Received: 29 September 2021 | Revised: 13 September 2022 | Accepted: 8 November 2022

IET Systems Biology

DOI: 10.1049/syb2.12055

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

# The effect of normalisation and error model choice on the distribution of the maximum likelihood estimator for a biochemical reaction

**Caterina Thomaseth** | **Nicole E. Radde**

University of Stuttgart, Institute for Systems Theory and Automatic Control, D-70569 Stuttgart, Germany

**Correspondence**

Nicole E. Radde, University of Stuttgart, Institute for Systems Theory and Automatic Control, Pfaffenwaldring 9, D-70569 Stuttgart, Germany.
Email: nicole.radde@ist.uni-stuttgart.de

**Funding information**

Deutsche Forschungsgemeinschaft, Grant/Award Number: EXC 2075–390740016
Open Access funding enabled and organized by Projekt DEAL.

**Abstract**

Sparse and noisy measurements make parameter estimation for biochemical reaction networks difficult and might lead to ill-posed optimisation problems. This is potentiated if the data has to be normalised, and only fold changes rather than absolute amounts are available. Here, the authors consider the propagation of measurement noise to the distribution of the maximum likelihood (ML) estimator in an in silico study. Therefore, a model of a reversible reaction is considered, for which reaction rate constants using fold changes is estimated. Noise propagation is analysed for different normalisation strategies and different error models. In particular, accuracy, precision, and asymptotic properties of the ML estimator is investigated. Results show that normalisation by the mean of a time series outperforms normalisation by a single time point in the example provided by the authors. Moreover, the error model with a heavy-tail distribution is slightly more robust to large measurement noise, but, beyond this, the choice of the error model did not have a significant impact on the estimation results provided by the authors.

## 1 | INTRODUCTION

Using chemical reaction kinetics to describe intracellular regulation and signalling processes results in systems of ordinary differential equations (ODEs). Often, kinetic parameters such as reaction rate constants are not known and have to be estimated from time course data. This is usually formulated as an inverse problem, in which an objective function that describes the discrepancy between the data and respective model predictions is optimised with respect to these unknown parameters [1, 2]. Solving this optimisation problem can be difficult, and different computational schemes and algorithms have been proposed, see for example, Degasperi et al. [3], Gábor and Banga [4], Hass et al. [5], Kreutz [6], Raue et al. [7], and Schmiester et al. [2]. In the context of intracellular process modelling, solutions are often additionally complicated by sparse and noisy data [3, 7].

Here, we consider the particular problem of parameter estimation from the relative concentration data. This problem arises if Western Blot (WB) measurements are used for model calibration. Western blotting is a technique to quantify protein amounts, and measured signal intensities are assumed to be proportional to respective protein amounts. Since proportionality factors depend on the membranes and the antibodies and are unique for each blot, the data have to be normalised in order to enable a comparison across different replicates. Hence WB data provide concentrations in terms of fold changes to a controlled condition. Compared to absolute concentrations, fold changes generally contain less information about the kinetic parameters which are to be estimated [8], and hence constitutes an additional difficulty for the inference of regulatory networks.

Different normalisation strategies (NS) are applied in practice [9], leading to different normalised data sets. Moreover, using the maximum likelihood (ML) estimation to determine the parameters requires the choice of an error model (EM). Common assumptions in this context are additive normally distributed noise [3, 5, 10] or multiplicative log-normally distributed noise sources [7, 11].

In this study, we investigate the combined impact of the NS and EM on the asymptotic properties of the ML estimators, exemplarily on a simple test-bed model of a reversible

phosphorylation reaction. We investigate in particular (i) the effects of increasing data set sizes on the accuracy and precision of the inference results, (ii) the robustness of the estimation to different noise levels, and (iii) how the choice of the NS and the EM affects the distribution of the ML estimates. Therefore, we design a Monte Carlo in silico study. Results show that in this example we need to measure a relatively large amount of data to obtain reliable estimates of model parameters. Moreover, normalisation to the mean of a time series outperforms normalisation to a single time point. And finally, while the choice of the EM does not lead to evident differences in the quality of the estimation results, it does in terms of computational costs.

## 2 | METHODS

### 2.1 | Modelling framework

Our in silico simulation workflow is shown in Figure 1. The general procedure is depicted in the left column; equations for the example system and respective output of all workflow steps are shown in the centre column and the right column, respectively.

#### 2.1.1 | Data simulation

We consider dynamical models of the form

$$\dot{x}(t,\theta) = f(x(t),\theta) \tag{1a}$$

$$z(t,\theta) = h(x(t),\theta), \tag{1b}$$

with state variables $x \in \mathbb{R}_+^N$, unknown kinetic parameters $\theta \in \mathbb{R}_+^M$ and output variables $z \in \mathbb{R}_+^Q$. The vector field $f$ and the function $h$ are continuous functions of $x$ and $\theta$.

The measured data are described by random variables (RVs) $\tilde{\mathbf{z}}_i(t_k)$, whose distributions $p_{\tilde{\mathbf{z}}_i(t_k)}(\tilde{z}_i(t_k))$ are functions of the state variables $x_i$ $(t_k, \theta_0)$ obtained via the numerical integration of Equation (1) with a 'true' parameter vector $\theta_0$ and a simulation EM. Western blotting provides signal intensities that are proportional to protein concentrations, with unknown scaling factors $\alpha_m$ that are specific for each replicate $m$ of the experiment corresponding to one blot,

$$\tilde{\mathbf{y}}_i^m(t_k) = \alpha_m \tilde{\mathbf{z}}_i(t_k). \tag{2}$$

#### 2.1.2 | Normalisation

Normalisation is required to compare across technical replicates, that is, to get rid of the blot specific index $m$. It constitutes a non-linear transformation $T_1$ of the original data set $\tilde{\mathbf{y}}_i^m(t_k)$ (Equation (2)):

$$\mathbf{y}_{i,\mathrm{NS}}(t_k) = T_1^{\mathrm{NS}}(\tilde{\mathbf{y}}_i^j(t_k)) = T_1^{\mathrm{NS}}(\tilde{\mathbf{z}}_i(t_k)). \tag{3}$$

