

PEPSDI: Scalable and flexible inference framework for stochastic dynamic single-cell models

Sebastian Persson¹, Niek Welkenhuysen¹, Sviatlana Shashkova², Samuel Wiqvist³, Patrick Reith^{1,4}, Gregor W. Schmidt⁵, Umberto Picchini^{1*}, Marija Cvijovic^{1*}

1 Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Sweden.

2 Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Sweden.

3 Centre for Mathematical Sciences, Lund University, Sweden.

4 Department of Biology and Biological Engineering, Chalmers University of Technology, Sweden.

5 Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

* Corresponding Authors:

Umberto Picchini, e-mail: picchini@chalmers.se

Marija Cvijovic, e-mail: marija.cvijovic@chalmers.se

Abstract

Mathematical modelling is an invaluable tool to describe dynamic cellular processes and to rationalise cell-to-cell variability within the population. This requires statistical methods to infer unknown model parameters from dynamic, multi-individual data accounting for heterogeneity caused by both intrinsic and extrinsic noise. Here we present PEPSDI, a scalable and flexible framework for Bayesian inference in state-space mixed-effects stochastic dynamic single-cell models. Unlike previous frameworks, PEPSDI imposes a few modelling assumptions when inferring unknown model parameters from time-lapse data. Specifically, it can infer model parameters when intrinsic noise is modelled by either exact or approximate stochastic simulators, and when extrinsic noise is modelled by either time-varying, or time-constant parameters that vary between cells. This allowed us to identify hexokinase activity as a source of extrinsic noise, and to deduce that sugar availability dictates cell-to-cell variability in the budding yeast *Saccharomyces cerevisiae* SNF1 nutrient sensing pathway.

Traditionally, investigations in the life sciences have focused on a population “ensemble average” level. On one side, such population approach reduces noise from atypical cells. However, any cellular population is in general heterogeneous, with a range of different physical, chemical, and biological properties. Thus, population methods smooth out and hence miss biologically relevant cell-to-cell variability [1]. For example, such approaches will overlook drug-resistant bacteria or cancer cells in a general cell populations. Furthermore, cell-to-cell variability plays an important role in the decision making of a population, such as quick adaptation to fluctuating environments [2]. The only way to identify all biologically relevant processes, and thus describe cell heterogeneity, is to study the whole population cell-by-cell [3].

To study life processes occurring in individual cells within the population, fluorescent time-lapse microscopy can be employed to track proteins in multiple cells over time [1]. Ideally, this can give a view of cell heterogeneity, and potentially help elucidate cellular reaction dynamics. However, to learn more from acquired data, dynamic modelling naturally complements time-lapse microscopy, and aids in deducing sources of cell-to-cell variability [4, 5, 6]. But to fully exploit dynamic modelling, unknown model parameters must typically be inferred/estimated from data [7]. This is non-trivial to perform from single-cell time-lapse data, mainly because models describing individual cells must account for cell-to-cell variability caused by both intrinsic (e.g. variations in chemical reactions) and extrinsic (e.g. variability in protein concentrations) noise [8].

Several inference methods exist that, to various degrees, account for cell-to-cell variability when deducing model parameters from time-lapse data. In common, they allow extrinsic noise to be modelled by letting model parameters, e.g protein synthesis rates, vary between cells [4, 6, 9, 10]. Methods based on ordinary differential equations (ODEs) [6] further assume that intrinsic noise is negligible. On the other side, the dynamic prior propagation (DPP) [4] and the stochastic differential equation mixed-effects models (SDEMEM) [9, 10] encode intrinsic noise via exact [11], or approximate [12] stochastic simulators, respectively.

Although useful, current inference methods have drawbacks. The fact that ODE based methods assume intrinsic noise to be negligible is often hard to justify. The DPP method imposes multiple model assumptions, such as time invariant rate constants. The SDEMEM methods employ approximate simulators to model intrinsic noise, that can be inaccurate when few molecules control the dynamics [13]. Overall, available frameworks only address specific questions. Moreover, there are scenarios where all existing methods are inadequate. For example, when studying a gene expression model with low numbers of molecules and a time-varying transcription rate using available inference approaches will impose unrealistic model assumptions, potentially leading to incorrect model prediction [14].

To fully exploit the power and facilitate usage of stochastic single-cell dynamic models, we propose, for the first time, a flexible Bayesian inference framework for stochastic dynamic mixed-effects models. By building upon a Bayesian inference framework originally thought for SDEMEMs [10], we introduce a novel, computationally more efficient inference method, that is capable of inferring unknown model parameters when intrinsic noise is modelled by either exact [11, 15], or approximate [12, 13] stochastic simulators. Moreover, by leveraging on the state-of-the-art statistical methods [16, 17], our framework allows for large flexibility in how extrinsic noise is modelled. Using synthetic examples, we show how this flexibility facilitates understanding of a stochastic gene expression model regulated by an extrinsic time-varying signal and cellular pathways where intrinsic noise causes cells to migrate between states. Further, by combining time-lapse microscopy with microfluidics, we employ our inference framework to distinguish between multiple network structures and identify sources of cell heterogeneity in the budding yeast *Saccharomyces cerevisiae* SNF1 nutrient sensing pathway.

Results

Inference framework for stochastic dynamic single-cell models

We developed a flexible modelling framework, PEPSDI (**P**articles **E**ngine for **P**opulation **S**tochastic **D**ynam**I**cs), that infers unknown model parameters from dynamic data for single-cell dynamic models that account for both intrinsic and extrinsic noise (Fig. 1). The latter can be modelled by letting parameters believed to vary between cells, such as protein translation rates, follow a probability distribution. Furthermore, the model parameters can incorporate extrinsic time variant signals, such as the circadian clock [18], and measured extrinsic data, such as cell volume. Intrinsic noise can be modelled by multiple stochastic simulators, specifically the exact SSA (Gillespie) and Extrande simulators [11, 15], and the approximate tau-leaping [13] and Langevin simulators [12]. Hence, PEPSDI is applicable for gene expressions models with low numbers of molecules [19], and signalling models where large numbers of molecules can make exact simulators unfeasible [13]. This framework can further infer the strength of the measurement error, and is suitable when either all, or a subset of the model components are observed. More formally, PEPSDI produces Bayesian inference for state-space models with latent dynamics incorporating mixed-effects, that is state-space mixed-effects model (SSMEM). It builds upon the schemes previously proposed for SDEMEMs [10].

From the methodological point of view, PEPSDI is a Gibbs sampler targeting the full posterior distribution of all unknowns. To allow large flexibility in the model construction, some of the Gibbs-steps can be targeted using Hamiltonian Monte Carlo (HMC) [17]. For example, it can be assumed that the synthesis and breakdown rates of a protein follow a log-normal distribution and if correlation between the rates is suspected, *a-priori*, the HMC sampler permits efficient inference of the correlation [20]. For the Gibbs-steps where the likelihood function is intractable, PEPSDI uses a pseudo-marginal approach employing particle filters [16]. This enables the user to select from a wide range of stochastic simulators [13]. Furthermore, for computational efficiency we employ, when possible, correlated particle filters [21], and tune the parameters proposal distributions using adaptive algorithms [22, 23, 24].

PEPSDI can be run with two options. The default is to allow cell constant parameters to vary weakly between cells, and the second option is to keep cell constant parameter fixed. The former speeds up the inference by a factor of 35 compared to the latter (Supplementary Fig. S3), and larger speed ups are likely attainable when the number of observed cells increase.

PEPSDI is written in Julia [25], and is available on GitHub (<https://github.com/cvijoviclab/PEPSDI>). To encourage usage of our framework, the provided code is flexible with regards to the model structure and all examples are available as notebooks (<https://github.com/cvijoviclab/PEPSDI/tree/main/Code/Examples>). Guidelines on how to run the inference schemes are presented in Supplementary Note 3, and a tutorial in Supplementary Note 4.

Application to simulated circadian clock gene expression model

We applied the developed inference framework on simulated data from a simple circadian clock [18] gene expression model (Fig. 2a). The circadian clock was modelled as a sine function regulating the transcription activity, causing the protein levels to oscillate (Fig. 2b). Moreover, to simulate strong intrinsic noise, the numbers of molecules were kept low.

Additional extrinsic noise was simulated by letting the cell-specific rate constants follow a multivariate log-normal distribution; $\mathbf{c}^{(i)} \sim \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are the mean and covariance matrix of the associated Gaussian distribution. Its multiplicative nature [26] makes the log-

normal distribution a common choice for modelling parameters believed to vary between cells [6]. We also correlated the rate parameters to emulate that protein synthesis and degradation rates might co-vary [5, 27] (Supplementary Note 5).

The intrinsic noise was modelled using the exact Extrande simulator [15]. Even though faster approximate algorithms can be used to model intrinsic noise, low numbers of molecules in the model make these impractical [13].

We ran PEPSDI for 50,000 iterations, and recovered the true model parameters of transcription rate, and how it varies within the cell population (Fig. 2d and Supplementary Fig. S1). There is a slight bias in the cell-to-cell variability of the breakdown rates (τ_2, τ_4), but, since the model accurately describes the data (Fig. 2c), this is likely because only 40 cells are observed.

The inferred model has predictive power. We considered an experiment where an additional gene is inserted into the model, and time-lapse data is collected for 40 new cells. The model accurately predicts an increase in both protein, and cell-to-cell variability levels (Fig. 2e).

