

Fitting real data by means of non-homogeneous lognormal diffusion processes

PATRICIA ROMÁN-ROMÁN^{*,†}, JUAN JOSÉ SERRANO-PÉREZ,
AND FRANCISCO TORRES-RUIZ

In order to achieve a good fit to real data that evolve over time and whose observed trend shows deviations with respect to an exponential shape, a non-homogeneous lognormal diffusion process with time dependent infinitesimal mean and variance is considered. Such model provides a more flexible structure of the variance than that of the non-homogeneous diffusion process only in its infinitesimal mean, allowing to reproduce the behaviour of the observed data more accurately and enable us to tackle problems in which data variability plays a fundamental role with a higher degree of reliability. A procedure for the estimation of the time functions included in the infinitesimal mean and variance is proposed and hypothesis testing to confirm or refute the need for considering non-homogeneous processes to fitting real data are designed. A simulation study corroborates the validity of the proposed estimation procedure. Finally, a real data application of a patient-derived xenograft (PDX) tumor model is performed.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 60G99, 62M99; secondary 62P10.

KEYWORDS AND PHRASES: Non-homogeneous lognormal diffusion process, Fitting data, Tumor growth.

1. INTRODUCTION

In a wide range of applied scientific fields, the available data originate from the observation of continuous random variables vary exponentially, at least for a limited time interval. In this context, lognormal diffusion processes have been widely used as a stochastic model for the fitting of data. The homogeneous version of this process, with an exponential trend which justifies using its mean function both for fitting and for forecasting purposes, has been employed in modeling neuron activity (Ricciardi and Lánsky, [12]), Consumer Price Index (Al-Eideh et al. [1]) and river flows (Lefebvre, [10]), among other phenomena. A non homogeneous version including a time dependent function in its infinitesimal mean allows for a more appropriate fit to data

whose observed trend shows deviations with respect to the exponential shape. These deviations can be due to factors that are intrinsic to the evolution of the phenomenon under study as well as to external influences. So far, some works have employed this version to consider the influence of certain economic variables on the evolution of the Gross National Product (Gutiérrez et al. [6]) or of housing prices in Spain (Gutiérrez Jáimez et al. [7]), as well as the influence of a therapy in the growth of specific tumors (Román-Román and Torres-Ruiz [14]).

This model is defined, in general, by means of a stochastic differential equation of the type

$$dX(t) = h(t)X(t)dt + \sigma X(t)dW(t), \quad t \geq t_0,$$

where $h(t)$ is a continuous function $\forall t \geq t_0$ and $\sigma > 0$. This equation is obtained from the ordinary differential equation of Malthusian growth with a dependent time fertility rate by including a random component such as a white noise with constant variance. The use of the appropriate functions $h(t)$ leads to models that fit acceptably the mean of the growth variable under study. However, it may be that these models do not reflect properly the variance of the sample paths of the process due to a poor flexibility of the variance structure. Concretely,

$$Var[X(t)] = (E[X(t)])^2 \left(e^{\sigma^2(t-t_0)} - 1 \right),$$

from where

$$(1) \quad \log \left(1 + \frac{Var[X(t)]}{(E[X(t)])^2} \right)$$

must be a linear function with positive slope. So when fitted to real data, the model may be good enough to reproduce mean behavior, but not variability.

The introduction of a time-dependent noise leads to a stochastic process with infinitesimal variance which is also dependent on time, thus resulting in a more flexible structure for the variance. Thereby, the obtained process also allows to reproduce the variability of observed data more reliably. Several models of this type have already been considered in various fields of application dealing with stochastic processes. Such has been the case, for instance, of the

^{*}This work was supported in part by the Ministerio de Economía y Competitividad, Spain, under Grant MTM2014-58061-P.

[†]Corresponding author.

study of queueing systems by means of diffusion approximations (Giorno et al. [5]), of the study of neuron models using an Ornstein-Uhlenbeck process (Buonocore et al., [4]), of the study of population growth using a lognormal process (Khodabin et al. [8]), or the use of a Gompertz process for modeling tumor growth in the presence of therapies (Román-Román et al. [15]).

In this paper, we will consider non homogeneous versions of the lognormal diffusion process both for the mean and for the infinitesimal variance. The resulting models allow for a more faithful reproduction of the behavior of observed data, and enable us to tackle problems in which data variability plays a fundamental role (like the first-passage-time problem) with a higher degree of reliability.

To begin with, we will introduce the general model, describing its main features and how they relate to our procedure for the estimation of the time functions included. By analyzing the afore mentioned procedure we will be able to set the conditions that the available data must meet for a good fitting to be possible.

Given that the model we introduce is a generalization of simpler models in which functions included in the mean and infinitesimal variance could be independent on time, we address the problem of the maximum likelihood (ML) estimation of the constant parameter included in each of the infinitesimal moments when the time function included in the other one is known or can be approximated beforehand. Keeping in mind our goal of fitting the model to real data, we test some hypotheses to find out whether or not the use of non homogeneous models is justified. Furthermore, studying the errors incurred in fitting the model to the observed means and variances may provide additional criteria for the selection of the more appropriate model.

A simulation study is used to corroborate the validity of the proposed estimation procedures. Finally, a real data application of a patient-derived xenograft (PDX) tumor model is performed.

2. THE GENERAL NON-HOMOGENEOUS MODEL

From the ordinary differential equation of Malthusian growth, limited to a bounded time interval¹,

$$\frac{dx(t)}{dt} = rx(t), \quad t_0 \leq t \leq T,$$

replacing the fertility rate r (deterministic and constant) by a deterministic fertility depending on time, $h(t)$, and then replacing this fertility with $h(t) + \Lambda(t)$, where $\Lambda(t)$ is a white noise with time-dependent variance $v(t)$, the following Langevin equation is obtained:

$$\frac{dX(t)}{dt} = h(t)X(t) + X(t)\Lambda(t), \quad t_0 \leq t \leq T$$

¹Since it is our goal to fit real data observed in a given time interval, the model will have to be defined in said interval.

which, rewritten in the usual form for stochastic differential equations, leads to

$$(2) \quad dX(t) = h(t)X(t)dt + \sqrt{v(t)}X(t)dW(t), \quad t_0 \leq t \leq T,$$

where $W(t)$ denotes the standard Wiener process.

The solution of (2) is a non-homogeneous diffusion process $\{X(t); t \in [t_0, T]\}$ taking values in R^+ and with infinitesimal moments

$$(3) \quad \begin{aligned} A_1(x, t) &= h(t)x \\ A_2(x, t) &= v(t)x^2, \end{aligned}$$

with $h(t)$ and $v(t)$ continuous functions in $[t_0, T]$, and $v(t) > 0, \forall t \geq t_0$.

By considering an appropriate transformation (following Ricciardi, [11]) that changes the Kolmogorov equation of this process into that of the Wiener process, the transition probability density function (p.d.f.) of the process can be obtained. Concretely,

$$X(t)/X(s) = y \sim \Lambda \left(\log(y) + \int_s^t h(\tau)d\tau - \frac{1}{2} \int_s^t v(\tau)d\tau, \int_s^t v(\tau)d\tau \right).$$

To ensure that all finite dimensional distributions of the process are lognormal, the initial distribution should be considered either degenerate or lognormal. Moreover, we should note that the process defined as $Y(t) = X(t)/X(t_0)$ has the same infinitesimal moments² as $X(t)$, and therefore the same transition p.d.f. The only difference between the processes is that $Y(t)$ has a degenerate initial distribution³ at 1. Therefore, for the purposes of this paper (to estimate functions $h(t)$ and $v(t)$) we need only focus on the degenerate case, that is $P[X(t_0) = x_0] = 1$ (in particular, the case $x_0 = 1$). In such a case, the mean and variance functions are

$$(4) \quad \begin{aligned} E[X(t)] &= E[X(t)|X(t_0) = x_0] = x_0 \exp \left(\int_{t_0}^t h(\tau)d\tau \right), \\ Var[X(t)] &= Var[X(t)|X(t_0) = x_0] \\ &= (E[X(t)])^2 \left[\exp \left(\int_{t_0}^t v(\tau)d\tau \right) - 1 \right]. \end{aligned}$$

Note that, in this case, function given by (1), or equivalently $Var[X(t)]/(E[X(t)])^2$, must be a increasing function (condition much less restrictive than those for the non-homogeneous diffusion process only in the infinitesimal mean).

²The proof can be performed by applying the generalized Itô's formula (see Theorem 3.4.2 in [9]) to $f(X(t), t) = X(t)/X(t_0)$, which yields the same stochastic differential equation that $X(t)$ verifies.

³Note that with real growth data, consideration of the process $Y(t)$ involves the use of relative growth data, which it is usual in the study of tumor growth.

Moreover,
(5)

$$Var[\log(X(t))] = Var[\log(X(t))|X(t_0) = x_0] = \int_{t_0}^t v(\tau)d\tau.$$

3. FITTING THE MODEL

The procedure described below considers a process with a degenerate initial distribution. For this reason, if the available data do not fulfill this condition a transformation should be made, for example, to relative data with respect to the initial value for each sample path.

