Statistics and Computing (2026) 36:126
https://doi.org/10.1007/s11222-026-10878-w

ORIGINAL PAPER

Likelihood-based inference for the Gompertz model with Poisson errors

Paolo Onorati[1] · **Sofia Ruiz-Suarez**[2] · **Radu V. Craiu**[2]

Received: 4 November 2025 / Accepted: 4 April 2026

Abstract
Population dynamics models play an important role in several fields, such as actuarial science, demography, and ecology, as they help explain past fluctuations and predict future populations. Statistical inference for these models can be difficult when, in addition to the process' inherent stochasticity, one also needs to account for sampling error. Ignoring the latter can lead to biases in the estimation, which in turn can produce erroneous conclusions about the system's behavior. The Gompertz model with Poisson errors is widely used to infer population abundance dynamics, but a full likelihood approach can be computationally prohibitive. We close this gap by developing efficient computational tools for statistical inference in the Gompertz model with Poisson sampling error, based on the full likelihood. The approach is illustrated using models in the Bayesian and frequentist paradigms. Performance is illustrated with simulations and data analysis.

Keywords Bayesian inference · EM algorithm · Gompertz model · MCMC · Sampling error

1 Introduction

Population dynamics models are crucial in both applied and theoretical mathematical biology because they help to explain past population fluctuations and project future population abundances (Newman et al. 2014). These models are used to manage species conservation, understand the dynamics of biological invasions, and assess species responses to environmental changes. The latter can range from those produced by human developments to those triggered by climate change (Beyer et al. 2013). The accuracy and reliability of these models are often affected by two major components of uncertainty: i) process' stochasticity connected to, say, demographic and environmental factors, and ii) sampling error.

Ecologists have long recognized the importance of separating observation and process errors in ecological modeling, and significant progress has been achieved through the use of state-space models to analyze time series of population fluctuations (Dennis et al. 2006; Hostetler and Chandler 2015; Auger-Méthé et al. 2021). However, including both process noise and observation error in the model remains challenging as it produces significant computational difficulties. To illustrate, Staples et al. (2004) note that the sampling error increases the variability in the data, leading to positively biased estimators of process variation, and propose a restricted maximum likelihood approach to estimate both the process variation and the error variance. Subsequently, (Lindén and Knape 2009) show empirically that observation errors can change the autocorrelation structure of a time series, leading to potential biases when such errors are ignored.

The Gompertz model (Gompertz 1825) is widely used to describe and characterize population dynamics. Its use permeates multiple disciplines, including actuarial science (e.g., Butt and Haberman 2004; Baione and Levantesi 2018), demography (e.g., Alexander et al. 2017; Tai and Noymer 2018), as well as life sciences where it can be used to model the growth patterns of animals, plants, tumors, and the volume of bacteria (Winsor 1932; Tjørve and Tjørve 2017). Most importantly, the Gompertz model can be modified to account for sampling variability and, in some clearly structured problems, it is possible to explicitly formulate the likelihood function in the presence of sampling error. For example, if observation noise is assumed to follow a log-normal distri-

✉ Paolo Onorati
paolo.onorati@dauphine.psl.eu

[1] CEREMADE, Université Paris Dauphine-PSL, Place du Maréchal de Lattre de Tassigny, Paris 75016, Île-de-France, France

[2] Department of Statistical Sciences, University of Toronto, 700 University Ave, Toronto M5G 1X6, Ontario, Canada

Published online: 25 April 2026

bution, then it is feasible to write the likelihood function (Staples et al. 2004), and the model can be transformed into a linear Gaussian state-space model that can be fitted using the Kalman filter (Dennis et al. 2006). However, if the sampling error is not log-normal, the expression of the likelihood becomes more complex. Relevant for this paper is the study by Lele (2006), which shows that likelihood-based inference for a stationary Gompertz model with Poisson sampling errors requires the calculation of a high-dimensional integral, resulting in a prohibitive computational burden. He proposes a composite likelihood approach that reduces the dimension of the integral and makes the computation manageable for longer time series. However, replacing the full likelihood with a composite one comes with an inferential cost because the resulting confidence or credible intervals will not have nominal coverage. Subsequently, Lele et al. (2007) employed a data augmentation strategy for both the Gompertz model with normal errors and the Ricker model with Poisson errors. They used a generic random walk Metropolis-Hastings algorithm, implemented through the WinBUGS library (Lunn et al. 2000), to sample from the posterior distribution of the parameters and latent variables.

The main contribution of this paper is the introduction of computational methods that allow for a full likelihood-based analysis of the Gompertz model with Poisson errors. Central to our approach is an efficient algorithm for computing the likelihood, which allows us to conduct both frequentist and Bayesian inference. The developed methods are available in a new R package published on GitHub. The paper is organized as follows. In Section 2 we introduce the notation and the statistical model. Section 3 contains the estimation procedure and the computational algorithm for the frequentist approach, which is based on a simulation-aided EM algorithm (Dempster et al. 1977; Wei and Tanner 1990). The Bayesian analysis relies on a new MCMC sampler which is described in Section 4. The paper continues with a Section 5 of numerical experiments that includes simulations and a data analysis. A discussion of open questions and future research directions closes the paper in Section 6.

2 Statistical Model

Consider a population observed at times $t = 1, 2, \ldots, T$. The Gompertz model with stochastic errors establishes a probabilistic model for the evolution of N_t, the abundance of the population at time t, for $1 \leq t \leq T$. Specifically, assume that for all $1 \leq t \leq T-1$,

$$N_{t+1} = N_t \exp(a + b \log(N_t) + \varepsilon_{t+1}), \tag{1}$$

where a is an individual intrinsic growth rate parameter, b links the current measure with the past, and ε_t are independent errors normally distributed: $\varepsilon_t \sim N(0, \sigma^2)$. The population abundance in (1) can clearly take non-integer values. We interpret N_t as a continuous non-negative random variable that approximates the true value. The logarithmic transformation $Z_t = \log(N_t)$ in (1) leads to

$$Z_{t+1} = a + (1+b)Z_t + \varepsilon_{t+1}. \tag{2}$$

Equation (2) implies that the Gompertz model corresponds, on the logarithmic scale, to an autoregressive process of order 1. In this paper, we assume that $|1+b| < 1$, which is a condition required for stationarity of the AR(1) model (2). Without stationarity, the statistical model falls outside the scope of this article as one cannot assume a stationary distribution for Z_t and a completely different approach to inference is needed.

Under a stationary regime, the n-dimensional marginal distribution of the log transformed population abundance, Z_t, has a multivariate normal distribution $(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_n})' \sim N_n(\boldsymbol{\mu}, \Sigma)$ where

$$\boldsymbol{\mu} = (\theta_1, \theta_1, \ldots, \theta_1)', \ \Sigma = \theta_2 B, \ B_{jk} = (1+b)^{|i_j - i_k|}, \tag{3}$$

where $\theta_1 = -a/b$ and $\theta_2 = -\sigma^2/\big[b(2+b)\big]$. In what follows, we denote $\boldsymbol{\theta} = (\theta_1, \theta_2, b)$.

In this paper, we consider the case where the exact population abundance N_t is unknown. Instead, we assume that at time $t \in \{1, \ldots, T\}$, N_t^* is the observed abundance with conditional density $\pi(N_t^*|N_t, \phi)$, indexed by the parameter ϕ. Following the lines of the Gompertz approximation (1), we allow statistical models that will treat N_t^* as a non-negative continuous random variable, in line with past formulations (see Dennis et al. 2006; 2006; Lele 2006, for example). We assume that the random variables $\{N_t^* | \ 1 \leq t \leq T\}$ are independent conditionally on $\{N_t : \ 1 \leq t \leq T\}$. The observed-data likelihood is obtained by taking advantage of the conditional distribution of $N_t^*|N_t$ and averaging the missing true population values $\{N_1, \ldots, N_T\}$.

$$L(\boldsymbol{\theta}, \phi|\boldsymbol{N}^*) = \int\int\ldots\int \pi(N_1^*|N_1, \phi)\pi(N_2^*|N_2, \phi) \ldots \pi(N_T^*|N_T, \phi)\pi(\boldsymbol{N}|\boldsymbol{\theta})\, dN_1\, dN_2 \ldots dN_T, \tag{4}$$

where $\pi(\boldsymbol{N}|\boldsymbol{\theta})$ indicates the joint distribution of the unobserved time series of exact population abundances.

