Computational Statistics and Data Analysis 198 (2024) 107996

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Latent event history models for quasi-reaction systems

Matteo Framba [a], Veronica Vinciotti [a,*], Ernst C. Wit [b]

[a] University of Trento, Italy
[b] Università della Svizzera italiana, Switzerland

ARTICLE INFO

Dataset link: http://dati.istat.it/Index.aspx?QueryId=18460

Dataset link: https://github.com/pcm-dpc/COVID-19

Keywords:
SDEs
Local linear approximation
Kalman filter
EM algorithm

ABSTRACT

Various processes, such as cell differentiation and disease spreading, can be modelled as quasi-reaction systems of particles using stochastic differential equations. The existing Local Linear Approximation (LLA) method infers the parameters driving these systems from measurements of particle abundances over time. While dense observations of the process in time should in theory improve parameter estimation, LLA fails in these situations due to numerical instability. Defining a latent event history model of the underlying quasi-reaction system resolves this problem. A computationally efficient Expectation-Maximization algorithm is proposed for parameter estimation, incorporating an extended Kalman filter for evaluating the latent reactions. A simulation study demonstrates the method's performance and highlights the settings where it is particularly advantageous compared to the existing LLA approaches. An illustration of the method applied to the diffusion of COVID-19 in Italy is presented.

1. Introduction

An increasing number of natural phenomena can be described by quasi-reaction systems of stochastic differential equations, as these are able to capture the inherent stochasticity of many processes. Examples include the stem cell differentiation process (Pellin et al., 2019, 2023), the dynamics of a biological system (Wilkinson, 2018) or of an infectious disease spreading (Britton and Pardoux, 2019), and the diverse applications of diffusion processes (Craigmile et al., 2023). The dynamics of these systems depend critically on parameters which are often unknown. Estimating these parameters is therefore important for characterizing and predicting the evolution of a dynamic system.

The likelihood of the intermittently observed process has rarely an explicit form (Wilkinson, 2018). To overcome this problem, Local Linear Approximation (LLA) methods provide an explicit approximation of the likelihood function under some assumptions (Shoji and Ozaki, 1998). Nevertheless, both in the case when observations are too spaced out in time and when the inter-observations times are too close, estimates based on the LLA are biased. Komorowski et al. (2011) present an extensive study on the effects of correlation between molecule concentrations on statistical inference, in the specific case of stochastic chemical kinetics models. Various approaches for reducing the variance of parameter estimators in a generic multi-response, non-linear model are available and could be used also in the case of dynamic systems. In the context of D-optimal designs, the most commonly used criterion assumes knowledge of the variance-covariance matrix (Fedorov, 2013). Although alternatives exist that use only an estimate of this matrix (Cooray-Wijesinha and Khuri, 1987), recent studies have observed that minimising the determinant of the information matrix is computationally efficient but not very robust (Hatzis and Larntz, 1992). An alternative approach is the use of Tikhonov

* Corresponding author at: Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Povo, Italy.
E-mail addresses: matteo.framba@unitn.it (M. Framba), veronica.vinciotti@unitn.it (V. Vinciotti), ernst.jan.camiel.wit@usi.ch (E.C. Wit).

https://doi.org/10.1016/j.csda.2024.107996
Received 25 October 2023; Received in revised form 27 May 2024; Accepted 28 May 2024

Available online 31 May 2024

regularisation techniques (Engl et al., 1996). However, if the measurements are taken very close together in time, the concentrations can be constant, leading to zero standard deviations and making also regularisation infeasible.

An approach to overcome these limitations is proposed. Intuitively, when the particles in the system are observed very close in time, one may be able to reconstruct which events of the stochastic process have taken place in order to result in a change of the system from the current to the next time point. Thus, the core element of the proposed approach involves integrating event history analysis into the framework of quasi-reaction systems. Originally conceived for sociological studies, event history models have been used on a range of applications, from engineering to medicine, economics, political science and psychology (Box-Steffensmeier and Jones, 2004). As the rates governing the evolutions of the state of the system and of the underlying event counting process are clearly linked, and they depend on the previous state of the system, the first contribution of the paper will be to formalise a joint statistical model that couples the two processes.

The second contribution of the paper is to develop an inferential procedure for the proposed model. As the occurrence of the events is not observed, an Expectation-Maximation (EM) algorithm for parameter estimation is derived. In particular, at the E-step, Kalman filter (Kalman, 1960) is used for the prediction of the latent events from the dynamics of the system observed on the entire time interval. The most popular version of Kalman filtering is for the case of Gaussian linear systems. However, an extension is required. Firstly, since the latent event counts are not Gaussian, the Poisson distribution is approximated with a continuous Gamma distribution, which is then transformed to a Gaussian distribution via a marginal transformation. Secondly, since the resulting system is non-linear, an extended Kalman filtering procedure is proposed for estimating the latent state of the event count process (Anderson and Moore, 2012). This allows the evaluation of the Q-function, which is then maximised at the M-step of the EM algorithm. In this way, the approach relates to other implementations of EM algorithms with an embedded Kalman filter, such as (Shumway and Stoffer, 1982; Ghahramani and Hinton, 1996) for dynamic linear systems, and more recent extensions for non-linear systems, such as the EM extended Kalman filter of Bar-Shalom et al. (2001), the EM unscented Kalman filter of Wan and Van Der Merwe (2000) and the EM particle filter of Zia et al. (2008).

The rest of the paper is organized as follows. In Section 2, the latent event history model for quasi-reaction systems is formalized. In Section 3, the EM algorithm for parameter estimation is described. In Section 4, a simulation study demonstrates the method's performance and highlights the settings where it is particularly advantageous compared to the existing LLA approaches. In Section 5, an illustration of the method on the modelling of the COVID-19 transmission dynamics in Italy is presented. Finally, in Section 6, conclusions and directions for future work are discussed.

2. Modelling quasi-reaction systems

Consider a closed system in which p substrates interact, each denoted as Q_l with $l = 1, \dots, p$. These substrates could represent the compartments of an infectious disease model, the cell types in a cell differentiation model, or the different molecules in a biochemical reaction system. The j-th chemical reaction can generally be described as

$$k_{1j}Q_1 + \dots + k_{pj}Q_p \xrightarrow{\beta_j} s_{1j}Q_1 + \dots + s_{pj}Q_p \qquad j \in 1, \dots, r, \tag{1}$$

where r indicates the number of reactions describing the dynamic system. Let $\mathcal{R}$ denote the set of possible reactions. The *stoichiometric coefficients* k_{lj} and s_{lj} are fixed integer values that describe the amount of substrate l, as reactant and product, respectively, that is needed for reaction j to occur, while $\theta_j = \exp(\beta_j) \in \mathbb{R}^+$ is the rate at which reaction j occurs.

The log-reaction rates $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ characterize the evolution of the dynamic system. These are the parameters that need to be estimated, given realizations of the state of the system over time. Let then $Y_l(t)$ denote the amount of the l-th particle at time t, with $t \in [0, T]$. Let $\mathbf{Y}(t) = (Y_1(t), \dots, Y_p(t))^T \in \mathbb{N}_0^p$ denote the state of the system at time t. Even if reaction equations like (1) are often used to represent kinetic models, as these facilitate a qualitative understanding of the dynamics, from a mathematical point of view chemical reactions are modelled primarily as systems of stochastic differential equations (Wilkinson, 2018). This methodology enables a quantitative interpretation of the dynamics, as it allows to study the temporal variation of the counts $\mathbf{Y}$ from the dynamics at the unit level.

According to the underlying dynamic system, particles encounter resulting in an instantaneous firing of one of the reactions. As a result, the system moves to the next state. Two viewpoints can be taken in the characterization of the stochastic process that induces changes in the state $\mathbf{Y}$ over time. The first viewpoint, presented in Section 2.1, models the stochastic process by which reactions occur and, as a by-product, the state of the system $\mathbf{Y}$ moves to a new configuration, deterministically. A second viewpoint, presented in Section 2.2, directly describes the change of the system based on the amount of particles available at a certain point in time and the hazard rate of each reaction at that point in time. As the occurrence of reactions is not observed, the second viewpoint is the most direct approach for modelling dynamic systems. Indeed, this is the approach considered in the literature and the one that results in the traditional LLA approaches for parameter estimation.

Instead, when the system is observed at small time intervals, one may be able to reconstruct the underlying process of reactions, leading to a more accurate characterization of the dynamic system. The main reason is the high temporal correlation between the states at small time scales. Motivated by this, Section 2.3 shows how the two viewpoints can be unified into a joint statistical model.

2.1. Event process

Let $e_j := (t_j, r_j)$ be the *event* that reaction $j \in \mathcal{R}$ occurs at time t_j. Associated with this marked point process and with each reaction j, there is a multivariate counting process

$$N_j(t) = \#\{\text{Reactions of type } j \text{ occurring in time interval } [0,t], j \in \mathcal{R}\}.$$

$N_j(t)$ is assumed to follow a non-homogeneous Poisson process

$$N_j(t) \sim \text{Poisson}(\Lambda_j(t)),$$

with cumulative rate

$$\Lambda_j(t) = E[N_j(t) \mid \mathcal{F}_{t^-}] = \int_0^t \lambda_j(\boldsymbol{Y}(u);\boldsymbol{\beta})du,$$

where $\mathcal{F}_{t^-}$ is the history of the process up to, but excluding, time t.

The hazard rate $\lambda_j(\boldsymbol{Y}(t);\boldsymbol{\beta})$ depends on the state of the system at time t as well as on the amount of particles of each type that are needed for each reaction to occur, i.e., the stoichiometric coefficients k_{lj} in (1). In particular, it holds that (Wilkinson, 2018)

$$\lambda_j(\boldsymbol{Y}(t);\boldsymbol{\beta}) = \exp(\beta_j)\prod_{l=1}^{p}\binom{Y_l(t)}{k_{lj}}, \tag{2}$$

where $\binom{Y_l(t)}{k_{lj}} = 0$, for all $Y_l(t) < k_{lj}$.

