Calatayud, J.; Cortés, J.; Jornet, M. (2020). Improvement of random coefficient differential models of growth of anaerobic photosynthetic bacteria by combining Bayesian inference and gPC. Mathematical Methods in the Applied Sciences. 43(14):7885-7904. https://doi.org/10.1002/mma.5546

Research Article

Mathematical Methods in the Applied Sciences

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000
MOS subject classification: 34A34; 34G20; 47J35; 60H35; 62F15; 62P10; 65C20; 65C60; 92D25

Improvement of random coefficient differential models of growth of anaerobic photosynthetic bacteria by combining Bayesian inference and gPC

J. Calatayud[a], J.-C. Cortés[a], M. Jornet[a,*]

The time evolution of microorganisms, such as bacteria, is of great interest in biology. In the article by D. Stanescu et al. [Electronic Transactions on Numerical Analysis, 34, 44–58 (2009)], a logistic model was proposed to model the growth of anaerobic photosynthetic bacteria. In the laboratory experiment, actual data for two species of bacteria were considered: R. capsulatus and C. vibrioforme. In this paper, we suggest a new nonlinear model by assuming that the population growth rate is not proportional to the size of the bacteria population, but to the number of interactions between the microorganisms, and by taking into account the beginning of the death phase in the kinetic curve. Stanescu et al. evaluated the effect of randomness into the model coefficients by using generalized Polynomial Chaos (gPC) expansions, by setting arbitrary distributions without taking into account the likelihood of the data. By contrast, we utilize a Bayesian inverse approach for parameter estimation to obtain reliable posterior distributions for the random input coefficients in both the logistic and our new model. Since our new model does not possess an explicit solution, we use gPC expansions to construct the Bayesian model and to accelerate the Markov Chain Monte Carlo algorithm for the Bayesian inference.

Keywords: Bacterial growth model; Population dynamics; Nonlinear biological model; Bayesian inverse problem; Generalized Polynomial Chaos

1. Introduction

The time evolution of microorganisms, such as bacteria, has been of great interest in biology for decades [1–3]. In this regard, mathematical models are important to understand and generalize laboratory experiments and to make predictions [4–10]. These models are usually continuous systems that involve ordinary or partial differential equations, which depend on input parameters (initial conditions, forcing term and/or coefficients, etc.) often with a biological interpretation (carrying capacity, growth rate, birth or death rate, concentration of nutrients, etc.). If experimental values are available for the model coefficients, we have a forward model to describe and forecast the main features of the biological system. But in general, to determine the model parameters, experimental data needs to be used. The process of adjusting the coefficients in virtue of collected data is called an inverse problem.

Deterministic differential equations have been widely studied from a theoretical and numerical point of view. To solve the inverse problem in this setting, one usually turns to optimization algorithms, for instance a least squares fitting.

However, deterministic models do not take into account the inherent uncertainty associated to biological processes. Inaccuracies in the measurements often arise due to errors in the laboratory experiments (human, mechanical, etc.), lack of information, missed data, etc. It thus becomes necessary to treat the input parameters in a random sense. This gives rise to random ordinary and partial differential equations [11–13].

[a] *Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*
[*] *Correspondence to: M. Jornet, Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*
[†] E-mail: jucagre@doctor.upv.es; jccortes@imm.upv.es; marjorsa@doctor.upv.es

The primary objective when dealing with random models is uncertainty quantification, i.e., understanding the main statistical features of the forward random model predictions. There are different techniques in the extant literature to handle stochastic systems: Monte Carlo simulations [14], method of moments [12], Random Variable Transformation technique [15–21], generalized Polynomial Chaos (gPC) [22–32], etc. When a closed form solution of the random model is not available, one of the best and computationally cheapest methods for uncertainty quantification is the gPC technique: the solution stochastic process is expressed as a mean square limit of Galerkin projections onto subspaces of orthogonal polynomials [22, 23].

To solve an inverse problem in a random setting, gPC by itself cannot determine suitable model coefficients. The Bayesian approach allows making statistical inference from prior probabilistic information of the parameters and the likelihood associated to the data. The output of Bayesian inference is a posterior probability distribution for each of the parameters, which permits quantifying uncertainty via the posterior predictive distribution [33–35].

When the solution of the random differential equation does not have a closed form expression, each sampling point of the Markov Chain Monte Carlo algorithm requires a numerical solution of the differential equation, which might be time consuming. The gPC and Galerkin projection techniques give approximations of the solution stochastic process, which may be used in the Bayesian model to accelerate the numerical simulations [22, 36, 37].

In this paper, we apply this technique to model the growth of anaerobic photosynthetic bacteria. The population of bacteria increases in size by using light energy to reduce CO_2. In the laboratory experiment, two species of bacteria were considered: *Rhodobacter capsulatus* (*R. capsulatus*) and *Chlorobium vibrioforme* (*C. vibrioforme*). Direct cell counts were made every two or three days until a stationary phase was achieved. This measurements give two sets of data for each one of the two populations under study. In [26], the authors considered a logistic model to explain bacterial growth in both populations, based on Malthusian exponential growth model and competitiveness when there is scarcity of nutrients (mainly light and CO_2). We suggest a new nonlinear model equation by assuming that the population growth rate is not proportional to the size of the bacteria population, but to the number of interactions between the microorganisms (the squared abundance), and by taking into account the start of the decline phase. In [26], uncertainty is put into the model by using arbitrary distributions for the coefficients. By contrast, we have utilized a Bayesian inverse approach for parameter estimation. Since our model does not have an explicit solution, we have combined gPC expansions together with the stochastic Galerkin projection technique to accelerate the Bayesian inference. Thus, we assess the effect of randomness and quantify the uncertainty in a rigorous way.

The structure of the paper is the following. In Section 2, we show and explain the experimental data, and we analyze empirically which should be a good differential equation model. In Section 3, we provide a comprehensive analysis of the logistic model and show its associated Bayesian model. In Section 4, we expose the main theoretical features of the combination of gPC expansions and Bayesian inference. We extend its applicability to models with random variance of the errors, and we use the logistic model as a test example. In Section 5, we expose the theoretical ideas to improve the modeling from [26]. Section 6 is devoted to numerical experiments for deterministic fittings and uncertainty quantification. Finally, Section 7 draws the conclusions.

2. Data on anaerobic photosynthetic bacterial growth

We will use experimental data from the growth of anaerobic photosynthetic bacteria under infrared lightning conditions. The population of bacteria increases in size by using light energy to reduce CO_2 (photosynthesis). In the laboratory experiment, two species of bacteria were considered: *Rhodobacter capsulatus* (*R. capsulatus*) and *Chlorobium vibrioforme* (*C. vibrioforme*). For further details about the experiment, we refer the reader to [26]. Table 1 shows laboratory data on the population sizes of *R. capsulatus* and *C. vibrioforme* under infrared lightning conditions in different mediums. The number of cells/mL has been rescaled by dividing by 10^6. Figure 1 plots the cell counts from Table 1.

R. capsulatus		*C. vibrioforme*	
Time (days)	Population (cells/mL, scale 10^6)	Time (days)	Population (cells/mL, scale 10^6)
0	0.583	0	0.986
2	0.635	14	2.41
4	1.08	16	2.24
7	3.20	18	4.21
9	5.23	21	5.72
11	5.28	23	5.99
14	5.30	25	7.86
		28	6.52

Table 1. Bacteria population sizes [26].

At the first days, when there is no competition between bacteria and no limitation of resources (light and CO_2), the population seems to increase with exponential growth. This was the model proposed by Thomas Malthus in 1798 in his essay [38]. A more

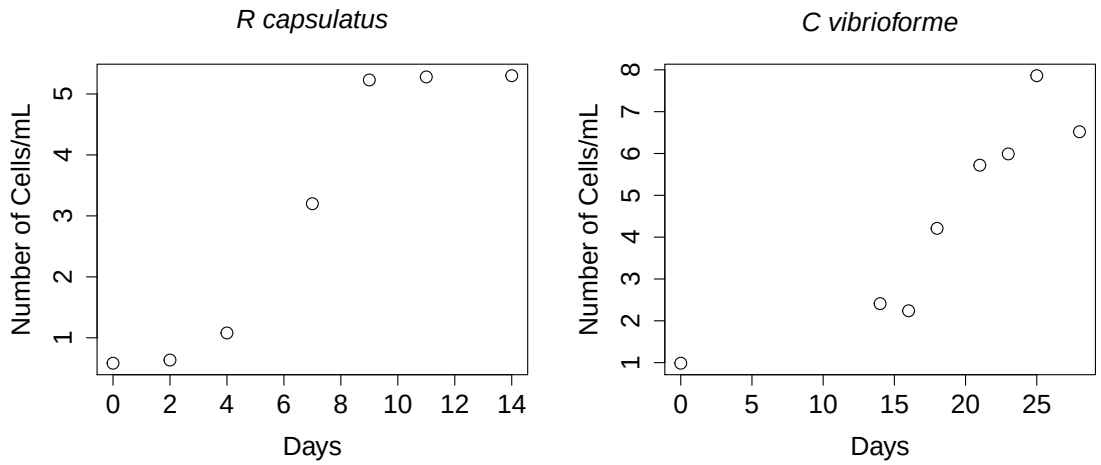

Figure 1. Population size of *R. capsulatus* (left) and *C. vibrioforme* (right).

modern formulation of the Malthusian growth model can be read at the introductory text [5]. The importance of the Malthusian model is evident as in the field of population ecology it is considered as the first law of population dynamics [39]. In his written essay, Malthus already described how non-abundance of sustenance would affect the growth of species. This contribution was developed by Verhulst in 1838 [40]: as time passes and the number of microorganisms augments, there is more competition for the limited food, so that the growth rate decreases with the size of the population. This fact is the basis of the so-called logistic model. Its modern formulation can be consulted in [5], for instance. In principle, the logistic model corresponds to the observed s-shape in Figure 1. This applies especially for *R. capsulatus*. By contrast, *C. vibrioforme* presents a drop of the amount of bacteria at the end, so the logistic model may not fit as expected. This descent might come from the commence of the death phase.

We will see that, in both groups of *R. capsulatus* and *C. vibrioforme*, better results are obtained if we consider that, for the first days, the rate of change is not proportional to the population size, but to the total number of interactions between the microorganisms, i.e., to the squared abundance. Also, the introduction of the death rate in the modeling will play a key role. Taking into account the competition for the limited resources as time goes on, we will obtain a variation of the logistic model that will allow a better modeling for the data from Figure 1. This sort of model formulation has not been widely used in the biological modeling literature, and the unique reference on utilizing squared abundance for the modeling of the growth rate that we have found has been [41, pp. 17–18]. Its justification from a dynamics standpoint is not as clear as the exponential or logistic growth models. However, in our particular database and from a mathematical point of view, this idea of employing squared abundance works better than the logistic model.

3. Logistic model

Let $y(t)$ be the total population at time t measured in days. If there is no competition or there is sustenance for life, and the rate of change is proportional to the total abundance, the Malthusian exponential model describes the population growth [5,38]: $y'(t) = ry(t)$. Given an initial condition $y(t_0) = y_0$, the unique solution of this model is given by $y(t) = y_0 e^{rt}$. In Figure 1, this model seems suitable for the first 9 days in the *R. capsulatus* group and for the first 21 days in the *C. vibrioforme* population. However, at the 9th and 21st days, respectively, an inflection point in the bacterial growth is observed, due to the limited abundance of resources (light and CO_2). If we take into account this scarcity of substances as time passes, then we obtain the logistic model [5,40]:

$$y'(t) = ry(t)\left(1 - \frac{y(t)}{K}\right). \tag{1}$$

The term K is the carrying capacity. In the logistic equation, it is assumed that the growth rate lags linearly with the population size. Under the initial condition $y(t_0) = y_0$, the ordinary differential equation (1) has a unique solution:

$$y(t) = \frac{y_0 K}{y_0 + (K - y_0)e^{-rt}}. \tag{2}$$

In the following subsections, we will review the fit of model (1)–(2) to the data from Table 1 done in [26]. Later, the coefficients in (1) will be randomized. In [26], a gPC approach was used to evaluate the effect of randomness in the coefficients.