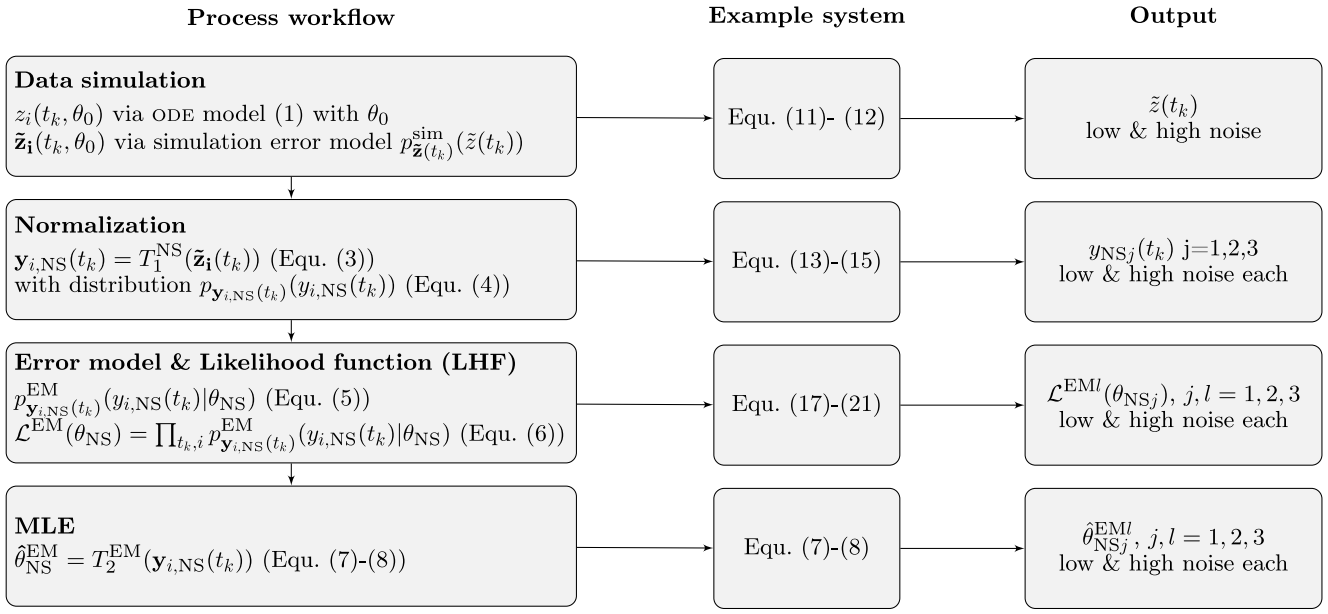


**FIGURE 1** Workflow for the in silico MC study of the distributions of the maximum likelihood (ML) estimates. Left: process workflow, Centre: Respective equations of the example system, Right: simulation output. Data $z_i$ $(t_k, \theta_0)$ are simulated via the ODE model (1) with parameter $\theta_0$. These data are corrupted by noise according to the simulation error model (EM) $p_{\tilde{\mathbf{z}}_i}(t_k)^{\mathrm{sim}}(\tilde{z}_i(t_k))$, leading to $\tilde{z}_i(t_k)$. Theoretically, measurement signals $\tilde{y}_i^m(t_k)$ are obtained by multiplying $\tilde{z}_i(t_k)$ with a replicate-specific proportionality factor $\alpha_m$ (Equation (2)). Choosing $\alpha_m$ is, however, not necessary due to normalisation in the next step. Output of this first step are two time series $\tilde{z}(t_k)$ with low and high noise levels. Since we have a single output in the example, we have omitted the index $i$ here. Normalised data $y_{i,\mathrm{NS}}(t_k)$ is obtained by a transformation $T_1^{\mathrm{NS}}(\tilde{\mathbf{z}}_i(t_k))$ with distribution $p_{\mathbf{y}_{i,\mathrm{NS}}(t_k)}(y_{i,\mathrm{NS}}(t_k))$. Outputs are $3 \times 2$ time series $y_{\mathrm{NS}j}$ $(t_k)$ (three NSs combined with two noise levels). Three different EMs $p^{\mathrm{EM}}_{\mathbf{y}_{i,\mathrm{NS}}(t_k)}(y_{i,\mathrm{NS}}(t_k)|\theta_{\mathrm{NS}})$ are used to formulate the likelihood function (LHF) $\mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}})$, which are combined with the six time series from the previous step. Optimisation can be interpreted as a second transformation $\hat{\theta}^{\mathrm{EM}}_{\mathrm{NS}} = T_2^{\mathrm{EM}}(\mathbf{y}_{i,\mathrm{NS}}(t_k))$ which transforms normalised data into parameters $\hat{\theta}^{\mathrm{EM}}_{\mathrm{NS}}$. Finally, this leads to an output of $2 \times 3 \times 3$ (low and high noise, 3 NSs, 3 EMs) parameter distributions. MC, Monte Carlo; NS, normalisation strategies.

Because of the non-linearity, the normalised data $\mathbf{y}_{i,\mathrm{NS}}(t_k)$ follow a distribution

$$p_{\mathbf{y}_{i,\mathrm{NS}}(t_k)}(y_{i,\mathrm{NS}}(t_k)) \tag{4}$$

which is different from that of the original concentrations, $p_{\tilde{\mathbf{z}}_i(t_k)}(\tilde{z}_i(t_k))$.

### 2.1.3 | Error model and likelihood function

The normalised data set $\mathbf{y}_{i,\mathrm{NS}}(t_k)$ (Equation (3)) is used for ML parameter estimation. Therefore, we formulate different EMs which allow us to evaluate the likelihood function (LHF) $\mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}})$. The chosen EM relates the experimental data with the model parameters $\theta_{\mathrm{NS}}$. For one data point $\mathbf{y}_{i,\mathrm{NS}}(t_k)$, the EM defines the density

$$p^{\mathrm{EM}}_{\mathbf{y}_{i,\mathrm{NS}}(t_k)}\left(y_{i,\mathrm{NS}}(t_k)|\theta_{\mathrm{NS}}\right). \tag{5}$$

Exploiting the independence of data points, this leads to the LHF

$$\mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}}) = \prod_{k\in\mathcal{I}_{\mathrm{NS}}}\prod_{i=1,\ldots,N} p^{\mathrm{EM}}_{\mathbf{y}_{i,\mathrm{NS}}(t_k)}\left(y_{i,\mathrm{NS}}(t_k)|\theta_{\mathrm{NS}}\right), \tag{6}$$

with $\mathcal{I}_{\mathrm{NS}}$ indicating a time index set that is dependent on the chosen normalisation strategy and is specified later on.

### 2.1.4 | Maximum likelihood estimation

The Maximum likelihood estimation (MLE) $\widehat{\theta}^{\mathrm{EM}}_{\mathrm{NS}}$ is obtained by maximising the LHF,

$$\widehat{\theta}^{\mathrm{EM}}_{\mathrm{NS}} = \arg\max_{\theta_{\mathrm{NS}}} \mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}}). \tag{7}$$

Under repeated data generation, the estimate $\widehat{\theta}^{\mathrm{EM}}_{\mathrm{NS}}$ is also a RV, whose distribution is obtained through a non-linear transformation $T^{\mathrm{EM}}_2$ of the normalised data set Equation (3):

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EM}}_{\mathrm{NS}} = T^{\mathrm{EM}}_2\left(\mathbf{y}_{i,\mathrm{NS}}(t_k)\right). \tag{8}$$

Under some regularity conditions, the MLE is often a good estimator in the sense that it is consistent, that is, it converges in probability to the true parameters $\theta^*$ as the number of samples increases, $\widehat{\theta}\xrightarrow{p}\theta^*$. This implies that $\widehat{\theta}$ is asymptotically unbiased. Furthermore, it is asymptotically normal, that is, $\sqrt{N}\left(\widehat{\theta}-\theta^*\right)\xrightarrow{d}\mathcal{N}\left(0, I(\theta^*)^{-1}\right)$ with Fisher information matrix $I(\theta^*)$, which guarantees that it converges fast enough (with a rate $1/\sqrt{N}$). And last, the MLE is asymptotically efficient, meaning that $\widehat{\theta}$ achieves the minimum possible variance among all unbiased estimators, or the Cramér–Rao lower bound for large sample sizes, making it a precise estimator. These properties of MLEs date back to Fisher [12] and can be found in any statistics textbooks (see e.g. Gelman et al. [13]).

Regularity conditions for consistency are smoothness of the LHF, identifiability of $\widehat{\theta}$ and existence of the mean value $\mathbb{E}_{\theta^*}\log p_{Y_k}\left(y_k|\theta\right)$. Furthermore, $\widehat{\theta}$ must not lie on the boundary of the defined domain $\Theta$. In addition, for asymptotic normality $\mathbb{V}_{\theta^*}\log p_{Y_k}\left(y_k|\theta\right)$ has to exist. If these are satisfied, then the distribution of $\widehat{\theta}$ is for large sample sizes that are approximately normal with a small variance.

In order to study the distribution of $\widehat{\boldsymbol{\theta}}^{\mathrm{EM}}_{\mathrm{NS}}$, we apply a Monte Carlo approach in which we generate experimental data according to Equation (2) and then propagate the noise via the transformations $T_1$ and $T_2$.

## 2.2 | A reversible reaction as a simple test-bed model

### 2.2.1 | Data simulation

As a simple test-bed model, we consider a reversible phosphorylation reaction,

$$\mathrm{P} \underset{\mathrm{k_2}}{\overset{\mathrm{k_1}}{\rightleftharpoons}} \mathrm{P}^*$$

where $P$ and $P^*$ denote unphosphorylated and phosphorylated protein and $p$ and $p^*$ the respective concentrations. Assuming mass action kinetics, the dynamics of the fraction

$$x(t) = \frac{p^*(t)}{p(t)+p^*(t)} \tag{9}$$

of phosphorylated proteins is given by

$$\dot{x}(t,\theta) = k_1 - (k_1+k_2)x(t) \tag{10a}$$

$$z(t,\theta) = x(t,\theta) \tag{10b}$$

with $\theta = (k_1, k_2)$ and solution

$$z(t,\theta) = \frac{k_1}{k_1+k_2}\left[1-e^{-(k_1+k_2)t}\right] + z_0 e^{-(k_1+k_2)t}, \tag{11}$$

where the initial condition $z_0$ is assumed to be known and $z_0 \neq 0$ to avoid structural non-identifiability [14]. Since $N = 1$, we neglect the index $i$ to enumerate components in the following.

For data simulation, we use a mixed EM $p^{\mathrm{sim}}_{\tilde{\mathbf{z}}(t_k)}(\tilde{z}(t_k))$ with a multiplicative and an additive error as suggested in Gábor and Banga [4], Kreutz et al. [11],

$$\tilde{\mathbf{z}}(t_k) = x(t_k,\theta_0)\eta + \epsilon \tag{12}$$

with noise parameters $\eta \sim \log\mathcal{N}\left(0,\sigma^2_\eta\right)$ and $\epsilon \sim \mathcal{N}\left(0,\sigma^2_\epsilon\right)$. Standard deviations (SD) $\sigma_\eta$ and $\sigma_\epsilon$ were set to realistic experimental values according to Ref. [15], where a reference value of 10% for the proportional component of the

measurement noise, introduced due to variability in pipetting the cellular lysates was recorded.

### 2.2.2 | Normalisation

We consider three commonly used NSs for time series data (see e.g. presentations of WB data by Degasperi et al. [3, 9], Jarajapu et al. [16], Jensch et al. [17], Santos et al. [18], Talavera et al. [19], and Wang et al. [20]), namely

1. Normalisation by the value at the first time point (**NS1**), which often also represents a control condition:

$$y_{\mathrm{NS1}}(t_k) = \frac{\tilde{y}^j(t_k)}{\tilde{y}^j(t_1)} = \frac{\tilde{z}(t_k)}{\tilde{z}(t_1)} \tag{13}$$

The set of time indices $k$ is given by $\mathcal{I}_{\mathrm{NS1}} = \{2, \ldots, K\}$, since all relative data at time point $t_1$ are equal to one.

2 Normalisation by the value at the last time point (**NS2**):

$$y_{\mathrm{NS2}}(t_k) = \frac{\tilde{y}^j(t_k)}{\tilde{y}^j(t_K)} = \frac{\tilde{z}(t_k)}{\tilde{z}(t_K)} \tag{14}$$

The set of time indices $k$ is here given by $\mathcal{I}_{\mathrm{NS2}} = \{1, \ldots, K-1\}$, since all data at time point $t_K$ are equal to 1.

