



# Identifiability analysis for models of the translation kinetics after mRNA transfection

Susanne Pieschner<sup>1,2</sup> · Jan Hasenauer<sup>1,2,3</sup> · Christiane Fuchs<sup>1,2,4</sup>

Received: 17 May 2021 / Revised: 25 March 2022 / Accepted: 26 March 2022 /

Published online: 17 May 2022

© The Author(s) 2022, corrected publication 2022

## Abstract

Mechanistic models are a powerful tool to gain insights into biological processes. The parameters of such models, e.g. kinetic rate constants, usually cannot be measured directly but need to be inferred from experimental data. In this article, we study dynamical models of the translation kinetics after mRNA transfection and analyze their parameter identifiability. That is, whether parameters can be uniquely determined from perfect or realistic data in theory and practice. Previous studies have considered ordinary differential equation (ODE) models of the process, and here we formulate a stochastic differential equation (SDE) model. For both model types, we consider structural identifiability based on the model equations and practical identifiability based on simulated as well as experimental data and find that the SDE model provides better parameter identifiability than the ODE model. Moreover, our analysis shows that even for those parameters of the ODE model that are considered to be identifiable, the obtained estimates are sometimes unreliable. Overall, our study clearly demonstrates the relevance of considering different modeling approaches and that stochastic models can provide more reliable and informative results.

**Keywords** Differential equation models · Stochastic modeling · Itô diffusion process · Chemical Langevin equation · Parameter identifiability · mRNA transfection

**Mathematics Subject Classification** 92-08 · 60J70 · 62F15

---

✉ Christiane Fuchs  
christiane.fuchs@uni-bielefeld.de

<sup>1</sup> Institute of Computational Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Oberschleißheim, Germany

<sup>2</sup> Department of Mathematics, Technical University Munich, Garching, Germany

<sup>3</sup> Faculty of Mathematics and Natural Sciences, University of Bonn, Bonn, Germany

<sup>4</sup> Faculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany

## 1 Introduction

mRNA transfection is the process of introducing mRNA into a living cell. mRNA delivery has become increasingly interesting for biomedical applications because it enables treatment of diseases by means of targeted expression of proteins and it is transient, avoiding the risk of permanently integrating into the genome (see e. g. Sahin et al. 2014). The most prominent applications of mRNA transfection are the mRNA-based vaccines that are successfully used to fight the COVID-19 pandemic (DeFrancesco 2020; Borah et al. 2021). These successful use cases demonstrate the importance of a precise understanding of the dynamics of the underlying processes.

We aim at facilitating further insights into the determinants of the mRNA delivery process and the translation kinetics through the use of mechanistic modeling and parameter inference for such models from experimental data. Therefore, we consider data from an mRNA transfection experiment using fluorescence reporters and fluorescence microscopy which is one of the few ways to measure quantities within a living cell over time (i. e. keeping it alive is necessary). Due to the discrete nature of the molecular species within a cell and due to the fact that random fluctuations play a key role (Elowitz et al. 2002; Raj and van Oudenaarden 2008), a continuous-time, discrete-space Markov process, also called a *Markov jump process (MJP)*, for which the dynamics are described by the so-called chemical master equation (CME), is widely accepted to be an appropriate stochastic description of the biochemical processes within a cell (Schnoerr et al. 2017; Gillespie 1992). However, parameter inference for MJPs is computationally very demanding and often infeasible (see e. g. Warne et al. 2019). Therefore, several other representations of the biochemical kinetics have been developed. Higham (2008) provides an accessible introduction to those. To some extent they can be considered as approximations to the corresponding MJP. The most commonly used representation is the *reaction rate equation (RRE)* which is a system of *ordinary differential equation (ODEs)* and thus provides a deterministic and state-continuous description of the kinetics. One approach that preserves the stochastic nature of the underlying process is the approximation by *Itô diffusion processes*. These are continuous-time, continuous-space stochastic processes described by *Itô-type stochastic differential equation (SDEs)*.

Single-cell fluorescence data from transfection experiments has been analyzed based on ODE modeling in several previous studies, e. g. in Ligon et al. (2014), Leonhardt et al. (2014), Fröhlich et al. (2018), and Reiser et al. (2019). In all of these studies, several parameters of the considered ODE models for the translation kinetics after mRNA transfection were not identifiable from the experimental data. Moreover, the quality of the parameter estimates, i. e. whether the true kinetic rates are adequately captured, is unclear. Using models that explicitly account for stochasticity can help improve our ability to determine kinetic parameters from experimental data (Munsky et al. 2009). There are two important notions with respect to parameter identifiability (Raue et al. 2009): *Structural identifiability* exclusively considers the structure of the model including the model of the observations and answers the question whether the parameters can be uniquely determined if we are given perfect data, i. e. an infinite amount of data observed without measurement error and continuously in time. Whereas *practical identifiability* is concerned with the question whether the

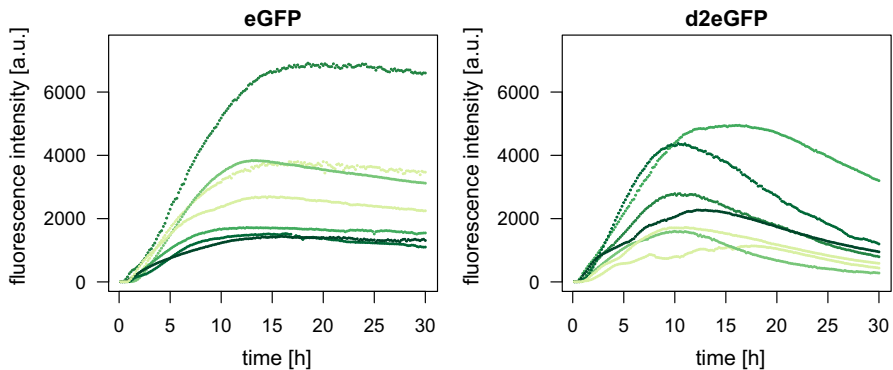
parameters can be determined from a specific data set (which is always finite and usually subject to measurement error). Browning et al. (2020) have recently compared parameter identifiability for ODE and SDE modeling approaches for example models from different contexts based on simulated data and showed that SDE modeling does improve the identifiability for the studied models.

Here, we formulate an SDE model of the translation kinetics after mRNA transfection and compare its parameter identifiability to that of the corresponding ODE model. This approach serves two purposes at the same time: First of all, we aim to get further biological insight into the kinetic processes. Second, we investigate and showcase the benefit of SDE models over ODE models with respect to parameter identifiability on the specific biological example. In doing so, we build on the work of others: While the same ODE model has already been considered e. g. by Fröhlich et al. (2018) and Ballnus et al. (2017), we contrast it with the corresponding SDE dynamics. Inference from fluorescence data for SDE models has also been conducted e. g. in Heron et al. (2007), Finkenstädt et al. (2008), and Komorowski et al. (2009), however for an experimental setup that also included the transcription process, other than in the experiments considered here. We add the focus on parameter identifiability and the exclusive use of diffuse priors as novel aspects.

This article is structured as follows: In Sect. 2, we describe the experimental data that motivated the study. In Sect. 3, we present the biochemical reaction network which we want to consider for the translation kinetics after mRNA transfection and formulate its ODE and SDE representation. After stating the assumed model of the observations and summarizing the parameters that we would like to infer from these observations in Sect. 4, we consider several approaches to assess structural identifiability of the parameters for both model types in Sect. 5. We define the parameter posteriors for both modeling approaches and study the practical identifiability of the parameters based on simulated and experimental data in Sect. 6, and conclude with a summary and discussion of our findings in Sect. 7. All relevant code to perform the presented analysis is available on Github: [https://github.com/fuchslab/Translation\\_kinetics\\_after\\_transfection](https://github.com/fuchslab/Translation_kinetics_after_transfection).

## 2 Experimental data

We consider data that has previously been analyzed (based on ODE modeling) and published in Fröhlich et al. (2018). The data was generated in an experiment where human hepatoma epithelial cells from the cell line HuH7 were transfected with mRNA encoding a green fluorescent protein (GFP). The cells were fixed on micro patterned protein arrays and time lapse microscopy images of the cells were taken every 10 minutes over the course of at least 30 hours (i.e. there are at least 180 measurements per cell). For the first hour, the cells were incubated with mRNA lipoplexes which consist of (cationic) lipid membrane encapsulating GFP coding mRNA. Afterwards, the cells were washed with cell culture medium such that no further lipoplex uptake occurred. The time point at which the lipoplexes were taken up, dissolved and released the mRNA as well as the number of mRNA molecules released are unknown. For the considered experimental setup, individual lipoplexes carry on average 350 mRNAs



**Fig. 1** Trajectories of the mean fluorescence intensity for seven cells from the mRNA transfection experiment in Fröhlich et al. (2018) for eGFP and d2eGFP (April 27, 2016), respectively

and the number of lipoplexes entering a single cell seems to be very low (Leonhardt et al. 2014).

The released mRNA was translated into a fluorescent protein which caused the cells to fluoresce. For each image taken during the experiment, the fluorescence intensity is integrated over squares occupied by one cell in order to obtain one value for the mean fluorescence intensity per cell and time point (see Fröhlich et al. 2018, for further details about the image analysis and Fig. 1 therein for an example image).

The experiment was conducted with two different types of GFP that differ in their protein lifetime: enhanced GFP (eGFP) and a destabilized enhanced GFP (d2eGFP). The data set contains measurements for more than 800 cells for each type of GFP. Some trajectories of the mean fluorescence intensity are displayed in Fig. 1.

It was shown before that ODE models of the translation kinetics of an individual cell are not globally identifiable with the available experimental data as described above (Ballnus et al. 2017). Several of the ODE model parameters cannot be uniquely determined based on one observed fluorescence trajectory. Fröhlich et al. (2018) use a mixed-effect ODE model in order to incorporate the translation kinetics of several cells and data for both different types of GFP (eGFP and d2eGFP). Through this approach, they are able to improve parameter identifiability (by breaking the symmetry between the degradation rate constants); however, their approach is computationally very intense, necessitates distribution assumptions about the random effects, required conducting the experiment with two types of GFP, and still leaves several parameters non-identifiable. Here, we are interested in the question whether the use of an SDE model can improve the parameter identifiability even when only one fluorescence trajectory is observed.

### 3 Mathematical models of the translation kinetics

Following the introduction of mRNA into a living cell, i. e. after *mRNA transfection*, the information contained in the mRNA molecules is used for the production of (green

fluorescent) proteins, which is the *translation* process. We will focus on modeling the translation kinetics of one cell in order to study parameter identifiability based on one observed fluorescence trajectory. Therefore, we only consider the (released) mRNA and the GFP molecules explicitly:

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \begin{array}{l} \text{amount of mRNA molecules,} \\ \text{amount of GFP molecules.} \end{array}$$

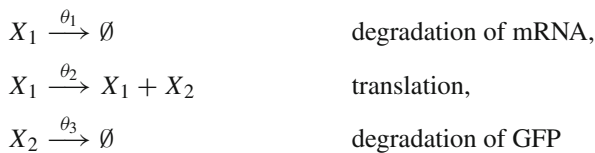
As the number of lipoplexes entering a cell is small (Leonhardt et al. 2014) and as most observed fluorescence trajectory do not show a bi-phasic behaviour, we assume that all mRNA molecules (within one cell) are released at once from the lipoplexes and denote this initial time point by  $t_0$ . Before  $t_0$ , there are neither mRNA nor GFP molecules, and at  $t_0$ , an amount of  $m_0$  mRNA molecules is released, i.e.

$$X(t) \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ for } t < t_0 \quad \text{and} \quad X(t_0) = \begin{pmatrix} m_0 \\ 0 \end{pmatrix}.$$

Conceivable extensions of this basic model are e. g. to include enzymatic degradation of the mRNA and/or the protein, ribosomal binding to the mRNA for translation, and a maturation step of the protein. However, we will only consider the basic configuration as described above.

### 3.1 Markov jump process

Assuming that the matter within the cell is well-stirred and in thermal equilibrium, an MJP is regarded to be the most adequate representation of this system after  $t_0$ . In our model, there are three possible reactions:



with kinetic rate parameters  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ . Given the state vector  $X(t) = (X_1(t), X_2(t))^T$ , the propensity functions of the three reactions are

$$f_1(X(t)) = \theta_1 X_1, \quad f_2(X(t)) = \theta_2 X_1, \quad \text{and} \quad f_3(X(t)) = \theta_3 X_2,$$

i. e. the probability that reaction  $i$  occurs within the next time interval of length  $\Delta t$  is approximately  $f_i(X(t))\Delta t$  for  $i = 1, 2, 3$ .

If we denote the probability distribution of the random variable  $X(t)$  by

$$P_{i,j}(t) = \mathbb{P}(X_1(t) = i, X_2(t) = j)$$

for a fixed initial condition on  $X(t_0)$ , the corresponding CME reads

$$\begin{aligned} \frac{\partial P_{i,j}(t)}{\partial t} = & \theta_1(i+1)P_{i+1,j}(t) + \theta_2iP_{i,j-1}(t) + \theta_3(j+1)P_{i,j+1}(t) \\ & - (\theta_1i + \theta_2i + \theta_3j)P_{i,j}(t). \end{aligned}$$

Although the system contains only first-order reactions, there is no known closed-form solution to the CME. Thus, the generation of model trajectories and the evaluation of the likelihood of data given parameter values requires computationally demanding stochastic simulations. This fact impedes efficient inference for this MJP and leads to the consideration of other modeling approaches.

