Supplemental Document for Patient-Specific Prediction of Abdominal Aortic Aneurysm Expansion using Bayesian Calibration

Liangliang Zhang[1], Zhenxiang Jiang[2], Jongeun Choi[3], Chae Young Lim[4], Tapabrata Maiti[1] and Seungik Baek[2]

[1]Department of Statistics & Probability, Michigan State University, East Lansing, Michigan, 48824-1027
E-mail address, L. Zhang: liangliangzhang.stat@gmail.com

[2]Mechanical Engineering, Michigan State University, East Lansing, Michigan, 48824-1027

[3]School of Mechanical Engineering, Yonsei University, Seoul, Republic of Korea
E-mail address, J. Choi: jongeunchoi@yonsei.ac.kr

[4]Department of Statistics, Seoul National University, Seoul, South Korea

S1 Methods

S1.1 Statistical models

Let $\mathcal{GP}(\boldsymbol{m}(\cdot), \boldsymbol{k}(\cdot,\cdot))$ be the Gaussian process with the mean function $\boldsymbol{m}(\cdot)$ and the covariance function $\boldsymbol{k}(\cdot,\cdot)$. $\mathcal{GP}$ is flexible and popularly used as a prior model for functions [1;2;3;4]. We introduce the following Gaussian processes as prior beliefs for the G&R computation model and the model error:

$$\begin{aligned} r(\boldsymbol{x}, \boldsymbol{\theta}) &\sim \mathcal{GP}(\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{k}_1(\boldsymbol{x}, \boldsymbol{\theta}; \boldsymbol{x}', \boldsymbol{\theta}')), \\ \delta(\boldsymbol{x}) &\sim \mathcal{GP}(\boldsymbol{m}_2(\boldsymbol{x}), \boldsymbol{k}_2(\boldsymbol{x}, \boldsymbol{x}')). \end{aligned} \tag{1}$$

To specify Gaussian process priors in (1) further, we introduce mean and covariance structures for both processes.

For a mean structure, we use a linear combination of basis functions to form the general mean structure. Let $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})$ be a vector of basis functions and $\boldsymbol{\beta}$ be a vector of corresponding coefficients such that $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}^T$, where $(\cdot)^T$ is the transpose of a matrix (or a vector). Since we need two mean functions, $\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta})$ and $\boldsymbol{m}_2(\boldsymbol{x})$, we introduce two sets of $\boldsymbol{h}$ and $\boldsymbol{\beta}$ such that $\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}_1^T$ and $\boldsymbol{m}_2(\boldsymbol{x}) = \boldsymbol{h}_2(\boldsymbol{x})\boldsymbol{\beta}_2^T$. In this paper, we consider the following linear mean structures for computational efficiency:

$$\boldsymbol{m}_1(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{\beta}_1^T = \beta_{10} + \beta_{11}t + \beta_{12}\theta_1 + \beta_{13}\theta_2, \tag{2}$$

where $\boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta}) = [1\ t\ \theta_1\ \theta_2]$ and $\boldsymbol{\beta}_1 = [\beta_{10}\ \beta_{11}\ \beta_{12}\ \beta_{13}]$, and

$$\boldsymbol{m}_2(\boldsymbol{x}) = \boldsymbol{h}_2(\boldsymbol{x})\boldsymbol{\beta}_2^T = \beta_{21}t + \beta_{22}s, \tag{3}$$

where $\boldsymbol{h}_2(\boldsymbol{x}) = [t\ s]$ and $\boldsymbol{\beta}_2 = [\beta_{21}\ \beta_{22}]$. These mean structures imply that the mean function of the G&R computation model is linear in time t and calibration parameters, $\{\theta_1, \theta_2\}$. The mean function of the model error is linear in time t and location s. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1\ \boldsymbol{\beta}_2]$ are hyperparameters for the mean functions in a Bayesian context.

For a covariance structure, we use the following exponential functions as follows.

$$\begin{aligned} \boldsymbol{k}_1(\boldsymbol{x}, \boldsymbol{\theta}; \boldsymbol{x}', \boldsymbol{\theta}') =& \sigma_1^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{\Omega}_x(\boldsymbol{x} - \boldsymbol{x}')^T\} \\ & \times \exp\{-(\boldsymbol{\theta} - \boldsymbol{\theta}')\boldsymbol{\Omega}_\theta(\boldsymbol{\theta} - \boldsymbol{\theta}')^T\}, \\ \boldsymbol{k}_2(\boldsymbol{x}, \boldsymbol{x}') =& \sigma_2^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{\Omega}_x^\star(\boldsymbol{x} - \boldsymbol{x}')^T\}, \end{aligned}$$

where $\boldsymbol{\Omega}_x$, $\boldsymbol{\Omega}_\theta$, $\boldsymbol{\Omega}_x^\star$ are diagonal matrices such that $\boldsymbol{\Omega}_x = \text{diag}(\omega_{x1}, \omega_{x2})$, $\boldsymbol{\Omega}_\theta = \text{diag}(\omega_{\theta 1}, \omega_{\theta 2})$, $\boldsymbol{\Omega}_x^\star = \text{diag}(\omega_{x1}^\star, \omega_{x2}^\star)$. The hyperparameters for the covariance functions are σ_1^2, σ_2^2 and ω_{xj}, $\omega_{\theta j}$, $\omega_{xj}^\star$'s. Note that the resulting covariance functions are not isotropic since they allow different scaling for each coordinate. $k1$ is a multiplicative kernel. The magnitude of correlation is greater when either x or theta are closer.

For the parameters controling the covariances, we introduce $\boldsymbol{\psi} = [\boldsymbol{\psi}_1\ \boldsymbol{\psi}_2]$, with

$$\boldsymbol{\psi}_1 = [\omega_{x1}\ \omega_{x2}\ \omega_{\theta 1}\ \omega_{\theta 2}\ \sigma_1^2],\ \boldsymbol{\psi}_2 = [\omega_{x1}^\star\ \omega_{x2}^\star\ \sigma_2^2]. \tag{4}$$

$\boldsymbol{\psi}_1$ is the set of hyperparameters related to the G&R computation model, i.e., $r(\cdot,\cdot)$ in (1). $\boldsymbol{\psi}_2$ is the set of hyperparameters for the model error, i.e., $\delta(\cdot)$ in (1). For the remainder of the paper, the AAA G&R computation model is simply referred to as the computation model.

S2 Bayesian Analysis for Calibration

Our approach considers computation model outputs as the data in addition to the observations, and then combines these two to calibrate $\boldsymbol{\theta}$ in a Bayesian calibration framework.

S2.1 Likelihood of computation model

We need computation model outputs generated at various sets of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ as the data for calibration. To avoid confusion, we use $\boldsymbol{\theta}^* = [\theta_1^* \ \theta_2^*]$ and $\boldsymbol{x}^*$ instead of $\boldsymbol{\theta}$ and $\boldsymbol{x}$, respectively, when they were used to generate computation model outputs. For example, at $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$, we obtain one computation model output $r(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$. Computation model outputs which correspond to a set of various $(\boldsymbol{x}^*, \boldsymbol{\theta}^*)$ will be used in Bayesian calibration as a part of the data set. We call $\boldsymbol{\theta}^*$ calibration inputs and $\boldsymbol{x}^*$ variable inputs.
Regarding the computation model outputs, let N be the total number of pairs of variable inputs and calibration inputs. For the i-th set of inputs, $(\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*)$, we then let y_i be the corresponding computation model output, i.e., $y_i = r(\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*)$. We further define the computation model output vector as $\boldsymbol{y} = [y_1 \ \cdots \ y_N]^T \in \mathbb{R}^{N\times 1}$ and the corresponding input matrix as

$$X_c = [(\boldsymbol{x}_1^*, \boldsymbol{\theta}_1^*)^T, \cdots, (\boldsymbol{x}_N^*, \boldsymbol{\theta}_N^*)^T]^T \in \mathbb{R}^{N\times 4}.$$

Gaussian process prior on the computation model then gives $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{H}_1(X_c)\boldsymbol{\beta}_1^T, \boldsymbol{V}_1(X_c))$, where $\boldsymbol{H}_1(X_c) = [\boldsymbol{h}_1(\boldsymbol{x}_1^*, \boldsymbol{\theta}_1^*)^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_N^*, \boldsymbol{\theta}_N^*)^T]^T$ and the (i,j) entry of $\boldsymbol{V}_1(X_c)$ is $\boldsymbol{k}_1((\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*), (\boldsymbol{x}_j^*, \boldsymbol{\theta}_j^*))$.

