994, Farrar et al. 1994, Farrar and Jauregui 1996, Salawu 1997, Ren and Roeck 2002, Peeters and De Roeck 2001, Teughels and De Roeck 2004). Typical

benchmark examples for damage detection of in-situ structures are the I-40 Bridge (Farrar and Cone 1994, Farrar et al. 1994 and 1995, Farrar and Jauregui 1996, Doebling and Farrar 1998) and the Z24 highway bridge (Peeters and De Roeck 2001, Teughels and De Roeck 2004). Low-frequency modal properties are often employed during the damage detection process due to measurement data incompleteness (Idichandy and Ganapathy 1990).

One of the major technical challenges associated with the physics-based damage detection is that damage is typically a local phenomenon and it may not significantly influence the low-frequency modal properties of a structure. In many real-world SHM applications, the physics-based damage detection is found to be insensitive to localized damage if only low-frequency modal properties are examined. Therefore, many efforts have been also made to find alternative features that have a higher sensitivity to localized damage such as modal shape curvature, modal strain energy and modal flexibility. Pandy et al. (1991) proposed to substitute mode shapes with mode shape curvatures in order to obtain spatial information about damage which have the effect of amplifying any discontinuities in the mode shape caused by localized damage. This approach was applied in assessing changes in mode shape curvature for the I-40 Bridge (Farrar et al. 1994), the Z24 bridge (Abdel Wahab and De Roeck 1999), an FRP highway bridge (Guan et al. 2006) and the plate-like structure (Rucevskis et al. 2015). Stubb et al. (1992) proposed the

concept of modal strain energy that is strain energy stored in a structure when it deforms as alternative features for local damage detection. The basic idea behind the modal strain energy-based damage detection lies in that damage will induce changes in the distribution of the strain energy stored in the undamaged and damaged structures, which is more appropriate to detect and locate local damage. Several successful applications of the physics-based damage detection based on modal strain energy can be found in Doebling et al. (1997), Cornwell et al. (1999), Shi et al. (1998 and 2000), Hu et al. (2006 and 2009), Yan et al. (2012).

Apart from the aforementioned alternative features extracted from basic modal properties, the modal flexibility is also proposed to detect and localize damage in structures. For example, Pandey and Biswas (1994 and 1995) identified damage in a wide-flange steel beam through the evaluation of the changes in the flexibility matrix of the structure. The modal flexibility was then used to assess the health condition of a three-span reinforced-concrete highway bridge (Toksoy and Aktan, 1994). Ni et al. (2008) conducted the damage detection in the cable-stayed Ting Kau Bridge based on modal flexibility analysis. Other damage detection applications using modal flexibility can be found in Koo et al. (2010) and Sung et al. (2014). In many cases, these derived features from basic modal properties are based on a physical interpretation of the relationship between the changes in modal properties and the changes to the structural properties.

There are pros and cons for these derived features; the interested reader can find detailed discussions in Farrar and Worden (2012).

Another major challenge associated with the physics-based damage detection is that the physics-based damage detection is in essence an inverse problem, which can be ill-posed and ill-conditioned. As a result, the existence, uniqueness and stability of a solution to this inverse problem cannot be guaranteed (Tikhonov 1995, Friswell and Mottershead 1995). Therefore, the inverse problem has to be solved in either an optimization procedure or some form of regularization. The optimization solution to the inverse problem often suffers from local minima and thus various global optimization algorithms have been proposed to obtain an optimum for the physics-based damage detection such as genetic algorithms (Chou and Ghaboussi 2001, Hao and Xia 2002, Perera and Torres 2006) and particle swarm optimization (Seyedpoor 2012).

Initial regularization methods for the physics-based damage detection (Fritzen and Jennewein 1998, Friswell 2006, Chen 2008, Weber and Paultre 2009, Entezami and Shariatmadar 2014, Rucevskis et al. 2015) make use of the ℓ2-norm regularization that is also termed Tikhonov regularization to produce a smooth solution to the inverse problem. However, the solution is sometimes found to be over-smoothed as Tikhonov regularization promotes smoothness especially when the number of the available sensors is limited (Zhang and Xu 2016). More recent regularization methods for the inverse

damage detection problem employ sparsity constraints including ℓ1-norm regularization (Hernandez 2014, Zhou et al. 2015 and 2018, Zhang and Xu 2016, Zhang et al. 2017, Wu and Zhou 2018, Hou et al. 2018 and 2019) or the equivalent but more flexible Bayesian regularization (Huang et al. 2017 and 2018, Hou et al. 2019). The concept of sparsity is inherent in many practical engineering applications and for example, in the content of structural health monitoring, damage occurs in part of the elements or substructures of the whole structure. The sparsity constraints are generally serving as a term to penalize physical parameters that are relevant to structural damage such that those relevant to structural damage can be better identified. The regularization methods using various sparsity constraints allow for more robust damage detection results especially when the measurement noise is present in the measured structural response.

The third major challenge confounded in the physics-based damage detection is that SHM data are often polluted by measurement noise, which may give rise to false damage detection results. Therefore, it is essential to analyze and quantify the uncertainty resulting from measurement noise in SHM data to determine the associated influence on the subsequent structural damage detection. Papadopoulos and Garcia (1998) developed a probabilistic framework to improve the robustness characteristics of current damage detection methodologies with the use of stochastic finite element models. Sohn and Law (1997) proposed a Bayesian probabilistic approach to locate and quantify damage using

estimated modal parameters when the measurement data are corrupted with measurement noise. This approach was then applied to identify damage of a reinforced-concrete bridge column (Sohn and Law 2000). A series of Bayesian methodologies for the physics-based damage detection are proposed by Beck and his colleagues (Vanik et al. 2000, Beck et al. 2001, Ching and Beck 2004, Yuen et al. 2004 and 2006, Lam et al. 2014, Huang et al. 2017 and 2018).

The fourth major challenge for the physics-based damage detection lies in the fact that the damage-sensitive features are also sensitive to the changing environment (for example temperature) and operational conditions (loading and boundary) and consequently this may have serious detrimental effects on damage detection performance due to the fact that changes in structural responses induced by the changing environment or operational conditions may be so significant as to mask those caused by damage (Peeters and Roeck 2001). To remove or alleviate the environmental or operational effect on structural damage detection, a vast of studies have been conducted. The most common solution to the damage detection in changing environmental and operational conditions is to correlate damage-sensitive features with the changing environmental or operational factors through regression or interpolation approaches before physics-based damage detection methods are applied (Peeters et al. 2001, Worden et al. 2002, Ni et al. 2005 and 2009, Sohn 2006, Hua et al. 2007, Cross et al. 2011, 2012 and 2013, Coletta et al. 2019).

There are still many other technical challenges for the physics-based damage detection. For example, the process of extracting basis modal parameters from measured input and response data involves fitting a linear model to the data obtained before and after damage. Yet, many types of damage introduce nonlinearities into the interested structure and these nonlinearities may cause the structure to violate the three basic assumptions in experimental modal analysis (Weber et al. 2009, Li and Law 2010, Ebrahimian et al. 2017, Xin et al. 2019). These assumptions are: the structure is linear; the structure is time invariant; the structure exhibits reciprocity.

2.2.2 Data-driven damage detection methods

The data-driven damage detection approach is often dependent on a statistical model for SHM data such as the use of time series models or machine learning algorithms.

a) ***Time series models***

Time series models can be employed to estimate the relationship between the measured dynamic response outputs and inputs when the monitored structure is known to be in its normal condition. The underlying assumption on damage detection using time series models is that damage causes the structure to response in a manner inconsistent with its undamaged relationship and thus the time series model would not be able to predict the subsequent monitoring data in its damage condition well. As a result,

significant changes in parameters associated with time series models or unexpectedly large prediction residuals can be attributed to the presence of damage.

Initially, linear time series models were popularly used to recover the data input-output relationship in a structure under healthy condition such as autoregressive (AR) model, autoregressive with exogenous (ARX) input model (Sohn and Farrar 2001 and Fugate et al. 2001), autoregressive moving average (ARMA) model (Nair et al. 2006, Nair and Kiremidjian 2007, Carden and Brownjohn 2008, Yao and Pakzad 2012) and autoregressive moving average with exogenous (ARMAX) model (Mei et al. 2016). Nonlinear time series models including nonlinear ARX (NARX) input model and nonlinear ARMAX (NARMAX) model were then proposed to better recover the input-output relationship in a healthy structure that responds in a nonlinear manner in its initial undamaged state (Wei et al. 2005, Oh and Sohn 2009, Peng et al. 2011, Yao and Pakzad 2012). It should be noted that in the application of time series models to damage detection, the model order and the associated model parameters have to be determined carefully through available techniques such as Akaike's information criterion (AIC) or Bayesian information criterion (BIC) for choosing the model order while least square or maximum likelihood methods for estimating model parameters.

Despite conventional time series models described above, there exist other types of time series models that are possible to derive a data-driven damage detection approach.

For example, Bernal (2013) proposed a novel damage detection method using a Kalman filter to deal with uncertainties in SHM data resulting from measurement noise and changing operational conditions. Yang et al. (2006 and 2007) made use of adaptive extended Kalman filter to identify structural damage without load information. Recently, a cointegration-based data model was proposed to remove the changing operational and environmental effects on long-term SHM data collected from in-situ structures (Cross et al. 2011). More recently, a Bayesian dynamic linear model was utilized to quantify uncertainty in model parameters and hidden state variables which tends to be more robust than conventional time series models (Nguyen et al. 2019) because the prior distributions placed on the model parameters provide an automatic "Occam's razor" effect, penalizing unnecessarily complex model.

b) ***Machine learning algorithms***

Another main stream of data-driven damage detection approach for individual structures is based on machine learning algorithms. For instance, Tsou and Shen (1994) identified structural damage characteristics (location and severity) from the change of its dynamic properties (eigenvalues and mode shapes) through an artificial neural network (ANN) model by which model damage location and severity could be identified. Zang and Imregun (2001) made use of the measured frequency response function data to train an ANN model such that principal component analysis could be applied to detect potential

structural damage.

However, the ANN training may suffer from overfitting problem and the resulting data model tends to present to a low generalization performance. To deal with this problem, Yuen and Lam (2006) proposed a Bayesian probabilistic method for selecting a proper number of hidden layers and neurons in training an ANN model, which was then improved using Ritz vectors as damage-sensitive features to characterize damage-induced changes (Lam et al. 2006). Bayesian neural network (BNN) was recently proposed for bridge integrity assessment (Arangio and Beck 2012). Jiang et al. (2007 and 2009) and Jiang and Mahadevan (2008) developed a Bayesian wavelet neural network (BWNN) method to investigate the effect of noise on the accuracy of structural damage detection. To identify and quantify structural damage, Bayesian hypothesis testing was utilized (Jiang and Mahadevan 2008, Sankararaman and Mahadevan 2011 and 2013, and Subramanian and Mahadevan 2019). It was found that from comparative studies that the BWNN-based damage detection approaches outperformed the existing wavelet denoising methods when SHM data were corrupted by the measurement noise.

Convolutional neural network (CNN) which is a more powerful machine learning algorithm was recently introduced in the data-driven damage detection (Jamshidi et al. 2018, Abdeljaber et al. 2017 and 2018). The outstanding benefit of it lies in that it enables the extraction of optimal damage-sensitive features automatically from SHM data,

whereas it needs massive training data and large computational resource. For example, Abdeljaber et al. (2017 and 2018) developed one-dimensional CNN for structural damage detection using vibrational data, which was then verified on SHM benchmark data from IASC-ASCE Structural Health Monitoring Task Group. Bao et al. (2019) and Tang et al. (2019) proposed to convert the raw one-dimensional vibrational data to images such that they can be used to train two-dimensional CNNs for classifying structural conditions more easily.

Apart from neural networks, many other machine learning algorithms are also possible to develop a data-driven damage detection approach. Support vector machine (SVM) that is computationally more efficient was utilized for structural damage detection (Worden and Lane 2001). Bornn et al. (2009) developed an autoregressive SVM for structural damage detection based on the measured time series data. Oh and Sohn (2009) made use of SVM to diagnose structural damage condition under the changing environmental and operation variations. A Bayesian extension to SVM was recently proposed by Wang et al. (2019) to develop a probabilistic damage detection method for rail condition assessment based on in-situ SHM data. Another extension to SVM, the relevant vector machine was employed for fast structural condition diagnostics and prognostics by Wang et al. (2012).

2.3 Damage Detection for Multiple Nominally Identical Structures (NISs)

The problem of damage detection in a population of nominally identical structures was considered by Papatheou et al. (2014, 2015) who employed two different variants of the tail wing of a Piper PA-28 aircraft to create two pairs of nominally identical structures by separating the tail wings in half. An experimental investigation was conducted to explore the diversity and complexity in modal properties such as natural frequencies, mode shapes and frequency response function (FRF) magnitudes for the considered nominally identical aircraft wings. The Mahalanobis squared distance was used to quantify uncertainty in modal properties and identify potential damaged aircraft wings, where a proper threshold must be carefully determined.

Chandrashekhar and Ganguli (2016) conducted numerically a damage detection study on a set of delaminated composite plates through Monte Carlo simulation. The effect of uncertainties due to material variability as well as measurement noise on statistical analysis of the modal frequencies for these nominally identical composite plates was investigated. It is found that the changes in modal frequencies due to high level of material variability may mask the ones induced by structural damage. A robust fuzzy logic system was then proposed to identify damaged composite plates under multiple sources of uncertainty.

Instead of an empirical statistical analysis of the FRF magnitude and phase measured

from a population of nominally identical structures, a closed-form probabilistic uncertainty quantification model of these extracted damage-sensitive features was developed by Mao and Todd (2016). A Bayesian recursive classification method was then proposed to statistically identify different damage types/locations in multiple ball bearings based on the sequential probability ratio test, that are viewed as nominally identical structures operating in rotating machinery.

The cross-dependence of structural health monitoring data from a population of nominally identical structures was then considered by Schubert et al. (2015) through the comparison of two distinct linear models (the mixed-effects linear model and the random intercept linear model) in modelling the probability of detection process for fatigue crack in aircraft components. In the mixed-effects model, uncertainties that exist in the monitoring data due to the measurement error and variability between nominally identical structures are mixed together leading to an independent data assumption, while the two types of uncertainties are considered separately through the introduction of a particular correlation data structure in the random intercept model. Simulated data and experimental data from three test specimens of a wing attachment lug that is a crucial component of the aircraft were used to demonstrate the advantages of the random intercept model in analyzing structural health monitoring data acquired from a population of nominally identical structures. Nevertheless, as the authors indicated, the random intercept linear

model is of limited usage due to the fact that the underlying model can be strongly nonlinear in many more complex applications. More importantly, the correlation data structure in the random intercept linear model is defined manually that can be not the real case.

The damage detection sensitivity characterization of the same structural health monitoring system applied to a population of nominally identical structures or crucial components under the same environmental conditions was investigated by Janapati et al. (2016). An experimental study was conducted on 30 identical composite plates by an active-sensing acousto-ultrasound-based SHM method and it was found that the position of each sensor–actuator pair with respect to a known damage location and the damage growth pattern reflecting the variability between nominally identical structures are the two most critical parameters influencing the reliability of the probability of damage detection.

The problem of damage detection in a population of nominally identical structures for vibration response-only cases was preliminarily explored in a series of studies by Vamvoudakis et al. (2014, 2015, 2016, 2018) with the use of conventional statistical time series methods (2015), a multiple model (MM) scheme (2014), an unsupervised MM-based method using autoregressive model (2015) and its enhanced version using principle component analysis (PCA) (2016). A more comprehensive study of the unsupervised

MM-based damage detection scheme using statistical time series methods was recently presented by the same authors (2018). The main idea behind the MM-based damage detection scheme is to establish a healthy sample space for nominally identical structures using vibrational response data measured from a number of nominally identical healthy structures, with which sample space the condition of the rest of structural units can be identified through statistical hypothesis testing such as likelihood ratio or Kullback-Leibler divergence tests. Experimental studies conducted on 31 nominally identical composite beams that are served as main components of commercial unmanned aerial vehicles demonstrated that the unsupervised PCA-enhanced MM-based damage detection was the optimal solution to the data-driven response-only damage detection problem in a population of nominally identical structures. This data-driven damage detection scheme using statistical time series methods was recently extended by Poulimenos et al. (2019) to investigate the performance of transmittance function-based damage detection in a set of composite beams with manufacturing variability subject to impact damage varying operating conditions. The problem of damage detection in a population of nominally identical railway structures was investigated by Oregui et al. (2015).

The effect of manufacturing uncertainty on the problem of damage detection in a population of composite airfoil structures was investigated by Teimouri et al. (2015 and 2017). Conventional neural network models were employed to predict the size and

location of damage with five composite airfoil samples in different thicknesses (Teimouri et al. 2015). Gaussian process models which enable probabilistic modelling and quantitative evaluation of the associated uncertainty in damage state assessment for the rest of composite airfoil structures were proposed to derive a more robust and reliable structural health monitoring system (Teimouri et al. 2017).

Studies not explicitly focused on the problem of damage detection in a population of nominally identical structures but addressing quantification of uncertainties due to inter-structure variability and measurement error, were conducted by Eerland et al. (2016) and Tabor et al. (2018) via Gaussian process learnings. Eerland et al. (2016) proposed a Gaussian process model to capture the probability distribution of a set of aircraft trajectories from historical measurement data. A single probabilistic model was derived which can be used for both modelling the dispersion of trajectories along the common flightpath and for measuring the difference of new trajectories from historical data.

Tabor et al. (2018) proposed an extension to standard Gaussian process regression (GPR) to model the data sets composed of only a few replicated specimens and displaying a heteroscedastic behavior. The heteroscedasticity in the experimental data is due to inter-structure variability as well as several uncontrollable factors in laboratory experiments. A global Gaussian process model was established for probabilistic modelling of heteroscedastic experimental data and this model can be used to study probabilistically

the quantity how adding fibers to high-performance concrete decreases water permeability. An application of the method to the data of high-performance fiber-reinforced concrete experiments highlights fiber-added benefits for reducing water permeability caused by macrocracks.

2.4 Summary

The problem of damage detection in individual structures has been well defined, whereas the counterpart in nominally identical structures remains challenging due to different sources of uncertainty present in SHM data. Taking into account the computational complexity in data modelling and uncertainty quantification in nominally identical structures, the data-driven damage detection approach would be more preferable than the physics-based approach. Nevertheless, the current data-driven damage detection approaches for nominally identical structures focuses on using conventional time series models, which is computationally inefficient. Moreover, the data-driven damage detection approaches based on conventional time series models can suffer from the overfitting problem as there exist many adjustable parameters in time series models. The essence of overfitting is that the elaborated time series models attempt to describe the effect of measurement noise instead of the underlying behavior of the structure of concern. To alleviate the overfitting problem in time series models, model selection techniques

such as AI and BIC have to be utilized, whereby the choice of optimal time series models has to be executed iteratively. As a result, additional SHM data and computational cost have to be taken on the determination of model order and the associated model parameters. Since damage detection approaches have to be executed for many times for a population of nominally identical structures, a computationally efficient and noise robust damage detection approach for nominally identical structures is highly desired.

Chapter 3

SSBL for Population Features of NISs

3.1 Introduction

The problem of damage detection in a population of nominally identical structures is very challenging in that there exist two different categories of uncertainty in SHM data. The first one called epistemic uncertainty is due to measurement noise in the measured data. This type of uncertainty is sometimes referred to as reducible uncertainty. The second type of uncertainty called aleatory uncertainty in SHM data is caused by inherent variability that exists between nominally identical structures, made in the same way of the same materials but with material, geometric or manufacturing variability. This type of uncertainty is sometimes referred to as irreducible uncertainty. It is therefore essential to analyze and quantify the uncertainties in SHM data to determine the associated influence on the subsequent structural diagnosis and prognosis for the population of nominally identical structures.

The simplest way in quantifying the associated uncertainties in SHM data is to simply combine the uncertainty within an individual structure due to measurement noise (intra-structure uncertainty) and uncertainty between nominally identical structures due to material or manufacturing variability (inter-structure uncertainty). The differences

between the damage-sensitive features measured from an individual structure and the associated population feature of all nominally identical structures are further assumed to be independent and Gaussian-distributed with zero mean. This simplifying assumption allows a data-driven response-only population feature model to be formulated for representing the normal (healthy) condition of all nominally identical structures and quantifying uncertainties in damage-sensitive features for nominally identical structures. It should be noted that a population of nominally identical structures typically operate in the same operational and environmental conditions. Therefore, it is made possible to separate the changes in SHM data caused by the varying operational and environmental conditions from the changes due to structural damage. If the changes in SHM data are found to appear among the whole group of NISs, they are believed to be caused by the varying operational and environmental conditions; if the changes arise only in a few of nominally identical structures but are not observable for most of NISs, they are believed to be caused by potential structural damage. Thus, this study deals with two main sources of uncertainties: measurement noise and structural variability.

We formulate the population feature model of nominally identical undamaged structures using standard sparse Bayesian learning (SSBL) in Section 3.2. SSBL is a fully probabilistic methodology, thus allowing for the consideration of various uncertainties in the population modelling of the damage-sensitive features for nominally identical

structures. Moreover, by assigning a hierarchical prior scheme, redundant basis functions employed in the standard sparse Bayesian modelling of the population damage-sensitive features can be automatically pruned, giving rise to a sparse population model that enables to maintain good reconstruction and prediction accuracies not only on training data but also on unseen data (the pruning procedure eliminates redundant basis functions in the standard sparse Bayesian modelling of the population damage-sensitive features while this procedure itself does not eliminate the damage-sensitive features). Furthermore, the sparse population model enables to make fast prediction and thus the subsequent damage detection performed on new structural units can be performed in a timely manner. We apply the type-II maximum likelihood algorithm to infer the model hyperparameters associated with the proposed data-driven population feature model. Section 3.3 illustrates the application of the proposed population model based on SSBL to a practical example. Section 3.4 presents a summary of the results of this chapter.

3.2 Problem Formulation

3.2.1 The training data

Assume that we have M nominally identical sample structures which are known undamaged. From the mth sample structure, we can extract a set of training data $\mathbf{D}_m =$

$\{(\mathbf{x}_{mn}, y_{mn}): n = 1, \dots, N\}$ to define the undamaged condition of the population of nominally identical structures, where $\mathbf{x}_{mn}$ is the vector-valued training input of dimension T, y_{mn} is the corresponding scalar-valued training output, and N is the number of the training input-output pairs from each sample structure. In the case of M nominally identical sample structures, we have in total M training datasets $\mathbf{D} = \{\mathbf{D}_m: m = 1, \dots, M\}$ and thus the total number of training data points is $K = M \times N$.

The training data can be the transformed responses in the frequency domain based on Fourier or wavelet transforms, such as Fourier amplitude spectrum and power spectrum density. They can be also specific components extracted from raw SHM data that are possible to define the normal (healthy) condition of nominally identical structures.

3.2.2 Model specification

As mentioned earlier, there are multiple types of uncertainty that exist in the training data and also it is difficult to consider these uncertainties separately. The simplest approach to take into account these uncertainties together is to define a single population model for representing the healthy status of all nominally identical structures. Before introducing the population model that defines healthy condition of nominally identical structure, the training inputs and outputs are first realigned as $\mathbf{x} = [\mathbf{x}_{11}, \dots, \mathbf{x}_{1N}, \dots, \mathbf{x}_{M1}, \dots, \mathbf{x}_{MN}]^{\mathrm{T}}$ and $\mathbf{y} = [y_{11}, \dots, y_{1N}, \dots, y_{M1}, \dots, y_{MN}]^{\mathrm{T}}$, where $\mathbf{x}$ is

the $K \times T$ training input matrix and $\mathbf{y}$ is the $K \times 1$ training output vector. By following a standard sparse Bayesian learning framework (Tipping 2001), the training output y_{mn} is assumed to be the population model output $f(\mathbf{x}_{mn})$ with the additive error ε_{mn},

$$y_{mn} = f(\mathbf{x}_{mn}) + \varepsilon_{mn} \tag{3.1}$$

where $f(\cdot)$ is the population feature model and the error term ε_{mn} is the difference between the training output y_{mn} and the population feature model output $f(\mathbf{x}_{mn})$, which is further assumed to be independently normally distributed with zero mean and constant variance σ^2.

A kernel method to infer the population model $f(\mathbf{x}_{mn})$ can be represented as a linear combination of a set of nonlinear basis functions given by

$$f(\mathbf{x}_{mn}) = \boldsymbol{\phi}^{\mathrm{T}}(\mathbf{x}_{mn})\mathbf{w} = \sum_{l=1}^{L} k(\mathbf{x}_{mn}, \mathbf{x}_{cl}) w_l \tag{3.2}$$

where $k(\mathbf{x}_{mn}, \mathbf{x}_{cl})$ is the basis (also called kernel) function centered on a pre-selected input location $\mathbf{x}_{cl}$, w_l is the associated weight and L is the number of the basis functions considered. The basis function vector $\boldsymbol{\phi}$ can be represented by

$$\boldsymbol{\phi}(\mathbf{x}_{mn}) = [k(\mathbf{x}_{mn}, \mathbf{x}_{c1}), k(\mathbf{x}_{mn}, \mathbf{x}_{c2}), \dots, k(\mathbf{x}_{mn}, \mathbf{x}_{cL})]^{\mathrm{T}} \tag{3.3}$$

which is centered respectively on input locations $\mathbf{x}_c = [\mathbf{x}_{c1}, \mathbf{x}_{c2}, \dots, \mathbf{x}_{cL}]^{\mathrm{T}}$ and the associated weight vector $\mathbf{w}$ is represented as $\mathbf{w} = [w_1, w_2 \dots, w_L]^{\mathrm{T}}$. Due to the simplifying assumption on the error term ε_{mn}, the likelihood of the training data

$p(\mathbf{y}|\mathbf{w},\sigma^2)$ can be written as

$$\begin{aligned} \mathrm{P}(\mathbf{y}|\mathbf{w},\sigma^2) &= \mathcal{N}(\mathbf{y}|\boldsymbol{\Phi}_1\mathbf{w},\sigma^2\mathbf{I}_K) \\ &= (2\pi\sigma^2)^{-K/2}\exp(-\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{w}\|_2^2/(2\sigma^2)) \end{aligned} \tag{3.4}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix, $\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{w}\|_2$ is the Euclidean distance between $\mathbf{y}$ and $\boldsymbol{\Phi}_1\mathbf{w}$, and $\boldsymbol{\Phi}_1$ is the $K \times L$ design matrix defined on all training points given by $\boldsymbol{\Phi}_1 = [\boldsymbol{\phi}(\mathbf{x}_{11}), \ldots, \boldsymbol{\phi}(\mathbf{x}_{1N}), \ldots, \boldsymbol{\phi}(\mathbf{x}_{M1}), \ldots, \boldsymbol{\phi}(\mathbf{x}_{MN})]^{\mathrm{T}}$.

The weight vector $\mathbf{w}$ and the error term σ^2 are unknown parameters we would like to estimate. The ordinary least squares estimate of these unknowns from the Euclidean squared distance $\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{w}\|_2^2$ or the equivalent maximum likelihood estimate from $p(\mathbf{y}|\mathbf{w},\sigma^2)$ may pose the overfitting problem as there are in general many weight parameters to be estimated. To avoid this, additional constraint has to be imposed on these unknown parameters to obtain a simpler model, resulting in the so-called regularized least squares or penalized maximum likelihood estimates. In a Bayesian treatment, this is achieved naturally by imposing a set of priors over the parameters $\mathbf{w}$ and σ^2. These priors control the generalization ability of the learning process. In the SSBL framework (Tipping 2001), a hierarchical prior scheme is used to offer good generalization performance and more importantly to deliver a sparse population feature model which depends only on a small number of important basis functions.

3.2.3 Prior selection

A popular choice of the prior over the weight parameter $\mathbf{w}$ is the zero-mean Gaussian prior distribution given by

$$\mathrm{P}(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{l=1}^{L} \mathcal{N}(w_l|0, \alpha_l^{-1}) \tag{3.5}$$

where the weight prior precision $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_L]^{\mathrm{T}}$ is an $L \times 1$ hyperparameter vector which independently controls the strength of the prior over their associated weights.

The zero-mean Gaussian prior itself is a conjugate distribution for the likelihood function and thus the posterior probability of the weight can be obtained in closed form. Moreover, most of the probability masses of the zero-mean Gaussian prior are concentrated around zero such that the posterior distribution of the weight is dominated by the data rather than the prior. If using a non-informative prior, too many of the probabilistic masses are placed on larger values of the weights that are unrealistic. In addition, the posterior distribution of the weight could become mathematically unmanageable. It is worth mentioning that zero-offset in the mean function of a GP prior does not imply that the resulting predictive distribution will also have a zero mean. In general, the mean function of the predictive distribution, which depends on the testing input data and the hyperparamters that are trained using the training data, is non-zero even when the GP prior is given with zero mean. The Gaussian prior itself does not ensure sparsity in the SSBL and it is achieved by placing a hierarchical prior over the

hyperparameters, given as

$$\mathrm{P}(\boldsymbol{\alpha}) = \prod_{l=1}^{L} \Gamma(a)^{-1} b^{a} \alpha_l^{a-1} e^{-b\alpha_l} \tag{3.6}$$

where $\Gamma(\cdot)$ is the gamma function, and a, b are hyperparameters of the hierarchical prior over the weight precision. In the case of the prior over the error variance σ^2, a suitable prior is given as

$$\mathrm{P}(1/\sigma^2) = \Gamma(c)^{-1} d^{c} \sigma^{2(1-\mathrm{c})} e^{-d/\sigma^2} \tag{3.7}$$

where c and d are hyperparameters of the prior over noise variance. To see how sparsity is introduced, one may marginalize out the weight $\mathrm{P}(w_l)$

$$\begin{aligned} \mathrm{P}(w_l) &= \int \mathrm{P}(w_l|\alpha_l)\mathrm{P}(\alpha_l) d\alpha_l \\ &= \frac{b^{a}\Gamma(a+1/2)}{(2\pi)^{1/2}\Gamma(a)} (b + w_l^2/2)^{-(a+1/2)} \end{aligned} \tag{3.8}$$

which corresponds to a Student's t-distribution. Thus, the overall marginal weight prior is a product of independent Student-t distribution over the weight vector $\mathbf{w}$. In the case of the uniform prior, with $a = b = 0$, we obtain the improper prior

$$\mathrm{P}(w_n) \propto 1/|w_n| \tag{3.9}$$

which is sharply peaked at zero, in charge of producing sparse population models. Therefore, the hyperparameters a and b should be set to as close as zero as machine precision allows. The hyperparameters c and d are also required to be fixed to small values to ensure that the unknown parameter posterior is dominated by the training data rather than the predefined prior. The Bayesian inference-based estimation of unknown

parameters is essentially the combination of information from the data and the priors about the unknown parameters. In sparse Bayesian learning, when the hyperparameters are set to as close as zero, the priors placed on the weights and the noise variance become more non-informative (the priors become uniform by setting the hyperparameters to zero) and thus the estimation results tend to be more objective as they are dominated by the data. There might exist the best options for tuning the hyperparameters, but in practical applications we just need to make sure that these hyperparameters are set to small values.

3.2.4 Parameter inference

With the likelihood function Equation (3.6) and the prior Equations (3.8) to (3.10), the posterior distribution over unknown parameters can be obtained using Bayes' theorem as,

$$\mathrm{P}(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}) = \mathrm{P}(\mathbf{y}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)\mathrm{P}(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)/\mathrm{P}(\mathbf{y}) \tag{3.10}$$

where $\mathrm{P}(\mathbf{y})$ is a normalizing constant independent of the unknown parameters, given by

$$\mathrm{P}(\mathbf{y}) = \int \mathrm{P}(\mathbf{y}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)\mathrm{P}(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2) d\mathbf{w} d\boldsymbol{\alpha} d\sigma^2 \tag{3.11}$$

Through the combination of the likelihood and the prior, the posterior distribution in Equation (3.10) captures everything we known about unknown parameters. However, this posterior distribution cannot be obtained analytically as the integral in Equation (3.11) is intractable. In general, one may resort to approximate fully Bayesian solutions such as

Markov chain Monte Carlo (MCMC) and variational inference which typically involve the high computational complexity. Alternatively, the joint posterior distribution over the unknown parameters $\mathrm{P}(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y})$ can be easily decomposed as

$$\mathrm{P}(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}) = \mathrm{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}, \sigma^2)\mathrm{P}(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}) \tag{3.12}$$

Given $(\boldsymbol{\alpha}, \sigma^2)$, the conditional posterior distribution over the weight $\mathrm{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}, \sigma^2)$ is analytically possible and thus SSBL becomes a search for the most plausible value of the hyperparameter posterior over $(\boldsymbol{\alpha}, \sigma^2)$. Using Bayes' theorem, the posterior distribution over the scale parameters $(\boldsymbol{\alpha}, \sigma^2)$,

$$\begin{aligned} \mathrm{P}(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}) &= \frac{\mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)\mathrm{P}(\boldsymbol{\alpha})\mathrm{P}(\sigma^2)}{\mathrm{P}(\mathbf{y})} \\ &\propto \mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)\mathrm{P}(\boldsymbol{\alpha})\mathrm{P}(\sigma^2) \end{aligned} \tag{3.13}$$

In the case of uniform hyperpriors ($a, \mathrm{b}, \mathrm{c}$ and d are set to small values), the priors $\mathrm{P}(\boldsymbol{\alpha})$ and $\mathrm{P}(\sigma^2)$ have almost no effect on the maximization of the product of the log marginal posterior over $(\boldsymbol{\alpha}, \sigma^2)$ and thus we need only maximize the term $\mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)$ with respect to $(\boldsymbol{\alpha}, \sigma^2)$, that is computable given by

$$\begin{aligned} \mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2) &= \int \mathrm{P}(\mathbf{y}|\mathbf{w}, \sigma^2)\, \mathrm{P}(\mathbf{w}|\boldsymbol{\alpha}) d\mathbf{w} \\ &= \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{K}_1) \end{aligned} \tag{3.14}$$

where the mean vector and covariance matrix are given by

$$\boldsymbol{\mu}_1 = \mathbf{0}_K \tag{3.15}$$

$$\mathbf{K}_1 = \sigma^2 \mathbf{I}_K + \boldsymbol{\Phi}_1 \mathbf{A}^{-1} \boldsymbol{\Phi}_1^{\mathrm{T}} \tag{3.16}$$

where $\mathbf{0}_K$ is the $L \times 1$ zero vector and $\mathbf{A}$ is the $L \times L$ diagonal matrix with $\mathbf{A} = \mathrm{diag}(\alpha_1, \alpha_2 \ldots, \alpha_L)$. The most plausible value $(\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}})$ of $(\boldsymbol{\alpha}, \sigma^2)$ is given as

$$(\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) = \mathrm{argmax}_{(\boldsymbol{\alpha},\sigma^2)} \, \mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2) \tag{3.17}$$

In Bayesian statistics, the quantity $\mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)$ is known as the marginal likelihood, and its maximization known as the type-II maximum likelihood method (Tipping 2001). Maximization with respect to $(\boldsymbol{\alpha}, \sigma^2)$ can be done numerically using Expectation-Maximization (EM) algorithm and incremental learning algorithm.

This procedure often leads to a significant number of hyperparameters α_l tending towards infinity, effectively removing or pruning the corresponding basis functions in Equation (3.2). The basic idea is that the basis functions that are not significantly contributing to explaining the training data are automatically removed, resulting in a sparse population model. The basis functions that survive are called relevance vectors and thus sparse Bayesian learning is also called relevance vector machine. Empirically it is often found that the number of relevance vectors is much smaller than the total number of training points.

Note that we can compute analytically the posterior distribution over the weights given by

$$\begin{aligned} \mathrm{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) &= \frac{\mathrm{P}(\mathbf{y}|\mathbf{w},\sigma^2_{\mathrm{MP}})\mathrm{P}(\mathbf{w}|\boldsymbol{\alpha}_{\mathrm{MP}})}{\mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}})} \\ &= \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{K}_2) \end{aligned} \tag{3.18}$$

with the covariance and the mean for the weight vector are given respectively by

$$\mathbf{K}_2 = (\mathbf{A}_{\mathrm{MP}} + \sigma_{\mathrm{MP}}^{-2}\boldsymbol{\Phi}_1^{\mathrm{T}}\boldsymbol{\Phi}_1)^{-1} \tag{3.19}$$

$$\boldsymbol{\mu}_2 = \sigma_{\mathrm{MP}}^{-2}\mathbf{K}_2\boldsymbol{\Phi}_1^{\mathrm{T}}\mathbf{y} \tag{3.20}$$

where $\mathbf{A}_{\mathrm{MP}} = \mathrm{diag}(\alpha_{1,\mathrm{MP}}, \alpha_{2,\mathrm{MP}} \dots, \alpha_{L,\mathrm{MP}})$.

3.2.5 Predictive posterior probability density function

Given a new test point $\mathbf{x}_{\mathrm{p}} = [\mathbf{x}_{\mathrm{p},1}, \dots, \mathbf{x}_{\mathrm{p},S}]^{\mathrm{T}}$, predictions can be made for the population feature model outputs $\mathbf{y}_{\mathrm{p}} = [y_{\mathrm{p},1}, \dots, y_{\mathrm{p},S}]^{\mathrm{T}}$ in terms of the posterior predictive probability density, given by

$$\begin{aligned} \mathrm{P}(\mathbf{y}_{\mathrm{p}}|\mathbf{y}) &\approx \int \mathrm{P}(\mathbf{y}_{\mathrm{p}}|\mathbf{w}, \sigma_{\mathrm{MP}}^2)\mathrm{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}_{\mathrm{MP}}, \sigma_{\mathrm{MP}}^2)d\mathbf{w} \\ &= \mathcal{N}(\boldsymbol{\mu}_{\mathrm{p}}, \mathbf{K}_{\mathrm{p}}) \end{aligned} \tag{3.21}$$

with

$$\boldsymbol{\mu}_{\mathrm{p}} = \boldsymbol{\Phi}_2^{\mathrm{T}}\boldsymbol{\mu}_2 \tag{3.22}$$

$$\mathbf{K}_{\mathrm{p}} = \boldsymbol{\Phi}_2\mathbf{K}_2\boldsymbol{\Phi}_2^{\mathrm{T}} + \sigma_{\mathrm{MP}}^2\mathbf{I}_S \tag{3.23}$$

where $\boldsymbol{\Phi}_2 = [\boldsymbol{\phi}(\mathbf{x}_{\mathrm{p},1}), \dots, \boldsymbol{\phi}(\mathbf{x}_{\mathrm{p},S})]^{\mathrm{T}}$ is the new design matrix for population feature prediction and $\mathbf{I}_S$ denotes an $S \times S$ identity matrix, and $\boldsymbol{\mu}_{\mathrm{p}}$ and $\mathbf{K}_{\mathrm{p}}$ are the posterior population feature mean and covariance, respectively. The predictive covariance in Equation (3.23) is the sum of two terms: the pooled uncertainty $\sigma_{\mathrm{MP}}^2\mathbf{I}_S$ from measurement noise and structural variability, and the modelling uncertainty $\boldsymbol{\Phi}_2\mathbf{K}_2\boldsymbol{\Phi}_2^{\mathrm{T}}$

induced by SSBL for the population features of NISs. It is apparently important to derive an accurate estimate of the signal noise term as it can effectively dictate how much of uncertainty in the prediction is governed by measurement noise, and how much of it is explained by actual modelling uncertainty against the given dictionary. This is effectively a problem of optimizing the hyperparameters.

3.2.6 The choice of kernel functions

The choice of kernel functions is the crucial ingredient in SSBL, as it encodes our assumptions about the population model we wish to learn. An advantage of SSBL lies in that the kernel function $k(\mathbf{x}, \mathbf{x}_c)$ is not required to satisfy Mercer's condition. Commonly used kernel functions include polynomial kernel, Gaussian kernel and Laplace kernel. In this thesis, the Gaussian kernel is preferred due to its good smoothness, given as

$$k(\mathbf{x}, \mathbf{x}_c) = \exp(-\|\mathbf{x} - \mathbf{x}_c\|_2^2 / 2\gamma^2) \tag{3.24}$$

where γ is kernel width.

The update rules for the hyperparameters depend on computing the posterior weight covariance matrix, which requires an inverse operation (in fact, Cholesky decomposition) of order $\mathcal{O}(L^3)$ complexity and $\mathcal{O}(L^3)$ memory storage, with L being the number of basis functions. In standard SBL, the basis function is typically centred on at all training

points, that is $\mathbf{x}_c = \mathbf{x}$. This leads to a computational complexity of $\mathcal{O}(K^3)$, with K being the number of the total training point, which is computationally prohibitively expensive in the case of nominally identical structures. Due to the fact that $\mathbf{x}_{1n} = \mathbf{x}_{2n} = \cdots = \mathbf{x}_{Mn}$ for any n, one may simply let $\mathbf{x}_c = [\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1N}]^{\mathrm{T}}$. As a result, the number of basis functions used in standard SBL is only N and the computational complexity of the standard SBL algorithm will be reduced considerably from order $\mathcal{O}(K^3)$ to $\mathcal{O}(N^3)$.

It should be noted that only the Gaussian kernel is used in the case study throughout the thesis. The optimal number of kernels is automatically determined by the hierarchical sparse Bayesian learning scheme and the optimal parameters associated with kernel functions in sparse Bayesian learning are determined by three model assessment indices, including root mean squared residual, mean standardized log loss and sparsity ratio which will be introduced later.

3.3 Case study

Nominally identical structures or components exist widely in the real world. For example, railway wheels are a typical example of nominally identical structures that are made of the same materials, manufactured in the same specifications and assembled in the same train. Damage in railway wheels is called wheel defects because of wheel out-

of-roundness such as wheel flat, wheel shell and wheel polygonization (Nielsen and Johansson 2000, Johansson 2006). Wheel flat can arise due to unintentional sliding of the wheel on the rail. The primary cause of wheel flat is that the braking force is too high in relation to the available wheel/rail friction. The reason for this might be that the brakes are poorly adjusted, frozen or defective. Another reason might be that there are regions where wheel/rail frication incidentally and locally becomes low. Wheel polygonization is believed to be a coupling to the frequencies of natural vibration of the railway wheelset. Wheel shell can arise due to loss of flakes of material from the wheel tread. Excessive vertical wheel/rail contact forces with respect to the diameter of the wheel is the primary cause for this particular form of rolling contact fatigue. Wheel defects can cause severe damage to both rail and vehicle components such as sleepers, rails, wheelsets and bearings. They also lead to increased impact and rolling noise levels and to discomfort for passengers owing to high vibration amplitudes. Therefore, wheel condition assessment is of great importance to ensure railway safety and to reduce the required cost for the maintenance of railway infrastructure.

Wheel condition assessment, however, suffers from wheel defect diversity, random vehicle dynamics, track irregularity and limitations in the number of measurement sensors and their arrangement. Measurement results can be of high variability and randomness. In addition, these nominally identical structures are not truly identical due to variability

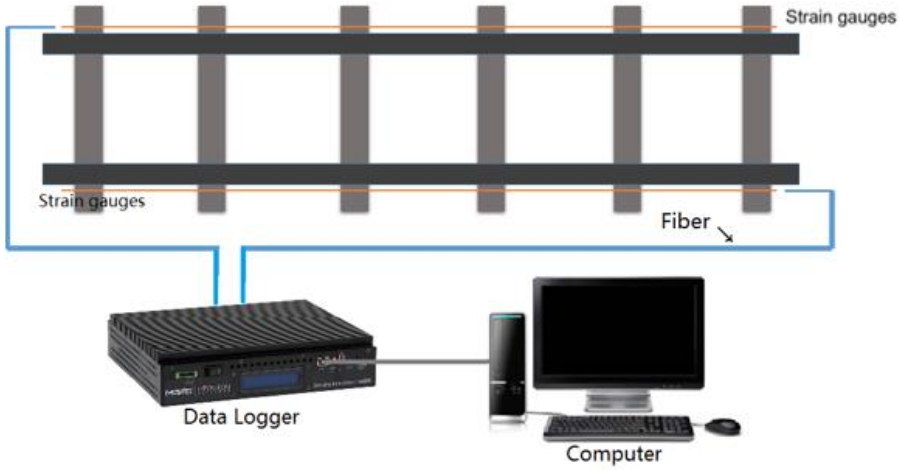
in the materials, manufacturing, assembly, boundary conditions, and so forth, which cause corresponding variability or uncertainty in the dynamics and the vibration response characteristics. Condition assessment of railway wheels within deterministic analysis is not reliable and improper use may lead to false estimation and prediction. Therefore, a Bayesian probabilistic model using SBL is proposed in this chapter to quantify the associated uncertainties in dynamic measurements for normal wheels, which can be used subsequently for wheel condition assessment.

3.3.1 Track-side monitoring system

To assess the condition of wheels, the monitoring data should be first collected. The monitoring data are collected from a track-side monitoring system (Liu and Ni 2018) that is based on fiber Bragg grating (FBG) sensors, mounted on the rail foot. The fiber sensing track-side monitoring system in this case study consists of two arrays of 42 FBG strain gauges, two optical cables, a high-speed optical interrogator, and a desktop computer. Each sensor array comprises 21 FBG gauges, evenly spaced at 0.15m intervals on rail foot of one single track and the total measurement range reaches 3.0m to cover the whole wheel circumference. By using two optical cables, these FBG sensors are connected to a high-speed optical interrogator, which real-timely reads their wavelength information. These wavelengths are then transformed into strain information in accordance with sensor

sensitivities. The high-speed optical interrogator is controlled by the desktop computer which is also used for data acquisition, data analysis and decision making for wheel maintenance scheduling. The optical interrogator and the desktop computer are both operated in an auxiliary office that is about 120m away from the monitoring area. Figure 3.1 shows the configuration of the FBG-based track-side monitoring system and Figure 3.2 gives the deployment of FBG strain gauges. The FBG sensors have been calibrated in laboratory before their installation, indicating a strain sensitivity of $1\mathrm{pm}/\mu\varepsilon$ and a temperature sensitivity of $10\mathrm{pm}/\ ^{o}\mathrm{C}$. In this application, temperature compensation is not necessary because the online monitoring activity for a train can be done in several seconds, during which environment temperature will not change dramatically. For the same reason, the dynamic strain change due to rail temperature evolution is also disregarded. Figure 3.3 shows the rail strain variation monitored by the FBG sensor SEN-D2 when a typical 8-car passenger train passes the instrumented rail section at a nominal speed of 10km/h. The sampling frequency f_s is set to be 5kHz to ensure that abnormal wheels can be sensed. It is observed that rail bending strain varies between -100με and 150με, wherein 32 peaks are consistent with 32 wheels on a single track. Figure 3.4 gives the corresponding Fourier amplitude spectrum (FAS) under a logarithmic scale and it is found that the strain signal is dominated by the low frequency components under 10Hz. Low-frequency components are believed to be dominated by axle loads, wheel bases and

long wave track irregularities, while high-frequency components are believed to be caused by wheel and rail surface roughness as well as measurement noises (Wei et al. 2012, Filograno et al. 2012 and 2013). The loading conditions such as fully loaded and non-loaded trains do not significantly influence the wheel defect-incurred impact on the railway track, which has been validated by both numerical modelling (Uzzal et al. 2008) and field test (Nielsen and Johansson 2000).



(a) Schematic diagram of the FBG-based track-side monitoring system

(b) Deployed FBG strain gauges

Figure 3.1 FBG-based track-side monitoring system

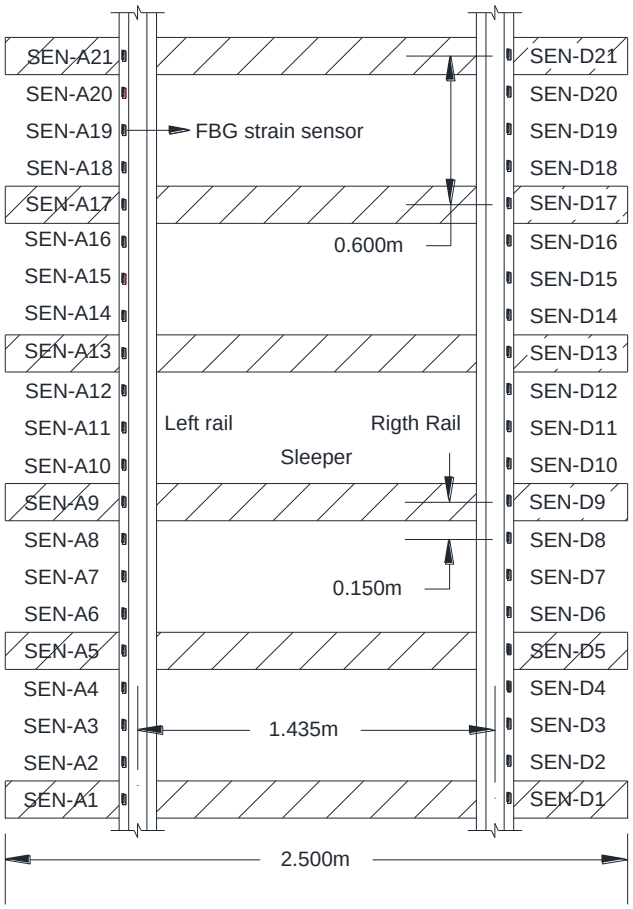


Figure 3.2 Deployment of FBG gauges

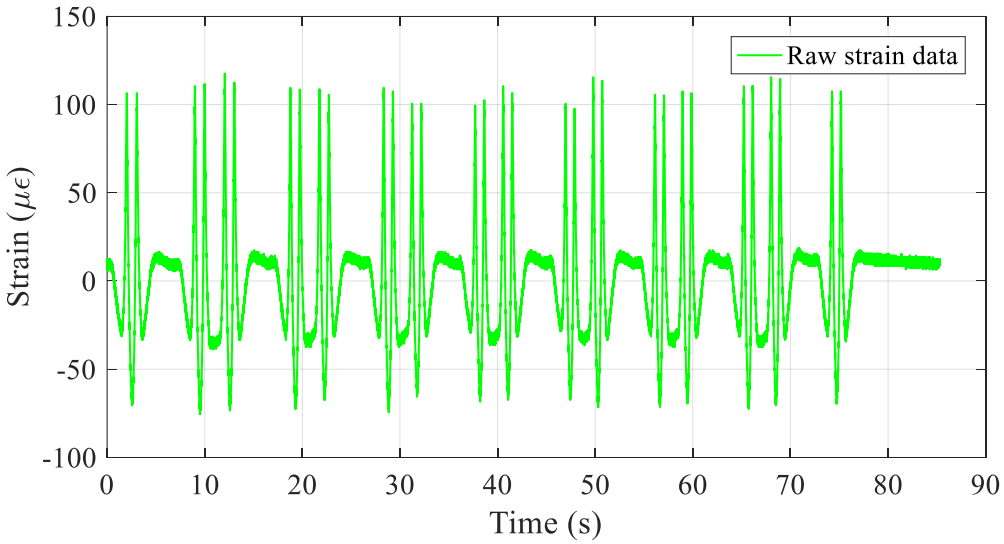


Figure 3.3 Rail foot strain recorded by SEN-D2

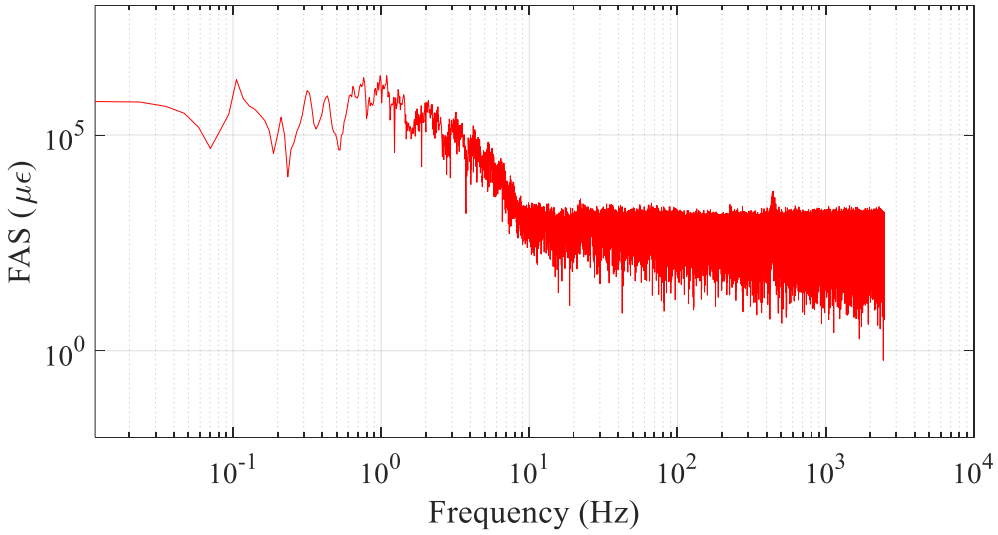


Figure 3.4 FAS of rail foot strain recorded by SEN-D2

3.3.2 Data preprocessing

To obtain useful information for wheel condition assessment, Filograno et al. (2013) have proposed an empirical formula to extract the high frequency components from the monitoring strain data, which is defined as

$$f_c = kv \tag{3.25}$$

where the cutoff frequency f_c is linearly proportional to the nominal train speed v with the coefficient k. The coefficient k is set to be 1.0Hz·h/km after many experiments and the train speed v is fixed to its nominal value. The high frequency components termed detrended data in this study, can be extracted from raw strain monitoring data by using an ideal high-pass filter with the cutoff frequency f_c determined by the train speed. The remaining part is the low frequency components, termed the trend. Figures 3.5 and 3.6 shows the trend and detrended data respectively with a cutoff frequency $f_c = 10\text{Hz}$. As shown in Figure 3.7, 32 data segments near strain peaks are separated from the detrended data and then allocated to each wheel, with each segment lasting 0.3s and consisting of 1500 data points. An upper frequency limit f_u is then utilized to exclude redundant high frequency components. The upper limit frequency f_u is the upper bound of the wheel defect-induced effect on the variation of rail bending strain in the frequency domain. It is related to the train speed and the targeted possible maximum wheel defect. Typically, it is extremely difficult to explicitly determine such a relationship, partially because there

is a lack of samples of the possible maximum wheel defect for field test but mainly because the wheel defect-induced impact load on the rail nonlinearly depends on the train speed (Nielsen and Johansson 2000, Johansson and Nielsen 2003). In this study, the considered train speed is $v = 10$km/h and the targeted maximum wheel defect length $L_{\max} = 150$mm. The upper limit frequency f_u is empirically set to be 300Hz, which guarantees that the chosen frequency-domain response components can deliver the richest information about wheel defect. The FASs of 32 detrended data segments are given in Figure 3.8.

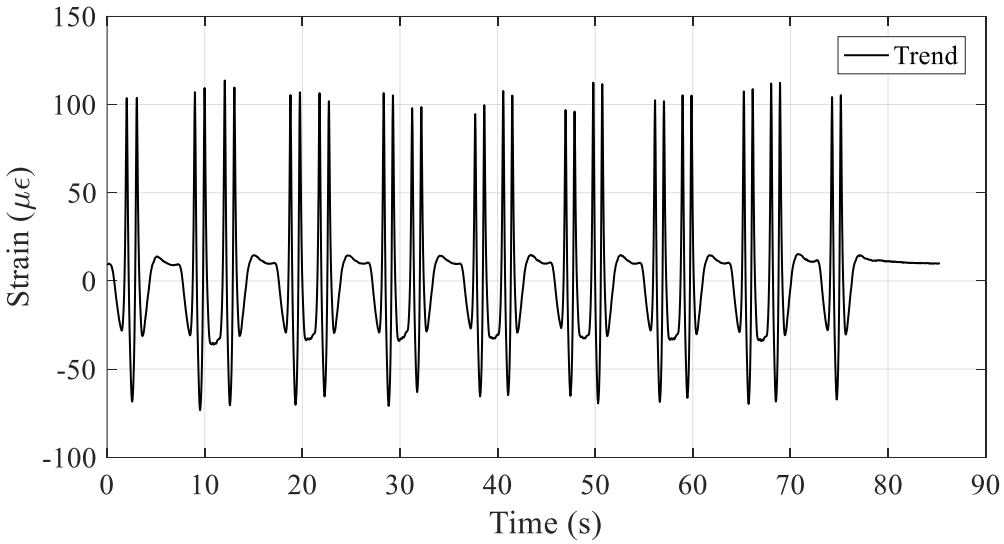


Figure 3.5 Low-frequency component of rail foot strain recorded by SEN-D2

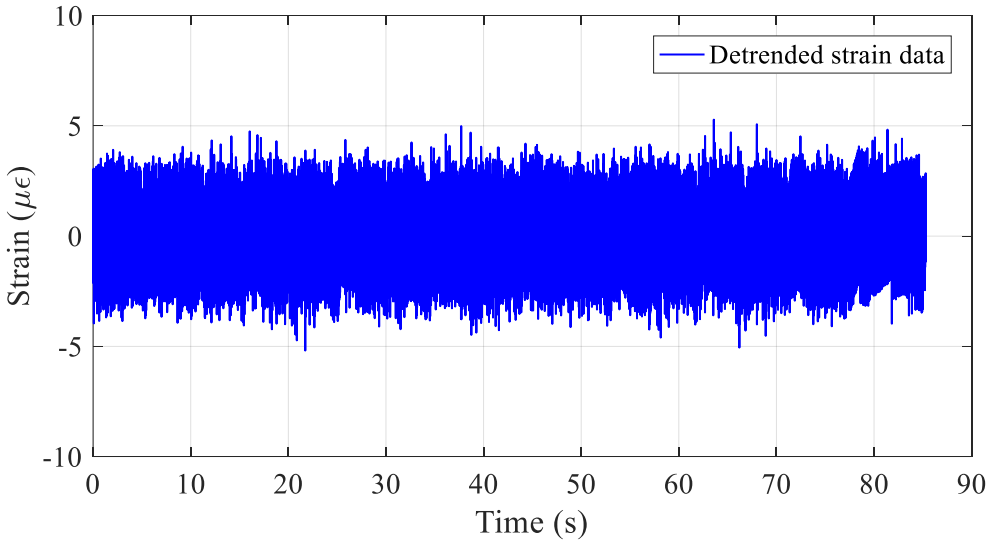


Figure 3.6 High-frequency component of rail foot strain recorded by SEN-D2

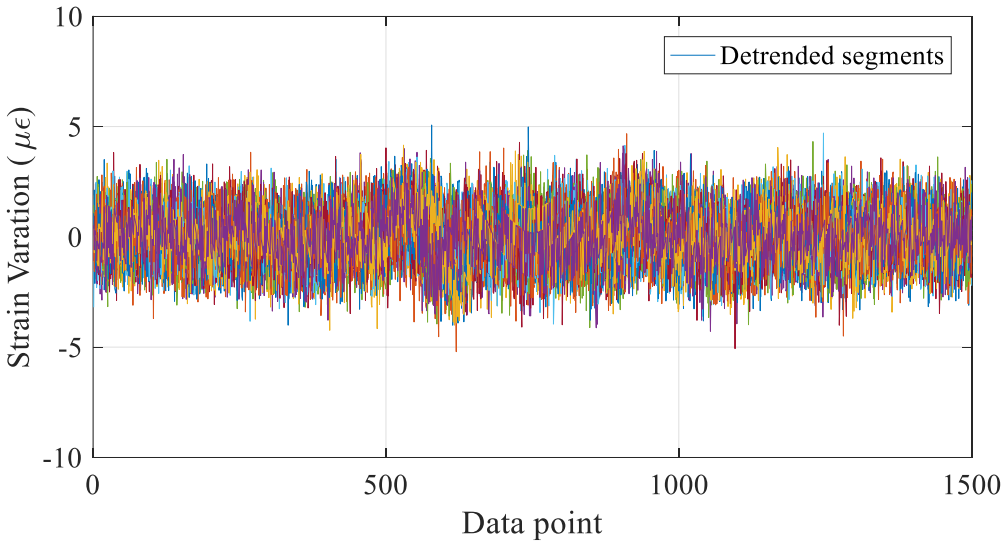


Figure 3.7 High-frequency components for 32 wheels recorded by SEN-D2

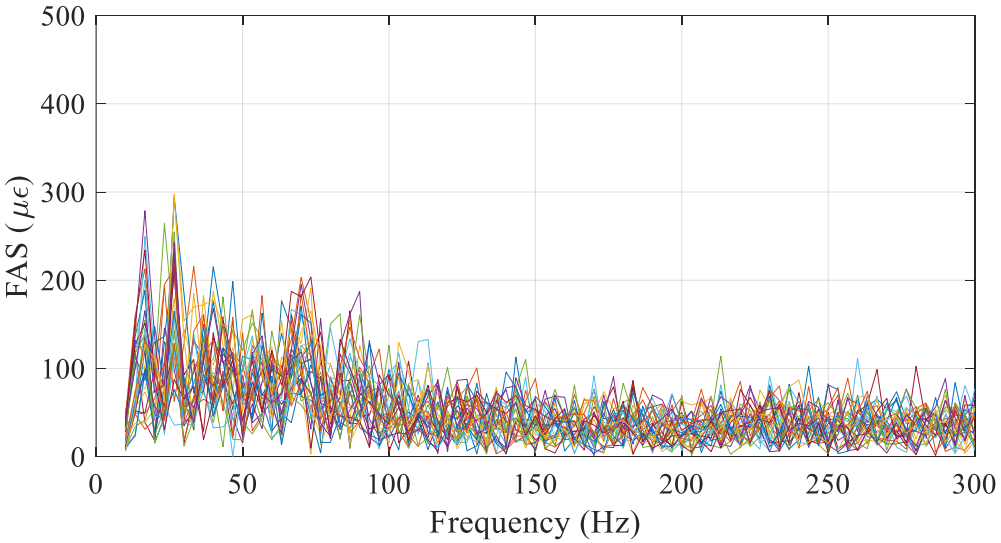


Figure 3.8 FASs of 32 detrended data segments

3.3.3 Feature extraction

Fourier spectrum analysis is a common practice for feature extraction that is an essential step for wheel condition assessment. In this example, Fourier amplitude spectrum $d(f)$ of high frequency components of raw strain monitoring data in the frequency band $[f_c, f_u]$ is first normalized by its integration d_I that is defined as

$$d_I = \int_{f_c}^{f_u} d(f)df \tag{3.26}$$

When a discrete Fourier transform (DFT) is employed, d_I can be approximated by using the trapezoidal rule as

$$d_I \approx \sum_{i=1}^{N-1} (d_i + d_{i+1})\Delta f/2 \tag{3.27}$$

where Δf is the frequency resolution of DFT and N is the number of frequency bins in $[f_c, f_u]$ determined by

$$N = round\{(f_u + f_c)/\Delta f\} \tag{3.28}$$

The round is a mathematical operation for taking integer. As a result, the normalized Fourier amplitude spectrum (NFAS) can be stated in a probabilistic logic and its CDF on discrete frequency bins can be derived as

$$y_i = \begin{cases} 0, & i = 1 \\ \sum_{i=1}^{N-1} (d_i + d_{i+1})\Delta f/2\,, & i \neq 1 \end{cases} \tag{3.29}$$

The values of all CDFs range from 0 to 1 in the frequency band $[f_c, f_u]$. The CDFs for 32 NFASs of detrended segments are given in Figure 3.9 with $\Delta f = 3.3$Hz and $N = 88$.

In the next stage, these CDFs will be utilized for training a probabilistic population model to define healthy wheels by means of standard SBL and this population model will be used as a baseline reference model in the phase of wheel condition assessment when new monitoring data from railway wheels are available.