Distributions for the coefficients were set by the authors just from an empirical point of view, without using the information given by the data, i.e., the likelihood. By contrast, we will determine appropriate posterior distributions for the input coefficients of (1), by utilizing Bayesian inference.

3.1. Deterministic curve fitting

Given the data from Table 1, the authors in [26] calculated deterministic coefficients r, y_0 and K in (2) such that the squared error is minimized. This method is usually called least squares fitting. Given a set of collected data $d_1, \ldots, d_N$ at times $t_1, \ldots, t_N$ and given the output of a model $M(\zeta, t)$, where ζ are the input parameters and t is the time, the squared error is expressed as

$$\sum_{i=1}^{N} (d_i - M(\zeta, t_i))^2.$$

A least squares fitting consists in finding the set of parameters ζ_0 such that

$$\min_{\zeta} \sum_{i=1}^{N} (d_i - M(\zeta, t_i))^2 = \sum_{i=1}^{N} (d_i - M(\zeta_0, t_i))^2. \tag{3}$$

This last expression (3) is called residual squared error. The best model should minimize the residual squared error. In Table 2, the estimates for the coefficients and the residual squared error are shown. In Figure 2, we plot the least squares fitting together with the measured data. These computations have been already done in [26]. We observe that the deterministic logistic model approximates well the data, although the last data from *C. vibrioforme* presents problems due to its unexpected decreasing behavior.

Parameters for *R. capsulatus*				Parameters for *C. vibrioforme*			
r	y_0	K	residual	r	y_0	K	residual
0.6157	0.1244	5.5623	0.600	0.3184	0.0292	7.4242	3.3127

Table 2. Parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the logistic model.

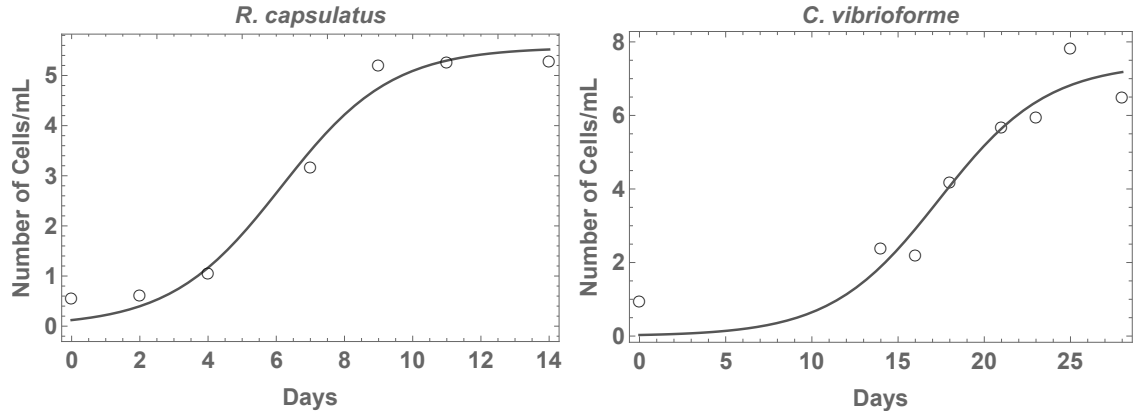

Figure 2. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the logistic model. The real data is denoted by ○ and the fitting is given by the black continuous line.

3.2. Random coefficients and Bayesian inference

Due to the variability involved in the measurements (errors in the collection of data, lack of information, etc.) and the inherent uncertainty associated to population dynamics phenomena, it would be better to treat the input coefficients in a random sense. That is, we suppose that the parameters r, y_0 and K from the logistic model (1) depend on an experiment ω: $r = r(\omega)$, $y_0 = y_0(\omega)$ and $K = K(\omega)$. We denote by Ω the set of all experiments ω, equipped with a σ-algebra $\mathcal{F}$ and a probability measure $\mathbb{P}$, so that we have an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ [22, Def. 2.4]. In this context, the solution $y(t)$ given by (2) becomes a stochastic process $y(t, \omega)$. The main objective thus becomes to quantify the uncertainty of $y(t, \omega)$, for example, with the computation of its mean and variance. When no analytical expression can be obtained, a computational approach needs to be used.

In [26], some probability distributions are given to r, y_0 and K, to assess the effect of randomness into the input parameters. The distributions are set empirically, without utilizing the information given by the data, that is to say, the likelihood. Thus, in order to improve the methodology from [26], we propose a Bayesian model to determine reliable (posterior) distributions for r, y_0 and K. Let $(t_1, \dots, t_N)$ be the times of interest: for *R. capsulatus* $(t_1, t_2, t_3, t_4, t_5, t_6, t_7) = (0, 2, 4, 7, 9, 11, 14)$, and for *C. vibrioforme* $(t_1, t_2, t_3, t_4, t_5, t_6, t_7) = (0, 14, 16, 18, 21, 23, 25, 28)$, respectively (see Table 1). Let y_i be the random variable that models the size of the population at time t_i. The Bayesian model takes the following form:

$$(y_1, \dots, y_N)|(r, y_0, K, \sigma) \sim \pi(y_1, \dots, y_N|r, y_0, K, \sigma) = \prod_{i=1}^{N} \pi(y_i|r, y_0, K, \sigma), \tag{4}$$

$$\pi(r, y_0, K, \sigma) = \pi(r)\pi(y_0)\pi(K)\pi(\sigma), \tag{5}$$

$$r \sim \pi(r), \quad y_0 \sim \pi(y_0), \quad K \sim \pi(K), \quad \sigma \sim \pi(\sigma). \tag{6}$$

Here, π denotes the corresponding probability density function. In (4), it is assumed that the errors caused by the logistic model for $t_1, \dots, t_N$ are independent random variables, with zero expectation and variance σ^2. In fact, bearing in mind expression (2) for the solution of the logistic equation, we will set

$$\pi(y_i|r, y_0, K, \sigma) \sim \text{Normal}\left(\frac{y_0 K}{y_0 + (K - y_0)\mathrm{e}^{-r t_i}}, \sigma\right),$$

i.e., the errors are taken Gaussian. In (5), we reflect the fact that the parameters are independent *a priori*. We have set this prior independence because we do not have any prior information about the covariances, and the forthcoming gPC theory will be exposed for independent random inputs. When computing the joint posterior distribution, the values of the posterior covariances will adapt to the data, so there will possibly be no independence. Finally, in (6), we set the prior distributions for r, y_0, K and σ.

The distributions for y_0 and K must be positive, as they measure bacteria abundance. The distribution for σ is also positive by definition of variance. Concerning r, looking at Figure 1 we deduce that r should be positive as well. Nonetheless, in practice, it could be possible to set prior probability distributions with support intersecting negative numbers, provided that we put a positive mean value and a very small variance.

On the other hand, the distributions of r, y_0 and K should have as mean value the deterministic estimates from Table 2. From an intuitive point of view, these deterministic estimates are the unique information available to set the prior distributions $\pi(r)$, $\pi(y_0)$ and $\pi(K)$. While from a mathematical standpoint, we know that, when the sample size N is large, the posterior distribution follows approximately a normal law with mean value given by the maximum likelihood estimator (combine [42, Th. 3.10] for the asymptotic behavior of the maximum likelihood estimator and [42, Th. 8.3] for the asymptotic limit of the Bayesian estimator). In this case, as the errors are supposed to be Gaussian and independent, the maximum likelihood estimator coincides with the least squares fitting performed in Table 2. Notice that, taking the deterministic estimates from Table 2 as the mean values for the prior distributions, we are employing the collected data to set the priors, so that we are actually using the so-called empirical Bayes method [33, Ch. 9], [43, 44].

The formula for the joint posterior density function of the parameters is given by

$$\pi(r, y_0, K, \sigma|y_1, \dots, y_N) = \frac{\pi(y_1, \dots, y_N|r, y_0, K, \sigma)\pi(r, y_0, K, \sigma)}{\iiiint \pi(y_1, \dots, y_N|r, y_0, K, \sigma)\pi(r, y_0, K, \sigma)\mathrm{d}r\,\mathrm{d}y_0\,\mathrm{d}K\,\mathrm{d}\sigma}.$$

To compute the posterior density function of a subset of the random vector (r, y_0, K, σ), just marginalize by integrating. The posterior predictive distribution is expressed as

$$\pi(\tilde{y}_1, \dots, \tilde{y}_N|y_1, \dots, y_N) = \iiiint \pi(\tilde{y}_1, \dots, \tilde{y}_N|r, y_0, K, \sigma)\pi(r, y_0, K, \sigma|y_1, \dots, y_N)\mathrm{d}r\,\mathrm{d}y_0\,\mathrm{d}K\,\mathrm{d}\sigma.$$

These formulas and more information on Bayesian inference are available in the standard reference book [33]. Usually, the posterior distribution of the model parameters is not analytically tractable, therefore simulation algorithms must be carried out to sample from the posterior distribution. This class of algorithms is encompassed under the name Markov Chain Monte Carlo simulation [33, Ch. 6–7], [45]. An algorithm for Bayesian simulation has been implemented in WinBUGS [33, Ch. 8], [46]. Other possible software are JAGS [33, p. 214], [47–49]; SAS (SAS Institute, Cary NC) [33, Ch. 8], [50]; etc.

In Section 6, we will specify prior distributions for the parameters r, y_0 and K. We will thus assess the effect of randomness into the inputs of the random logistic model.

4. Combining Bayesian inference and gPC

In this section we will show how gPC expansions can be used to perform Bayesian inference when the solution of the differential equation model does not have an explicit expression. There are results in the extant literature that combine Bayesian inference and gPC expansions when the error from the model is assumed to be Gaussian, with zero expectation and constant

variance [36, 37], [22, Ch. 8]. We will extend these results to a random variance, so that one does not have to make point estimate guesses on the variability of the model error. We will show how these theoretical results work with the random logistic model. This will be a test example, since we saw in the previous section that the logistic equation has a closed form solution, to which Bayesian inference can be directly applied. By contrast, in Section 5, in which we improve the logistic equation to a more suitable model, we will need gPC expansions to carry out accelerated Bayesian inference, since the solution of the new model equation will not be explicitly known.

4.1. Theoretical results

Consider an ordinary differential equation model $y'(t) = F(t, y(t))$. Suppose that both F and the solution $y(t)$ depend on some random input parameters $\zeta_1, \dots, \zeta_s$ (the initial condition is among these inputs). The probabilistic properties of these random inputs are: mutual independence, absolute continuity and finiteness of all moments. Let $\zeta = (\zeta_1, \dots, \zeta_s)$ be the joint vector of coefficients. Suppose that we have times of interest, $t_1, \dots, t_N$, in which we have collected data $d_1, \dots, d_N$. We suppose that $y'(t) = F(t, y(t))$ is the suitable deterministic model to explain $d_1, \dots, d_N$, and from it we construct the following Bayesian model: if we denote by y_i the random variable that models d_i, then

$$(y_1, \dots, y_N)|\zeta \sim \prod_{i=1}^{N} \text{Normal}(y(t_i), \sigma), \tag{7}$$

$$\zeta \sim \pi(\zeta) = \prod_{i=1}^{N} \pi(\zeta_i), \quad \sigma \sim \pi(\sigma). \tag{8}$$

We are assuming that σ is either absolutely continuous with density function $\pi(\sigma)$ or a constant. In the case of being a constant, the subsequent development is applicable by considering a Dirac delta function as its density function. Let

$$\pi(\zeta, \sigma|d_1, \dots, d_N) = \frac{\pi(d_1, \dots, d_N|\zeta, \sigma)\pi(\zeta, \sigma)}{\iint \pi(d_1, \dots, d_N|\zeta, \sigma)\pi(\zeta, \sigma)\, \mathrm{d}\zeta\, \mathrm{d}\sigma}$$

be the joint posterior density function of the parameters. This Bayesian model was proposed in the previous section for the random logistic model, and it makes sense when the explicit solution $y(t)$ of the differential equation model is available.