S2.2 Likelihood of real observations

Let n be the number of observations and $\boldsymbol{z} = [z_1 \ \cdots \ z_n]^T \in \mathbb{R}^{n\times 1}$ be the set of observations corresponding to the variable input matrix $X_o = [\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T]^T \in \mathbb{R}^{n\times 2}$. Note that X_o is not necessarily the same as the set of variable inputs for the computation model outputs. In general, the number of observations is smaller than that of the computation model outputs since we can control the amount of the computation model outputs. To calibrate $\boldsymbol{\theta}$ from the observations, we augment variable inputs X_o with $\boldsymbol{\theta}$ such that $X_o(\boldsymbol{\theta}) = [(\boldsymbol{x}_1, \boldsymbol{\theta})^T, \cdots, (\boldsymbol{x}_n, \boldsymbol{\theta})^T]^T$. From the calibration model, we then have

$$\boldsymbol{z} \sim \mathcal{N}\Big(\boldsymbol{H}_1(X_o(\boldsymbol{\theta}))\boldsymbol{\beta}_1^T + \boldsymbol{H}_2(X_o)\boldsymbol{\beta}_2^T, \lambda\boldsymbol{I}_n + \boldsymbol{V}_1(X_o(\boldsymbol{\theta})) + \boldsymbol{V}_2(X_o)\Big),$$

where $\boldsymbol{H}_1(X_o(\boldsymbol{\theta})) = [\boldsymbol{h}_1(x_1, \boldsymbol{\theta})^T, \cdots, \boldsymbol{h}_1(\boldsymbol{x}_n, \boldsymbol{\theta})^T]^T$ and $\boldsymbol{H}_2(X_o) = [\boldsymbol{h}_2(x_1)^T, \cdots, \boldsymbol{h}_2(\boldsymbol{x}_n)^T]^T$. $\boldsymbol{V}_1(X_o(\boldsymbol{\theta}))$ is defined in a similar way to define $\boldsymbol{V}_1(X_c)$. The (i,j) entry of $\boldsymbol{V}_2(X_o)$ is $\boldsymbol{k}_2(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

S2.3 Joint likelihood

We combine the computation model outputs and observations, $\boldsymbol{d} = [\boldsymbol{y}^T\ \boldsymbol{z}^T]^T \in \mathbb{R}^{(N+n)\times 1}$, which we call a data vector.

$$\boldsymbol{d} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{m}_d(\boldsymbol{\theta}), \boldsymbol{V}_d(\boldsymbol{\theta})), \tag{5}$$

where $\boldsymbol{m}_d(\boldsymbol{\theta}) := \mathbb{E}(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{\beta}^T$, with

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{H}_1(X_c) & \boldsymbol{0} \\ \boldsymbol{H}_1(X_o(\boldsymbol{\theta})) & \boldsymbol{H}_2(X_o) \end{pmatrix}, \quad \text{and}$$

$$\begin{aligned} \boldsymbol{V}_d(\boldsymbol{\theta}) &: = \text{var}(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi}) \\ &= \begin{pmatrix} \boldsymbol{V}_1(X_c) & \boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta}))^T \\ \boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta})) & \lambda \boldsymbol{I}_n + \boldsymbol{V}_1(X_o(\boldsymbol{\theta})) + \boldsymbol{V}_2(X_o) \end{pmatrix}, \end{aligned} \tag{6}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{C}_1(X_c, X_o(\boldsymbol{\theta}))$ is a cross-covariance matrix whose (i, j) entry is $\boldsymbol{k}_1((\boldsymbol{x}_i^*, \boldsymbol{\theta}_i^*), (\boldsymbol{x}_j, \boldsymbol{\theta}_j))$. We note that $\boldsymbol{m}_d(\boldsymbol{\theta})$ and $\boldsymbol{V}_d(\boldsymbol{\theta})$ also depend on $\boldsymbol{\beta}$, λ and $\boldsymbol{\psi}$. We drop them to reduce notational complexity.

S2.4 Calibration

To estimate $(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi})$ under the Bayesian framework, we consider the following prior distributions and assumptions.
A.1 $\boldsymbol{\beta}_1$ in (2) and $\boldsymbol{\beta}_2$ in (3) have non-informative priors, i.e., $p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto 1$.
A.2 $\boldsymbol{\theta}$ is independent of the other parameters.
A.3 $\boldsymbol{\theta}$ follows a normal distribution.
A.4 $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ in (4) follow lognormal distributions.
A.5 $\log(\lambda)$ has a non-informative prior, i.e., $p(\log(\lambda)) \propto 1$. Note that based on **A.1** and **A.2**, we get the joint prior distribution in the following form $p(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi}) \propto p(\boldsymbol{\theta})p(\lambda)p(\boldsymbol{\psi})$. For **A.3**, we use the sample mean and variance of the calibration parameter inputs that were used to generate computation model outputs as mean and variance for normal prior density. **A.4** guarantees that $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ in (4) are positive values in the calculation since they are all hyperparameters in covariance functions. Together with the prior specification given above, we have the following full joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi})$ given $\boldsymbol{d}$:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi} \mid \boldsymbol{d}) &\propto p(\boldsymbol{\theta})p(\lambda)p(\boldsymbol{\psi})f(\boldsymbol{d}; \boldsymbol{m}_d(\boldsymbol{\theta}), \boldsymbol{V}_d(\boldsymbol{\theta})) \\ &\propto p(\boldsymbol{\theta})p(\lambda)p(\boldsymbol{\psi})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\{(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))\}\right], \end{aligned} \tag{7}$$

where $f(\boldsymbol{d}; \boldsymbol{m}, V)$ is the multivariate Gaussian density function with mean m and variance V. Note that the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\psi})$ given $\boldsymbol{d}$ is proper even if we assume non-informative priors for $\boldsymbol{\beta}$.

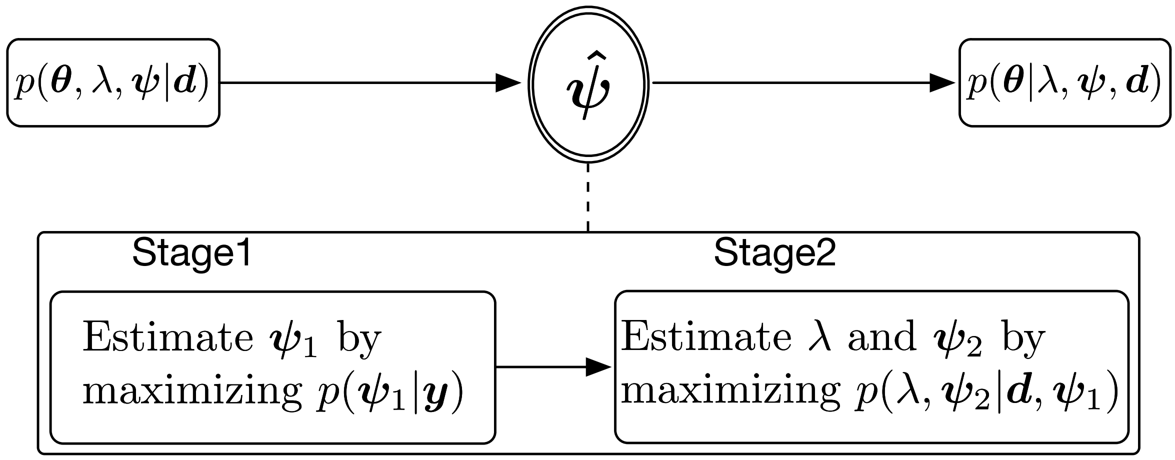


Figure S1: Two stages to estimate λ and $\boldsymbol{\psi}$. $\boldsymbol{\psi}_1$ describes covariance hyperparameters in computational model, so it can be estimated using the computational output $\boldsymbol{y}$.