Estimation of $h(t)$ and $v(t)$

From (4), we can obtain the following expressions for $h(t)$

$$h(t) = H'(t),$$

with $H(t) = \log(E[X(t)])$. Similarly, from (5)

$$v(t) = V'(t),$$

with $V(t) = Var[\log(X(t))]$.

From values x_{ij} ($i = 1, \dots, d, j = 1, \dots, n$, with $x_{i1} = x_i, \forall i = 1, \dots, d$) corresponding to a discrete sampling of d sample paths of the process at time instants t_1, t_2, \dots, t_n , following to Albano et al. [2], [3], and Román-Román and Torres-Ruiz [14], the derivatives of functions H and V can be approximated from values $\log(x_j)$ (being x_j the mean growth data at time t_j) and b_j (the variance of the logarithms of the values of growth data at time t_j), $j = 1, \dots, n$. However, some remarks should be made:

- On the one hand, since function $v(t)$ must be positive, function $Var[\log(X(t))]$ must be increasing. However, the sequence of values $b_j, j = 1, 2, \dots, n$, obtained from real or simulated data, can not meet such condition. This can be due to the existence of random sampling fluctuations in the case of simulated or observed data of the model, or because the observed data can not be fitted by the model considered.
- On the other hand, the numerical approximation of the derivative of a function from observed values amplifies small deviations in the observations (see [13]).

In order to avoid errors in the approximation of the derivative and to preserve data for which the sequence of values b_j shows a growing trend, albeit with fluctuations that could lead to negative estimates of function $v(t)$ at some time instants, we propose smoothing out the data in order to produce an increasing function where possible. Whenever that can not be done, the present model should not be used for fitting such data. Similarly with data leading to a sequence of values of the function $Var[X(t)]/(E[X(t)])^2$ without an increasing trend, a good fit could not be achieved although the estimation procedure can be used.

Before approximating function $h(t)$, and, in this case, only in order to avoid errors of the derivative approximation, we also suggest smoothing out (in this case without restrictions) the sequence of values $\log(x_j), j = 1, \dots, n$.

On the basis of the above considerations, the proposed procedure to approximate functions $h(t)$ and $v(t)$ is the following:

- Approximate the values of functions $H(t)$ and $V(t)$ at each time instant by the corresponding values of a smoothing function fitting values $H_j = \log(x_j)$ and an increasing smoothing function⁴ fitting values $V_j = b_j, j = 1, \dots, n$, respectively.
- Determine values h_j and v_j as approximations of $H'(t_j)$ and $V'(t_j)$ from values H_j and $V_j, j = 1, \dots, n$, respectively, by using symmetric derivative at intermediate time instants, right derivative at the initial observation instant and left derivative at the final one.
- Finally, obtain approximations of functions $h(t)$ and $v(t)$ from values h_j and $v_j, j = 1, \dots, n$, respectively.

4. TESTING HYPOTHESES ABOUT TIME DEPENDENT FUNCTIONS

The model considered in this paper generalizes those characterized by infinitesimal moments

$$(6) \quad \begin{aligned} A_1(x, t) &= h(t)x \\ A_2(x, t) &= \sigma^2 x^2, \end{aligned}$$

and

$$(7) \quad \begin{aligned} A_1(x, t) &= mx \\ A_2(x, t) &= v(t)x^2. \end{aligned}$$

In this section we consider the problem of testing whether it is really necessary, in order to fit the available data, to employ a model of the type (3) or if it would suffice to go with one of the (6) or (7) type models. To address this we must first study, for model (6), the ML estimation of parameter σ^2 when function $h(t)$ is known or can be approximated beforehand. Similarly, for model (7), we must look into the estimation of parameter m when function $v(t)$ is known or can be approximated beforehand. In both cases, such estimations are based, in general, on observations $\mathbf{x} = \{x_{ij}, i = 1, \dots, d, j = 1, \dots, n_i\}$ corresponding to a discrete sampling of d sample paths of the process at times t_{ij} , with $t_{i1} = t_1, i = 1, \dots, d$, and by considering a degenerate initial distribution $P[X(t_1) = x_1] = 1$.

Estimating σ^2

For model (6) the transition p.d.f. is

⁴In both cases we have considered a smoothing spline employing the same number of equivalent parameters (degrees of freedom) as a previous local quadratic fit.

$$f(x, t|y, s) = \frac{1}{x\sqrt{2\pi\sigma^2(t-s)}} \times \exp\left(-\frac{\left[\log\left(\frac{x}{y}\right) - \int_s^t h(\tau)d\tau + \frac{\sigma^2}{2}(t-s)\right]^2}{2\sigma^2(t-s)}\right)$$

and the log-likelihood, $\log L_{\mathbf{x}}(\sigma^2)$, apart from constants, is

$$(8) \quad \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[\log\left(\frac{x_{ij}}{x_{i,j-1}}\right) - \int_{t_{i,j-1}}^{t_{ij}} h(\tau)d\tau + \frac{\sigma^2}{2}(t_{ij} - t_{i,j-1})\right]^2}{\sigma^2(t_{ij} - t_{i,j-1})} + (k-d)\log(\sigma^2),$$

where $k = \sum_{i=1}^d n_i$. So the ML estimation of σ^2 , $\hat{\sigma}_h^2$, is the solution to the equation

$$\frac{\sigma^4}{2} \left(\sum_{i=1}^d (t_{in_i} - t_{i1}) - \frac{1}{2} \sum_{i=1}^d \sum_{j=2}^{n_i} (t_{ij} - t_{i,j-1})^2 \right) + \sigma^2(k-d) - \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[\log\left(\frac{x_{ij}}{x_{i,j-1}}\right) - \int_{t_{i,j-1}}^{t_{ij}} h(\tau)d\tau\right]^2}{t_{ij} - t_{i,j-1}} = 0,$$

that maximizes (8).

Estimating m

For model (7) the transition p.d.f. is

$$f(x, t|y, s) = \frac{1}{x\sqrt{2\pi \int_s^t v(\tau)d\tau}} \times \exp\left(-\frac{\left[\log\left(\frac{x}{y}\right) - m(t-s) + \frac{1}{2} \int_s^t v(\tau)d\tau\right]^2}{2 \int_s^t v(\tau)d\tau}\right)$$

and the log-likelihood, $\log L_{\mathbf{x}}(m)$, apart from constants, is

$$x \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[\log\left(\frac{x_{ij}}{x_{i,j-1}}\right) - m(t_{ij} - t_{i,j-1}) + \frac{1}{2} \int_{t_{i,j-1}}^{t_{ij}} v(\tau)d\tau\right]^2}{\int_{t_{i,j-1}}^{t_{ij}} v(\tau)d\tau},$$

from which the ML estimation of m is

$$\hat{m}_v = \frac{\sum_{i=1}^d \sum_{j=2}^{n_i} (t_{ij} - t_{i,j-1}) \left[\frac{\log(x_{ij}/x_{i,j-1})}{\int_{t_{i,j-1}}^{t_{ij}} v(\tau)d\tau} + \frac{1}{2} \right]}{\sum_{i=1}^d \sum_{j=2}^{n_i} \frac{(t_{ij} - t_{i,j-1})^2}{\int_{t_{i,j-1}}^{t_{ij}} v(\tau)d\tau}}.$$

4.1 Testing hypotheses about function v(t)

We must find out whether model (6) suffices to fit the available data. More specifically, we will test hypothesis H_0 : $v(t) = \sigma^2$ for model (3).

The usual method of testing a hypothesis of equality between two functions is to consider a statistic measuring of

the distance between them, such as:

$$\int_{t_0}^T (\hat{v}(t) - \sigma^2)^2 dt \quad \text{or} \quad \int_{t_0}^T |\hat{v}(t) - \sigma^2| dt,$$

where $\hat{v}(t)$ is the estimation of function $v(t)$. Since in the context of experimental studies of tumor growth the available data are observations of the tumor volume at discrete time instants t_1, t_2, \dots, t_n , we consider the discrete versions of such statistics

$$\sum_{j=1}^n (\hat{v}(t_j) - \sigma^2)^2 \quad \text{or} \quad \sum_{j=1}^n |\hat{v}(t_j) - \sigma^2|,$$

and we propose a Monte Carlo procedure for computing a parametric bootstrap p value. After doing the estimations for model (3) (obtaining $\hat{h}(t)$ and $\hat{v}(t)$) and calculating the σ^2 value to be tested (ML estimation of σ^2 for model (6)) from the observed growth data (d sample paths at time instants t_1, \dots, t_n), the procedure, described for the second of the proposed distances, is as follows:

- Generate r bootstrap samples of d sample paths at time instants t_1, t_2, \dots, t_n of a process $\{X^*(t); t_1 \leq t \leq T\}$ of type (6) with $h(t) = \hat{h}(t)$ and $\sigma^2 = \hat{\sigma}_h^2$, that is, with infinitesimal moments

$$A_1^*(x, t) = \hat{h}(t) x \\ A_2^*(x, t) = \hat{\sigma}_h^2 x^2,$$

and initial distribution $P[X^*(t_1) = 1] = 1$.