If the sampling error distribution $\pi(N_t^*|N_t, \phi)$ is assumed to be log-normal, then it is possible to explicitly formulate the complete likelihood function as in Dennis et al. (2006). However, for other sampling error distributions, the likelihood (4) cannot be expressed in closed form using elementary functions. Here we focus on the case of a Poisson sampling error distribution, i.e. $N_t^* \sim \text{Poisson}(N_t)$, for all $1 \leq t \leq T$. In this case, no additional parameters ϕ are associated with

the sampling error distribution. However, the use of a Poisson distribution leads to computational challenges because the likelihood is not available in closed form, as discussed by Lele (2006). In Section 3 we present a computational method that uses the expression (4) to perform statistical inference. Our approach is very general and can also be implemented in a Bayesian framework, as we show in Section 4. The performance of the proposed algorithms is examined in Section 5.1.

3 Maximum likelihood based inference

We describe a frequentist inferential method for the Gompertz model with Poisson observation errors. In particular, we are interested in computing the maximum likelihood estimator (MLE) for $\boldsymbol{\theta}$ and producing confidence intervals. The proposed algorithm takes advantage of the data augmentation strategy implied by (4). It is apparent that if the population abundances $\boldsymbol{N}$ were available, then the inference would be straightforward, since it would be based on the analysis of a first-order autoregressive model for $Z_t = \log N_t$. However, since population abundances $N_t, t \in \{1, 2, ..., T\}$ are unobserved latent variables, we adopt a data augmentation approach.

The expectation-maximization (EM) algorithm (Dempster et al. 1977) is a powerful iterative procedure that is used to estimate the parameters of models with missing data. In our case, $\boldsymbol{N}^*$ is the vector of observed data, $\boldsymbol{Z} = \log \boldsymbol{N}$ is the vector of missing data, and $\boldsymbol{\theta}^{(k)}$ is the value of the parameter estimate in the k-th iteration. The algorithm iteratively computes $\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ where the Q function is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathbb{E}[\log \pi(\boldsymbol{N}^*, \boldsymbol{Z}|\boldsymbol{\theta})|\boldsymbol{N}^*, \boldsymbol{\theta}^{(k)}]\,. \tag{5}$$

Calculating the conditional expectation in (5) represents the E-step and its optimization as a function of $\boldsymbol{\theta}$ is the M-step. The algorithm cycles between the E- and M-steps until the difference between two consecutive $\boldsymbol{\theta}^{(k)}$'s becomes negligible.

In the specific case of the Gompertz model with Poisson sampling error, the complete data log-likelihood can be separated as

$$\log \pi(\boldsymbol{N}^*, \boldsymbol{Z}|\boldsymbol{\theta}) = \log \pi(\boldsymbol{N}^*|\boldsymbol{Z}) + \log \pi(\boldsymbol{Z}|\boldsymbol{\theta})\,,$$

and only the second term depends on $\boldsymbol{\theta}$. Nevertheless, we cannot compute in closed form the expectation required in the E-step, so a Monte Carlo strategy must be used. The approach was introduced by Wei and Tanner (1990) and was further expanded by McCulloch (1994); Chan and Ledolter (1995); Booth and Hobert (1999) and Caffo et al. (2005). The algorithm proposed here falls within the class of Markov chain Monte Carlo EM (MCMC-EM), since the draws required to complete the E-step are obtained using a Gibbs sampler. Specifically, we approximate the expectation with the average

$$\mathbb{E}[\log \pi(\boldsymbol{Z}|\boldsymbol{\theta})|\boldsymbol{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}] \approx \frac{1}{J}\sum_{j=1}^{J} \log \pi(\widetilde{\boldsymbol{Z}}_j|\boldsymbol{\theta})\,, \tag{6}$$

where $\widetilde{\boldsymbol{Z}}_j$ are sampled from $\boldsymbol{Z}|\boldsymbol{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^k$ via MCMC. The MCMC estimate is maximized in the M-step. For the latter, we use the L-BFGS-B algorithm because it allows for constrained optimization, as $-2 < b < 0$. The likelihood of $\pi(\boldsymbol{Z}|\boldsymbol{\theta})$ approaches zero as b approaches its bounds, unless Z_1 is treated as a constant value when the autoregressive process becomes nonstationary.

We start the MCMC-EM algorithm with a value of $J = 10^3$ in (6) and, as prescribed by Booth and Hobert (1999) and Caffo et al. (2005), we increase the value of J with the number of iterations in order to increase precision in the estimate of the expected complete data log-likelihood. Specifically, we use the method proposed in Caffo et al. (2005) to determine whether an increase in J is necessary or not. When an increase is needed, we double the current value of J. Due to limitations of our computing environment, we do not allow J to be greater than 2×10^4, but such a limit can be relaxed. The conditional distribution $\boldsymbol{Z}|\boldsymbol{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ is nonstandard, so we must customize a Gibbs sampler to obtain the necessary draws. In the following, we omit the superscript $^{(k)}$ for ease of notation.

According to the Gibbs sampler design, we need to obtain samples from the full conditional distributions of each latent variable given the observed data, the current parameter values and the remaining unobserved population abundances. For brevity, we denote $\boldsymbol{Z}_{-t}$ the vector $\boldsymbol{Z}$ from which the t-th component, Z_t, is excluded. Thus, we need to sample from

$$Z_t|\boldsymbol{Z}_{-t}, \boldsymbol{N}^*, \boldsymbol{\theta} = \begin{cases} Z_1|Z_2, N_1^*, \boldsymbol{\theta} & \text{if} \quad t = 1 \\ Z_t|Z_{t-1}, Z_{t+1}, N_t^*, \boldsymbol{\theta} & \text{if } 1 < t < T \\ Z_T|Z_{T-1}, N_T^*, \boldsymbol{\theta} & \text{if} \quad t = T \end{cases}. \tag{7}$$

The expression above exploits the independence of the components of $\boldsymbol{N}^*$ conditionally on $\boldsymbol{Z}$ and the sparse structure of the inverse of the correlation matrix B, which are both propitious to the mixing of the Gibbs sampler. In addition, we note that (7) greatly improves computational time through vectorization. Indeed, the odd-index components of $\boldsymbol{Z}$ are mutually independent conditionally on the even-index ones, and vice versa. Consequently, the subroutines used to update the odd- and even-index components can be vectorized.

From expression (7), we obtain the unnormalized density for every full conditional distribution of Z_t,

$$\pi(Z_t|\boldsymbol{Z}_{-t}, \boldsymbol{N}^*, \boldsymbol{\theta}) \propto \begin{cases} \pi(N_1^*|Z_1)\,\pi(Z_1|Z_2, \boldsymbol{\theta}) & \text{if} \quad t = 1 \\ \pi(N_t^*|Z_t)\,\pi(Z_t|Z_{t-1}, Z_{t+1}, \boldsymbol{\theta}) & \text{if } 1 < t < T \\ \pi(N_T^*|Z_T)\,\pi(Z_T|Z_{T-1}, \boldsymbol{\theta}) & \text{if} \quad t = T \end{cases}. \tag{8}$$

The second factor in the above expression is always a Gaussian density with mean μ_t and variance τ_t^2 given by

$$\mu_t = \begin{cases} \frac{\theta_1\sigma^2+(Z_2-a)(1+b)\theta_2}{\sigma^2+(1+b)\theta_2} & \text{if} \quad t = 1 \\ \frac{a+(1+b)(Z_{t+1}+Z_{t-1}-a)}{1+(1+b)^2} & \text{if } 1 < t < T \\ a + (1+b)Z_{T-1} & \text{if} \quad t = T \end{cases},$$

$$\tau_t^2 = \begin{cases} \frac{\sigma^2\theta_2}{\sigma^2+(1+b)\theta_2} & \text{if} \quad t = 1 \\ \frac{\sigma^2}{1+(1+b)^2} & \text{if } 1 < t < T \\ \sigma^2 & \text{if} \quad t = T \end{cases}. \tag{9}$$

Therefore, the expression (8) is equal to the product between a Gaussian density and a Poisson mass probability, so that

$$\pi(Z_t|\boldsymbol{Z}_{-t}, N^*, \boldsymbol{\theta}) \propto \exp\left(-e^{Z_t} - \frac{(Z_t-\mu_t)^2}{2\tau_t^2}\right), \tag{10}$$

where μ_t and τ_t^2 are given by (9) and channel the dependence between Z_t and $\boldsymbol{Z}_{-t}$.

We sample from the above density using an accept-reject algorithm with the $N(\xi, \omega^2)$ distribution serving as the proposal density. The following proposition provides the upper bound between the target and the proposal and, in the special case of $\omega^2 = \tau_t^2$, the best value for ξ.