To calibrate (estimate) $\boldsymbol{\theta}$ from the posterior distribution of $\boldsymbol{\theta} \mid \boldsymbol{d}$, we need to integrate out $\boldsymbol{\beta}$, λ and $\boldsymbol{\psi}$. As shown explicitly in (5), $\boldsymbol{m}_d(\boldsymbol{\theta})$ depends on $\boldsymbol{\beta}$, while $\boldsymbol{V}_d(\boldsymbol{\theta})$ depends on λ and $\boldsymbol{\psi}$. Clearly, $\boldsymbol{m}_d(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\beta}$, so we can find

$$\boldsymbol{\beta} \mid \boldsymbol{\theta}, \lambda, \boldsymbol{\psi}, \boldsymbol{d} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{W}(\boldsymbol{\theta})), \tag{8}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{d}$, $\boldsymbol{W}(\boldsymbol{\theta}) = (\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{H}(\boldsymbol{\theta}))^{-1}$. Thus, we can integrate out $\boldsymbol{\beta}$ in (7) and get

$$p(\boldsymbol{\theta}, \lambda, \boldsymbol{\psi} \mid \boldsymbol{d})$$
$$\propto p(\boldsymbol{\theta})p(\lambda)p(\boldsymbol{\psi})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}}|\boldsymbol{W}(\boldsymbol{\theta})|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\{(\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\}\right].$$

While $\boldsymbol{\beta}$ is integrated out easily, λ and $\boldsymbol{\psi}$ are not. Numerical integration can be considered but may not give an accurate result since it is the multi-dimensional integration. Thus, we use the conditional posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{d}$ and a plausible estimate of λ and $\boldsymbol{\psi}$ plugged in, as proposed by Kennedy and O'Hagan (2001)[5]. The plausible estimate of λ and $\boldsymbol{\psi}$ is obtained in a two-step approach. First, we estimate $\boldsymbol{\psi}_1$ by maximizing $p(\boldsymbol{\psi}_1|\boldsymbol{y})$ (Stage 1). Then, we estimate λ and $\boldsymbol{\psi}_2$ by maximizing $p(\lambda, \boldsymbol{\psi}_2|\boldsymbol{d}, \boldsymbol{\psi}_1)$ after plugging in the estimated $\boldsymbol{\psi}_1$ (Stage 2). In this study, numerical maximization is used with 500 iterations for Stage 1 and 100 iterations for Stage 2.
Having estimated λ and $\boldsymbol{\psi}$ and plugging it into the distribution of $\boldsymbol{\theta} \mid \boldsymbol{d}, \boldsymbol{\psi}$, we obtain the posterior distribution of the calibration parameter $\boldsymbol{\theta}$ to be

$$\begin{aligned} p(\boldsymbol{\theta} \mid \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}) &\propto p(\boldsymbol{\theta})|\boldsymbol{V}_d(\boldsymbol{\theta})|^{-\frac{1}{2}}|\boldsymbol{W}(\boldsymbol{\theta})|^{\frac{1}{2}} \\ &\times \exp\left[-\frac{1}{2}\{(\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\}\right]. \end{aligned} \tag{9}$$

and use (9) to calibrate and make an inference of $\boldsymbol{\theta}$.

S2.5 Prediction

We introduce variable inputs $X_p = [\boldsymbol{x}_{n+1}^T, \cdots, \boldsymbol{x}_{n+m}^T]^T \in \mathbb{R}^{m\times 2}$ to be used for prediction. The corresponding prediction of the QoI at given variable inputs X_p is denoted as $\boldsymbol{P} = [P_1 \ \cdots \ P_m]^T \in \mathbb{R}^{m\times 1}$. For quality of the prediction, we can consider variable inputs in X_c that cover X_p.
Under the Bayesian framework, a prediction can be made using the predictive distribution of the (unobserved) true process $\zeta(\boldsymbol{x})$ given full data $\boldsymbol{d}$. More specifically, we consider the predictive distribution of $\zeta(\boldsymbol{x})$ given $\boldsymbol{d}$ and $\hat{\boldsymbol{\psi}}$ since we fix λ and $\boldsymbol{\psi}$ using the estimate $\hat{\lambda}$ and $\hat{\boldsymbol{\psi}}$ in our approach. We first analytically obtain the distribution of $\zeta(\boldsymbol{x}) \mid \boldsymbol{\theta}, \lambda, \boldsymbol{\psi}, \boldsymbol{d}, \boldsymbol{\beta}$, which is conditionally normal since $\zeta(\boldsymbol{x})$ and $\boldsymbol{d}$ are normally distributed. Using the laws of total expectation and total variance to integrate out $\boldsymbol{\beta}$, we obtain the distribution of $\zeta(\cdot)$ conditional on $\boldsymbol{\theta}$, $\hat{\lambda}$ and $\hat{\boldsymbol{\psi}}$, which is also normal. Therefore, its mean function is given by

$$\mathbb{E}(\zeta(\boldsymbol{x})|\boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})), \tag{10}$$

where

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta}) \\ \boldsymbol{h}_2(\boldsymbol{x}) \end{pmatrix},$$

and

$$\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{V}_1((\boldsymbol{x}, \boldsymbol{\theta}), X_o) \\ \boldsymbol{V}_1\{(\boldsymbol{x}, \boldsymbol{\theta}), X_o(\boldsymbol{\theta})\} + \boldsymbol{V}_2(\boldsymbol{x}, X_o) \end{pmatrix}.$$

Additionally, its covariance function is given by

$$\begin{aligned} \mathrm{cov}(\zeta(\boldsymbol{x}), \zeta(\boldsymbol{x}') \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}) = \ & k_1((\boldsymbol{x}, \boldsymbol{\theta}), (\boldsymbol{x}', \boldsymbol{\theta})) + k_2(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} \boldsymbol{v}(\boldsymbol{x}', \boldsymbol{\theta}) \\ + (\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta}))^T & \boldsymbol{W}(\boldsymbol{\theta})(\boldsymbol{h}(\boldsymbol{x}', \boldsymbol{\theta}) \boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} \boldsymbol{v}(\boldsymbol{x}', \boldsymbol{\theta})). \end{aligned} \tag{11}$$

We then can obtain the predictive distribution of $\zeta(\boldsymbol{x})$ given $\boldsymbol{d}$, $\hat{\lambda}$ and $\hat{\boldsymbol{\psi}}$ by integrating $\boldsymbol{\theta}$ out from the distribution of $\zeta(\boldsymbol{x}) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}$ with respect to the posterior distribution of $\boldsymbol{\theta}$ given in (9). From (10) and (11), we obtain the predictive expectation and variance of $\zeta(\cdot)$ evaluated at inputs X_p as follows:

$$\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}] = \mathbb{E}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\}, \tag{12}$$

and

$$\mathrm{var}[\zeta(X_p) \mid \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}] = \mathbb{E}_{\boldsymbol{\theta}}\{\mathrm{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} + \mathrm{var}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\}, \tag{13}$$

where $\boldsymbol{\zeta}(X_p) = [\zeta(\boldsymbol{x}_{n+1}) \ \cdots \ \zeta(\boldsymbol{x}_{n+m})]^T$.
Based on the posterior density (9) of the calibration parameter $\boldsymbol{\theta}$, we can draw Markov Chain Monte Carlo (MCMC) samples for $\boldsymbol{\theta}$ and thus predictive expectation and variance (12) and (13) can be approximated as follows. First, we draw M samples $\{\boldsymbol{\theta}(1), \cdots, \boldsymbol{\theta}(M)\}$

from (9) using the Metropolis-Hastings algorithm[6] and plug these samples into (10) and (11). Next, we get the predictions $\boldsymbol{P}$, which are the MCMC estimates of (12) written as

$$\boldsymbol{P} := \widehat{\mathbb{E}}[\boldsymbol{\zeta}(X_p) \mid \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}] = \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(i), \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}].$$

Similarly, the outermost expectation and variance on the right hand side of (13) are estimated by their corresponding sample mean and sample variance with respect to $\boldsymbol{\theta}$. Then, we obtain the predictive variances $\boldsymbol{V}_p$ as

$$\boldsymbol{V}_p : = \widehat{\text{var}}[\boldsymbol{\zeta}(X_p) \mid \hat{\lambda}, \hat{\psi}, \boldsymbol{d}] = \widehat{\mathbb{E}}_{\boldsymbol{\theta}}\{\text{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} + \widehat{\text{var}}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\},$$

where

$$\widehat{\mathbb{E}}_{\boldsymbol{\theta}}\{\text{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} = \frac{1}{M}\sum_{i=1}^{M} \text{var}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(\boldsymbol{i}), \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}],$$

$$\widehat{\text{var}}_{\boldsymbol{\theta}}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}, \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\} = \frac{1}{M-1}\sum_{i=1}^{M}\{\mathbb{E}[\boldsymbol{\zeta}(X_p) \mid \boldsymbol{\theta}(i), \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}] - \widehat{\mathbb{E}}[\boldsymbol{\zeta}(X_p) \mid \hat{\lambda}, \hat{\boldsymbol{\psi}}, \boldsymbol{d}]\}^2.$$

We will construct the 95% credible intervals based on these predictions $\boldsymbol{P}$ and their predictive variances $\boldsymbol{V}_p$.

