CALIBRATED COMPUTER MODELS USING MARKOV CHAIN MONTE CARLO AND VARIATIONAL BAYES

By

Mookyong Son

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Statistics—Doctor of Philosophy

2023

ABSTRACT

Computer models are used to solve complex problems in many scientific applications, such as nuclear physics and climate research. Although a popular approach, Markov chain Monte Carlobased Bayesian calibration of computer models has not been investigated much regarding theoretical properties until relatively recently. Hence, this work focuses on the theory of computer model calibration through a proof of posterior consistency of the estimated physical process.

In Chapter 1, we review the general framework of computer model calibration, Gaussian Processes, and Posterior Consistency.

In Chapter 2, we prove the posterior consistency of the estimated physical process in the Bayesian model calibration framework using Markov chain Monte Carlo. We used the extension of Schwartz's theorem to show the posterior contraction rate using GP priors.

In Chapter 3, we propose a fast and scalable posterior approximation algorithm for Bayesian computer model calibration via Variational Inference. Variational Inference is an optimization-based method alternative to the time-consuming Markov chain Monte Carlo approximation of posterior distributions popularized by machine learners. We provide the statistical guarantee of the proposed algorithm in the form of a posterior consistency theorem for the estimated physical process under regularity assumptions on the variational family. The main results are shown in the two widely used classes of Gaussian Process priors, the Squared Exponential covariance class and the Matern covariance class. We also provide a simulation study to demonstrate the proposed method's time efficiency and fidelity compared to the standard Markov chain Monte Carlo method.

In Chapter 4, we applied the Bayesian model calibration framework to β decay calculation and compared different calibration approaches. We adopt χ^2 results as a benchmark to show the advantage of Bayesian approaches.

Finally, we conclude the thesis with future directions for further research in Chapter 5.

Copyright by MOOKYONG SON 2023

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my advisors, Dr. Tapabrata Maiti and Dr. Shrijita Bhattacharya, for their invaluable guidance throughout my research journey. Their insightful suggestions have consistently propelled me forward and led to significant progress.

I also want to acknowledge the members of my dissertation committee, Dr. Witold Nazarewicz and Dr. Chih-Li Sung, whose feedback and comments have been instrumental in shaping the quality of my research. I am particularly grateful for the weekly meetings with physicists at the Facility for Rare Isotope Beams (FRIB), which have broadened my perspective on the practical applications of Statistics. Additionally, the BAND research collaboration provided me with wonderful networking opportunities with fellow physicists. Furthermore, I extend my gratitude to my mentors, Stefan Wild and Arindam Fadikar, during my remote internship at Argonne National Lab amidst the pandemic. Our discussions greatly honed my expertise in my field.

I am thankful for all the friends I made during my time at Michigan State University. Special thanks go to Arkaprabha Ganguli, Anirban Samaddar, and Anriana Manousidaki, with whom I shared powerful memories while studying for prelim during the summer. Passing prelim without their support would have been nearly impossible. I am also indebted to Tong Li and Vojtech Kejzlar for the profound discussions that greatly enriched my thesis.

Last but not least, my gratitude goes to my family. The unwavering support of my wife, Jin Young Kwak, and my daughter, Lia, has made the completion of my Ph.D. possible. Moreover, I am truly grateful for the unconditional encouragement I have received from my parents, parents-in-law, brother, and brother-in-law throughout this journey. Their constant belief in me has been a driving force behind my accomplishments.

Throughout my five years at Michigan State University, the courses and seminars have imparted valuable knowledge, which I have successfully applied to my research. I eagerly look forward to embracing my future as a statistician with enthusiasm and dedication.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1 Computer Model Calibration	1
1.2 Gaussian Processes	3
1.3 Posterior Consistency	5
BIBLIOGRAPHY	7
CHAPTER 2 CALIBRATED COMPUTER MODEL USING MARKOV CHAIN	0
MONTE CARLO	9
2.1 Bayesian Hierarchical Formulation of Kennedy and O'Hagan Model	9
2.2 Preliminaries	1
2.3 Posterior Consistency in Kenndy and O'Hagan Model	7
2.4 Conclusion and Discussion	8
BIBLIOGRAPHY	0
APPENDIX APROOF OF THEOREM 1 <td>2</td>	2
APPENDIX BCHECKING ASSUMPTION 2.2.1	7
APPENDIX C CHECKING ASSUMPTION 2.2.2 1	1
APPENDIX D CHECKING ASSUMPTION 2.2.2 2	9
CHAPTER 3 CALIBRATED COMPUTER MODEL USING VARIATIONAL BAYES	
(VARIATIONAL INFERENCE)	-1
3.1 Variational Bayes formulation	-1
3.2 Variational Bayes Algorithm for Discrepancy Model	.4
3.3 Posterior Contraction	6
3.4 Simulation Study: Ball Dropping Experiment	.9
3.5 Simulation Study: Logistic Function	0
3.6 Conclusion and Discussion	3
BIBLIOGRAPHY	5
APPENDIX A CHECKING ASSUMPTION 3.3.2 (I)	7
APPENDIX B CHECKING ASSUMPTION 3.3.2 (II)	8
APPENDIX CPROOF OF THEOREM 3	1
CHAPTER 4 BAYESIAN MODEL CALIBRATION FOR BETA DECAY	
CALCULATIONS	6
4.1 Data Collection	8
4.2 Extension of the calibration model to two observable types 6	9
4.3 Prior Structure	0
4.4 Calibration Results	2
4.5 Prediction Results	7
4.6 Conclusion and Discussion	9
BIBLIOGRAPHY	3
APPENDIX A LATIN HYPERCUBE DESIGN 8	5
APPENDIX B CHI-SQUARE OPTIMIZATION FRAMEWORK 8	6

CHAPTER 1

INTRODUCTION

1.1 Computer Model Calibration

Computer models are an essential part of today's scientific process. With the increasing access to high-performance computing, computer models provide much faster and economical environment for scientists to rapidly develop their domain fields such as Biology, Chemistry, Epidemiology, Engineering, and Physics to just name a few. However, computer models are canonically imperfect representation of physical systems. The landmark publication of Kennedy and O'Hagan (2001) approached this issue by introducing discrepancy function between a physical process and a computer model. This work, up to the present time, accumulated more than 3000 citations and led to countless applications. For instance, see King et al. (2019); Gatt et al. (2020) for examples in nuclear physics, Seede et al. (2020); Honarmandi and Arróyave (2020) for examples in material sciences, and Colosimoa et al. (2018); Hüllen et al. (2020); Choi et al. (2018) for engineering applications.

The theoretical properties of the original framework have not been investigated until relatively recently. There are two properties of the calibration model we can think of, the one is identifiability of calibration parameters and the other is the consistency of the estimated physical process to the true physical process. Non-identifiability issue of calibration parameters in the KOH model stems from the original paper Kennedy and O'Hagan (2001), which concerns the distribution of observed physical data induced by the computer model, which does not uniquely determine the calibration parameter. There is a couple of papers to handle this problem by modifying the original model, for example take a look at Plumlee (2017) and Xie and Xu (2020). In this paper, however, our goal is to show the consistency property of the original KOH model in Fully Bayesian setting. There are also some papers to deal with this issue such as Xie and Xu (2020) or Li and Xiong (2022). But again these papers are based on the modified version of the model such as defining the calibration parameters as L_2 minimization between the computer model and the physical process or decomposing the discrepancy term.

The KOH model is built based on some assumptions. Firstly, Kennedy and O'Hagan (2001) assumes that although the computer model is not fully available, one does have exact realizations from the computer model at a finite discrete number of points. To facilitate this, a Gaussian process is assumed on the computer model indexed by (t, θ) with nuisance parameters ϕ_f . θ is often referred to as the calibration parameter. Additionally, Kennedy and O'Hagan (2001) assumes a discrepancy term between the true physical process and the computer model. It assumes a Gaussian process on the discrepancy term with index t with nuisance parameters ϕ_δ . The aim of the Kennedy and O'Hagan (2001) model is to study the relationship between the computer model f and the physical process ζ by learning about calibration parameter θ , and eventually to do predictions $\mathbf{y}^* = (y_1^*, \dots, y_j^*)$ of the physical process ζ at new inputs (t_1^*, \dots, t_j^*) . The current theoretical developments have three main drawbacks (1) they assume the computer model is explicitly available (2) θ is an unknown parameter characterizing the computer model (3) ϕ_f and ϕ_δ are unknown parameters.

In this paper, we provide the posterior consistency of the estimated physical process under the full Bayesian treatment of the Kennedy and O'Hagan (2001) model. Following Kejzlar et al. (2020), we consider an equivalent representation of Kennedy and O'Hagan (2001) model as a hierarchical model. Consequently, we prove the posterior consistency of the estimated physical process using an original extension of Schwartz's theorem for non-parametric regression problems with GP priors. Unlike all the previous works in the literature, we stick to the assumption that computer model is not fully available and one has only the evaluations from the computer model at finite number of points. Secondly, we assume that θ , ψ_1 and ψ_2 are not just parameters with simple plug-in estimators but random variables with their own prior distributions.

1.1.1 Kennedy O'Hagan Model

Let us consider observations $y = (y_1, ..., y_n)$ of a physical process ζ depending on a known set of inputs $t_i \in \Omega \subset \mathbb{R}^p$, $i = 1, ..., n, p \ge 1$ given by

$$y_i = \zeta(t_i) + \sigma \epsilon_i, \tag{1.1}$$

where σ represents the scale of observational error, and $\epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

Let f denote a computer model. Kennedy and O'Hagan (2001) introduced a discrepancy term δ to represent the unknown systematic error between the computer model f and the physical process ζ . Hence the physical process can be represented as a summation of imperfect computer model plus the discrepancy term.

$$\zeta(\mathbf{t}_i) = f(\mathbf{t}_i, \boldsymbol{\theta}) + \delta(\mathbf{t}_i), \quad i = 1, \dots, n$$
(1.2)

Note, the above expression involves known inputs t_i and unknown parameters θ , also referred to as the *calibration parameter*. The calibration parameter is not controllable in the physical process ζ and thus may be used to represent aspects of the computer model which can not be measured in the physical observations. In some cases, the calibration parameter may also carry a real physical meaning. For example, consider a famous ball drop example by Plumlee (2017). If one constructs a computer model to check the time for a ball to hit the ground from a certain height. Then the gravitational force could be used as a calibration parameter to build a computer model. Further discussion of physical meaning in calibration parameters could be checked at Brynjarsdóttir and O'Hagan (2014).

Next, we assume that the evaluation of computer model f is either expensive or not directly available, which are often the case in a real-world application. The common practice in this situation is to fit a surrogate for a computer model based on the simulation data. Hence, let $z = (z_1, ..., z_s)$ denote outputs from the computer model f under a known set of inputs $\tilde{t}_i \in \Omega$, $p \ge 1$ and $\tilde{\theta}_i \in \Theta$.

$$z_i = f(\tilde{\boldsymbol{t}}_i, \tilde{\boldsymbol{\theta}}_i) \quad i = 1, \dots, s,$$
(1.3)

Without loss of generality, we assume $\Omega = [0, 1]^p$ for the rest of the paper.

1.2 Gaussian Processes

Gaussian Process is a collection of random variables, any finite number of which have a joint multivariate Gaussian distribution. Hence we can think of it as a infinite-dimensional multivariate Gaussian distribution. A Gaussian process f(x) is completely specified by it's mean function m(x)



Figure 1.1 (Gaussian Process priors) The left figure shows realization of three random functions from Gaussian Processes prior, and the right figure shows realization of three random functions from posterior distribution updated by five training data, which is shown in "+" signs. This figure is taken from Rasmussen and Williams (2006).

and covariance function k(x, x'), where

$$m(x) = \mathbb{E}[f(x)]$$
$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$$

A commonly adopted mean function is a constant, while there are many types of covariance functions widely used. Two widely used covariance functions are 'Squared Exponential' (SE) and 'Matérn'. Squared Exponential covariance functions has the following form.

$$k(x, x'; \eta, l) = \eta \exp\left(\frac{-1}{2l^2}(x - x')^2\right)$$

where $\eta > 0$ is the scale parameter and l > 0 is the length scale parameter. Matérn covariance functions has the following form.

$$k(x, x'; \eta, l, \nu) = \eta \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x - x'|}{l} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}|x - x'|}{l} \right)$$

where K_{ν} is the modified Bessel function and $\nu > 0$ is the roughness parameter.



Figure 1.2 (Covariance functions of GP) The left figure shows Matérn covariance as a function of input distance r = |x - x'| while η and l are fixed at 1. Different colors represent different roughness parameter value v. The red line corresponds to $v = \infty$, which is equivalent to SE covariance. The right figure shows realization of random functions from GP of each covariance function.

The connection between these two classes of covariance function is that the SE covariance is a special case of Metérn covariance where v equals to ∞ . When the roughness parameter becomes infinity, the sample function from GP with SE covariance function becomes infinitely differentiable, in other words, very smooth. This could be checked at the Figure 1.2. The sample function from the SE covariance function is colored red, and it looks very smooth compared to other wiggly sample functions.

This smoothness assumption is often too restrictive to model the response surface of physical processes. Hence, in practice, it is often advised to use the Matérn covariance, Stein (1999). However, SE covariance is often easier to derive the sample function properties of GP. Therefore, we will consider both cases in developing large sample theory in Bayesian model calibration in this thesis.

1.3 Posterior Consistency

The consistency of estimators is usually discussed in the context of frequentist statistics. In a frequentist framework, a consistent estimator converges in probability to the true parameter as the number of data points approaches infinity. This idea can be extended to the Bayesian framework as well. For Bayesian methods, convergence properties are discussed using the entire posterior distribution rather than a single estimate.

Assume that a set of data $y^n = (y_1, \dots, y_n)$ are randomly sampled from $f_0 \in \mathcal{F}$. Then the posterior distribution is a random measure defined as follows,

$$\Pi(B|\mathbf{y}^n) = \frac{\int_B \prod_{i=1}^n f(y_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n f(y_i) d\Pi(f)}$$

where *B* is the measurable set in \mathcal{F} .

The natural question we have is what happens to the posterior distribution as the sample size n increases to infinity. If the posterior distribution of f concentrates around some small neighborhood of true parameter f_0 , then we say that the *posterior distribution is consistent*. In other words, as we see more data, data provide more information on the parameter and ideally the posterior distribution of the parameter converges to the truth, which is a point mass.

There has been a line of research to show the posterior consistency, and the first approach was done by Doob (1949). In this work, Doob showed that posterior distribution is consistent almost everywhere on prior support, but this was not useful to check consistency at a particular density. Later in 1965, Schwartz introduced a test function framework to show the posterior consistency Schwartz (1965). In particular, he showed that in order to show that the posterior is consistent, it is necessary that the true parameter and the complement set of the neighborhood of the true parameter can be separated. This idea of separation is conveniently formalized through the existence of appropriate tests for testing $H_0: f = f_0$, $H_A: f \in U^c$, where U is some neighborhood of f_0 .

When posterior consistency holds, quantifying the rate at which the posterior contracts to the truth is often possible. This is called the *posterior contraction rate*. Hence, in the following chapters, we show not only the posterior is consistent but also the posterior contraction rates in the KOH model setup using MCMC and Variational Bayes.

BIBLIOGRAPHY

- Brynjarsdóttir, J. and O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30:114007.
- Choi, W., Menberg, K., Kikumoto, H., Heo, Y., Choudhary, R., and Ooka, R. (2018). Bayesian inference of structural error in inverse models of thermal response tests. *Applied Energy*, 228:1473–1485.
- Colosimoa, B. M., Huangb, Q., Dasguptac, T., and Tsung, F. (2018). Opportunities and challenges of quality engineering for additive manufacturing. *Journal of Quality Technology*, 50:233–252.
- Doob, J. L. (1949). Application of the theory of martingales. *in Le Calcul des Probabilites et ses Applications*, 40:23–27.
- Gatt, D., Yousif, C., Cellura, M., Camilleri, L., and Guarino, F. (2020). Assessment of building energy modelling studies to meet the requirements of the new energy performance of buildings directive. *Renewable and Sustainable Energy Reviews*, 127:1–16.
- Honarmandi, P. and Arróyave, R. (2020). Uncertainty quantification and propagation in computational materials science and simulation assisted materials design. *Integrating Materials and Manufacturing Innovation*, 9:103–143.
- Hüllen, G., and Sun-Hye Kim, J. Z., Sinha, A., J.Realff, M., and Boukouvala, F. (2020). Managing uncertainty in data-driven simulation-based optimization. *Computers and Chemical Engineering*, 136:1–16.
- Kejzlar, V., Son, M., Bhattacharya, S., and Maiti, T. (2020). A fast, scalable, and calibrated computer model emulator: An empirical bayes approach.
- Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:425–464.
- King, G. B., Lovell, A. E., Neufcourt, L., and Nunes, F. M. (2019). Direct comparison between Bayesian and frequentist uncertainty quantification for nuclear reactions. *Physical Review Letters*, 122:232502.
- Li, Y. and Xiong, S. (2022). Physical parameter calibration. arXiv:2208.00124 [stat.ME].
- Plumlee, M. (2017). Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112:1274–1285.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

- Schwartz, L. (1965). On bayes procedures. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 4:10–26.
- Seede, R., Shoukr, D., Zhang, B., Whitt, A., Gibbons, S., Flater, P., Elwany, A., Arroyave, R., and Karaman, I. (2020). An ultra-high strength martensitic steel fabricated using selective laser melting additive manufacturing: Densification, microstructure, and mechanical properties. *Acta Materialia*, 186:199–214.
- Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.
- Xie, F. and Xu, Y. (2020). Bayesian projected calibration of computer models. *Journal of the American Statistical Association*, 0(0):1–18.

CHAPTER 2

CALIBRATED COMPUTER MODEL USING MARKOV CHAIN MONTE CARLO

In Section 1, we restate the original Kennedy O'Hagan (KOH) model framework Kennedy and O'Hagan (2001) as a hierarchical regression model with GP prior. In Section 2, we introduce the general consistency theorem and assumptions we need for the full Bayesian calibration model to be consistent. In Section 3 we check the assumptions of the consistency theorem in the context of the KOH model. Finally, we conclude the chapter with a discussion in section 4.

2.1 Bayesian Hierarchical Formulation of Kennedy and O'Hagan Model

We present the KOH model described in Chapter 1 as a Bayesian hierarchical model with GP priors. This representation is adopted from Kejzlar et al. (2020) and is crucial for the theoretical results. It reframes the Bayesian model as a version of a non-parametric regression problem with a GP prior for $\zeta(t)$ and additive noise.

To be specific, we consider the following GP prior structure:

$$\begin{aligned} \zeta(t)|f(t,\theta), \delta(t) &\sim f(t,\theta) + \delta(t), \\ \delta(t) &\sim \mathcal{GP}_{\delta}(m_{\delta}(t), k_{\delta}(t,t'); \phi_{\delta}), \\ f(t,\theta) &\sim \mathcal{GP}_{f}(m_{f}(t,\theta), k_{f}((t,\theta), (t',\theta')); \phi_{f}). \end{aligned}$$
(2.1)

where ϕ_{δ} and ϕ_f denote the hyperparameters associated with the Gaussian prior on δ and f respectively. For example, $\phi_f = (\eta_f, l_f)$ comprise of η_f , the scale parameter of the covariance function and l_f , the length scale parameters of the covariance function, l_f . We first consider a prior distribution for the scale parameter σ^2 as

$$\sigma^2 \sim \Pi(\sigma^2) \tag{2.2}$$

With $\phi = (\phi_{\delta}, \phi_f) \in A_{\phi}$, we assume the following prior distribution of ϕ

$$\boldsymbol{\phi} \sim \Pi(\boldsymbol{\phi}) \tag{2.3}$$

Finally, we consider the following prior distribution for the calibration parameter $\theta \in \Theta$

$$\boldsymbol{\theta} \sim \Pi(\boldsymbol{\theta}) \tag{2.4}$$

Based on (1.3) and (2.1) the joint distribution $p(\zeta, z | \theta, \phi)$ is a multivariate normal distribution with mean

$$M_{\zeta,z|\theta,\phi} = \begin{pmatrix} M_f(T_y(\theta)) + M_{\delta}(T_y) \\ M_f(T_z(\widetilde{\theta})) \end{pmatrix}, \qquad (2.5)$$

and covariance

$$K_{\zeta,z|\theta,\phi} = \begin{pmatrix} K_f(T_y(\theta), T_y(\theta)) + K_{\delta}(T_y, T_y) & K_f(T_y(\theta), T_z(\widetilde{\theta})) \\ K_f(T_z(\widetilde{\theta}), T_y(\theta)) & K_f(T_z(\widetilde{\theta}), T_z(\widetilde{\theta})) \end{pmatrix}$$
(2.6)

Here, $M_f(T_y(\theta))$ is a column vector with j^{th} element $m_f(t_j, \theta)$, $M_{\delta}(T_y)$ is a column vector with j^{th} element $m_{\delta}(t_j)$, and $M_f(T_z(\widetilde{\theta}))$ is a column vector with j^{th} element $m_f(\widetilde{t}_j, \widetilde{\theta}_j)$. $K_f(T_y(\theta), T_y(\theta))$ is the matrix with (i, j) element $k_f((t_i, \theta), (t_j, \theta))$, $K_{\delta}(T_y, T_y)$ is the matrix with (i, j) element $k_{\delta}(t_i, t_j)$, and $K_f(T_z(\widetilde{\theta}), T_z(\widetilde{\theta}))$ is the matrix with (i, j) element $k_f((\widetilde{t}_i, \widetilde{\theta}_i), (\widetilde{t}_j, \widetilde{\theta}_j))$. We can define the matrix $K_f(T_y(\theta), T_z(\widetilde{\theta}))$ similarly with the kernel k_f . To separate out the scale parameters, let's define new notations for the covariance matrix as follows: $K_f(T_y(\theta), T_y(\theta)) = \eta_f \widetilde{\Sigma}_{11}, K_{\delta}(T_y, T_y) = \eta_{\delta} \widetilde{\Sigma}_{11'}, K_f(T_y(\theta), T_z(\widetilde{\theta})) = \eta_f \widetilde{\Sigma}_{12}$, and $K_f(T_z(\widetilde{\theta}), T_z(\widetilde{\theta})) = \eta_f \widetilde{\Sigma}_{22}$.

Consequently, the conditional distribution $p(\zeta | z, \theta, \phi)$ is also a multivariate normal with the mean and covariance functions:

$$m_{\zeta|z,\theta,\phi} = m_f(t,\theta) + m_{\delta}(t) + \sum_{i=1}^s \sum_{j=1}^s k_f((t,\theta), (\widetilde{t}_j, \widetilde{\theta}_j)) (\frac{1}{\eta_f} \widetilde{\Sigma}_{22(j,i)}^{-1}(l_f, l_\theta)) [z_i - m_f(\widetilde{t}_i, \widetilde{\theta}_i))]$$

$$(2.7)$$

$$k_{\zeta|z,\theta,\phi} = k_f((\boldsymbol{t},\theta),(\boldsymbol{t}',\theta')) + k_{\delta}(\boldsymbol{t},\boldsymbol{t}') - \sum_{i=1}^s \sum_{j=1}^s k_f((\boldsymbol{t},\theta),(\widetilde{\boldsymbol{t}}_j,\widetilde{\boldsymbol{\theta}}_j)) \left(\frac{1}{\eta_f} \widetilde{\Sigma}_{22(j,i)}^{-1}(l_f,l_{\delta})\right) k_f((\boldsymbol{t},\theta),(\widetilde{\boldsymbol{t}}_j,\widetilde{\boldsymbol{\theta}}_j))$$
(2.8)

where $(1/\eta_f)\tilde{\Sigma}_{22(j,i)}^{-1}(l_f, l_{\delta})$ is the (j, i) element of $[K_f(T_z(\tilde{\theta}), T_z(\tilde{\theta}))]^{-1}$ which depends on l_f, l_{δ} . To make our argument simple and efficient, without loss of generality, we are going to assume zero mean for f_m and δ .

In what follows, we are interested in the posterior density of the parameter $S = (\zeta, \sigma)$

$$\Pi(S|\mathbf{y}, \mathbf{z}) = \frac{p(S, \mathbf{y}, \mathbf{z})}{p(\mathbf{y}, \mathbf{z})} = \frac{p(\mathbf{y}|S)p(S|\mathbf{z})}{p(\mathbf{y}, \mathbf{z})} = \frac{1}{p(\mathbf{y}, \mathbf{z})} \left[\prod_{i}^{n} p(y_i|\zeta_i, \sigma) \right] \frac{d\Pi(S|\mathbf{z})}{dS}$$
(2.9)

where the marginal density $p(\mathbf{y}, \mathbf{z})$ is given by $p(\mathbf{y}, \mathbf{z}) = \int \left[\prod_{i=1}^{n} p(y_i | \zeta_i, \sigma)\right] d\Pi(S|\mathbf{z})$ and the prior density $d\Pi(S|\mathbf{z})/dS$ is given by

$$\begin{aligned} \frac{d\Pi(S|z)}{dS} &= \int_{\Theta} \int_{A_{\phi}} p(S,\theta,\phi|z) d\phi d\theta = \int_{\Theta} \int_{A_{\phi}} p(S|z,\theta,\phi) p(\phi,\theta|z) d\phi d\theta \\ &= \left(\int_{\Theta} \int_{A_{\phi}} p(\zeta|z,\theta,\phi) p(\phi|z) p(\theta) d\phi d\theta \right) \times p(\sigma) \\ &= \left(\int_{\Theta} \int_{A_{\phi_{\delta}}} \int_{A_{\phi_{f}}} p(\zeta|z,\theta,\phi) p(\phi_{f}|z) p(\phi_{\delta}) p(\theta) d\phi_{f} d\phi_{\delta} d\theta \right) \times p(\sigma) \\ &= \frac{d\Pi_{1}(\zeta|z)}{d\zeta} \times \frac{d\Pi_{2}(\sigma)}{d\sigma} \end{aligned}$$

2.2 Preliminaries

For the purposes of this section, we shall assume that $\sigma^2 = 1$ and $\Pi_1 = \Pi_{\zeta}$. The results can be easily extended to unknown σ^2 with minor modifications. The basic idea of posterior consistency is that the posterior probability of an arbitrary neighborhood around the true parameter goes to 1 as the sample size *n* goes to infinity. Consider the following neighborhood,

$$\mathcal{U}_{\epsilon_n} = \left\{ \zeta : \frac{1}{n} \sum_{i=1}^n |\zeta(t_i) - \zeta_0(t_i)| \le \epsilon_n \right\}$$
(2.10)

Then in the KOH model setup we described in the previous section, posterior consistency means that as *n* goes to ∞ ,

$$\Pi_{\zeta}\left\{\zeta\in\mathcal{U}_{\epsilon_n}|\mathbf{y},\mathbf{z}\right\}\to 0 \qquad a.s.\ [P_{\zeta_0}]$$

where the posterior distribution is given by (2.9).