Proposition 1 *Using the above notation the following hold:*

i) The ratio between expression (10) and the density of $N(\xi, \omega^2)$ is unbounded if $\omega^2 < \tau_t^2$.

ii) If $\omega^2 \geq \tau_t^2$ then there exists a unique maximum equal to

$$\widehat{Z}_t = \begin{cases} \log\left(\frac{N_t^*\tau_t^2-\xi+\mu_t}{\tau_t^2}\right) & , \text{if } \omega^2 = \tau_t^2 \\ \frac{N_t^*\omega^2\tau_t^2-\xi\tau_t^2+\mu\omega^2}{\omega^2-\tau_t^2}\mathrm{W}_0\left(\frac{\omega^2\tau^2}{\omega^2-\tau_t^2}\exp\left\{\frac{N_t^*\omega^2\tau_t^2-\xi\tau_t^2+\mu\omega^2}{\omega^2-\tau_t^2}\right\}\right) & , \text{if } \omega^2 > \tau_t^2 \end{cases},$$

where $\mathrm{W}_0(\cdot)$ denotes the upper branch of the Lambert function W.

iii) When $\omega^2 = \tau_t^2$ the value of ξ that minimizes the maximum $\widehat{Z}_t$ is

$$\widehat{\xi} = N_t^*\tau_t^2 + \mu_t - \mathrm{W}_0\left(\tau_t^2\exp(N_t^*\tau_t^2 + \mu_t)\right).$$

Proof i) The ratio between the target and the proposal density is proportional to

$$h(Z_t) = \exp\left(Z_t N_t^* - e^{Z_t} - \frac{(Z_t-\mu_t)^2}{2\tau_t^2} + \frac{(Z_t-\xi)^2}{2\omega^2}\right). \tag{11}$$

It is easy to show that it is unbounded if $\omega^2 < \tau_t^2$.

ii) We compute the first and second log-derivatives and obtain

$$\frac{\mathrm{d}\log h(Z_t)}{\mathrm{d}Z_t} = N_t^* + \frac{Z_t-\xi}{\omega^2} - \frac{Z_t-\mu_t}{\tau_t^2} - e^{Z_t},$$

$$\frac{\mathrm{d}^2\log h(Z_t)}{\mathrm{d}Z_t^2} = \frac{1}{\omega^2} - \frac{1}{\tau_t^2} - e^{Z_t}.$$

Therefore, the second derivative is always negative as long as $\omega^2 \geq \tau_t^2$; in this case $h(\cdot)$ is log-concave and the unique maximum occurs at the value $Z_t = \widehat{Z}_t$ where the first derivative is equal to 0. When $\omega^2 \geq \tau_t^2$, setting the first derivative to 0 yields the following:

$$e^{\widehat{Z}_t} = \frac{N_t^*\omega^2\tau_t^2 - \xi\tau_t^2 + \mu_t\omega^2}{\omega^2\tau_t^2} - \widehat{Z}_t\,\frac{\omega^2-\tau_t^2}{\omega^2\tau_t^2}.$$

If we settle on $\omega^2 > \tau_t^2$, after some algebra we obtain

$$\widehat{Z}_t = \frac{N_t^*\omega^2\tau_t^2-\xi\tau_t^2+\mu\omega^2}{\omega^2-\tau_t^2}\mathrm{W}_0\left(\frac{\omega^2\tau^2}{\omega^2-\tau_t^2}\exp\left(\frac{N_t^*\omega^2\tau_t^2-\xi\tau_t^2+\mu\omega^2}{\omega^2-\tau_t^2}\right)\right),$$

where W_0 denotes the upper branch of the Lambert function W. However, this solution is cumbersome to work with and, more importantly, does not guarantee a more efficient solution.

iii) If we set $\omega^2 = \tau_t^2$ we obtain

$$\widehat{Z}_t = \log\left(\frac{N_t^*\tau_t^2 - \xi + \mu_t}{\tau_t^2}\right).$$

Replacing $\omega^2 = \tau_t^2$ in (11) we obtain

$$\log h(\widehat{Z}_t) = \widehat{Z}_t N_t^* + \frac{(\widehat{Z}_t-\xi)^2 - (\widehat{Z}_t-\mu_t)^2}{2\tau_t^2} - e^{\widehat{Z}_t}.$$

The value $\widehat{\xi}$ that minimizes $\log h(\widehat{Z}_t)$ is found using

$$\begin{aligned} \frac{\mathrm{d}\log h(\widehat{Z}_t)}{\mathrm{d}\xi} &= \left.\frac{\mathrm{d}\log h(x)}{\mathrm{d}x}\right|_{x=\widehat{Z}_t}\frac{\mathrm{d}\widehat{Z}_t}{\mathrm{d}\xi} + \left.\frac{\mathrm{d}\log h(x)}{\mathrm{d}\xi}\right|_{x=\widehat{Z}_t} \\ &= \left.\frac{\mathrm{d}\log h(x)}{\mathrm{d}\xi}\right|_{x=\widehat{Z}_t} \\ &= -\frac{\widehat{Z}_t-\xi}{\tau_t^2}. \end{aligned}$$

Since the second derivative is always positive, $\log h(\widehat{Z}_t)$ is a log-convex function with a unique minimum at $\xi = \widehat{Z}_t$. However, because $\widehat{Z}_t$ is itself a function of ξ and

$$\xi = \log\left(\frac{N_t^*\tau_t^2 - \xi + \mu_t}{\tau_t^2}\right),$$

after some algebra, we obtain the final expression

$$\widehat{\xi} = N_t^*\tau_t^2 + \mu_t - \mathrm{W}_0\left(\tau_t^2 \exp(N_t^*\tau_t^2 + \mu_t)\right).$$

□

For the efficiency of the algorithm, it is important to carefully choose the initialization point, $\boldsymbol{\theta}^{(0)}$, which we set equal to the method of moments estimate provided by the following proposition.

Proposition 2 *In the Gompertz model with Poisson sampling error distribution, the mean, variance and covariances for* $\{N_t^*, 1 \le t \le T\}$ *are:*

i) $\mathbb{E}(N_t^*) = \exp\left(\theta_1 + \frac{\theta_2}{2}\right)$,

ii) $Var(N_t^*) = \exp\left(\theta_1 + \frac{\theta_2}{2}\right) + \left(\exp(\theta_2) - 1\right)\exp(2\theta_1 + \theta_2)$,

iii) $Cov(N_t^*, N_{t+h}^*) = \left[\exp\left(\theta_2(1+b)^h\right) - 1\right]\exp(2\theta_1 + \theta_2)$.

Proof Equation (3) implies

$$\mathbb{E}[Z_t] = \theta_1\,,\ \mathrm{Var}(Z_t) = \theta_2\,,\ \mathrm{Cov}(Z_t, Z_{t+h}) = \theta_2(1+b)^{|h|}\,.$$

On the other hand, $N_t = \exp(Z_t)$ and based on the expression for the moment generating function of a Gaussian random variable, we get

$$\mathbb{E}\big[\exp(cZ_t)\big] = \exp\left(c\theta_1 + \frac{c^2\theta_2}{2}\right),$$

and thus the following expressions hold

$$\begin{aligned}
\mathbb{E}[N_t] &= \mathbb{E}\big[\exp(Z_t)\big] = \exp\left(\theta_1 + \frac{\theta_2}{2}\right),\\
\mathrm{Var}(N_t) &= \mathbb{E}\big[\exp(2Z_t)\big] - \mathbb{E}^2\big[\exp(Z_t)\big]\\
&= \exp(2\theta_1 + \theta_2)(\exp(\theta_2) - 1)\,,\\
\mathrm{Cov}(N_t, N_{t+h}) &= \mathbb{E}\big[\exp(Z_t + Z_{t+h})\big] - \mathbb{E}^2\big[\exp(Z_t)\big]\\
&= \exp\big(2\theta_1 + \theta_2\big)\big[\exp(\theta_2(1+b)^h) - 1\big].
\end{aligned}$$

The last expression is valid because $Z_t + Z_{t+h} \sim N\big(2\theta_1, 2\theta_2[1+(1+b)^h]\big)$ and $\mathbb{E}\big[\exp(Z_t)\big] = \mathbb{E}\big(\exp(Z_{t+h})\big)$. Furthermore, since $N_t^*|N_t \overset{ind}{\sim} \mathrm{Po}(N_t)$ we have $\mathbb{E}[N_t^*|N_t] = \mathrm{Var}(N_t^*|N_t) = N_t$ and $\mathrm{Cov}(N_t^*, N_{t+h}^*|N_t, N_{t+h}) = 0$ due to conditional independence. Using the law of total expectation, the law of total variance, and the law of total covariance is easy to obtain the first two moments for N_t^*, they are

$$\begin{aligned}
\mathbb{E}[N_t^*] &= \mathbb{E}\big[\mathbb{E}[N_t^*|N_t]\big]\\
&= \exp\left(\theta_1 + \frac{\theta_2}{2}\right),\\
\mathrm{Var}(N_t^*) &= \mathbb{E}\big[\mathrm{Var}(N_t^*|N_t)\big] + \mathrm{Var}\big(\mathbb{E}[N_t^*|N_t]\big)\\
&= \exp\left(\theta_1 + \frac{\theta_2}{2}\right) + \exp\big(2\theta_1 + \theta_2\big)\\
&\quad \times \left(\exp\big(\theta_2\big) - 1\right),\\
\mathrm{Cov}(N_t^*, N_{t+h}^*) &= \mathbb{E}\big[\mathrm{Cov}(N_t^*, N_{t+h}^*|N_t, N_{t+h})\big]\\
&\quad + \mathrm{Cov}\big(\mathbb{E}[N_t^*|N_t], \mathbb{E}[N_{t+h}^*|N_{t+h}]\big)\\
&= \exp\big(2\theta_1 + \theta_2\big)\Big[\exp\big(\theta_2(1+b)^h\big) - 1\Big].
\end{aligned}$$

□

We set the starting point $\boldsymbol{\theta}^{(0)}$ in order to match the first two moments provided by the above proposition with their sample counterparts. For covariance, we set $h = 1$ because the first lag is the most efficient to estimate.