Posterior is proper

A proper density in Bayesian analysis implies that the density is integrable. In equation 7, the joint mean $\boldsymbol{m_d}(\boldsymbol{\theta})$ is a linear combination of $\boldsymbol{\beta}$ in the Gaussian function. When the prior of $\boldsymbol{\beta}$ is noninformative, one can show the posterior of $\boldsymbol{\beta}$ follows a normal distribution as shown in equation 8, which implies that the posterior of $\boldsymbol{\beta}$ is proper. Therefore, the posterior density of linear coefficients in the Gaussian mean function is proper when the prior density of these coefficients is noninformative. This proposition can be illustrated through a simple example as follows.
Consider a standard linear regression problem

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\epsilon \sim \mathcal{N}(0, I_n)$. This corresponds to the following likelihood function:

$$p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})^T(\boldsymbol{Y} - \boldsymbol{X\beta})\right).$$

We assume that $\boldsymbol{\beta}$ follows a noninformative prior, i.e. $\pi(\boldsymbol{\beta}) \propto 1$. This means the prior density is improper. However, the posterior of $\boldsymbol{\beta}$ can be derived as

$$p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{X}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y})^T(\boldsymbol{X}^T\boldsymbol{X})(\boldsymbol{\beta} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y})\right),$$

which implies that the posterior of $\boldsymbol{\beta}$ follows a normal distribution

$$p(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{X}) = \mathcal{N}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y},(\boldsymbol{X}^T\boldsymbol{X})^{-1}),$$

so that it is proper.

After integrating out $\boldsymbol{\beta}$, the posterior distribution of $(\boldsymbol{\theta},\lambda,\boldsymbol{\psi})$ given $\boldsymbol{d}$ becomes the multiplication of a noninformative prior ($p(\log(\lambda)) \propto 1$ which means $p(\lambda) \propto \frac{1}{\lambda}I_{\lambda>0}$), proper priors ($p(\boldsymbol{\theta})$ and $p(\boldsymbol{\psi})$ are assumed in A.1-A.5) and proper likelihood (multivariate Gaussian density $f(\boldsymbol{d};\boldsymbol{m},\boldsymbol{V})$). So the posteriors of $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\psi})$ are proper as well. λ is a hyperparameter in the covariance function of a multivariate Gaussian density $f(\boldsymbol{d};\boldsymbol{m},\boldsymbol{V})$. Then we conclude that the posterior of λ is proper, too. This proposition can be illustrated through a simple example as follows. Consider the Gaussian likelihood function with the variance parameter λ

$$p(d|\lambda) = \frac{1}{\sqrt{2\pi\lambda}}\exp\left(-\frac{d^2}{\lambda}\right),\ \lambda > 0.$$

The prior density is assumed as $\pi(\lambda) \propto \frac{1}{\lambda}$. Then the posterior of λ is proper, since $\int_0^{+\infty}\frac{1}{\lambda^{1+1/2}}\exp(-\frac{1}{\lambda}) = \frac{1}{\Gamma(\frac{1}{2})}$ (the integration of inverse gamma function).

S3 Programming steps for simulated observations

Fig. 6 shows us the whole process of Bayesian analysis which combines the information of priors, computer model, statistical model and observations to calibrate the parameter $\boldsymbol{\theta}$ to make predictions. To be more specific, the programming steps for the simulation study are given as follows. The sections, tables, equations referred here are from the main paper.

Step 0: Specify the statistical models for $\boldsymbol{m}_1(\boldsymbol{x},\boldsymbol{\theta})$ and $\boldsymbol{m}_2(\boldsymbol{x})$ (See Sections 2.(d)).

Step 1: Set the initials including true values of $\boldsymbol{\psi}_2$ and $\boldsymbol{\beta}_2$, normal priors of $\boldsymbol{\theta}$, log-normal priors of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. See prior assumptions in Section 3.

Step 2: Determine inputs X_c, X_o and X_p as described in Table 2.

Step 3: Simulate the full data $\boldsymbol{d} = [\boldsymbol{y}^T\ \boldsymbol{z}^T]^T$ by using X_c, X_o and true values of $\boldsymbol{\psi}_2$ and $\boldsymbol{\beta}_2$ as described in Section 4.

Step 4: Run Stage 1 with 500 iterations to obtain the estimate $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$. See Section 3.

Step 5: Run Stage 2 with 100 iterations to obtain the estimate $\hat{\boldsymbol{\psi}}_{\mathbf{2}}$. See Section 3.

Step 6: Draw samples of $\boldsymbol{\theta}$ from the posterior distribution $p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\psi}},\boldsymbol{d})$ with estimated $\hat{\boldsymbol{\psi}} = [\hat{\boldsymbol{\psi}}_{\mathbf{1}}^T,\hat{\boldsymbol{\psi}}_{\mathbf{2}}^T]^T$ plugged in. See Section 3.

Step 7: Plug samples of $\boldsymbol{\theta}$, estimated $\boldsymbol{\psi}$, full data $\boldsymbol{d}$ and X_p into (3.10) and (3.11) to get predictions and their corresponding variances. See Section 3.

S4 Results

Table S1: True values of the maximum diameter at different years in simulated observation cases

Year	Case 1a	Case 1b	Year	Case 2a	Case 2b	Case 3a	Case 3b
3.5	3.3556	3.3752	4.5	3.8930	3.8898	3.8306	3.8716
4.0	3.5990	3.6210	5.0	4.1906	4.1570	4.1044	4.1338
4.5	3.8586	3.8770	5.5	4.4870	4.4264	4.3884	4.4046
5.0	4.1300	4.1438	6.0	4.7810	4.7020	4.6820	4.6874
5.5	4.4132	4.4202	6.5	5.0680	4.9908	4.9826	4.9856
6.0	4.7104	4.7052	7.0	5.3598	5.2994	5.2886	5.3032

Table S2: Relative errors at 8 locations across time in simulated observation cases

	Case 1a	Case 1b
$T[3.5]$	[0.0181 0.0254 0.0215 0.0145 0.0046 0.0050 0.0021 0.0012]	[0.0193 0.0214 0.0141 0.0001 0.0037 0.0147 0.0221 0.0113]
$T[4.0]$	[0.0083 0.0183 0.0261 0.0228 0.0075 0.0019 0.0072 0.0008]	[0.0241 0.0126 0.0012 0.0057 0.0030 0.0066 0.0158 0.0146]
$T[4.5]$	[0.0018 0.0194 0.0353 0.0313 0.0113 0.0137 0.0197 0.0041]	[0.0310 0.0004 0.0154 0.0156 0.0144 0.0039 0.0057 0.0229]
$T[5.0]$	[0.0018 0.0276 0.0437 0.0371 0.0164 0.0245 0.0306 0.0092]	[0.0373 0.0114 0.0278 0.0252 0.0244 0.0128 0.0040 0.0327]
$T[5.5]$	[0.0004 0.0366 0.0439 0.0354 0.0188 0.0257 0.0340 0.0060]	[0.0363 0.0140 0.0279 0.0275 0.0256 0.0148 0.0078 0.0356]
$T[6.0]$	[0.0181 0.0358 0.0289 0.0209 0.0124 0.0095 0.0248 0.0188]	[0.0148 0.0035 0.0104 0.0160 0.0120 0.0050 0.0001 0.0173]
	Case 2a	Case 2b
$T[4.5]$	[0.0509 0.0560 0.0353 0.0140 0.0054 0.0136 0.0016 0.0208]	[0.0058 0.0021 0.0037 0.0003 0.0047 0.0007 0.0054 0.0033]
$T[5.0]$	[0.0717 0.0520 0.0265 0.0117 0.0033 0.0090 0.0167 0.0690]	[0.0106 0.0135 0.0124 0.0068 0.0041 0.0052 0.0049 0.0128]
$T[5.5]$	[0.0853 0.0441 0.0201 0.0094 0.0011 0.0048 0.0267 0.1021]	[0.0151 0.0197 0.0166 0.0105 0.0099 0.0092 0.0053 0.0257]
$T[6.0]$	[0.0735 0.0400 0.0238 0.0152 0.0083 0.0051 0.0359 0.0990]	[0.0031 0.0138 0.0092 0.0038 0.0030 0.0019 0.0012 0.0204]
$T[6.5]$	[0.0003 0.0501 0.0434 0.0372 0.0326 0.0282 0.0545 0.0279]	[0.0630 0.0104 0.0153 0.0184 0.0228 0.0231 0.0242 0.0393]
$T[7.0]$	[0.1637 0.0844 0.0804 0.0786 0.0749 0.0691 0.0923 0.1328]	[0.2117 0.0583 0.0576 0.0573 0.0673 0.0658 0.0714 0.1824]
	Case 3a	Case 3b
$T[4.5]$	[0.0388 0.0028 0.0173 0.0207 0.0208 0.0266 0.0193 0.0132]	[0.0041 0.0058 0.0042 0.0108 0.0066 0.0017 0.0099 0.0035]
$T[5.0]$	[0.0316 0.0009 0.0169 0.0192 0.0188 0.0245 0.0177 0.0115]	[0.0015 0.0013 0.0011 0.0068 0.0047 0.0041 0.0027 0.0077]
$T[5.5]$	[0.0243 0.0003 0.0138 0.0179 0.0174 0.0208 0.0141 0.0066]	[0.0029 0.0022 0.0023 0.0024 0.0026 0.0058 0.0038 0.0089]
$T[6.0]$	[0.0265 0.0005 0.0115 0.0196 0.0179 0.0158 0.0127 0.0106]	[0.0017 0.0059 0.0051 0.0025 0.0005 0.0067 0.0058 0.0165]
$T[6.5]$	[0.0302 0.0155 0.0126 0.0250 0.0201 0.0107 0.0238 0.0172]	[0.0084 0.0168 0.0068 0.0064 0.0037 0.0068 0.0034 0.0244]
$T[7.0]$	[0.0301 0.0239 0.0149 0.0287 0.0193 0.0051 0.0262 0.0219]	[0.0108 0.0153 0.0038 0.0035 0.0008 0.0081 0.0021 0.0278]