### 3.2 ODE model

The RRE is a system of ODEs that provides a deterministic and state-continuous description of biochemical kinetics and can be derived as the large-volume limit of the corresponding MJP (see e.g. Kurtz 1972). The following system of ODEs is a deterministic approximation of the MJP modeling the dynamics as described above:

$$\frac{dX(t)}{dt} = \begin{pmatrix} -\theta_1 X_1(t) \\ \theta_2 X_1(t) - \theta_3 X_2(t) \end{pmatrix} \quad \text{for } t \geq t_0. \quad (1)$$

This system has previously been formulated and analyzed e.g. in Leonhardt et al. (2014). It admits the solution

$$\begin{aligned} X_1(t) &= m_0 \exp(-\theta_1(t-t_0)), \\ X_2(t) &= \begin{cases} \frac{\theta_2 m_0}{\theta_3 - \theta_1} (e^{-\theta_1(t-t_0)} - e^{-\theta_3(t-t_0)}) & , \text{ for } \theta_1 \neq \theta_3, \\ \theta_2 m_0 (t-t_0) e^{-\theta_3(t-t_0)} & , \text{ for } \theta_1 = \theta_3. \end{cases} \end{aligned} \quad (2)$$

Note that the solution for  $X_2(t)$  is symmetric in the parameters  $\theta_1$  and  $\theta_3$ . Moreover, note that while it is common for the RRE to describe molecule concentrations; here, we let  $X(t)$  denote molecule numbers. However, since the model that we consider here contains only first-order reactions, this does not affect the interpretation of the kinetic rate parameters (i. e. Eq. (1) would look the same for concentrations).

### 3.3 SDE model

While the ODE model of the translation kinetics after mRNA transfection had been considered before, we are the first to formulate the stochastic but state-continuous approximation to the MJP in Sect. 3.1 based on an  $(It\delta)$  diffusion process. The approximating diffusion process is described by an SDE which is known as the chemical Langevin equation (CLE) in the context of biochemical kinetics (Gillespie 2000). The CLE is derived based on the changes in molecule numbers caused by the three possible

reactions and on the corresponding propensities and reads:

$$d\mathbf{X}(t) = \begin{pmatrix} -\theta_1 X_1(t) \\ \theta_2 X_1(t) - \theta_3 X_2(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{\theta_1 X_1(t)} & 0 \\ 0 & \sqrt{\theta_2 X_1(t) + \theta_3 X_2(t)} \end{pmatrix} d\mathbf{B}(t) \tag{3}$$

for  $t \geq t_0$  and where  $\mathbf{B}(t)$  is a 2-dimensional standard Brownian motion. The solution of Eq. (3) is not explicitly known and will be approximated based on the Euler-Maruyama scheme as explained in Sect. A.1 of the supplementary material. See Fuchs (2013) for a concise introduction to diffusion processes and several approaches to derive general diffusion approximations.

## 4 Model parameters and observations

### 4.1 Model of the observations

In the experiment described in Sect. 2, neither the amount of mRNA molecules nor that of GFP molecules can be measured directly. Instead, a fluorescence signal is observed which is composed of a background fluorescence and a signal which is proportional to the amount of GFP molecules. Moreover, previous studies assumed a constant offset and multiplicative measurement noise in the recorded fluorescence trajectories (Fröhlich et al. 2018). Therefore, denoting a trajectory of mean fluorescence intensity observed at time points  $t_k$ , for  $k = 1, \dots, K$ , by  $\{y_k\}_{k=1, \dots, K}$ , we assume that

$$y_k = (\text{scale} \cdot X_2(t_k) + \text{offset}) \cdot \epsilon_k, \quad \epsilon_k \sim \log \mathcal{N}(0, \sigma^2),$$

where the random variables  $\epsilon_k$  are independent.

Note that the observations depend only on the amount  $X_2$  of GFP molecules, but not directly on the amount  $X_1$  of mRNA molecules.

### 4.2 Model parameters

Based on the observations  $\{y_k\}_{k=1, \dots, K}$ , we aim to infer the following unknown parameters:

- The three kinetic parameters  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$  that denote the rate constants for mRNA degradation, translation, and GFP degradation,
- The initial amount  $m_0$  of mRNA molecules and the time point  $t_0$  at which it is released,
- The scaling factor *scale* and the *offset* for the fluorescence signal,
- And the standard deviation  $\sigma$  of the measurement errors.

In the remainder of this work, we investigate how far such inference is possible in theory and practice.

## 5 Structural identifiability analysis

Our main interest lies in the question which of the model parameters for our two model types (ODE and SDE) can be inferred from the experimental data as described in Sects. 2 and 4. Here, we first focus on the parameters  $\theta$ ,  $m_0$ , scale, and offset that drive the dynamics of the process and the fluorescence signal. We analyze the structural identifiability which only considers the model equations of the process dynamics and the observation equation (not the actual data) and assumes that we are in a perfect data situation, i. e. we have an infinite amount of data observed without measurement error (Raue et al. 2009). Plainly speaking, structural identifiability analysis answers the question whether different parameter combinations can lead to the same model output. While for ODE models, there are analytical methods to assess structural identifiability, no such methods exist for SDE models. Therefore, we consider several approaches that do not directly assess the structural identifiability for our SDE model, but answer related questions. In the following subsections, we consider a transformed version of both model types, we make use of a surrogate model and the open source software DAISY as has recently been suggested by Browning et al. (2020), and finally we also study simulations of both model types.

### 5.1 Transformed models

We can reformulate the differential equations for both model types by setting

$$\mathbf{Z}(t) = \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix} := \begin{pmatrix} \frac{X_1(t)}{m_0} \\ \text{scale} \cdot X_2(t) + \text{offset} \end{pmatrix},$$

which means that

$$\mathbf{Z}(t) \equiv \begin{pmatrix} 0 \\ \text{offset} \end{pmatrix} \text{ for } t < t_0, \quad \text{and } \mathbf{Z}(t_0) = \begin{pmatrix} 1 \\ \text{offset} \end{pmatrix}.$$

Hence, the second component of the transformed process models the fluorescence signal which we assume to be observed, meaning  $y = Z_2$ .

#### Transformed ODE model

For the ODE model in Eq. (1), the transformed model reads

$$\frac{d\mathbf{Z}(t)}{dt} = \begin{pmatrix} -\theta_1 Z_1(t) \\ \text{scale} \theta_2 m_0 Z_1(t) - \theta_3 (Z_2(t) - \text{offset}) \end{pmatrix} \text{ for } t \geq t_0, \quad (4)$$



and has the solution

$$\begin{aligned}
 Z_1(t) &= \exp(-\theta_1(t - t_0)), \\
 Z_2(t) &= \begin{cases} \frac{\text{scale} \theta_2 m_0}{\theta_3 - \theta_1} (e^{-\theta_1(t-t_0)} - e^{-\theta_3(t-t_0)}) + \text{offset} & , \text{ for } \theta_1 \neq \theta_3, \\ \text{scale } \theta_2 m_0 (t - t_0) e^{-\theta_3(t-t_0)} + \text{offset} & , \text{ for } \theta_1 = \theta_3. \end{cases}
 \end{aligned}$$

The parameters *scale*, *m*<sub>0</sub>, and  $\theta_2$  appear only as a product. Thus, it can already be deduced from this equation that at most the product of the three parameters will be identifiable but not the three parameters individually (Fröhlich et al. 2018). Moreover, since only  $Z_2(t)$  is observed and it is symmetric in the parameters  $\theta_1$  and  $\theta_3$  (i. e. switching their values will lead to the same model output), these two parameters can at most be locally identifiable if no further constraints are imposed on them as done in Leonhardt et al. (2014).

**Transformed SDE model**

For the SDE model in Eq. (3), we apply the Itô formula (as stated in Sect. A.1 of the supplementary material) to obtain the transformed model

$$\begin{aligned}
 d\mathbf{Z}(t) &= \begin{pmatrix} -\theta_1 Z_1(t) \\ \text{scale } \theta_2 m_0 Z_1(t) - \theta_3 (Z_2(t) - \text{offset}) \end{pmatrix} dt \\
 &+ \begin{pmatrix} \sqrt{\frac{\theta_1}{m_0}} Z_1(t) & 0 \\ 0 & \sqrt{\text{scale} \sqrt{\text{scale } \theta_2 m_0 Z_1(t) + \theta_3 (Z_2(t) - \text{offset})}} \end{pmatrix} dB_t \text{ for } t \geq t_0. \quad (5)
 \end{aligned}$$

Note that here, the parameters *scale* and *m*<sub>0</sub> also appear outside the product *scale*  $\theta_2 m_0$ . Furthermore, there is no apparent symmetry between  $\theta_1$  and  $\theta_3$ . Therefore, it might be possible to gain more information about the individual parameters from data for the SDE model than for the ODE model.

**5.2 Using a surrogate model**

Next, we want to use the open source software DAISY (Differential Algebra for Identifiability of SYstems, Bellu et al. 2007) to assess the structural identifiability of the parameters in the two models of the translation kinetics. DAISY is a software tool that implements a differential algebra algorithm to perform structural identifiability analysis for systems of polynomial or rational ODEs and that also allows to include unknown initial conditions. We use the transformed models from the previous subsection for the identifiability analysis. For the ODE model in Eq. (4), the analysis with DAISY is straight forward since it is intended for the use for ODE models. After applying DAISY, the obtained output shows that when considering the set of parameters  $\{\theta, m_0, \text{scale}, \text{offset}\}$ , the model is non-identifiable. The DAISY output also reveals that this non-identifiability is due to the fact that the parameters  $\theta_1$  and  $\theta_3$  are only locally identifiable and the parameters  $\theta_2, m_0$ , and *scale* are not individually identifiable, but only their product is. This confirms the assertions from the previous

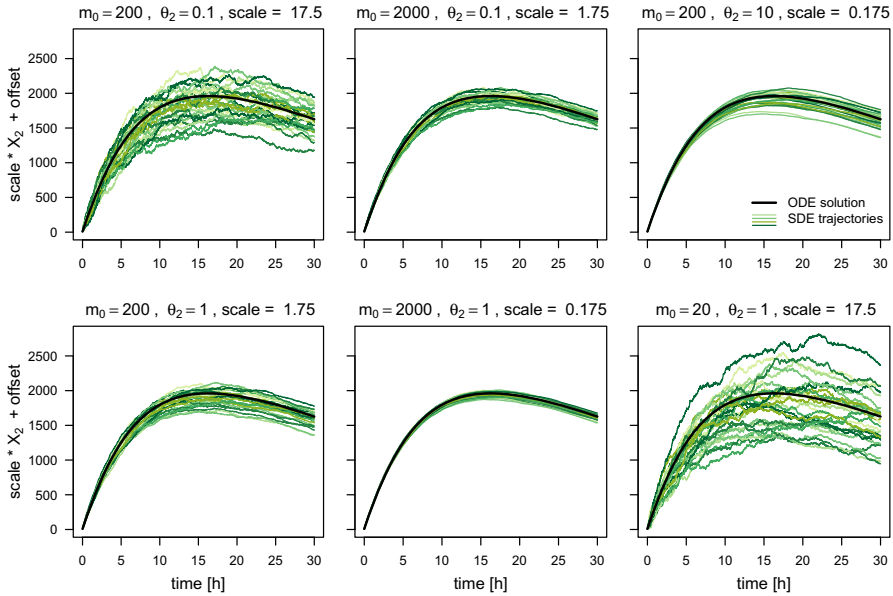
subsection. Moreover, we obtain that the remaining parameter offset is structurally identifiable.

For SDE models, Browning et al. (2020) suggest to formulate a surrogate model based on the moment equations of the diffusion process. The moment equations are a system of ODEs, and thus, DAISY can be applied to this system. For the SDE (5), let  $M_{ij}(t) = \mathbb{E}[(Z_1(t))^i (Z_2(t))^j]$  be the (mixed) moment of the diffusion process of order  $i$  and  $j$ . The moments are obtained by applying the Itô formula (as stated in Sect. A.1 of the supplementary material) to  $(Z_1(t))^i (Z_2(t))^j$  and then taking the expectation. Considering the first and the second moments of the process states results in the following system of ODEs:

$$\begin{aligned} \frac{dM_{10}(t)}{dt} &= -\theta_1 M_{10}(t), & M_{10}(t_0) &= 1, \\ \frac{dM_{01}(t)}{dt} &= \text{scale } \theta_2 m_0 M_{10}(t) - \theta_3 M_{01}(t) + \theta_3 \text{offset}, & M_{01}(t_0) &= \text{offset}, \\ \frac{dM_{20}(t)}{dt} &= \frac{\theta_1}{m_0} M_{10}(t) - 2\theta_1 M_{20}(t), & M_{20}(t_0) &= 1, \\ \frac{dM_{02}(t)}{dt} &= \text{scale}^2 \theta_2 m_0 M_{10}(t) + \theta_3 (\text{scale} \\ &\quad + 2\text{offset}) M_{01}(t) - 2\theta_3 M_{02}(t) \\ &\quad + 2\text{scale } \theta_2 m_0 M_{11}(t) - \text{scale } \theta_3 \text{offset}, & M_{02}(t_0) &= \text{offset}^2, \\ \frac{dM_{11}(t)}{dt} &= \theta_3 \text{offset } M_{10}(t) + \text{scale } \theta_2 m_0 \\ &\quad M_{20}(t) - (\theta_1 + \theta_3) M_{11}(t), & M_{11}(t_0) &= \text{offset}, \end{aligned}$$

where the equations for the two first moments  $M_{10}$  and  $M_{01}$  coincide with the ODE model in Eq. (4). Since in the experiment, only the fluorescence signal is observed, we consider the moments that only depend on the second component of the process, i. e.  $M_{01}$  and  $M_{02}$ , as output states for the identifiability analysis. Using DAISY, we obtain that the surrogate model is globally identifiable, i. e. all six parameter values could be uniquely determined if we were able to observe the moments  $M_{01}$  and  $M_{02}$  directly, infinitely long over time, and without measurement error. While this implies that the parameters are structurally identifiable from moment measurements, it does not directly answer our original question of whether the parameters of the SDE model can be identified from perfect observations of a single fluorescence trajectory.

