THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Cox process representation and inference for stochastic reaction-diffusion processes

OPEN ACCESS

Cox process representation and inference for stochastic reaction–diffusion processes

David Schnoerr[1,2,3], Ramon Grima[1,3] & Guido Sanguinetti[2,3]

Complex behaviour in many systems arises from the stochastic interactions of spatially distributed particles or agents. Stochastic reaction–diffusion processes are widely used to model such behaviour in disciplines ranging from biology to the social sciences, yet they are notoriously difficult to simulate and calibrate to observational data. Here we use ideas from statistical physics and machine learning to provide a solution to the inverse problem of learning a stochastic reaction–diffusion process from data. Our solution relies on a non-trivial connection between stochastic reaction–diffusion processes and spatio-temporal Cox processes, a well-studied class of models from computational statistics. This connection leads to an efficient and flexible algorithm for parameter inference and model selection. Our approach shows excellent accuracy on numeric and real data examples from systems biology and epidemiology. Our work provides both insights into spatio-temporal stochastic systems, and a practical solution to a long-standing problem in computational modelling.

[1] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, UK. [2] School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK. [3] SynthSys, University of Edinburgh, Edinburgh EH9 3JD, UK. Correspondence and requests for materials should be addressed to G.S. (email: gsanguin@inf.ed.ac.uk).

Many complex behaviours in several disciplines originate from a common mechanism: the dynamics of locally interacting, spatially distributed agents. Examples arise at all spatial scales and in a wide range of scientific fields, from the microscopic interactions of low-abundance molecules within cells, to ecological and epidemic phenomena at the continental scale. Frequently, stochasticity and spatial heterogeneity play a crucial role in determining the process dynamics and the emergence of collective behaviour[1–8].

Stochastic reaction–diffusion processes (SRDPs) constitute a convenient mathematical framework to model such systems. SRDPs were originally introduced in statistical physics[9,10] to describe the collective behaviour of populations of point-wise agents performing Brownian diffusion in space and stochastically interacting with other, nearby agents according to predefined rules. The flexibility afforded by the local interaction rules has led to a wide application of SRDPs in many different scientific disciplines where complex spatio-temporal behaviours arise, from molecular biology[4,11,12], to ecology[13], and to the social sciences[14].

Despite their popularity, SRDPs pose considerable challenges, as analytical computations are only possible for a handful of systems[8]. Thus, many analytical techniques that are widely used for non-spatial stochastic systems cannot be used for SRDPs. From the practical point of view, perhaps the single most important outstanding problem is inference in SRDP models: given observations of the system, can we reconstruct the interaction rules/local dynamic parameters? Solving this inverse problem would be important, as it would allow to quantitatively assess model fit to data and to compare different models/hypotheses in the light of observations.

SRDPs are generally analysed by either Brownian dynamics simulations of individual particles or by resorting to spatial discretization, leading to the so-called 'reaction–diffusion master equation' (RDME)[15,16]. The computational complexity of the RDME obviously increases as the spatial discretization becomes finer, and in many cases the limiting process does not lead to the original SRDP[17]. Significant effort has been spent to improve the performance of the two types of simulations[18–24]; however, the computational costs are still high and quickly become prohibitive for larger systems. More importantly, the lack of an analytical alternative to simulations means that evaluating the fit of a model to observations (the likelihood function) is computationally extremely expensive, effectively ruling out statistical analyses. As far as we are aware, the few attempts at statistical inference for SRDPs either used simulation-based likelihood free methods[13], inheriting the intrinsic computational difficulties discussed above, or abandoned the SRDP framework by adopting a coarse space discretization[25] or neglecting the individual nature of agents using a linear-noise approximation[26].

In this paper, we propose an approximate solution to the problem of computing the likelihood function in SRDPs, thus providing a principled solution to the inverse problem of model calibration. Using the classical theory of the Poisson representation (PR) for stochastic reaction processes[27], we show that marginal probability distributions of SRDPs can be approximated in a mean-field sense by spatio-temporal Cox point processes, a class of models widely used in spatio-temporal statistics[28]. Cox processes model the statistics of point patterns via an unobserved intensity field, a random function that captures the local mean of the observed point process. This relationship between SRDPs and Cox processes is surprising, as SRDPs are mechanistic, microscopic descriptions of spatio-temporal systems, while Cox processes are employed phenomenologically to explain regularities in point patterns. This novel link between these two classes of models enables us to formally associate an SRDP with a continuous evolution equation on the local statistics of the process in terms of (stochastic) partial differential equations (SPDEs). Crucially, this novel representation of SRDPs allows us to efficiently approximate multiple-time marginals and thus data likelihoods associated with observed point patterns, enabling us to leverage the rich statistical literature on spatio-temporal point processes for parameter estimation and/or Bayesian inference[28,29].
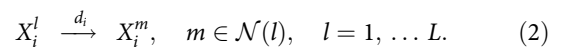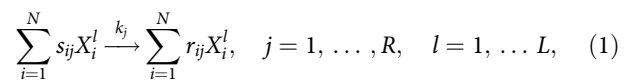
We demonstrate the efficiency and accuracy of our approach for the problem of parameter inference and model selection by means of a number of numerical and real data examples, using non-trivial models from systems biology and epidemiology. Our results provide both a valuable resource for performing statistical inference and assessing model fit in this important class of models, and a novel conceptual perspective on spatio-temporal stochastic systems.

