Statistical inference of peroxisome dynamics, preprint February 2018

Cyril Galitzine[1], Pierre M. Jean Beltran[2], Ileana M. Cristea[2], and Olga Vitek[1]

[1] College of Science, College of Computer and Information Science
Northeastern University, Boston, 02115, USA
[2] Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University,
Washington Road, Princeton, NJ 08544, USA

Abstract. The regulation of organelle abundance sustains critical biological processes, such as metabolism and energy production. Biochemical models mathematically express these temporal changes in terms of reactions, and their rates. The rate parameters are critical components of the models, and must be experimentally inferred. However, the existing methods for rate inference are limited, and not directly applicable to organelle dynamics.
This manuscript introduces a novel approach that integrates modeling, inference and experimentation, and incorporates biological replicates, to accurately infer the rates. The approach relies on a biochemical model in form of a stochastic differential equation, and on a parallel implementation of inference with particle filter. It also relies on a novel microscopy workflow that monitors organelles over long periods of time in cell culture. Evaluations on simulated datasets demonstrated the advantages of this approach in terms of increased accuracy and shortened computation time. An application to imaging of peroxisomes determined that fission, rather than *de novo* generation, is predominant in maintaining the organelle level under basal conditions. This biological insight serves as a starting point for a system view of organelle regulation in cells.

Keywords: Bayesian inference, stochastic differential equation, stochastic process, particle filter, organelles, replicate, peroxisomes

1 Introduction

Eukaryotic cells are organized into subcellular membrane-bound structures, such as the mitochondria, peroxisomes, and endosomes, known as *organelles* (Fig. 1). Dynamic control of organelle abundance is fundamental for cellular homeostasis, allowing cells to adapt to their environmental, metabolic, and energetic needs [21,25,42,48]. Genetic mutations that affect organelle dynamics are known to cause severe diseases in humans, such as developmental and neurological disorders linked to peroxisome regulation [43].

The understanding of organelle dynamics has been central in basic cell biology research. The processes inducing changes in organelle abundance are well-known. These include organelle production by fission and/or *de novo* biogenesis, as well as organelle destruction by fusion and/or degradation [10,29,41].

However, the integration of these individual processes into the overall control of organelle abundance remains unclear. Although genetic mutations and pharmacological interventions have provided insight into individual mechanisms, the uncovered pathways shared components, thus complicating the integration [21,42,48]. Development of a biochemical model of organelle dynamics is therefore a valuable approach for gaining biological insight into organelle regulation.

Biochemical models express temporal changes in organelle abundance in terms of basic mechanistic processes called *reactions*. Since organelle abundances are typically low (tens to hundreds), a stochastic biochemical model [51] is best suited to model their temporal evolution [34].

Stochastic biochemical models characterize reactions with *rate parameters*, which relate the speed of occurrence of the reaction to organelle counts. In complex biological systems, the rate parameters cannot be determined from first principles, and have to be inferred from experimental measurements collected over time. Here, we propose an integrated microscopy and computational method to infer the rates that regulate organelle abundance from time course organelle counts in cell culture, as we demonstrate for peroxisomes.

Peroxisomes are critical organelles required for cell detoxification and lipid metabolism [42]. Fluorescence microscopy allows us to simultaneously count peroxisomes from multiple cells in the course of time in a minimally invasive manner. However, technological limitations restrict the experiments to less than 100 time points per cell, which for inference purposes is considered *sparse*. The counts are furthermore contaminated by biological and technological variation [50].

Here we argue that, similarly to any other area of data-driven research, rate inference in sparse settings is improved by conducting experiments with multiple cell replicates. Although extending the biochemical models of organelle regulation to replicated experiments is straightforward in theory, it is challenging in practice. First, the replicates complicate modeling and inference of rate parameters, as expressing cellular heterogeneity dramatically increases the computational cost. Second, long-term imaging of organelles (for over 8 hours) is required to observe consistent changes in counts across cells. This is difficult to do for a single cell, and even more so for multiple cells. To our knowledge, there are no reports of peroxisome imaging for this length of time. As a result, previous studies are limited, focused on simulated data [2] or on transcription [54]. They are not applicable to studies of organelle dynamics.

To overcome the limitations above, we describe an algorithm for inferring rate parameters in biochemical models from replicated experiments, and an imaging method that supports the inference by long-term monitoring of peroxisome counts in multiple live cells. This algorithm takes as input peroxisome counts, acquired from fluorescent microscopy images by a commercial software. We demonstrate that this approach provides new biological insight into the mechanisms of peroxisome regulation.

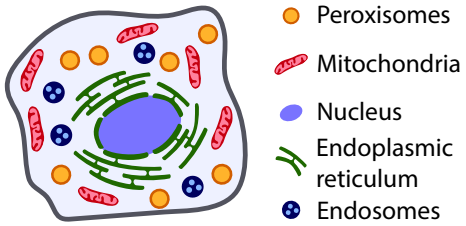

Fig. 1. Illustration of organelles in a eukaryotic cell.

2 Background

2.1 Organelle dynamics via fluorescence microscopy

Organelle dynamics is defined as the process that regulates organelle shape and numbers. Studies of organelle dynamics commonly use fluorescent probes and microscopy (abbreviated to fluorescence microscopy). The technology induces cells to produce a fluorescent protein fusion that is targeted to the organelle. Using fluorescent probes as markers, organelle structures are identified [25] and used to count organelle numbers [24].

Peroxisomes are particularly well suited for studies of dynamics, as their round punctate structure (Fig. 1) facilitates counting from microscopy images [42]. In addition, peroxisomes do not undergo fusion. The biochemical model that governs peroxisome counts in a cell is simplified to only three stochastic processes, each with its own rate parameter, as in Fig. 2. For example, an increase in peroxisome counts implies that the joint rate of processes that control biogenesis (i.e., *de novo* generation and fission) exceeded the rate of degradation. This sheds light into their involvement in cellular events that require changes in peroxisome numbers, such as during cell growth in human cells or in response to different nutrient conditions in yeast [24,34].

For accurate rate inference, important data considerations include the availability of accurate counts, multiple replicates, and time lapse acquisition. However, live peroxisome imaging is commonly performed over small periods of time (a few minutes) [9], and high-throughput peroxisome imaging has been limited to the use of fixed cells [40]. Moreover, high magnification objectives (60X or higher) used to resolve peroxisomes ($0.5 - 1\mu$ m in size) [24] limit imaging to individual cells. This manuscript addresses these challenges by developing a dedicated experimental approach that allows imaging of 100 time points per cell and up to 20 replicate cells per experiment over a time period of over 8 hours. This in turn enables the accurate inference of the rates.

2.2 Modeling and inference of organelle dynamics

Modeling Mukherji *et al* have proposed a stochastic model of organelle dynamics in yeast [34], which we review in this section in the case of peroxisomes. We

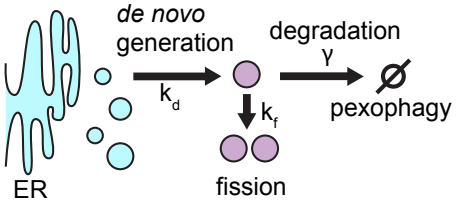

Fig. 2. The biochemical model that governs peroxisome count in a cell. Peroxisomes are created *de novo* at rate k_d or via fission at a rate k_f, and degraded at rate γ. k_d is in units of time^{-1}, while k_f and γ are in units of peroxisome^{-1}time^{-1}.

denote by $x\,(t) \in \mathbb{N}$ the count of peroxisomes in a cell at time $t \in \mathbb{R}^+$. Given the joint effect of the three stochastic processes, the probability $p\,(x,t)$ that the count equals x at time t is governed by

$$\frac{dp\,(x,t)}{dt} = [k_d + k_f\,(x-1)]\,p\,(x-1,t) + [\gamma\,(x+1)]\,p\,(x+1,t) - [k_d + (k_f+\gamma)\,x]\,p\,(x,t) \tag{1}$$

where $p\,(x_0,0) = 1$, and $p\,(x \neq x_0,0) = 0$. The equation describes the Markov jump process [16], and is used in many areas of research, e.g. to describe a birth-death immigration process [16,46] in ecological systems [1,20]. The rate parameter k_d is in units of time^{-1}, while k_f and γ are in units of peroxisome^{-1}time^{-1}. In Eq. (1) k_d, k_f and γ are unknown and must be inferred from the data.