We use a sampling importance resampling strategy for the initialization of $\boldsymbol{Z}$. The importance density is equal to

$$\begin{aligned}
q(\boldsymbol{Z}) &= \pi\big(Z_1|N_1^*, Z_2 = \log(N_2^*)\big)\pi\big(Z_2|N_2^*, Z_1, Z_3\\
&= \log(N_3^*)\big)\ldots\pi\big(Z_{t-1}|N_{t-1}^*, Z_{t-2}, Z_t\\
&= \log(N_t^*)\big)\pi\big(Z_t|N_t^*, Z_{t-1}\big).
\end{aligned}$$

Thus, the importance density $q(\boldsymbol{Z})$ is equal to the distribution of a complete update of the Gibbs sampler starting from $\boldsymbol{Z} = \log \boldsymbol{N}^*$. The resampling strategy allows $\boldsymbol{Z}^{(0)}$ to be an approximate sample from $\pi(\boldsymbol{Z}|\boldsymbol{N}^*, \boldsymbol{\theta}^{(0)})$. We draw 10,000 values of $\boldsymbol{Z}$ from $q(\cdot)$ and compute the importance weights. The initial value is then sampled with probability proportional to these weights.

The proposed MCMC-EM method can also be used to calculate the asymptotic variance of the MLE. As noted in Caffo et al. (2005), one can use the output of the sampling algorithm to calculate the inverse of the Fisher information matrix using the method of Louis (1982). This allows the estimation of the asymptotic variance of $\widehat{\boldsymbol{\theta}}$.

4 Bayesian Inference

The Bayesian paradigm offers another probabilistic mechanism to estimate finite sample variances for the estimators of interest, allows principled ways to incorporate prior knowledge when it is available, and integrates model uncertainty

into the predictions' variability via model averaging. The crux of the approach is the posterior distribution which encodes all the uncertainty after observing the data. In the Gompertz model with Poisson errors, the posterior distribution is analytically intractable, so we must study it using MCMC sampling. The data augmentation strategy presented in Section 3 also plays a central role in the design of the MCMC algorithm. Perhaps surprisingly, the numerical experiments show that the algorithm for sampling the Bayesian posterior is much more efficient than the MCMC-EM in terms of computational time.

To perform a fair comparison with the MLE, we propose using a weakly informative prior. We use an uniform prior for b, that is, $b \sim U(-2, 0)$, and normal-inverse gamma priors for θ_1 and θ_2,

$$\theta_2 \sim \text{Inv} - \text{Gamma}(\varphi_1, \varphi_2) \text{ and } \theta_1|\theta_2 \sim N(\eta_1, \eta_2\theta_2)\,.$$

We also assume that θ_1, θ_2 are a priori independent of b. We choose values for the hyperparameters $\varphi_1, \varphi_2, \eta_1$, and η_2 that lead to a weakly informative prior. We set $\varphi_1 = \varphi_2 = 0.1$, $\eta_1 = 0$, and $\eta_2 = 100$. The sampling algorithm's steps do not depend on the particular values we choose for the hyperparameters, so we describe them in terms of generic values.

The dependent draws from the conditional distribution of all parameters and augmented data $\pi(\theta_1, \theta_2, b, \boldsymbol{Z}, |N^*)$ are obtained using a Gibbs sampler.

Let $\theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}, \boldsymbol{Z}^{(k)}$ be the sample values in the k-th iteration of the MCMC algorithm. The starting values $\theta_1^{(0)}, \theta_2^{(0)}, b^{(0)}$ are set equal to the method of moments estimates and $\boldsymbol{Z}^{(0)}$ is drawn by sampling importance resampling, as in the initialization of MCMC-EM.

Given $\theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}$, and $\boldsymbol{Z}^{(k)}$, we obtain $\theta_1^{(k+1)}, \theta_2^{(k+1)}$, $b^{(k+1)}$, and $\boldsymbol{Z}^{(k+1)}$ using the following update scheme:

1. For $t = 1, 3, 5, \ldots$, sample $Z_t^{(k+1)}|N_t^*, \theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}, \boldsymbol{Z}_{-t}^{(k)}$.
2. For $t = 2, 4, 6, \ldots$, sample $Z_t^{(k+1)}|N_t^*, \theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}, \boldsymbol{Z}_{-t}^{(k+1)}$.
3. Sample $\theta_1^{(k+1)}, \theta_2^{(k+1)}, b^{(k+1)}|\boldsymbol{Z}^{(k+1)}$.
   (a) Sample $b^{(k+1)}|\boldsymbol{Z}^{(k+1)}$.
   (b) Sample $\theta_2^{(k+1)}|b^{(k+1)}, \boldsymbol{Z}^{(k+1)}$.
   (c) Sample $\theta_1^{(k+1)}|\theta_2^{(k+1)}, b^{(k+1)}, \boldsymbol{Z}^{(k+1)}$.

In steps (1) and (2), we use the acceptance-rejection algorithm described in section 3. In steps (3b) and (3c) the full conditionals are standard, since

$$\theta_2|b, \boldsymbol{Z} \sim \text{Inv} - \text{Gamma}\left(\phi_1 + \frac{T}{2}, \phi_2 + \frac{1}{2}(\boldsymbol{Z} - \eta_1 1_T)' \times (\eta_2 1_T 1_T' + B)^{-1}(\boldsymbol{Z} - \eta_1 1_T)\right),$$

$$\theta_1|\theta_2, b, \boldsymbol{Z} \sim N\left(\frac{\eta_1 + \eta_2 1_T' B^{-1}\boldsymbol{Z}}{1 + \eta_2 1_T' B^{-1} 1_T}, \frac{\eta_2\theta_2}{1 + \eta_2 1_T' B^{-1} 1_T}\right),$$

where 1_T denotes the T-dimensional vector of ones. In contrast, step (3a) is not standard. To sample from the conditional density of b given $\boldsymbol{Z}$ we use another accept-reject algorithm. The proposal is the uniform prior itself; thus, the upper bound between the target density and the proposal density is the maximum of $\pi(\boldsymbol{Z}|b)$. Since the prior of (θ_1, θ_2) is a normal-inverse gamma distribution, straightforward calculation yields $Z|\theta_2, b \sim N_T\big(\eta_1 1_T, \theta_2(\eta_2 1_T 1_T' + B)\big)$ and then

$$\begin{aligned}
\pi(\boldsymbol{Z}|b) &= \int_{\theta_2}\int_{\theta_1} \pi(\boldsymbol{Z}|\theta_1, \theta_2, b)\pi(\theta_1, \theta_2)\mathrm{d}\theta_1\mathrm{d}\theta_2 \\
&\propto \det{}^{-\frac{1}{2}}(\eta_2 1_T 1_T' + B) \int_0^{+\infty} \theta_2^{-\frac{2\phi_1+T}{2}-1} \\
&\quad \times \exp\left(-\frac{1}{\theta_2}\left(\phi_2 + \frac{1}{2}(\boldsymbol{Z} - \eta_1 1_T)'(\eta_2 1_T 1_T' + B)^{-1}(\boldsymbol{Z} - \eta_1 1_T)\right)\right)\mathrm{d}\theta_2 \\
&\propto \det{}^{-\frac{1}{2}}(\eta_2 1_T 1_T' + B) \\
&\quad \times \left(1 + \frac{1}{2\phi_2}(\boldsymbol{Z} - \eta_1 1_T)'(\eta_2 1_T 1_T' + B)^{-1}(\boldsymbol{Z} - \eta_1 1_T)\right)^{-\frac{2\phi_1+T}{2}}.
\end{aligned} \tag{12}$$

The expression above can be further simplified using the matrix determinant lemma, Sherman-Morrison formula, and the closed-form expression for B^{-1}. The latter is available because B is the correlation matrix of a stationary autoregressive process of order 1.