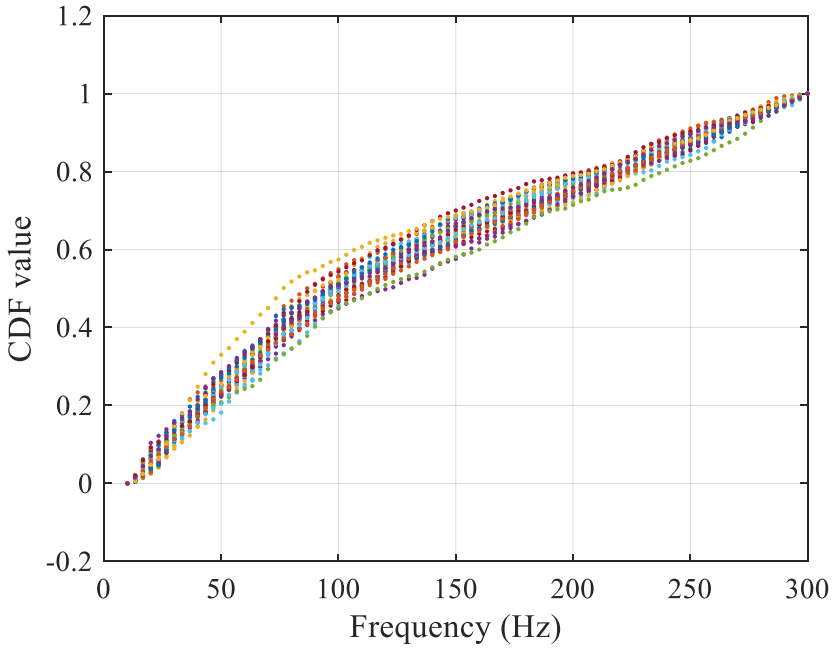


Figure 3.9 CDFs of FASs in the frequency range of 10-300Hz for 32 wheels recorded by SEN-D2

3.3.4 SSBL for population features of nominally identical wheels

In this section, raw monitoring data from the sensor SEN-D2 are taken as an example to illustrate the SSBL method for modelling the population features of nominally identical wheels. As shown in Figure 3.3, raw monitoring data from the sensor SEN-D2 contain condition assessment information of 32 wheels, in which high-frequency components sensitive to wheel condition assessment have been extracted and assigned to each wheel, as shown in Figure 3.4. Typically, the measured strain variation of rail bending can be classified into two parts (Wei et al. 2012; Filograno et al. 2012 and 2013): the low-frequency component and the high-frequency component, by using the cutoff frequency f_c in Eq. (3.25) which is dependent on the train speed. The low-frequency component is found to be primarily controlled by axle loads, wheel bases and thus it is often used for the dynamic weighing, axle counting and identification of trains. By contrast, the high-frequency component is mainly caused by wheel and rail surface roughness and measurement noises and thus it is often used for wheel condition assessment. The low-frequency component (including the DC component at 0 Hz in Figure 3.4) does not affect the accuracy of sparse Bayesian learning results because the model training uses only damage-sensitive features extracted from the high-frequency component. Then, 32 CDFs that characterize wheel conditions have been derived, with 26 CDFs ($M = 26$) of healthy wheels used for learning a probabilistic population feature model of all healthy wheels

and the remaining 6 CDFs ($Q = 6$) used for testing the performance of the learned population feature model. The total training set is composed of 2288 ($K = 2288$) samples points evenly distributed on 88 ($N = 88$) frequency bins. The Gaussian kernel function is employed here because of its good adaptability discussed before. In the framework of standard SBL (Tipping 2000 and 2003), the kernel width needs to be predefined. Although adaptive algorithms (Wipf and Bhaskar 2004, Mohsenzadeh and Sheikhzadeh 2014) have been proposed for automatic selection of kernel width, they can be computationally expensive especially when the training set is large. In this case study, a random initialization strategy is utilized to explore the optimal kernel width γ and three indices are employed to investigate the performance of the learned population feature model when different kernel widths are used. The first index is the root mean square residual (RMSR) given by

$$\mathrm{RMSR} = \sqrt{\sum_{q=1}^{Q}\sum_{n=1}^{N}\frac{(y_{qn} - \mu_{\mathrm{p},n})^2}{QN}} \tag{3.30}$$

where N is the number of frequency bins for each CDF, Q is the number of the remaining CDFs used for testing the performance of the learned population feature model, y_{qn} is the observed value of the qth CDF on the nth frequency bin and $\mu_{\mathrm{p},n}$ is its expectation value predicted by the learned population feature model. The second index is the mean standardized log loss (MSLL), given by

$$\begin{aligned}\mathrm{MSLL} &= -\frac{1}{QN}\log \mathrm{P}(\mathbf{y}) \\ &= \frac{1}{2QN}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left[\log(2\pi\mathbf{K}_{\mathrm{p},nn}) + \mathbf{K}_{\mathrm{p},nn}^{-1}(y_{qn}-\mu_{\mathrm{p},n})^2\right]\end{aligned} \tag{3.31}$$

where $\mathbf{y}$ denotes the testing outputs from the remaining sample structures and $\mathbf{K}_{\mathrm{p},nn}$ is the nth diagonal element of the posterior covariance matrix of the population features $\mathbf{K}_{\mathrm{p}}$. The third index is the sparsity ratio $\mathcal{K}$, defined as

$$\mathcal{K} = \frac{N_{RV}}{MN}\times 100\% \tag{3.32}$$

where N_{RV} is the number of non-zero weights in the learned population feature model. In this case study, the kernel widths $\gamma = 1$ and 100 are selected to be two extreme cases by tentative calculations.

Then the SSBL-based population feature model is trained by successively increasing the kernel width γ from 1 to 100. Figures 3.10 to 3.13 show four SSBL population feature models in conjunction with their predicted expectations and the associated 95% confidence intervals when the kernel width is γ is equal to 1, 26, 41 and 100, respectively. Figures 3.14 and 3.15 show the variation of RMSR, MSLL and the sparse ratio $\mathcal{K}$ against kernel width γ. From Figure 3.10, it is seen that the trained model is relatively complicated when the kernel width is too small and as a result, 79 relevance vectors (basis functions) are needed to represent the CDFs. Figures 3.11 and 3.12 provide two better alternative models, which are simpler and sparser with only 6 and 4 relevance vectors

used respectively. When the kernel width continues to increase, the probabilistic population feature model of nominally identically undamaged wheels will lose expressive ability as evidenced in Figure 13, where only 3 relevance vectors remain in the model but they are insufficient to characterize the CDFs. As shown in Figures 3.14 and 3.15, both RMSR and MSSL increase in general with the kernel width γ. Within the kernel width range considered, there exist two different local optimal kernel widths, which are $\gamma = 26$ and 41 for RMSR, and $\gamma = 25$ and 40 for MSLL, in relation to different explanations about the complexity of the population feature model. By contrast, the sparsity ratio $\mathcal{K}$ almost deceases continuously with the kernel width γ as shown in Figure 3.16. As a compromise between expressive ability and sparseness, the kernel width γ=40 is considered here as the optimal value to construct the population feature model that is fairly simple and favorably consistent with the likelihood of the training data. There exist only 4 relevance vectors (basis functions) in the probabilistic model when γ=40, justifying a sparse representation of the built model. Table 3.1 show the 4 relevance vectors including the distribution of active (nonzero) weights and the associated kernel functions. The kernel functions defined in Eq. (3.24) are isotropic and thus they have the same variance. They are centered at the input locations and in this study, the input locations are the 88 frequency bins, evenly distributed on the band $x_f = [10\text{Hz}, 300\text{Hz}]$, that has been scaled to integers among $x = [1, 88]$ by using a linear transformation $x =$

$3x_f/10 - 2$. Therefore, these kernel functions have integer mean values.

Figure 3.17 shows the quantification of multiple sources of uncertainty in the obtained population feature model. It is seen that the posterior uncertainty of the population feature model of nominally identically undamaged wheels is dominated by the mixed uncertainty (the pooling of measurement noise-induced uncertainty and structural variability-induced uncertainty). The modelling-induced uncertainty in the SSBL of the population features is insignificant as there are as many as 2288 training points.

It should be mentioned that the group of nominally identical structures to be analyzed, in the case of wheels are assumed to operate in the same operational and environmental condition. This is because the track-side monitoring system is often installed at a specific location before the trains arriving at the railway station, for the convenience of facilitating its use, management and maintenance. Therefore, this allows the trains, including not only normal-speed trains but also high-speed trains to pass the monitoring area at a fixed speed (the same operational and environmental condition).

Table 3.1 Active weights and associated kernel functions

Weight number	Weight distribution	Associated kernel function
w_1	$\mathcal{N}(-5.35, 0.05^2)$	$\exp\left(-\frac{\|x-1\|^2}{2\times 40^2}\right)$
w_{209}	$\mathcal{N}(5.65, 0.048^2)$	$\exp\left(-\frac{\|x-9\|^2}{2\times 40^2}\right)$
w_{1769}	$\mathcal{N}(-1.73, 0.020^2)$	$\exp\left(-\frac{\|x-69\|^2}{2\times 40^2}\right)$
w_{2263}	$\mathcal{N}(2.24, 0.016^2)$	$\exp\left(-\frac{\|x-88\|^2}{2\times 40^2}\right)$

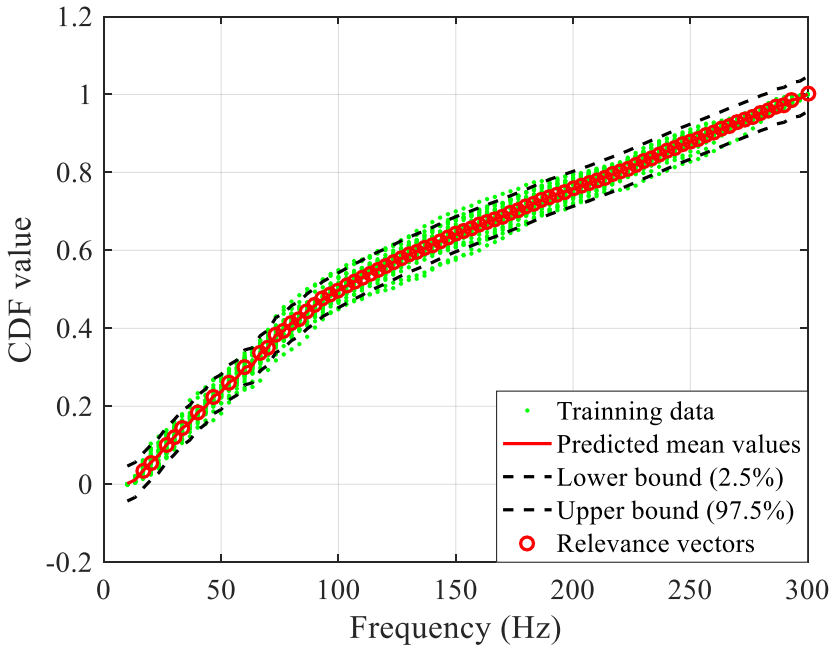


Figure 3.10 CDF learning with kernel $\gamma = 1$

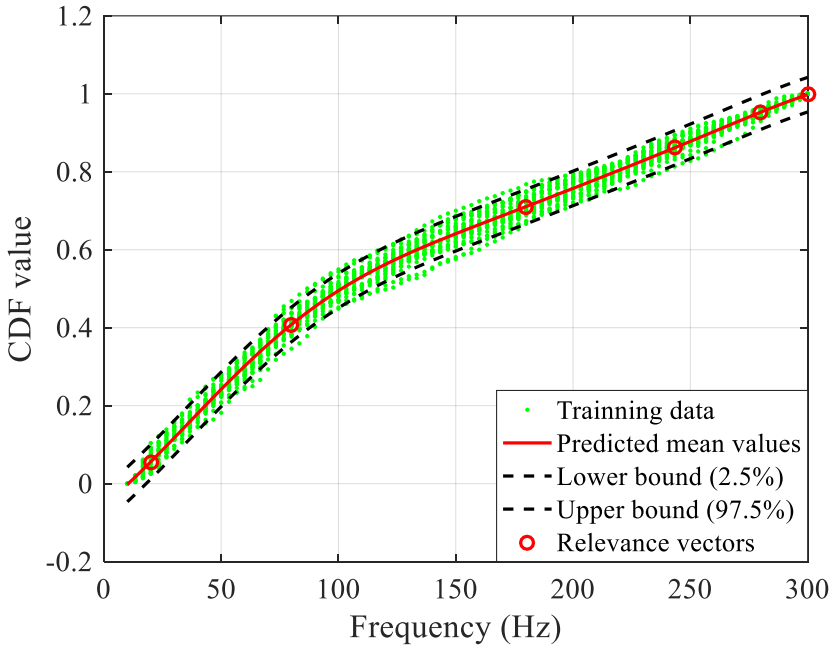


Figure 3.11 CDF learning with kernel $\gamma = 26$

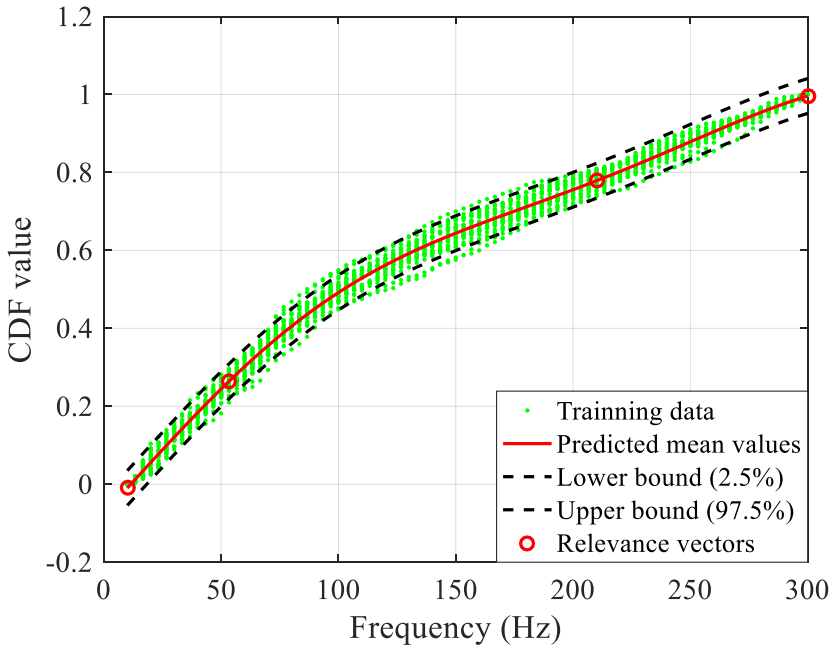


Figure 3.12 CDF learning with kernel $\gamma = 41$

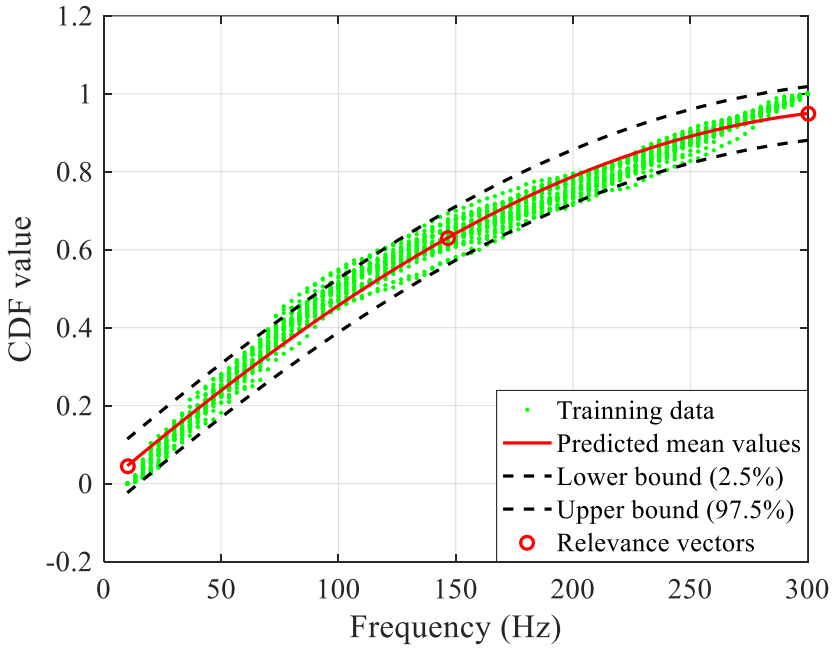


Figure 3.13 CDF learning with kernel $\gamma = 100$

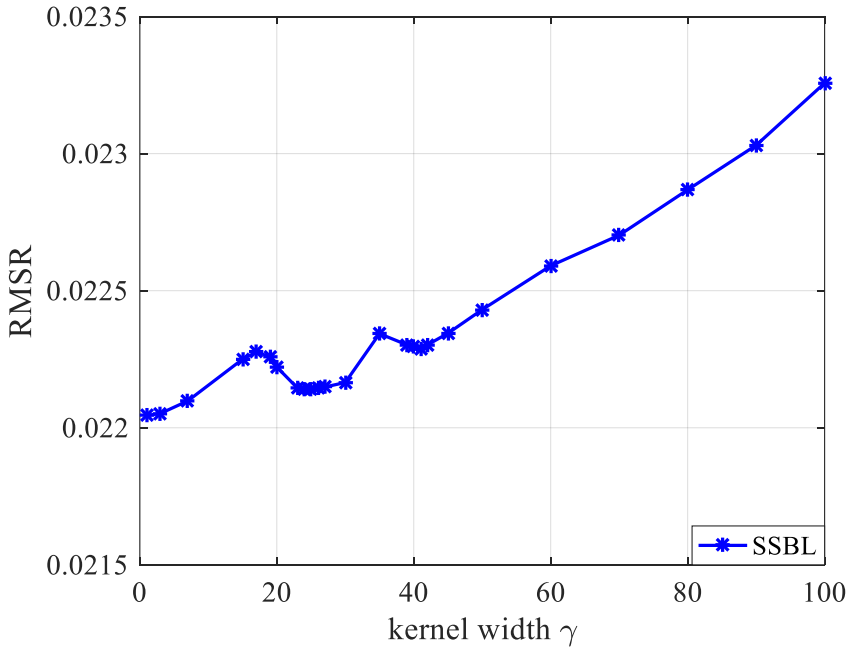


Figure 3.14 RMSR against kernel width γ

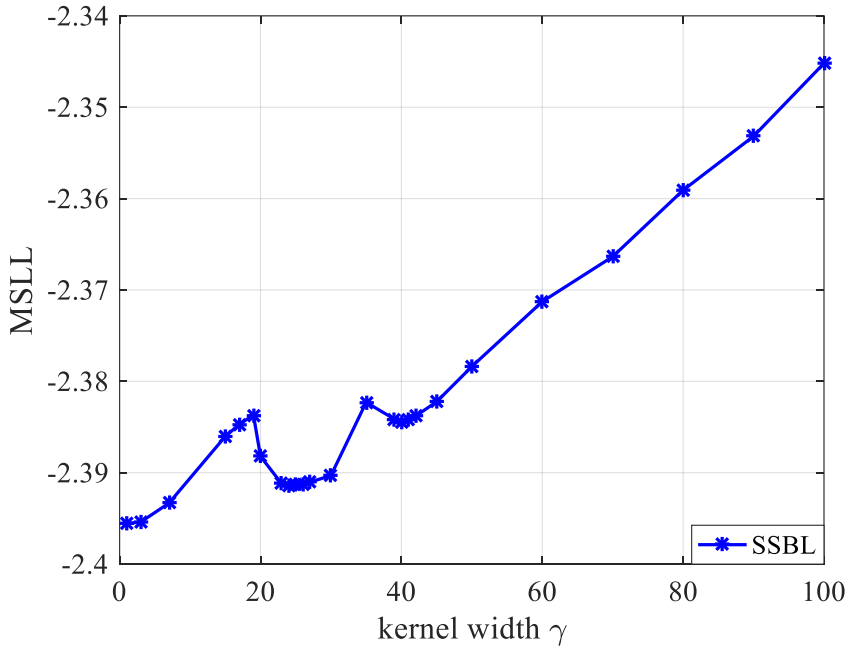


Figure 3.15 MSLL against kernel width γ

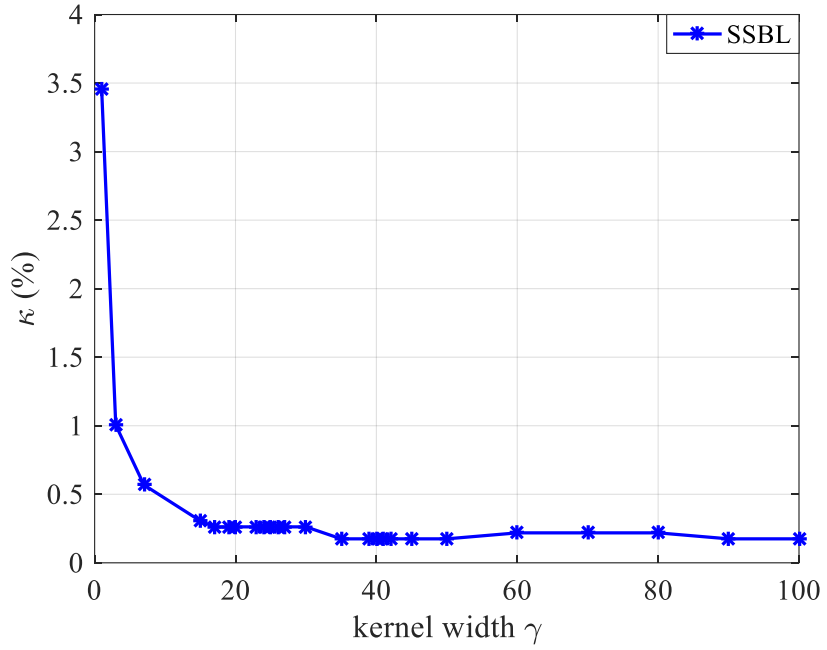


Figure 3.16 Sparsity ratio $\mathcal{K}$ against kernel width γ

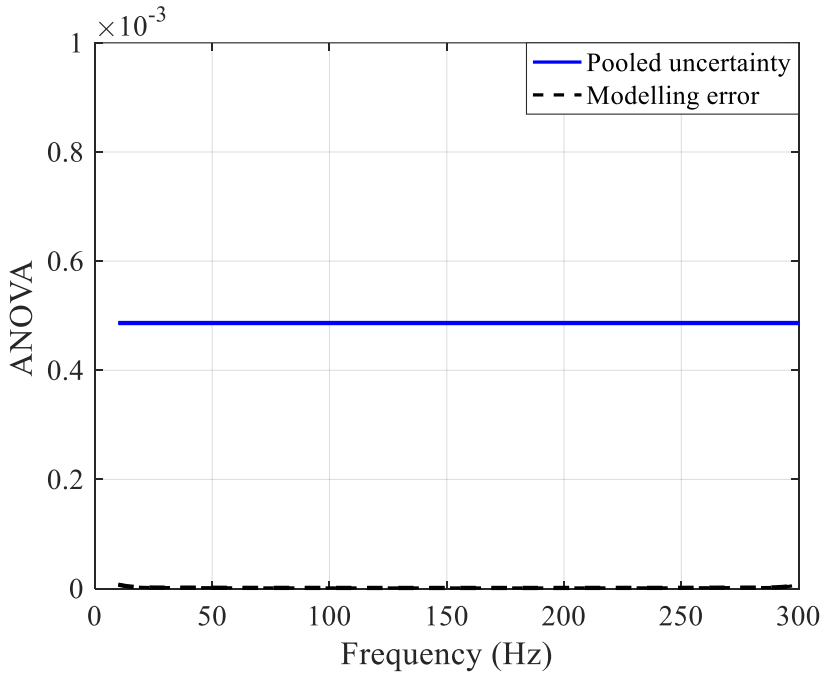


Figure 3.17 ANOVA of multiple sources of uncertainty in the homoscedastic population feature model

3.4 Summary

This chapter presents a Bayesian probabilistic approach for modelling the population features in nominally identical structures. In this approach, the uncertainties in SHM data resulting from measurement noise and inter-structure variability are simply pooled together such that SSBL can be employed to model the population features of nominally identical structures and quantify the associated uncertainties. The advantage to this approach lies in the fact that SSBL is formulated in a fully probabilistic framework which easily takes into account uncertainties in damage-sensitive features for nominally identical structures. A further advantage of this approach is that redundant basis functions utilized in sparse Bayesian modelling of the population damage-sensitive features can be automatically pruned, giving rise to a sparse population model that enables to maintain good prediction accuracy not only on training data but also on unseen data. More importantly, the resulting data-driven population feature model using SSBL enables fast prediction such that the subsequent damage detection procedure can be performed on nominally identical structures in a more efficient and effective way. To validate the performance of SSBL in modelling the population features of nominally identical structure, a case study has been conducted using real-world SHM data.

In addition, we have elaborated on how to extract damage-sensitive features from real-world structural health monitoring data. Based on the extracted damage-sensitive

features, we have also demonstrated the capacity of the standard sparse Bayesian learning on establishing a data-driven population feature model for nominally identical healthy structures. This lays a solid foundation for further studies conducted in the next two chapters for establishing better population feature models by using the variants of standard sparse Bayesian learning.

To validate the performance of SSBL in modelling the population features of nominally identical structures, a case study has been conducted using real-world SHM data. As the case study shows, the pooling of different sources of uncertainty can cause heteroscedasticity in the training data and the homoscedastic baseline model learned by SSBL cannot correctly quantify the uncertainty associated with the population features of NISs. The heteroscedastic training data in the modelling of the population features of NISs will be considered in the next two chapters.

Chapter 4

HSBL for Population Features of NISs

4.1 Introduction

A data-driven statistical framework has been pursued in the previous chapter to model the population features of nominally identical structures (NISs). In this framework, SHM data acquired from multiple NISs are simply pooled together and assumed to be homoscedastic such that standard sparse Bayesian learning (SSBL) can be implemented. In many cases, however, the pooled SHM data are often found to be heteroscedastic due to the existence of multiple sources of uncertainty, including the epistemic uncertainty resulting from the measurement noise and the aleatoric uncertainty from structural variability. The existence of heteroscedasticity can pose a big challenge when modelling the population features of NISs as it can make a statistical model inefficient to illustrate the true uncertainty of it. More importantly, it may invalidate statistical tests when an inefficient model is applied for damage detection.

Currently, there exist a number of statistical modelling techniques available to deal with the heteroscedastic data and most of them are realized in heteroscedastic Gaussian process (HGP) regression. In the typical HGP regression, two standard Gaussian process (GP) models are often employed, with the one for modelling the function mean value and

the other for estimating the input-dependent error level. However, the combination of two GPs can give rise to a joint posterior distribution of the function value and the input-dependent error, that is non-Gaussian and no longer analytically intractable. As a result, various approximate inference methods are employed which include sampling approaches (Goldberg et al. 1998, Wang 2014) and analytical approximations (Le et al. 2005, Kersting et al. 2007, Lázaro-Gredilla and Titsias 2011, McHutchon and Rasmussen 2011, McHutchon 2014, Muñoz-González et al. 2014 and 2016, McHutchon 2015, Urban et al. 2015). Nevertheless, it will often be computationally too expensive to train and make predictions using an HGP.

In this chapter, we borrow the idea of Kersting et al. (2007) and make an extension to SSBL, termed heteroscedastic sparse Bayesian learning (HSBL) to address the heteroscedastic data when training a data-driven population feature model for NISs. Two SSBL models will be utilized in an HSBL, with one for estimating the posterior mean of population features of NISs and the other for updating the posterior uncertainty of them. We first introduce the basic theory of the HSBL in Section 4.2, followed by a comparative case study conducted in Section 4.3 to demonstrate the superiority of HSBL over SSBL in modelling the pooled data showing heteroscedasticity. Concluding remarks are made in Section 4.4.

4.2 Problem Formulation

As described in the previous chapter, assume that we have a group of nominally identical structures being monitored, where M structures are known undamaged. From any undamaged structure, we have a training dataset $\mathbf{D}_m = \{(\mathbf{x}_{mn}, y_{mn})\}_{n=1}^{n=N}$, where y_{mn} is the scalar-valued training output, $\mathbf{x}_{mn}$ is the vector-valued training input, m is the structure serial number, and n is the serial number in the dataset of size N. In the case of M undamaged structures, we have an assembled dataset $\mathbf{D} = \{\mathbf{D}_m\}_{m=1}^{m=M} = \{(\mathbf{x}_{mn}, y_{mn})\}_{m=1,n=1}^{m=M,n=N}$ of size $K = M \times N$ when developing a data-driven statistical model for the interpretation of healthy population features of all NISs.

4.2.1 Model specification

Following the pooled modelling framework in SSBL, we assume that the training output y_{mn} can be represented as

$$y_{mn} = f_{\mathrm{p}}(\mathbf{x}_{mn}) + \varepsilon_{mn} \tag{4.1}$$

with

$$\varepsilon_{mn} \sim \mathcal{N}(0, g^2(\mathbf{x}_{mn})) \tag{4.2}$$

where the error term ε_{mn} is the discrepancy of the training output y_{mn} from the population feature model output $f_{\mathrm{p}}(\mathbf{x}_{mn})$ at the nth input point $\mathbf{x}_{mn}$ from the mth structure. The error ε_{mn} is assumed as a random variable that is independent and

normally distributed with zero mean as in SSBL, whereas its standard deviation is varying across the input in HSBL, which is described by a slowly-varying function $g(\mathbf{x}_{mn})$. The slowly-varying function $g(\mathbf{x}_{mn})$ is referred to the fact that it should be simpler than the population feature model $f_{\mathrm{p}}(\mathbf{x}_{mn})$. Otherwise, we cannot determine the error function $g(\mathbf{x}_{mn})$ using the idea of Kersting et al. (2007). Thus, two unknown functions are required to infer from the training data, including the population feature model $f_{\mathrm{p}}(\mathbf{x}_{mn})$ and the non-negative error level function $g(\mathbf{x}_{mn})$.

4.2.2 Hierarchical Bayesian learning

The proposed HSBL makes use of two SSBL models, with the first for estimating the mean population features of NISs and the other for adjusting the input-dependent error level. Assume that the heteroscedastic population feature model of NISs is also a generalization of the Gaussian probability distribution and the joint distribution of its population feature predictions $\mathbf{y}_{\mathrm{p},1} = [y_{\mathrm{p},1}, \dots, y_{\mathrm{p},S}]^{\mathrm{T}}$ on the test points $\mathbf{x}_{\mathrm{p}} = [\mathbf{x}_{\mathrm{p},1}, \dots, \mathbf{x}_{\mathrm{p},S}]^{\mathrm{T}}$ with S being the number of the predictions, is given by

$$\mathrm{P}(\mathbf{y}_*|\mathbf{D}) = \mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{HSBL}}, \mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}}\right) \tag{4.3}$$

where the posterior population feature mean $\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{HSBL}}$ and covariance $\mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}}$ are both required to be estimated. In the proposed hierarchical sparse Bayesian learning scheme, the posterior population feature mean $\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{HSBL}}$ is approximated by the corresponding

quantity in the homoscedastic population feature model, given by

$$\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{HSBL}} \approx \boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{SSBL}} = \boldsymbol{\Phi}_2^{\mathrm{T}} \boldsymbol{\mu}_2 \tag{4.4}$$

where $\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{SSBL}}$ is the posterior population feature mean from the first SSBL given in Equation (3.22), $\boldsymbol{\mu}_2$ is the marginal posterior mean of the weight parameters in the first SSBL defined in Equation (3.20), and $\boldsymbol{\Phi}_2$ is the new design matrix for population feature prediction defined in the preceding chapter. The population feature covariance $\mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}}$, however, is generally inconsistent with the one from the first SSBL

$$\mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}} \neq \mathbf{K}_{\mathrm{p}}^{\mathrm{SSBL}} = \boldsymbol{\Phi}_2 \mathbf{K}_2 \boldsymbol{\Phi}_2^{\mathrm{T}} + \sigma_{\mathrm{MP}}^2 \mathbf{I}_S \tag{4.5}$$

where $\mathbf{K}_2$ is the marginal posterior covariance of the weight parameters in the first SSBL defined in Equation (3.19), σ_{MP}^2 is the most plausible estimate of the error variance in the first SSBL, and $\mathbf{I}_S$ is the $S \times S$ identity matrix. This is mainly due to the fact that $\sigma_{\mathrm{MP}}^2 \neq g^2(\mathbf{x}_{\mathrm{p},s})$. Therefore, the core problem in HSBL is to estimate the changing error level.

In classical regression analysis, the regression residual is commonly used for the evaluation of the constant noise level σ^2 (Draper and Smith 2014). In light of this, we make use of the regression residual to infer the changing error level $g(\mathbf{x}_{mn})$. In the Bayesian paradigm, the residual is referred as the deviation of the training output from the corresponding estimated mean (Carlin and Louis 2010, Carlin et al. 2013). When developing a data-driven statistical model for representing the population features of NISs

based on SSBL, the residual is given as

$$r_{mn} = y_{mn} - \mu_{\mathrm{p},n}^{\mathrm{SSBL}} \tag{4.6}$$

where $\mu_{\mathrm{p},n}^{\mathrm{SSBL}}$ is the posterior predictive population feature mean at the training input $\mathbf{x}_{mn}$ in the first SSBL. By rewriting the residual r_{mn} as

$$\begin{aligned} r_{mn} &= y_{mn} - f_{\mathrm{p}}(\mathbf{x}_{mn}) + f_{\mathrm{p}}(\mathbf{x}_{mn}) - \mu_{\mathrm{p},n}^{\mathrm{SSBL}} \\ &= \varepsilon_{mn} + \left[f_{\mathrm{p}}(\mathbf{x}_{mn}) - \mu_{\mathrm{p},n}^{\mathrm{SSBL}}\right] \end{aligned} \tag{4.7}$$

we can see that the residual r_{mn} is composed of two items: the first term ε_{mn} is the error term and the other $f_{\mathrm{p}}(\mathbf{x}_{mn}) - \mu_{\mathrm{p},n}^{\mathrm{SSBL}}$ is the modelling error. The error ε_{mn} is a random variable, while the modelling error $f_{\mathrm{p}}(\mathbf{x}_{mn}) - \mu_{\mathrm{p},n}^{\mathrm{SSBL}}$ is a deterministic variable. If the first SSBL model is well established, we may have

$$f_{\mathrm{p}}(\mathbf{x}_{mn}) - \mu_{\mathrm{p},n}^{\mathrm{SSBL}} \approx 0 \tag{4.8}$$

The residual r_{mn} is thus approximately a zero-mean Gaussian random variable, given as

$$r_{mn} \approx \varepsilon_{mn} \sim \mathcal{N}(0, g^2(\mathbf{x}_{mn})) \tag{4.9}$$

The absolute residual $|r_{mn}|$ at the input point $\mathbf{x}_{mn}$ follows a half-normal distribution and its mean is given by (Leone et al. 1961)

$$\mathbb{E}\{|r_{mn}|\} = g(\mathbf{x}_{mn})/\sqrt{\pi/2} \tag{4.10}$$

Thus, we may estimate the standard deviation $g(\mathbf{x}_{mn})$ of the input-dependent error at the training input location $\mathbf{x}_{mn}$ as

$$g(\mathbf{x}_{mn}) = \sqrt{\pi/2}\, \mathbb{E}\{|r_{mn}|\} \tag{4.11}$$

Obviously, $\sqrt{\pi/2}\,\mathbb{E}\{|r_{mn}|\}$ is an approximate unbiased estimate of the input-dependent error level $g(\mathbf{x}_{mn})$.

Therefore, a new training data $\mathbf{D}' = \{\mathbf{x}_{mn}, z_{mn}\}_{m=1,n=1}^{m=M,n=N}$ with $z_{mn} = |r_{mn}|$ can be, therefore, created to train a second SSBL model to estimate the most likely noise level $\tilde{g}(\mathbf{x}_{\mathrm{p},s}) = \sqrt{\pi/2}\,\mu_{\mathrm{z}_{\mathrm{p},s}}$ at the population feature test point $\mathbf{x}_{\mathrm{p},s}$. It should be mentioned that the most likely noise level is required to refine $\tilde{g}(\mathbf{x}_{\mathrm{p},s}) = \max\left(0, \sqrt{\pi/2}\,\mu_{\mathrm{z}_{\mathrm{p},s}}\right)$ to ensure a nonnegative noise level. The heteroscedastic sparse Bayesian learning is presented in Algorithm 1.

The uncertainty associated with the population features of NISs is thus updated to

$$\mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}} = \mathbf{\Phi}_2 \mathbf{K}_2 \mathbf{\Phi}_2^{\mathrm{T}} + \mathbf{G}_{\mathrm{p}} \tag{4.12}$$

where

$$\mathbf{G}_{\mathrm{p}} = \mathrm{diag}(\tilde{g}^2(\mathbf{x}_{\mathrm{p},1}), \ldots, \tilde{g}^2(\mathbf{x}_{\mathrm{p},S})) \tag{4.13}$$

with $\tilde{g}^2(\mathbf{x}_{\mathrm{p},s}) = \max\left(0, \pi\mu_{\mathrm{z}_{\mathrm{p},s}}^2/2\right)$. It should be mentioned that one may also update the posterior mean of the population features of NISs in Equation (4.3) using the updated noise level, whereas it is not necessary to improve the population model due to the fact that the first SSBL for estimating population mean is more reliable than the other for estimating the input-dependent error function.

Algorithm 1 Heteroscedastic Sparse Bayesian Learning (HSBL)

1. Train the first SSBL model on the training dataset $\mathbf{D} = \{\mathbf{x}_{mn}, y_{mn}\}_{m=1,n=1}^{m=M,n=N}$ and estimate the posterior distribution of the population features over training locations $y_{mn}|\mathbf{D} \sim \mathcal{N}\left(\mu_{y_{mn}}, \sigma_{y_{mn}}^2\right)$;
2. Calculate the residuals $r_{mn} = y_{mn} - \mu_{y_{mn}}$ and build a new training dataset $\mathbf{D}' = \{\mathbf{x}_{mn}, z_{mn}\}_{m=1,n=1}^{m=M,n=N}$ with $z_{mn} = |r_{mn}|$;
3. Train the second SSBL model on the new training dataset $\mathbf{D}'$ and estimate the input-dependent noise standard deviation $\tilde{g}\left(\mathbf{x}_{\mathrm{p},s}\right) = \sqrt{\pi/2}\,\mu_{\mathrm{z}_{\mathrm{p},s}}$;
4. Update the most likely error standard deviations $\tilde{g}(\mathbf{x}_{\mathrm{p},s}) = \max\left(0, \sqrt{\pi/2}\,\mu_{\mathrm{z}_{\mathrm{p},s}}\right)$;
5. Make prediction on future observations $\mathbf{y}_{\mathrm{p}}|\mathbf{D} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{p},s}^{\mathrm{HSBL}}, \mathbf{K}_{\mathrm{p},ss}^{\mathrm{HSBL}}\right)$.

4.3 Case Study

To demonstrate the advantages of HSBL over SSBL in developing a data-driven statistical model for the population features of NISs when the training data are found heteroscedastic, the case study conducted in the preceding chapter is examined again in this section. In order for a good model to interpret the training data, one may have to carefully select the optimal hyperparameters (kernel widths) of the kernel functions in sparse Bayesian machine learning. Three quality indices are used here for assessing such selection, including the root mean square residual (RMSR), the mean standardized log loss (MSLL) and the sparsity ratio $\mathcal{K}$. The RMSR index is defined as

$$\mathrm{RMSR} = \sqrt{\sum_{q=1}^{Q}\sum_{n=1}^{N}\frac{(y_{qn}-\mu_{\mathrm{p},n}^{\mathrm{HSBL}})^2}{QN}} \tag{4.14}$$

where y_{qn} is the nth testing output from the qth testing sample structure, $\mu_{\mathrm{p},n}^{\mathrm{HSBL}}$ is the corresponding posterior mean of the population feature predicted by HSBL, is the number of data points from each testing sample structure, and Q is the number of the remaining sample structures for model testing. The MSLL index is given by

$$\begin{aligned}\mathrm{MSLL} &= -\frac{1}{QN}\log \mathrm{P}(\mathbf{y}) \\ &= \frac{1}{2QN}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left[\log(2\pi\mathbf{K}_{\mathrm{p},nn}^{\mathrm{HSBL}}) + (\mathbf{K}_{\mathrm{p},nn}^{\mathrm{HSBL}})^{-1}(y_{qn}-\mu_{\mathrm{p},n}^{\mathrm{HSBL}})^2\right]\end{aligned} \tag{4.15}$$

where $\mathbf{y}$ denotes the testing outputs from all testing sample structures, and $\mathbf{K}_{\mathrm{p},nn}^{\mathrm{HSBL}}$ is the nth diagonal element of the posterior covariance matrix $\mathbf{K}_{\mathrm{p}}^{\mathrm{HSBL}}$ in the HSBL population

feature model. The sparsity ratio $\mathcal{K}$ is defined as

$$\mathcal{K} = \frac{N_{\mathrm{RV}}^{\mathrm{HSBL}}}{MN} \times 100\% \tag{4.16}$$

with $N_{\mathrm{RV}}^{\mathrm{HSBL}}$ the number of relevance vectors (the basis functions with nonzero weights) in the HSBL population feature model, calculated by

$$N_{\mathrm{RV}}^{\mathrm{HSBL}} = N_{\mathrm{RV}}^{\mathrm{SSBL}(f)} + N_{\mathrm{RV}}^{\mathrm{SSBL}(g)} \tag{4.17}$$

where $N_{\mathrm{RV}}^{\mathrm{SSBL}(f)}$ and $N_{\mathrm{RV}}^{\mathrm{SSBL}(g)}$ denote the number of relevance vectors in the first and second SSBL models, respectively. The SMSR, SMLL and sparsity ratio $\mathcal{K}$ are three quantities to measure learning losses when some trivial model is preferred. Typically, lower values indicate better performance for the trivial model.

4.3.1 HSBL for population features of nominally identical wheels

An SSBL is first performed on the same training data used in the previous chapter in order to obtain the posterior mean of the population features of nominally identically healthy wheels, shown in Figure 4.1. Based on the first SSBL model, the regression residuals are then calculated and another SSBL is performed on the absolute values of the obtained residuals to update the posterior uncertainty of the population features of NISs, as shown in Figure 4.2. Finally, the combination of the above two SSBL models gives rise to an HSBL model for the population features for NISs, as plotted in Figure 4.3. Intuitionally, the heteroscedastic population feature model from HSBL is much more

consistent with the training data than the homoscedastic one from SSBL. This is achieved at the price of twice the learning effort of an SSBL model. Fortunately, the obtained heteroscedastic population feature model remains very simple benefitting from the mechanism of sparse Bayesian learning and it requires only three additional relevance vectors, as shown in Figure 4.3.

In the process of learning the heteroscedastic population feature model, three indices are used to monitor the model performance and select the optimal kernel width, including RMSR, MSLL and sparsity ratio $\mathcal{K}$ that have been defined before. The RMSR, MSLL and $\mathcal{K}$ against with the kernel width r are shown from Figures 4.4 to 4.6, respectively. It is found that the RMSR index of the heteroscedastic population feature model from an HSBL is exactly the same as that of the homoscedastic model from SSBL. This is because the posterior population feature means are not updated in the current HSBL framework. In general, the RMSR increases with the kernel width. Two local minima of the RMSR index can be found at around $\gamma = 26$ and 40, respectively, corresponding to different model explanations and complexities.

Nevertheless, this is not the case for the MSLL and the sparsity ratio $\mathcal{K}$. It is seen from Figure 4.5 that the MSLL of the heteroscedastic population feature model is much smaller than that of the homoscedastic model for the considered kernel widths. This is because the updated posterior uncertainty of the population features is more consistent

with the training data, thus significantly reducing the MSLL regression loss. There exist also two local minima for the MSLL at around $\gamma = 24$ and 45, respectively. Since additional relevance vectors are required in an HSBL to update the posterior uncertainty of the population features of NISs, the sparsity ratio $\mathcal{K}$ of the heteroscedastic population feature model is a little worse than that of the homoscedastic one, as shown in Figure 4.6.

If there is no additional information about the model complexity, we usually prefer to make use of a larger kernel width that will give rise to a simpler model. However, it should be noted that the larger local optimal kernel width is different from the index RMSR, MSLL and $\mathcal{K}$. In this study, we employ the larger local optimal kernel width associated with the MSLL index $(\gamma = 45)$ as this index is more compatible with data likelihood.

Figures 4.7 and 4.8 provide a detailed comparison of the population feature model derived from SSBL and HSBL. It can be seen that the posterior means of the population features of NISs from SSBL and HSBL are exactly the same as those, not refined in the proposed HSBL. Yet, the posterior uncertainties of the population features obtained from the two different modelling frameworks are distinctly different. The posterior standard deviations of the heteroscedastic population features from HSBL are concave downwards and varying across the considered damage-sensitive frequency band, with smaller variability at both ends but larger variability in the middle. This is in agreement with the

input-dependent variability pattern present in the raw training data. By contrast, the posterior standard deviations of the homoscedastic population features from SSBL are nearly constant across the damage-sensitive frequency band as the fundamental assumption on SSBL is that the pooled uncertainties from measurement noise and structural variability remain homoscedastic across the damage-sensitive frequency bins. As a result, the homoscedastic population feature model stands to locally overestimate the posterior standard deviations in their low-uncertainty area, and is likely to underestimate them in their high-uncertainty area. Interestingly, it is found that the averaged posterior standard deviation of the population feature model from the SSBL is almost identical to that of the HSBL ($\bar{\sigma}_y^{\mathrm{SSBL}} = 0.0224 \approx \bar{\sigma}_y^{\mathrm{HSBL}} = 0.0221$). This phenomenon is not a coincidence. In fact, the assumed constant standard deviation of the errors in the SSBL can be proved to be an approximately unbiased estimator for the averaged standard deviation of the heteroscedastic errors in the HSBL.

Figure 4.9 shows the quantification of multiple sources of uncertainty in the obtained HSBL population feature model. It is seen that the posterior uncertainty of the population feature model for nominally identically healthy wheels remains dominated by the pooled uncertainty (the pooling of measurement noise-induced uncertainty and structural variability-induced uncertainty), whereas the updated one is found to be more consistent with the uncertainty present in the training data. The modelling-induced uncertainty in

the HSBL of the population features is also insignificant as there are as many as 2288 training points.

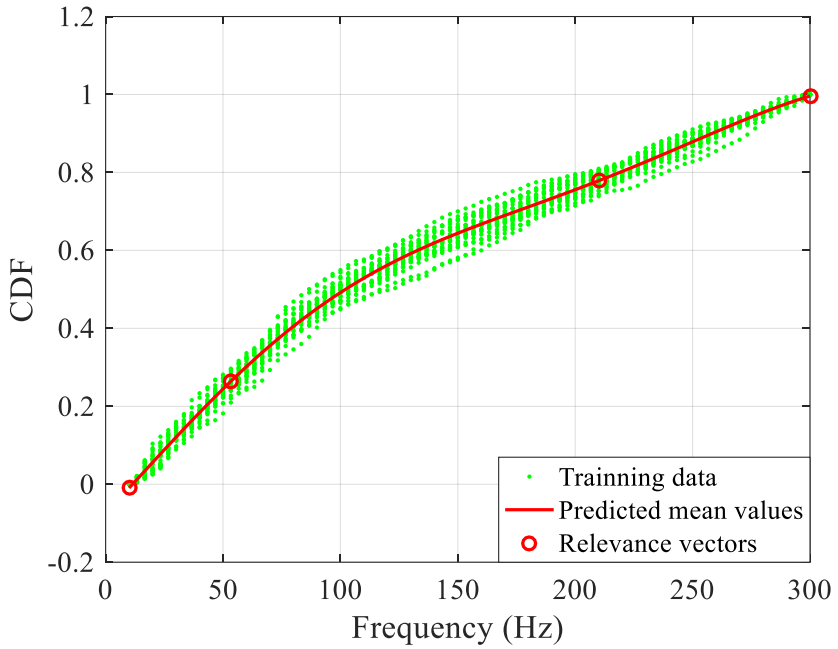


Figure 4.1 Posterior means of the population features of NISs from SSBL(f)

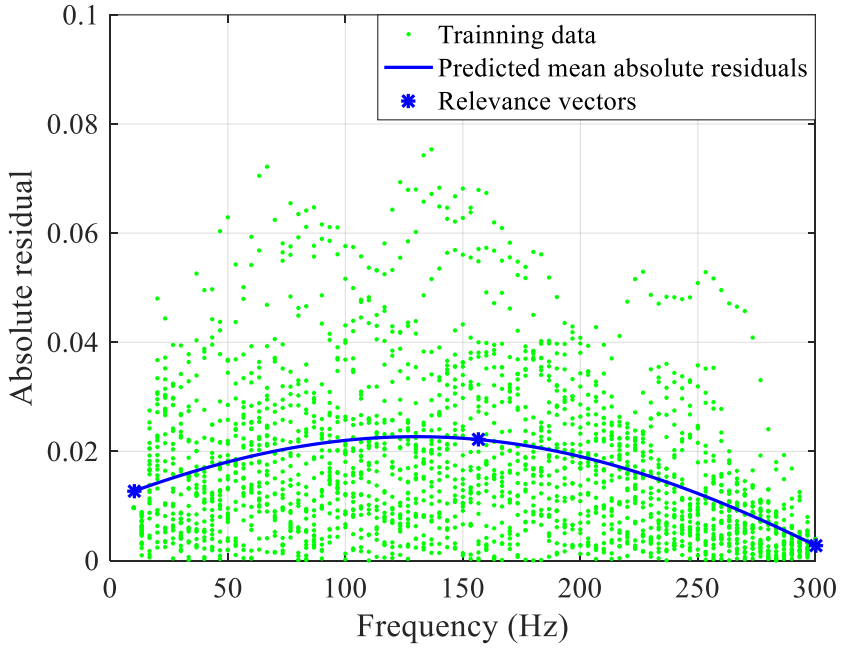


Figure 4.2 Posterior standard deviations of the population features of NISs from SSBL(g)

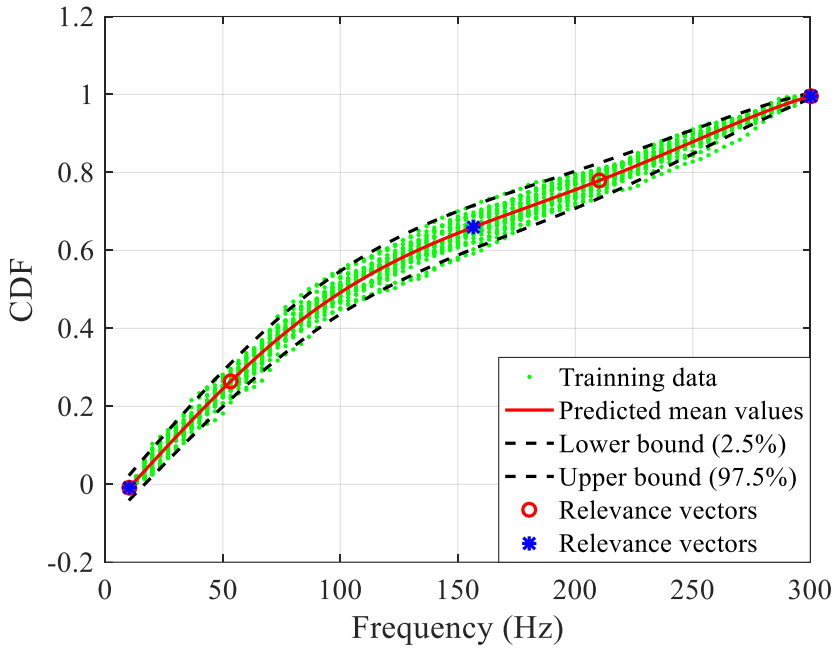


Figure 4.3 Updated population feature model from HSBL

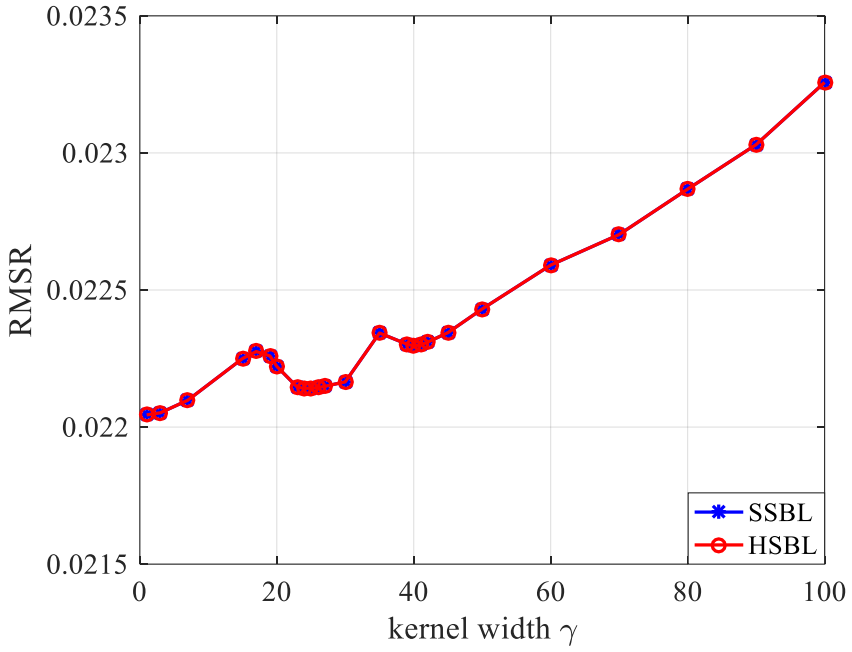


Figure 4.4 RMSR against γ

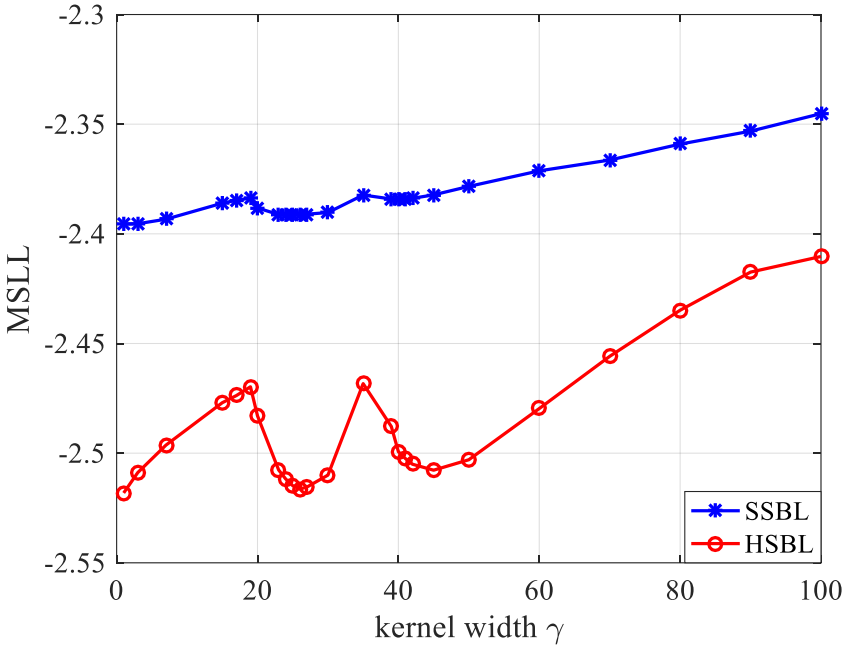


Figure 4.5 MSLL against γ

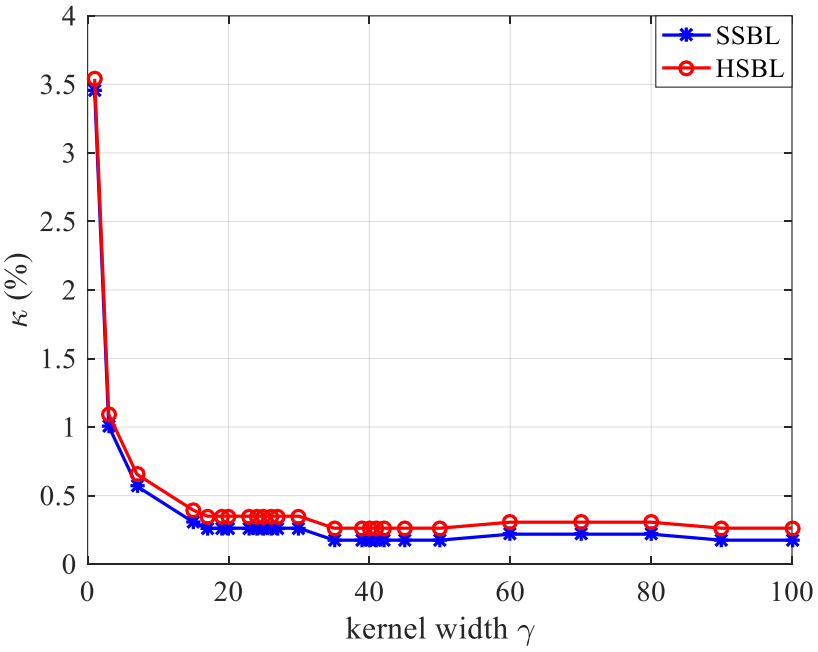


Figure 4.6 Sparsity ratio $\mathcal{K}$ against γ

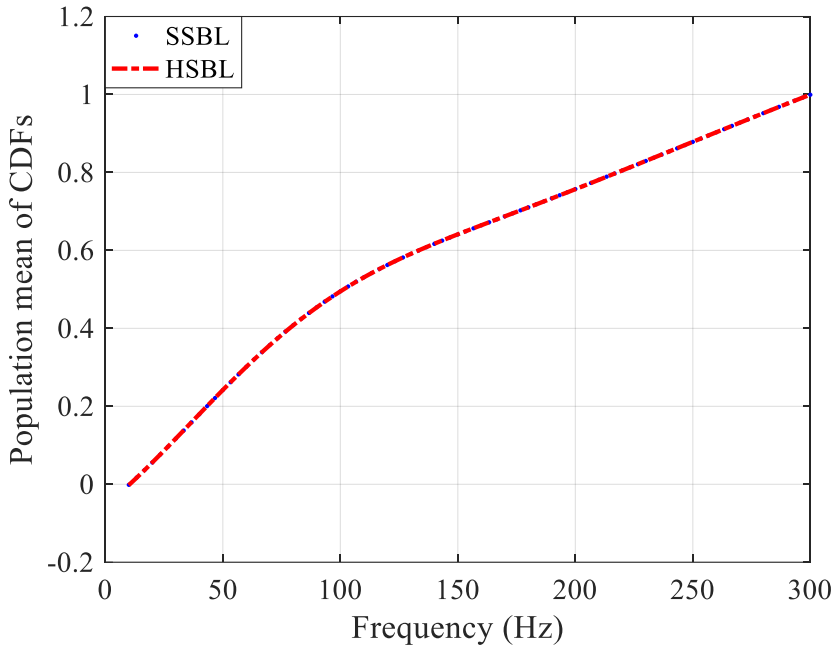


Figure 4.7 Posterior means of the population features of NISs

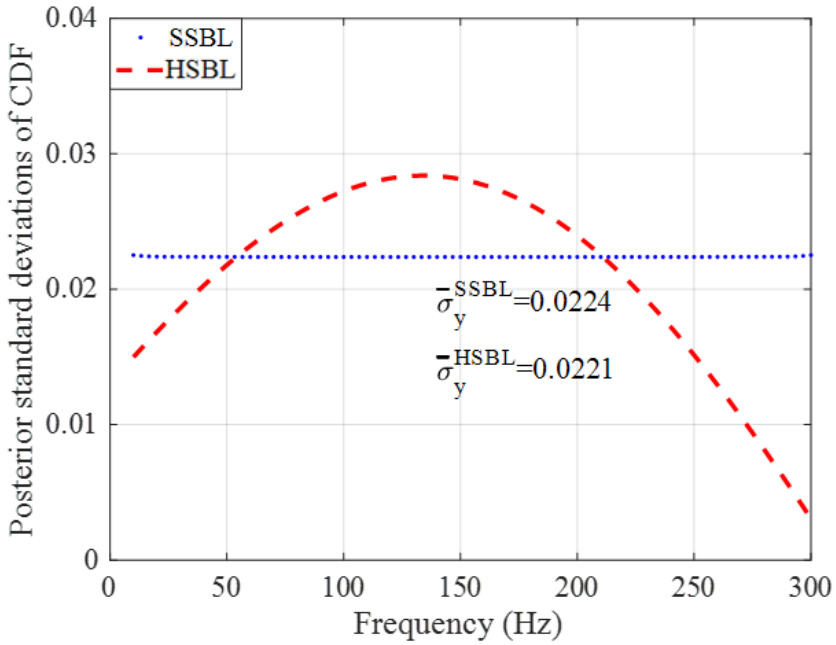


Figure 4.8 Posterior standard deviations of the population features of NISs

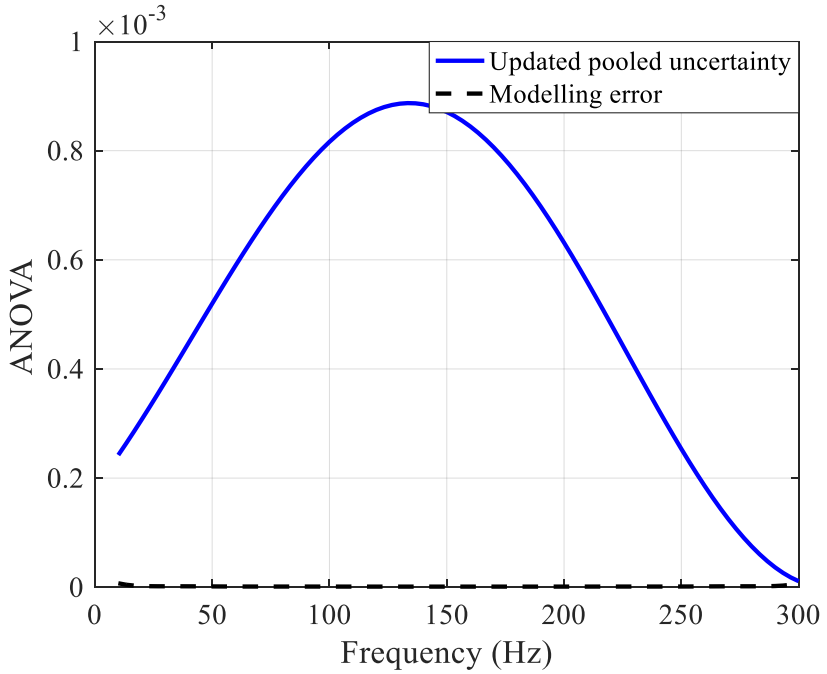


Figure 4.9 ANOVA of multiple sources of uncertainty in the heteroscedastic population feature model

4.4 Summary

This chapter presents another data-driven Bayesian probabilistic framework for modelling the population features of NISs. In this framework, a heteroscedastic variant to SSBL, termed heteroscedastic sparse Bayesian learning (HSBL) is first proposed to deal with the training data with the input-dependent variability, resulting from the pooling of different sources of uncertainty such as measurement noise and structural variability in SHM data. The proposed HSBL is a hierarchical learning-based approach consisting of two SSBL models, with the first for estimating the posterior means of the population features of NISs and the other for adjusting its biased posterior uncertainty. To quantify the input-dependent variability in the training data, we infer it from the absolute residuals in the first SSBL model by the method of moment estimation, which enables to provide an unbiased estimate of it. To demonstrate the advantages of HSBL over SSBL in the modelling of the population features of NISs, the same case study in the previous chapter is then conducted. The comparison results show that the heteroscedastic population feature model inferred from HSBL interprets the training data from NISs much better than that from SSBL, whereas this is achieved at the price of the additional computational effort.

Up to now, we have presented two distinct statistical frameworks for the modelling of the population features of NISs, based on SSBL and HSBL, respectively. The two

frameworks share the same strategy for handling uncertainty in SHM data from multiple nominally identical structures by simply pooling the two different sources of uncertainty, resulting from measurement noise and structural variability. In this strategy, the data dependence is discounted, which may overestimate the information in SHM data, leading to an overly narrow confidence interval for the population feature model. In the next chapter, we will show another novel Bayesian probabilistic framework for modelling the population features of NISs, in which different sources of uncertainty in SHM data can be modelled separately, thus giving rise to a more reliable population feature model.

Chapter 5

PSBL for Population Features of NISs

5.1 Introduction

In Chapter 3, two different types of uncertainty in SHM data are mixed together such that SSBL can be used for modelling and quantifying the combined uncertainty in the population features of nominally identical structures (NISs). A disadvantage caused by the pooled uncertainty is that it may lead to heteroscedasticity in the training data, which has been dealt with in Chapter 4 based on PSBL. Another more important disadvantage of the pooling of different types of uncertainty lies in the information overestimation of data interpretation to some extent (Schubert et al. 2015, Tabor et al. 2018). This may result in an underestimated uncertainty, giving rise to an overly narrow confidence. Therefore, it is important that the data-driven model is able to separately account for different sources of uncertainty in the modelling of population features of NISs.

In this chapter, another novel data-driven population model is established by panel sparse Bayesian learning (PSBL) with which two different sources of uncertainty can be modelled and quantified separately and thus information overestimation is avoided. A group of sub-models will be used in PSBL to model the individual structural behaviors which results in a model space representing the population features of nominally identical

structures. The rest of the chapter is organized as follows. Theoretical aspects of the PSBL methodology is formulated in Section 5.2, followed by a case study carried out in Section 5.3 to demonstrate the superiority of PSBL over SSBL and HSBL in the modelling and quantification of the data-driven population features of nominally identical structures. Concluding remarks are given in Section 5.4.