The data $\mathcal{D} = \bigcup_{t=1}^{T}(t_t, y_t)$ are time points $t_1 < t_2 < \cdots < t_T$ and organelle counts $y_1, \ldots, y_T$ observed in a same cell. In presence of measurement error, the observed counts differ from the true (*hidden*) counts x_t governed by Eq. (1). The Normally distributed measurement error is frequently assumed $p\,(y_t\,|x_t\,) = \mathcal{N}\left(x_t, \sigma^2\right)$ [2] [52].

Inference To infer the rate parameters in Eq. (1), traditional inference methods, such as those based on a particle filter [11], simulate many different trajectories from the equation via an exact simulation method, e.g. the Gillespie algorithm [15]. Unfortunately, these methods are computationally expensive. In this manuscript we propose to reformulate Eq. (1) in terms of an equivalent stochastic differential equation, which leads to Bayesian formulation and to inference with parallelization. It reduces the computational time by a large fraction (~ 30), thus enabling rapid feedback for follow-up biological investigations.

2.3 Bayesian rate inference in unreplicated systems

Inference of stochastic biochemical systems is challenging because the likelihood $p\,(\mathcal{D}\,|\theta\,)$ is usually unavailable in closed form. Although frequentist modeling and inference has been proposed [8] [23] [49], it is less suited to experiments with sparse time-course measurements where the inferred rates are subject to relatively high uncertainty. Frequentist inference is therefore rarely used.

To our knowledge, the Bayesian formulation of Eq. (1) has never been considered. However, similar equations modeling other stochastic biochemical systems have received a great deal of attention in, e.g. [13,52]. We briefly overview the approaches developed in these other contexts, as they form the basis of the proposed method for systems with multiple replicates.

Modeling The Bayesian formulation of Eq. (1) requires the specification of a joint prior distribution of $\theta = (k_d, k_f, \gamma, \sigma)$, and the posterior

$$p(\theta \,|\mathcal{D}) \propto p(\mathcal{D} \,|\theta)\, p(\theta) \tag{2}$$

Assuming a memoryless process where the increments $x_{t+1} - x_t$ are statistically independent, and an error model where the measured count only depends on the hidden count, Eq. (2) becomes [52]

$$p(\theta \,|\mathcal{D}) \propto \int \prod_{t=1}^{T} p(y_t \,|x_t, \theta)\, p(x_t \,|x_{t-1}, \theta)\, p(\theta)\, dx_t \tag{3}$$

The integration is over all the possible hidden states at each time point.

Inference Since the likelihood is unavailable in closed form, Bayesian inference is performed by numerically sampling from the posterior distribution. Most approaches are based on a Metropolis Hastings (MH) algorithm, but vary in methods that approximate the likelihood and update the parameters. For example, the exact stochastic process approximates the likelihood and an update scheme in [7]. A similar approach with the moment closure approximation is in [32]. However, these methods are inapplicable in presence of measurement error.

In presence of measurement error, the posterior distribution is most often sampled using a particle filter [11,12], which combines a Markov chain Monte Carlo (MCMC) sampler with a sequential Monte Carlo. It relies on the sequential propagation and reweighing of N computational particles $\boldsymbol{p}_{1 \leq i \leq N}$. Each particle has a weight $\boldsymbol{p}_i(w)$ and a value $\boldsymbol{p}_i(x)$ along the time points. The particle filter method propagates and reweighs the particles along the time course, such that the likelihood at time t in Eq. (3) is the product of particle weight sums over all time points.

Several variants of particle filter aim to improve its computational efficiency. For example, the Particle Marginal Metropolis Hastings (PMMH) [4,18] simultaneously targets both the parameters and the hidden counts, i.e. $p(x, \theta \,|\mathcal{D})$. This manuscript takes an approach similar to PMMH. However, since we are not interested in inference for the hidden counts, we target $p(\theta \,|\mathcal{D})$.

Particle filter is computationally expensive, particularly when used for complex equations such as Eq. (1). As such, they are often parallelized and run on distributed memory systems [39] (although, to the best of our knowledge, never for stochastic biochemical models). Most implementations split the computational particles between multiple processes, and iteratively propagate and reweigh the particles locally within each process [44]. Particles (or other information) is exchanged between the processes to avoid infrequent or local weight

normalization. Different such schemes have been proposed, e.g. distributed resampling with non-proportional allocation (DRNA) [6] or local selection (LS) [31].

2.4 Bayesian rate inference in replicated systems

To the best of our knowledge, replicated experiments have not been previously used to infer rate parameters of organelle dynamics. Here we briefly discuss related methods proposed in the context of other stochastic biochemical systems.

Modeling In experiments with replicate time courses, the data are a collection of time points and organelle counts across $k = 1, \ldots, K$ cells. In the notation of this manuscript, $\mathcal{D} = \bigcup_{k=1}^{K} \mathcal{D}^k$, where $\mathcal{D}^k = \bigcup_{t=1}^{T_k} \left(t_t^k, y_t^k\right)$. The time steps $t_{t+1}^k - t_t^k$ can vary between the cells.

Zechner *et al* model transcriptional and post-transcriptional processes in heterogeneous cell populations, where rates vary between cells [28,53,54]. Assuming that the replicate cells are governed by the same rates (*homogeneous* rates) and are statistically independent, the posterior in Eq. (3) becomes [54]:

$$p\left(\theta \,|\mathcal{D}\right) \propto \prod_{k=1}^{K} \int \prod_{t=1}^{T_k} p\left(y_t^k \,|x_t^k, \theta\right) p\left(x_t^k \,|x_{t-1}^k, \theta\right) p\left(\theta\right) dx_t^k \tag{4}$$

Since no information about rate values is known *a priori*, $p\left(\theta\right)$ is a weakly informative prior (e.g. a Lognormal distribution). Unfortunately, the method cannot handle situations where numbers of data points or measurement time steps differ between the cells. Therefore, this approach is unsuitable for inference of rate parameters of peroxisome dynamics from microscopy data. The likelihood of Eq. (4) can also be extended [54,**?**] to a situations where the rates vary between cells (i.e., are *heterogeneous*) and statistically independent. In such case, the variation of rates between cells (i.e. the *intrinsic* biological variation) is described in terms of distribution $p\left(\theta|\alpha\right)$ with hyperparameters α. Expressing the posterior of Eq. (4) in terms of α, we obtain:

$$p\left(\alpha \,|\mathcal{D}\right) \propto \int \prod_{k=1}^{K} \int \prod_{t=1}^{T_k} p\left(y_t^k \,|x_t^k, \theta\right) p\left(x_t^k \,|x_{t-1}^k, \theta\right) p\left(\theta|\alpha\right) p\left(\alpha\right) dx_t^k \; d\theta \tag{5}$$

The rates are often assumed to follow a gamma distribution [**?**] which ensures their positivity and can well approximate the Normal distribution.

Inference Zechner *et al* [54] aimed to reconstruct promoter activation and transcription. Therefore, they were interested in the distribution over hidden counts $\left(x_1^1, \cdots, x_T^1 \cdots, x_1^K, \cdots, x_T^K\right)$. They jointly inferred the hidden counts and the rate parameters by sampling from $p\left(x, \theta \,|\mathcal{D}\right)$. Since targeting this distribution via Metropolis Hastings was intractable, the authors introduced a recursive Bayesian

procedure where (ignoring cell to cell variations for the rates) the posterior distribution at time t_t was computed from the posterior distribution at time t_{t-1}

$$p\left(x_{1:t}^1, \cdots, x_{1:t}^K, \theta \left| y_{1:t}^1, \cdots, y_{1:t}^K \right.\right) \propto$$
$$\left[\prod_{k=1}^K p\left(y_t^k \left| x_t^k, \theta \right.\right) p\left(x_t^k \left| x_{t-1}^k, \theta \right.\right)\right] p\left(x_{1:t-1}^1, \cdots, x_{1:t-1}^K, \theta \left| y_{1:t-1}^1, \cdots, y_{1:t-1}^K \right.\right)$$

until all the T time points of the K cell replicates have been used.