2.2. Particle count process

The state of the system $\boldsymbol{Y}(t)$ is itself also a continuous time discrete Markov process. In the particular setting of a quasi-reaction system, it is possible to establish the temporal evolution of the probability distribution $P(\mathbf{Y};t)$, i.e., the probability that $\mathbf{Y}$ is the state of the system at time t. This will again depend on the state of the system just before time t. In particular, the distribution can be shown to satisfy the chemical master equations (Wilkinson, 2018)

$$\frac{dP(\boldsymbol{Y};t)}{dt} = \sum_{j\in\mathcal{R}}\left[\lambda_j\left(\boldsymbol{Y}(t) - V_{\cdot j};\boldsymbol{\beta}\right)P\left(\boldsymbol{Y} - V_{\cdot j};t\right) - \lambda_j(\boldsymbol{Y}(t);\boldsymbol{\beta})P(\boldsymbol{Y};t)\right], \tag{3}$$

where V denotes the net effect matrix, with (l,j) entry given by $v_{lj} = s_{lj} - k_{lj}$.

A solution of (3) gives the full transition probability kernel for the system dynamics. The master equations, however, can be solved analytically only in a small number of cases, due to the vast spectrum of conceivable state configurations (McQuarrie, 1967). On the other hand, from the master equations, one can derive the conditional expectation and variance of the rate of changes of the system. These are given, respectively, by

$$\begin{aligned}\frac{\mathbb{E}[\mathbf{Y}(t+dt) - \mathbf{Y}(t) \mid \mathbf{Y}(t)]}{dt} &= V\lambda(\boldsymbol{Y}(t);\boldsymbol{\beta}),\\ \frac{\mathbb{V}[\mathbf{Y}(t+dt) - \mathbf{Y}(t) \mid \mathbf{Y}(t)]}{dt} &= V\,\text{diag}(\lambda(\boldsymbol{Y}(t);\boldsymbol{\beta}))V^T.\end{aligned} \tag{4}$$

These two moments form the basis of the LLA solution to the master equations via a generalised least-squares approach (Pellin et al., 2019). Alternative approximations based on the van Kampen expansion have been proposed within a Bayesian inferential approach (Capistrán et al., 2012).

2.3. Latent event history model

The two characterizations described above are now merged into one joint model based on realizations of the process at discrete time points. Let then $\mathbf{Y}_i = \mathbf{Y}(t_i)$, $i = 0,\dots,N$, be the state of the process at $N+1$, not necessarily equispaced, time points. Under the non-homogeneous Poisson process described in Section 2.1 and assuming that the hazard rates remain constant within the N time intervals, the increments of event counts follow a Poisson distribution, conditional on the history of the process. In particular,

$$\Delta N_{ij} = N_j(t_i) - N_j(t_{i-1}) \mid \mathcal{F}_{t_{i-1}} \sim \text{Poisson}(\mu_{ij}(\mathbf{Y}_{i-1};\boldsymbol{\beta})), \quad j = 1,\dots,r, \tag{5}$$

where

$$\mu_{ij}(\mathbf{Y}_{i-1};\boldsymbol{\beta}) = (t_i - t_{i-1})\lambda_j(\mathbf{Y}_{i-1};\boldsymbol{\beta}), \tag{6}$$

with $\lambda_j(\mathbf{Y}_{i-1};\boldsymbol{\beta})$ defined as in Equation (2). For the rest of the manuscript, $\Delta\mathbf{N}_i$ and $\boldsymbol{\mu}_i$ denote the vectors of reaction counts ΔN_{ij} and rates $\mu_{ij}(\mathbf{Y}_{i-1};\boldsymbol{\beta})$, respectively, in the interval $(t_{i-1},t_i]$ across the r reactions.

It is clear how knowledge of the increments $\Delta\mathbf{N}_i$ would allow for perfect prediction of the state of the system at time t_i, since $\mathbf{Y}_i - \mathbf{Y}_{i-1} = V\Delta\mathbf{N}_i$. Combined with (5), this implies that $\mathbf{Y}_i - \mathbf{Y}_{i-1}$ is a linear combination of Poisson random variables, conditional on the history of the process. However, this linear combination does not have an explicit distribution in itself, and, more importantly, the increments for different particle types are not independent, leading to a further complication in the likelihood. For this reason, an approximate state-space formulation of the process that circumvents a direct full likelihood approach is proposed.

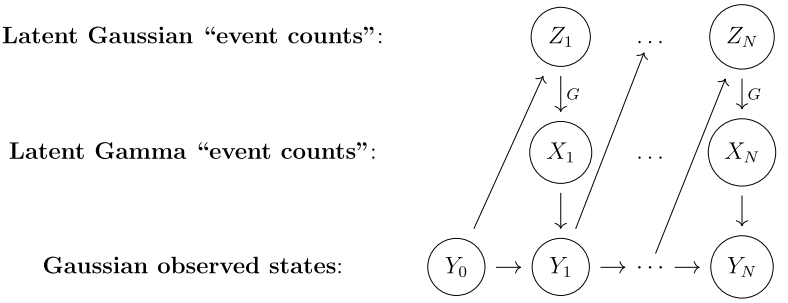

Fig. 1. Latent event history model DAG. The state of the system at time t_i, $\mathbf{Y}_i$, depends on the previous state, $\mathbf{Y}_{i-1}$, and on the number of reactions $\mathbf{X}_i$ that occur in the interval $(t_{i-1}, t_i]$. The latent $\mathbf{X}_i$ is approximated with a Gamma distribution and connected deterministically with a Gaussian random vector $\mathbf{Z}_i$, via a marginal transformation G. Notice how $\mathbf{Z}_i$ is independent of future states, $\mathbf{Y}_{(i+1):N}$, conditional on current and past states, $\mathbf{Y}_{0:i}$.

To this end, an approximation of the Poisson distribution of $\Delta\mathbf{N}_i$ with a continuous distribution is proposed. In particular, a Gamma distribution with a mean and variance matching that of $\Delta\mathbf{N}_i$, and with a similar skewness, is considered. In this way, the process can be rewritten as a Gaussian state space model. More in detail, the discrete increments ΔN_{ij} in Equation (5) are associated to the continuous random variable $X_{ij} = F_{ij}^{-1}(\Phi_{ij}(Z_{ij}))$, where F_{ij} is the CDF of a Gamma distribution with scale parameter 1 and shape parameter $\mu_{ij}(\mathbf{Y}_{i-1}; \boldsymbol{\beta})$ from Equation (6), and Φ_{ij} is the CDF of a Gaussian distribution with mean and variance both equal to $\mu_{ij}(\mathbf{Y}_{i-1}; \boldsymbol{\beta})$. So Z_{ij} is the Gaussian random variable that is uniquely associated to the Gamma distributed random variable X_{ij}, and with the same conditional mean and variance as the original ΔN_{ij} variable. In the remaining of the paper, $\mathbf{Z}_i$ will denote the r-dimensional vector of Gaussian random variables associated to the event counts in the interval $(t_{i-1}, t_i]$, $\mathbf{X}_i$ the corresponding Gamma random variables and G the function that transforms $\mathbf{Z}_i$ into $\mathbf{X}_i$, namely

$$\mathbf{X}_i = G(\mathbf{Z}_i) = \left(F_{i1}^{-1}(\Phi_{i1}(Z_{i1})), \ldots, F_{ir}^{-1}(\Phi_{ir}(Z_{ir}))\right).$$

With the latent event counts $\Delta\mathbf{N}_i$ approximated by $\mathbf{X}_i$, it follows that, approximately, $\mathbf{Y}_i - \mathbf{Y}_{i-1} = V\mathbf{X}_i = VG(\mathbf{Z}_i)$. In the following, the state space model will be formulated more generally, so as to account also for possible measurement error in the observations $\mathbf{Y}_i$, which may be relevant in some applied settings. In particular, the following latent event history model is proposed:

$$\begin{cases} \mathbf{Z}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, & \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \mathrm{diag}(\boldsymbol{\mu}_i)\right), \\ \mathbf{Y}_i = \mathbf{Y}_{i-1} + VG(\mathbf{Z}_i) + \boldsymbol{\psi}_i, & \boldsymbol{\psi}_i \sim \mathcal{N}(0, \Sigma), i = 1, \ldots, N, \end{cases} \tag{7}$$

where $\boldsymbol{\psi}_i$ is a Gaussian noise vector with mean zero and variance-covariance $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. The case of no measurement error in $\mathbf{Y}_i$, which will be considered in the simulations, will correspond to the special case of $\sigma_l^2 = 0$, $l = 1, \ldots, p$. Fig. 1 summarizes the dependence structure associated to the proposed model.

3. Inference

This section discusses statistical inference of the latent event history model (7). Denoting with $\mathbf{Y}$ the $(N+1) \times p$ matrix of observations at the $N+1$ time points and $\mathbf{Z}$ the $(N+1) \times r$ matrix of latent variables, estimation of $\boldsymbol{\beta}$ and Σ requires the optimization of the marginal log-likelihood

$$\ell_{\mathbf{Y}}(\boldsymbol{\beta}, \Sigma) = \log \int_{\mathbf{Z}} L_{\mathbf{Z},\mathbf{Y}}(\boldsymbol{\beta}, \Sigma) d\mathbf{Z}. \tag{8}$$

As common in the presence of latent variables, an Expectation-Maximisation (EM) algorithm for parameter estimation is derived (Dempster et al., 1977). To this end, the complete log-likelihood, conditional on the initial state $\mathbf{Y}_0$ and assuming some measurement error in $\mathbf{Y}_i$ ($\Sigma \neq 0$), can be factorized into

$$\ell_{\mathbf{Z},\mathbf{Y}}(\boldsymbol{\beta}, \Sigma) = \sum_{i=1}^{N} \left[\ell_{\mathbf{Z}_i|\mathbf{Y}_{i-1}}(\boldsymbol{\beta}) + \ell_{\mathbf{Y}_i|\mathbf{Z}_i,\mathbf{Y}_{i-1}}(\Sigma)\right], \tag{9}$$

where

$$\ell_{\mathbf{Z}_i|\mathbf{Y}_{i-1}}(\boldsymbol{\beta}) = -\frac{1}{2}\left\{ r\log(2\pi) + \log(|\mathrm{diag}(\boldsymbol{\mu}_i)|) + \left[\mathbf{Z}_i - \boldsymbol{\mu}_i\right]^T (\mathrm{diag}(\boldsymbol{\mu}_i))^{-1} \left[\mathbf{Z}_i - \boldsymbol{\mu}_i\right] \right\}$$

and

$$\ell_{\mathbf{Y}_i|\mathbf{Z}_i,\mathbf{Y}_{i-1}}(\Sigma) = -\frac{1}{2}\left\{ p\log(2\pi) + \log(|\Sigma|) + \left[\mathbf{Y}_i - \mathbf{Y}_{i-1} - VG(\mathbf{Z}_i)\right]^T \Sigma^{-1} \left[\mathbf{Y}_i - \mathbf{Y}_{i-1} - VG(\mathbf{Z}_i)\right] \right\}.$$

The EM algorithm will then consist in the following two steps, which are iterated until convergence:

- *E-step*: Setting $\boldsymbol{\beta}$ and Σ to the current estimate of the parameters, $\boldsymbol{\beta}^*$ and Σ^*, respectively, compute the expected value of the complete log-likelihood (9) with respect to the distribution of the latent variables given the observations:

$$Q(\boldsymbol{\beta}, \Sigma | \boldsymbol{\beta}^*, \Sigma^*) = \mathbb{E}_{\mathbf{Z}|\mathbf{Y}, \boldsymbol{\beta}^*, \Sigma^*}[\ell_{\mathbf{Z},\mathbf{Y}}(\boldsymbol{\beta}, \Sigma)]. \tag{10}$$

- *M-step*: Find the optimal $\boldsymbol{\beta}$ and Σ by maximising the objective function (10) with respect to $\boldsymbol{\beta}$ and Σ.