There are many pieces of literature regarding posterior consistency. An early result of posterior consistency is by Doob (1949), which shows that the posterior distribution is consistent almost everywhere on prior support. Many Bayesians were satisfied with the result at the time, but it was not useful to check the consistency at a particular density. Then the breakthrough was made by Schwartz (1965), who proved a theorem for posterior consistency of parameters from i.i.d random variables. The conditions for this theorem are the prior positivity of neighborhoods of the true parameters and the existence of uniformly consistent test functions. This introduction

of test function conditions made Schwartz's theorem a powerful tool for checking the posterior consistency in many applications. Please refer to Barron (1999) for further insight into Schwartz's theorem. Choudhuri et al. (2004) extended this result to a triangular array of independent but non-identically distributed observations in terms of convergence in probability. Choi and Schervish (2007) strengthened this result to almost sure convergence. The proof of KOH model consistency is based on this extension of the Schwartz theorem, and note that the priors are updated with computer model evaluations, which makes the proof complicated.

2.2.1 Posterior Consistency Theorem

Theorem 1. Let $\{y_i\}_{i=1}^{\infty}$ be independently and normally distributed with mean $\zeta(t_i)$ and standard deviation $\sigma > 0$ with respect to a common σ -finite measure. Let $\zeta \in \mathcal{F}$, where \mathcal{F} denotes the space of continuously differentiable functions on $\Omega = [0, 1]^p$. Let $\zeta \in \mathcal{F}$ denote the parameter of interest. Let the prior of ζ depending on a known set of inputs $z = (z_1, \dots, z_s)$ be given by the product measure $d\Pi_{\zeta}(\zeta|z)/d\zeta$. For $\zeta_0 \in \mathcal{F}$, P_{ζ_0} denote the joint conditional distribution of $\{y_i\}_{i=1}^{\infty}$ given the true ζ_0 . Let $\{U_n\}_{n=1}^{\infty}$ be a sequence of subsets of \mathcal{F} . Now under the following two main assumptions,

Assumption 2.2.1. Existence of tests. Suppose there exist test functions $\{\Phi_n\}_{n=1}^{\infty}$, a sequence of sets $\{\mathcal{F}_n\}_{n=1}^{\infty}$ and constants $C_1, C_2, c_1, c_2 > 0$ such that

- 1. $\sum_{n=1}^{\infty} E_{\zeta_0} \Phi_n < \infty$
- 2. $\sup_{\zeta \in U_n^c \cap \mathcal{F}_n} E_S(1 \Phi_n) \leq C_1 \exp(-c_1 n)$

Assumption 2.2.2. Prior positivity of neighborhoods. The marginal prior $\Pi_{\zeta}(\zeta)$ satisfies

- 1. $\Pi_{\zeta}(\mathcal{F}_n^c) \leq \exp(-Cn\epsilon_n^2)$
- 2. $\Pi_{\zeta}(\|\zeta \zeta_0\|_{\infty} < \epsilon_n) \ge \exp(-Cn\epsilon_n^2)$

The posterior distribution as in (2.9) satisfies

$$\Pi\{\zeta \in U_n^c | \mathbf{y}, \mathbf{z}\} \to 0 \qquad a.s. \ [P_{\zeta_0}] \tag{2.11}$$

We now simplify the posterior distribution in (2.9) as follows

$$\Pi(\zeta \in U_{n}^{c}|\mathbf{y}, \mathbf{z}) = \frac{\int_{U_{n}^{c}} \left[\prod_{i}^{n} p(y_{i}|\zeta_{i}, \sigma) \right] d\Pi(\zeta|\mathbf{z})}{\int \left[\prod_{i}^{n} p(y_{i}|\zeta_{i}, \sigma) \right] d\Pi(\zeta|\mathbf{z})}$$

$$\leq \Phi_{n} + \frac{(1 - \Phi_{n}) \int_{U_{n}^{c} \cap \mathcal{F}_{n}} \left[\prod_{i=1}^{n} (p(y_{i}|\zeta_{i}, \sigma)/p(y_{i}|\zeta_{0,i}, \sigma_{0})) \right] d\Pi(\zeta|\mathbf{z})}{\int \left[\prod_{i=1}^{n} (p(y_{i}|\zeta_{i}, \sigma)/p(y_{i}|\zeta_{0,i}, \sigma_{0})) \right] d\Pi(\zeta|\mathbf{z})}$$

$$+ \frac{\int_{U_{n}^{c} \cap \mathcal{F}_{n}^{c}} \left[\prod_{i=1}^{n} (p(y_{i}|\zeta_{i}, \sigma)/p(y_{i}|\zeta_{0,i}, \sigma_{0})) \right] d\Pi(\zeta|\mathbf{z})}{\int \left[\prod_{i=1}^{n} (p(y_{i}|\zeta_{i}, \sigma)/p(y_{i}|\zeta_{0,i}, \sigma_{0})) \right] d\Pi(\zeta|\mathbf{z})}$$

$$= \Phi_{n} + \frac{\mathbf{I}_{(1,n)}(\mathbf{y})}{\mathbf{I}_{(3,n)}(\mathbf{y})} + \frac{\mathbf{I}_{(2,n)}(\mathbf{y})}{\mathbf{I}_{(3,n)}(\mathbf{y})}$$

$$(2.12)$$

Assumption 2.2.1. 1 and Assumption 2.2.1. 2 requires uniformly consistent sequence of tests for testing $H_0: \zeta = \zeta_0$ versus $H_1: \zeta \in U_n^c$. These conditions are used to show that the numerator, $\mathbf{I}_{(1,n)}(\mathbf{y})$ in (2.12) goes to 0 as $n \to \infty$. The construction of these test functions for GPs has been studied in Choi and Schervish (2007) (see Section 3.5.3) and uses Assumption 2.2.3 on the design points.

Note, Assumption 2.2.2. 1 assumes that the prior gives negligible probability outside the set \mathcal{F}_n , shall be used to show that the numerator, $\mathbf{I}_{(2,n)}(\mathbf{y})$ in (2.12) goes to 0 as $n \to \infty$. Finally, Assumption 2.2.2. 2 states that the marginal prior Π_{ζ} gives enough mass on the set $B_n = \{\zeta : \|\zeta - \zeta_0\|_{\infty} \le \epsilon_n\}$ where B_n denotes a set around the true process ζ_0 , and Assumption 2.2.2. 2 give a positive prior on the neighborhood of the true parameter ζ_0 , and it allows us to show that the denominator, $\mathbf{I}_{(3,n)}(\mathbf{y})$ in (2.12) goes to ∞ .

Assumption 2.2.2. 2 is a problem well-known in the GP community as *small ball probability*, and it has been studied by many in the probability literature (see Ghosal and Roy (2006); Li and Shao (2001); Tokdar and Ghosh (2007); Vaart and Zanten (2008)). The small ball probability in GP prior is closely related to posterior contraction through the *c*oncentration function

$$\phi_{\zeta_0}(\epsilon_n) = \inf_{h \in \mathbb{H}: \|h - \zeta_0\| < \epsilon_n} \|h\|_{\mathbb{H}}^2 - \log \Pi(\|\zeta\| < \epsilon_n)$$

where, \mathbb{H} is the Reproducing Kernel Hilbert Space (RKHS) associated to the prior, and $\|\cdot\|_{\mathbb{H}}$ is the corresponding RKHS norm. The first infimum part of the concentration function denotes the decrease in prior mass when the ball is shifted from the origin to ζ_0 . The second part represents prior mass in a ball of radius ϵ_n around zero function (see Ghosal and van der Vaart (2017), Vaart and Zanten (2008), and van der Vaart and van Zanten (2011) for further description of this function). By the Theorem 2.1 of Vaart and Zanten (2008), $\Pi(\|\zeta - \zeta_0\|_{\infty} < \epsilon_n) \ge \exp(-\phi_{\zeta_0}(\epsilon_n)/2)$. In the Supplement, we show that $\phi_{\zeta_0}(\epsilon_n)/2 \le n\epsilon_n^2$ for the Squared Exponential covariance and Matérn covariance class.

2.2.2 Assumptions for the KOH model

In order to apply consistency Theorem 1 to KOH model, we need assumptions on the smoothness of the Gaussian process prior $\zeta | z, \theta, \phi$ (see equations (2.7) and (2.8)) together with a set of assumptions on the priors for ϕ and θ (see equations (1.1), (2.3) and (2.4)). Additionally, we need an assumption on the rate of design points filling out the interval $[0, 1]^p$.

2.2.2.1 Assumptions for design points under KOH

Assumption 2.2.3. Non random design points.

For each hypercube H in $[0, 1]^p$, let $\lambda(H)$ be its Lebesgue measure. Suppose that there exists a constant $0 < K_d < 1$ such that whenever $\lambda(H) \ge 1/(nK_d)$, H contains at least one design point.

Note that equally spaced design satisfies Assumption 2.2.3. Suppose $t_i = (t_{i,1}, \dots, t_{i,p})$, $i = 1, \dots, n$ are spaced such that $|t_{i+1,k} - t_{i,k}| = n^{-1/d}$, $i, j = 1, \dots, n$, $k = 1, \dots, p$ and $0 = t_{0,k} < t_{1,k} \le \dots \le t_{n,k} < t_{n+1,k} = 1$. Then each hypercube *H* with a Lebesgue measure of at least 1/n contains at least one design point. By the same logic, space-filling designs such as Latin Hypercube Sampling with maxmin distance Morris and Mitchell (1995), which is commonly used in the computer model fields, also satisfy this assumption. Please refer to Santner et al. (2003) for further information regarding designs in computer experiments.

2.2.2.2 Assumptions on the prior for KOH

The prior assumption comprises two parts. One is for the smoothness of sample paths of Gaussian Processes f and δ . (Assumption 2.2.4. 1 and Assumption 2.2.4. 2). The other is for technical assumptions on priors of hyperparameters and calibration parameters (Assumption 2.2.5 1, and Assumption 2.2.5. 2, Assumption 2.2.5. 3).

Assumption 2.2.4. Gaussian process.

1. For the Gaussian processes \mathcal{GP}_f and \mathcal{GP}_{δ} , assume zero mean $m_f = m_{\delta} = 0$. In addition, assume the scale parameters $\eta_f, \eta_{\delta} > 0$, length scale parameters $l_f = (l_{f_1}, \dots, l_{f_{p+q}}) \in \mathbb{R}^{p+q}$, $l_{\delta} = (l_{\delta_1}, \dots, l_{\delta_p}) \in \mathbb{R}^p$, inputs $\mathbf{t} = (t^1, t^2, \dots, t^p) \in \mathbb{R}^p$, $\mathbf{t}' = (t'^1, t'^2, \dots, t'^p) \in \mathbb{R}^p$, and calibration parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{p'}), \quad \widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \dots, \widetilde{\theta}_{p'}) \in \boldsymbol{\Theta}$, let the covariance functions $k_f((\mathbf{t}, \boldsymbol{\theta}), (\mathbf{t}', \boldsymbol{\theta}'))$ and $k_{\delta}(\mathbf{t}, \mathbf{t}')$ be a product of isotropic and integrable covariance functions, one for each dimension.

$$k_{f}((t,\theta),(t',\theta')) = \eta_{f}R(t^{1},t'^{1};l_{f_{1}})\cdots R(t^{p},t'^{p};l_{f_{p}})R(\theta_{1},\widetilde{\theta}_{1};l_{f_{p+1}})\cdots R(\theta_{p'},\widetilde{\theta}_{p'};l_{f_{p+p'}})$$
$$k_{\delta}(t,t') = \eta_{\delta}R(t^{1},t'^{1};l_{\delta_{1}})R(t^{2},t'^{2};l_{\delta_{2}})\cdots R(t^{p},t'^{p};l_{\delta_{p}})$$

where each $R(t^{i}, t'^{i}; l_{i}) = R\left(\frac{|t^{i}-t'^{i}|}{l_{i}}\right)$, and R is the positive multiple of density.

2. For the Gaussian Process \mathcal{GP}_f and \mathcal{GP}_δ , the mean functions m_f , m_δ are continuously differentiable with respect to input t and covariance functions k_f and k_δ have partial derivatives up to order 2p + 2.

Assumption 2.2.5. *Hyperparameters, calibration parameter and observational error scale parameter.*

- 1. η_f and η_δ have prior distributions with support \mathbb{R}^+ , and there exists $0 < h < \frac{1}{2}$ and $b_1, b_2, b_3, b_4 > 0$ such that $\mathbb{P}[\eta_f > n^h] \le b_1 \exp(-b_2 n)$, and $\mathbb{P}[\eta_\delta > n^h] \le b_3 \exp(-b_4 n)$, $\forall n \ge 1$.
- 2. l_f and l_{δ} have prior distributions with bounded support on \mathbb{R}^+ . That is for the bounds, $B_f, B_{\delta} > 0$, set $\mathbb{P}[l_{f_i} > B_f] = 0$, for $i = 1, \dots, p + p'$, and $\mathbb{P}[l_{\delta_j} > B_{\delta}] = 0$, for $j = 1, \dots, p$
- 3. θ has a continuous density on the support Θ such that $\int_{\Theta} p(\theta) d\theta = 1$.

Note that in Assumption 2.2.4. 1, the product of the isotropic covariance functions is also an isotropic covariance function. The functional form of $R(\cdot)$ in the covariance kernel in Assumption 2.2.4. 1 affects the proof of showing the Assumption 2.2.2. 1. Choi (2005) avoids functional form by considering the continuity of supremum of variance terms (the same assumption was also used

in Lemma 1 in Kejzlar et al. (2020)). In this paper, we are going to prove Assumption 2.2.2. 1 for general case, which encompasses the widely used covariance functions such as Squared Exponential covariance. Assumption 2.2.4. 1 and Assumption 2.2.4. 2 together with Assumption 2.2.5. 1 and Assumption 2.2.5. 2 are sufficient conditions to ensure that the support of GP prior for $\zeta | z, \theta, \phi$ contains every continuously differentiable function based on the arguments by Theorem 4 of Ghosal and Roy (2006) and Theorem 4.5 of Tokdar and Ghosh (2007). Particularly, Assumption 2.2.4. 2 implies that the same condition holds for the conditional process $\zeta | z, \theta, \phi$, which is a function of f and δ . This enables us to apply Theorem 4 of Ghosal and Roy (2006) and Theorem 4.5 of Tokdar and Ghosh (2007) in KOH model setup. Many covariance functions satisfy this assumption, including the product of squared exponential covariance functions. For detailed descriptions of the analytic properties of covariance functions, please check Cramer and Leadbetter (1967), Adler (1981), and Adler (1990). Furthermore, Assumption 2.2.4. 1 and Assumption 2.2.4. 2 guarantees the existence of continuous sample derivatives with probability 1, p.171 and p.185 of Cramer and Leadbetter (1967). In addition, by the Theorem 5 of Ghosal and Roy (2006), these Assumption 2.2.4. 1 and Assumption 2.2.4. 2 guarantee that the GP prior $\zeta | z, \theta, \phi$ gives non zero probability to the set of continuous functions with bounded derivatives.

Assumption 2.2.5. 1 relaxes the assumption of compact support for all hyperparameters as used in the empirical Bayes approach Kejzlar et al. (2020) of KOH model. We relaxed that assumption in such a way that scale parameters for GP covariances, n_f and n_{δ} , have an exponentially small probability of blowing up. Assumption 2.2.5. 2 indicates that the length scale parameters have bounded support. In the algorithm, we used log transformation to have the support of length scale parameters to be \mathbb{R}^+ . However, we can always use different transformation. For example, sigmoid transformation could be used to have a support on [0, 1]. Hence, without loss of generality let's assume they are bounded for theoretical investigation. Then, this condition allows us to upper bound the quantity $\sum_{i=1}^{s} \sum_{j=1}^{s} \left| \tilde{\Sigma}_{22(j,i)}^{-1} (l_f, l_{\delta}) \right|$ by a constant \bar{C} in a bounded support. This assumption is a technical assumption to ensure that the conditional prior $\Pi_{\zeta}(\cdot|z)$ assigns exponentially small probability to \mathcal{F}_n^C . Assumption 2.2.5. 3 assumes continuous density on the support of calibration parameters. In practice, the support of calibration parameters is determined by domain experts.

2.3 Posterior Consistency in Kenndy and O'Hagan Model

We have already presented the KOH model in Section 2.1. We shall first assume that the design points $\mathbf{t} = (t_1, \dots, t_n)$ on which the physical process ζ is defined (see (1.1)) follows the non random design as discussed under Assumption 2.2.3. As described in Section 2.1, we have considered a hierarchical Gaussian prior on the physical process ζ as in (2.1). With \mathbf{z} defined in (1.3), we obtain an induced conditional multivariate Gaussian prior on the physical process ζ whose mean and covariance kernels are given by (2.7) and (2.8) respectively. We first assume that this conditional prior $\zeta | \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\phi}$ satisfies all the conditions under Assumption 2.2.4. We also assume that the prior on hyperparameter $\boldsymbol{\phi} = (n_f, n_\delta, l_f, l_\delta)$ satisfies Assumption 2.2.5. 1 and Assumption 2.2.5. 2. Under these assumptions, we next present the main result on the joint consistency of the physical process ζ and the scale parameter σ under the KOH model.

2.3.1 Posterior Consistency Theorem of KOH Model

Theorem 2. Let $\{y_i\}_{i=1}^{\infty}$ be independently and normally distributed with mean $\zeta(t_i)$ and standard deviation $\sigma > 0$ with respect to a common σ -finite measure. Let ζ denote the parameter of interest. Let the design points $t_i \in [0,1]^p$ satisfy Assumption 2.2.3. With reference to Section 2.1, let $z = (z_1, \dots, z_s)$ be as defined in (1.3). We assume that the conditional prior $\zeta | z, \theta, \phi$ satisfies Assumption 2.2.4. 1, Assumption 2.2.4. 2. We also assume that the prior on ϕ satisfies Assumption 2.2.5. 1 and 2.2.5. 2. In addition, assume the calibration parameter satisfies 2.2.5. 3. Let P_{ζ_0} denote the joint conditional distribution of $\{y_i\}_{i=1}^{\infty}$ given the true ζ_0 . Let the function ζ_0 be continuously differentiable, then for a sequence ϵ_n that goes to 0 and for the neighborhood \mathcal{U}_{ϵ_n} of ζ as defined in equation (2.10),

$$\Pi_{\zeta}\{\zeta \in \mathcal{U}_{\epsilon_n}^c | \mathbf{y}, \mathbf{z}\} \to 0 \qquad a.s. \ [P_{\zeta_0}] \tag{2.13}$$

where $\Pi_{\zeta}\{|\mathbf{y}, \mathbf{z}\}$ denotes the posterior distribution of ζ given the data $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{z} = (z_1, \dots, z_s)$.

The proof of Theorem 2 is based on the Theorem 1. In order to prove Theorem 2, we shall show

that Assumption 2.2.2 and Assumption 2.2.1 hold under the KOH model. We present the proofs of Assumption 2.2.2 and Assumption 2.2.1 under the KOH model in the Appendix.

2.4 Conclusion and Discussion

The Kennedy O'Hagan model assumes both physical data and computer model evaluation data are available and make use of the Gaussian Process to calibrate the parameters and to predict physical process at new input space. Although the identifiability issue of θ is still a concern in the KOH model, it has been widely used in diverse fields. At the same time, researchers tried to provide a theoretical justification for this model. For example, you can look at Tuo and Wu (2015, 2016) and Plumlee (2017), among others.

However, the theoretical properties of the fully Bayesian approach of the KOH model have not been investigated much. The most relevant paper is Xie and Xu (2020), in which they proposed a Bayesian projected calibration method following the frequentist L_2 -projected calibration method in Tuo and Wu (2015) and showed theoretical justification. An additional relevant study conducted by Gramacy and Apley (2015) employed a localized approach with parallelization techniques to approximate a large GP emulator. However, it is important to note that this particular work focused on GP computer emulation and did not operate within the KOH model framework. Hence the main contribution of this paper is the theoretical justification of using the original full Bayesian KOH model. Our approach is based on the extended use of Schwartz's theorem in our model setup to prove the consistency of the fitted emulator.

One issue we haven't discussed in this paper is the identifiability of θ . This is because it is not an issue for the prediction. However, this would be a great next research topic. As Kennedy and O'Hagan (2001) stated, the parameter is the one best explains the difference between the true physical model and the computer model. So it lacks physical meaning. But when it comes to applications, as discussed in Higdon et al. (2005) and Han et al. (2003), there are both cases where the parameter has little or no physical meaning (tuning parameter) and the cases where it does have a physical meaning (calibration parameter). This is an integral part for the domain scientists because the calibration parameters themselves are the interest to them sometimes, and it helps for scientific understanding of the model and future scientific extension. Hence developing the new definition of θ , which contains the concept of both approximating the true physical model and physical interpretation would be a big breakthrough in this field.

BIBLIOGRAPHY

- Abrahamsen, P. (1997). A review of gaussian random fields and correlation functions, volume 917. Norwegian Computing Center.
- Adler, R. J. (1981). The Geometry of Random Fields. Wiley.
- Adler, R. J. (1990). An introduction to continuity, extrema, and related topics for Gaussian process, volume 12. IMS Lecture notes-monograph series.
- Barron, A. R. (1999). Information-theoretic characterization of bayes performance and the choice of priors in parametric and nonparametric problems. *In Bayesian Statistics*, 6:27–52.
- Choi, T. (2005). Posterior Consistency in Nonparametric Regression Problems under Gaussian Process Priors. PhD thesis, Carnegie Mellon University.
- Choi, T. and Schervish, M. J. (2007). Posterior Consistency in Nonparametric Regression Problems under Gaussian Process Priors. *Journal contribution*.
- Choudhuri, N., Ghosal, S., and Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050 1059.
- Cramer, H. and Leadbetter, M. R. (1967). Stationary and related stochastic processes. Wiley.
- Doob, J. L. (1949). Application of the theory of martingales. *in Le Calcul des Probabilites et ses Applications*, 40:23–27.
- Ghosal, S. and Roy, A. (2006). Posterior consistency of gaussian process prior for nonparametric binary regression. *Ann. Statist.*, 34(5):2413–2429.
- Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press.
- Gramacy, R. B. and Apley, D. W. (2015). Local gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24(2):561–578.
- Han, G., Santner, T. J., and Rawlinson, J. J. (2003). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics*, 51(4):464 474.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2005). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26:448–466.
- Kejzlar, V., Son, M., Bhattacharya, S., and Maiti, T. (2020). A fast, scalable, and calibrated computer model emulator: An empirical bayes approach.

- Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:425–464.
- Li, W. V. and Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. *Stochastic Processes: Theory and Methods. Handbook of Statistics*, 19:533–598.
- Morris, M. D. and Mitchell, T. J. (1995). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Journal of Statistical Planning and Inference*, 43:381–402.
- Plumlee, M. (2017). Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112:1274–1285.
- Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003). *The design and analysis of computer experiments*, volume 1. Springer.
- Schwartz, L. (1965). On bayes procedures. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 4:10–26.
- Tokdar, S. T. and Ghosh, J. K. (2007). Posterior consistency of logistic gaussian process priors in density estimation. *Journal of Statistical Planning and Inference*, 137(1):34 42.
- Tuo, R. and Wu, C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals* of *Statistics*, 43:2331–2352.
- Tuo, R. and Wu, C. F. J. (2016). A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. SIAM/ASA Journal on Uncertainty Quantification, 4:767–795.
- Vaart, A. W. V. D. and Zanten, J. H. V. (2008). Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, 36:1435–1463.
- van der Vaart, A. W. and van Zanten, J. H. (2011). Information rates of nonparametric gaussian process methods. *Journal of Machine Learning Research*, pages 2095–2119.
- Xie, F., Jin, W., and Xu, Y. (2019). Rates of contraction with respect to L_2 -distance for Bayesian nonparametric regression. *Electronic Journal of Statistics*, 13(2):3485 3512.
- Xie, F. and Xu, Y. (2020). Bayesian projected calibration of computer models. *Journal of the American Statistical Association*, 0(0):1–18.