- Calculate values

$$D_l = \sum_{j=1}^n |\tilde{v}_l(t_j) - \hat{\sigma}_h^2|, \quad l = 1, \dots, r,$$

where $\tilde{v}_l(t_j)$, $j = 1, \dots, n$ is the estimation of $v(t)$ at each t_j for a model of type (3) and for the l -th bootstrap sample.

- Since the test statistic takes positive values, calculate the p -value of the test statistic as the proportion of values D_l that are greater than or equal to $\sum_{j=1}^n |\hat{v}(t_j) - \hat{\sigma}_h^2|$.

4.2 Testing hypotheses about function h(t)

We must find out whether model (7) is enough to fit the available data. To this end, we will test hypothesis H_0 : $h(t) = m$ for model (3).

As in the previous case, we will start by considering statistics

$$\sum_{j=1}^n (\hat{h}(t_j) - m)^2 \quad \text{or} \quad \sum_{j=1}^n |\hat{h}(t_j) - m|,$$

and we propose a Monte Carlo procedure for computing a parametric bootstrap p value. After estimating model (3)

Table 1. Functions $h(t)$ and $v(t)$ for cases considered in the simulation study

	$h(t)$	$v(t)$
Case 1	0.05	0.04^2
Case 2	0.05	$(0.07(0.5 + t/40))^2$
Case 3	0.05	$0.001 + 0.0001t$
Case 4	$0.03 + 0.7e^{-0.4t}$	0.04^2
Case 5	$0.07 - 0.0005t$	0.04^2
Case 6	$0.05(1 + 0.125\sin(0.125\pi t)) + 0.05(0.125^2)\pi t\cos(0.125\pi t)$	$0.0025 + 0.00005t$
Case 7	$2(10 - t)/625$	$\frac{0.1 - 0.1(t-1)\left[(t-20)(40+(t-20)^2)^{-1} - 4(t-10)/625\right]}{0.1(t-1) + \sqrt{10(40+(t-20)^2)\exp(2(81-(t-10)^2)/625)}}$

Table 2. MRAEs for estimations of functions $h(t)$ and $v(t)$ for the models considered for fitting data in all cases

	Non-homogeneous model (3)		Non-homogeneous only in the infinitesimal mean model (6)		Non-homogeneous only in the infinitesimal variance model (7)		Homogeneous model	
	$h(t)$	$v(t)$	$h(t)$	$v(t)$	$h(t)$	$v(t)$	$h(t)$	$v(t)$
Case 1	0.0014	0.0359	0.0014	0.0261	0.0023	0.0359	0.0023	0.0261
Case 2	0.0113	0.0385			0.0035	0.0385		
Case 3	0.0051	0.0436			0.0051	0.0436		
Case 4	0.0224	0.0137	0.0224	0.0147				
Case 5	0.0057	0.0113	0.0057	0.0151				
Case 6	0.0634	0.0188						
Case 7	0.0445	0.0449						

(obtaining $\hat{h}(t)$ and $\hat{v}(t)$) and calculating the m value to be tested (ML estimation of m for model (7) with $v(t) = \hat{v}(t)$) from the observed growth data (d sample paths at time instants t_1, \dots, t_n), the procedure, described for the second of the proposed distances, is as follows:

- Generate r bootstrap samples of d sample paths at time instants t_1, t_2, \dots, t_n of a process $\{X^{**}(t); t_1 \leq t \leq T\}$ of type (7) with $v(t) = \hat{v}(t)$ and $m = \hat{m}_{\hat{v}}$, that is, with infinitesimal moments

$$\begin{aligned} A_1^{**}(x, t) &= \hat{m}_{\hat{v}} x \\ A_2^{**}(x, t) &= \hat{v}(t) x^2, \end{aligned}$$

and initial distribution $P[X^{**}(t_1) = 1] = 1$.

- Calculate values

$$D_l = \sum_{j=1}^n |\tilde{h}_l(t_j) - \hat{m}_{\hat{v}}|, \quad l = 1, \dots, r,$$

where $\tilde{h}_l(t_j)$, $j = 1, \dots, n$ is the estimation of $h(t)$ at each t_j for a model of type (3) and for the l -th bootstrap sample.

- Since the test statistic takes positive values, calculate the p -value of the test statistic as the proportion of values D_l that are greater than or equal to $\sum_{j=1}^n |\hat{h}(t_j) - \hat{m}_{\hat{v}}|$.

5. SIMULATION STUDY

In order to validate the estimation procedures proposed above, a simulation study has been carried out for several

combinations of functional forms of functions $h(t)$ and $v(t)$, including the constant case for each one (see Table 1). The functions $h(t)$ and $v(t)$ have been chosen so that the mean and variance functions of the resulting processes show interesting behaviors of several types.

In the cases in which one of the functions is constant, we can use the procedures described in Section 3 (without using the hypotheses mentioned therein and assuming the function is time dependent) and the corresponding one of Section 4. Moreover, in Case 1, the estimation of m and σ^2 for a homogeneous lognormal diffusion process has also been performed.

In all cases the estimation of the functions was replicated 150 times (by using 25 simulated sample paths with simulation step 0.1) and the MRAEs (the mean of relative absolute errors in the observation times) were calculated. The results are displayed in Table 2. Graphs (a) and (b) of Figures 1 to 7 show the different estimations of functions $h(t)$ and $v(t)$ respectively, along with the real functions in all cases.

In applications to real data, we would not be able to compare the estimation of functions $h(t)$ and $v(t)$ with the real functions, since the latter are unknown. For this reason, this simulation study includes the calculation of the MRAEs incurred in the estimation of the mean and variance functions of the fitted processes (see Table 3). In applications with real data the errors will be calculated with respect to the sample mean and variance. Graphs (c) and (d) of Figures 1 to 7 show the estimations of the mean and variance functions respectively, along with their real values.

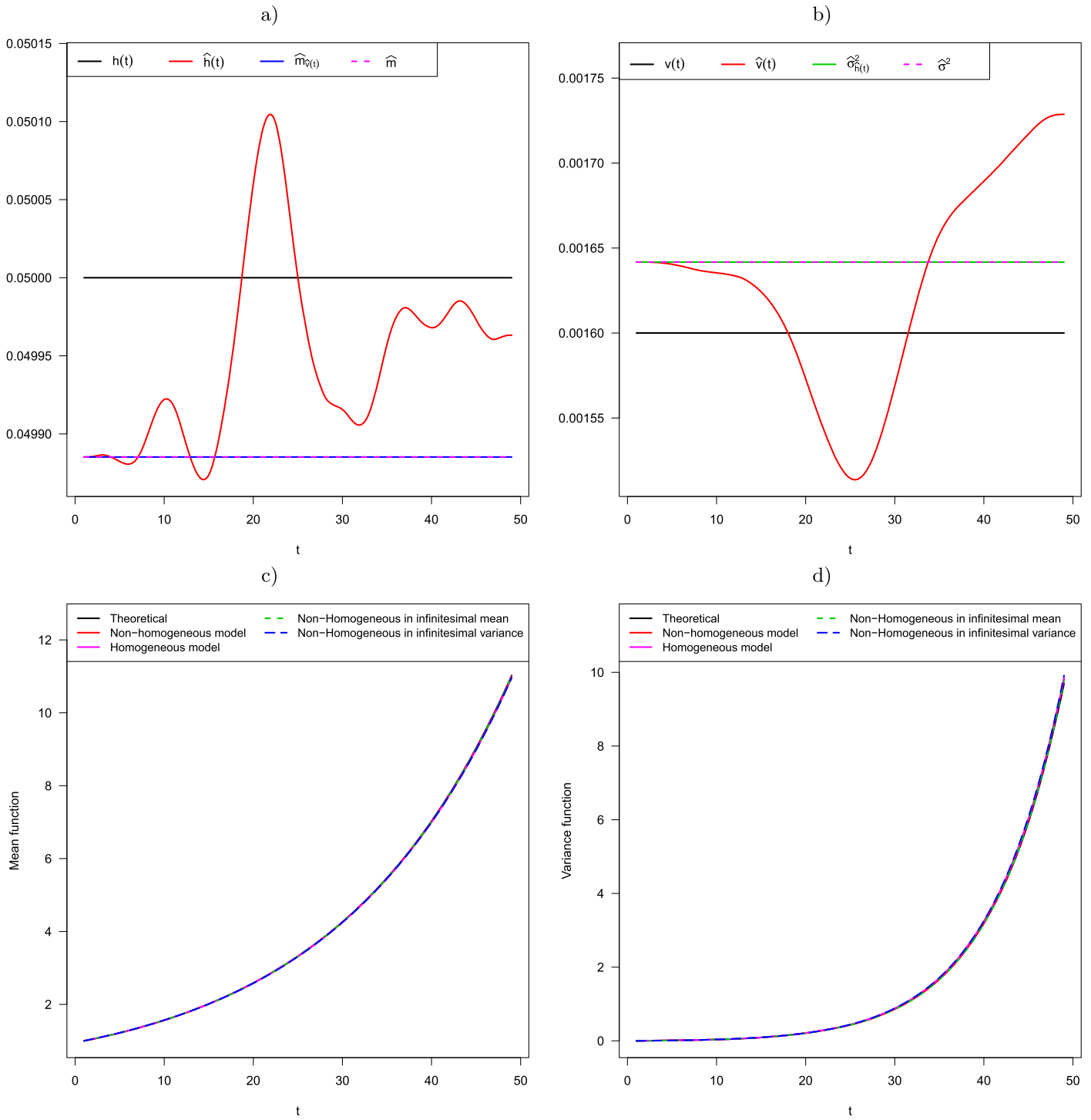


Figure 1. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 1.