In the next two sections, the computational aspects associated to the two steps, respectively, are discussed in detail.

3.1. E-step: extended Kalman filtering

With the complete log-likelihood written as in (9), the Q-function (10) with slight abuse of notation is given by

$$Q(\boldsymbol{\beta}, \Sigma | \boldsymbol{\beta}^*, \Sigma^*) = \mathbb{E}[\ell_{\mathbf{Z}|\mathbf{Y}}(\boldsymbol{\beta}) | \mathbf{Y}_{0:N}] + \mathbb{E}[\ell_{\mathbf{Y}|\mathbf{ZY}}(\Sigma) | \mathbf{Y}_{0:N}], \tag{11}$$

where $\mathbf{Y}_{0:N}$ denotes the data across all time points. Under model (7), the first term involves the following expectation

$$\begin{aligned}
\mathbb{E}[\ell_{\mathbf{Z}|\mathbf{Y}}(\boldsymbol{\beta}) | \mathbf{Y}_{0:N}] =& \mathbb{E}\left[-\frac{1}{2} \sum_{i=1}^{N} \left\{ r \log(2\pi) + \log(|\text{diag}(\boldsymbol{\mu}_i)|) + \left[\mathbf{Z}_i - \boldsymbol{\mu}_i\right]^T \text{diag}(\boldsymbol{\mu}_i)^{-1} \left[\mathbf{Z}_i - \boldsymbol{\mu}_i\right] \right\} | \mathbf{Y}_{0:N} \right] \\
&\propto -\frac{1}{2} \sum_{i=1}^{N} \left\{ \mathbb{E}\left[\mathbf{Z}_i^T | \mathbf{Y}_{0:N}\right] \text{diag}(\boldsymbol{\mu}_i)^{-1} \mathbb{E}\left[\mathbf{Z}_i | \mathbf{Y}_{0:N}\right] + \text{Tr}\left[\text{diag}(\boldsymbol{\mu}_i)^{-1} \mathbb{V}[\mathbf{Z}_i | \mathbf{Y}_{0:N}]\right] - 2\mathbb{E}\left[\mathbf{Z}_i^T | \mathbf{Y}_{0:N}\right] \cdot \mathbf{1} + \boldsymbol{\mu}_i^T \cdot \mathbf{1} \right\},
\end{aligned}$$

while expectation of the second term results in

$$\begin{aligned}
\mathbb{E}[\ell_{\mathbf{Y}|\mathbf{ZY}}(\Sigma) | \mathbf{Y}_{0:N}] =& \mathbb{E}\left[-\frac{1}{2} \sum_{i=1}^{N} \left\{ p \log(2\pi) + \log(|\Sigma|) + \left[\Delta\mathbf{Y}_i - V G(\mathbf{Z}_i)\right]^T \Sigma^{-1} \left[\Delta\mathbf{Y}_i - V G(\mathbf{Z}_i)\right] \right\} | \mathbf{Y}_{0:N} \right] \\
&\propto -\frac{1}{2} \sum_{i=1}^{N} \Big\{ -2\Delta\mathbf{Y}_i^T \Sigma^{-1} V \mathbb{E}\left[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}\right] + \mathbb{E}\left[G(\mathbf{Z}_i)^T | \mathbf{Y}_{0:N}\right] V^T \Sigma^{-1} V \mathbb{E}\left[G(\mathbf{Z}_i) \mid \mathbf{Y}_{0:N}\right] \\
&\quad + \text{Tr}(\Sigma^{-1} V \mathbb{V}\left[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}\right] V^T) \Big\},
\end{aligned}$$

with $\Delta\mathbf{Y}_i = \mathbf{Y}_i - \mathbf{Y}_{i-1}$ and keeping only the terms dependent on the latent variables.

In particular, the calculation of the Q-function requires the evaluation of the following first and second moments: $\mathbb{E}[\mathbf{Z}_i | \mathbf{Y}_{0:N}]$, $\mathbb{V}[\mathbf{Z}_i | \mathbf{Y}_{0:N}]$, $\mathbb{E}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}]$, and $\mathbb{V}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}]$. To this end, a Kalman filter approach is considered. Firstly, notice how the dependences implied by model (7) are such that

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}_i | \mathbf{Y}_{0:N}] &= \mathbb{E}[\mathbf{Z}_i | \mathbf{Y}_{0:i}], \\
\mathbb{V}[\mathbf{Z}_i | \mathbf{Y}_{0:N}] &= \mathbb{V}[\mathbf{Z}_i | \mathbf{Y}_{0:i}], \\
\mathbb{E}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}] &= \mathbb{E}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:i}], \\
\mathbb{V}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:N}] &= \mathbb{V}[G(\mathbf{Z}_i) | \mathbf{Y}_{0:i}],
\end{aligned}$$

since $\mathbf{Z}_i$ is independent of future states, $\mathbf{Y}_{(i+1):N}$, conditional on current and past states, $\mathbf{Y}_{0:i}$ (Fig. 1). In the following, the first two quantities are denoted with $\hat{\mathbf{z}}_{i|i}$ and $V_{i|i}$, respectively. This means that the smoothing step of a traditional Kalman filtering procedure is not needed, and only the prediction and update steps are. Secondly, the non-linearity in $\mathbf{Z}_i$ induced by the marginal transformation G means that a standard Kalman filter approach is not applicable. Thus, in order to calculate the first and second moments of $G(\mathbf{Z}_i)$, an extended Kalman filter is considered, where the function G is approximated with a second order Taylor expansion.

According to the derivations in Appendix A, the first two expectations are given by

$$\begin{aligned}
\hat{\mathbf{z}}_{i|i} &= \mathbb{E}\left[\mathbf{Z}_i \mid \mathbf{Y}_{0:i}\right] = \hat{\mathbf{z}}_{i|i-1} + K_i \left[\mathbf{Y}_i - \mathbf{Y}_{i-1} - V\left(\mathbf{g}_{i|i-1} + \frac{1}{2}\text{vect}(V_{i|i-1} H_{i|i-1})\right)\right], \\
V_{i|i} &= \mathbb{E}\left[\left(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i}\right)\left(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i}\right)^T \mid \mathbf{Y}_{0:i}\right] = \left(\mathbb{I}_r - K_i V J_{i|i-1}\right) V_{i|i-1},
\end{aligned}$$

Algorithm 1 Extended Kalman Filter (E-step).

Require: $\mathbf{Y}, \boldsymbol{\beta}^*, \Sigma^*, V$
for $i = 1, \dots, N$ **do**
1. *Prediction step*
$$\hat{\mathbf{z}}_{i|i-1} = \boldsymbol{\mu}_i$$
$$V_{i|i-1} = \text{diag}(\boldsymbol{\mu}_i)$$
2. *Update step*
$$\hat{\mathbf{z}}_{i|i} = \hat{\mathbf{z}}_{i|i-1} + K_i \left[\mathbf{Y}_i - \mathbf{Y}_{i-1} - V\left(\mathbf{g}_{i|i-1} + \tfrac{1}{2}\text{vect}(V_{i|i-1} H_{i|i-1})\right)\right]$$
$$V_{i|i} = \left(\mathbb{I} - K_i V J_{i|i-1}\right) V_{i|i-1}$$
with
$$K_i = (V V_{i|i-1} J_{i|i-1})^T (V J_{i|i-1} V_{i|i-1} J_{i|i-1}^T V^T + \Sigma)^{-1}$$
$$\mu_{ij} = \exp(\beta_j) \prod_{l=1}^{p} \binom{Y_{t_{i-1}}}{k_{ij}} (t_i - t_{i-1}) \qquad j = 1, \dots, r$$
end for

where

$$K_i = (V V_{i|i-1} J_{i|i-1})^T (V J_{i|i-1} V_{i|i-1} J_{i|i-1}^T V^T + \Sigma)^{-1},$$

and where the various quantities predicted from data up to time t_{i-1}, which are formally defined in Appendix A, are dependent on a current estimate of parameters $\boldsymbol{\beta}^*$ and Σ^*. As for the moments of $G(\mathbf{Z}_i)$, these are approximated by

$$\mathbb{E}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}\right] \approx \mathbf{g}_{i|i} + \frac{1}{2}\text{vect}(V_{i|i} H_{i|i}),$$
$$\mathbb{V}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}\right] \approx J_{i|i} V_{i|i} J_{i|i}^T,$$

with

$$\mathbf{g}_{i|i} = G(\mathbf{Z})|_{\hat{\mathbf{z}}_{i|i}}, \qquad J_{i|i} = \frac{\partial G(\mathbf{Z})}{\partial \mathbf{Z}}|_{\hat{\mathbf{z}}_{i|i}}, \qquad H_{i|i} = \frac{\partial^2 G(\mathbf{Z})}{\partial \mathbf{Z}^2}|_{\hat{\mathbf{z}}_{i|i}}.$$

In particular, note how these moments depend on the moments of $\mathbf{Z}_i$ derived above, i.e., $\hat{\mathbf{z}}_{i|i}$ and $V_{i|i}$, so the latter are the main quantities that need to be calculated at the E-step.