APPENDIX A

PROOF OF THEOREM 1

Note z is implicitly conditioned on the prior. If we decompose the following term.

$$\begin{split} \Pi_{\zeta}(\mathcal{U}_{\epsilon_{n}}^{c}|\mathbf{y},z) &= \int_{\Theta} \int_{A_{\phi}} \Pi(\mathcal{U}_{n}^{c},\theta,\phi|\mathbf{y},z)d\phi d\theta \\ &= \frac{\int_{\mathcal{U}_{n}^{c}} \int_{\Theta} \int_{A_{\phi}} p(\zeta,\mathbf{y},z,\theta,\phi)d\phi d\theta d\zeta}{\int_{\mathcal{F}} \int_{\Theta} \int_{A_{\phi}} p(\mathbf{y}|\zeta) \Big(\int_{\Theta} \int_{A_{\phi}} p(\zeta,\theta,\phi,z)d\phi d\theta \Big) d\zeta} \\ &= \frac{\int_{\mathcal{U}_{n}^{c}} p(\mathbf{y}|\zeta) \Big(\int_{\Theta} \int_{A_{\phi}} p(\zeta,\theta,\phi,z)d\phi d\theta \Big) d\zeta}{\int_{\mathcal{F}} p(\mathbf{y}|\zeta) d\Pi(\zeta|z)} \\ &= \frac{\int_{\mathcal{U}_{n}^{c}} p(\mathbf{y}|\zeta) d\Pi(\zeta|z)}{\int_{\mathcal{F}} \prod_{i=1}^{n} (p(y_{i}|\zeta_{i},\sigma)/p(y_{i}|\zeta_{0,i},\sigma_{0}))d\Pi(\zeta|z)} \\ &\leq \Phi_{n} + \frac{(1-\Phi_{n}) \int_{\mathcal{U}_{n}^{c}\cap\mathcal{F}_{n}} \prod_{i=1}^{n} (p(y_{i}|\zeta_{i},\sigma)/p(y_{i}|\zeta_{0,i},\sigma_{0}))d\Pi(\zeta|z)}{\int_{\mathcal{F}} \prod_{i=1}^{n} (p(y_{i}|\zeta_{i},\sigma))/(p(y_{i}|\zeta_{0,i},\sigma_{0}))d\Pi(\zeta|z)} \\ &+ \frac{\int_{\mathcal{U}_{n}^{c}\cap\mathcal{F}_{n}^{c}} \prod_{i=1}^{n} (p(y_{i}|\zeta_{i},\sigma))/(p(y_{i}|\zeta_{0,i},\sigma_{0}))d\Pi(\zeta|z)}{\int_{\mathcal{F}} \prod_{i=1}^{n} (p(y_{i}|\zeta_{i},\sigma))/(p(y_{i}|\zeta_{0,i},\sigma_{0}))d\Pi(\zeta|z)} \\ &= \Phi_{n} + \frac{\mathbf{I}_{(1,n)}(\mathbf{y})}{\mathbf{I}_{(3,n)}(\mathbf{y})} + \frac{\mathbf{I}_{(2,n)}(\mathbf{y})}{\mathbf{I}_{(3,n)}(\mathbf{y})} \\ &\leq \exp(-C_{1}n\varepsilon_{n}^{2}) + \exp(-(C_{2}n\varepsilon_{n}^{2}/2)} + \frac{\exp(-C_{3}n\varepsilon_{n}^{2}/2)}{\exp(-(C+C_{4})n\varepsilon_{n}^{2})} \\ &= \exp(-C_{1}n\varepsilon_{n}^{2}) + \exp(-(C_{2}/2+C-C_{4})n\varepsilon_{n}^{2}) + \exp(-(C_{3}/2+C-C_{4})n\varepsilon_{n}^{2}) \end{split}$$

Where

$$\mathbf{I}_{1,n} = (1 - \Phi_n) \int_{\mathcal{U}_n^c \cap \mathcal{F}_n} \prod_{i=1}^n (p(y_i | \zeta_i) / p(y_i | \zeta_{0,i})) d\Pi(\zeta | z)$$

$$\mathbf{I}_{2,n} = \int_{\mathcal{U}_n^c \cap \mathcal{F}_n^c} \prod_{i=1}^n (p(y_i | \zeta_i)) / (p(y_i | \zeta_{0,i})) d\Pi(\zeta | z)$$

$$\mathbf{I}_{3,n} = \int_{\mathcal{F}} \prod_{i=1}^n (p(y_i | \zeta_i)) / (p(y_i | \zeta_{0,i})) d\Pi(\zeta | z)$$

It suffices to show the following four results,

$$\Phi_n \le \exp(-C_1 n \varepsilon_n^2) \tag{A.1}$$

$$\mathbf{I}_{1,n} \le \exp(-C_2 n \varepsilon_n^2 / 2) \tag{A.2}$$

$$\mathbf{I}_{2,n} \le \exp(-C_3 n \varepsilon_n^2 / 2) \tag{A.3}$$

$$\mathbf{I}_{3,n} \le \exp(-(C+C_4)n\varepsilon_n^2) \tag{A.4}$$

Mainly, (A.1) involves Markov inequality, Borel-cantelli lemma, and Assumption. 2.2.1. (i) in the main paper. Proofs of (A.2) and (A.3) follows from Theorem 2 of Choudhuri et al. (2004) and proof of (A.4) follows from Lemma 2.4 of Xie et al. (2019). All the proofs need a slight modification from the original papers since we have prior on ζ updated with simulation data *z*. These modifications are discussed next.

Proof of (A.1): Using Assumption 2.2.1 and Markov's Inequality,

$$\implies \mathbb{P}_{0}[\Phi_{n} > \exp(-C_{1}n\varepsilon_{n}^{2})] \leq \frac{1}{\exp(-C_{1}n\varepsilon_{n}^{2})}\mathbb{E}_{0}[\Phi_{n}]$$

$$\implies \sum_{n=1}^{\infty} \mathbb{P}_{0}[\Phi_{n} > \exp(-C_{1}n\varepsilon_{n}^{2})] \leq \sum_{n=1}^{\infty} \frac{1}{\exp(-C_{1}n\varepsilon_{n}^{2})}\mathbb{E}_{0}[\Phi_{n}] \leq C\sum_{n=1}^{\infty} \mathbb{E}_{0}[\Phi_{n}] < \infty$$

$$\implies \mathbb{P}_{0}[\Phi_{n} > \exp(-C_{1}n\varepsilon_{n}^{2}) \quad i.o.] = 0$$

$$\implies \mathbb{P}_{0}[\Phi_{n} \leq \exp(-C_{1}n\varepsilon_{n}^{2})] \rightarrow 1$$

Proof of (A.2): For every non-negative function ψ_n and any set A, by Fubini's theorem,

$$\mathbb{E}_{\zeta_0}\left[\psi_n(\mathbf{y})\int_A \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)}d\Pi(\zeta|\mathbf{z})\right] = \int_A \int \psi_n(\mathbf{y})\frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)}p(\mathbf{y}|\zeta_0)d(\mathbf{y})d\Pi(\zeta|\mathbf{z})$$
$$= \int_A \mathbb{E}_{\zeta}\left[\psi_n(\mathbf{y})\right]d\Pi(\zeta|\mathbf{z})$$

Now with $A = \mathcal{U}_n^c \cap \mathcal{F}_n$ and $\psi_n = \Phi_n$, we have,

$$\mathbb{E}_{\zeta_0} \left[\mathbf{I}_{(1,n)}(\mathbf{y}) \right] = \int_{\mathcal{U}_n^c \cap \mathcal{F}_n} \mathbb{E}_{\zeta} \left[(1 - \Phi_n) \right] d\Pi(\zeta | \mathbf{z})$$
$$\leq \sup_{\zeta \in \mathcal{U}_n^c \cap \mathcal{F}_n} \mathbb{E}_{\zeta} \left[(1 - \Phi_n) \right]$$
$$\leq \exp(-C_2 n \epsilon_n^2)$$

where the last inequality follows as a consequence of Assumption. 2.2.1. Thus,

$$\mathbb{P}_{\zeta_0}\left[\mathbf{I}_{(1,n)}(\mathbf{y}) \ge \exp\left(-C_2 n\epsilon_n^2/2\right)\right] \le \exp\left(C_2 n\epsilon_n^2/2\right)\exp(-C_2 n\epsilon_n^2) = \exp\left(-C_2 n\epsilon_n^2/2\right)$$

The proof follows as a consequence of Borel-Cantelli lemma.

Proof of (A.3): Note that

$$\mathbb{E}_{\zeta_0}[\mathbf{I}_{(2,n)}(\mathbf{y})] = \mathbb{E}_{\zeta_0}\left[\int_{U_n^c \cap \mathcal{F}_n} \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z})\right]$$
$$= \int_{\mathcal{U}_n^c \cap \mathcal{F}_n^c} \mathbb{E}_{\zeta}[1] d\Pi(\zeta|\mathbf{z})$$
$$\leq d\Pi(\mathcal{F}_n^c|\mathbf{z})$$
$$\leq \exp(-C_3 n\epsilon_n^2)$$

By Assumption 2.2.2 (i) in the main paper, the second inequality holds. The proof follows as a consequence of the Borel-Cantelli Lemma. **P**roof of (A.4): Denote $\Pi_{\zeta}(\cdot|B_n, z) = \Pi_{\zeta}(\cdot \cap B_n|z)/\Pi_{\zeta}(B_n|z)$ to be the renormalized restriction of $\Pi_{\zeta}(\cdot|z)$ on B_n , where B_n is defined as,

$$B_n = \{ \zeta : \| \zeta - \zeta_0 \|_{\infty} < \epsilon_n \}$$

Let,

$$V_{n,i} = \zeta_0(t_i) - \int \zeta(t_i) d\Pi_{\zeta}(\zeta|B_n), \qquad W_{n,i} = \frac{1}{2} \int (\zeta(t_i) - \zeta_0(t_i))^2 d\Pi_{\zeta}(\zeta|B_n)$$
$$\mathcal{H}_n = \left\{ \int \prod_{i=1}^n \frac{p(y_i|\zeta_i)}{p(y_i|\zeta_{0,i})} d\Pi_{\zeta}(\zeta|z) > \Pi_{\zeta}(B_n|z) \exp(-C_4 n \varepsilon_n^2) \right\}$$

Note $V_{n,i}^2 \leq \varepsilon_n^2$ and $W_{n,i} \leq \frac{1}{2}\varepsilon_n^2$, because

$$\begin{aligned} V_{n,i}^2 &= \left(\zeta_0(\boldsymbol{x}_i) - \int \zeta(\boldsymbol{x}_i) d\Pi_{\zeta}(\zeta | B_n, \boldsymbol{z})\right)^2 \\ &\leq \int \left(\zeta_0(\boldsymbol{x}_i) - \zeta(\boldsymbol{x}_i)\right)^2 d\Pi_{\zeta}(\zeta | B_n, \boldsymbol{z}) \\ &\leq \int \|\zeta_0(\boldsymbol{x}_i) - \zeta(\boldsymbol{x}_i)\|_{\infty}^2 d\Pi_{\zeta}(\zeta | B_n, \boldsymbol{z}) \\ &\leq \varepsilon_n^2 \end{aligned}$$

and because,

$$W_{n,i} = \frac{1}{2} \int \left(\zeta(\mathbf{x}_i) - \zeta_0(\mathbf{x}_i) \right)^2 d\Pi_{\zeta}(\zeta | B_n, \mathbf{z})$$

$$\leq \frac{1}{2} \int \| \zeta(\mathbf{x}_i) - \zeta_0(\mathbf{x}_i) \|_{\infty}^2 d\Pi_{\zeta}(\zeta | B_n, \mathbf{z})$$

$$\leq \frac{1}{2} \varepsilon_n^2$$

Also note,

$$\begin{split} \int \prod_{i=1}^{n} \frac{p(y_{i}|\zeta_{0,i})}{p(y_{i}|\zeta_{i})} \Pi_{\zeta}(\zeta|B_{n}, z) \\ &= \int \sum_{i=1}^{n} \left\{ -\frac{1}{2} \log \left(2\pi \right) - \frac{1}{2} (y_{i} - \zeta_{0,i})^{2} + \frac{1}{2} \log \left(2\pi \right) + \frac{1}{2} (y_{i} - \zeta_{i})^{2} \right\} \Pi_{\zeta}(\zeta|B_{n}, z) \\ &= \int \sum_{i=1}^{n} \left\{ -\frac{1}{2} (\zeta_{0,i} + e_{i} - \zeta_{0,i})^{2} + \frac{1}{2} (\zeta_{0,i} + e_{i} - \zeta_{i})^{2} \right\} \Pi_{\zeta}(\zeta|B_{n}, z) \\ &= \int \sum_{i=1}^{n} \left\{ \frac{1}{2} (\zeta_{i} - \zeta_{0,i})^{2} + e_{i} (\zeta_{i} - \zeta_{0,i}) \right\} \Pi_{\zeta}(\zeta|B_{n}, z) \\ &= \sum_{i=1}^{n} W_{n,i} + \sum_{i=1}^{n} e_{i} V_{n,i} \\ &\leq \frac{1}{2} \varepsilon_{n}^{2} + \sum_{i=1}^{n} e_{i} V_{n,i} \end{split}$$

Hence

$$\begin{aligned} \mathcal{H}_{n}^{C} &= \left\{ \int \prod_{i=1}^{n} \frac{p(y_{i}|\zeta_{i})}{p(y_{i}|\zeta_{0,i})} \Pi_{\zeta}(\zeta|z) \leq \Pi_{\zeta}(\tilde{B}_{n}|z) \exp(-C_{4}n\varepsilon_{n}^{2}) \right\} \\ &\subset \left\{ \int \prod_{i=1}^{n} \frac{p(y_{i}|\zeta_{i})}{p(y_{i}|\zeta_{0,i})} \Pi_{\zeta}(\zeta|\tilde{B}_{n},z) \leq \exp(-C_{4}n\varepsilon_{n}^{2}) \right\} \\ &\subset \left\{ \int \sum_{i=1}^{n} \log \frac{p(y_{i}|\zeta_{i})}{p(y_{i}|\zeta_{0,i})} \Pi_{\zeta}(\zeta|B_{n},z) \leq -C_{4}n\varepsilon_{n}^{2} \right\} \\ &= \left\{ \int \sum_{i=1}^{n} \log \frac{p(y_{i}|\zeta_{0,i})}{p(y_{i}|\zeta_{i})} \Pi_{\zeta}(\zeta|B_{n},z) > C_{4}n\varepsilon_{n}^{2} \right\} \\ &\subset \left\{ \frac{1}{2}\varepsilon_{n}^{2} + \sum_{i=1}^{n} e_{i}V_{n,i} > C_{4}n\varepsilon_{n}^{2} \right\} \\ &= \left\{ \sum_{i=1}^{n} e_{i}V_{n,i} > C_{5}n\varepsilon_{n}^{2} \right\} \end{aligned}$$

Now we can use the Chernoff bound for the Gaussian random variables,

$$\mathbb{P}_{0}\left\{\sum_{i=1}^{n} e_{i}V_{n,i} > C_{5}n\epsilon_{n}^{2}\right\} \leq \exp\left(-\left(C_{5}n\epsilon_{n}^{2}\right)^{2}\frac{1}{\sum_{i=1}^{n}V_{n,i}^{2}}\right)$$
$$= \exp\left(-C_{5}^{2}\frac{n\epsilon_{n}^{4}}{\frac{1}{n}\sum_{i=1}^{n}V_{n,i}^{2}}\right)$$
$$= \exp\left(-C_{5}^{2}\frac{n\epsilon_{n}^{4}}{\epsilon_{n}^{2}}\right)$$
$$= \exp\left(-C_{5}^{2}n\epsilon_{n}^{2}\right)$$
$$\to 0$$

Therefore,

$$\mathbb{P}_0\{\mathcal{H}_n^C\} \leq \mathbb{P}_0\left\{\sum_{i=1}^n e_i V_{n,i} > Cn\varepsilon_n^2\right\} \to 0$$

With Assumption 2.2.2 (ii) in the main paper, this completes the proof.

APPENDIX B

CHECKING ASSUMPTION 2.2.1

Assumption 2.2.1. (i) and (ii) require the existence of a uniformly consistent sequence of tests for testing $H_0: \zeta = \zeta_0$ versus $H_1: \zeta \in \mathcal{U}_{\epsilon_n}^c \cap \mathcal{F}_n$. The construction of these test functions for GPs has been studied in Choi and Schervish (2007) (see Section 3.5.3) and uses Assumption 2.2.3 in the main paper about the design points. Since these test functions mainly depend on the properties of ζ and the model $y_i = \zeta(t_i) + \sigma \epsilon_i$, the adaptation of the test functions in Choi and Schervish (2007) to our case is immediate with known σ . For the completeness of the paper, the proof is adopted here based on the following five lemmas from Choi (2005).

Lemma B.0.1. Let ζ_1 be a continuous function in $[0, 1]^p$, and define $\zeta_{ij} = \zeta_i(t_j)$ for i = 0, 1 and $j = 1, \dots, n$. Let $\epsilon > 0$, and let r > 0. Let $c_n = n^{r_1}$ for $\alpha_1/2 < \tau < 1/2$ and $1/2 < \alpha_1 < 1$. Let $b_j = 1$ if $\zeta_{1j} > \zeta_{0j}$ and -1 otherwise. Let $\Psi_{1n}[\zeta_1, \epsilon]$ be the indicator of the set A_1 , where A_1 is defined as

$$A_1 = \left\{ \sum_{j=1}^n b_j \left(\frac{Y_j - \zeta_{0j}}{\sigma_0} \right) > 2c_n \sqrt{n} \right\}$$

Then there exists a constant C_3 such that for all ζ_1 that satisfy $\sum_{j=1}^n |\zeta_{1j} - \zeta_{0j}| > rn$,

$$E_{P_{\zeta_0}}(\Psi_{1n}[\zeta_1,\epsilon]) < C_3 \exp(-2c_n^2)$$

Also there exists constants C_4 and C_5 such that for all sufficiently large n and all ζ satisfying $||\zeta - \zeta_0||_{\infty} < r/4$ and for all $\sigma \le \sigma_0(1 + \epsilon)$

$$E_{P_{\zeta}}(1-\Psi_{1n}[\zeta_1,\epsilon]) \le C_4 \exp(-nC_5)$$

where P_{ζ} is the joint distribution of $\{Y_n\}_{n=1}^{\infty}$ assuming parameters to be ζ .

Proof: See Lemma 3.5.9 in Choi (2005).

Lemma B.0.2. Let ζ_1 be a continuous function in $[0, 1]^p$, and define $\zeta_{ij} = \zeta_i(t_j)$ for i = 0, 1 and $j = 1, \dots, n$. Let Ψ_{2n} be the indicator of the set A_2 , where A_2 is defined as

$$A_2 = \left\{ \sum_{j=1}^n b_j \left(\frac{Y_j - \zeta_{0j}}{\sigma_0} \right) > n(1 + \epsilon) \right\}$$

Then there exists a constant C_6 such that for all ζ_1 ,

$$E_{P_{\ell_0}}(\Psi_{2n}) < \exp(-nC_6)$$

Also there exists constants C_7 such that for all sufficiently large n and all ζ satisfying $||\zeta - \zeta_0||_{\infty} < r/4$ and for all $\sigma > \sigma_0(1 + \epsilon)$

$$E_{P_{\zeta}}(1-\Psi_{2n}[\zeta_1,\epsilon]) \le \exp(-nC_7)$$

where P_{ζ} is the joint distribution of $\{Y_n\}_{n=1}^{\infty}$ assuming parameters as ζ . **P**roof: See Lemma 3.5.10 in Choi (2005).

Lemma B.0.3. Let ζ_1 be a continuous function in $[0, 1]^p$, and define $\zeta_{ij} = \zeta_i(t_j)$ for i = 0, 1 and $j = 1, \dots, n$. Let $\epsilon > 0$, and let $0 < r < 4\sigma_0\sqrt{\epsilon - \epsilon^2}$. Let $\Psi_{3n}[\zeta_1, \epsilon]$ be the indicator of the set A_3 , where A_3 is defined as

$$A_3 = \left\{ \sum_{j=1}^n b_j \left(\frac{Y_j - \zeta_{1j}}{\sigma_0} \right) \le c(1 - \epsilon^2) \right\}$$

Then there exists a constant C_8 such that for all ζ_1 that satisfy $\sum_{j=1}^n |\zeta_{1j} - \zeta_{0j}| > rn$,

$$E_{P_{\zeta_0}}(\Psi_{3n}[\zeta_1,\epsilon]) < \exp(-nC_8)$$

Also there exists constants C_9 such that for all sufficiently large n and all ζ satisfying $||\zeta - \zeta_0|| < r/4$ and for all $\sigma \leq \sigma_0(1 - \epsilon)$

$$E_{P_{\zeta}}(1 - \Psi_{3n}[\zeta_1, \epsilon]) \le \exp(-nC_9)$$

where P_{ζ} is the joint distribution of $\{Y_n\}_{n=1}^{\infty}$ assuming parameters to be ζ .

Proof: See Lemma 3.5.11 in Choi (2005). The following two lemmas show that under Assumption 2.2.3. and Assumption 2.2.4., there exist enough number of design points to acquire $\sum_{j=1}^{n} |\zeta_{1j} - \zeta_{0j}| > rn$ in *p* dimensions.

Lemma B.0.4. Assume Assumption 2.2.3. and Assumption 2.2.4. holds. Let λ be the Lebesgue measure in \mathbb{R}^p . Let K_d be the constants mentioned in the Assumptions 2.2.3 and V be a constant such that $V > \left\|\frac{\partial}{\partial t^j}\zeta_0(t^1, \dots, t^p)\right\|_{\infty}$. Let A_V be the set of all continuous functions γ such that $\forall t_1, t_2 \in [0, 1]^p$, $|\gamma(t_1) - \gamma(t_2)| \leq V||t_1 - t_2||$. For each such function γ and ϵ , define $B_{\epsilon, \gamma} = \{t : t_1 \in [0, 1]^p\}$.

 $|\gamma(t)| > \epsilon$. Then for each $\epsilon > 0$, there exists an integer N such that for all $n \ge N$ and all $\gamma \in A_V$,

$$\sum_{i=1}^{n} |\gamma(t_i)| \ge n\lambda(B_{\epsilon,\gamma})\frac{\epsilon}{2}$$

Proof: See Lemma 4.2.1 in Choi (2005).

Lemma B.0.5. Assume Assumption 2.2.3. and Assumption 2.2.4. holds. Let A_V be the set of all continuous functions ζ such that $||\zeta||_{\infty} < M_n$ and $\left\|\frac{\partial}{\partial t_j}\zeta(t_1, \cdots, t_d)\right\|_{\infty} < V, j = 1, \cdots, d$. Then for each $\epsilon > 0$, there exist an integer N and r > 0 such that, for all $n \ge N$ and all $\zeta \in A_V$ such that $||\zeta - \zeta_0||_1, \sum_{j=1}^n |\zeta_{1j} - \zeta_{0j}| \ge rn$

Proof: See Lemma 4.2.2 in Choi (2005).

Now we are going to prove the existence of test functions in the KOH model setup.

Proof: First, define

$$\Psi_{n}[\zeta_{1},\epsilon] = 1 - (1 - \Psi_{1n}[\zeta_{1},\epsilon])(1 - \Psi_{2n}[\zeta_{1},\epsilon])(1 - \Psi_{3n}[\zeta_{1},\epsilon])$$

From Lemmas B.0.1-B.0.3, it is clear that $E_{P_{\zeta_0}}\Psi_n < \exp(-2c_n^2)$ and $E_{P_{\zeta}}(1-\Psi_n) \leq \exp(-nC_{10})$ for each $\zeta_1 \in \mathcal{U}_{\epsilon_n}^C$. To create a test that doesn't depend on a specific choice of ζ_1 in Lemma B.0.1, we make use of the covering number of the sieve. Let r be the same number that appears in Lemmas B.0.1, B.0.2, B.0.3. Let $\varepsilon = \min\{\epsilon/2, r/4\}$. Let N_{ε} be the ε -covering number of \mathcal{F}_n in the supremum norm. Recall that given $0 < \delta < 1/2$, $(2\delta + 1)/2 < \alpha_1 < 1$ and $\alpha/2 < \tau_1 < 1/2$. Thus, with $M_n = O(n^{\alpha_1})$ and $c_n = n^{\tau_1}$, we have $\log(N_{\varepsilon}) = o(c_n^2)$. Let $\zeta^1, \dots, \zeta^{N_{\varepsilon}} \in \mathcal{F}_n$ be such that for each $\zeta \in \mathcal{F}_n$, there exists j such that $||\zeta - \zeta^j||_{\infty} < \varepsilon$. If $||\zeta - \zeta^j||_1 > \epsilon$, then $||\zeta^j - \zeta_0||_1 > \epsilon/2$. Finally, define

$$\Phi_n = \max_{1 \le j \le N_{\varepsilon}} \Psi_n[\zeta^j, \frac{\epsilon}{2}]$$

Since we verified that there exists *r* such that $\sum_{j=1}^{n} |\zeta_{1j} - \zeta_{0j}| > rn$ holds for every such ζ^{j} in Lemma B.0.5, then

$$E_{P_{\zeta_0}} \Phi_n \le \sum_{j=1}^{N_{\varepsilon}} E_{P_{\zeta_0}} \Psi_n[\zeta^j, \frac{\epsilon}{2}]$$
$$\le C_3 N_{\varepsilon} \exp(-2c_n^2)$$
$$= C_3 \exp(\log(N_{\varepsilon}) - 2c_n^2)$$
$$\le C_3 \exp(-c_n^2)$$

For $\zeta \in \mathcal{U}_{\epsilon_n}^C \cap \mathcal{F}_n$, the type 2 error probability of Ψ_n is no larger than the minimum of the individual type 2 error probabilities of the $\Psi_n[\zeta^j, \epsilon/2]$ tests. Hence we have the results.
APPENDIX C

CHECKING ASSUMPTION 2.2.2 1

We construct our sieve \mathcal{F}_n such that it increases to the space of all continuously differentiable functions with bounded derivatives on $[0, 1]^p$. To this end, we make use of an increasing sequence of constants M_n . To be specific, let $\{M_n\}_{n=1}^{\infty}$ be a sequence of numbers such that $M_n \to \infty$ as $n \to \infty$ and $M_n = O(n^{\alpha})$ for some $\alpha \in (1/2, 1)$. Also, define,

$$\mathcal{F}_n = \left\{ \zeta(\cdot) : \ ||\zeta||_{\infty} < M_n, \ ||\frac{\partial}{\partial t_i}\zeta||_{\infty} < M_n, \ i = 1, \cdots, p \right\}$$
(C.1)

In lieu of (C.1), for the sequence of sieves \mathcal{F}_n , $\Pi_{\zeta}(\mathcal{F}_n^c|z)$ is exponentially small as long as the following two probabilities are exponentially small

$$\Pi_{\zeta} \left\{ \zeta : ||\zeta||_{\infty} > M_n | z \right\} \tag{C.2}$$

$$\Pi_{\zeta} \left\{ \zeta : || \frac{\partial}{\partial t_i} \zeta ||_{\infty} > M_n \bigg| z \right\}, \quad i = 1, \cdots, p$$
(C.3)

This result is stated in the Lemma below.