3 Normalisation by the mean value of all time points in a time series (**NS3**):

$$y_{\mathrm{NS3}}(t_k) = \frac{\tilde{y}^j(t_k)}{\frac{1}{K}\sum_{k=1}^{K}\tilde{y}^j(t_k)} = \frac{\tilde{z}(t_k)}{\frac{1}{K}\sum_{k=1}^{K}\tilde{z}(t_k)}, \tag{15}$$

with $\mathcal{I}_{\mathrm{NS3}} = \{1, \ldots, K\}$.

The corresponding normalised model outputs, to be used in the following step for the definition of the LHF, are defined as follows:

$$z_{\mathrm{NS1}}(t_k, \theta) = \frac{z(t_k, \theta)}{z(t_1, \theta)} \tag{16a}$$

$$z_{\mathrm{NS2}}(t_k, \theta) = \frac{z(t_k, \theta)}{z(t_K, \theta)} \tag{16b}$$

$$z_{\mathrm{NS3}}(t_k, \theta) = \frac{z(t_k, \theta)}{\frac{1}{K}\sum_{k=1}^{K} z(t_k, \theta)}. \tag{16c}$$

### 2.2.3 | Error models and maximum likelihood estimation

We combine these NSs with three different EMs, which are then used for the parameter estimation. We note here that we cannot derive an analytical expression of the distribution (4) for the EM used for the simulation (Equation (12)) and any of the three NSs. Thus, we are not able to compare the inference results in our settings with the true or gold standard model. However, if none of the NS and EM combinations reveal the true underlying model, this provides grounds for a fair comparison across all combinations.

Using the two common assumptions of normally (see e.g. Fröhlich et al. [10], Kreutz and Timmer [21], Raue et al. [7], Weber et al. [22]) or log-normally (see e.g. Kreutz et al. [11], Limpert et al. [23], Thomaseth et al. [24]) distributed WB raw data, we consider three classes of EMs for the normalised data set, namely normal, log-normal, or Gaussian ratio (GR) distributions, which define $p^{\mathrm{EM}}_{\mathbf{y}_{\mathrm{NS}}(t_k)}$ and the LHF $\mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}})$:

1. Normal error model (**N-EM**):

$$\mathbf{y}_{\mathrm{NSj}}(t_k) \sim \mathcal{N}\left(z_{\mathrm{NSj}}\left(t_k, \theta_{\mathrm{NSj}}\right), \sigma^2\right), \; j = 1, 2, 3. \tag{17}$$

2 Log-normal error model (**LN-EM**):

$$\mathbf{y}_{\mathrm{NSj}}(t_k) \sim \log \mathcal{N}\left(\log z_{\mathrm{NSj}}\left(t_k, \theta_{\mathrm{NSj}}\right), \sigma^2\right), \; j = 1, 2, 3 \tag{18}$$

3 Gaussian ratio error model (**GR-EM**):

$$\mathbf{y}_{\mathrm{NS1}}(t_k) \sim \frac{\mathcal{N}(z(t_k, \theta_{\mathrm{NS1}}), \sigma^2)}{\mathcal{N}(z(t_1, \theta_{\mathrm{NS1}}), \sigma^2)}, \; \rho = 0 \tag{19}$$

$$\mathbf{y}_{\mathrm{NS2}}(t_k) \sim \frac{\mathcal{N}(z(t_k, \theta_{\mathrm{NS2}}), \sigma^2)}{\mathcal{N}(z(t_K, \theta_{\mathrm{NS2}}), \sigma^2)}, \; \rho = 0 \tag{20}$$

$$\mathbf{y}_{\mathrm{NS3}}(t_k) \sim \frac{\mathcal{N}(z(t_k, \theta_{\mathrm{NS3}}), \sigma^2)}{\mathcal{N}\left(\frac{1}{K}\sum_{k=1}^{K} z(t_k, \theta_{\mathrm{NS3}}), \frac{\sigma^2}{K}\right)}, \; \rho = \frac{1}{\sqrt{K}} \tag{21}$$

with $k \in \mathcal{I}_{\mathrm{NSj}}$ for each specific $j$.

Concerning the GR-EM, we will restrict ourselves to the case of independent Gaussian RVs at the numerator and denominator for different time points, leading to a correlation coefficient $\rho = 0$ for the first and second NS and $\rho = 1/\sqrt{K}$ for the third one, as demonstrated in SI, Section 1. From the assumption that all $\tilde{\mathbf{x}}(t_k)$ are independent RVs given $\theta_{\mathrm{NSj}}$, it follows that

$$\mathrm{Var}\left(\frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{z}}(t_k)\right) = \frac{\sigma^2}{K}, \tag{22}$$

as assumed in the EM Equation (21).

The GR distribution is characterised by four parameters $a$, $b$, $r$ and $s$ (see e.g. Hayya et al. [25], Hinkley [26, 27], Marsaglia [28, 29]), which in this context are related to the simulated quantities and therefore functions of the unknown ODE model parameters. In SI, Section 2 we show the definition of this parametrisation of the GR distributions for all three NSs

(Equations (19)–(21)), which were used for the implementation of the LHF in our simulation study.

For all three EMs we assume that the parameter $\sigma \in \mathbb{R}_{>0}$, related to the SD of the considered distributions, is the same for all $k$ of the data set. In the case of the GR-EM, we assume that $\sigma^2$ is the variance of the Gaussian RVs at the numerator and denominator $\tilde{\mathbf{x}}(t_k), \forall k \in \{1, \ldots, K\}$. Two options are used in practice to assign these values for $\sigma$: it can be either a priori empirically determined from experimental data or it can be estimated simultaneously with the model parameters $\theta$. Several studies hint to the fact that the empirical estimation is unreliable and should be avoided due to a low number of technical replicates available for estimation, and simultaneous estimation should be preferred [3, 7]. Many parametric models are suggested for the estimation of the SD $\sigma$ [3, 5, 17]. Therefore, in this study we implemented the estimation of $\sigma$ simultaneously to $\theta = (k_1, k_2)$ and decided to consider the most basic model, for which a single parameter value $\widehat{\sigma}_{\mathrm{MLE}}$ is estimated from the available experimental data set.

## 3 | RESULTS

### 3.1 | Normalisation to the mean of the simulated time series leads to lower uncertainty compared to normalisation to a single time point

In silico-generated time series data are shown in Figure 2. Figure 2a shows box plots of the simulated time series data $\tilde{\mathbf{z}}(t_k)$ for four exemplary time points $t_k$, $k \in \{1, 2, 3, 4\}$, obtained with the test-bed model Equation (11) and the mixed EM Equation (12). Distributions of the corresponding means of the four time points are shown on the right. The parameters and initial condition were set to $\theta_0 = (4, 1)$ and $z_0 = 0.3$. In order to compare the effects for low and high measurement noise, we simulated data with low (Figure 2a left) and high (Figure 2a right) noise levels. Distributions of the normalised data $\mathbf{y}_{\mathrm{NS}}(t_k)$ are shown in Figure 2b for the three NSs and low (left column) and high (right column) noise levels. As a consequence of the different variances and coefficients of variation (CV) of the quantities used as reference conditions for normalisation, the statistical properties of the obtained normalised data sets also differ. In particular, the data normalised with NS1 (upper row, reference condition has largest CV) have the largest uncertainty, followed by NS2 (lower CV of reference condition compared to NS1) and finally by NS3, which has the lowest uncertainty. This comes from the fact that the variance of the mean is reduced by the factor $1/K$ (Equation (22)). Furthermore, NS1 is the most sensitive to higher noise levels (right column), showing many more outliers in the right tails of the distributions, which were cut off for representative reasons. These facts are in line with Degasperi et al. [9], who suggested to avoid choosing normalisation points with low quantified intensities for hypothesis testing studies, since this strategy results in large CVs for normalised data.

### 3.2 | A realistic number of measurements leads to boundary effects of the estimationand sloppy parameters

We solved the inference problem via MLE for all combinations of the three NSs, the three EMs, and low and high noise levels. The resulting distributions of the ML estimates $\widehat{\mathbf{k}}_1$ and $\widehat{\mathbf{k}}_2$ are illustrated in Figure 3 exemplarily for N-EM and NS1. Shown here are box plots (Figure 3a), as well as 1D marginal distribution histograms and 2D scatter plots for low (left) and high (right) noise levels and $K = 4$ (bright colours) and $K = 8$ (darker colours) time points (Figure 3b). Boundaries for the optimization have been set to [0,10] for both rate parameters.

The distribution of $\widehat{\mathbf{k}}_2$ accumulates at the lower bound zero, a boundary effect which can be observed in the scatterplots of all EM and NS combinations and both noise levels (see SI, Section 3.1). It causes bimodal distributions for both parameters. For our particular example, the boundary effect is directly visible in the solution Equation (11) of the ODE system: Here, the parameter $k_2$ only appears as a summand together with $k_1$, thus in many cases in which the optimiser wants to assign negative values for $k_2$, which is suppressed by the lower boundary, it assigns values $\widehat{k}_1 > k_1^* = 4$.

When doubling the amount of time points $K$ from four to eight, while keeping the same noise level (left and right parts in Figure 3b), some of the samples of $\widehat{\mathbf{k}}_2$ estimated close to zero are released, and the rest of the distribution moves towards the true parameter value.