Table S3: True values of the maximum diameter at different years in real observation cases

Years	Patient 1	Years	Patient 2	Years	Patient 3
2.3	4.6126	3.0	4.2576	8.5	5.7309
3.2	5.4224	4.0	4.5067	9.0	5.8023

Table S4: Relative errors at 8 locations across time in real observation cases

Methods	Years	Patient 1
Bayesian Calibration	2.3	[0.0502 0.0220 0.0054 0.0084 0.0214 0.0087 0.0409 0.0407]
	3.2	[0.0806 0.0846 0.0039 0.0360 0.0145 0.0268 0.0033 0.0138]
Gaussian Process	2.3	[0.0135 0.0159 0.0067 0.0254 0.0469 0.0050 0.0097 0.0003]
	3.2	[0.0165 0.0485 0.1143 0.1094 0.1476 0.1186 0.0315 0.0554]
Methods	Years	Patient 2
Bayesian Calibration	3.0	[0.1262 0.0260 0.0122 0.0168 0.0194 0.0319 0.0236 0.0052]
	4.0	[0.0387 0.0035 0.0157 0.0077 0.0218 0.0374 0.0261 0.0068]
Gaussian Process	3.0	[0.0220 0.0125 0.0102 0.0241 0.0153 0.0289 0.0272 0.0245]
	4.0	[0.0674 0.0683 0.0531 0.0430 0.0497 0.0478 0.0349 0.0306]
Methods	Years	Patient 3
Bayesian Calibration	8.5	[0.0288 0.0199 0.0127 0.0095 0.0035 0.0193 0.0153 0.0471]
	9.0	[0.0190 0.0077 0.0180 0.0067 0.0549 0.0646 0.0670 0.1099]
Gaussian Process	8.5	[0.0417 0.0614 0.0462 0.0346 0.0131 0.0044 0.0006 0.0146]
	9.0	[0.0402 0.0468 0.0317 0.0460 0.0773 0.0860 0.0722 0.0386]

Table S5: Priors and Estimates of Hyperparameters for Case 1a

$\boldsymbol{\psi}_1$			ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean		1	1	1	1	1
	variance		1.1	1.1	1.1	1.3	1.2
Estimates from Stage1			0.641	6.552	0.044	0.725	0.355
$\boldsymbol{\psi}_2$				λ	$\omega_{x1}^{\star}$	$\omega_{x2}^{\star}$	σ_2^2
True values				0.001	1	1	0.005
Lognormal Prior	mean			0.01	1	1	0.005
	variance			0.1	1.1	1.1	0.2
Estimates from Stage2				0.0007	0.270	0.185	0.022
$\boldsymbol{\beta}$		β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates		0.008	0.626	0.096	0.290	0.016	-0.048

Table S6: Priors and Estimates of Hyperparameters for Case 1b

$\boldsymbol{\psi}_1$			ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean		1	1	1	1	1
	variance		1.1	1.1	1.1	1.3	1.2
Estimates from Stage1			0.703	6.406	0.046	0.702	0.290
$\boldsymbol{\psi}_2$				λ	$\omega_{x1}^{\star}$	$\omega_{x2}^{\star}$	σ_2^2
True values				0.001	1	1	0.001
Lognormal Prior	mean			0.01	1	1	0.001
	variance			0.1	1.1	1.1	0.2
Estimates from Stage2				0.002	0.442	0.791	0.0007
$\boldsymbol{\beta}$		β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates		0.0008	0.628	0.096	0.291	0.024	0.017

Table S7: Priors and Estimates of Hyperparameters for Case 2a

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.803	6.815	0.059	0.675	0.261
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ_2^2
True values			0.001	1	1	0.005
Lognormal Prior	mean		0.01	1	1	0.005
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.384	0.272	0.122
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.009	0.637	0.101	0.315	-0.086	-0.016

Table S8: Priors and Estimates of Hyperparameters for Case 2b

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.700	6.836	0.054	0.990	0.304
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.004	0.331	1.103	0.003
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.030	0.638	0.102	0.313	-0.014	-0.009

Table S9: Priors and Estimates of Hyperparameters for Case 2c

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.794	6.467	0.054	0.818	0.283
ψ_2			λ	$\omega^\star_{x1}$	$\omega^\star_{x2}$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.353	2.925	0.008
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.009	0.641	0.101	0.312	0.006	-0.003

Table S10: Priors and Estimates of Hyperparameters for Case 3a

ψ_1		ω_{x1}	ω_{x2}	$\omega_{\theta 2}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.898	15.920	0.250	1.640	0.533
ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.005
Lognormal Prior	mean		0.01	1	1	0.005
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.003	0.210	0.054	0.032
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.070	0.621	0.094	0.341	-0.030	0.004

Table S11: Priors and Estimates of Hyperparameters for Case 3b

ψ_1		ω_{x1}	$\omega_{\theta 1}$	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.951	15.786	0.217	1.971	0.526
ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
True values			0.001	1	1	0.001
Lognormal Prior	mean		0.01	1	1	0.001
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.001	0.637	1.705	0.0006
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.066	0.616	0.094	0.345	-0.020	0.015

Table S12: Priors and Estimates of Hyperparameters for real observations of Patient 1

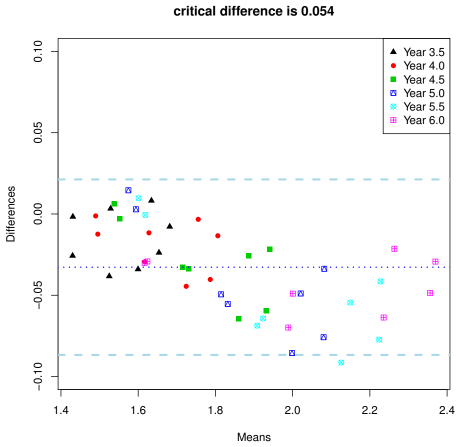
ψ_1		ω_{x2}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	1.1	1.1	1.1	1.3	1.2
Estimates from Stage1		0.242	1.702	0.082	0.008	0.729
ψ_2			λ	$\omega_{x1}^\star$	$\omega_{x2}^\star$	σ_2^2
Lognormal Prior	mean		0.01	1	1	0.1
	variance		0.1	1.1	1.1	0.2
Estimates from Stage2			0.01	1	1	0.001
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	0.182	0.537	0.163	-0.189	0.021	-0.058

Table S13: Priors and Estimates of Hyperparameters for real observations of Patient 2

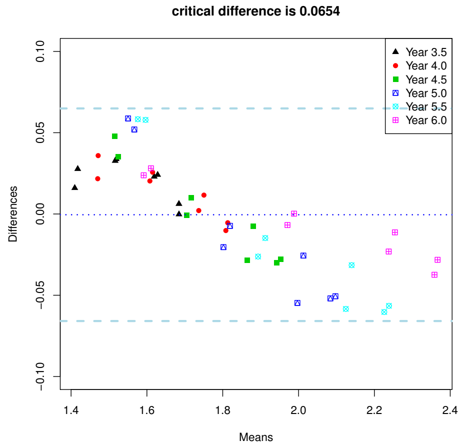
ψ_1		ω_{x2}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	10	10	10	10	10
Estimates from Stage1		0.137	3.110	0.011	0.011	0.146
ψ_2			λ	$\omega_{x1}^{\star}$	$\omega_{x2}^{\star}$	σ_2^2
Lognormal Prior	mean		0.01	1	1	0.1
	variance		10	10	10	10
Estimates from Stage2			0.01	1.791	1.182	0.0003
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	-0.515	0.405	0.185	0.200	-0.071	0.228