Lemma C.0.1. Let $M_n = O(n^{\alpha})$, where $\alpha \in (1/2, 1)$ and $2\alpha - h \ge 1$ where h is as defined in Assumption 2.2.5. 1. Then, there exists positive constants D_1, D_2, d_1, d_2 such that,

$$\Pi_{\zeta} \left\{ \zeta : ||\zeta||_{\infty} > M_n |z \right\} \le D_1 \exp(-d_1 n)$$

$$\Pi_{\zeta} \left\{ \zeta : ||\frac{\partial}{\partial t_i} \zeta||_{\infty} > M_n |z \right\} \le D_2 \exp(-d_2 n), \quad i = 1, \cdots, p$$

Proof: For the proof of the lemma, we show the results step by step. First, note that for $x \ge 0$, $0 \le R(x) \le 1, -T_1 \le R'(x) \le 0, -T_2 \le R''(x) \le T_3$ for positive T_1, T_2 , and T_3 . In addition, if we evaluate at x = 0, then R(0) = 1 and $R''(0) = -T_4 < 0$, where $T_4 > 0$. The widely used Squared Exponential covariance function, Matérn covariance function, and Cauchy covariance function satisfy these equalities and inequalities. Please refer to Chapter 4 of Abrahamsen (1997) for further description. Then,

$$\begin{split} \mathbb{E}_{\zeta|\theta,\phi,z}[\zeta(t)] &= \sum_{k=1}^{s} \sum_{j=1}^{s} k_{f}((t,\theta), (\tilde{t}_{j}, \tilde{\theta}_{j})) \left(\frac{1}{\eta_{f}} \tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\theta})\right) z_{k} \\ &= \sum_{k=1}^{s} \sum_{j=1}^{s} \eta_{f} R(t, \tilde{t}_{j}; l_{f_{l}}) R(\theta, \tilde{\theta}_{j}; l_{\theta}) \left(\frac{1}{\eta_{f}} \tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right) z_{k} \\ &= \sum_{k=1}^{s} \sum_{j=1}^{s} R(t^{1}, \tilde{t}_{j}^{1}; l_{f_{1}}) \cdots R(t^{p}, \tilde{t}_{j}^{p}; l_{f_{p}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{j,1}; l_{f_{p+1}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'} l_{f_{p+p'}}) \left(\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right) z_{k} \\ &\leq \sum_{k=1}^{s} \sum_{j=1}^{s} \left(\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right) z_{k} \\ &\leq \left(\max_{k \in \{1, \cdots, s\}} z_{k}\right) \sum_{k=1}^{s} \sum_{j=1}^{s} \left|\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right| \\ &\leq \left(\max_{k \in \{1, \cdots, s\}} z_{k}\right) \bar{C} \\ &= C \end{split}$$
 (C.4)

$$\mathbb{E}_{\zeta|\theta,\phi,z}\left[\frac{\partial}{\partial t^{k}}\zeta(t)\right] = \sum_{k=1}^{s} \sum_{j=1}^{s} R(t^{1}, \tilde{t}_{j}^{1}; l_{f_{1}}) \cdots R'(t^{k}, \tilde{t}_{j}^{k}; l_{f_{k}}) \left(-l_{f_{k}} \frac{t^{k} - \tilde{t}_{j}^{k}}{|t^{k} - \tilde{t}_{j}^{k}|}\right) \cdots R(t^{p}, \tilde{t}_{j}^{p}; l_{f_{p}}) \\ \times R(\theta_{1}, \tilde{\theta}_{j,1}; l_{f_{p+1}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'} l_{f_{p+p'}}) \left(\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right) z_{k} \\ \leq T_{1}\left(\max_{k \in \{1, \cdots, s\}} z_{k}\right) \sum_{k=1}^{s} \sum_{j=1}^{s} l_{f_{k}} \left|\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f}, l_{\delta})\right|$$
(C.5)

Step 1) Let, $\sigma_0^2(\zeta) = \sup_{t \in \Omega \subset \mathbb{R}^p} \operatorname{Var}(\zeta(t)|z, \theta, \phi)$. By applying Lemma.3.5.6 from Choi (2005) at the 1st inequality below, we can bound the probability of the given set from above.

$$\mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\zeta(t)\right| > M_{n}\left|z\right) = \int_{\Theta}\int_{A_{\phi}}\mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\zeta(t)\right| > M_{n}\left|z,\theta,\phi\right)p(\theta,\phi|z)d\phi d\theta$$
$$\leq \int_{\Theta}\int_{A_{\phi}}\exp\left(-\frac{1}{4}\frac{M_{n}^{2}}{\sigma_{0}^{2}(\zeta) + \mathbb{E}^{2}(\zeta(t)|z,\theta,\phi)}\right)p(\theta,\phi|z)d\phi d\theta$$
$$= \int_{\Theta}\int_{A_{\phi}}\exp\left(-\frac{1}{4}\frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f}\sum_{i=1}^{s}\sum_{j=1}^{s}\left|K_{j,i}^{-1}(l_{f},l_{\theta})\right| + C^{2}\right)p(\theta,\phi|z)d\phi d\theta$$

Hence, using Assumption 8. 1,

$$\begin{split} & \mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}} \left|\zeta(t)\right| > M_{n} \left|z\right) \\ &= \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{\infty} \int_{0}^{\infty} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{i=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &= \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{n^{b}} \int_{0}^{n^{b}} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{i=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &+ \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{\infty} \int_{n^{b}}^{\infty} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{i=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &+ \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{n^{b}}^{\infty} \int_{0}^{\infty} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{i=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &- \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{n^{b}}^{n^{b}} \int_{n^{b}}^{\infty} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{i=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &\leq \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{n^{b}} \int_{n^{b}}^{n^{b}} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{j=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &+ \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{n^{b}} \int_{n^{b}}^{n^{b}} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{j=1}^{s} \sum_{j=1}^{s} \left|K_{j,i}^{-1}(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &+ \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{n^{b}} \int_{n^{b}}^{n^{b}} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{j=1}^{s} \left|K_{j,i}-1(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ &+ \int_{\Theta} \int_{A_{\Phi^{-}(\eta_{f},\eta_{\delta})}} \int_{0}^{n^{b}} \int_{0}^{n^{b}} \exp\left(-\frac{1}{4} \frac{M_{n}^{2}}{\eta_{f} + \eta_{\delta} + \eta_{f} \sum_{j=1}^{s} \left|K_{j,i}-1(l_{f}, l_{\theta})\right| + C^{2}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{\delta}|z)p(\phi_{\delta}|z)p(\phi_{\delta}|z)p(\phi_{\delta}$$

Therefore,

$$\mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\zeta(t)\right| > M_{n}\left|z\right) \leq \int_{\Theta}\int_{A_{\phi_{2}}}\int_{0}^{n^{h}}\int_{0}^{n^{h}}\exp\left(-\frac{1}{4}\frac{M_{n}^{2}}{\eta_{f}+\eta_{\delta}+\eta_{f}\sum_{i=1}^{s}\sum_{j=1}^{s}\left|K_{j,i}^{-1}(l_{f},l_{\theta})\right|+C^{2}\right)\right)$$

$$p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta$$

$$+\mathbb{P}(\eta_{f}>n^{h}|z)+\mathbb{P}(\eta_{\delta}>n^{h}|z)-0$$

$$\leq \exp\left(-\frac{1}{4}\frac{n^{2\alpha_{1}}}{n^{h}+n^{h}+n^{h}\bar{C}+C^{2}}\right)+b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n)$$

$$\leq \exp\left(-Cn^{2\alpha_{1}-h}\right)+b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n)$$

$$\leq D_{1}\exp(-d_{1}n)$$

Step 2) Let $\sigma_k^2(\zeta) = \sup_{t \in \Omega \subset \mathbb{R}^p} \operatorname{Var}\left(\frac{\partial}{\partial t^k} \zeta(t) \middle| z, \theta, \phi\right)$. Further let, $l_{f_k}^*, l_{\delta_k}^*$ are the maximizer of $\exp\left(-M_n^2/4(T_4\eta_f l_{f_k}^2 + T_4\eta_\delta l_{\delta_i}^2 + \eta_f l_{f_k}^2 T_1^2 \bar{C} - T_1 C l_{f_k} \bar{C})\right)$ in the bounded support. By applying the same logic as in step 1), we have the upper bound of the probability.

$$\begin{split} & \mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\frac{\partial}{\partial t^{k}}\zeta(t)\right| > M_{n}\Big|z\right) \\ &= \int_{\Theta}\int_{A_{\phi}}\mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\frac{\partial}{\partial t^{k}}\zeta(t)\right| > M_{n}\Big|z,\theta,\phi\right)p(\theta,\phi|z)d\phi d\theta \\ &= \int_{\Theta}\int_{A_{\phi}}\exp\left(-\frac{1}{4}\times\frac{M_{n}^{2}}{\sigma_{k}^{2}(\zeta)+\mathbb{E}_{\zeta|\theta,\phi,z}^{2}\left[\frac{\partial}{\partial t^{k}}\zeta(t)\right]}\right)p(\theta,\phi|z)d\phi d\theta \\ &\leq \int_{\Theta}\int_{0}^{n^{h}}\int_{0}^{n^{h}} \\ & \exp\left(-\frac{1}{4}\times\frac{M_{n}^{2}}{T_{4}\eta_{f}l_{f}^{2}+T_{4}\eta_{\delta}l_{\delta}^{2}+\eta_{f}l_{f}^{2}T_{1}^{2}\bar{C}+T_{1}\left(\max_{k\in\{1,\cdots,s\}}z_{k}\right)\sum_{k=1}^{s}\sum_{j=1}^{s}l_{f_{k}}\left|\tilde{\Sigma}_{22(j,k)}^{-1}(l_{f},l_{\delta})\right|\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\phi_{2}|z)p(\theta)d\phi d\theta \\ & +\mathbb{P}(\eta_{f}>n^{h}|z)+\mathbb{P}(\eta_{\delta}>n^{h}|z) \end{split}$$

Hence,

$$\begin{split} & \mathbb{P}\left(\sup_{t\in\Omega\subset\mathbb{R}^{p}}\left|\frac{\partial}{\partial t^{k}}\zeta(t)\right| > M_{n}\bigg|z\right) \\ & \leq \int_{\Theta}\int_{0}^{n^{h}}\int_{0}^{n^{h}}\exp\left(-\frac{1}{4}\times\frac{M_{n}^{2}}{T_{4}\eta_{f}l_{f_{k}}^{*2}+T_{4}\eta_{\delta}l_{\delta_{k}}^{*2}+\eta_{f}l_{f_{k}}^{*2}T_{1}^{2}\bar{C}+T_{1}Cl_{f_{k}}^{*}\bar{C}}\right) \\ & p(\eta_{f}|z)p(\eta_{\delta}|z)p(\theta)d\phi d\theta \\ & +b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n) \\ & \leq \exp\left(-\frac{1}{4}\times\frac{n^{2\alpha_{1}}}{T_{4}n^{h}l_{f_{k}}^{*2}+T_{4}n^{h}l_{\delta_{k}}^{*2}+n^{h}l_{f_{k}}^{*2}T_{1}^{2}\bar{C}+T_{1}Cl_{f_{k}}^{*}\bar{C}}\right)+b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n) \\ & \leq \exp\left(-\frac{n^{2\alpha_{1}-h}}{4(T_{4}l_{f_{k}}^{*2}+T_{4}l_{\delta_{k}}^{*2}+l_{f_{k}}^{*2}T_{1}^{2}\bar{C})}\right)+b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n) \\ & =\exp\left(-Cn^{2\alpha_{1}-h}\right)+b_{1}\exp(-b_{2}n)+b_{3}\exp(-b_{4}n) \\ & \leq D_{2}\exp\left(-d_{2}n\right) \end{split}$$

Note that for the second inequality, we used the assumption that the length scale parameters have bounded support, which leads to specific values, l_f^* and l_{δ}^* , that maximize the term in the exponential. In addition the bound of σ_0^2 and σ_i^2 used in the proof of step 1) and step 2) are derived below.

Bounds used for the proof of step 1) and step 2) above Remember from Assumption 2.2.5. 2, we have $\sum_{i=1}^{s} \sum_{j=1}^{s} \left| \tilde{\Sigma}_{22(j,i)}^{-1}(l_f, l_\delta) \right| \leq \bar{C}$ Also, $\frac{1}{\eta_f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_f, l_\theta)$ is the (j,i)-th element of $[K_f(T_z(\tilde{\theta}), T_z(\tilde{\theta}))]^{-1}$, and it depends on l_f, l_δ . Under the assumed covariance structure, the covariance kernel of ζ conditioned on z, θ, ϕ is as follows. Let $l_f = (l_{f_t}, l_\theta)$, where $l_{f_t} = (l_{f_1}, \cdots, l_{f_p})$ and $l_\theta = (l_{f_{p+1}}, \cdots, l_{f_p+p'})$, and $l_\delta = (l_{\delta_1}, \cdots, l_{\delta_p})$

$$\begin{aligned} k_{\zeta|z,\theta,\phi}(t,t') &= k_f((t,\theta),(t',\theta')) + k_{\delta}(t,t') \\ &- \sum_{i=1}^s \sum_{j=1}^s \eta_f R(t,\tilde{t}_i;l_{f_t}) R(\theta,\tilde{\theta}_i;l_{\theta}) \left(\frac{1}{\eta_f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_f,l_{\delta})\right) \eta_f R(t',\tilde{t}_j;l_{f_t}) R(\theta,\tilde{\theta}_j;l_{\theta}) \\ &- \sum_{i=1}^s \sum_{j=1}^s \eta_f \tilde{\Sigma}_{22(j,i)}^{-1}(l_f,l_{\delta}) R(t,\tilde{t}_i;l_{f_t}) R(\theta,\tilde{\theta}_i;l_{\theta}) R(t',\tilde{t}_j;l_{f_t}) R(\theta,\tilde{\theta}_j;l_{\theta}) \end{aligned}$$

Hence,

$$\begin{aligned} k_{\zeta|z,\theta,\phi}(t,t') &= \eta_f R(t,t';l_{f_t}) + \eta_{\delta} R(t,\tilde{t}';l_{\delta}) \\ &= \eta_f R(t^1,t'^1;l_{f_1}) R(t^2,t'^2;l_{f_2}) \cdots R(t^p,t'^p;l_{f_p}) \\ &+ \eta_{\delta} R(t^1,t'^1;l_{\delta_1}) R(l_{t^2,t'^2;\delta_2}) \cdots R(t^p,t'^p;l_{\delta_p}) \\ &- \sum_{i=1}^s \sum_{j=1}^s \eta_f \tilde{\Sigma}_{22(j,i)}^{-1}(l_f,l_{\delta}) R(t^1,\tilde{t}_i^1;l_{f_1}) R(t^2,\tilde{t}_i^2;l_{f_2}) \cdots R(t^p,\tilde{t}_i^p;l_{f_p}) \\ &\times R(\theta_1,\tilde{\theta}_{i,1};l_{f_{p+1}}) R(\theta_2,\tilde{\theta}_{i,2};l_{f_{p+1}}) \cdots R(\theta_{p'},\tilde{\theta}_{i,p'};l_{f_{p+p'}}) \\ &\times R(\theta_1,\tilde{\theta}_{j,1};l_{f_p+1}) R(\theta_2,\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p'}}) \\ &\times R(\theta_1,\tilde{\theta}_{j,1};l_{f_{p+1}}) R(\theta_2,\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p'}}) \end{aligned}$$

By letting t^* the maximizer of $Var(\zeta(t)|z, \theta, \phi)$, we can bound the following term.

$$\begin{split} \sigma_{0}^{2}(\zeta) &= \sup_{t \in \Omega \subset \mathbb{R}^{p}} \operatorname{Var}(\zeta(t)|z, \theta, \phi) \\ &= \eta_{f} R(0)^{p} + \eta_{\delta} R(0)^{p} \\ &- \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f}, l_{\delta}) R(t^{*1}, \tilde{t}_{1}^{i}; l_{f_{1}}) R(t^{*2}, \tilde{t}_{i}^{2}; l_{f_{2}}) \cdots R(t^{*p}, \tilde{t}_{i}^{p}; l_{f_{p}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{i,1}; l_{f_{p+1}}) R(\theta^{2}, \tilde{\theta}_{i}^{2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{i,p'}; l_{f_{p+p'}}) \\ &\times R(t^{*1}, \tilde{t}_{j}^{1}; l_{f_{1}}) R(t^{*2}, \tilde{t}_{j}^{2}; l_{f_{2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'}; l_{f_{p+p'}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{i,1}; l_{f_{p+1}}) R(\theta_{2}, \tilde{\theta}_{j,2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'}; l_{f_{p+p'}}) \\ &\leq \eta_{f} R(0)^{p} + \eta_{\delta} R(0)^{p} \\ &+ \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \left| \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f}, l_{\delta}) \right| R(t^{*1}, \tilde{t}_{i}^{1}; l_{f_{1}}) R(t^{*2}, \tilde{t}_{i}^{2}; l_{f_{2}}) \cdots R(t^{*p} \tilde{t}_{i}^{p}; l_{f_{p}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{i,1}; l_{f_{p+1}}) R(\theta_{2}, \tilde{\theta}_{i,2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{i,p'}; l_{f_{p+p'}}) \\ &\times R(t^{*1}, \tilde{t}_{j}^{1}; l_{f_{1}}) R(t^{*2}, \tilde{t}_{j}^{2}; l_{f_{2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{i,p'}; l_{f_{p+p'}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{i,1}; l_{f_{p+1}}) R(\theta_{2}, \tilde{\theta}_{i,2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'}; l_{f_{p+p'}}) \end{split}$$

Since R(0) = 1 and $R(x) \le 1$ for widely used covariance functions such as Squared Exponential covariance, Matérn covariance, and Cauchy covariance and using Assumption 8. 2,

$$\sigma_0^2(\zeta) \le \eta_f + \eta_\delta + \eta_f \bar{C}$$

For bounding $\sigma_i^2(\zeta) = \sup_{t \in \Omega \subset \mathbb{R}^p} \operatorname{Var}\left(\frac{\partial}{\partial t_i}\zeta(t) | z, \theta, \phi\right)$, we are going to use the fact that the process constructed by taking the partial derivative of $\zeta | z, \theta, \phi$ with respect to the k^{th} component is again a GP with continuous sample paths and covariance function with $\frac{\partial^2}{\partial t^k \partial t'^k} k_{\zeta | z, \theta, \phi}(t, t')$. For proof of this fact, please refer to Ghosal and Roy (2006).

$$\begin{split} \frac{\partial}{\partial t^{k}} \frac{\partial}{\partial t^{k}} k_{\zeta|z,\theta,\phi}(t,t') \\ &= \frac{\partial}{\partial t^{k}} \frac{\partial}{\partial t^{k}} \Big[\eta_{f} R(t^{1},t'^{1};l_{f_{1}}) R(t^{2},t'^{2};l_{f_{2}}) \cdots R(t^{p},t'^{p};l_{f_{p}}) \\ &+ \eta_{\delta} R(t^{1},t'^{1};l_{\delta_{1}}) R(t^{2},t'^{2};l_{\delta_{1}}) \cdots R(t^{p},t'^{p};l_{\delta_{p}}) \\ &- \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f},l_{\delta}) R(t^{1},\tilde{l}_{i}^{1};l_{f_{1}}) R(t^{2},\tilde{l}_{i}^{2};l_{f_{2}}) \cdots R(t^{p},\tilde{l}_{i}^{p};l_{f_{p}}) \\ &\times R(\theta_{1},\tilde{\theta}_{i,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{i,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{i,p'};l_{f_{p+p'}}) \\ &\times R(\theta_{1},\tilde{\theta}_{i,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p'}}) \\ &\times R(\theta_{1},\tilde{\theta}_{j,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p'}}) \Big] \\ &= \frac{\partial}{\partial t^{k}} \Big[\eta_{f} R(t^{1},t'^{1};l_{f_{1}}) R(t^{2},t'^{2};l_{f_{2}}) \cdots R'(t^{k},t'^{k};l_{f_{k}}) \left(-l_{f_{k}} \frac{t^{k}-t'^{k}}{|t^{k}-t'^{k}|} \right) \cdots R(t^{p},t'^{p};l_{f_{p}}) \\ &+ \eta_{\delta} R(t^{1},t'^{1};l_{\delta_{1}}) R(t^{2},t'^{2};l_{\delta_{2}}) \cdots R'(t^{k},t'^{k};l_{\delta_{k}}) \left(-l_{\delta_{k}} \frac{t^{k}-t'^{k}}{|t^{k}-t'^{k}|} \right) \cdots R(t^{p},t'^{p};l_{\delta_{p}}) \\ &- \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f},l_{\delta}) R(t^{1},\tilde{t}_{i}^{1};l_{f_{1}}) R(t^{2},\tilde{t}_{i}^{2};l_{f_{2}}) \cdots R(t^{p},\tilde{\theta}_{i,p'};l_{f_{p}}) \\ &\times R(\theta_{1},\tilde{\theta}_{i,1};l_{f_{p}}) R(\theta_{2},\tilde{\theta}_{i,2};l_{f_{2}}) \cdots R(\theta_{p'},\tilde{\theta}_{i,p'};l_{f_{p'}}) \\ &\times R(t'^{1},\tilde{t}_{j}^{1};l_{f_{1}}) R(t'^{2},\tilde{t}_{j}^{2};l_{f_{2}}) \cdots R(t'^{k},\tilde{t}_{i}^{k};l_{f_{k}}) \left(-l_{f_{k}} \frac{t'^{k}-\tilde{t}_{j}^{k}}{|t^{k}-\tilde{t}_{j}|} \right) \cdots R(t'^{p},\tilde{t}_{p}^{p};l_{f_{p}}) \\ &\times R(\theta_{1},\tilde{\theta}_{i,1};l_{f_{p}}) R(\theta_{2},\tilde{\theta}_{i,2};l_{f_{p}}) \cdots R(\theta_{p'},\tilde{\theta}_{i,p'};l_{f_{p'}}) \\ &\times R(\theta_{1},\tilde{\theta}_{j,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p}})] \end{split}$$

Hence,

$$\begin{split} \frac{\partial}{\partial t^{k}} \frac{\partial}{\partial t'^{k}} k_{\zeta|z,\theta,\phi}(t,t') \\ &= -\eta_{f} l_{f_{k}}^{2} R(t^{1},t'^{1};l_{f_{1}}) R(t^{2},t'^{2};l_{f_{2}}) \cdots R''(t^{k},t'^{k};l_{f_{k}}) \cdots R(t^{p},t'^{p};l_{f_{p}}) \\ &- \eta_{\delta} l_{\delta_{k}}^{2} R(t^{1},t'^{1};l_{\delta_{1}}) R(t^{2},t'^{2};l_{\delta_{2}}) \cdots R''(t^{k},t'^{k};l_{\delta_{k}}) \cdots R(t^{p},t'^{p};l_{\delta_{p}}) \\ &- \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f},l_{\delta}) R(t^{1},\tilde{t}_{i}^{1};l_{f_{1}}) \cdots R'(t^{k},\tilde{t}_{i}^{k};l_{f_{k}}) \left(-l_{f_{k}} \frac{t^{k}-\tilde{t}_{i}^{k}}{|t^{k}-\tilde{t}_{i}^{k}|} \right) \cdots R(t^{p},\tilde{t}_{i}^{p};l_{f_{p}}) \\ &\times R(\theta_{1},\tilde{\theta}_{i,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{i,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{i,p'};l_{f_{p+p'}}) \\ &\times R(t'^{1},\tilde{t}_{j}^{1};l_{f_{1}}) R(t'^{2},\tilde{t}_{j}^{2};l_{f_{1}}) \cdots R'(t'^{k},\tilde{t}_{i}^{k};l_{f_{k}}) \left(-l_{f_{k}} \frac{t'^{k}-\tilde{t}_{i}^{k}}{|t'^{k}-\tilde{t}_{j}^{k}|} \right) \cdots R(t'^{p},\tilde{t}_{j}^{p};l_{f_{p}}) \\ &\times R(\theta_{1},\tilde{\theta}_{j,1};l_{f_{p+1}}) R(\theta_{2},\tilde{\theta}_{j,2};l_{f_{p+2}}) \cdots R(\theta_{p'},\tilde{\theta}_{j,p'};l_{f_{p+p'}}) \end{split}$$

Assume that t^{**} is the maximizer of $\frac{\partial^2}{\partial (t^k)^2} \operatorname{Var}\left(\zeta(t) | z, \theta, \phi\right)$. Then, we can simplify the bound by using the following facts coming from the covariance assumption for R.

$$\begin{split} \sigma_{k}^{2}(\zeta) &= \sup_{t \in \Omega \subset \mathbb{R}^{p}} \frac{\partial^{2}}{\partial (t^{k})^{2}} \operatorname{Var} \left(\zeta(t) \middle| z, \theta, \phi \right) \\ &= -\eta_{f} l_{f_{k}}^{2} R(0)^{p-1} R''(0) - \eta_{\delta} l_{\delta_{k}}^{2} R(0)^{p-1} R''(0) \\ &- \sum_{i=1}^{s} \sum_{j=1}^{s} \eta_{f} \tilde{\Sigma}_{22(j,i)}^{-1}(l_{f}, l_{\delta}) \\ &\times R(t^{**1}, \tilde{t}_{i}^{1}; l_{f_{1}}) \cdots R'(t^{**k}, \tilde{t}_{i}^{k}; l_{f_{k}}) \left(l_{f_{k}} \frac{t^{**k} - \tilde{t}_{i}^{k}}{|t^{**k} - \tilde{t}_{i}^{k}|} \right) \cdots R(t^{**p}, \tilde{t}_{i}^{p}; l_{f_{p}}) \\ &\times R(\theta_{1}, \tilde{\theta}_{i,1}; l_{f_{p+1}}) R(\theta_{2}, \tilde{\theta}_{i,2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{i,p'}; l_{f_{p+p'}}) \\ &\times R(t^{**1}, \tilde{t}_{j}^{1}; l_{f_{1}}) \cdots R'(t^{**k}, \tilde{t}_{j}^{k}; l_{f_{k}}) \left(l_{f_{k}} \frac{t^{**k} - \tilde{t}_{i}^{k}}{|t^{**k} - \tilde{t}_{j}^{k}|} \right) \cdots R(t^{**p}, \tilde{t}_{j}^{p}; l_{f_{p}}) \\ &\times R(\theta_{1} - \tilde{\theta}_{j,1}; l_{f_{p+1}}) R(\theta_{2}, \tilde{\theta}_{j,2}; l_{f_{p+2}}) \cdots R(\theta_{p'}, \tilde{\theta}_{j,p'}; l_{f_{p+p'}}) \\ &\leq T_{4} \eta_{f} l_{f_{k}}^{2} + T_{4} \eta_{\delta} l_{\delta_{k}}^{2} + \eta_{f} l_{f_{k}}^{2} T_{1}^{2} \bar{C} \end{split}$$

APPENDIX D

CHECKING ASSUMPTION 2.2.2 2

First, recall two classical spaces of finite smoothness: Hölder and Sobolev. Hölder space $C^{\alpha}(X)$, where $\alpha > 0$, is the space of all functions whose partial derivatives of order up to $\lceil \alpha \rceil - 1$ exist and are uniformly bounded, where $\lceil \alpha \rceil$ represents the largest integer that is smaller than α . Sobolev space $H^{\alpha}(X)$ is the set of functions that are restriction to X of functions $f_0 : \mathbb{R}^p \to \mathbb{R}$ with Fourier transform $\hat{f}_0(\lambda) = (2\pi)^{-p} \int \exp(i\lambda^T t) f_0(t) dt$ satisfying

$$\int (1+\|\lambda\|^2)^{\alpha} |\hat{f}_0|^2 d\lambda < \infty.$$

van der Vaart and van Zanten (2011) showed behavior of concentration function in two different classes of GP. First, consider Matérn class of GP prior, which corresponds to the mean zero GP $W = (W_t : t \in [0, 1]^p)$ with covariance function

$$\mathbb{E}[W_s W_t] = \int_{\mathbb{R}^p} \exp(i\lambda^T (s-t)) m(\lambda) d\lambda$$
(D.1)

where spectral densities $m : \mathbb{R}^p \to \mathbb{R}$ is given by, for $\alpha > 0$,

$$m(\lambda) = \frac{1}{(a + \|\lambda\|^2)^{\alpha + p/2}}$$
(D.2)

Combining the results of Lemma 3 and Lemma 4 ofvan der Vaart and van Zanten (2011), we have the following results,

Lemma D.0.1. For Matérn class of GP, if $\zeta_0 \in C^{\beta}[0,1]^p \cap H^{\beta}[0,1]^p$ for $0 < \beta \leq \alpha$, the concentration function satisfies,

$$\phi_{\zeta_0}^M(\epsilon_n) \lesssim \epsilon_n^{\frac{-(2\alpha+p-2\beta)}{\beta}} + \epsilon_n^{\frac{-p}{\alpha}}$$
(D.3)

For the second case, consider the Square Exponential class of GP prior. The squared exponential class corresponds to the mean zero GP with covariance function.

$$\mathbb{E}[W_s W_t] = \exp(-\|s - t\|^2), \qquad s, t \in [0, 1]^p$$
(D.4)

where spectral densities $m : \mathbb{R}^p \to \mathbb{R}$ is given by,

$$m(\lambda) = \frac{1}{2^p \pi^{p/2}} \exp\left(\frac{-\|\lambda\|^2}{4}\right)$$
(D.5)

Combining the results of Lemma 6 and Lemma 7 ofvan der Vaart and van Zanten (2011), we have the following results,

Lemma D.0.2. For Squared Exponential class of GP, if $\zeta_0 \in H^{\beta}[0,1]^p$ for $d/2 < \beta$, then for a constant C that depends only on ζ_0 , the concentration function satisfies,

$$\phi_{\zeta_0}^{SE}(\epsilon_n) \leq \exp(C\epsilon_n^{-2/(\beta - p/2)}) + (-\log\epsilon_n)^{1+p}$$
(D.6)

Note both the right hand sides of $\phi_{\zeta_0}^M(\epsilon_n)$ and $\phi_{\zeta_0}^{SE}(\epsilon_n)$ tend to infinity as *n* goes to infinity. However, since slowly decreasing sequence ϵ_n satisfy $n\epsilon_n^2 \to \infty$ as *n* goes to infinity, for large enough *n* we have $\phi_{\zeta_0}^M(\epsilon_n) \leq n\epsilon_n^2$ and $\phi_{\zeta_0}^{SE}(\epsilon_n) \leq n\epsilon_n^2$.