When $y(t)$ does not have a closed form expression and one has to use Markov Chain Monte Carlo algorithms [33, Ch. 6–7], [45], the main computational drawback is that each sampling point requires a solution of the underlying stochastic system $y'(t) = F(t, y(t))$. The idea to speed up the Bayesian inference is to approximate $y(t)$ via another function in an $\mathrm{L}^2(\Omega)$ sense, and then to put the approximation in the mean of the normal distribution from (7).

To approximate $y(t)$, we use the gPC technique. We work in the Hilbert space $(\mathrm{L}^2(\Omega), \langle , \rangle)$ of random variables with finite variance, where the inner product is defined as $\langle X, Y\rangle = \mathbb{E}[XY]$. Suppose that the random inputs $\zeta_1, \dots, \zeta_s$ are independent and absolutely continuous random variables with finite moments and density functions $\pi(\zeta_1), \dots, \pi(\zeta_s)$ (the prior density functions defined in (8)). There are different gPC approaches, which are kept to a minimum below:

(i) (Classical) gPC: Suppose that $\zeta_1, \dots, \zeta_s$ are functions of random variables with distributions that belong to the Askey-Wiener scheme. That is, $\zeta = h(\xi)$, where $h : \mathbb{R}^r \to \mathbb{R}^s$ is a Borel measurable function and $\xi = (\xi_1, \dots, \xi_r)$ is a random vector with independent components, such that ξ_i is Gaussian, Gamma, Beta or uniform distributed. Take the univariate orthogonal polynomials from the Askey-Wiener scheme associated to the distribution of ξ_i, $1 \leq i \leq r$, and compute a simple tensor product to obtain multivariate orthogonal polynomials in ξ. Let $\{\phi_i(\xi)\}_{i=1}^{\infty}$ be the sequence of orthogonal polynomials with respect to $\langle , \rangle$. As $y(t)$ is a function of ξ, we can expand $y(t) = \sum_{i=1}^{\infty} \tilde{y}_i(t)\phi_i(\xi)$ in $\mathrm{L}^2(\Omega)$, where $\tilde{y}_i(t) = \mathbb{E}[y(t)\phi_i(\xi)]/\mathbb{E}[\phi_i(\xi)^2]$ is the i-th Fourier coefficient. This approach is based on [22, 23].

(ii) Adaptive gPC: In this case, the distributions of $\zeta_1, \dots, \zeta_s$ do not necessarily belong to the Askey-Wiener scheme. Let $\mathcal{C}_i^p = \{1, \zeta_i, \dots, \zeta_i^p\}$ be the canonical basis of polynomials in ζ_i up to degree p. We orthonormalize this basis with respect to $\langle , \rangle$ and thus get $\Xi_i^p = \{\phi_0^i(\zeta_i), \dots, \phi_p^i(\zeta_i)\}$. Using a simple tensor product, we define an orthonormal basis with respect to $\langle , \rangle$ in the space of multivariate polynomials in ζ up to degree p: $\Xi = \{\phi_1(\zeta), \dots, \phi_P(\zeta)\}$, where $P = (p + s)!/p!s!$. If we let $p, P \to \infty$, we obtain an orthonormal sequence $\{\phi_i(\zeta)\}_{i=1}^{\infty}$. We expand $y(t)$ as $y(t) = \sum_{i=1}^{\infty} \tilde{y}_i(t)\phi_i(\zeta)$, where $\tilde{y}_i(t) = \mathbb{E}[y(t)\phi_i(\zeta)]$ is the i-th Fourier coefficient. This approach is based on [24, 25] and is referred to as adaptive gPC. An advantage of this strategy is that we are not restricted to standard probability distributions, because of the Random Variable Transformation technique [18]. However, using the Gram-Schmidt procedure may lead to a loss of orthogonality for large p [51], which may ruin the computations. Nonetheless, adaptive gPC usually converges in an algebraic or exponential rate (spectral convergence), so a small p usually suffices and no loss of orthogonality problems appear.

For convergence issues on the (classical) gPC and adaptive gPC expansions, we refer the reader to [52, Th. 3.6], which completely determines the problem of convergence: (i) if $y(t) \in \mathrm{L}^2(\Omega)$, then $y(t) = \sum_{i=1}^{\infty} \tilde{y}_i(t)\phi_i(\xi)$ if the moment problem is uniquely solvable for each random variable $\xi_1, \dots, \xi_r$; (ii) if $y(t) \in \mathrm{L}^2(\Omega)$, then $y(t) = \sum_{i=1}^{\infty} \tilde{y}_i(t)\phi_i(\zeta)$ if the moment problem is uniquely solvable for each random variable $\zeta_1, \dots, \zeta_s$.

We will adopt the adaptive gPC approach (ii), as it allows more general probability distributions for $\zeta_1, \dots, \zeta_s$ for the stochastic Galerkin projection technique, without having to compute the inverse of their cumulative distribution functions,

see [22, expr. (5.15) (5.16)], [26, expr. (4.8) (4.9)]. The stochastic Galerkin projection technique is based on approximating $y(t) \approx \sum_{i=1}^P \hat{y}_i^P(t)\phi_i(\zeta) = \hat{y}^P(t)$, by imposing $\hat{y}^P(t)$ to be a solution of the differential equation. Using the orthonormality of $\phi_1(\zeta), \ldots, \phi_P(\zeta)$, we obtain a deterministic system of differential equations for the coefficients $\hat{y}_1^P(t), \ldots, \hat{y}_P^P(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{y}_k^P(t) = \langle F\left(t, \sum_{i=1}^P \hat{y}_i^P(t)\phi_i(\zeta)\right), \phi_k(\zeta)\rangle,$$

$$\hat{y}_k^P(0) = \mathbb{E}[y_0\phi_k(\zeta)],$$

for $k = 1, \ldots, P$. Under certain conditions, the Galerkin projection $\hat{y}^P(t)$ tends in $\mathrm{L}^2(\Omega)$ as $P \to \infty$ to $y(t)$, see [53].

Consider the Bayesian model

$$(y_1, \ldots, y_N)|\zeta \sim \prod_{i=1}^N \mathrm{Normal}(\hat{y}^P(t_i), \sigma), \tag{9}$$

$$\zeta \sim \pi(\zeta) = \prod_{i=1}^N \pi(\zeta_i), \quad \sigma \sim \pi(\sigma). \tag{10}$$

Let

$$\pi_P(\zeta, \sigma|d_1, \ldots, d_N) = \frac{\pi_P(d_1, \ldots, d_N|\zeta, \sigma)\pi(\zeta, \sigma)}{\iint \pi_P(d_1, \ldots, d_N|\zeta, \sigma)\pi(\zeta, \sigma)\,\mathrm{d}\zeta\,\mathrm{d}\sigma}$$

be the joint posterior density of the parameters, where $\pi_P(d_1, \ldots, d_N|\zeta, \sigma)$ is the likelihood from (9) and $\pi(\zeta, \sigma)$ is the prior from (10) which coincides with (8). In [37], the authors proved that, if $\hat{y}^P(t) \to y(t)$ in $\mathrm{L}^2(\Omega)$ as $P \to \infty$ and σ is constant, then $\pi_P(\zeta|d_1, \ldots, d_N)$ tends to $\pi(\zeta|d_1, \ldots, d_N)$ as $P \to \infty$ in the sense of the Kullback-Leibler divergence:

$$D(\pi_P||\pi) = \int \pi_P(\zeta|d_1, \ldots, d_N) \log \frac{\pi_P(\zeta|d_1, \ldots, d_N)}{\pi(\zeta|d_1, \ldots, d_N)}\,\mathrm{d}\zeta \overset{P\to\infty}{\longrightarrow} 0.$$

Moreover, if $\hat{y}^P(t) \to y(t)$ in $\mathrm{L}^2(\Omega)$ algebraically/exponentially, then $D(\pi_P||\pi) \to 0$ algebraically/exponentially.

Let us see that this result can be extended to a random σ that possesses a prior distribution $\pi(\sigma)$. In [37, Lemma 4.2, expr. (4.11)], we need to add an integration with respect to $\pi(\sigma)$. If we impose

$$\mathbb{E}_{\pi(\sigma)}\left[\frac{1}{\sigma^{3N}}\right] = \int_0^\infty \frac{\pi(\sigma)}{\sigma^{3N}}\,\mathrm{d}\sigma < \infty,$$

then the same conclusion from [37, Lemma 4.2] holds:

$$D(\pi_P||\pi) = \iint \pi_P(\zeta, \sigma|d_1, \ldots, d_N) \log \frac{\pi_P(\zeta, \sigma|d_1, \ldots, d_N)}{\pi(\zeta, \sigma|d_1, \ldots, d_N)}\,\mathrm{d}\zeta\,\mathrm{d}\sigma \overset{P\to\infty}{\longrightarrow} 0.$$

Thus, under general assumptions, $\pi_P(\zeta, \sigma|d_1, \ldots, d_N) \to \pi(\zeta, \sigma|d_1, \ldots, d_N)$ in the sense of the Kullback-Leibler divergence. Formally, the posterior predictive distribution computed from the Galerkin projection (9)–(10) tends to the posterior predictive distribution from (7)–(8):

$$\pi_P(\tilde{d}_1, \ldots, \tilde{d}_N|d_1, \ldots, d_N) = \iint \pi_P(\tilde{d}_1, \ldots, \tilde{d}_N|\zeta, \sigma)\pi_P(\zeta, \sigma|d_1, \ldots, d_N)\,\mathrm{d}\zeta\,\mathrm{d}\sigma$$

$$\overset{P\to\infty}{\longrightarrow} \pi(\tilde{d}_1, \ldots, \tilde{d}_N|d_1, \ldots, d_N) = \iint \pi(\tilde{d}_1, \ldots, \tilde{d}_N|\zeta, \sigma)\pi(\zeta, \sigma|d_1, \ldots, d_N)\,\mathrm{d}\zeta\,\mathrm{d}\sigma.$$

4.2. Application to the random logistic model

Consider the Bayesian model (4)–(6). Instead of using the solution (2) from the logistic differential equation (1), we approximate it via a stochastic Galerkin procedure as a test of the previous theory. Let $\zeta = (r, y_0, K)$ and $\hat{y}^P(t) = \sum_{i=1}^P \hat{y}_i^P(t)\phi_i(\zeta)$. The system of differential equations for the coefficients $\hat{y}_1^P(t), \ldots, \hat{y}_P^P(t)$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{y}_k^P(t) = \sum_{i=1}^P \hat{y}_i^P(t)\mathbb{E}[r\phi_i(\zeta)\phi_k(\zeta)] - \sum_{i,j=1}^P \hat{y}_i^P(t)\hat{y}_j^P(t)\mathbb{E}\left[\frac{r}{K}\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)\right],$$

$$\hat{y}_k^P(0) = \mathbb{E}[y_0\phi_k(\zeta)],\ k = 1, \ldots, P.$$

By means of standard numerical techniques, the Galerkin coefficients $\hat{y}_1^P(t), \ldots, \hat{y}_P^P(t)$ can be computed at the times of interest $t_1, \ldots, t_N$. To do statistical inference, we use the model (9)–(10). In Section 6, we show numerical experiments. We will see that the posterior distribution $\pi_P(\zeta, \sigma|d_1, \ldots, d_N)$ is similar to the true posterior $\pi(\zeta, \sigma|d_1, \ldots, d_N)$ from (4)–(6), even for small p and P, due to the spectral convergence.