Algorithm 1 summarizes the calculations required for the Kalman filter at the E-step of the algorithm, based on a current estimate of parameters, $\boldsymbol{\beta}^*$ and Σ^*. The Kalman filter predictions of the latent states are used in the evaluation of $Q(\boldsymbol{\beta}, \Sigma|\boldsymbol{\beta}^*, \Sigma^*) = \mathbb{E}[\ell_{\mathbf{Z}|\mathbf{Y}}(\boldsymbol{\beta})|\mathbf{Y}_{0:N}] + \mathbb{E}[\ell_{\mathbf{Y}|\mathbf{ZY}}(\Sigma)|\mathbf{Y}_{0:N}]$, with

$$\mathbb{E}[\ell_{\mathbf{Z}|\mathbf{Y}}(\boldsymbol{\beta})|\mathbf{Y}_{0:N}] = -\frac{1}{2} N r \log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left\{ \log(|\text{diag}(\boldsymbol{\mu}_i)|) + \hat{\mathbf{z}}_{i|i}^T \text{diag}(\boldsymbol{\mu}_i)^{-1} \hat{\mathbf{z}}_{i|i} + \text{Tr}\left[\text{diag}(\boldsymbol{\mu}_i)^{-1} \cdot V_{i|i}\right] - 2\hat{\mathbf{z}}_{i|i}^T \cdot \mathbf{1} + \boldsymbol{\mu}_i^T \cdot \mathbf{1} \right\}, \quad (12)$$

$$\mathbb{E}[\ell_{\mathbf{Y}|\mathbf{ZY}}(\Sigma)|\mathbf{Y}_{0:N}] = -\frac{1}{2} N \log\left((2\pi)^p \prod_{l=1}^{p} \sigma_l^2\right) - \frac{1}{2}\sum_{i=1}^{N}\left\{ \Delta\mathbf{Y}_i^T \Sigma^{-1} \Delta\mathbf{Y}_i - 2\Delta\mathbf{Y}_i^T \Sigma^{-1} V\left(\mathbf{g}_{i|i} + \frac{1}{2}\text{vect}(V_{i|i} H_{i|i})\right)\right.$$
$$\left. + \left(\mathbf{g}_{i|i} + \frac{1}{2}\text{vect}(V_{i|i} H_{i|i})\right)^T V^T \Sigma^{-1} V \left(\mathbf{g}_{i|i} + \frac{1}{2}\text{vect}(V_{i|i} H_{i|i})\right) + \text{Tr}\left(\Sigma^{-1} V J_{i|i} V_{i|i} J^T V^T\right) \right\}, \quad (13)$$

and $\boldsymbol{\beta}^*$ and Σ^* the current values of the parameters used for the Kalman filter quantities $\hat{\mathbf{z}}_{i|i}$ and $V_{i|i}$.

3.2. M-step

The M-step maximizes the conditional expectation of the complete log-likelihood with respect to the parameters. Thus, the M-step involves the maximisation of the Q-function (11) with respect to $\boldsymbol{\beta}$ and Σ. Since the first term (12) does not depend on Σ, while the second term (13) does not depend directly on $\boldsymbol{\beta}$, the M-step results in the optimization of the first term (12) for the estimation of $\boldsymbol{\beta}$ and of the second term for the estimation of Σ. The latter is in fact available in closed form and is given by

$$\hat{\Sigma} = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}(\Delta Y_i - V G(\mathbf{Z}_i))(\Delta Y_i - V G(\mathbf{Z}_i))^T \middle| \mathbf{Y}_{0:N}\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left\{ \Delta Y_i Y_i^T - 2\Delta Y_i (g_{i|i} + \frac{1}{2} V_{i|i} H_{i|i})^T V^T + V(g_{i|i} + \frac{1}{2} V_{i|i} H_{i|i})(g_{i|i} + \frac{1}{2} V_{i|i} H_{i|i})^T V^T + V J_{i|i} V_{i|i} J_{i|i}^T V^T \right\}. \quad (14)$$

The optimal values of $\boldsymbol{\beta}$ and Σ from the M-step are used as the new $\boldsymbol{\beta}^*$ and Σ^*, respectively, for computing a new expected log-likelihood at the E-step. This iterative procedure is repeated until convergence, e.g., until the estimates of $\boldsymbol{\beta}$ do not change significantly. Algorithm 2 summarizes the proposed EM algorithm.

Algorithm 2 EM algorithm.

Require: $\mathbf{Y}, V, \boldsymbol{\beta}_{ini}, \Sigma_{ini}, \sigma^2, tol, maxit$
  while $err \geq tol$ & $it < maxit$ **do**
    for $i = 1, \dots, N$ **do**
  1. E-step:
      Extended Kalman Filter: calculate $\hat{\mathbf{z}}_{i|i}$, $V_{i|i}$ from $\mathbf{Y}$, V, $\boldsymbol{\beta}_{old}$ Σ_{old}
  2. M-step:
      $\boldsymbol{\beta}_{new}, \Sigma_{new} \leftarrow \arg\max_{\beta,\Sigma} Q(\boldsymbol{\beta}, \Sigma | \boldsymbol{\beta}_{old}, \Sigma_{old}, \hat{\mathbf{z}}_{i|i}, V_{i|i})$
      $err \leftarrow \max ||\boldsymbol{\beta}_{new} - \boldsymbol{\beta}_{old}||_1^1$
      $\boldsymbol{\beta}_{old} \leftarrow \boldsymbol{\beta}_{new}$
      $\Sigma_{old} \leftarrow \Sigma_{new}$
      $it \leftarrow it + 1$.
    end for
  end while

3.3. Computational cost

The computational cost of the proposed EM algorithm is the combination of the computational cost of the E- and M-steps. At the E-step, the latent variables $\mathbf{Z}_i$ across the N time intervals are of dimension r, with r the number of reactions, and their covariance $V_{i|i}$ requires the inversion of a $p \times p$ matrix, where p is the number of substrates. Thus, the total complexity of the E-step is $\mathcal{O}(Nrp^3)$. On the other hand, the M-step concerns the optimization of an r-dimensional vector of parameters $\boldsymbol{\beta}$ and involves N inversions of an $r \times r$ matrix for the calculation of the objective function. Thus, the total complexity of the M-step is $\mathcal{O}(Nr^3p)$. This results in a computational cost of the full algorithm of the order $\mathcal{O}(Nr^3p^3)$, although this may vary depending on the speed of convergence of the numerical algorithm used for the optimization of the Q-function at the M-step.

3.4. Standard errors of reaction rates

Estimates of the reaction rates $\boldsymbol{\theta} = \exp(\boldsymbol{\beta})$ are the main output of the EM inference. Uncertainties on these point estimates can be summarised by their standard errors. Since the marginal log-likelihood in (8) is not a direct result of the EM algorithm, the standard errors are calculated from the Fisher information matrix associated to the Q-function, evaluated at the point estimates of $\boldsymbol{\theta}$ and Σ (Oakes, 1999). In particular, this is given by

$$I(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2}{\partial \boldsymbol{\theta}^2} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}, \hat{\Sigma})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{15}$$

The variance of $\hat{\theta}_j$ is then given by $(I(\hat{\boldsymbol{\theta}})^{-1})_{jj}$.

If necessary, standard errors can be constructed also on $\boldsymbol{\beta}$. In particular, using the Delta method (Dorfman, 1938), the variances of $\boldsymbol{\beta}$ can be approximated by

$$\mathbb{V}(\hat{\beta}_j) = \frac{2}{\sum_{i=1}^{N} \left(\frac{2(\hat{z}^2_{i|i,j} + V_{i|i,j})}{\hat{\mu}_{ij}} - 1 \right)},$$

with $\hat{\mu}_{ij}$ and $\hat{z}_{i|i,j}$ denoting the j-th elements of the vectors $\hat{\boldsymbol{\mu}}_i$ and $\hat{\mathbf{z}}_{i|i}$, respectively, and $V_{i|i,j}$ the diagonal entry of $V_{i|i}$, all evaluated at the optimal point estimates of the parameters.

3.5. Model selection

In empirical settings, one may be interested in comparing different quasi-reaction systems of possibly varying complexity. Similarly to the derivation of the standard errors, a modified version of standard model selection criteria is considered, where the log-likelihood is replaced by the Q-function, which is instead a direct output of the EM algorithm (Ibrahim et al., 2008). In particular, the optimal model is taken as the one that minimizes the information criterion

$$IC = -2Q(\hat{\boldsymbol{\beta}}, \hat{\Sigma} | \hat{\boldsymbol{\beta}}, \hat{\Sigma}) + P(\hat{\boldsymbol{\beta}}), \tag{16}$$

with $Q(\hat{\boldsymbol{\beta}}, \hat{\Sigma} | \hat{\boldsymbol{\beta}}, \hat{\Sigma})$ the Q-function (11) evaluated upon convergence of the EM algorithm and $P(\hat{\boldsymbol{\beta}})$ a term penalizing model complexity. In the real application, the Bayesian Information Criterion (BIC) will be considered, where $P(\hat{\boldsymbol{\beta}}) = r \log(N)$, with r the number of reaction rates in the model and N the number of time intervals.