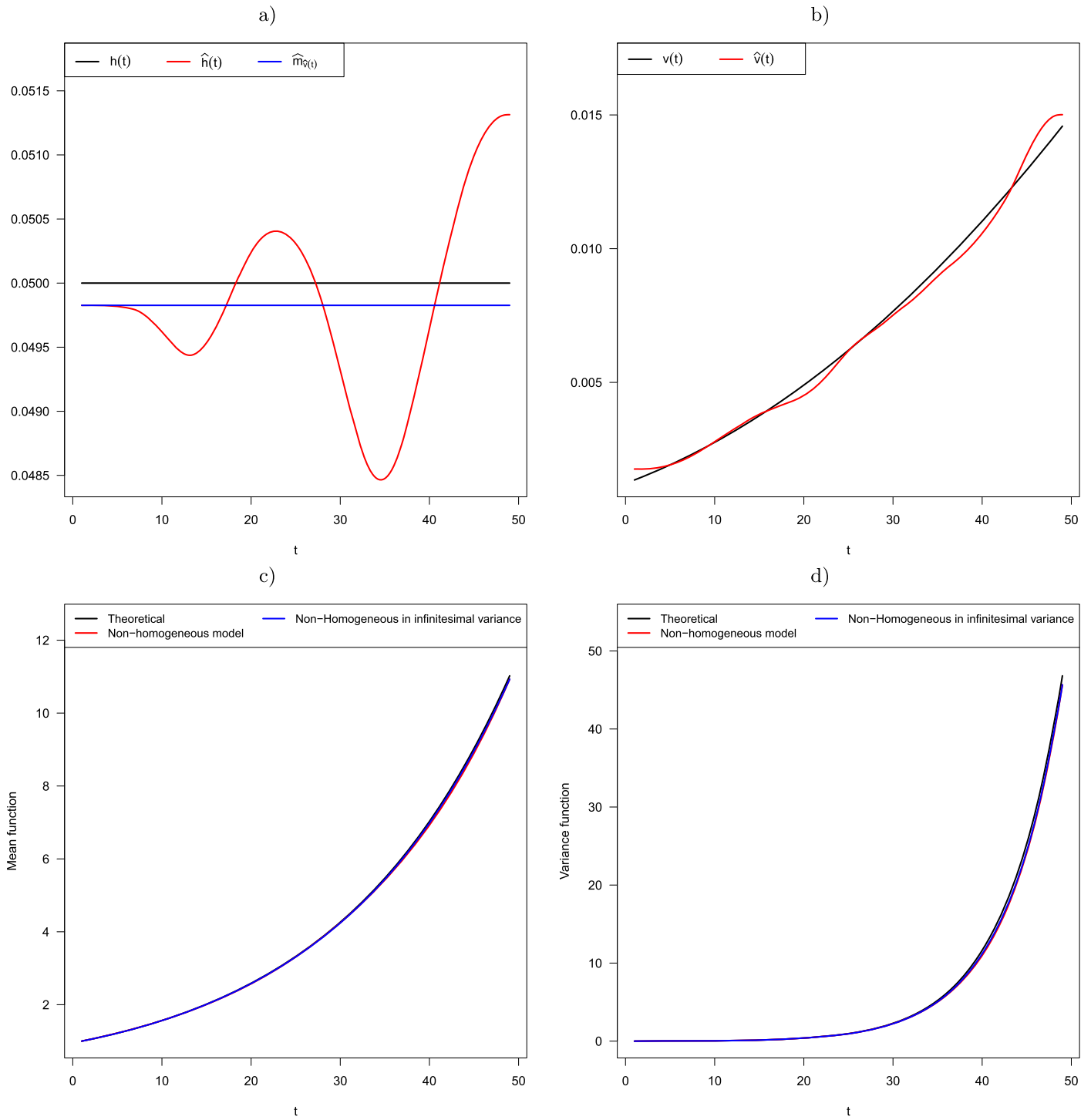


Figure 2. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 2.

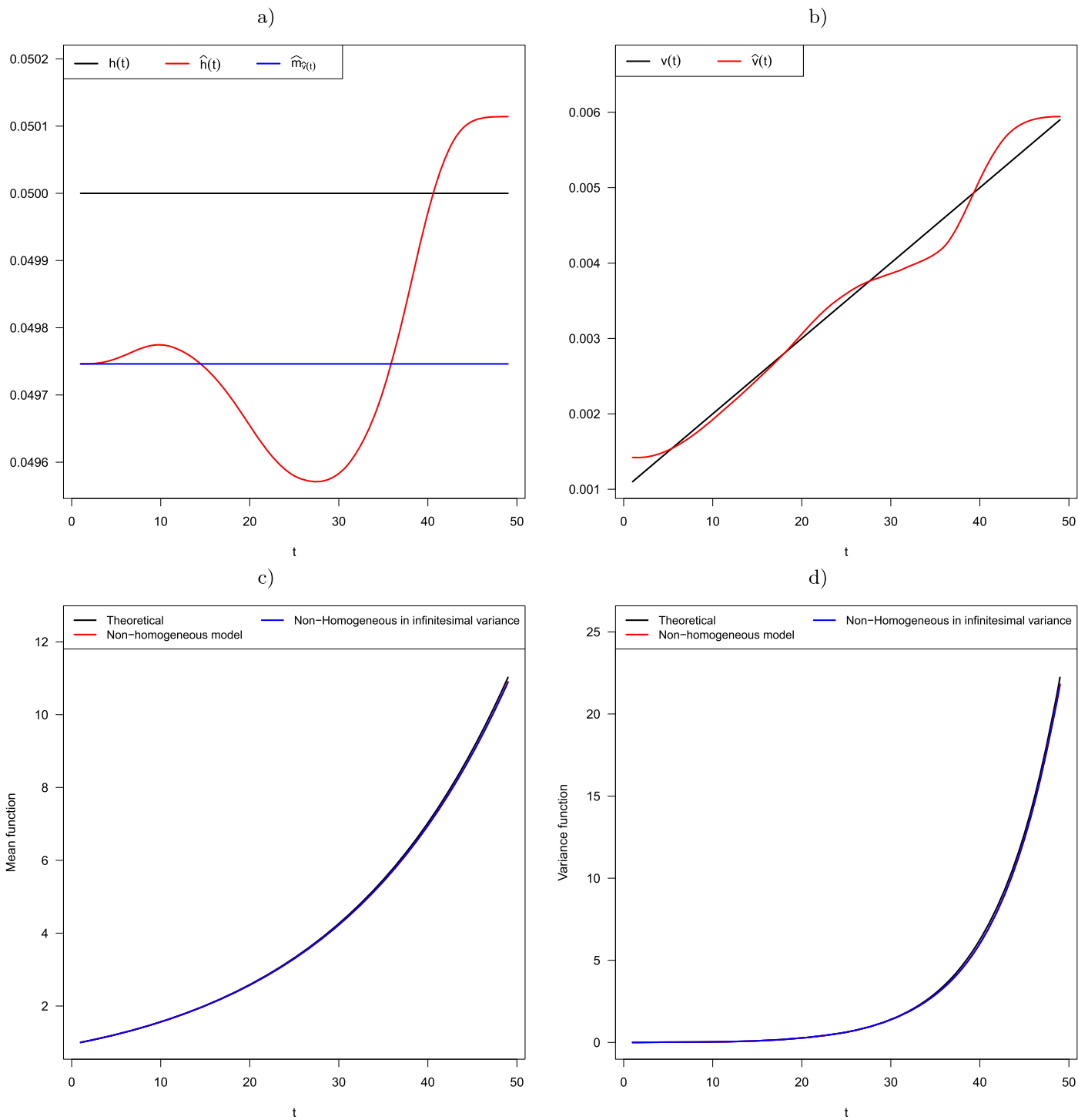


Figure 3. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 3.