5. Improvement of the logistic model

Suppose that, under non-scarcity of nourishment, the growth rate $y'(t)$ is not proportional to the population size $y(t)$, but to the total number of interactions, i.e., to the squared abundance $y(t)^2$. In this case, the model becomes a variation of the Malthusian growth model: $y'(t) = ry(t)^2$. If we take into account competition inside the test tubes because of limited resources, mainly light and CO_2, then the growth rate constant decays linearly with the population size:

$$y'(t) = ry(t)^2\left(1 - \frac{y(t)}{K}\right). \tag{11}$$

The coefficient K is the carrying capacity. Unlike the logistic differential equation, given an initial condition $y(t_0) = y_0$, the ordinary differential equation (11) does not have an explicit form for the solution.

This model formulation (11) has not been extensively used in the biological modeling literature, and the unique reference on squared abundance for the modeling of the growth rate that we have found has been [41, pp. 17–18]. Its biological justification is debatable, nonetheless, in our particular database and from a mathematical standpoint, (11) works better than the logistic model.

Model (11) may be improved if we take into account the death rate:

$$y'(t) = ry(t)^2\left(1 - \frac{y(t)}{K}\right) - \delta y(t). \tag{12}$$

After the stationary phase, bacteria population enters into the so-called death phase. Due to the lack of nutrients, the population size starts to decline. In this model, K is interpreted as the carrying capacity under no mortality.

At this point, we can compare models (1), (11) and (12) from the following mathematical point of view. Consider a general model $y'(t) = f(y(t))$. If we assume that f is sufficiently smooth, then we can express f as a Taylor power series: $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots$. If we suppose that there is no spontaneous generation in the population, then $a_0 = 0$. Therefore, the model equation becomes $y'(t) = a_1y(t) + a_2y(t)^2 + a_3y(t)^3 + \dots$. If we truncate at the first term, then we obtain Malthus model. If we keep the first and second term, the logistic equation appears. If we put $a_1 = 0$ and keep a_2 and a_3, we get model (11). Finally, if we take the first three terms, the differential equation becomes (12). This provides an intuition on why (12) should make the most significant improvement. Indeed, we will see that model (12) improves the fitting of both the logistic equation (1) and model (11).

5.1. Deterministic curve fitting

Given the data from Table 1, we get deterministic estimates for r, y_0, K and δ in (11) and (12) via a least squares procedure (3). In this case, the least squares fitting has to be performed without an explicit solution of the differential equation model. In Section 6, we will show the least squares fitting and the residual squared error. We will observe that the error is much smaller for (11) and (12) than for the logistic model (1), especially for the *R. capsulatus* population. Therefore, from a deterministic point of view, taking into account interactions instead of total population allows a better modeling. The best modeling will be achieved with (12). This highlights the importance of adding the effect of mortality in the equation.

5.2. Random coefficients and combination of Bayesian inference and gPC

To randomize both (11) and (12), we consider that the output depends on an experiment ω, which belongs to the sample space Ω of an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Thereby, the parameters are random variables: $r = r(\omega)$, $y_0 = y_0(\omega)$, $K = K(\omega)$, $\delta = \delta(\omega)$; and the solution $y(t)$ becomes a stochastic process $y(t, \omega)$. The goal is to quantify its uncertainty computationally, by approximating its mean and variance statistics.

We model the bacterial growth with Bayesian model (7)–(8). As no explicit solution of (11) is available, we approximate $y(t)$ in $\mathrm{L}^2(\Omega)$ using the stochastic Galerkin projection technique. Let $\{\phi_i(\zeta)\}_{i=1}^P$ be the orthonormal sequence from the adaptive gPC approach, where $\zeta = (r, y_0, K)$. Let $\hat{y}^P(t) = \sum_{i=1}^P \hat{y}_i^P(t)\phi_i(\zeta)$ be the Galerkin projection. The deterministic coefficients are computed by solving numerically the following system of deterministic differential equations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\hat{y}_k^P(t) = &\sum_{i,j=1}^P \hat{y}_i^P(t)\hat{y}_j^P(t)\mathbb{E}[r\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)]\\ &- \sum_{i,j,l=1}^P \hat{y}_i^P(t)\hat{y}_j^P(t)\hat{y}_l^P(t)\mathbb{E}\left[\frac{r}{K}\phi_i(\zeta)\phi_j(\zeta)\phi_l(\zeta)\phi_k(\zeta)\right],\end{aligned}$$

$$\hat{y}_k^P(0) = \mathbb{E}[y_0\phi_k(\zeta)],\ k = 1, \dots, P.$$

In the case of model (12), letting $\zeta = (r, y_0, K, \delta)$ and using a similar reasoning, we get the following system of deterministic differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{y}_k^P(t) = \sum_{i,j=1}^{P} \hat{y}_i^P(t)\hat{y}_j^P(t)\mathbb{E}[r\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)]$$
$$- \sum_{i,j,l=1}^{P} \hat{y}_i^P(t)\hat{y}_j^P(t)\hat{y}_l^P(t)\mathbb{E}\left[\frac{r}{K}\phi_i(\zeta)\phi_j(\zeta)\phi_l(\zeta)\phi_k(\zeta)\right] - \sum_{i=1}^{P} \hat{y}_i^P(t)\mathbb{E}[\delta\phi_i(\zeta)\phi_k(\zeta)],$$
$$\hat{y}_k^P(0) = \mathbb{E}[y_0\phi_k(\zeta)],\ k = 1, \ldots, P.$$

Bayesian model (9)–(10) permits assessing the effect of randomness in (11) and (12) using the likelihood of the data.

We remark that the prior distributions for r, y_0, K, δ and σ must be positive, although in practice it could be possible to set prior distributions taking negative values as long as we put a positive mean value and a very small variance.

In the numerical experiments, the prior distributions will be strongly centered around the deterministic value obtained in the least squares fitting. As we already explained, this is because for a large sample size N, the posterior distribution follows approximately a normal law with mean value given by the maximum likelihood estimator. In Section 6, we show the simulation results from WinBUGS.

6. Numerical experiments

In this section, we perform numerical experiments of the models presented before. First, we will specify prior distributions for r, y_0 and K to carry out Bayesian inference in the logistic model (1). As a checking performance, we will see that expressing the solution (2) of the logistic equation (1) via gPC expansions gives good approximations of the posterior and posterior predictive distributions. On the other hand, we will fit the new models proposed, (11) and (12), in a deterministic manner to compare the results with those achieved in [26]. Finally, we will combine Bayesian inference and gPC expansions to simulate from the posterior distributions of the random parameters with a cheap computational expense.

6.1. Random logistic model

Consider the logistic model (1) and its solution (2) in a randomized setting. To model the data from Table 1 using the Bayesian approach (4)–(6), we have set

$$r \sim \text{Gamma}(\alpha_r, \beta_r),\ y_0 \sim \text{Gamma}(\alpha_{y_0}, \beta_{y_0}),\ K \sim \text{Gamma}(\alpha_K, \beta_K),\ \sigma \sim \text{Unif}(a_\sigma, b_\sigma) \tag{13}$$

(recall that these four parameters are positive), where we have employed the shape-rate notation for the gamma distribution. As we want the mean of these random variables to be equal to the least square estimates (maximum likelihood estimators) from Table 2, we impose:

$$\textit{R. capsulatus}: \ \frac{\alpha_r}{\beta_r} = 0.6157,\ \frac{\alpha_{y_0}}{\beta_{y_0}} = 0.1244,\ \frac{\alpha_K}{\beta_K} = 5.5623,$$
$$\textit{C. vibrioforme}: \ \frac{\alpha_r}{\beta_r} = 0.3184,\ \frac{\alpha_{y_0}}{\beta_{y_0}} = 0.0292,\ \frac{\alpha_K}{\beta_K} = 7.4242.$$

For the *R. capsulatus* group, we suppose that approximately 68% of the variability of r, y_0 and K is 0.05, 0.05 and 0.2, respectively (this is the subjective part of the modeling), and we take this as the typical deviation (by the 68-95-99.7 rule [54]). Then

$$\textit{R. capsulatus}: \ \frac{\alpha_r}{\beta_r^2} = 0.0025,\ \frac{\alpha_{y_0}}{\beta_{y_0}^2} = 0.0025,\ \frac{\alpha_K}{\beta_K^2} = 0.04.$$

For the *C. vibrioforme* population, we take

$$\textit{C. vibrioforme}: \ \frac{\alpha_r}{\beta_r^2} = 0.0009,\ \frac{\alpha_{y_0}}{\beta_{y_0}^2} = 0.0009,\ \frac{\alpha_K}{\beta_K^2} = 0.01.$$

This gives, for the *R. capsulatus* population, the values $\alpha_r = 151.6346$, $\beta_r = 246.28$, $\alpha_{y_0} = 6.190144$, $\beta_{y_0} = 49.76$, $\alpha_K = 773.4795$ and $\beta_K = 139.0575$; and for the *C. vibrioforme* population, the parameters $\alpha_r = 112.64$, $\beta_r = 353.78$, $\alpha_{y_0} = 0.947$, $\beta_{y_0} = 32.44$, $\alpha_K = 5511.87$ and $\beta_K = 742.42$. For σ, assuming that the error in the modeling is at most 1, we set $\sigma \sim \text{Unif}(0, 1)$.

With this information, we have the prior distributions for the parameters. The posterior distribution of the model parameters is not analytically tractable, so simulation algorithms must be carried out to sample from the posterior distribution (Markov Chain Monte Carlo simulation) [33, Ch. 6–7], [45]. The Bayesian model has been implemented in WinBUGS [33, Ch. 8], [46]. We fixed a burnin period of 75,000 iterations and simulated 150,000 samples of the parameters. We executed two chains with different initial values to assess convergence. The computer timing was 47 seconds for the burnin period, plus 94 seconds for the later 150,000 samples. In Table 3, we show a descriptive analysis of the posterior distributions. In Figure 3 and Figure 4, the posterior density function is plotted for each of the parameters. Figure 5 presents the means and credible intervals from the posterior predictive distributions at times $t_1, \ldots, t_N$. We observe that the means provide good estimations for the data. Moreover, the credible intervals contain all data points.

Posterior distributions for *R. capsulatus*				Posterior distributions for *C. vibrioforme*			
Parameter	Mean	sd	0.95 interval	Parameter	Mean	sd	0.95 interval
r	0.619	0.039	$(0.544, 0.699)$	r	0.321	0.025	$(0.274, 0.373)$
y_0	0.126	0.034	$(0.069, 0.203)$	y_0	0.031	0.015	$(0.010, 0.069)$
K	5.560	0.164	$(5.243, 5.888)$	K	7.421	0.098	$(7.229, 7.614)$

Table 3. Descriptive table for the posterior distributions of the parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1).