Application to simulated stochastic bistable model

Intrinsic noise can have a strong impact on cellular dynamics [28, 29]. Cellular processes have been shown to exhibit stochastic oscillations in gene regulation [28] and stochastic bistability as reported for the *lac*-operon regulation [29]. To study the performance of PEPSDI for such a process, we implemented the Schlögl model (Fig. 3a) [30], where cells stochastically migrate between states of high and low gene expression (Fig. 3b).

To simulate extrinsic noise, we let one of the synthesis rates, c_3 , follow a log-normal distribution. To emulate that certain model parameters, such as protein dissociation rates, can have a neglectable variability [4], we kept a synthesis (c_1) and dissociation rate (c_2) constant between individuals. The synthesis rate c_4 was assumed to be known (Supplementary Note 5).

To account for large numbers of molecules, we modelled intrinsic noise via the fast, approximate, Langevin simulator. We then used so-called guided proposals [9], directing simulations towards observed values, making the inference more efficient for models with stochastic events.

We ran PEPSDI for 500,000 iterations, and posterior distributions of the modelled-quantities were recovered (Fig. 3c), allowing for better understanding of mechanistic properties of the model. For example, by simulating the inferred model, it becomes apparent that low values of the synthesis rate c_1 commit cells to a first cell-state, e.g. of low gene-expression, while for large values of c_1 , cells commit to, and jump between, two different states of gene expression (Supplementary Fig. S2).

Performance evaluation of adaptive MCMC proposals

To propose unknown model quantities that rely on a pseudo-marginal step in our Gibbs sampler (see Methods), PEPSDI employs adaptive Markov chain Monte Carlo (MCMC) schemes [23]. However, these schemes were developed for problems where the likelihood function is not approximated via Monte Carlo scheme, as is the case for pseudo-marginal methods. Thus, we benchmarked three adaptive MCMC proposal schemes, namely adaptive metropolis (AM) [22], the AM with global scaling (Alg. 4 in [23]), and the robust AM (RAM) samplers [24]. For computational reasons, single time-series inference for two different models is considered.

For the Schlögl model (Fig. 3), the stochastic bistability (Fig. 3b) causes the likelihood approximation used in the pseudo-marginal method to have a large variance despite the usage of

guided proposals. To investigate if this impacts the performance of adaptive MCMC schemes, we launched multiple inference runs. Overall, the RAM sampler on average had the highest multiple effective sample size (MultiESS) value [31], and thus most efficiently explored the posterior surface (Fig. 4a). This sampler also had the smallest variability in MultiESS, suggesting it is robust against, for example, bad start guesses (start guess 3 and 4 in Fig. 4a). Moreover, the RAM sampler typically required fewer particles, implying higher computational efficiency.

For the Ornstein-Uhlenbeck stochastic differential equation model (Supplementary Note S5), the likelihood approximation has a small variance. Using the same setup as for the Schlögl model, the RAM sampler in this setting also performed the best.

Application to glucose repression pathway

After considering synthetic examples, we set out to study the SNF1 pathway in *S.cerevisiae*. The SNF1 complex, and its mammalian homolog AMPK, play a major role in both metabolic regulation and maintenance of cellular homeostasis. In response to stress, such as ageing and nutrients limitation, SNF1 mediates the signal transduction to transcription factors. Mig1 is a transcriptional repressor which the SNF1 complex deactivates when energy-rich carbon sources are limited. This is followed by Mig1 relocation to the cytoplasm and release of repression of genes responsible for utilisation of alternative carbon sources [32, 33]. If the amount of energy-rich nutrients is elevated, Mig1 translocates to the nucleus [34]. This process is accompanied by Mig1 dephosphorylation where the Reg1 phosphatase plays the main role [1]. While this signalling cascade is well studied, it has also been indicated that the SNF1 pathway dynamics exhibits a high range of cell-to-cell variability [35].

To deduce the dynamics of the SNF1 pathway at a cell level, time-resolved data is required. We utilised fluorescent microscopy to follow Mig1 on a single-cell level and observe its localisation over time (Fig. 5a-b). This was coupled with microfluidics systems enabling high control of the cell environment. Mig1 localisation was followed during 15 min after cells were switched from carbon sourced depleted conditions to two different fructose conditions (w/v): 0.05%, and 2% (Fig. 5c).

Modelling of Mig1 nuclear dynamics

To elucidate both the source of cell heterogeneity and reaction mechanisms behind the Mig1 dynamics data (Fig. 5c), we formulated and analysed two network structures (Fig. 6a and b) and two extrinsic noise sources (Fig. 6c and d), resulting in four plausible models. In both network structures, Mig1 shuttles between the cytosol and nucleus in a carbon source-dependent [36] and -independent manner [37]. The carbon source-dependent response has been observed to occur in two phases, a transient initial Mig1 nuclear entry followed by nucleocytoplasmic shuttling [38]. This can be explained by two independent pathways regulating glucose derepression. One pathway activating the Snf1 kinase and the other pathway directing Snf1 towards Mig1 [39]. We investigated whether the first pathway promotes Mig1 nuclear entry via a fast signal (Fig. 6a), or if Mig1 nuclear entry is delayed due to Reg1 activation (Fig. 6b). The second pathway was modelled via an unknown metabolic component due to the fact that Mig1 dynamics are closely intertwined with metabolic activity [35, 38] (Supplementary Note 6).

To model the extrinsic noise sources, we assume that the Mig1 initial values ($Mig1c_{t_0}$, $Mig1n_{t_0}$) and the strength of the Mig1 regulation (c_1 , c_2) vary between cells (Fig. 6c), as the latter incorporates multiple processes in the initial glycolysis. Since the rates (c_1 , c_2) also encompass fructose abundance we assume and infer fructose-dependent log-mean values. The second ex-

intrinsic noise source proposal (Fig. 6d) takes into account that the cell-to-cell variability might be fructose-dependent.

We employed literature supported priors (Supplementary Note 6) and ran PEPSDI multiple times for each proposed model. Models were compared using posterior visual check [40], that is the capability to capture both the observed trend and cell-to-cell variability (Fig. 6e, f and Supplementary Fig. S4). Overall, the model with a delayed fructose activation of Mig1 nuclear export and a fructose-dependent cell-to-cell variability in Mig1 regulation best described the data (model 2B Fig. 6f).

Cell-to-cell variability in Mig1 nuclear dynamics is tightly regulated by fructose availability

After selecting the best model, we investigated the characteristics of the inferred model parameters. The coefficient of variations for the Mig1 regulation (c_1, c_2) shows that these rates substantially differ between cells (Fig. 6h). This suggests that extrinsic noise has a dominating role in Mig1 localisation dynamics. Furthermore, the magnitude of the variation, particularly for rate of activation of Reg1 via fructose (c_1), is larger for low fructose conditions.

Our results show a correlation between c_1, c_2 rates (Fig. 6h), implying that cells with a strong fructose-dependent nuclear import of Mig1, also have a stronger nuclear export. By simulating the model without correlation we confirm that this correlation regulates cell-to-cell variability (Fig. 6g). Thus, our result suggests that co-regulation of two fructose-dependent pathways controls cell heterogeneity in Mig1 localisation dynamics.

The log-mean value of the rate parameter activating the metabolic component ($\log\text{-mean}(c_2)$) shows that the magnitude of the long-term nuclear export of Mig1 is weaker in high fructose (Fig. 6h). This is consistent with previous reports that nuclear export of Mig1 is primarily an effect of sugar depletion [36].

The hexokinase Hxk1 is a source of cell heterogeneity in Mig1 localisation dynamics

In model simulations, the cell-varying rate constant c_1 linearly correlates with the short-term (0-15 min) Mig1 localisation upon fructose addition to sugar-starved cells (Supplementary Fig. S4). Thus, since c_1 captures a signalling process acting on Reg1 from the initial hexose metabolism via the hexokinase Hxk1, our modelling suggests a relationship between Hxk1 and such cell-to-cell variability.

To validate this prediction, we collected single-cell time-lapse microscopy data from a *hxx1Δhxx2Δ* strain carrying Hxk1-expressing plasmids where both Mig1 localisation and Hxk1 expression are monitored upon fructose addition. In line with model predictions, the observed Hxk1 expression linearly correlates with Mig1 localisation, and the magnitude of variability in localisation explained by Hxk1 (linear regression R^2 -value) matches the predicted magnitude (Supplementary Fig. S4). We hence conclude Hxk1 as a source of cell heterogeneity in Mig1 dynamics.

Discussion

Understanding the inherited nature of how biological processes dynamically change over time, and exhibit intra- and inter-individual variability, has been a major focus of systems biology.

The rise of single-cell fluorescent microscopy has enabled the study of those phenomena, but further progress is limited by the availability of methods that facilitate modelling, the essential follow up for rationalisation of such data. To address this, we developed PEPSDI, a versatile framework for Bayesian inference for dynamic stochastic single-cell models. We used PEPSDI to recover true model quantities for a circadian clock stochastic gene expression model, and deduce mechanistic details for a model where cells stochastically move between two states of gene expression. We hereafter studied the SNF1 signalling in yeast and identified hexokinase activity as a source of extrinsic noise, and deduced that sugar availability dictates cell-to-cell variability.