5.2 Problem Formulation

5.2.1 The training data

As described in the previous two chapters, assume that we have a group of nominally identical structures being monitored, where M structures are known undamaged. From any undamaged structure, we have a training dataset $\mathbf{D}_m = \{(\mathbf{x}_{mn}, y_{mn})\}_{n=1}^{n=N}$, where y_{mn} is the scalar-valued training output, $\mathbf{x}_{mn}$ is the vector-valued training input, m is the serial number of the sample structure for model training and n is the serial number in the training dataset of size N. In the case of M undamaged structures, we have an assembled dataset $\mathbf{D} = \{\mathbf{D}_m\}_{m=1}^{m=M} = \{(\mathbf{x}_{mn}, y_{mn})\}_{m=1,n=1}^{m=M,n=N}$ of size $K = M \times N$ when developing a data-driven statistical model for the interpretation of healthy features for all NISs.

5.2.2 Model specification

Different from SSBL and HSBL that define a single population feature model in the training phase, the proposed PSBL makes use of a group of sub-models as

$$y_{mn} = f_m(\mathbf{x}_{mn}) + \varepsilon_{mn}, \text{ with } \varepsilon_{mn} \sim \mathcal{N}(0, \sigma^2) \tag{5.1}$$

where ε_{mn} is the difference between the training output y_{mn} and the associated individual sub-model output $f_m(\mathbf{x}_{mn})$ from the mth structure at the training input location $\boldsymbol{x}_{mn}$. ε_{mn} is assumed to be mutually independent and normally distributed with zero mean and a constant variance σ^2. Using the kernel-based learning method, the individual sub-model output $f_m(\mathbf{x}_{mn})$ can be represented by a set of nonlinear basis functions as

$$f_m(\mathbf{x}_{mn}) = \boldsymbol{\phi}_m^{\mathrm{T}}(\mathbf{x}_{mn})\mathbf{w}_m = \sum_{l=1}^{L} k(\mathbf{x}_{mn}, \mathbf{x}_{cl}) w_{ml} \tag{5.2}$$

where $\boldsymbol{\phi}_m(\boldsymbol{x}_{mn}) = [k(\mathbf{x}_{mn}, \mathbf{x}_{c1}), \dots, k(\mathbf{x}_{mn}, \mathbf{x}_{cL})]^{\mathrm{T}}$ is a vector of size L. Based on the white noise assumption, the likelihood of the training data can be written as

$$\mathrm{P}(\mathbf{y}_m) = \mathcal{N}(\mathbf{f}_m, \sigma^2\mathbf{I}_N) = \mathcal{N}(\boldsymbol{\Phi}_{1m}\mathbf{w}_m, \sigma^2\mathbf{I}_N) \tag{5.3}$$

where $\mathbf{y}_m = [y_{m1}, \dots, y_{mN}]^{\mathrm{T}}$ is an $N \times 1$ vector, $\mathbf{f}_m = [f_{m1}, \dots, f_{mN}]^{\mathrm{T}}$ is the associated model output and $\mathbf{I}_N$ is the identity matrix of size $N \times N$, and $\boldsymbol{\Phi}_{1m}$ is an $N \times L$ matrix, given as

$$\begin{aligned}\boldsymbol{\Phi}_{1m} &= [\boldsymbol{\phi}_m(\mathbf{x}_{m1}), \dots, \boldsymbol{\phi}_m(\boldsymbol{x}_{mN})]^{\mathrm{T}} \\ &= \begin{bmatrix} k(\mathbf{x}_{m1}, \mathbf{x}_{\mathrm{C}1}) & \cdots & k(\mathbf{x}_{m1}, \mathbf{x}_{CL}) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_{mN}, \mathbf{x}_{\mathrm{C}1}) & \cdots & k(\mathbf{x}_{mN}, \mathbf{x}_{CL}) \end{bmatrix}\end{aligned} \tag{5.4}$$

Based on the independent assumption on the error ε_{mn}, we can apply the product rule of probability theory to obtain the probability of all the training data as

$$\begin{aligned}\mathrm{P}(\mathbf{y}) &= \prod_{m=1}^{M} \mathrm{P}(\mathbf{y}_m) \\ &= \prod_{m=1}^{M} \mathcal{N}(\mathbf{f}_m, \beta^{-1}\mathbf{I}_N) \\ &= \prod_{m=1}^{M} \mathcal{N}\left(\boldsymbol{\Phi}_{1,m}\mathbf{w}_m, \beta^{-1}\mathbf{I}_N\right) \\ &= \mathcal{N}(\boldsymbol{\Phi}_1\mathbf{w}, \beta^{-1}\mathbf{I}_K)\end{aligned} \tag{5.5}$$

where $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \dots, \mathbf{y}_M^{\mathrm{T}}]^{\mathrm{T}}$ is a vector of size MN, $\mathbf{w} = [\mathbf{w}_1^{\mathrm{T}}, \dots, \mathbf{w}_M^{\mathrm{T}}]^{\mathrm{T}}$ is the associated weight vector of size ML, $\mathbf{I}_K$ is the identity matrix of size $MN \times MN$, $\boldsymbol{\Phi}_1$ is the assembled design matrix given as

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \boldsymbol{\Phi}_{1,1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Phi}_{1,M} \end{bmatrix} \tag{5.6}$$

which is a block-diagonal matrix of size $MN \times ML$.

5.2.3 Prior specification

By following the SBL framework, a prior over the weight parameters $\mathbf{w}$ is the zero-

mean Gaussian prior distribution given by

$$\begin{aligned} \mathrm{P}(\mathbf{w}|\boldsymbol{\alpha}) &= \prod_{m=1}^{M} \mathrm{P}(\mathbf{w}_m|\boldsymbol{\alpha}_m) \\ &= \prod_{m=1}^{M}\prod_{l=1}^{L} \mathrm{P}(w_{ml}|\alpha_{ml}) \\ &= \prod_{m=1}^{M}\prod_{l=1}^{L} \mathcal{N}(0, \alpha_{ml}^{-1}) \end{aligned} \tag{5.7}$$

where the hyperparameters $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_M^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{\alpha}_m = [\alpha_{m1}, \ldots, \alpha_{mL}]^{\mathrm{T}}$ are vectors of size ML and L, respectively. A hierarchical prior over the hyperparameters, given as

$$\begin{aligned} \mathrm{P}(\boldsymbol{\alpha}) &= \prod_{m=1}^{M} \mathrm{P}(\boldsymbol{\alpha}_m) \\ &= \prod_{m=1}^{M}\prod_{l=1}^{L} \mathrm{P}(\alpha_{ml}) \\ &= \prod_{m=1}^{M}\prod_{l=1}^{L} \mathrm{Gamma}(\alpha_{ml}|a, b) \end{aligned} \tag{5.8}$$

In the case of the prior over the error term σ^2, a suitable prior is given as

$$\mathrm{P}(1/\sigma^2) = \mathrm{Gamma}(1/\sigma^2|c, d) \tag{5.9}$$

where c and d are hyperparameters of the prior over noise term.

5.2.4 Parameter inference

For the case of uniform hyperpriors, we only need to maximize the term $\mathrm{P}(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)$, which is analytically computable and given by

$$\begin{aligned} \mathrm{P}(\mathbf{y}|\boldsymbol{\alpha},\sigma^2) &= \prod_{m=1}^{M} \mathrm{P}(\mathbf{y}_m|\boldsymbol{\alpha}_m,\sigma^2) \\ &= \prod_{m=1}^{M} \int \mathrm{P}(\mathbf{y}_m|\mathbf{w}_m,\sigma^2)\, p(\mathbf{w}_m|\boldsymbol{\alpha}_m) d\mathbf{w}_m \\ &= \prod_{m=1}^{M} \mathcal{N}\left(\boldsymbol{\mu}_{1,m}, \mathbf{K}_{1,m}\right) \end{aligned} \tag{5.10}$$

with the mean vector $\boldsymbol{\mu}_{1,m}$ and covariance matrix $\mathbf{K}_{1,m}$ given by

$$\boldsymbol{\mu}_{1,m} = \mathbf{0}_N \tag{5.11}$$

$$\mathbf{K}_{1,m} = \boldsymbol{\Phi}_{1,m}\mathbf{A}_{1,m}^{-1}\boldsymbol{\Phi}_{1,m}^{\mathrm{T}} + \sigma^2\mathbf{I}_N \tag{5.12}$$

where $\mathbf{0}_N$ is the zero vector of size L and $\mathbf{A}_{1,m}$ is the diagonal matrix of size $L \times L$ with $\mathbf{A}_{1,m} = \mathrm{diag}(\alpha_{m1}, \alpha_{m2} \ldots, \alpha_{mL})$. The most plausible value of $(\boldsymbol{\alpha}, \sigma^2)$ is denoted by $(\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}})$, derived from

$$(\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) = \mathrm{argmax}_{(\boldsymbol{\alpha},\sigma^2)}\, \mathrm{P}(\mathbf{y}|\boldsymbol{\alpha},\sigma^2) \tag{5.13}$$

Given $(\boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}})$, we can compute analytically the posterior distribution over the weights given by

$$\begin{aligned} \mathrm{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}_{\mathrm{MP}}, \sigma^2_{\mathrm{MP}}) &= \prod_{m=1}^{M} \mathrm{P}\left(\mathbf{w}_m\middle|\mathbf{y}_m, \boldsymbol{\alpha}_{m,\mathrm{MP}}, \sigma^2_{\mathrm{MP}}\right) \\ &= \prod_{m=1}^{M} \frac{\mathrm{P}(\mathbf{y}_m|\mathbf{w}_m, \sigma^2_{\mathrm{MP}})\mathrm{P}(\mathbf{w}_m|\boldsymbol{\alpha}_{m,\mathrm{MP}})}{\mathrm{P}(\mathbf{y}_m|\boldsymbol{\alpha}_{m,\mathrm{MP}}, \sigma^2_{\mathrm{MP}})} \\ &= \prod_{m=1}^{M} \mathcal{N}(\boldsymbol{\mu}_{2,m}, \mathbf{K}_{2,m}) \end{aligned} \tag{5.14}$$

where the mean vector $\boldsymbol{\mu}_{2,m}$ and the covariance matrix $\mathbf{K}_{2,m}$ are given respectively,

$$\mathbf{K}_{2,m} = (\sigma_{\mathrm{MP}}^{-2}\mathbf{\Phi}_{1,m}^{\mathrm{T}}\mathbf{\Phi}_{1,m} + \mathbf{A}_{1,m})^{-1} \tag{5.15}$$

$$\boldsymbol{\mu}_{2,m} = \sigma_{\mathrm{MP}}^{-2}\mathbf{K}_{2,m}\mathbf{\Phi}_{1,m}^{\mathrm{T}}\mathbf{y}_m \tag{5.16}$$

where $\mathbf{A}_{1,m,\mathrm{MP}} = \mathrm{diag}(\alpha_{m1,\mathrm{MP}}, \alpha_{m2,\mathrm{MP}} \dots, \alpha_{mL,\mathrm{MP}})$.

5.2.5 Posterior probability distribution prediction

Given a set of new test points $\mathbf{x}_{\mathrm{p}} = [\mathbf{x}_{\mathrm{p},1}, \dots, \mathbf{x}_{\mathrm{p},S}]^{\mathrm{T}}$ of size S, predictions are made on the population feature model outputs $\mathbf{y}_{\mathrm{p},m} = [y_{\mathrm{p},m1}, \dots, y_{\mathrm{p},mS}]^{\mathrm{T}}$, in terms of the posterior predictive distribution

$$\begin{aligned} \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathbf{y}) &\approx \int \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathbf{w}_m, \sigma_{\mathrm{MP}}^2)\mathrm{P}(\mathbf{w}_m|\boldsymbol{\alpha}_{m,\mathrm{MP}}, \sigma_{\mathrm{MP}}^2)d\mathbf{w}_m \\ &= \mathcal{N}(\boldsymbol{\mu}_{3,m}, \mathbf{K}_{3,m}) \end{aligned} \tag{5.17}$$

where the mean vector $\boldsymbol{\mu}_{3,m}$ and the covariance matrix $\mathbf{K}_{3,m}$ are given respectively as

$$\boldsymbol{\mu}_{3,m} = \mathbf{\Phi}_2\boldsymbol{\mu}_{2,m} \tag{5.18}$$

$$\mathbf{K}_{3,m} = \mathbf{\Phi}_2\mathbf{K}_{2,m}\mathbf{\Phi}_2^{\mathrm{T}} + \sigma_{\mathrm{MP}}^2\mathbf{I}_N \tag{5.19}$$

with $\mathbf{\Phi}_2$ the prediction matrix of size $N \times L$ matrix, given as

$$\begin{aligned} \mathbf{\Phi}_2 &= \left[\boldsymbol{\phi}(\mathbf{x}_{\mathrm{p},1}), \dots, \boldsymbol{\phi}(\mathbf{x}_{\mathrm{p},S})\right]^{\mathrm{T}} \\ &= \begin{bmatrix} k(\mathbf{x}_{\mathrm{p},1}, \mathbf{x}_{\mathrm{C}1}) & \cdots & k(\mathbf{x}_{\mathrm{p},1}, \mathbf{x}_{\mathrm{C}L}) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_{\mathrm{p},S}, \mathbf{x}_{\mathrm{C}1}) & \cdots & k(\mathbf{x}_{\mathrm{p},S}, \mathbf{x}_{\mathrm{C}L}) \end{bmatrix} \end{aligned} \tag{5.20}$$

Therefore, we have the model space $\boldsymbol{\mathcal{M}} = \{\mathcal{M}_1, \cdots, \mathcal{M}_M\}$ with each individual model $\mathcal{M}_m$ consisting of a set of probability densities for random prediction, given as

$$\mathcal{M}_m : \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{3,m}, \mathbf{K}_{3,m}) \tag{5.21}$$

The associated model prior is

$$\mathrm{P}(\mathcal{M}_m) = \pi_m \in (0,1) \quad \forall m = 1, \dots, M \tag{5.22}$$

which satisfies

$$\sum_{m=1}^{M} \mathrm{P}(\mathcal{M}_m) = \sum_{m=1}^{M} \pi_m = 1 \tag{5.23}$$

Using the law of total probability, the population model prediction is

$$\mathrm{P}(\mathbf{y}_{\mathrm{p}}) = \sum_{m=1}^{M} \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathcal{M}_m)\mathrm{P}(\mathcal{M}_m) \stackrel{\text{def}}{=} \sum_{m=1}^{M} \mathrm{P}(\mathbf{y}_{\mathrm{p}}|\mathbf{y})\pi_m \tag{5.24}$$

where the population feature prediction $\mathbf{y}_{\mathrm{p}} = [y_{\mathrm{p},1}, \dots, y_{\mathrm{p},S}]^{\mathrm{T}}$ is a vector of size S and each component distribution is a multivariate Gaussian distribution.

Since they are a linear combination of Gaussian densities, they inherit some of the advantages of the Gaussian distribution: they are analytically tractable for many types of computations; they have desirable asymptotic properties (e.g. the central limit theorem), and thus scale well with the data dimensionality. Furthermore, many natural data sets occur in clusters which are approximately Gaussian. Then, the predictive posterior mean and covariance of the population model are:

$$\boldsymbol{\mu}_{\mathrm{p}} = \mathbb{E}_{\mathrm{P}(\mathbf{y}_{\mathrm{p}})}\{\mathbf{y}_{\mathrm{p}}\} = \sum_{m=1}^{M} \boldsymbol{\mu}_{3,m}\pi_m = \sum_{m=1}^{M} \pi_m \boldsymbol{\Phi}_2 \boldsymbol{\mu}_{2,m} \tag{5.25}$$

$$
\begin{aligned}
\mathbf{K}_{\mathrm{p}} &= \mathbb{E}_{\mathrm{P}(\mathbf{y}_{\mathrm{p}})}\left\{\left(\mathbf{y}_{\mathrm{p}}-\boldsymbol{\mu}_{\mathrm{p}}\right)\left(\mathbf{y}_{\mathrm{p}}-\boldsymbol{\mu}_{\mathrm{p}}\right)^{\mathrm{T}}\right\} \\
&= \mathbb{E}_{\mathrm{P}(\mathbf{y}_{\mathrm{p}})}\left\{\mathbf{y}_{\mathrm{p}}\mathbf{y}_{\mathrm{p}}^{\mathrm{T}}\right\}-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}+\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}} \\
&= \sum_{m=1}^{M}\mathrm{P}(\mathcal{M}_m)\mathbb{E}_{\mathrm{P}(\mathbf{y}_{\mathrm{p}}|\mathcal{M}_m)}\left\{\mathbf{y}_{\mathrm{p}}\mathbf{y}_{\mathrm{p}}^{\mathrm{T}}\right\}-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}} \\
&= \sum_{m=1}^{M}\pi_m\mathbb{E}_{\mathrm{P}(\mathbf{y}_{\mathrm{p}}|\mathbf{y})}\left\{\mathbf{y}_{\mathrm{p},m}\mathbf{y}_{\mathrm{p},m}^{\mathrm{T}}\right\}-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}} \\
&= \sum_{m=1}^{M}\pi_m\left(\mathbf{K}_{3,m}+\boldsymbol{\mu}_{3,m}\boldsymbol{\mu}_{3,m}^{\mathrm{T}}\right)-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}} \\
&= \sum_{m=1}^{M}\pi_m\mathbf{K}_{3,m}+\sum_{m=1}^{M}\pi_m\left(\boldsymbol{\mu}_{3,m}\boldsymbol{\mu}_{3,m}^{\mathrm{T}}-\boldsymbol{\mu}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\right) \\
&= \sum_{m=1}^{M}\pi_m\left[\mathbf{K}_{3,m}+\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)^{\mathrm{T}}\right] \\
&= \sum_{m=1}^{M}\pi_m\left[\boldsymbol{\Phi}_2\mathbf{K}_{2,m}\boldsymbol{\Phi}_2^{\mathrm{T}}+\sigma_{\mathrm{MP}}^2\mathbf{I}_N+\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)^{\mathrm{T}}\right]
\end{aligned}
\tag{5.26}
$$

As shown in Equation (5.25), the mean of population output is the average of means of posterior predictive distributions under a model set, using model probabilities as weights. The covariance of the population model output can be decomposed into the sum of three terms. As shown in Equation (5.26), the first term $\boldsymbol{\Phi}_2\mathbf{K}_{2,m}\boldsymbol{\Phi}_2^{\mathrm{T}}$ is the training error, representing the modelling uncertainty in each model output $\boldsymbol{y}_{\mathrm{p},m}$; the second term $\sigma_{\mathrm{MP}}^2\mathbf{I}_N$ is due to the measurement noise, giving rising to the prediction uncertainty in each model output $\boldsymbol{y}_{\mathrm{p},m}$ and the third term $\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)\left(\boldsymbol{\mu}_{3,m}-\boldsymbol{\mu}_{\mathrm{p}}\right)^{\mathrm{T}}$ is due to structural variability, giving rising to prediction uncertainty on each model $\boldsymbol{y}_{\mathrm{p},m}$ and the

population model output $\boldsymbol{y}_{\mathrm{p},m}$. The sum of the first two terms are called the intra-structure uncertainty, while the last term is called the inter-structure uncertainty.

If there is no additional information about the preference of the candidate models, one can simply assume prior ignorance about which model is preferred through a vague prior, that assumes no model is more favored than any other and the observed data carry all the information. Thus, we can assign prior probabilities to the models as $p(\mathcal{M}_m) \overset{\text{def}}{=} \pi_m = 1/M$. Therefore, the population model output in Equation (5.18) becomes

$$\mathrm{P}(\mathbf{y}_{\mathrm{p}}) = \frac{1}{M}\sum_{m=1}^{M} \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathcal{M}_m) = \frac{1}{M}\sum_{m=1}^{M} \mathrm{P}(\mathbf{y}_{\mathrm{p},m}|\mathbf{y}) \tag{5.27}$$

Thus, the mean vector and covariance matrix of the population feature outputs become

$$\boldsymbol{\mu}_{\mathrm{p}} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\Phi}_2 \boldsymbol{\mu}_{2,m} \tag{5.28}$$

$$\mathbf{K}_{\mathrm{p}} = \frac{1}{M}\sum_{m=1}^{M} \left[\boldsymbol{\Phi}_2 \mathbf{K}_{2,m} \boldsymbol{\Phi}_2^{\mathrm{T}} + \sigma_{\mathrm{MP}}^2 \mathbf{I}_N + \left(\boldsymbol{\mu}_{3,m} - \boldsymbol{\mu}_{\mathrm{p}}\right)\left(\boldsymbol{\mu}_{3,m} - \boldsymbol{\mu}_{\mathrm{p}}\right)^{\mathrm{T}}\right] \tag{5.29}$$

It should be noted that the population model output is a multivariate Gaussian mixture distribution (Titterington et al. 1985), which is not necessarily Gaussian. In particular, they can be sometimes multimodal distributions.

5.3 Case Study

To demonstrate the superiority of PSBL over SSBL and HSBL in developing a data-driven statistical model for the population features of NISs, the case study conducted in

the previous two chapters is examined again in this section. In order for the best statistical model to interpret the training data, one should carefully select the optimal widths of the kernel functions in sparse Bayesian learning. In the section, three model quality indices are employed for such selection, including the root mean square residual (RMSR), the mean standardized log loss (MSLL), and the sparsity ratio $\mathcal{K}$. The RMSR index in the PSBL framework is defined as

$$\text{RMSR} = \sqrt{\sum_{q=1}^{Q}\sum_{n=1}^{N}\frac{(y_{qn}-\mu_{\text{p},n})^2}{QN}} \tag{5.30}$$

where N is the number of the testing data from each individual sample structure for model testing, Q is the total number of the testing sample structures, y_{qn} is the nth testing point in the qth testing sample structure, and $\mu_{\text{p},n}$ is the corresponding HSBL population feature model prediction. The MSLL index is given by

$$\begin{aligned}\text{MSLL} &= -\frac{1}{QN}\log \text{P}(\mathbf{y}) \\ &= \frac{1}{2QN}\sum_{m=1}^{Q}\sum_{n=1}^{N}\left[\log[2\pi\mathbf{K}_{\text{p},nn}] + \mathbf{K}_{\text{p},nn}^{-1}(y_{qn}-\mu_{\text{p},n})^2\right]\end{aligned} \tag{5.31}$$

where $\mathbf{y}$ denotes the testing outputs from all sample structures for model testing, and $\mathbf{K}_{\text{p},nn}$ is the nth diagonal element of the posterior covariance matrix $\mathbf{K}_{\text{p}}$ in the PSBL population feature model. The sparsity ratio $\mathcal{K}$ is defined in PSBL as

$$\mathcal{K} = \frac{N_{RV}}{MN}\times 100\% \tag{5.32}$$

where N_{RV} is the total number of kernel functions with non-zero weights in the panel

sparse Bayesian learning framework. The SMSR, SMLL and sparsity ratio $\mathcal{K}$ are three quantities for measuring the regression losses when some trivial model is preferred and typically, lower values indicate better performance for the trivial model.

5.3.1 PSBL for population features of nominally identical wheels

We perform PSBL on the training data that have been used in the previous two chapters in order to obtain the new statistical model for representing the population features of nominally identical railway wheels. To assess the resulting model performance and choose the best kernel width, three indices are employed in the learning process, including RMSR, MSLL and sparsity ratio $\mathcal{K}$, that are defined in the last section. Figures 5.1 to 5.3 show the change of the RMSR, MSLL and sparsity ratio $\mathcal{K}$ against the kernel width γ, respectively. From Figures 5.1 and 5.2, it can be observed that the RMSR and MSLL regression losses increase noticeably with the kernel width γ. Within the kernel width range considered, there exist two different local optimal kernel widths, with $\gamma = 24$ and 45 for the RMSR, but $\gamma = 24$ and 42 for the MSLL, corresponding to different explanations about the complexity of the population feature model for NISs. By contrast, the sparsity ratio $\mathcal{K}$ almost deceases gradually with the kernel width γ, giving rise to a population feature model increasingly simple (Figure 5.3). As a compromise between expressive ability and sparseness, the kernel width $\gamma = 42$ is considered in this

study as the optimal value to construct the probabilistic population feature model that is fairly simple and favorably consistent with the likelihood of the training data. If the kernel width $\gamma = 24$ is used, the resulting population feature model tends to be more complex, but with a reduced generalization performance.

Comparing the regression losses caused by the three different modelling frameworks when developing a data-driven statistical model for representing the population features for the monitored railway wheels, it can be found that within the kernel width $\gamma = 0{\sim}70$, the SMSR and MSLL losses caused by PSBL are much less than those by SSBL and HSBL. This is mainly because the separate modelling scheme of different sources of uncertainty in PSBL has eliminated the effect of structural variability on the two regression loss items. By taking advantage of the separate modelling scheme, it is possible to calculate the regression loss in PSBL that is only caused by the measurement noise and modelling error. By contrast, the regression loss in the pooled modelling schemes including SSBL and HSBL is caused by the measurement noise, structural variability and modelling error, with the first two components pooled. However, when we increase the kernel width to 100, the PSBL may lose the ability to interpret the training data, leading to the SMSR and MSLL losses even larger than those in SSBL and HSBL. The sparsity of the population feature model given by PSBL is typically much worse than that by SSBL and HSBL (Figure 5.3), due to the fact that the population feature model in PSBL is in

fact a collection of many sub-models and for each sub-model, a sufficient number of relevance vectors have to be used to interpret the training data. As a result, the total number of relevance vectors in PSBL is far more than that in SSBL and PSBL.

Figure 5.4 gives the obtained population feature model from PSBL, that is much more sophisticated than that from SSBL (Figure 3.12) and HSBL (Figure 4.3). In the PSBL, 153 relevance vectors are required, whereas 3 and 5 relevance vectors are only required in SSBL and in HSBL, respectively. This suggests that the population feature model from PSBL is much more complicated, and different sources of uncertainty will greatly increase the modelling complexity of the population feature model. Fortunately, since these models are all data-driven, the resulting population feature model given by PSBL remains quite simple and fast damage diagnostics for nominally identical structures can be achieved.

A more detailed comparison of the three SBL frameworks in developing data-driven statistical models for representing the population features of nominally identical railway wheels is given in Figures 5.5 and 5.6. It is found that the three CDF curves in Figure 5.5 for representing the posterior means of the population features of railway wheels almost coincide, which suggests that the three different SBL frameworks do not have a serious impact on the modelling of the mean population features of NISs. However, this is not the case for the estimated posterior uncertainties associated with the population features

of NISs. From Figure 5.6, it can be observed the standard deviations of the population features of railway wheels estimated from SSBL are approximately a straight line in the damage-sensitive frequency band, due to the assumption of homoskedasticity in SSBL. This does not match the heteroscedasticity in the training data. By contrast, those estimated from HSBL stand to be a concave curve, with smaller variability at both ends of the damage-sensitive frequency band but larger variability in the middle, which is more consistent with the heteroscedastic training data. The inferred standard deviations of the population features of railway wheels from PSBL share the same heteroscedastic pattern in the damage-sensitive frequency band as those from HSBL, with smaller variability at both ends but larger variability in the middle, though the corresponding curve is the most complicated. In addition, the averaged standard deviations of the population features of railway wheels from PSBL is obviously larger than that from SSBL or HSBL $(\bar{\sigma}_y^{\mathrm{PSBL}} = 0.0247 > \bar{\sigma}_y^{\mathrm{SSBL}} = 0.0224 \approx \bar{\sigma}_y^{\mathrm{HSBL}} = 0.0222)$. This advocates that the pooling of different types of uncertainty utilized in both SSBL and HSBL, indeed, may lead to an underestimation of the uncertainty associated with the population features for NISs. It should be mentioned that if the nominally identical structures are monitored under strong measurement noise, the multiple sources of uncertainty in the monitoring data could be underestimated significantly due to the fact that the pooling scheme neglects the intrinsic dependence within the monitoring data that is collected from some individual structures.

As such, the posterior uncertainty of the population feature model learned from SSBL and HSBL will be underestimated. Therefore, the separate modelling strategy of different types of uncertainty used in PSBL should be preferred when considering a data-driven statistical model for characterizing the population features of NISs.

Another remarkable advantage of the separate modelling strategy in PSBL is that it allows us to conduct the analysis of variance (ANOVA) and determine separate contributions of different sources of uncertainty to the diversity of the learned population features of NISs. In this case study, structural variability contributes the most uncertainty in the population features of nominally identical railway wheels as shown in Figure 5.7. Although the measurement noise contributes a considerable amount of uncertainty in the population features, whereas it is significantly less than that induced by structural variability. The uncertainty induced by the modelling error in the population features is barely visible due to the fact that we make use of as many as 2288 training points in developing the data-driven model for the population features of NISs.

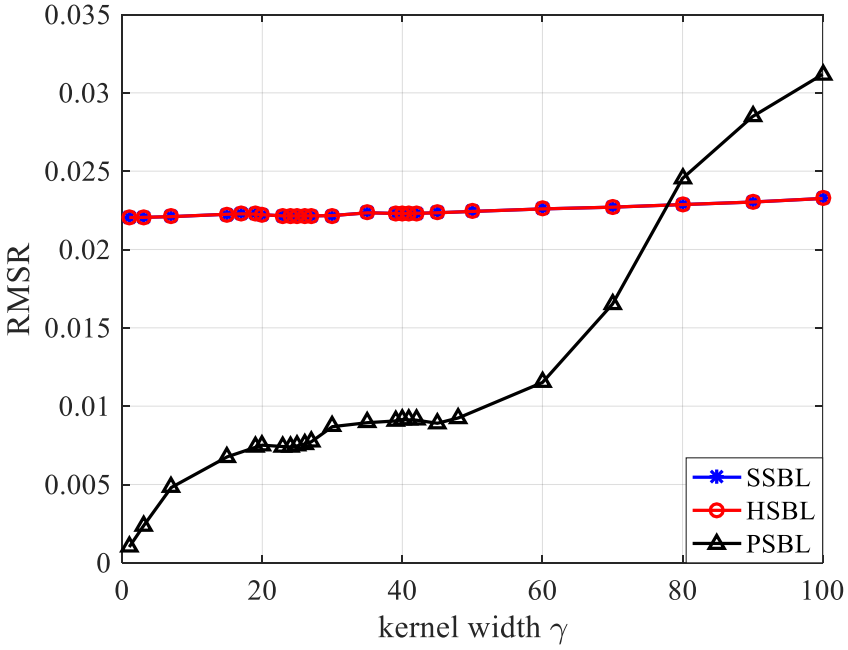


Figure 5.1 SMSR against kernel width γ

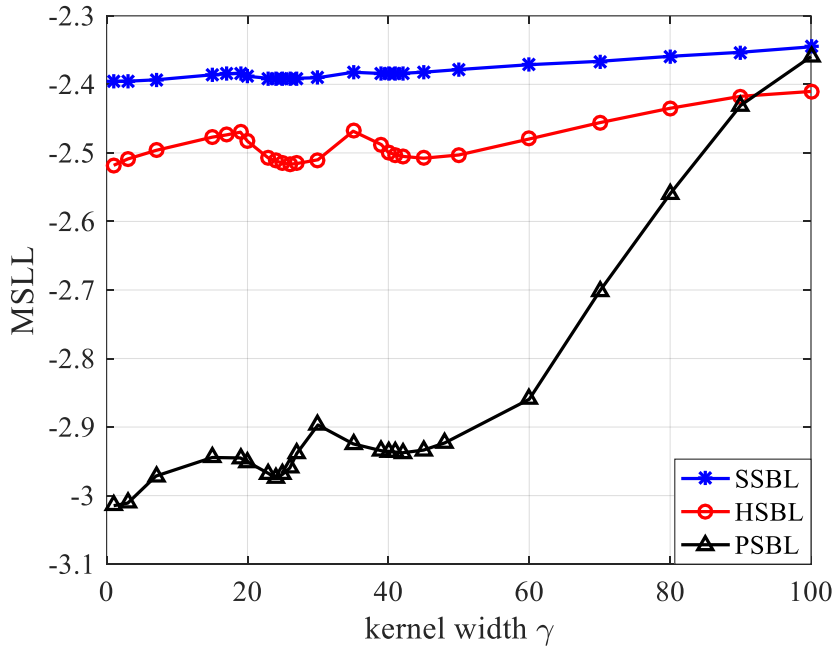


Figure 5.2 MSLL against kernel width γ

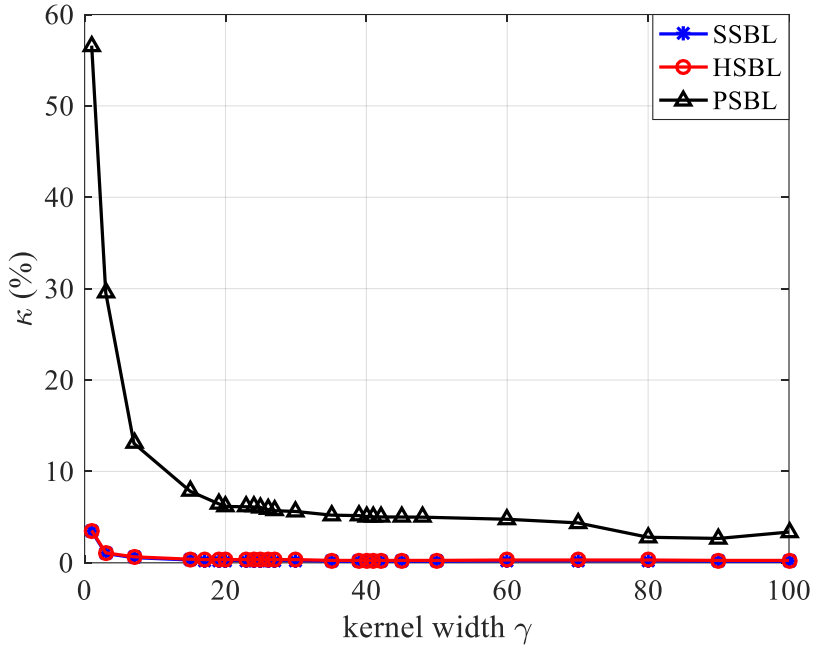


Figure 5.3 Sparsity ratio $\mathcal{K}$ against kernel width γ

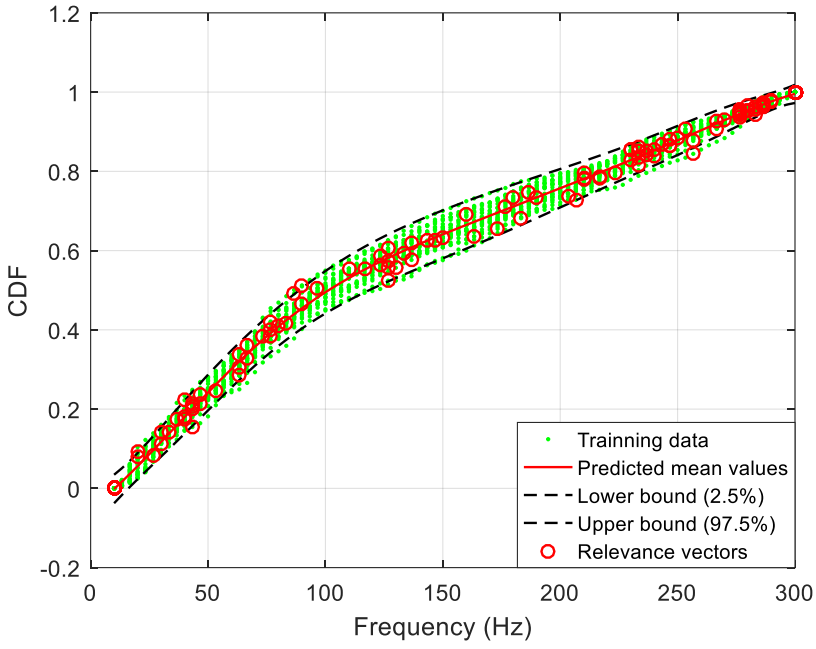


Figure 5.4 Population feature model derived from PSBL

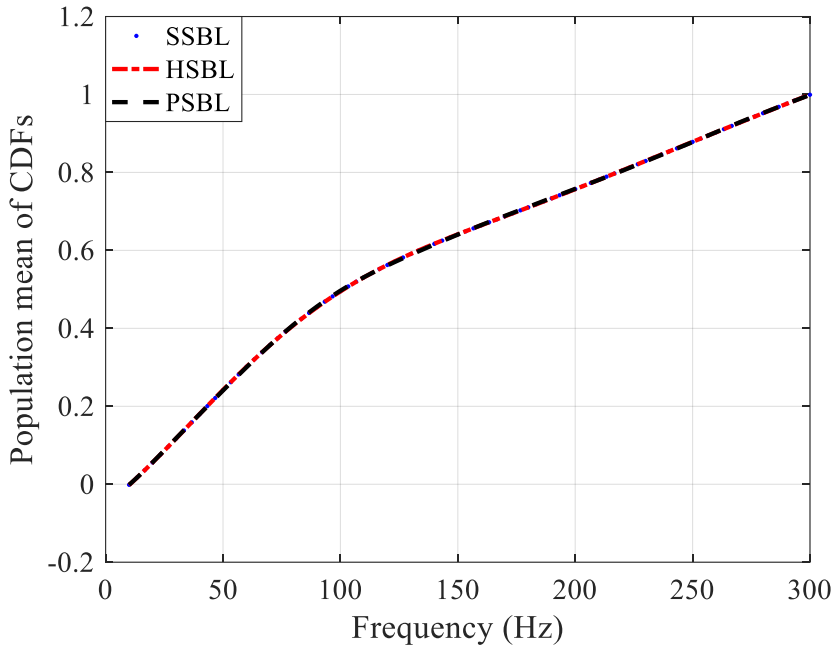


Figure 5.5 Posterior means of the population features of NISs

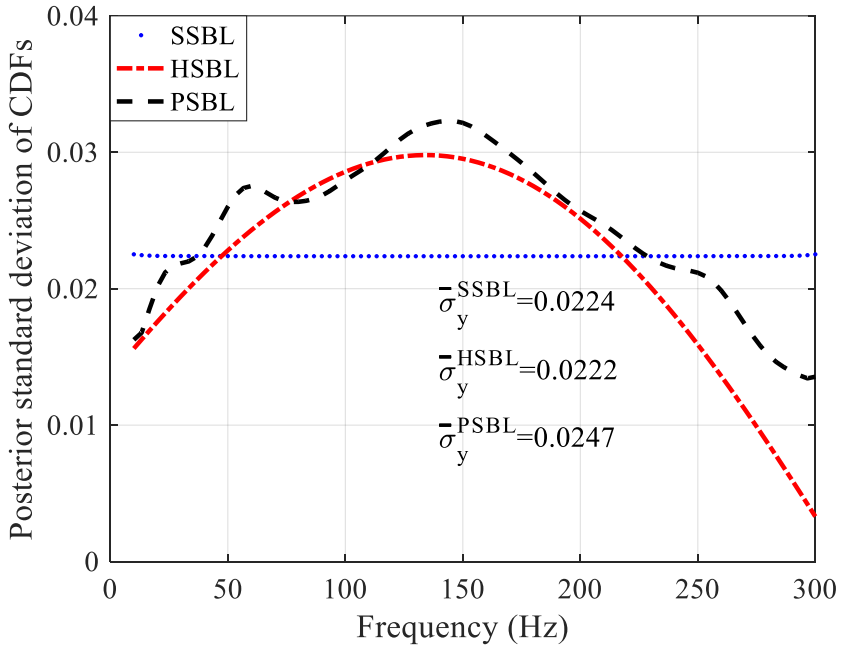


Figure 5.6 Posterior standard deviations of the population features of NISs

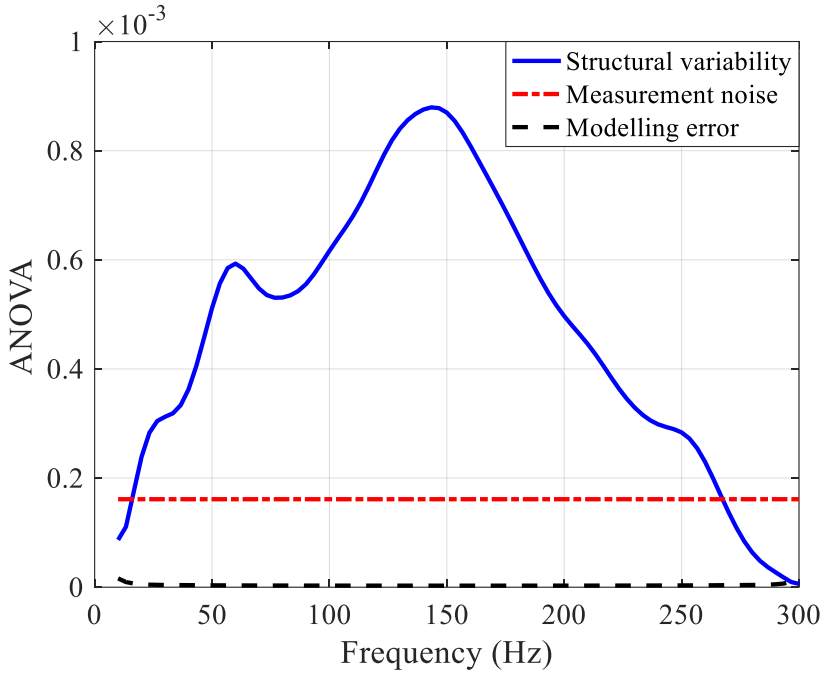


Figure 5.7 ANOVA of multiple sources of uncertainty in the panel population feature model

5.4 Summary

This chapter presents another novel data-driven probabilistic framework for modelling the population features of nominally identical structures, based on panel sparse Bayesian learning. In the framework, different sources of uncertainty in the acquired SHM data are made possible to be dealt with separately, including the measurement noise within each individual structure and structural variability between multiple nominally identical structures. The obtained population feature model of nominally identical structures from panel sparse Bayesian learning becomes a collection of sparse Bayesian sub-models, with each sub-model representing the feature behavior of each sample structure monitored. With the aid of the panel sparse Bayesian learning framework, the problem of information overestimation in both standard and heteroscedastic sparse Bayesian learnings is avoided, resulting from the simple pooling of these different categories of uncertainty. More importantly, these different categories of uncertainty can be quantified separately in panel sparse Bayesian learning.

To validate the superiority of the PSBL over SSBL and PSBL in developing a data-driven population feature model for nominally identical structures, the case study throughout the thesis is carried out again. It can be found that no matter which framework of sparse Bayesian learning is used, the posterior means of the population feature for nominally identical structures are almost the same. Nevertheless, the estimated posterior

uncertainties associated with the population feature of nominally identical structures from PSBL stand to be larger than those inferred from SSBL and HSBL, which indicates the necessity of panel sparse Bayesian learning framework.

In the next chapter, we will further examine the impact of the three SBL modelling frameworks on the subsequent damage diagnostics for the concerned nominally identical structures.

Chapter 6

Statistical Tests for Damage Diagnostics of NISs

6.1 Introduction

This thesis aims to develop a data-driven damage detection method specifically targeted at a population of nominally identical structures (NISs). The method is formulated in an unsupervised learning scheme which makes only use of response data acquired from a number of undamaged structures being nominally identical. The scheme consists of two phases: the baseline phase and the inspection phase. In the baseline phase, response-only data are employed to build a data-driven statistical model for representing population features of all nominally identical undamaged structures. This has been pursued in the preceding three chapters to obtain the optimal population model based on three possible modelling frameworks: standard sparse Bayesian learning (SSBL), heteroscedastic sparse Bayesian learning (HSBL) and panel sparse Bayesian learning (PSBL). In the inspection phase, new measurements will be compared to predictions from the population feature model established in the baseline phase and, as a result, the identification and quantification of damage in NISs can be attained by following some objective diagnostic criteria, which is the main focus of this chapter.

There exist a vast of diagnostic criteria available for structural damage identification

and quantification. The most commonly used criterion is to use Euclidean distance to measure the differences or residuals between new measurements and model predictions. Oftentimes, a certain threshold is empirically predetermined to classify the structures of interest as damaged or undamaged (Chang et al. 2000, Sohn and Law 2001, Titurus et al. 2003, Yuen et al. 2004). A preferable measure is to use the Mahalanobis distance in order to take into account the changing precisions in model predictions (Mosavi et al. 2012, Figueiredo et al. 2014, Vamvoudakis-Stefanou et al. 2018, Yeager et al. 2019, Villani et al. 2019). Yet, the distance-based diagnostic methods are relatively subjective. More importantly, when an inappropriate threshold value is selected, the induced risk of false damage detection results cannot be illustrated. Alternatively, one may resort to statistical hypothesis tests to seek a more scientific procedure for structural damage identification and quantification. For example, classical null hypothesis significant testing (NHST) can be used to not only identify the damage but also to illustrate the induced risk of such a diagnostic procedure, in terms of the significance level (Mao and Todd 2013, Yuen and Ortiz 2017, Jamshidi et al. 2017). Yet, it is difficult to quantify the damage extent in a direct manner. By contrast, Bayesian point null hypothesis testing (PNHT) that allows structural damage to be identified and quantified in terms of Bayes factor has been recently employed (Jiang and Mahadevan 2008, Sankararaman and Mahadevan 2011 and 2013, Wang et al. 2018). The resulting risk of it is averaged over the priors for unknown

parameters in the hypotheses and thus it is more robust than the distance-based diagnostic methods. Nevertheless, Bayesian PNHT is very sensitive to the unknown parameter priors, giving rise to the so-called Jeffreys-Lindley paradox. A detailed interpretation and discussion of statistical hypothesis tests for structural damage identification and quantification will be presented in the next two sections.

In this chapter, a novel damage diagnostic logic is proposed based on Bayesian null hypothesis significance testing (NHST) with which these disadvantages of the damage diagnostic criteria mentioned above can be avoided. The rest of the chapter is organized as follows. Two types of Bayesian residuals including the raw Bayesian residual and the standardized Bayesian residual are introduced in Section 6.2. Upon the standardized Bayesian residual, three categories of statistical hypothesis tests served as structural damage diagnostics are then presented in Section 6.3, including classical NHST, Bayesian PNHT and the newly proposed Bayesian NHST. A case study that has been conducted in the last three chapters is investigated in Section 6.4 to verify the effectiveness of the Bayesian NHST on damage diagnostics in a population of nominally identical structures. The concluding remarks are presented in Section 6.5.

6.2 Bayesian Residuals

In the preceding three chapters, we have made used of M healthy sample structures

being nominally identical to derive a population feature model for the representation of all nominally identical healthy structures. In this chapter, we are required to diagnose whether or not a structure of concern is in an undamaged state. The concerned structure can be one of the nominally identical sample structures used in the baseline phase, with its initial state known undamaged but its current state unknown. It can be also one of the structures, while not used in the baseline phase, in the same category as the sample structures, operating in the same operational environment). From the concerned structure, we may also extract a set of damage-sensitive feature outputs $\mathbf{y}_{\mathrm{u}} = [y_{\mathrm{u},1}, \dots, y_{\mathrm{u},S}]^{\mathrm{T}}$ at the feature inputs $\mathbf{x}_{\mathrm{u}} = [x_{\mathrm{u},1}, \dots, x_{\mathrm{u},S}]^{\mathrm{T}}$, with the S being the number of feature input-output pairs. From the population feature model, we can obtain the predicted population feature outputs $\mathbf{y}_{\mathrm{p}} = [y_{\mathrm{p},1}, \dots, y_{\mathrm{p},S}]^{\mathrm{T}}$ at the feature inputs $\mathbf{x}_{\mathrm{u}}$. As a result, damage diagnostics of the concerned structure can be attained by examining the discrepancies (also called residuals) between the newly extracted damage-sensitive feature outputs $\mathbf{y}_{\mathrm{u}}$ and the corresponding population feature model predictions $\mathbf{y}_{\mathrm{p}}$. If the residuals are as small as expected, the concerned structure is diagnosed undamaged. By contrast, if the residuals are extraordinarily large, the concerned structure is diagnosed damaged. Before introducing statistical tests for damage diagnostics in NISs, we first present three types of Bayesian residuals in order to meet with the fundamental assumption among any statistical hypothesis test: samples to be tested are mutually independent and identically

distributed from the same population.

6.2.1 Raw Bayesian residual

In the context of Bayesian regression analysis, residuals are referred to as the differences between the observed outputs and the corresponding posterior means from a probabilistic prediction model that is learned from Bayesian inference (Carlin et al. 2010, Gelman et al. 2013). For the problem of damage diagnostics of nominally identical structures considered in this thesis, the residuals are referred to as the differences between the damage-sensitive feature outputs of the concerned structure and the corresponding posterior predictive means from some population feature model. Though the population feature model can be pursued in distinct modelling frameworks (SSBL, HSBL or PSBL), the corresponding posterior predictive means can be described by a joint Gaussian distribution, given by

$$\mathbf{y}_{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{p}}, \mathbf{K}_{\mathrm{p}}) \tag{6.1}$$

where $\boldsymbol{\mu}_{\mathrm{p}}$ and $\mathbf{K}_{\mathrm{p}}$ are, respectively, the posterior mean vector and covariance matrix of the predicted population feature outputs $\mathbf{y}_{\mathrm{p}}$, corresponding to the extracted damage-sensitive feature outputs $\mathbf{y}_{\mathrm{u}}$ at the damage-sensitive feature inputs $\mathbf{x}_{\mathrm{u}}$ from the concerned structure. The posterior mean vector $\boldsymbol{\mu}_{\mathrm{p}}$ and the posterior covariance matrix $\mathbf{K}_{\mathrm{p}}$ are given, respectively, in Equations (3.22) and (3.33), equations (4.4) and (4.12),

and Equations (5.25) and (5.26). Using the definition of the residuals in the problem of damage diagnostics for NISs, we have

$$\mathbf{r} = \mathbf{y}_{\mathrm{u}} - \mathbb{E}\{\mathbf{y}_{\mathrm{p}}\} = \mathbf{y}_{\mathrm{u}} - \boldsymbol{\mu}_{\mathrm{p}} \tag{6.2}$$

where $\mathbf{r} = [r_1, \dots, r_S]^{\mathrm{T}}$ are termed Bayesian raw residuals, without transformation. If the population feature model is well defined and the concerned structure happens to be undamaged, the residuals, according to the Gaussian identity, is jointly Gaussian with zero means and the covariance $\mathbf{K}_{\mathrm{p}}$, given by

$$\mathbf{r} = \mathbf{y}_{\mathrm{u}} - \boldsymbol{\mu}_{\mathrm{p}} \sim \mathcal{N}(\mathbf{0}_S, \mathbf{K}_{\mathrm{p}}) \tag{6.3}$$

whereas $\mathbf{0}_S$ denotes the zero vector of size S. It can be found from Equation (6.3) that the residuals $\mathbf{r}$ are not necessarily mutually independent and identically distributed as the covariance matrix $\mathbf{K}_{\mathrm{p}}$ is unnecessarily a scalar matrix, whose off-diagonal elements are not all zero and diagonal elements are not all equal. For instance, when the population feature model is formulated in the SSBL framework, the posterior covariance matrix $\mathbf{K}_{\mathrm{p}}$ of the predicted population features is composed of two items: $\boldsymbol{\Phi}_2 \mathbf{K}_2 \boldsymbol{\Phi}_2^{\mathrm{T}}$ and $\sigma_{\mathrm{MP}}^2 \mathbf{I}_S$. The first item $\boldsymbol{\Phi}_2 \mathrm{K}_2 \boldsymbol{\Phi}_2^{\mathrm{T}}$, typically, is not a scalar matrix, though the second item $\sigma_{\mathrm{MP}}^2 \mathbf{I}_S$ is a scalar. The sum of the two items may result in the correlation and heteroscedasticity among the raw Bayesian residuals $\mathbf{r}$. Similar conclusions can be made in the case of the population feature model established by HSBL or PSBL. Therefore, generally, the raw Bayesian residuals cannot be directly tested if pursuing a more accurate diagnostic result

of the health state about the concerned structure.

6.2.2 Standardized Bayesian residual

The raw Bayesian residuals should be decorrelated and standardized before conducting statistical hypothesis tests to make certain the health state of the concerned structure. There exist a number of techniques to decorrelate and standardize the raw Bayesian residuals such that they can satisfy the independent and identically distributed assumption among statistical hypothesis tests. Due to the fact that the posterior covariance matrix $\mathbf{K}_{\mathrm{p}}$ is always positive definite, the raw Bayesian residuals $\mathbf{r}$ can be linearly transformed into

$$\boldsymbol{\xi} = \mathbf{K}_{\mathrm{p}}^{-1/2}\mathbf{r} = \mathbf{K}_{\mathrm{p}}^{-1/2}(\mathbf{y}_{\mathrm{u}} - \boldsymbol{\mu}_{\mathrm{p}}) \sim \mathcal{N}(\mathbf{0}_S, \mathbf{I}_S) \tag{6.4}$$

where $\mathbf{0}_S$ is the zero vector of size S, $\mathbf{I}_S$ is the identity matrix of size S, and $\boldsymbol{\xi} = [\xi_1, \dots, \xi_S]^{\mathrm{T}}$ are termed the standardized Bayesian residuals which are mutually independent and identically distributed with each $\xi_s \sim \mathcal{N}(0,1)$. This linear transformation is known as the Mahalanobis transformation which is in charge of the decorrelation and standardization of random variables (Härdle and Simar 2015). The Mahalanobis transformation is believed the optimal in the sense that the transformed random variables (the standardized Bayesian residuals $\boldsymbol{\xi}$) have the maximum similarity to the original variables (the raw Bayesian residuals $\mathbf{r}$) but the minimal additional adjustment, though

there remain other linear transformation techniques to eliminate the correlation and heteroscedasticity among random variables (e.g. Kessy et al. 2018).

6.3 Statistical Tests for Damage Diagnostics

To develop a scientific procedure for damage diagnostics of nominally identical structures, statistical hypothesis tests should be employed. Without loss of generality, we conduct statistical hypothesis tests on the standardized Bayesian residuals $\boldsymbol{\xi}$, which are mutually independent and identically distributed standard normal random variables. To be specific, we examine whether or not the mean of the standardized Bayesian residuals is around zero. If their mean is found around zero, there is no discrimination between the newly extracted damage-sensitive feature outputs and the corresponding predictions from the population feature model, that is, the concerned structure is classified as healthy; discrimination exists if otherwise, that is, the concerned structure is diagnosed as damaged. Three categories of statistical hypothesis tests, including classical NHST, Bayesian PNHT and Bayesian NHST are introduced in this section to illustrate the benefits of the Bayesian NHST over frequentist NHST and Bayesian PNHT in developing a scientific damage diagnostic method for nominally identical structures.

6.3.1 Frequentist null hypothesis significance testing

The frequentist NHST can be the most common practice when pursuing a scientific methodology for structural damage diagnostics. In the case of the standardized Bayesian residuals ξ_s, we may assume that $\xi_s \sim \mathcal{N}(\mu_\xi, \sigma^2)$, where $\sigma^2 = 1$ is known. To determine whether or not the concerned structure is undamaged, two mutually exclusive hypotheses H_0 (null hypothesis) and H_1 (alternative hypothesis) can be made for the concerned structure in its undamaged state and damaged state, given by

$$\begin{aligned} H_0 &: \mu_\xi = 0 \ \text{(the concerned structure is undamaged)} \\ H_1 &: \mu_\xi \neq 0 \ \text{(the concerned structure is damaged)} \end{aligned} \tag{6.5}$$

In frequentist NHST, a usual z statistic is often calculated to test the two competing hypotheses. For the standardized Bayesian residuals, the z statistic is given as

$$z = \frac{\left|\bar{\xi} - 0\right|}{1/\sqrt{S}} = \sqrt{S}\left|\bar{\xi}\right| \tag{6.6}$$

where $\bar{\xi} = \frac{1}{S}\sum_{s=1}^{S} \xi_s$ .Then, one can easily obtain the usual p-value in the above frequentist NHST as

$$p = 2[1 - \Phi(|z|)] \tag{6.7}$$

where Φ is the cumulative distribution function of standard normal random variable and the p-value is the probability that a statistical summary would be equal to or more extreme than the observed value, when the null hypothesis H_0 is true (Wasserstein and Lazar 2016). When the p-value exceeds a certain threshold, called the significance level

of the test, traditionally 0.05 or 0.01 and denoted as α, that suggests that the null hypothesis H_0 should not be rejected. In other words, the concerned structure will be diagnosed as healthy when $p > \alpha$. Otherwise, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 should be accepted. Thus, the concerned structure is judged as damaged when $p < \alpha$.

Typically, there exist two categories of testing errors among statistical hypothesis tests: type-I error and type-II error (DeGroot and Schervish 2002, Lehmann and Romano 2006). The type-I error, called the significance level α (also known as the false-positive error) is the rejection of a true null hypothesis, while the type-II error (also known as false-negative error, denoted as β) is the non-rejection of a false null hypothesis. Consequently, there are two types of diagnostic risks associated with the procedure of structural damage diagnostics. The type-I diagnostic risk is the error that a healthy structure is falsely judged as damaged with the error rate being no more than α, while the type-II diagnostic risk is the error that a damaged structure is incorrectly diagnosed as healthy, with the error rate being no more than β. Many of statistical tests desire to minimize both of the two testing or diagnostic errors, which appears impractical due to the fact that the real alternative to the null hypothesis is not available but supposed. In the case of unsupervised structural damage detection, for example, the statistical model of the concerned structure in its damaged state is assumed rather than inferred from its

damaged data. Consequently, most frequentist tests, in essence, is controlling the type-I testing or diagnostic error rate (the significance level α). The testing errors associated with structural damage diagnostics are illustrated in Table 6.1.

Although frequentist NHST is helpful for the identification of damage in a structure, a number of important limitations attached to it can be found. First, the p-value in frequentist NHST is a conditional probability that only figures out the likelihood of the data (or any mor extreme result), given that the null hypothesis is correct: $\mathrm{P}(D|H_0)$. It is incapable of providing the direct evidence for the acceptance or rejection of the null hypothesis H_0, that is the probability of the null hypothesis being true, given the data: $\mathrm{P}(H_0|D)$. Second, frequentist NHST is only utilized to make a choice of the null and alternative hypotheses and the p-value itself does not offer an assessment of the strength of the evidence in favor of the null hypothesis (Kass and Raftery 1995). As a result, the identification of damage in the concerned structure is possible, but the quantification of damage is difficult. Third, there may be other useful information that could be used to construct an alternative hypothesis; that is, there was prior information. The prior information, however, is difficult to be encoded in the alternative hypothesis in frequentist NHST.

Table 6.1 Diagnostic errors in structural damage detection

Decision	H_0 is true (the concerned structure is undamaged)	H_0 is false (the concerned structure is damaged)
Accept H_0	Correct decision (the undamaged structure is correctly judged as healthy)	Type-II error β (the damaged structure is falsely judged as undamaged)
Reject H_0	Type-I error α (the undamaged structure is judged as damaged)	Correct decision (the damaged structure is correctly judged as damaged)

6.3.2 Bayesian point null hypothesis testing

To overcome the limitations associated with classical NHST, Bayesian hypothesis testing was introduced by Harold Jeffreys in 1939. Bayesian hypothesis testing through Bayes factor provides researchers with several key practical advantages. First, the Bayes factor quantifies evidence for and against the two competing statistical hypotheses. It does not matter whether one of the two hypotheses under testing is a null hypothesis. Hence, the strength of evidence in favor of the null hypothesis can be quantified, something that is impossible using the p-value in frequentist NHST. Second, the posterior probabilities of the two hypotheses $\mathrm{P}(H_0|D)$ and $\mathrm{P}(H_1|D)$ can be obtained, that are the ultimate goal of statistical tests. Third, Bayesian hypothesis testing provides a means of including other information when assessing the evidence for a hypothesis. Prior information about unknown parameters in the testing is easily incorporated.

Over the past decade, Bayesian hypothesis testing had been applied to facilitating structural damage diagnostics in a probabilistic inference procedure. For example, a probabilistic damage identification method, for example, was established by Jiang and Mahadevan (2008) by implementing Bayesian hypothesis testing on the residuals between the measured responses and predictions from a nonparametric fuzzy wavelet neural network model. This method was then applied to facilitating the updating of the uncertainty (Sankararaman and Mahadevan 2011) and the quantification of damage

(Sankararaman and Mahadevan 2013) in a structural frame and a hydraulic actuation system. Recently, Bayesian hypothesis testing was proposed by Wang et al. (2018) to establish a fully Bayesian inference framework for the identification, localization and quantification of damage in the railway turnout in operation. The novel diagnostic logic enables the damage identification and quantification in a probabilistic sense, while it is difficult to indicate the associated diagnostic risk.

To illustrate the Bayesian hypothesis testing procedure for structural damage diagnostics, we consider the same hypotheses made in Equation (6.5). Different from frequentist NHST, the unknown mean parameter μ_ξ in Bayesian hypothesis testing is considered as a random variable. For the standardized Bayesian residuals, a prior distribution is placed on the unknown mean parameter μ_ξ in the alternative hypothesis H_1,

$$\mathrm{P}(\mu_\xi) = \mathcal{N}(0, \tau^2) \tag{6.8}$$

where τ is the standard deviation of the prior distribution, chosen to reflect the believed plausible range of μ_ξ if the alternative hypothesis H_1 were true. Then, the Bayes factor is a ratio of the marginal likelihood of the null hypothesis H_0 and the alternative hypothesis H_1, given by

$$
\begin{aligned}
BF_{01} &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)}{\mathrm{P}(\boldsymbol{\xi}|H_1)} \\
&= \frac{\mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi = 0\right)}{\int \mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi\right)\mathrm{P}\left(\mu_\xi\right)d\mu_\xi} \\
&= \frac{\prod_{s=1}^{S}(2\pi)^{-1/2}\exp(-\xi_s^2/2)}{\int_{-\infty}^{+\infty}(2\pi\tau^2)^{-1/2}\exp\left(-\mu_\xi^2/2\tau^2\right)\prod_{s=1}^{S}(2\pi)^{-1/2}\exp(-\xi_s^2/2)\,d\mu_\xi} \\
&= \sqrt{1+S\tau^2}\exp\left[-\frac{1}{2}z^2/(1+1/S\tau^2)\right]
\end{aligned}
\tag{6.9}
$$

where $z = \sqrt{S}\left|\bar{\xi}\right|$ is the same as the usual test statistic for the standard normal random variable in frequentist NHST and the subscript of the Bayes factor BF_{01} identifies which hypotheses are being compared, and the order denotes which hypothesis is in the numerator and which is in the denominator. Such a testing procedure is often known as Bayesian point null hypothesis testing (Berger and Sellke 1987, Berger and Delampady 1987, Aitkin et al. 2005). Due to the fact that the Bayes factor is always positive, it is more useful to consider twice the natural logarithm of the Bayes factor that is on the same scale the familiar deviance and likelihood ratio test statistics (Jeffreys 1961, Kass and Raftery, 1995), given by

$$2\ln(BF_{01}) = \ln(1+N\tau^2) - z^2/[1+1/N\tau^2] \tag{6.10}$$

The Bayes factor has a natural and straightforward interpretation. It is constructed as a rational means to measure the evidence brought by the data in favor of the null hypothesis relative to the alternative hypothesis. A decision about which hypothesis to select is then based on the numerical value of BF_{01}, the default boundary between null and alternative being $BF_{01} = 1$, as the data then brings the same evidence in favor of both hypotheses.

One of the most common interpretations of the Bayes factor is shown in Table 6.2, first proposed by Jeffreys in 1961 and slightly modified by Kass and Raftery in 1995.

Table 6.2 Interpretation of Bayes factors

BF_{01}	2ln (BF_{01})	Strength of Evidence
>150	>10	Very strong evidence for H_0
20~150	6~10	Strong evidence for H_0
3~20	2~6	Positive evidence for H_0
1~3	0~2	Not worth more than a bare mention evidence for H_0
1	0	No evidence
1/3~1	-2~0	Not worth more than a bare mention evidence for H_1
1/20~1/3	-6~-2	Positive evidence for H_1
1/150~1/20	-10~-6	Strong evidence for H_1
<1/150	<-10	Very strong evidence for H_1

While the Bayes factor is useful to quantify the strength of evidence, one should keep in mind that the posterior probabilities of the two hypotheses are our ultimate goal. To derive the posterior probabilities of the null and alternative hypotheses, we are required to define prior probabilities over the hypothesis H_0 and H_1 being true, denoted $\mathrm{P}(H_0)$ and $\mathrm{P}(H_1)$, respectively. As there are only two hypotheses in the testing, the two priors hold that

$$\mathrm{P}(H_0) + \mathrm{P}(H_1) = 1 \tag{6.11}$$

The prior odds ratio is given by

$$\pi_{01} = \frac{\mathrm{P}(H_0)}{\mathrm{P}(H_1)} \tag{6.12}$$

Using Bayes' law, the posterior probability of the null hypothesis p_0 is given by

$$\begin{aligned} p_0 = \mathrm{P}(H_0|\boldsymbol{\xi}) &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)\mathrm{P}(H_0)}{\mathrm{P}(\boldsymbol{\xi})} \\ &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)\mathrm{P}(H_0)}{\mathrm{P}(\boldsymbol{\xi}|H_0)\mathrm{P}(H_0) + \mathrm{P}(\boldsymbol{\xi}|H_1)\mathrm{P}(H_1)} \\ &= \frac{BF_{01}\mathrm{P}(H_0)}{BF_{01}\mathrm{P}(H_0) + \mathrm{P}(H_1)} \\ &= \frac{BF_{01}\pi_{01}}{BF_{01}\pi_{01} + 1} \end{aligned} \tag{6.13}$$

and the posterior probability of the alternative hypothesis p_1 is given by

$$\begin{aligned} p_1 = \mathrm{P}(H_1|\boldsymbol{\xi}) &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_1)\mathrm{P}(H_1)}{\mathrm{P}(\boldsymbol{\xi})} \\ &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_1)\mathrm{P}(H_1)}{\mathrm{P}(\boldsymbol{\xi}|H_0)\mathrm{P}(H_0) + \mathrm{P}(\boldsymbol{\xi}|H_1)\mathrm{P}(H_1)} \\ &= \frac{\mathrm{P}(H_1)}{BF_{01}\mathrm{P}(H_0) + \mathrm{P}(H_1)} \\ &= \frac{1}{BF_{01}\pi_{01} + 1} \end{aligned} \tag{6.14}$$

The posterior probabilities of the null and alternative hypotheses explicitly derived in Bayesian hypothesis testing, providing us with the direct evidence in favor of the two hypotheses, that is impossible in frequentist NHST.