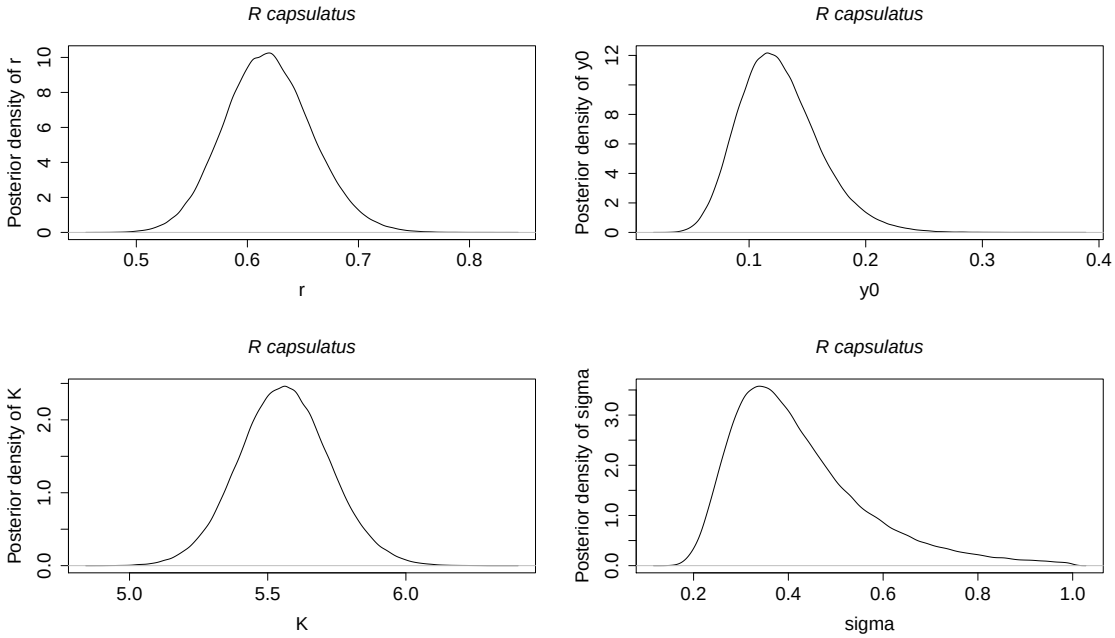

Figure 3. Posterior distributions for the model parameters of *R. capsulatus* under the random logistic model (1).

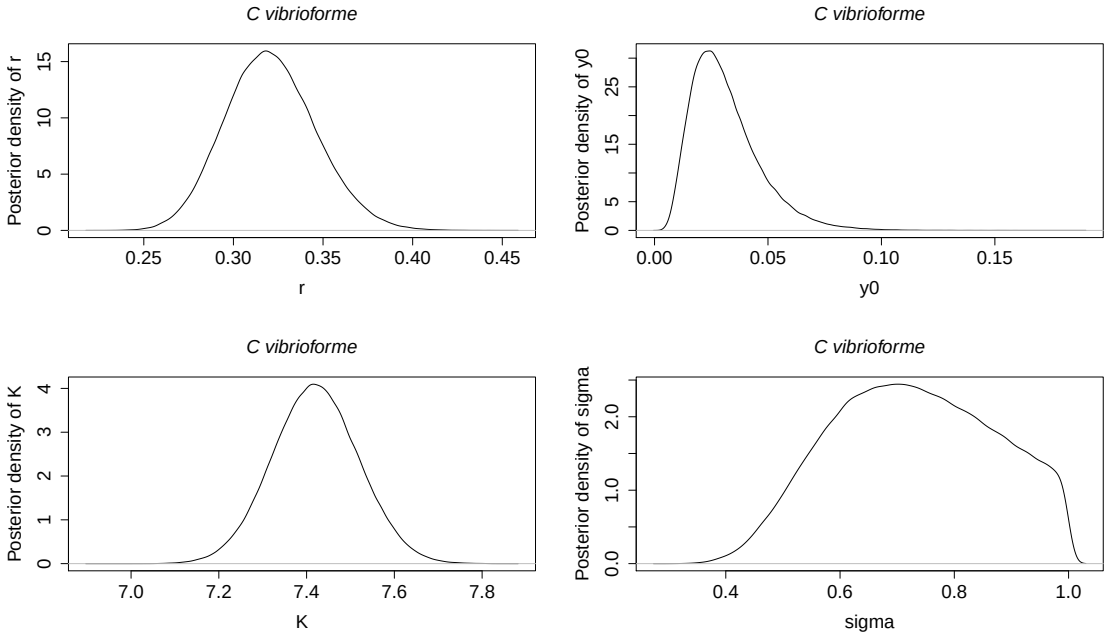

Figure 4. Posterior distributions for the model parameters of *C. vibrioforme* under the random logistic model (1).

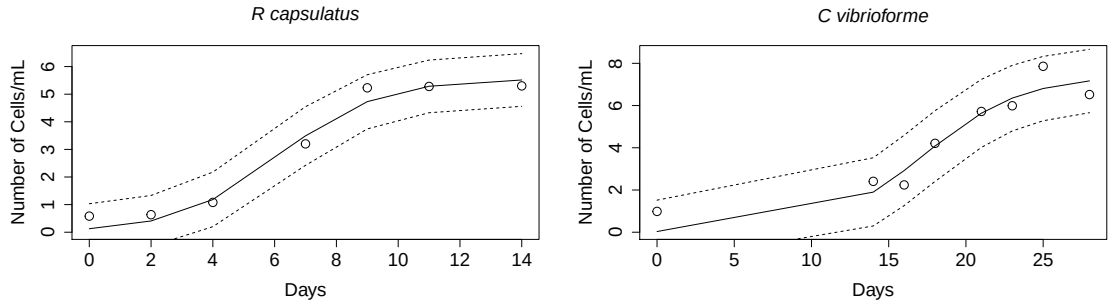

Figure 5. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1). The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

6.2. Random logistic model and gPC expansions

Consider the logistic model (1) and its randomized solution (2). We use Bayesian model (9)–(10) as an approximation of the Bayesian approach (4)–(6), with the same prior distributions as in (13), just as a test example of the theory exposed. As before, the posterior distribution of the model parameters is not analytically tractable, so we have used WinBUGS to simulate samples from the posterior distributions. We simulated 150,000 samples of the model parameters, after having removed the first 75,000 iterations (burnin period). We executed two chains with distinct initial iterates to evaluate convergence. The execution timing was 47 seconds for the burnin period, plus 94 seconds for the later 150,000 samples. In Table 4, Figure 6, Figure 7 and Figure 8, we show the results from the Bayesian inference with $p = 2$. Observe that the estimates are similar to those from Table 3, Figure 3, Figure 4 and Figure 5, although $p = 2$ is a small order of truncation. This is due to the spectral convergence of the stochastic Galerkin projection.

Posterior distributions for *R. capsulatus*				Posterior distributions for *C. vibrioforme*			
Parameter	Mean	sd	0.95 interval	Parameter	Mean	sd	0.95 interval
r	0.6215	0.04073	(0.5431, 0.7028)	r	0.3338	0.02593	(0.2837, 0.3842)
y_0	0.1262	0.03524	(0.06632, 0.2043)	y_0	0.02825	0.01403	(0.003839, 0.05829)
K	5.557	0.1652	(5.238, 5.888)	K	7.418	0.0989	(7.226, 7.613)

Table 4. Descriptive table for the posterior distributions of the parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1) with gPC expansions.

6.3. Improvement of the logistic model

Consider the new model (11), in which the Malthusian growth rate is substituted by taking into account squared abundance. We perform a least squares fitting (3) to find the optimal estimates for r, y_0 and K. In Table 5, we show the estimates and the residual squared error. We observe that the error is much smaller for (11) than for the logistic model (1), especially for the *R. capsulatus* population. Figure 9 shows how accurate is the approximation with model (11). Hence, from a deterministic standpoint, taking into account interactions instead of total abundance allows a better modeling.

Concerning the new model (12), in which we take into account the death rate, we also perform a least squares fitting (3) to get the optimal estimates for r, y_0, K and δ. Table 6 and Figure 10 present the results. The deterministic fitting improves that of model (11).

Parameters for *R. capsulatus*				Parameters for *C. vibrioforme*			
r	y_0	K	residual	r	y_0	K	residual
0.327079	0.479572	5.3322	0.0238877	0.0847533	0.690599	7.0967	1.85554

Table 5. Parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11).

6.4. Random new model and gPC expansions

Consider the new model (11) with random input coefficients. Using the Galerkin projection technique, we use Bayesian model (9)–(10) to quantify the uncertainty of the solution process. We have set the prior distributions (13) for the parameters,

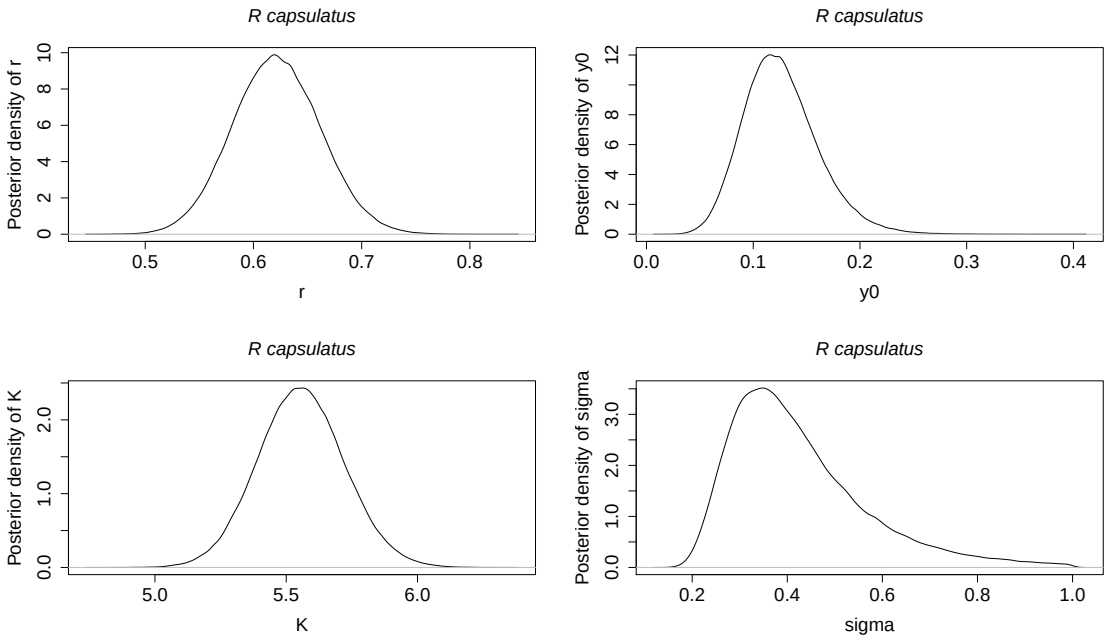

Figure 6. Posterior distributions for the model parameters of *R. capsulatus* under the random logistic model (1) with gPC expansions.

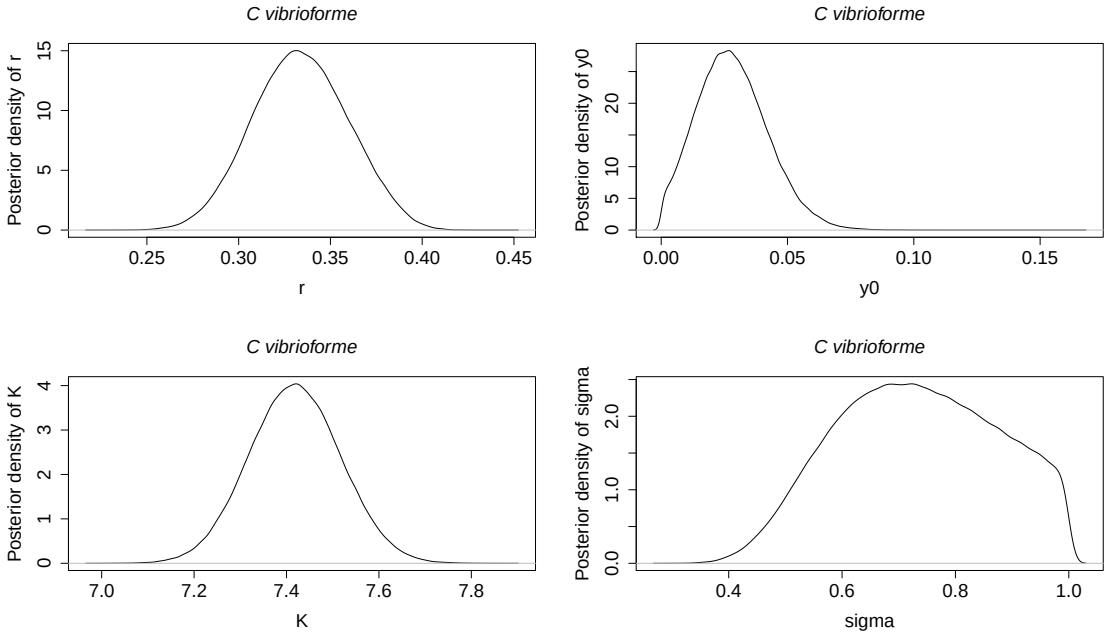

Figure 7. Posterior distributions for the model parameters of *C. vibrioforme* under the random logistic model (1) with gPC expansions.