Results

SRDPs and the PR. In the classical Doi interpretation[9,10], which we adopt here, SRDPs describe the evolution of systems of point particles performing Brownian diffusion in a spatial domain $\mathcal{D}$. While SRDPs are used in a variety of disciplines, we will use the language of chemical reactions to describe them in the following. We assume the existence of N different types of particles X_i, or species, which can interact through a set of predefined rules or reactions. We assume the structure of the model, that is, which reactions are possible, to be known exactly; later, we will relax this assumption to allow for the existence of a (finite) number of possible alternative mechanisms. Each particle of a particular species X_i performs Brownian diffusion with a species-specific microscopic diffusion constant D_i. Bimolecular reactions between particles occur with a certain rate whenever two particles are closer than some specified reaction range. In principle, both reaction and microscopic diffusion rate constants may be space dependent, for example, to account for geometric constraints; for simplicity, we will only describe the homogeneous case here.

SRDPs are frequently analysed via coarse graining by discretizing space and assuming dilute and well-mixed conditions in each compartment; in this case the dynamics in each compartment is described by the chemical master equation[30]. Modelling diffusion of particles between neighbouring compartments as unimolecular reactions leads to the RDME[15], which describes the dynamics of a continuous-time Markov jump process. For systems with only zero- or first-order reactions, the RDME converges to Brownian dynamics in the continuum limit. For non-linear systems, however, this is not the case in two or more dimensions because the rate of bimolecular reactions converges to zero, independently of the scaling of the corresponding rate constant[17].

Consider now a set of chemical species X_i in a finite volume divided into L cubic compartments of edge length h, interacting via the following R reactions:

$$\sum_{i=1}^{N} s_{ij} X_i^l \xrightarrow{k_j} \sum_{i=1}^{N} r_{ij} X_i^l, \quad j = 1, \ldots, R, \quad l = 1, \ldots L, \quad (1)$$

$$X_i^l \xrightarrow{d_i} X_i^m, \quad m \in \mathcal{N}(l), \quad l = 1, \ldots L. \quad (2)$$

Here, s_{ij} and r_{ij} are the number of reactants and product particles of species X_i in the jth reaction, respectively, k_j is the rate constant of the jth reaction, X_i^m denotes species X_i in the mth compartment and d_i is the diffusion rate constant of species X_i. The latter is related to the microscopic diffusion constant D_i via $d_i = D_i/h^2$. We assume homogenous diffusion here, that is, d_i is independent of the compartment position, but it would be straightforward to extend the following analysis to space-dependent diffusion. $\mathcal{N}(l)$

denotes all the adjacent compartments of the lth compartment. Equation (1) describes chemical reactions happening in single compartments, whereas equation (2) describes diffusion between adjacent compartments. We confine our analysis to reactions with at most two reactant and at most two product particles, that is, $\sum_{i=1}^{N} s_{ir}$, $\sum_{i=1}^{N} r_{ir} \leq 2$, $r = 1, \ldots, R$, since higher-order reactions rarely occur in nature. We call a reaction linear if $\sum_{i=1}^{N} s_{ir} \leq 1$ and bimolecular if $\sum_{i=1}^{N} s_{ir} = 2$. Under well-mixed and dilute conditions in each compartment, the evolution of marginal probabilities of this system is given by the RDME that is given in general form in the Methods section.

In the case of only a single compartment, that is, $L = 1$, the state of the system is given by $\mathbf{n} = (n_1, \ldots, n_N)$, where n_i is the number of X_i particles in the system, and the time evolution of the single-time probability distribution $P(\mathbf{n}, t)$ is determined by the chemical master equation (see the Methods section). Gardiner derived an alternative description of the dynamics of such a system by making the ansatz of writing $P(\mathbf{n}, t)$ as a Poisson mixture[27,31]:

$$P(\mathbf{n}, t) = \int d\mathbf{u} \prod_{i=1}^{N} \mathcal{P}(n_i; u_i) p(\mathbf{u}, t), \quad u_i \in \mathbb{C}, \quad (3)$$

where $\mathbf{u} = (u_1, \ldots, u_N)$ and $\mathcal{P}(n_i; u_i) = (e^{-u_i} u_i^{n_i})/n_i!$ is a Poisson distribution in n_i with mean u_i, and the u_i are complex-valued in general. Using this ansatz in the chemical master equation one can derive an exact Fokker–Planck equation for $p(\mathbf{u}, t)$ or equivalently a Langevin equation for $\mathbf{u}(t)$[27] (see Methods section for more details). Gardiner derived this result for the non-spatial chemical master equation and applied it to the RDME to study the corresponding continuum limit. While the PR provides an elegant analytical tool to study reaction systems, its applicability is severely hindered by the fact that the Poisson variables u are in general complex valued and hence lack a clear physical interpretation; in particular, all bimolecular reactions and all linear reactions with two non-identical product particles give rise to a complex-valued PR (for a taxonomy of which reaction systems become complex valued, see Supplementary Methods).

Cox process representation. While in the classical view of the PR, the auxiliary variables u_i are simply introduced as a mathematical device, we can make some progress by considering a joint process over the u_i and n_i variables. Formally, this is equivalent to what in statistics is called demarginalization: a complex process is replaced by a (simpler) process in an augmented state space, such that the marginals of the augmented process return exactly the initial process. To formalize this idea, we first introduce some concepts from spatial statistics.

A (spatial) Poisson process[32] is a measure on the space of zero-dimensional subsets of a domain $\mathcal{D}$; in this work, we consider Poisson processes that admit an intensity function $u(x)$, which gives the rate of finding a point in an infinitesimally small spatial region. The number of points in a finite spatial region is then a Poisson random variable, with mean given by the integral of the intensity function over that region. A Cox process (also called 'doubly stochastic Poisson process') is a generalization of a Poisson process, where the intensity field is itself a random process. Conditioned on a realization of the intensity field, the Cox process reduces to a Poisson process (see the Methods section for a more detailed definition of Poisson and Cox processes). We will consider families of spatial Poisson (Cox) process indexed by a time variable; importantly, in this case the intensity field can be thought of as the state variable of the system, with the actual spatial points being noisy realizations of this state (see Fig. 1 for a graphical explanation). Our first observation

follows directly from Gardiner's analysis of the continuum limit of the RDME (see Supplementary Methods for a proof).