In contrast, studies of peroxisome dynamics are not interested in the hidden counts, and only need to sample $p\left(\theta \left| \mathcal{D} \right.\right)$ or $p\left(\alpha \left| \mathcal{D} \right.\right)$. This reduced dimensionality allows us to directly sample from Eq. (4) or Eq. (5), without resorting to the complications of Eq. (6).

3 Methods

3.1 Expressing the biochemical model of peroxisome dynamics as a stochastic differential equation

We propose to reformulate the model in Eq. (1) in terms of an equivalent stochastic differential equation (SDE)[35]

$$dx\left(t\right) = \left[k_d + \left(k_f - \gamma\right) x\left(t\right)\right] dt + \left[k_d + \left(k_f + \gamma\right) x\left(t\right)\right] dW\left(t\right) \text{ with } x\left(0\right) = x_0 \tag{6}$$

were $W\left(t\right)$ is Brownian motion, and $x\left(t\right) \in \mathbb{R}$ is the continuous approximation of the discrete peroxisome count $x\left(t\right) \in \mathbb{N}$. Obtained by the diffusion approximation, this equation has the same solution as Eq. (1) [3,17], but is less expensive to solve.

We solve this equation with the Euler-Maruyama method [27]. The solution advances with time step $\Delta t = t_{t+1} - t_t$ following:

$$x_{t+1} = \left[k_d + \left(k_f - \gamma\right) x_t\right] \Delta t + \left[k_d + \left(k_f + \gamma\right) x_t\right] \sqrt{\Delta t} Z \text{ with } Z \sim \mathcal{N}\left(0, 1\right) \tag{7}$$

to obtain x_{t+1} from x_t. In the following, the numerical solution of the SDE from t_t to t_{t+1} is abbreviated $x_{t+1} \sim p_{t_t \to t_{t+1}}\left(x_{t+1} \left| x_t, \theta \right.\right)$. Since the solution is not deterministic, solving the equation between two time step amounts to sampling from the transition density between the steps.

Our experience indicates limited rate variation between the cells. We therefore first assume that all replicate cells in an experiment are homogeneous, i.e. have the same peroxisome regulation rates. We further assume no prior information about the rates, and specify a flat, uninformative prior $p\left(\theta\right) = 1$. In a second time, we relax the homogeneous rate assumption by considering heterogenous rates, this time assuming a flat prior for the hyperparameters $p\left(\alpha\right) = 1$.

In fluorescence microscopy a variety of experimental factors, e.g. the luminosity of the fluorescent tag or the topology of each cell, impact the measurement error. We express this with a Normal measurement error, i.e. :

$$p_\varepsilon\left(y_t^k \left| x_t^k, \theta \right.\right) = \frac{1}{\sqrt{2\pi}\sigma^k} e^{-\frac{1}{2\left(\sigma^k\right)^2}\left(y_t^k - x_t^k\right)^2} \tag{8}$$

where the standard deviation σ^k depends on the cell replicate k, but is constant in time.

Considering both the SDE model and the measurement error, and marginalizing the hidden states, the posterior analogous to Eq. (4) becomes:

$$p\left(\theta\left|\mathcal{D}\right.\right) \propto \prod_{k=1}^{K} \int \prod_{t=1}^{T_k} \left[p_{\varepsilon}\left(y_t^k \left| x_t^k, \theta\right.\right) p_{t_t^k \to t_{t+1}^k}\left(x_t^k \left| x_{t-1}^k, \theta\right.\right)\right] dx_t^k \tag{9}$$

while in the case of heterogeneous rates, the posterior analogous to Eq. (5) is:

$$p\left(\alpha\left|\mathcal{D}\right.\right) \propto \int \prod_{k=1}^{K} \int \prod_{t=1}^{T_k} p\left(\theta|\alpha\right) \left[p_{\varepsilon}\left(y_t^k \left| x_t^k, \theta\right.\right) p_{t_t^k \to t_{t+1}^k}\left(x_t^k \left| x_{t-1}^k, \theta\right.\right)\right] dx_t^k \; d\theta \tag{10}$$

3.2 Parallel inference for replicated experiments with homogeneous rates

MCMC sampling The reformulation of the model in Eq. (1) in terms of a SDE in Eq. (8) reduces the computational cost of parameter estimation. Specifically, we propose to sample the posterior distribution $p\left(\theta\left|\mathcal{D}\right.\right) = p\left(k_d, k_f, \gamma, \sigma^1, \cdots, \sigma^K\left|\mathcal{D}\right.\right)$ in Eq. (10) with the Metropolis Hastings algorithm.

The algorithm requires us to calculate the log likelihood LogLik_k $= \log\left[p\left(\theta\left|\mathcal{D}_k\right.\right)\right] = \log\left[p\left(\mathcal{D}_k\left|\theta\right.\right)\right]$ for each cell replicate k, and the overall log likelihood $\text{LogLik} = \sum_{k=1}^{K} \text{LogLik}_k$. The advantage of the algorithm is its ability to carry out the inference in a distributed memory multicore environment, and in a parallel manner. While traditional implementations of particle filter approximate each LogLik_k in a single core, here we propose to simultaneously calculate LogLik_k using multiple computing cores or CPUs (called *processes* in what follows). The parallelization along each replicate is fairly natural and straightforward for the calculation of the overall log likelihood. The parallel calculation of each replicate log likelihood with a particle filter is, however, more involved due to the need to exchange particle between processes. This will reduce the computation time of each MCMC step, and in turn drastically reduce the overall computation time.

The proposed sampling is a modification of a standard Metropolis Hastings algorithm, as detailed in Algorithm 1. Global operations involving all cell replicates, such as the generation/acceptance of MH samples (lines 5 and 15), or the sum of LogLik_k (line 12) are standard, and performed by the master process Proc^0. A Lognormal proposal distribution (lines 5 and 15) enforces the positivity condition for θ. The magnitude of each rate step (line 5) is proportional to the value of the rate.

Parallel particle filter The calculation of LogLik_k with parallel particle filter (Algorithm 2) is the core of the proposed algorithm. It is an instance of distributed resampling with non-proportional allocation (DRNA), with global

Algorithm 1 *Metropolis Hastings Sampler*

Inputs: data $\mathcal{D}$
Params: # of MCMC samples S
# of burn in samples S_b
initial values θ^0
random walk parameter σ_{MH}
Functions: Algorithm 2
Output: samples $\left\{\theta^{S_b}, \cdots, \theta^{S-1}\right\}$

procedure MCMCs($\mathcal{D}$)
  for s in $0 : S-1$ **do**
    Process Proc_0 **does:**
      ▶ *Generate proposal parameter*
      $\theta^* \sim \text{Lognormal}\left(\log\theta^s, \sigma^2_{\text{MH}}\right)$
    All processes of $\mathcal{P}^k$ **collectively do:**
      $\theta^{*,k} \leftarrow \theta^*$
      ▶ *Calculate replicate log-likelihood*
      $\text{LogLik}_k\left(\theta^{*,k}\right) \leftarrow \text{PPF}\left(\theta^{*,k}, \mathcal{D}^k, \mathcal{P}^k\right)$
    Process Proc_0 **does:**
      ▶ *Sum all replicate log-likelihoods*
      $\text{LogLik}\left(\theta^*\right) \leftarrow \sum_{k=1}^{K} \text{LogLik}_k\left(\theta^{*,k}\right)$
      ▶ *Calculate MH acceptance ratio*
      $\text{LogA} \leftarrow \text{LogLik}\left(\theta^*\right) - \text{LogLik}\left(\theta^s\right)$
      $\text{LogA} \leftarrow \text{LogA} + \log\frac{\theta^*}{\theta^s}$
      ▶ *Accept/reject proposal*
      $r \leftarrow \min(0, \text{LogA})$
      $u \sim U\left(0,1\right)$
      if $\log u < r$ **then**
        $\theta^{s+1} \leftarrow \theta^*$
      else
        $\theta^{s+1} \leftarrow \theta^s$

reweighing at each step [6,30,55]. Unlike the existing algorithms, we distribute the particles of a LogLik_k between multiple processes, and allow each process to resample its own particles. To facilitate mixing, a fraction of particles are exchanged between a process and its neighbors. We describe this in more detail below.