Parameters for *R. capsulatus*					Parameters for *C. vibrioforme*				
r	y_0	K	δ	residual	r	y_0	K	δ	residual
0.43931	0.557972	5.60822	0.126042	0.0099536	0.116177	0.961639	7.6808	0.0694181	1.81952

Table 6. Parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12).

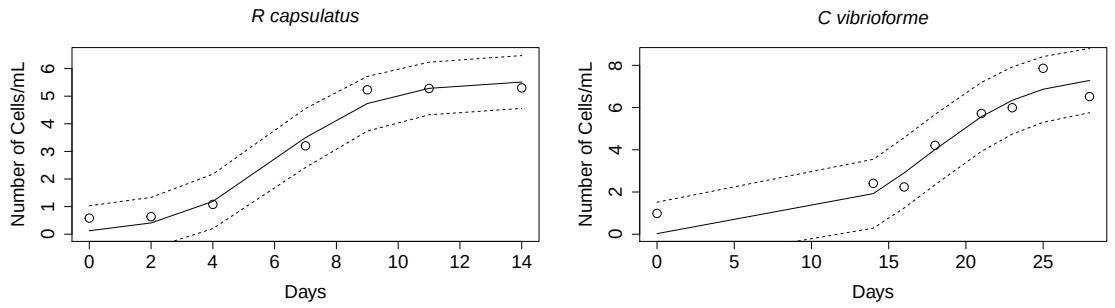

Figure 8. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

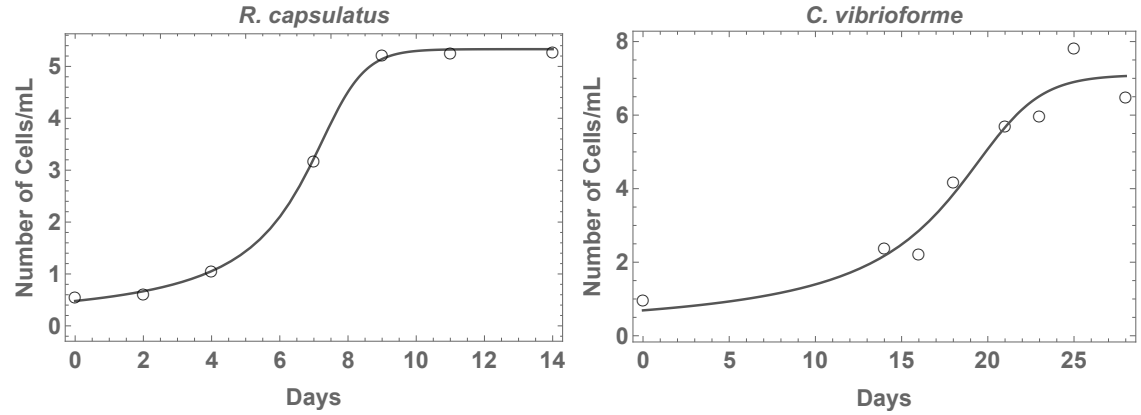

Figure 9. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11). The real data is denoted by ∘ and the fitting is given by the black continuous line.

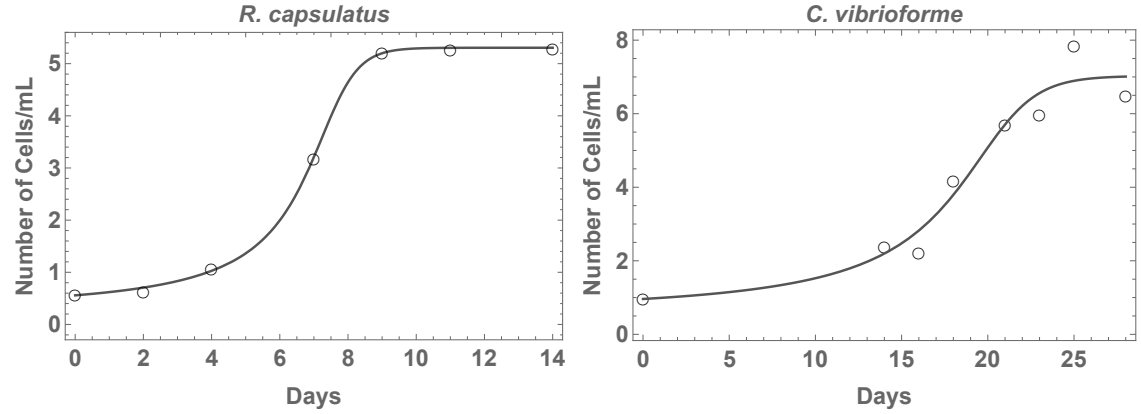

Figure 10. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12). The real data is denoted by ∘ and the fitting is given by the black continuous line.

with the values $\alpha_r = 42.7927$, $\beta_r = 130.8316$, $\alpha_{y_0} = 91.99572$, $\beta_{y_0} = 191.8288$, $\alpha_K = 710.8089$ and $\beta_K = 133.305$ for the *R. capsulatus* population; $\alpha_r = 2.873249$, $\beta_r = 33.90132$, $\alpha_{y_0} = 190.7708$, $\beta_{y_0} = 276.2396$, $\alpha_K = 1259.079$ and $\beta_K = 177.4175$ for the *C. vibrioforme* population; and the error $\sigma \sim \text{Unif}(0, 1)$. These values may be calculated as in the logistic model, by imposing the mean of the gamma distribution, α/β, to be the least squares fitting estimate from Table 5, and the variance of the gamma distribution, α/β^2, to be the desired dispersion for the parameter (this is more subjective). With a burnin period of 75,000 iterations, plus 150,000 samples for the model coefficients, and with two chains to assess convergence, the computational time was 47 seconds for the burnin period, plus 94 seconds for the latter 150,000 samples. In Table 7, Figure 11, Figure 12

and Figure 13, we show the results from the Bayesian inference with $p = 2$. Observe that the solid lines from the figures behave similarly to the deterministic fittings from Figure 9. Moreover, the credible regions contain all data points, therefore the Bayesian model is appropriate for our data set.

Finally, consider the new model (12) with uncertainty. We combine the stochastic Galerkin projection technique and the Bayesian model (9)–(10) to quantify the uncertainty for the solution stochastic process. As prior distributions for r, y_0, K, δ and σ, we have set (13) for r, y_0, K and σ, and $\delta \sim \text{Gamma}(\alpha_\delta, \beta_\delta)$. For the numerical experiments, we have set $\alpha_r = 77.19731$, $\beta_r = 175.724$, $\alpha_{y_0} = 124.5331$, $\beta_{y_0} = 223.1888$, $\alpha_K = 786.3033$, $\beta_K = 140.2055$, $\alpha_\delta = 6.354634$ and $\beta_\delta = 50.4168$ for the *R. capsulatus* population; $\alpha_r = 5.398838$, $\beta_r = 46.4708$, $\alpha_{y_0} = 369.8998$, $\beta_{y_0} = 384.6556$, $\alpha_K = 1474.867$, $\beta_K = 192.02$, $\alpha_\delta = 48.18873$ and $\beta_\delta = 694.181$ for the *C. vibrioforme* population; and the error $\sigma \sim \text{Unif}(0, 1)$. As usual, we set a burnin period of 75,000 iterations, plus 150,000 samples for the model parameters, and with two chains to check convergence. The computational time was 47 seconds for the burnin period, plus 94 seconds for the latter 150,000 samples (the same time as the previous models, despite having one more parameter). In Table 8, Figure 14, Figure 15 and Figure 16, we present the results from the Bayesian inference with order of truncation $p = 2$. Observe that the credible intervals contain all data measurements.

Posterior distributions for *R. capsulatus*				Posterior distributions for *C. vibrioforme*			
Parameter	Mean	sd	0.95 interval	Parameter	Mean	sd	0.95 interval
r	0.3468	0.03955	$(0.2661, 0.3993)$	r	0.05636	0.01112	$(0.03632, 0.07967)$
y_0	0.4404	0.03182	$(0.4037, 0.4867)$	y_0	0.7038	0.05447	$(0.602, 0.815)$
K	5.352	0.162	$(5.253, 5.733)$	K	7.369	0.1535	$(7.008, 7.608)$

Table 7. Descriptive table for the posterior distributions of the parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11) with gPC expansions.

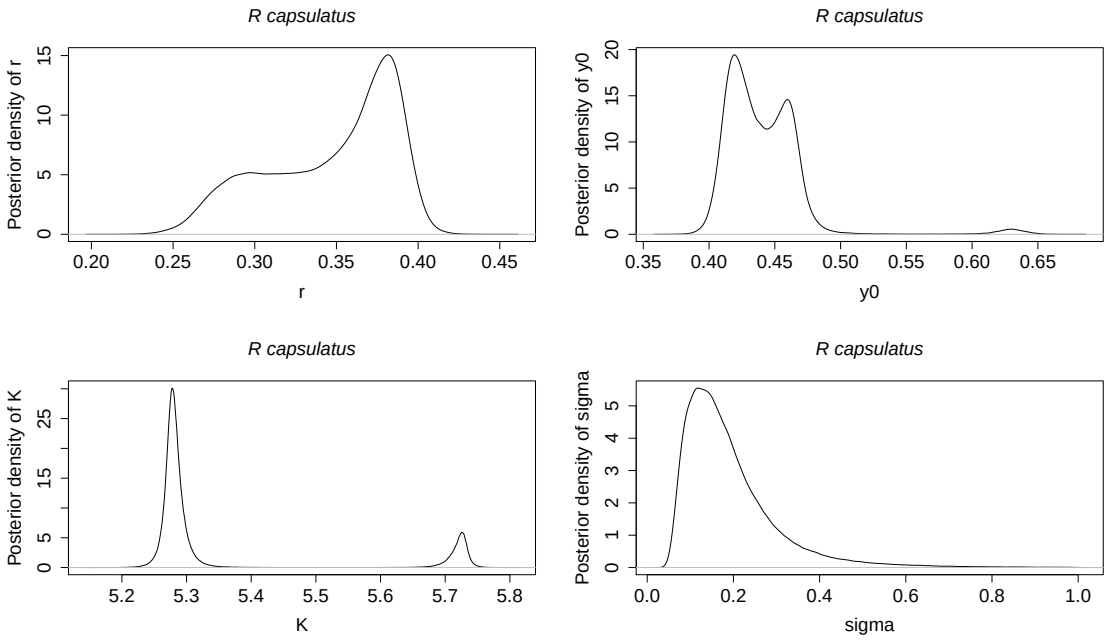

Figure 11. Posterior distributions for the model parameters of *R. capsulatus* under the new model (11) with gPC expansions.