When the two hypotheses H_0 and H_1 are equally probable, we may have

$$\mathrm{P}(H_0) = \mathrm{P}(H_1) = 0.5 \tag{6.15}$$

and the prior odds ratio is

$$\pi_{01} = 1 \tag{6.16}$$

Then, the posterior probability of the null hypothesis p_0 is reduced to

$$p_0 = \mathrm{P}(H_0|\boldsymbol{\xi}) = \frac{BF_{01}}{BF_{01} + 1} \tag{6.17}$$

and the posterior probability of the alternative hypothesis p_1 becomes

$$p_1 = \mathrm{P}(H_1|\boldsymbol{\xi}) = \frac{1}{BF_{01} + 1} \tag{6.18}$$

Bayesian PNHT has been increasingly advocated to develop a probabilistic framework for structural damage diagnostics as it enables us not only to identify potential damage in the concerned structure (that can be also attained in frequentist NHST), but also to quantify the extent of damage in the concerned structure (that cannot be attained in frequentist NHST), in terms of Bayes factor and posterior probabilities of structural conditions. Despite the fact that the diagnostic risk associated with Bayesian PNHT is not be explicitly available, the likelihood of the data is averaged over the prior of the unknown in the alternative hypothesis and thus, if the prior is well defined, the damage diagnostic results should be much more robust than those from frequentist NHST, which uses an significance level α, empirically prespecified.

While the merits of Bayesian hypothesis testing have been reported repeatedly (Kass and Raftery, 1995, Morey et al. 2016, Wagenmakers et al. 2018), the Bayes factor is very sensitive to the defined prior in the alternative hypothesis. As $\tau^2 \to +\infty$ or $S \to +\infty$, it is immediate that $2\ln(BF_{01}) \to +\infty$, $BF_{01} \to +\infty$ and $\mathrm{P}(H_0|\boldsymbol{\xi}) \to 1$, providing overwhelming evidence for H_0, even though the usual test statistic $z = \sqrt{S}\bar{\xi}$ for the standard normal random variable in frequentist NHST was any large value. This indicates that the testing result always supports the null hypothesis H_0, regardless whether or not it holds true (in that case, the concerned structure would be diagnosed as healthy no matter

whether it is undamaged or damaged). This gives rise to the so-called Jeffreys-Lindley paradox (Lindley 1957, Bartlett 1957, Berger and Sellke 1987, Robert 1993, Cousins 2017), in the sense that Bayesian and non-Bayesian methods can reach quite different conclusions in testing. The Jeffreys-Lindley paradox illustrates a counterintuitive situation in statistical inference where frequentist and Bayesian approaches to a hypothesis testing problem give inconsistent results. When testing structural health state in the light of monitoring data, we may find that the diagnostic results from frequentist and Bayesian approaches can be at odds. The structure of concern could be diagnosed as damaged by the frequentist approach, while be diagnosed as undamaged by the Bayesian approach. As such, we will be at a loss as to what the true health state of the structure of concern is.

Some statisticians have argued that an arbitrarily diffuse prior is not appropriate for the alternative hypothesis as the marginal likelihood of the alternative hypothesis is the weighted average of the likelihood over all possible point hypotheses, where the prior serves as the weight. As τ^2 is increased, a greater relative weight is assumed on larger values of μ_ξ, that are impossible. Unreasonably large values of μ_ξ in the alternative hypothesis reversely provide increased support for the null hypothesis H_0. When these unreasonably large values of μ_ξ have increasing weight, the average favors the null to a greater extent. As a consequence, specifications of alternatives that weight unreasonably

large values of μ_ξ heavily will yield the Bayes factor that too heavily supports the null hypothesis.

A reasonable setting is $\tau^2 = \sigma^2$, giving a unit-information prior on the unknown μ_ξ in the alternative hypothesis, that has been extensively employed in the context of probabilistic damage detection algorithms (Jiang and Mahadevan 2008, Sankararaman and Mahadevan 2013, Wang et al. 2018). The Bayes factor in Equation (6.9) is reduced to

$$BF_{01} = \sqrt{1+S}\exp\left[-\frac{1}{2}z^2/(1+1/S)\right] \tag{6.19}$$

The setting is believed reasonable as the prior distribution $\mathrm{P}(\mu_\xi)$ placed on the unknown μ_ξ in the alternative hypothesis H_1 does not include much mass on highly implausible regions. With this setting, the prior $\mathrm{P}(\mu_\xi)$ has only a small amount of information in the alternative hypothesis. However, it still can be found that as $S \to +\infty$, $2\ln(BF_{01}) \to +\infty$, $BF_{01} \to +\infty$ and $\mathrm{P}(H_0|\boldsymbol{\xi}) \to 1$. This indicates that the Jeffreys-Lindley paradox cannot be fully avoided in the current Bayesian PNHT by simply adjusting the prior. As a result, one has to make use of the customized Bayes factor according to the sample size to be tested (Sellke et al. 2001, Rouder et al. 2009).

6.3.3 Bayesian null hypothesis significance testing

To overcome the so-called Jeffreys-Lindley paradox in Bayesian PNHT, a novel type

of Bayesian hypothesis testing, termed Bayesian null hypothesis significance testing is proposed in this section.

a) ***Simple alternative hypothesis***

Different from the alternative hypothesis that is composite, made in frequentist NHST and Bayesian PNHT, a simple point alternative hypothesis is proposed in Bayesian NHST, in the light of the idea proposed by West and Harrison (1997), given by

$$\begin{aligned} H_0&: \mu_\xi = 0 \ \text{(the concerned structure is undamaged)} \\ H_1 &= \begin{cases} H_1^{-h}: \mu_\xi = -h \ \text{or} \\ H_1^{+h}: \mu_\xi = +h \end{cases} \ \text{(the concerned structure is damaged)} \end{aligned} \tag{6.20}$$

where h is the chosen mean shift value, describing the expected difference between the population means of the standardized Bayesian residuals when the concerned structure is respectively in its undamaged state and damaged state. The shift value h is related to the false positive diagnostic risk α, which will be discussed later. The choice of h reflects the understanding of professionals regarding the presumed discrepancies in the features for classifying structures with and without damage. Its choice is essentially a balance between the false positive and false negative error rates. If the value of h is smaller, a higher rate of false alarm of damage on structures occurs; while a larger value of h corresponds to a higher rate of missing alarm. For different application domains, one may use a customized value of h, according to the total loss caused by the two types of errors. The shift value h is assumed to have the same absolute value as the standardized residuals

are the standard normal variables. The concept of shift can be still effective for multiple dimensions by using a shift vector. The alternative H_1 is composed of two potential cases, with a negative mean shift value $-h$ and a positive mean shift value $+h$. Typically, the alternative hypothesis, either H_1^{-h} or H_1^{+h}, with a higher probability should be preferred and thus we have

$$\mathrm{P}(\boldsymbol{\xi}|H_1) = \max\left\{\mathrm{P}\left(\boldsymbol{\xi}|H_1^{-h}\right), \mathrm{P}\left(\boldsymbol{\xi}|H_1^{+h}\right)\right\} \tag{6.21}$$

where $\boldsymbol{\xi} = [\xi_1, \dots, \xi_S]^{\mathrm{T}}$ are the standardized Bayesian residuals, defined in Section 6.2.2. This type of hypothesis testing is termed Bayesian null hypothesis significance testing because the mean shift value h is directly related to the significance level α (type-I diagnostic risk) which will be discussed later in this section. It can be also called Bayesian simple hypothesis testing (HST) due to the fact that the null and alternative hypotheses are both conditional on a single point hypothesis. Such a hypothesis test was used by Lipowsky et al. (2010) to facilitating the change detection of gas turbine performance and then by Wang (2017) to identify potential outliers in SHM data.

b) ***Intrinsic Bayes factor***

The Bayes factor with a negative mean shift value denoted as BF_{01}^{-h}, is given by

$$\begin{aligned} BF_{01}^{-h} &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)}{\mathrm{P}\left(\boldsymbol{\xi}\middle|H_1^{-h}\right)} \\ &= \frac{\mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi = 0\right)}{\mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi = -h\right)} \\ &= \frac{\prod_{s=1}^{S}(2\pi)^{-1/2}\exp(-\xi_s^2/2)}{\prod_{s=1}^{S}(2\pi)^{-1/2}\exp[-(\xi_s+h)^2/2]} \\ &= \exp\left[Nh\left(h/2+\bar{\xi}\right)\right] \end{aligned} \tag{6.22}$$

where $\bar{\xi} = \frac{1}{S}\sum_{s=1}^{S}\xi_s$. Then, the Bayes factor with a positive mean shift value denoted as BF_{01}^{+h}, is given by

$$\begin{aligned} BF_{01}^{+h} &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)}{\mathrm{P}\left(\boldsymbol{\xi}\middle|H_1^{+h}\right)} \\ &= \frac{\mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi = 0\right)}{\mathrm{P}\left(\boldsymbol{\xi}\middle|\mu_\xi = +h\right)} \\ &= \frac{\prod_{s=1}^{S}(2\pi)^{-1/2}\exp[-\xi_s^2/2]}{\prod_{s=1}^{S}(2\pi)^{-1/2}\exp[-(\xi_s-h)^2/2]} \\ &= \exp\left[Sh\left(h/2-\bar{\xi}\right)\right] \end{aligned} \tag{6.23}$$

Hence, the preferable Bayes factor is given by

$$\begin{aligned} BF_{01} &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)}{\mathrm{P}(\boldsymbol{\xi}|H_1)} \\ &= \frac{\mathrm{P}(\boldsymbol{\xi}|H_0)}{\max\left\{\mathrm{P}\left(\boldsymbol{\xi}\middle|H_1^{-h}\right), \mathrm{P}\left(\boldsymbol{\xi}\middle|H_1^{+h}\right)\right\}} \\ &= \min\{BF_{01}^{-h}, BF_{01}^{+h}\} \\ &= \exp\left[Sh(h/2-|\bar{\xi}|)\right] \end{aligned} \tag{6.24}$$

Twice the natural logarithm of the Bayes factor in Bayesian NHST is given by

$$2\ln(BF_{01}) = Sh(h-2|\bar{\xi}|) \tag{6.25}$$

As shown in Equations (6.24) and (6.25), the Bayes factor BF_{01} in Bayesian NHST remains heavily dependent on the testing sample size S (the number of the standardized Bayesian residuals), which could also cause the so-called Jeffreys-Lindley paradox. Fortunately, it is found that the Bayes factor in Bayesian NHST is easy to be normalized by the sample size S, thus possibly avoiding the Jeffreys-Lindley paradox. By taking the Sth root of the Bayes factor in Equation (6.24). we have

$$IBF_{01} = (BF_{01})^{1/S} = \exp\left[h\left(h/2 - |\bar{\xi}|\right)\right] \tag{6.26}$$

where IBF_{01} is termed the intrinsic Bayes factor, that was coined by Berger and Pericchi (1996) in the investigation of the problem of multiple model comparison and prediction. Two forms of intrinsic Bayes factors were proposed in their paper, with one averaged arithmetically and the other averaged geometrically on the number of potential models to be selected. We prefer the geometric intrinsic Bayes factor as it better follows the law of probability (the probability of intersection). Twice the natural logarithm of the geometric intrinsic Bayes factor is given by

$$2\ln(IBF_{01}) = h\left(h - 2|\bar{\xi}|\right) \tag{6.27}$$

From Equations (6.26) and (6.27), it is observed that the intrinsic Bayes factor and its its twice the natural logarithm are both independent on the sample size S and thus the novel Bayesian NHST does not suffer from the Jeffreys-Lindley paradox. By contrast, the Bayes factor in Bayesian PNHT cannot be easily normalized due to the complicated relationship

between the Bayes factor and the sample size S, as shown in Equations (6.9) and (6.10).

c) *Intrinsic prior*

To derive the posterior probabilities of the two competing hypotheses associated with the intrinsic Bayes factor in geometric average, one has to define two novel priors over the null and alternative hypotheses H_0 and H_1, termed the intrinsic priors, wherein the intrinsic prior of the null hypothesis H_0 is denoted as π_0^I and the intrinsic prior of the alternative hypothesis H_1 is denoted as π_1^I. The two intrinsic priors satisfy

$$\pi_0^I + \pi_1^I = 1 \tag{6.28}$$

The odds ratio of the two intrinsic priors, termed the intrinsic prior odds and denoted as π_{01}^I is given by

$$\pi_{01}^I = \pi_0^I / \pi_1^I = (\pi_{01})^{1/S} \tag{6.29}$$

Based on the above two equations, we have the intrinsic prior probability π_0^I over the null hypothesis H_0

$$\pi_0^I = \frac{\pi_{01}^I}{1 + \pi_{01}^I} \tag{6.30}$$

and the intrinsic prior probability π_1^I over the alternative hypothesis H_1

$$\pi_1^I = \frac{1}{1 + \pi_{01}^I} \tag{6.31}$$

It should be mentioned that one cannot get $\pi_0^I = [\pi_0]^{1/S}$ and $\pi_1^I = [\pi_1]^{1/S}$, necessarily.

d) *Intrinsic posterior*

Similarly, we denote the intrinsic posterior probability of the null hypothesis H_0 as

p_0^I and the intrinsic posterior probability of the alternative hypothesis H_1 as p_1^I, which satisfy

$$p_0^I + p_1^I = 1 \tag{6.32}$$

The odds ratio of the two intrinsic posteriors, termed the intrinsic posterior odds and denoted by p_{01}^I is given by

$$p_{01}^I = p_0^I / p_1^I = (p_{01})^{1/S} \tag{6.33}$$

Thus, the intrinsic posterior probability p_0^I of the null hypothesis H_0 is given by

$$p_0^I = \frac{p_{01}^I}{1 + p_{01}^I} \tag{6.34}$$

and the intrinsic posterior probability p_1^I of the alternative hypothesis H_1 is given by

$$p_1^I = \frac{1}{1 + p_{01}^I} \tag{6.35}$$

Also, one cannot also get $p_0^I = (p_0)^{1/S}$ and $p_1^I = (p_1)^{1/S}$, necessarily. From Equations (6.34) and (6.35), it can be found that the key to derive the intrinsic posterior probability p_0^I of the null hypothesis H_0 and the intrinsic posterior probability p_1^I of the alternative hypothesis H_1 is to obtain the intrinsic posterior odds p_{01}^I of the two hypotheses.

Due to the fact that the Bayes factor BF_{01} is the ratio of the posterior odds p_{01} to its prior odds π_{01} of H_0 and H_1 (Kass and Raftery, 1995), we have

$$p_{01} = BF_{01}\pi_{01} \tag{6.36}$$

Similarly, by taking the Sth root of both sides of the above equation, we have

$$p_{01}^I = (p_{01})^{1/S} = (BF_{01}\pi_{01})^{1/S} = (BF_{01})^{1/S}(\pi_{01})^{1/S} = IBF_{01}\pi_{01}^I \tag{6.37}$$

Interestingly, it can be found that Equations (6.36) and (6.37) share the same form of expression. Thus, the intrinsic Bayes factor IBF_{01} can be stated as the ratio of the intrinsic posterior odds p_{01}^I to its prior odds π_{01}^I of H_0 and H_1.

By substituting the intrinsic posterior odds p_{01}^I with $BF_{01}^I \pi_{01}^I$ in Equations (6.34) and (6.35), we have the intrinsic posterior probability p_1^I of the null hypothesis H_0 as

$$p_0^I = \frac{IBF_{01}\pi_{01}^I}{1 + IBF_{01}\pi_{01}^I} \tag{6.38}$$

and the intrinsic posterior probability p_1^I over the alternative hypothesis H_1 as

$$p_1^I = \frac{1}{1 + IBF_{01}\pi_{01}^I} \tag{6.39}$$

If the two hypotheses H_0 and H_1 are equally probable, we have

$$\mathrm{P}(H_0) = \mathrm{P}(H_1) = 0.5 \tag{6.40}$$

The intrinsic prior odds ratio of the two hypotheses H_0 and H_1 are

$$\pi_{01}^I = (\pi_{01})^{1/S} = [\mathrm{P}(H_0)/\mathrm{P}(H_1)]^{1/S} = 1 \tag{6.41}$$

Therefore, the intrinsic posterior probability p_0^I of the null hypothesis H_0 and the intrinsic posterior probability p_1^I of the alternative hypothesis H_1 are given respectively by

$$p_0^I = \frac{IBF_{01}}{1 + IBF_{01}} \tag{6.42}$$

$$p_1^I = \frac{1}{1 + IBF_{01}} \tag{6.43}$$

In the newly proposed Bayesian NHST, a very clear interpretation of the measure of evidence for supporting the two competing hypotheses is given, in terms of the intrinsic

Bayes factor and intrinsic posterior probabilities. More importantly, the novel intrinsic Bayes factor averaged geometrically on the number of the testing sample and the resulting posterior probabilities of the two competing hypotheses become scale-invariant quantities such that the so-called Jeffreys-Lindley paradox is avoided. This paves the way for practical engineering applications of Bayesian hypothesis testing.

e) ***Diagnostic risk***

The choice of the mean shift value h affects the evaluated value of intrinsic Bayes factor (or its twice the natural logarithm) and thus has a crucial influence on the structural damage diagnostic result. Table 6.3 shows the influence of the mean shift value h on intrinsic Bayes factor and diagnostic results for testing the standardized Bayesian residuals. It can be found that the condition diagnostic result of a structure depends on both the absolute mean value $|\bar{\xi}|$ of the standardized Bayesian residuals and the prespecified mean shift value h. If a larger value of the mean shit h is chosen, it turns out be difficult to reject the null hypothesis H_0 such that a damaged structure may be falsely judged healthy. By contrast, if a smaller mean shit value h is adopted, the null hypothesis H_0 can be easily rejected, giving rising to the type-I diagnostic risk (a healthy structure is inclined to be incorrectly classified damaged).

Table 6.3 Influence of the mean shift value h on intrinsic Bayes factor and diagnostic results

Decision interval	IBF_{01}	$2\ln(IBF_{01})$	Decision
$\lvert\bar{\xi}\rvert < h/2$	IBF_{01}>1	$2\ln(IBF_{01}) > 0$	H_0 is accepted; (The concerned structure is diagnosed as healthy)
$\lvert\bar{\xi}\rvert = h/2$	IBF_{01} =1	$2\ln(IBF_{01}) = 0$	No evidence (More SHM data are required)
$\lvert\bar{\xi}\rvert > h/2$	IBF_{01} <1	$2\ln(IBF_{01}) < 0$	H_0 is rejected (The concerned structure is diagnosed as damaged)

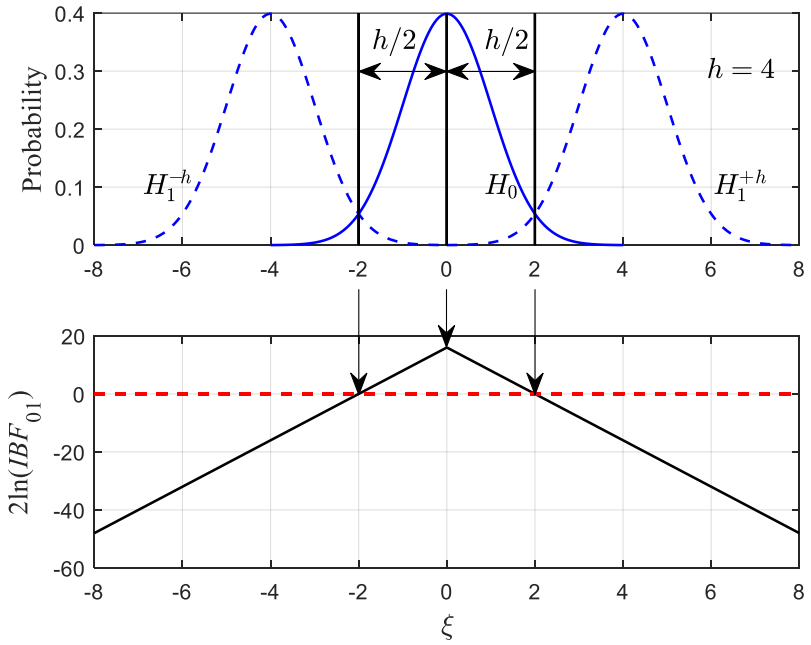


Figure 6.1 Intrinsic Bayes factor ($h = 4$)

Figure 6.1 shows the value of twice the natural logarithm of the intrinsic Bayes factor when the chosen mean shit value h is 4. It can be found that if $|\bar{\xi}| < 2$, we have $2\ln(IBF_{01}) > 0$ and H_0 will be accepted (the concerned structure will be diagnosed as healthy); if $|\bar{\xi}| > 2$, we have $2\ln(IBF_{01}) < 0$ and H_0 will be rejected (the concerned structure will be diagnosed as damaged); if $|\bar{\xi}| = 2$, no decision can be made due to the fact that the data support the two hypotheses equally (in other words, the health condition of the concerned structure is not clear, based on the limited data being tested). To quantify structural health condition in terms of intrinsic Bayes factor in a more specific way, one can refer to Table 6.2.

The value of h is directly related to the significance level α, given by

$$\alpha = 2[1 - \Phi(h/2)] \tag{6.44}$$

where Φ is the cumulative distribution function of the standard normal random variable. The significance level α in Bayesian NHST is a risk-controlling quantity, the same as in classical NHST, referred to as the probability of rejecting the null hypothesis H_0 when it is true. When Bayesian NHST is used to identify structural damage, it is referred to as the probability that a healthy structure is falsely judged damaged.

In the implementation of Bayesian NHST for structural damage diagnostics, the first step that we are required to take is to choose a proper significance level α in order to control the type-I diagnostic risk (the same as in frequentist NHST). Then, given the

significance level α, one can easily derive the corresponding mean shift h in Bayesian NHST as

$$h = 2\Phi^{-1}(1-\alpha/2) \tag{6.45}$$

where Φ^{-1} is the inverse of cumulative distribution function of the standard normal random variable. Typical significance levels and corresponding mean shift values are given in Table 6.4.

Table 6.4 Typical significance levels and corresponding mean shift values

Significance level α	61.71%	31.73%	13.36%	4.55%	1.24%	0.27%
Mean shift value h	1	2	3	4	5	6

6.4 Case Study

The campaign of online condition assessment of railway wheels is served as a case study in this section to compare, in detail, the performance of the proposed unsupervised damage diagnostic methods, based on three categories of population feature models (SSBL, HSBL and PSBL used in the baseline phase) and three categories of damage diagnostic logics (frequentist NHST, Bayesian PNHT and Bayesian NHST used in the inspection phase). A typical 8-car high-speed passenger train equipped with sixty-four undefective wheels is first run on the rail instrumented with the track-side monitoring system that has been described in Chapter 3. Online condition monitoring data are collected to establish a healthy population feature model for all nominally identical undefective wheels. This population feature model is served as a baseline to identify wheels, potentially, with defects in the inspection phase. To obtain the optimal population feature model, three categories of modelling frameworks have been compared based on SSBL (Chapter 3), HSBL (Chapter 4) and PSBL (Chapter 5), respectively. Then, a blind test is conducted by replacing some healthy wheels by defective wheels and running the train equipped with several defective wheels on the same rail. New monitoring data are acquired to validate and compare the performance of the proposed unsupervised damage diagnostics methods of NISs, based on the three different population feature models and the three different diagnostic logics. After the blind test, all tested wheels are delivered to

a maintenance workshop for offline wheel profile measurement. As such, a comparison between the online diagnostic results by the proposed methods and the offline inspection results from wheel profile measurement system can be made.

It is worth noting that the monitoring data collected by a single sensor might be unable to capture the defect-relevant information in case the minor defective tread (e.g. a small flat) didn't roll over the rail section deployed with the sensor; it would result in a false negative if using only the monitoring data from the single sensor. When using the monitoring data from all the deployed sensors, more reliable defect detection results would be obtained as the effect of minor defective tread must be sensed by at least one sensor if the sensors are densely deployed along a rail segment longer than the wheel perimeter. In the following, both the wheel defect detection results by using the monitoring data from a single sensor (SEN-A2 deployed on the left rail track and SEN-D2 deployed on the right rail track) and from all the sensors are presented.

In the implementation of frequentist NHST and Bayesian NHST on the condition assessment of the nominally identical wheels, the chosen significance level is α=1.24%, at price of a false-positive diagnostic error rate being no more than 1.24%. The corresponding mean shift value h in Bayesian NHST is 5, according to Equation (6.45). To derive the intrinsic posterior probabilities of the two hypotheses made for the blindly tested wheels with and without defects, an equal prior probability is assigned to the null

hypothesis H_0 and the alternative hypothesis H_1, with $\mathrm{P}(H_0) = \mathrm{P}(H_1) = 0.5$. When conducting Bayesian PNHT for the condition assessment of these nominally identical wheels, a unit-information prior $\mathrm{P}(\mu_\xi) = \mathcal{N}(0,1)$ is assigned to the unknown parameter associated with the alternative hypothesis. To the posterior probabilities of the blindly tested wheels with and without defects, an equal prior probability is also placed on the null hypothesis H_0 and the alternative hypothesis H_1 in Bayesian PNHT.

6.4.1 Diagnostic results of wheel defects using a single senor

The wheel defect diagnostics is first pursued by using the monitoring data from a single sensor. The sensors SEN-A2 and SEN-D2 deployed respectively on the left and right trail tracks are taken as an example to illustrate the influence of three categories of population feature models (established by SSBL, HSBL and PSBL, respectively) on the defect diagnostic results of nominally identical wheels. The advantages and limitations of the three categories of statistical diagnostic tests on structural damage identification are investigated, including frequentist NHST, Bayesian PNHT and the novel Bayesian NHST proposed in this thesis.

a) ***Frequentist NHST results***

The condition assessment results of left wheels are shown in Figures 6.2 to 6.4 by using frequentist NHST on the monitoring data from the sensor-A2 deployed on the left

rail track, with the population feature models established by SSBL, HSBL and PSBL, respectively. The condition assessment results of the right wheels are shown in Figures 6.5 to 6.7 by using frequentist NHST on the monitoring data form the sensor-D2 installed on the right rail track, with the population feature models built by SSBL, HSBL and PSBL, respectively. The magenta lines in the six figures are the chosen significance level α=1.24%, served as a threshold for classifying wheels with and without defect at a price of the false-positive error rate no more than 1.24%. It is observed that the p-values of the 27th left wheel, and 24th and 27th right wheels are under the magenta line (the case $p < \alpha$ in frequentist NHST) and thus the three left wheels are diagnosed as defective. By contrast, the p-values of other left wheels are all above the chosen significance level (the case $p > \alpha$ in frequentist NHST) and therefore, these left wheels are diagnosed as undefective. It is found that when using different categories of population feature models, the calculated p-values are varied, they have no decisive impact on the ultimate condition classification of these wheels. Besides, the p-values in frequentist NHST are found difficult to quantify the extents of wheel defects as shown in the six figures.

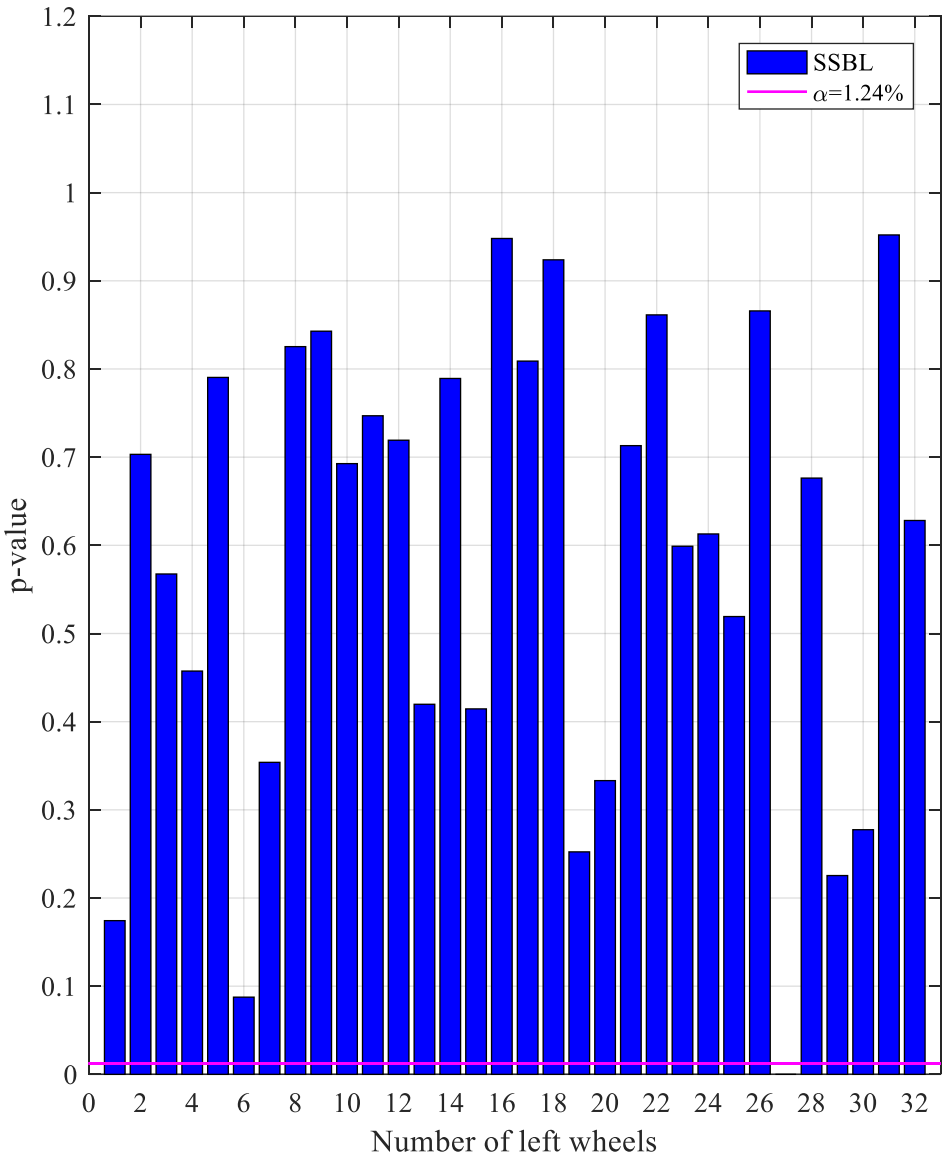


Figure 6.2 p-values of left wheels using monitoring data from SEN-A2 deployed on left rail track (SSBL and Frequentist NHST)

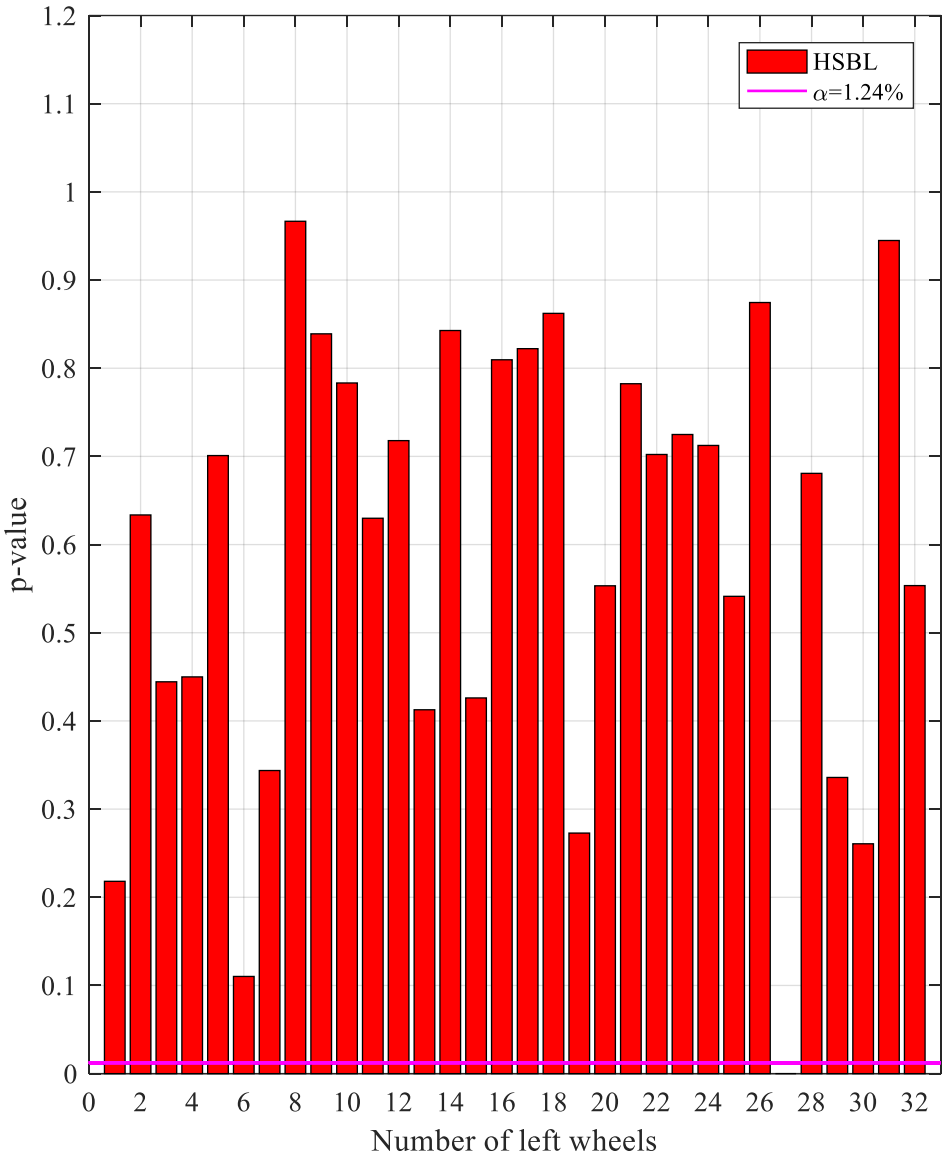


Figure 6.3 p-values of left wheels using monitoring data from SEN-A2 deployed on left rail track (HSBL and Frequentist NHST)

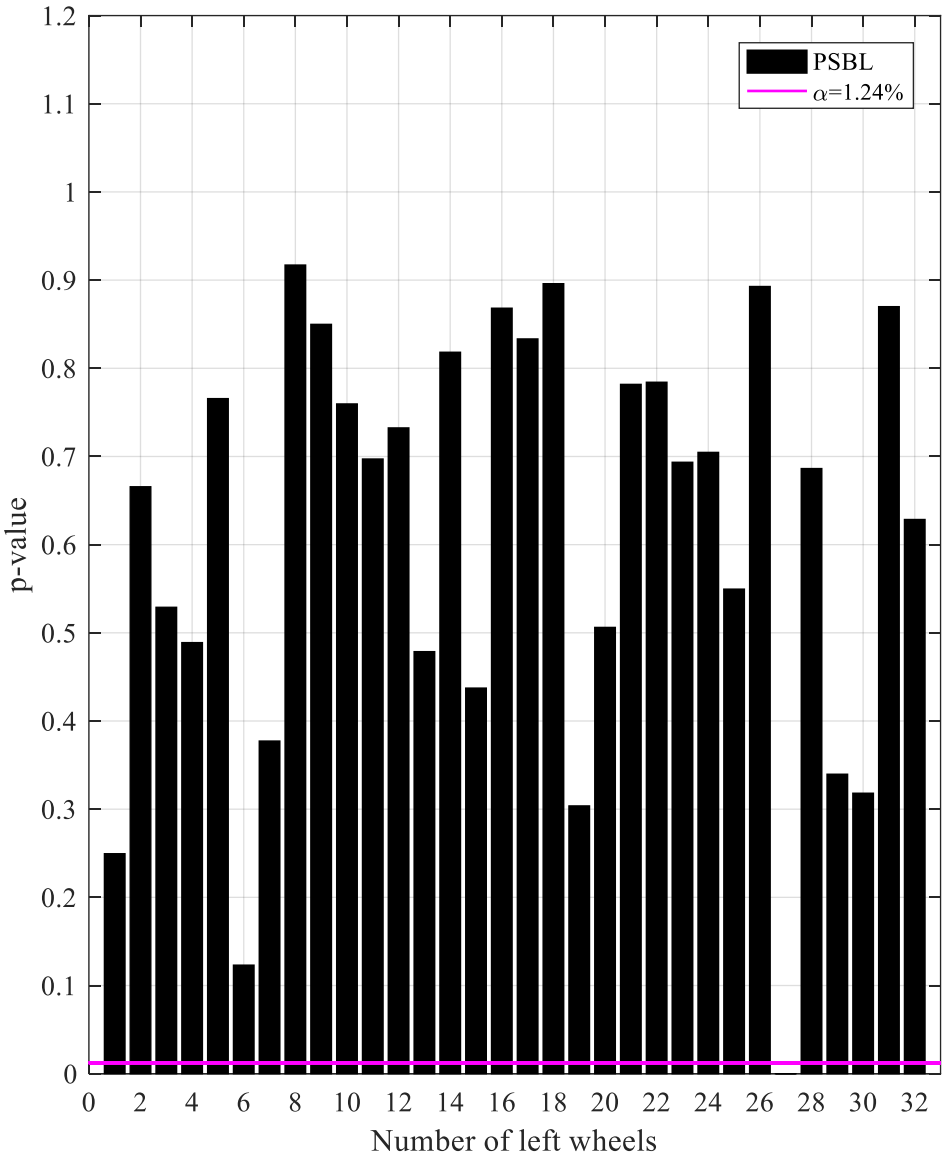


Figure 6.4 p-values of left wheels using monitoring data from SEN-A2 deployed on left rail track (PSBL and Frequentist NHST)

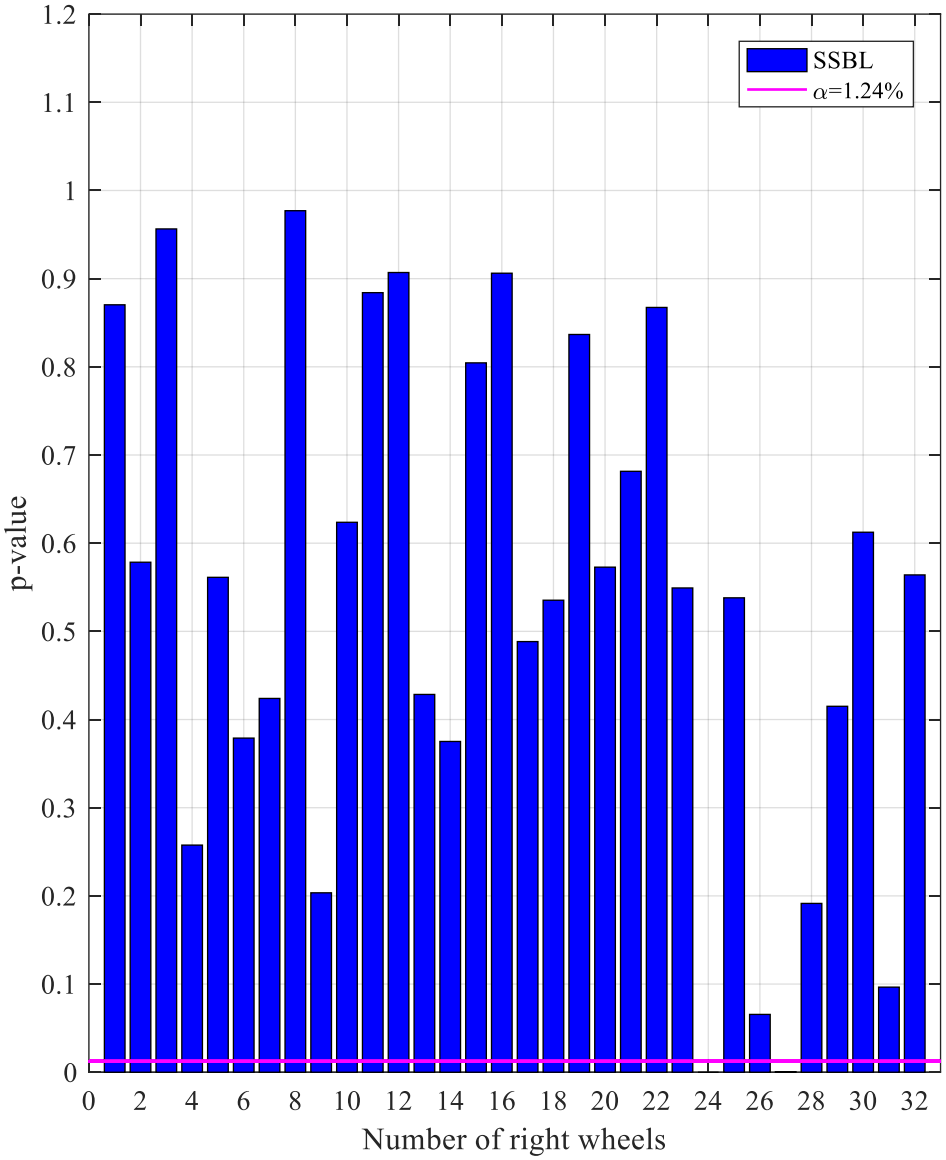


Figure 6.5 p-values of right wheels using monitoring data from SEN-D2 deployed on right rail track (SSBL and Frequentist NHST)

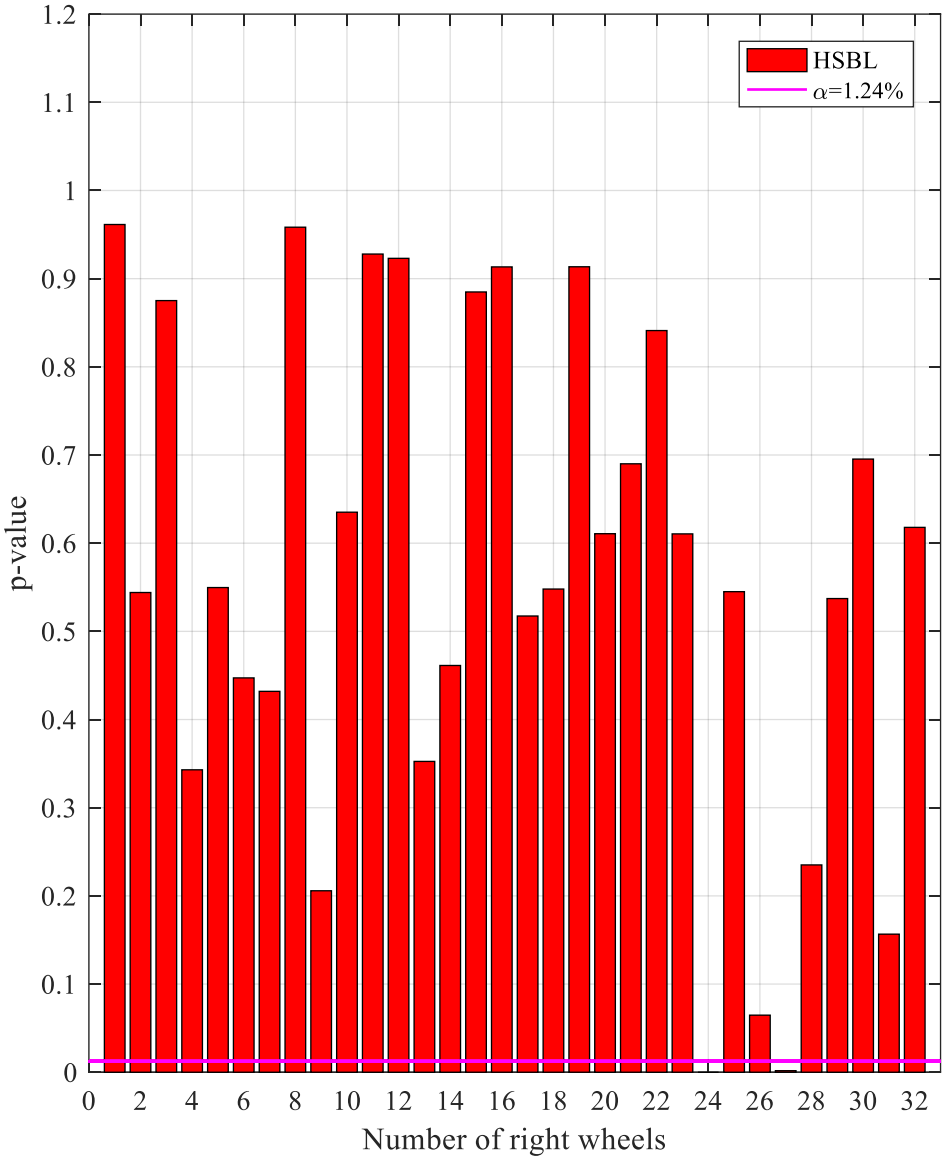


Figure 6.6 p-values of right wheels using monitoring data from SEN-D2 deployed on right rail track (HSBL and Frequentist NHST)

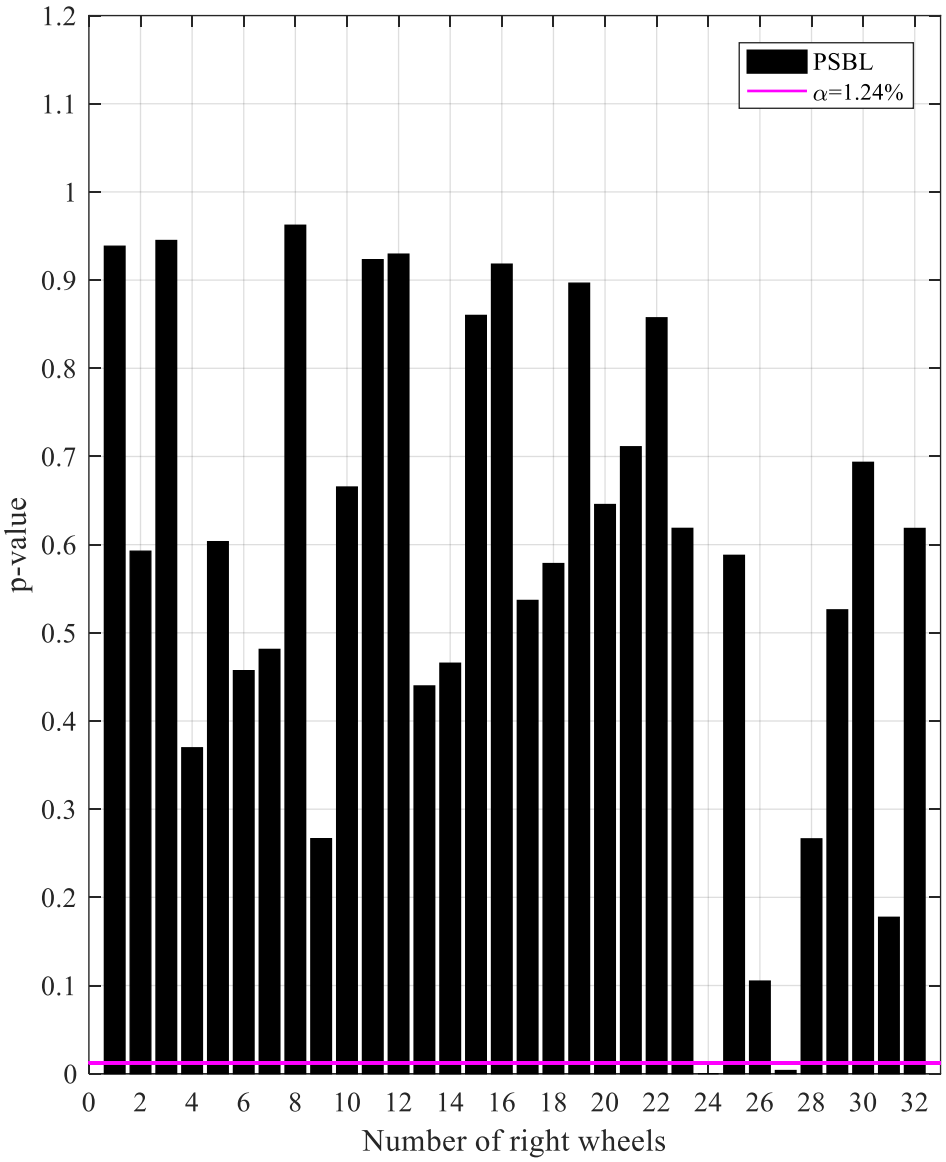


Figure 6.7 p-values of right wheels using monitoring data from SEN-D2 deployed on right rail track (PSBL and Frequentist NHST)

b) ***Bayesian PNHT results***

The condition assessment results of left wheels using Bayesian PNHT on the monitoring data from the sensor SEN-A2 are shown in Figures 6.8 to 6.13, with the population feature models established by SSBL, HSBL and PSBL, respectively. The condition assessment results of right wheels using Bayesian PNHT on the monitoring data from the sensor SEN-D2 are shown in Figures 6.14 to 6.19, with the population feature models established by SSBL, HSBL and PSBL, respectively. The magenta lines in the twelve figures are the thresholds for classifying wheels with and without defects, whereas the resulting diagnostic risk is not clear. The lengths of the bars in Figure 6.9, 6.11, 6.13, 6.15, 6.17 and 6.19 are all one due to the fact that $\Pr(H_0|\mathbf{D}) + \Pr(H_1|\mathbf{D}) = 1$.

When using the SSBL population feature model, the Bayes factor (the Bayes factor in this case study is, by default, referred to as its twice the natural logarithm) for the 27th left wheel is -22.0 (Figure 6.8) with a posterior probability of defect is almost 100% (Figure 6.9). This advocates very strongly that the 27th left wheel is heavily defected. The Bayes factors for the 1st, 6th, 13th, 19th, 20th, 29th and 30th left wheels are all between -2 and 0, and their corresponding posterior probabilities of defect are between 50.0% and 73.1%, providing evidence that these wheels are weakly defected. Other left wheels are diagnosed as undefective because of having Bayes factors being all positive and posterior probabilities of defect being all under 50.0%. Despite the eight left wheels, one more left

wheel (the 3rd left wheel) is diagnosed as defective (weakly) when using the HSBL population feature model, with a Bayes factor being -0.08 (Figure 6.10) and a posterior probability of defect being 51.0% (Figure 6.11). On the contrary, one less left wheel (the 13th left wheel) is diagnosed as defected if using the PSBL population feature model, with its Bayes factor being positive (Figure 6.12) and its posterior probability of defect less than 50.0% (Figure 6.13).

Regardless of which type of population feature models used, the 24th right wheel is diagnosed as seriously defective because its Bayes factors are all less than -10 (Figure 6.14, 6.16 and 6.18) and its posterior probabilities of defect are all more than 99.3% (Figure 6.15, 6.17 and 6.19). The Bayes factors of the 27th right wheels are respectively -7.2, -5.3 and -4.8 based on the SSBL, HSBL and PSBL population feature models, and the corresponding posterior probabilities of defect are respectively 97.3%, 93.4% and 91.7%. Hence, the 27 right wheel is diagnosed respectively as moderately, mildly and mildly defected if the population feature model in the baseline phase is established by SSBL, HSBL and PSBL. For the SSBL and PSBL population feature models, the Bayes factors of the 4th, 6th, 9th, 26th, 28th and 31st right wheels are all between -2 and 0 and their posterior probabilities of defect are between 50.0% and 73.1%, which suggests that the six right wheels are weakly defected. Other right wheels are diagnosed as undefective because of having a Bayes factor being positive and a probability of defect less than 50%.

By contrast, if using the HSBL population feature model, one more right wheel (the 13th right wheel) is diagnosed as defective (weakly), with its Bayes factor being -0.1 and a probability of defect being 51.2%.

In comparison to the diagnostic results from frequentist NHST, in which only the 27th left wheel and the 24th and 27th right wheels are identified as defected, more wheels are diagnosed as defected in Bayesian PNHT. This indicates that the diagnostic results from frequentist NHST and Bayesian PNHT are inconsistent in this case study, triggering the well-known Jeffrey-Lindley paradox in statistical tests. The inconsistency between the frequentist NHST and the Bayesian PNHT diagnostic results could be alleviated to some extent by adjusting the significance level α in frequentist NHST or the prior assigned on the unknown parameter associated with the alternative hypothesis in Bayesian PNHT. However, it can never be eliminated in the current Bayesian hypothesis testing. In addition, it is found that evidence to support the null hypothesis (the wheel is without defect) and the alternative hypothesis (the wheel is with defect) is out of proportion in Bayesian PNHT. For example, there is decisive evidence to support the claim that the 27th left wheel and the 24th right wheel are seriously defected no matter which type of population feature models used. Yet, no overwhelming evidence is found to support that the wheels identified as undefective are definitely without defect.

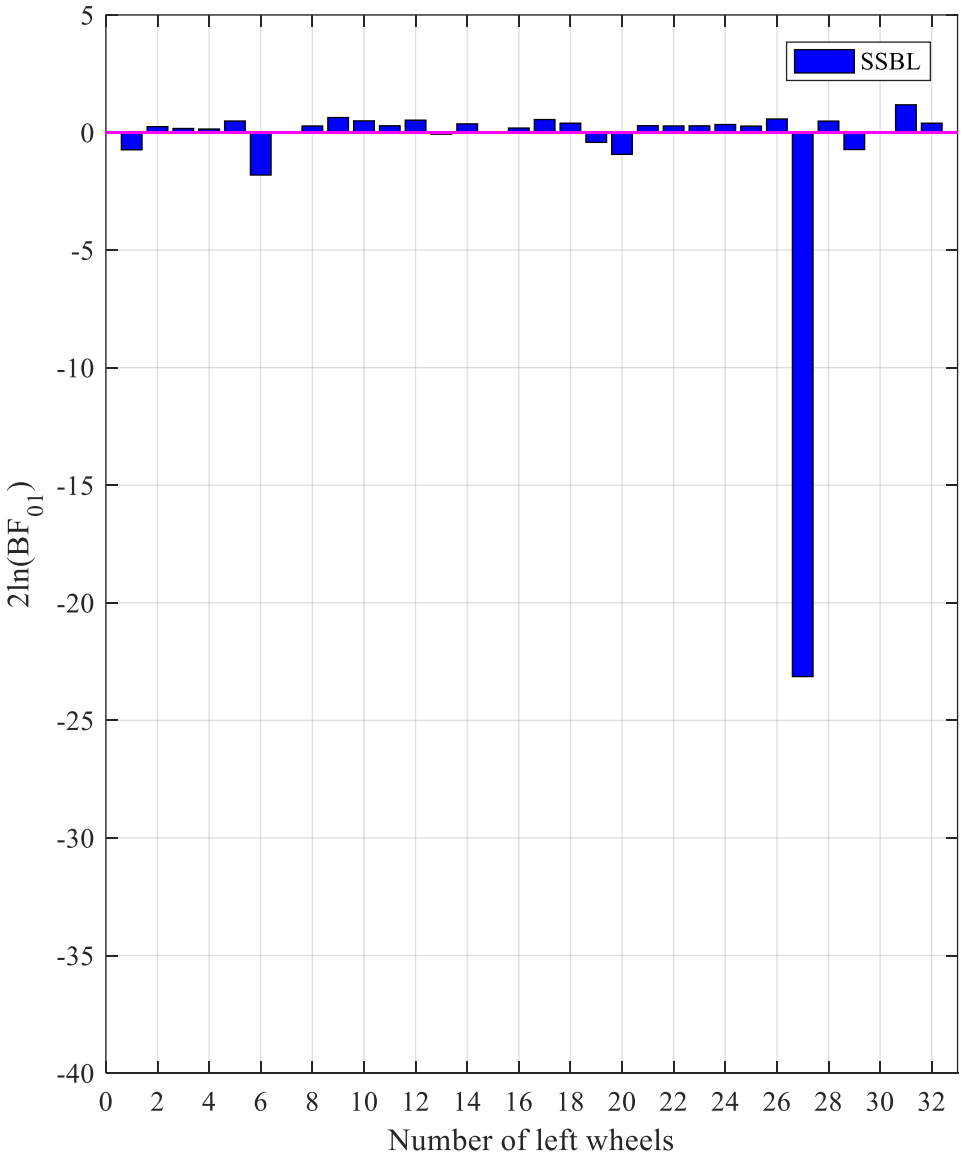


Figure 6.8 Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (SSBL and Bayesian PNHT)

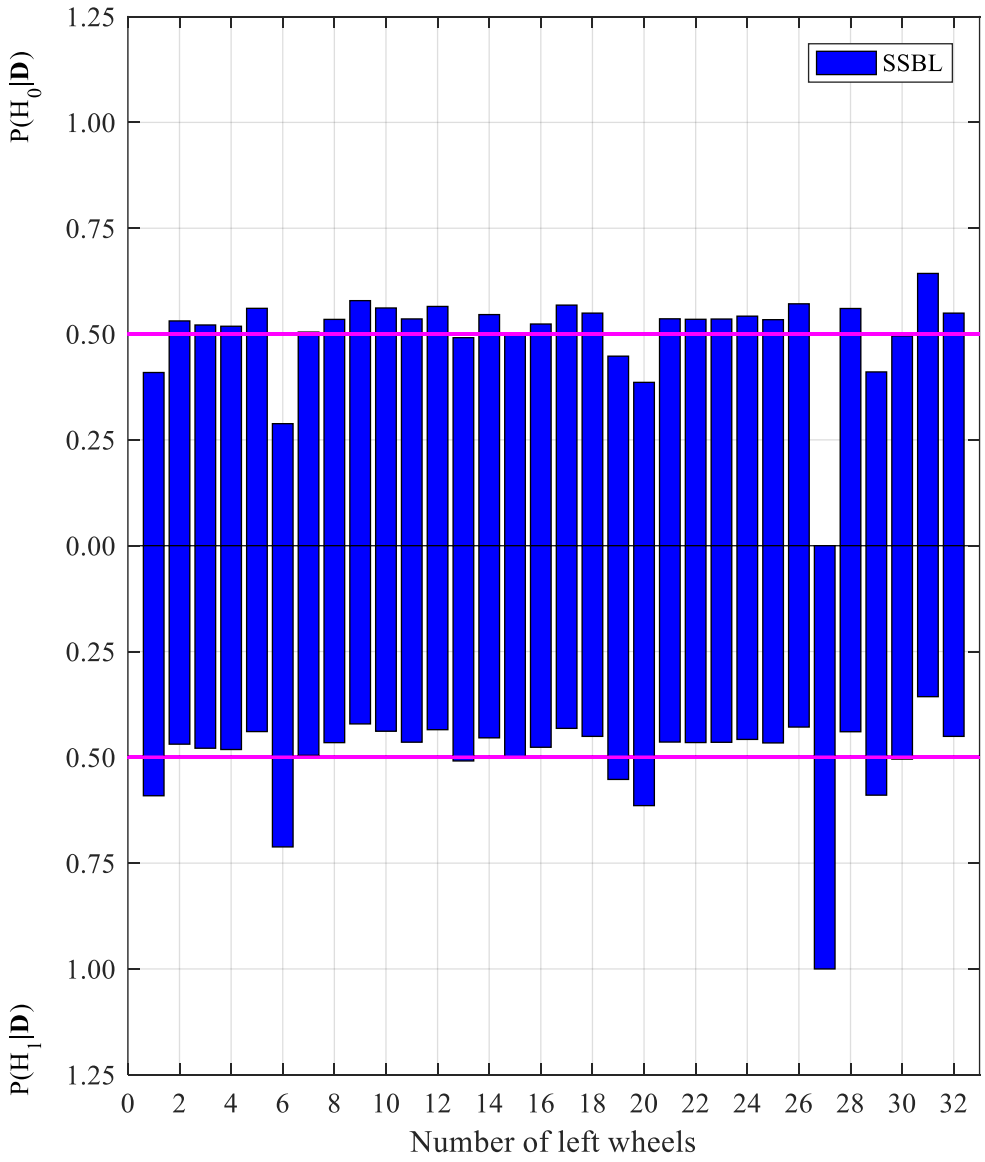


Figure 6.9 Posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (SSBL and Bayesian PNHT)

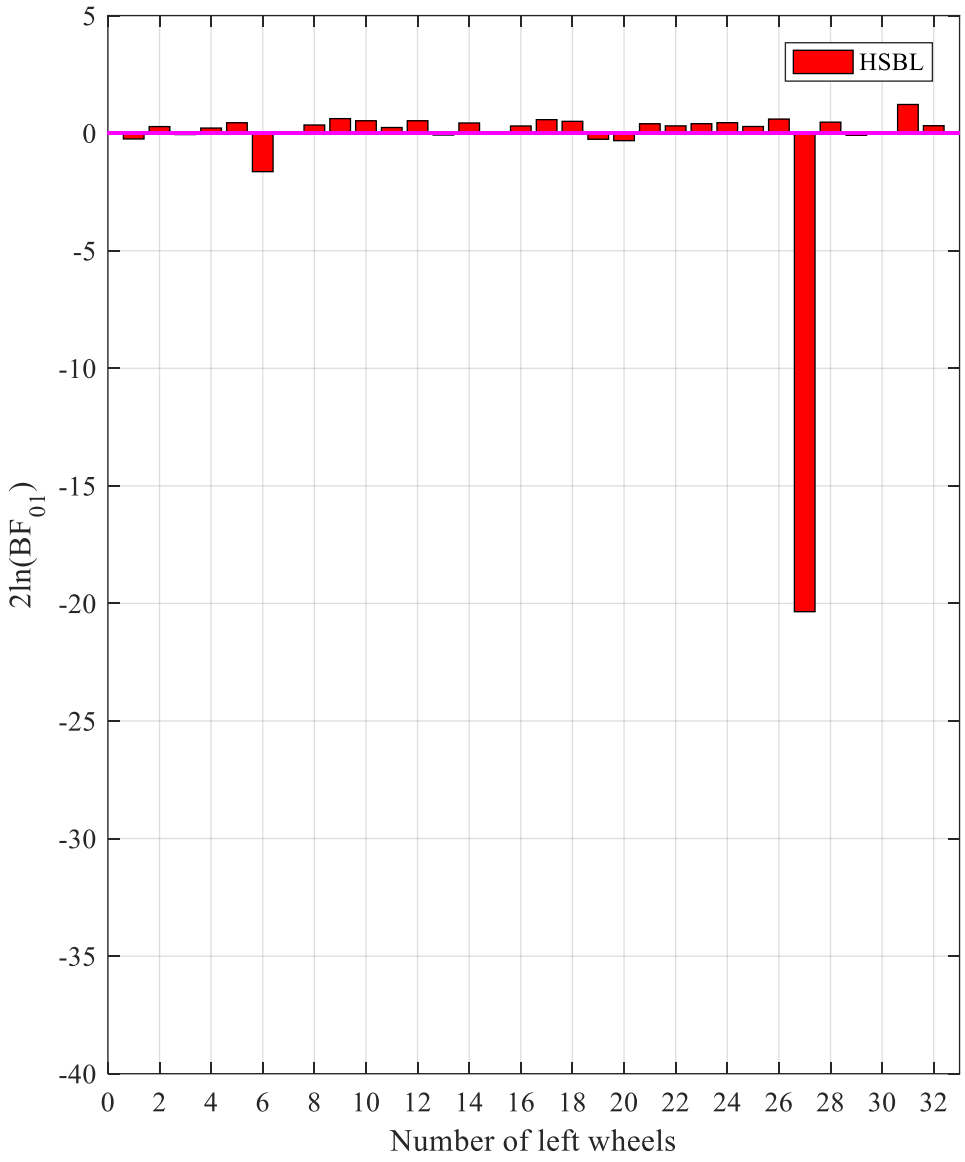


Figure 6.10 Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (HSBL and Bayesian PNHT)