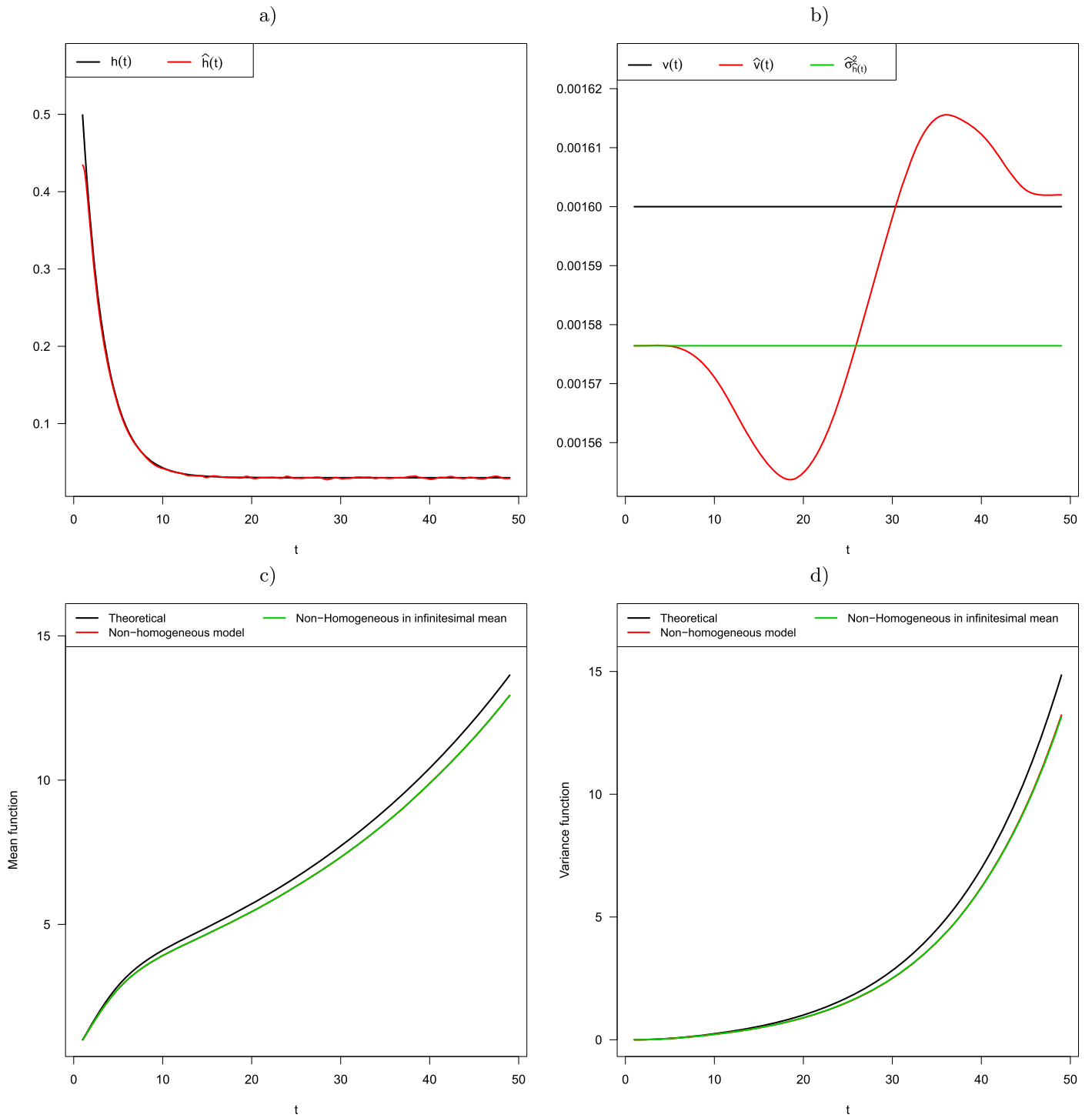


Figure 4. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 4.

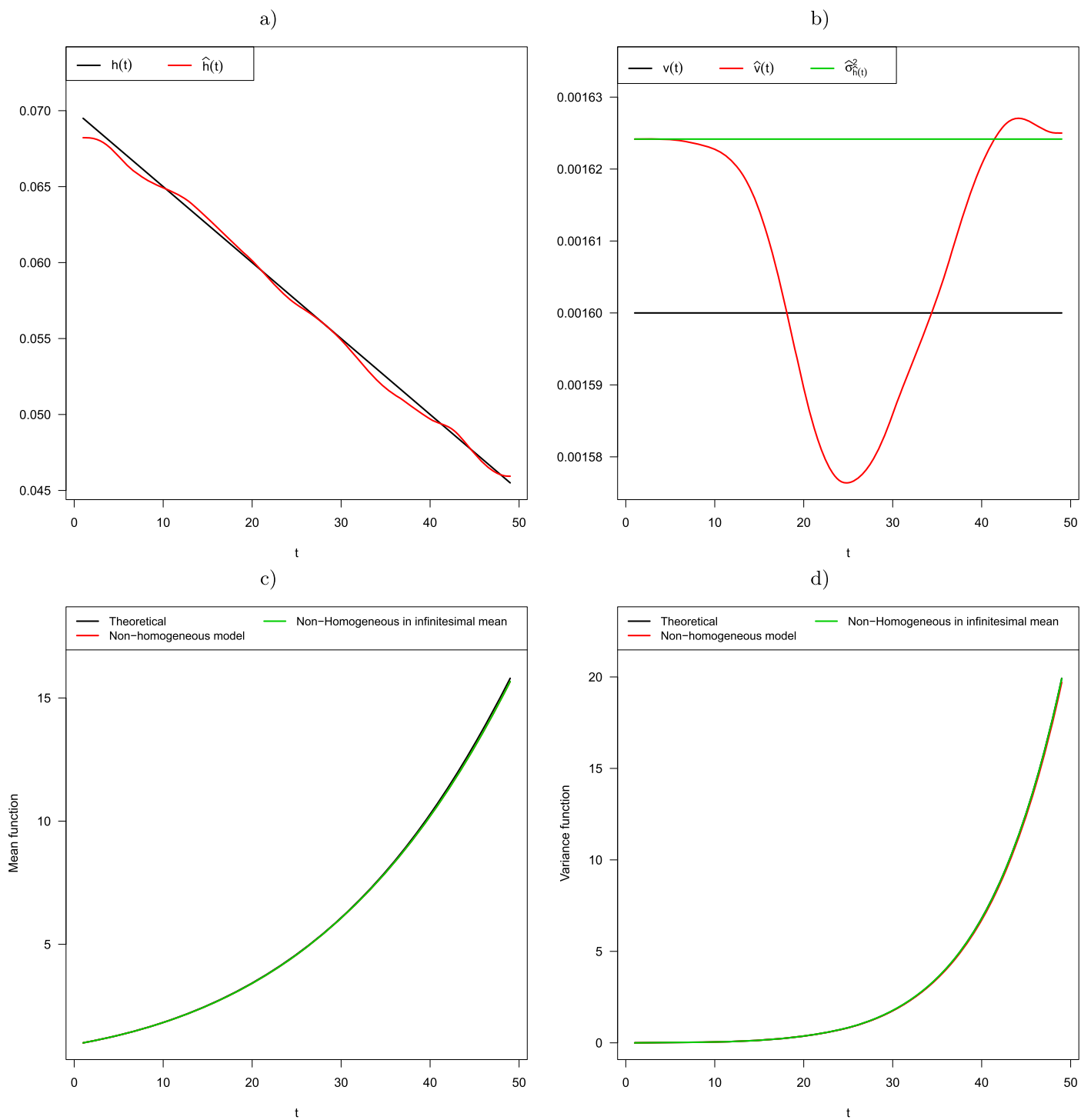


Figure 5. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 5.