Proposition 3 *In the Gompertz model with Poisson sampling error distribution, the density of $\boldsymbol{Z}|b$ is*

$$\begin{aligned}
\pi(\boldsymbol{Z}|b) \propto{}& (1 - r^2)^{1-\frac{T}{2}}\left((\eta_2(T-2)-1)r^2 - 2\eta_2(T-1)r + \eta_2 T + 1\right)^{-\frac{1}{2}} \\
&\times \left(1 + \frac{r^2\sum_{s=2}^{T-1} W_s^2 - 2r\sum_{h=1}^{T-1} W_h W_{h+1} + \sum_{t=1}^{T} W_t^2}{2\phi_2(1-r^2)}\right. \\
&- \frac{\eta_2(1-r^2)}{(\eta_2(T-2)-1)r^2 - 2\eta_2(T-1)r + \eta_2 T + 1} \\
&\left.\times \frac{r^2\left(\sum_{s=2}^{T-1} W_s\right)^2 - 2r\sum_{s=2}^{T-2} W_s \sum_{t=1}^{T} W_t + \left(\sum_{t=1}^{T} W_t\right)^2}{2\phi_2(1+r^2)}\right)^{-\phi_1-\frac{T}{2}},
\end{aligned} \tag{13}$$

where $W_t = Z_t - \eta_1$ and $r = 1 + b$.

Proof Let $r = 1 + b$ and $W = Z - \eta_1 1_T$. Then Hamilton (1994)[Section 5.2]

$$B = \begin{bmatrix} 1 & r & r^2 & \dots & r^{T-1} \\ r & 1 & r & \dots & r^{T-2} \\ \dots & \dots & \dots & \dots & \dots \\ r^{T-1} & r^{T-2} & r^{T-3} & \dots & 1 \end{bmatrix},$$

$$\det(B) = (1-r^2)^{T-1},$$

$$B^{-1} = \frac{1}{1-r^2} \begin{bmatrix} 1 & -r & & & \\ -r & 1+r^2 & -r & & \\ & \ddots & \ddots & \ddots & \\ & & -r & 1+r^2 & -r \\ & & & -r & 1 \end{bmatrix}.$$

Using the matrix determinant lemma and Sherman-Morrison formula, it is obtained

$$\det(\eta_2 1_T 1_T' + B) = \det(B)(1 + \eta_2 1_T' B^{-1} 1_T),$$

$$(\eta_2 1_T 1_T' + B)^{-1} = B^{-1} - \eta_2 \frac{B^{-1} 1_T 1_T' B^{-1}}{1 + \eta_2 1_T' B^{-1} 1_T}.$$

This implies

$$\begin{aligned} \pi(Z|b) &\propto \det^{-\frac{1}{2}}(B)(1 + \eta_2 1_T' B^{-1} 1_T)^{-\frac{1}{2}} \\ &\times \left(1 + \frac{W' B^{-1} W}{2\phi_2} - \frac{\eta_2}{(1 + \eta_2 1_T' B^{-1} 1_T)}\right. \\ &\times \left.\frac{W' B^{-1} 1_T 1_T' B^{-1} W}{2\phi_2}\right). \end{aligned} \tag{14}$$

Consider separately $1 + \eta_2 1_T' B^{-1} 1_T$. It is straightforward that $1_T' B^{-1} 1_T$ is the sum of all components of the matrix B^{-1}; in this matrix there are $(T-2)$, $2(T-1)$, and two times $1 + r^2$, $-r$, and one respectively. Then

$$\begin{aligned} 1 + \eta_2 1_T^\top B^{-1} 1_T &= 1 + \frac{\eta_2}{1-r^2}\left(2 + (1+r^2)(T-2) - 2r(T-1)\right) \\ &= \frac{1 - r^2 + \eta_2\left(r^2(T-2) - 2r(T-1) + T\right)}{1-r^2} \\ &= \frac{\left(\eta_2(T-2) - 1\right)r^2 - 2\eta_2(T-1)r + \eta_2 T + 1}{1-r^2}. \end{aligned} \tag{15}$$

We now focus our attention on $W' B^{-1} W$, which can be simplified as follows:

$$W' B^{-1} W = \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1 & -r & 0 & \cdots & 0 \\ -r & 1+r^2 & -r & \cdots & 0 \\ 0 & -r & 1+r^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -r \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}$$

$$\begin{aligned} &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} W_1 - rW_2 \\ (1+r^2)W_2 - r(W_1 + W_3) \\ (1+r^2)W_3 - r(W_2 + W_4) \\ \vdots \\ (1+r^2)W_t - r(W_{t-1} + W_{t+1}) \\ \vdots \\ (1+r^2)W_{T-1} - r(W_{T-2} + W_T) \\ W_T - rW_{T-1} \end{bmatrix} \\ &= \frac{1}{1-r^2}\left(W_1^2 - rW_1W_2 + W_T^2 - rW_{T-1}W_T \right. \\ &\left. + \sum_{s=2}^{T-1}\left[(1+r^2)W_s^2 - rW_s(W_{s-1} + W_{s+1})\right]\right) \\ &= \frac{1}{1-r^2}\left(\sum_{t=1}^{T} W_t^2 - rW_1W_2 - r\sum_{s=2}^{T-1} W_sW_{s+1} \right. \\ &\left. - rW_TW_{T_1} - r\sum_{s=2}^{T-1} W_sW_{s-1} + r^2\sum_{s=2}^{T-1} W_s^2\right) \\ &= \frac{1}{1-r^2}\left(r^2\sum_{s=2}^{T-1} W_s^2 - 2r\sum_{s=2}^{T-1} W_sW_{s+1} + \sum_{t=1}^{T} W_t^2\right). \end{aligned} \tag{16}$$

Similarly, $W' B^{-1} 1_T$ yields

$$\begin{aligned} W' B^{-1} 1_T &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1 & -r & 0 & \cdots & 0 \\ -r & 1+r^2 & -r & \cdots & 0 \\ 0 & -r & 1+r^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -r \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1-r \\ r^2 - 2r + 1 \\ \cdots \\ r^2 - 2r + 1 \\ 1-r \end{bmatrix} \\ &= \frac{(W_1 + W_T)(1-r) + (1-r)^2\sum_{s=2}^{T-1} W_s}{(1-r)(1+r)} \\ &= \frac{\sum_{t=1}^{T} W_t - r\sum_{s=2}^{T-1} W_s}{1+r}. \end{aligned} \tag{17}$$

Therefore,

$$\begin{aligned} W' B^{-1} 1_T 1_T' B^{-1} W &= (W' B^{-1} 1_T)^2 \\ &= \frac{\left(\sum_{t=1}^{T} W_t - r\sum_{s=2}^{T-1} W_s\right)^2}{(1+r)^2} \\ &= \frac{r^2(\sum_{s=2}^{T-1} W_s)^2 - 2r\sum_{s=2}^{T-1} W_s\sum_{t=1}^{T} W_t + (\sum_{t=1}^{T} W_t)^2}{(1+r)^2}. \end{aligned} \tag{18}$$

Finally, the form (13) is obtained by substituting into (14) the expressions (15), (16), and (18). □

It is important to note that equation (13) is inexpensive to evaluate, as it is just a combination of ratios and products of polynomials in b. This allows us to adopt the following two-step strategy. We first perform a grid search for the maximum of $\pi(\boldsymbol{Z}|b)$, as a function of b. To this end, we compute (13) over a dense sequence of values of b, say $-1.99, -1.98, \ldots, -0.01$, and set

$$b_0 = \arg \max_{b=-1.99,-1.98,\ldots,-0.01} \pi(\boldsymbol{Z}|b)\,.$$

It is reasonable that the value of b_0 is close to the global maximum of $\pi(\boldsymbol{Z}|b)$. The L-BFGS-B algorithm is initialized at b_0 in order to find the global maximum. Although this optimization must be repeated at each iteration of the Gibbs sampler, the algorithm remains fast due to the simplicity of expression (13). In fact, as reported in Section 5.1, the Gibbs sampler is much faster than the MCMC-EM optimizer. Furthermore, as reported in the Supplementary Material, our proposed Gibbs sampler is also faster than the state-of-the-art provided by Stan (Carpenter et al. 2017).

5 Numerical experiments

The purpose of running the numerical experiments in this section is two-fold. First, we examine the difference between the frequentist and Bayesian inferences, which can be informative when the data has a modest size and we want to gauge the influence of the prior. Second, we provide proof-of-concept for the algorithms proposed and examine their performance on data generated under different scenarios, including one in which the sampling error distribution is misspecified. We also consider a comparison between the Bayesian inference produced using the sampler we design here and the one using a generic Stan implementation. The latter can be considered state of the art, since, to our knowledge, no other MCMC sampler has been designed for this problem, except for the generic implementation of Lele et al. (2007).