Moreover, as has already been pointed out, structural identifiability analysis assumes a perfect data situation. From its results, we cannot conclude that the parameters will be identifiable from the actual experimental data (which is always finite and usually subject to measurement error). In contrast, the notion of practical identifiability is concerned with the question whether the parameters can be determined from a specific data set (Raue et al. 2009) and we will address this question in Sect. 6.



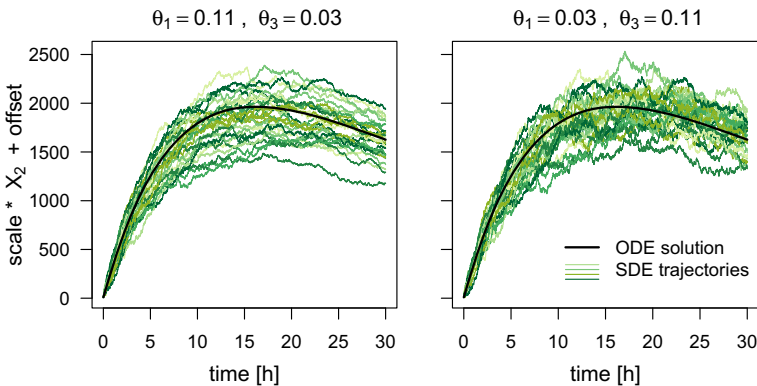
**Fig. 2** The ODE trajectory and 20 SDE trajectories of the fluorescence intensity simulated for different values of  $m_0$ ,  $\theta_2$ , and scale, while keeping their product constant at  $\text{scale}\theta_2m_0 = 350$ . The remaining parameters are set to  $\theta_1 = 0.11$ ,  $\theta_3 = 0.03$ , offset = 8.9, and  $t_0 = 0$

### 5.3 Simulation-based assessment of parameter influence

Another attempt to assess parameter identifiability is to simulate from both model types for different parameter settings and compare whether we see differences in the simulation output. To obtain simulations from the ODE model, we use its solution in Eq. (2). Since the ODE model is deterministic, each parameter setting yields one unique output trajectory; whereas for the SDE model, we simulate several trajectories for each parameter setting using the Euler-Maruyama scheme with a time step of 0.01 hours.

#### Keeping the product scale $\theta_2m_0$ constant

As already pointed out in Sect. 5.1, the trajectories of the fluorescence intensity for the ODE model are identical if the product scale  $\theta_2m_0$  and the remaining parameters are fixed, even when the individual factors scale,  $\theta_2$ , and  $m_0$  vary. To assess whether this is (at least approximately) also the case for the SDE model, we simulate several trajectories with different values for scale,  $\theta_2$ , and  $m_0$  while keeping their product constant. For each parameter setting, we set the same random seed at the beginning of the simulation. To ensure that the analysis is performed in a relevant regime, we chose the parameters similar to the parameters estimated by Fröhlich et al. (2018):  $\text{scale}\theta_2m_0 = 350$ ,  $\theta_1 = 0.11$ ,  $\theta_3 = 0.03$ , offset = 8.9, and  $t_0 = 0$ . Fig. 2 displays the simulated trajectories.



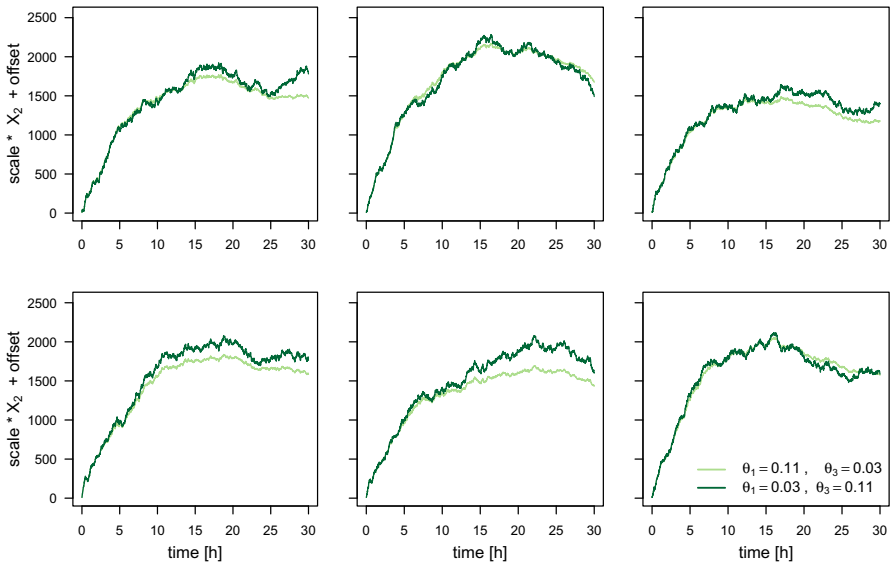
**Fig. 3** The ODE trajectory and 20 SDE trajectories of the fluorescence intensity simulated for two parameter combinations where the values of  $\theta_1$  and  $\theta_3$  are swapped. The remaining parameters are set to  $\text{scale} = 17.5$ ,  $\theta_2 = 0.1$ ,  $m_0 = 200$ ,  $\text{offset} = 8.9$ , and  $t_0 = 0$

It is evident that the SDE trajectories behave differently for different combinations of scale,  $\theta_2$ , and  $m_0$  which yield the same values of the product  $\text{scale}\theta_2m_0$ . For example, when we keep  $m_0$  fixed while increasing scale and decreasing  $\theta_2$ , the variation between but also within the trajectories increases. When we keep scale fixed while decreasing  $m_0$  and increasing  $\theta_2$ , especially the variation between trajectories seems to increase. And finally, when we keep  $\theta_2$  fixed while decreasing  $m_0$  and increasing scale, the variation between and within the trajectories increases. Our focus is on estimating the parameters from individual observed trajectories. In this context, especially the difference in the variation within the trajectories is relevant.

### Swapping the degradation rate constants $\theta_1$ and $\theta_3$

The trajectories of the fluorescence intensity for the ODE model are also identical if the values for  $\theta_1$  and  $\theta_3$  are swapped while the remaining parameters are fixed. We simulate trajectories for the parameter combinations  $(\theta_1, \theta_3) = (0.11, 0.03)$  and  $(\theta_1^*, \theta_3^*) = (0.03, 0.11)$ , respectively, while setting the remaining parameters to  $\text{scale} = 17.5$ ,  $\theta_2 = 0.1$ ,  $m_0 = 200$ ,  $\text{offset} = 8.9$ , and  $t_0 = 0$ . For the SDE model, we again simulate several trajectories for both parameter settings and set the same random seed at the beginning of the simulation.

Figure 3 shows the ODE trajectory and several SDE trajectories in one panel for each of the two parameter combinations separately. Whereas, Fig. 4 presents one SDE trajectory for each of the two parameter combinations together in one panel. Again, the SDE trajectories do behave differently for the different parameter combinations. While there seems to be only little difference in the variation between the trajectories, the variation within the trajectories is clearly higher for lower  $\theta_1$  and higher  $\theta_3$ . This indicates that it may be possible to uniquely determine the values of  $\theta_1$  and  $\theta_3$  even when estimating from only one observed trajectory.



**Fig. 4** One trajectory of the fluorescence intensity for the SDE model simulated for each of the two parameter combinations where the values of  $\theta_1$  and  $\theta_3$  are swapped. The remaining parameters are set to  $scale = 17.5$ ,  $\theta_2 = 0.1$ ,  $m_0 = 200$ ,  $offset = 8.9$ , and  $t_0 = 0$

## 6 Posterior properties and credible intervals

After having studied the structural properties of the ODE and SDE model in the previous section; next, we would like to assess the practical parameter identifiability by trying to estimate the parameters from data as described in Sect. 4. We take a Bayesian approach to parameter estimation because it allows for uncertainty assessment of the parameter estimates and also for handling unobserved process components and measurement error by using Markov chain Monte Carlo (MCMC) methods to sample from the parameter posterior distribution. Therefore, in this section, we define the parameter posterior densities for the two model types and study their properties based on MCMC sampling results for simulated data generated without and with (simulated) measurement error as well as for experimental data.

### 6.1 Formulation of the inference problem

In Bayesian statistics, we can formulate our assumptions and general knowledge about the model parameters  $\xi \in \Xi \subseteq \mathbb{R}^p$  in terms of a prior distribution with probability density  $p(\xi)$ . After having observed data  $\mathcal{D}$  about the phenomenon which we are trying to model, we update our knowledge about the parameter and describe it by the posterior distribution with density  $\pi(\xi | \mathcal{D})$ . In our case, we consider data  $\mathcal{D} = \{y_k\}_{k=1, \dots, K}$  and the vector of all unknown parameters is  $\xi = (\theta, m_0, scale, offset, t_0, \sigma)$ . The relation between the prior and the posterior density resulting from Bayes' theorem is given by:

$$\pi(\xi | \mathcal{D}) \propto p(\mathcal{D} | \xi) p(\xi),$$

where  $p(\mathcal{D} | \xi)$  denotes the density of the distribution of  $\mathcal{D}$  conditioned on  $\xi$  and is determined by the considered model. Viewed as a function of the parameter,  $\mathcal{L}(\xi | \mathcal{D}) := p(\mathcal{D} | \xi)$  is called the likelihood (function). Comprehensive introductions to Bayesian statistics can be found e. g. in Lee (2012) and Gelman et al. (2013).

To sample from the subsequently formulated posterior densities for the two model types, we use the open source software Stan (Carpenter et al. 2017) which provides an efficient C++ implementation of the Hamiltonian Monte Carlo (HMC) based No-U-Turn Sampler (NUTS). See Sect. A.3 of the supplementary material for a brief introduction to this topic. We use Stan through its R interface `rstan` (Stan Development Team 2019).

### Posterior of the ODE model

For the ODE model, there is a deterministic relationship between the process values  $X(t)$  and the parameters  $\theta$ ,  $m_0$  and  $t_0$  (or between the fluorescence signal and the parameters additionally including scale and offset, respectively).

Define the index  $k^* := \min\{k \in \{1, \dots, K\} | t_k \geq t_0\}$  of the first observation time point after the mRNA molecules are released, then the posterior density  $\pi$  from which we would like to sample is proportional to

$$\begin{aligned} & \pi(\theta, m_0, \text{scale}, \text{offset}, \sigma^2, t_0 | \{y_k\}_{k=1, \dots, K}) \\ & \propto \left( \prod_{k=k^*}^K \phi\left(\log(y_k) \mid \log\left(\text{scale} \frac{\theta_2 m_0}{\theta_3 - \theta_1} \left(e^{-\theta_1(t-t_0)} - e^{-\theta_3(t-t_0)}\right) + \text{offset}\right), \sigma^2\right) \right) \\ & \quad \cdot \left( \prod_{k=1}^{k^*-1} \phi\left(\log(y_k) \mid \log(\text{offset}), \sigma^2\right) \right) \\ & \quad \cdot p(\theta_1) p(\theta_2) p(\theta_3) p(m_0) p(t_0) p(\text{scale}) p(\text{offset}) p(\sigma^2), \end{aligned} \quad (6)$$

where  $\phi(\cdot | \mu, \eta^2)$  denotes the density of the normal distribution with mean  $\mu$  and variance  $\eta^2$  and the  $p(\cdot)$  denote the parameter prior densities which we assume to be independent.

If the priors  $p(\theta_1)$  and  $p(\theta_3)$  are symmetric to each other, then the posterior is also symmetric with respect to the two degradation rate constants.

The scaling factor  $\text{scale}$ , the translation rate constant  $\theta_2$ , and the initial amount of mRNA  $m_0$  appear only as a product in the likelihood function; therefore, as pointed out before, at most their product  $\text{scale} \theta_2 m_0$  is identifiable.