Figure 3c shows confidence bounds for the corresponding inferred model trajectories, along with the box plots of the normalised data $\mathbf{y}_{NS1}(t_k)$, $k = 1, \ldots K$, and the noiseless trajectory $z_{\mathrm{NS1}}(t, \theta_0)$ (red line). Specifically, the 5th and 95th percentiles were evaluated: The totality of the trajectories is represented inside the blue-coloured area of Figure 3c, where 90% of the trajectories is contained in the darker blue region and the cut off 5% from top and bottom are contained in the light blue-coloured areas. The Inferred confidence intervals are in good agreement with the noiseless trajectory and the box plots of the normalised data. Since the distributions for the different time points are not independent but coupled via the ODE model, we do anyway not expect a perfect fit between the box plots and the inferred confidence intervals. This mainly affects time points near the time point used for normalisation, where confidence intervals for the inferred trajectories are smaller than those of the noisy normalised data. As expected, the uncertainty increases with higher noise levels (from left to right). Moreover, comparing $K = 4$ to $K = 8$ time points (top to bottom row), the uncertainty does not become smaller, which is due to the fact that the concentration at time point $t_1$ used for normalisation is smaller for $K = 8$, which causes a larger spread in the normalised data. Corresponding figures obtained for the other two NS NS2 and NS3 as in Figure 3c are given in SI, Section 6, Figures 14 and 15. In these other two cases, the uncertainty of model outputs significantly improves in the initial phase of the trajectories (near $t_0$) comparing $K = 4$ to $K = 8$ time points (top to bottom row).

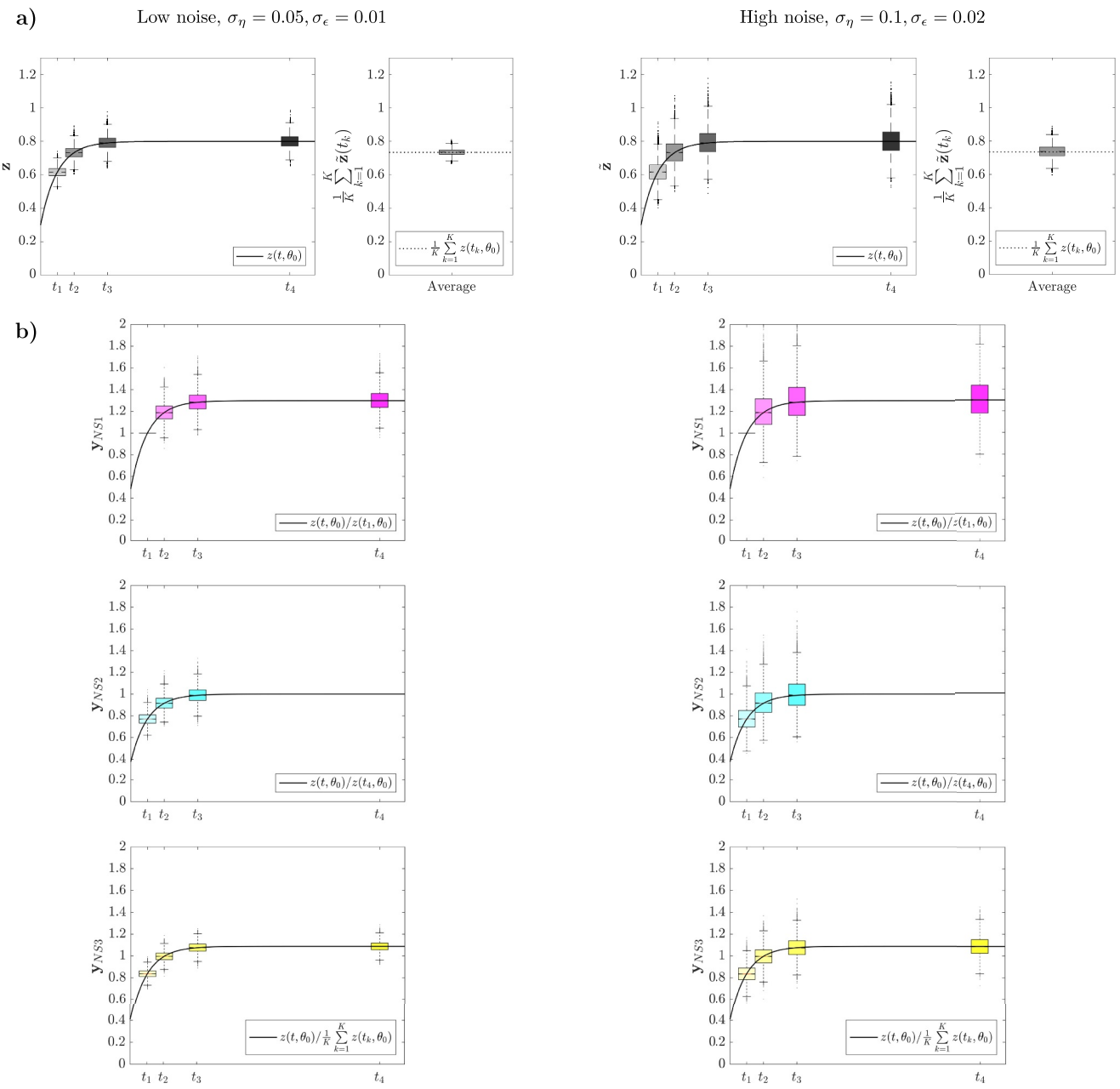


**FIGURE 2** (a) Noisy simulated time series data. Distributions of the state variable $\tilde{\mathbf{z}}(t_k)$ at four exemplary time points $(t_1, t_2, t_3, t_4) = (0.2, 0.4, 0.8, 3)$, and distribution of the mean of the corresponding samples at the four time points. We generated $n = 10,000$ realisations via Monte Carlo simulations from the noise model Equation (12). The continuous line represents the noise-free time course of the state variable $z(t, \theta_0)$, obtained for $z_0 = 0.3$ and $\theta_0 = (4, 1)$. (*Left*) Low noise level, (*Right*) high noise level. (b) Noisy normalised time series data. The shown distributions (box plots) are obtained from the sampled noisy realisations of the normalised variables $\mathbf{y}_{NSj}(t_k)$, $j = 1, 2, 3$, for the three different NSs, marked with different colours and are given for low (left) and high (right) noise. NS, normalisation strategies.

Most importantly, for all scenarios shown here, the uncertainty in the inferred trajectories (i.e. the confidence interval length relative to the nominal value) is consistently smaller than the uncertainty of the ML estimators, indicating that the very different parameter combinations can lead to similar model outputs. This phenomenon is sometimes referred to as sloppy parameters [30].

Overall, the results show that $K = 4$ and even $K = 8$ data points are not sufficient to estimate model parameters accurately, since the distribution of the ML estimates has a large variance, and many estimates deviate substantially from the true values. This causes the described boundary effects and leads to a visible bias of both estimates in the box plots. Increasing the data set size is a possible solution to overcome this problem. However, boundary effects had still a considerable impact after a six-fold increase in the number of replicates from $M = 1$ to $M = 6$ in all scenarios (see SI, Section 3.2).