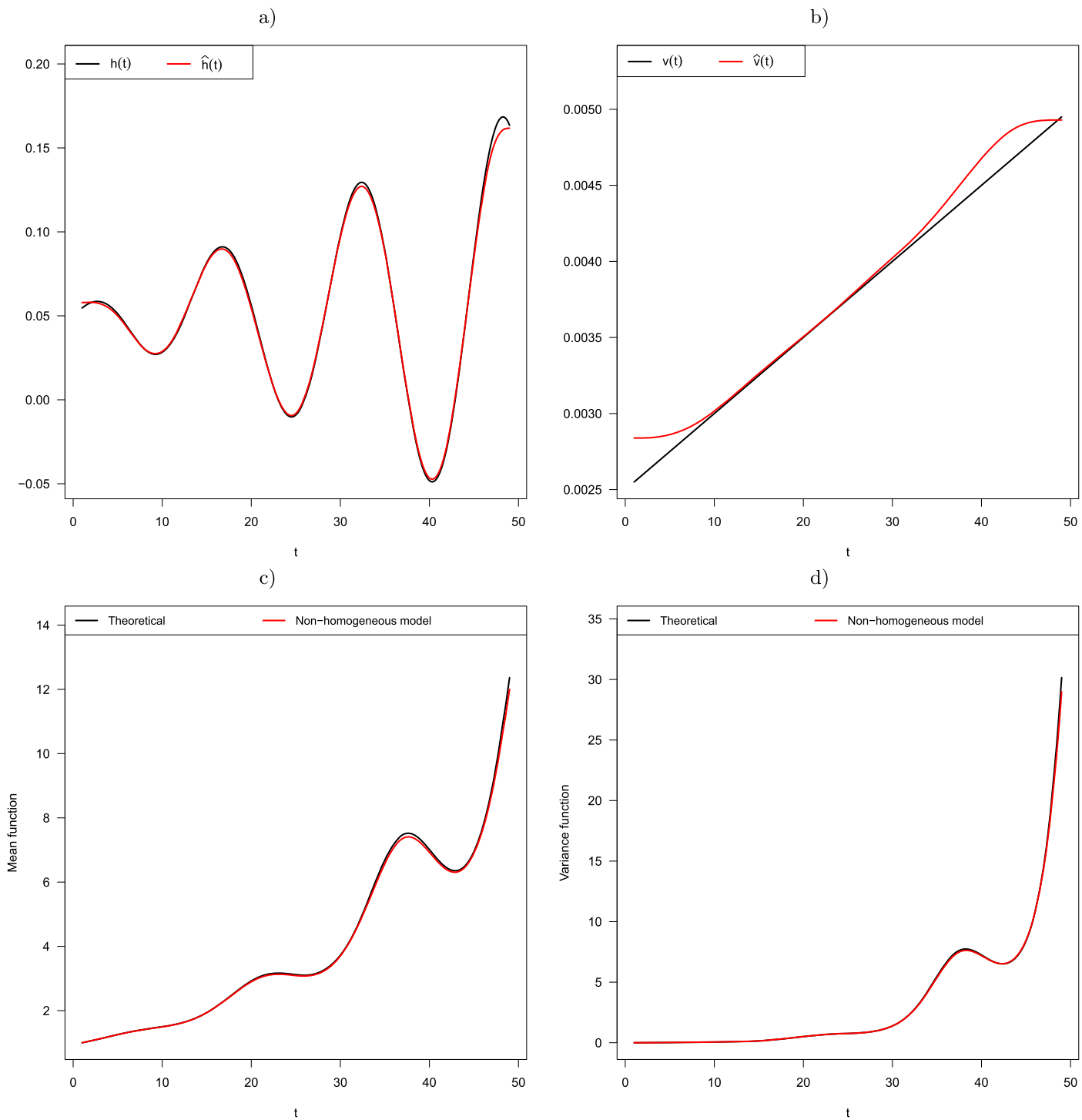


Figure 6. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 6.

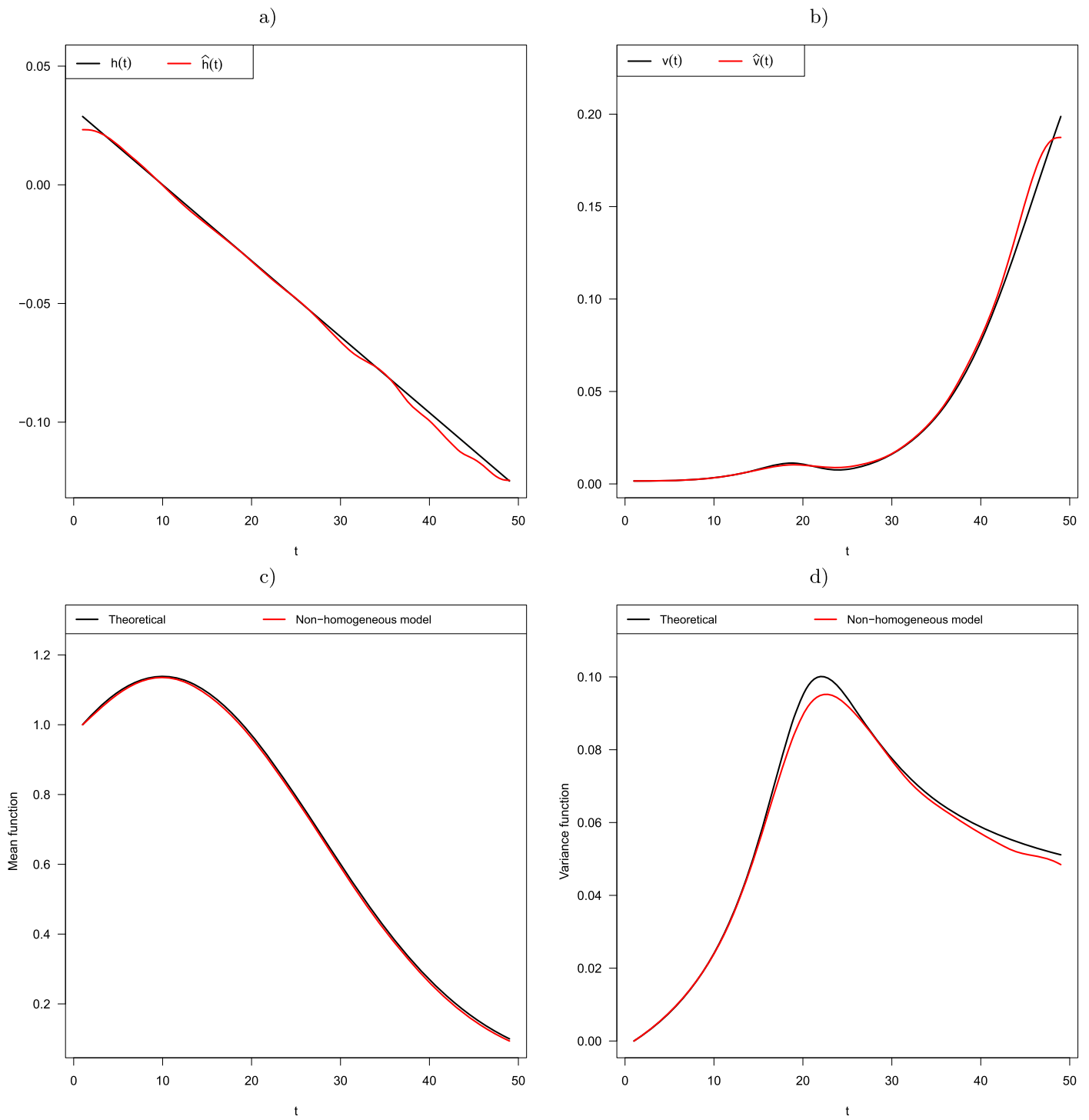


Figure 7. $h(t)$, $v(t)$, mean and variance functions and their approximations for the models considered for fitting data of Case 7.

Table 3. MRAEs for estimations of the mean and variance functions for the models considered for fitting data in all cases

	Non-homogeneous model (3)		Non-homogeneous only in the infinitesimal mean model (6)		Non-homogeneous only in the infinitesimal variance model (7)		Homogeneous model	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Case 1	0.0016	0.0122	0.0016	0.0234	0.0028	0.0121	0.0028	0.0210
Case 2	0.0063	0.0459			0.0042	0.0431		
Case 3	0.0066	0.0361			0.0061	0.0349		
Case 4	0.0468	0.1065	0.0468	0.1052				
Case 5	0.0045	0.0079	0.0045	0.0067				
Case 6	0.0079	0.0223						
Case 7	0.0176	0.0246						

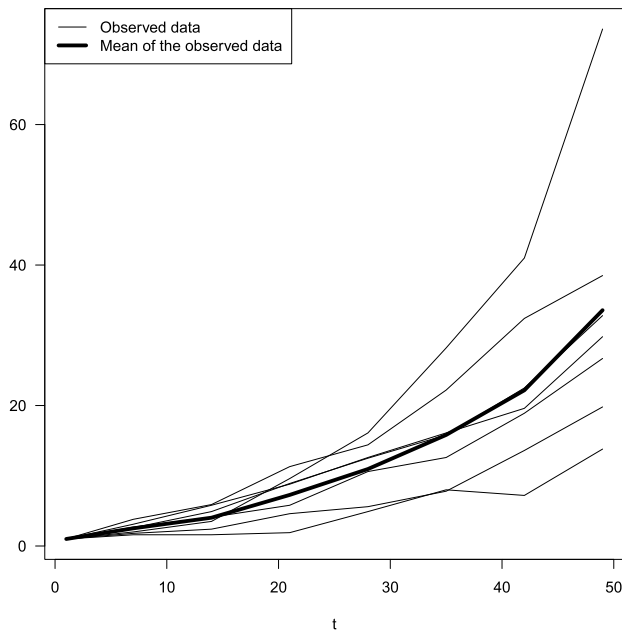


Figure 8. Tumor growth for experimental group.

The results show that the estimation procedures proposed are valid both for the case of general non-homogeneous process and for models in which one or both infinitesimal moments are homogeneous (when it makes sense your consideration), resulting in errors very small and very similar for different models.