Table S14: Priors and Estimates of Hyperparameters for real observations of Patient 3

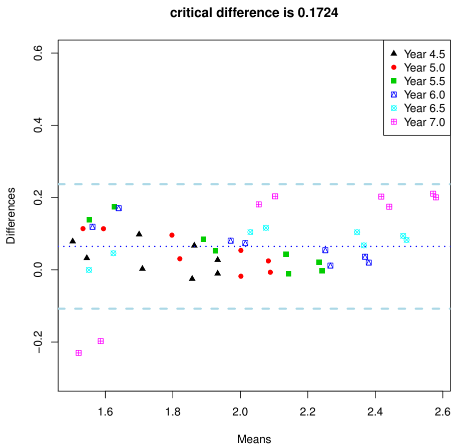
ψ_1		ω_{x2}	ω_{x2}	$\omega_{\theta 1}$	$\omega_{\theta 2}$	σ_1^2
Lognormal Prior	mean	1	1	1	1	1
	variance	10	10	10	10	10
Estimates from Stage1		0.050	21.712	0.060	0.009	0.202
ψ_2			λ	$\omega_{x1}^{\star}$	$\omega_{x2}^{\star}$	σ_2^2
Lognormal Prior	mean		0.01	1	1	0.1
	variance		10	10	10	10
Estimates from Stage2			0.01	11.828	1.365	0.061
β	β_{10}	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}
Estimates	-0.531	0.280	0.138	0.161	0.015	0.069



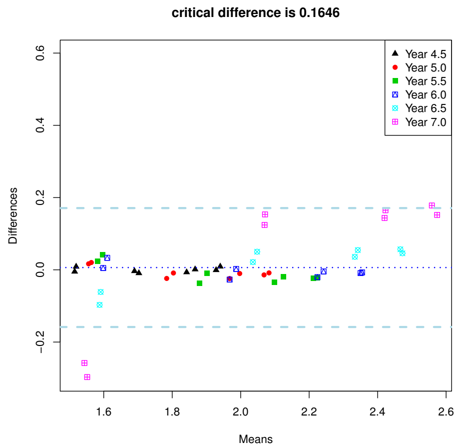
(a) Case 1a



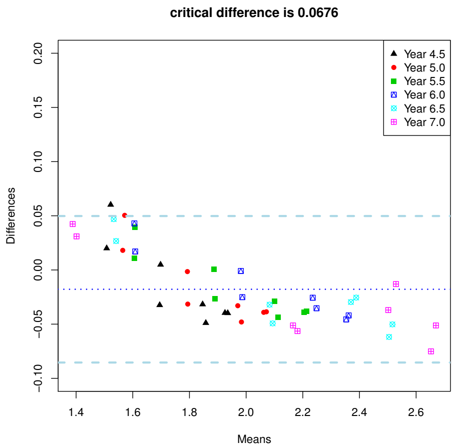
(b) Case 1b



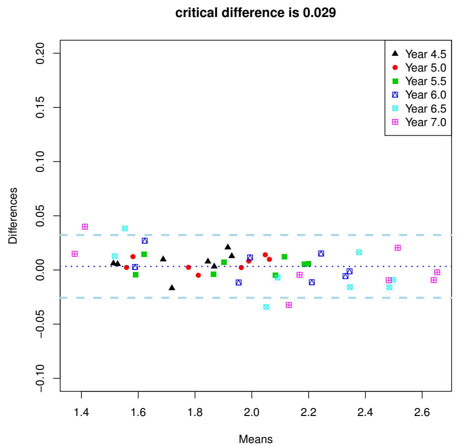
(c) Case 2a



(d) Case 2b

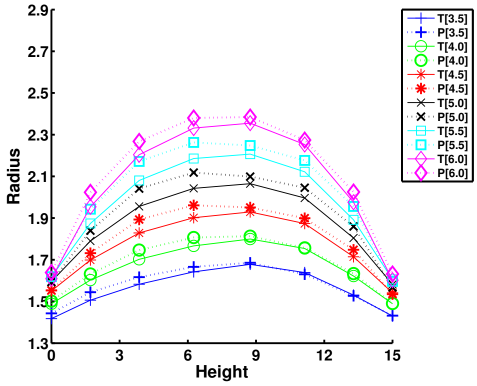


(e) Case 3a

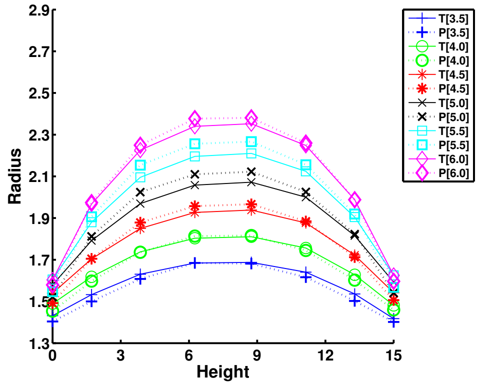


(f) Case 3b

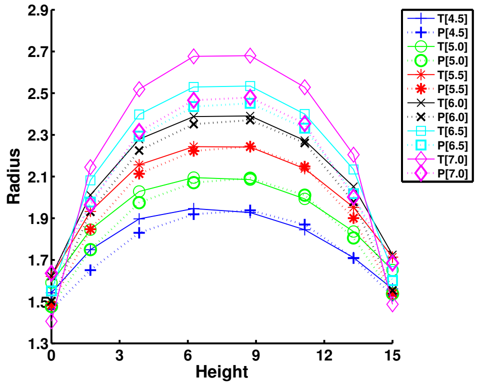
Figure S2: Bland-Altman plots of predictions (A) and simulated observations (B). The x-axis is the mean of A and B, while y-axis is A minus B. "Critical difference" equals half the difference of lower limit and upper limit. The blue dashed line denotes the two standandard deviations (95% level). The dotted blue line in the middle denotes the mean.



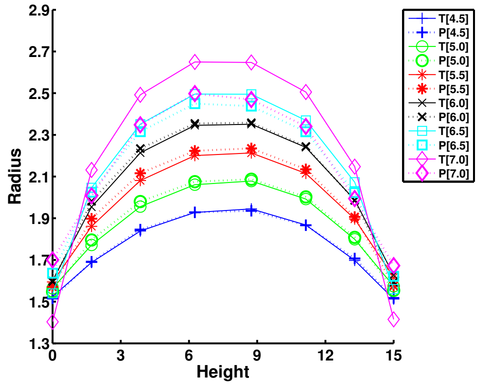
(a) Case 1a



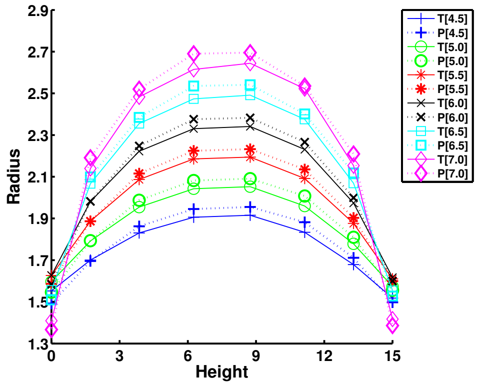
(b) Case 1b



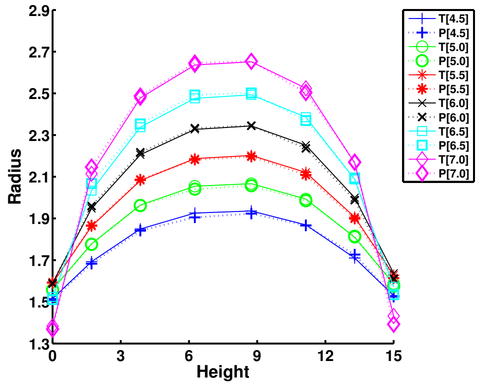
(c) Case 2a



(d) Case 2b



(e) Case 3a



(f) Case 3b

Figure S3: Predictions of the true processes. $T[t]$ denotes the true values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the true values, while the dashed line denotes the predictions.