4. Simulation study

In this section, a simulation study is provided to evaluate the performance of the proposed method under different settings and to highlight those where it is particularly advantageous compared to the existing LLA approaches. For the simulation, a dynamic system with a low number of particles ($p = 4$) and reactions ($r = 6$) is considered, in order to mimic a setting that is common in many applications, such as the cell differentiation process studied by Pellin et al. (2023). In the specifics, the 6 reactions contain one

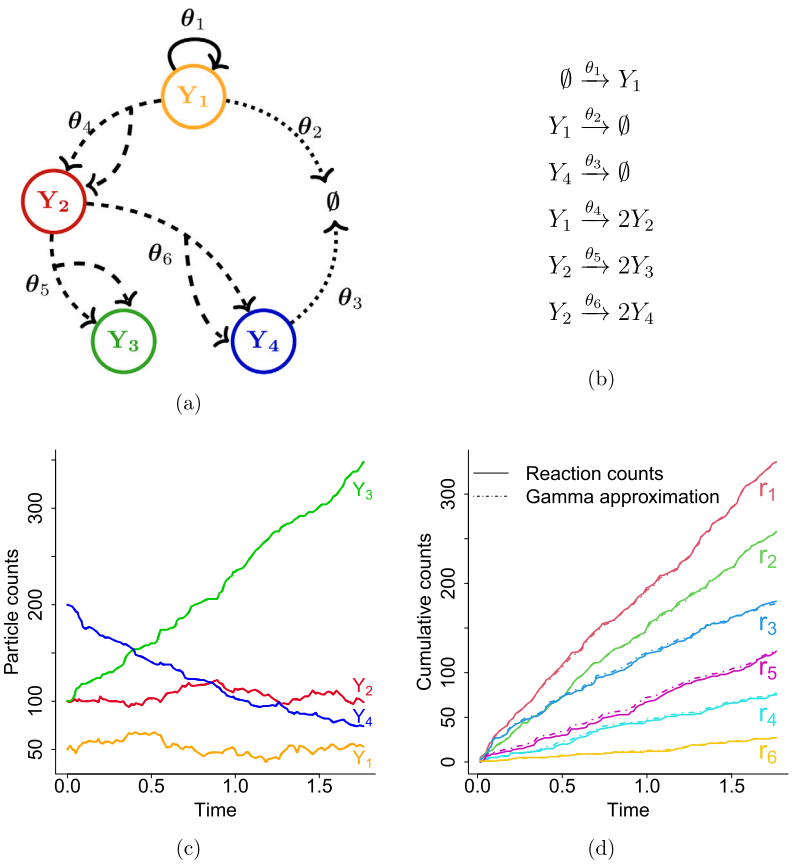

Fig. 2. Specifications of the cell differentiation process used in the simulation study. (a) Structure of the process with $p = 4$ particles. Each substrate is represented by a coloured node, whereas birth, death and differentiation reactions are denoted by full, dotted and dashed edges, respectively. (b) The corresponding quasi-reaction system. (c) An example of trajectories generated by means of a Gillespie algorithm. (d) Cumulative counts of the reactions increments $\Delta\mathbf{N}_i$ (full lines) and of their Gamma approximations $\mathbf{X}_i$ (dotted lines).

duplication, two death and three differentiation reactions. Fig. 2a provides a graphical representation of the system, while Fig. 2b reports the 6 reactions. These correspond to the net effect matrix

$$V = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 & 2 \end{pmatrix}.$$

The parameters are set to

$$\boldsymbol{\beta}_{true} = \log(\boldsymbol{\theta}_{true}) = (5.30, 1.10, -0.11, -0.22, -0.22, -1.61)^T,$$

and no measurement error is considered ($\Sigma = 0$). Starting with initial particle counts set to $\mathbf{y}_0 = (50, 100, 100, 200)$, a Gillespie algorithm is used to generate the stochastic process over time (Gillespie, 1977). Fig. 2c reports one run of the algorithm, while Fig. 2d shows how the reaction counts $\mathbf{N}_i$ are close to those based on the Gamma approximation $\mathbf{X}_i$ of $\Delta\mathbf{N}_i$, with the Gamma distribution defined using the true parameters $\boldsymbol{\beta}$.

Improvement over local linear approximation approach. In the first simulation study, the performance of the algorithm is compared against the existing LLA approach in terms of parameter estimation. Given the motivation behind the proposed methodology, one would expect an improvement when the interval between consecutive observations is particularly small, as this generates a strong temporal correlation among the particle counts. Moreover, one would expect the difference to be more pronounced at low sample sizes, i.e., a small number of time points, as this will make statistical inference more challenging in general and may amplify the effect of strong temporal correlations.

In order to test these hypotheses, a subset of the trajectories generated by the Gillespie algorithm is considered. In particular, observations are retained at every 10, 15, 20, 25 and 30 time points out of the originally sampled trajectories. These are referred to as *jumps*. The larger the jump is, the larger the gap between consecutive time points where the process is observed. This will generally

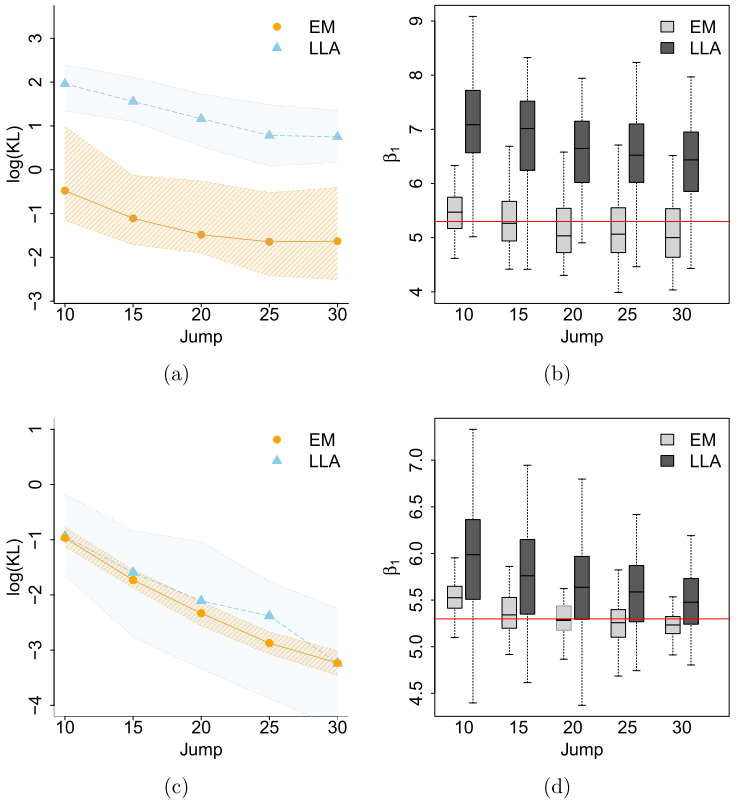

Fig. 3. Comparison between EM and LLA methods. On the left, the KL measure, in the log scale, shows that parameter estimation with the EM algorithm is closer to the true parameters than with the LLA approach. On the right, the plots show how, for one of the parameters (β_1), estimates are more accurate with the EM than with the LLA approach. The true value is indicated by the horizontal red line. All plots show how the effects are more pronounced with $N = 5$ (3a, 3b) than with $N = 50$ (3c, 3d) time intervals. The boxplots are obtained across 100 simulations.

translate into a large number of reactions that may have occurred between one time point and the next, although this will depend also on the dynamics of the process at the specific time interval. In order to test also the effect of sample size, in each of the settings, two scenarios are considered: one where the first $N = 5$ time intervals are considered and a second one where the first $N = 50$ time intervals are considered, generated as above.

Parameter estimation is conducted for each of the datasets using LLA and the proposed EM algorithm. LLA uses the moments in (4) as the basis of a generalised least-squares approach given the particle count data $\mathbf{Y}$. For the EM algorithm described in Algorithm 2, the LLA solution is set as starting value for $\boldsymbol{\beta}$ ($\boldsymbol{\beta}_{ini}$), $\Sigma = 0$, the net effect matrix is set to V defined as above, the tolerance for convergence to $tol = 0.002$ and the maximum number of iterations to $maxit = 300$. Upon convergence, the quality of the estimation is evaluated by calculating the Kullback-Leibler divergence between the estimated and the true parameters. In particular, this is defined by

$$KL(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_{true}) = \mathbb{E}_{\mathbf{y}^+}[\log p(\mathbf{y}^+|\boldsymbol{\beta}_{true}) - \log p(\mathbf{y}^+|\hat{\boldsymbol{\beta}})],$$

where $\mathbf{y}^+$ indicates an additional dataset with the same characteristics as the one used for inference, and generated from the same underlying process defined by $\boldsymbol{\beta}_{true}$. The lower this value is, the closer the inferred process is to the true one.

Fig. 3 reports the results in the form of boxplots across the 100 simulations for each of the settings. The results show how parameter estimation with the proposed EM algorithm is better than with the existing LLA approach, both in terms of the KL divergence (left panel) and estimation of one of the parameters (β_1, right panel). All plots show how the effects are more pronounced with small sample sizes ($N = 5$, Fig. 3a and 3b) than with larger sample sizes ($N = 50$, Figs. 3c and 3d). Finally, Fig. 3d in particular shows how the two approaches tend to converge to a similar performance for larger time intervals (i.e., a large *jump*). This is to be expected, since temporal correlation will become less strong the larger the time interval. At the same time, the reconstruction of the reactions that have taken place within that time interval will also be less accurate. However, Fig. 3d shows how, even in this case, estimation from the EM algorithm appears to be less biased and more accurate than with the LLA approach.

Computational cost in terms of number of time points, reactions, particles. A second simulation study explores how the computational cost of the algorithm varies with respect to the number of time points (N), the number of reactions (r) and the number of particles (p).

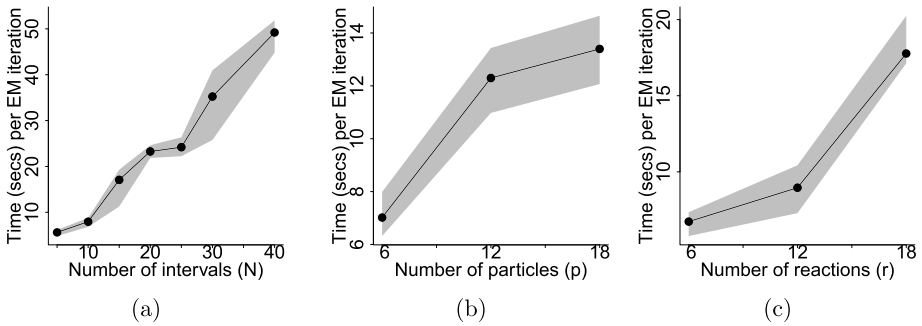

Fig. 4. Computational cost of EM algorithm. Average computational time (in seconds) of one iteration of the EM algorithm in terms of (a) the number of time intervals (N), (b) the number of particles (p), (c) the number of reactions (r). Median, first and third quartiles are shown across 100 simulations.