Modelling Mig1 dynamics (Fig. 6) suggests larger cell heterogeneity in the fructose activation of Reg1 (c_1) and the component that regulates Mig1 shuttling (c_2) upon fructose limitation (Fig. 6h). We hypothesise the presence of a sugar-dependent biological switch, triggered when fructose is present. This is consistent with the fact that under low fructose, only a few cells due to extrinsic noise activate the switch leading to larger cell-to-cell variability.

Our modelling also suggests cell variability in the Mig1 nuclear export. In particular, in the rate constant c_2 which activates the metabolic component regulating Mig1 nuclear export. Although this process is slower than the Reg1 activation (c_1), they are strongly correlated. Moreover, results show that the magnitude of c_2 is larger under fructose limitation (Fig. 6h). It has been reported that rapid activation of the SNF1 complex upon glucose starvation is ensured by increased levels of ADP [41]. We therefore suggest that within the SNF1 pathway, ADP is a central part of the c_2 rate constant, facilitating nuclear export of Mig1 via activation of the SNF1 complex. This is consistent with elevated ADP levels upon limited glucose or fructose concentrations in the cellular environment. Moreover, hexokinases that we identified to be a part of the c_1 rate, regulate Reg1, which in turn also requires the abundant presence of a hexose sugar for full activity of the phosphatase [42]. Taken together, we hypothesise that c_1 and c_2 incorporate hexokinases and ADP, respectively, and are both controlled via the metabolism. This creates a tight co-regulation of the SNF1 pathway, which regulates cell-to-cell variability (Fig. 6g).

To study the single-cell behaviour of the Mig1 dynamics, we leveraged on the modifiable modular nature of PEPSDI. This modularity facilitates modelling of intrinsic noise by either the SSA [11], Extrande [15], tau-leaping [43] or Langevin [12] stochastic simulators. Additionally, new modules such as the hybrid-simulators [44] used to study NF κ B-pathway [45] can be incorporated. Likewise, new pseudo-marginal modules can be added, and considering that these schemes are actively researched [46, 47], the performance of the framework can improve further.

Besides being modifiable, we aimed to make PEPSDI accessible, by providing extensive tutorial notebooks on both how to use it, and how to leverage the underlying pseudo-marginal framework to model single time-series data. Additionally, to help users of pseudo-marginal inference we evaluated adaptive Markov chain Monte Carlo (MCMC) proposals [22, 23, 24], resulting in the best performance of the RAM sampler [24] (Fig. 4). However, the evaluation is performed on two models, and further analysis is required before generalising our conclusions.

In summary, we developed and employed PEPSDI to deduce the reaction dynamics, and sources of cell heterogeneity behind single-cell time-lapse data. Since PEPSDI is an inference framework for dynamic state-space mixed-effects models, it can also be applied for problems arising in ecology [9], in neuroscience [10] and in pharmacokinetics and pharmacodynamics (PKPD) [48]. With the versatility of applications for state-space mixed-effects models, and with our framework that does not impose strict model assumptions, it is easy to envision additional applications such as modelling the behaviour of cancer cell populations [49]. We believe that this framework will play an increasingly important role in addressing challenging biological ques-

tions that cannot be answered by experimental approaches only, thus providing novel insights for better understanding life processes.

Materials and methods

Stochastic simulations

Under the assumption that the system is well mixed, the dynamics of a stochastic reaction-network of d species, $\mathbf{X} = (X_1(t), \dots, X_d(t))$, is described by the chemical master equation [13, 43]. The propensity function $h(\mathbf{x}, \mathbf{c}, t)$ is a measure of the probability that one, out of R , reactions will occur depending on \mathbf{X} and the rate-constants \mathbf{c} . Considering that the master-equation can rarely be solved due to its probabilistic nature [13, 43], our inference framework relies on simulating from it.

The SSA direct method [11] produces exact stochastic simulations. Similarly, the Extrande simulator [15] produces exact solutions when the rate-constants \mathbf{c} are time variant. However, both approaches simulate each reaction event and can thus be slow for large propensities [50]. Assuming that the propensities do not change noticeably during the time interval $[t, t + \tau]$ (leap condition 1), the reactions will be close to independence of each other. Hence, the tau-leap approach with fixed step-length can be used to update the states-vector from time t to $t + \tau$ [12, 51]. Further assuming that the propensities are sufficiently large in $[t, t + \tau]$ (leap conditions 2), the state-vector can be updated via the chemical Langevin stochastic differential equation [12, 43]. Typically, the number of molecules must be large for leap condition 1 and 2 to hold.

Our framework allows models to use the SSA method and the Extrande method when neither leap conditions holds. When one or both conditions holds, our framework allows usage of either tau-leaping or the Langevin approach.

PEPSDI: Bayesian inference for single-cell dynamic models

PEPSDI performs Bayesian inference for state-space models with latent dynamics incorporating mixed-effects, shortly "state-space mixed-effects models" (SSMEMs). A state-space model is a discrete-time, stochastic model that contains two sets of equations: (i) one describing how a latent Markov process transitions in time (the state equation) and (ii) another one describing how an observer measures the latent process at each discrete time-point (the observation equation), assuming conditional independence between observations and latent states. Thus, a state-space model can describe a stochastic chemical reaction network. Here, we outline PEPSDI, and a more complete description also showing pseudo-algorithms is provided in Supplementary Note S1.

To perform inference, PEPSDI requires single-cell time-lapse measurements $\mathbf{y}_l^{(i)}$ for the i -th individual collected at $(l = 1, \dots, n_i)$ discrete time-points $t^{(i)} = (t_1^{(i)}, \dots, t_{n_i}^{(i)})$ for $i = 1, \dots, M$ individuals. Note, from now we denote with "individual" the measurements from a single cell. However since PEPSDI can be used in other applied areas (e.g. ecology, PKPD, etc), an individual is more generally a unit from the population of interest. The individual data is assumed to be noise-corrupted as $\mathbf{y}_l^{(i)} = \mathbf{g}(\mathbf{x}_l^{(i)}, \boldsymbol{\epsilon}_l^{(i)})$, where we use the shorthand notation $z_l^{(i)}$ to denote a variable z observed at time $t_l^{(i)}$. Here, $\boldsymbol{\epsilon}_l^{(i)}$ is the measurement error which follows an error distribution $\boldsymbol{\epsilon}_l^{(i)} \sim \pi_{\boldsymbol{\epsilon}}(\boldsymbol{\xi})$, and $\mathbf{g}(\cdot)$ is a (possibly non-linear) function of its arguments. To infer parameters in the nutrient-sensing Mig1 pathway, the observed Mig1 data represents

a ratio of nuclear to cytosolic intensity, and thus $g(\mathbf{x}_l^{(i)}, \epsilon_l^{(i)}) = Mig1N(t)/Mig1C(t) + \epsilon_l^{(i)}$, with independent $\epsilon_l^{(i)} \sim \mathcal{N}(0, \sigma^2)$.

For a SSMEM, PEPSDI infers the individual rate-constants $\mathbf{c} = (\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)})$, cell-constant rate-parameters $\boldsymbol{\kappa}$, the parameters $\boldsymbol{\xi}$ for the measurement error, and the population parameters $\boldsymbol{\eta}$. The population parameters describe the distribution of the individual parameters, $\mathbf{c}^{(i)} \sim \pi(\mathbf{c}|\boldsymbol{\eta})$. For example if $\mathbf{c}^{(i)}$ follows a log-normal distribution, the population parameter corresponds to $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Omega})$ and individual rate constants to $\mathbf{c}^{(i)} \sim \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Omega})$. Overall, the posterior distribution we target is:

$$\pi(\mathbf{c}, \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}|\mathbf{y}) \propto \pi(\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)}, \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}) \prod_{i=1}^M p(\mathbf{y}^{(i)}|\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}). \quad (1)$$

Note, in Eq. 1 we have assumed that measurements from different subjects are conditionally independent, given the individual-specific $\mathbf{c}^{(i)}$ and the population parameters.

The posterior (Eq. 1) is high-dimensional, and ideally we could sample from it via a Gibbs-sampler [52] by looping through the following three steps:

$$\begin{aligned} 1. \pi(\mathbf{c}|\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \mathbf{y}) &\propto \prod_{i=1}^M \pi(\mathbf{c}^{(i)}|\boldsymbol{\eta}) \pi(\mathbf{y}^{(i)}|\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}) \\ 2. \pi(\boldsymbol{\kappa}, \boldsymbol{\xi}|\mathbf{c}, \boldsymbol{\eta}, \mathbf{y}) &\propto \pi(\boldsymbol{\kappa}, \boldsymbol{\xi}) \prod_{i=1}^M \pi(\mathbf{y}^{(i)}|\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}) \\ 3. \pi(\boldsymbol{\eta}|\mathbf{c}, \boldsymbol{\kappa}, \boldsymbol{\xi}, \mathbf{y}) &\propto \pi(\boldsymbol{\eta}) \prod_{i=1}^M \pi(\mathbf{c}^{(i)}|\boldsymbol{\eta}). \end{aligned} \quad (2)$$

Notice that it is possible to sample from the first step by independently sampling for each of the $\mathbf{c}^{(i)}$ separately from the other ones. That is step 1 can be written as

$$1. \pi(\mathbf{c}^{(i)}|\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \mathbf{y}) \propto \pi(\mathbf{c}^{(i)}|\boldsymbol{\eta}) \pi(\mathbf{y}^{(i)}|\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}), \quad i = 1, \dots, M \quad (3)$$

and hence the sampling step for each $\mathbf{c}^{(i)}$ only needs to access the corresponding individual-specific $\pi(\mathbf{y}^{(i)}|\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi})$.