CHAPTER 3

CALIBRATED COMPUTER MODEL USING VARIATIONAL BAYES (VARIATIONAL INFERENCE)

3.1 Variational Bayes formulation

This section introduces a general Variational Bayesian (VB) approach that we are going to use. Let a set of parameters of interest Ψ have a prior distribution $p(\Psi)$. Given a data set \mathcal{D} , the goal of the Bayesian method is to find the posterior distribution. Using Bayes' theorem, that is to say, find the following distribution.

$$\pi(\boldsymbol{\Psi}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\Psi})p(\boldsymbol{\Psi})}{\int_{\boldsymbol{\Psi}} p(\mathcal{D}|\boldsymbol{\Psi})p(\boldsymbol{\Psi})d\boldsymbol{\Psi}}$$

However, the core problem of posterior calculation comes from the denominator part of the posterior distribution, *evidence*. Modern Bayesian statisticians have been focused on approximating this difficult-to-calculate density Blei et al. (2017). There are diverse approaches to approximating the posterior distributions. To name a few, there are Markov Chain Monte Carlo (MCMC), Hamiltonian Monte Carlo (HMC), and Approximate Bayesian Computation (ABC). An alternative machine learning method has been getting popular recently, *Variational Bayes* (VB) or *Variational Inference* (VI). The advantage of VB over other approaches is clear; it is faster and easier to scale to large data sets. This fact is also very appealing to the computer simulation setup where we want to build a GP surrogate for a large simulation data set.

To apply VB, we first need a divergence criteria between two probability measures. The most widely used divergence criterion is the *Kullback-Leibler (KL) divergence* or *KL distance*.

Definition 1. (*Kullback-Leibler* (*KL*) *divergence*) For two probability measures P and Q on a measurable space X, let P is absolutely continuous with respect to Q. Further if μ is any measure on X for which densities p and q with $P(dx) = p(x)\mu(dx)$ and $Q(dx) = q(x)\mu(dx)$ exist. Then the Kullback-Leibler divergence is defined as

$$KL(P||Q) = \int_{\mathcal{X}} \log \frac{P(dx)}{Q(dx)} P(dx) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \mu(dx) = \mathbb{E}_p \left[\log \frac{p}{q} \right]$$

Definition 1 shows the standard definition of KL divergence. However, since we are working on infinite-dimensional function spaces, there is no sensible infinite-dimensional Lebesgue measure. Therefore, we need a more general definition of KL divergence to work with. Kolmogorov formulated a general definition of KL divergence as a supremum of KL divergence over all finite measurable partitions on the space of the index set Kolmogorov et al. (1992). This formulation has been adopted in the functional approach of variational inference in other papers such as de G. Matthews et al. (2016) and Sun et al. (2019). Some early works related to KL divergence between stochastic processes include Csató and Opper (2002), Csató (2002). Seeger (2003a), and Seeger (2003b).

Definition 2. (*Kullback-Leibler (KL) divergence for Stochastic Processes*) For two stochastic processes P and Q, the Kullback-Leibler divergence is defined as

$$KL(P||Q) = \sup_{X \in \mathcal{X}} KL(P_X||Q_X)$$

where X denotes measurement sets and P_X, Q_X are the marginal distribution of function values at X.

Note that this definition of KL divergence depends on measurement set *X*. Sun et al. (2019) proposed sampling-based measurement sets where the measurements set are sampled from both training inputs and random points from the domain. In this paper, however, we are going to assume the measurement set contains all the training data points for theoretical analysis. Then, for the same class of stochastic process, the logarithm term cancels out for KL divergence because we are considering a finite measurement set. This simplifies our proof of the consistency of variational posterior.

The first step to building a VB approach is to predefine a family of distributions we are going to optimize. This family is called *variational family*, Q. The common practice is to consider a family of distributions where a product form for each dimension of parameters is assumed, *mean field variational family* Q_{MF} . After setting up the variational family, the next step is to find the closest distribution to the true posterior in terms of KL distance over Q_{MF} . This can be mathematically

written as follows.

$$Q^* = \underset{Q \in Q_{MF}}{\arg\min} \operatorname{KL}(Q || \Pi(\cdot | \mathcal{D}))$$

The problem with this optimization is that we cannot calculate the evidence, $p(\mathcal{D})$. If we take a closer look, we can check the dependence of KL distance to log evidence.

$$\begin{aligned} \mathsf{K}L(Q||\Pi(\cdot|\mathcal{D})) &= \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log p(\theta|\mathcal{D})] \\ &= \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log p(\theta, \mathcal{D})] + \log p(\mathcal{D}) \end{aligned}$$

Hence we optimize an equivalent objective function, the *evidence lower bound*(ELBO), $\mathcal{L}(q)$ Jordan et al. (1999). The name comes from the fact that it lower bounds the log evidence.

$$\mathcal{L}(q) = \mathbb{E}_q[\log p(\mathcal{D}, \theta)] - \mathbb{E}_q[\log q(\theta)]$$
$$\leq \log p(\mathcal{D})$$

3.1.1 Black Box Variational Inference (BBVI)

VB method has been widely adopted for approximating the posterior distributions. But the problem of VB is that it requires significant model-specific analysis in order to develop the VB algorithm. This boils down to calculating ELBO, $\mathbb{E}_q[\log p(\mathcal{D}, \theta)] - \mathbb{E}_q[\log q(\theta)]$. This expectation often does not have a closed-form expression. Ranganath et al. (2014) proposed Black Box Variational Inference (BBVI) method. This is basically a stochastic optimization method by using an unbiased noisy gradient to maximize ELBO. Consider a set of variational parameters λ , then

$$\nabla_{\lambda} \mathcal{L}(q) = \mathbb{E}_q [\nabla_{\lambda} \log q(\Psi_s | \lambda) \times (\log p(\Psi_s, \mathcal{D}) - \log q(\Psi_s | \lambda)]$$
(3.1)

where $\Psi_s \sim q(\Psi|\lambda)$ The unbiased Monte Carlo estimate of this gradient is

$$\widehat{\nabla_{\lambda} \mathcal{L}(q)} = \frac{1}{S} \sum_{s=1}^{S} \left[\nabla_{\lambda} \log q(\Psi_s | \lambda) \times (\log p(\Psi_s, \mathcal{D}) - \log q(\Psi_s | \lambda)) \right]$$
(3.2)

Using this noisy gradient, BBVI update the variational parameters $\lambda^{t+1} \leftarrow \lambda^t + \rho^t \widehat{\nabla_{\lambda} \mathcal{L}(q)}$, until ELBO converges. If the sequence of step sizes $\{\rho^t\}_{t=1}^{\infty}$ satisfies following Robbins-Monro conditions, then the algorithm guaranteed to converge to a local maximum.

$$\sum_{i=1}^{\infty} \rho^t = \infty, \qquad \sum_{i=1}^{\infty} (\rho^t)^2 < \infty$$
(3.3)

3.1.2 Reparametrization Trick

There is one issue implementing the above algorithm. That is samples of parameters are not differentiable with respect to variational parameters λ . Kingma and Welling (2013) and Rezende et al. (2014) approached this problem by reparametrizing samples in a clever way, such that the stochasticity is independent of the parameters. By separating out the stochasticity, samples become deterministically depend on the parameters of the distribution, hence we could differentiate the samples with respect to the variational parameters. For example, consider a sample z from normal distribution with mean μ and variance σ^2 . Equivalently, this could be written as $z = \sigma \times \epsilon + \mu$, where $\epsilon \sim N(0, 1)$. This reparametrization trick allows us to sample from a standard normal distribution and then scaled by the mean and standard deviation to get a sample from $N(\mu, \sigma)$. In this way, we could keep back propagate the mean and standard deviation through the samples.

3.2 Variational Bayes Algorithm for Discrepancy Model

Coming back to the discrepancy model setup, we intend to find the variational approximation to the posterior $p(\theta, \phi, \sigma | d)$. First, let $\phi = (\phi_1, \phi_2)$ where ϕ_1 corresponds to real-valued parameters and ϕ_2 corresponds to positive-valued parameters. We posit a variational family of the form,

$$\tilde{Q} = Q_{\zeta|\theta,\phi} \times Q_{\theta} \times Q_{\phi_1} \times Q_{\phi_2} \times Q_{\sigma}$$

Here,

Assumption 3.2.1. We assume the following structure on the variational family

$$\begin{aligned} \mathcal{Q}_{\zeta|\theta,\phi} &= \{\zeta|\theta,\phi \sim GP(\boldsymbol{m}_{\zeta|\theta,\phi}^{q},\boldsymbol{k}_{\zeta|\theta,\phi}^{q})\},\\ \mathcal{Q}_{\theta} &= \{\theta \sim \mathcal{N}(\boldsymbol{m}_{\theta},\boldsymbol{S}_{\theta})\}, \qquad \mathcal{Q}_{\phi_{1}} = \{\phi_{1} \sim \mathcal{N}(\boldsymbol{m}_{\phi_{1}},\boldsymbol{S}_{\phi_{2}})\}\\ \mathcal{Q}_{\phi_{2}} &= \{\tilde{\phi}_{2} \sim \mathcal{N}(\boldsymbol{m}_{\phi_{2}},\boldsymbol{S}_{\phi_{2}})\}, \qquad \mathcal{Q}_{\sigma} = \{\tilde{\sigma} \sim \mathcal{N}(\boldsymbol{m}_{\sigma},\boldsymbol{s}_{\sigma})\}, \end{aligned}$$

where $\tilde{\phi}_2 = \log \phi_2$ is the log operator applied elementwise and $\tilde{\sigma} = \log \sigma$. Also, under the mean field variational family, one assumes that the matrices S_{θ} , S_{ϕ_1} and S_{ϕ_2} are diagonal matrices.

Note that the variational distribution of ζ depends on the calibration parameter θ and hyperparameters ϕ . Hence this is not the mean field family. Now if we assume using the same covariance function for variational distribution of ζ as prior of ζ , such as squared exponential covariance or Matérn covariance function, then the KL divergence between variational distribution for ζ and prior for ζ become zero as we discussed. Now we can simplify the objective KL function.

$$KL(q(\zeta, \theta, \phi, \sigma)||p(\zeta, \theta, \phi, \sigma|d)) = \int \log\left(\frac{q(\theta, \phi, \sigma)}{p(\theta, \phi, \sigma|d)}\right) q(\theta, \phi, \sigma) d\theta d\phi d\sigma$$
(3.4)

Therefore the optimal variational distribution q^* is given by

$$q^* = \underset{q \in Q}{\arg\min} \operatorname{KL}(q(\theta, \phi, \sigma | \lambda) || p(\theta, \phi, \sigma | d)).$$
(3.5)

where $\lambda = (\boldsymbol{m}_{\theta}, \boldsymbol{S}_{\theta}, \boldsymbol{m}_{\phi_1}, \boldsymbol{S}_{\phi_1}, \boldsymbol{m}_{\phi_2}, \boldsymbol{S}_{\phi_2}, \boldsymbol{m}_{\sigma}, \boldsymbol{S}_{\sigma})$ denotes the set of all variational parameters taken together, and $\boldsymbol{Q} = \boldsymbol{Q}_{\theta} \times \boldsymbol{Q}_{\phi_1} \times \boldsymbol{Q}_{\phi_2} \times \boldsymbol{Q}_{\sigma}$. The optimal variational distribution q^* is obtained in practice by maximizing an ELBO.

$$\mathcal{L}(q) = \mathbb{E}_{q(\theta, \phi, \sigma \mid \lambda)} \Big[\log p(\boldsymbol{d} \mid \boldsymbol{\theta}, \phi, \sigma) + \log p(\boldsymbol{\theta}, \phi, \sigma) - \log q(\boldsymbol{\theta}, \phi, \sigma \mid \lambda) \Big]$$
(3.6)

We can adopt Black Box Variational Inference (BBVI) algorithm to solve this optimization problem. That is, find the unbiased Monte Carlo estimate of the derivative of ELBO.

$$\nabla_{\lambda} \mathcal{L}(q) = \mathbb{E}_q[\nabla_{\lambda} \log q(\theta, \phi, \sigma | \lambda) (\log p(d | \theta, \phi, \sigma) + \log p(\theta, \phi, \sigma) - \log q(\theta, \phi, \sigma | \lambda))] \quad (3.7)$$

$$\approx \frac{1}{S} \sum_{s=1}^{S} \nabla_{\lambda} \log q(\boldsymbol{\theta}^{s}, \boldsymbol{\phi}^{s}, \sigma^{s} | \boldsymbol{\lambda})$$
(3.8)

$$\times \left(\log p(\boldsymbol{d}|\boldsymbol{\theta}^{s}, \boldsymbol{\phi}^{s}, \sigma^{s}) + \log p(\boldsymbol{\theta}^{s}, \boldsymbol{\phi}^{s}, \sigma^{s}) - \log q(\boldsymbol{\theta}^{s}, \boldsymbol{\phi}^{s}, \sigma^{s}|\boldsymbol{\lambda})\right)$$
(3.9)

$$= \hat{\nabla}_{\lambda} \mathcal{L}(q) \tag{3.10}$$

Then use the stochastic gradient descent (SGD) algorithm. The algorithm is summarized in the next algorithm 1 table. Let $g_1(z; \mu, \sigma) = z \odot \sigma + \mu$ and $g_2(z; \mu, \sigma) = \exp(g_1(z; \mu, \sigma))$.

Algorithm 1 BBVI algorithm in Computer Calibration Model

Input: Model, tolerance τ , batch size L, learning rate sequence $\{\rho^t\}_{t=1}^{\infty}$ **Output:** $\lambda^* = \lambda^t$ 1: Randomly initialize λ^0 2: Set $t \leftarrow 0, \Delta \leftarrow 0$ 3: while increament of ELBO > τ do 4: $t \leftarrow t + 1$ $(\boldsymbol{\theta}^1, \cdots, \boldsymbol{\theta}^L) = g_1(\boldsymbol{w}^1, \cdots, \boldsymbol{w}^L; \boldsymbol{m}_{\theta}, \boldsymbol{S}_{\theta}), \text{ where } (\boldsymbol{w}^1, \cdots, \boldsymbol{w}^L) \stackrel{i.i.d.}{\sim} N(\boldsymbol{0}, I_{p' \times p'})$ 5: $(\boldsymbol{\phi}_{1}^{1}, \cdots, \boldsymbol{\phi}_{1}^{L}) = g_{1}(\boldsymbol{w}^{'1}, \cdots, \boldsymbol{w}^{'L}; \boldsymbol{m}_{\phi_{1}}, \boldsymbol{S}_{\phi_{2}}), \text{ where } (\boldsymbol{w}^{'1}, \cdots, \boldsymbol{w}^{'L}) \stackrel{i.i.d.}{\sim} N(\boldsymbol{0}, I_{2\times 2})$ $(\boldsymbol{\phi}_{2}^{1}, \cdots, \boldsymbol{\phi}_{2}^{L}) = g_{2}(\boldsymbol{w}^{''1}, \cdots, \boldsymbol{w}^{''L}; \boldsymbol{m}_{\phi_{2}}, \boldsymbol{S}_{\phi_{2}}),$ 6: 7: where $(\boldsymbol{w}^{''1}, \cdots, \boldsymbol{w}^{''L}) \stackrel{i.i.d.}{\sim} N(\boldsymbol{0}, I_{2p+p'\times 2p+p'})$ 8: 9: 10: 11: $\lambda^t \leftarrow \lambda^{t-1} + \rho^t \hat{\nabla}_\lambda \mathcal{L}(q)$ 12: 13: end while

3.3 Posterior Contraction

Just as MCMC case, we shall assume that $\sigma^2 = 1$. The results can be extended to unknown σ^2 with minor modifications as well. To statistically validate a Bayesian approach, one needs to establish posterior consistency which states that the posterior probability of an arbitrary neighborhood around the true parameter goes to 1 as the sample size *n* goes to infinity. In this section, we first establish the posterior consistency of the true posterior in (2.9). Next, we establish the posterior consistency of the variational posterior in (3.5). We shall use the notation \mathbb{E}_0 and \mathbb{P}_0 to denote the expectation and probability of y_i with respect to the true parameters ζ_0 . In this direction, define

$$Q^* = \underset{Q \in Q}{\arg\min} \int \log \frac{dQ(\zeta, \theta, \phi)}{d\Pi(\zeta, \theta, \phi | \mathbf{y}, z)} dQ(\zeta, \theta, \phi)$$
$$dQ^*_{\zeta}(\zeta) = \int_{\Theta} \int_{\Phi} dQ^*(\zeta, \theta, \phi),$$
$$d\Pi_{\zeta}(\zeta) = \int_{\Theta} \int_{\Phi} d\Pi(\zeta, \theta, \phi),$$
$$d\Pi_{\zeta}(\zeta | \mathbf{y}, z) = \int_{\Theta} \int_{\Phi} d\Pi(\zeta, \theta, \phi | \mathbf{y}, z)$$

Further, $Q_{\zeta}^{*}(\zeta)$, $\Pi_{\zeta}(\zeta)$ and $\Pi_{\zeta}(\zeta|\mathbf{y}, \mathbf{z})$ denote the marginal distribution of ζ under the variational posterior, the prior, and the true posterior respectively.

3.3.1 Assumptions

Let ϵ_n be a sequence satisfying $\epsilon_n \to 0$ and $n\epsilon_n^2 \to \infty$. Consider the following neighborhood of the true physical process as follows

$$\mathcal{U}_{\epsilon_n} = \left\{ \zeta : \frac{1}{n} \sum_{i=1}^n |\zeta(t_i) - \zeta_0(t_i)| \le \epsilon_n \right\}$$

We consider a sequence of sets $\{\mathcal{F}_n\}_{n=1}^{\infty}$ such that it increases to include the space of all continuously differentiable functions with bounded derivatives on $[0, 1]^p$. The following assumption describes the prior distributions on all the unknown parameters under consideration.

Assumption 3.3.1. We assume the following structure on the prior distributions of θ , ϕ_1 , ϕ_2 and σ^2

$$\Pi(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{m}_{\theta}^{0}, \boldsymbol{S}_{\theta}^{0}), \ \Pi(\boldsymbol{\phi}_{1}) \sim \mathcal{N}(\boldsymbol{m}_{\phi_{1}}^{0}, \boldsymbol{S}_{\phi_{2}}^{0}),$$
$$\Pi(\tilde{\boldsymbol{\phi}}_{2}) \sim \mathcal{N}(\boldsymbol{m}_{\phi_{2}}^{0}, \boldsymbol{S}_{\phi_{2}}^{0}), \ \Pi(\tilde{\sigma}) \sim \mathcal{N}(\boldsymbol{m}_{\sigma}^{0}, (\boldsymbol{s}_{\sigma}^{0})^{2}),$$

where $\tilde{\phi}_2 = \log \phi_2$ with the log operator being applied element wise and $\tilde{\sigma} = \log \sigma$. For simplicity S^0_{θ} , $S^0_{\phi_1}$, and $S^0_{\phi_2}$ are assumed to be diagonal matrices.

Assumption 3.3.2. For some $Q \in Q$

- (*i*) $\sup_{i=1,\dots,n} \mathbb{E}_Q \left[\int \zeta(t_i) d\Pi(\zeta | \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{z}) \zeta_0(t_i) \right]^2 \le \epsilon_n^2$
- (*ii*) $KL(Q(\theta, \phi)||\Pi(\theta, \phi, z)) \le Cn\epsilon_n^2 \log n$

Assumption 3.3.2 (i) states that there exists a variational member Q such that under Q the expectation of difference between the conditional expectation of the physical process ζ under the prior and the true physical process ζ_0 is negligible. This happens when Q places overwhelming mass on those values of θ and ϕ for which the conditional expectation of the physical process ζ approaches the true process ζ_0 . For the Squared Exponential covariance and Matérn covariance class, we establish that Assumption 3.3.2 (i) holds as long as there exists θ_n^* and ϕ_n^* satisfying

$$\sup_{i=1,\cdots,n} \left| \int \zeta(t_i) d\Pi(\zeta | \boldsymbol{\theta}_n^*, \boldsymbol{\phi}_n^*, \boldsymbol{z}) - \zeta_0(t_i) \right| \le \epsilon_n$$
(3.11)

for some $||\theta_n^*||_2 \leq B_\theta$ and $||\phi_n^*||_2 \leq B_\phi$ with B_θ and B_ϕ independent of *n*. Note, (3.11) assumes there exist good choices of θ and ϕ such that the conditional expectation of the physical process ζ under the parameters approaches the true physical process ζ_0 . Assumption 3.3.2 (ii) states that the KL distance between the variational member Q and the prior Π is bounded up to an order of $n\epsilon_n^2 \log n$. Although Q needs to give overwhelming probabilities to good values of θ and ϕ , it cannot be too far from the prior Π if variational posterior consistency needs to hold. For the Squared Exponential covariance and Matérn covariance class, this will hold as long as θ_n^* and ϕ_n^* satisfying (3.11) are bounded in L_2 norm up to an order of \sqrt{n} .

Indeed the GP priors as designed in (2.1) and Assumption 3.3.1 on the priors of θ and ϕ together with Assumption 3.2.1 on the variational family Q will be essential components towards the proof Assumption 3.3.2 (i) and (ii).

3.3.2 Main Results

Theorem 3. (Variational Posterior Consistency) Suppose Assumption 3.3.2 holds in addition to Assumption 2.2.1 and Assumption 2.2.2, then

$$Q^*_{\zeta}(\mathcal{U}^c_{\varepsilon_n})\longrightarrow 0$$

in \mathbb{P}_0^n probability as $n \to \infty$ and $\varepsilon_n = \epsilon_n \sqrt{M_n \log n}, M_n \to \infty$.

Note M_n can be any sequence increasing slowly to ∞ . Theorem 3 shows that the variational posterior of ζ concentrates around the true parameter value ζ_0 at the rate ε_n which is slightly higher than the rate ϵ_n for the true posterior. The detailed proof of the Theorem 3 has been provided in the Appendix. However, we next give a brief overview.

Proof of Theorem 3: By Corollary 4.15 in Boucheron et al. (2013), we have

$$\int f dQ_{\zeta}^* \leq \mathrm{K}L(Q_{\zeta}^*, \Pi_{\zeta}(.|\mathbf{y}, \mathbf{z})) + \log \int e^f d\Pi_{\zeta}(\cdot|\mathbf{y}, \mathbf{z})$$
(3.12)

Using (3.12) with $f = Cn\varepsilon_n^2 \mathbf{1}_{\mathcal{U}_{\varepsilon_n}^C}$ with **1** being the indicator function of a set, we get

$$Cn\varepsilon_{n}^{2}Q_{\zeta}^{*}(\mathcal{U}_{\varepsilon_{n}}^{C}) \leq \mathrm{K}L(Q_{\zeta}^{*}||\Pi_{\zeta}(|\mathbf{y},\mathbf{z})) + \log\left[1 + e^{Cn\varepsilon_{n}^{2}}\Pi_{\zeta}(\mathcal{U}_{\varepsilon_{n}}^{C}|\mathbf{y},\mathbf{z})\right]$$
(3.13)

By Assumptions 2.2.1 and 2.2.2, it can be shown that $\Pi_{\zeta}(\mathcal{U}_{\epsilon_n}^C|\mathbf{y}, \mathbf{z}) \leq \exp(-C'n\epsilon_n^2)$ where C' > C. Additionally, under Assumption 3.3.2, it can be shown that $KL(Q_{\zeta}^*||\Pi_{\zeta}(|\mathbf{y}, \mathbf{z})) \leq C''n\epsilon_n^2 \log n$ for some C'' > 0. Using these results in Relation (3.13), the proof of Theorem 3 follows.

3.4 Simulation Study: Ball Dropping Experiment



Figure 3.1 (Functions and data of Ball dropping experiment) The left figure shows the true physical process and the computer model with true calibration parameter. The right figure shows the data we obtained from these functions.

Let's consider a famous ball-dropping experiment from Plumlee (2017). A ball is dropped from a chosen vertical position 8 meters above the ground with initial velocity -0.1 m/s. Then the ball's vertical position at time *x* is given by following the physical process.

$$\zeta_0(t) = \frac{5}{2} \log \left\{ \frac{50}{49} - \frac{50}{49} \tanh \left[\tanh^{-1}(\sqrt{0.02}) + \sqrt{2}x \right]^2 \right\} + 8, \qquad x \in [0, 1]$$
(3.14)

We are going to consider an inexact computer model for this physical process with unknown calibration parameter θ as

$$f_m(t,\theta) = 8 - x - \frac{\theta x^2}{2}$$
 (3.15)

Note that the calibration parameter represents gravitational acceleration and let the true value of θ to be 10. The left figure in Figure 3.1 shows these two functions in the domain $x \in [0, 1]$. We can see that the computer model behaves very similarly to the true physical process, but there is a systematic difference, which will be modeled using GP.