### Posterior of the SDE model

For the SDE model, the states  $X(t_k)$ , for  $k = 1, \dots, K$ , of the process conditioned on the parameters  $\theta$ ,  $m_0$  and  $t_0$  are random numbers (for  $t_k \geq t_0$ ). Hence, we have to marginalize over the process states to obtain the posterior density of the parameters

which we want to infer:

$$\begin{aligned} &\pi \left( \theta, m_0, \text{scale}, \text{offset}, \sigma^2, t_0 | \{y_k\}_{k=1, \dots, K} \right) \\ &= \int_{\mathbb{R}_+^{2 \times K}} \pi \left( \theta, m_0, \text{scale}, \text{offset}, \sigma^2, t_0, \{\mathbf{X}(t_k)\}_{k=1, \dots, K} | \{y_k\}_{k=1, \dots, K} \right) \\ &\quad d\mathbf{X}(t_1) \dots d\mathbf{X}(t_K). \end{aligned}$$

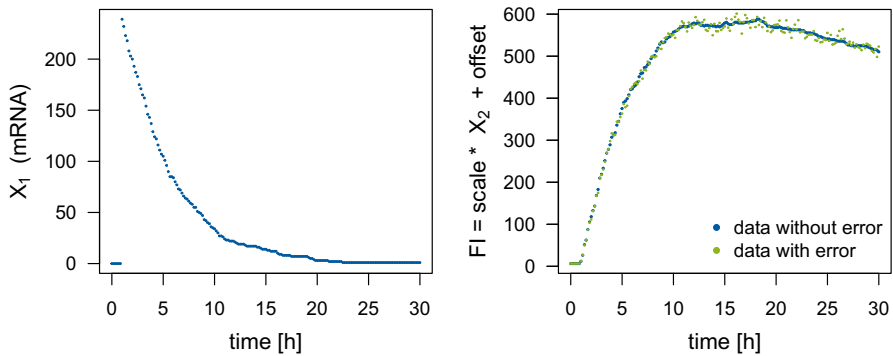
Therefore, again defining  $k^* := \min\{k \in \{1, \dots, K\} | t_k \geq t_0\}$ , we would need to sample from

$$\begin{aligned} &\pi \left( \theta, m_0, \text{scale}, \text{offset}, \sigma^2, t_0, \{\mathbf{X}(t_k)\}_{k=1, \dots, K} | \{y_k\}_{k=1, \dots, K} \right) \\ &\propto \left( \prod_{k=1}^K \phi \left( \log(y_k) | \log(\text{scale} \cdot X_2(t_k) + \text{offset}), \sigma^2 \right) \right) \\ &\cdot \left( \prod_{k=k^*}^{K-1} \pi \left( \mathbf{X}(t_{k+1}) | \mathbf{X}(t_k), \theta \right) \right) \pi \left( \mathbf{X}(t_{k^*}) | \theta, m_0, t_0 \right) \left( \prod_{k=1}^{k^*-1} \delta(\|\mathbf{X}(t_k) - (0, 0)^T\|) \right) \\ &\cdot p(\theta) p(m_0) p(t_0) p(\text{scale}) p(\text{offset}) p(\sigma^2), \end{aligned}$$

where  $\phi(\cdot | \mu, \eta^2)$  denotes the density of the normal distribution with mean  $\mu$  and variance  $\eta^2$ ,  $\delta(\cdot)$  denotes the Dirac delta function,  $\|\cdot\|$  denotes a norm (e.g. the  $l_2$ -norm), and the factors  $\pi(\mathbf{X}(t_{k+1}) | \mathbf{X}(t_k), \theta), k = k^*, \dots, K - 1$ , denote the transition density of the process. However, the fact that the process  $\mathbf{X}$  switches from a deterministic regime before  $t_0$  to a stochastic one after  $t_0$  complicates the estimation of  $t_0$  together with the remaining parameters. Therefore, we will assume that  $t_0$  is determined beforehand, e.g. based on the estimates for the ODE model. Consequently, we sample from

$$\begin{aligned} &\pi \left( \theta, m_0, \text{scale}, \text{offset}, \sigma^2, \{\mathbf{X}(t_k)\}_{k=1, \dots, K} | \{y_k\}_{k=1, \dots, K}, t_0 \right) \\ &\propto \left( \prod_{k=1}^K \phi \left( \log(y_k) | \log(\text{scale} \cdot X_2(t_k) + \text{offset}), \sigma^2 \right) \right) \\ &\cdot \left( \prod_{k=k^*}^{K-1} \pi \left( \mathbf{X}(t_{k+1}) | \mathbf{X}(t_k), \theta \right) \right) \pi \left( \mathbf{X}(t_{k^*}) | \theta, m_0, t_0 \right) \left( \prod_{k=1}^{k^*-1} \delta(\|\mathbf{X}(t_k) - (0, 0)^T\|) \right) \\ &\cdot p(\theta) p(m_0) p(\text{scale}) p(\text{offset}) p(\sigma^2). \tag{7} \end{aligned}$$

While for the ODE model, the posterior distribution is only 8-dimensional and can be sampled from directly; for the SDE model, we need to sample from a  $(7 + 2K)$ -dimensional distribution and then marginalize over the  $2K$  dimensions of the process states to obtain the posterior distribution of the parameters of interest. Moreover,



**Fig. 5** One trajectory used in the simulation study that was simulated with Gillespie's algorithm with parameters  $\theta = (0.2, 0.32, 0.01)$ ,  $m_0 = 240$ ,  $t_0 = 0.96$ , scale = 1.8, and offset = 8.5, and for the green dotted line, multiplicative measurement error with  $\sigma = 0.02$  was added to the fluorescence intensity (FI)

there is no explicit exact expression for the transition density  $\pi(X(t_{k+1})|X(t_k), \theta)$ ; wherefore, it will be approximated by a normal density based on the Euler-Maruyama scheme as explained in Sect. A.1 of the supplementary material. For this approximation to be appropriate, we have to ensure that the time steps between observations are small enough. We do so in Sect. A.2 of the supplementary material. For more complex models or lower-resolution data, approaches like pseudo-marginal methods would be required instead of the Euler-Maruyama scheme.

## 6.2 Estimation based on simulated data

In order to assess how well we can recover the model parameters for both model types from individually observed trajectories, we first work with simulated data that is generated with Gillespie's algorithm (Gillespie 1976, 1977).

### 6.2.1 Sampling results for simulated data without measurement error

For now, we assume the fluorescence intensity to be observed without measurement error. We consider synthetic data generated with parameters  $\theta = (0.2, 0.32, 0.01)$ ,  $m_0 = 240$ ,  $t_0 = 0.96$ , scale = 1.8, and offset = 8.5, which are based on the results by Fröhlich et al. (2018) and thus represent realistic values. The data was simulated with Gillespie's algorithm over a time interval of 30 hours which yields trajectories similar to the experimental data, and we use the fluorescence intensity at 181 equidistant time points as observations to mirror the structure of the experimental data. The simulated fluorescence intensity (without measurement error) is depicted by the blue dotted line on the right hand side of Fig. 5. We use Stan to sample from the posterior distributions of the ODE model and the SDE model given the simulated data. Since we assume the data to be observed without measurement error, the parameter offset can be determined directly from the first observation. Therefore, we do not include measurement error (and thus the parameter  $\sigma$ ) and the parameter offset in the posterior distribution of the SDE model. Whereas for the ODE model, deviations of the observed data from the



deterministic ODE trajectory have to be attributed to measurement error; therefore, the parameter  $\sigma$  has to be included in the posterior distribution of the ODE model. We also include the parameter offset for the ODE model in order to avoid degeneracy of the posterior. For the SDE model, we use the mean estimate of  $t_0$  obtained from the sample for the ODE model as the fixed value of  $t_0$ . We assume the following prior distributions:  $\theta_i \sim \mathcal{N}_{\geq 0}(0, 5^2)$  for  $i = 1, 2, 3$ ,  $m_0 \sim \mathcal{N}_{\geq 0}(300, 300^2)$ ,  $\text{scale} \sim \mathcal{U}(0, 30)$ , where  $\mathcal{N}_{\geq a}(\mu, \eta^2)$  denotes the normal distribution truncated from below by  $a$ , and additionally for the ODE model,  $\text{offset} \sim \mathcal{U}(0, 30)$ ,  $\sigma \sim \mathcal{U}(0.001, 10)$ , and  $t_0 \sim \mathcal{U}(0, 30)$ . These diffuse priors reflect our little prior knowledge.

For both model types, we generate 8 HMC chains of 5000 iterations and discard the first half of the iterations as warm-up. Thus, we use a posterior sample of size 20,000 for each model type in the subsequent analysis. Tables 1 and 2 summarize the Stan output of the posterior samples for the ODE and the SDE model, respectively, and also include the true parameter values that were used to simulate the data for comparison. The tables also contain the 2.5%-, 50%-, and 97.5%-quantiles of the samples. We use the interval between the 2.5%- and the 97.5%-quantile as an estimate of the 95%-credible interval (CI). This choice of CI will lead to one wide interval in case of a bimodal density rather than the union of two narrow intervals as it would be the case for a highest posterior density region. This is in line with relating the width of a CI to parameter identifiability and becomes relevant for  $\theta_1$  and  $\theta_3$  in the ODE case, see Fig. 6. For the ODE model, we see that the parameters offset and  $t_0$  are well estimated since mean and median of the sample correspond to the true value, the CIs are very narrow, the effective sample size (ESS)  $n_{\text{eff}}$  is high and  $\hat{R}$  is equal to 1. As expected, the measurement error parameter  $\sigma$  is estimated to be higher than the true value of zero. Of greater interest are the remaining parameters as we can compare the results for them between the two model types.

We first focus on the two degradation rate constants  $\theta_1$  and  $\theta_3$ . Our analysis in Sect. 5 already showed that for the ODE model, these two parameters are only locally identifiable and the likelihood function landscape is symmetric with respect to them in the case of identical priors for both parameters. This symmetry is for the rather uninformative prior distributions also preserved in the posterior distribution and apparent in the density plots in Fig. 6. The density estimates of the posterior sample for the ODE model are clearly bimodal. The reason that the two modes are not exactly symmetric here is that HMC chains usually are only able to explore one mode and in our example 5 out of the 8 chains happen to end up in the mode where  $\theta_1$  is higher than  $\theta_3$  while only 3 chains converge to the other mode. For sampling methods which facilitate switching between modes, e.g. adaptive parallel tempering, this was not observed (results not shown). The fact that each HMC chain only samples from one of the modes is also the reason for the extremely low ESSs and the very high values of  $\hat{R}$  for  $\theta_1$  and  $\theta_3$  in Table 1. Moreover, note that neither of the modes and not even the ranges of all values in the posterior sample cover the true parameter values of  $\theta_1$  and  $\theta_3$ .

For the SDE model on the other hand, Fig. 6 and Table 2 show that the posterior density is clearly unimodal with respect to  $\theta_1$  and  $\theta_3$ , and mean and median of the sample are close or equal to the true values. High ESSs and  $\hat{R}$  values equal to 1 are achieved, from which we derive reliability of the 95% CI. These are narrow and cover

**Table 1** Summary of the Stan output for the ODE model given simulated data without measurement error; the true parameter values that were used to simulate the data; and the prior distributions used for the Stan sampling, where  $\mathcal{N}_{\geq 0}(\mu, \eta^2)$  denotes the normal distribution truncated from below by  $a$  and  $\mathcal{U}(l, u)$  denotes the uniform distribution

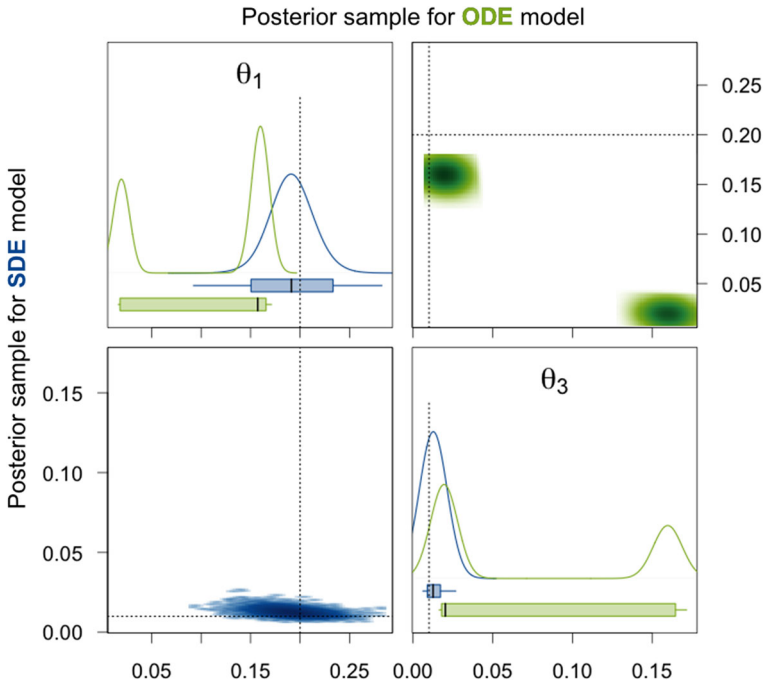
	True value	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	0.20	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.11	0.634	0.02	0.16	0.17	4	26.65
$\theta_2$	0.32	$\mathcal{N}_{\geq 0}(0, 5^2)$	1.52	1.370	0.02	0.64	7.56	12168	1.00
$\theta_3$	0.01	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.07	0.943	0.02	0.02	0.16	4	33.00
$m_0$	240.00	$\mathcal{N}_{\geq 0}(300, 300^2)$	204.62	1.017	2.26	135.79	724.38	10984	1.00
scale	1.80	$\mathcal{U}(0, 30)$	7.02	1.137	0.07	3.46	27.28	9806	1.00
offset	6.50	$\mathcal{U}(0, 30)$	6.50	0.011	6.37	6.50	6.64	17113	1.00
$t_0$	0.96	$\mathcal{U}(0, 30)$	0.96	0.002	0.96	0.96	0.97	18718	1.00
$\sigma$	0.00	$\mathcal{U}(0.001, 10)$	0.03	0.054	0.02	0.03	0.03	16091	1.00
$\theta_2 m_0$	76.80	-	213.08	2.487	4.57	35.99	1668.06	11087	1.00
$\theta_2 \text{scale}$	0.58	-	6.46	2.875	0.17	0.92	55.00	7704	1.00
$m_0 \text{scale}$	432.00	-	1033.30	2.181	16.48	195.96	7975.22	7899	1.00
$\theta_2 m_0 \text{scale}$	138.24	-	124.67	0.007	122.96	124.66	126.38	13299	1.00

Moreover, c.v. denotes the coefficient of variation and the columns headed by percentages contain the quantiles of the respective percentage value for the posterior sample

**Table 2** Summary of the Stan output for the SDE model given simulated data without measurement error, the true parameter values that were used to simulate the data; and the prior distributions used for the Stan sampling

	True value	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	0.20	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.19	0.108	0.15	0.19	0.23	3120	1.00
$\theta_2$	0.32	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.39	0.999	0.09	0.26	1.48	206	1.04
$\theta_3$	0.01	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.01	0.167	0.01	0.01	0.02	2514	1.00
$m_0$	240.00	$\mathcal{N}_{\geq 0}(300, 300^2)$	344.37	0.589	57.21	313.25	800.67	184	1.05
scale	1.80	$\mathcal{U}(0, 30)$	1.66	0.172	1.19	1.62	2.30	3062	1.00
$\theta_2 m_0$	76.80	-	82.60	0.178	55.68	81.69	113.53	5296	1.00
$\theta_2 \text{scale}$	0.58	-	0.61	0.923	0.17	0.43	2.21	178	1.05
$m_0 \text{scale}$	432.00	-	576.69	0.634	86.08	511.66	1440.41	232	1.04
$\theta_2 m_0 \text{scale}$	138.24	-	133.18	0.083	112.47	132.74	156.40	5829	1.00

The parameters  $\sigma$ , offset, and  $t_0$  are not estimated here. For the simulated data without measurement error, no measurement error is assumed for the SDE model. Thus, there is no  $\sigma$  to be estimated, and the offset is determined directly from the data and not estimated. Moreover, the initial time point  $t_0$  is predetermined based on the mean estimate of the sample for the ODE model



**Fig. 6** Density estimates of the posterior samples for parameters  $\theta_1$  and  $\theta_3$  for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given simulated data without measurement error. *Diagonal panels:* Marginal densities for the respective parameter and boxplots showing the 95% CI as box, the range of the sample as whiskers, and the median as thick black line. *Off-diagonal panels:* Smoothed scatter plots of the two-dimensional projections of the samples where darker hues signify higher density values. The dotted lines represent the true parameter values that were used to simulate the data

the true parameter values. Thus, we can conclude that the parameters  $\theta_1$  and  $\theta_3$  are identifiable for the SDE model here.