The results are shown in Fig. 4. The first scenario (Fig. 4a) considers the same generative process as before, fixing $jump = 30$ and letting the number of time intervals vary in $N = 5, 10, 15, 20, 25, 30, 40$. The plot shows how the average computational time of the EM algorithm is approximately linear in N.

The second scenario (Fig. 4b) evaluates how the computational time varies with respect to the number of particles p. The three systems in Table B.1 of Appendix B are considered, with $jump = 40$ and $N = 10$. The systems are characterized by the same number of reactions as before ($r = 6$), but an increasing number of particles, namely $p = 6, 12, 18$, respectively. Fig. 4b does not show the cubic dependence in p, that was anticipated. Given that the computational time is the combined time from the E- and the M-steps, this suggests a much slower M-step.

Finally, the third scenario (Fig. 4c) evaluates how the computational time varies with respect to the number of reactions r. As before the parameters are set to $jump = 40$ and $N = 10$, but the three systems in Table B.2 of Appendix B are now considered. These are characterized by the same number of particles as before ($p = 6$), but an increasing number of reactions, namely $r = 6, 12, 18$, respectively. The plot shows a super-linear dependence in r.

5. Illustration on Italian COVID-19 data

This section provides a real data illustration focusing on the COVID-19 pandemic. Worldwide, more than 700 million infections and almost 7 million deaths were recorded as of August 16, 2023 (WHO, 2020). As a result, significant efforts have been made in order to understand the phenomenon and find strategies to control the spreading of the disease. Italy has been one of the countries of interest during the pandemic, being the first European country to experience a significant outbreak of the disease (Liao et al., 2020). The first case was confirmed on 31 January 2020. Since then, for more than two years, data were collected daily. The sufficiently close interval between observations is ideal for the application of the method, as it generates strong temporal correlations. The following analysis focusses on daily data within three specific time intervals, characterized by three different levels of contagion:

- **Phase 1**: 9^{th} *March* - 4^{th} *May,* 2020. Strong restrictions on travel throughout the country, banning all forms of gathering in private and public places (Conte, 2020c);
- **Phase 2**: 4^{th} *May* - 7^{th} *October,* 2020. Containment measures were relaxed, allowing the travelling for visits to relatives (within a region) and the restart of several production activities (Conte, 2020b);
- **Phase 3**: 8^{th} *October,* 2020 - 14^{th} *January,* 2021. Wearing of masks became compulsory both outdoor and indoor, and assemblages were restricted (Conte, 2020a).

Within each of the three phases and using data from all 21 Italian regions, the proposed EM algorithm is used to fit the parameters of the following two dynamic systems:

Model A	**Model B**
$I_k \xrightarrow{\theta_{1k}} 2I_k$	$I_k \xrightarrow{\theta_{1k}} 2I_k$
$I_k \xrightarrow{\theta_2} R_k$	$I_k \xrightarrow{\theta_{2k}} R_k$
$I_k \xrightarrow{\theta_3} D_k$	$I_k \xrightarrow{\theta_{3k}} D_k$

The systems correspond to simple SIR compartmental models, where I is the number of infectious individuals, R the number of recovered individuals and D the number of deceased individuals (Simon, 2020). In particular, the first reaction models the creation of one infectious individual once a susceptible individual meets an infectious one. Note how the number of susceptible individuals S has been omitted, as it is almost constant throughout the observation period. The second and third reactions correspond to the cases of an infectious individual recovering and dying, respectively. The index k denotes the region. Thus, model B is characterized by

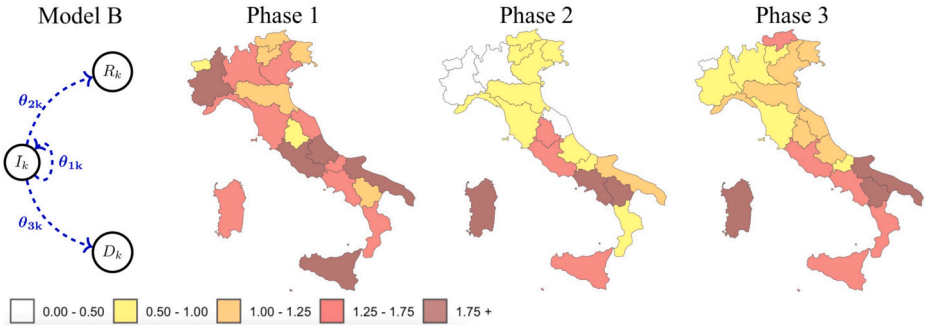

Fig. 5. Visualization of the estimated R_0 values in Italy. On the left, the system of kinetic reactions for the k-th region. On the right, a visualization of the estimated values of the basic reproductive numbers for each Italian region, in the 3 phases of interest, categorised according to the colour coding used by the Italian government. An additional shade of dark red has been added to highlight the regions with the most critical R_0 values.

region-specific rates for all three reactions. This results in a system with $p = 63$ particles and $r = 63$ rates. On the other hand, model A hypothesizes a simpler model where the recovery and death rates are assumed to be the same across Italy, under an assumption that these depend primarily on the specifics of the virus and are not as affected by the level of contagion in the population.

The proposed EM algorithm is used, with a tolerance $tol = 10^{-5}$, for the estimation of the reaction rates $\boldsymbol{\theta}$ and of the noise Σ. Using Equation (16) with a BIC penalty term, model A and model B result in $9.66 \cdot 10^5$ and $2.64 \cdot 10^5$ BIC values, respectively. This leads to the choice of the more complex model B, with region-specific recovery and death reaction rates. Since the BIC tends to select sparser models compared to other model selection criteria, this suggests that other model selection criteria would have led to the same conclusion. As for the parameter estimates, the error variances were generally far from zero, suggesting that some of the recorded cases were subject to a measurement error.

Fig. 5 visualises the results in terms of the basic reproduction number R_0, which is the number of new infections that each infected individual produces on average (Wood and Wit, 2021). This can be estimated from the fitted models by calculating $\theta_{1k}/(\theta_{2k} + \theta_{3k})$ In Fig. 5, these values are categorized according to the colour coding used by the Italian government to evaluate the severity of the disease spread. In particular, values below the boundary of $R_0 = 1$ are associated to a long-term decrease of the epidemic, while values above 1 indicate a long-term increase of the epidemic, with larger values (darker colours) associated to progressively more severe scenarios. The estimated R_0 values are in line with those from other studies (Giordano et al., 2020; Remuzzi and Remuzzi, 2020; Mingliang et al., 2022). The results in Fig. 5 show how during Phase 2 the infection was limited, as a consequence of the containment measures implemented in Phase 1. During Phase 3, a revival of the disease spread is observed, in particular in the southern regions of Italy. The standard errors of R_0, calculated using the Delta method from the standard errors of the estimated reaction rates $\boldsymbol{\theta}$ given by (15), show significant differences in R_0 values between two consecutive phases at a 95% significance level, with the only exception of the Autonomous Province of Bolzano between phase 1 and phase 2, and Molise and Sicily between phase 2 and phase 3. Moreover, although not assumed by the model, the R_0 estimates show some geographical clustering, which is to be expected given the movements of individuals between neighbouring regions.

6. Conclusions

A novel procedure for the statistical inference of quasi-reaction systems has been proposed. Local linear approximation methods tend to perform poorly when the system is observed at fine time intervals. This is due to numerical instability caused by strong correlations in the observations from one time point to the next. The proposed method focuses instead on reconstructing the underlying process of latent reactions. To this end, a latent event history model of the observed count process driven by a latent process of reactions is developed. A computationally efficient EM algorithm for parameter estimation is proposed, incorporating an extended Kalman filtering procedure for predicting the latent states. A simulation study demonstrates how the proposed method performs better than the existing LLA approach, particularly when the time intervals between consecutive observations are small and the number of time points is low.

The method is illustrated by an application on the Italian Covid 19 data during the critical phase of the pandemic, between March 2020 and January 2021. The basic reproduction number R_0 of the 21 Italian regions estimated by the method in three consecutive phases of the pandemic shows higher values at the beginning and at the end of the time period. This is to be expected given the evolution of the disease and the societal restrictions that were imposed by the Italian government during this period.

The simple epidemic model considered is clearly a simplification of the pandemic process. The model does not consider inter-regional infections, nor effects from outside Italy or heterogeneity in the population. Most likely, ignoring this type of effects means that R_0 has been over-estimated by the models (Gomes et al., 2022). Future work will consider applying the same methodology to fit more complex models, such as the compartmental model of Wood and Wit (2021), which includes hospital infections and other types of interactions.

Data availability

The data used in this paper are available from the Dipartimento della Protezione Civile and the Istituto Nazionale di Statistica (ISTAT) via the webpage of the department (http://dati.istat.it/Index.aspx?QueryId=18460) and a Github repository (https://github.com/pcm-dpc/COVID-19).

Appendix A. Kalman filtering (E-step)

This section discusses the extended Kalman filtering procedure that was developed for the evaluation of $\mathbb{E}[\mathbf{Z}_i|\mathbf{Y}_{0:i}]$, $\mathbb{V}[\mathbf{Z}_i|\mathbf{Y}_{0:i}]$, $\mathbb{E}[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}]$, and $\mathbb{V}[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}]$.