The algorithm partitions all the available processes (except the master Proc^0) into K groups. Every group $\mathcal{P}^k = \left\{\text{Proc}_0^k, \text{Proc}_1^k, \cdots, \text{Proc}_{N_{\text{proc},k}}^k\right\}$ is dedicated to calculating LogLik_k. Proc_0^k is the master process used for global group operations, while the rest $N_{\text{proc},k}$ processes are slave processes.

Each slave process Proc_j^k stores N particles of the filter related to cell replicate k, denoted by $\boldsymbol{p}_i^{j,k}$, $1 \le i \le N$. Each particle has a weight, which characterizes the plausibility of its representation of the hidden state. The particle values $\boldsymbol{p}_i^{j,k}\left(x\right)$ are initialized from a Poisson distribution centered around the observed

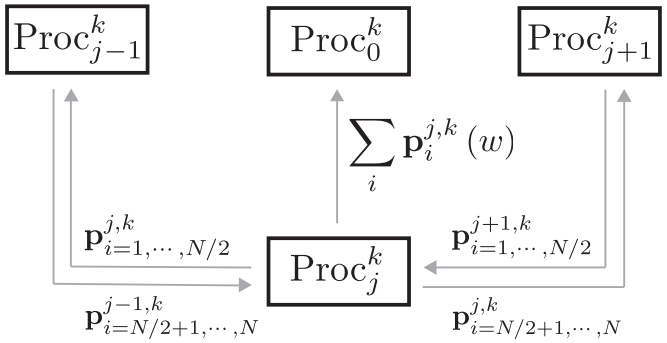

Fig. 3. Communication between Proc_j^k and its neighbors, at one time step, for cell replicate k. Proc_j^k sends the sum of particle weights $\sum_i \mathbf{p}_i^{j,k}(w)$ to the master node. Half of the particles are then exchanged between the neighboring slave processes.

organelle counts at t_1, and the particle weights $\boldsymbol{p}_i^{j,k}(w)$ from a Uniform distribution (lines 5-6). At each observed time point t the particles are propagated to $t+1$ according to the Euler scheme (line 11 and Eq. (8)). This is the most computationally expensive part, due to the large number of particles considered.

After the update, each particle is re-weighted following the Normal error model (line 13 and Eq. (9)). Finally, the algorithm sums all the particle weights into the quantity SW (line 14), and increments the LogLik_k of cell k (line 18).

To prevent the loss of accuracy, the slave processes exchange particles in a circular manner, as illustrated in Fig. 3 (lines 21-22). For each process j of cell replicate k, $N/2$ particles are sent to the process Proc_{j-1}^k to its left, while the remaining half are sent to the process Proc_{j+1}^k to its right. The first process Proc_1^k is viewed as the neighbor of the last process $\text{Proc}_{N_{proc,k}}^k$. This ring topology minimizes the communication between the processes, and maximizes the efficiency of parallelization. Finally, after within-process weight normalization (line 25), the particles are sampled according to their weights using stochastic universal sampling [5] (line 27). This ensures that only highly plausible particles are retained for the next time step.

Since the calculation of each replicate likelihood is independent of the others, replicates with different number of data points and time discretization are trivially handled. If one cell replicate is acquired in a longer time course than the rest, it receives more processes to minimize the idle time of the other replicates waiting for the calculation.

Model-based conclusions The inferred distribution of the rates are obtained from the output samples $(\theta^{S_b}, \cdots, \theta^{S_b})$ of Algorithm 1. Since the samples are highly correlated, they are thinned by a factor (determined from their autocorrelation spectrum) before estimating the posterior distributions.

Algorithm 2 *Parallel Particle Filter*

Inputs: parameters θ
data $\mathcal{D}^k$ and processes $\mathcal{P}^k$ for cell k
Params: # of particles per process N
Output: LogLik_k of cell k

procedure $\text{PPF}(\theta, \mathcal{D}^k, \mathcal{P}^k)$
  Each $\text{Proc}_j^k \;\; 1 \leq j < N_{\text{proc},k}$ **does:**
    ▶ *Initialize particle values and weights*
    for i in 1 to N **do**
      $\mathbf{p}_i^{j,k}(x) \sim \text{Poiss}\left(y_1^k\right)$
      $\mathbf{p}_i^{j,k}(w) \leftarrow \frac{1}{N \times N_{\text{proc},k}}$
  $\text{LogLik}_k \leftarrow 0$
  for t in 0 to $T^k - 1$ **do**
    Each $\text{Proc}_j^k \;\; 1 \leq j < N_{\text{proc},k}$ **does:**
      for i in 1 to N **do**
        ▶ *Propagate particles, Eq. (8)*
        $\mathbf{p}_i^{j,k}(x) \sim p_{t_t \to t_{t+1}}(\cdot|\mathbf{p}_i^{j,k}(x), \theta)$
        ▶ *Calculate particle weights, Eq. (9)*
        $\mathbf{p}_i^{j,k}(w) \leftarrow p_\varepsilon\left(y_{t+1}^k|\mathbf{p}_i^{j,k}(x), \theta\right)$
      Send $\sum_i \mathbf{p}_i^{j,k}(w)$ to Proc_0^k
    Process Proc_0^k **does:**
      ▶ *Increment LogLik$_k$*
      $\text{SW} \leftarrow \sum_{j=1}^{N_{\text{proc},k}-1} \sum_i \mathbf{p}_i^{j,k}(w)$
      $\text{LogLik}_k \leftarrow \text{LogLik}_k + \log\left(\frac{\text{SW}}{N}\right)$
    Each $\text{Proc}_j^k \;\; 1 \leq j < N_{\text{proc},k}$ **does:**
      ▶ *Exchange particles between processes*
      $\left\{\mathbf{p}_i^{j,k}\right\}_{i=N/2+1}^{N} \leftrightarrow \left\{\mathbf{p}_i^{\text{Right}^{j,k},k}\right\}_{i=1}^{N/2}$
      $\left\{\mathbf{p}_i^{j,k}\right\}_{i=1}^{N/2} \leftrightarrow \left\{\mathbf{p}_i^{\text{Left}^{j,k},k}\right\}_{i=N/2+1}^{N}$
      ▶ *Normalize weights for each process*
      for i in 1 to N **do**
        $\mathbf{p}_i^{j,k}(w) \leftarrow \frac{\mathbf{p}_i^{j,k}(w)}{\sum_i \mathbf{p}^i(w)}$
      ▶ *Resample particles by weight*
      Sample $\mathbf{p}_i^{j,k} \sim \mathbf{p}_i^{j,k}(w)$ N times
  return LogLik_k

The units of k_f and γ differ from the units of k_d, and the values of the rates are not comparable directly. On the other hand, the ratios $k_d : k_f\overline{N} : \gamma\overline{N}$ (where $\overline{N}$ is the average number of peroxisomes per cell) are the relative prevalence of

each reaction in numbers of reaction per unit time. Therefore, to facilitate the interpretation, we report the results in terms of k_d, $k_f\overline{N}$ and $\gamma\overline{N}$ in what follows.