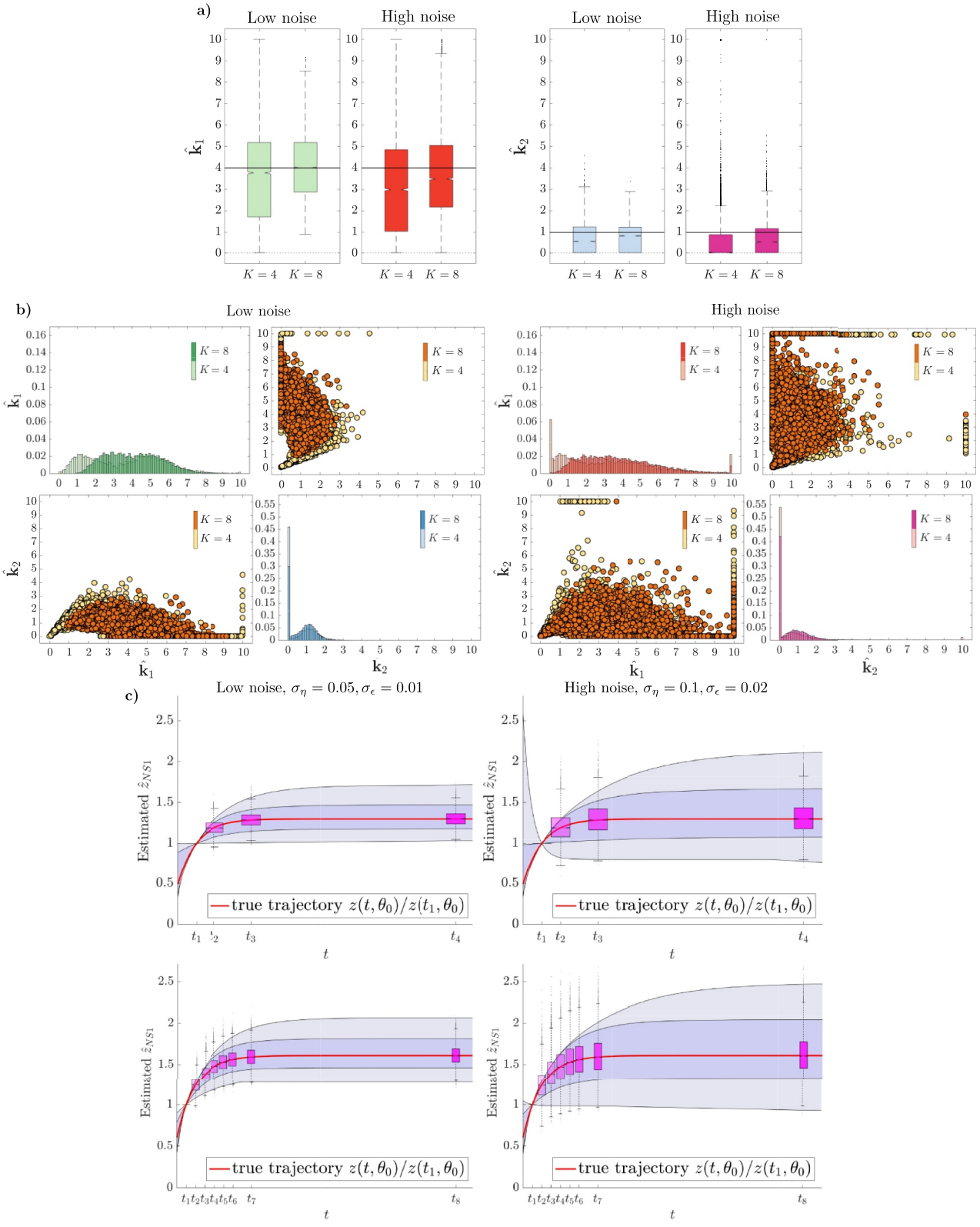


**FIGURE 3** Boundary effects appear in a broad range under realistic experimental settings. (a) Box plots of the estimated parameters $\widehat{\mathbf{k}}_1$ and $\widehat{\mathbf{k}}_2$ obtained with $K = 4$ or $K = 8$ measured time points of the phosphorylated protein concentration. The considered time points are $t_i$, $i = 1, \ldots, 4 = (0.2, 0.4, 0.8, 3)$ for $K = 4$ and $t_i$, $i = 1, \ldots, 8 = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 3)$ for $K = 8$. (b) Marginal distributions (histograms) and scatter plots in the two-dimensional parameter space of $\widehat{\mathbf{k}}_1$ and $\widehat{\mathbf{k}}_2$, obtained for $M = 1$ replicates for $K \in \{4, 8\}$ time points each. (c) Inferred normalised model trajectories $\widehat{z}_{NS1}(t)$ corresponding to parameter estimates shown in (a) and (b) along with the box plots of the normalised data $\mathbf{y}_{NS1}(t_k)$, $k = 1, \ldots K$, for $K = 4$ (upper part) and $K = 8$ (lower part) and the noiseless trajectory $z_{\mathrm{NS1}}(t, \theta_0)$ (red line). The results were obtained using the Normal error model (N-EM), the first set of normalised data (NS1) and two noise levels (low on the left: $\sigma_\eta = 0.05$, $\sigma_\epsilon = 0.01$, high on the right: $\sigma_\eta = 0.1$, $\sigma_\epsilon = 0.02$).

### 3.3 | A large number of data points is needed for good maximum likelihood estimates

This motivated us to increase the number of data points even further. We decided to compare three numbers of time points, $K \in \{4, 8, 12\}$ and to consider $M = 10$ replicates, resulting in a total number of 40, 80 and 120 simulated measurements, respectively. In addition, we widened the search space and in particular allowed the optimiser also to search for negative values. This was only possible for N-EM and GR-EM, while it could not be implemented for the LN-EM, since as a consequence the output functions $y_{\mathrm{NSj}}(t, \theta_0), j = 1, 2, 3$ may assume negative values that cause errors in the optimization of the LHF $\mathcal{L}^{\mathrm{EM}}(\theta_{\mathrm{NS}})$, which is not defined for negative $y(t_k)$ values.

Figure 4 shows the estimation results for the three EMs and NS1 for low (left) and high (right) noise levels. Despite the large amount of data used for parameter estimation, the boundary effects could not be eliminated in the case of the LN-EM (Figure 4b), for both noise levels. Concerning N-EM and GR-EM (Figures 4a and 4c), the boundary effects disappear entirely for the low noise level (left column). Instead, for high noise, we regard only the case $K = 12$ to be almost unaffected. The same behaviour can be observed when using NS2 or NS3 (see SI, Section 3.3).

Since the SDs of the EMs used for estimation have a completely different meaning from the SDs of the EM used for data simulation, there is no 'true' value for $\sigma$ for a comparison with the estimated values $\widehat{\sigma}$, hence their distributions are not discussed and for completeness are only shown exemplarily in SI, Section 4.

Overall, our analysis shows that in our test-bed scenario a very large number of data points is needed for the distributions of ML estimates to behave approximately normal with a small variance.

### 3.4 | Normalisation causes a large increase in parameter uncertainties

We asked the question whether the observed large uncertainty in ML estimates is mainly caused by the normalisation step. Therefore, we repeated the estimation procedure with the unnormalised data $\tilde{\mathbf{z}}$ (see data in Figure 2a) assuming that we can quantify the fraction of phosphorylated protein directly without an unknown scaling factor. The resulting distribution of the ML estimates are illustrated in Figure 5b exemplarily for low and high noise levels, $K = 4$ (light colours) and $K = 8$ (darker colours) time points and $M = 10$ replicates. These settings equal those in Figure 4a (displayed again in Figure 5a to allow a better comparison), where normalised data with NS1 had been used. Comparing both ML distributions, the one in Figure 5b is much smaller and tightly distributed around the true parameter values compared to Figure 5a. The 2D distribution can also well be approximated by a multivariate normal distribution for both $K = 4$ and $K = 8$ time points. Overall, this clearly demonstrates that data normalisation causes a large loss in information about the parameters of our model.

### 3.5 | The Gaussian ratio error model is less biased than the Normal error model especially for high noise levels

To further evaluate the quality of the obtained inference results, we considered the statistical measure of bias of the median to quantify the *accuracy* of the estimation, that is, the closeness of the considered value to the true parameter value. Additionally, we considered the interquartile range (IQR), representing a standard measure of the dispersion of a distribution, as an indicator for *precision*, that is, a characterisation of the variability of the estimate. Figure 6 visualises the accuracy and precision of the inference results for both the estimated model parameters $\widehat{\boldsymbol{\theta}} = \left(\widehat{\mathbf{k}}_1, \widehat{\mathbf{k}}_2\right)$. Here, the results are only shown for those scenarios that were not affected by the boundary effects, that is, N-EM (first row) and GR-EM (second row), in combination with all three different normalised data sets (different columns). Increasing data set sizes are visualised by larger dots, and different colours were used for the two noise levels. We can observe that increasing the data set size has a small effect when the noise in the data is low, as green dots ($\widehat{\mathbf{k}}_1$, low noise) and blue dots ($\widehat{\mathbf{k}}_2$, low noise) of all sizes are close to each other. Thus, for a very large amount of measured data, for which the distributions of the ML estimators tend to the asymptotic behaviour, the benefit of doubling or tripling the amount of measured data is rather minimal with low input noise, while this is not the case for high noise. Moreover, especially the bias is much lower for the GR-EM than for the N-EM for high noise levels and NS1 and NS2. Thus, the GR-EM is less biased than the N-EM in these settings.

### 3.6 | Impact of normalisation strategies on the uncertainty of maximum likelihood estimates

We pose the question of how the three considered NSs affect noise propagation from raw concentration measurements to estimated model parameters, while keeping fixed other features of the inference process.

Results in terms of bias of the median versus IQR values are shown in Figure 7a for low noise and $K = 4$ time points and Figure 7b for high noise levels and $K = 12$ time points. In this analysis, we focus only on the results obtained with N-EM and GR-EM. These considered scenarios relate to all cases in which the estimation results did not present boundary effects.

First, we observe that the impact of the three NSs is different for the two estimated parameters, hence we cannot derive a universal statement concerning the impact of the three NSs on the estimation results. For low noise level (Figure 7a) the different NSs affect mainly the accuracy of the estimation, while