Remark 1. Consider an SRDP on a spatial domain $\mathcal{D}$ and temporal domain $\mathcal{T}$, and let all reactions involve production or decay of at most one particle. Then, for appropriate initial conditions, $\forall t \in \mathcal{T}$ the single-time-point spatial probability distribution of the SRDP is exactly the same as of a spatial Poisson process.

General SRDPs. We can build on this point process representation to develop novel, mathematically consistent approximation schemes for SRDPs in general. Consider, for example, a bimolecular reaction of the type $A + B \xrightarrow{k} C$ with propensity function $f(n_A, n_B) = k n_A n_B / \Omega$, where n_A and n_B are the number of A and B particles in the system, respectively, and Ω is the system volume. While the PR for such systems is complex valued, we can formally obtain a real system by applying a mean-field approximation that replaces the bimolecular reaction $A + B \xrightarrow{k} C$ by the two reactions $A \xrightarrow{k \langle n_B \rangle / \Omega} C$ and $B \xrightarrow{k \langle n_A \rangle / \Omega} C$ with propensity functions $f(n_A, n_B) = k n_A \langle n_B \rangle / \Omega$ and $f(n_A, n_B) = k n_B \langle n_A \rangle / \Omega$, respectively. Here, $\langle \cdot \rangle$ denotes the expectation of a random variable with respect to its marginal distribution. The proposed approximation hence replaces the direct interaction of the particles by an effective interaction of A with the mean field of B and vice versa. Other bimolecular reactions and linear reactions with two non-identical product particles can be approximated in a similar manner. This leads to a real-valued evolution equation for the u_i variables, see Methods and Supplementary Methods for details and examples.

Applying this approximation to a general RDME, and subsequently the PR and taking the continuum limit gives the following set of N coupled SPDEs (see Methods section for a derivation)

$$du_i(x, t) = \left[D_i \Delta u_i(x, t) + \sum_{r=1}^{R} S_{ir} g_r(\mathbf{u}(x, t)) \right] dt$$
$$+ \sum_{r'} \sqrt{2 g_{r'}(\mathbf{u}(x, t))} dW_{r'}(x, t), \quad (4)$$

where the sum over r' runs over all reactions with two product particles of species X_i. In particular, this means that in the absence of reactions with two identical product molecules the diffusion term in equation (4) vanishes and equation (4) reduces to a partial differential equation (PDE), that is, the $u_i(x, t)$ are deterministic. x in equation (4) is a spatial location, $D_i = h^2 d_i$ is the microscopic diffusion constant of species X_i, Δ is the Laplace operator, $\mathbf{u}(x, t) = (u_1(x, t), \ldots, u_N(x, t))$, where $u_i(x, t)$ is the intensity field of species X_i, $dW_{r'}(x, t)$ is spatio-temporal Gaussian white noise and we have defined the propensity functions $g_r(\mathbf{u}(x, t))$ in PR space. The latter are obtained by applying the mean-field approximation to the propensity functions $f_r(\mathbf{n})$ and subsequently replacing $n_i \to u_i(x, t)$ and $\langle n_i \rangle \to \langle u_i(x, t) \rangle$. Note that the latter denotes a local expectation of the stochastic random field $u_i(x, t)$, rather than a spatial averaging. See Methods and Supplementary Methods for more details and examples.

To obtain equation (4), we approximated bimolecular reactions by linear reactions. Note, however, that the propensity functions of the latter reactions depend on the mean fields of certain species. This means that the resulting SPDEs in equation (4) are generally non-linear and hence in principle are able to capture non-linear dynamical behaviours.

Equation (4) looks similar to the spatial chemical Langevin equation[33], but has a different interpretation here, since it
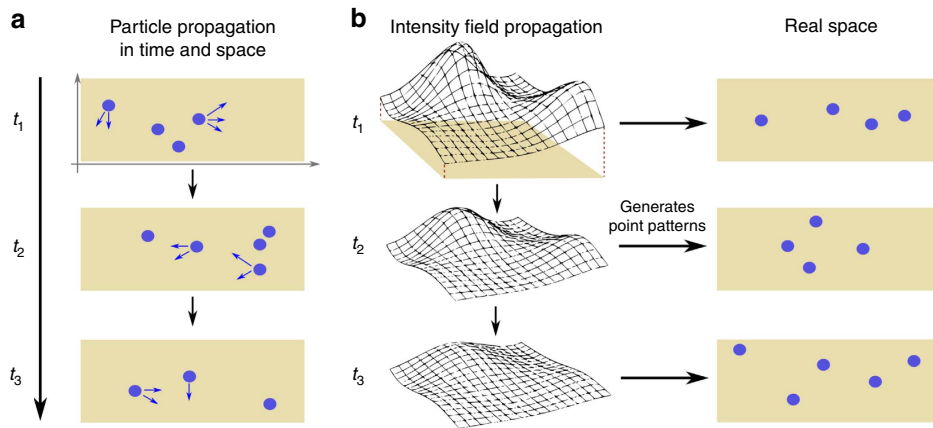
Figure 1 | Visualization of Cox process approximation of SRDPs for multi-time points. (a) Time evolution of the true SRDP in space. Particles diffuse in space, may decay or are created and react with each other. (**b**) Time evolution of a Cox process. Here, the intensity field evolves in time, rather than the points in real space. The latter are merely noisy realizations of the intensity field. In particular, the point patterns at two different time points are independent of each other conditioned on the intensity field.

describes the intensity in PR space. In particular, just as any other PDE or SPDE description in real space, the spatial chemical Langevin equation does not provide a generative model for the actual location of the events, and thus would not allow us to directly model statistically particle locations. Notice that, as a consequence of the mean-field approximation, the mean values of the u_i fields are the same as in a deterministic rate equation description; however, the dynamics of the observed variable, that is, the points in space, remain stochastic even when the intensity field evolves deterministically. We can therefore extend Remark 1 to obtain the following result (see Supplementary Methods for a proof).