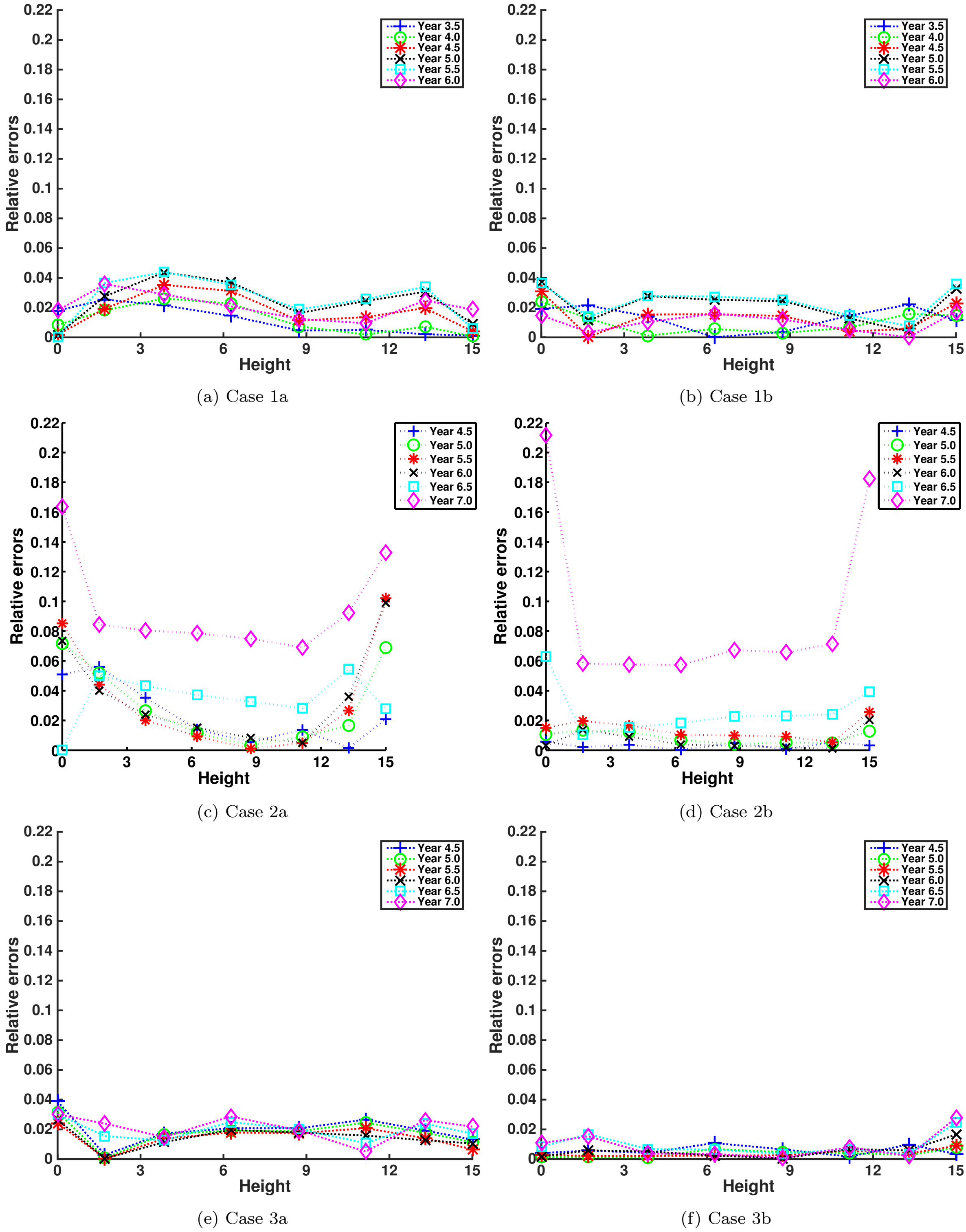
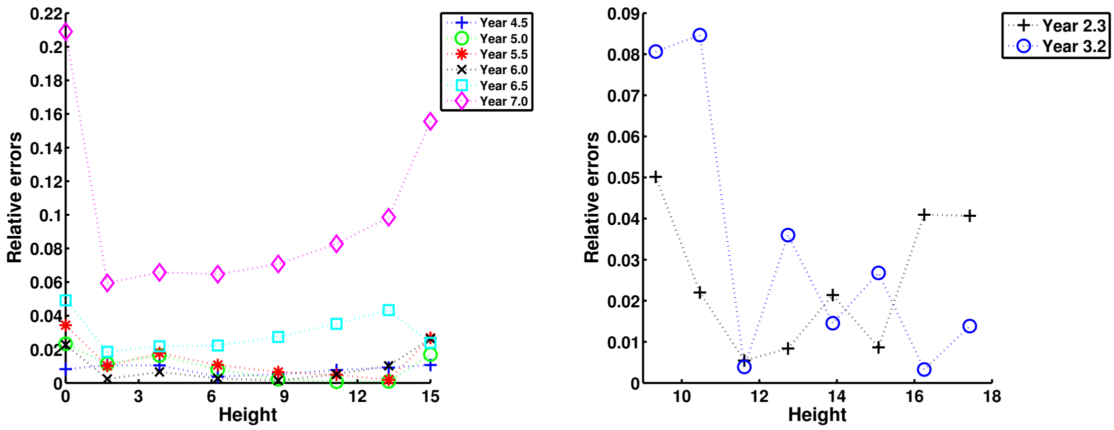
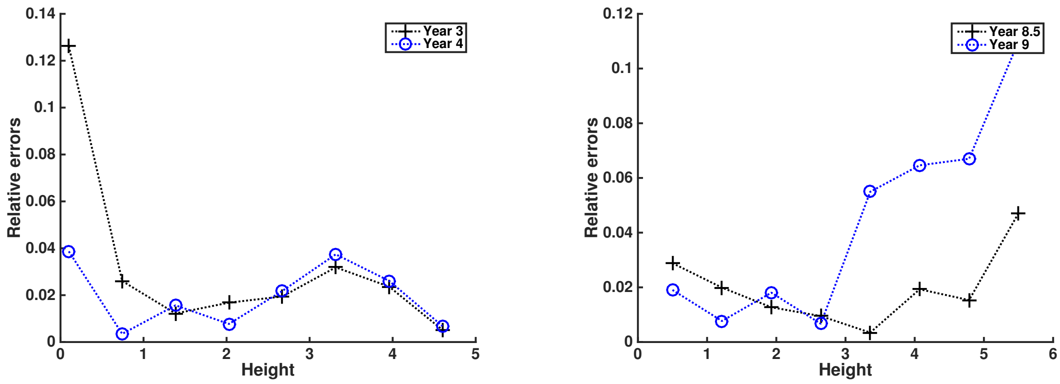


Figure S4: Relative errors between predictions and true values.



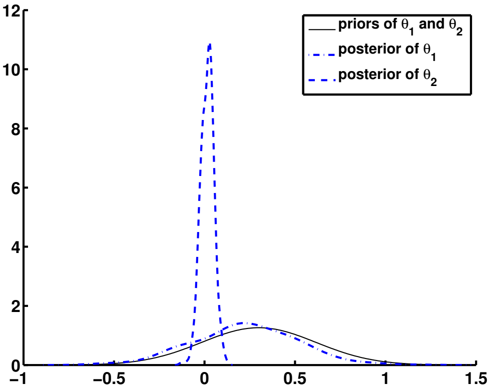
(a) Case 2c (b) Real observation of Patient 1

Figure S5: Relative errors for Case 2c (a) and for real observation of Patient 1 case (b).

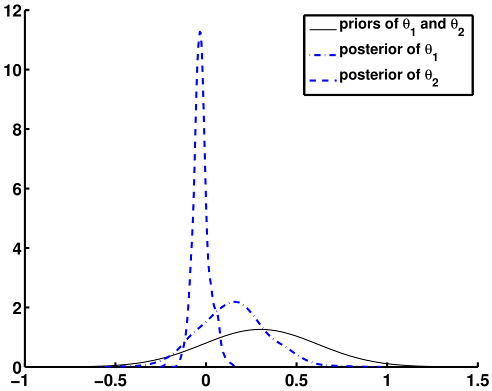


(a) Real observation of Patient 2 (b) Real observation of Patient 3

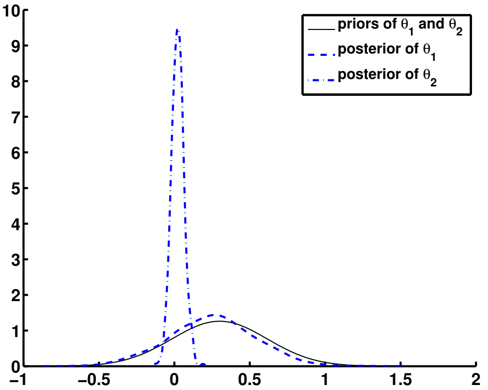
Figure S6: Relative errors for Case 2c (a) and for real observation case (b).