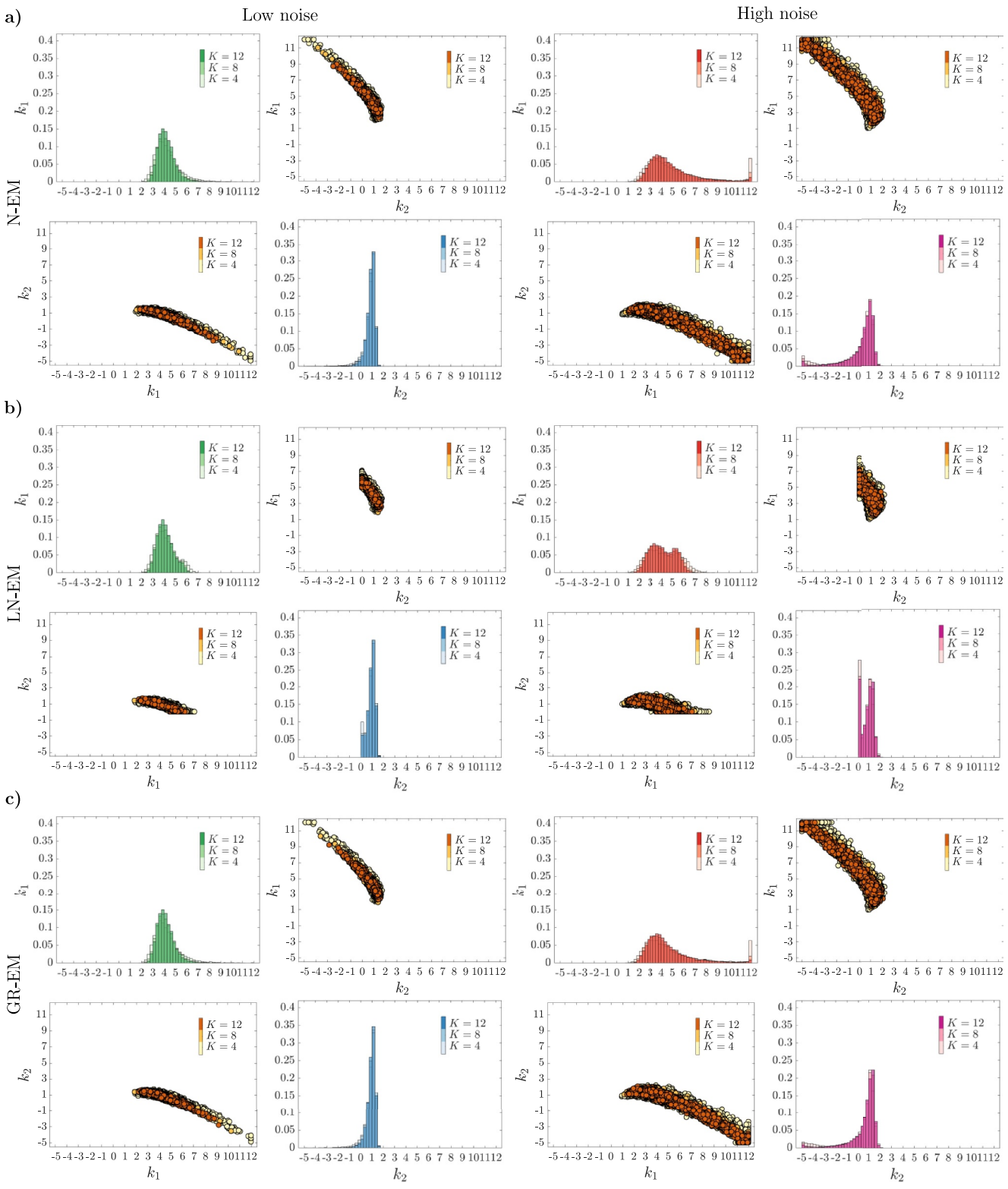


**FIGURE 4** Effect of the amount of time points on the MLE distributions. Marginal distributions (histograms) and 2D scatter plots of $\widehat{\mathbf{k}}_1$ and $\widehat{\mathbf{k}}_2$, obtained for $K \in \{4, 8, 12\}$ time points and $M = 10$ replicates each. These results were obtained using the first set of normalised data (NS1) and (a) Normal error model (N-EM), (b) Log-normal error model (LN-EM), (c) Gaussian ratio error model (GR-EM), for two noise levels, respectively (low on the left: $\sigma_\eta = 0.05$, $\sigma_\epsilon = 0.01$, high on the right: $\sigma_\eta = 0.1$, $\sigma_\epsilon = 0.02$). The considered time points are $t_i$, $i = 1, \ldots, 4 = (0.2, 0.4, 0.8, 3)$ for $K = 4$, $t_i$, $i = 1, \ldots, 8 = (0.1, 0.16, 0.27, 0.44, 0.72, 1.18, 1.93, 3.16)$ for $K = 8$ and $t_i$, $i = 1, \ldots, 12 = (0.1, 0.14, 0.19, 0.26, 0.35, 0.48, 0.66, 0.9, 1.23, 1.69, 2.31, 3.16)$ for $K = 12$. MLE, Maximum likelihood estimation.

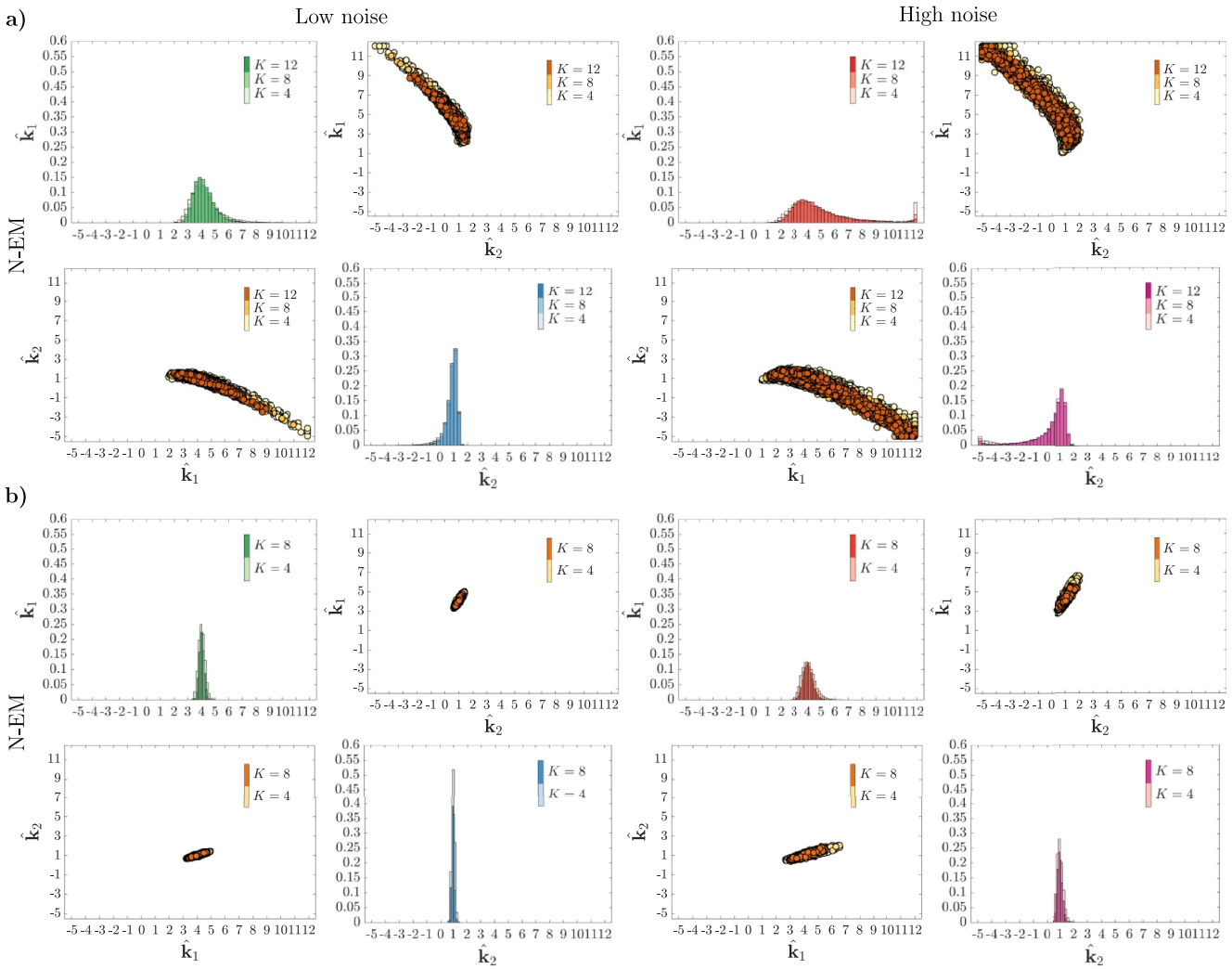


**FIGURE 5** Parameter estimation results from the unnormalised data. Marginal distributions (histograms) and 2D scatter plots of $\widehat{\mathbf{k}}_1$ and $\widehat{\mathbf{k}}_2$, obtained for $K \in \{4, 8\}$ time points and $M = 10$ replicates each. These results were obtained using (a) the first set of normalised data $\mathbf{y}_{NS1}$ and (b) the unnormalised data $\tilde{\mathbf{z}}$ and Normal error model (N-EM) for two noise levels, respectively (low on the left: $\sigma_\eta = 0.05$, $\sigma_\epsilon = 0.01$, high on the right: $\sigma_\eta = 0.1$, $\sigma_\epsilon = 0.02$). The considered time points are $t_i$, $i = 1, \ldots, 4 = (0.2, 0.4, 0.8, 3)$ for $K = 4$ and $t_i$, $i = 1, \ldots, 8 = (0.1, 0.16, 0.27, 0.44, 0.72, 1.18, 1.93, 3.16)$ for $K = 8$

the IQR is nearly unaffected for both parameters. For high noise (Figure 7b) they impact both accuracy and precision. The trend is, however, roughly maintained when considering the same EM and the same parameter with the increasing noise.

A surprising result is that the NS1 (magenta dots) does not always lead to the worst results (i.e. both higher bias and higher IQR) even if the corresponding normalised data used for estimation show the largest variability (Figure 2). Instead, NS2 (blue dots) causes the largest IQR and also the largest bias for $\widehat{\mathbf{k}}_1$, for both noise levels. This is probably due to the relative error contribution for data simulation. Since $z$ is smaller for earlier time points, its absolute measurement error is also smaller.

In summary, despite the large amount of data used for the estimation, the impact of different NSs is not univocal on all estimated parameters under realistic noise settings. Therefore, choosing NS3, which generally shows the lowest bias also for higher noise levels, seems to be a good compromise.

An aggregated comparison of the goodness of the estimation results of N-EM and GR-EM across the three NSs in terms of the Bayesian information criterion values is conducted in SI, Section 5. Here, N-EM and GR-EM behave similar for all three NSs. This similarity is probably due to the validity of the condition for the approximation of the GR distribution with a Gaussian distribution, which holds in the case of uncorrelated signals for sufficiently large CV of the nominator RV of the ratio distribution, as discussed in the study by Hayya et al. [25], Marsaglia [28, 29]. Based on these results, we suggest to select the N-EM, since it is computationally faster to optimise than the GR-EM, while leading to similar results.