However, in practice, steps 1-2 cannot trivially be sampled from, due to the intractability of the likelihood for the i -th individual $\pi(\mathbf{y}^{(i)}|\boldsymbol{\kappa}, \boldsymbol{\xi}, \mathbf{c}^{(i)})$, and here sampling is performed using a pseudo-marginal approach following [10]. The posterior targeted in step 3 is tractable, and thus $\boldsymbol{\eta}$ is sampled using Hamiltonian Monte Carlo [17, 53, 54].

PEPSDI can be run with two flavours. Both sample the conditionals in Eq. 2 via, when required, pseudo-marginal approaches. However, the default option is to slightly perturb the SSMEM. This prevents the need of step 2 in the Gibbs-sampler, resulting in substantially shorter run-time. To properly motivate this perturbation, we first cover pseudo-marginal particles-based inference.

Pseudo-Marginal particles-based inference

The pseudo-marginal Metropolis-Hastings scheme samples from the desired posterior by considering the marginal of an augmented posterior [16, 55]. More details are given in Supplementary

Note 1. Briefly, the pseudo-marginal approach considers

$$\pi(\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \mathbf{u}^{(i)} | \mathbf{y}^{(i)}) \propto \pi(\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}) \pi(\mathbf{c}^{(i)} | \boldsymbol{\eta}) \hat{\pi}_{\mathbf{u}^{(i)}}(\mathbf{y}^{(i)} | \mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}) \pi(\mathbf{u}^{(i)}) \quad (4)$$

where often (though not necessarily) parameters can be a-priori independent $\pi(\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}) = \pi(\boldsymbol{\kappa})\pi(\boldsymbol{\eta})\pi(\boldsymbol{\xi})$, and the $\mathbf{u}^{(i)} \sim \pi(\mathbf{u}^{(i)})$ are auxiliary variables used to obtain an unbiased estimate $\hat{\pi}_{\mathbf{u}^{(i)}}(\mathbf{y}^{(i)} | \mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi})$ of $\pi(\mathbf{y}^{(i)} | \mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi})$, that is

$$\mathbb{E}_{\mathbf{u}^{(i)}} [\hat{\pi}_{\mathbf{u}^{(i)}}(\mathbf{y}^{(i)} | \mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi})] = \pi(\mathbf{y}^{(i)} | \mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\xi}), \quad (5)$$

where $\mathbb{E}_{\mathbf{u}^{(i)}}(\cdot)$ means that the expectation is taken with respect to the distribution of the $\mathbf{u}^{(i)}$.

Thanks to the (assumed) unbiasedness of the estimated likelihood, the marginal of the augmented posterior is the *exact* posterior of interest

$$\int \pi(\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi}, \mathbf{u}^{(i)} | \mathbf{y}^{(i)}) d\mathbf{u}^{(i)} = \pi(\mathbf{c}^{(i)}, \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{y}^{(i)}), \quad (6)$$

even though an estimated likelihood term has been employed inside Eq. 4. An efficient way to obtain an (non-negative) unbiased likelihood estimate for state-space models is to use a sequential Monte Carlo procedure known as the particle filter [56]. The particle filter approximates unbiasedly [57, 58] the expectation in Eq. 5 by using N Monte Carlo draws that in this context are named “particles”. The variance of the estimated likelihood decreases when increasing N , and typically the success of pseudo-marginal approaches relies on having a small likelihood-variance [47, 59, 60]. That is using too few particles yields inefficient inference, but on the other hand run-time increases with N . However, the result that makes pseudo-marginal powerful is that, from a theoretical point of view, it provides *exact* Bayesian inference [16, 55], regardless the number of particles employed, thanks to Eq. 6.

To employ as few particles as possible, the following three strategies are considered. Firstly, we induce correlation in the particles between subsequent iterations, and this still preserves exact Bayesian inference [21, 61]. However, this is only feasible for Poisson or Langevin integrators. Secondly, our framework implements a particles tuning scheme. Thirdly, guided particle proposals [9, 62] are used when possible (for details see Supplementary Note S1).

From the considerations above, PEPsDI performs exact Bayesian inference for the model parameters. However, to further reduce the computational requirements to run the inference scheme, we developed a Gibbs-sampler that targets a slightly perturbed parameterisation of a SSMEM-model, which we now describe.

Inference for perturbed SSMEM (default option in PEPsDI)

The “perturbed SSMEM” treats cell-constant parameters $(\boldsymbol{\kappa}, \boldsymbol{\xi})$ as parameters that instead vary, with a small fixed variance, between cells. For example, for $\boldsymbol{\xi}$ this means that for the perturbed SSMEM we assume to have $\boldsymbol{\xi}^{(i)} \sim \mathcal{N}(\boldsymbol{\xi}_{pop}, \delta \cdot \mathbf{I})$, with $\delta > 0$ a fixed constant selected by the researcher (and using a similar reasoning for $\boldsymbol{\kappa}^{(i)}$). In some sense $(\boldsymbol{\kappa}, \boldsymbol{\xi})$ have been “perturbed” to artificially vary between cells. This parameterisation, inspired by how cell-constant parameters can be treated in Monolix [63], allows $(\boldsymbol{\xi}_{pop}, \boldsymbol{\kappa}_{pop})$ to be inferred alongside with the population parameters $\boldsymbol{\eta}$ in step 3, and $(\boldsymbol{\xi}^{(i)}, \boldsymbol{\kappa}^{(i)})$ to be inferred alongside $\mathbf{c}^{(i)}$ in step 1. Step 2 is thus avoided.

Avoiding step 2 is desirable from the computational point of view. This is because step 2 of Gibbs-sampler (Eq. 2) requires a stochastic approximation of the sum of the individual

log-likelihoods (for numerical stability it is preferable to work on log-transformed quantities), enabled by a particle filter. The variance of this sum is typically large, since each element of the sum is a stochastically approximated log-likelihood. To achieve a small-variance for this log-likelihood, and thus efficient inference, many particles are required which can cause substantial run-time (Supplementary Fig. S3).

As seen for the Ornstein-Uhlenbeck model, the Gibbs sampler corresponding to the perturbed SSMEM occasionally produces slightly wider credibility intervals for $(\kappa_{\text{pop}}, \xi_{\text{pop}})$ (Supplementary Fig. S5). However, we consider this a worthwhile compromise since we do not observe any bias, while reaching a speed sometimes larger than a factor 30 (Supplementary Fig. S3). Moreover, if necessary PEPSDI can be run with the perturbed SSMEM to rapidly obtain preliminary results from pilot-runs, and the latter can be used to inform the setup (e.g. starting parameter values) to launch the inference for the unperturbed SSMEM.

Single-cell microscopy data

A glass bottom petri-dish (GWST-5030, WillCo Wells) was treated with Poly-L-Lysine solution (P4832, Sigma-Aldrich) for 15 min at room temperature. The Poly-L-Lysine solution was removed, and the petri-dish was washed with MQ water for 2 times and left to dry overnight. Yeast cells (W303(202) NRD1-mCherry- Hph MIG1-GFP-KanMX) were grown overnight to mid-log phase at 30 °C in YNB synthetic complete medium (formedium) (6.7 g/l yeast nitrogen base with ammonium sulphate (formedium), 790 mg/l complete supplement mix (formedium) and supplied with 3% ethanol). These mid-log phase cells were added to the petri-dish and left to sediment, cell which did not adhere to the surface were removed by washing with growth media. Exposure of cells to YNB media with different concentrations of fructose was performed by using a BioPen system with BioPen prime pipette tip (Fluicell AB) as described in previous work [64]. Imaging was performed on an inverted microscope a Leica DMi8 inverted fluorescence microscope (Leica microsystems). The microscope was equipped with a HCX PL APO 40 × /1.30 oil objective (Leica microsystems), Lumencor SOLA SE (Lumencor) led light and Leica DFC9000 GT sCMOS camera (Leica microsystems). 3 datapoints were taken before the cells were exposed to fructose, after exposure a data point was taken every 30 second for 15 min. Every time point 5 images with an axial distance of 0.5 μm in between the images were acquired in transmission and fluorescent light to ensure an in-focus image for all cells. These images in transmission were taken with 10 ms exposure and 190 LED intensity. Mig1-GFP was observed with an filtercube with a excitation: 450/490, dichroic: 495 and emission: 500-550 filtercube at 300 ms exposure. Nrd1-mCherry was observed with an excitation: 540/580, dichroic: 585 and emission: 592-668 filtercube at 160 ms exposure at 30% LED intensity.

The Hexokinase experiment was performed in W3030(202) *hxx1* Δ *hxx2* Δ , Nrd1-mCherry, Mig1-GFP, TDH3p-HXK1-CYC1t, TDH3p-mAmetrine-CYC1t cells were exposed to an upshift in Fructose from 0 to 2.0%, experimental setup was as described in [38]. Analysis of fluorescence intensity was performed with the ImageJ distribution FIJI [65].