3.3 Inference of cell to cell rate variations

The method presented in the previous section can readily be extended to account for cell to cell variations in the rates. We assume that k_d, k_f and γ each follow a Gamma distribution with its own shape and rate parameters: i.e. $k_d \sim \text{Gamma}\left(\alpha_{k_d}, \beta_{k_d}\right), k_f \sim \text{Gamma}\left(\alpha_{k_f}, \beta_{k_f}\right)$ and $\gamma \sim \text{Gamma}\left(\alpha_\gamma, \beta_\gamma\right)$. Instead of directly sampling the shape and scale of the Gamma distributions, we sample their mean μ and standard deviation σ. This approach is equivalent (since e.g. for k_d, $\alpha_{k_d} = \mu_{k_d}^2/\sigma_{k_d}^2$ and $\beta_{k_d} = \sigma_{k_d}^2/\mu_{k_d}$) but it allows a better interpretation of the inferred parameters, and reduces sampling variation. We propose to sample $\alpha = \left(\mu_{k_d}, \sigma_{k_d}, \mu_{k_f}, \sigma_{k_f}, \mu_\gamma, \sigma_\gamma, \sigma^1, \cdots, \sigma^K\right)$ from the posterior distribution $p\left(\alpha \left| \mathcal{D} \right.\right)$ in Eq. (11) with the Metropolis Hastings algorithm detailed in the previous section. We approximate the integral over all θ parameters by using N_r single importance samples ${}^r\theta \sim p\left(\theta|\alpha\right), 1 \leq r \leq N_r$ (which provides an unbiased estimator of the integral) so that:

$$p\left(\alpha \left| \mathcal{D} \right.\right) \propto\!\!\!\propto \frac{1}{N_r}\sum_{r=1}^{N_r}\prod_{k=1}^{K}\int\prod_{t=1}^{T_k}\left[p_\varepsilon\left(y_t^k \left| x_t^k, {}^r\theta\right.\right) p_{t_t^k \to t_{t+1}^k}\left(x_t^k \left| x_{t-1}^k, {}^r\theta\right.\right)\right] dx_t^k \tag{11}$$

We use the exact same particle filer detailed in algorithm 2 while the difference in the MCMC sampler resides in that we sample α instead of θ and need to integrate the likelihood over θ. As such, in algorithm 1, $\theta^*, \theta^s, \theta^{s+1}$ are replaced by $\alpha^*, \alpha^s, \alpha^{s+1}$ respectively while line 7 (the generation of rates from hyperparameters) becomes $\theta^{*,k} \sim p\left(\theta|\alpha\right)$. At each MCMC step s, and for each replicate k, N_r distinct $\theta^{*,k,r} \sim p\left(\theta|\alpha\right), 1 \leq r \leq N_r$ proposals are furthermore generated. The corresponding likelihoods, ${}^r\text{LogLik}_k$ are then summed following:

$${}^r\text{LogLik}\left({}^r\theta^*\right) \leftarrow \sum_{k=1}^{K} {}^r\text{LogLik}_k\left({}^r\theta^{*,k}\right) \tag{12}$$

and the log of the integral over θ (i.e. sum over r) is obtained through:

$$\text{LogLik}\left(\theta^*\right) \leftarrow {}^1\text{LogLik}\left({}^1\theta^*\right) + \log\sum_{r=2}^{N_r}\exp\left[{}^r\text{LogLik}\left({}^r\theta^*\right) - {}^1\text{LogLik}\left({}^1\theta^*\right)\right] \tag{13}$$

where ${}^r\text{LogLik}$ has been reindexed such such that ${}^1\text{LogLik} > {}^2\text{LogLik} > \cdots > {}^{N_r}\text{LogLik}$ (Log-Sum-Exp trick). Equations (13) and (14) both replace line 12 of algorithm 1.

3.4 Implementation

We implemented the procedure in C++ for speed, and parallelized it using the MPI-2.2 (Message Passing Interface) [36] libraries. The source code and documentation is available at `github.com/cyrilgalitzine/Organelle`.

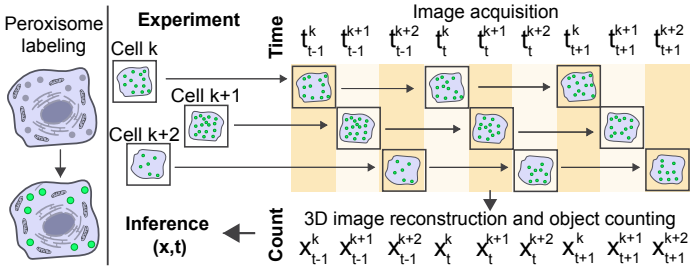

Fig. 4. Schematic representation of peroxisome imaging and counting in live human cells.

3.5 Imaging and counting peroxisomes by confocal microscopy

First, peroxisomes in human liver cells (HepG2) were labeled by expression of the fluorescent protein, EGFP, tagged with the peroxisome targeting sequence, PTS1, as in [38]. Transfection conditions were optimized to avoid enlarged aberrant peroxisomes from overexpression, as well as reduce background cytosolic fluorescence while maintaining peroxisome-specific fluorescent signals. At 24 hours following transfection, live cells were imaged with a 60X objective using a Nikon Ti-E confocal microscope equipped with a Yokogawa spinning disc (CSU-21), NIDAQ Piezo Z-axis stage for fast Z acquisition, and temperature-controlled humidified chamber set at 37°C (Princeton University). Z-stacks were acquired with $0.2\mu m$ steps for $22\mu m$ at 50ms exposure per step to limit laser exposure to < 10 seconds per cell. Image acquisition was automated for sequential imaging of individual cells (Fig. 4). Overall, this workflow maintained instrument use to a reasonable timeframe, and improved cell viability by avoiding continuous laser exposure. It maximized data collection at time intervals that allow detection of changes in peroxisome counts without oversampling. Using this instrumentation, we could image 20 cells with 6 minutes data point intervals for a total of 10 hours.

To count peroxisomes, images were processed using the Nikon NIS-Elements AR v5.0. Image Z-stacks were deconvolved [47,19], and individual peroxisomes were detected semi-automatically using the 3D Spot Detection feature (50-300 manually adjusted contrast, 1:2 Z-axis elongation, $0.7\mu m$ diameter spot, and variable size spots) (Fig. 5). Organelle abundance was quantified as the number of objects detected per cell and per time point.

4 Datasets

4.1 Experimental datasets

We acquired a total of three experimental datasets, called Day 1, Day 2, and Day 3. The final datasets consisted of 13 replicate cells for Day 1, 10 replicate cells for Day 2, and 20 replicate cells for Day 3. The count results for two of

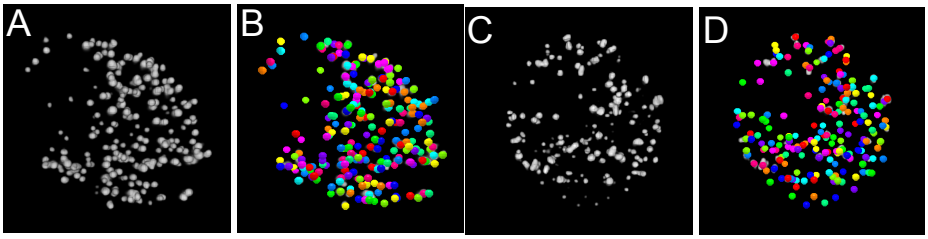

Fig. 5. A: peroxisome fluorescence signal from a cell. B: Output of the 3D spot detection algorithm for the same cell. Each colored sphere indicates a peroxisome. C, D: same as A, B, for another cell with fewer peroxisomes.

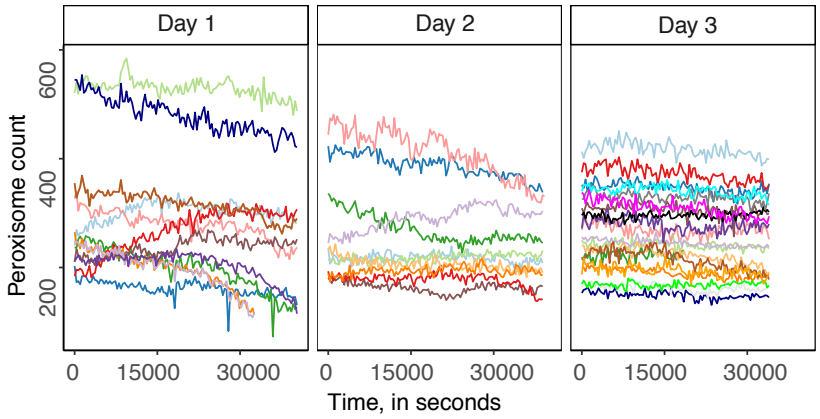

Fig. 6. Peroxisome counts in the time course experiments. Colors indicate cell replicates.

these experiments are shown in Fig. 6. Between-cell and between-day variability was observed for both the average number of peroxisomes in a cell and the slope of the trace throughout the experiments. The number of cells per experiment and time points per cell varied as some cells moved out of focus or died before completion of the experiment. The cell heterogeneity and incomplete data were important considerations of the rate parameter inference.