Prediction step. The prediction step calculates the first and second moments of $\mathbf{Z}_i$ conditional on $\mathbf{Y}_{0:i-1}$. According to model (7), these are in fact the conditional moments of $\mathbf{Z}_i$ given the state of the system at the previous time point, $\mathbf{Y}_{i-1}$. Thus

$$\begin{aligned}\hat{\mathbf{z}}_{i|i-1} &= \mathbb{E}\left[\mathbf{Z}_i \mid \mathbf{Y}_{i-1}\right] = \boldsymbol{\mu}_i,\\ V_{i|i-1} &= \mathbb{V}\left[\mathbf{Z}_i \mid \mathbf{Y}_{i-1}\right] = \mathrm{diag}(\boldsymbol{\mu}_i).\end{aligned} \tag{A.1}$$

Update step. Following from the prediction step, the update step refines these predictions by comparing them to the observed values at time i. In particular, the conditional distribution of $\mathbf{Z}_i$ is updated from the past with the information coming from $\mathbf{Y}_i$ by first deriving the joint distribution of $\mathbf{Y}_i$ and $\mathbf{Z}_i$ conditional on $\mathbf{Y}_{0:i-1}$. According to model (7), this is a multivariate Gaussian distribution, which can be written generically as

$$\begin{bmatrix}\mathbf{Z}_i\\ \mathbf{Y}_i\end{bmatrix}\Big|\mathbf{Y}_{0:i-1} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{bmatrix}, \begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}\right). \tag{A.2}$$

From the prediction step (A.1), $\mathbf{m}_1$ and S_{11} are already known. As regards to the other elements of the mean and covariance,

$$\begin{aligned}\mathbf{m}_2 &= \mathbb{E}[\mathbf{Y}_i|\mathbf{Y}_{0:i-1}] = \mathbb{E}[\mathbf{Y}_{i-1} + VG(\mathbf{Z}_i) + \boldsymbol{\psi}_i|\mathbf{Y}_{0:i-1}] = \mathbf{Y}_{i-1} + V\mathbb{E}[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}],\\ S_{12} &= \mathbb{C}ov[\mathbf{Z}_i, \mathbf{Y}_i|\mathbf{Y}_{0:i-1}] = \mathbb{C}ov[\mathbf{Z}_i, \mathbf{Y}_{i-1} + VG(\mathbf{Z}_i) + \boldsymbol{\psi}_i|\mathbf{Y}_{0:i-1}] = V\mathbb{C}ov[\mathbf{Z}_i, G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}]V^T,\\ S_{22} &= \mathbb{V}[\mathbf{Y}_i|\mathbf{Y}_{0:i-1}] = \mathbb{V}[\mathbf{Y}_{i-1} + VG(\mathbf{Z}_i) + \boldsymbol{\psi}_i|\mathbf{Y}_{0:i-1}] = V\mathbb{V}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}\right]V^T + \Sigma.\end{aligned} \tag{A.3}$$

In order to calculate the first and second moments of $G(\mathbf{Z}_i)$, the non-linear function G is approximated with its Taylor expansion of order 2 centred at $\hat{\mathbf{z}}_{i|i-1}$, i.e.,

$$G(\mathbf{Z}_i) \approx \mathbf{g}_{i|i-1} + J_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1}) + \frac{1}{2}\mathrm{diag}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})H_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1}).$$

The first term of the expansion is the deterministic vector of size r

$$\mathbf{g}_{i|i-1} = G(\hat{\mathbf{z}}_{i|i-1}).$$

The other terms have a simplified form due to the fact that the j-th element of the function G is a function only of the j-th element of $\mathbf{Z}_i$. Thus, the $r \times r$ matrix of first derivatives is a diagonal matrix, with (j, j) element given by

$$\left[J_{i|i-1}\right]_{jj} = \left[\frac{\partial G(\mathbf{Z})}{\partial \mathbf{Z}}\Big|_{\hat{\mathbf{z}}_{i|i}}\right]_{jj} = \frac{\partial G_j}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}} = \frac{\partial (F_{ij}^{-1}(\Phi_{ij}(Z_{ij})))}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}} = \left(\frac{\partial F_{ij}}{\partial x_{ij}}\Big|_{G(\hat{z}_{ij|i-1})}\right)^{-1}\frac{\partial \Phi_{ij}}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}},$$

where, using the functional form of the Normal and Gamma CDFs

$$\frac{\partial \Phi_{ij}}{\partial z_{ij}}(z) = \frac{e^{-\frac{(z-\mathbb{E}[Z_{ij}])^2}{2\mathbb{V}[Z_{ij}]}}}{\sqrt{2\pi\mathbb{V}[Z_{ij}]}}, \qquad \frac{\partial F_{ij}}{\partial x_{ij}}(x) = \frac{e^{-x}x^{\mathbb{E}[X_{ij}]-1}}{\Gamma(\mathbb{E}[X_{ij}])}\mathbb{1}_{[x>0]}.$$

Similarly, the $r \times r \times r$ Hessian matrix, can be written as an $r \times r$ diagonal matrix with second derivatives on the diagonal, namely

$$\begin{aligned}\left[H_{i|i-1}\right]_{jj} &= \frac{\partial}{\partial z_{ij}}\left[\left(\frac{\partial F_{ij}}{\partial x_{ij}}\Big|_{G(\hat{z}_{ij|i-1})}\right)^{-1}\frac{\partial \Phi_{ij}}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}}\right]\\ &= \frac{-\frac{\partial^2 F_{ij}}{(\partial x_{ij})^2}\Big|_{G(\hat{z}_{ij|i-1})}\frac{\partial G_j}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}}\frac{\partial \Phi_{ij}}{\partial z_{ij}}\Big|_{\hat{z}_{ij|i-1}} + \frac{\partial F_{ij}}{\partial x_{ij}}\Big|_{G(\hat{z}_{ij|i-1})}\frac{\partial^2 \Phi_{ij}}{(\partial z_{ij})^2}\Big|_{\hat{z}_{ij|i-1}}}{\left(\frac{\partial F_{ij}}{\partial x_{ij}}\Big|_{G(\hat{z}_{ij|i-1})}\right)^2},\end{aligned}$$

where

$$\frac{\partial^2 \Phi_{ij}}{\partial z_{ij}^2}(z) = \frac{(z - \mathbb{E}[Z_{ij}])e^{-\frac{(z-\mathbb{E}[Z_{ij}])^2}{2\mathbb{V}[Z_{ij}]}}}{\sqrt{2\pi}\mathbb{V}[Z_{ij}]^{3/2}}, \quad \frac{\partial^2 F_{ij}}{\partial x_{ij}^2}(x) = \frac{e^{-x}x^{\mathbb{E}[X_{ij}]-2}(\mathbb{E}[X_{ij}] - x - 1)}{\Gamma(\mathbb{E}[X_{ij}])}\mathbb{1}_{[x>0]}.$$

Going back to (A.3), the Taylor approximation can now be used to calculate the required conditional expectations. In particular,

$$\begin{aligned}
\mathbb{E}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}\right] \approx & \mathbb{E}\left[\mathbf{g}_{i|i-1} + J_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1}) + \frac{1}{2}\text{diag}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})H_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] \\
= & \mathbf{g}_{i|i-1} + J_{i|i-1}\mathbb{E}\left[\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1}|\mathbf{Y}_{0:i-1}\right] + \frac{1}{2}\mathbb{E}\left[\text{diag}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})H_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] \\
= & \mathbf{g}_{i|i-1} + \frac{1}{2}\text{vect}(V_{i|i-1}H_{i|i-1}),
\end{aligned}$$

$$\begin{aligned}
\mathbb{C}ov[\mathbf{Z}_i, G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}] \approx & \mathbb{C}ov\left[\mathbf{Z}_i, \mathbf{g}_{i|i-1}|\mathbf{Y}_{0:i-1}\right] + \mathbb{C}ov\left[\mathbf{Z}_i, J_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] \\
& + \mathbb{C}ov\left[\mathbf{Z}_i, \frac{1}{2}\text{diag}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})H_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] \\
= & \mathbb{C}ov\left[\mathbf{Z}_i, J_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] = \mathbb{V}ar[\mathbf{Z}_i|\mathbf{Y}_{0:i-1}]J_{i|i-1} = V_{i|i-1}J_{i|i-1},
\end{aligned}$$

$$\mathbb{V}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i-1}\right] \approx \mathbb{V}\left[\mathbf{g}_{i|i-1} + J_{i|i-1}(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1})|\mathbf{Y}_{0:i-1}\right] = J_{i|i-1}\mathbb{V}\left[\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i-1}|\mathbf{Y}_{0:i-1}\right]J_{i|i-1}^T = J_{i|i-1}V_{i|i-1}J_{i|i-1}^T.$$

Finally, plugging these expressions into (A.1), it follows that

$$\begin{aligned}
\mathbf{m}_2 &\approx \mathbf{Y}_{i-1} + V\left[\mathbf{g}_{i|i-1} + \frac{1}{2}\text{vect}(V_{i|i-1}H_{i|i-1})\right], \\
S_{22} &\approx V[J_{i|i-1}V_{i|i-1}J_{i|i-1}^T]V^T + \Sigma, \\
S_{12} &\approx VV_{i|i-1}J_{i|i-1},
\end{aligned}$$

which, together with $\mathbf{m}_1$ and S_{11} derived previously, define the joint distribution (A.2) of $\mathbf{Z}_i$ and $\mathbf{Y}_i$ conditional on $\mathbf{Y}_{i-1}$. From this, using the formulae for the conditional distributions from a jointly Gaussian random vector, it follows that $\mathbf{Z}_i$, conditional on $\mathbf{Y}_{0:i}$, has a multivariate Gaussian distribution, with mean and covariance given, respectively, by

$$\begin{aligned}
\hat{\mathbf{z}}_{i|i} &= \mathbb{E}\left[\mathbf{Z}_i \mid \mathbf{Y}_{0:i}\right] = \hat{\mathbf{z}}_{i|i-1} + K_i\left[\mathbf{Y}_i - \mathbf{Y}_{i-1} - V\left(\mathbf{g}_{i|i-1} + \frac{1}{2}\text{vect}(V_{i|i-1}H_{i|i-1})\right)\right], \\
V_{i|i} &= \mathbb{E}\left[\left(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i}\right)\left(\mathbf{Z}_i - \hat{\mathbf{z}}_{i|i}\right)^T \mid \mathbf{Y}_{0:i}\right] = \left(\mathbb{I}_r - K_iVJ_{i|i-1}\right)V_{i|i-1},
\end{aligned}$$

where

$$K_i = (VV_{i|i-1}J_{i|i-1})^T(VJ_{i|i-1}V_{i|i-1}J_{i|i-1}^TV^T + \Sigma)^{-1}.$$

Note how the update step refines the conditional expectation found in the prediction step in proportion to the difference between the actual and estimated observations, i.e., the prediction error. Moreover, this is directly proportional to the magnitude of the Kalman *gain matrix* K_i, which captures the linear relationship between the noise and the variance of the latent variable (Kim and Bang, 2018).