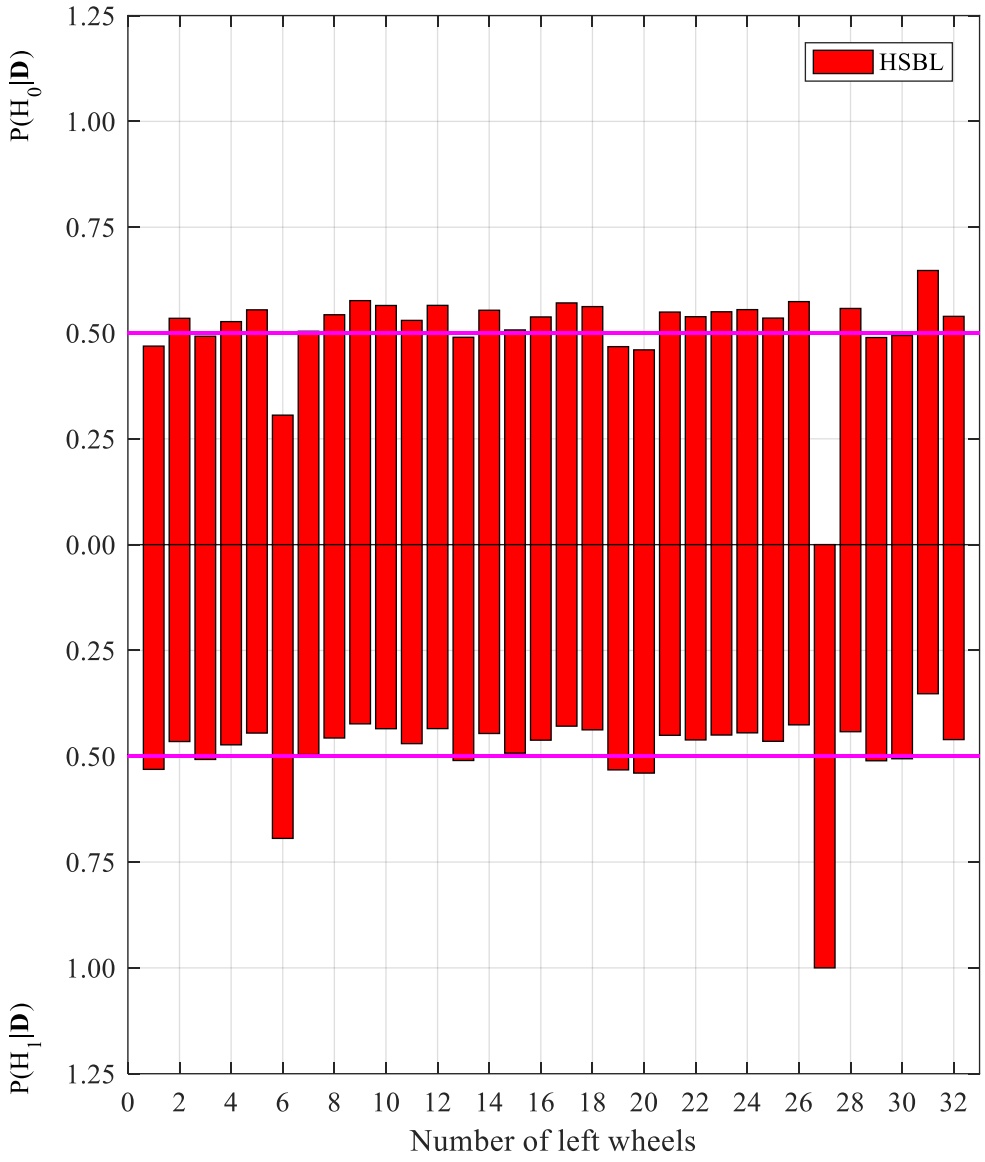


Figure 6.11 Posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (HSBL and Bayesian PNHT)

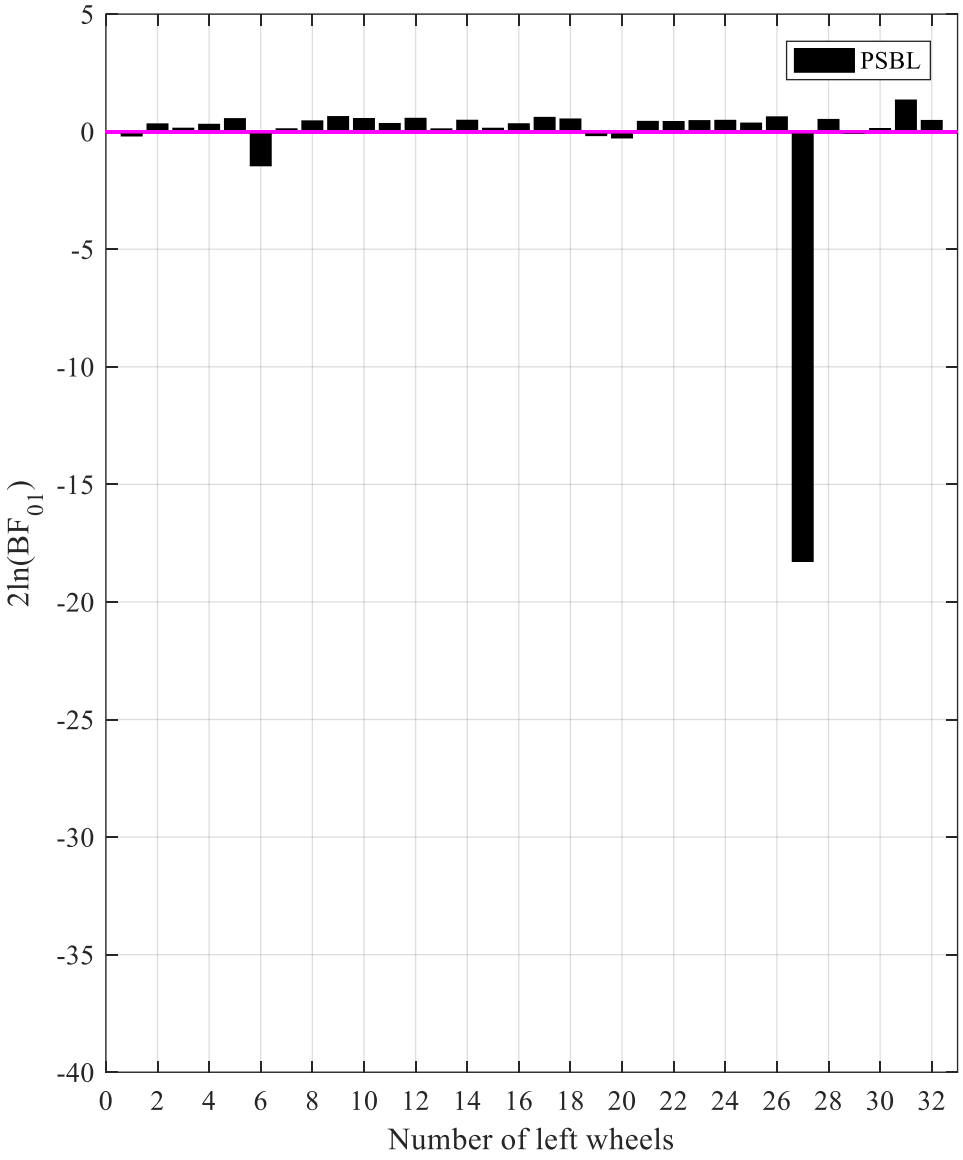


Figure 6.12 Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (PSBL and Bayesian PNHT)

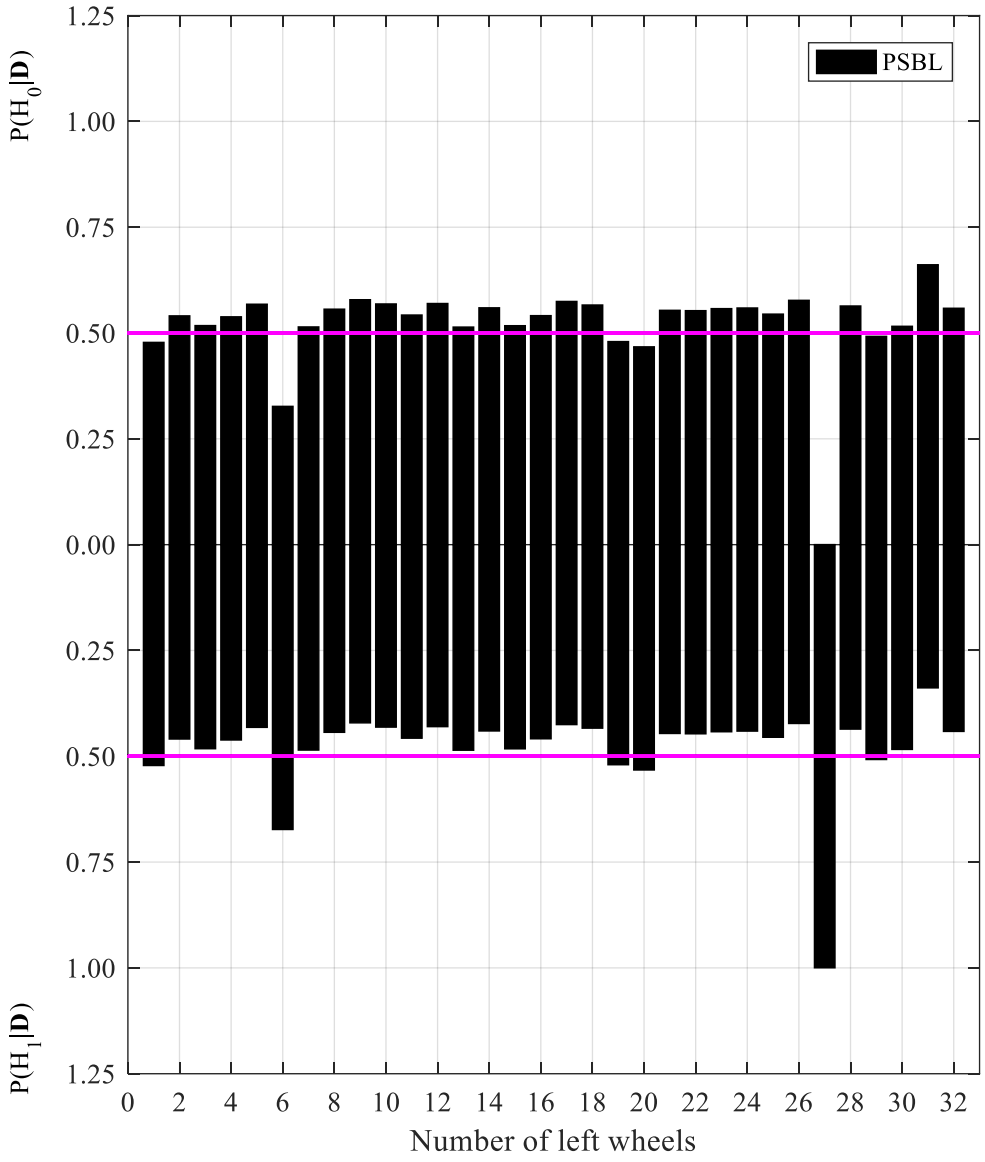


Figure 6.13 Posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (PSBL and Bayesian PNHT)

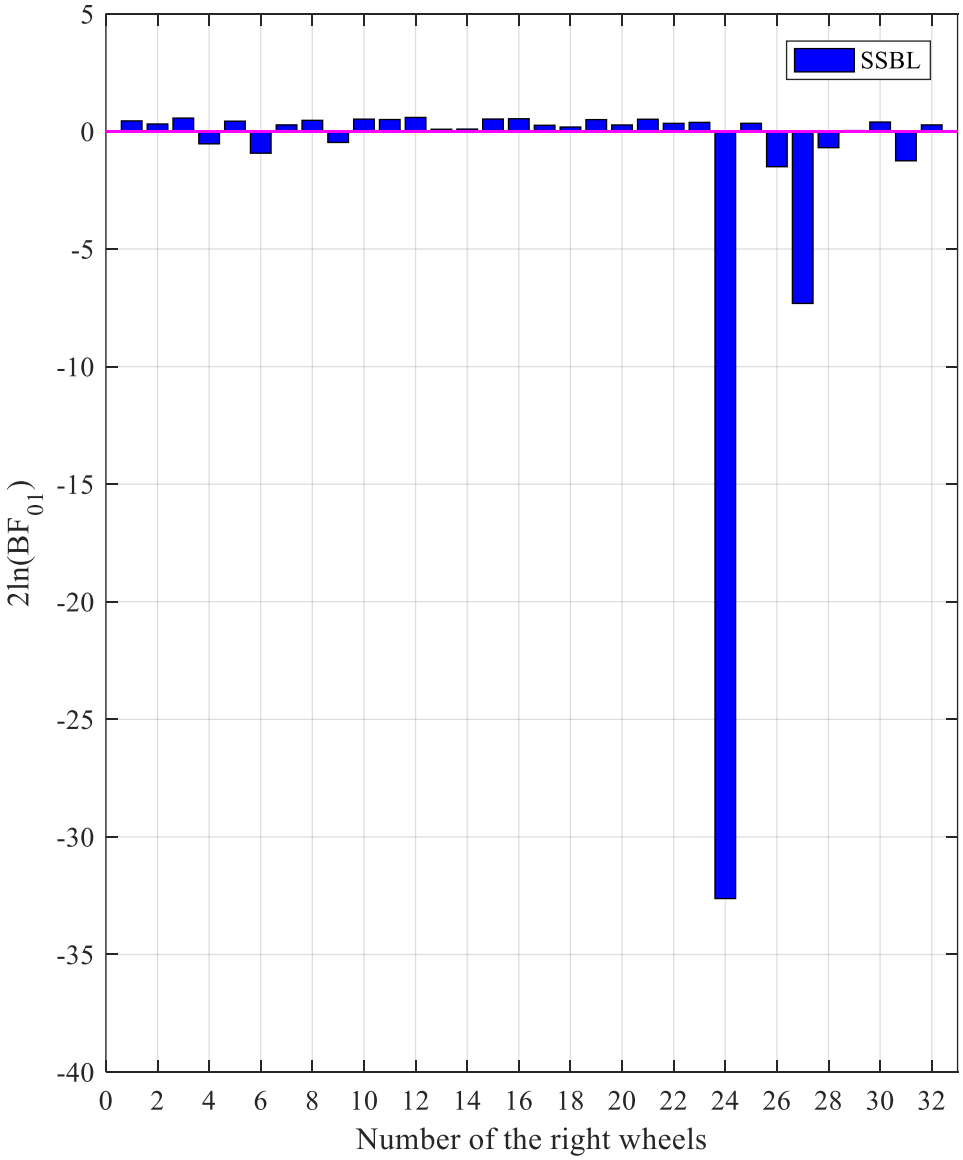


Figure 6.14 Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (SSBL and Bayesian PNHT)

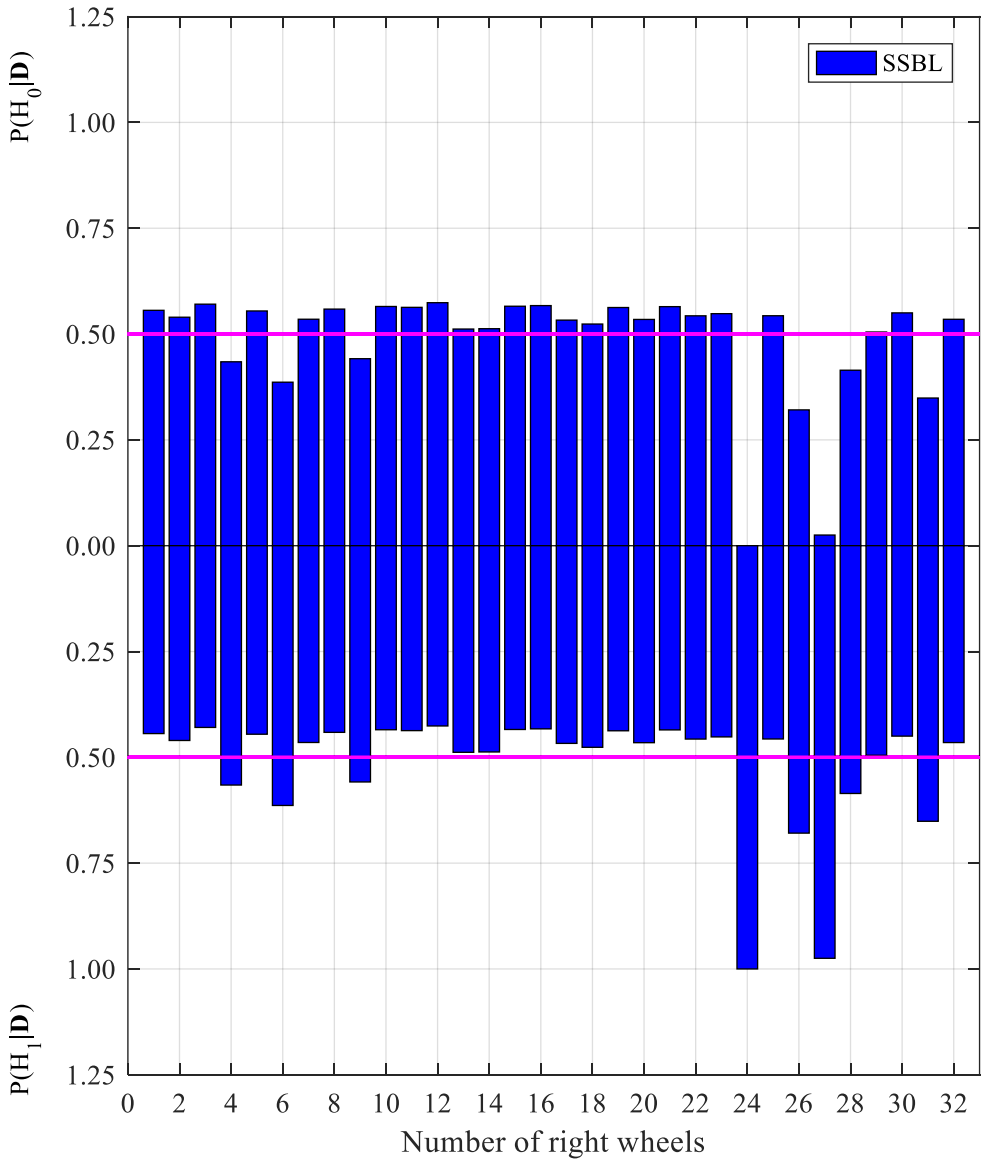


Figure 6.15 Posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (SSBL and Bayesian PNHT)

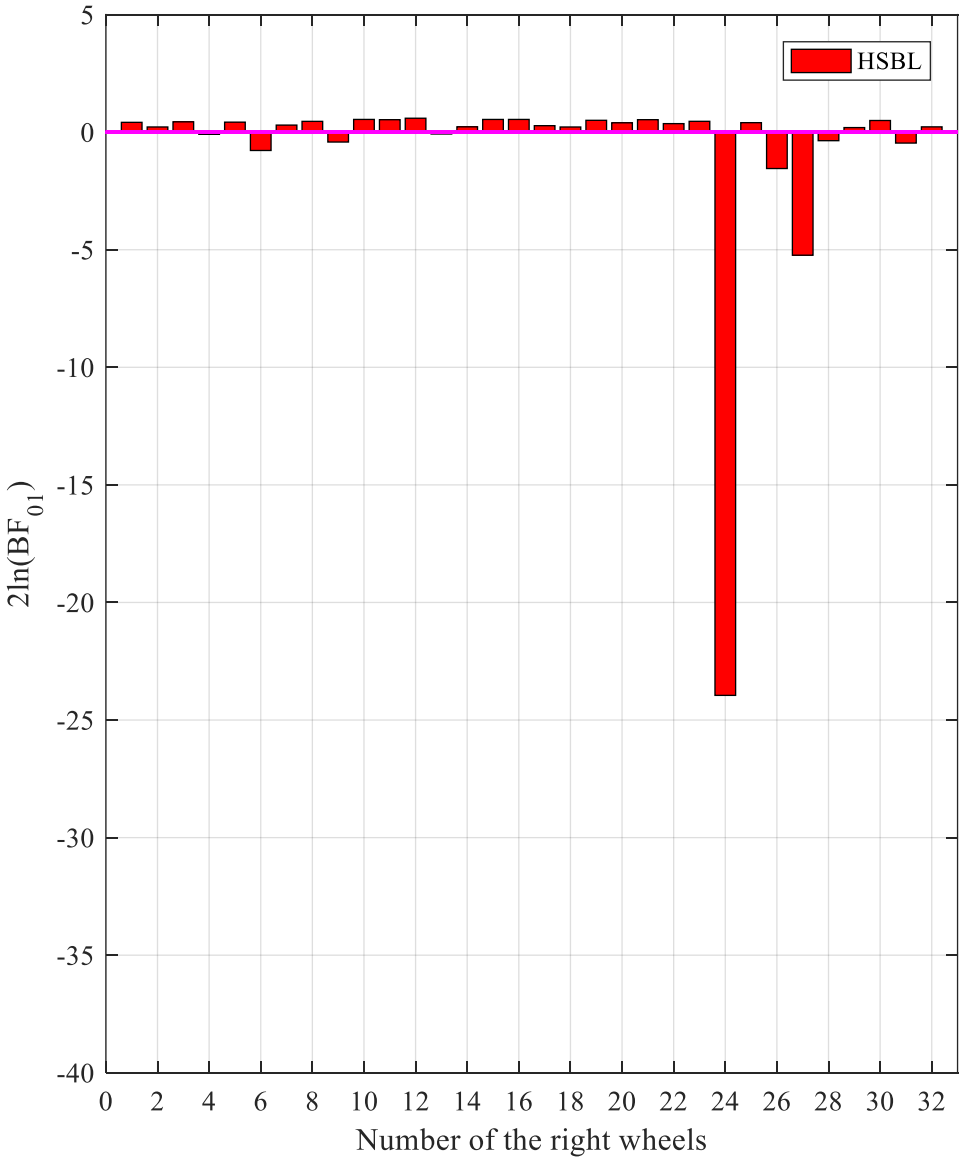


Figure 6.16 Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (HSBL and Bayesian PNHT)

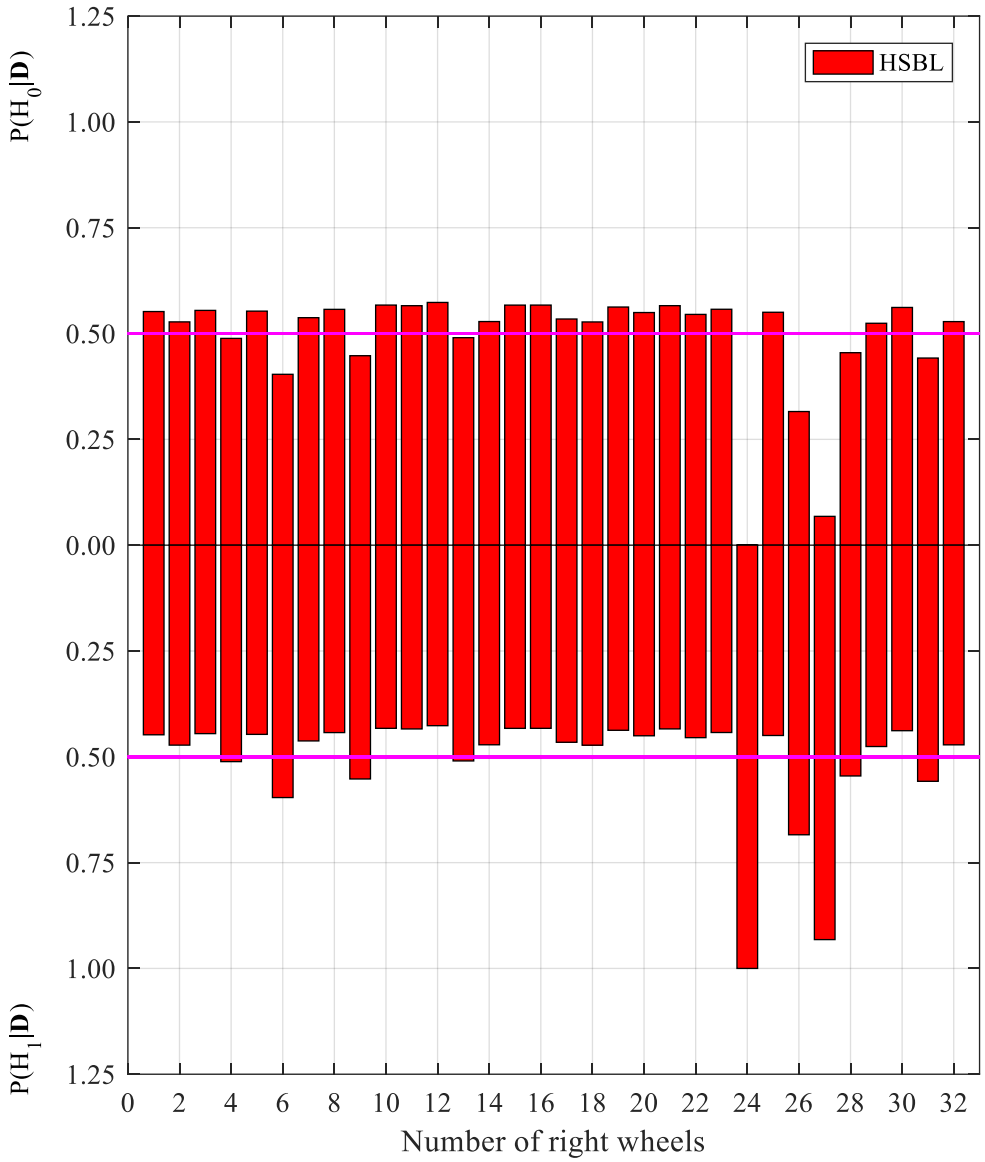


Figure 6.17 Posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (HSBL and Bayesian PNHT)

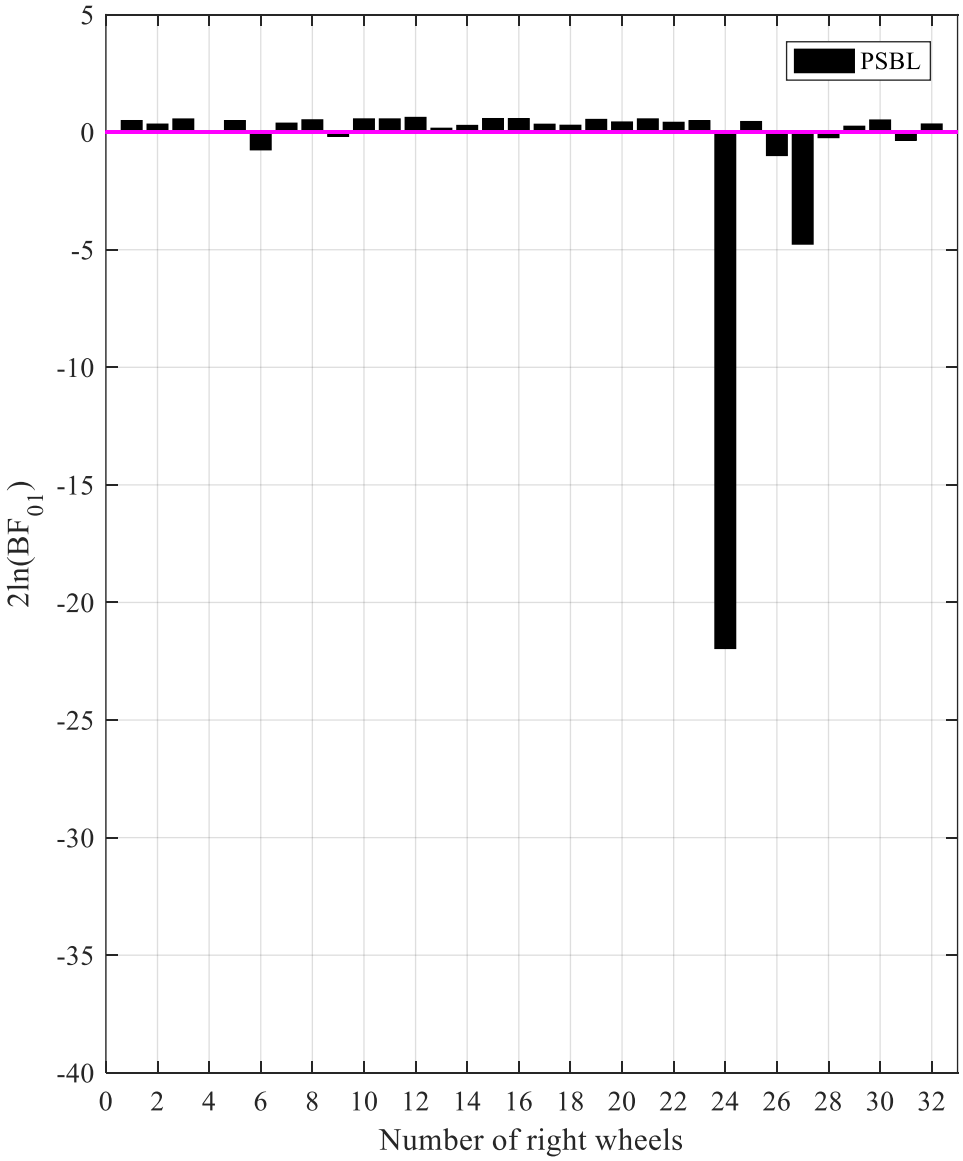


Figure 6.18 Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (PSBL and Bayesian PNHT)

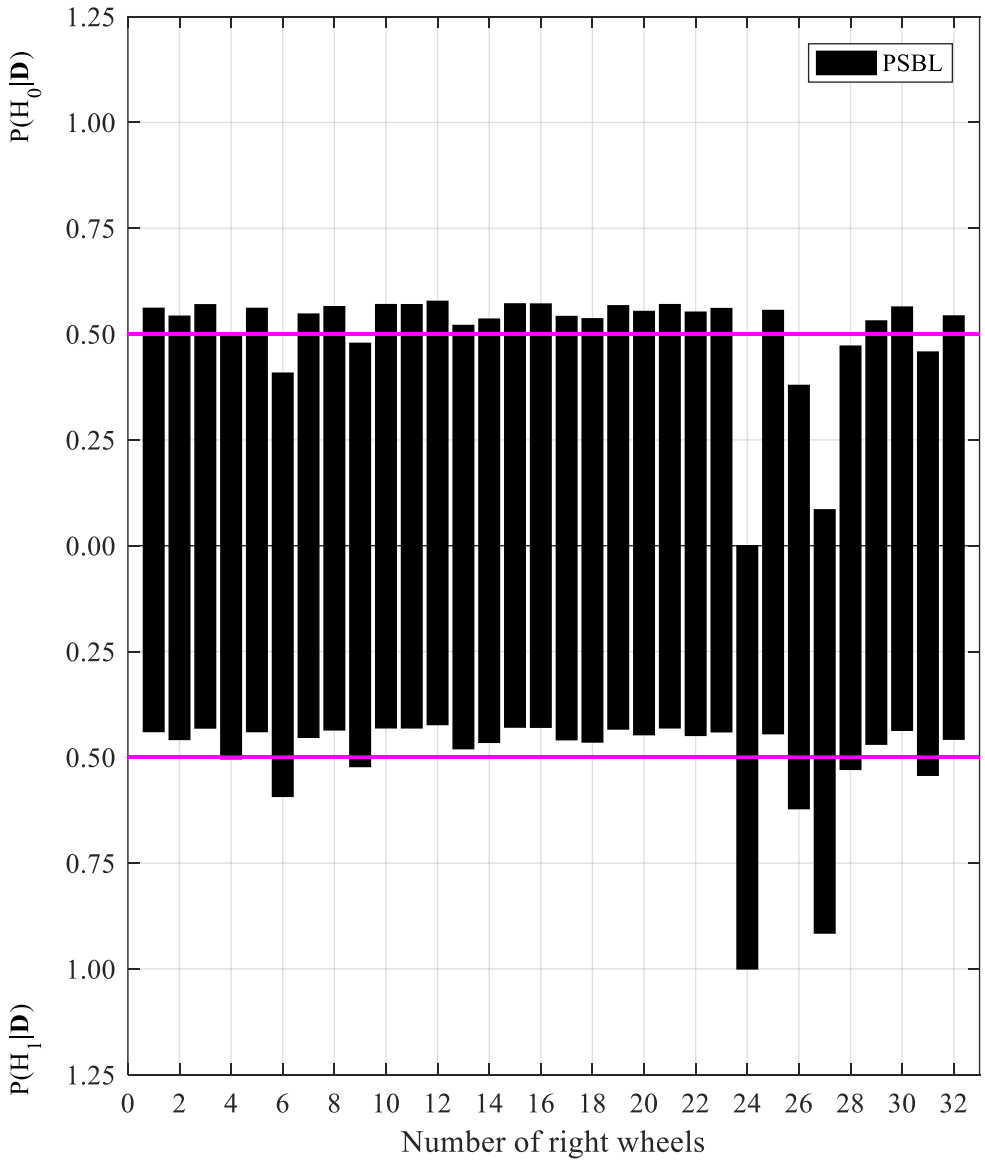


Figure 6.19 Posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (PSBL and Bayesian PNHT)

c) ***Bayesian NHST results***

Figures 6.20 to 6.25 show the diagnostic results for left wheels using Bayesian NHST on the monitoring data from the SEN-A2, with the population feature models built in SSBL, HSBL and PSBL, respectively. Figures 6.26 to 6.31 show the diagnostic results for right wheels using Bayesian NHST on the monitoring data from the SEN-D2, with the population feature models established in SSBL, HSBL and PSBL, respectively. The magenta lines in the twelve figures are the thresholds for classifying the wheels with and without defect, under the significance level α =1.24%. The lengths of the bars in Figure 6.21, 6.23, 6.25, 6.27, 6.29 and 6.31 are all one as $\mathrm{P}(H_0|\mathbf{D}, \alpha = 1.24\%) + \mathrm{P}(H_1|\mathbf{D}, \alpha = 1.24\%) = 1$.

It is observed that the intrinsic Bayes factors of the 27th left wheel are -32.9 (Figure 6.20), -32.6 (Figure 6.22) and -28.5 (Figure 6.24), respectively, based on the SSBL, HSBL and PSBL population feature models. The corresponding posterior probabilities of defect are all nearly 100% (Figures 6.21, 6.23 and 6.25). Hence, the 27th left wheel is diagnosed as heavily defected, regardless of which category of population feature models employed. Other left wheels are diagnosed as undefective due to the fact that their intrinsic Bayes factors are all positive and posterior probabilities of defect are all less than 50%.

The intrinsic Bayes factors of the 24th right wheel are -46.3 (Figure 6.26), -40.2 (Figure 6.28) and -36.0 (Figure 6.30), based on the SSBL, HSBL and PSBL population

feature models, respectively. The corresponding posterior probabilities of defect are all almost 100% (Figures 6.27, 6.29 and 6.31). Hence, the 24th right wheel is diagnosed as heavily defected, regardless of the kinds of population feature models.

The intrinsic Bayes factors associated with the 27th right wheel are respectively -9.1, -6.3 and -4.1 based on the three categories of population feature models and their corresponding posterior probabilities of defect are respectively 98.9%, 95.9% and 88.6%. Consequently, the 27th right wheel is identified respectively as moderately, moderately and mildly defected based on the SSBL, HSBL and PSBL population feature models. Other right wheels are identified as undefective because their intrinsic Bayes factors are all positive and their posterior probabilities with defect are all less than 50%.

Different from Bayesian PNHT, the diagnostic results from frequentist NHST and Bayesian NHST are found consistent. For example, when using the monitoring data from a single sensor, only the 27th left wheel, and the 24th and 27th right wheels are identified as defected in both frequentist NHST and Bayesian NHST. By contrast, more wheels are diagnosed as defective in Bayesian PNHT in addition to the three wheels identified as defective in frequentist and Bayesian NHSTs. Therefore, it is concluded that the diagnostic results from non-Bayesian and Bayesian hypothesis tests could be consistent if making a more reasonable alternative hypothesis in Bayesian hypothesis testing. The reasonability lies in that the refined alternative hypothesis in the novel Bayesian

hypothesis testing makes it possible to explicitly control the type-I diagnostic risk, the same as it in frequentist hypothesis testing. Nevertheless, this may not be possible in Bayesian PNHT as the type-I diagnostic risk is averaged over the prior of the unknown parameter associated with the alternative hypothesis such that the diagnostic risk is not explicitly controllable.

In comparison to frequentist NHST, Bayesian NHST provides direct evidence for the condition assessment of nominally identical wheels, in terms of the intrinsic posterior probabilities of defect. More importantly, the extent of defect on wheels can be quantitatively quantified in Bayesian NHST, in terms of intrinsic Bayes factors. Besides, evidence to support the null hypothesis and alternative hypothesis is found comparable. In this case study, strong evidence exists to support both the claims that wheels are with and without defects.

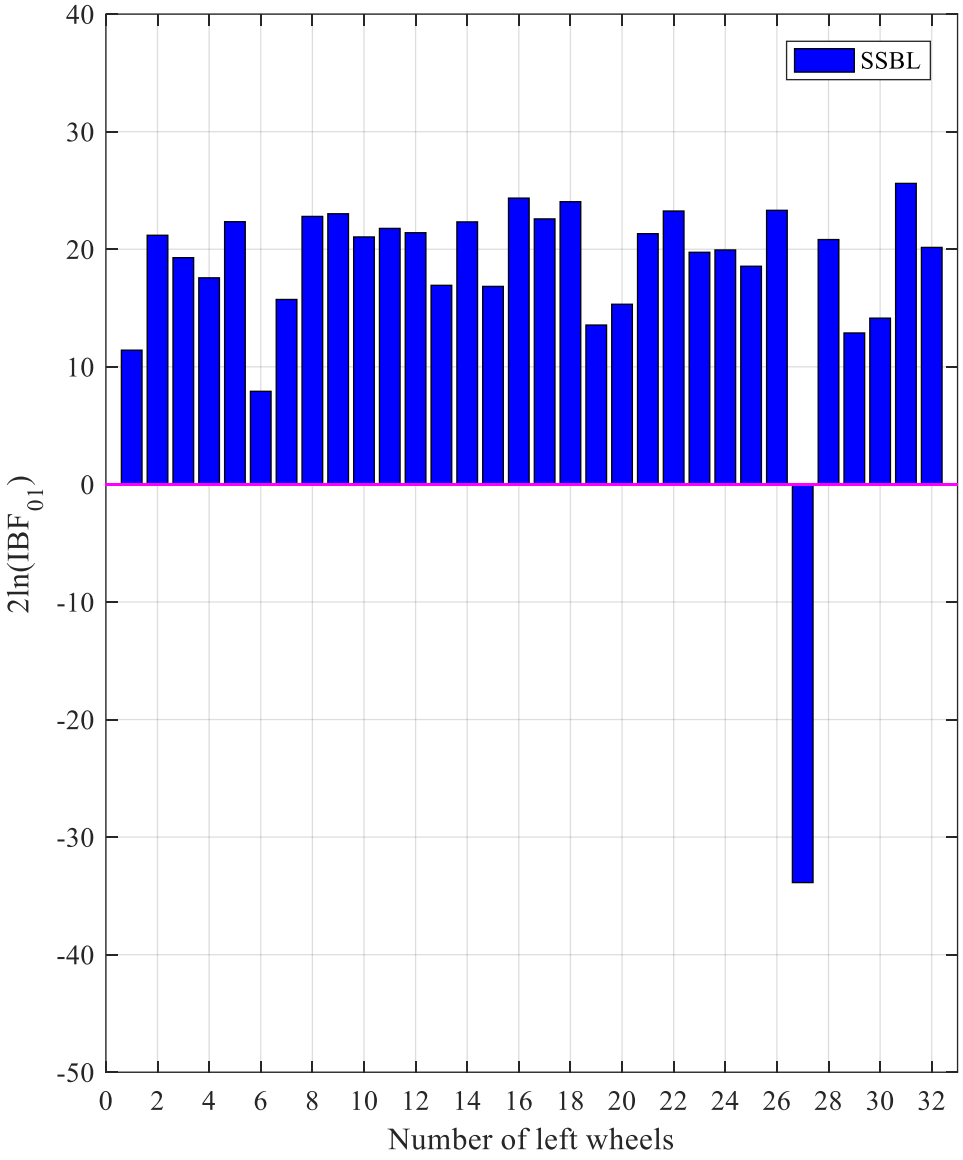


Figure 6.20 Intrinsic Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (SSBL and Bayesian NHST)

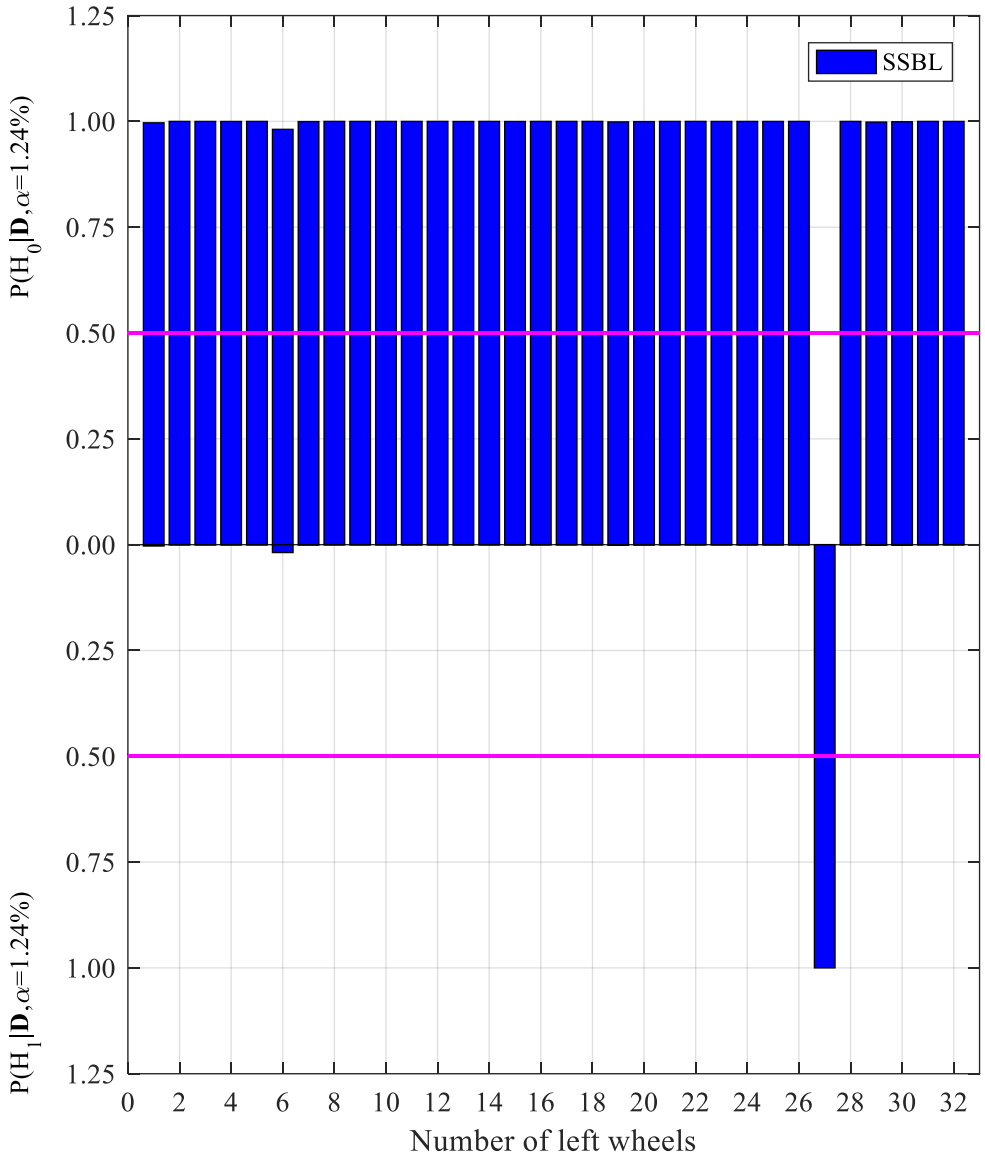


Figure 6.21 Intrinsic posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (SSBL and Bayesian NHST)

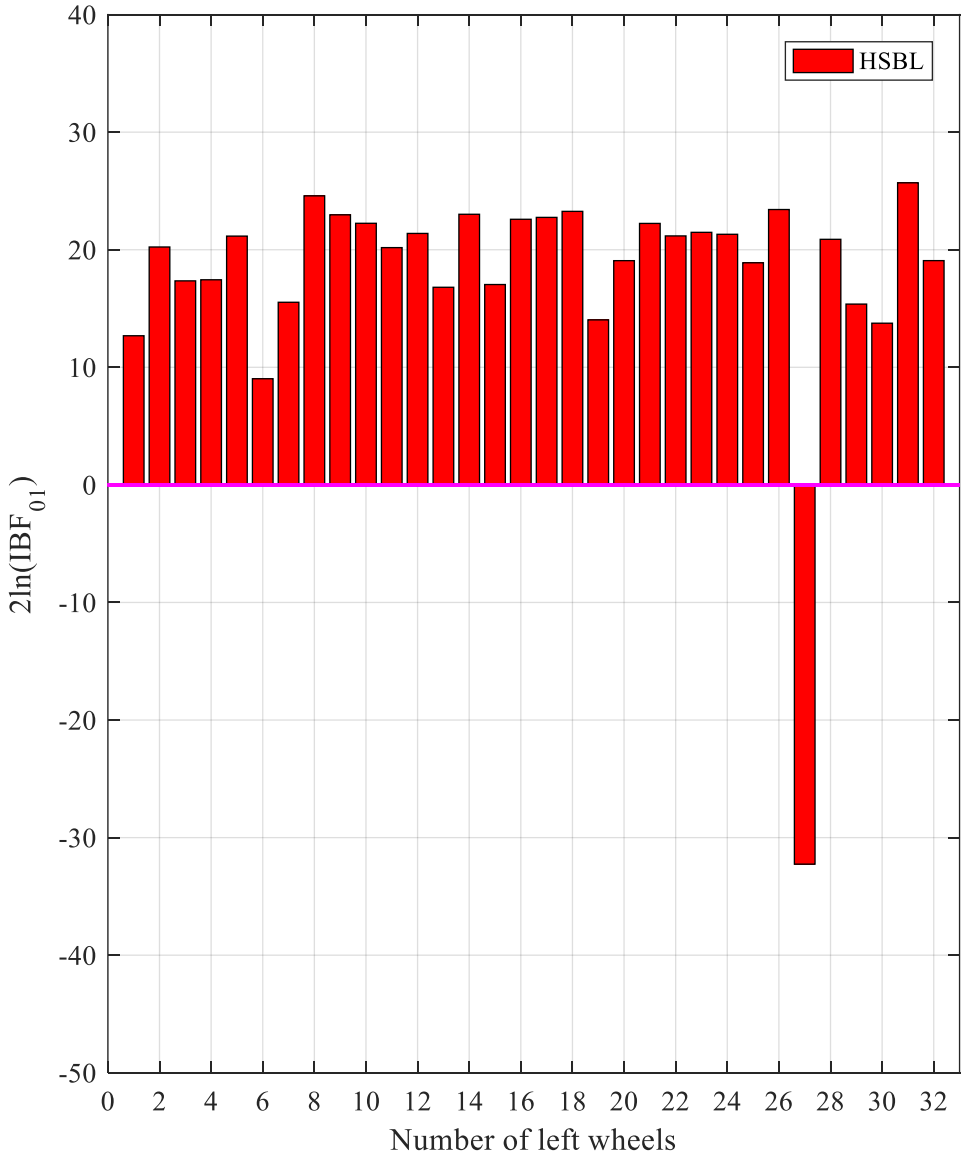


Figure 6.22 Intrinsic Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (HSBL and Bayesian NHST)

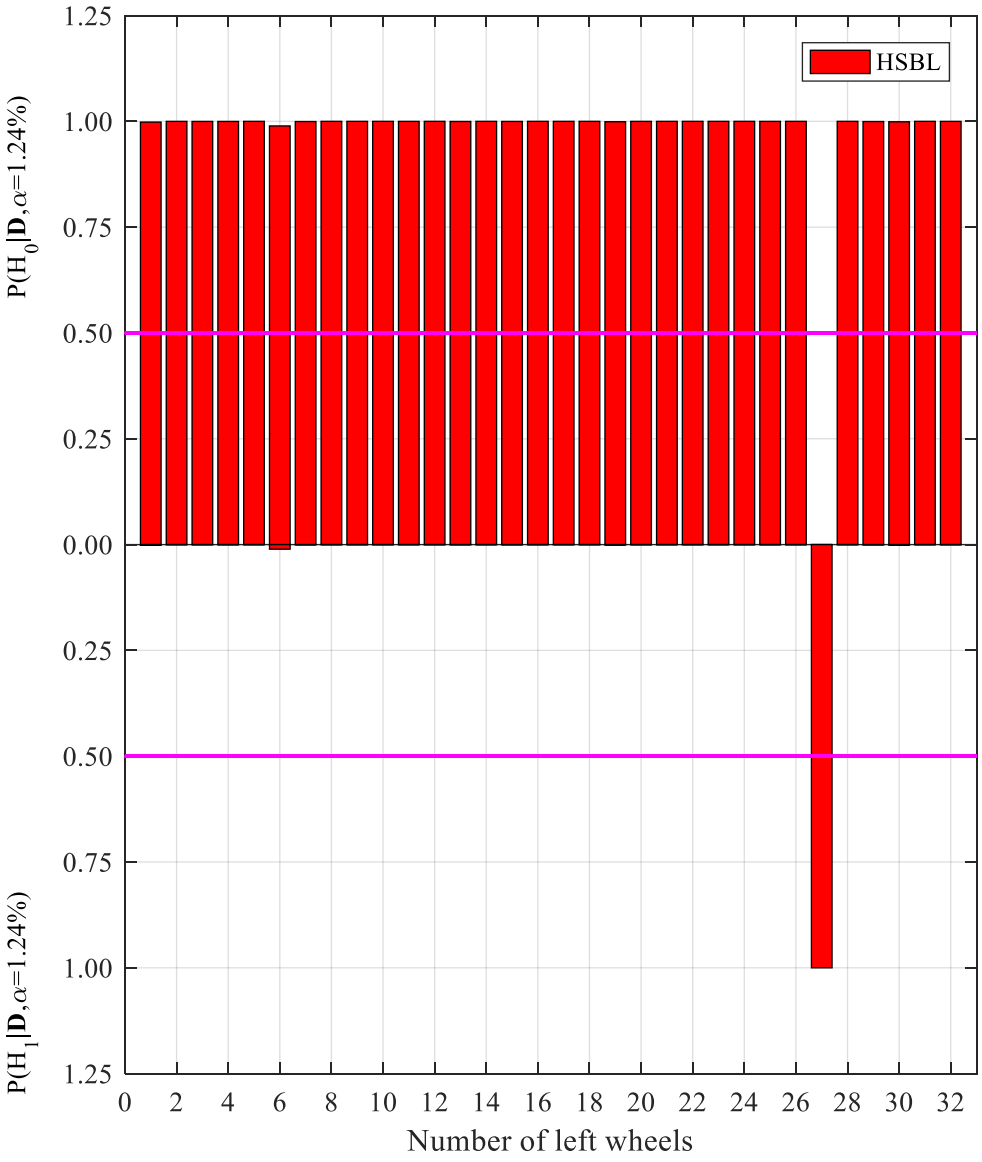


Figure 6.23 Intrinsic posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (HSBL and Bayesian NHST)

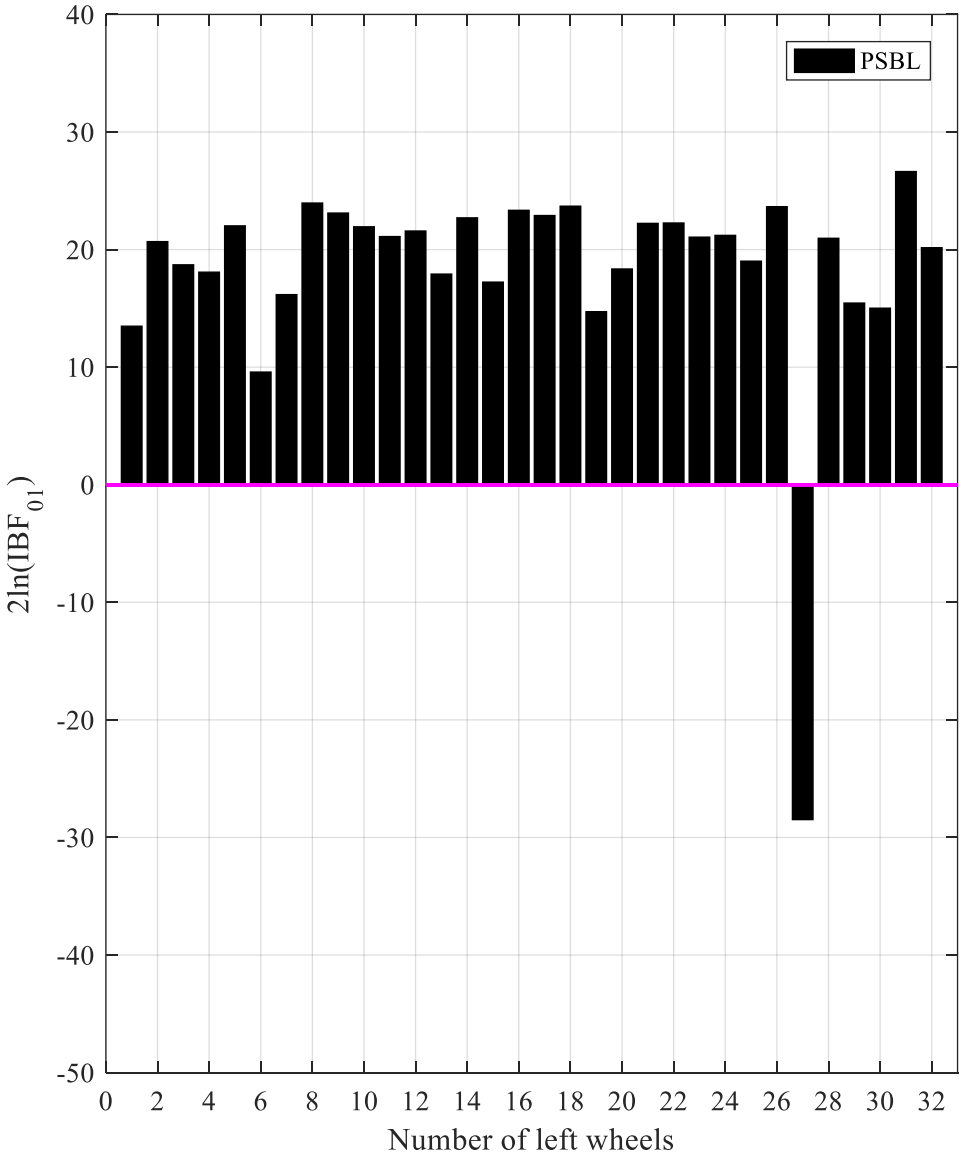


Figure 6.24 Intrinsic Bayes factors of left wheels using monitoring data from SEN-A2 deployed on left rail track (PSBL and Bayesian NHST)

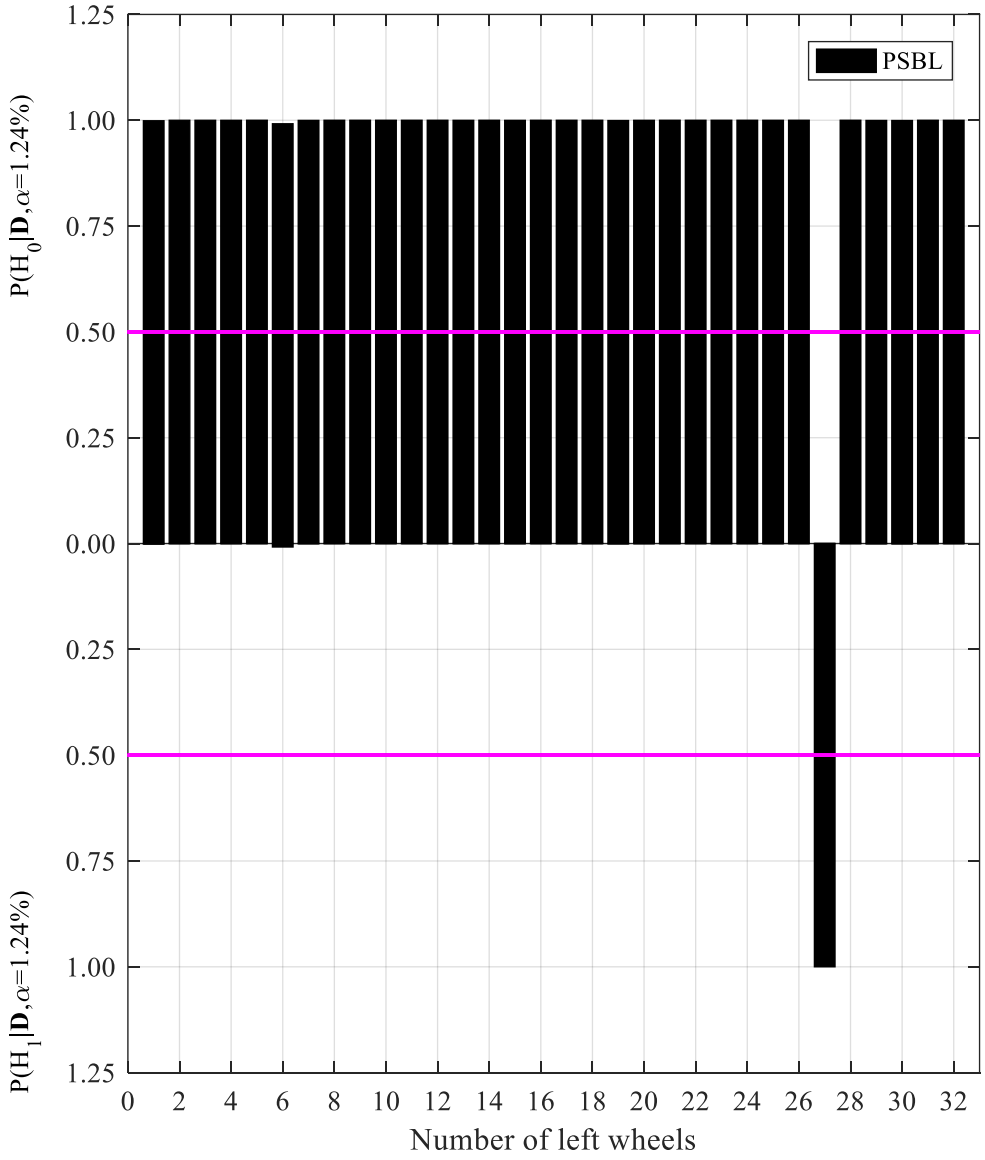


Figure 6.25 Intrinsic posterior probabilities of left wheels using monitoring data from SEN-A2 deployed on left rail track (PSBL and Bayesian NHST)

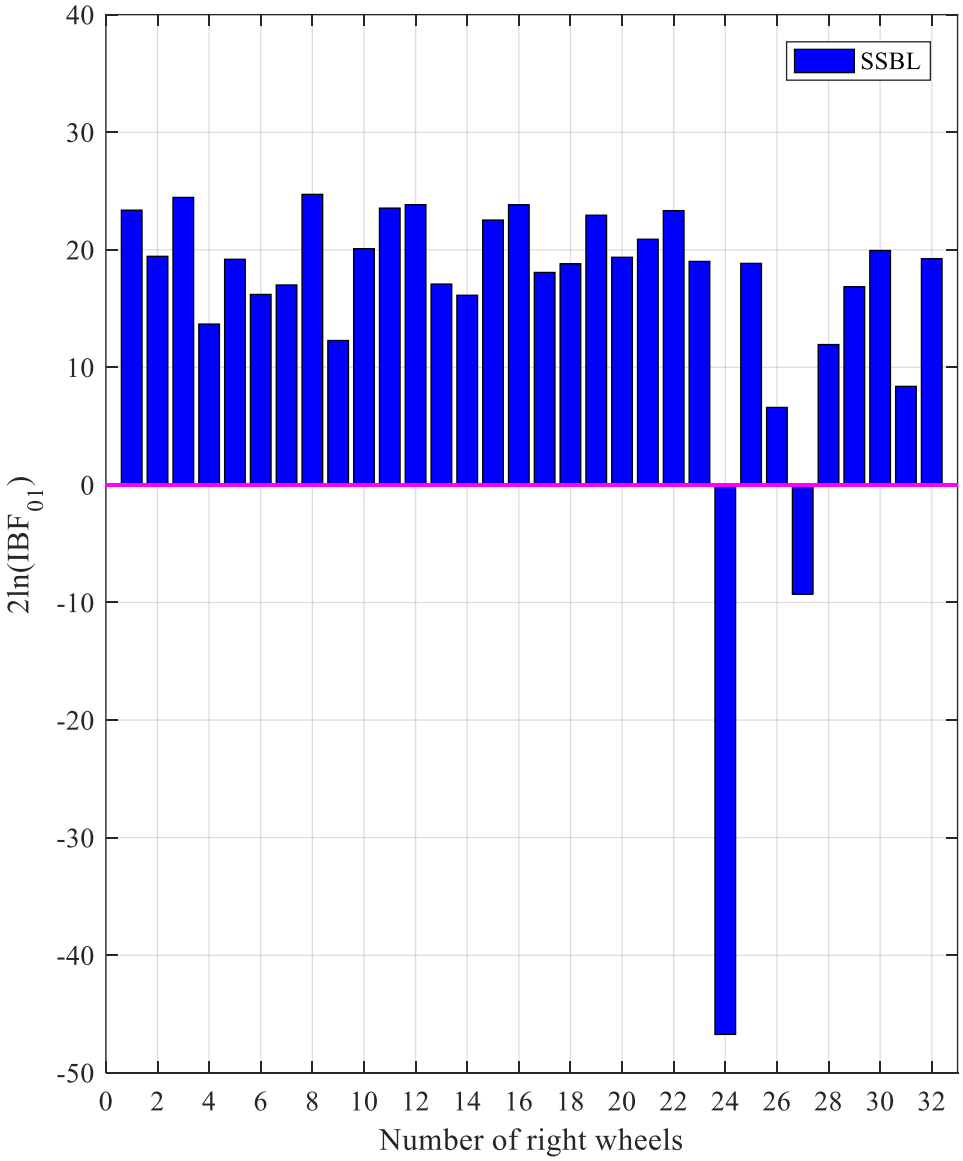


Figure 6.26 Intrinsic Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (SSBL and Bayesian NHST)

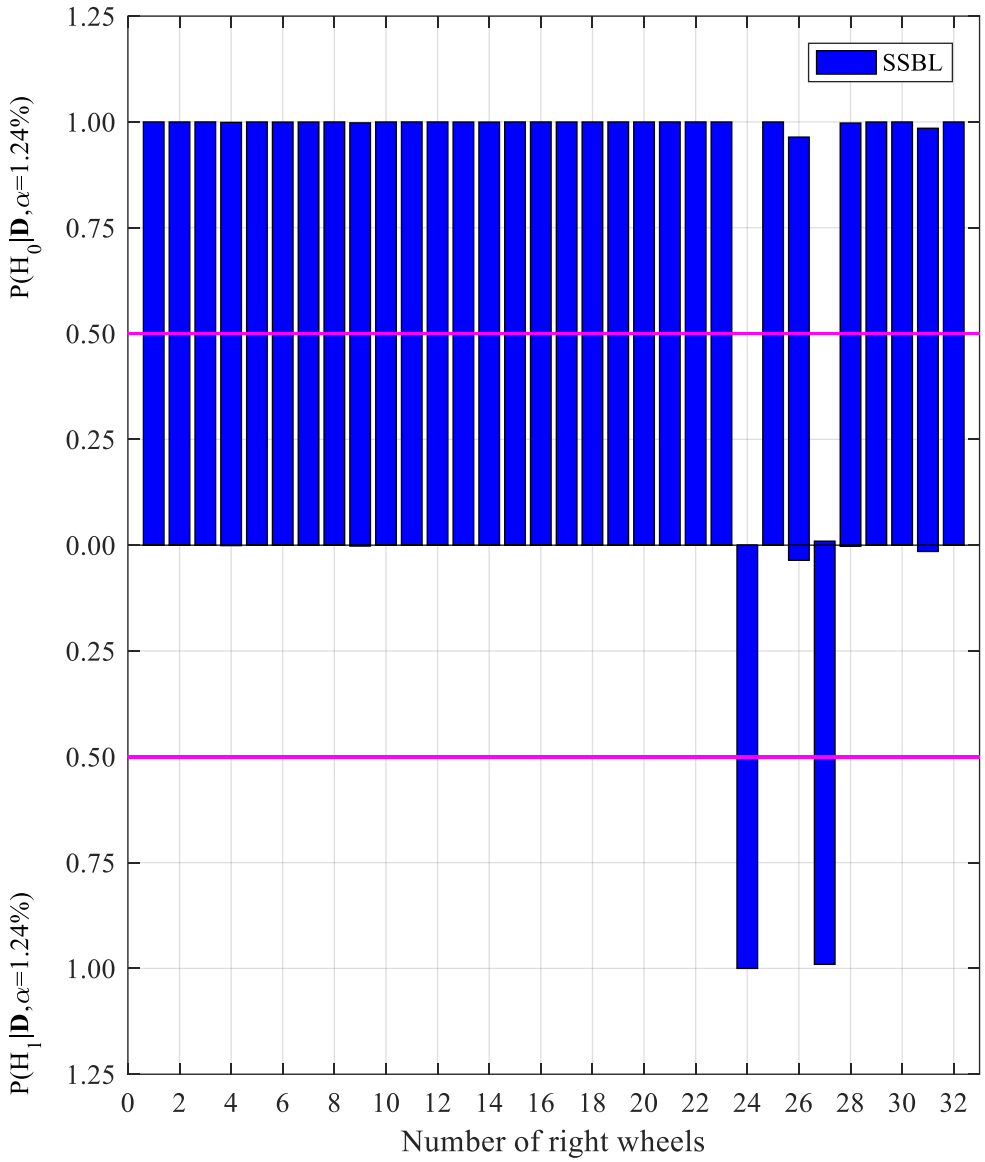


Figure 6.27 Intrinsic posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (SSBL and Bayesian NHST)