## 4 | DISCUSSION AND CONCLUSIONS

In this study, we presented the results of a statistical analysis of the combined effects of different NS, different EM and low and high noise levels for WB time series data on the parameter estimation for a biochemical reaction. We developed a statistical framework to investigate the noise propagation from the

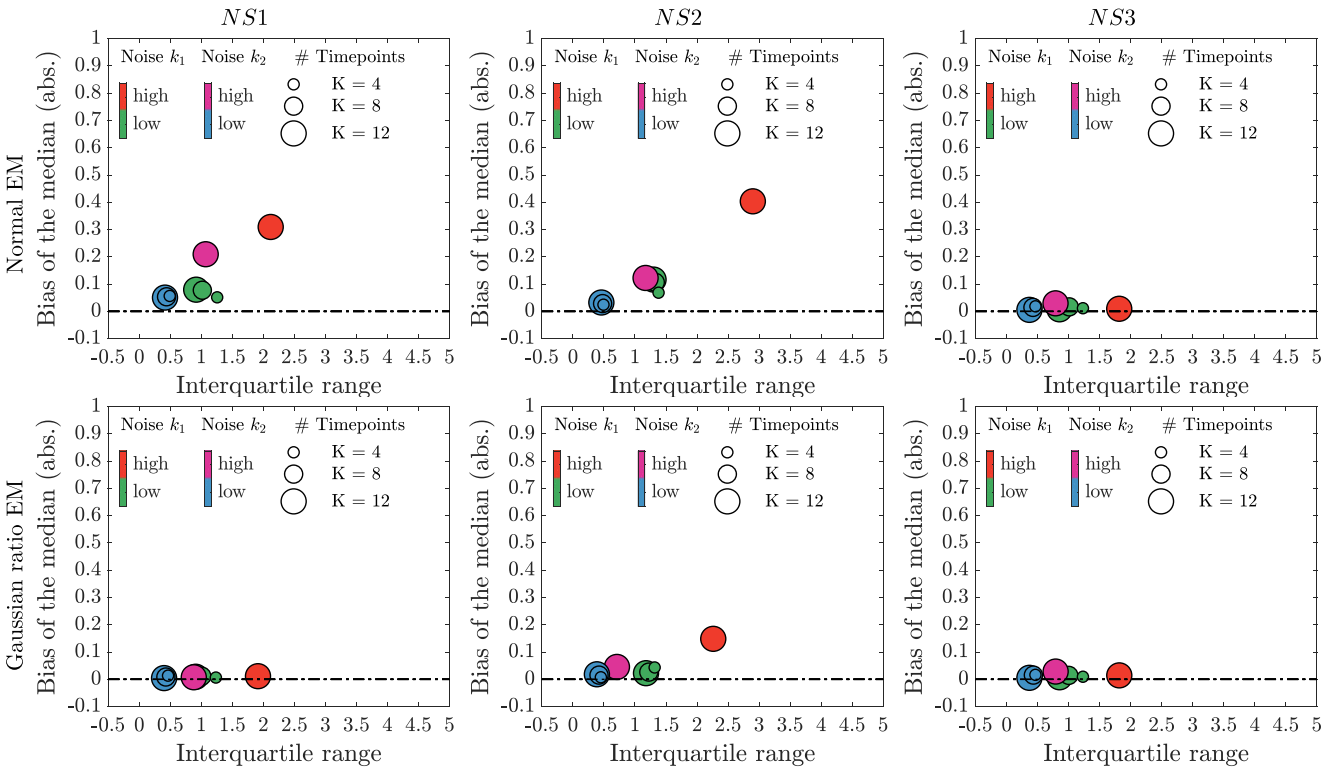


**FIGURE 6** Effect of normalisation strategies (NS) and error model (EM) choice on precision and accuracy of the MLE for varying numbers of time points. Absolute values of the bias of the median versus interquartile range (IQR) values for both the estimated parameter values obtained with $K = 4$, 8 and 12 time points and $M = 10$ replicates each, given for low noise level ($\sigma_\eta = 0.05$ and $\sigma_\epsilon = 0.01$). For high noise ($\sigma_\eta = 0.1$ and $\sigma_\epsilon = 0.02$) these statistics are given only for the case $K = 12$. Green and red dots refer to the parameter $\hat{\mathbf{k}}_1$, while blue and magenta refer to $\hat{\mathbf{k}}_2$. MLE, Maximum likelihood estimation.

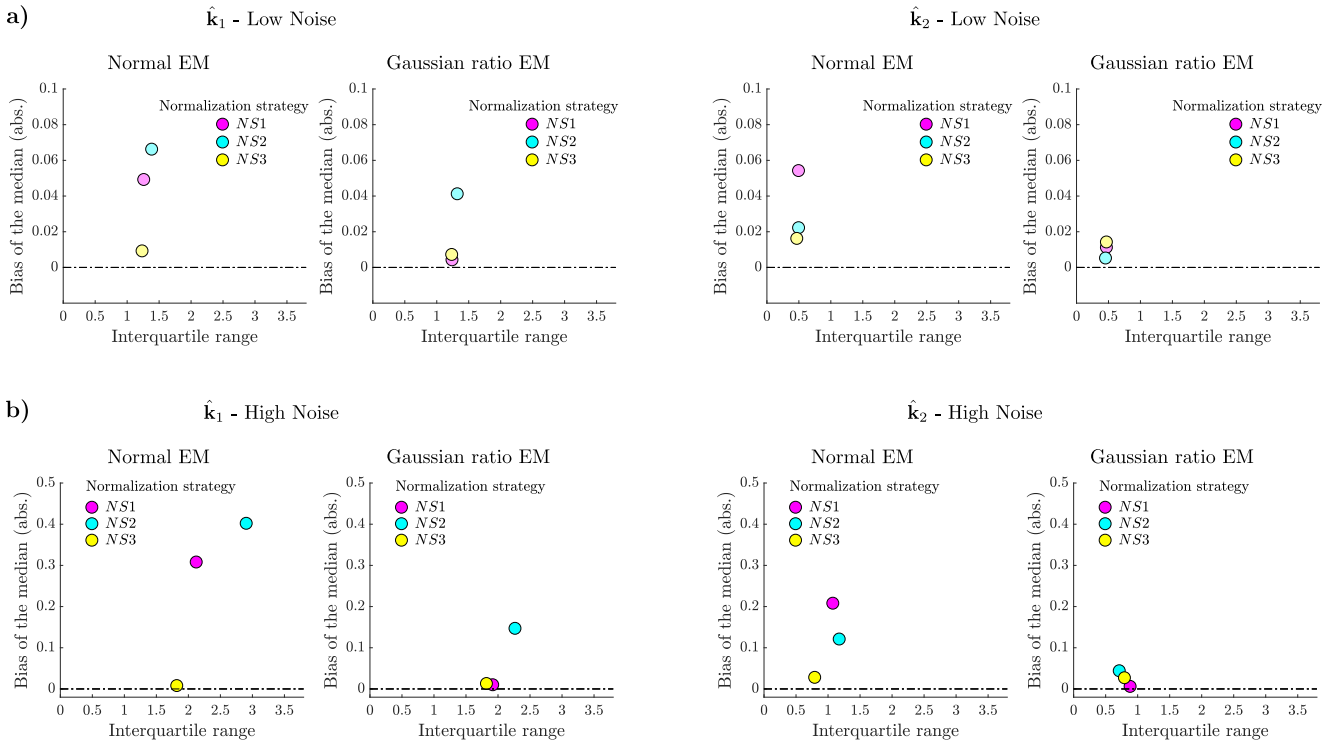


**FIGURE 7** Effects of three alternative NSs on accuracy and precision of parameter estimates. Absolute values of the bias of the median versus interquartile range (IQR) values are given for both the estimated parameters $\hat{\mathbf{k}}_1$ (left) and $\hat{\mathbf{k}}_2$ (right), obtained with the three considered NSs. These statistics were obtained assuming either the Normal error model (N-EM) or Gaussian ratio error model (GR-EM) as likelihood function (LHF) for the optimisation problem, for $M = 10$ replicates and (a) $K = 4$ time points and low noise level ($\sigma_\eta = 0.05$, $\sigma_\epsilon = 0.01$) and (b) $K = 12$ time points and high noise level ($\sigma_\eta = 0.1$, $\sigma_\epsilon = 0.02$). NS, normalisation strategies.

measured data to the estimated parameters via Monte Carlo simulations. The methodology was exemplified in an in silico study in which we used a reversible protein phosphorylation reaction as a test-bed model to generate time series data by using a realistic noise model for WB data [11, 15]. For the inference problem, we used three different NS to normalise WB signals, which were combined with three EMs for ML estimation. Since no analytic expression of the LHF is available for the realistic noise model used for simulation, we could not compare with the 'gold standard', that is, the 'true' model which was actually used for data generation. This setup resembles a real case scenario in which the true statistical process generating the data is in general different from the EM assumed in the inference problem. In our in silico study, we made use of statistical measures such as bias of the median and IQR to evaluate the accuracy, precision, and asymptotic properties of the estimation results and to compare the goodness of results.

From the results, we derived some interesting findings which give rise to guidelines for the integration of WB data into mathematical models. First, concerning the choice of the normalisation strategy, we got a clear recommendation for applying the third strategy, that is, normalisation by the mean value, instead of normalisation by a fixed point. Opting for this normalisation strategy, in fact, leads to the best inference results in terms of accuracy and precision of the estimated parameters, independent from the chosen experimental design and noise level in the input data. We anticipate that this is a generalisable result, which results from the fact that the variance of the mean of datapoints is often smaller than the variance of the individual data points (Figure 2a).