7. Conclusions

Mathematical models for biological population growth are important to understand and generalize the results to other situations and to make predictions. Due to the inherent uncertainty associated to biological phenomena (errors in the laboratory experiments, lack of information, missed data, etc.), randomness must be introduced in the model. In this paper, we have studied a random differential model of growth of anaerobic photosynthetic bacteria. In the laboratory experiment, actual measurements for two species of bacteria were collected: *R. capsulatus* and *C. vibrioforme*. A previous article by D. Stanescu et al. [*Electronic*

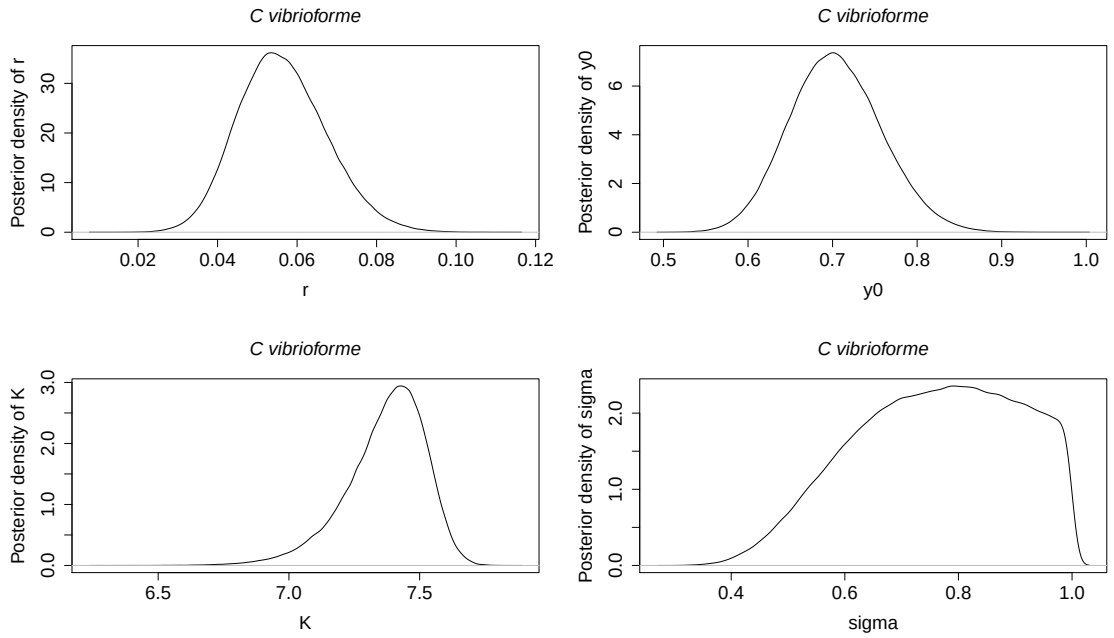

Figure 12. Posterior distributions for the model parameters of *C. vibrioforme* under the new model (11) with gPC expansions.

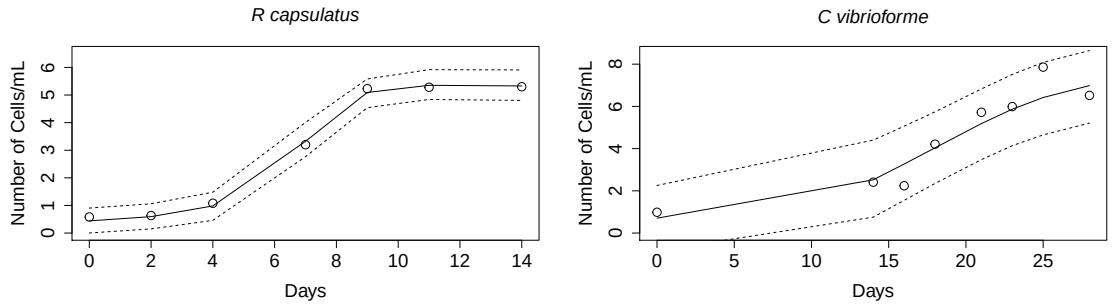

Figure 13. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

Posterior distributions for *R. capsulatus*				Posterior distributions for *C. vibrioforme*			
Parameter	Mean	sd	0.95 interval	Parameter	Mean	sd	0.95 interval
r	0.4291	0.03444	(0.3621, 0.4986)	r	0.07133	0.01208	(0.04899, 0.09645)
y_0	0.5322	0.05639	(0.467, 0.6763)	y_0	0.9604	0.04884	(0.8672, 1.058)
K	5.618	0.121	(5.339, 5.815)	K	7.908	0.1513	(7.554, 8.152)
δ	0.1068	0.04146	(0.04051, 0.1954)	δ	0.06487	0.008284	(0.04913, 0.08149)

Table 8. Descriptive table for the posterior distributions of the parameters for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12) with gPC expansions.

Transactions on Numerical Analysis, 34, 44–58 (2009)] considered a logistic model to explain bacterial growth in both populations, based on Malthusian exponential growth model and competitiveness when there is scarcity of nutrients (mainly light and CO_2). In our article, we have improved the fit of the deterministic logistic model by assuming that the growth rate is proportional to the squared abundance of microorganisms, and by taking into account the start of the death phase in the kinetic curve. Instead of introducing uncertainty into the model by using arbitrary distributions for the coefficients, we have utilized a Bayesian inverse approach for parameter estimation. Since our model does not have an explicit solution, one would need to solve it for each sampling point of the Markov Chain Monte Carlo algorithm. However, gPC expansions together with the stochastic Galerkin

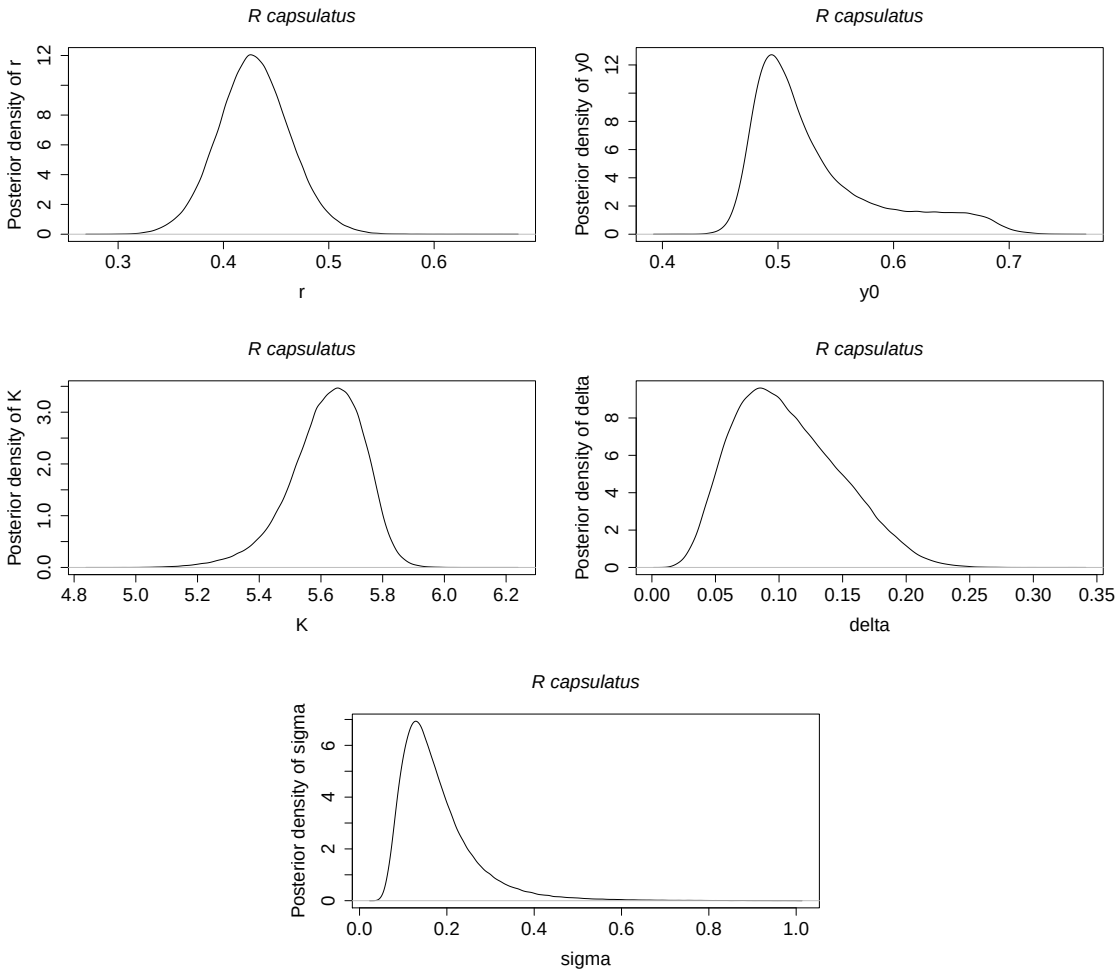

Figure 14. Posterior distributions for the model parameters of *R. capsulatus* under the new model (12) with gPC expansions.

projection technique have allowed accelerating the Bayesian inference. Spectral convergence of the Galerkin projection implies exponential convergence rate for the corresponding prior distributions in the sense of the Kullback-Leibler divergence, even when the variance of the error is supposed random with a prior distribution. This fact has permitted obtaining reliable results for the posterior distributions of the coefficients and the posterior predictive distribution, so that it is possible to computationally quantify the uncertainty for the bacteria population growth.

Acknowledgements

This work has been supported by Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. Marc Jornet acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

Conflict of Interest Statement

The authors declare that there is no conflict of interests regarding the publication of this article.

References

1. Monod J. The growth of bacterial cultures. *Annu Rev Microbiol.* 1949;3(1):371-394.

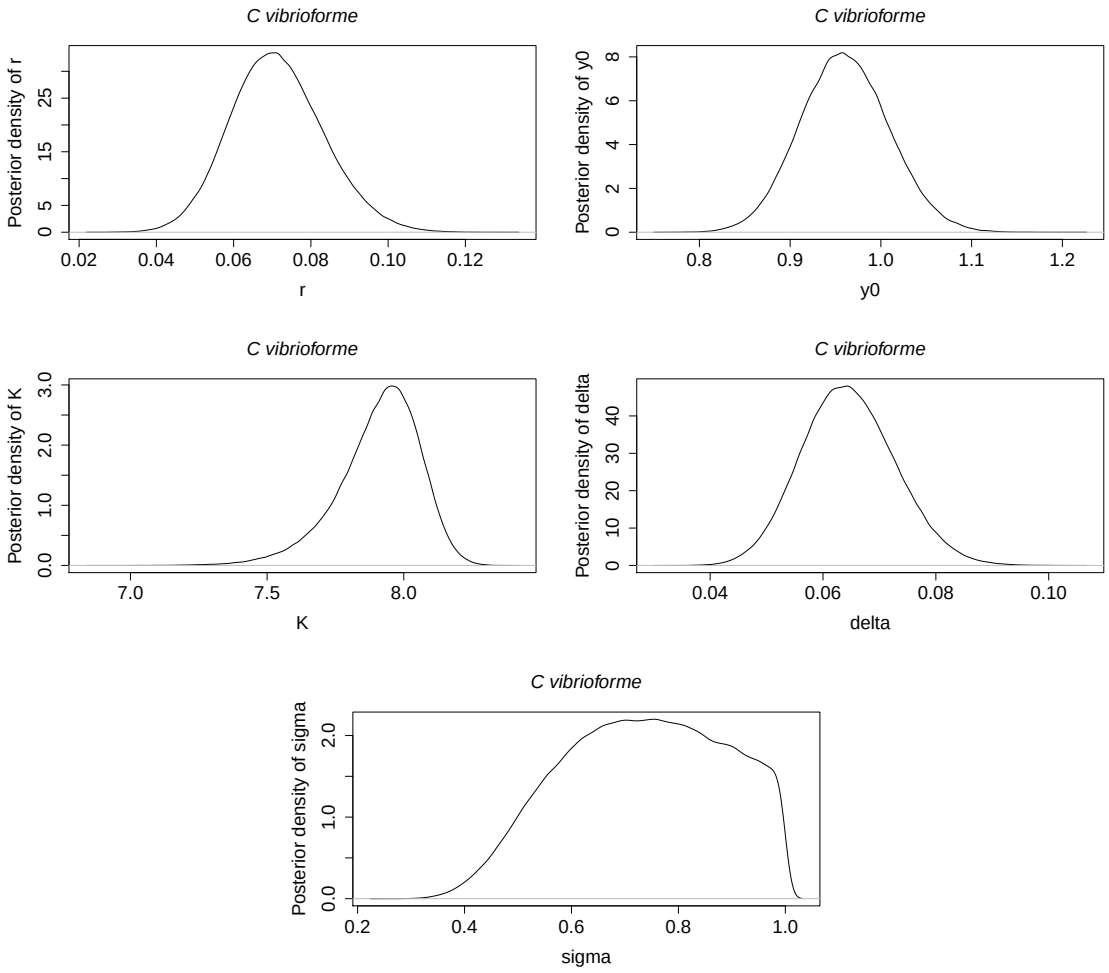

Figure 15. Posterior distributions for the model parameters of *C. vibrioforme* under the new model (12) with gPC expansions.