Next, we consider the translation rate constant  $\theta_2$ , the initial amount  $m_0$  of mRNA molecules, and the factor scale. For the ODE, at most the product  $\theta_2 m_0 \text{scale}$  is identifiable. This is also apparent from the results presented in Table 1 and Fig. 7. For the individual parameters and also for all products of two out of the three parameters, the 95% CIs are extremely broad and the mean and median as point estimates are not at all close to the true values. The reason why there are nevertheless quite high ESSs and  $\hat{R}$  values equal to 1 achieved is that the variation within each of the HMC chains is very high and thus does not differ substantially from the variation between the chains for these parameters as can be seen in the traceplots of the chains in Sect. A.4.1 of the supplementary material. For the product  $\theta_2 m_0 \text{scale}$  of all three parameters, the 95% CI is very narrow for the posterior sample of the ODE model and also the ESS is high and  $\hat{R}$  equal to 1. Without knowing the true parameter values, one would assume that this product is well estimated. However, the 95% CI and even the whole range of the sample do not cover the true value. For the SDE model, the 95% CI for the product  $\theta_2 m_0 \text{scale}$  is broader, but it covers the true parameter value and also the mean and

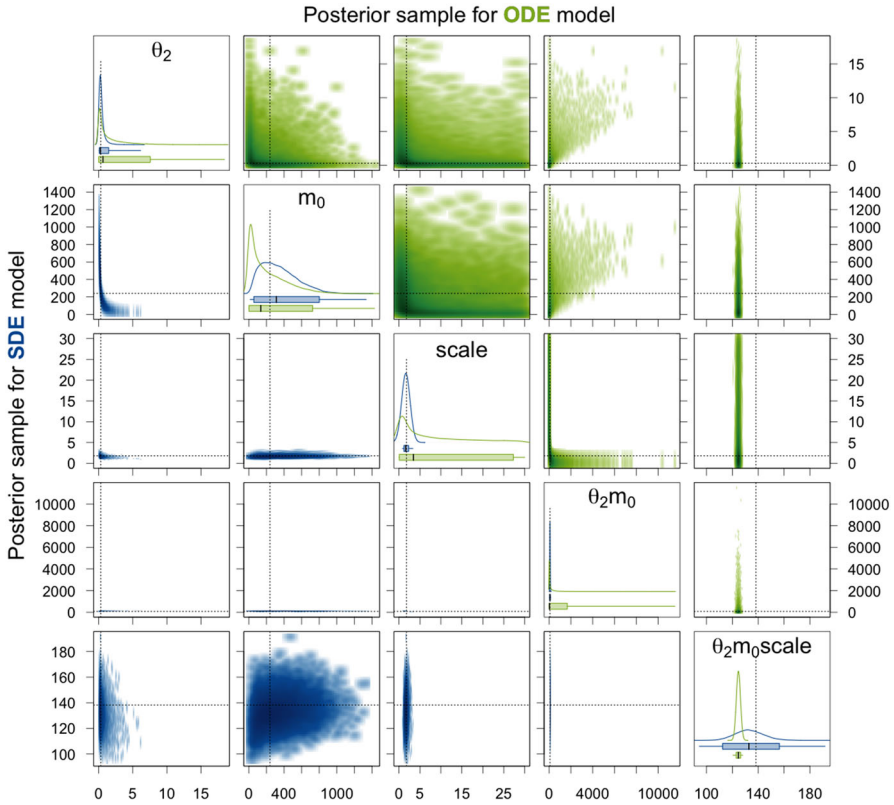


Fig. 7 Density estimates of the posterior samples for parameters  $\theta_2$ ,  $m_0$ , scale, and their products for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given simulated data without measurement error. For a detailed description of the figure’s elements, see Fig. 6

the median as point estimates are closer to the true value than the mean and the median for the ODE model. Moreover, the ESS is quite high and  $\hat{R}$  is equal to 1 for the SDE model. We therefore conclude that the product  $\theta_2 m_0 \text{scale}$  is identifiable. The generally lower ESSs are due to the fact that for the SDE model, we sample from a distribution of much larger dimension as explained in Sect. 6.1. Additionally, for the SDE model, the parameters scale and  $\theta_2 m_0$  have narrow 95% CIs (especially compared to those for the ODE model) that include the true parameter values, high ESSs, and  $\hat{R}$  values of 1 and can therefore be considered identifiable. The remaining parameters  $\theta_2$ ,  $m_0$ ,  $\theta_2 \text{scale}$ , and  $m_0 \text{scale}$  have rather broad 95% CIs and only achieve low ESSs and  $\hat{R}$  values higher than 1.02. Hence, they seem to be non-identifiable. Notice, however, that at least for the parameters  $\theta_2$  and  $\theta_2 \text{scale}$ , the 95% CIs are substantially more narrow for the SDE model compared to the ODE model.

We have simulated another 99 trajectories with the same parameters and performed Stan sampling in the same way as described in the beginning of this subsection. For each model type and each posterior sample of the different simulated trajectories, we calculate the length of the 95% CI and determine the median and the coefficient of

variation (c.v.) over these lengths for each model type. Also, we rescale the lengths of the 95% CI by dividing by the true parameter value and again determine the median of the normalized quantities. The rescaling is done to transfer the values to a more similar scale. Note, however, that the values are nevertheless not directly comparable between different parameters. Moreover, we check whether the true parameter value that was used to simulate the data is included in the 95% CI. Table 3 shows the aggregated results for the posterior samples of all 100 trajectories and also includes the length of the interval between the 2.5%- and the 97.5%-quantile of the prior distributions. Except for the parameters  $m_0$  and  $\theta_2 m_0 \text{scale}$ , the median length of the 95% CIs for the SDE model is always smaller than for the ODE model. For parameter  $\theta_2 m_0 \text{scale}$ , the CI lengths are a lot smaller for the ODE model; however, the CIs cover the true parameter value only 13 out of 100 times while for the SDE model, the true value is covered 93 times. For the other parameters that we classified as identifiable for the SDE model in the analysis of the individual trajectory (i.e.  $\theta_1$ ,  $\theta_3$ ,  $\text{scale}$ , and  $\theta_2 m_0$ ), the median length of the 95% CIs is clearly smaller for the SDE model than for the ODE model and the true parameter value is covered at least 91 out of 100 times for the SDE model. For parameter  $m_0$ , the CI lengths are high for both model types because the parameter is not identifiable for either model type. For the other parameters that we classified as not identifiable for both model types in the analysis of the individual trajectory (i.e.  $\theta_2$ ,  $\theta_2 \text{scale}$ , and  $m_0 \text{scale}$ ), the median length of 95% CIs is clearly smaller for the SDE model than for the ODE model, at least by a factor of 4.

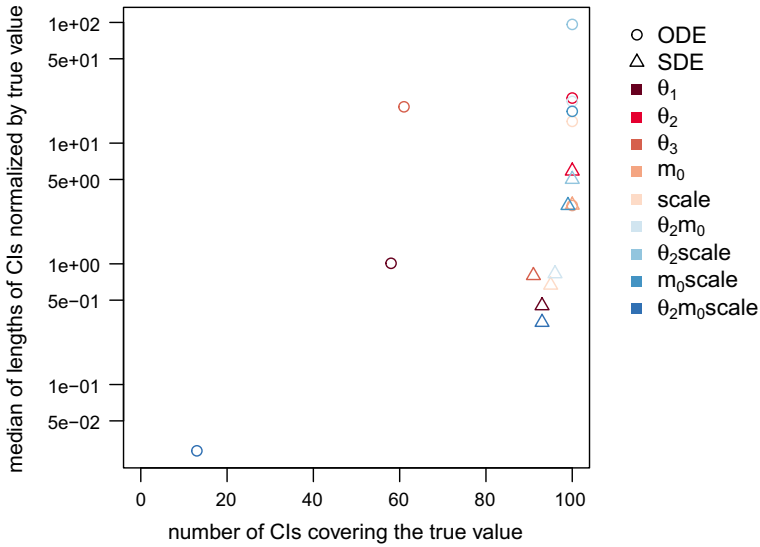
The last two columns of Table 3 are visualized in Fig. 8 where we plot the median of the rescaled CI lengths against the number of CIs that cover the true parameter value. The desirable region of value combinations is in the bottom right corner of the graph where the number of CIs covering the true value is high and the median rescaled CI length is small. Note that, clearly, more importance should be given to high numbers of CIs covering the true value as it is useless to be very certain about a parameter estimate (indicated by a short CI) while the correct value is not included in the CI. However, even for parameters that are identifiable, we do not expect to obtain a coverage of the true value of 100% since we are considering 95% CIs. Therefore, values of 100 rather tend to hint at non-identifiability. In Fig. 8, we can see that for the majority of the parameters, the triangles representing the value combinations for the SDE model are closer to the desirable region. Only for parameter  $m_0$  (which is not identifiable for either model type), the value combinations are almost the same for both model types. And as we already pointed out for the product  $\theta_2 m_0 \text{scale}$ , the median length of the 95% CIs is smaller for the ODE model; however, a lot fewer CIs cover the true parameter value for the ODE model than for the SDE model. Thus, the result obtained for the SDE model is to be preferred.

We provide further Stan-specific diagnostics in Appendix A.5. Those mostly show poorer values for the sampling output for the SDE model than for the ODE model. This is not surprising as we sample from a much higher-dimensional distribution for the SDE model. We do not consider the poor diagnostics as a disadvantage of the procedure as they provide information that we do not even have for other MCMC algorithms and thus cannot compare to them.

**Table 3** Statistics of posterior samples for the two model types aggregated over 100 simulated trajectories without measurement error

	Prior		Posterior		Median of length of CIs rescaled by true value	Number of CIs covering true value
	Length of prior 95% center interval	Length of prior 95% interval	Median length of 95% CIs	c.v. of lengths of 95% CIs		
$\theta_1$	ODE	11.05	0.20	0.009	1.01	58
	SDE	11.05	0.09	0.002	0.45	93
$\theta_2$	ODE	11.05	7.56	0.002	23.63	100
	SDE	11.05	1.88	0.712	5.87	100
$\theta_3$	ODE	11.05	0.20	0.010	20.00	61
	SDE	11.05	0.01	0.000	0.80	91
$m_0$	ODE	884.82	730.41	0.057	3.04	100
	SDE	884.82	735.98	9.868	3.07	100
scale	ODE	28.50	27.27	0.001	15.15	100
	SDE	28.50	1.20	0.158	0.67	95
$\theta_2 m_0$	ODE	6056.48	1701.92	2.524	22.16	100
	SDE	6056.48	63.60	2.768	0.83	96
$\theta_2$ scale	ODE	228.08	55.47	0.164	96.29	100
	SDE	228.08	2.89	0.762	5.01	100
$m_0$ scale	ODE	19271.13	7923.38	8.603	18.34	100
	SDE	19271.13	1315.63	86.806	3.05	99
$\theta_2 m_0$ scale	ODE	113232.70	3.90	41.595	0.03	13
	SDE	113232.70	45.28	0.624	0.33	93

We also include the length of the interval between the 2.5%- and the 97.5%-quantile of the prior distribution in order to compare it to the length of the 95% CIs and thus to assess whether the posterior distribution contains more information about the parameters' values



**Fig. 8** Statistics of posterior samples for the two model types aggregated over 100 simulated trajectories without measurement error. The desirable region of value combinations is in the bottom right corner of the graph

### 6.2.2 Sampling results for simulated data with measurement error

In this section, we use the same simulated data as in the previous section, but for each of the 100 trajectories, we add multiplicative measurement error with parameter  $\sigma = 0.02$ . Again, we use Stan to sample from the posterior distributions of the ODE model (6) and the SDE model (7) for each of the simulated trajectories and use the same priors as stated in the previous section. We generate 8 HMC chains of 5000 iterations, discard the first half of the iterations as warm-up, and thus use a posterior samples of size 20,000 in the subsequent analysis.