Data availability

The .tif files of the the experimental data are available on figshare (<https://figshare.com/s/2d56e0a6a928ef1dd7ac>). The time-lapse data used to fit the Mig1-model are available on GitHub; https://github.com/cvijoviclab/PEPSDI/tree/main/Intermediate/Experimental_data/Data_fructose.

Code availability

PEPSDI is available on GitHub (<https://github.com/cvijoviclab/PEPSDI>). The code for producing the results in this paper are further available in the GitHub repository (<https://github.com/cvijoviclab/PEPSDI/tree/main/Code/Examples>).

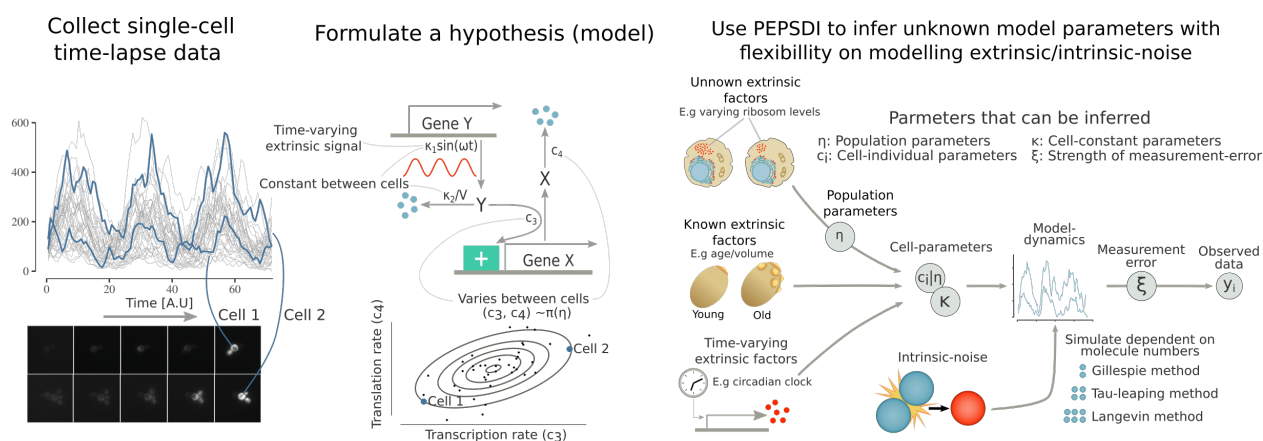


Figure 1: PEPSDI: A Bayesian inference framework for single-cell stochastic dynamic models. Single-cell time-lapse data obtained via fluorescent microscopy often exhibits considerable cell-to-cell variability (left). Dynamic modelling (middle) can help elucidate both the reaction dynamics, and sources of cell-to-cell variability behind such data. PEPSDI (right) is a flexible inference framework for dynamic stochastic single-cell models that imposes few model assumptions. For example, extrinsic noise can be modelled by letting cell-individual quantities i) be modelled probabilistically as $\mathbf{c}^{(i)} \sim \pi(\mathbf{c}^{(i)}|\boldsymbol{\eta})$ (unknown extrinsic factors), ii) be combined with known extrinsic data (known extrinsic factors), and iii) be time-variant (time varying extrinsic factors). Furthermore, PEPSDI includes multiple stochastic algorithms for modelling intrinsic noise, and assumes that the observed data $\mathbf{y}^{(i)}$ is acquired with a measurement error. Overall, based on the observed data PEPSDI can infer the cell individual parameters $\mathbf{c}^{(i)}$, the cell constant parameters $\boldsymbol{\kappa}$, the population parameters $\boldsymbol{\eta}$ and the strength of the measurement error $\boldsymbol{\xi}$.

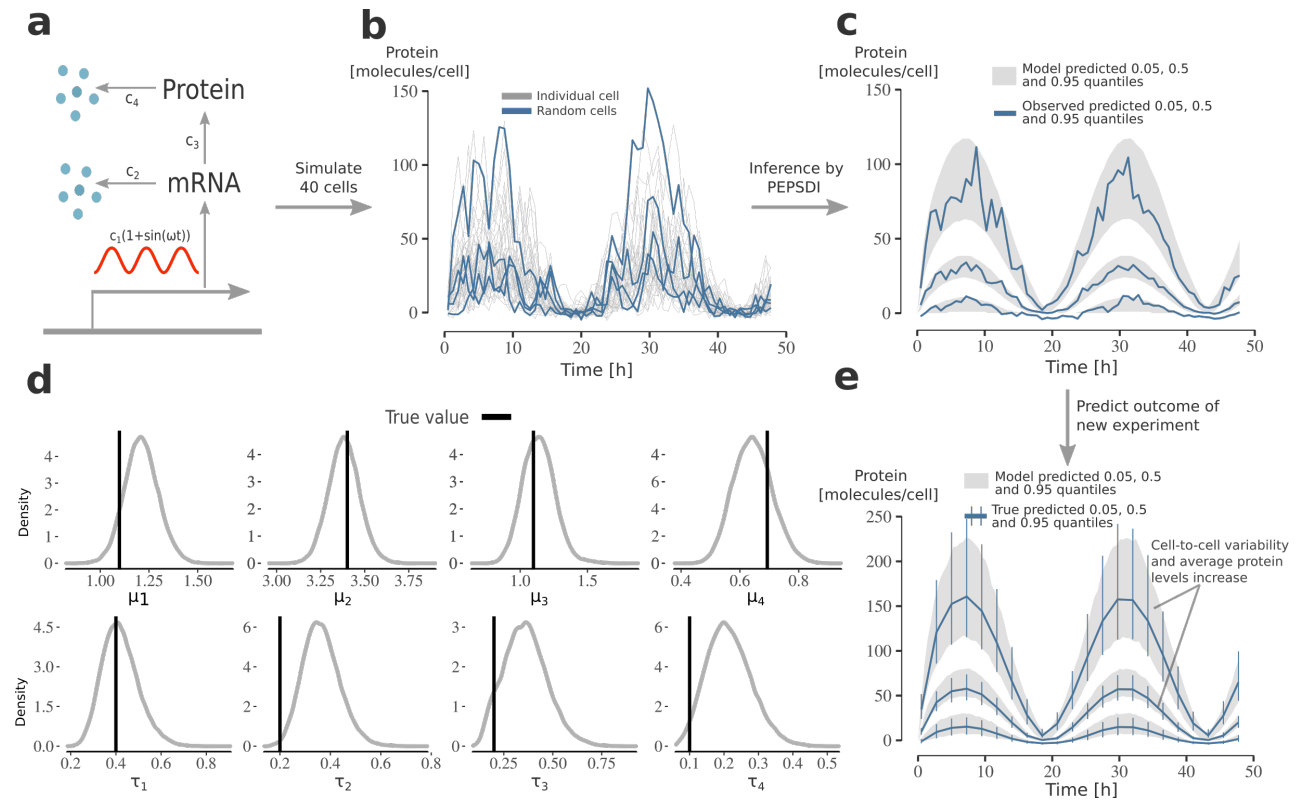


Figure 2: Applying PEPSDI on synthetic data from circadian clock gene expression model. **a)** Schematic representation of the gene expression model. The model consists of two states (mRNA, Protein), and four reactions with associated rate constants $\mathbf{c} = (c_1, \dots, c_4)$. The circadian clock, modelled as a sine function with a period of 24 hours, regulates the transcriptions activity c_1 . Extrinsic noise was simulated by assuming that rate-constants jointly follow a log-normal distribution: $\mathbf{c} \sim \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Omega})$. **b)** Protein count per cell for 40 cells simulated using the gene-network model. Data was simulated using the exact Extrande algorithm with an additive Gaussian measurement noise. **c)** Posterior visual checks [40]. The plot was generated as follows: i) from the inferred posterior distribution simulate 40 cells, and ii) for the 40 cells compute the 0.05, 0.5, 0.95 quantiles. This was repeated 10,000 times yielding 95% credibility intervals. The blue lines correspond to the observed quantiles (from the data in (b)). **d)** Inference results using the data in b). The plots shows the marginal posterior for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_4)$ and $\boldsymbol{\tau} = \text{diag}\{\boldsymbol{\Omega}\}^{1/2}$. **e)** Using the inferred model to predict the outcome of adding an extra gene in the model. An extra gene was modelled by doubling the transcription rate c_1 . The grey lines represent 95% credibility intervals for the 0.05, 0.5 and 0.95 quantiles (obtained as in subplot (c)) when simulating the inferred model with c_1 doubled. The blue lines and error bars represents the true 95 % credibility intervals when observing 40 cells.