Result 1. Consider the same setting as in Remark 1. Under appropriate initial conditions, if there is at least one linear reaction with two product particles of the same species, the system's single-time-point distribution is exactly the same as of a Cox process, whose intensity fulfils the SPDE given in equation (4). If the system involves other types of reactions, including bimolecular reactions, the single-time probability distribution of the SRDP is approximated in a mean-field sense by that of a Poisson (Cox) process, whose intensity fulfils equation (4).

The likelihood function. Result 1 provides an efficient means to calculate statistics such as expected number of agents within a certain volume, without the need to perform extensive Monte Carlo simulations, since it only requires to solve a (S)PDE for which a rich literature of numerical methods exists[28,29]. The numerical methods used in this paper are presented in Supplementary Methods. Importantly, we can use Result 1 to approximate the likelihood function of a configuration of points arising from an SRDP, by using the well-known Cox process likelihood: if $u(x,t)$ is the intensity of a spatio-temporal Cox process with distribution $p(u(x,t))$ and $\mathbf{y}$ a given measurement at time t_0, the corresponding likelihood is given by[28]

$$p(\mathbf{y}) = \int \mathcal{D}u(x,t_0) \prod_{s \in \mathbf{y}} u(s,t_0) e^{-\int dx u(x,t_0)} p(u(x,t_0)). \quad (5)$$

This function can be easily optimized to yield statistical estimates of kinetic parameters from single-time observations.

Remark 2. We would like to emphasize that in the case of a Cox process, that is, a stochastic intensity field, the number of particles in two non-overlapping spatial regions are correlated random

variables in general. The reason is that the PR ansatz in equation (3) is not merely a product of Poisson distributions, but rather an integral over such a product weighted by a corresponding mixing distribution. In the case of a Poisson process, that is, a deterministic intensity field, in contrast, the numbers of particles in two non-overlapping spatial regions are always uncorrelated.

Time-series observations. We next consider the problem of approximating the joint distribution of point patterns arising from an SRDP at different time points. This is important when we have time-series observations, that is, spatial measurements $\mathbf{y} = (\mathbf{y}_{t_0}, \ldots, \mathbf{y}_{t_n}), \mathbf{y}_{t_i} \subset \mathcal{D}$ at discrete time points $t_0,\ldots,t_n$, and we want to compute the likelihood $p(\mathbf{y}|\Theta)$ of the data given a model Θ. Since the system is Markovian the likelihood factorizes as

$$p(\mathbf{y}|\Theta) = p\left(\mathbf{y}_{t_0}|\Theta\right) \prod_{i=1}^{n} p\left(\mathbf{y}_{t_i}|\mathbf{y}_{t_{i-1}}, \Theta\right). \quad (6)$$

We would like to approximate this likelihood, using the relation to Cox processes established in Result 1. While this is in principle straightforward, computing the terms $p(\mathbf{y}_{t_i}|\mathbf{y}_{t_{i-1}}, \Theta)$ involves determining the distribution over the associated $u_i(x,t)$ fields in PR space. This would involve inverting the PR transformation in equation (3), which is computationally inconvenient. Instead, we opt for an approximation strategy: assume that we have determined the PR distribution $p(\mathbf{u}_{t_{i-1}})$ at time t_{i-1}, where we introduced the shorthand $\mathbf{u}_{t_i} = \mathbf{u}(x, t_i)$. By definition of the intensity of a Poisson process, $\mathbf{u}_{t_{i-1}}$ represents the expectation of the random configuration of points $\mathbf{y}_{t_{i-1}}$ at time t_{i-1}. We then approximate $p(\mathbf{y}_{t_i}|\mathbf{y}_{t_{i-1}}, \Theta)$ in a mean-field way by replacing the explicit dependence on $\mathbf{y}_{t_{i-1}}$ with its expectation $p(\mathbf{y}_{t_i}|\mathbf{y}_{t_{i-1}}, \Theta) \approx \langle p(\mathbf{y}_{t_{i-1}}|\mathbf{y}_{t_{i-1}}, \Theta) \rangle_{p(\mathbf{y}_{t_{i-1}}|\mathbf{u}_{t_{i-1}})} = p(\mathbf{y}_{t_i}|\mathbf{u}_{t_{i-1}}, \Theta)$. Figure 1 visualizes this approximation. Figure 1a shows the time evolution in an SRDP, while Fig. 1b shows the time evolution of a corresponding approximating Cox process. This leads to a new interpretation of the measured points $\mathbf{y} = (\mathbf{y}_{t_0}, \ldots, \mathbf{y}_{t_n})$: while they are snapshots of the actual state in the true system, they correspond to the noisy realizations of the state $\mathbf{u}(x,t)$ in the Cox process picture. We thus have the following result.

Result 2. The joint n-time-point marginal distribution of an SRDP can be approximated in a mean-field sense by the joint

probability of a Poisson (Cox) process, with intensity governed by the (S)PDE in equation (4).