At first, we again focus on the results for one of the trajectories, namely the trajectory represented by the green dotted line in Fig. 5. Tables 4 and 5 summarize the Stan output of the posterior samples for the ODE and the SDE model, respectively. The parameter  $t_0$  is estimated very accurately based on the posterior sample for the ODE model. Also, the parameter offset is well estimated for both model types but with a more narrow 95% CI for the SDE model. The parameter  $\sigma$  is accurately determined for the SDE model as well. For the ODE model,  $\sigma$  is again overestimated. Figure 9 visualizes the components of the posterior samples for parameters  $\theta_1$ ,  $\theta_3$ , offset, and  $\sigma$ . Again, the bimodality of the posterior with respect to  $\theta_1$  and  $\theta_3$  is apparent for the ODE model and neither the 95% CIs nor the ranges of the sample cover the true parameter values. For the SDE model on the other hand, the distribution is unimodal and the 95% CIs do cover the true parameter values for  $\theta_1$  and  $\theta_3$ . However, their 2-dimensional smoothed scatter plot in Fig. 9 does not show a simple elliptic shape (as for the simulated data without measurement error) but almost a banana-like shape. This may also be the



**Table 4** Summary of the Stan output for the ODE model given simulated data with measurement error, the true parameter values that were used to simulate the data, and the prior distributions used for the Stan sampling

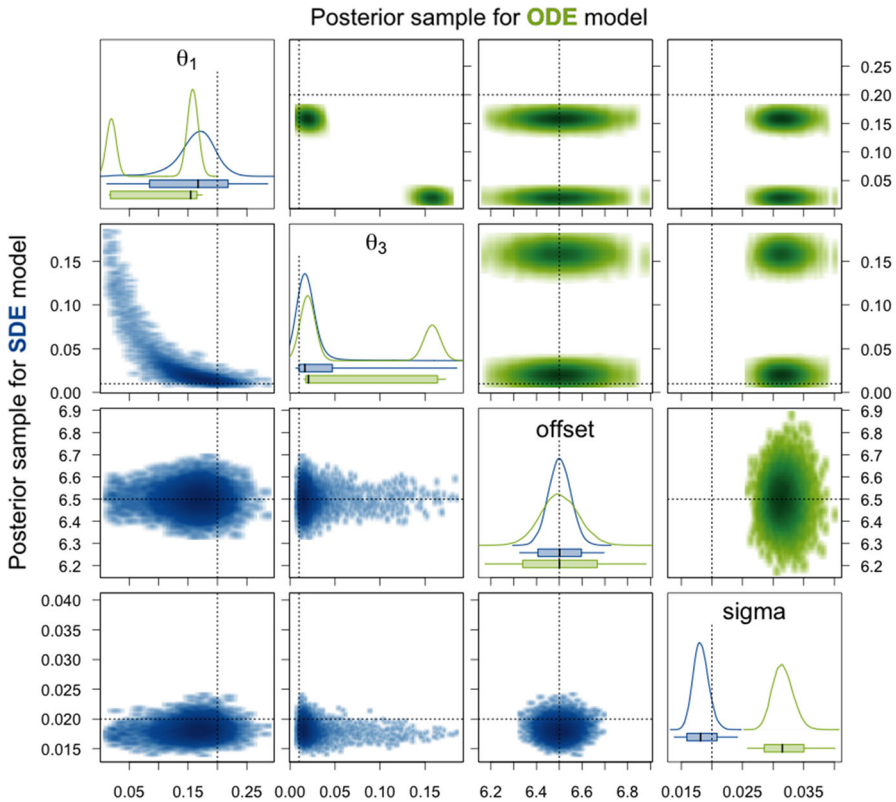
	True value	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	0.20	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.11	0.632	0.02	0.15	0.17	4	20.94
$\theta_2$	0.32	$\mathcal{N}_{\geq 0}(0, 5^2)$	1.54	1.364	0.02	0.65	7.63	11619	1.00
$\theta_3$	0.01	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.938	0.02	0.02	0.16	4	26.81	1.00
$m_0$	240.00	$\mathcal{N}_{\geq 0}(300, 300^2)$	205.41	1.024	2.25	133.96	738.49	10649	1.00
scale	1.80	$\mathcal{U}(0, 30)$	6.84	1.152	0.07	3.29	27.06	8778	1.00
offset	6.50	$\mathcal{U}(0, 30)$	0.013	6.34	6.50	6.67	14496	1.00	1.00
$t_0$	0.96	$\mathcal{U}(0, 30)$	0.96	0.003	0.96	0.96	0.97	17129	1.00
$\sigma$	0.00	$\mathcal{U}(0.001, 10)$	0.03	0.053	0.03	0.03	0.04	13581	1.00
$\theta_2 m_0$	76.80	-	217.59	2.441	4.56	37.40	1683.06	9479	1.00
$\theta_2 \text{scale}$	0.58	-	6.31	2.841	0.17	0.92	54.94	7209	1.00
$m_0 \text{scale}$	432.00	-	983.07	2.151	16.20	189.75	7514.28	8054	1.00
$\theta_2 m_0 \text{scale}$	138.24	-	123.47	0.009	121.43	123.46	125.58	12035	1.00

c.v. denotes the coefficient of variation and the columns headed by percentages contain the quantiles of the respective percentage value

**Table 5** Summary of the Stan output for the SDE model given simulated data with measurement error, the true parameter values that were used to simulate the data, and the prior distributions used for the Stan sampling

	True value	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	0.20	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.16	0.201	0.08	0.17	0.22	304	1.03
$\theta_2$	0.32	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.73	1.448	0.07	0.37	3.93	296	1.02
$\theta_3$	0.01	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.02	0.660	0.01	0.02	0.05	176	1.04
$m_0$	240.00	$\mathcal{N}_{\geq 0}(300, 300^2)$	274.72	0.742	28.30	224.46	777.96	225	1.04
scale	1.80	$\mathcal{U}(0, 30)$	1.67	0.430	0.65	1.54	3.36	415	1.02
offset	6.50	$\mathcal{U}(0, 30)$	6.50	0.007	6.41	6.50	6.60	15599	1.00
$\sigma$	0.00	$\mathcal{U}(0.001, 10)$	0.02	0.069	0.02	0.02	0.02	2106	1.00
$\theta_2 m_0$	76.80	-	89.78	0.473	35.58	80.76	191.24	349	1.02
$\theta_2 \text{scale}$	0.58	-	0.93	1.145	0.16	0.56	4.15	247	1.03
$m_0 \text{scale}$	432.00	-	477.62	0.944	29.55	338.17	1697.86	216	1.04
$\theta_2 m_0 \text{scale}$	138.24	-	124.39	0.093	102.85	123.92	148.63	2119	1.01

The initial time point  $t_0$  is not estimated here, but predetermined based on the mean estimate of the sample for the ODE model

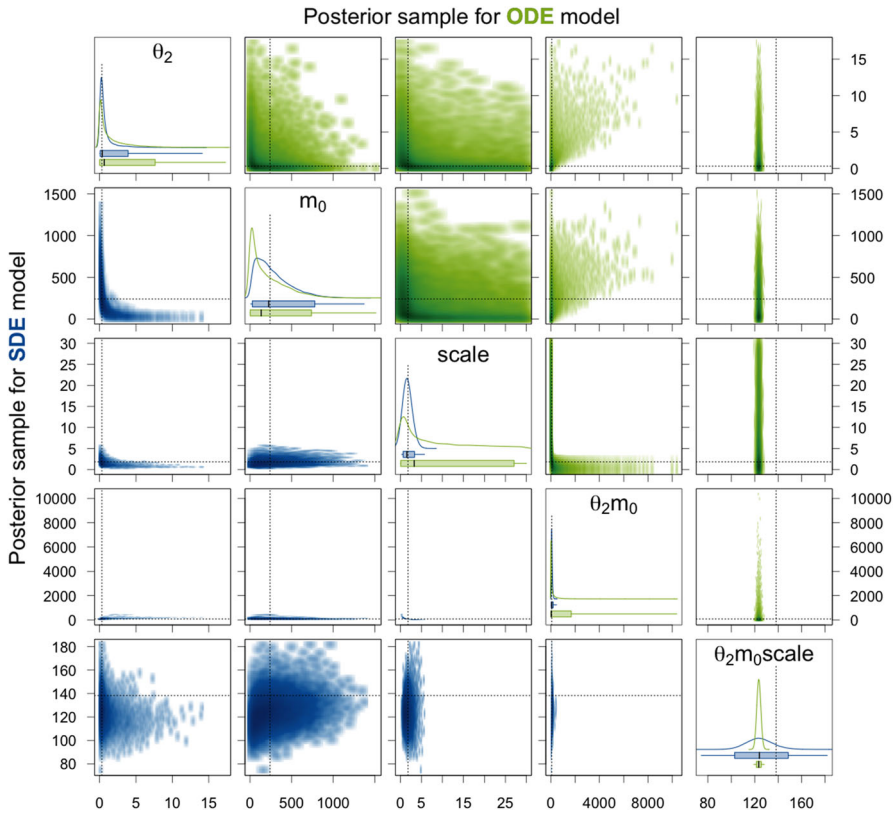


**Fig. 9** Density estimates of the posterior samples for parameters  $\theta_1$ ,  $\theta_3$ , offset, and  $\sigma$  for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given simulated data with measurement error. For a detailed description of the figure's elements, see Figure 6

reason for the deteriorated sampling efficiency discernible from the low ESSs and higher  $\hat{R}$ -values in Table 5.

Figure 10 visualizes the components of the posterior samples for parameters  $\theta_2$ ,  $m_0$ , scale, and their products. For the ODE model, again only the product  $\theta_2 m_0 \text{scale}$  is identifiable in the sense that the corresponding 95% CI is very narrow, the ESS is high, and the  $\hat{R}$ -value is equal to 1. However, the 95% CI again does not cover the true parameter value. For the SDE model, the 95% CI for  $\theta_2 m_0 \text{scale}$  is broader but it does contain the true value. Also, the ESS is high and the  $\hat{R}$ -value is close to 1. Moreover, the parameters scale and  $\theta_2 m_0$  have narrow 95% CIs, high ESSs, and  $\hat{R}$ -values close to 1 for the SDE model, and thus, we conclude that they are identifiable. Note that also for  $\theta_2$ ,  $m_0 \text{scale}$ , and  $\theta_2 \text{scale}$ , the 95% CIs are much narrower for the SDE model than for the ODE model (Fig. 10).

In Sect. A.4.1 of the supplementary material, we also include trace plots and further figures of the sampling output for the trajectory displayed in Fig. 5. There, we present the same posterior samples as used in this and the previous subsection. But instead of comparing the posterior samples between the two model types, the poste-



**Fig. 10** Density estimates of the posterior samples for parameters  $\theta_2$ ,  $m_0$ ,  $scale$ , and their products for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given simulated data with measurement error. For a detailed description of the figure’s elements, see Fig. 6

rior samples are compared between the simulated data without and with measurement error for each model type separately. In summary, we find that for the SDE model, the 95% CIs increase for almost all parameters except  $m_0$  for data with measurement error. Whereas for the ODE model, there is hardly any difference for most of the parameters between the posterior samples for the data without and with measurement error since the majority of the parameters is not identifiable anyway. The marginal posterior samples for the parameters offset,  $t_0$ , and  $\theta_2 m_0 scale$  are only slightly affected by the measurement error. Only the marginal posterior sample of the measurement error parameter  $\sigma$  is substantially affected and, as expected, consists of higher values for data with measurement error.