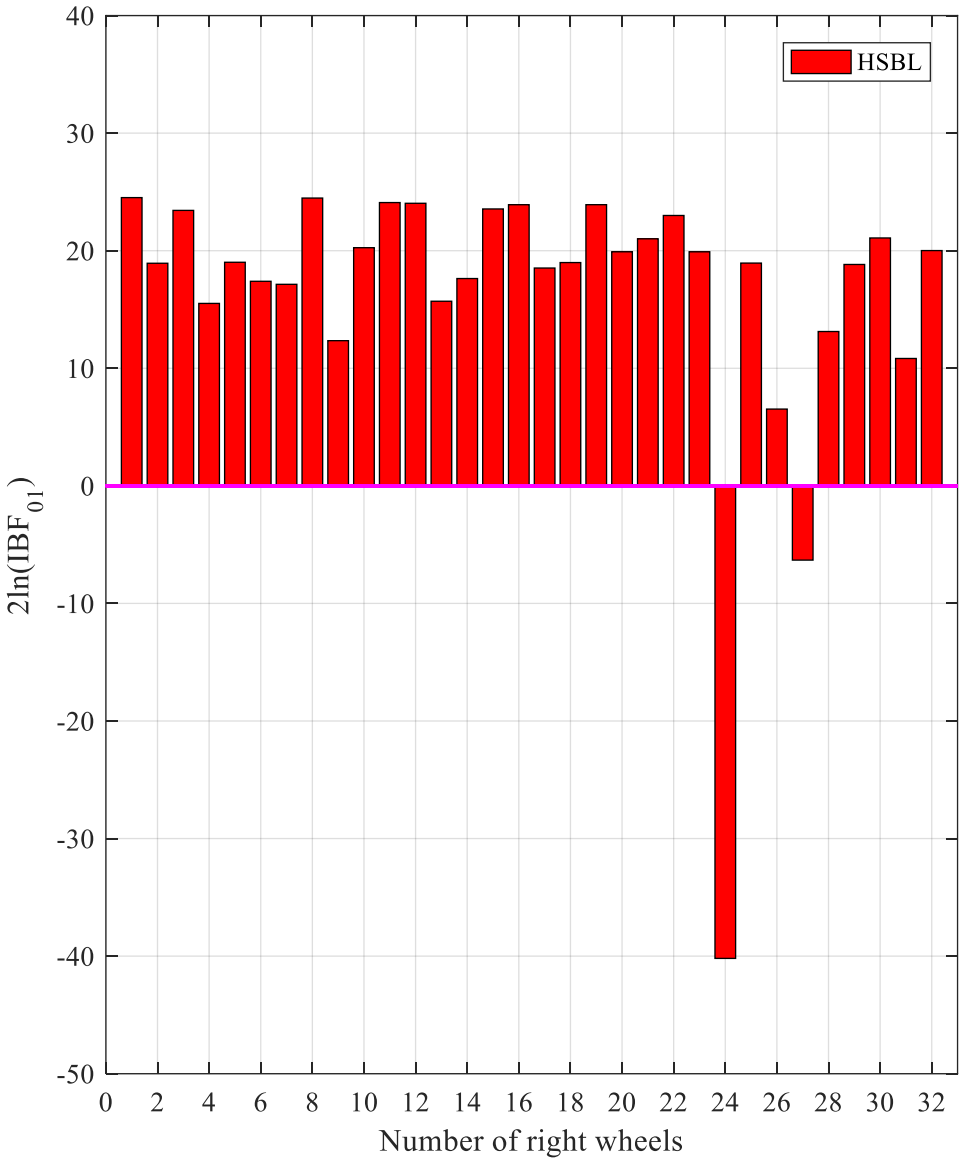


Figure 6.28 Intrinsic Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (HSBL and Bayesian NHST)

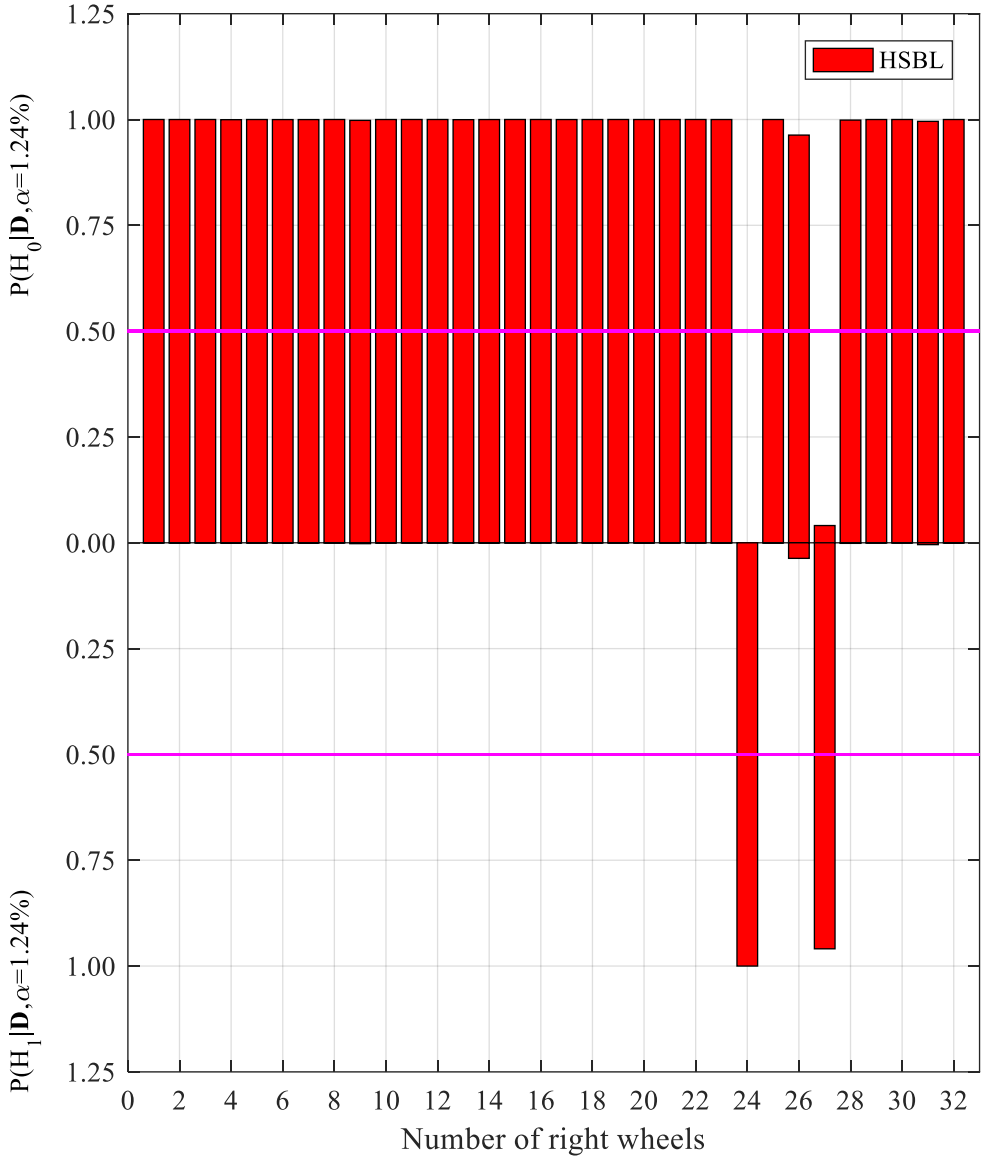


Figure 6.29 Intrinsic posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (HSBL and Bayesian NHST)

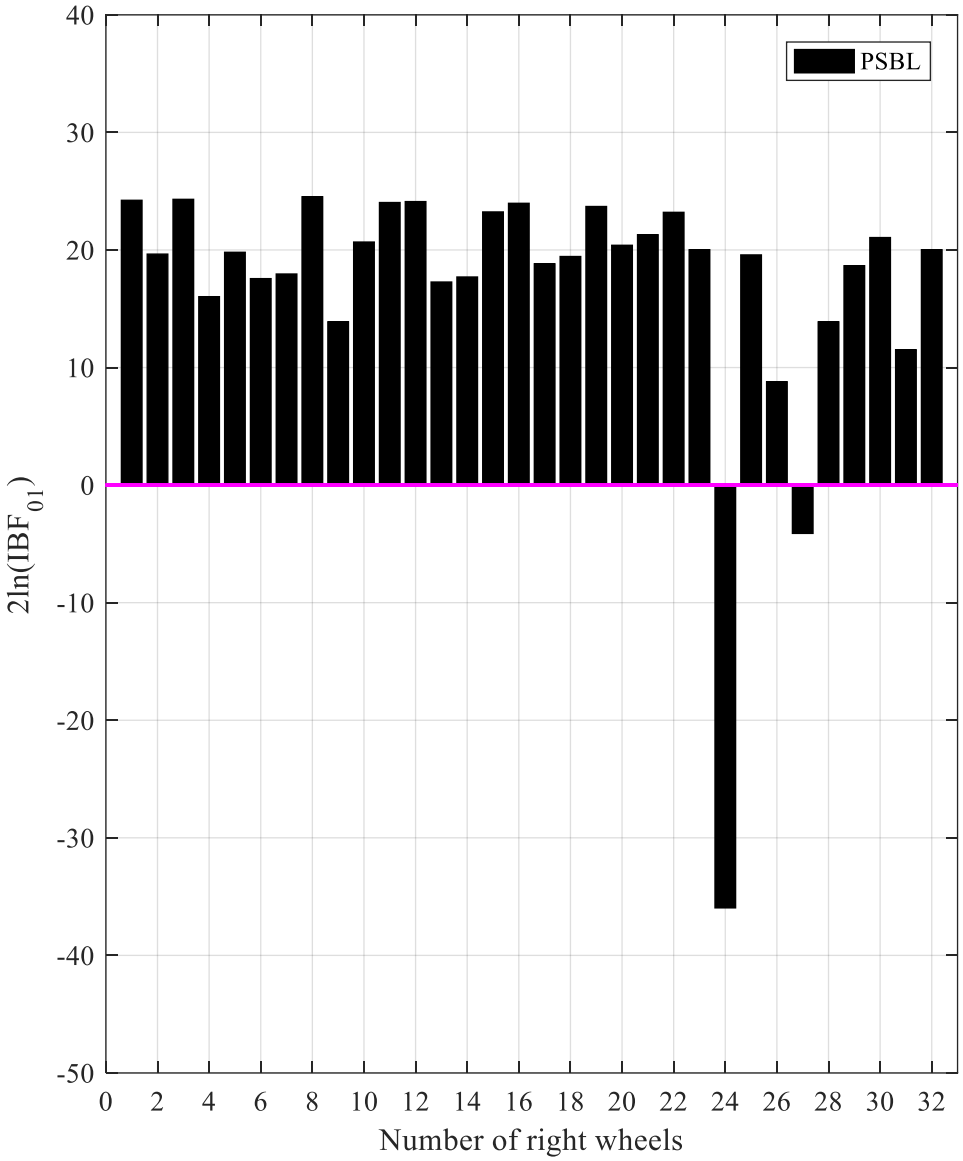


Figure 6.30 Intrinsic Bayes factors of right wheels using monitoring data from SEN-D2 deployed on right rail track (PSBL and Bayesian NHST)

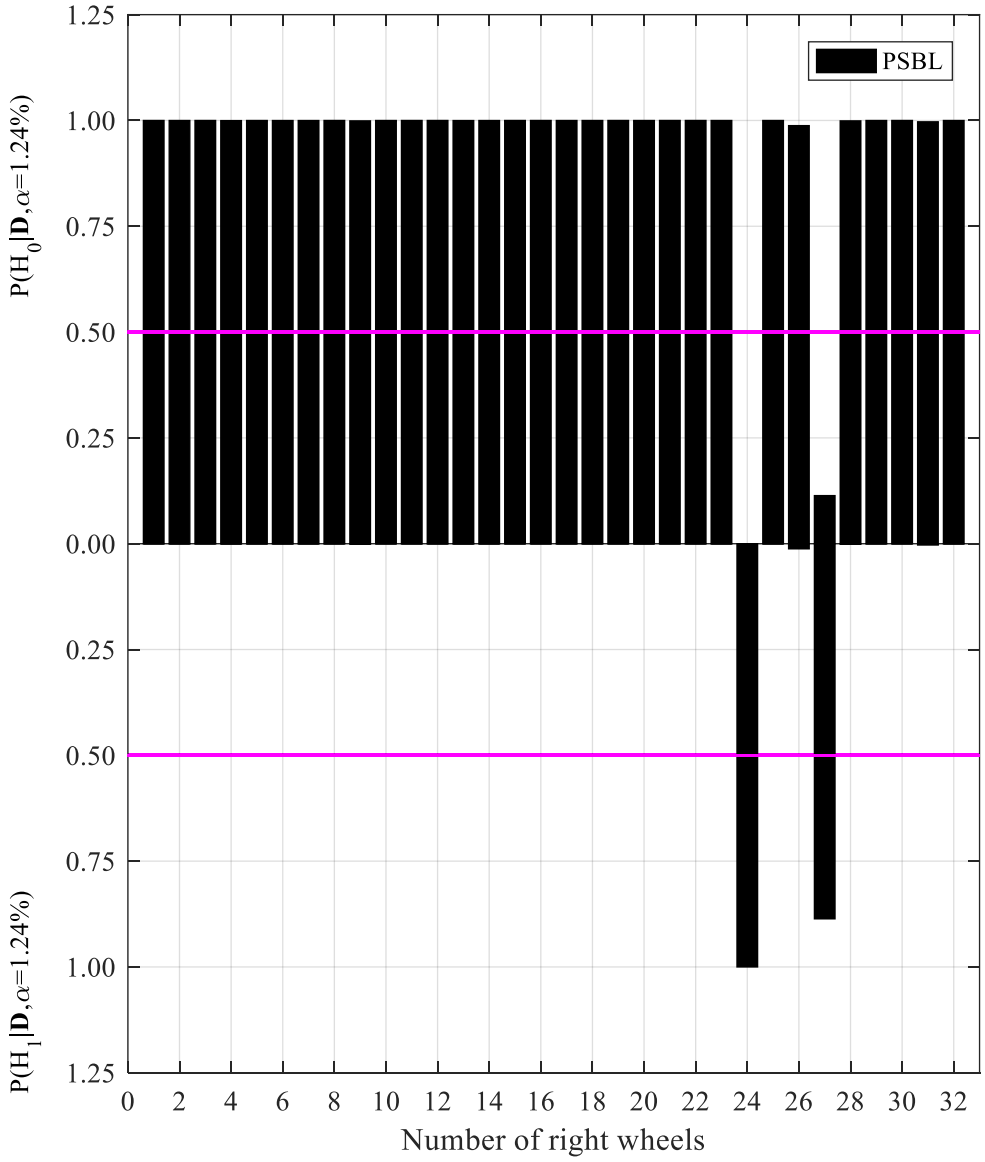


Figure 6.31 Intrinsic posterior probabilities of right wheels using monitoring data from SEN-D2 deployed on right rail track (PSBL and Bayesian NHST)

6.4.2 Diagnostic results of wheel defects by integrating all sensors

The wheel defect diagnostics is the pursued by using the monitoring data from all sensors deployed on two sides of rail tracks. As we are interested in finding defective wheels, the worst wheel condition indices are preferred, including the smallest p-values in frequentist NHST, the smallest Bayes factors and the largest posterior probabilities of defect in Bayesian PNHT, and the smallest intrinsic Bayes factors and the largest intrinsic posterior probabilities of defect in Bayesian NHST. These indices are given, respectively, by

$$p = \min\{p(1), \dots, p(N_s)\} \tag{6.46}$$

$$2\ln(BF_{01}) = \min\{2\ln(BF_{01}(1)), \dots, 2\ln(BF_{01}(N_s))\} \tag{6.47}$$

$$p_0 = \min\{p_0(1), \dots, p_0(N_s)\} \tag{6.48}$$

$$p_1 = \max\{p_1(1), \dots, p_1(N_s)\} \tag{6.49}$$

$$2\ln(IBF_{01}) = \min\{2\ln(IBF_{01}(1)), \dots, 2\ln(IBF_{01}(N_s))\} \tag{6.50}$$

$$p_0^I = \min\{p_0^I(1), \dots, p_0^I(N_s)\} \tag{6.51}$$

$$p_1^I = \max\{p_1^I(1), \dots, p_1^I(N_s)\} \tag{6.52}$$

where N_s is the number of sensors deployed on each side of the rail track. For the track-side monitoring system in this thesis, $N_s = 21$.

a) ***Frequentist NHST results***

Figures 6.32 to 6.34 illustrate the condition assessment results of left wheels by using

frequentist NHST on the monitoring data from all 21 sensors deployed on the left rail track, with the population feature model established from SSBL, HSBL and PSBL, respectively. It is observed that regardless of which type of population feature models used, the 1st, 6th and 27th left wheels are diagnosed as defected as their p-values are below the magenta lines ($\alpha = 1.24\%$) and other left wheels are diagnosed as undefective as their associated p-values are above the magenta lines.

Figures 6.35 to 6.37 show the condition assessment results for right wheels by using frequentist NHST on the monitoring data from all 21 sensors deployed on the right rail track with the population feature model established from SSBL, HSBL and PSBL, respectively. It is found that when using the SSBL population feature model, five right wheels are diagnosed as defective, including the 1st, 6th, 24th, 27th and 31st wheels. Other right wheels are diagnosed as undefective because their p-values are larger than the prespecified significance level $\alpha = 1.24\%$. However, if using the HSBL or PSBL population feature models, only four right wheels are identified as defective, including the 1st, 6th, 24th and 27th wheels. Thus, a qualitative difference is found in the condition classification of right wheels from the three categories of population feature models when using the monitoring data from all 21 sensors deployed on the right rail track.

By comparing Figures 6.2 to 6.7 with Figures 6.32 to 6.37, it is also found that using only the monitoring data from a single sensor may fail to identify potentially defective

wheels. For instance, using only the monitoring data from the sensor SEN-A2 fails to identify potential defects on the 1st and 6th left wheels, while using only the monitoring data from the sensor SEN-D2 fails to identify potential defects on the 1st and 6th right wheels, regardless of the kinds of population feature models employed.

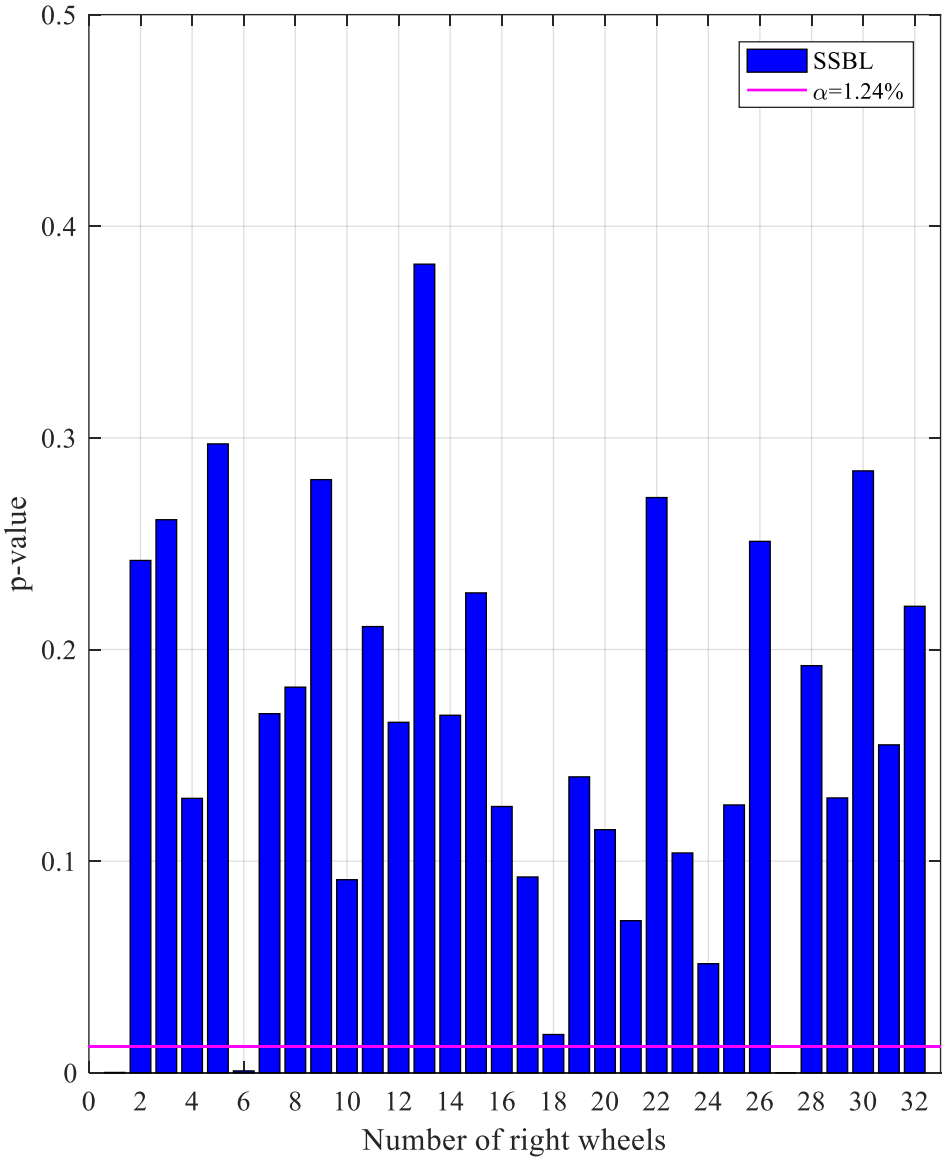


Figure 6.32 p-values of left wheels using monitoring data from all sensors deployed on left rail track (SSBL and Frequentist NHST)

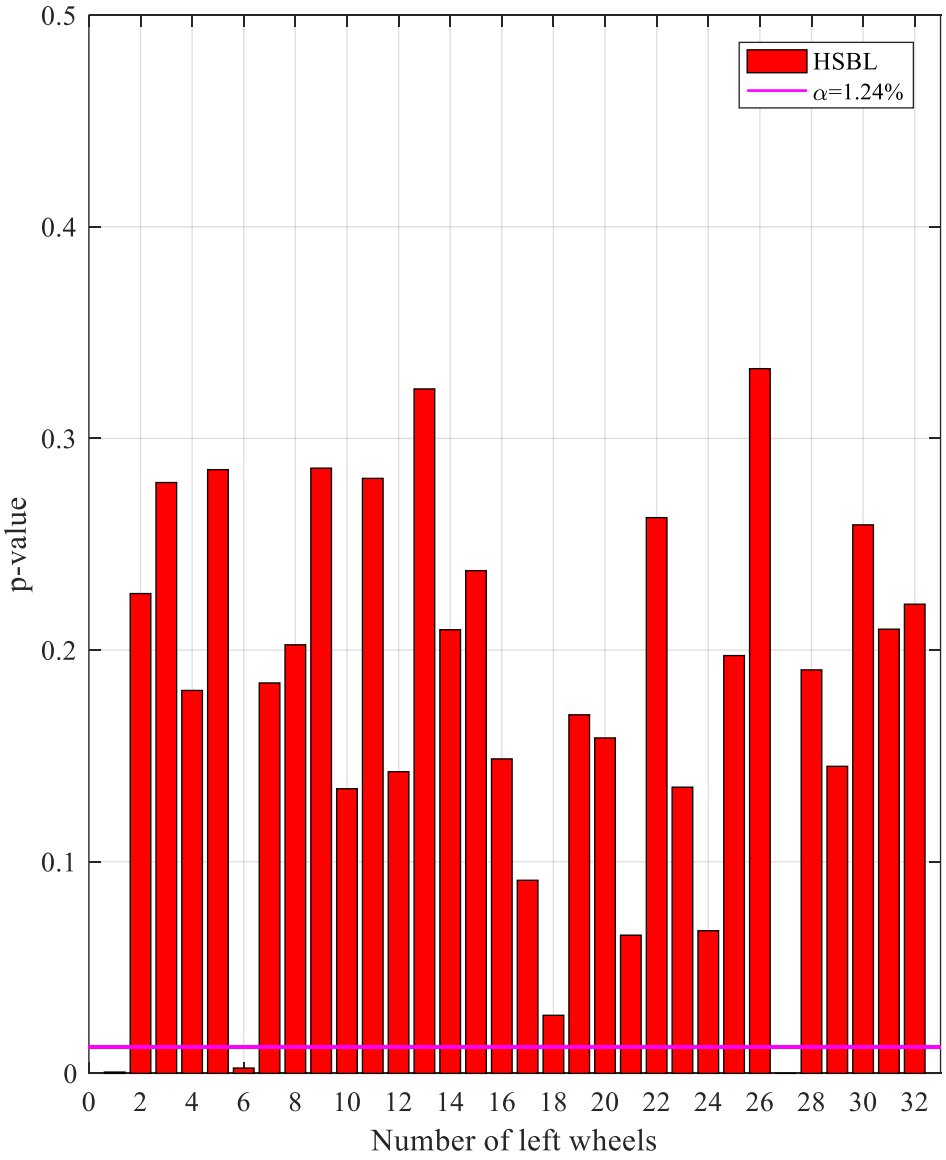


Figure 6.33 p-values of left wheels using monitoring data from all sensors deployed on left rail track (HSBL and Frequentist NHST)

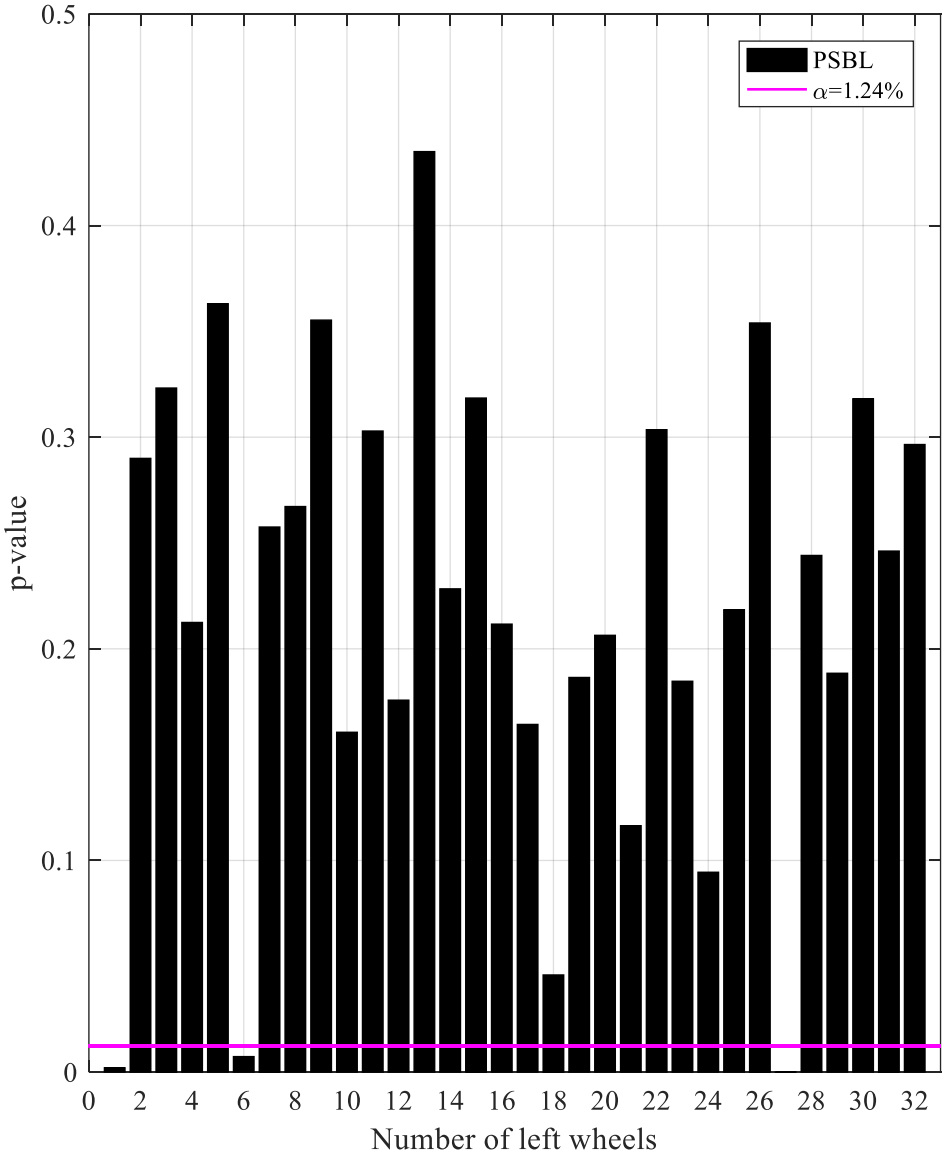


Figure 6.34 p-values of left wheels using monitoring data from all sensors deployed on left rail track (PSBL and Frequentist NHST)

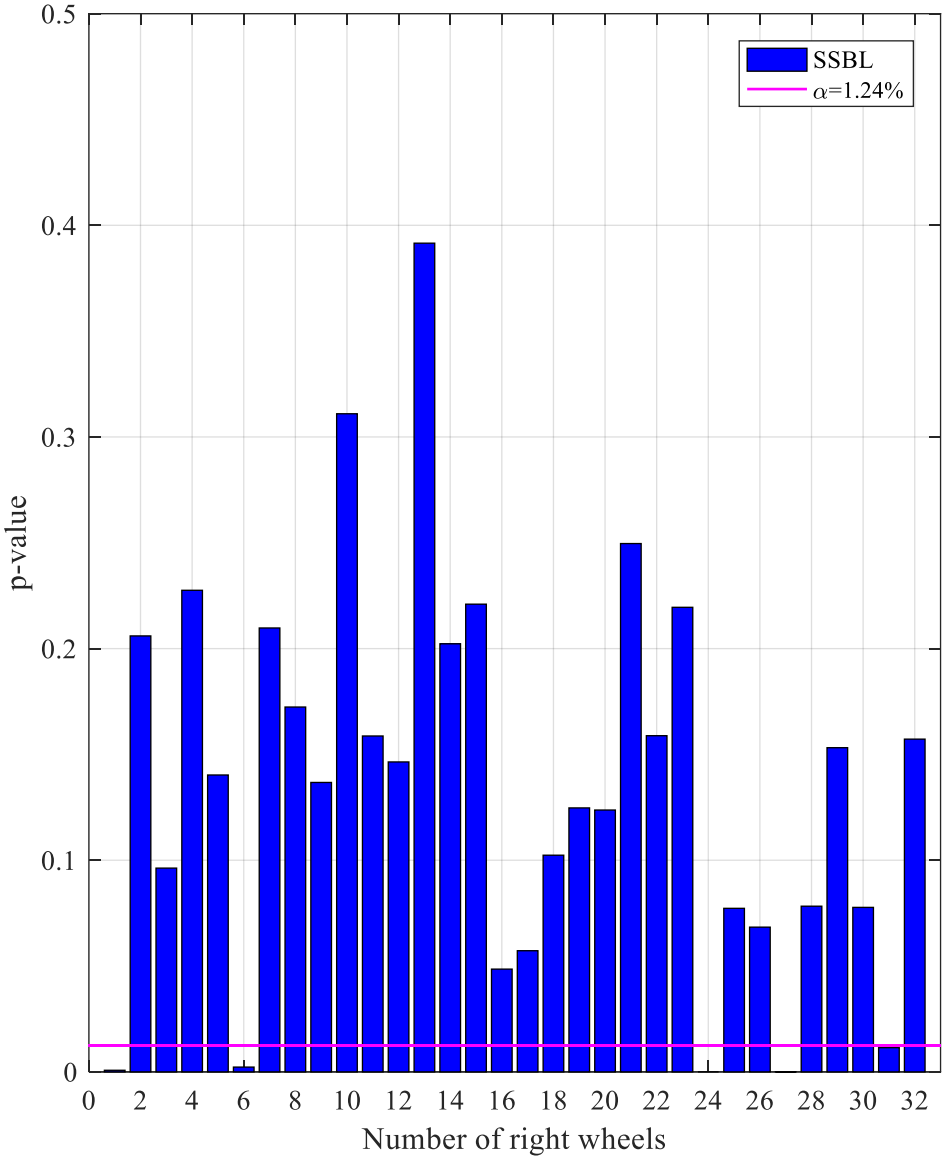


Figure 6.35 p-values of right wheels using monitoring data from all sensors deployed on right rail track (SSBL and Frequentist NHST)

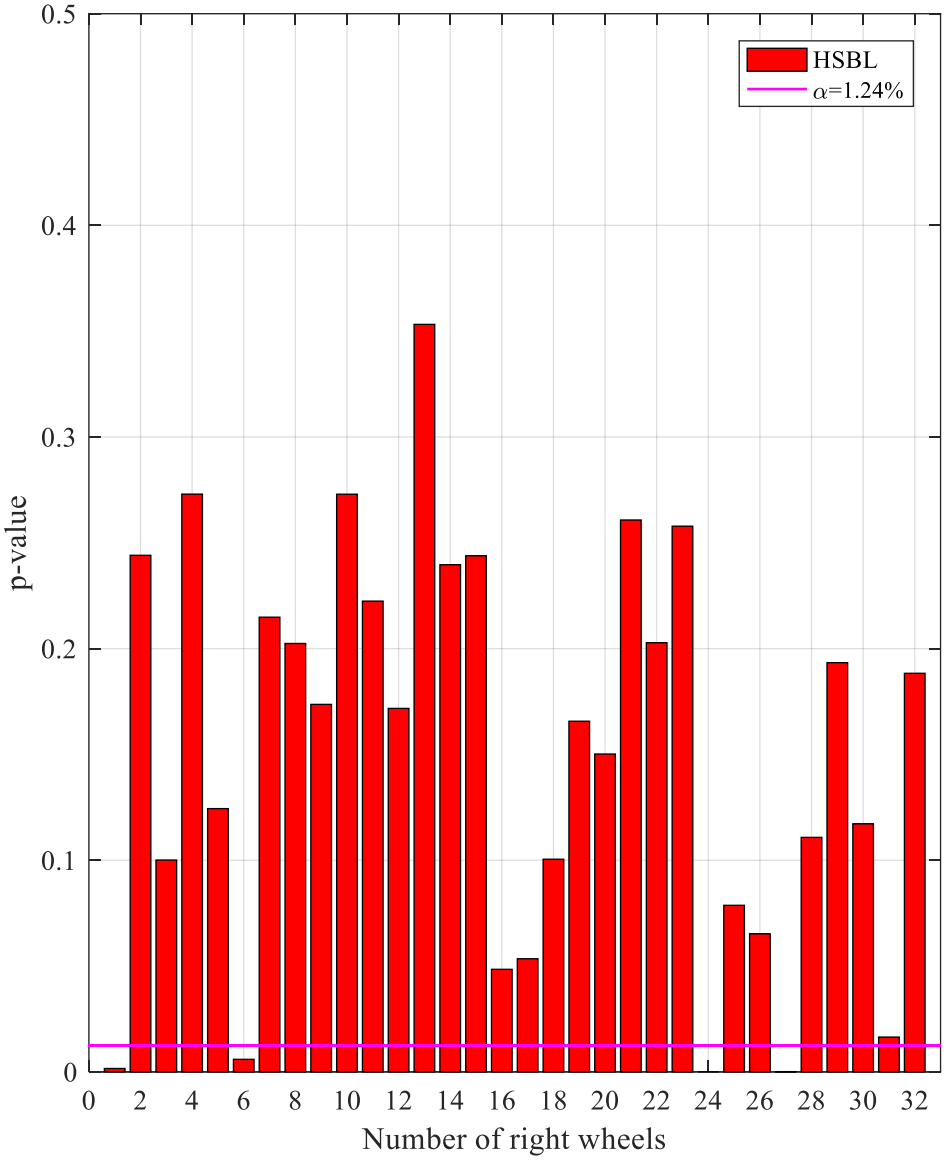


Figure 6.36 p-values of right wheels using monitoring data from all sensors deployed on right rail track (HSBL and Frequentist NHST)

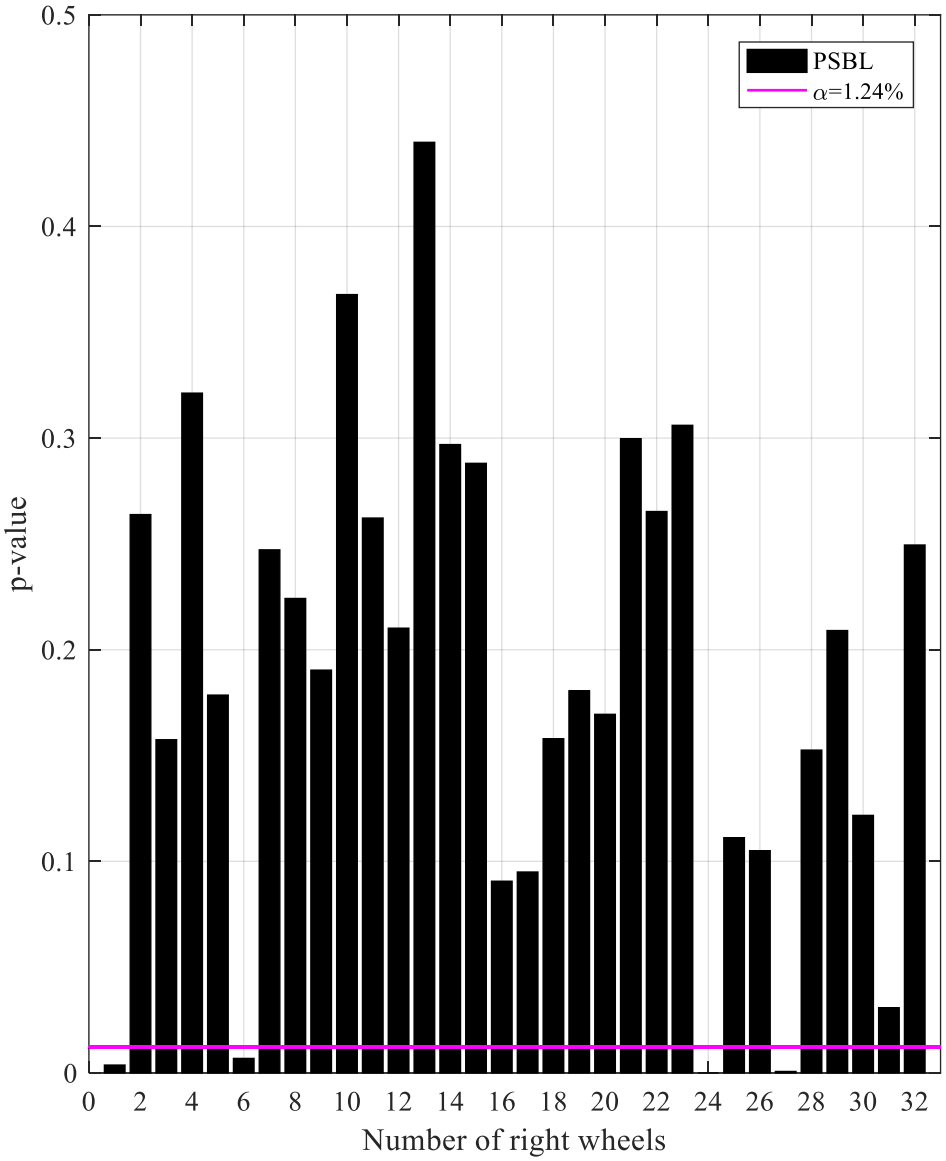


Figure 6.37 p-values of right wheels using monitoring data from all sensors deployed on right rail track (PSBL and Frequentist NHST)

b) ***Bayesian PNHT results***

The diagnostic results of left wheels using Bayesian PNHT on the monitoring data from all 21 sensors deployed on the left rail track are shown in Figures 6.38 to 6.43, with the population feature models established by SSBL, HSBL and PSBL, respectively. It is found that when using the SSBL or HSBL population feature model, all left wheels are diagnosed as defective as their associated Bayes factors are all negative and their associated posterior probabilities of defect are all more than 50.0%. If using the PSBL population feature model, except the 9th, 26th and 30th left wheels, all other left wheels are identified as defective. The diagnostic results of right wheels using Bayesian PNHT on the monitoring data from all 21 sensors deployed on the right rail track are shown in Figures 6.44 to 6.49, with the population feature models established by SSBL, HSBL and PSBL, respectively. It is found that the Bayes factors of right wheels are all negative and their corresponding posterior probabilities of defect are all more than 50%. Thus, all right wheels are identified as defective regardless of the kinds of population feature models.

Comparing Figures 6.38, 6.40, 6.42, 6.44, 6.46 and 6.48, it is seen that among the three types of population feature models, the Bayes factors are the smallest, calculated from the SSBL population feature model. This is mainly attributed to the fact that multiple sources of uncertainty are pooled together in the SSBL population feature model, including measurement noise in the online monitoring activity of railway wheels (intra-

structure uncertainty) as well as variability in wheel material and/or manufacturing process (inter-structure uncertainty). The pooling of multiple sources of uncertainty yields an underestimation of uncertainty associated with the population features of all nominally identical undefective wheels. Consequently, the obtained Bayes factors are included to be underestimated, while the extents of defects on the wheels are overestimated. Although the HSBL population feature model can be useful to alleviate the overestimation of the extent of defect on the wheels, the most reliable diagnostic results are given by the PSBL population feature model, in which multiple sources of uncertainty can be modelled separately.

Comparing the diagnostic results on the wheels from frequentist NHST (Figures 6.32 to 6.37), it is found that more wheels are identified as defective if using Bayesian PNHT on the monitoring data from all sensors deployed on each side of the rail. For example, if using frequentist NHST, only three left wheels are diagnosed as defective, while no more than five right wheels are identified as defected no matter which category of the employed population feature models. By contrast, nearly all wheels (except the 26th and 30th left wheels) are diagnosed defective when using Bayesian PNHT. The Jeffreys-Lindley takes place again in this case study even though the monitoring data from all deployed sensors are utilized.

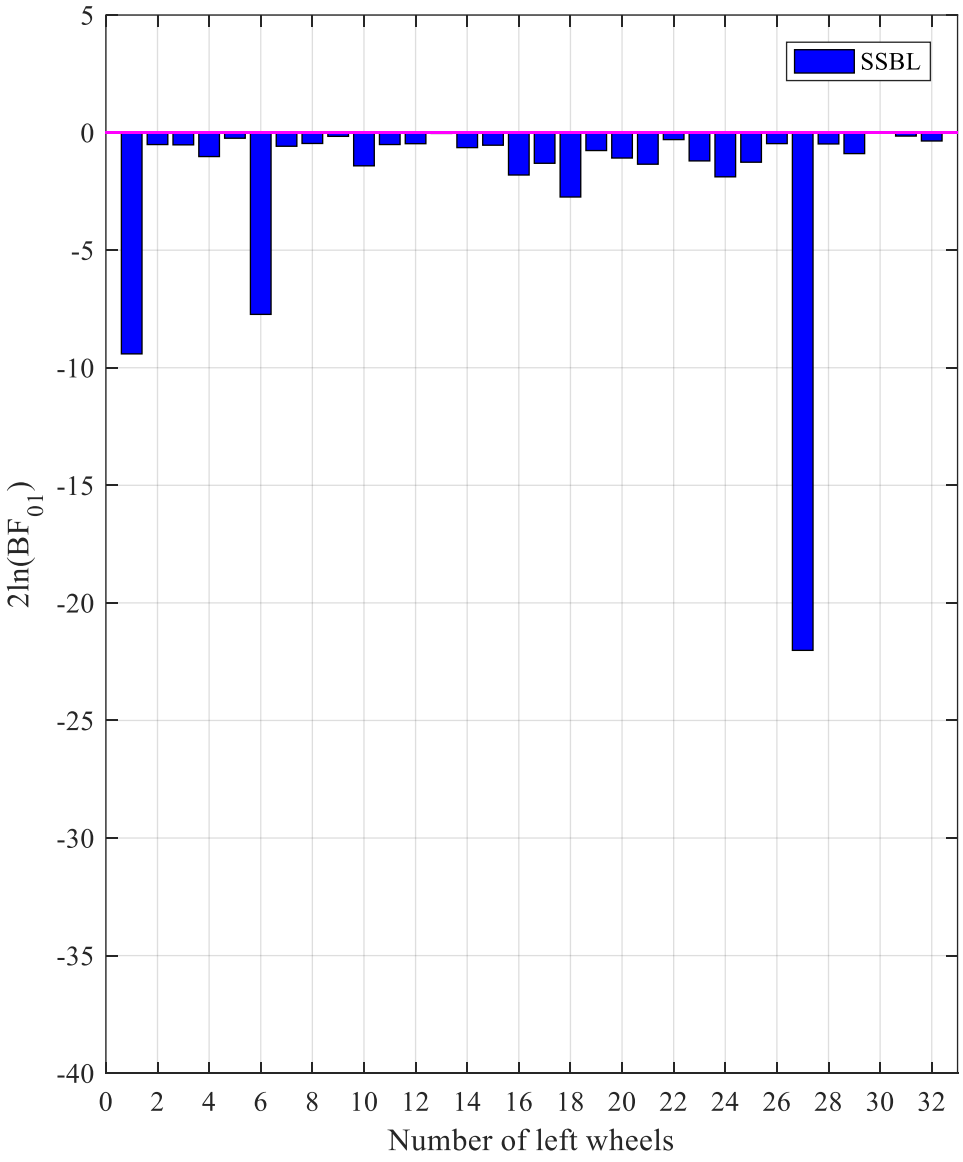


Figure 6.38 Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (SSBL and Bayesian PNHT)

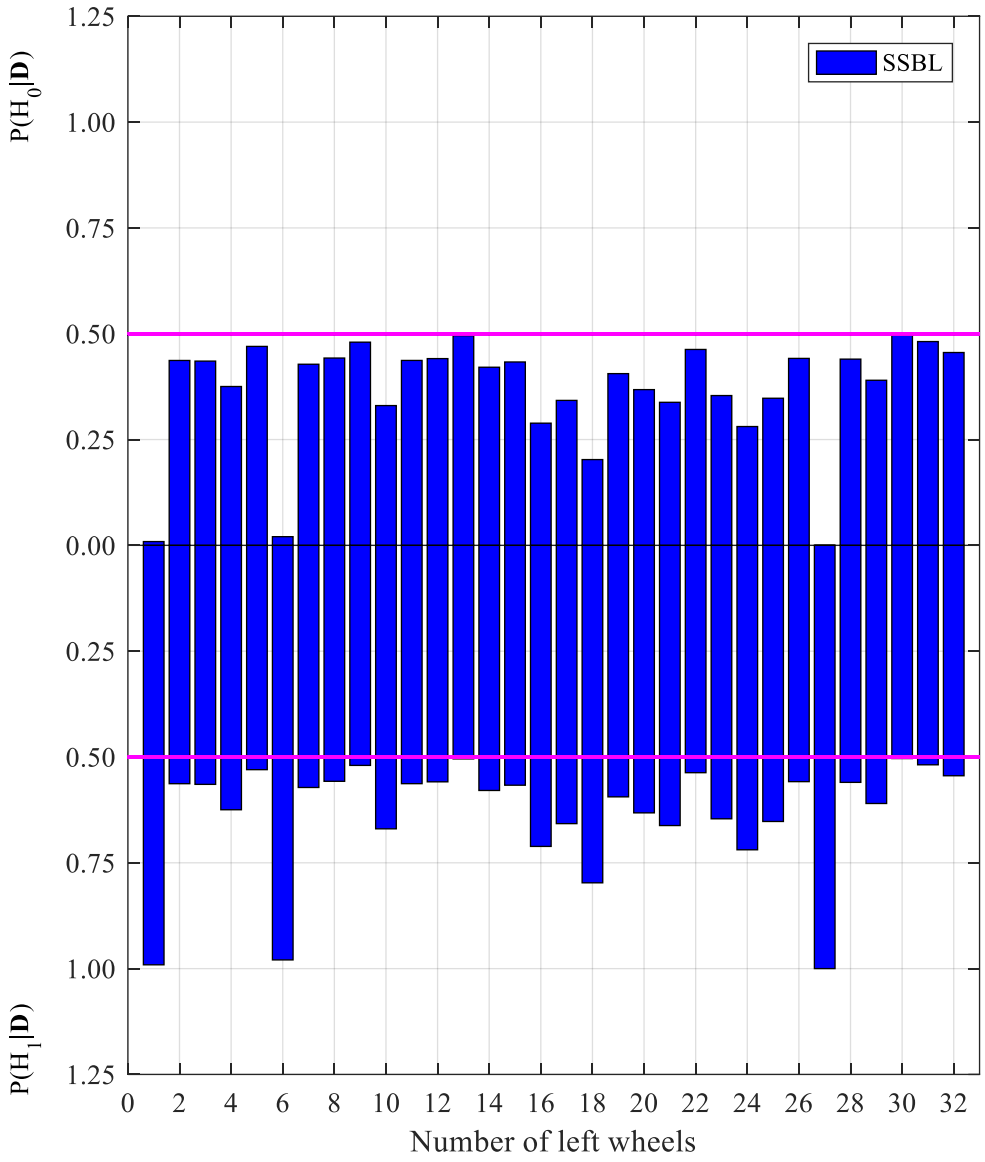


Figure 6.39 Posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (SSBL and Bayesian PNHT)

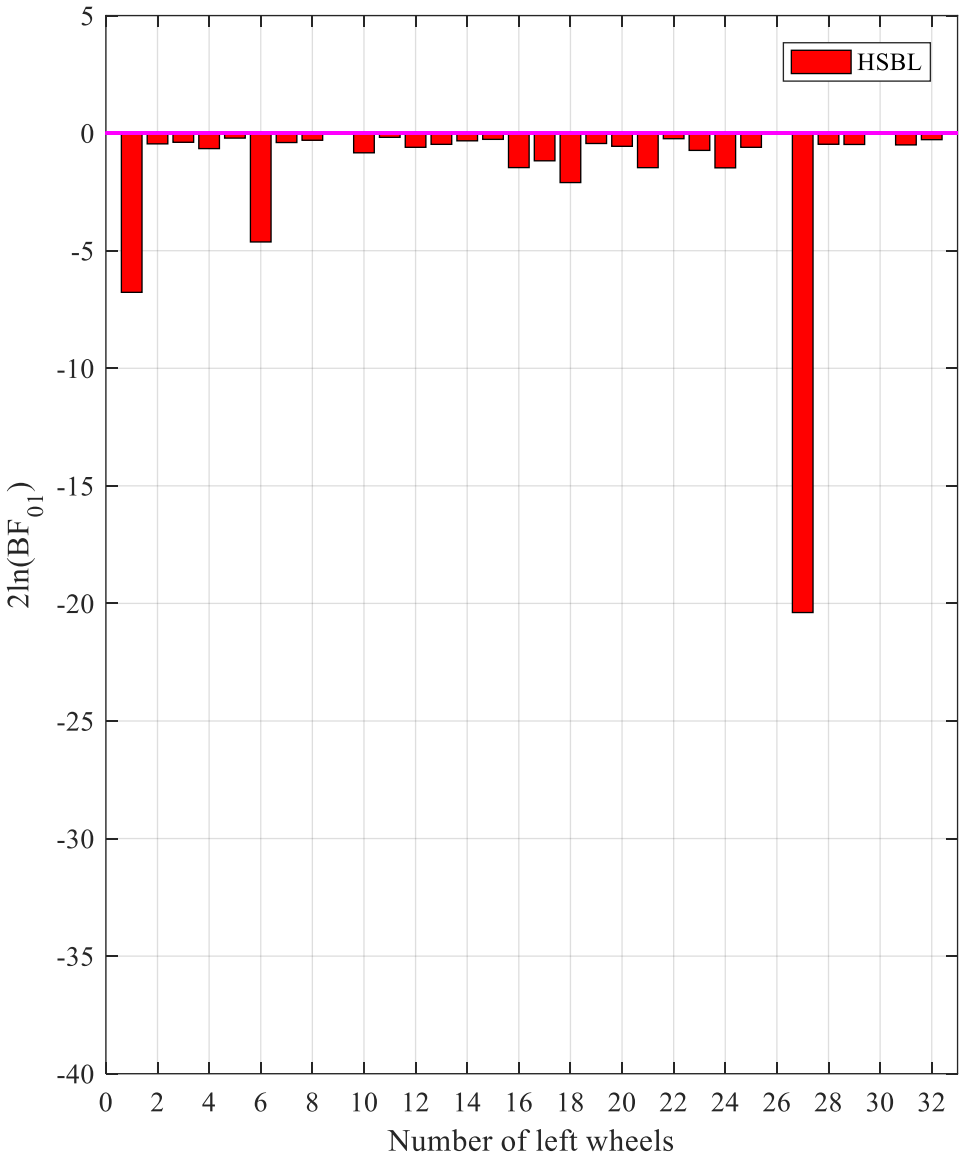


Figure 6.40 Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (HSBL and Bayesian PNHT)

Figure 6.41 Posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (HSBL and Bayesian PNHT)

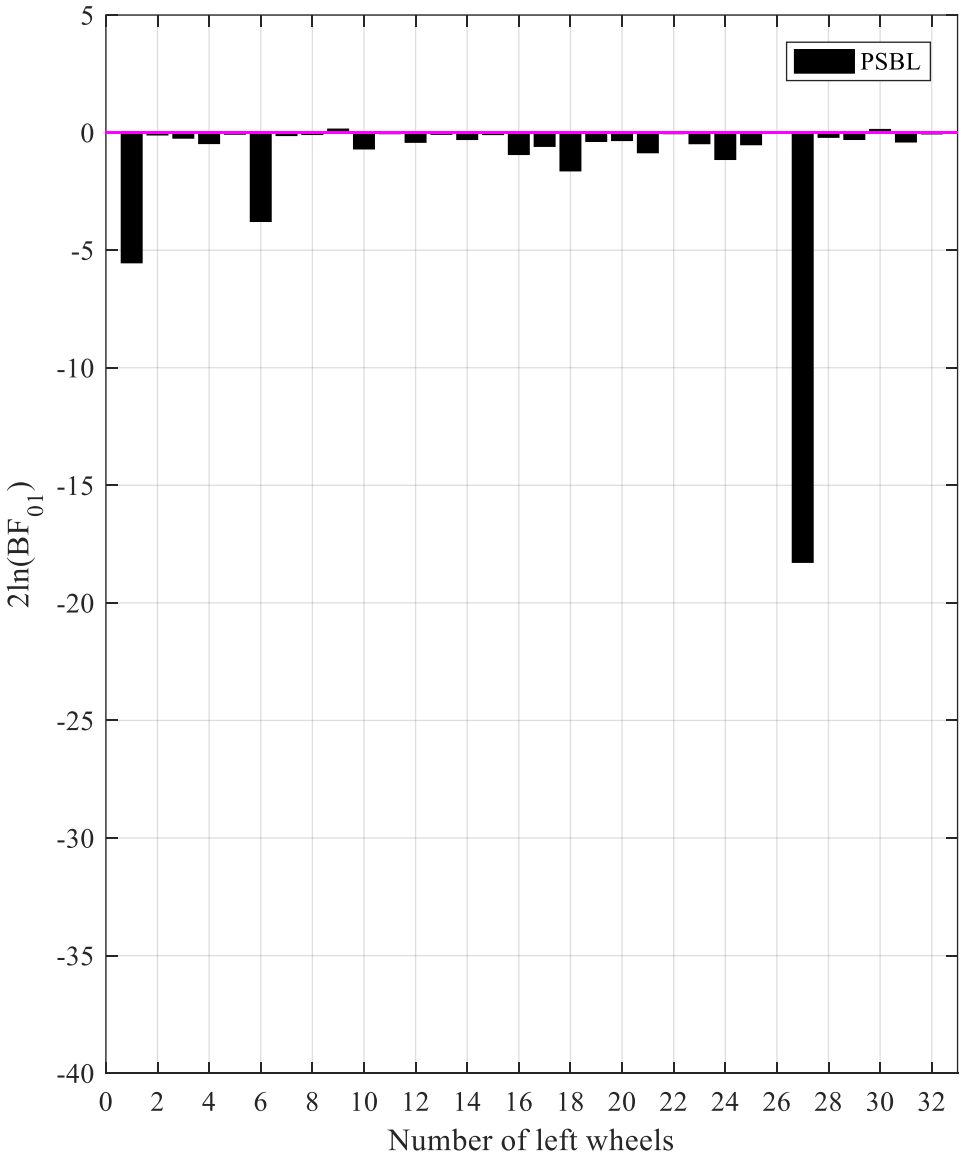


Figure 6.42 Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (PSBL and Bayesian PNHT)

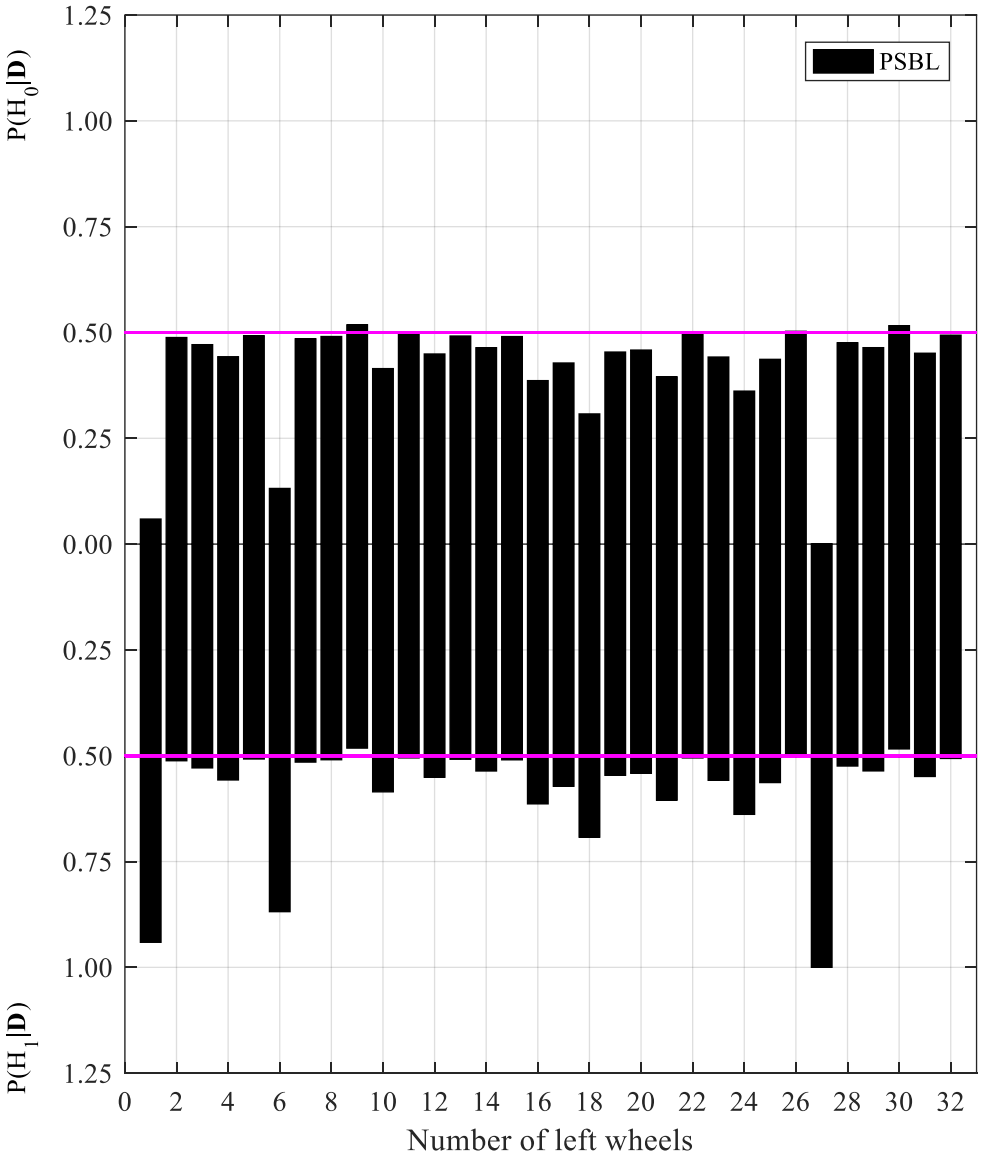


Figure 6.43 Posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (PSBL and Bayesian PNHT)

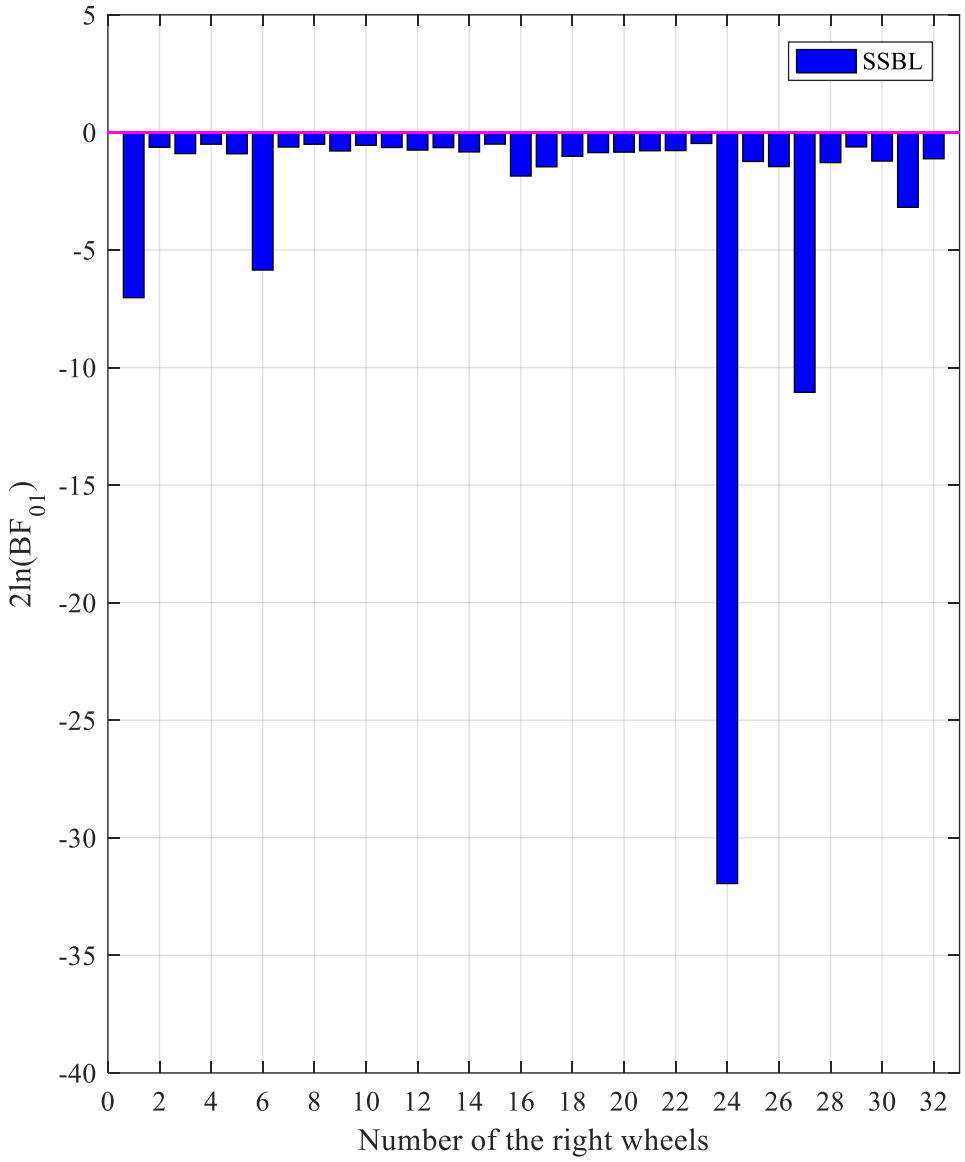


Figure 6.44 Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (SSBL and Bayesian PNHT)