4.2 Simulated datasets

To evaluate the proposed approach in the case of homogeneous rates we simulated three additional datasets SIM A, SIM B and SIM C. The datasets were simulated with the Gillespie algorithm, with $K = 14$ and also with $K = 1$ cell replicates. Each cell replicate was initialized with a different count y_0^k (taken to be identical to the experimental counts of the first 14 cells on Day 3), but simulated with the same duration $T^k = 33936$s, frequency (336s^{-1}) and stan-

SIM	Dataset	$k_d \times 10^4$	$k_f \times 10^5$	$\gamma \times 10^5$	σ^1
A	**True**	**7.75**	**4**	**4**	**6**
	14 cells	29% 10 (7.65)	8% 4.32 (0.85)	12% 4.47 (0.85)	0.7% 5.95 (0.79)
	1 cell	1700% 140 (148)	4% 4.18 (3.05)	80% 7.17 (3.96)	-5% 5.68 (0.91)
B	**True**	**2.75**	**3**	**4**	**7.5**
	14 cells	224% 8.92 (5.50)	-9% 2.71 (0.93)	-1% 4.90 (0.91)	4% 7.8 (0.92)
	1 cell	758% 23.60 (26.5)	8% 3.01 (3.00)	13% 5.65 (3.20)	3% 7.70 (0.87)
C	**True**	**50**	**1**	**5**	**6**
	14 cells	-5% 47.2 (1.14)	30% 1.29 (0.63)	-4% 3.30 (0.58)%	-4% 5.77 (0.92)
	1 cell	8% 53.80 (51.6)	198% 2.98 (2.27)	73% 5.19 (2.34)	-10% 5.44 (0.71)

Table 1. Experimental design and parameter estimation in simulated datasets. True parameter values in each simulation are in bold. Table entries report parameter estimates, standard deviation of the posterior distribution (in parentheses), and the % error. k_d is expressed in second^{-1}, while k_f and γ in peroxisome^{-1}second^{-1}.

dard deviation of the measurement error σ. At the inference stage, the standard deviation of the measurement error σ^k were inferred separately for each replicate.

The values of the rate parameters were inspired from those the experimental datasets, and reported in Table 1. In SIM A and SIM B the values of k_d were low, corresponding to a realistic situation where the *de novo process* is less prevalent than the fission or degradation, making it hard to detect. In SIM C k_d was relatively high, and the three reactions occurred relatively often. The values of k_f and γ were identical in SIM A, but differed slightly in SIM B and SIM C. The datasets were used to evaluate the ability of the proposed approach to estimate these different parameter configurations.

5 Results

5.1 Parallel inference shortened computation time

We inferred the rates in the experimental and the simulated datasets using 4800 particles and $\sigma_{\text{MH}} = 0.04$, resulting in a MH acceptance rate between 0.2 and 0.3. We set $N_{\text{proc},k} = 2$, and thinned the original 100,000 MCMC sampling iterations to every 500. Each iteration lasted on average 7.5 seconds (wall clock time) for Day 1 (13 replicates on 39 CPUs), 6 seconds for Day 2 (11 replicates on 34 CPUs) and 9 seconds for Day 3 (20 replicates on 71 CPUs). The overall inference took between 9 and 13 days.

In contrast, the existing modeling and inference procedures required substantially longer computation time. For example, the use of the original stochastic model in Eq. (1) with Gillespie algorithm increased the time per iteration by

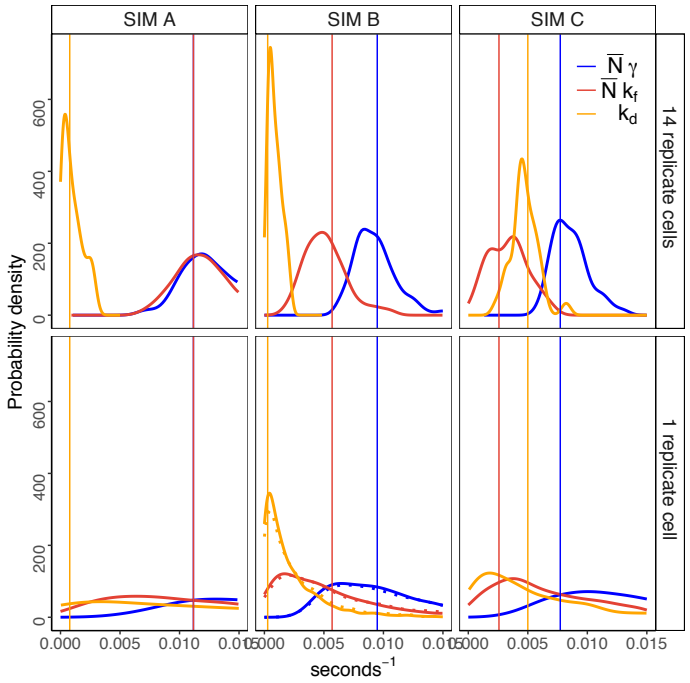

Fig. 7. Posterior distributions of the rates in the simulated datasets. To make the rates comparable, k_f and γ are multiplied by the average peroxisome count $\overline{N}$ in each dataset. Vertical lines are the true parameter values. A dotted line denotes rate distributions obtained with Eq. (1) and the Gillespie algorithm instead of the SDE (solid line)
.

about 30% with identical inference results. In the case of Day 1, this results in each iteration taking 10 seconds, for an overall total of about 14 days.

Similarly, the use of the SDE model in Eq. (8) with serial inference per cell increased the time per iteration by a factor of $K\,(N_{\mathrm{proc},k}-1)$ (the parallel overhead is negligible as compared to the particle movement). In the case of Day 1, representative iterations lasted 195 seconds, estimating the overall inference time of 270 days. Therefore, the biological insights from multiple replicates were out of reach without the proposed parallel procedure.

5.2 The approach accurately inferred the rates

Fig. 7 shows the posterior distributions, and Table 1 summarizes the properties of the inferred rates in the simulated datasets. Since the inferred σ^k were similar across replicates, the table only reports the value for the first replicate σ^1. The proposed approach accurately inferred the rates in experiments with 14 cell replicates. SIM A and SIM B with low k_d challenged the estimation of this rate, as evidenced by its skewed and variable posterior distribution. A larger *de novo*

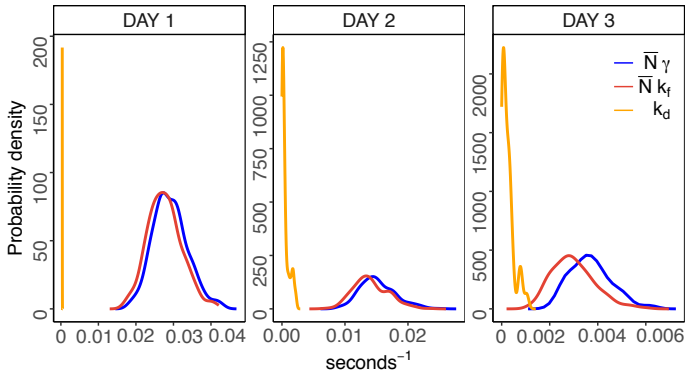

Fig. 8. As in Fig. 7, for the experimental datasets.

rate in SIM C led to more accurate estimates. Table 1 shows that the inferred posterior distributions of $\overline{N}k_f$ and $\overline{N}\gamma$ had little variance. Their relatively large breadth in Fig. 7 was due to the multiplication by $\overline{N}$. We obtained identical inference results with the SDE model of Eq. (8) as with the Gillespie algorithm applied to Eq. (1) as shown in the case of SIM B (1 replicate) in Fig. 7. This, combined with the fact that simulation results were generated using Eq. (1), demonstrates that the SDE approximation reduced computational cost without compromising the accuracy of the results.

5.3 Replicate cells improved inference of the rates

Fig. 7 compares the inferred posterior distributions with one versus 14 cell replicates in the simulated datasets, and Table 1 summarizes the results. Inference from unreplicated experiments had high uncertainty in all the experiments, and led to broader posterior distributions. In particular, rates associated with rare events (such as *de novo* in SIM A and SIM B) could not be accurately estimated with only one replicate, and had an ∼10-fold error for the mean. The uncertainty diminished in experiments with 14 cell replicates, and the standard deviation of the posterior distributions of k_f and γ was reduced by a factor of 3 to 4. This result emphasized the importance of incorporating replicate cells into the rate inference procedure.