Table 6 and Fig. 11 display the statistics of the posterior samples aggregated over the 100 simulated trajectories. Similar to the results for the simulated data without measurement error, the median length of the 95% CIs for the SDE model is always smaller than for the ODE model, except for the parameters  $m_0$  and  $\theta_2 m_0 scale$  and additionally  $\sigma$  (which was not included for the SDE model in the previous subsection). Again, for the majority of the parameters, the results for the SDE model represented

**Table 6** Statistics of posterior samples for the two model types aggregated over 100 simulated trajectories with measurement error

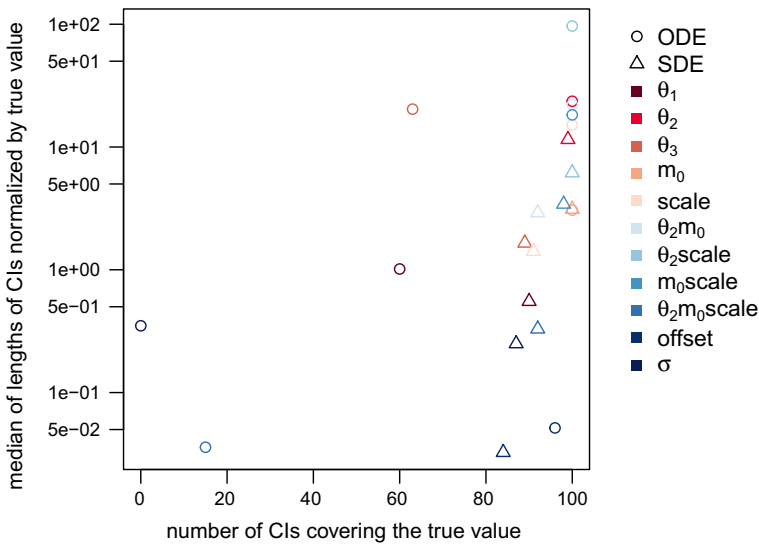
		Prior	Posterior			
		Length of prior 95% center interval	Median length of 95% CIs	c.v. of lengths of 95% CIs	Median of length of CIs rescaled by true value	Number of CIs covering true value
$\theta_1$	ODE	11.05	0.20	0.008	1.01	60
	SDE	11.05	0.11	0.016	0.55	90
$\theta_2$	ODE	11.05	7.55	0.002	23.59	100
	SDE	11.05	3.69	0.996	11.52	99
$\theta_3$	ODE	11.05	0.20	0.008	20.35	63
	SDE	11.05	0.02	0.094	1.65	89
$m_0$	ODE	884.82	733.85	3.584	3.06	100
	SDE	884.82	746.74	5.909	3.11	100
scale	ODE	28.50	27.25	0.002	15.14	100
	SDE	28.50	2.54	0.255	1.41	91
$\theta_2 m_0$	ODE	6056.48	1702.37	2.714	22.17	100
	SDE	6056.48	223.22	428.350	2.91	92
$\theta_2 scale$	ODE	228.08	55.70	0.125	96.70	100
	SDE	228.08	3.55	0.639	6.16	100
$m_0 scale$	ODE	19271.13	7896.07	667.494	18.28	100
	SDE	19271.13	1479.13	253.530	3.42	98
$\theta_2 m_0 scale$	ODE	113232.70	4.96	38.038	0.04	15
	SDE	113232.70	45.56	1.492	0.33	92
Offset	ODE	28.50	0.33	0.987	0.05	96
	SDE	28.50	0.21	0.016	0.03	84
$\sigma$	ODE	9.50	0.01	0.325	0.35	0
	SDE	9.50	0.01	0.000	0.25	87

We also include the length of the interval between the 2.5%- and the 97.5%-quantile of the prior distribution

by triangles in Fig. 11 are closer to the desirable region of value combinations in the bottom right corner of the graph, except for the parameters  $m_0$ ,  $\theta_2 m_0 scale$ , and offset. For the parameter offset, the median CI length is slightly higher for the ODE model than for the SDE model, however, the CIs for the ODE model also contain the true parameter value more often. So for this parameter, the ODE model, for once, shows the preferable result.

### 6.3 Estimation based on experimental data

Since our results in the previous section have shown that the SDE model yields more reliable parameter estimates (assuming an MJP is an adequate description for the translation kinetics after mRNA transfection) than the ODE model even for those



**Fig. 11** Statistics of posterior samples for the two model types aggregated over 100 simulated trajectories with measurement error. The desirable region of value combinations is in the bottom right corner of the graph

parameters that are identifiable for both model types, we reanalyze the experimental data published in Fröhlich et al. (2018) and described in Sect. 2. For each type of GFP (eGFP and d2eGFP), we randomly select 100 observed trajectories for our analysis and again use Stan to sample from the posterior distributions of the ODE model (6) and the SDE model (7) for each of the trajectories using the same priors as stated in Sect. 6.2.1. We generate 8 HMC chains of 5000 iterations, discard the first half of the iterations as warm-up, and thus use a posterior samples of size 20,000 in the subsequent analysis. For each type of GFP, we first analyze the sampling output for one observed trajectory in detail and then summarize results for all 100 observed trajectories. Here in the main part of the article, we only show the results for the eGFP data and provide the results for the d2eGFP data in Appendix A.4.3. The comparison between the eGFP and the d2eGFP trajectories demonstrate how practical identifiability may differ between seemingly similar datasets. Moreover, we provide further Stan-specific diagnostics in Appendix A.5.

### Sampling results for experimental dataset 1 (for eGFP)

Tables 7 and 8 present a summary of the Stan output for the posterior sample of one observed trajectory for the ODE and the SDE model, respectively, and Figs. 12 and 13 compare the density estimates of these two posterior samples. Moreover, we provide trace plots for these two posterior samples in Section A.4.2 in the supplementary material. The results look qualitatively *very* similar (almost identical) to those obtained for the simulated data with measurement error in Sect. 6.2.2. Therefore, we do not repeat the detailed description but only point out that the range of values sampled for

**Table 7** Summary of the Stan output for the ODE model given experimental data for eGFP and the prior distributions used for the sampling. c.v. denotes the coefficient of variation and the columns headed by percentages contain the quantiles of the respective percentage value

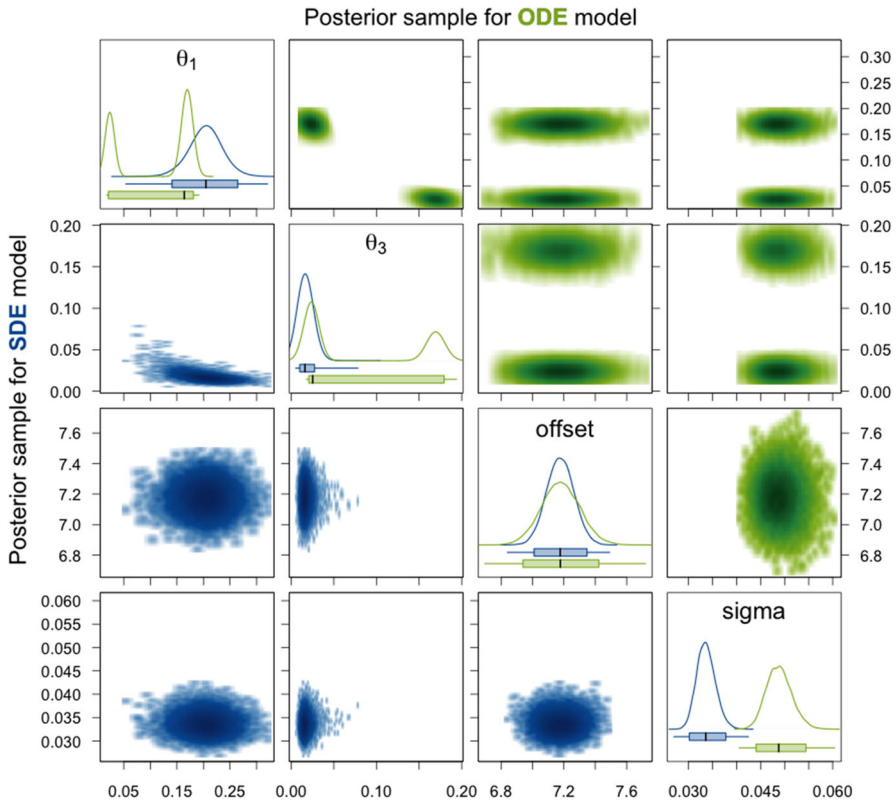
	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.11	0.617	0.02	0.16	0.18	4	13.60
$\theta_2$	$\mathcal{N}_{\geq 0}(0, 5^2)$	1.44	1.407	0.01	0.56	7.39	12661	1.00
$\theta_3$	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.08	0.903	0.02	0.03	0.18	4	16.98
$m_0$	$\mathcal{N}_{\geq 0}(300, 300^2)$	198.99	1.046	1.75	127.89	731.84	12008	1.00
scale	$\mathcal{U}(0, 30)$	6.64	1.186	0.05	3.01	27.00	10321	1.00
offset	$\mathcal{U}(0, 30)$	7.18	0.017	6.94	7.18	7.42	17800	1.00
$t_0$	$\mathcal{U}(0, 30)$	1.46	0.004	1.44	1.46	1.47	15996	1.00
$\sigma$	$\mathcal{U}(0.001, 10)$	0.05	0.054	0.04	0.05	0.05	16833	1.00
$\theta_2 m_0$	–	200.04	2.654	3.17	28.49	1627.75	11121	1.00
$\theta_2 \text{scale}$	–	5.53	3.043	0.12	0.67	48.94	8535	1.00
$m_0 \text{scale}$	–	942.49	2.307	11.63	153.30	7610.72	9270	1.00
$\theta_2 m_0 \text{scale}$	–	85.74	0.014	83.35	85.73	88.19	12577	1.00

**Table 8** Summary of the Stan output for the SDE model given experimental data for eGFP and the prior distributions used for the sampling. The initial time point  $t_0$  is not estimated here, but predetermined based on the mean estimate of the sample for the ODE model

	Prior distribution	Mean	c.v.	2.5%	50%	97.5%	$n_{\text{eff}}$	$\hat{R}$
$\theta_1$	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.20	0.152	0.14	0.20	0.26	946	1.01
$\theta_2$	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.33	1.366	0.04	0.19	1.52	305	1.02
$\theta_3$	$\mathcal{N}_{\geq 0}(0, 5^2)$	0.02	0.269	0.01	0.02	0.03	894	1.01
$m_0$	$\mathcal{N}_{\geq 0}(300, 300^2)$	298.35	0.705	34.46	250.62	809.42	218	1.03
scale	$\mathcal{U}(0, 30)$	2.12	0.309	1.14	2.02	3.69	772	1.01
offset	$\mathcal{U}(0, 30)$	7.18	0.012	7.01	7.18	7.35	20589	1.00
$\sigma$	$\mathcal{U}(0.001, 10)$	0.03	0.058	0.03	0.03	0.04	17224	1.00
$\theta_2 m_0$	–	47.92	0.327	23.41	46.00	83.31	857	1.01
$\theta_2 \text{scale}$	–	0.60	1.182	0.11	0.37	2.54	256	1.02
$m_0 \text{scale}$	–	650.64	0.831	57.98	498.32	2049.93	283	1.02
$\theta_2 m_0 \text{scale}$	–	92.71	0.115	73.32	92.15	115.16	2711	1.00

the parameters  $\theta_1$  and  $\theta_3$  for the SDE model is slightly smaller for the experimental trajectory here. Thus, we do not see the banana-like shape in the two-dimensional smoothed scatter plot of the two parameters for the SDE model in Fig. 12 as for the simulated trajectory in Fig. 9 and the sampling efficiency increases as indicated by higher ESSs and lower  $\hat{R}$  values for the two parameters in Table 8.

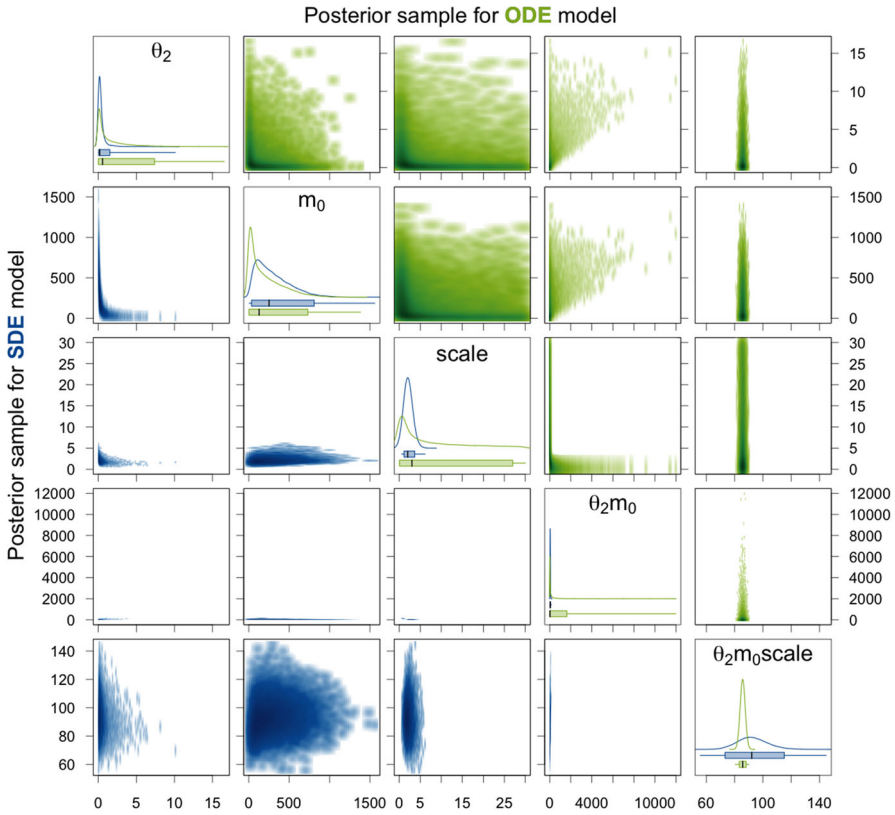
The statistics of posterior samples aggregated for 100 experimental trajectories for eGFP in Table 9 are also qualitatively similar to those for the simulated trajectories in Table 6. For the majority of the parameters, the median length of the 95% CI is



**Fig. 12** Density estimates of the posterior samples for parameters  $\theta_1$ ,  $\theta_3$ , offset, and  $\sigma$  for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given experimental data for eGFP. *Diagonal panels:* Marginal densities for the respective parameter and boxplots showing the 95% CI as box, the range of the sample as whiskers, and the median as thick black line. *Off-diagonal panels:* Smoothed scatter plots of the two-dimensional projections of the samples where darker hues signify higher density values

smaller for the posterior samples for the SDE model than for those for the ODE model. Only for parameters  $\theta_1$ ,  $\theta_2$ , and  $\theta_2 m_0 \text{scale}$ , this is not the case. Note in particular that for the parameters  $\theta_2 m_0$  and  $\text{scale}$ , which are non-identifiable for the ODE model (also apparent from the very long CIs here), the median length of the 95% CI for the SDE model is again much smaller compared to that of the ODE and to that of the prior. This indicates that these two parameters are identifiable for the SDE model also for the experimental data. That the uncertainty of the parameter estimate for  $\theta_2 m_0 \text{scale}$  is greater for the SDE than for the ODE model is consistent with our results for the simulated data. The parameter  $\theta_2$  is considered to be non-identifiable for both model types and the difference between the median CI lengths is relatively small. Finally, for parameter  $\theta_1$ , we see that the result is more or less the same as for  $\theta_3$  for the ODE model due to the symmetry of the posterior distribution with respect to these two parameters. Whereas for the SDE model there is no symmetry and there is more variance in the posterior samples with respect to  $\theta_1$  than to  $\theta_3$  (indicated by a greater median CI length). The smaller median CI length of  $\theta_1$  for the ODE model compared





**Fig. 13** Density estimates of the posterior samples for parameters  $\theta_2$ ,  $m_0$ , scale, and their products for the SDE (blue, lower triangle) and ODE (green, upper triangle) model given experimental data for eGFP. For a detailed description of the figure’s elements, see Fig. 12

to the SDE model is due to the fact that for many of the observed trajectories the values of  $\theta_1$  and  $\theta_3$  seem to be very close together. In this case, the posterior distribution of the ODE model appears to be unimodal (as the two modes overlap) and the posterior variance with respect to the two parameters is small (and equal due to the symmetry). Thus, overall this variance is smaller than the posterior variance with respect to  $\theta_1$  for the SDE model.