Relation to Gardiner's work. As mentioned before, Gardiner had already derived similar equations as equation (4) for single-time marginals of SRDPs[27]. Crucially, however, an approximation scheme for multiple-time joint marginals was, to our knowledge, never proposed; multi-time joint marginals are necessary for inference from time-series observations, hence the importance of Result 2. Furthermore, Gardiner's approach generally leads to a complex-valued PR; this motivates the novel approximation schemes of certain reactions that we introduced here. It is this novel real-valued PR, together with the interpretation of the PR variables as state variables, which allows us to derive the novel relation between SRDPs and spatio-temporal Poisson (Cox) processes.

Inference. Result 2 is particularly powerful statistically, because it enables us to analytically approximate the exact (intractable) likelihood $p(\mathbf{y}|\Theta)$ in equation (6), by the likelihood of a spatio-temporal Cox process with intensity $u(x,t)$. The intensity itself follows the dynamics imposed by the PR in equation (4); importantly, the PR explicitly links the dynamic parameters governing the evolution of the intensity function to the microscopic diffusion and reaction rate constants of the SRDP.

Parameter estimation can therefore be performed efficiently by maximizing the Cox process likelihood. In the simpler case where the intensity function evolves deterministically, the likelihood can be evaluated numerically via the solution of a system of PDEs, and the dynamic parameters can be numerically recovered using standard optimization algorithms. In the case where the intensity function evolves stochastically, we evaluate the likelihood by an approximate filtering approach, as commonly used in many statistical and engineering applications (see Supplementary Methods for algorithmic details).

The availability of a likelihood function enables us to provide a statistically meaningful, data-driven assessment of how well a model describes the data. This is particularly important when there is uncertainty as to the precise mechanism underlying the data, for example, the exact reactions or species involved. Likelihood estimates, appropriately penalized to account for model complexity, can then be used to select models according to their support from the data.

It is important to notice that our approach directly optimizes the kinetic parameters of the model, rather than fitting an intensity function to the observed points and then fitting the dynamics. Since kinetic parameters are usually much fewer than the number of observations available, the risk of over-fitting is generally low in our approach.

Next, we apply Result 1 and Result 2 to several examples, and perform parameter inference by maximizing the data likelihood. We solve the corresponding (S)PDEs numerically by projecting them onto a finite set of spatial basis functions, see Supplementary Methods for details. The software used is available as Supplementary Data 1–4.

Parameter estimation for a gene expression system. To demonstrate the accuracy of our method, we first consider simulated time-series data in this section. Consider a gene expression system as illustrated in Fig. 2. For simplicity, we consider a one-dimensional version of this system here, with the nucleus located at one side of the cell. A gene located in the nucleus is transcribed into messenger RNA (mRNA) molecules. The latter decay and diffuse across the whole cell and are translated into proteins in the cytosol. The protein molecules also
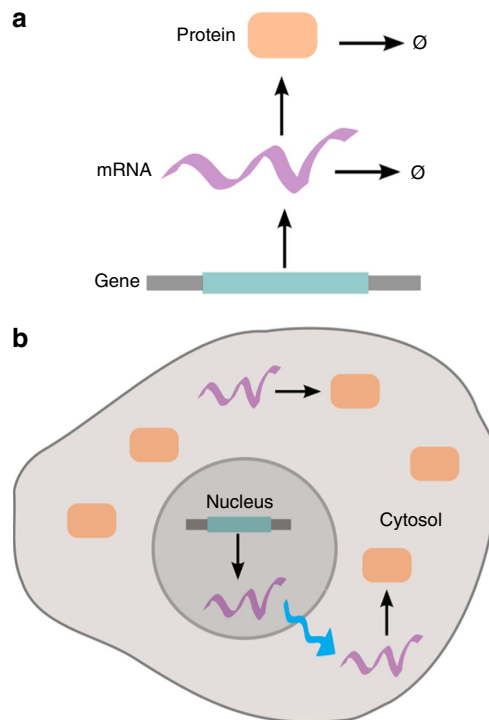
Figure 2 | Gene expression system. (**a**) Chemical reactions taking place. (**b**) Illustration of system. The mRNA becomes transcribed in the nucleus, and becomes translated to proteins in the cytosol. mRNA and protein molecules decay stochastically and undergo Brownian diffusion across the whole cell.

diffuse across the whole cell and decay. For simplicity, we do not model the gene explicitly, but assume that mRNA becomes transcribed with a certain fixed rate m_1 homogeneously in the nucleus. The corresponding reactions are

$$\varnothing \xrightarrow{m_1} M, \quad M \xrightarrow{m_2} \varnothing, \tag{7}$$

$$M \xrightarrow{p_1} M + P, \quad P \xrightarrow{p_2} \varnothing, \tag{8}$$
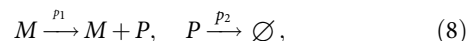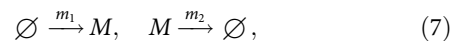
where M and P denote the mRNA and protein, respectively. For this system, the SPDE of our method in equation (4) becomes deterministic and thus corresponds to a Poisson process.

In addition to the reaction parameters m_1, m_2, p_1 and p_2, we have to infer the nucleus size r, as well as the diffusion rates d_m and d_p of the mRNA and protein, respectively, summing up to a total number of seven parameters. We assume that the positions of the protein molecules are observed at 30 time points, while the mRNA is unobserved. The results for one parameter set are shown in Table 1. Considering that we observe the protein at only 30 time points with unobserved mRNA and that we have seven unknown parameters, the inferred average values are remarkably close to the exact values. Moreover, the s.d.'s of the inferred results for single data sets are small, demonstrating the accuracy and precision of our method.