Similarly to the earlier derivations,

$$\begin{aligned}
\mathbb{E}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}\right] &\approx \mathbf{g}_{i|i} + \frac{1}{2}\text{vect}(V_{i|i}H_{i|i}), \\
\mathbb{V}\left[G(\mathbf{Z}_i)|\mathbf{Y}_{0:i}\right] &\approx J_{i|i}V_{i|i}J_{i|i}^T,
\end{aligned}$$

with

$$\mathbf{g}_{i|i} = G|_{\hat{\mathbf{z}}_{i|i}}, \qquad J_{i|i} = \frac{\partial G(\mathbf{Z})}{\partial \mathbf{Z}}|_{\hat{\mathbf{z}}_{i|i}}, \qquad H_{i|i} = \frac{\partial^2 G(\mathbf{Z})}{\partial \mathbf{Z}^2}|_{\hat{\mathbf{z}}_{i|i}}.$$

Appendix B. Dynamic systems used for the simulation study

This section reports the systems of reactions that were used in Section 4 for evaluating the computational complexity of the algorithm with respect to the number of particles p (Table B.1) and the number of reactions r (Table B.2).

Table B.1
Three dynamic systems with $r = 6$ reactions, and an increasing number of particles ($p = 6, 12, 18$).

$p = 6$	$p = 12$	$p = 18$
$Y_1 \xrightarrow{\theta_1} Y_2$	$Y_1 + Y_7 \xrightarrow{\theta_1} Y_2 + Y_8$	$Y_1 + Y_7 + Y_{13} \xrightarrow{\theta_1} Y_2 + Y_8 + Y_{14}$
$Y_1 \xrightarrow{\theta_2} Y_3$	$Y_1 + Y_7 \xrightarrow{\theta_2} Y_3 + Y_9$	$Y_1 + Y_7 + Y_{13} \xrightarrow{\theta_2} Y_3 + Y_9 + Y_{15}$
$Y_2 \xrightarrow{\theta_3} Y_4$	$Y_2 + Y_8 \xrightarrow{\theta_3} Y_4 + Y_{10}$	$Y_2 + Y_8 + Y_{14} \xrightarrow{\theta_3} Y_4 + Y_{10} + Y_{16}$
$Y_1 \xrightarrow{\theta_4} Y_4$	$Y_1 + Y_7 \xrightarrow{\theta_4} Y_4 + Y_{10}$	$Y_1 + Y_7 + Y_{13} \xrightarrow{\theta_4} Y_4 + Y_{10} + Y_{16}$
$Y_4 \xrightarrow{\theta_5} Y_6$	$Y_4 + Y_{10} \xrightarrow{\theta_5} Y_6 + Y_{12}$	$Y_4 + Y_{10} + Y_{16} \xrightarrow{\theta_5} Y_6 + Y_{12} + Y_{18}$
$Y_5 \xrightarrow{\theta_6} Y_5$	$Y_5 + Y_{11} \xrightarrow{\theta_6} Y_5 + Y_{11}$	$Y_5 + Y_{11} + Y_{17} \xrightarrow{\theta_6} Y_5 + Y_{11} + Y_{17}$

Table B.2
Three dynamic systems with $p = 6$ particles, and an increasing number of reactions ($r = 6, 12, 18$).

$r = 6$	$r = 12$	$r = 18$
$\mathcal{R}_6 :$	$\mathcal{R}_{12} : \mathcal{R}_6 \cup$	$\mathcal{R}_{18} : \mathcal{R}_{12} \cup$
$Y_2 \xrightarrow{\theta_1} Y_1 + Y_3$	$Y_8 \xrightarrow{\theta_1} Y_7 + Y_9$	$Y_{14} \xrightarrow{\theta_1} Y_{13} + Y_{15}$
$Y_3 \xrightarrow{\theta_2} Y_2 + Y_4$	$Y_9 \xrightarrow{\theta_2} Y_8 + Y_{10}$	$Y_{15} \xrightarrow{\theta_2} Y_{14} + Y_{16}$
$Y_4 \xrightarrow{\theta_3} Y_3 + Y_5$	$Y_{10} \xrightarrow{\theta_3} Y_9 + Y_{11}$	$Y_{16} \xrightarrow{\theta_3} Y_{15} + Y_{17}$
$Y_5 \xrightarrow{\theta_4} Y_4 + Y_6$	$Y_{11} \xrightarrow{\theta_4} Y_{10} + Y_{12}$	$Y_{17} \xrightarrow{\theta_4} Y_{16} + Y_{18}$
$Y_5 \xrightarrow{\theta_5} Y_6$	$Y_{12} \xrightarrow{\theta_5} Y_1 1$	$Y_{18} \xrightarrow{\theta_5} Y_{17}$
$Y_6 \xrightarrow{\theta_6} Y_1$	$Y_7 \xrightarrow{\theta_6} Y_1 2$	$Y_{13} \xrightarrow{\theta_6} Y_{18}$

References

Anderson, B., Moore, J., 2012. Optimal Filtering. Courier Corporation.

Bar-Shalom, Y., Li, X., Kirubarajan, T., 2001. Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. John Wiley & Sons.

Box-Steffensmeier, J., Jones, B., 2004. Event History Modeling: A Guide for Social Scientists. Cambridge University Press.

Britton, T., Pardoux, E., 2019. Stochastic Epidemic Models with Inference. Lecture Notes in Mathematics. Springer.

Capistrán, M., Christen, J., Velasco-Hernández, J., 2012. Towards uncertainty quantification and inference in the stochastic SIR epidemic model. Math. Biosci. 240 (2), 250–259.

Conte, G., 2020a. Dpcm 1 gennaio 2021. https://www.gazzettaufficiale.it/eli/id/2021/01/05/21G00001/sg.

Conte, G., 2020b. Dpcm 26 aprile 2020. https://www.gazzettaufficiale.it/eli/id/2020/04/27/20A02352/sg.

Conte, G., 2020c. Dpcm 9 marzo 2020. https://web.archive.org/web/20201001105941/http://www.governo.it/it/articolo/firmato-il-dpcm-9-marzo-2020/14276.

Cooray-Wijesinha, M., Khuri, A., 1987. The sequential generation of multiresponse d-optimal designs when the variance-covariance matrix is not known. Commun. Stat., Simul. Comput. 16 (1), 239–259.

Craigmile, P., Herbei, R., Liu, G., Schneider, G., 2023. Statistical inference for stochastic differential equations. WIREs: Comput. Stat. 15 (2), e1585.

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

Dorfman, R., 1938. A note on the delta-method for finding variance formulae. Biom. Bull. 1 (92), 129–137.

Engl, H., Hanke, M., Neubauer, A., 1996. Regularization of Inverse Problems. Mathematics and Its Applications. Springer, Netherlands.

Fedorov, V., 2013. Theory of Optimal Experiments. Elsevier.

Ghahramani, Z., Hinton, G., 1996. Parameter Estimation for Linear Dynamical Systems.

Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., Colaneri, M., 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. 26 (6), 855–860.

Gomes, M., Ferreira, M., Corder, R., King, J., Souto-Maior, C., Penha-Gonçalves, C., Gonçalves, G., Chikina, M., Pegden, W., Aguas, R., 2022. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. J. Theor. Biol. 540, 111063.

Hatzis, C., Larntz, K., 1992. Optimal design in nonlinear multiresponse estimation: Poisson model for filter feeding. Biometrics, 1235–1248.

Ibrahim, J., Zhu, H., Tang, N., 2008. Model selection criteria for missing-data problems using the EM algorithm. J. Am. Stat. Assoc. 103 (484), 1648–1658.

Kalman, R., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45.

Kim, Y., Bang, H., 2018. Introduction to Kalman filter and its applications. Introd. Implement. Kalman Filter 1, 1–16.

Komorowski, M., Finkenstadt, B., Rand, D., Gillespie, C., Wilkinson, D., 2011. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. Proc. Natl. Acad. Sci. 108 (21), 8645–8650.

Liao, Z., Lan, P., Liao, Z., Zhang, Y., Liu, S., 2020. TW-SIR: time-window based SIR for COVID-19 forecasts. Sci. Rep. 10 (1), 22454.

McQuarrie, D., 1967. Stochastic approach to chemical kinetics. J. Appl. Probab. 4 (3), 413–478.

Mingliang, Z., Simos, T.E., Tsitouras, C., 2022. R0 estimation for COVID-19 pandemic through exponential fit. Math. Methods Appl. Sci. 45 (3), 1632–1639.

Oakes, D., 1999. Direct calculation of the information matrix via the EM algorithm. J. R. Stat. Soc., Ser. B, Stat. Methodol. 61 (2), 479–482.

Pellin, D., Biasco, L., Aiuti, A., Di Serio, C., Wit, E., 2019. Penalized inference of the hematopoietic cell differentiation network via high-dimensional clonal tracking. Appl. Netw. Sci. 4 (115).

Pellin, D., Biasco, L., Scala, S., di Serio, C., Wit, E., 2023. Tracking hematopoietic stem cell evolution in a Wiskott-Aldrich clinical trial. Ann. Appl. Stat. 17.

Remuzzi, A., Remuzzi, G., 2020. COVID-19 and Italy: what next? Lancet 395 (10231), 1225–1228.

Shoji, I., Ozaki, T., 1998. Estimation for nonlinear stochastic differential equations by a local linearization method. Stoch. Anal. Appl. 16 (4), 733–752.

Shumway, R., Stoffer, D., 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3 (4), 253–264.

Simon, C.M., 2020. The SIR dynamic model of infectious disease transmission and its analogy with chemical kinetics. PeerJ Phys. Chem. 2, e14.

Wan, E.A., Van Der Merwe, R., 2000. The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373), pp. 153–158.
WHO, 2020. WHO COVID-19 Dashboard. World Health Organization, Geneva.
Wilkinson, D.J., 2018. Stochastic Modelling for Systems Biology, third edition. Chapman and Hall/CRC.
Wood, S., Wit, E., 2021. Was $R < 1$ before the English lockdowns? On modelling mechanistic detail, causality and inference about Covid-19. PLoS ONE 16 (9), e0257455.
Zia, A., Kirubarajan, T., Reilly, J., Yee, D., Punithakumar, K., Shirani, S., 2008. An EM algorithm for nonlinear state estimation with model uncertainties. IEEE Trans. Signal Process. 56 (3), 921–936.