In numerical experiments presented in this paper to compare algorithms, we replaced the now deprecated WinBUGS with Stan, the latter being the current state of the art off-the-shelf algorithm for Bayesian computation and is generally known to exhibit superior efficiency compared to WinBugs. The code implementing the Stan algorithm is publicly available on the GitHub repository and is also reported in the Supplementary Material. The Stan algorithm relies on the same latent variable representation as the one used for the Gibbs and MLE approaches. It also exploits the sparse structure of the inverse of B, which allows us to avoid computationally expensive calculation of matrix inverses and determinants. As Stan software is implemented in C++, to ensure a fair comparison with our proposed algorithms, we translated, with the help of Cursor (Anysphere 2025) and Copilot (Microsoft 2025), some of our subroutines written in R into C++ via the Rcpp framework.

5.1 Simulations

5.1.1 Simulation scenarios

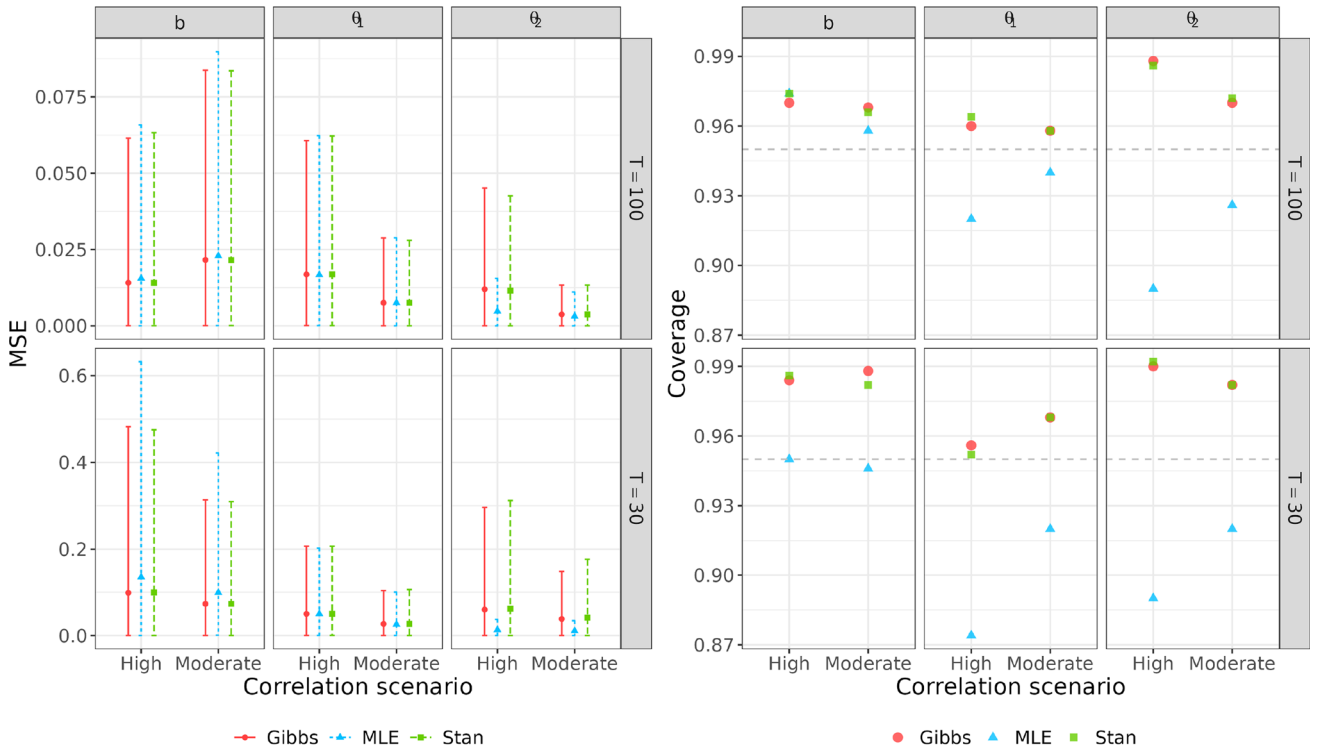
The simulation study evaluates the performance of the proposed algorithms when the sampling model is correct and when it is misspecified. Specifically, we generated time series of varying lengths under the Poisson noise model and under a model with sampling errors that have a negative binomial distribution. In addition, we varied the model parameters to examine the performance of the algorithms under different scenarios. One of the scenarios has parameter values close to the estimates obtained in the Redstart data analysis.

The data are fitted using the frequentist approach described in Section 3, and using the Bayesian methods described in Section 4. In addition, Bayesian inference is also produced using an off-the-shelf Stan implementation.

We analyze data generated under eight simulation scenarios. Scenario **S1** yields a moderate level of serial correlation given by $b^\dagger = -0.5$, while setting **S2** exhibits high levels of correlation given by $b^\dagger = -0.22$. Based on the parameter estimates obtained from the Redstart data analysis, we fixed $\theta_1^\dagger = 2$ and $\theta_2^\dagger = 0.22$, for both **S1** and **S2**. We simulated these settings considering time series of lengths $T = 30$. Scenarios **S3** and **S4** use the same parameters as, respectively, **S1** and **S2**, but have $T = 100$.

Scenarios **S5** - **S8** are produced using the same parameter values as **S1** - **S4**, respectively, but with a misspecified sampling error model. Specifically, the sampling error was simulated using a negative binomial distribution with mean $\exp(Z_t)$ and variance $2\exp(Z_t)$. These values were chosen to provide a realistic comparison with the Poisson model. Note that a negative binomial with the mean and variance specified as above corresponds to one with a success probability of 0.5 and a dispersion (size) parameter selected so that the mean is equal to $\exp(Z_t)$. It is known that as the probability of success p approaches zero, the negative binomial distribution converges to a Poisson distribution with the same mean. Thus, setting $p = 0.5$ ensures that the model is sufficiently different to illustrate the impacts of misspecification. For each setting, we computed the mean square error (MSE), the coverage of the 95% confidence and credible intervals, for every parameter in each scenario.

The simulation analysis is conducted with 500 independent replicates, and for the Bayesian methods, each replicate is based on an MCMC sample of size 10, 000. In addition, we report the effective sample sizes obtained by our sam-



(a) MSE (b) Coverage

Fig. 1 Results for the correctly specified model with Poisson distributed errors: Mean square error and 95% coverage by correlation setting for each parameter and time series length T for scenarios **S1** - **S4**. Within each sub-figure, columns represent parameters and rows represent the different time series length T. Results form the Gibbs sampler are shown with points and solid lines results, results for the MLE approach are indicated by triangles and fine dashed lines, and results for Stan are displayed with squares and bold dashed lines. In the MSE plots (left panels), the central points/triangles indicate the mean values, and the error bars represent the 5th and 95th percentiles. The coverage plots (right panels) show the coverage obtained using credible intervals approximated from Gibbs samples (red points), Stan samples (green squares), and using Wald confidence intervals obtained using the MLE and is asymptotic variance produced by MCMC-EM