Next, we extend the system in Fig. 2 by adding an autocatalytic reaction for the protein,

$$P \xrightarrow{p_3} P + P. \tag{9}$$

Including this reaction leads to a non-vanishing noise term in equation (4) and the system corresponds to a Cox process. We note that the system has a steady state only if $p_3 < p_2$, with an otherwise exponentially growing mean protein number. On the mean level only the difference $p_2 - p_3$ is identifiable, and we fix

Table 1 | Inference results for gene expression system.

	r	d_m	d_p	m_1	m_2	p_1	p_2
Exact	0.3	0.1	0.1	20	0.5	20	0.2
Inferred	0.31 (0.06)	0.12 (0.08)	0.14 (0.06)	23 (12)	0.51 (0.4)	26 (18)	0.25 (0.1)

The table shows the inferred parameter values for the gene expression system illustrated in Fig. 2, with reactions in equations (7) and (8). We assume the measurements of the protein, while the mRNA is unobserved. The inference is carried out by maximizing the likelihood of simulated data for 30 measurement points separated by $\Delta t = 0.5$. This procedure is carried out for 100 simulated data sets and the mean value and s.d. (in parenthesis) of the inferred results are displayed.

Table 2 | Inference results for gene expression system with additional autocatalytic reaction.

	r	d_m	d_p	m_1	m_2	p_1	p_2
Exact	0.3	0.1	0.1	20	0.5	20	0.2
Inferred	0.30 (0.05)	0.14 (0.08)	0.088 (0.03)	27 (17)	0.57 (0.3)	24 (21)	0.19 (0.08)

The table shows the inferred parameter values for the gene expression system illustrated in Fig. 2, with reactions in equations (7) and (8), and the additional autocatalytic reaction in equation (9). Since, only the difference $p_2 - p_3$ is identifiable, we fix $p_3 = 0.01$ and infer the other seven parameters. The table shows the average and s.d. (in parenthesis) of the inference results for 100 simulated data set.

Table 3 | Inference results for SIRS system.

	$10^3 \times d$	$10 \times r$	$10 \times s$	$10^3 \times k^{PR}$	k	w	S_{ini}
Exact	1	0.2	2	—	100	0.01	200
Inferred	0.8 (0.3)	0.19 (0.09)	1.8 (1.2)	2.5 (0.5)	—	—	
Exact	1	0.2	2	—	100	0.01	300
Inferred	0.9 (0.4)	0.15 (0.06)	1.4 (0.9)	2.4 (0.5)	—	—	
Exact	1	2	2	—	100	0.02	200
Inferred	1.0 (0.6)	1.6 (0.7)	1.5 (1.0)	3.4 (1.1)	—	—	
Exact	1	0.2	2	—	1000	0.005	200
Inferred	0.8 (0.4)	0.21 (0.11)	2.2 (1.6)	2.4 (0.5)	—	—	
Exact	2	0.2	2	—	100	0.01	100
Inferred	1.6 (0.8)	0.19 (0.09)	1.9 (1.2)	4.6 (1.1)	—	—	

The table shows the results for parameter inference for the SIRS system, with reactions given in equation (10). The inference is carried out by maximizing the likelihood of the simulated data for 40 measurement points. This procedure is carried out for 200 simulated data sets, and the mean value and s.d. (in parenthesis) of the inferred results are displayed.

$p_3 = 0.01$. We thus infer the same parameters as in the previous case, but this time modelled as a Cox process. Table 2 shows the results, indicating the accuracy of our method. See Supplementary Methods for more information on the used equations and algorithmic details.

Parameter estimation for an SIRS model. We next consider the SIRS system, a popular model for describing the dynamics of an infection spreading through a population. Such systems are frequently modelled as SRDPs[34] or discretized versions thereof[35]. We consider a system in the two-dimensional square $[0,1] \times [0,1]$. The system comprises a susceptible (S), an infected (I) and a recovered species (R), which perform Brownian diffusion and interact via the reactions

$$S + I \xrightarrow{k,w} 2I, \quad I \xrightarrow{r} R, \quad R \xrightarrow{s} S, \qquad (10)$$
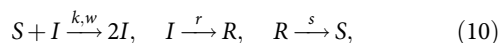
where the bimolecular infection is characterized by the microscopic reaction rate k and the reaction range w. We assume that all three species diffuse with the same diffusion rate d. We assume further that initially there are no recovered (R) particles, S_{ini} susceptible (S) particles placed uniformly over the whole area and one infected (I) particle located at [0.05, 0.05]. We consider the case that all three species are observed and perform inference for 40 simulated data points using Result 2, thereby replacing k and w by an effective bimolecular reaction parameter k^{PR}. The model thus has four parameters that need to be inferred:

the diffusion rate d, the recovery rate r, the susceptible rate s and the bimolecular infection rate k^{PR}. Table 3 shows the corresponding results, demonstrating the accuracy and precision of our method. The computational efficiency of our method in comparison with stochastic simulations is particularly pronounced here. For the first parameter set in Table 3, for example, the Brownian dynamics simulation of a single realization of the system takes ~ 250 s, while the whole inference procedure for the four parameters using our method takes only ~ 10 s for one simulated data set on a 3.1 GHz core. See Supplementary Methods for more details.