5.4 Inference from replicate cells revealed maintenance of peroxisome counts by prominent fission, compared to low *de novo* generation rates

Fig. 8 compares the posterior distribution for the three experimental datasets, and Table 2 summarizes the results. We consistently observed similar values of rate parameters of peroxisomes degradation and fission. Moreover, we consistently observed a 5 to 100 times smaller value of the rate of *de novo* generation.

Dataset	$k_d \times 10^4$	$k_f \times 10^5$	$\gamma \times 10^5$
Day 1	2.51 (3.66)	10.0 (1.36)	10.5 (1.38)
Day 2	6.45 (5.60)	5.57 (1.12)	6.02 (1.11)
Day 3	1.05 (1.72)	1.19 (0.37)	1.43 (0.32)

Table 2. Results for the experimental datasets with the homogeneous rate model. The reports parameter estimates, standard deviation of the posterior distribution (in parentheses). k_d is expressed in second^{-1} and k_f and γ in peroxisome^{-1}second^{-1}.

Dataset	$\mu_{k_d} \times 10^4$	$\sigma_{k_d} \times 10^4$	$\mu_{k_f} \times 10^5$	$\sigma_{k_f} \times 10^5$	$\mu_\gamma \times 10^5$	$\sigma_\gamma \times 10^5$
Day 1	7.6 (4.8)	8.9 (8.9)	8.99 (1.4)	0.39 (0.2)	9.75 (1.4)	0.49 (0.3)
Day 2	11.3 (10.5)	12.2 (10.9)	4.99 (1.03)	0.39 (0.16)	5.61(1.06)	0.38 (0.18)
Day 3	4.6 (3.2)	3.5 (3.5)	1.21 (0.38)	0.12 (0.07)	1.55 (0.38)	0.096 (0.08)

Table 3. Results for the experimental datasets for the heterogeneous rate model. The reports parameter estimates, standard deviation of the posterior distribution (in parentheses). k_d is expressed in second^{-1} and k_f and γ in peroxisome^{-1}second^{-1}.

The results indicate that *de novo* peroxisome generation in mammalian cells is a relatively rare event, occurring approximately 8 to 45 times per day.

5.5 Rates varied little between cells

We obtained in Section 5.4 fairly narrow inferred distribution for the rates which allowed us to make important biological conclusions. We would, however, like to distinguish how much of the variance of inferred rates is caused by possible rate heterogeneity between cells and how much is caused by statistical uncertainty (i.e. a too low number of replicates or data points). This is achieved by using the heterogeneous rate model which models rate cell to cell variations. Fig. 9 shows the posterior hyperparameter distributions obtained with the heterogeneous rate model, and Table 3 summarizes the results. On average, for k_f and γ, the intrinsic rate standard deviation was about 5-10%, indicating relatively small intrinsic biological variations as compared to the rate values. The intrinsic variance of rates k_f and γ, i.e. σ_{k_f} and σ_γ, reported in Table 3 were less than 30% of the variance of the rate mean (in parentheses next to the mean value). This indicates that most of the variance obtained with the homogeneous rate model reported in Table 2 arised from statistical uncertainty (i.e. a too low number of replicate) instead of biological variation. In contrast, for days 1 and 3, σ_{k_d} was relatively large (about twice the rate mean standard deviation). This shows that, for some conditions, the uncertainty associated with cell to cell variation was more important than the statistical uncertainty in the *de novo* rate value.

6 Discussion

In contrast to other organelles, peroxisome are constantly recycled in healthy cell populations, and degraded to remove old or damaged peroxisomes [22,37].

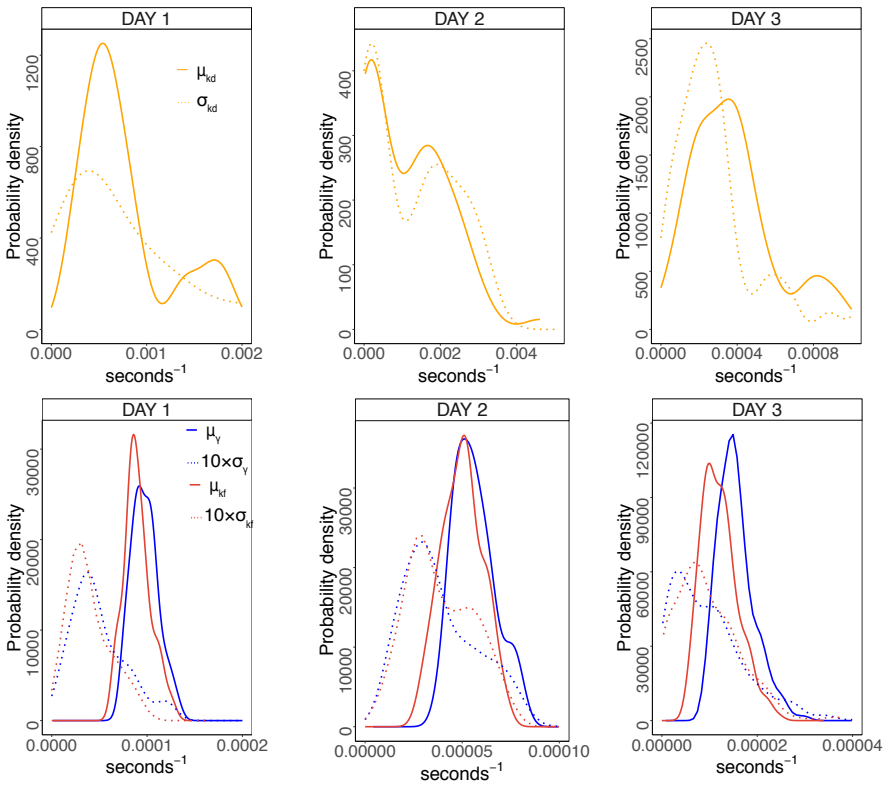

Fig. 9. As in Fig. 8, with the heterogeneous rate model. The rate standard deviations for k_f and γ (dotted lines) are multiplied by 10 to plot their distributions on the same plots as the rate means.

Defining the predominant process of peroxisome production is a current topic of debate [14,26,33,45]. The stochastic model proposed by Mukherji *et al* was tested in yeast cells [34]. The authors observed a switch from the predominance of *de novo* generation to fission, occurring under conditions that increase peroxisome numbers. While yeast cells only have 5-20 peroxisomes per cell, humans and other mammals need larger number of peroxisomes (100-500). It is therefore possible that mammals evolved to use fission as a primary mechanism for peroxisome proliferation [22,24].

Here, we used experimental data to directly infer the rates governing peroxisome abundance. While the inferred rates for fission and degradation were similar, *de novo* generation was less frequent. The infrequent *de novo* generation is in line with previous studies estimating low numbers (30) of new peroxisomes per day [26]. The inference of peroxisome rates helps us reconcile previous conflicting evidence. It leads to a new model, where peroxisome population undergoes

recycling via two opposing processes, fission and degradation, in addition to a basal *de novo* generation rate.

The results indicated that accuracy of rate inference depends on the value of the rates. In particular, rates associated with rare events, such as the *de novo* rate, are difficult to infer. This can be mitigated by imaging more cell replicates, or by extending the imaging time.

The inferred rates varied between instances of experiments repeated on multiple days. The variation in the rates across days could be explained as biological effects of the cell batch analyzed, such as confluency and age, which are known to affect the mechanisms of peroxisome biogenesis [**?**].

The cell to cell variation for the rates was limited in the case of the fission and degradation rates but more pronounced for the *de novo* rate which further compounded the uncertainty in its estimation.

The overall rate inference, and the assessment of the uncertainty, may be improved by analyzing the combined data from all the days. This will require extending the model to include inter-day variation, and accommodating the extra computational cost. Since the proposed SDE-based modeling is flexible, and since the inference algorithm supports parallelization, the proposed approach is in principle extendable to such situations. However, measurements on more days will be required to establish a model of inter-day variation.

This proposed inference procedure can also be extended to other organelles. Imaging tools for other organelles are available and widely used [38]. The modeling and inference procedure can include additional reactions, such as fusion.