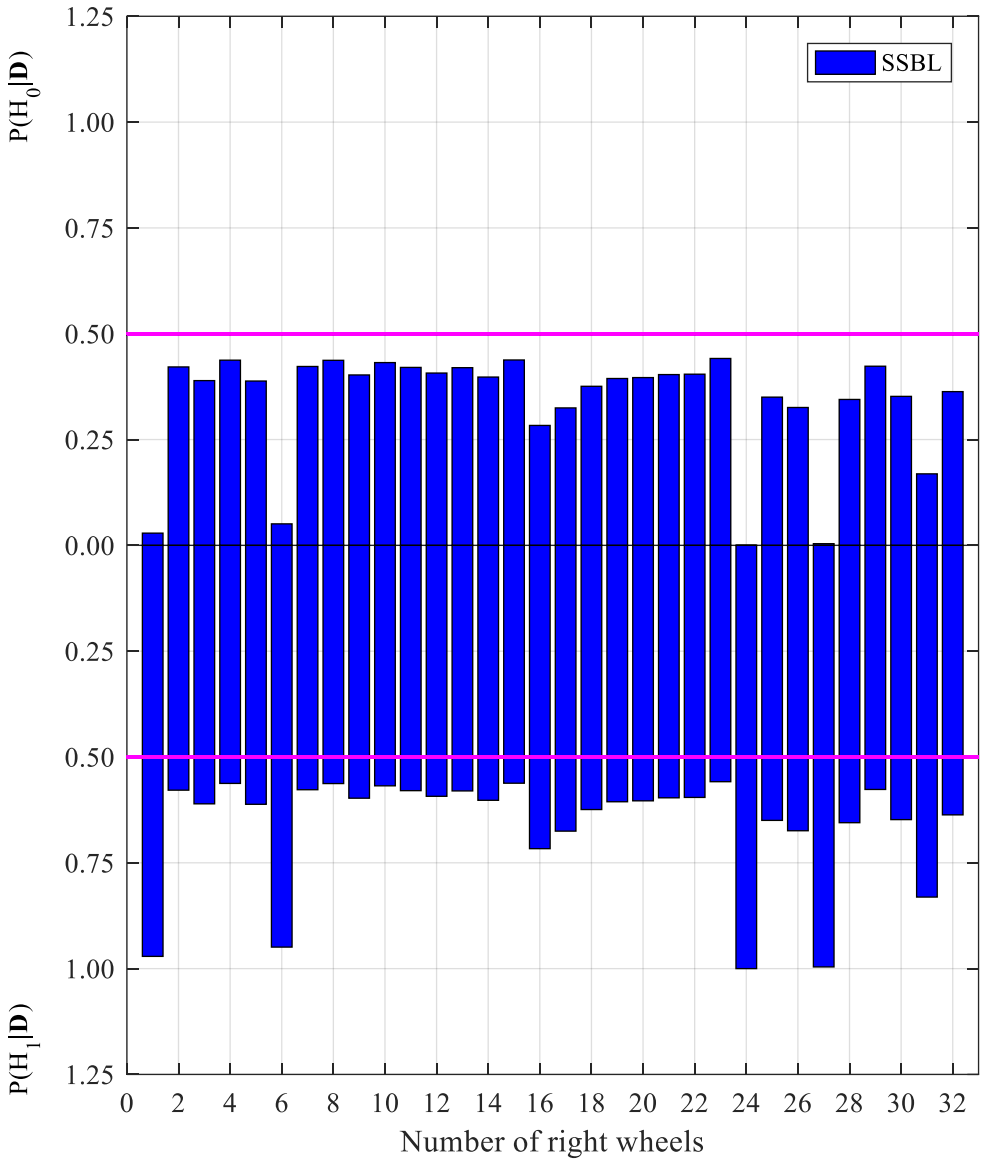


Figure 6.45 Posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (SSBL and Bayesian PNHT)

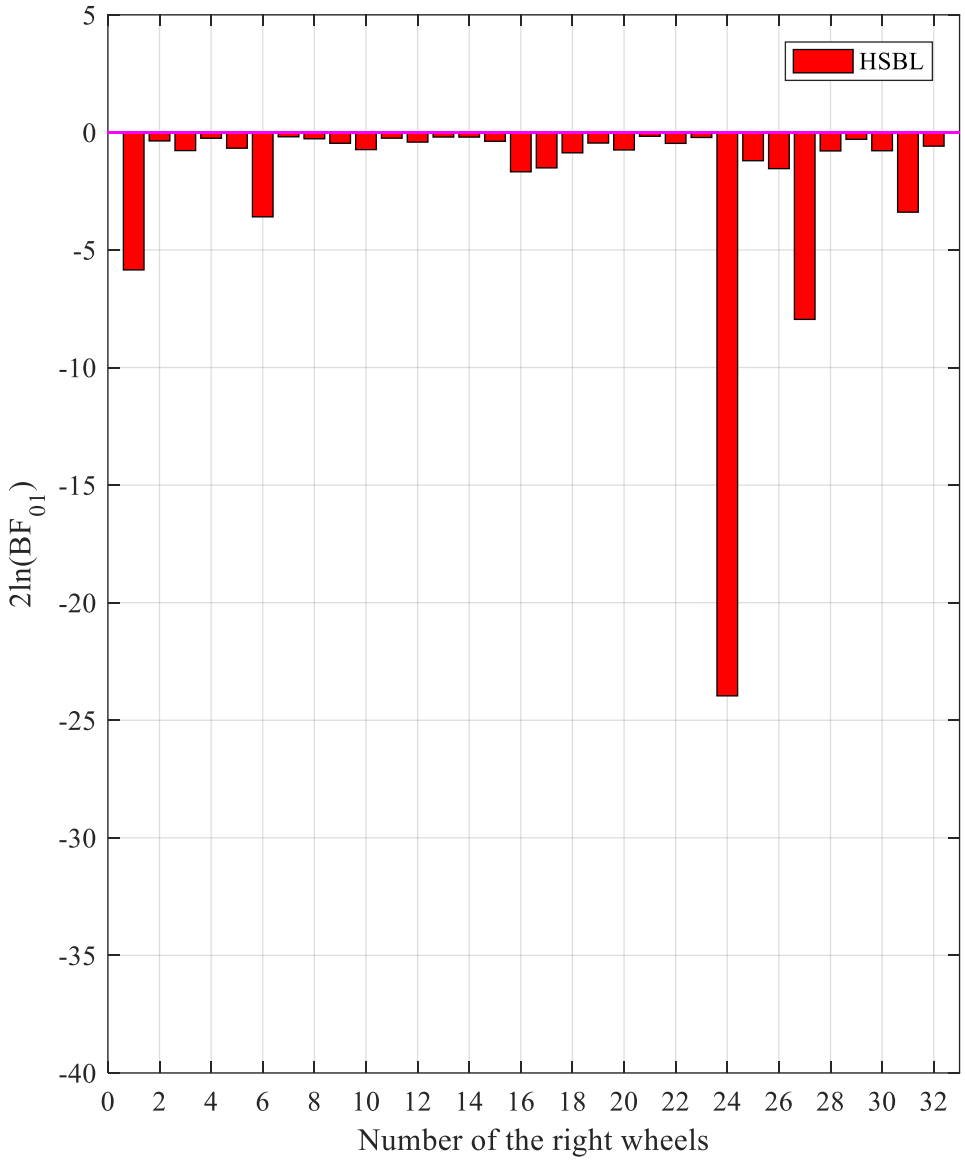


Figure 6.46 Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (HSBL and Bayesian PNHT)

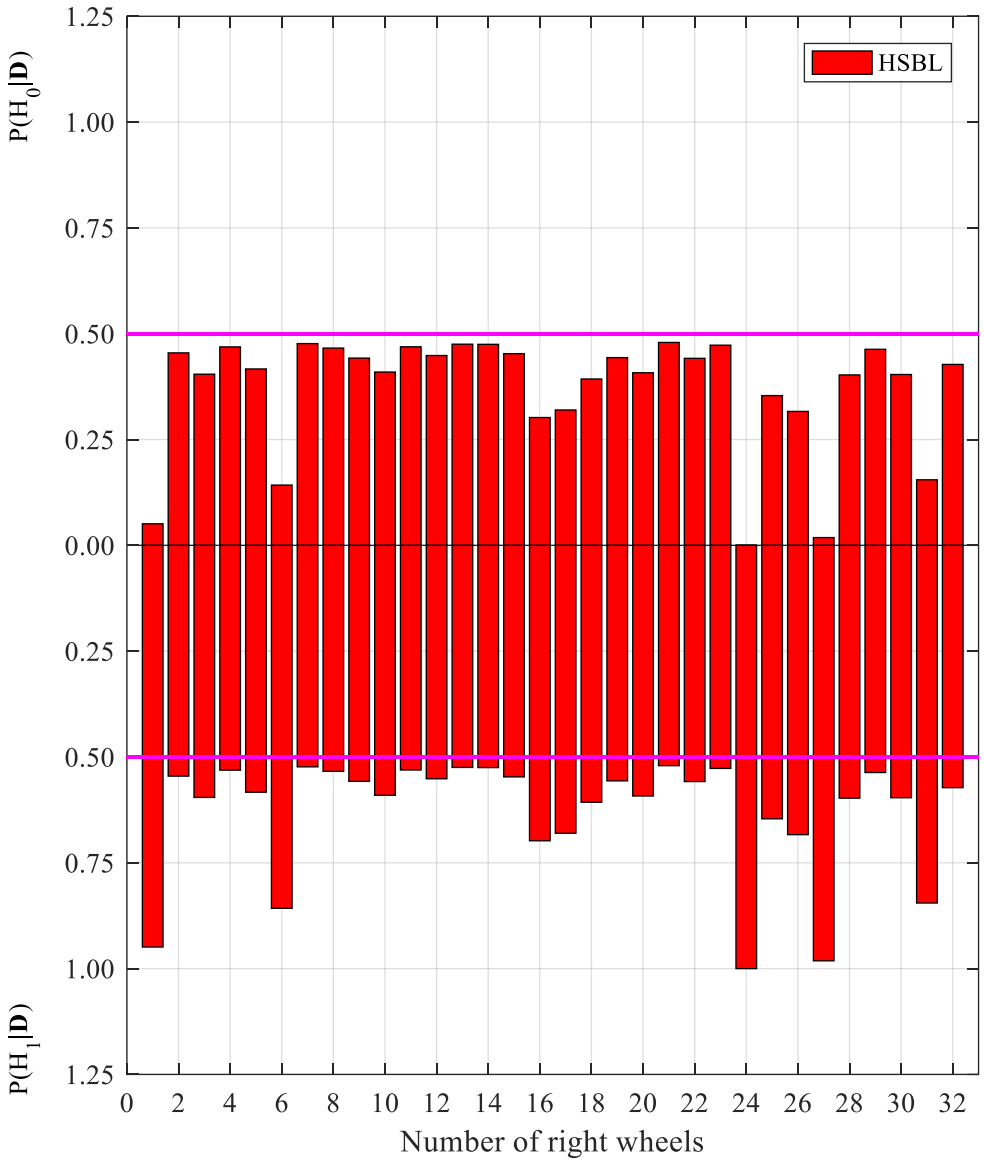


Figure 6.47 Posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (HSBL and Bayesian PNHT)

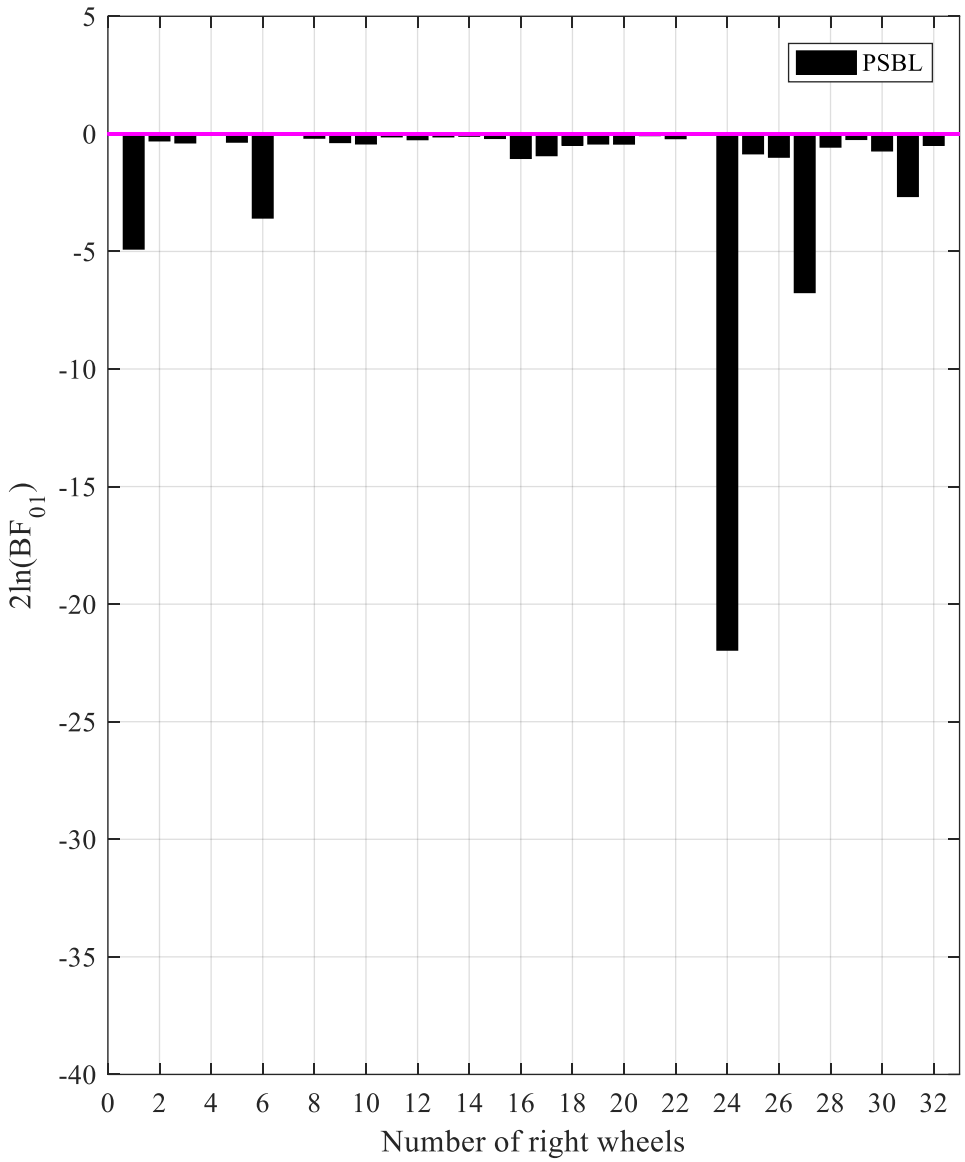


Figure 6.48 Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (PSBL and Bayesian PNHT)

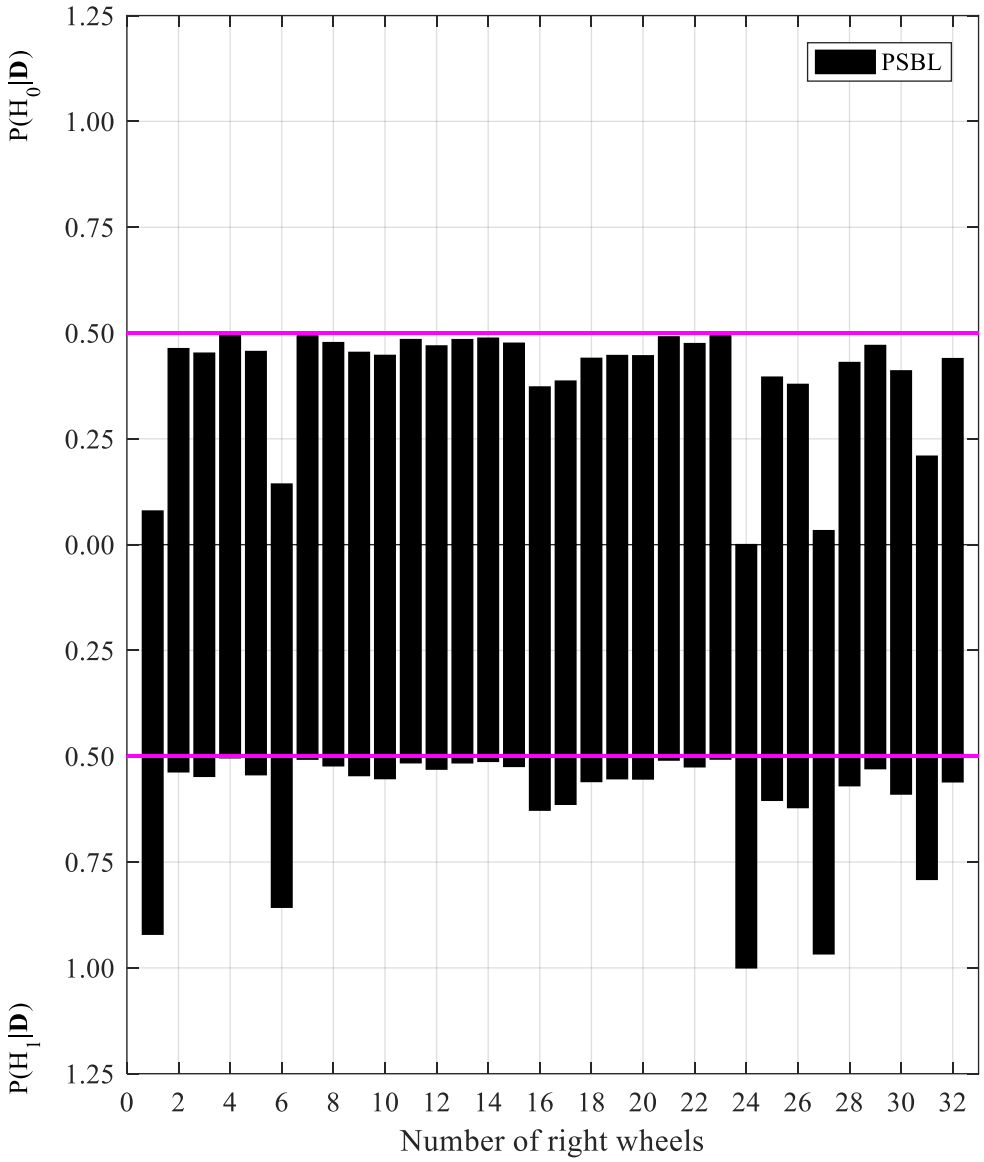


Figure 6.49 Posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (PSBL and Bayesian PNHT)

c) ***Bayesian NHST results***

The diagnostic results of left wheels using the Bayesian NHST on the monitoring data from all 21 sensors deployed on the left rail track are shown in Figures 6.50 to 6.55, with the population feature models established by SSBL, HSBL and PSBL, respectively. It is found that regardless of which type of population feature models, the intrinsic Bayes factors associated with the 1st, 6th and 27th left wheels are all negative and their corresponding posterior probabilities of defect are more than 50.0%. Thus, the three left wheels are diagnosed as defective. Other left wheels are diagnosed as undefective as their intrinsic Bayes factors are all positive and their posterior probabilities of defect are all less than 50.0%.

To be specific, when using the SSBL population feature model, the intrinsic Bayes factors associated with the 1st, 6th and 27th left wheels are respectively -12.6, -8.2 and -33.6, and their corresponding posterior probabilities of defect are 99.8%, 98.4% and 100.0%, respectively. Hence, the three left wheels are diagnosed respectively as heavily, moderately and heavily defected. If using the HSBL population feature model, the intrinsic Bayes factors of the three left wheels are respectively -9.2, -5.2 and -32.6, and their corresponding intrinsic posterior probabilities of defect are 99.0%, 93.0% and 100.0%, respectively. Thus, the three left wheels are diagnosed as moderately, mildly and heavily defected, respectively. If using the PSBL population feature model, the intrinsic

Bayes factors with respect to the 1st, 6th and 27th left wheels are respectively -5.8, -1.8 and -28.5, and their posterior probabilities of defect are 94.9%, 71.2% and 100.0%, respectively. Therefore, the three left wheels are diagnosed as mildly, weakly and heavily defected, respectively. Though quantitative differences can be found in the diagnostic results for left wheels based on different categories of population feature models, no qualitative differences exist in the ultimate condition classification of the left wheels.

The diagnostic results of right wheels using the Bayesian NHST on the monitoring data from all 21 sensors deployed on the right rail track are shown in Figures 6.56 to 6.61, with the population feature models established by SSBL, HSBL and PSBL, respectively. It is seen that when using the SSBL population feature model, the intrinsic Bayes factors of the 1st, 6th, 24th, 27th and 31st right wheels are respectively -8.6, -5.7, -46.0, -16.7 and -0.44, with the corresponding intrinsic posterior probabilities of defect being 98.7%, 94.5%, 100.0%, 100.0% and 55.5%, respectively. Thus, the five right wheels are diagnosed respectively as moderately, mildly, heavily, heavily and weakly defected. Other right wheels are diagnosed as undefective as their intrinsic Bayes factors are positive and their intrinsic posterior probabilities of defect are less than 50.0%. By contrast, if using the HSBL population feature model, only four right wheels are diagnosed as defected, including the 1st, 6th, 24th and 27th wheels. The intrinsic Bayes factors for the four wheels are respectively -6.6, -2.5, -40.0, and -13.5 and their corresponding intrinsic

posterior probabilities of defect are respectively 96.4%, 77.7%, 100.0% and 99.9%. Thus, the four right wheels are diagnosed respectively as moderately, mildly, heavily and heavily defected. Similarly, if using the PSBL population feature model, only the 1st, 6th, 24th and 27th right wheels are diagnosed as defected, whereas their associated intrinsic Bayes factors and posterior probabilities of defect increase to -4.0, -2.0, -36.0 and -8.8, and 88.1%, 73.3%, 100.0% and 98.8%, respectively. Therefore, the four right wheels are identified respectively as mildly, slightly, heavily and moderately defected when using the PSBL population feature model. Both quantitative and qualitative differences are found in the condition assessment of right wheels if using different types of population feature models.

Different from Bayesian PNHT, the diagnostic results from frequentist NHST and Bayesian NHST are again coincided in the condition classification of wheels even though the monitoring data from all deployed sensors are used. For instance, the 1st, 6th and 27th left wheels are diagnosed as defected in both frequentist NHST (Figures 6.32 to 6.34) and Bayesian NHST (Figures 6.50 to 6.55), regardless of the types of the population feature models. On the contrary, almost all left wheels, except the 26th and 30th left wheels (Figures 6.38 to 6.43), are identified as defected based on Bayesian PNHT. The diagnostic results of right wheels are also consistent between frequentist NHST and Bayesian NHST, with the 1st 6th 24th 27th and 31st right wheels identified as defected when using the

SSBL population feature model (Figure 6.35, and Figures 6.56 and 6.57), and with the 1st 6th 24th and 27th right wheels diagnosed as defected if using the non-standard population feature models (Figures 6.36 and 6.37, and Figure 6.58 to 6.61). However, all right wheel wheels are diagnosed as defected if using the Bayesian PNHT on the three categories of population feature models. As this case study shows, the Jeffrey-Lindley paradox can be indeed eliminated by making a refined alternative hypothesis in Bayesian hypothesis testing. The refined alternative hypothesis allows the significance level α (the type-I error or diagnostic risk) associated with Bayesian hypothesis testing to be explicitly controlled as the same as the level in frequentist NHST. By contrast, the Jeffrey-Lindley paradox cannot be avoided in Bayesian PNHT as the induced diagnostic risk it gives rise to is not explicitly controllable.

In addition, it is found that the strength of evidence to support the null hypothesis (the wheel is with defect) and alternative hypothesis (the wheel is without defect) is comparable in Bayesian NHST. There is both very strong evidence to support the hypotheses that the wheel is with or without defect. By contrast, there is no overwhelming evidence to support the claim that the wheel is without defect in Bayesian PNHT.

Comparing Figures 6.32 to 6.37 with Figures 6.50 to 6.61, it is seen that direct evidence (the intrinsic posterior probability of defect) in Bayesian NHST is obtained for the condition assessment of wheels, whereas p-values in frequentist NHST can only be

served as indirect evidence to support claims about wheel conditions. More importantly, the extent of defect can be quantified more easily in Bayesian NHST, in terms of intrinsic Bayes factors, which is difficult in frequentist NHST.

By comparing Figures 6.20 to 6.31 with Figures 6.50 to 6.61, it is found that using only the monitoring data from a single sensor fails to identify some potentially defective wheels. For example, using only monitoring data from the sensor SEN-A2 fails to identify the suspected defects on the 1st and 6th left wheels, while using only the monitoring data from the sensor SEN-D2 fails to detect the suspected defects on the 1st and 6th right wheels.

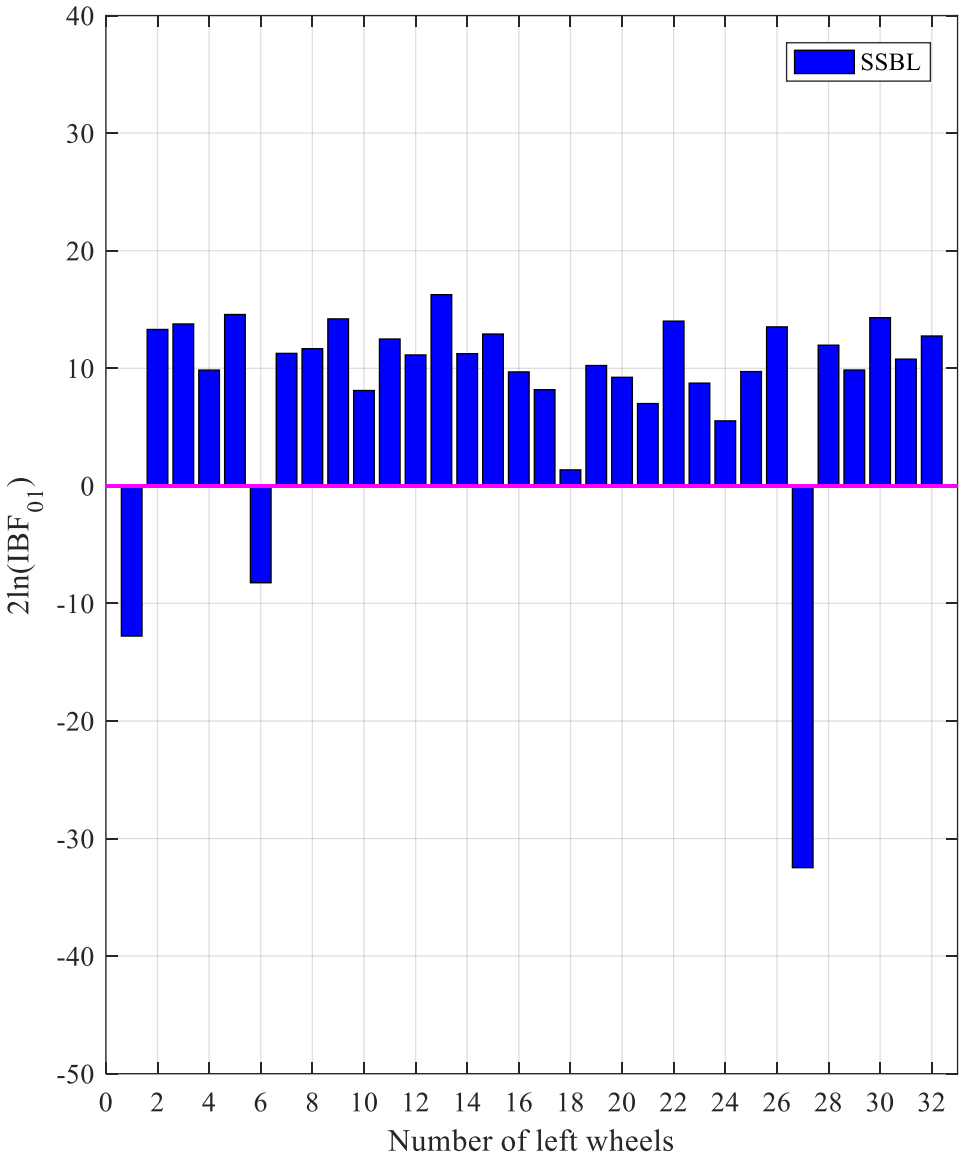


Figure 6.50 Intrinsic Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (SSBL and Bayesian NHST)

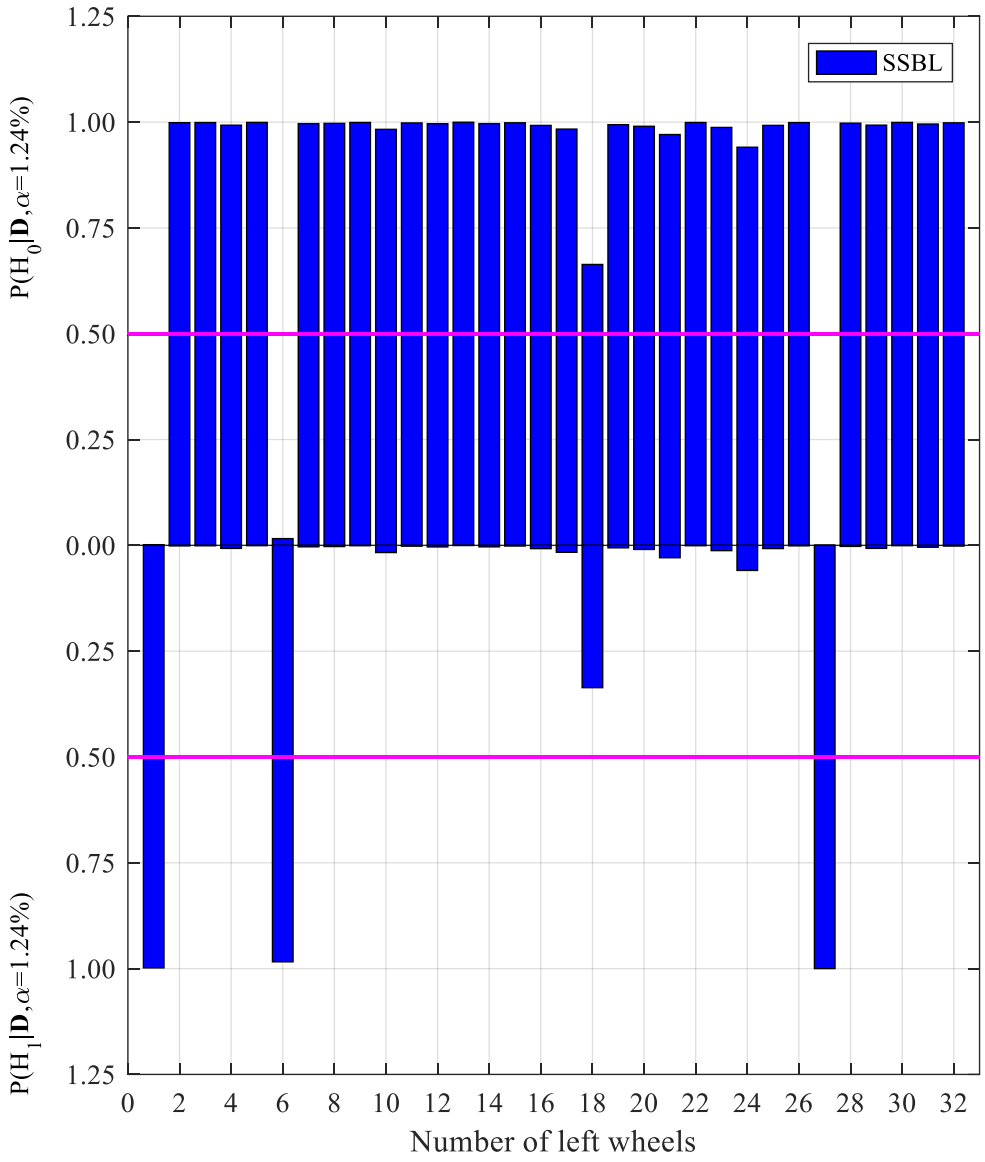


Figure 6.51 Intrinsic posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (SSBL and Bayesian NHST)

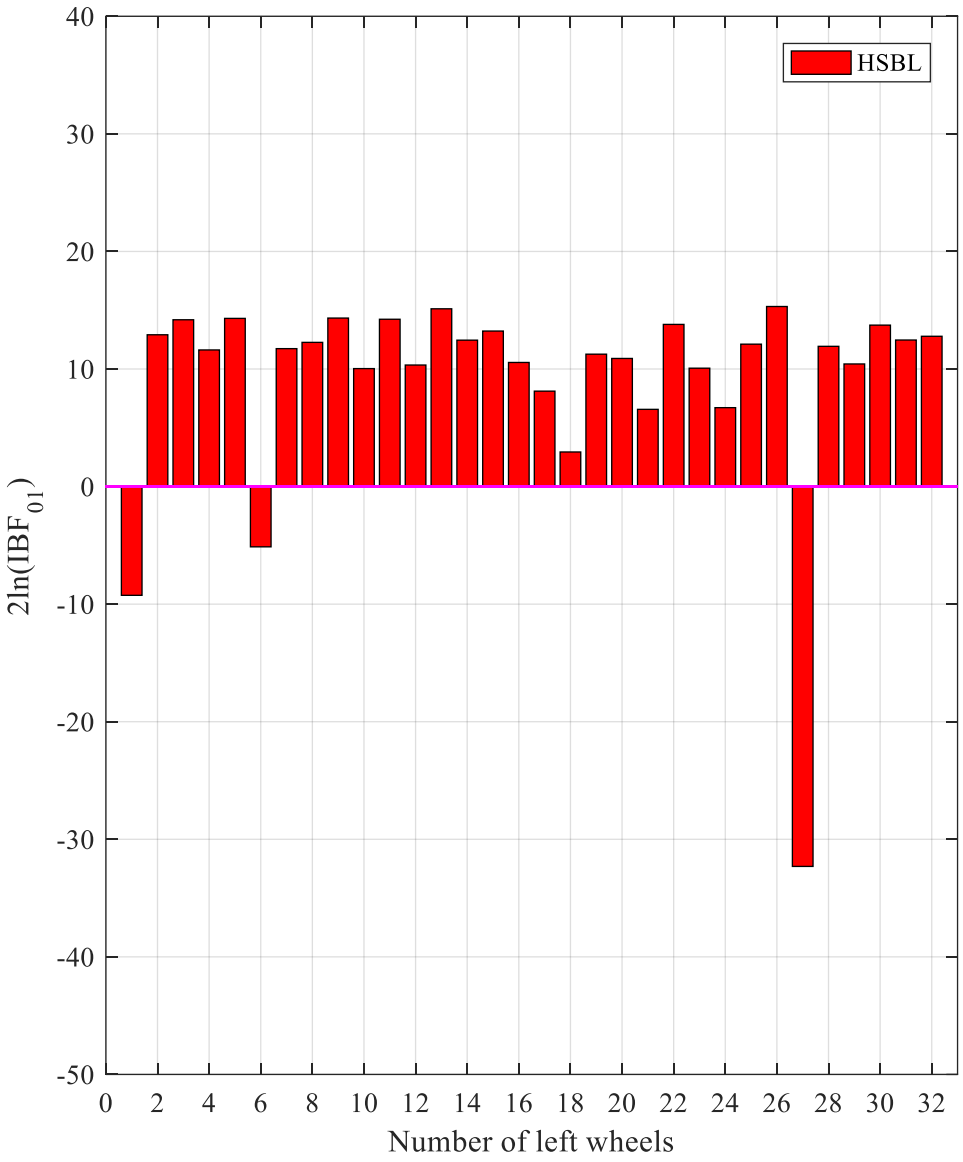


Figure 6.52 Intrinsic Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (HSBL and Bayesian NHST)

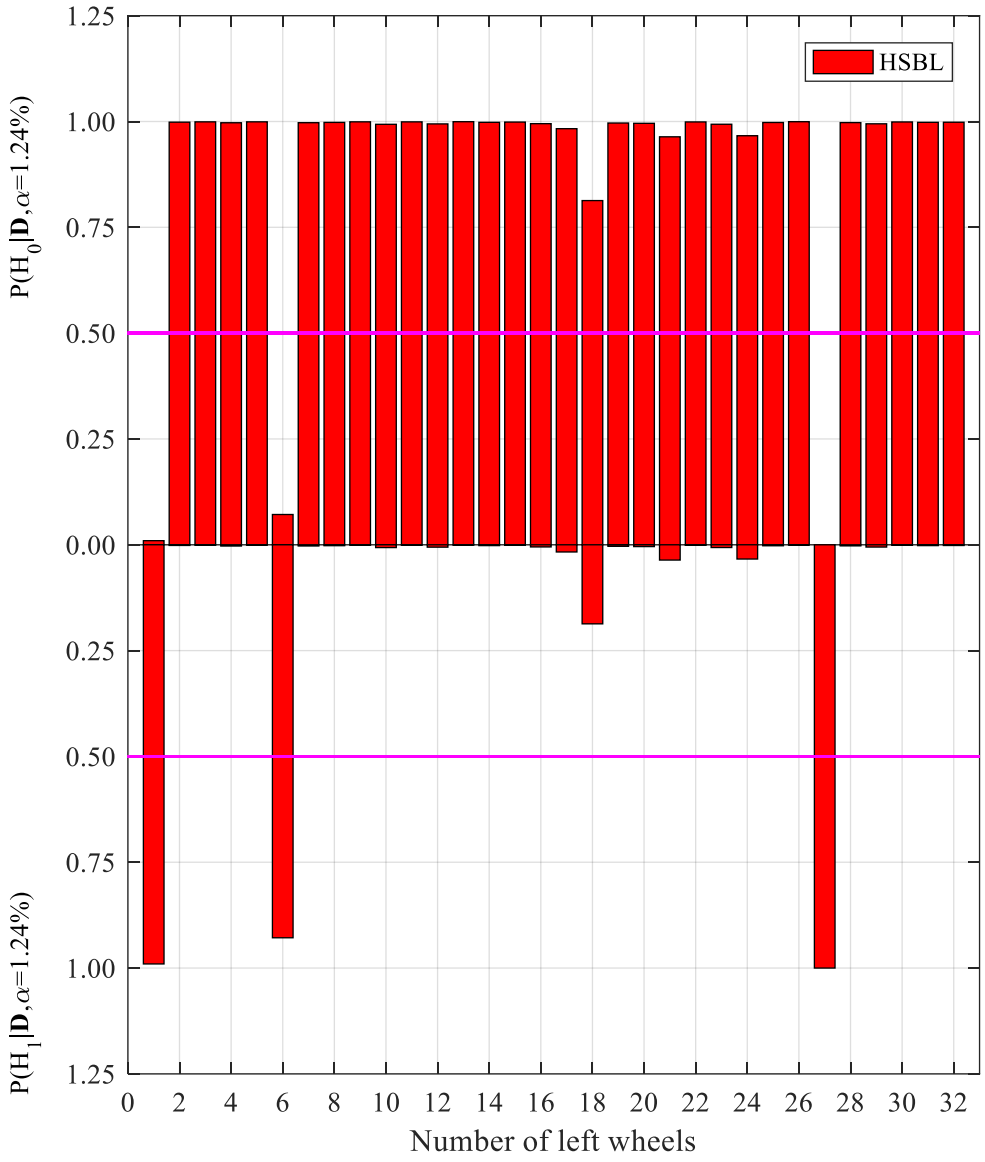


Figure 6.53 Intrinsic posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (HSBL and Bayesian NHST)

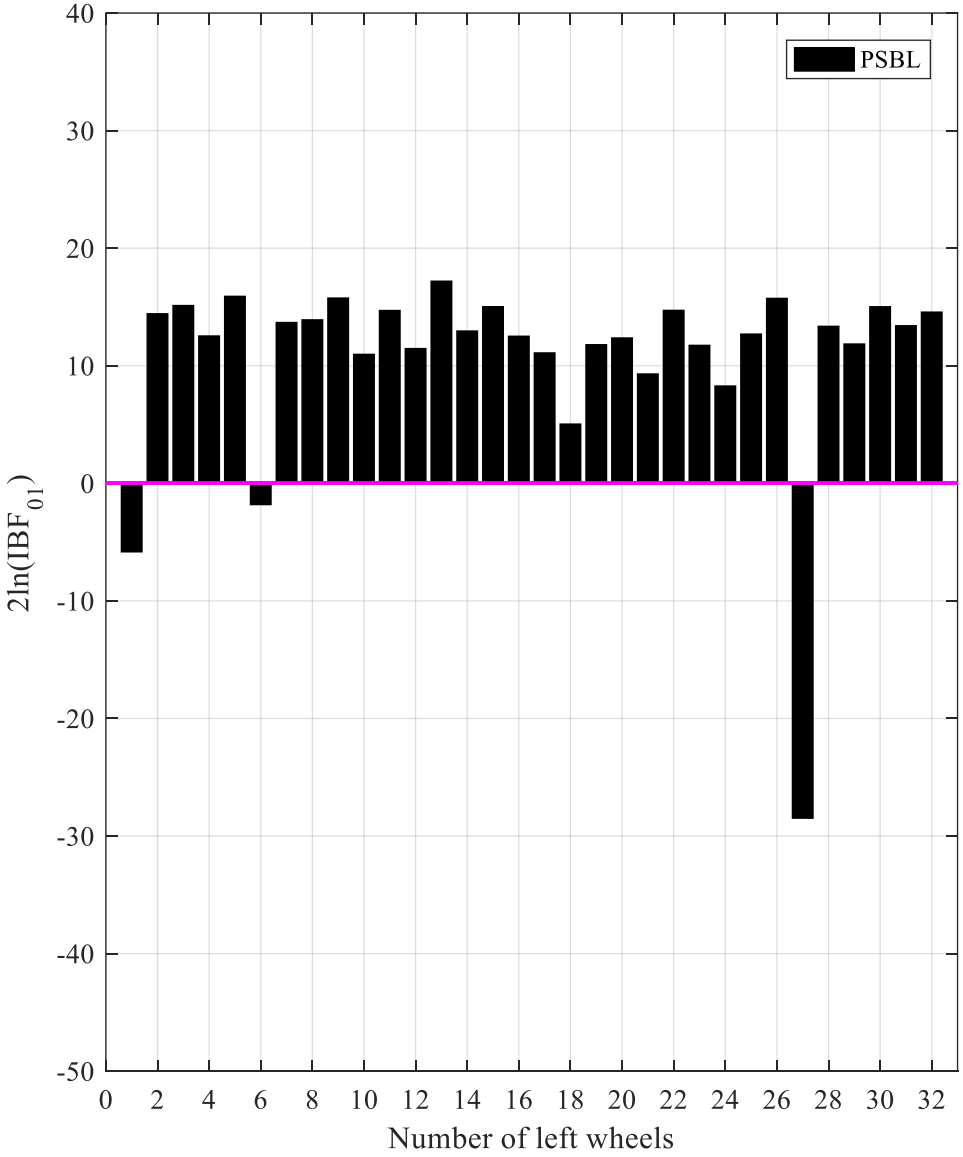


Figure 6.54 Intrinsic Bayes factors of left wheels using monitoring data from all sensors deployed on left rail track (PSBL and Bayesian NHST)

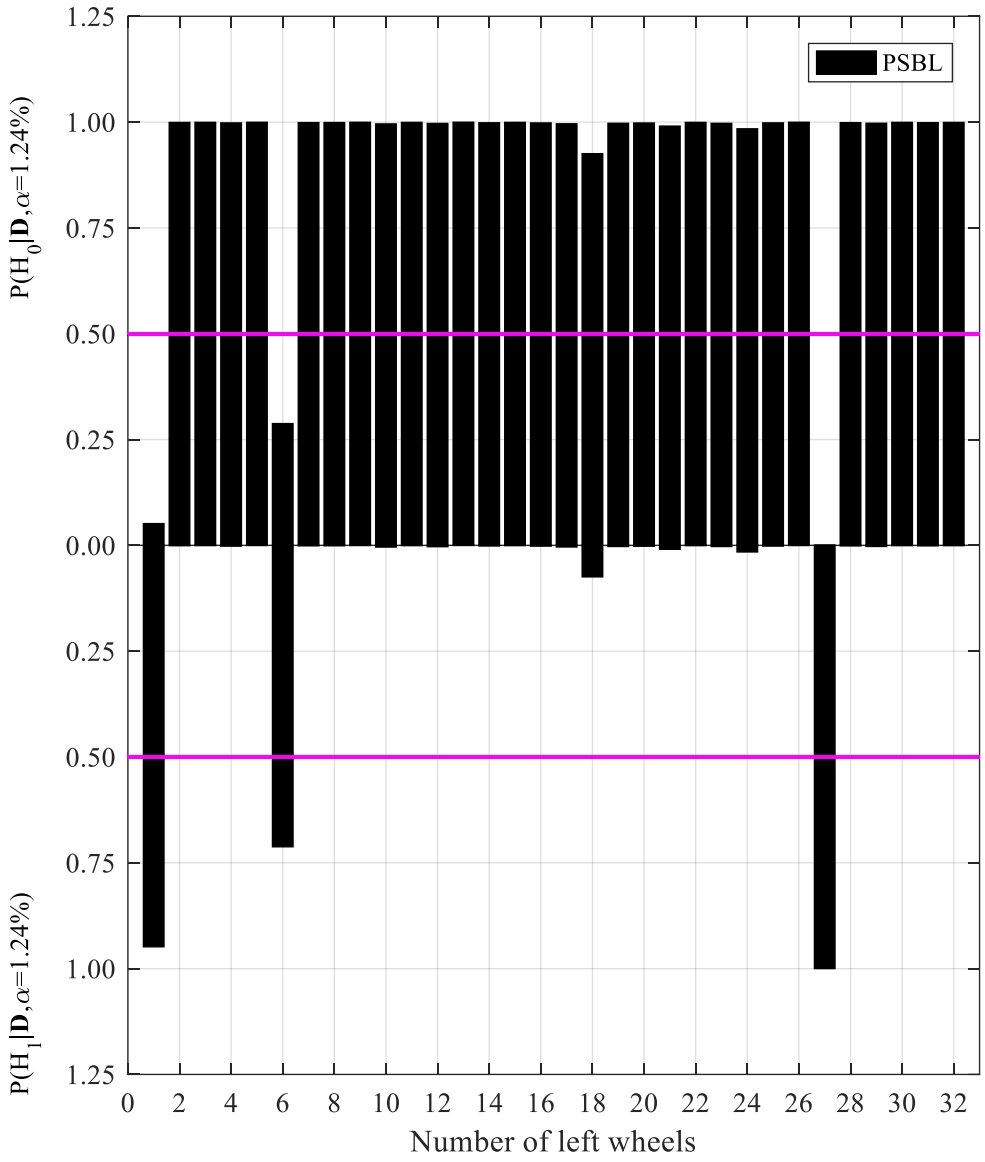


Figure 6.55 Intrinsic posterior probabilities of left wheels using monitoring data from all sensors deployed on left rail track (PSBL and Bayesian NHST)

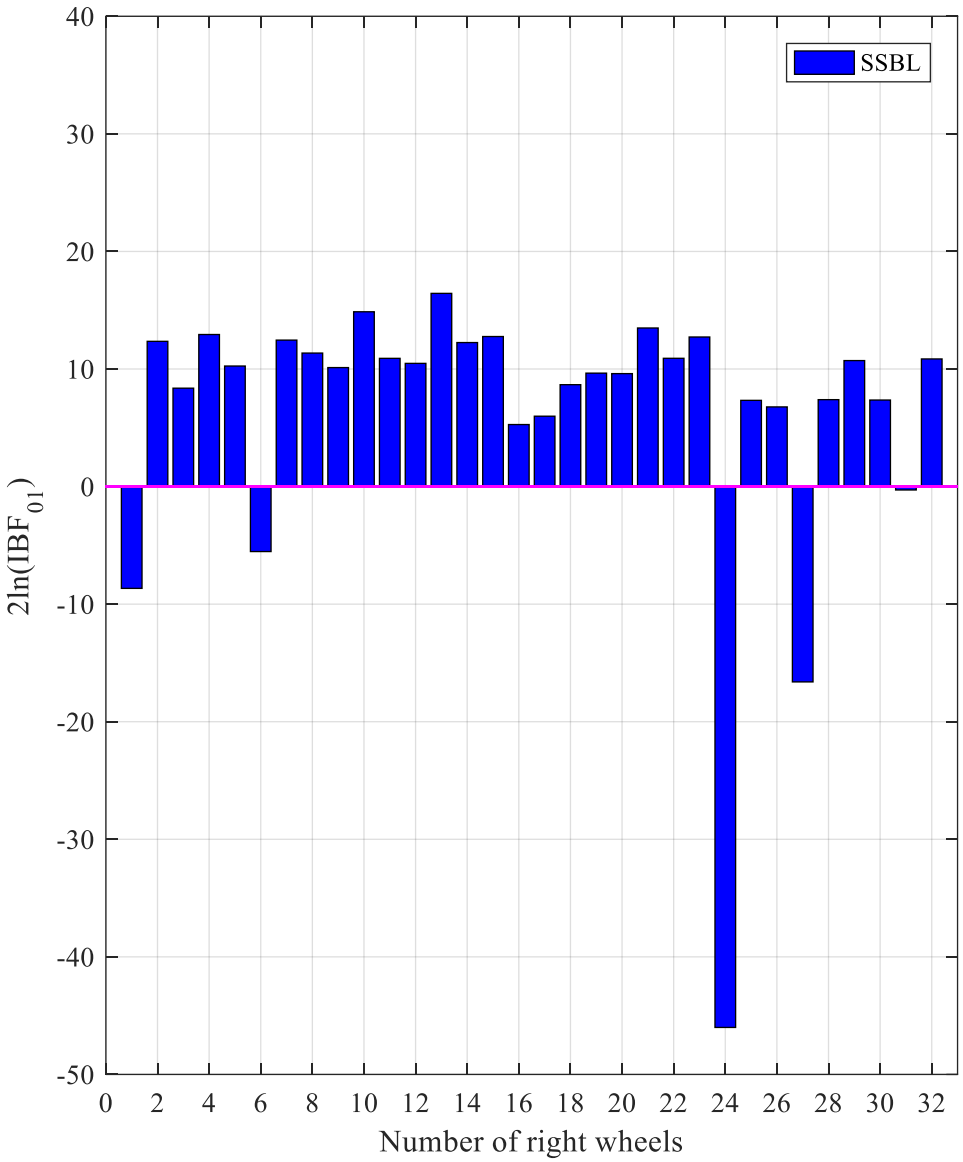


Figure 6.56 Intrinsic Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (SSBL and Bayesian NHST)

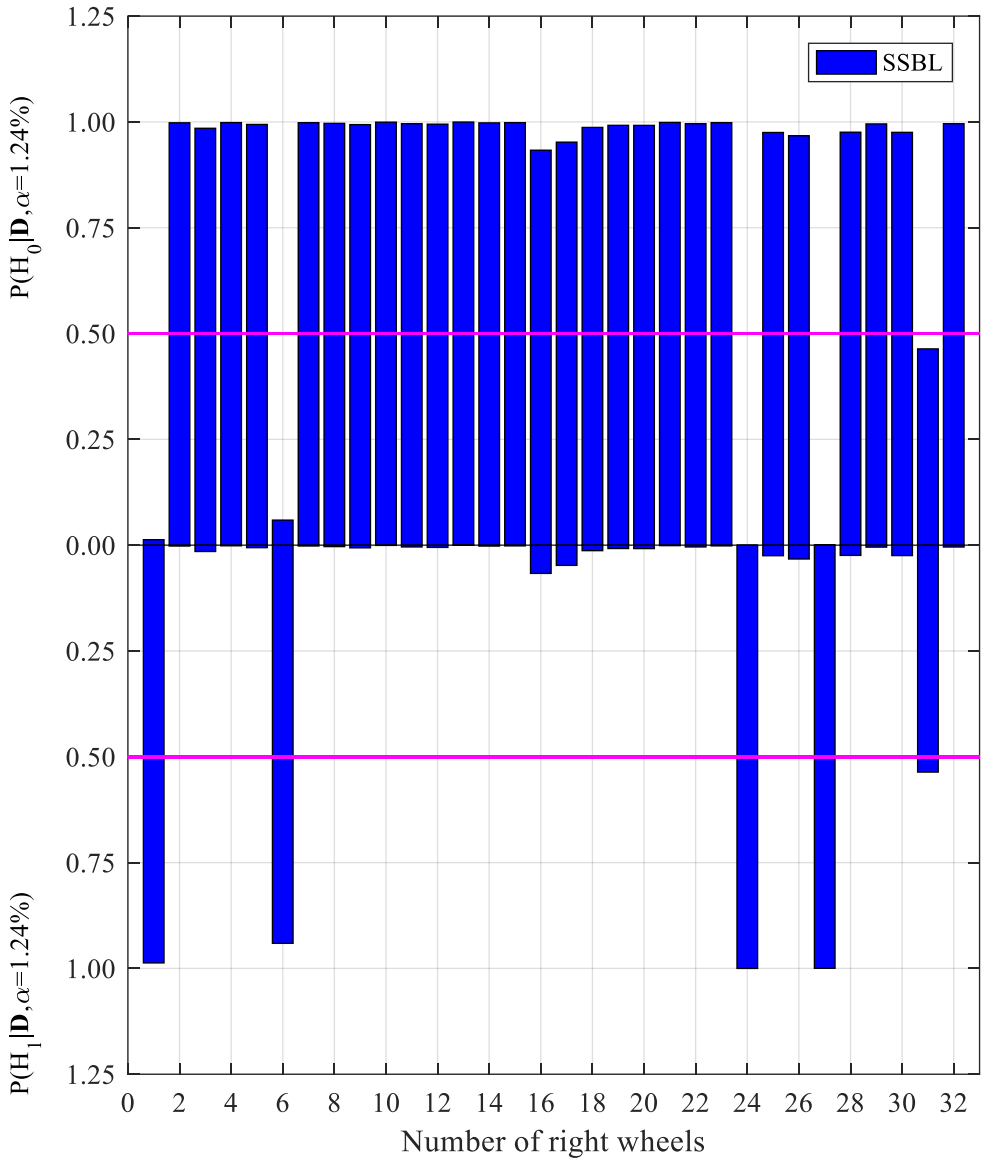


Figure 6.57 Intrinsic posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (SSBL and Bayesian NHST)

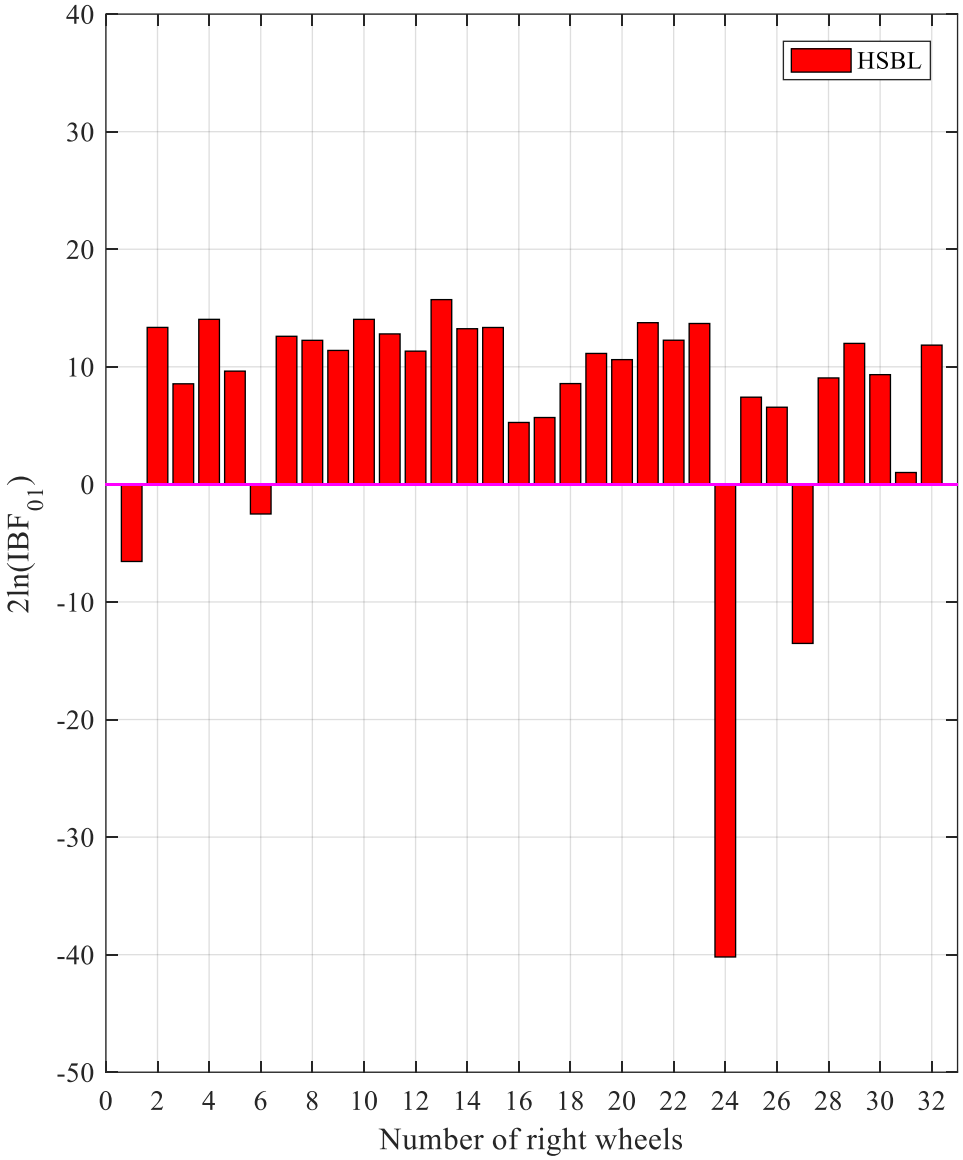


Figure 6.58 Intrinsic Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (HSBL and Bayesian NHST)

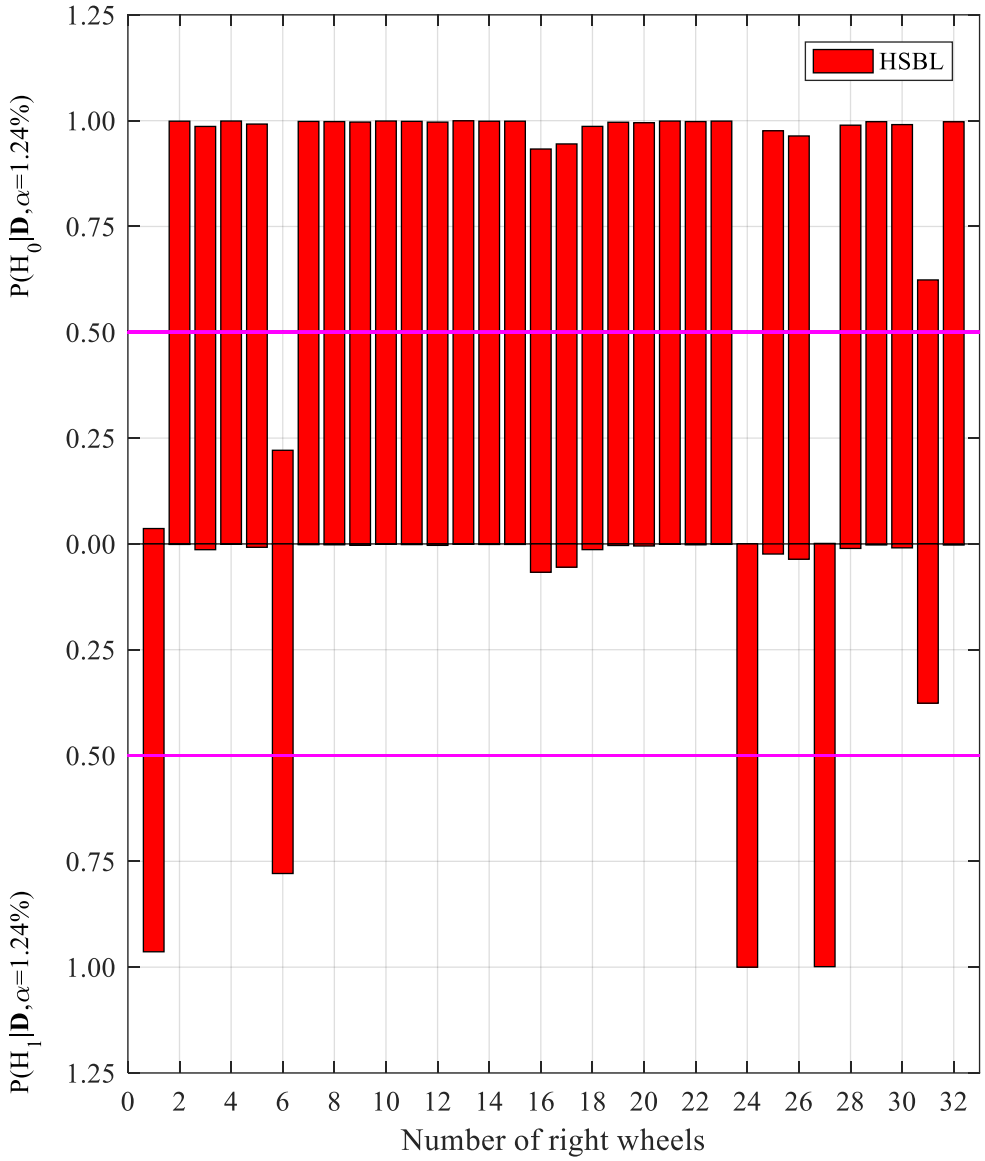


Figure 6.59 Intrinsic posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (HSBL and Bayesian NHST)

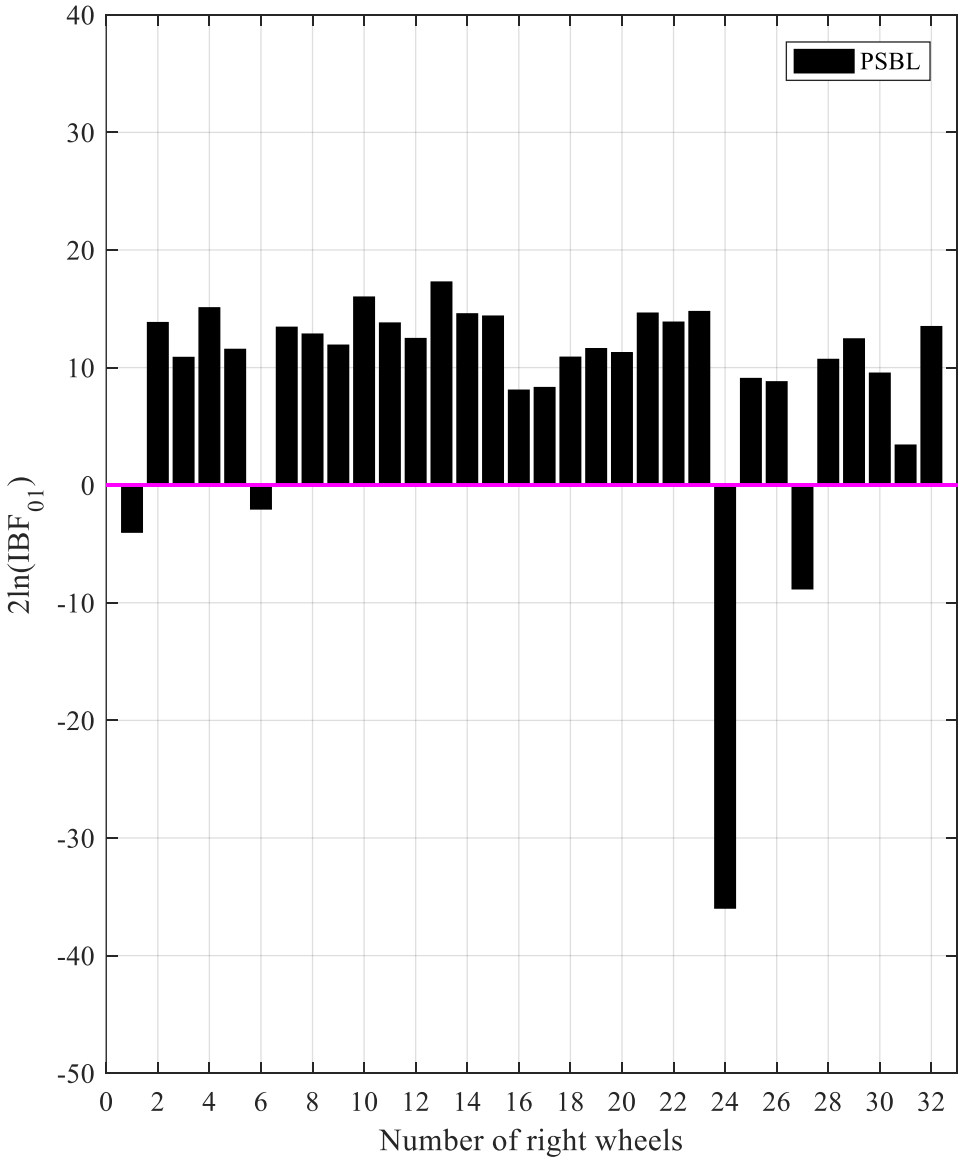


Figure 6.60 Intrinsic Bayes factors of right wheels using monitoring data from all sensors deployed on right rail track (PSBL and Bayesian NHST)

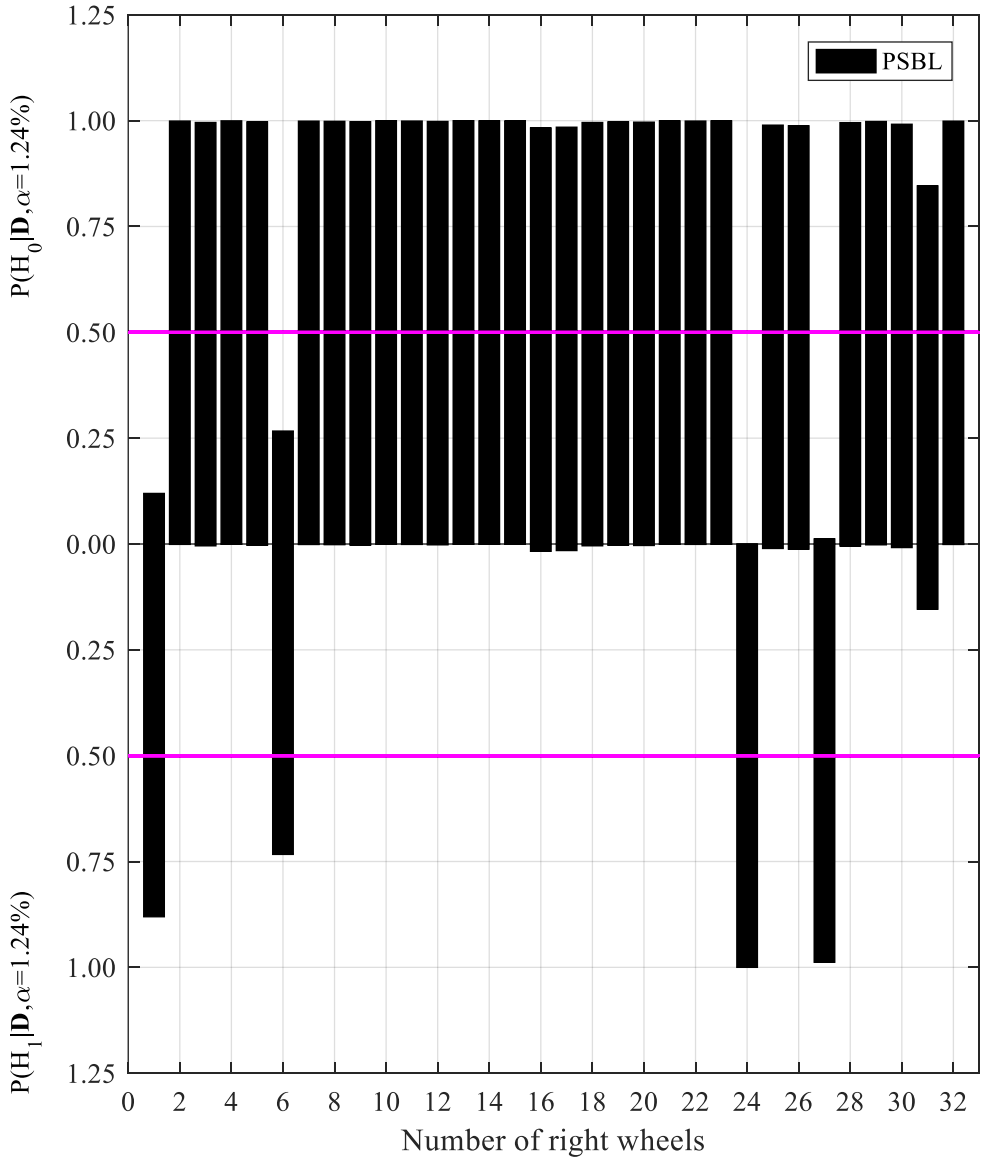


Figure 6.61 Intrinsic posterior probabilities of right wheels using monitoring data from all sensors deployed on right rail track (PSBL and Bayesian NHST)

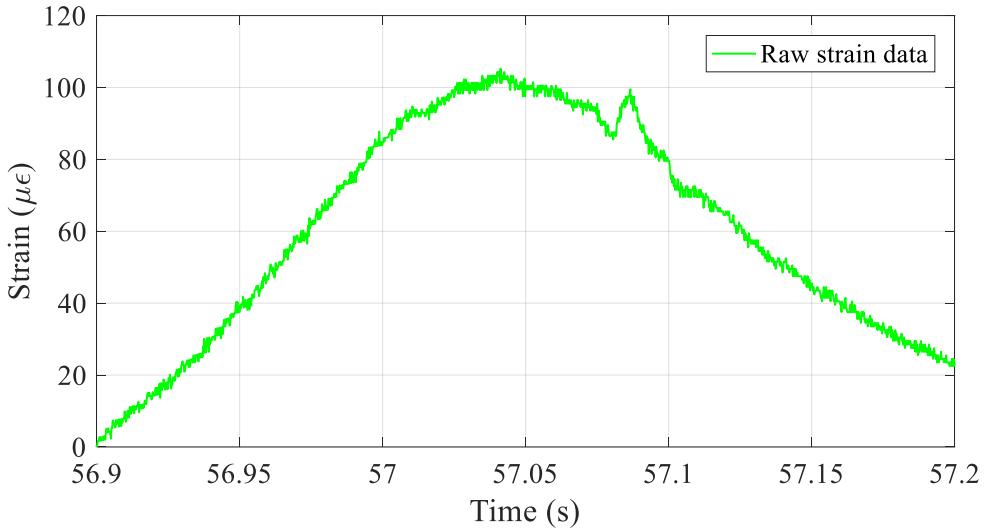


Figure 6.62 Measured signal when the 24th right wheel passes through the location of the sensor SEN-D2.