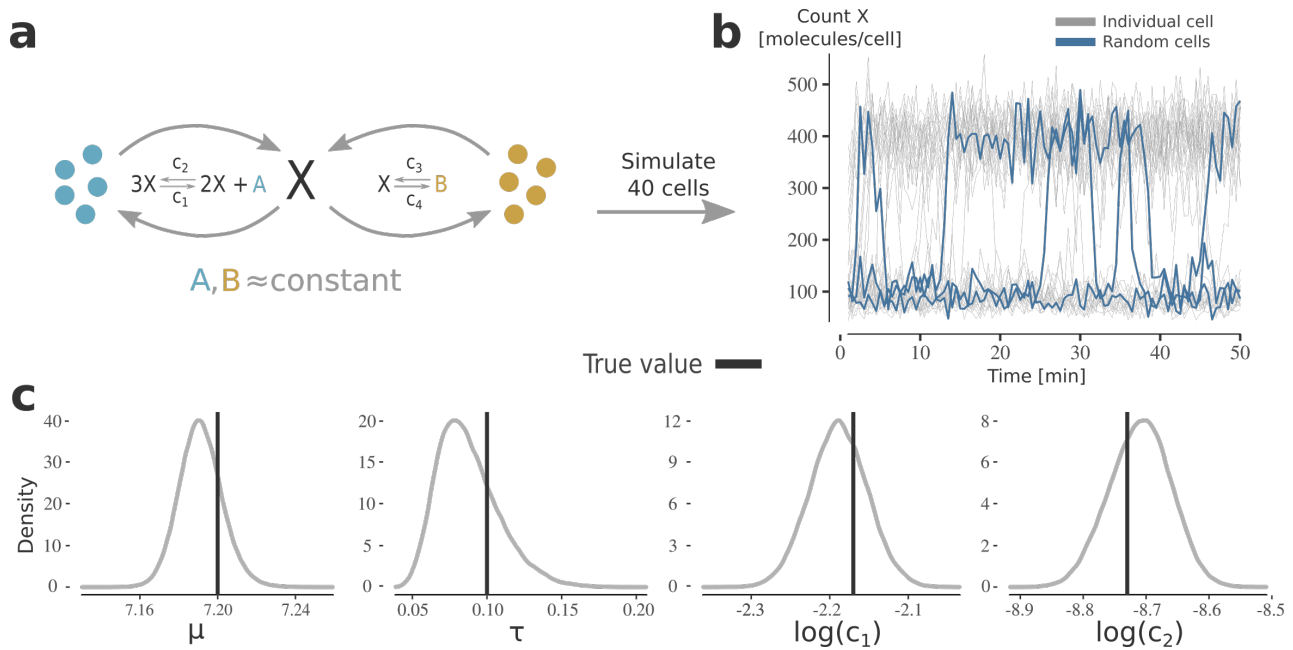


Figure 3: Applying PEPSDI on synthetic data from stochastic bi-stable model. **a)** Schematic representation of the stochastic bi-stable Schögl model. Since the species (A, B) are assumed to be available in excess, the model consists of one state (X) and four reactions with associated rate constants $\mathbf{c} = (c_1, \dots, c_4)$. Extrinsic noise was simulated by assuming that c_1 follows a log-normal distribution; $c_1 \sim \mathcal{LN}(\mu, \tau^2)$. The remaining parameters are assumed to be cell-constant and c_4 is assumed to be known. **b)** Molecule count of X per cell for 40 cells simulated using the Schögl model. Data was simulated using the SSA algorithm with an additive Gaussian measurement noise. Noticeably, a subset of cells stochastically migrates between two cell-states. **c)** Inference results using the data in (b)). To efficiently infer the posterior distribution the model was simulated using the Langevin simulator. The plots shows the marginal posterior for the population parameters (μ, τ), and the cell-constant parameters (c_2, c_3).

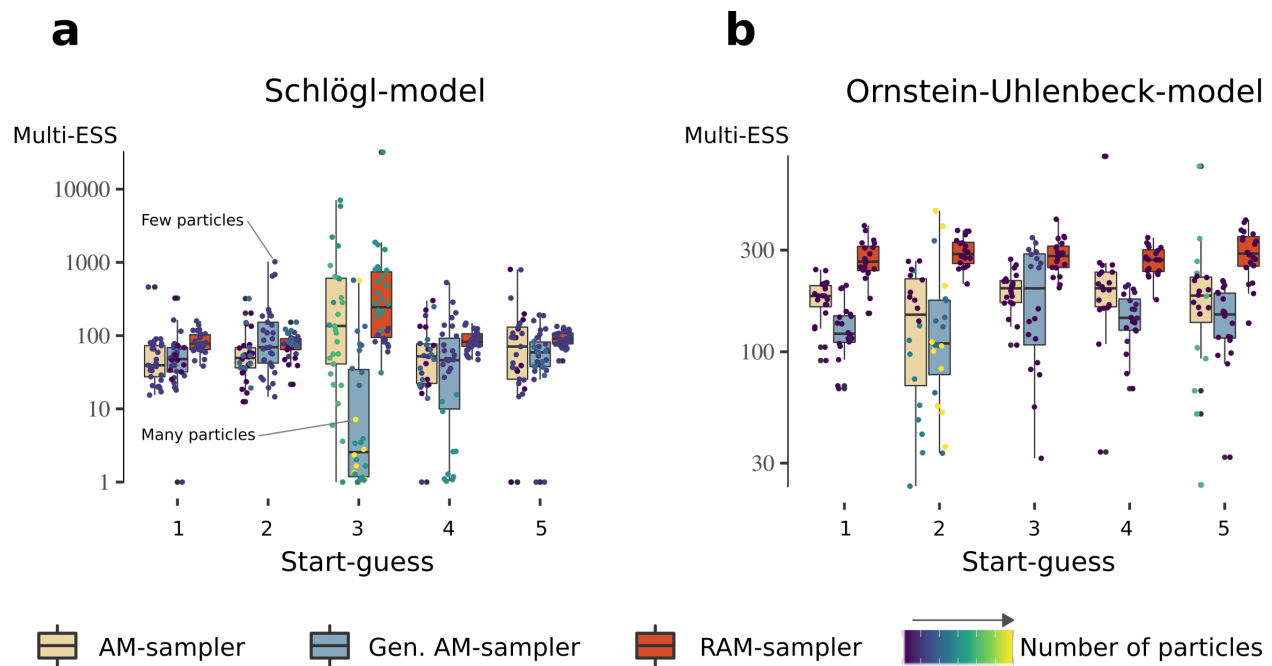


Figure 4: Benchmarking adaptive MCMC-proposals for pseudo-marginal inference. a) Benchmark results for the Schlögl-model. The quality of the adaptive schemes was measured using the MultiESS-criterium [31], where higher values are better. For computational reasons the benchmark was performed for a single-individual (single time-series data). Overall we simulated three datasets. For each dataset we ran five pilot runs with initial parameter values set at randomly chosen prior locations, and then tuned the number of particles (Supplementary Note 1). Starting from the last drawn parameter value in each pilot-run 10 further inference runs were independently launched. The colours denote the several adaptive proposal schemes, and the colour bar represents the number of particles selected by the tuning scheme. A high number of particles implies longer run-times, and an inefficient pilot run. b) Results for the Ornstein-Uhlenbeck model. The benchmark conditions were the same as in a).