Regarding the experimental design, we analysed the effects of increasing the total amount of the measured data. As expected, increasing the number of time points and/or the number of replicates improves the quality of the estimation. A less expected result was that typical numbers of time points in WB data lead to very broad distributions of the ML estimates, resulting in boundary effects also for large boundaries of the search space. Even a huge increase in the number of data points (up to 120 measurements) in combination with larger boundaries (also in the negative half plane) could not in all cases remove these boundary effects. In particular, this was not possible for LN-EM, hence this EM was neglected in the subsequent analysis. For the same reason, the data set with the high noise level was in the subsequent analysis considered in combination with the largest data set of 120 data points.

The GR-EM seems to be slightly more robust than N-EM for high noise levels, which is probably caused by the heavy-tailedness of the GR distribution, which makes the estimation robust against the outliers in the data [31]. Except from that we observe no significant differences between the N-EM and GR-EM.

Our simulation results highlight the importance of taking noise transformation into account when dealing with data processing techniques like normalisation. This is also supported by Degasperi et al. [3], who compare two approaches to scale model simulations to relative measured data: First, introducing scaling factors to convert the simulated data to the scale of the experimental data (SF approach = scaling factor), or second, normalising simulated variables in the same way as the data (DNS approach = data-driven normalisation of the simulation). The authors tested both methods with different objective functions and optimisation algorithms for the parameter estimation of dynamical systems and concluded that the DNS approach is favourable in terms of identifiability and convergence speed of the optimisation algorithms and should therefore be the preferred method in dynamic modelling studies.

Finally, our results indicate that the transformation of raw data such as normalisation may lead to significant uncertainties in the estimated model parameters, even for large data sets.

## AUTHOR CONTRIBUTION

**Caterina Thomaseth**: Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualisation; Writing – original draft; Writing – Review & Editing. **Nicole E. Radde**: Conceptualisation; Formal analysis; Funding acquisition; Methodology; Project administration; Resources; Supervision; Validation; Writing – original draft; Writing - Review & Editing.

## ACKNOWLEDGEMENTS

This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC2075-390740016. Regarding finite size and boundary effects in inverse problems, we acknowledge input from Prof. Jens Timmer, University of Freiburg.

Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

## DATA AVAILABILITY STATEMENT

Data available on request from the authors.

## ORCID

*Nicole E. Radde* https://orcid.org/0000-0002-5145-0058

## REFERENCES

1. Ashyraliyev, M., et al.: Systems biology: parameter estimation for biochemical models. FEBS J. 276(4), 886–902 (2009). https://doi.org/10.1111/j.1742-4658.2008.06844.x
2. Schmiester, L., et al.: Petab—interoperable specification of parameter estimation problems in systems biology. PLoS Comp. Biol. 17(1), e1008646 (2021). https://doi.org/10.1371/journal.pcbi.1008646
3. Degasperi, A., Fey, D., Kholodenko, B.N.: Performance of objective functions and optimisation procedures for parameter estimation in system biology models. NPJ Syst. Biol. Appl. 3(20), 1–9 (2017). https://doi.org/10.1038/s41540-017-0023-2
4. Gábor, A., Banga, J.R.: Robust and efficient parameter estimation in dynamic models of biological systems. BMC Syst. Biol. 9(74), 74 (2015). https://doi.org/10.1186/s12918-015-0219-2
5. Hass, H., et al.: Benchmark problems for dynamic modeling of intracellular processes. Bioinformatics 35(17), 3073–3082 (2019). https://doi.org/10.1093/bioinformatics/btz020

6. Kreutz, C.: Guidelines for benchmarking of optimization-based approaches for fitting mathematical models. Genome Biol. 20(1), 281 (2019). https://doi.org/10.1186/s13059-019-1887-9
7. Raue, A., et al.: Lessons learned from quantitative dynamical modeling in Systems Biology. PLoS One 8(9), 1–17 (2013). https://doi.org/10.1371/journal.pone.0074335
8. Kirch, J., et al.: The effect of model rescaling and normalization on sensitivity analysis on an example of a MAPK pathway model. EPJ Nonlinear Biomed. Physi. 4(3), 1–23 (2016). https://doi.org/10.1140/epjnbp/s40366-016-0030-z
9. Degasperi, A., et al.: Evaluating strategies to normalize biological replicates of Western blot data. PLoS One 9(1), 1–11 (2014)
10. Fröhlich, F., Loos, C., Hasenauer, J.: Scalable inference of ordinary differential equation models of biochemical processes. In: Sanguinetti, G., Huynh-Thu, V.A. (eds.) Gene Regulatory Networks. Methods and Protocols., Volume 1883 of Methods in Molecular Biology, 1st ed. Humana Press, New York (2019)
11. Kreutz, C., et al.: An error model for protein quantification. Bioinformatics 23(20), 2747–2753 (2007). https://doi.org/10.1093/bioinformatics/btm397
12. Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. London, Ser. A 222, 309–368 (1922)
13. Gelman, A., et al.: Bayesian Data Analysis. Texts in Statistical Science, 2 edition. Chapman & Hall, CRC (2004)
14. Wang, Q.: Analysis of the Effects of Experimental Data Normalization on Model Calibration. Technical report. University of Stuttgart, Institute for Systems Theory and Automatic Control (2018). Study thesis
15. Schilling, M., et al.: Computational processing and error reduction strategies for standardized quantitative data in biological networks. FEBS J. 272(24), 6400–6411 (2005). https://doi.org/10.1111/j.1742-4658.2005.05037.x
16. Jarajapu, Y.P., et al.: Protection of blood retinal barrier and systemic vasculature by insulin-like growth factor binding protein-3. PLoS One 7(7), e39398 (2012). https://doi.org/10.1371/journal.pone.0039398
17. Jensch, A., et al.: The tumor suppressor protein DLC1 maintains protein kinase D activity and Golgi secretory function. J. Biol. Chem. 293(37), 14407–14416 (2018). https://doi.org/10.1074/jbc.RA118.003787
18. Santos, S., Verveer, P., Bastiaens, P.: Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. Nat. Cell Biol. 9(3), 324–330 (2007). https://doi.org/10.1038/ncb1543
19. Talavera, M.M., et al.: Mitogen-activated protein kinase phosphatase-1 prevents lipopolysaccharide-induced apoptosis in immature rat intestinal epithelial cells. Pediatr. Res. 78(2), 128–136 (2015). https://doi.org/10.1038/pr.2015.88
20. Wang, W., et al.: Role of TLR4-p38 MAPK-Hsp27 signal pathway in LPS-induced pulmonary epithelial hyperpermeability. BMC Pulm. Med. 18(178), 178 (2018). https://doi.org/10.1186/s12890-018-0735-0
21. Kreutz, C., Timmer, J.: Systems biology: experimental design. FEBS J. 276(4), 923–942 (2009). https://doi.org/10.1111/j.1742-4658.2008.06843.x
22. Weber, P., et al.: Parameter estimation and identifiability of biological networks using relative data. Proc. 18th World Congress the Inter. Fed. Automat. Contr. 44(1), 11648–11653 (2011). https://doi.org/10.3182/20110828-6-it-1002.01007
23. Limpert, E., Stahel, W.A., Abbt, M.: Log-normal distributions across the sciences: keys and clues. Bioscience 51(5), 341–352 (2001). https://doi.org/10.1641/0006-3568(2001)051[0341:lndats]2.0.co;2
24. Thomaseth, C., et al.: Modeling sphingomyelin synthase 1 driven reaction at the Golgi apparatus can explain data by inclusion of a positive feedback mechanism. J. Theor. Biol. 337, 174–180 (2013). https://doi.org/10.1016/j.jtbi.2013.08.022
25. Hayya, J., Armstrong, D., Gressis, N.: A note on the ratio of two normally distributed variables. Manag. Sci. 21(11), 1338–1341 (1975). https://doi.org/10.1287/mnsc.21.11.1338
26. Hinkley, D.V.: On the ratio of two correlated normal random variables. Biometrika 56(3), 635–639 (1969). https://doi.org/10.1093/biomet/56.3.635
27. Hinkley, D.V.: Correction: on the ratio of two correlated normal random variables. Biometrika 57(3), 683 (1970). https://doi.org/10.2307/2334796
28. Marsaglia, G.: Ratios of normal variables and ratios of sums of uniform variables. J. Amer. Stat. Assoc. 60(309), 193–204 (1965). https://doi.org/10.1080/01621459.1965.10480783
29. Marsaglia, G.: Ratios of normal variables. J. Stat. Software 16(4), 1–10 (2006). https://doi.org/10.18637/jss.v016.i04
30. Gutenkunst, R.N., et al.: Universally sloppy parameter sensitivities in systems biology models. PLoS Comput. Biol. 3(10), e189 (2007). https://doi.org/10.1371/journal.pcbi.0030189
31. Maier, C., Loos, C., Hasenauer, J.: Robust parameter estimation for dynamical systems from outlier-corrupted data. Bioinformatics 33(5), 718–725 (2017). https://doi.org/10.1093/bioinformatics/btw703


## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Thomaseth, C., Radde, N.E.: The effect of normalisation and error model choice on the distribution of the maximum likelihood estimator for a biochemical reaction. IET Syst. Biol. 17(1), 1–13 (2023). https://doi.org/10.1049/syb2.12055