For the data generation, we will use Latin hypercube samples for (x, θ) in the bounded space $[0, 1] \times [7, 12]$. In addition, we are going to add normal random error with a variance 0.01 to the experimental data. Then the location of the training data of size (N, S) = (30, 200) could



Figure 3.2 (Ball dropping example: posterior results of MCMC approach and VI approach) The left figure shows the prior distribution and the posterior distributions of calibration parameters using two different approaches. The right figure presents Root Mean Squared Error for 100 out of the sample physical data.

be checked in Figure 3.1, where N is the number of experimental data and S is the number of simulation data.

Figure 3.2 shows the results of the proposed method in terms of posterior distributions and the Root Mean Squared Error (RMSE) for 100 out of the sample physical data. The posterior distributions of the calibration parameter are centered around the truve value, and the proposed VI method shows less uncertainty. This is anticipated since the mean field variational family tends to underestimate the variance. Even in RMSEs the proposed VI method well captures the performance of MCMC approach. The main difference is coming from the running time of each algorithm. The proposed VI methods converged in less than 1 min. This convergence is remarkable compared to MCMC result, which took around 13 minutes for 10000 posterior samples.

3.5 Simulation Study: Logistic Function

Let's consider a simple computer model in the form of a logistic growth function:

$$f_m((t,x),\theta) = \frac{1+\theta_1}{1+\exp\left(-[1+x+\theta_2 t]\right)}$$
(3.16)

where we assume that the calibration parameter θ takes values in the space $[0, 2]^2$ and the model inputs (t, x) take values in the space $[-10, 10]^2$. The true physical process is modeled according to

$$\zeta_0(t,x) = f_m((t,x), \theta) + \delta(t,x) = \frac{1+\theta_1}{1+\exp\left(-\left[1+x+\theta_2 t\right]\right)} + \beta,$$
(3.17)

where $\beta = 1$ is a constant systematic error of the model and the true value of $\theta = (\theta_1, \theta_2)$ are arbitrarily set to be (1, 1.8). Since we are going to have 2 calibration parameters, we are going to consider three cases of sample sizes to compare the time gain and the fidelity of the variational approximation. The position of the data in the 2-dimensional input space is shown in Figure 3.3 for the case (n, s) = (500, 1000).

The posterior distributions of calibration parameters are drawn in Figure 3.4 using MCMC and VI to compare the results. As one can see, MCMC often has dented shapes, but the VI method approximates the posterior density well with round contour plots on top of the MCMC result. As the sample size increases, the uncertainty of MCMC results decreases with less standard deviations in both θ_1 and θ_2 , and the VI results show very similar behavior. We can also check this fact from Table 3.1 with the mean and standard deviation of each sample size.







Figure 3.4 (Posterior distributions of two calibration parameters) The blue lines represents posterior distribution coming from MCMC method, orange color is the variational approximation, and the red dot represents the true value of calibration parameters. We consider three cases of sample size to compare the performance of variational approximation.

(n,s)	(125,250)		(250,500)		(500,1000)	
Method	MCMC	VB	MCMC	VB	MCMC	VB
θ_1	0.89(0.08)	0.89(0.05)	0.95(0.04)	0.93(0.04)	0.97(0.03)	0.96(0.02)
θ_2	1.77(0.09)	1.76(0.08)	1.72(0.03)	1.72(0.03)	1.76(0.02)	1.75(0.01)
Condition (3.11)	1.326		1.045		1.027	

Table 3.1 Posterior means and standard deviations (in parenthesis) of calibration parameters. Here, n corresponds to the number of experimental observations and s corresponds to the number of computer model evaluations.

Table 3.1 also shows the condition (3.11) of variational family, the supremum of difference of the mean of GP prior given θ and ϕ and the true physical processes.

$$\sup_{i=1,\cdots,n} \left| \int \zeta(t_i) d\Pi(\zeta | \boldsymbol{\theta}_n^*, \boldsymbol{\phi}_n^*, \boldsymbol{z}) - \zeta_0(t_i) \right| \leq \epsilon_n$$

The left side of the equation is calculated in this example and the value is decreasing to 1 as the number of sample sizes increases. The values are not decreasing to 0 in this case, because the simulation data in this setup has a constant bias of value 1, hence it is decreasing to the value 1. If the simulation equation approximates the real physical process well so that the difference is negligible, then the mean of GP prior updated by the simulation data will be very close to the true physical process and the upper bound of the left side of the equation will decrease to 0.

The big difference between the two approaches is shown in each algorithm's running time. The upper figures in Figure 3.5 show the convergence of ELBO for the VI approach in terms of time in hours and the posterior sampling time for the MCMC approach. One can see that ELBO convergence is quite fast compared to the MCMC method. For example, for (n, s) = (500, 1000) case, the MCMC algorithm took 14 hours for 10,000 posterior samples, while the VI approach took around 30 minutes for ELBO convergence. In other words, when ELBO converges, MCMC methods could only produce 450 posterior samples at maximum among all three cases of different sample sizes.

We can also check that the Root Mean Squared Error (RMSE) for the 100 out-of-sample test data set also converges quickly for VI method which shows quite close results to MCMC result. MCMC approach also converges quite quickly in this simple simulation setup, but acquiring enough posterior samples for inference is very time-consuming compared to the VI approach.

In short, we can see that the proposed VI method is consistent with MCMC in this simple example with a bit of smaller uncertainty. On the other hand, the computational time is very different; the VI approach is faster up to 23 times in this example compared to the MCMC approach.



Figure 3.5 (Time comparison for Logistic growth function example between posterior results of MCMC approach and VI approach) Upper figures show the change of ELBO in terms of time for VI approach and the Lower figures show the change of Root Mean Squared Error(RMSE) for 100 out of the sample physical data with regard to time.

3.6 Conclusion and Discussion

The original KOH model is a powerful approach for computer model calibration and estimation of physical processes. However, the standard MCMC implementation is not trivial and computa-

tionally expensive due to the use of GPs. We addressed the scalability issue of KOH model by replacing MCMC with VB approximation which lead to a significant reduction in the code's runtime. Our simulation study shows that the VB method produces posterior approximation with high fidelity with much less computation time compared to the MCMC. In the future, we plan to deploy the proposed VB approach on a wide range of validation studies including a real data example. On the theoretical end, we first established the posterior contraction rate of the true posterior followed by the posterior contraction rate for the variational approximation. As reflected in the theoretical analysis, the posterior contraction rate of the proposed VB approach is only a bit slower than the true posterior. In the theoretical section, we assumed the parameter σ^2 to be known. Extending to the case of unknown σ^2 will be an interesting line of work. Further, our theoretical development relies on the existence of suitable choices of θ and ϕ such that the conditional expectation of the computer model approaches the true physical process ζ_0 .

BIBLIOGRAPHY

- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.
- Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.
- Csató, L. (2002). Gaussian processes: iterative sparse approximations. *Ph.D. thesis, Aston University*.
- Csató, L. and Opper, M. (2002). Sparse on-line gaussian processes. *Neural Computation*, 14(3):641–668.
- de G. Matthews, A. G., Hensman, J., Turner, R. E., and Ghahramani, Z. (2016). On sparse variational methods and the kullback-leibler divergence between stochastic processes. *Artificial Intelligence and Statistics*, pages 231–239.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233.
- Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint* arXiv:1312.6114.
- Kolmogorov, A. N., Shiryaev, A. N., and Lindquist, G. (1992). *Selected works of A. N. Kolmogorov. Volume II.* Kluwer Academic Publishers.
- Lee, H. (2000). Consistency of posterior distributions for neural networks. *Neural Network*, 13:629–642.
- Plumlee, M. (2017). Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112:1274–1285.
- Ranganath, R., Gerrish, S., and Blei, D. (2014). Black box variational inference. In *Proceedings* of the Seventeenth International Conference on Artificial Intelligence and Statistics, volume 33 of *Proceedings of Machine Learning Research*, pages 814–822. PMLR.
- Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR.
- Seeger, M. (2003a). Bayesian gaussian process models: Pac-bayesian generalisation error bounds and sparse approximations. *Ph.D. thesis, University of Edinburgh.*

Seeger, M. (2003b). Pac-bayesian generalisation error bounds for gaussian process classification.

Journal of Machine Learning Research, 3:233–269.

Sun, S., Zhang, G., Shi, J., and Grosse, R. (2019). Functional variational bayesian neural networks. *International Conference on Learning Representations*.

APPENDIX A

CHECKING ASSUMPTION 3.3.2 (I)

Under Assumption 3.3.2 (i) using (3.11), θ^* , ϕ^* depends on *n*. However, for notation simplicity we suppress *n* in the remaining proof. Now, let

$$|\mathbb{E}_{\zeta|\boldsymbol{\theta},\boldsymbol{\phi}}\zeta(t_i) - \zeta_0(t_i)| = f_i(\boldsymbol{\theta},\boldsymbol{\phi})$$

Further define $\mathcal{A} = \{ |\theta_j - \theta_j^*| \le \epsilon_n, |\phi_{1,j} - \phi_{1,j}^*| \le \epsilon_n, |\log \phi_{2,j} - \log \phi_{2,j}^*| \le \epsilon_n \},\$

$$\begin{split} & \mathbb{E}_{q(\theta,\phi)} | (\zeta_0(t_i) - \mathbb{E}_{\zeta|\theta,\phi,z}(\zeta(t_i)) |^2 = \int f_i^2(\theta,\phi) q(\theta,\phi) d\theta d\phi \\ &= \int_{\mathcal{A}} f_i^2(\theta,\phi) q(\theta,\phi) d\theta d\phi + \int_{\mathcal{A}^c} f_i^2(\theta,\phi) q(\theta,\phi) d\theta d\phi \\ &\leq \int_{\mathcal{A}} (f_i^2(\theta^*,\phi^*) + C\epsilon_n^2) q(\theta,\phi) d\theta d\phi + \left(\max_{k \in \{1,\cdots,s\}} z_k \bar{C} + |\zeta_0(t_i)| \right)^2 Q(\mathcal{A}^c) \\ &\leq C\epsilon_n^2 + C \exp(-n\epsilon_n^2) \end{split}$$

The first inequality above holds since $f_i^2(\theta, \phi)$ is a continuous function in θ and ϕ . Thus, for every $\delta_{1n} > 0$, there exists $\delta_{2n} > 0$ such that for every $(\theta, \phi) \in \mathcal{B}((\theta^*, \phi^*), \delta_{2n}), f_i^2(\theta, \phi) \leq f_i^2(\theta^*, \phi^*) + \delta_{1n}$ where $\mathcal{B}((\theta^*, \phi^*), \delta_n)$ is a ball of radius δ_n around (θ^*, ϕ^*) . Let $\epsilon_n = \min(\sqrt{\delta_{1n}}, \delta_{2n})$.

To prove the second inequality $Q(\mathcal{A}^c) \leq \exp(-n\epsilon_n^2)$, we use the tail bounds for normal distribution which states if X is normally distributed, $\mathbb{P}(|X - \mu|/\sigma \geq x) \leq \exp(-x^2)$ as $x \to \infty$ in addition to the fact that $Q((A_1 \cap \cdots \cap A_r)^c) = \sum_r Q(A_i^c)$. To be specific,

$$\prod_{j=1}^{p_{\theta}} Q(\sqrt{n}|\theta_j - \theta_j^*| \ge \sqrt{n}\epsilon_n) \le 2p_{\theta} \exp(-n\epsilon_n^2)$$
$$\prod_{j=1}^{p_{\phi_1}} Q(|\phi_{1,j} - \phi_{1,j}^*| \ge \epsilon_n) \le 2p_{\phi_1} \exp(-n\epsilon_n^2)$$
$$\prod_{j=1}^{p_{\phi_2}} Q(|\log \phi_{2,j} - \log \phi_{2,j}^*| \ge \epsilon_n) \le 2p_{\phi_2} \exp(-n\epsilon_n^2)$$

Finally, $|f_i(\theta, \phi)| \le \max_{k \in \{1, \dots, s\}} z_k \overline{C} + |\zeta_0(t_i)|$ follows as a consequence of (C.4).

APPENDIX B

CHECKING ASSUMPTION 3.3.2 (II)

Assume independent priors for θ and ϕ as follows

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{p_{\boldsymbol{\theta}}} p(\theta_{i}) = \prod_{i=1}^{p_{\boldsymbol{\theta}}} \mathcal{N}(\mu_{\theta,i}, \sigma_{\theta,i}^{2}), q(\boldsymbol{\theta}) = \prod_{i=1}^{p_{\boldsymbol{\theta}}} q(\theta_{i}) = \prod_{i=1}^{p_{\boldsymbol{\theta}}} \mathcal{N}\left(m_{\theta,i}, s_{\theta,i}^{2}\right)$$

$$p(\boldsymbol{\phi}_{1}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{1}}} p(\phi_{1,i}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{1}}} \mathcal{N}\left(\mu_{\phi_{1},i}, \sigma_{\phi_{1},i}^{2}\right), q(\boldsymbol{\phi}_{1}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{1}}} q(\phi_{1,i}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{1}}} \mathcal{N}\left(m_{\phi_{1},i}, s_{\phi_{1},i}^{2}\right)$$

$$p(\boldsymbol{\phi}_{2}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{2}}} p(\phi_{2,i}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{2}}} \mathcal{L}\mathcal{N}\left(\log \mu_{\phi_{2,i}}, \sigma_{\phi_{2,i}}^{2}\right), q(\boldsymbol{\phi}_{2}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{2}}} q(\phi_{2,i}) = \prod_{i=1}^{p_{\boldsymbol{\phi}_{2}}} \mathcal{L}\mathcal{N}\left(\log m_{\phi_{2,i}}, s_{\phi_{2,i}}^{2}\right)$$

Note that, $Q(\theta, \phi) = Q(\theta)Q(\phi_1)Q(\phi_2), \Pi(\theta, \phi|z) = \Pi(\phi_1|z)\Pi(\phi_2|z)\Pi(\theta)$. Therefore,

$$\begin{split} & \operatorname{KL}(Q(\zeta,\theta,\phi)||\Pi(\zeta,\theta,\phi|z)) \\ &= \operatorname{KL}(Q(\phi_1)||\Pi(\phi_1|z)) + \operatorname{KL}(Q(\phi_2)||\Pi(\phi_2|z)) + \operatorname{KL}(Q(\theta)||\Pi(\theta)) \end{split}$$

Each KL term can be calculated as follows. Let

$$p(\theta_i) = \mathcal{N}(\mu_{\theta,i}, \sigma_{\theta,i}^2), q(\theta_i) = \mathcal{N}(m_{\theta,i}, s_{\theta,i}^2)$$

then we can simplify $KL(Q(\theta)||\Pi(\theta))$ as

$$\begin{split} E_{Q(\theta)} \left[-\frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(2\pi s_{\theta,i}^{2}\right) - \left(\sum_{i=1}^{p_{\theta}} \frac{(\theta_{i} - m_{\theta,i})^{2}}{2s_{\theta,i}^{2}}\right) + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(2\pi \sigma_{\theta,i}^{2}\right) + \left(\sum_{i=1}^{p_{\theta}} \frac{(\theta_{i} - \mu_{\theta,i})^{2}}{2\sigma_{\theta,i}^{2}}\right) \right] \\ &= -\frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(2\pi s_{\theta,i}^{2}\right) - \frac{1}{2} \sum_{i=1}^{p_{\theta}} E_{Q(\theta)} \left[\left(\frac{\theta_{i} - m_{\theta,i}}{s_{\theta,i}}\right)^{2} \right] \\ &+ \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(2\pi \sigma_{\theta,i}^{2}\right) \right) + \frac{1}{2} E_{Q(\theta)} \left[\left(\sum_{i=1}^{p_{\theta}} \frac{(\theta_{i} - \mu_{\theta,i})^{2}}{\sigma_{\theta,i}^{2}}\right) \right] \\ &= -\frac{1}{2} p_{\theta} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(s_{\theta,i}^{2}\right) - \frac{1}{2} p_{\theta} + \frac{1}{2} p_{\theta} \log 2\pi + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(\sigma_{\theta,i}^{2}\right) \\ &+ \frac{1}{2} \sum_{i=1}^{p_{\theta}} \left(\frac{s_{\theta,i}^{2}}{\sigma_{\theta,i}^{2}} + \left(\frac{m_{\theta,i} - \mu_{\theta,i}}{\sigma_{\theta,i}}\right)^{2}\right) \\ &= -\frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(s_{\theta,i}^{2}\right) - \frac{1}{2} p_{\theta} + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(\sigma_{\theta,i}^{2}\right) + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \left(\frac{s_{\theta,i}^{2}}{\sigma_{\theta,i}^{2}} + \left(\frac{m_{\theta,i} - \mu_{\theta,i}}{\sigma_{\theta,i}}\right)^{2} \right) \end{split}$$

Choose $m_{\theta,i} = \theta_{0,i}$ and $s_{\theta,i}^2 = 1/n$. Then, for a positive constant $C_1 > 0$, we can simplify $KL(Q(\theta)||\Pi(\theta))$ as follows

$$= -\frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(s_{\theta,i}^{2}\right) - \frac{1}{2} p_{\theta} + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(\sigma_{\theta,i}^{2}\right) + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \left(\frac{s_{\theta,i}^{2}}{\sigma_{\theta,i}^{2}} + \left(\frac{m_{\theta,i} - \mu_{\theta,i}}{\sigma_{\theta,i}}\right)^{2}\right)$$
$$= \frac{1}{2} p_{\theta} \log n - \frac{1}{2} p_{\theta} + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \log\left(\sigma_{\theta,i}^{2}\right) + \frac{1}{2} \sum_{i=1}^{p_{\theta}} \left(\frac{1}{n\sigma_{\theta,i}^{2}} + \left(\frac{\theta_{0,i} - \mu_{\theta,i}}{\sigma_{\theta,i}}\right)^{2}\right)$$
$$= o(C_{1} n \epsilon_{n}^{2} \log n)$$

where the above line follows by Assumption 8. (i).

By similar logic, let $p(\phi_{1,i}) = \mathcal{N}(\mu_{\phi_1,i}, \sigma_{\phi_1,i}^2), q(\phi_{1,i}) = \mathcal{N}(m_{\phi_1,i}, s_{\phi_1,i}^2)$ then with $m_{\phi_1,i} = \phi_{0,1,i}$ and $s_{\phi_1,i}^2 = 1/n$. Then, for a positive constant $C_1 > 0$

$$\mathrm{K}L(Q(\boldsymbol{\phi}_1)||\Pi(\boldsymbol{\phi}_1|\boldsymbol{z})) = o(C_1 n \epsilon_n^2 \log n)$$

Let $p(\log \phi_{2,i}) = \mathcal{N}(\log \mu_{\phi_{2},i}, \sigma_{\phi_{2},i}^2), q(\log \phi_{2,i}) = \mathcal{N}(\log m_{\phi_{2},i}, s_{\phi_{2},i}^2)$, then

$$\begin{split} & \mathsf{KL}(\mathcal{Q}(\boldsymbol{\phi}_2)||\Pi(\boldsymbol{\phi}_2|z)) \\ &= E_{\mathcal{Q}(\boldsymbol{\phi}_2)} \left[-\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left((s_{\phi_2,i}\phi_{2,i}^{\frac{3}{2}})\right) - \left(\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \frac{(\log \phi_{2,i} - \log m_{\phi_2,i})^2}{2s_{\phi_2,i}^2}\right) \right] \\ &\quad + E_{\mathcal{Q}(\boldsymbol{\phi}_2)} \left[\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left((\sigma_{\phi_2,i}\phi_{2,i}^{\frac{3}{2}})\right) + \left(\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \frac{(\log \phi_{2,i} - \log \mu_{\phi_2,i})^2}{2\sigma_{\phi_2,i}^2}\right) \right] \\ &= E_{\mathcal{Q}(\boldsymbol{\phi}_2)} \left[-\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left(s_{\phi_2,i}\right) - \frac{3}{2} \sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left(\phi_{2,i}\right) - \frac{1}{2} \sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \left(\frac{(\log \phi_{2,i} - \log m_{\phi_2,i})}{s_{\phi_2,i}}\right)^2 \right] \\ &\quad + E_{\mathcal{Q}(\boldsymbol{\phi}_2)} \left[\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left(\sigma_{\phi_2,i}\right) + \frac{3}{2} \sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left(\phi_{2,i}\right) + \frac{1}{2} \left(\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \frac{(\log \phi_{2,i} - \log \mu_{\phi_2,i})^2}{\sigma_{\phi_2,i}^2} \right) \right] \\ &= -\sum_{i=1}^{p_{\boldsymbol{\phi}_2}} \log\left(s_{\phi_2,i}\right) - \frac{3}{2} \sum_{i=1}^{p_{\boldsymbol{\phi}_2}} m_{\phi_2,i} - \frac{1}{2} p_{\phi_2} + \sum_{i=1}^{p_{\phi_2}} \log(\sigma_{\phi_2,i}) + \frac{3}{2} \sum_{i=1}^{p_{\phi_2}} m_{\phi_2,i} \\ &\quad + \frac{1}{2} \sum_{i=1}^{p_{\phi_2}} \left(\frac{s_{\phi_2,i}^2}{\sigma_{\phi_2,i}^2} + \left(\frac{\log m_{\phi_2,i} - \log \mu_{\phi_2,i}}{\sigma_{\phi_2,i}} \right)^2 \right) \end{split}$$

Hence,

$$\begin{split} \mathsf{KL}(Q(\phi_2)||\Pi(\phi_2|z)) \\ &= -\sum_{i=1}^{p_{\phi_2}} \log\left(s_{\phi_{2,i}}\right) - \frac{3}{2} \sum_{i=1}^{p_{\phi_2}} m_{\phi_{2,i}} - \frac{1}{2} p_{\phi_2} + \sum_{i=1}^{p_{\phi_2}} \log(\sigma_{\phi_{2,i}}) + \frac{3}{2} \sum_{i=1}^{p_{\phi_2}} m_{\phi_{2,i}} \\ &+ \frac{1}{2} \sum_{i=1}^{p_{\phi_2}} \left(\frac{s_{\phi_{2,i}}^2}{\sigma_{\phi_{2,i}}^2} + \left(\frac{\log m_{\phi_{2,i}} - \log \mu_{\phi_{2,i}}}{\sigma_{\phi_{2,i}}} \right)^2 \right) \\ &= \frac{1}{2} p_{\phi_2} \log n - \frac{3}{2} p_{\phi_2} \phi_{0,2,i} - \frac{1}{2} p_{\phi_2} + \sum_{i=1}^{p_{\phi_2}} \log(\sigma_{\phi_{2,i}}) + \frac{3}{2} \sum_{i=1}^{p_{\phi_2}} m_{\phi_{2,i}} \\ &+ \frac{1}{2} \sum_{i=1}^{p_{\phi_2}} \left(\frac{1}{n \sigma_{\phi_{2,i}}^2} + \left(\frac{\log \phi_{0,2,i} - \log \mu_{\phi_{2,i}}}{\sigma_{\phi_{2,i}}} \right)^2 \right) \\ &= o(C_1 n \epsilon_n^2 \log n) \end{split}$$

where we choose $m_{\phi_2,i} = \phi_{0,2,i}$ and $s_{\phi_2,i}^2 = 1/n$. Then, for a positive constant $C_1 > 0$ where the above line follows by Assumption 8. (i). This completes the proof.

APPENDIX C

PROOF OF THEOREM 3

Based on the sketch of the proof of Theorem 3 in the main paper, the complete proof boils down to show $KL(Q_{\zeta}^*||\Pi_{\zeta}(\cdot|y, z)) \leq C'' n\epsilon_n^2 \log n$ for some C'' > 0. This depends on the following lemmas. Again, remember that z is implicitly conditioned on the prior.

Lemma C.0.1. *Inequality for KL divergence*

$$KL(Q_{\zeta}^{*}||\Pi_{\zeta}(\cdot|\mathbf{y}, \mathbf{z})) \le KL(Q(\boldsymbol{\theta}, \boldsymbol{\phi})||\Pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}, \mathbf{z}))$$
(C.1)

Lemma C.0.2. There exists a constant C_2 such that for any $\epsilon_n \to 0$ with $n\epsilon_n^2 \to \infty$,

$$\mathbb{P}_0\left\{ \left| \log \frac{p(\mathbf{y}|\mathbf{z})}{p(\mathbf{y}|\zeta_0)} \right| < C_2 n \epsilon_n^2 \log n \right\} \longrightarrow 1$$
(C.2)

Lemma C.0.3. There exists a constant C_3 such that for any $\epsilon_n \to 0$ with $n\epsilon_n^2 \to \infty$,

$$\mathbb{P}_{0}\left\{\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{\phi})}\left[\log\frac{p(\boldsymbol{y}|\boldsymbol{\zeta}_{0})}{p(\boldsymbol{y}|\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\phi})}\right] \leq C_{3}n\epsilon_{n}^{2}\log n\right\} \longrightarrow 1$$
(C.3)

Lemma C.0.4. By using the Lemma C.0.2, Lemma C.0.3, and Assumption 3.3.2 (ii) in the main paper it can be established with dominating probability for any $C_9 > 0$, as $n \to \infty$,

$$KL(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{y}, z)) \le C_9 n\epsilon_n^2 \log n$$
(C.4)

Using Lemma C.0.4, we show variational posterior and the true posterior is bounded.