## 7 Discussion and conclusion

We have modeled the translation kinetics after mRNA transfection using a two-dimensional Itô diffusion process described by an SDE and compared this modeling approach to one using ODEs. We have studied the parameter identifiability for both modeling approaches for the case that we observe a fluorescence signal which we assume to be a linear transformation of the amount of protein molecules (corrupted by multiplicative measurement error). Our analysis using the structural identifiability

**Table 9** Statistics of posterior samples aggregated for 100 experimental trajectories for eGFP

		Prior	Posterior	
		Length of prior 95% center interval	Median length of 95% CIs	c.v. of length of 95% CIs
$\theta_1$	ODE	11.05	0.11	0.058
	SDE	11.05	0.14	0.012
$\theta_2$	ODE	11.05	7.91	0.016
	SDE	11.05	9.91	0.998
$\theta_3$	ODE	11.05	0.11	0.057
	SDE	11.05	0.06	0.039
$m_0$	ODE	884.82	747.71	0.212
	SDE	884.82	456.80	172.338
Scale	ODE	28.50	27.50	0.004
	SDE	28.50	4.61	5.288
$\theta_2 m_0$	ODE	6056.48	2032.46	23.287
	SDE	6056.48	230.89	219.006
$\theta_2 \text{scale}$	ODE	228.08	68.38	1.879
	SDE	228.08	22.01	33.157
$m_0 \text{scale}$	ODE	19271.13	9093.22	69.292
	SDE	19271.13	1392.46	4601.374
$\theta_2 m_0 \text{scale}$	ODE	113232.70	24.10	57.316
	SDE	113232.70	138.03	78.659
Offset	ODE	28.50	0.96	2.661
	SDE	28.50	0.38	1.086
$\sigma$	ODE	9.50	0.01	0.244
	SDE	9.50	0.01	0.004

of the moment-equations-based surrogate model and a simulation-based assessment suggest that the SDE model might lead to better parameter identifiability. The most systematic approach is the one based on the surrogate model and DAISY as suggested by Browning et al. (2020). However, it cannot help us *confirm* a difference in the parameter identifiability between the SDE and the ODE model, especially because we are interested in the parameter identifiability based on a single observed trajectory and the moment-equations-based approach assumes that we are able to observe the first and the second moment of the fluorescence signal. Even when we take into account that we have several observed trajectories available from the experiment, these do not provide information about the moments because the initial time point  $t_0$  of mRNA release is different for every trajectory and also for the other parameters, in particular for  $m_0$ , assuming that they are equal for all observed cells does not seem reasonable.

By simulating from the SDE model, we were able to assess the differences in the variation within individual trajectories for different parameter combinations. Again our results suggest that the SDE model provides better parameter identifiability. Unlike for the ODE model, the degradation rates  $\theta_1$  and  $\theta_3$  for the SDE model appear to

be structurally globally identifiable. Also, scale and the product  $\theta_2 m_0$  seem to be structurally identifiable, but the individual parameters  $\theta_2$  and  $m_0$  do not. Identifiable parameter combinations sometimes hint towards effective reparameterisations for more efficient sampling, but this was not the case here. While the simple simulation approach worked out well for the model considered here, one of its weak points is, of course, the somewhat subjective visual assessment of the variation within trajectories. A more quantitative approach to this would be to simulate a large number of trajectories (with very small time step) for every considered parameter combination, to approximate the quadratic variation for each trajectory, and then, to compare these values between individual trajectories started with the same seed for different parameter combinations and to compare also the distributions of these values for different parameter combinations. Another drawback of both simulation-based approaches is the fact that the analysis is based on a finite set of parameter combinations that can be considered; and thus, drawing general conclusions for the entire parameter space may be problematic.

Moreover, we have assessed the practical parameter identifiability for both model types by sampling from the parameter posterior distribution given simulated data with and without measurement error and the experimental data published in Fröhlich et al. (2018). Section A.4.1 provides additional sampling results for the first case. We observe that CIs and ranges of posterior samples become visibly larger in the presence of measurement error for some but not all parameters. While our focus was on inference from individual fluorescence trajectories, using SDE mixed-effect modeling as done in Wqvist et al. (2021) would be a meaningful extension, especially e. g. to obtain a common estimate for the parameter scale. We found that the parameters  $\theta_1$  and  $\theta_3$  are indeed globally identifiable for the SDE model given individual trajectories, unlike for the ODE model. And not only the product  $\theta_2 m_0$  scale but also the parameter scale and the product  $\theta_2 m_0$  are globally identifiable for the SDE model. Moreover, for the simulated data, the 95% CIs for the identifiable parameters for the SDE model covered the true parameter value adequately many times. Whereas for the ODE model, the true parameter values for the parameters  $\theta_1$ ,  $\theta_3$ , and  $\theta_2 m_0$  scale were not covered by the 95% CIs for many of the posterior samples and were sometimes not even included in the range of values in the sample. The fact that the parameters  $\theta_1$  and  $\theta_3$  can be adequately determined using the SDE modeling approach given an individual trajectory renders the multi-experiment approach with different mRNA constructs and the computationally intense hierarchical optimization algorithm used in Fröhlich et al. (2018) unnecessary in the case that the determination of these parameters is the main objective. Besides, assuming that an MJP is the most appropriate description of the underlying dynamics, we saw that the estimated parameter values for a single cell trajectory based on the ODE model cannot be trusted even when narrow 95% CIs suggest low uncertainty. This has implications for previous studies using this ODE model (e.g. (Leonhardt et al. 2014)). The ODE-based nonlinear mixed-effect model used by Fröhlich et al. (2018) might have partially resolved the problem by allowing the integration of the datasets for eGFP and d2eGFP, thereby eliminating the symmetry. Yet, as the SDE model provides overall more appropriate results for the fitting of individual trajectories, an SDE-based nonlinear mixed-effect model would probably be an even better alternative.

While the SDE model is clearly superior in terms of the information that we are able to extract from a single trajectory about the parameters that determine the dynamics of the underlying process, it has nevertheless several disadvantages. First of all, we were not able to include the estimation of the initial time point  $t_0$  of mRNA release into the Stan sampling procedure due to technical/methodological limitations. Other sampling approaches such as particle MCMC (Golightly and Wilkinson 2011) might alleviate this problem, but to our knowledge, no examples of inferring an unknown time point for SDE models have been investigated so far and would thus require further work. Another drawback of the SDE model are the higher computational costs as we need to sample from a higher-dimensional distribution (due to the random process values) than for the ODE model. For the SDE model, the sampling in our study takes on average almost 5.5 hours while for the ODE model, it averages at about 20 minutes. Note, however, that for the ODE model, the NUTS method implemented in Stan generated only chains that sample only from one of the posterior modes. Therefore, one might argue that sampling did not properly converge in this example. In general, estimation procedures for SDE models are more complex and unlike for ODE models, publicly available software tools are rare and usually not generally applicable. There is a clear need to further develop such tools for SDE models in order to harness their full potential, especially with regard to better identifiability of kinetic parameters. On the other hand, combining both modeling approaches as we have done here by first determining  $t_0$  based on the ODE model and then estimating the kinetic parameters based on the SDE model is certainly also meaningful. Overall, our study clearly demonstrates the relevance of considering different modeling approaches and of selecting the appropriate one. We expect to see stochastic models being used more frequently as increasing computing power becomes available and facilitates inference for such models.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00285-022-01739-x>.

**Author Contributions** All authors designed the study. SP implemented the used methods, carried out the analysis and drafted the manuscript. CF and JH provided supervision. All authors contributed to the final version of the manuscript and gave final approval for publication.

**Funding** Open Access funding enabled and organized by Projekt DEAL. Our research was supported by the German Federal Ministry of Education and Research under Grant Numbers 01DH17024 and 01KI20271, by the European Commission under Grant Number 10101616, by the German Research Foundation under Grant Number HA 7376/3-1 and under Germany's Excellence Strategy EXC 2047/1 - 390685813 and EXC 2151 - 390873048, by the VolkswagenStiftung under Grant Number 99 450, and by the Helmholtz Association's pilot project "Uncertainty Quantification".

**Data availability** The experimental data used in this study had previously been published with the article Fröhlich et al. (2018). The simulated data can be generated with the code referenced below.

## Declarations

**Conflict of interest** The authors declare that there are no conflicts of interest regarding the publication of this paper.

**Code availability** The source code used to produce the computational results is available on GitHub at [https://github.com/fuchslab/Translation\\_kinetics\\_after\\_transfection](https://github.com/fuchslab/Translation_kinetics_after_transfection).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ballnus B, Hug S, Hatz K, Görlitz L, Hasenauer J, Theis FJ (2017) Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Syst Biol* 11(63):63
- Bellu G, Saccomani MP, Audoly S, D'Angiò L (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput Methods Programs Biomed* 88(1):52–61
- Borah P, Deb PK, Al-Shar'i NA, Dahabiyeh LA, Venugopala KN, Singh V, Shinu P, Hussain S, Deka S, Chandrasekaran B, Jaradat DMM (2021) Perspectives on RNA vaccine candidates for COVID-19. *Front Mol Biosci* 8:30
- Browning AP, Warne DJ, Burrage K, Baker RE, Simpson MJ (2020) Identifiability analysis for stochastic differential equation models in systems biology. *J Royal Soc Interface* 17(173):20200652
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: a probabilistic programming language. *J Stat Softw Articles* 76(1):1–32
- DeFrancesco L (2020) Whither COVID-19 vaccines? *Nat Biotechnol* 38(10):1132–1145
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297(5584):1183–1186
- Finkenstädt B, Heron EA, Komorowski M, Edwards K, Tang S, Harper CV, Davis JRE, White MRH, Millar AJ, Rand DA (2008) Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics* 24(24):2901–2907
- Fröhlich F, Reiser A, Fink L, Woschée D, Ligon T, Theis FJ, Rädler JO, Hasenauer J (2018) Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *NPJ Syst Biol Appl* 4(1):42
- Fuchs C (2013) *Inference for diffusion processes*. Springer, Berlin Heidelberg
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian data analysis*. Chapman and Hall/CRC, Boca Raton
- Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22(4):403–434
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81(25):2340–2361
- Gillespie DT (1992) A rigorous derivation of the chemical master equation. *Phys A Stat Mech Appl* 188(1):404–425
- Gillespie DT (2000) The chemical Langevin equation. *J Chem Phys* 113(1):297–306
- Golightly A, Wilkinson DJ (2011) Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* 1(6):807–820
- Heron EA, Finkenstädt B, Rand DA (2007) Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics* 23(19):2596–2603
- Higham DJ (2008) Modeling and simulating chemical reactions. *SIAM Rev* 50(2):347–368
- Komorowski M, Finkenstädt B, Harper CV, Rand DA (2009) Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinform* 10(1):343
- Kurtz TG (1972) The relationship between stochastic and deterministic models for chemical reactions. *J Chem Phys* 57(7):2976–2978
- Lee PM (2012) *Bayesian statistics: an introduction*, 4th edn. Wiley Publishing, Hoboken

- Leonhardt C, Schwake G, Stögbauer TR, Rappl S, Kuhr J-T, Ligon TS, Rädler JO (2014) Single-cell mRNA transfection studies: delivery, kinetics and statistics by numbers. *Nanomedicine* 10(4):679–688
- Ligon TS, Leonhardt C, Rädler JO (2014) Multi-level kinetic model of mRNA delivery via transfection of lipoplexes. *PLoS One* 9(9):e107148
- Munsky B, Trinh B, Khammash M (2009) Listening to the noise: random fluctuations reveal gene network parameters. *Mol Syst Biol* 5:318
- Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135(2):216–226
- Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25(15):1923–1929
- Reiser A, Woschée D, Mehrotra N, Krzysztoń R, Strey HH, Rädler JO (2019) Correlation of mRNA delivery timing and protein expression in lipid-based transfection. *Integr Biol* 11(9):362–371
- Sahin U, Karikó K, Türeci Ö (2014) mRNA-based therapeutics—Developing a new class of drugs. *Nat Rev Drug Discov* 13(10):759–780
- Schnoerr D, Sanguinetti G, Grima R (2017) Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J Phys A Math Theor* 50(9):093001
- Stan Development Team (2019). RStan: the R interface to Stan. R package version 2.19.1, <http://mc-stan.org/>
- Warne DJ, Baker RE, Simpson MJ (2019) Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *J Royal Soc Interface* 16(151):20180943
- Wiqvist S, Golightly A, McLean AT, Picchini U (2021) Efficient inference for stochastic differential equation mixed-effects models using correlated particle pseudo-marginal algorithms. *Comput Stat Data Anal* 157:107151

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.