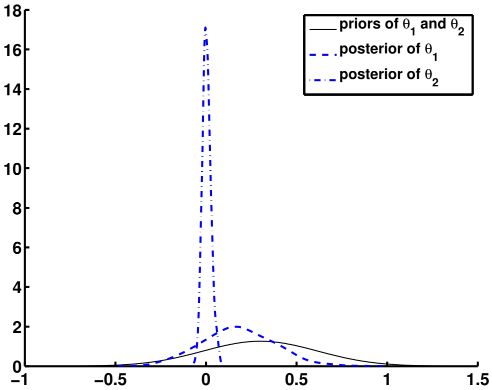
(a) Case 1a



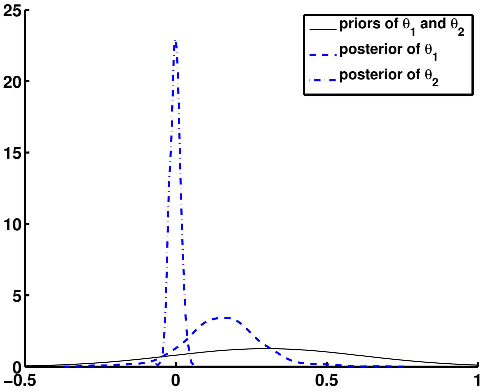
(b) Case 1b



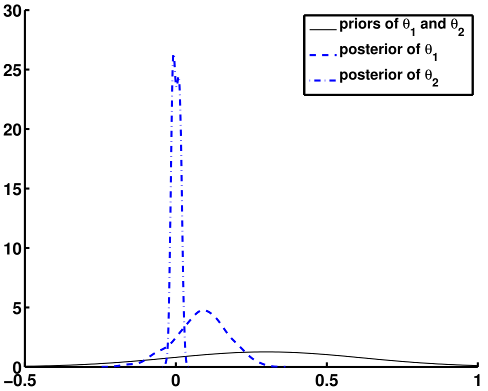
(c) Case 2a



(d) Case 2b

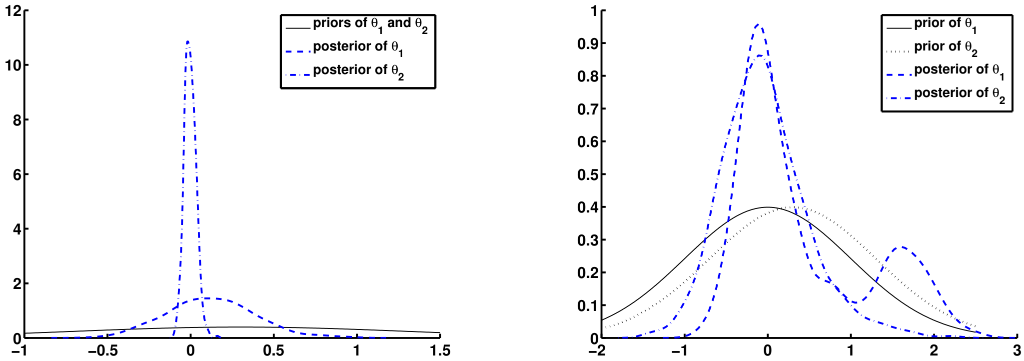


(e) Case 3a



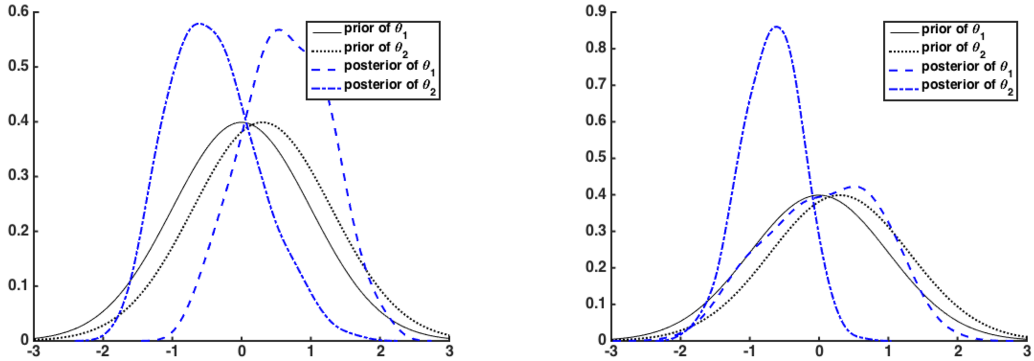
(f) Case 3b

Figure S7: **Prior and posterior densities of θ_1 and θ_2**



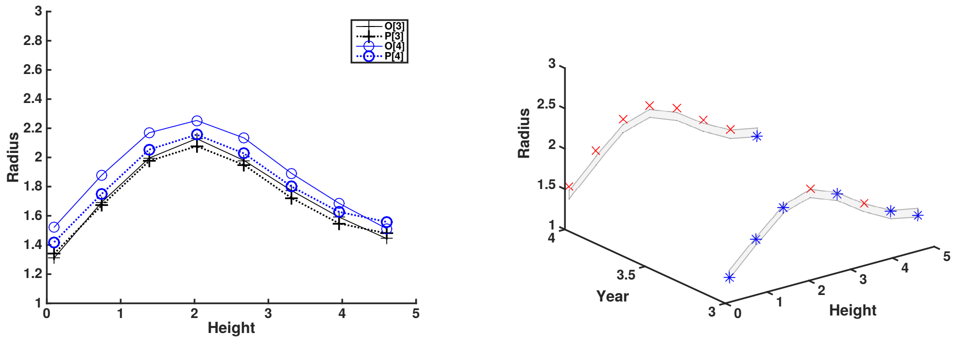
(a) Case 2c (b) Real observation of Patient 1

Figure S8: Prior and posterior densities of θ_1 and θ_2.



(a) Real observation of Patient 2 (b) Real observation of Patient 3

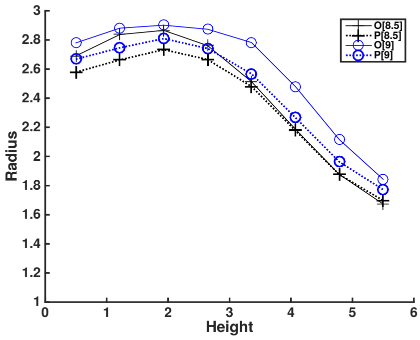
Figure S9: Prior and posterior densities of θ_1 and θ_2.



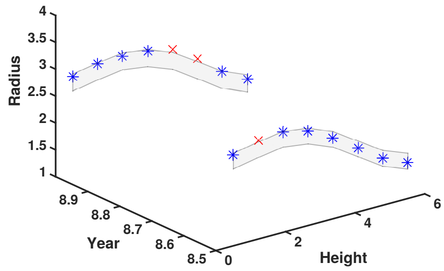
(a) Predictions of the observation processes. $O[t]$ denotes the observation values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the observations, while the dashed line denotes the predictions.

(b) 3D 95% credible band of predictions. The blue stars denote the observations lying inside the credible band. The red marks denote an observation value lying outside the credible band.

Figure S10: Gaussian Process Predictions and credible bands for Real observation of Patient 2.



(a) Predictions of the observation processes. $O[t]$ denotes the observation values at year t, while $P[t]$ denotes the predictions at year t. The solid line denotes the observations, while the dashed line denotes the predictions.



(b) 3D 95% credible band of predictions. The blue stars denote the observations lying inside the credible band. The red marks denote an observation value lying outside the credible band.

Figure S11: Gaussian Process Predictions and credible bands for Real observation of Patient 3.

References

[1] C. K. Williams and C. E. Rasmussen, "Gaussian processes for machine learning," *MIT Press*, 2006.

[2] Y. Xu, J. Choi, S. Dass, and T. Maiti, "Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 8, pp. 2078–2084, 2012.

[3] ——, *Bayesian prediction and adaptive sampling algorithms for mobile sensor networks: Online environmental field reconstruction in space and time.* Springer, 2016.

[4] H. N. Do, A. Ijaz, H. Gharahi, B. Zambrano, J. Choi, W. Lee, and S. Baek, "Prediction of Abdominal Aortic Aneurysm Growth using Dynamical Gaussian Process Implicit Surface," *IEEE Transactions on Biomedical Engineering*, 2018.

[5] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[6] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.