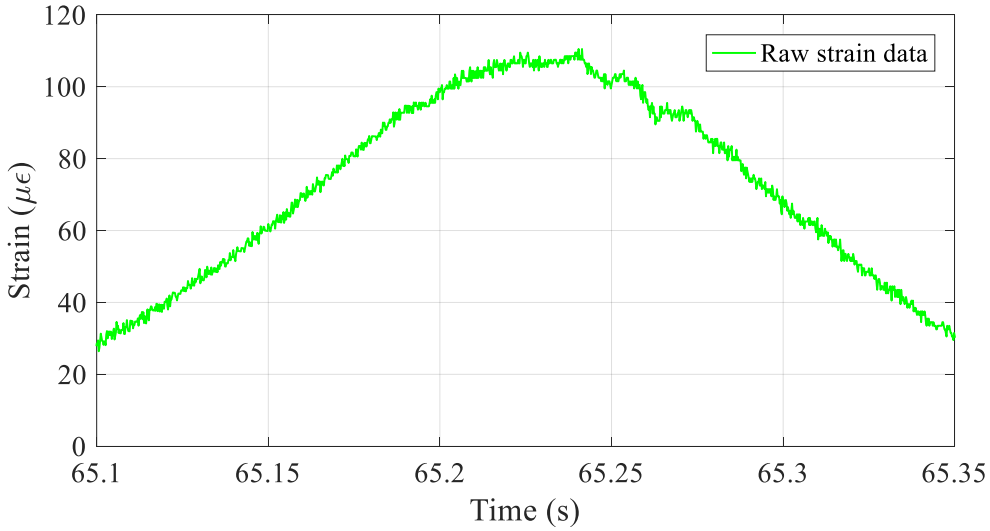


Figure 6.63 Measured signal when the 27th right wheel passes through the location of the sensor SEN-D2.

Figures 6.62 and 6.63 show the measurement segments of rail strain recorded by SEN-D2, when the 24th and the 27th right wheels pass through the location of the sensor. Visible disturbances are found in the measured signals of rail strain, due to the wheel defect-incurred impact on the rail track, with a more serious disturbance observed in the measured signal of rail strain for the 24th right wheel.

d) ***Comparison to blind test results***

To validate the effectiveness of the proposed online diagnostic methods for nominally identical railway wheels, offline inspection is conducted afterwards in a train depot. The offline wheel tread profile measurement indicates that among the sixty-four wheels, seven wheels are in fact defected, including the 1st, 6th and 27th left wheels, and the 1st, 6th, 24th and 27th right wheels. The defective wheels are listed in Tables 6.5 and 6.6, with the measured defect lengths shown in parentheses.

It is found that the 27th left wheel and 24th right wheel are with larger flats, with defect lengths being longer than the allowed tolerance limit (30mm in length), regulated by *Operation and Maintenance Regulations for Railway Electric Multiple Units* (TG/CL 127-2017). Thus, the two railway wheels should be reprofiled, immediately. Though an immediate reprofiling is not required to be conducted on other defective wheels with smaller flats (less than 30mm in length), close attention should be paid on them to monitor the deterioration of their conditions. No special measures are required to be taken on the undefective wheels.

By comparing the diagnostic results from the offline inspection and the online monitoring, it is found that the condition assessment results of these wheels from Bayesian PNHT are generally non-informative as it gives rise to too many diagnostic errors (many undefective wheels are falsely identified as defective), whenever using the

monitoring data from a single sensor or from all sensors deployed on the rail track. This is mainly attributed to theoretical problems in Bayesian PNHT (e.g. inability to use the same prior as used for inference; the Jeffreys-Lindley paradox) and therefore, it is concluded that the current Bayesian PNHT is not a good diagnostic methodology for the development of a general unsupervised damage detection framework, specially targeted at a group of nominally identical structures.

On the contrary, the condition assessment results from frequentist NHST and Bayesian NHST are quite informative. When using the monitoring data from all sensors deployed on the rail track, all defective wheels are successively identified, at the price of a very low false-positive error rate. Among the fifty-seven undefective wheels, only one wheel (the 31st right wheel) is falsely diagnosed as defected when using the SSBL population feature model. If using the non-standard SBL population feature models, no false-positive error occurs. It should be mentioned that using only the monitoring data from a single sensor fails to identify all defective wheels and more accurate diagnostic results can be obtained when using the monitoring data from all the deployed sensors.

Though the condition classification results (qualitative assessment) of the sixty-four wheels are exactly the same between frequentist NHST and Bayesian NHST, the Bayesian diagnostic methodology provides us with a more useful tool to evaluate the extents of wheel defects (quantitative assessment). For example, the extents of the wheel defects are

difficult to measure in frequentist NHST, where they can be quantitatively quantified easily in terms of intrinsic Bayes factors in Bayesian NHST (the extents of defects on the suspected defective wheels are given in parentheses in Tables 6.1 and 6.2). As a result, a more pertinent maintenance strategy can be followed if using Bayesian NHST to assess the extents of defects on the concerned wheels, which is impossible if using frequentist NHST for the identification of defects on wheels.

It should be mentioned the using the classification criterion on discrimination strength as shown in Table 6.2 to quantitatively assess the extents of wheel defects is a little bit subjective because of no physical implication. However, we may find in this case study that the classification intervals of the obtained intrinsic Bayes factors happen to perfectly match the severities of wheel defects if using the PSBL population feature models. When using the PSBL population feature models, the online diagnostic results suggest that the 27th left wheel and 24th right wheel are heavily defected, while other suspected defective wheels are diagnosed as defected in lower severity grades. The online diagnostic results are found to perfectly match the offline inspection results, with the 27th left wheel and 24th right wheel found to be with flats over 30mm in length and other suspected defective wheels to be with smaller flats (under 30mm in length). The other two categories of population feature models turn out to overestimate the extents of defects to some extent. Therefore, we may conclude that the PSBL population feature model used in the baseline

phase and the Bayesian NHST used in the inspection phase can be the best option for developing a general unsupervised damage detection framework, specially targeted at a group of nominally identical structures. Even if the classification intervals do not well match the extents of wheel defects, they can be calibrated and pertinently refined to provide more precise mapping between the ranges of intrinsic Bayes factors and the extents of defects.

It also should be mentioned that the choice of the level of significance α is critical to defect detection results. In the present study, the significance level α is set to be 1.24% in both frequentist NHST and Bayesian NHST, at the 1.24% diagnostic risk that a healthy wheel is falsely diagnosed as defective. When more offline measurement results of defective wheels with different defect extents are available, a more appropriate choice of the significance level α is probably made.

Table 6.5 Comparison of defective wheels identified from offline inspection and online monitoring (on the left rail track)

Offline inspection results	Wheel profile measurement system			27th (34.6mm); 1st (23.7mm); 6th (18.3mm);
Online monitoring results	A single sensor deployed on the left rail track (e.g. SEN-A2)	Frequentist NHST	SSBL	27th;
			HSBL	27th;
			PSBL	27th;
		Bayesian PNHT	SSBL	27th (heavily); 1st, 6th, 13th, 19th, 20th, 29th and 30th (weakly);
			HSBL	27th (heavily); 1st, 3rd 6th, 13th, 19th, 20th, 29th and 30th (weakly);
			PSBL	27th (heavily); 1st, 6th, 19th, 20th, 29th and 30th (weakly);
		Bayesian NHST	SSBL	27th (heavily);
			HSBL	27th (heavily);
			PSBL	27th (heavily);
	All 21 sensors deployed on the left rail track	Frequentist NHST	SSBL	1st, 6th and 27th;
			HSBL	1st, 6th and 27th;
			PSBL	1st, 6th and 27th;
		Bayesian PNHT	SSBL	27th (heavily); 1st and 6th (moderately); 18th (mildly); All others (Weakly);
			HSBL	27th (heavily); 1st (moderately); 4th and 18th (mildly); All others (weakly);
			PSBL	27th (heavily); 1st and 6th (mildly); All others excluding the 9th, 26th and 30th (weakly);
		Bayesian NHST	SSBL	1st and 27th (heavily); 6th (moderately);
			HSBL	27th (heavily); 1st (moderately); 6th (mildly);
			PSBL	27th (heavily); 1st (mildly); 6th (weakly);

Table 6.6 Comparison of defective wheels identified from offline inspection and online monitoring (on the right rail track)

Offline inspection results	Wheel profile measurement system			24th (36.0mm); 27th (26.9mm); 1st (21.4mm); 6th (17.5mm);
Online monitoring results	A single sensor deployed on the left rail track (e.g. SEN-A2)	Frequentist NHST	SSBL	24th and 27th;
			HSBL	24th and 27th;
			PSBL	24th and 27th;
		Bayesian PNHT	SSBL	24th (heavily); 27th (moderately); 4th, 6th, 9th, 26th, 28th and 31st (weakly);
			HSBL	24th (heavily); 27th (mildly); 4th, 6th, 9th, 13th, 26th, 28th and 31st (weakly);
			PSBL	24th (heavily); 27th (mildly); 4th, 6th, 9th, 26th, 28th and 31st (weakly);
		Bayesian NHST	SSBL	24th (heavily); 27th (moderately);
			HSBL	24th (heavily); 27th (moderately);
			PSBL	24th (heavily); 27th (mildly);
	All 21 sensors deployed on the left rail track	Frequentist NHST	SSBL	1st, 6th, 24th, 27th and 31st
			HSBL	1st, 6th, 24th and 27th
			PSBL	1st, 6th, 24th and 27th
		Bayesian PNHT	SSBL	24th and 27th (heavily); 1st and 6th (moderately); 31st (mildly); All others (weakly);
			HSBL	24th (heavily); 27th (moderately); 1st, 6th and 31st (mildly); All others (weakly);
			PSBL	24th (heavily); 27th (moderately); 1st, 6th and 31st (mildly); All others (weakly);

(To be continued)

Online monitoring results	All 21 sensors deployed on the left rail track	Bayesian NHST	SSBL	24th and 27th (heavily); 1st (moderately); 6th (mildly); 31st (weakly);
			HSBL	24th and 27th (heavily); 1st (moderately); 6th (mildly);
			PSBL	24th (heavily); 27th (moderately); 1st (mildly); 6th (weakly);

6.5 Summary

This chapter proposes a novel unsupervised diagnostic methodology for the identification and quantification of damage in a population of nominally identical structures. The diagnostic logic is based on Bayesian null hypothesis significance testing which allows the unsupervised damage detection procedure to be formulated in a fully probabilistic inference scheme, ranging from the development of statistical reference model for all nominally identical undamaged structures in the baseline phase to the probabilistic damage diagnostics of the concerned identical structures in the inspection phase. To validate its performance, a set of SHM strain data acquired from an FBG-based track-side monitoring system are used for online wheel condition assessment. Main findings are as follows: (1) the proposed damage diagnostics method allows much diagnostic capability beyond the simple identification of damage, additionally including the quantification of damage extent (in terms of the intrinsic Bayes factor) and the evaluation of the diagnostic risk (the significance level α that is the probability of a healthy structure falsely identified as damaged) it gives rise to; (2) it avoids the so-called Jeffreys-Lindley paradox in the current Bayesian damage diagnostic approaches based on Bayesian point null hypothesis testing and the attained damage assessment index is a normalized quantity, which no longer depends on the number of sample data to be tested; (3) in this case study, the diagnostic results show the choice of population feature models

for NISs in the baseline phase is critical to the subsequent damage identification in the inspection phase. The uncertainty associated with the population features of all nominally identical undamaged structures turns out to be underestimated in the SSBL population feature model, due to the pooling of multiple sources of uncertainty in SHM data, including measurement noise and structural variability. This, in turn, will overestimate the extent of damage in NISs. In serious cases, undamaged structures would be falsely diagnosed as damaged, giving rise to the false-positive diagnostic error. Although this detrimental effect can be alleviated by using the HSBL population feature model, the most reliable diagnostic results is given by the PSBL population feature model, in which multiple sources of uncertainty can be modelled separately and thus the uncertainty associated with the population features of all nominally identical undamaged structures is correctly evaluated. Though the damage detection results from SSBL proposed by Tipping (2001) show acceptable performance, they are indeed improved by the two variants of SSBL including HSBL and PSBL to some extent, as shown in Tables 6.5 and 6.6. Moreover, this study addresses the general problem of damage detection in nominally identical structures by using sparse Bayesian learning approaches, which is not limited to the studied case in which the two sources of uncertainty in wheel condition monitoring data are dominated by structural variability. Therefore, it is concluded that the PSBL population feature model the Bayesian NHST used in the baseline and inspection phases

respectively are the option in the development of an unsupervised damage detection method for nominally identical structures.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

This thesis has focused on the issue of damage detection in a population of nominally identical structures. It arises in many practical SHM applications to civil, mechanical and aerospace structures such as wind turbines, railway vehicles and aircraft fleets. This issue can be very challenging due to multiple sources of uncertainty present in SHM data including measurement noise in the monitoring of each individual structure (intra-structure uncertainty) and structural variability, for example, in the materials and/or manufacturing and assembling processes (inter-structure uncertainty). The case study conducted throughout the thesis is the monitoring campaign of railway wheels that are a typical example of nominally identical structures, made of the same steel material, manufactured with the same specification and assembled in the same train. The damage in railway wheels considered is geometrically manifested by out-of-roundness of wheel circumference such as wheel flat, shell and polygonization. A distributed optical fiber sensing track-side monitoring system has been deployed to online monitor, analyze and visualize SHM data for wheel condition assessment, with which multiple railway wheels or even multiple trains are made possible to be monitored in real time.

The major contribution of the thesis is the development of a general damage detection framework specifically targeted at a group of nominally identical structures rather than only a particular structure. Within this framework, a number of nominally identical structures or components are made possible to be monitored online using the same device and assessed simultaneously using the same diagnostic procedure. The developed damage detection methodology is formulated in an unsupervised learning scheme, which only makes use of response measurements from undamaged structures. It is composed of two successive phases: the baseline phase and the inspection phase. In the baseline phase, response measurements from multiple undamaged structures are employed to establish a data-driven statistical baseline model for modelling the population features of all nominally identical healthy structures. The statistical model is a probabilistic one such that multiple sources of uncertainties in SHM data can be modelled and quantified. Three modelling frameworks have been in turn proposed in this thesis with differences in the ways of dealing with multiple sources of uncertainties in the baseline model. All three modelling frameworks are based on sparse Bayesian learning (SBL), one of the state-of-the-art probabilistic machine learning methodologies following a probabilistic inference procedure. SBL is chosen here due to the advantages as follows: (1) It uses a probabilistic inference technique and thereby multiple sources of uncertainties in SHM data can be easily incorporated in the establishment of the data-driven reference model; (2) These

basis functions that are not essential in modelling the undamaged population features will be pruned automatically such that the resulting reference model will have a good prediction performance not only on the training SHM data, but also the unseen ones; (3) Relying on the sparse baseline model, the subsequent damage diagnostics in the inspection phase on the concerned structures can be achieved at the earliest possible time (the concerned structures can be the sample structures used in the baseline phase with their initial state being known undamaged but current state unknown; they can be also the structures unused in the baseline phase, but are of the same category as the sample structures).

In the inspection phase, new measurements are compared to the estimated population features from the data-driven reference model built in the baseline phase. Large discrepancies or residuals indicating possible damage in the monitored structures are signaled based on Bayesian null hypothesis significance testing, avoiding the limits in the current unsupervised damage detection approaches based on Euclidean distance, Mahalanobis squared-distance, classical null hypothesis significance testing or Bayesian point null hypothesis testing. Damage extent in the monitored group structures can be evaluated quantitatively and qualitatively in a probabilistic sense.

Chapter 3 introduced the first modelling framework simply by pooling the two types of uncertainties in SHM data such that standard sparse Bayesian learning (SSBL) can be

implemented to model the population features of all undamaged identical structures, with model parameters inferred from the Type-II maximum likelihood estimate. Chapter 4 presented the second modelling framework based on heteroscedastic sparse Bayesian learning (HSBL), an extension to SSBL in order to address the heteroscedasticity in training data, resulting from the mixed effects of two types of uncertainties. The HSBL consists of two modelling procedures, with the one for estimating the mean population features and the other for updating the associated uncertainty of it. Chapter 5 illustrated the third modelling framework in which two sources of uncertainties in SHM data are modelled separately using panel sparse Bayesian learning (PSBL), another extension to SSBL. Chapter 6 interpreted important limits in the current most popular unsupervised approaches for structural damage detection, followed by a novel diagnostic procedure based on Bayesian null hypothesis significance testing.

The major findings are as follows: (1) The first modelling framework is the simplest and the most computationally efficient, whereas its performance can be the worst in modelling the undamaged population features because different sources of uncertainties in SHM data are simply pooled together in SSBL and thus the uncertainty associated with the population features is often underestimated. Another adverse impact of the pooled uncertainties is that it gives rise to the heteroscedastic training data, which cannot be addressed in SSBL. Although the heteroscedasticity in the training data can be settled in

the aid of PSBL at the price of twice the computational effort of SSBL, the second modelling framework is still likely to underestimate the uncertainty of the population features. The third modelling framework is the most computationally intensive as many sub-models are required to recover each individual structural behavior. However, it can be the most reliable as different sources of uncertainties can be separately modeled in PSBL, which is most consistent with the information that the SHM data can deliver. More importantly, the quantification of different sources of uncertainties in the baseline model can be tracked and the respective contribution of them to the posterior uncertainty of the population features can be quantified. This essential merit can never be achieved in the first two modelling frameworks; (2) Regarding the impact on structural damage detection, the first two modelling frameworks may cause biased diagnostic results as the uncertainty in the healthy population features is underestimated and thus healthy structures are likely to be wrongly judged as damaged. Therefore, it is concluded that the third modelling framework is the best option in the establishment of a data-driven baseline model for nominally identical structures undamaged; (3) The Bayesian null hypothesis significance testing-based diagnostic logic makes it possible structural damage detection achieved in both qualitative and quantitative manners, which cannot be obtained in the current unsupervised approaches based on distance or classical significance test. In the meantime, it avoids the so-called Jeffrey-Lindely's paradox in Bayesian point null hypothesis testing

that has had popular use in the unsupervised damage detection. It should be noted, however, that the significance level in the newly developed diagnostic approach has to be carefully determined and it should be adjusted to the value that is consistent with SHM data from samples of damaged structures if available.

7.2 Recommendations for Further Research

The thesis has laid a foundation on the investigation of damage detection in a population of nominally identical structures based on unsupervised statistical machine learning approaches, to be specific, sparse Bayesian learnings. Yet, various future works can be still pursued in the areas: (1) Despite the fact that the kernel-based sparse Bayesian learnings are computationally competitive and robust in establishing a data-driven population reference model in the baseline phase, its sparsity and performance heavily depend on the selection of the appropriate kernel functions. Typically, this can be achieved by the aid of the model quality indices such as RMSR, MSLL and $\mathcal{K}$ employed in the thesis or alternatively a cross-validation procedure. A preferable methodology is to automatically select kernels during the model learning process (Tzikas et al. 2009); (2) The unknown parameters associated with the three SBL-based population reference models are all inferenced from the Type-II maximum likelihood estimation (Tipping 2001 and 2003). This might open up the possibility of overfitting, especially if there are many

hyperparameters in SBL-based population models. A more robust parameter estimation technique is to use a fully Bayesian inference framework such as MCMC sampling (Gilks et al. 1995) or variational inference methods (Bishop and Tipping 2000, Shutin et al. 2011), but it typically requires additional computational efforts; (3) In the practice of SHM, the nominally identical structures can be in service in the changing operational and environmental conditions rather than some certain condition. In the case study throughout the thesis, for example, the railway train can run at different velocities corresponding to a changing operational condition. The changing operational condition can produce the third source of uncertainty in the recorded SHM data and consequently it can cause a serious detrimental effect on damage detection performance due to the fact that the change in the structural response characteristics may be so significant as to nearly mask that caused by structural damage (Vamvoudakis-Stefanou et al. 2018, Poulimenos and Sakellariou 2019). Therefore, a considerable body of future research should be devoted to overcoming or alleviating this problem posed by the changing operational and environmental conditions in the damage detection in a population of nominally identical structures; (4) In the thesis, the structures of concern are presumed to be time-invariant and future research may include time-variant structures.

References

Abdeljaber O., Avci O., Kiranyaz S., Gabbouj M., and Inman D. J. (2017). "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks." *Journal of Sound and Vibration*, 388, 154-170.

Abdeljaber O., Avci O., Kiranyaz M. S., Boashash B., Sodano H., and Inman D. J. (2018). "1-D CNNs for structural damage detection: verification on a structural health monitoring benchmark data." *Neurocomputing*, 275, 1308-1317.

Aitkin M., Boys R. J., and Chadwick T. (2005). "Bayesian point null hypothesis testing via the posterior likelihood ratio." *Statistics and Computing*, 15(3), 217-230.

Arangio S. and Beck J. L. (2012). "Bayesian neural networks for bridge integrity assessment." *Structural Control and Health Monitoring*, 19(1), 3-21.

Adeli H. and Jiang X. (2006). "Dynamic fuzzy wavelet neural network model for structural system identification." *Journal of Structural Engineering*, 132(1), 102-111.

Bao Y., Tang Z., Li H., and Zhang Y. (2019). "Computer vision and deep learning–based data anomaly detection method for structural health monitoring." *Structural Health Monitoring*, 18(2), 401-421.

Bartlett M. S. (1957). "Comment on "A Statistical Paradox" by D. V. Lindley." *Biometrika*, 44, 533-534.

Beck J. L., Au S. K., and Vanik M. W. (2001). "Monitoring structural health using a probabilistic measure." *Computer-Aided Civil and Infrastructure Engineering*, 16(1), 1-11.

Beck J. L. (2010). "Bayesian system identification based on probability logic." *Structural Control and Health Monitoring*, 17(7), 825-847.

Bernal D. (2013). "Kalman filter damage detection in the presence of changing process and measurement noise." *Mechanical Systems and Signal Processing*, 39(1-2), 361-371.

Berger J. O., and Sellke T. (1987). "Testing a point null hypothesis: The irreconcilability of p values and evidence." *Journal of the American Statistical Association*, 82(397), 112-122.

Berger J. O., and Delampady M. (1987). "Testing precise hypotheses." *Statistical Science*, 317-335.

Berger J. O. and Pericchi L. R. (1996). "The intrinsic Bayes factor for model selection and prediction." *Journal of the American Statistical Association*, 91(433), 109-122.

Boogaard M. A., Li Z., and Dollevoet R. P. B. J. (2018). "In situ measurements of the crossing vibrations of a railway turnout." *Measurement*, 125, 313-324.

Bornn L., Farrar C. R., Park G., and Farinholt K. (2009). "Structural health monitoring with autoregressive support vector machines." *Journal of Vibration and*

Acoustics, 131(2), 021004.

Carden E. P. and Brownjohn J. M. (2008). "ARMA modelled time-series classification for structural health monitoring of civil infrastructure." *Mechanical Systems and Signal Processing*, 22(2), 295-314.

Carlin B. P. and Louis T. A. (2010). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC.

Cawley P. and Adams R. D. (1979). "The location of defects in structures from measurements of natural frequencies." *The Journal of Strain Analysis for Engineering Design*, 14(2), 49-57.

Chandrashekhar M. and Ganguli R. (2016). "Damage assessment of composite plate structures with material and measurement uncertainty." *Mechanical Systems and Signal Processing*, 75, 75-93.

Chang C. C., Chang T. Y. P., Xu Y. G., and Wang M. L. (2000). "Structural damage detection using an iterative neural network." *Journal of intelligent material systems and structures*, 11(1), 32-42.

Chen H. P. (2008). "Application of regularization methods to damage detection in large scale plane frame structures using incomplete noisy modal data." *Engineering Structures*, 30(11), 3219-3227.

Ching J. Y. and Beck J. L. (2004). "Bayesian analysis of the phase II IASC–ASCE

structural health monitoring experimental benchmark data." *Journal of Engineering Mechanics*, 130(10), 1233-1244.

Ching J. and Beck J. L. (2004). "New Bayesian model updating algorithm applied to a structural health monitoring benchmark." *Structural Health Monitoring*, 3(4), 313-332.

Chou J. H. and Ghaboussi J. (2001). "Genetic algorithm in structural damage detection." *Computers & structures*, 79(14), 1335-1353.

Cornwell P., Doebling S. W., and Farrar C. R. (1999). "Application of the strain energy damage detection method to plate-like structures." *Journal of Sound and Vibration*, 224(2), 359-374.

Cousins R. D. (2017). "The Jeffreys–Lindley paradox and discovery criteria in high energy physics." *Synthese*, 194(2), 395-432.

Cross E. J., Worden K., and Chen Q. (2011). "Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data." *Proceedings of the Royal Society A*: *Mathematical*, *Physical and Engineering Sciences*, 467(2133), 2712-2732.

Cross E. J., Manson G., Worden K., and Pierce S. G. (2012). "Features for damage detection with insensitivity to environmental and operational variations." *Proceedings of the Royal Society A*: *Mathematical*, *Physical and Engineering Sciences*, 468(2148),

4098-4122.

Cross E. J., Koo K. Y., Brownjohn J. M. W., and Worden K. (2013). "Long-term monitoring and data analysis of the Tamar Bridge." *Mechanical Systems and Signal Processing*, 35(1-2), 16-34.

Cross E. J., Worden K., and Chen Q. (2011). "Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2133), 2712-2732.

DeGroot M. H. and Schervish M. J. (2002). *Probability and Statistics*, 3rd ed. Addison-Wesley, Boston, USA.

Doebling S. W. and Farrar C. R. (1998). "Statistical damage identification techniques applied to the I-40 bridge over the Rio Grande River." *Proceeding of the 16th International Modal Analysis Conference*, Santa Barbara, California, USA.

Doebling S. W., Hemez F. M., Peterson L. D., and Farhat C. (1997). "Improved damage location accuracy using strain energy-based mode selection criteria." *AIAA Journal*, 35(4), 693-699.

Draper N. R. and Smith H. (2014). *Applied Regression Analysis*, 3rd. John Wiley & Sons.

Ebrahimian H., Astroza R., Conte J. P., and de Callafon R. A. (2017). "Nonlinear finite element model updating for damage identification of civil structures using batch

Bayesian estimation." *Mechanical Systems and Signal Processing*, 84, 194-222.

Eerland W. J., Box S., and Sóbester A. (2016). "Modeling the dispersion of aircraft trajectories using Gaussian processes." *Journal of Guidance, Control, and Dynamics*, 2661-2672.

Entezami A. and Shariatmadar H. (2014). "Damage detection in structural systems by improved sensitivity of modal strain energy and Tikhonov regularization method." *International Journal of Dynamics and Control*, 2(4), 509-520.

Farrar C. R., Baker W. E., Bell T. M., Cone K. M., Darling T. W., Duffey T. A., and Migliori A. (1994). "Dynamic characterization and damage detection in the I-40 bridge over the Rio Grande." *Technical Report LA-12767-MS*, Los Alamos, Los Alamos National Laboratory, NM, USA.

Farrar C. R. and Cone K. M. (1994). "Vibration testing of the I-40 bridge before and after the introduction of damage" *Technical Report LA-UR-94-3470*, Los Alamos, Los Alamos National Laboratory, NM, USA.

Farrar C. and Jauregui D. (1996). "Damage detection algorithms applied to experimental modal data from the I-40 bridge." *Technical Report LA-13074-MS*, Los Alamos, Los Alamos National Laboratory, NM, USA.

Farrar C. R. and Worden K. (2006). "An introduction to structural health monitoring." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and*

Engineering Sciences, 365(1851), 303-315.

Farrar C. R. and Worden K. (2012). *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, West Sussex, UK.

Figueiredo E., Radu L., Worden K., and Farrar C. R. (2014). "A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability." *Engineering Structures*, 80, 1-10.

Filograno M. L., Guillén P. C., Rodríguez-Barrios A., Martín-López S., Rodríguez-Plaza M., Andrés-Alguacil Á., and González-Herráez M. (2012). "Real-time monitoring of railway traffic using fiber Bragg grating sensors." *IEEE Sensors Journal*, 12(1): 85-92.

Filograno M. L., Corredera P., Rodriguez-Plaza M., Andres-Alguacil A., and Gonzalez-Herraez M. (2013). "Wheel flat detection in high-speed railway systems using fiber Bragg gratings." *IEEE Sensors Journal*, 13(12), 4808-4816.

Friswell M. I. (2006). "Damage identification using inverse methods." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 393-410.

Friswell M. and Mottershead J. E. (1995). *Finite Element Model Updating in Structural Dynamics*. Dordrecht: Kluwer Academic Publishers, Boston, USA.

Fritzen C. P., Jennewein D., and Kiefer T. (1998). "Damage detection based on model

updating methods." *Mechanical Systems and Signal Processing*, 12(1), 163-186.

Fugate M. L., Sohn H., and Farrar C. R. (2001). "Vibration-based damage detection using statistical process control." *Mechanical Systems and Signal Processing*, 15(4), 707-721.

Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., and Rubin D. B. (2013). *Bayesian Data Analysis*, 3rd. Chapman and Hall/CRC.

Goldberg P. W., Williams C. K., and Bishop C. M. (1998). "Regression with input-dependent noise: A Gaussian process treatment," *Advances in Neural Information Processing Systems*, 10, 493-499.

Guan H., Karbhari V. M., and Sikorsky C. S. (2006). "Web-based structural health monitoring of an FRP composite bridge." *Computer-Aided Civil and Infrastructure Engineering*, 21(1), 39-56.

Hao H. and Xia Y. (2002). "Vibration-based damage detection of structures by genetic algorithm." *Journal of Computing in Civil Engineering*, 16(3), 222-229.

Härdle W. and Simar L. (2015). *Applied Multivariate Statistical Analysis*, 4th ed. Springer, Berlin, Germany.

Hernandez E. M. (2014). "Identification of isolated structural damage from incomplete spectrum information using ℓ1-norm minimization." *Mechanical Systems and Signal Processing*, 46(1), 59-69.

Hios J. D. and Fassois S. D. (2009). "Statistical damage detection in a smart structure under different temperatures via vibration testing: a global model-based approach." *Smart Materials and Structures*, 413–414, 261-268.

Hou R., Xia Y., Bao Y., and Zhou X. (2018). "Selection of regularization parameter for ℓ1-regularized damage detection." *Journal of Sound and Vibration*, 423, 141-160.

Hou R., Xia Y., and Zhou X. (2018). "Structural damage detection based on ℓ1 regularization using natural frequencies and mode shapes." *Structural Control and Health Monitoring*, 25(3), e2107.

Hou R., Xia Y., Zhou X., and Huang Y. (2019). "Sparse Bayesian learning for structural damage detection using expectation–maximization technique." *Structural Control and Health Monitoring*, 26(5), e2343.

Hou Z., Noori M., and Amand R. S. (2000). "Wavelet-based approach for structural damage detection." *Journal of Engineering Mechanics*, 126(7), 677-683.

Hu H., Wang B. T., Lee C. H., and Su J. S. (2006). "Damage detection of surface cracks in composite laminates using modal analysis and strain energy method." *Composite Structures*, 74(4), 399-405.

Hu H. and Wu C. (2009). "Development of scanning damage index for the damage detection of plate structures using modal strain energy method." *Mechanical Systems and Signal Processing*, 23(2), 274-287.

Hua X. G., Ni Y. Q., Ko J. M., and Wong K. Y. (2007). "Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique." *Journal of Computing in Civil Engineering*, 21(2), 122-135.

Huang Y., Beck J. L., and Li H. (2017). "Hierarchical sparse Bayesian learning for structural damage detection: Theory, computation and application." *Structural Safety*, 64, 37-53.

Huang Y., Beck J. L., and Li H. (2017). "Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment." *Computer Methods in Applied Mechanics and Engineering*, 318, 382-411.

Huang Y., Beck J. L., and Li H. (2017). "Multi-task sparse Bayesian learning for model updating in structural health monitoring." *Proceedings of the 11th International Workshop on Structural Health Monitoring*, Stanford, California, USA.

Huang Y., Beck J. L., and Li, H. (2018). "Multitask sparse Bayesian Learning with Applications in Structural Health Monitoring." *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 712-730.

Idichandy V. G. and Ganapathy C. (1990). "Modal parameters for structural integrity monitoring of fixed offshore platforms." *Experimental Mechanics*, 30(4), 382-391.

Jamshidi A., Faghih-Roohi S., Hajizadeh S., Núñez A., Babuska R., Dollevoet R., Li Z., and De Schutter B. (2017), "A big data analysis approach for rail failure risk assessment." *Risk analysis*, 37(8), 1495-1507.

Jamshidi A., Hajizadeh S., Su Z., Naeimi M., Núnez A., Dollevoet R., .De Schutter B., and Li, Z. (2018). "A decision support approach for condition-based maintenance of rails based on big data analysis." *Transportation Research Part C: Emerging Technologies*, 95, 185-206.

Janapati V., Kopsaftopoulos F., Li F., Lee S. J., and Chang F. K. (2016). "Damage detection sensitivity characterization of acousto-ultrasound-based structural health monitoring techniques." *Structural Health Monitoring*, 15(2), 143-161.

Jeffreys H. (1961). *The Theory of Probability*. Oxford: Oxford University Press.

Jiang X. and Adeli H. (2007). "Pseudospectra, MUSIC, and dynamic wavelet neural network for damage detection of highrise buildings." *International Journal for Numerical Methods in Engineering*, 71(5), 606-629.

Jiang X., Mahadevan S., and Adeli H. (2007). "Bayesian wavelet packet denoising for structural system identification." *Structural Control and Health Monitoring*, 14(2), 333-356.

Jiang X. and Mahadevan S. (2008). "Bayesian wavelet methodology for structural damage detection." *Structural Control and Health Monitoring*, 15(7), 974-991.

Jiang X. and Mahadevan S. (2008). "Bayesian probabilistic inference for nonparametric damage detection of structures." *Journal of Engineering Mechanics*, 134(10), 820-831.

Jiang X., Mahadevan S., and Guratzsch R. (2009). "Bayesian wavelet methodology for damage detection of thermal protection system panels." *AIAA Journal*, 47(4), 942-952.

Johansson A. (2006). "Out-of-round railway wheels-assessment of wheel tread irregularities in train traffic." *Journal of Sound and Vibration*, 293(3-5), 795-806.

Johansson A. and Nielsen J. C. O. (2003), "Out-of-round railway wheels-wheel-rail contact forces and track response derived from field tests and numerical simulations." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 217(2): 135-146.

Kass R. E. and Raftery A. E. (1995). "Bayes factors." *Journal of the American Statistical Association*, 90(430), 773-795.

Kersting K., Plagemann C., Pfaff P., and Burgard W. (2007). "Most likely heteroscedastic Gaussian process regression." *Proceedings of the 24th International Conference on Machine learning*, Oregon, USA.

Kessy A., Lewin A., and Strimmer K. (2018). "Optimal whitening and decorrelation." *The American Statistician*, 72(4), 309-314.

Kim H. and Melhem H. (2004). "Damage detection of structures by wavelet analysis." *Engineering Structures*, 26(3), 347-362.

Koo K. Y., Sung S. H., Park J. W., and Jung H. J. (2010). "Damage detection of shear buildings using deflections obtained by modal flexibility." *Smart Materials and Structures*, 19(11), 115026.

Lam H. F., Hu Q., and Wong M. T. (2014). "The Bayesian methodology for the detection of railway ballast damage under a concrete sleeper." *Engineering Structures*, 81, 289-301.

Lam H. F., Yuen K. V., and Beck J. L. (2006). "Structural health monitoring via measured Ritz vectors utilizing artificial neural networks." *Computer-Aided Civil and Infrastructure Engineering*, 21(4), 232-241.

Lázaro-Gredilla M. and Titsias M. K. (2011). "Variational heteroscedastic Gaussian process regression." *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA.

Le Q. V., Smola A. J., and Canu S. (2005). "Heteroscedastic Gaussian process regression." *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany.

Lehmann E. L. and Romano J. P. (2006). *Testing Statistical Hypotheses*, 3rd ed. Springer Science & Business Media, New York, USA.

Leone F. C., Nelson L. S., and Nottingham R. B. (1961). "The folded normal distribution." *Technometrics*, 3(4), 543-550.

Li X. Y. and Law S. S. (2010). "Adaptive Tikhonov regularization for damage detection based on nonlinear model updating." *Mechanical Systems and Signal Processing*, 24(6), 1646-1664.

Lindley D. V. (1957). "A statistical paradox." *Biometrika*, 44(1/2), 187-192.

Lipowsky H., Staudacher S., Bauer M., and Schmidt, K. J. (2010). "Application of Bayesian forecasting to change detection and prognosis of gas turbine performance." *Journal of Engineering for Gas Turbines and Power*, 132(3), 031602.

Liu X. Z. and Ni Y. Q. (2018). "Wheel tread defect detection for high-speed trains using FBG-based online monitoring techniques." *Smart Structures and Systems*, 21(5), 687-694.

Mao Z. and Todd M. (2013). "Statistical modeling of frequency response function estimation for uncertainty quantification." *Mechanical Systems and Signal Processing*, 38(2), 333-345.

McHutchon A. J. (2015). *Nonlinear Modelling and Control using Gaussian Processes*. Doctoral dissertation, University of Cambridge, UK.

McHutchon A. and Rasmussen C. E. (2011). "Gaussian process training with input noise." *Advances in Neural Information Processing Systems*, 1341-1349.

Mei L., Mita A., and Zhou J. (2016). "An improved substructural damage detection approach of shear structure based on ARMAX model residual." *Structural Control and Health Monitoring*, 23(2), 218-236.

Mohsenzadeh Y. and Sheikhzadeh H. (2014). "Gaussian kernel width optimization for sparse Bayesian learning." *IEEE Transactions on Neural Networks and Learning Systems*, 26(4), 709-719.

Morey R. D., Romeijn J. W., and Rouder J. N. (2016). "The philosophy of Bayes factors and the quantification of statistical evidence." *Journal of Mathematical Psychology*, 72, 6-18.

Mosavi A. A., Dickey D., Seracino R., and Rizkalla S. (2012). "Identifying damage locations under ambient vibrations utilizing vector autoregressive models and Mahalanobis distances." *Mechanical Systems and Signal Processing*, 26, 254-267.

Muñoz-González L., Lázaro-Gredilla M., and Figueiras-Vidal A. R. (2014). "Divisive Gaussian processes for nonstationary regression." *IEEE Transactions on Neural Network and Learning Systems*, 25(11), 1991-2003.

Muñoz-González L., Lázaro-Gredilla M., and Figueiras-Vidal A. R. (2016). "Laplace approximation for divisive Gaussian processes for nonstationary regression." *IEEE Transactions Pattern Analysis on Machine Intelligence*, 38(3), 618-624.

Nair K. K., Kiremidjian A. S., and Law K. H. (2006). "Time series-based damage

detection and localization algorithm with application to the ASCE benchmark structure." *Journal of Sound and Vibration*, 291(1-2), 349-368.

Nair K. K. and Kiremidjian A. S. (2007). "Time series based structural damage detection algorithm using Gaussian mixtures modeling." *Journal of Dynamic Systems, Measurement, and Control*, 129(3), 285-293.

Nguyen L. H., Gaudot I., and Goulet J. A. (2019). "Uncertainty quantification for model parameters and hidden state variables in Bayesian dynamic linear models." *Structural Control and Health Monitoring*, 26(3), e2309.

Ni Y. Q. and Chen S. X. (2018). "Compressive sensing for vibration signals in highspeed rail monitoring." *Proceedings of the 9th European Workshop on Structural Health Monitoring*, Manchester, UK.

Ni Y. Q., Hua X. G., Fan K. Q., and Ko J. M. (2005). "Correlating modal properties with temperature using long-term monitoring data and support vector machine technique." *Engineering Structures*, 27(12), 1762-1773.

Ni Y. Q., Zhou H. F., Chan K. C., and Ko J. M. (2008). "Modal flexibility analysis of cable-stayed Ting Kau Bridge for damage identification." *Computer-Aided Civil and Infrastructure Engineering*, 23(3), 223-236.

Ni Y. Q., Zhou H. F., and Ko J. M. (2009). "Generalization capability of neural network models for temperature-frequency correlation using monitoring data." *Journal of*

Structural Engineering, 135(10), 1290-1300.

Nielsen J. C. and Johansson A. (2000). "Out-of-round railway wheels-a literature survey." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 214(2), 79-91.

Oh C. K. and Sohn H. (2009). "Damage diagnosis under environmental and operational variations using unsupervised support vector machine." *Journal of Sound and Vibration*, 325(1-2), 224-239.

Oregui M., Molodova M., Núñez A., Dollevoet R. P. B. J., and Li Z. (2015). "Experimental investigation into the condition of insulated rail joints by impact excitation." *Experimental Mechanics*, 55(9), 1597-1612.

Ovanesova A. V. and Suarez L. E. (2004). "Applications of wavelet transforms to damage detection in frame structures." *Engineering Structures*, 26(1), 39-49.

Pandey A. K., Biswas M., and Samman M. M. (1991). "Damage detection from changes in curvature mode shapes." *Journal of Sound and Vibration*, 145(2), 321-332.

Pandey A. K. and Biswas M. (1994). "Damage detection in structures using changes in flexibility." *Journal of Sound and Vibration*, 169(1), 3-17.

Pandey A. K. and Biswas M. (1995). "Experimental verification of flexibility difference method for locating damage in structures." *Journal of Sound and Vibration*, 184(2), 311-328.

Papadopoulos L. and Garcia E. (1998). "Structural damage identification: a probabilistic approach." *AIAA Journal*, 36(11), 2137-2145.

Papatheou E., Rahman T. A. Z., Barthorpe R. J., Park J., and Worden K. (2014). "An experimental investigation of feature complexity and diversity in nominally similar test structures." *Proceeding of the ISMA International Conference on Noise and Vibration Engineering*, Leuven, Belgium.

Papatheou E., Barthorpe R. J., and Worden K. (2015). "An experimental investigation of feature availability in nominally identical structures for population-based SHM." *Proceedings of the Society for Experimental Mechanics Series*, 7, 185-191.

Peeters B. and De Roeck G. (2001). "One-year monitoring of the Z24-Bridge: environmental effects versus damage events." *Earthquake Engineering & Structural Dynamics*, 30(2), 149-171.

Peeters B., Maeck J., and De Roeck G. (2001). "Vibration-based damage detection in civil engineering: excitation sources and temperature effects." *Smart Materials and Structures*, 10(3), 518.

Peng Z. K., Lang Z. Q., Wolters C., Billings S. A., and Worden K. (2011). "Feasibility study of structural damage detection using NARMAX modelling and Nonlinear Output Frequency Response Function based analysis." *Mechanical Systems and Signal Processing*, 25(3), 1045-1061.

Perera R. and Torres R. (2006). "Structural damage detection via modal data with genetic algorithms." *Journal of Structural Engineering*, 132(9), 1491-1501.

Poulimenos A. G. and Sakellariou J. S. (2019). "A transmittance-based methodology for damage detection under uncertainty: An application to a set of composite beams with manufacturing variability subject to impact damage and varying operating conditions." *Structural Health Monitoring*, 18(1), 318-333.

Ren W. X. and De Roeck G. (2002). "Structural damage identification using modal data. I: Simulation verification." *Journal of Structural Engineering*, 128(1), 87-95.

Ren W. X. and De Roeck G. (2002). "Structural damage identification using modal data. II: Test verification." *Journal of Structural Engineering*, 128(1), 96-104.

Robert C. P. (1993). "A note on Jeffreys-Lindley paradox." *Statistica Sinica*, 601-608.

Rouder J. N., Speckman P. L., Sun D., Morey R. D., and Iverson G. (2009). "Bayesian t tests for accepting and rejecting the null hypothesis." *Psychonomic Bulletin & Review*, 16(2), 225-237.

Rucevskis S., Sumbatyan M. A., Akishin P., and Chate A. (2015). "Tikhonov's regularization approach in mode shape curvature analysis applied to damage detection." *Mechanics Research Communications*, 65, 9-16.

Rucka, M. and Wilde K. (2006). "Application of continuous wavelet transform in vibration-based damage detection method for beams and plates." *Journal of Sound*

and Vibration, 297(3-5), 536-550.

Salawu O. S. (1997). "Detection of structural damage through changes in frequency: a review." *Engineering Structures*, 19(9), 718-723.

Sankararaman S. and Mahadevan S. (2011). "Uncertainty quantification in structural damage diagnosis." *Structural Control and Health Monitoring*, 18(8), 807-824.

Sankararaman S. and Mahadevan S. (2013). "Bayesian methodology for diagnosis uncertainty quantification and health monitoring." *Structural Control and Health Monitoring*, 20(1), 88-106.

Schubert Kabban C. M., Greenwell B. M., DeSimio M. P., and Derriso M. M. (2015). "The probability of detection for structural health monitoring systems: Repeated measures data." *Structural Health Monitoring*, 14(3), 252-264.

Sellke T., Bayarri M. J., and Berger J. O. (2001). "Calibration of ρ values for testing precise null hypotheses." *The American Statistician*, 55(1), 62-71.

Seyedpoor S. M. (2012). "A two-stage method for structural damage detection using a modal strain energy-based index and particle swarm optimization." *International Journal of Non-Linear Mechanics*, 47(1), 1-8.

Shi Z., Law S. S., and Zhang L. M. (1998). "Structural damage localization from modal strain energy change." *Journal of Sound and Vibration*, 218(5), 825-844.

Shi Z., Law S. S., and Zhang L. (2000). "Structural damage detection from modal strain

energy change." *Journal of Engineering Mechanics*, 126(12), 1216-1223.

Sohn H. (2006). "Effects of environmental and operational variability on structural health monitoring." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 539-560.

Sohn H. and Farrar C. R. (2001). "Damage diagnosis using time series analysis of vibration signals." *Smart Materials and Structures*, 10(3), 446.

Sohn H., Farrar C. R., Hemez F. M., Shunk D. D., Stinemates D. W., Nadler B. R., and Czarnecki J. J. (2003). "A review of structural health monitoring literature: 1996-2001." *Technical Report LA-13976-MS*, Los Alamos National Laboratory, NM, USA.

Sohn H. and Law K. H. (1997). "A Bayesian probabilistic approach for structure damage detection." *Earthquake Engineering & Structural Dynamics*, 26(12), 1259-1281.

Sohn H. and Law K. H. (2000). "Bayesian probabilistic damage detection of a reinforced-concrete bridge column." *Earthquake Engineering & Structural Dynamics*, 29(8), 1131-1152.

Sohn H. and Law K. H. (2000). "Application of load-dependent Ritz vectors to Bayesian probabilistic damage detection." *Probabilistic Engineering Mechanics*, 15(2), 139-153.

Sohn H. and Law K. H. (2001). "Damage diagnosis using experimental Ritz vectors." *Journal of Engineering Mechanics*, 127(11), 1184-1193.

Stubbs N., Kim J. T., and Topole K. (1992). "An efficient and robust algorithm for damage localization in offshore platforms" *Proceedings of the ASCE 10th Structures Congress*, 1, 543-546.

Subramanian A. and Mahadevan S. (2019). "Bayesian estimation of discrepancy in dynamics model prediction." *Mechanical Systems and Signal Processing*, 123, 351-368.

Sung S. H., Koo K. Y., and Jung H. J. (2014). "Modal flexibility-based damage detection of cantilever beam-type structures using baseline modification." *Journal of Sound and Vibration*, 333(18), 4123-4138.

Tabor L., Goulet J. A., Charron J. P., and Desmettre C. (2018). "Probabilistic Modeling of Heteroscedastic Laboratory Experiments Using Gaussian Process Regression." *Journal of Engineering Mechanics*, 144(6), 04018038.

Tang Z., Chen Z., Bao Y., and Li H. (2019). "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring." *Structural Control and Health Monitoring*, 26(1), e2296.

Teughels A. and De Roeck G. (2004). "Structural damage identification of the highway bridge Z24 by FE model updating." *Journal of Sound and Vibration*, 278(3), 589-610.

TG/CL 127-2017. *Operation and Maintenance Regulations for Railway Electric Multiple Units*. China Railway Corporation.

Tikhonov A. N. (1995). *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer Academic Publishers.

Tipping M. E. (2001). "Sparse Bayesian learning and the relevance vector machine." *Journal of Machine Learning Research*, 1, 211-244.

Tipping M. E. and Faul A. C. (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." *Proceedings of the 9th International Workshop Artificial Intelligence and Statistics*, Scotland, UK.

Titurus B., Friswell M. I., and Starek L. (2003). "Damage detection using generic elements: Part I. Model updating." *Computers & structures*, 81(24-25), 2273-2286.

Urban S., Ludersdorfer M., and Van Der Smagt P. (2015). "Sensor calibration and hysteresis compensation with heteroscedastic gaussian processes." *IEEE Sensors Journal*, 15(11), 6498-6506.

Uzzal R. U. A., Ahmed W., and Rakheja S. (2008), "Dynamic analysis of railway vehicle-track interactions due to wheel flat with a pitch-plane vehicle model." *Journal of Mechanical Engineering*, 39(2), 86-94.

Wagenmakers E. J., Marsman M., Jamil T., Ly, A., Verhagen J., Love J., and Matzke D. (2018). "Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications." *Psychonomic Bulletin & Review*, 25(1), 35-57.

West M. and Harrison J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed.

Springer Science & Business Media, New York, USA.

Wipf D. P. and Rao B. D. (2004). "Sparse Bayesian learning for basis selection." *IEEE Transactions on Signal Processing*, 52(8), 2153-2164.

Teimouri H., Milani A. S., Loeppky J., and Seethaler R. (2017). "A Gaussian process–based approach to cope with uncertainty in structural health monitoring." *Structural Health Monitoring*, 16(2), 174-184.

Teimouri H., Milani A. S., Seethaler R., and Heidarzadeh A. (2015). "On the impact of manufacturing uncertainty in structural health monitoring of composite structures: a signal to noise weighted neural network process." *Journal of Composite Materials*, 6(1), 28-39.

Titterington D. M., Smith A. F., and Makov U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley, New York.

Toksoy T. and Aktan A. E. (1994). "Bridge-condition assessment by modal flexibility." *Experimental Mechanics*, 34(3), 271-278.

Tsou P. and Shen M. H. (1994). "Structural damage detection and identification using neural networks." *AIAA Journal*, 32(1), 176-183.

Vamvoudakis-Stefanou K. J. and Fassois S. D. (2014). "Vibration-based damage detection for a population of like structures via a multiple model framework." *Proceedings of the 6th International Symposium on NDT in Aerospace*, Madrid, Spain.

Vamvoudakis-Stefanou K. J., Sakellariou J. S., and Fassois S. D. (2014). "Random vibration response-only damage detection for a set of composite beams." *Proceedings of the ISMA International Conference on Noise and Vibration Engineering*, Leuven, Belgium.

Vamvoudakis-Stefanou K., Sakellariou J., and Fassois S. (2016). "Random Coefficient model based damage detection for a population of nominally identical structures: An exploratory study." *Proceedings of the International Conference on Noise and Vibration Engineering*, Leuven, Belgium.

Vamvoudakis-Stefanou K. J. and Fassois S. D. (2017). "Vibration-based damage detection for a population of nominally identical structures via Random Coefficient Gaussian Mixture AR model-based methodology." *Procedia Engineering*, 199, 1888-1893.

Vamvoudakis-Stefanou K. J., Sakellariou J. S., and Fassois S. D. (2018). "Vibration-based damage detection for a population of nominally identical structures: Unsupervised Multiple Model (MM) statistical time series type methods." *Mechanical Systems and Signal Processing*, 111, 149-171.

Vanik M. W., Beck J. L., and Au S. (2000). "Bayesian probabilistic approach to structural health monitoring." *Journal of Engineering Mechanics*, 126(7), 738-745.

Villani L. G., Da Silva S., and Cunha Jr A. (2019). "Damage detection in uncertain

nonlinear systems based on stochastic Volterra series." *Mechanical Systems and Signal Processing*, 125, 288-310.

Wahab M. A. and De Roeck G. (1999). "Damage detection in bridges using modal curvatures: application to a real damage scenario." *Journal of Sound and Vibration*, 226(2), 217-235.

Wang C. Y. (2014). *Gaussian Process Regression with Heteroscedastic Residuals and Fast MCMC Methods*. Doctoral dissertation, Toronto University, Toronto, Canada.

Wang J. F., Liu X. Z., and Ni Y. Q. (2018). "A Bayesian probabilistic approach for acoustic emission-based rail condition assessment." *Computer-Aided Civil and Infrastructure Engineering*, 33(1), 21-34.

Wang P., Youn B. D., and Hu C. (2012). "A generic probabilistic framework for structural health prognostics and uncertainty management." *Mechanical Systems and Signal Processing*, 28, 622-637.

Wang Y. W. (2017). *Bayesian-based Methodology for Progressive Structural Health Evaluation and Prediction by Use of Monitoring Data*. Doctoral dissertation, The Hong Kong Polytechnic University, Hong Kong.

Wasserstein R. L. and Lazar N. A. (2016) "The ASA Statement on p-Values: Context, Process, and Purpose." *The American Statistician*, 70:2, 129-133.

Weber B. and Paultre P. (2009). "Damage identification in a truss tower by regularized

model updating." *Journal of Structural Engineering*, 136(3), 307-316.

Weber B., Paultre P., and Proulx J. (2009). "Consistent regularization of nonlinear model updating for damage identification." *Mechanical Systems and Signal Processing*, 23(6), 1965-1985.

Wei Z., Yam L. H., and Cheng L. (2005). "NARMAX model representation and its application to damage detection for multi-layer composites." *Composite Structures*, 68(1), 109-117.

Wei C., Xin Q., Chung W. H., Liu S. Y., Tam H. Y. and Ho S. L. (2012). "Real-time train wheel condition monitoring by fiber Bragg grating sensors." *International Journal of Distributed Sensor Networks*, 8(1): 409048.

Worden K., Farrar C. R., Manson G., and Park G. (2007). "The fundamental axioms of structural health monitoring." *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2082), 1639-1664.

Worden K. and Lane A. J. (2001). "Damage identification using support vector machines." *Smart Materials and Structures*, 10(3), 540.

Worden K., Sohn H., and Farrar C. R. (2002). "Novelty detection in a changing environment: regression and interpolation approaches." *Journal of Sound and Vibration*, 258(4), 741-761.

Wu X., Ghaboussi J., and Garrett Jr J. H. (1992). "Use of neural networks in detection of structural damage." *Computers & Structures*, 42(4), 649-659.

Wu Y. H. and Zhou X. Q. (2018). "ℓ1 Regularized Model Updating for Structural Damage Detection." *Journal of Structural Stability and Dynamics*, 18(12), 1850157.

Xin Y., Hao H., Li J., Wang Z. C., Wan H. P., and Ren W. X. (2019). "Bayesian based nonlinear model updating using instantaneous characteristics of structural dynamic responses." *Engineering Structures*, 183, 459-474.

Yan W. J., Ren W. X., and Huang T. L. (2012). "Statistic structural damage detection based on the closed-form of element modal strain energy sensitivity." *Mechanical Systems and Signal Processing*, 28, 183-194.

Yang J. N., Lin S., Huang H., and Zhou L. (2006). "An adaptive extended Kalman filter for structural damage identification." *Structural Control and Health Monitoring*, 13(4), 849-867.

Yang J. N., Pan S., and Huang H. (2007). "An adaptive extended Kalman filter for structural damage identifications II: unknown inputs." *Structural Control and Health Monitoring*, 14(3), 497-521.

Yao R. and Pakzad S. N. (2012). "Autoregressive statistical pattern recognition algorithms for damage detection in civil structures." *Mechanical Systems and Signal Processing*, 31, 355-368.

Yeager M., Gregory B., Key C., and Todd M. (2019). On using robust Mahalanobis distance estimations for feature discrimination in a damage detection

scenario. *Structural Health Monitoring*, 18(1), 245-253.

Yin T., Lam H. F., and Chow H. M. (2010). "A Bayesian probabilistic approach for crack characterization in plate structures." *Computer-Aided Civil and Infrastructure Engineering*, 25(5), 375-386.

Yuen K. V., Au S. K., and Beck J. L. (2004). "Two-stage structural health monitoring approach for phase I benchmark studies." *Journal of Engineering Mechanics*, 130(1), 16-33.

Yuen K. V., Beck J. L., and Au S. K. (2004). "Structural damage detection and assessment by adaptive Markov chain Monte Carlo simulation." *Structural Control and Health Monitoring*, 11(4), 327-347.

Yuen K. V., Beck J. L., and Katafygiotis L. S. (2006). "Unified probabilistic approach for model updating and damage detection." *Journal of Applied Mechanics*, 73(4), 555-564.

Yuen K. V., and Lam H. F. (2006). "On the complexity of artificial neural networks for smart structures monitoring." *Engineering Structures*, 28(7), 977-984.

Yuen K. V. and Ortiz G. A. (2017). "Outlier detection and robust regression for correlated data." *Computer Methods in Applied Mechanics and Engineering*, 313, 632-646.

Zang C., and Imregun M. (2001). "Structural damage detection using artificial neural networks and measured FRF data reduced via principal component projection."

Journal of Sound and Vibration, 242(5), 813-827.

Zhang C. D. and Xu Y. L. (2016). "Comparative studies on damage identification with Tikhonov regularization and sparse regularization." *Structural Control and Health Monitoring*, 23(3), 560-579.

Zhang C., Huang J. Z., Song G. Q., and Chen L. (2017). "Structural damage identification by extended K alman filter with ℓ1-norm regularization scheme." *Structural Control and Health Monitoring*, 24(11), e1999.

Zhou X., Hou R., and Wu Y. (2018). "Structural damage detection based on iteratively reweighted ℓ1 regularization algorithm." *Advances in Structural Engineering*, 22(6) 1479-1487.

Zhou X. Q., Xia Y., and Weng S. (2015). "ℓ1 regularization approach to structural damage detection using frequency data." *Structural Health Monitoring*, 14(6), 571-582.