C.0.1 Proof of Lemma C.0.1

Note,

$$\begin{split} \mathrm{K}L(Q_{\zeta}^{*}||\Pi_{\zeta}(\cdot|\mathbf{y},z)) &= \int \log \frac{dQ_{\zeta}^{*}(\zeta)}{d\Pi_{\zeta}(\zeta|\mathbf{y},z)} dQ_{\zeta}^{*}(\zeta) \\ &\leq \int \log \frac{dQ^{*}(\zeta,\theta,\phi)}{d\Pi(\zeta,\theta,\phi|\mathbf{y},z)} dQ^{*}(\zeta,\theta,\phi) \\ &\leq \int \log \frac{dQ(\zeta,\theta,\phi)}{d\Pi(\zeta,\theta,\phi|\mathbf{y},z)} dQ(\zeta,\theta,\phi) \\ &= \log \frac{q(\zeta|\theta,\phi)q(\theta,\phi)}{\pi(\zeta|\theta,\phi,\mathbf{y},z)\pi(\theta,\phi|\mathbf{y},z)} dQ(\zeta,\theta,\phi) \\ &= \mathrm{K}L(Q(\theta,\phi)||\Pi(\theta,\phi|\mathbf{y},z)) \end{split}$$

where the last line of equality follows since $q(\zeta | \theta, \phi) = \pi(\zeta | \theta, \phi, y, z)$. The first inequality above is due to the argument below and the second inequality above is because Q^* is the minimizer of $KL(Q||\Pi(\cdot|y, z))$. Proof of the 1st inequality is as follows.

$$\int \log \frac{dQ_{\zeta}^{*}(\zeta)}{d\Pi_{\zeta}(\zeta|\mathbf{y}, \mathbf{z})} dQ_{\zeta}^{*}(\zeta) = \int \log \frac{q_{\zeta}^{*}(\zeta)}{\pi_{\zeta}(\zeta|\mathbf{y}, \mathbf{z})} q_{\zeta}^{*}(\zeta) d\zeta$$

Now,

$$\log \frac{q_{\zeta}^{*}(\zeta)}{\pi_{\zeta}(\zeta|\mathbf{y},z)} q_{\zeta}^{*}(\zeta) = \int q^{*}(\zeta,\theta,\phi) d\theta d\phi \log \frac{\int q^{*}(\zeta,\theta,\phi) d\theta d\phi}{\int \pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi}$$
$$= \int \pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi \frac{\int (q^{*}(\zeta,\theta,\phi)/\pi(\zeta,\theta,\phi|\mathbf{y},z))\pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi}{\int \pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi}$$
$$\times \zeta \log \frac{\int q^{*}(\zeta,\theta,\phi)/\pi(\zeta,\theta,\phi|\mathbf{y},z)\pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi}{\int \pi(\zeta,\theta,\phi|\mathbf{y},z) d\theta d\phi}$$

Consider density function for fixed ζ : $\tilde{\pi}_{\zeta}(\theta, \phi) = \pi(\zeta, \theta, \phi|y, z)) / \int \pi(\zeta, \theta, \phi|y, z) d\theta d\phi$. Then,

$$\log \frac{q_{\zeta}^{*}(\zeta)}{\pi_{\zeta}(\zeta|\mathbf{y}, \mathbf{z})} q_{\zeta}^{*}(\zeta) = \int \pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z}) d\theta d\phi$$

$$E_{\tilde{\pi}_{\zeta}}(q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z})) \log E_{\tilde{\pi}_{\zeta}}(q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z}))$$

$$\leq \int \pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z}) d\theta d\phi$$

$$E_{\tilde{\pi}_{\zeta}}((q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z})) \log(q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z})))$$

$$= \int \pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z}) d\theta d\phi$$

$$\frac{\int q^{*}(\zeta, \theta, \phi) \log(q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z})) d\theta d\phi}{\int \pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z}) d\theta d\phi}$$

$$= \int q^{*}(\zeta, \theta, \phi) \log(q^{*}(\zeta, \theta, \phi)/\pi(\zeta, \theta, \phi|\mathbf{y}, \mathbf{z})) d\theta d\phi$$

where the second step follows by using Jensen's inequality on the convex function $x \to x \log x$. Thus,

$$\begin{split} \mathrm{K}L(Q_{\zeta}^{*}||\Pi_{\zeta}(\cdot|\mathbf{y}, \mathbf{z})) &= \int \log \frac{q_{\zeta}^{*}(\zeta)}{\pi_{\zeta}(\zeta|\mathbf{y}, \mathbf{z})} q_{\zeta}^{*}(\zeta) d\zeta \\ &\leq \int q^{*}(\zeta, \boldsymbol{\theta}, \boldsymbol{\phi}) \log(q^{*}(\zeta, \boldsymbol{\theta}, \boldsymbol{\phi}) / \pi(\zeta, \boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}, \mathbf{z}))) d\boldsymbol{\theta} d\boldsymbol{\phi} d\zeta \end{split}$$

which completes the proof.

C.0.2 Proof of Lemma C.0.2

Let
$$L^* = p(\mathbf{y}|\mathbf{z}) = \int_{\mathbb{S}} p(\mathbf{y}|\zeta) d\Pi(\zeta|\mathbf{z})$$
 and $L(\zeta_0) = p(\mathbf{y}|\zeta_0)$. By Markov's inequality, we have

$$\mathbb{P}_0 \left\{ \left| \log \int_{\zeta} \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) \right| > C_2 n \epsilon_n^2 \log n \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \mathbb{E}_0 \left\{ \left| \log \int \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) \right| \right\}$$

$$= \frac{1}{C_2 n \epsilon_n^2 \log n} \mathbb{E}_0 \left\{ \left| \log \frac{L^*}{L(\zeta_0)} \right| \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \mathbb{E}_0 \left\{ \left| \log \frac{L(\zeta_0)}{L^*(\zeta)} \right| \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \mathbb{E}_0 \left\{ \log \frac{p(\mathbf{y}|\zeta_0)}{f(\mathbf{y}|\zeta)} \right\} + \frac{2}{e} \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \left\{ \mathbb{E}_0 \left\{ \log \frac{p(\mathbf{y}|\zeta_0)}{f(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) \right\} + \frac{2}{e} \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \left\{ \mathbb{E}_0 \left\{ -\log \int \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) \right\} + \frac{2}{e} \right\}$$

$$\leq \frac{1}{C_2 n \epsilon_n^2 \log n} \mathbb{E}_0 \left\{ \mathbb{E}_0 \left\{ -\log \int \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) \right\} + \frac{2}{e} \right\}$$

$$= \frac{1}{C_2 n \epsilon_n^2 \log n} \left\{ \mathbb{E}_0 \left\{ -\log \Pi_\zeta (B_n|\mathbf{z}) + C_4 n \epsilon_n^2 \right\} + \frac{2}{e} \right\}$$

$$= \frac{1}{C_2 n \epsilon_n^2 \log n} \left\{ \mathbb{E}_0 \left\{ C n \epsilon_n^2 + C_4 n \epsilon_n^2 \right\} + \frac{2}{e} \right\}$$

Note the first inequality from the bottom, we applied log function to the result of (A.4). For the third inequality from the top, we used Lemma 4 from Lee (2000).

C.0.3 Proof of Lemma C.0.3

Note

$$\mathbb{E}_{q(\theta,\phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] = \mathbb{E}_{q(\theta,\phi)} \left[-\log \frac{p(\mathbf{y}|\mathbf{z}, \theta, \phi)}{p(\mathbf{y}|\zeta_0)} \right]$$
$$= \mathbb{E}_{q(\theta,\phi)} \left[-\log \frac{\int p(\mathbf{y}|\mathbf{z}, \theta, \phi, \zeta) p(\zeta|\theta, \phi, \mathbf{z}) d\zeta}{p(\mathbf{y}|\zeta_0)} \right]$$
$$= \mathbb{E}_{q(\theta,\phi)} \left[-\log \int \frac{p(\mathbf{y}|S)}{p(\mathbf{y}|\zeta_0)} p(\zeta|\theta, \phi, \mathbf{z}) d\zeta \right]$$

Now, using Markov's inequality, Jensen's inequality and Fubini's theorem,

$$\mathbb{P}_{0}\left[\mathbb{E}_{q(\theta,\phi)}\left[-\log\int\frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_{0})}p(\zeta|\theta,\phi,z)d\zeta\right] \geq C_{3}n\epsilon_{n}^{2}\log n\right]$$

$$\leq \frac{1}{C_{3}n\epsilon_{n}^{2}\log n}\mathbb{E}_{0}\left[\mathbb{E}_{q(\theta,\phi)}\left[-\log\int\frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_{0})}p(\zeta|\theta,\phi,z)d\zeta\right]\right]$$

$$\leq \frac{1}{C_{3}n\epsilon_{n}^{2}\log n}\mathbb{E}_{0}\left[\mathbb{E}_{q(\theta,\phi)}\left[\int-\log\frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_{0})}p(\zeta|\theta,\phi,z)d\zeta\right]\right]$$

$$= \frac{1}{C_{3}n\epsilon_{n}^{2}\log n}\left[\mathbb{E}_{q(\theta,\phi)}\mathbb{E}_{\zeta|\theta,\phi,z}\mathbb{E}_{0}\left[-\log\frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_{0})}\right]\right]$$

$$\leq \frac{1}{C_{3}n\epsilon_{n}^{2}\log n}\left[-\frac{n}{2}+\frac{n\sigma_{0}^{2}}{2}+C\epsilon_{n}^{2}+C\epsilon_{n}^{2}\exp(-n\epsilon_{n}^{2})\right] \rightarrow 0$$

where the second inequality above is due to Jensen's inequality In the last inequality,

$$\begin{split} & \mathbb{E}_{q(\theta,\phi)} \mathbb{E}_{\zeta|\theta,\phi,z} \mathbb{E}_{0} \left[-\log \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_{0})} \right] \\ &= \mathbb{E}_{q(\theta,\phi)} \mathbb{E}_{\zeta|\theta,\phi,z} \mathbb{E}_{0} \left[-\frac{1}{2} \sum_{i=1}^{n} (y_{i} - \zeta_{0}(t_{i}))^{2} + \frac{1}{2} \sum_{i=1}^{n} (y_{i} - \zeta(t_{i}))^{2} \right] \\ &= \mathbb{E}_{q(\theta,\phi)} \mathbb{E}_{\zeta|\theta,\phi,z} \mathbb{E}_{0} \left[-\frac{n}{2} + \frac{1}{2} \sum_{i=1}^{n} (y_{i} - \zeta(t_{i}))^{2} \right] \\ &= \mathbb{E}_{q(\theta,\phi)} \mathbb{E}_{\zeta|\theta,\phi,z} \left[-\frac{n}{2} + \frac{1}{2} \sum_{i=1}^{n} (1 + \zeta_{0}^{2}(t_{i}) - 2\zeta(t_{i})\zeta_{0}(t_{i}) + \zeta^{2}(t_{i})) \right] \\ &= \mathbb{E}_{q(\theta,\phi)} \left[-\frac{n}{2} + \frac{n}{2} \right] + \mathbb{E}_{q(\theta,\phi)} \left[\frac{1}{2} \sum_{i=1}^{n} (\zeta_{0}(t_{i}) - \mathbb{E}_{\zeta|\theta,\phi,\sigma,z}[\zeta(t_{i})])^{2} \right] \\ &\leq C\epsilon_{n}^{2} + C\epsilon_{n}^{2} \exp(-n\epsilon_{n}^{2}) \end{split}$$

where the last inequality above is a consequence of Assumption 3.3.2 (i).

C.0.4 Proof of Lemma C.0.4

Note,

$$\begin{split} & \operatorname{KL}(q(\theta, \phi) || p(\theta, \phi | \mathbf{y}, z)) \\ &= E_{q(\theta, \phi)} [\log q(\theta, \phi) - \log p(\mathbf{y} | z, \theta, \phi) - \log p(\theta, \phi | z) - \log p(z) + \log p(\mathbf{y}, z)] \\ &= \operatorname{KL}(Q(\theta, \phi) || \Pi(\theta, \phi | z)) + \log \frac{p(\mathbf{y}, z)}{p(z)} + E_{q(\theta, \phi)} \left[\log \frac{1}{p(\mathbf{y} | z, \theta, \phi)} \right] \end{split}$$

Hence,

$$\begin{split} & \mathsf{KL}(q(\theta, \phi)||p(\theta, \phi|\mathbf{y}, \mathbf{z})) \\ &= \mathsf{KL}(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{z})) + \log \frac{p(\mathbf{y}, \mathbf{z})}{p(\mathbf{z})p(\mathbf{y}|\zeta_0)} + E_{q(\theta, \phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] \\ &= \mathsf{KL}(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{z})) + \log \int_{\zeta} \frac{p(\mathbf{y}, \mathbf{z}, \zeta)}{p(\mathbf{z})p(\mathbf{y}|\zeta_0)} d\zeta + E_{q(\theta, \phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] \\ &= \mathsf{KL}(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{z})) + \log \int_{\mathbb{S}} \frac{p(\mathbf{y}|\zeta)p(\zeta|\mathbf{z})p(\mathbf{z})}{p(\mathbf{y}|\zeta_0)} d\zeta + E_{q(\theta, \phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] \\ &= \mathsf{KL}(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{z})) + \log \int_{\mathbb{S}} \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) + E_{q(\theta, \phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] \\ &= \mathsf{KL}(Q(\theta, \phi)||\Pi(\theta, \phi|\mathbf{z})) + \log \int_{\mathbb{S}} \frac{p(\mathbf{y}|\zeta)}{p(\mathbf{y}|\zeta_0)} d\Pi(\zeta|\mathbf{z}) + E_{q(\theta, \phi)} \left[\log \frac{p(\mathbf{y}|\zeta_0)}{p(\mathbf{y}|\mathbf{z}, \theta, \phi)} \right] \\ &\leq C_1 n \epsilon_n^2 \log n + C_2 n \epsilon_n^2 \log n + C_3 n \epsilon_n^2 \log n \\ &\leq C n \epsilon_n^2 \log n \end{split}$$

CHAPTER 4

BAYESIAN MODEL CALIBRATION FOR BETA DECAY CALCULATIONS



Figure 4.1 Calibration framework in β^{-} -decay calculation.

Accurate forecasts regarding the β -decay rates of neutron-rich nuclei are crucial for investigations of r-process studies. The recent development of the proton-neutron finite amplitude method (pnFAM) makes global β -decay studies feasible within the nuclear density functional theory. In pnFAM calculations with the Skyrme functional, time-odd terms, isoscalar pairing, and the axial-vector coupling strongly impact the results. However, they are not constrained by the properties of even-even nuclei. Thus, model calibration and the uncertainty quantification of β -decay predictions are necessary. In this chapter, we conduct Bayesian model calibration with selected experimental data and simulation data, and obtain the posterior distribution of parameters with reliable predictions. We will conduct various calibration schemes and compare the result with that of χ^2 optimization.

From a statistical point of view, this problem could be summarized as follows. Consider a physics model f_m which takes two inputs, t and θ . t is the known input for the model, and it is the number of protons Z and the number of neutrons N. The model also has another set of inputs called *calibration parameter* θ . This will consist of dimensionless Landau-Migdal parameters g'_0 , dimensionless Iso-scalar pairing strength v_0 , effective axial vector coupling (quenching effect) g_A . If we plug in all five values in f_m , then the physics model will spit out two outputs, GT resonance (GTR) energy, and the β decay rates. This relationship is also visualized in Figure 4.1.

The original framework proposed by Kennedy and O'Hagan (2001) can be computationally
demanding since it estimates the emulator f, discrepancy δ , and unknown parameters θ at simultaneously using the joint dataset d = (y, z). At the same time, this approach is flexible and generally yields a good fit to the experimental data at the expense of potential identifiability issue of the calibration parameters. For example, a poor fit of the emulator can be compensated by the discrepancy term which can lead to non-physical values of the calibration parameters while having a statistical model that fits well the training data Bayarri et al. (2009). We shall refer to this simultaneous estimation as the *integrated approach*.

In contrast to the integrated approach, one can fit the emulator first using only the computer model evaluations z and then find the posterior distribution of the unknown parameters using the experimental observations y. This two-stage estimation is known as the *modular approach* Bayarri et al. (2009). The modular approach tends to be computationally less challenging compared to the integrated approach as it does not require the evaluation of the joint likelihood $p(d|\theta)$. Additionally, while the fitted emulator obtained via the integrated approach could be somewhat inconsistent with the computer model, modular approach guarantees the performance of emulator since it is fitted using computer model evaluations only. From a Physical point of view, the modular approach approach appears more natural since one is often concerned with the performance of the emulator altogether with the estimation of the calibration parameters. Hence, we will compare and contrast these two different estimation approaches in the context of calibration with GTR energies and β -decay rates.

In addition, often model discrepancy term adds another level of complexity in terms of interpreting the results. Hence, we are going to consider a case where there is no model discrepancy term. To be specific, we are going to consider the following models: Integrated (*I*), Integrated without model discrepancy $(I - \delta)$, Modular (*M*), Modular without discrepancy $(M - \delta)$.

$$(I): y_i = f(t_i, \theta) + \delta(t_i) + \sigma \epsilon_i, \tag{4.1}$$

$$(I - \delta): y_i = f(t_i, \theta) + \sigma \epsilon_i, \qquad (4.2)$$

$$(M): y_i = \bar{f}(t_i, \theta) + \delta(t_i) + \sigma \epsilon_i, \qquad (4.3)$$

$$(M - \delta): y_i = f(t_i, \theta) + \sigma \epsilon_i, \qquad (4.4)$$

where \bar{f} represents a pretrained GP emulator using simulation data only.

4.1 Data Collection

4.1.1 Experimental data

As shown in Tables 4.1 and 4.2 below, there are two types of observables employed in the calibration: GTR energies E_{GTR} and β^- -decay half-lives $T_{1/2}$. The half-lives are given in ascending order, and their logarithms $\lg T_{1/2}$ (base 10) are adopted in the fit as the half-lives can vary by several orders of magnitude.

No.	Nucleus	E _{GTR}	Error	No.	Nucleus	$E_{\rm GTR}$	Error
1	²⁰⁸ Pb	15.6	0.2	3	⁹⁰ Zr	8.7	_
2	¹³² Sn	16.3	0.6	4	¹¹² Sn	8.94	0.25

Table 4.1 GTR energies and their experimental errors (in MeV) taken from Refs. Yasuda et al. (2018); Akimune et al. (1995); Gaarde et al. (1981); Pham et al. (1995) for the four nuclei selected in this work.

No.	Nucleus	$T_{1/2}$	Error	No.	Nucleus	$T_{1/2}$	Error
5	⁹⁸ Kr	0.043	0.004	18	¹⁶⁶ Gd	4.8	1.0
6	⁵⁸ Ti	0.058	0.009	19	¹⁵⁶ Nd	5.26	0.2
7	102 Sr	0.069	0.006	20	²⁰⁴ Pt	10.3	1.4
8	⁸² Zn	0.166	0.011	21	⁷⁴ Zn	95.6	1.2
9	⁴⁸ Ar	0.475	0.04	22	⁵² Ti	102	6
10	⁶⁰ Cr	0.49	0.01	23	¹⁸⁰ Yb	144	30
11	¹²⁶ Cd	0.515	0.017	24	114 Pd	145.2	3.6
12	114 Ru	0.54	0.03	25	²⁴² U	1008	30
13	¹³⁴ Sn	1.05	0.011	26	¹³⁴ Te	2508	48
14	¹⁵² Ce	1.4	0.2	27	⁹² Sr	9399.6	61.2
15	⁷⁸ Zn	1.47	0.15	28	¹⁵⁶ Sm	33840	720
16	⁷² Ni	1.57	0.05	29	²⁰⁰ Pt	45360	1080
17	⁹² Kr	1.84	0.008				

Table 4.2 β^- -decay half-lives and their experimental errors (in second) taken from National Nuclear Data Center for the 25 nuclei selected in this work. half-lives are listed in ascending order.

4.1.2 Simulation data

In the context of nuclear density functional theory (DFT), the finite amplitude method (FAM) provides an efficient solution for the (quasiparticle) random phase approximation (QRPA) Nakat-

sukasa et al. (2007); Avogadro and Nakatsukasa (2011). Recent advancements in the Skyrme proton-neutron FAM (PNFAM) have made it possible to compute charge-changing transitions in deformed nuclei and perform extensive calculations for β -decay rates Shafer et al. (2016); Mustonen et al. (2014); Mustonen and Engel (2016); Ney et al. (2020). In the PNFAM, the time-odd isovector Skyrme couplings, isoscalar pairing strength (v_0), and effective axial-vector coupling (g_A) have a strong impact on β -decay transitions. In the actual calculation, the dimensionless Landau-Migdal parameter (g'_0) was used instead of the time-odd isovector Skyrme couplings.

4.2 Extension of the calibration model to two observable types

In order to estimate all the unknown quantities including the calibration parameters using both GTR energies and β -decay half-lives, we extend the original framework described in Chapter 1 by modeling each observable type independently. Specifically, let y_1 denote a column vector of size n_1 representing the experimental observations of the 1st observable type and let z_1 denote a column vector of size m_1 consisting of the computer model evaluations of the 1st observable type. Overall, for the 1st observable type, we have $d_1 = (y_1, z_1)$. Similarly, for the 2nd observable type, let y_2 denote n_2 experimental observable type. Consequently, d_2 is the joint dataset for the 2nd observable type. As described in the section 4.1.1, the fit observables consist of 4 GTR energies and 25 β -decay half-lives. To emulate the the PNFAM with the Skyrme energy density functional, we generated 1000 data points for each data type using the Latin Hypercube Design (see the appendix/supplementary text for more details).

By denoting sets of unknown parameters for each data type as $\Psi_1 = (\theta_1, \phi_1, \sigma_1)$ and $\Psi_2 = (\theta_2, \phi_2, \sigma_2)$, we get the joint conditional data likelihood for the integrated approach as

$$\boldsymbol{d}_{1}, \boldsymbol{d}_{2} | \boldsymbol{\Psi}_{1}, \boldsymbol{\Psi}_{2} \sim N \left(\begin{pmatrix} \boldsymbol{0}_{1} \\ \boldsymbol{0}_{2} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{2} \end{pmatrix} \right), \tag{4.5}$$

where Σ_1 and Σ_2 are the covariance matrices consistent with the KOH model formulation in Chapter 2. The likelihood (4.5) can be used directly for the integrated approach's posterior sampling. Note that the calibration parameters generally don't need to be shared among the observable types. In

our case, however, $\theta_1 \cap \theta_2 \neq \emptyset$ as $\theta_1 = (g'_0, v_0)$ and $\theta_1 = (g'_0, v_0, g_A)$.

The joint conditional likelihood for the modular approach is given by the likelihood of experimental observations y_1 and y_2 as all the computer model outputs are used to train the emulator independently. Namely,

$$\mathbf{y}_{1}, \mathbf{y}_{2} | \tilde{\mathbf{\Psi}}_{1}, \tilde{\mathbf{\Psi}}_{2} \sim N \left(\begin{pmatrix} \hat{S}_{1}(\boldsymbol{\theta}_{1}) \\ \hat{S}_{2}(\boldsymbol{\theta}_{2}) \end{pmatrix}, \begin{pmatrix} K_{\delta_{1}}(T_{d_{1}}, T_{d_{1}}) + \sigma_{1}^{2}I_{n_{1}} & 0 \\ 0 & K_{\delta_{2}}(T_{d_{2}}, T_{d_{2}}) + \sigma_{2}^{2}I_{n_{2}} \end{pmatrix} \right)$$
(4.6)

where $\hat{S}_1(\theta_1)$ is the posterior mean of GP emulator at the design points $\{(\tilde{t}_{1,i})\}_{i=1}^{m_1}$ with calibration parameters θ_1 . The vector $\tilde{\Psi}_1$ consists of the GP hyperparameters for δ_1 , the scale of observational error for the 1st data type, and the calibration parameters θ_1 . $K_{\delta_1}(T_{d_1}, T_{d_1})$ is the matrix with elements $k_{\delta_1}(\tilde{t}_{1,i}, \tilde{t}_{1,j})$. The components of likelihood (4.6) corresponding to the 2nd data type are then defined analogically.

4.3 **Prior Structure**

It is often the case with Bayesian modeling that the careful elicitation of prior distributions is neglected. This is because if one has enough data, the effect of the prior becomes small Ghossal and Van der Vaart (2017). Thoughtful prior selection is very important in computer model calibration, because it can help mitigate the potential identifiability issue of calibration parameters. In particular, strong priors for GPs' hyperparameters are needed to distinguish between the systematic discrepancy δ and the observational error σ Brynjarsdóttir and O'Hagan (2014).

In order to improve the computational stability of MCMC sampler, we first reparameterize the GP length scales to the correlation hyperparameters. Specifically, for the length scales of GP emulator of $f(t, \theta)$, we set $\rho_i^{s_j} = exp(-1/4l_i^{s_j})$, $i = 1, \dots, 2 + q_j$ and j = 1, 2, where q_j is the dimension of calibration parameters in observable type j. Consider similar reparameterization for the length scale parameters for δ function. Since the original length scale parameters are positive real values, the MCMC sampling is slow since it has to sample potentially from all the real positive values. With this reparameterization, the sampling only needs to occur on the [0, 1] domain. Therefore, we consider a Beta distribution for the reparametrized length scales with shape parameters a = 1 and b = 0.3. This choice is informed by the fact that we want to check which of inputs are relevant for the outputs. For example, we found that among 4 parameters g'_1 could be removed from analysis based on previous sections. Similarly, in Bayesian analysis, since prior masses are clustered around 1 (length scales are very large and the effect of inputs is small), if an input is important, then we expect its posterior distribution moves toward 0.

For the precision parameters (inverse of scale parameters in covariance functions), first note that they have nice interpretations. They determine how much of the data's variance comes from each term (f, δ , or ϵ). Recall that all the outputs were standardized to have unit variance. Therefore, we specify the precision parameters of the covariance functions for f_1 and f_2 to be close to 1 by assigning them a Gamma prior with shape parameter $\alpha = 10$ and rate parameter $\beta = 10$. This leads to the expectation of each precision parameter for the covariance functions to be 1.

For the model discrepancy term, we expect to have a relatively small bias, which also helps to mitigate the potential confounding issues. This is because the computer emulator is expected to explain most of the variance in experimental observations. Hence, we set Gamma prior with shape parameter $\alpha = 10$ and rate parameter $\beta = 0.3$ for the precision parameters for the discrepancy term. This leads to the expectation of each precision parameter of the covariance functions for δ_1 and δ_2 to be 33. Hence the relative size of the covariances from the discrepancy term would be $3\%(=\frac{1}{33})$ of those from f_1 and f_2 .

Finally, for the observational error terms, we set Gamma prior with shape parameter $\alpha = 10$ and rate parameter $\beta = 0.0015$ for the observational error terms. By the same logic, the relative size of the covariances from the observational error term would be $0.015\%(=\frac{1}{6666.6667})$ of those from f_1 and f_2 . This hierarchical order of variances explained by each term is important to distinguish between discrepancy and observational error. In addition, the choice of prior on observational error terms is consistent with the χ^2 optimization results. More discussion of suitable priors on discrepancy terms can be found at Brynjarsdóttir and O'Hagan (2014) and Chong and Menberg (2018).

When it comes to the calibration parameters, we considered two cases. In the first case, we

consider a uniform prior on each parameter space $(g'_0 \in [0, 4], v_0 \in [-2, 0], g_A \in [-2, 0])$. In the second case, we consider weakly informative priors for the calibration parameters, which are taken to be Truncated Normal (TN) distributions centered around the empirical values $(g'_0 = 1.6, v_0^* = -1.1, g_A^* = -1)$ with small standard deviations truncated at the boundaries of each parameter space.

4.4 Calibration Results

The number of effective parameters is chosen based on the results of Li (2022). Therefore, we will consider two cases where g_A is free or fixed at value 1.