Figure 3 visualizes the dynamics of the SIRS system for one parameter set. Individual points from a simulation are shown in different colours (turquoise for S, bronze for I and blue for R), while the background colours represent a superposition of the respective intensity fields with optimized parameters. Notice how the PR approximation is able to capture the emerging behaviour of a wave of infection sweeping through the domain from bottom left to top right, before the establishment of a dynamic equilibrium between the three populations. Such a phenomenon is clearly due to the spatial aspect of the system, and could not have been recovered using an inference method that does not incorporate spatial information. Indeed, the overall number of infected individuals rises rapidly and remains essentially constant between time 20 and 35 before dropping to steady state, a behaviour that is simply not possible in a non-spatial SIRS model.
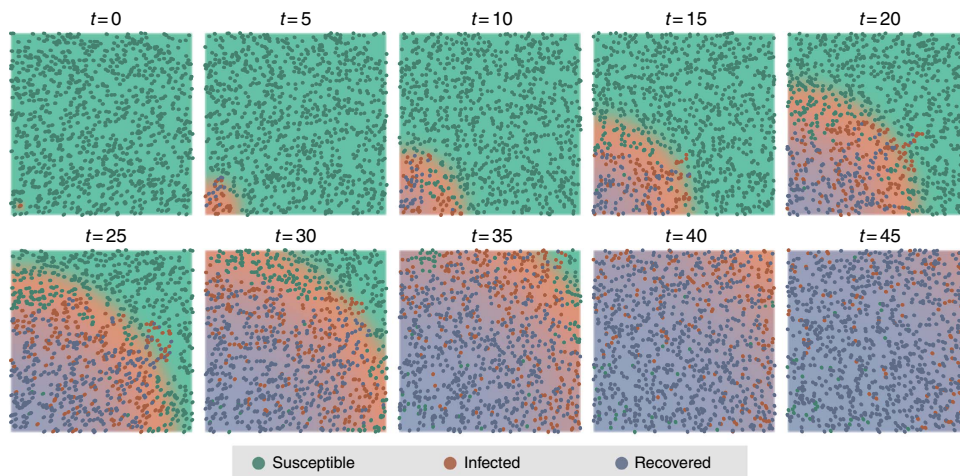
Figure 3 | Dynamics of SIRS system. The figure shows the time evolution of a single simulated realization (points) and of the prediction of our method (background colours) for the SIRS system with reactions in equation (10) for time $t = 0$–45, with steps of $\Delta t = 5$. For the simulation, we use the parameters $(S_{\text{ini}}, k, w, d, r, s) = (10^3, 10^4, 0.02, 0.0002, 0.3, 0.01)$ and for the point process prediction the corresponding inferred parameters. The background is an RGB image with the three colour components being proportional to the intensity fields of the three species S (turquoise), I (bronze) and R (blue). Notice how the mean-field approximation captures the complex behaviour of a wave of infection, spreading through the domain from the bottom left corner.
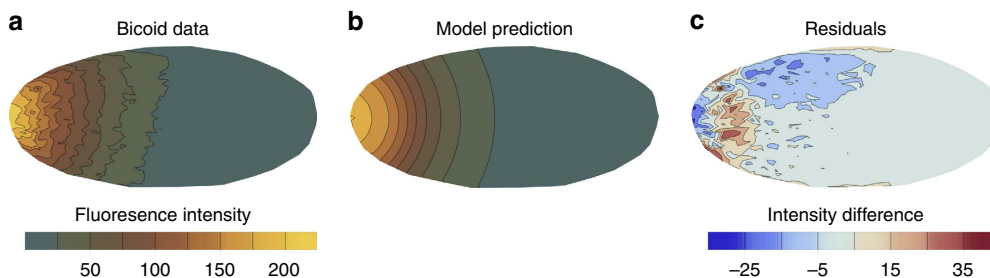
Figure 4 | Results for the *Drosophila* embryo bicoid data. (**a**) Measurement data of bicoid fluorescence intensity across a single embryo. (**b**) Corresponding prediction of our point process model. (**c**) Difference of the experimental data and point process prediction. We observe the point process prediction agrees well with the experimental data. The point process prediction is obtained by solving equation (4) numerically for the inferred parameter values maximizing the data likelihood.

Parameter estimation for *Drosophila* embryo data. Finally, we apply our method to the real gene expression data for the bicoid protein at cleavage stage 13 in the *Drosophila* embryo. The data for 17 embryos can be obtained from the FlyEx database[36]. The data consists of fluorescence intensity measurements on a spatial grid and is shown for one embryo in Fig. 4a. The protein becomes expressed in some region at the left end of the embryo, and then diffuses across the embryo and decays. The system is typically modelled by a linear birth–death process[25,26], and we assume the protein to be expressed within a certain distance r from the left end of the embryo. At cleavage stage 13, the system is supposed to be in steady state and we can perform inference using Result 1 and equation (4). For simplicity, we project the data to one dimension (see Supplementary Methods for more details).

The system has four parameters: the creation range r, the diffusion rate d, the production rate c_1 and decay rate c_2 of the bicoid protein. For steady-state data not all parameters, but only certain ratios are identifiable. We thus infer the creation range r, the diffusion rate d and the ratio $c = c_1/c_2$. For the average of the inferred parameters and their s.d.'s (shown here in parentheses) across the ensemble of embryos we obtain

$$r = 0.26(0.09), \quad d = 0.023(0.005), \quad c = 1.3(0.2) \times 10^4, \quad (11)$$

with s.d.'s of ~ 20–30%. We find that these results do not change significantly under variations of the initial parameter values used in the likelihood optimiser.

Figure 4 illustrates the inference result for one embryo. Figure 4a,b shows the bicoid density across the whole embryo for experimental data and the PR prediction, respectively. We observe good agreement between the measurement data and the point process approximation. Figure 4c shows a plot of the model residuals (difference between model predictions and real data); as can be seen, these are generally comparatively small. As could be expected, the larger errors are concentrated around the steeply changing gradient region between the anterior segments and the main body of the embryo.