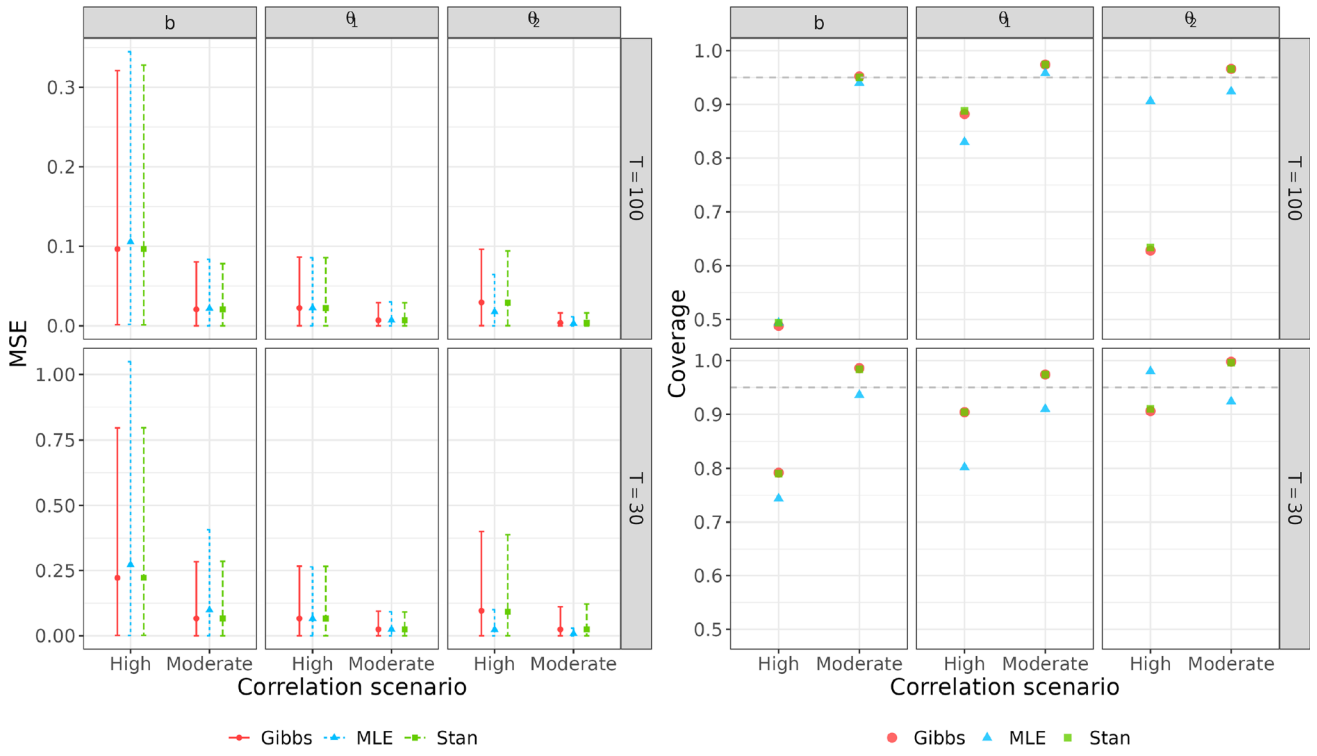
pler and Stan. Finally, we compute and report summaries of computation times.

5.1.2 Simulation results

As expected, the inference based on all three algorithms improves as the length of the time series increases. Figure 1 presents the results for the correctly specified model with Poisson distributed errors. In most cases, the MSE values across the three computational methods are quite similar. An exception occurs for θ_2 where the MLE shows a better MSE compared to the Bayesian algorithms, especially when $T = 30$. However, the Bayesian credible intervals are more conservative and tend to provide better coverage than the confidence intervals built around the MLE.

The results for the misspecified cases are shown in Figure 2. In this setting, performances differ noticeably between the high and moderate correlation scenarios. When the correlation is lower, the analyzes based on all three algorithms show better results, with a lower MSE and a more accurate coverage across all parameters and both sample sizes. Neither frequentist nor Bayesian analysis dominates in terms of MSE and coverage across all scenarios and parameters.

Figures 1 and 2 show that the Gibbs sampler and the Stan algorithm yield essentially the same inferential results. This is not surprising, as the two methods target the same distribution. Consequently, comparing them in terms of computational efficiency is more meaningful. The computational times for scenario **S4** are reported in Table 1, the computational times for the other scenarios are reported in the Supplementary Material. The Gibbs sampler is around four to eight times faster than the Stan algorithm in all simulation settings. In addition, the Gibbs sampler typically produces samples with lower autocorrelation. This remains true even with the accept-reject schemes for $\pi(Z_t|\boldsymbol{\theta}, \boldsymbol{Z}_{-t})$ and one for $\pi(b|\boldsymbol{Z})$. The empirical acceptance rates averaged over the components of $\boldsymbol{Z}$ are 83.78%, 88.66%, 83.84%, 90.61%, 83.75%, 80.87%,



(a) MSE (b) Coverage

Fig. 2 Results for the misspecified model with Negative binomial distributed errors: Mean square error and 95% coverage by correlation scenario for each parameter and time series length T for scenarios **S5** - **S8**. Within each sub-figure, columns represent parameters and rows represent the different time series length T. Results form the Gibbs sampler are shown with points and solid lines results, results for the MLE approach are indicated by triangles and fine dashed lines, and results for Stan are displayed with squares and bold dashed lines. In the MSE plots (left panels), the central points/triangles indicate the mean values, and the error bars represent the 5th and 95th percentiles. The coverage plots (right panels) show the coverage obtained using credible intervals approximated from Gibbs samples (red points), Stan samples (green squares), and using Wald confidence intervals obtained using the MLE and is asymptotic variance produced by MCMC-EM

83.78%, and 80.93% for scenarios **S1** to **S8**, respectively. The empirical acceptance rates for b are 40.92%, 36.15%, 27.34%, 22.80%, 40.97%, 40.48%, 27.34%, and 26.76% for scenarios **S1** to **S8**, respectively. Indeed, shifting our attention from the raw computational times to the *effective sample size per second* (ESSps), which is defined as the effective sample size divided by the computational time, we notice that the gap between the two algorithms increases significantly. The boxplots of the ESSps for each scenario and for the two algorithms are also reported in the Supplementary Materials. In particular, the mean ESSps values for the Gibbs sampler are approximately 2.5 to 22 times higher than those for the Stan algorithm. Consequently, the Gibbs sampler presented here outperforms Stan in computational efficiency in all experimental scenarios. This is not surprising, since the algorithm developed here has been designed specifically for this problem, while Stan is a sampling algorithm which is able to tackle a wide variety of models. Its general-purpose nature prevents Stan from taking advantage of the specific characteristic of a given target distribution.

Table 1 Summary statistics of the computational times of each algorithm (in seconds) under scenario **S4**

	1st Quantile	Mean	Median	3rd Quantile
Gibbs	2.75	3.13	2.99	3.41
MLE	93.51	127.26	122.35	151.37
Stan	19.87	25.74	24.32	29.96

Finally, it is important to note that the algorithm that implements the MLE approach is the slowest. It takes approximately 26.5-42.5 times longer than the Gibbs sampler and about 5-8.75 times longer than Stan. We suspect that this behavior is due to the fact that computing the MLE requires multiple MCMC runs.

5.2 Real data analysis

We analyze here the data set of American Redstart counts that was previously discussed by Lele (2006) and Dennis

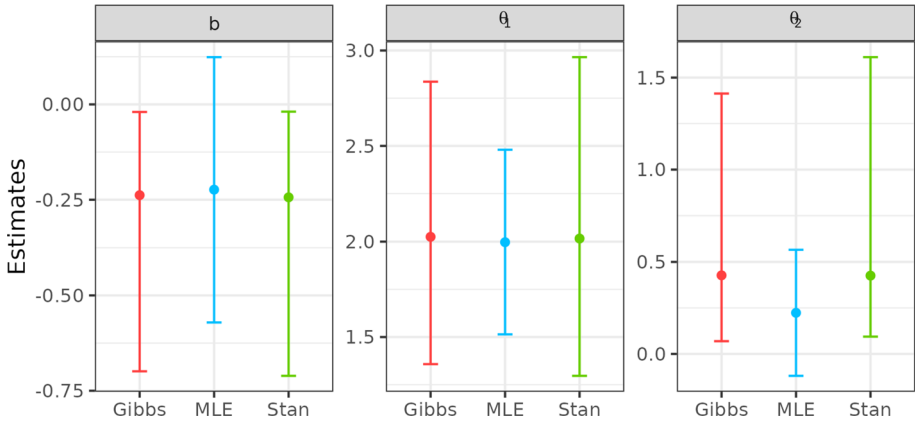


Fig. 3 Point estimates for the three parameters obtained using each inference method. For the two Bayesian methods, the error bars represent the 95% credible intervals, while for the MLE, they indicate the Wald 95% confidence intervals

Table 2 Effective sample sizes for the three parameters from the analysis of the American Redstart for both Bayesian inference methods

	b	θ_1	θ_2
Gibbs	1337.64	8250.73	10000
Stan	1552.83	1884.42	1568.34

et al. (2006). This data set is recorded with the number 0214332808636 in the North American Breeding Bird Survey (Peterjohn 1994; Robbins et al. 1986) and contains the number of specimens observed from 1966 to 1995 at a survey location. We fit the Gompertz model with Poisson-distributed errors to this dataset using Bayesian and frequentist approaches supported by the algorithms described in the previous sections. Stan was also used as a baseline for comparison within the Bayesian framework.

We computed point estimates for the three parameters, along with the 95% credible intervals for the Bayesian methods and the 95% confidence intervals based on Gaussian approximation for the frequentist approach (Figure 3).

Although Bayesian approaches show greater uncertainty in the estimation of θ_2, the point estimates for the three methods remain comparable. However, when examining the effective sample size and autocorrelation functions, the Gibbs sampler exhibits superior performance. Table 2 and Figure 4, show the effective sample sizes and, respectively, the autocorrelation functions for all parameters. The effective sample size for θ_2 and θ_1 is nearly six and four times larger, respectively, for the Gibbs sampler. Stan exhibits slightly better mixing for b, but requires longer computational time. When combining ESS and running time, we see that the Gibbs sampler had much better ESSps for all parameters as reported in Table 3.