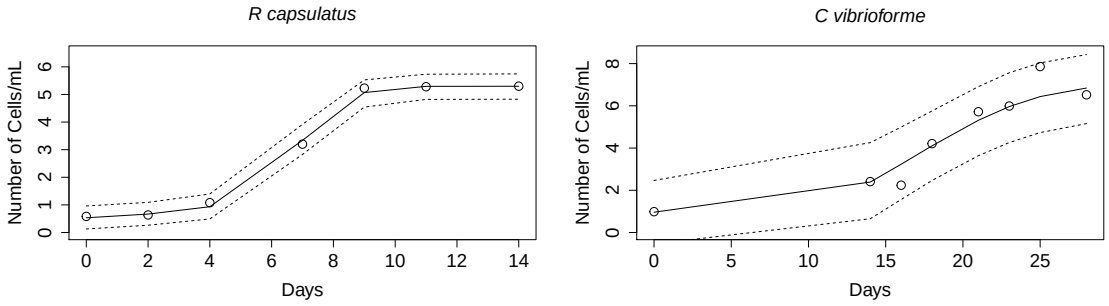

Figure 16. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

2. Novick A. Growth of bacteria. *Annu Rev Microbiol.* 1955;9(1):97-110.
3. Dutta R. *Fundamentals of Biochemical Engineering*. India: Springer; 2008.
4. Hethcote HW. The mathematics of infectious diseases. *SIAM Rev.* 2000;42(4):599-653.
5. Murray JD. *Mathematical Biology I*. Springer; 2002.
6. Levin SA, Hallam TG, Gross LJ. *Applied Mathematical Ecology*. Volume 18, Springer Science & Business Media; 2012.
7. Zwietering MH, Jongenburger I, Rombouts FM, Van't Riet K. Modeling of the bacterial growth curve. *Appl Environ Microb.* 1990;56(6):1875-1881.

8. McKellar RC. Development of a dynamic continuous-discrete-continuous model describing the lag phase of individual bacterial cells. *J Appl Microbiol*. 2001;90(3):407-413.
9. Fujikawa H, Kai A, Morozumi S. A new logistic model for bacterial growth. *J Food Hyg Soc Jpn*. 2003;44(3):155-160.
10. Juška A, Gedminienè G, Ivanec R. Growth of microbial populations. Mathematical modeling, laboratory exercises, and model-based data analysis. *Biochem Mol Biol Edu*. 2006;34(6):417-422.
11. Strand JL. Random Ordinary Differential Equations. *J Differ Equations*. 1970;7:538-553.
12. Soong TT. *Random Differential Equations in Science and Engineering*. Academic Press; 1973.
13. Øksendal B. *Stochastic Differential Equations*. Springer; 2003.
14. Fishman G. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Science & Business Media; 2013.
15. Cortés JC, Navarro-Quiles A, Romero JV, Roselló MD. Probabilistic solution of random autonomous first-order linear systems of ordinary differential equations. *Rom Rep Phys*. 2016;68(4):1397-1406.
16. Dorini FA, M. S. Cecconello, and M. B. Dorini. On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density. *Commun Nonlinear Sci*. 2016;33:160-173.
17. Dorini FA, Cunha MCC. Statistical moments of the random linear transport equation. *J Comput Phys*. 2008;227:8541-8550.
18. Casabán MC, Cortés JC, Romero JV, Roselló MD. Probabilistic solution of random SI-type epidemiological models using the Random Variable Transformation technique. *Commun Nonlinear Sci*. 2015;24(1-3):86-97.
19. Hussein A, Selim MM. Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique. *Appl Math Comput*. 2012;218(13):7193-7203.
20. Calatayud J, Cortés JC, Jornet M. The damped pendulum random differential equation: A comprehensive stochastic analysis via the computation of the probability density function. *Physica A*. 2018;512:261-279.
21. Calatayud J, Cortés JC, Jornet M. Uncertainty quantification for random parabolic equations with non-homogeneous boundary conditions on a bounded domain via the approximation of the probability density function. *Math Method Appl Sci*. 2018. DOI: 10.1002/mma.5333.
22. Xiu D. *Numerical Methods for Stochastic Computations. A Spectral Method Approach*. Princeton University Press, New York: Cambridge Texts in Applied Mathematics; 2010.
23. Xiu D, Karniadakis GE. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput*. 2002;24(2):619-644.
24. Chen-Charpentier BM, Cortés JC, Licea JA, Romero JV, Roselló MD, Santonja FJ, Villanueva RJ. Constructing adaptive generalized polynomial chaos method to measure the uncertainty in continuous models: A computational approach. *Math Comput Simulat*. 2015;109:113–129.
25. Cortés JC, Romero JV, Roselló MD, Villanueva RJ. Improving adaptive generalized polynomial chaos method to solve nonlinear random differential equations by the random variable transformation technique. *Commun Nonlinear Sci*. 2017;50:1-15.
26. Stanescu D, Chen-Charpentier BM, Jensen BJ, Colberg PJS. Random coefficient differential models of growth of anaerobic photosynthetic bacteria. *Electron T Numer Ana*. 2009; 34:44-58.
27. Chen-Charpentier BM, Stanescu D. Epidemic models with random coefficients. *Math Comput Model*. 2010; 52(7-8):1004-1010.
28. Stanescu D, Chen-Charpentier BM. Random coefficient differential equation models for bacterial growth. *Math Comput Model*. 2009; 50(5-6):885-895.
29. Santonja F, Chen-Charpentier BM. Uncertainty quantification in simulations of epidemics using polynomial chaos. *Comput Math Method M*. 2012; 2012.
30. González-Parra G, Chen-Charpentier BM, Arenas AJ. Polynomial chaos for random fractional order differential equations. *Appl Math Comput*. 2014;26:123-130.
31. Calatayud J, Cortés JC, Jornet M, Villanueva RJ. Computational uncertainty quantification for random time-discrete epidemiological models using adaptive gPC. *Math Method Appl Sci*. 2018;41:9618-9627.
32. Calatayud J, Cortés JC, Jornet M. On the convergence of adaptive gPC for non-linear random difference equations: Theoretical analysis and some practical recommendations. *J Nonlinear Sci App*. 2018; 11(9):1077-1084.
33. Lesaffre E, Lawson AB. *Bayesian Biostatistics*. John Wiley & Sons; 2012.
34. Mohammad-Djafari A. Bayesian inference for inverse problems. *AIP Conference Proceedings*. 2002; 617:477-496.
35. Corberán-Vallet A, Santonja FJ, Jornet-Sanz M, Villanueva RJ. Modeling chickenpox dynamics with a discrete time bayesian stochastic compartmental model. *Complexity*. 2018; 2018.
36. Marzouk YM, Najm HN, Rahn LA. Stochastic spectral methods for efficient bayesian solution of inverse problems. *J Comput Phys*. 2007;224(2):560-586.
37. Marzouk Y, Xiu D. A stochastic collocation approach to bayesian inference in inverse problems. *Commun Comput Phys*. 2009;6(4):826-847.
38. Malthus TR. *An Essay on the Principal of Population*. Oxford: Oxford World's Classics Paperbacks, Oxford University Press; 1999.
39. Turchin P. Does population ecology have general laws? *Oikos*. 2001;94(1):17-26.
40. Verhulst PF. Notice sur la loi que la population suit dans son accroissement. *Corr Math et Phys*. 1838; 10:113-121.
41. Roos AM. *Modeling Population Dynamics*. Netherlands: Notes from the University of Amsterdam; 2014.
42. Lehmann EL, Casella G. *Theory of Point Estimation*. Springer Science & Business Media; 2006.
43. Casella G. An introduction to empirical Bayes data analysis. *Am Stat*. 1985;39(2):83-87.
44. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC; 2010.
45. Tarantola A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, volume 89; 2005.
46. Lunn DJ, Thomas A, Best N, Spiegelhalter D. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput*. 2000;10(4):325-337.
47. Depaoli S, Clifton JP, Cobb PR. Just Another Gibbs Sampler (JAGS): Flexible Software for MCMC Implementation. *J Educ Behav Stat*. 2016;41(6):628-649.
48. Johnson TR, Kuhn KM. Bayesian Thurstonian models for ranking data using JAGS. *Behav Res Methods*. 2013;45(3):857-872
49. Plummer M. Jags: A program for analysis of bayesian graphical models using gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*. Volume 124, Vienna, Austria; 2003.
50. Stokes M, Chen F, Gunes F. An introduction to Bayesian analysis with SAS/STATR software. *Proceedings of the SAS Global Forum*

2014 Conference, SAS Institute Inc, Cary, USA (available at https://support.sas.com/resources/papers/proceedings14/SAS400-2014.pdf). Citeseer; 2014.
51. Giraud L, Langou J, Rozloznik M. The loss of orthogonality in the Gram-Schmidt orthogonalization process. *Comput Math Appl.* 2005;50(7):1069-1075.
52. Ernst OG, Mugler A, Starkloff HJ, Ullmann E. On the convergence of generalized polynomial chaos expansions. *ESAIM-Math Model Num.* 2012;46(2):317-339.
53. Shi W, Zhang C. Error analysis of generalized polynomial chaos for nonlinear random ordinary differential equations. *Appl Numer Math.* 2012;62(12):1954-1964.
54. Pukelsheim F. The three sigma rule. *Am Stat.* 1994;48(2):88-91.

Figure legends.

Figure 1. Population size of *R. capsulatus* (left) and *C. vibrioforme* (right).

Figure 2. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the logistic model. The real data is denoted by ∘ and the fitting is given by the black continuous line.

Figure 3. Posterior distributions for the model parameters of *R. capsulatus* under the random logistic model (1).

Figure 4. Posterior distributions for the model parameters of *C. vibrioforme* under the random logistic model (1).

Figure 5. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1). The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

Figure 6. Posterior distributions for the model parameters of *R. capsulatus* under the random logistic model (1) with gPC expansions.

Figure 7. Posterior distributions for the model parameters of *C. vibrioforme* under the random logistic model (1) with gPC expansions.

Figure 8. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the random logistic model (1) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

Figure 9. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11). The real data is denoted by ∘ and the fitting is given by the black continuous line.

Figure 10. Least squares fitting of *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12). The real data is denoted by ∘ and the fitting is given by the black continuous line.

Figure 11. Posterior distributions for the model parameters of *R. capsulatus* under the new model (11) with gPC expansions.

Figure 12. Posterior distributions for the model parameters of *C. vibrioforme* under the new model (11) with gPC expansions.

Figure 13. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (11) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.

Figure 14. Posterior distributions for the model parameters of *R. capsulatus* under the new model (12) with gPC expansions.

Figure 15. Posterior distributions for the model parameters of *C. vibrioforme* under the new model (12) with gPC expansions.

Figure 16. Model fitting for *R. capsulatus* (left) and *C. vibrioforme* (right) under the new model (12) with gPC expansions. The real data is denoted by ∘, the fitting is given by the black continuous line and the 0.95 credible interval is drawn with dashed lines.