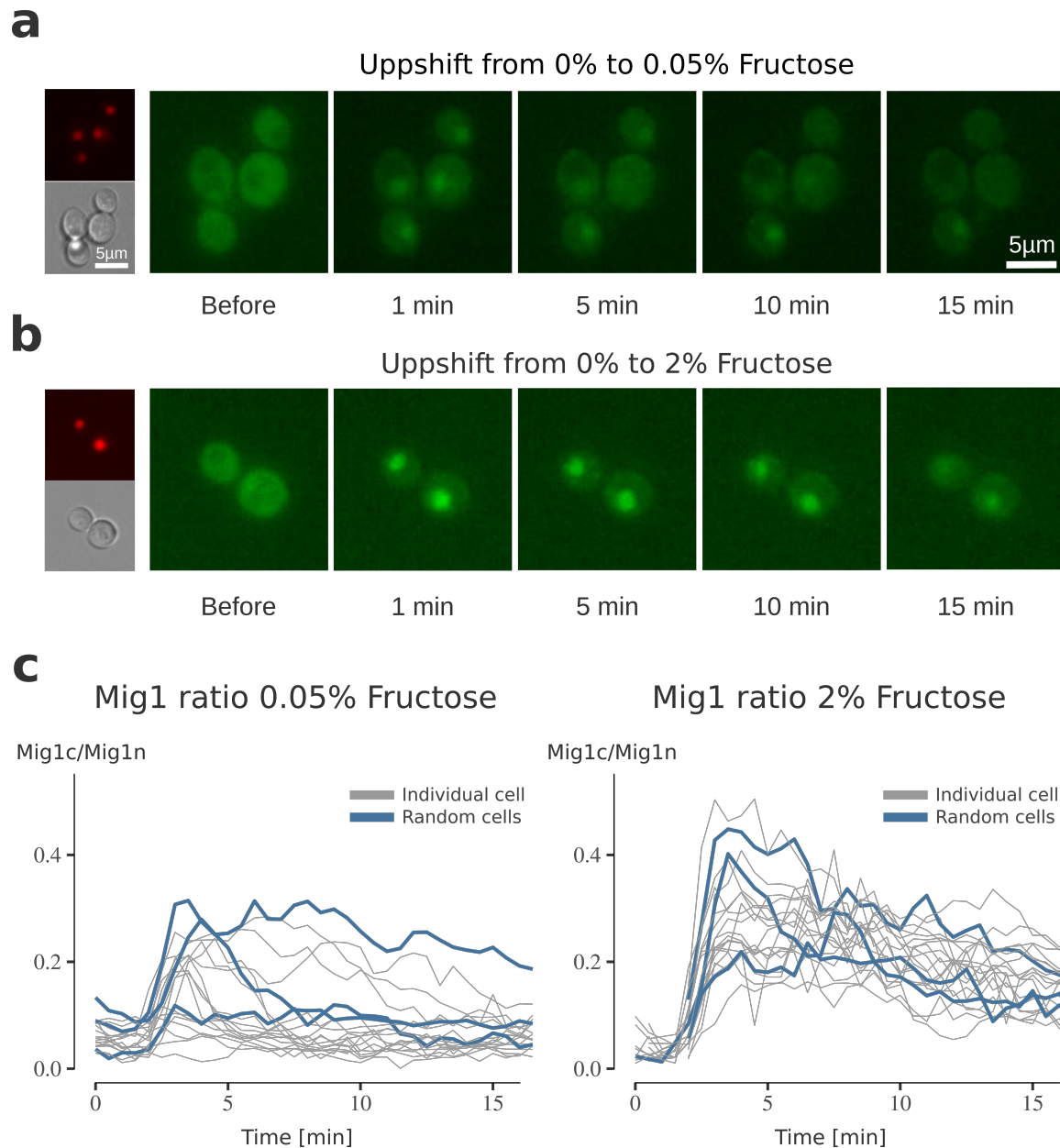


Figure 5: The impact of fructose on Mig1 dynamics. Mig1 localisation in response to the media exchange from media containing 0% fructose (no fast-fermentable carbon sources present) to media containing **a)** 0.05 % fructose and **b)** 2% fructose. The white lines corresponds to 5 μ m scale bars. Exchange of media was achieved through an open microfluidic system. Green fluorescent protein (GFP) images depict Mig1 localisation before and after switching of the media at the noted times. Brightfield images taken as control for the cell localisation. Red fluorescent protein (RFP) images depict Nrd1, a protein which is stationary in the nucleus and used as nuclear marker. **c)** The nuclear intensity of Mig1 for each single-cell in the experiment is given by the localisation index of Mig1 over time (minutes). Localisation index is determined by (Mig1n-Mig1c)/Mig1c (for short called Mig1n / Mig1c in the paper) with Mig1n being the intensity of Mig1 in the nucleus and Mig1c the intensity in the complete cell. All cell traces are grey, three random selected cells are given in blue. Overall, the data-sets combined consist of 37 individuals.

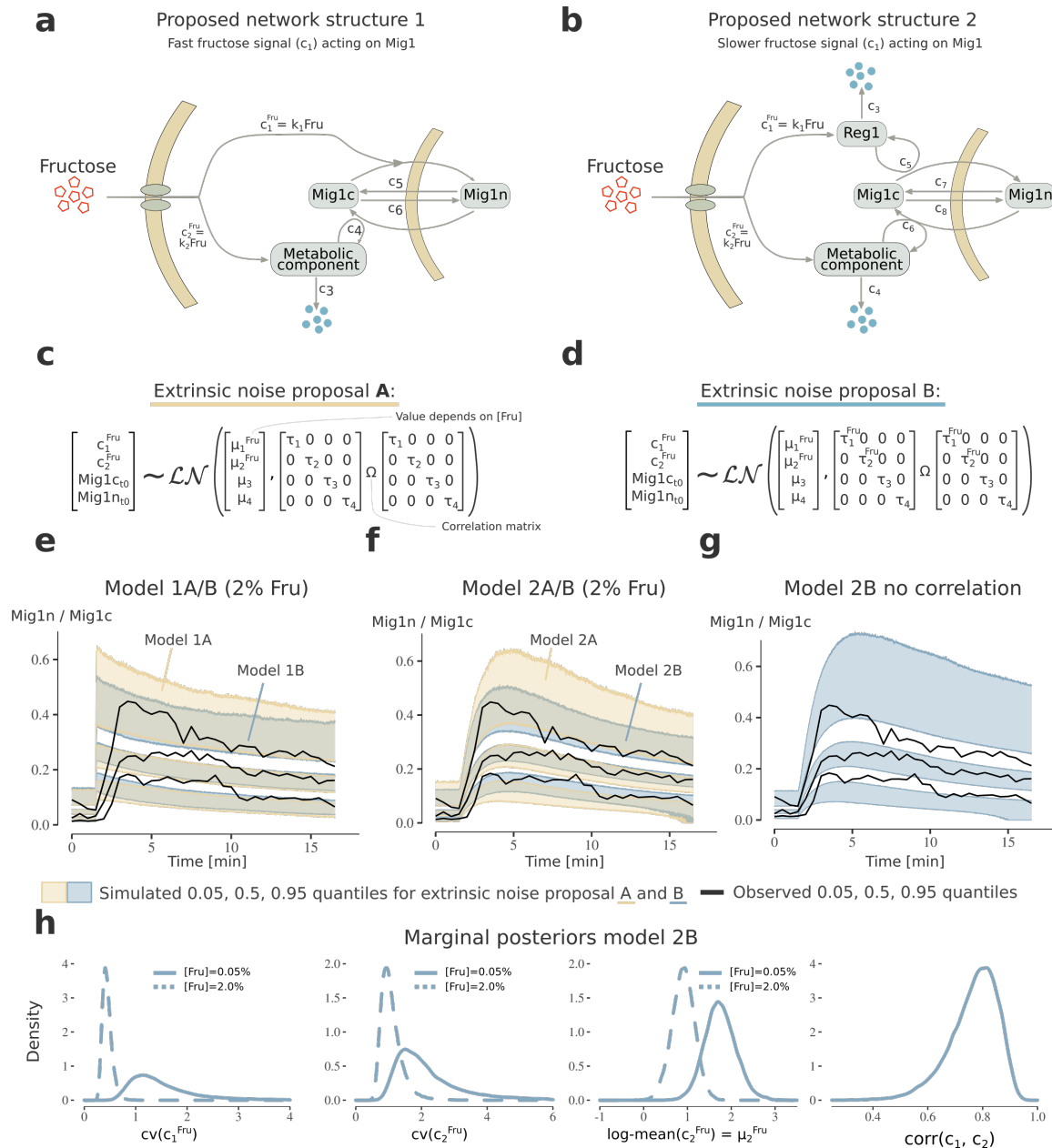


Figure 6: Modelling Mig1 dynamics in response to fructose addition. **a)** Proposed network structure 1 of Mig1 localisation dynamics. The model consists of three states, nuclear Mig1 (Mig1n), cytosolic Mig1 (Mig1c) and a metabolic component. The model proposes that a fructose signal directs Mig1 nuclear import, and that nuclear export is directed by a metabolic signal. **b)** Proposed network structure 2. In addition to structure 1 Mig1 nuclear entry is modelled as delayed due to Reg1 activation. **c)** Modelling proposal A of extrinsic noise. Since (c_1, c_2) encompass multiple cellular processes, and initial Mig1-levels are cell-varying, extrinsic noise was modelled by letting these follow a full log-normal distribution. Since (c_1, c_2) include the external fructose we model different log-mean values for the 2% and 0.05% conditions. **d)** Extrinsic noise proposal B. Additionally to proposal A, we model fructose dependent variability (τ -values) for (c_1, c_2) . **e-f)** Posterior visual check for the 2% fructose data for model structure 1 and 2, using extrinsic noise-proposal A and B. The credibility intervals (bands) were obtained as in Fig. 2. Black lines are the observed quantiles. Model structure 2 with extrinsic noise proposal B (model 2B) best describes the observed trend, and cell heterogeneity. **g)** Posterior visual check as in f) for model 2B with no correlation between $(c_1, c_2, Mig1n, Mig1c)$ (diagonal Ω). This yields an increase in cell-to-cell variability seen by the 0.05, and 0.95 quantile credibility intervals. **h)** Marginal posterior distribution for a subset of parameters in model 2B. The log-normal coefficient of variations $cv(c_i) = \exp(\tau_i^2) - 1$ are fructose dependent, and (c_1, c_2) are strongly correlated.

References

- [1] Shashkova S, Leake MC. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience Reports*. 2017 aug;37(4):20170031.
- [2] Hunting down heterogeneity. Nature Publishing Group; 2010.
- [3] Perkel JM. Single-cell biology: The power of one. *Science*. 2015 nov;350(6261):696–698.
- [4] Zechner C, Unger M, Pelet S, Peter M, Koepl H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*. 2014 feb;11(2):197–202.
- [5] Llamasi A, Gonzalez-Vargas AM, Versari C, Cinquemani E, Ferrari-Trecate G, Hersen P, et al. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. *PLoS Computational Biology*. 2016 feb;12(2).
- [6] Loos C, Hasenauer J. Mathematical modeling of variability in intracellular signaling. *Current Opinion in Systems Biology*. 2019 aug;16:17–24.
- [7] Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, et al. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE*. 2013 sep;8(9):e74335.
- [8] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002 aug;297(5584):1183–1186.
- [9] Whitaker GA, Golightly A, Boys RJ, Sherlock C. Bayesian inference for diffusion-driven mixed-effects models. *Bayesian Analysis*. 2017 jun;12(2):435–463.
- [10] Wiqvist S, Golightly A, McLean AT, Picchini U. Efficient inference for stochastic differential equation mixed-effects models using correlated particle pseudo-marginal algorithms. *Computational Statistics and Data Analysis*. 2021 may;157:107151.
- [11] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. In: *Journal of Physical Chemistry*. vol. 81. American Chemical Society; 1977. p. 2340–2361.
- [12] Gillespie DT. Chemical Langevin equation. *Journal of Chemical Physics*. 2000 jul;113(1):297–306.
- [13] Gillespie DT. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*. 2007 apr;58:35–55.
- [14] Munsky B, Li G, Fox ZR, Shepherd DP, Neuert G. Distribution shapes govern the discovery of predictive models for gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*. 2018 jul;115(29):7533–7538.
- [15] Voliotis M, Thomas P, Grima R, Bowsher CG. Stochastic Simulation of Biomolecular Networks in Dynamic Environments. *PLOS Computational Biology*. 2016 jun;12(6):e1004923.
- [16] Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010 jun;72(3):269–342.
- [17] Xu K, Ge H, Tebbutt W, Tarek M, Trapp M, Ghahramani Z. AdvancedHMC.jl: A robust, modular and efficient implementation of advanced HMC algorithms; 2019.