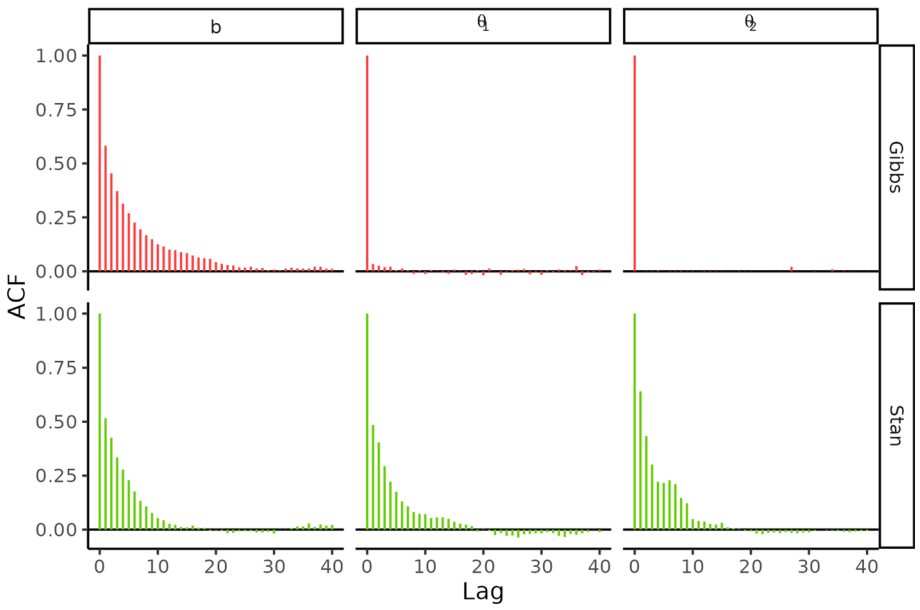


Fig. 4 Autocorrelation functions of the posterior samples for the three parameters under the two algorithms

Table 3 Effective sample sizes per second for the three parameters from the analysis of the American Redstart for both Bayesian inference methods

	b	θ_1	θ_2
Gibbs	1315.76	8115.74	9836.38
Stan	250.43	303.91	252.94

6 Discussion

We develop likelihood-based inference within the frequentist and Bayesian paradigms for the Gompertz model with Poisson sampling errors. The proposed approaches remove the need to consider pseudo-likelihood methods that mitigate computational challenges at the price of reducing the information provided by the data.

In our future work, we would like to investigate whether similar developments can be produced to modified versions of the model considered here. The latter can be created by modifying the growth curve by adding parameters that allow curvature and long-term behavior (Asadi et al. 2023) or when the population dynamics is determined by a stochastic differential equation as in Donnet et al. (2010).

Supplementary information

Supplementary Material with additional figures and tables is available. In addition, the developed methods are implemented in a new R package: `gse`. The new R package and the scripts of Section 5 are available online on GitHub at the following link: https://github.com/sofiar/GMLossF.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-026-10878-w.

Acknowledgements We thank three reviewers, the Associate Editor and the Editor for their suggestions and comments. We also thank Dr. Monica Alexander and Dr. Vianey Leos-Barajas for comments that have improved the paper. The first author was supported by MUR - Prin 2022 - Grant no. 2022FJ3SLA, funded by the European Union - Next Generation EU and by European Union under the ERC Synergy Grant 101071601 (OCEAN, 2023–2030). Views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The third author was supported by NSERC of Canada discovery grant RGPIN-2024-04506.

Author contributions RC and SRS formulated the research problem. PO developed the methodology, theoretical framework and carried out the analytical calculations. PO and SRS implemented the software, ran the simulations and performed the analysis. RC supervised the work, and all the authors discussed the results and contributed to the writing of the final manuscript.

Data availability The American Redstart counts dataset (ID: 02143328 08636) is part of the North American Breeding Bird Survey and is publicly available through the R package PVAClone on CRAN.

Declarations

Competing interests The authors declare no competing interests.

References

Alexander, M., Zagheni, E., Barbieri, M.: A flexible Bayesian model for estimating subnational mortality. Demography **54**(6), 2025–2041 (2017)

Anysphere.: Cursor - the ai code editor. https://cursor.com (2025)

Asadi, M., Di Crescenzo, A., Sajadi, F.A., et al.: A generalized Gompertz growth model with applications and related birth-death processes. Ricerche mat. **72**(2), 1–36 (2023)

Auger-Méthé, M., Newman, K., Cole, D., et al.: A guide to state-space modeling of ecological time series. Ecol. Monogr. **91**(4), e01470 (2021)

Baione, F., Levantesi, S.: Pricing critical illness insurance from prevalence rates: Gompertz versus Weibull. North American Actuarial Journal **22**(2), 270–288 (2018)

Beyer, H.L., Ung, R., Murray, D.L., et al.: Functional responses, seasonal variation and thresholds in behavioural responses of moose to road density. J. Appl. Ecol. **50**(2), 286–294 (2013)

Booth, J.G., Hobert, J.P.: Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J. R. Stat. Soc. Ser. B Stat Methodol. **61**(1), 265–285 (1999)

Butt, Z., Haberman, S.: Application of frailty-based mortality models using generalized linear models. ASTIN Bulletin: The Journal of the IAA **34**(1), 175–197 (2004)

Caffo, B., Jank, W., Jones, G.: Ascent-based Monte Carlo expectation-maximization. J. R. Stat. Soc. Ser. B Stat Methodol. **67**(2), 235–251 (2005)

Carpenter, B., Gelman, A., Hoffman, M.D., et al.: Stan: a probabilistic programming language. J. Stat. Softw. **76**, 1–32 (2017)

Chan, K.S., Ledolter, J.: Monte Carlo EM estimation for time series models involving counts. J. Am. Stat. Assoc. **90**(429), 242–252 (1995)

Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc.: Ser. B (Methodol.) **39**(1), 1–22 (1977)

Dennis, B., Ponciano, J.M., Lele, S.R., et al.: Estimating density dependence, process noise, and observation error. Ecol. Monogr. **76**(3), 323–341 (2006)

Donnet, S., Foulley, J.L., Samson, A.: Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. Biometrics **66**(3), 733–741 (2010)

Gompertz, B.: On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS. Philos. Trans. R. Soc. Lond. **115**, 513–583 (1825)

Hamilton, J.: Time Series Analysis. Princeton University Press, Princeton, NJ (1994)

Hostetler, J.A., Chandler, R.B.: Improved state-space models for inference about spatial and temporal variation in abundance from count data. Ecology 96(6):1713–1723. https://arxiv.org/abs/https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/14-1487.1 (2015)

Lele, S.R.: Sampling variability and estimates of density dependence: A composite-likelihood approach. Ecology **87**(1), 189–202 (2006)

Lele, S.R., Dennis, B., Lutscher, F.: Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. **10**(7), 551–563 (2007)

Lindén, A., Knape, J.: Estimating environmental effects on population dynamics: consequences of observation error. Oikos **118**(5), 675–680 (2009)

Louis, T.: Finding the observed information matrix when using the EM algorithm. J. R. Stat. Soc. Ser. B Stat Methodol. **44**(2), 226–233 (1982)

Lunn, D.J., Thomas, A., Best, N., et al.: Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. **10**(4), 325–337 (2000)

McCulloch, C.E.: Maximum likelihood variance components estimation for binary data. J. Am. Stat. Assoc. **89**(425), 330–335 (1994)

Microsoft.: Microsoft copilot. https://copilot.microsoft.com/ (2025)

Newman, K., Buckland, S., Morgan, B., et al.: Modelling Population Dynamics: Model Formulation. Springer, Fitting and Assessment using State-Space Methods (2014)

Peterjohn, B.: The North American breeding bird survey. Birding **26**(6), 386–398 (1994)

Robbins, C.S., Bystrak, D., Geissler, P.H.: The Breeding Bird Survey: its first fifteen years, 1965–1979. Tech. rep, US Fish and Wildlife Service (1986)

Staples, D.F., Taper, M.L., Dennis, B.: Estimating population trend and process variation for pva in the presence of sampling error. Ecology **85**(4), 923–929 (2004)

Tai, T.H., Noymer, A.: Models for estimating empirical Gompertz mortality: with an application to evolution of the gompertzian slope. Popul. Ecol. **60**(1–2), 171–184 (2018)

Tjørve, K.M.C., Tjørve, E.: The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the unified-Richards family. PLoS ONE **12**(6), 1–17 (2017)

Wei, G.C., Tanner, M.A.: A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. J. Am. Stat. Assoc. **85**(411), 699–704 (1990)

Winsor, C.P.: The Gompertz curve as a growth curve. Proc Natl Acad Sci U S A **18**(1), 1–8 (1932)