4.4.1 g_A free case

For the two different observable types (E_{GTR} , $T_{1/2}$), we fit the independent Gaussian Process with Matérn covariance class with roughness parameter 3/2, which shares the calibration parameters; (g'_0 , v_0) for GTR energies and (g'_0 , v_0 , g_A) for β^- -decay half-lives.

Figure 4.2 shows the posterior distribution of calibration parameters. The Figure's first row shows when using TN priors on the calibration parameters. MCMC results of 'I' and 'M' show multimodality in the posterior distribution of g'_0 . The VI method can not capture this multimodality because we predefined normal distribution as the variational family of the calibration parameters. However, the VI result is centered around the χ^2 result, a standard approach in the Physics community. The same thing holds for the models without discrepancy term.

For v_0 , VI approach well approximated MCMC posteriors, all are slightly off from the χ^2 result. The results without discrepancy term have smaller concentrations but more or less centered at the same values as posteriors of models with discrepancy term.

For g_A , while χ^2 result is around 0.5 area, MCMC posteriors without model discrepancy has moved slightly to the right. MCMC posteriors with model discrepancy center around a value 1.0. The approximated VI posterior is moved slightly right of the MCMC results. This might be the compensation from approximating the multimodality of the posterior of g'_0 with normal distribution.

The sensitivity of posterior distributions coming from using two different priors is minimal for MCMC methods. The second row of the figure shows the results when we use the uniform prior distribution on each parameter space. The posterior shape is more or less the same for g'_0 and v_0 ,

but the posterior distribution of g_A has more mass toward the value 0.5. For the VI, g'_0 has slightly moved to the left, while other parameters are more concentrated around the same values as in TN prior cases. The mean and standard deviation for each case are summarized in Table 4.3.

Methods	g'_0	v ₀	<i>g</i> _A
I (U)	1.653(0.104)	-0.941(0.064)	0.940(0.140)
I- δ (U)	1.584(0.037)	-0.916(0.080)	0.708(0.125)
M (U)	1.654(0.101)	-0.943(0.066)	0.925(0.148)
M- δ (U)	1.585(0.036)	-0.920(0.080)	0.705(0.123)
VI (U)	1.404(0.028)	-0.941(0.029)	1.089(0.049)
I (TN)	1.655(0.098)	-0.957(0.056)	0.975(0.097)
I- δ (TN)	1.590(0.036)	-0.950(0.073)	0.804(0.129)
M (TN)	1.658(0.098)	-0.958(0.058)	0.967(0.099)
M- δ (TN)	1.589(0.036)	-0.952(0.072)	0.800(0.124)
VI (TN)	1.600(0.049)	-0.955(0.045)	1.147(0.055)
χ^2 opt 1	1.592(0.034)	-1.197(0.179)	0.503(0.143)
I (U)	1.465(0.171)	-1.179(0.008)	1
I- δ (U)	2.383(0.014)	-1.802(0.191)	1
M (U)	1.456(0.142)	-1.519(0.008)	1
M- δ (U)	1.588(0.036)	-0.898(0.166)	1
VI (U)	1.419(0.038)	-0.778(0.010)	1
I (TN)	1.674(0.118)	-0.942(0.064)	1
I- δ (TN)	1.586(0.038)	-0.780(0.013)	1
M (TN)	1.568(0.158)	-1.119(0.099)	1
M- δ (TN)	1.589(0.036)	-0.922(0.163)	1
VI (TN)	1.422(0.040)	-0.778(0.009)	1
χ^2 opt 2	1.596(0.039)	-1(0.178)	1

Table 4.3 (Posterior summaries of calibration parameters using both observable types.) 'U' represents uniform prior and 'TN' represents truncated normal prior. We divided the case where g_A is either free or fixed at 1.

4.4.2 g_A fixed case

Figure 4.3 shows the posterior distribution of calibration parameters when g_A is fixed at 1. The posterior distributions of g'_0 behaves similar to the g_A free case. Posterior distributions of models with discrepancy term show multimodality, and posterior distributions of the models without discrepancy term center around χ^2 optimization results. VI results now have moved slightly to the left. On the other hand, posterior distributions of v_0 have different results for each model and for



Figure 4.2 (Posterior distribution of calibration parameters using both observable types with free g_A). The columns of the figure represent the posterior distribution of each calibration parameter. In addition, the upper three figures are posterior distributions using TN prior, and the lower three figures are those from a uniform prior. The purple solid lines represent the prior distributions (either uniform prior or TN prior) over each parameter space. Each line with different colors and different line types shows different estimation approaches; blue dashed lines for 'I', orange dotted lines for 'I- δ ', green dash-dotted lines for 'M', red dash-dot-dotted lines for 'M- δ ', and the brown solid line is for 'VI' approach we proposed in Chapter2. The black solid vertical line represents the χ^2 optimization results with two sigma error bands in black vertical dotted lines.

each prior distribution we used except for VI. Especially, $I - \delta$ with uniform prior case shows the posterior distribution pushed toward the boundary value near -2. This is very different from other results, and we tried to understand why this is happening.

One possible explanation could be found by looking at the χ^2 surface in Figure 4.4. It shows that the χ^2 has two regions where the χ^2 value is low. Hence we restricted the parameter space of v_0 to avoid those unphysical regions where v_0 is less than -2. However, by fitting the emulator and the observational error together ($I - \delta$ case), the posterior distribution of v_0 moved toward this unphysical area. This means that a good amount of proportion in the variance of the experimental was absorbed in the observational error. This fact could be confirmed by looking at the performance of the emulator using 100 GTR and 500 β -decay out of the sample simulation data set; Table 4.4. In both tables, the RMSE for GTR data is not much different whether one fixes g_A or not. However,



Figure 4.3 Posterior distribution of calibration parameters using both observable types with $g_A = 1$.



Figure 4.4 χ^2 surface for fixed g_A

RMSE for β data tends to be higher if we fix g_A . This is because the β decay data controls v_0 and g_A . In fact, if we look at the ' $I - \delta$ ' model with a uniform prior case, the emulator's performance on β decay data is significantly worse than other methods.

As we explained, the posterior distributions of observational error terms are quite different in $I - \delta'$ case. Figure 4.5 shows posterior distributions of hyperparameters for the $I - \delta'$ case with

Priors	(TN Prior)		(Unif Prior)		
Method	GTR RMSE	β^- RMSE	GTR RMSE	β^{-} RMSE	
Ι	0.057	0.707	0.056	0.707	
I-δ	0.057	0.703	0.057	0.703	
М	0.056	0.701	0.057	0.701	
$M-\delta$	0.057	0.700	0.057	0.701	
VI	5.085	1.011	5.086	1.011	
VI(mean)	0.033	0.538	0.033	0.535	
$I(g_A = 1)$	0.057	1.158	0.057	1.123	
I- δ ($g_A = 1$)	0.056	1.152	0.057	1.545	
M ($g_A = 1$)	0.057	1.130	0.057	1.165	
M- δ ($g_A = 1$)	0.057	1.151	0.057	1.150	
$VI (g_A = 1)$	6.368	7.765	6.525	7.746	
VI ($g_A = 1$, mean)	0.088	0.776	0.088	0.776	

Table 4.4 Average of RMSEs for 100 GTR and 500 β -decay out of sample computer outputs.

fixed g_A using two different priors. The three figures show the difference of posterior distributions in $1/\eta_{f_2}$, $1/\sigma_1^2$, and $1/\sigma_2^2$, while all the other hyperparameters have posterior distributions overlapping each other. This implies that assigning distinct priors to the calibration parameters had an impact on the learning of other parameters. This connection arises because various components, including the emulator, calibration parameters, and observational error, are interconnected within the framework of the '*I*' or '*I* – δ approach. The VI approach is also an approximation of MCMC in the framework of '*I*', and it shows poor performance of the emulator too. However, if we use the mean of variational posterior for prediction instead of full variational distribution, the performance of the emulator is quite comparable to MCMC methods.

This poor performance of the emulator could be avoided by training the emulator first ('M' or 'M – δ ' cases) using the simulation data. By doing so, we could avoid the confounding issue. Bayarri et al. (2009) also mentioned five reasons for using the modularization approach, which could be summarized as follows.

- 1. Keep a good module separate from a suspect module to avoid contamination.
- 2. Scientific understanding and development can require modularization.
- 3. Identifiability concerns or confounding might require modularization.

- 4. Mixing of MCMC analyses can greatly improve under modularization.
- 5. The computation is not otherwise possible.

In short, we have a certain expectation of the behavior of the emulator, and we would like to avoid the behavior of the emulator being contaminated by the experimental data by keeping the modules separate.



Figure 4.5 Posterior distribution of hyperparameters for fixed g_A

4.5 Prediction Results

In this section, we are going to compare the RMSEs for 179 out of the sample Beta decay data based on each model. Table 4.5 shows the performance of each model to predict 179 experimental data. There are two main features in the results. The first is that the model discrepancy term reduces the RMSEs, and the other is freeing g_A also reduces RMSEs. In addition, the VI approach shows significantly lower RMSE for fixed g_A . However, the Average RMSE for Bayesian approaches tends to be higher than the χ^2 optimization results. Then why should we adopt Bayesian calibration methods?

In Figure 4.7, we can check the advantage of the Bayesian calibration approach. The x-axis of the figure is the logarithm of beta decay rates (position of 179 nuclei in terms of beta decay rates), and the y-axis is the log-ratio of prediction over experimental values of beta decay rates. Therefore if the prediction is close to experimental values, the ratio becomes one, and by taking the logarithm, the value becomes zero. For this reason, the value zero on the y-axis is the benchmark and is represented as a dotted line.

The χ^2 results show that the prediction error is small overall. Hence the prediction interval can not capture the value zero. On the other hand, the MCMC prediction interval well captures

zero values, with similar behavior to the χ^2 results with accurate prediction in small values of beta decay rates and bad prediction at large values of beta decay rates. VI approach well approximated MCMC with smaller uncertainty. The prediction behavior for each model is very similar. For free g_A , the larger the x-value, the log ratio value tends to be small. This is due to the physical simulator, which can also be seen in χ^2 results. For fixed g_A , unlike χ^2 results, model prediction based on each model tends to systematically off from 0.

Methods	β^{-} RMSE (TN prior)	β^{-} RMSE (Unif prior)	
Ι	1.702	1.701	
Ι-δ	1.989	1.970	
Μ	1.646	1.641	
$M-\delta$	1.939	1.916	
VI	1.457	1.515	
χ^2 opt 1	0.829		
$I(g_A = 1)$	1.991	1.961	
I- δ ($g_A = 1$)	2.105	2.189	
$M(g_A = 1)$	1.893	1.882	
M- δ ($g_A = 1$)	2.124	2.119	
$\mathrm{VI}\left(g_{A}=1\right)$	1.810	1.807	
χ^2 opt 2 ($g_A = 1$)	0.9	974	

Table 4.5 Average of RMSEs for 179 Beta decay experimental data.

Another way to look at the advantage of Bayesian model calibration compared χ^2 results could be seen from Figure 4.8. In this figure, the x-axis represents the theoretical coverage of true values by controlling the length of the prediction interval, and the y-axis represents how many of those intervals actually include the value 0 among 179 prediction points (empirical coverage). We can clearly see that χ^2 prediction interval can not include the true values even if you increase the intervals. But all the Bayesian approach well captures the true values with increasing prediction intervals, either based on MCMC or VI.

The main difference between each model comes from the code's running time. The code's run time is 80 hours for '*I*' approach with MCMC, 20 hours for '*M*' and 3 hours for '*VI*'. Models without discrepancy term took slightly less time than those with discrepancy term.



Figure 4.6 Prediction of Beta decay rates of 179 nuclei with one standard deviation error bar for calibration models without discrepancy terms.

4.6 Conclusion and Discussion

In this work, we found that choosing an appropriate covariance function is of fundamental importance in modeling the physical process. In that sense, we would like to recommend Matérn class of covariance function instead of SE covariance all the time. Also, setting relatively strong priors in the hyperparameters of GPs is necessary to correctly distinguish the discrepancy term from observational error Brynjarsdóttir and O'Hagan (2014). In addition, instead of the naive application of the Bayesian calibration model, by keeping the modules separate, we could get a time advantage with similar results as the '*I*' approach with a reliable emulator performance. Exclusion of discrepancy term lead to a similar performance in terms of prediction but it could lead to biased results for calibration parameters. '*VI*' Approach gives us a fast and efficient result for calibration parameters, the researchers should be careful when it comes to interpreting calibration parameters or the performance of the emulator.



Figure 4.7 Prediction of Beta decay rates of 179 nuclei with one standard deviation error bar for calibration models with discrepancy terms.



Figure 4.8 Coverage of test data set using prediction intervals.

Lastly, further investigation of the relationship between χ^2 optimization and Bayesian calibration would be an interesting research topic. The current χ^2 result of RMSE is better than all the Bayesian approaches. Considering the fact that χ^2 is to minimize the scaled version of the squared difference between experimental data and simulation data, the performance of models without discrepancy is expected to be close to χ^2 results.

However, despite this expectation, the RMSE values for models without discrepancies are rather poor. One key reason for this discrepancy lies in the difference in how χ^2 and Bayesian approaches utilize the physics model. While χ^2 directly employs the physics model, Bayesian methods utilize a Gaussian Process (GP) emulator to approximate it. As a result, the Bayesian approach tends to perform worse than χ^2 .

In cases where the difference between the GP emulator and the physics model becomes negligible due to having sufficient simulation data, the models without discrepancies can closely approach χ^2 results. Furthermore, with an increase in experimental data, the Bayesian approach is expected to outperform χ^2 optimization. This is because Bayesian models converge to the true physical processes with increasing experimental data, as discussed in previous chapters, while χ^2 optimization remains in constrained world of physics models. Nevertheless, this offers a potential rationale for the present findings, and further investigation through theoretical analysis and extensive simulation studies is needed.

BIBLIOGRAPHY

- Akimune, H., Daito, I., Fujita, Y., Fujiwara, M., Greenfield, M. B., Harakeh, M. N., Inomata, T., Jänecke, J., Katori, K., Nakayama, S., Sakai, H., Sakemi, Y., Tanaka, M., and Yosoi, M. (1995). Direct proton decay from the Gamow-Teller resonance in ²⁰⁸Bi. *Phys. Rev. C*, 52(2):604–615.
- Avogadro, P. and Nakatsukasa, T. (2011). Finite amplitude method for the quasiparticle randomphase approximation. *Phys. Rev. C*, 84(1):014314.
- Balay, S., Abhyankar, S., Adams, M. F., Benson, S., Brown, J., Brune, P., Buschelman, K., Constantinescu, E., Dalcin, L., Dener, A., Eijkhout, V., Gropp, W. D., Hapla, V., Isaac, T., Jolivet, P., Karpeev, D., Kaushik, D., Knepley, M. G., Kong, F., Kruger, S., May, D. A., McInnes, L. C., Mills, R. T., Mitchell, L., Munson, T., Roman, J. E., Rupp, K., Sanan, P., Sarich, J., Smith, B. F., Zampini, S., Zhang, H., Zhang, H., and Zhang, J. (2021a). PETSc/TAO users manual. Technical Report ANL-21/39 Revision 3.16, Argonne National Laboratory.
- Balay, S., Abhyankar, S., Adams, M. F., Benson, S., Brown, J., Brune, P., Buschelman, K., Constantinescu, E. M., Dalcin, L., Dener, A., Eijkhout, V., Gropp, W. D., Hapla, V., Isaac, T., Jolivet, P., Karpeev, D., Kaushik, D., Knepley, M. G., Kong, F., Kruger, S., May, D. A., McInnes, L. C., Mills, R. T., Mitchell, L., Munson, T., Roman, J. E., Rupp, K., Sanan, P., Sarich, J., Smith, B. F., Zampini, S., Zhang, H., Zhang, H., and Zhang, J. (2021b). PETSc Web page.
- Balay, S., Gropp, W. D., McInnes, L. C., and Smith, B. F. (1997). Efficient management of parallelism in object oriented numerical software libraries. In Arge, E., Bruaset, A. M., and Langtangen, H. P., editors, *Mod. Softw. Tools Sci. Comput.*, pages 163–202. Birkhäuser Press.
- Bayarri, M. J., Berger, J. O., and Liu, F. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119 150.
- Brynjarsdóttir, J. and O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30:114007.
- Chong, A. and Menberg, K. (2018). Guidelines for the bayesian calibration of building energy models. *Energy and Buildings*, 174:527 547.
- Gaarde, C., Rapaport, J., Taddeucci, T. N., Goodman, C. D., Foster, C. C., Bainum, D. E., Goulding, C. A., Greenfield, M. B., Hören, D. J., and Sugarbaker, E. (1981). Excitation of giant spin-isospin multipole vibrations. *Nucl. Phys. A*, 369(2):258–280.
- Ghossal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.
- Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:425–464.

- Li, T. (2022). Application of density functional theory in nuclear structure. *Ph.D. thesis, Michigan State University*.
- Mustonen, M. T. and Engel, J. (2016). Global description of β^- decay in even-even nuclei with the axially-deformed skyrme finite-amplitude method. *Phys. Rev. C*, 93(1):014304.
- Mustonen, M. T., Shafer, T., Zenginerler, Z., and Engel, J. (2014). Finite-amplitude method for charge-changing transitions in axially deformed nuclei. *Phys. Rev. C*, 90(2):024308.
- Nakatsukasa, T., Inakura, T., and Yabana, K. (2007). Finite amplitude method for the solution of the random-phase approximation. *Phys. Rev. C*, 76(2):024318.
- Ney, E. M., Engel, J., Li, T., and Schunck, N. (2020). Global description of β^- decay with the axially deformed skyrme finite-amplitude method: Extension to odd-mass and odd-odd nuclei. *Phys. Rev. C*, 102(3):034326.
- Pham, K., Jänecke, J., Roberts, D. A., Harakeh, M. N., Berg, G. P. A., Chang, S., Liu, J., Stephenson, E. J., Davis, B. F., Akimune, H., and Fujiwara, M. (1995). Fragmentation and splitting of Gamow-Teller resonances in Sn(³He,t)Sb charge-exchange reactions, A=112–124. *Phys. Rev. C*, 51(2):526–540.
- Shafer, T., Engel, J., Fröhlich, C., McLaughlin, G. C., Mumpower, M., and Surman, R. (2016). β decay of deformed *r*-process nuclei near A = 80 and A = 160, including odd-A and odd-odd nuclei, with the skyrme finite-amplitude method. *Phys. Rev. C*, 94(5):055802.
- Wild, S. M. (2017). Chapter 40: POUNDERS in TAO: Solving Derivative-Free Nonlinear Least-Squares Problems with POUNDERS. In *Advances and Trends in Optimization with Engineering Applications*, MOS-SIAM Series on Optimization, pages 529–539. Society for Industrial and Applied Mathematics.
- Yasuda, J., Sasano, M., Zegers, R. G. T., Baba, H., Bazin, D., Chao, W., Dozono, M., Fukuda, N., Inabe, N., Isobe, T., Jhang, G., Kameda, D., Kaneko, M., Kisamori, K., Kobayashi, M., Kobayashi, N., Kobayashi, T., Koyama, S., Kondo, Y., Krasznahorkay, A. J., Kubo, T., Kubota, Y., Kurata-Nishimura, M., Lee, C. S., Lee, J. W., Matsuda, Y., Milman, E., Michimasa, S., Motobayashi, T., Muecher, D., Murakami, T., Nakamura, T., Nakatsuka, N., Ota, S., Otsu, H., Panin, V., Powell, W., Reichert, S., Sakaguchi, S., Sakai, H., Sako, M., Sato, H., Shimizu, Y., Shikata, M., Shimoura, S., Stuhl, L., Sumikama, T., Suzuki, H., Tangwancharoen, S., Takaki, M., Takeda, H., Tako, T., Togano, Y., Tokieda, H., Tsubota, J., Uesaka, T., Wakasa, T., Yako, K., Yoneda, K., and Zenihiro, J. (2018). Extraction of the Landau-Migdal parameter from the Gamow-Teller Giant Resonance in ¹³²Sn. *Phys. Rev. Lett.*, 121(13):132501.

APPENDIX A

LATIN HYPERCUBE DESIGN

For the Bayesian calibration model, we need a set of model evaluations for each observable type. In order to evaluate the model using a GP emulator, design points are required. Hence we used *Latin Hypercube Design* in the scaled input space of calibration parameters $[0, 1]^3$. The figures below represent the points we used for the evaluation of the computer model. An interesting point is that in the right Figure of A.1, the points in g_A are very regular compared to other parameters. This is because the calculation of a computer model with different g_A is cheap once one point is calculated.



Figure A.1 (Latin Hypercube Design points) Left figure is for GTR simulation and the right figure is for β decay simulation.

APPENDIX B

CHI-SQUARE OPTIMIZATION FRAMEWORK

The χ^2 optimization framework, based on Dobaczewski et al. (2014), seeks to determine the optimal model parameters by minimizing the weighted sum of squared residuals. In here, we briefly introduce the important results following the presentation of Li (2022).

$$\chi^{2}(\mathbf{x}) = \frac{\sum_{k=1}^{n_{d}} \varepsilon_{k}^{2}(\mathbf{x})}{n_{d} - n_{x}} = \frac{1}{n_{d} - n_{x}} \sum_{k=1}^{n_{d}} \left[\frac{s_{k}(\mathbf{x}) - d_{k}}{w_{k}} \right]^{2},$$
(B.1)

The objective function $\chi^2(\mathbf{x})$ represents the difference between model predictions $s_k(\mathbf{x})$ and experimental values d_k for each observable k, scaled by their respective weights w_k . The weight w_k is essential to make the residuals dimensionless, especially when dealing with multiple observable types.

The assumption is made that all weighted residuals ε_k are independent and follow a common normal distribution with a zero mean and a variance of σ^2 . Consequently, the χ^2 value at the optimal point \hat{x} is approximately equal to this variance σ^2 .

To meet this assumption, it becomes essential to select weights w_k that closely match the errors associated with the model predictions s_k , encompassing theoretical, numerical, and experimental uncertainties. This strategy is employed to bring the $\chi^2(\hat{x})$ close to a value of 1.

Although each data point can have its unique weight, it is common practice for data points of the same type to be assigned the same weight since their errors are expected to be similar. This ensures consistency in the treatment of similar observations within the optimization process.

The relative weights among different observable types are the key consideration because a global scale factor *s*, known as the Birge factor (Birge, 1932), can be introduced to scale the entire χ^2 value and ensure it approaches 1.

$$\chi^2(\hat{\boldsymbol{x}}) \to \tilde{\chi}^2(\hat{\boldsymbol{x}}) = \chi^2(\hat{\boldsymbol{x}})/s = 1, \ w_k \to \tilde{w}_k = w_k \sqrt{s}.$$
(B.2)

Then, for consistency between weights and residual distributions, the scaled weight for a given

observable type \tilde{w}_{typ} should be close to

$$r_{\rm typ} = \sqrt{\frac{n_d}{n_{\rm typ}(n_d - n_x)} \sum_{k \in \rm typ} [s_k(\hat{x}) - d_k]^2},$$
(B.3)

where n_{typ} is the number of points of the given type.

By utilizing the linear expansions of weighted residuals $\varepsilon_k(\mathbf{x})$ around the optimal point $\hat{\mathbf{x}}$, we can convert the original nonlinear optimization problem into a linear one, allowing us to apply the linear-regression framework. Here are some important conclusions. For more rigorous mathematical details, please refer to Seber (1989).

Let x^* be the true parameter vector. Then the difference $(\hat{x} - x^*)$ approximately follows a multivariate normal distribution: $\hat{x} - x^* \sim N(\vec{0}, \text{Cov}(\hat{x}))$. The covariance matrix is

$$\operatorname{Cov}(\hat{\boldsymbol{x}}) \approx \chi^2(\hat{\boldsymbol{x}}) \left[J^T(\hat{\boldsymbol{x}}) J(\hat{\boldsymbol{x}}) \right]^{-1}, \qquad (B.4)$$

where the $n_d \times n_x$ Jacobian matrix $J(\mathbf{x})$ is defined by $J_{kl} = \frac{\partial \varepsilon_k}{\partial x_l}$.

As for the optimization routine, we employ POUNDERS of Wild (2017) in PETSc/TAO toolkit, Balay et al. (1997, 2021a,b).

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, we have explored some theoretical properties of the original KOH model framework, specifically focusing on the posterior consistency of the estimated physical processes. This work was guided by the lack of theory backing up the method, which has been widely used in science and engineering problems with over 4000 citations.

We first started to look at the KOH model as a hierarchical GP regression model and adopted the extension of Schwartz theorem's test function framework to calculate the posterior contraction rate using true posterior. We further proposed a new variational algorithm to overcome the computational issue of the MCMC-based KOH model. In addition, we showed the posterior contraction rate of the proposed method under reasonable assumptions on GP prior and regularity conditions on the variational family. This was supported by simulation studies and the real data application in β -decay calculation. We have also examined a variety of issues concerning the implementation of both methods, MCMC and VB: Choosing covariance class for modeling the physical process, prior structure to mitigate the identifiability issue of calibration parameters, and modularizing the training of the emulator for a better understanding of the model and a better interpretation.

Our findings and discussions have significant implications of understanding the KOH model. The findings presented in this thesis will serve as a foundation for future theoretical investigations in this area. Other interesting directions of research from here could be summarized as follows.

Improving the performance of the emulator: The newly proposed algorithm in this thesis remains rooted in Gaussian Processes. As a consequence, the optimization process necessitated addressing the issue of nonpositive semi-definite covariance. However, there is potential to explore alternative avenues where advanced Machine Learning techniques, such as Deep learning models, diffusion models, or generative models, could replace Gaussian Processes. In particular, the use of Deep Gaussian processes for constructing emulators has gained attention from researchers recently. This presents a captivating and relevant topic for investigation, especially in the current era of AI,

enticing researchers from various domains.

Calibration parameters carrying physical meaning: An alternative avenue of investigation pertains to the calibration parameters. In the original KOH model, all the unknowns are estimated simultaneously, which raises concerns about the identifiability of calibration parameters. Therefore, an intriguing project would involve establishing a novel definition for calibration parameters that not only ensures identifiability but also preserves their physical significance. Interested people may explore Li and Xiong's (2022) fascinating method of decomposing the discrepancy term through Taylor expansion.

Theoretical investigation of Modular approach: The primary focus of the original KOH model is to excel in handling experimental data. Nevertheless, domain scientists also place significant emphasis on the emulator. Therefore, maintaining a separate emulator module by training it independently could prove beneficial in enhancing the comprehension and interpretation of the model. Demonstrating theoretical backing for the Modular approach presents an engaging path for future research.