- [18] Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*. 2002 may;109(3):307–320.
- [19] Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nature Publishing Group; 2008.
- [20] Alvarez I, Niemi J, Simpson M. Bayesian inference for a covariance matrix. *The Annals of Statistics*. 2014 aug;20(4):1669–1696.
- [21] Deligiannidis G, Doucet A, Pitt MK. The Correlated Pseudo-Marginal Method. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2018 nov;80(5):839–870.
- [22] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli*. 2001;7(2):223–242.
- [23] Andrieu C, Thoms J. A tutorial on adaptive MCMC. *Statistics and Computing*. 2008 dec;18(4):343–373.
- [24] Vihola M. Robust adaptive Metropolis algorithm with coerced acceptance rate. *Stat Comput*. 2012;22:997–1008.
- [25] Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM Review*. 2017 feb;59(1):65–98.
- [26] Limpert E, Stahel WA, Abbt M. Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into v. *BioScience*. 2001 may;51(5):341–352.
- [27] Persson S, Welkenhuysen N, Shashkova S, Cvijovic M. Fine-Tuning of Energy Levels Regulates SUC2 via a SNF1-Dependent Feedback Loop. *Frontiers in Physiology*. 2020 aug;11:954.
- [28] Ma L, Wagner J, Rice JJ, Hu W, Levine AJ, Stolovitzky GA. A plausible model for the digital response of p53 to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 oct;102(40):14266–14271.
- [29] Ozbudak EM, Thattai M, Lim HH, Shraiman BI, Van Oudenaarden A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 2004 feb;427(6976):737–740.
- [30] F Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Z Physik*. 1972;1972:147–161.
- [31] Vats D, Flegal JM, Jones GL. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*. 2019 jun;106(2):321–337.
- [32] Wu J, Trumbly RJ. Multiple Regulatory Proteins Mediate Repression and Activation by Interaction with the Yeast Mig1 Binding Site. *Yeast*. 1998;14:985–1000.
- [33] Treitel MA, Kuchin S, Carlson M. Snf1 Protein Kinase Regulates Phosphorylation of the Mig1 Repressor in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*. 1998 nov;18(11):6273–6280.
- [34] Wollman AJM, Shashkova S, Hedlund EG, Friemann R, Hohmann S, Leake MC. Transcription factor clusters regulate genes in eukaryotic cells. *eLife*. 2017 aug;6.

- [35] Welkenhuysen N, Borgqvist J, Backman M, Bendrioua L, Goksör M, Adiels CB, et al. Single-cell study links metabolism with nutrient signaling and reveals sources of variability. *BMC Systems Biology*. 2017 jun;11(1):59.
- [36] Devit MJ, Waddle JA, Johnston M. Regulated nuclear translocation of the Mig1 glucose repressor. *Molecular Biology of the Cell*. 1997 oct;8(8):1603–1618.
- [37] Bendrioua L, Smedh M, Almquist J, Cvijovic M, Jirstrand M, Goksör M, et al. Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *Journal of Biological Chemistry*. 2014 may;289(18):12863–12875.
- [38] Schmidt GW, Welkenhuysen N, Ye T, Cvijovic M, Hohmann S. Mig1 localization exhibits biphasic behavior which is controlled by both metabolic and regulatory roles of the sugar kinases. *Molecular Genetics and Genomics*. 2020 nov;295(6):1489–1500.
- [39] García-Salcedo R, Lubitz T, Beltran G, Elbing K, Tian Y, Frey S, et al. Glucose de-repression by yeast AMP-activated protein kinase SNF1 is controlled via at least two independent steps. *FEBS Journal*. 2014 apr;281(7):1901–1917.
- [40] Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2019 feb;182(2):389–402.
- [41] Mayer FV, Heath R, Underwood E, Sanders MJ, Carmena D, McCartney RR, et al. ADP regulates SNF1, the *saccharomyces cerevisiae* homolog of AMP-activated protein kinase. *Cell Metabolism*. 2011 nov;14(5):707–714.
- [42] Sanz P, Alms GR, Haystead TAJ, Carlson M. Regulatory Interactions between the Reg1-Glc7 Protein Phosphatase and the Snf1 Protein Kinase. *Molecular and Cellular Biology*. 2000 feb;20(4):1321–1328.
- [43] Gillespie DT, Hellander A, Petzold LR. Perspective: Stochastic algorithms for chemical kinetics. *Journal of Chemical Physics*. 2013;138(17).
- [44] Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*. 2002;117(15):6959–6969.
- [45] Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, et al. Pulsatile stimulation determines timing and specificity of NF- κ B – dependent transcription. *Science*. 2009 apr; 324(5924) : 242 – 246.
- [46] Golightly A, Sherlock C. Augmented pseudo-marginal Metropolis-Hastings for partially observed diffusion processes. 2020 sep.
- [47] Schmon SM, Deligiannidis G, Doucet A, Pitt MK. Large-sample asymptotics of the pseudo-marginal method. *Biometrika*. 2021 mar;108(1):37–51.
- [48] Donnet S, Samson A. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced Drug Delivery Reviews*. 2013 jun;65(7):929–939.
- [49] Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al.. A compendium of mutational cancer driver genes. *Nature Research*; 2020.
- [50] Wilkinson DJ. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*. 2009 feb;10(2):122–133.

- [51] Golightly A, Gillespie CS. Simulation of stochastic kinetic models. *Methods in Molecular Biology*. 2013;1021:169–187.
- [52] Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. New York, NY: Springer New York; 2013.
- [53] Hoffman MD, Gelman A. *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*; 2014.
- [54] Ge H, Xu K, Ghahramani Z. Turing: a language for flexible probabilistic inference. In: *International Conference on Artificial Intelligence and Statistics, {AISTATS}* 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain; 2018. p. 1682–1690.
- [55] Andrieu C, Roberts GO. The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*. 2009 apr;37(2):697–725.
- [56] Doucet A, Freitas N, Gordon N. *An Introduction to Sequential Monte Carlo Methods*. In: *Sequential Monte Carlo Methods in Practice*. Springer New York; 2001. .
- [57] del Moral P. *Feynman-Kac formulae. Genealogical and interacting particle systems, with applications*. Springer Verlag New York, Series {\S} Probability and its Applications; 2004.
- [58] Pitt MK, Silva RDS, Giordani P, Kohn R. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. In: *Journal of Econometrics*. vol. 171. North-Holland; 2012. p. 134–151.
- [59] Sherlock C, Thiery AH, Roberts GO, Rosenthal JS. On the efficiency of pseudo-marginal random walk metropolis algorithms. *Annals of Statistics*. 2015 feb;43(1):238–275.
- [60] Doucet A, Pitt MK, Deligiannidis G, Kohn R. Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*. 2015 jun;102(2):295–313.
- [61] Golightly A, Bradley E, Lowe T, Gillespie CS. Correlated pseudo-marginal schemes for time-discretised stochastic kinetic models. *Computational Statistics and Data Analysis*. 2019 aug;136:92–107.
- [62] Golightly A, Wilkinson DJ. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*. 2005 sep;61(3):781–788.
- [63] Lixoft. Monolix version 2019R2. Antony, France: Lixoft SAS; <http://lixoft.com/products/monolix/>. 2019. Available from: <http://lixoft.com/products/monolix/>.
- [64] Ainla A, Jeffries GDM, Brune R, Orwar O, Jesorka A. A multifunctional pipette. *Lab on a Chip*. 2012 apr;12(7):1255–1261.
- [65] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al.. Fiji: An open-source platform for biological-image analysis. *Nature Publishing Group*; 2012.

Acknowledgements

This work was supported by the Swedish Research Council (VR2019-03924 to UP and VR2017-05117 to MC), the Chalmers AI Research Centre (CHAIR) to UP, the Swedish Foundation for Strategic Research (FFL15-0238 to MC) and the Marie Skłodowska-Curie grant agreement No 764591 to PR.

Author contributions

S.P. and M.C. conceived the study. S.P. developed the method and Mig1 model, analysed and interpreted the data. S.S. and N.W. experimental design, interpretation of Mig1 data. N.W., P.R., G.S. carried out Mig1 experiments. S.W. assisted with method development, and U.P. co-supervised method development. M.C. supervised the work. S.P., N.W., S.S., U.P., M.C. wrote the paper, and all authors contributed to reviewing, editing and providing additional text for the manuscript.

Competing interests

Authors declare no conflicts of interest