6. APPLICATION TO REAL DATA

We have considered data about the growth of *BC297MONp5* from an experimental group of 7 mice. The tumor volume was measured at days 1, 7, 14, 21, 28, 35, 42 and 49. Subsequently, the relative volume of tumor with respect to the initial volume was calculated.

Figure 8 shows the evolution and the mean relative tumor volume as a function of the days after starting treatment.

For a non-homogeneous model of type (3), the estimates of functions $h(t)$ and $v(t)$ are shown in Figure 9a) and 9b),

Table 4. MRAEs in the estimations of the mean and variance for real data

	Mean	Variance
Non-homogeneous model (3)	0.0000149	0.2218521
Non-homogeneous only in the infinitesimal mean model (6)	0.0000148	0.3720499
Non-homogeneous only in the infinitesimal variance model (7)	0.4572298	0.7474404

respectively. In order to confirm that a model of this type (including a time-dependent function both in its infinitesimal mean and in its infinitesimal variance) is really necessary, we must solve the following hypothesis testing problems: $H_0 : h(t) = m$ and $H_0 : v(t) = \sigma^2$.

In the first case, by applying the method described in Subsection 4.2, the value of m to be tested is $\hat{m}_{\hat{v}(t)} = 0.05746798$, and the value of statistic $\sum_{j=1}^n |\hat{h}(t_j) - \hat{m}_{\hat{v}(t)}|$ is 0.2241380, which yields a p-value of 0 and leads to the rejection of the null hypothesis.

In the second case, by applying the method described in Subsection 4.1, the value of σ^2 to be tested is $\hat{\sigma}_{\hat{h}(t)}^2 = 0.08680918$ and the value of statistic $\sum_{j=1}^n |\hat{v}(t_j) - \hat{\sigma}_{\hat{h}(t)}^2|$ is 0.04573504, which yields a p-value of 0.08. This does not provide clear evidence for the rejection of the null hypothesis.

Figures 10 to 12 show the approximations of the mean and variance provided by the three models under consideration: general non-homogeneous (3), non-homogeneous only in its infinitesimal mean (6), and non-homogeneous only in its infinitesimal variance (7) respectively, along with their observed values. In addition, Table 4 displays the MRAEs incurred in the estimation of mean and variance for the three models. The graphs of Figure 12 and the MRAEs in the third row of Table 4 show how inadequate a type (7) model is for fitting the data, as already proven by testing the hypothesis $H_0 : h(t) = 0.05746798$. The comparison of the graphs in Figures 10 and 11 and the MRAEs for models (3) and

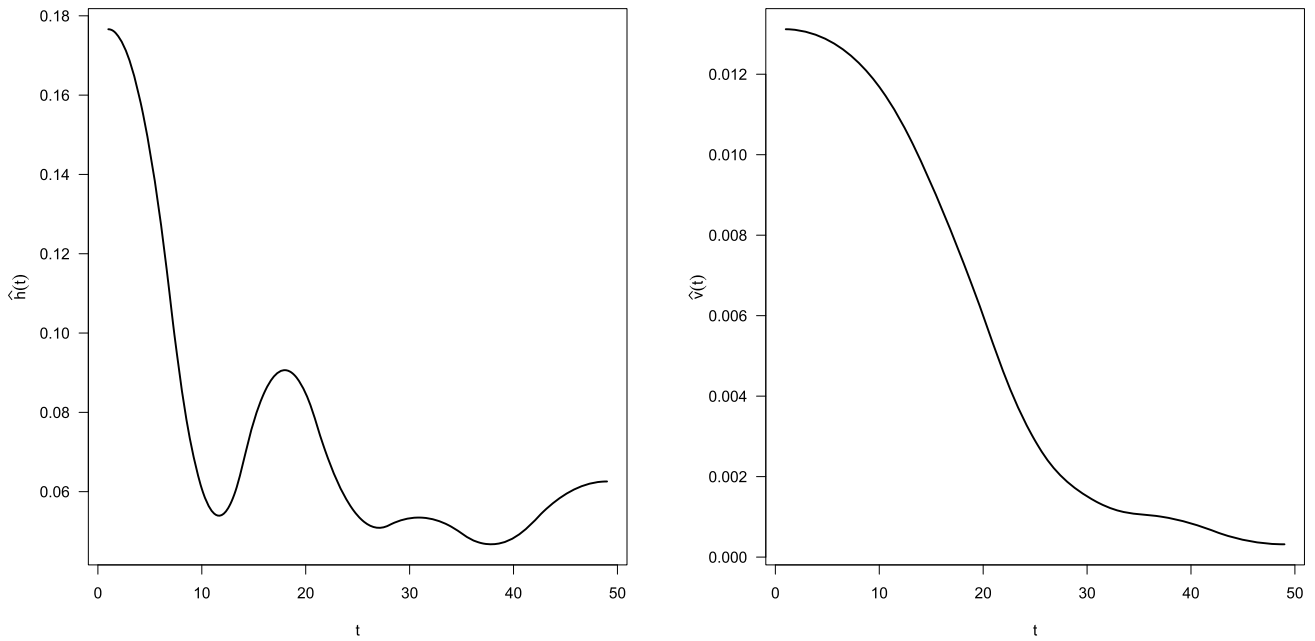


Figure 9. Approximated $h(t)$ and $v(t)$ for the non-homogenous model fitted to real data.

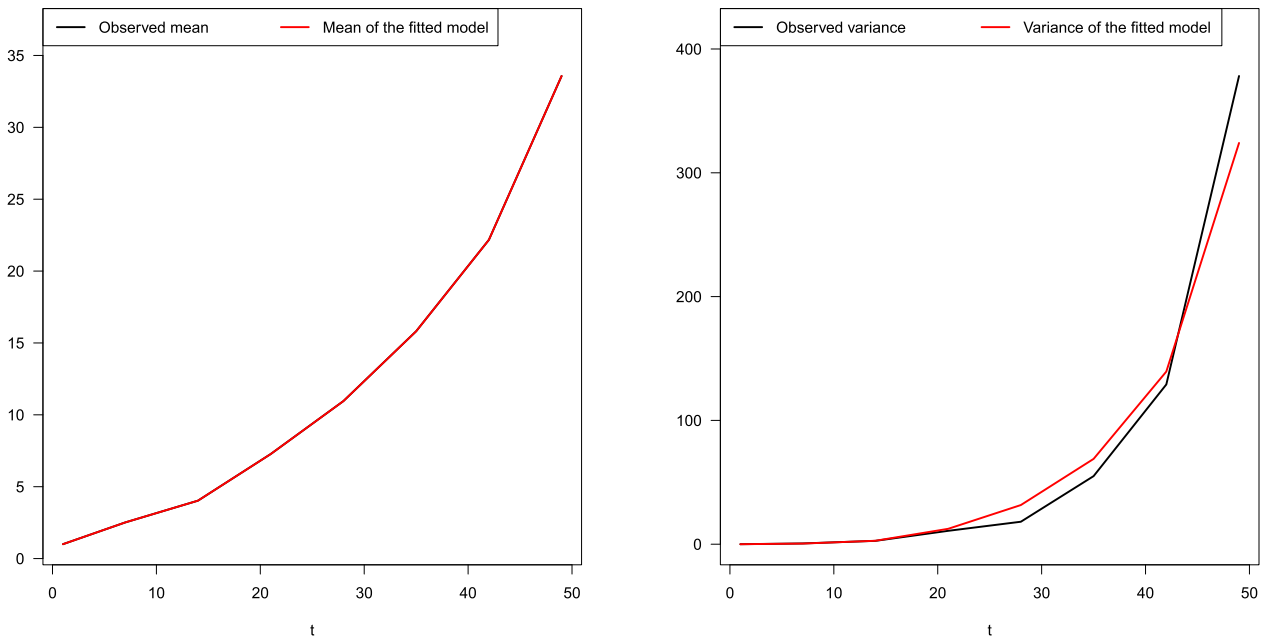


Figure 10. Approximated $E(X(t))$ and $Var(X(t))$ for the non-homogenous model fitted to real data.

(6) shows that, despite a significance level less than 0.08 would result in accepting a type (6) model for the purpose of fitting the data, the non-homogeneous model proposed in this paper significantly improves the fitting to the variance.

ACKNOWLEDGEMENTS

The authors are very grateful to Dr Didier Decaudin at the Translational Research Department of Institut Curie, Paris, for providing the data used in this research.