Organelle dynamics are subject to alteration and regulation upon extracellular and intracellular cues. For example, peroxisomes increase in numbers when cells grow to undergo division. This occurs by either increasing the rate parameters of one of the production processes, or by decreasing the rate parameter of the degradation. The proposed approach can be used to assess this process of cell adaptation. Therefore, this work serves as a starting point for achieving a system view of the biophysical properties, used by the cell to regulate its organelle content.

Acknowledgements

This work was supported in part by a Burroughs Wellcome travel grant (to C.G.), a Dodds Fellowship (to P.M.J.B), NIH grants GM114141, HL127640, Mallinckrodt Scholar Award (to I.M.C), and Sy and Laurie Sternberg award (to O.V.).

References

1. L. J. Allen. *Stochastic Population and Epidemic Models.* Persistence and Extinction. Springer, Cham, 2015.
2. M. Amrein and H. R. Künsch. Rate estimation in partially observed Markov jump processes with measurement errors. *Statistical Computing*, 22:513, 2011.

3. D. F. Anderson and T. G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, page 3. Springer, 2011.
4. C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of Royal Statistical Society: Series B*, 72:269, 2010.
5. J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the second international conference on genetic algorithms*, page 14, 1987.
6. M. Bolic, P. M. Djuric, and S. Hong. Resampling algorithms and architectures for distributed particle filters. *IEEE Transactions on Signal Processing*, 53:2442, 2005.
7. R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistical Computing*, 18:125, 2007.
8. C. Bretó, D. He, E. L. Ionides, and A. A. King. Time series analysis via mechanistic models. *Annals of Applied Statistics*, 3:319, 2009.
9. J. L. Costello, I. G. Castro, C. Hacker, T. A. Schrader, J. Metz, D. Zeuschner, A. S. Azadi, L. F. Godinho, V. Costina, P. Findeisen, A. Manner, M. Islinger, and M. Schrader. ACBD5 and VAPB mediate membrane associations between peroxisomes and the ER. *Journal of Cell Biology*, 216:331, 2017.
10. C. Denesvre and V. Malhotra. Membrane fusion in organelle biogenesis. *Current Opinions in Cell Biology*, 8:519, 1996.
11. A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, 2001.
12. A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12:656, 2009.
13. C. Fuchs. *Inference for diffusion processes: with applications in life sciences*. Springer, 2013.
14. Y. Fujiki, K. Okumoto, S. Mukai, M. Honsho, and S. Tamura. Peroxisome biogenesis in mammalian cells. *Frontiers in Physiology*, 5, 2014.
15. D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403, 1976.
16. D. T. Gillespie. *Markov processes: an introduction for physical scientists*. Elsevier, 1991.
17. D. T. Gillespie. The chemical Langevin equation. *Journal of Chemical Physics*, 113:297, 2000.
18. A. Golightly and D. J. Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface focus*, 1:807, 2011.
19. M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 30:1649, 2011.
20. T. G. Hallam and S. A. Levin. *Mathematical ecology: an introduction*. Springer, 2012.
21. J. Huotari and A. Helenius. Endosome maturation. *EMBO Journal*, 30:3481, 2011.
22. S. J. Huybrechts, P. P. Van Veldhoven, C. Brees, G. P. Mannaerts, G. V. Los, and M. Fransen. Peroxisome dynamics in cultured mammalian cells. *Traffic*, 10:1722, 2009.
23. E. L. Ionides, C. Bretó, and A. A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103:18438, 2006.
24. M. Jauregui and P. K. Kim. Probing peroxisome dynamics and biogenesis by fluorescence imaging. *Current protocols in cell biology*, page 21, 2014.

25. P. M. Jean Beltran, R. A. Mathias, and I. M. Cristea. A portrait of the human organelle proteome in space and time during cytomegalovirus infection. *Cell Systems*, 3:361, 2016.
26. P. K. Kim, R. T. Mullen, U. Schumann, and J. Lippincott-Schwartz. The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER. *Journal of Cell Biology*, 173:521, 2006.
27. P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.
28. H. Koeppl, C. Zechner, A. Ganguly, S. Pelet, and M. Peter. Accounting for extrinsic variability in the estimation of stochastic rate constants. *International Journal of Robustness and Nonlinear Control*, 22:1103, 2012.
29. M. Lowe and F. A. Barr. Inheritance and biogenesis of organelles in the secretory pathway. *Nature Reviews Molecular and Cellular Biology*, 8:429, 2007.
30. J. Míguez. Analysis of parallelizable resampling algorithms for particle filtering. *Signal Processing*, 87:3155, 2007.
31. J. Míguez, M. F. Bugallo, and P. M. Djuric. A new class of particle filters for random dynamic systems with unknown statistics. *EURASIP Journal on Advances in Signal Processing*, 2004:303619, 2004.
32. P. Milner, C. S. Gillespie, and D. J. Wilkinson. Moment closure based parameter inference of stochastic kinetic models. *Statistical Computing*, 23:287, 2012.
33. A. M. Motley and E. H. Hettema. Yeast peroxisomes multiply by growth and division. *The Journal of Cell Biology*, 178:399, 2007.
34. S. Mukherji and E. K. O'Shea. Mechanisms of organelle biogenesis govern stochastic fluctuations in organelle abundance. *eLife*, 3:e02678, 2014.
35. B. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer, 2014.
36. P. S. Pacheco. *Parallel programming with MPI*. Morgan Kaufmann, 1997.
37. B. Poole, T. Higashi, and C. de Duve. The synthesis and turnover of rat liver peroxisomes: Iii. the size distribution of peroxisomes and the incorporation of new catalase. *The Journal of Cell Biology*, 45:408, 1970.
38. M. A. Rizzo, M. W. Davidson, and D. W. Piston. Fluorescent protein tracking and detection: Fluorescent protein structure and color variants. *Cold Spring Harbor Protocols*, 4, 2009.
39. O. Rosén, A. Medvedev, and M. Ekman. Speedup and tracking accuracy evaluation of parallel particle filter algorithms implemented on a multicore architecture. In *Control Applications (CCA), 2010 IEEE International Conference on*, page 440. IEEE, 2010.
40. J. Z. Sexton, Q. He, L. J. Forsberg, and J. E. Brenman. High content screening for non-classical peroxisome proliferators. *International Journal of High Throughput Screening*, 2010:127, 2010.
41. V. Sica, L. Galluzzi, J. M. Bravo-San Pedro, and V. Izzo. Organelle-specific initiation of autophagy. *Molecular Cell*, 59:522, 2015.
42. J. J. Smith and J. D. Aitchison. Peroxisomes take shape. *Nature Reviews Molecular and Cellular Biology*, 14:803, 2013.
43. S. J. Steinberg, G. V. Raymond, N. E. Braverman, and A. B. Moser. *Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum*. U. of Washington, Seattle, 1993.
44. I. Strid. Parallel particle filters for likelihood evaluation in dsge models: An assessment. Technical report, Society for Computational Economics, 2006.

45. A. van der Zand, J. Gent, I. Braakman, and H. F. Tabak. Biochemically distinct vesicles from the endoplasmic reticulum fuse to form peroxisomes. *Cell*, 149:397, 2012.
46. N. G. Van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 1992.
47. C. Vonesch and M. Unser. A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Transactions on Image Processing*, 17:539, 2008.
48. T. Wai and T. Langer. Mitochondrial dynamics and metabolic regulation. *Trends in Endocrinology & Metabolism*, 27:105, 2016.
49. Y. Wang, S. Christley, E. Mjolsness, and X. Xie. Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent. *BMC Systems Biology*, 4:1, 2010.
50. J. C. Waters. Accuracy and precision in quantitative fluorescence microscopy. *The Journal of Cell Biology*, 185:1135, 2009.
51. D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10:122, 2009.
52. D. J. Wilkinson. *Stochastic modelling for systems biology, Second edition*. CRC Press, 2011.
53. C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences*, 109:8340, 2012.
54. C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11:197, 2014.
55. S. Zenker. Parallel particle filters for online identification of mechanistic mathematical models of physiology from monitoring data: performance and real-time scalability in simulation scenarios. *Journal of Clinical Monitoring and Computing*, 24:319, 2010.