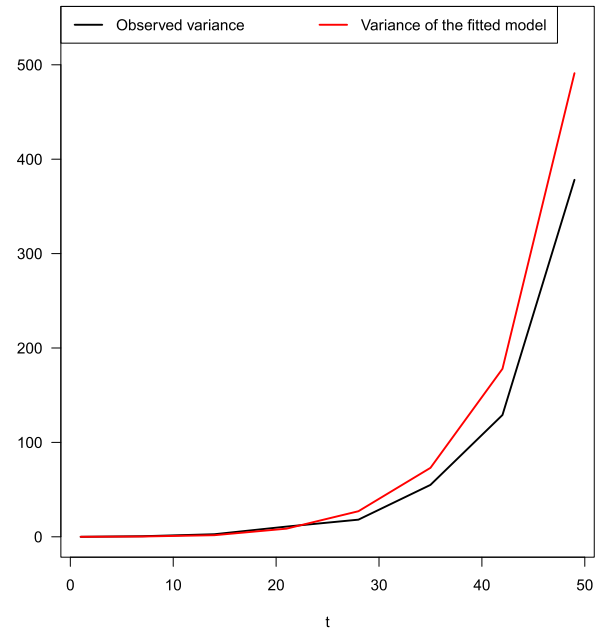
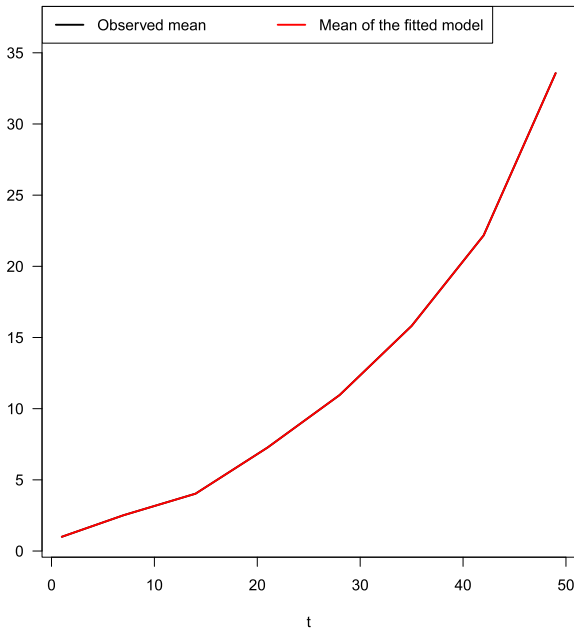


Figure 11. Approximated $E(X(t))$ and $Var(X(t))$ for the model non-homogenous only in its infinitesimal mean fitted to real data.

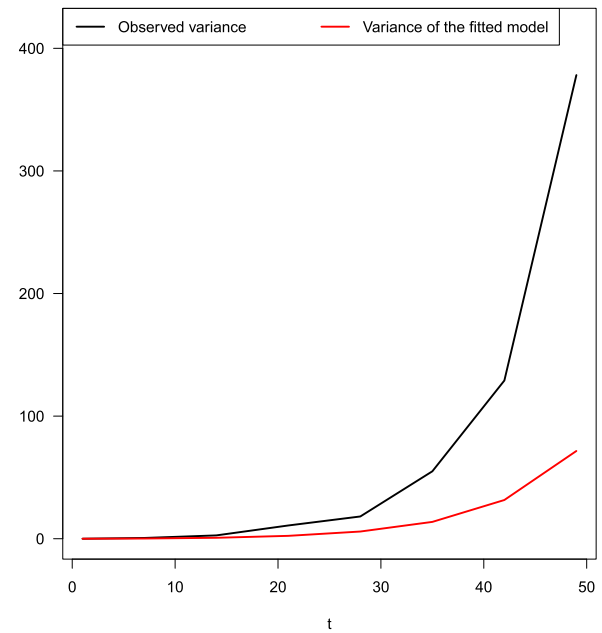
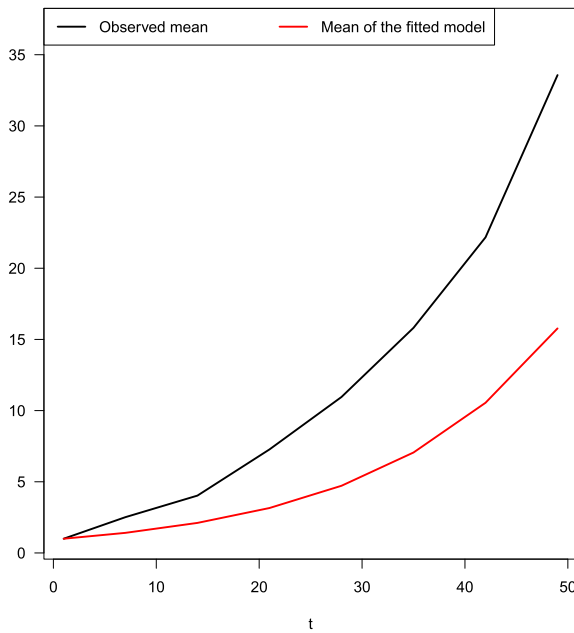


Figure 12. Approximated $E(X(t))$ and $Var(X(t))$ for the model non-homogenous only in its infinitesimal variance fitted to real data.

Received 15 April 2016

REFERENCES

- [1] AL-EIDEH, B. M., AL-REFAI, A. S. A, SBEITI, W. M. (2004). Modelling the CPI using a lognormal diffusion process and implications on forecasting inflation. *IMA J. Manag. Math.* **15** 39–51. [MR2064070](#)
- [2] ALBANO, G., GIORNO, V., ROMÁN-ROMÁN, P., TORRES-RUIZ, F (2011). Inferring the effect of therapy on tumors showing stochastic Gompertzian growth. *J. Theor. Biol.* **276** 67–77. [MR2974973](#)
- [3] ALBANO, G., GIORNO, V., ROMÁN-ROMÁN, P., ROMÁN-ROMÁN, S., TORRES-RUIZ, F. (2015). Estimating and determining the effect of a therapy on tumor dynamics by means of a modified Gompertz diffusion process. *J. Theor. Biol.* **364** 206–219.

- [4] BUONOCORE, A., CAPUTO, L., NOBILE, A. G., PIROZZI, E. (2015). Restricted Ornstein-Uhlenbeck process and applications in neuronal models with periodic input signals. *J. Comput. Appl. Math.* **285** 59–71. [MR3325252](#)
- [5] GIORNO, V., NOBILE, A. G., RICCIARDI, L. M. (1987). On some time-non-homogeneous diffusion approximations to queueing systems. *Adv. Appl. Prob.* **19** 974–994. [MR0914602](#)
- [6] GUTIÉRREZ, R., ROMÁN, P., TORRES, F. (1999). Inference and first-passage-times for the lognormal diffusion process with exogenous factors. Application to modelling in economics. *Appl. Stoch. Model. Bus.* **15** 325–332.
- [7] GUTIÉRREZ JÁIMEZ, R., GUTIÉRREZ SÁNCHEZ, R., NAFIDI, A., ROMÁN ROMÁN, P., TORRES RUIZ, F. (2005). Inference in Gompertz type non homogeneous stochastic systems by means of discrete sampling. *Cybernet. Syst.* **36** 203–216.
- [8] KHODABIN, M., MALEKNEJAD, K., ROSTAMI, M., NOURI, M. (2012). Interpolation solution in generalised stochastic exponential population growth model. *Appl Math. Model.* **36** 1023–1033. [MR2872875](#)
- [9] LAWLER, G. F. (2014). *Stochastic calculus: an introduction with applications*, <http://www.math.uchicago.edu/~lawler/finbook2.pdf> (accessed 17 March 2016).
- [10] LEFEBVRE, M. (2002). Using a lognormal diffusion process to forecast river flows. *Water Resour. Res.* **38** 12-1–12-8.
- [11] RICCIARDI, L. M. (1976). On the transformation of diffusion processes into the Wiener processes. *J. Math. Anal. Appl.* **54** 185–199. [MR0405608](#)
- [12] RICCIARDI, L. M. AND LÁNSKY, P. (2002). Diffusion models of neuron activity. In *The handbook of brain theory and neural networks*, edited by M. A. Arbib. Cambridge, MA: The MIT Press, pp. 343–348.
- [13] ROBERTS, S. Computing derivatives and integrals. Engineering Computation, Lecture 6, <http://www.robots.ox.ac.uk/~sjrob/Teaching/EngComp/ec16.pdf> (accessed 17 March 2016).
- [14] ROMÁN-ROMÁN, P., TORRES-RUIZ, F. (2012). Inferring the effect of therapies on tumor growth by using diffusion processes. *J. Theor. Biol.* **314** 34–56. [MR2984442](#)
- [15] ROMÁN-ROMÁN, P., ROMÁN-ROMÁN, S., SERRANO-PÉREZ, J. J., TORRES-RUIZ, F. (2016). Modeling tumor growth in the presence of a therapy with an effect on rate growth and variability by means of a modified Gompertz diffusion process. *J. Theor. Biol.* **407** 1–17. [MR3541905](#)

Patricia Román-Román
 Dpto. Estadística e Investigación Operativa
 Universidad de Granada
 Spain
 E-mail address: proman@ugr.es

Juan José Serrano-Pérez
 Dpto. Estadística e Investigación Operativa
 Universidad de Granada
 Spain
 E-mail address: jjberra@ugr.es

Francisco Torres-Ruiz
 Dpto. Estadística e Investigación Operativa
 Universidad de Granada
 Spain
 E-mail address: fdeasis@ugr.es