Model selection for an SIRS model. Next, we use Result 2 to perform model selection. Specifically, we use our method to decide which of two given microscopic models is more likely to be the true model underlying some given data set. To this end, we use the Bayesian information criterion (BIC)[37]. The BIC for a model is the negative log-likelihood penalized by a term depending on the number of inferred parameters and number
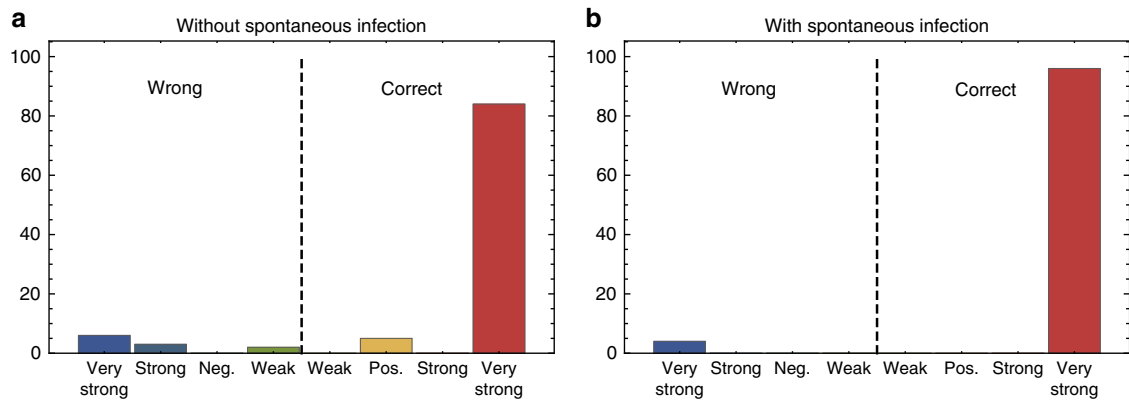
Figure 5 | Model selection results for SIRS system. We use the Bayesian information criterion (BIC) for model selection for the SIRS system, with reactions in equation (10) and the additional spontaneous infection reaction in equation (12). The sign of the difference in the BIC numbers of the two models determines if the correct model is selected, and the corresponding magnitude how confident this choice is. We simulated 20 experiments for 5 parameter sets each. The figures show the combined results of these 100 experiments. (**a**) The true system used to generate the data does not include spontaneous infection. The parameter sets are the same as in Table 3. (**b**) The true system used to generate the data does include spontaneous infection. The parameter sets are the same as in Table 3, but with modified bimolecular infection rate that we set to $k = 10$, 5, 10, 100 and 10, respectively. The spontaneous infection rate is set to $v = 0.002$ for all parameter sets. In both cases, we observe that our method selects the correct model in $>80\%$ of the cases, and in most of these cases with 'very strong' confidence. This demonstrates the strong performance of our method for this model selection problem.

of measurement points. The model with the lower BIC is then chosen to be the true model.

As an example, we modified the SIRS model of equation (10) by including the possibility of a spontaneous, spatially homogeneous infection of susceptible agents, according to the reaction

$$S \xrightarrow{v} I. \tag{12}$$
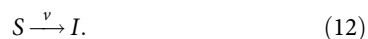
We consider two scenarios: the true microscopic model used to generate the data does or does not contain the spontaneous infection reaction. In either case, we use our method to select the true model. To this end, we optimize the likelihood with respect to the parameters using our method for both models, and compare the corresponding BICs. Figure 5a,b shows the results for the two scenarios of true model without and with spontaneous infection, respectively. The figures show how often our method selected the right or wrong model, and with which confidence level. Each of the figures shows the combined results for 5 different parameter sets and 20 independent simulations for each parameter set. We find that our method chose the correct model in the vast majority of the cases (89% where the true system does not include spontaneous infection and 96% where the true system does contain spontaneous infection). Moreover, our method chooses the correct model with a 'very strong' confidence in most of the cases. This shows that our method is well suited for the problem of model selection. The effectiveness of our model selection approach is remarkable, since the two mechanisms (spontaneous infection and contact infection) can lead to identifiability problems. Such problems are particularly acute when spatial heterogeneities even out rapidly, as in the case of fast diffusion: the few mistakes that our model selection approach makes are primarily due to random samples of the SRDP when the infection spreads particularly fast, so that, for most time points, the process is effectively equilibrated.

Discussion

We considered two popular classes of models for studying stochasticity in spatio-temporal systems; SRDPs and spatio-temporal point processes. The two classes of models are both commonly used in many disciplines, such as epidemiology[14,38]

and social sciences[39]; however, they are widely perceived as conceptually distinct. SRDPs are microscopic, mechanistic descriptions used to predict the dynamics of spatially interacting particles, whereas point processes are typically used empirically to perform inference tasks for systems for which no microscopic description exists. The two approaches therefore seem to be orthogonal to each other.

However, in this paper, we have shown that the two methods are intimately related. By using the PR we established a Cox process representation of SRDPs, which is exact for certain classes of systems and approximate for others. This novel representation enables us to apply a wide range of statistical inference methods to SRDPs, which has not been possible so far. We applied the developed method to several example systems from systems biology and epidemiology, and obtained remarkably accurate results.

Since our method agrees with a deterministic rate equation description on the mean level, bimolecular reactions may lead to deviations from the true mean, which is known to be the case in some non-spatial scenarios[40]. Since in our method distributions are given as real Poisson mixtures, sub-Poissonian fluctuations cannot be captured. However, Gardiner showed that fluctuations of SRDPs are dominated by diffusion on small length scales and therefore Poissonian[33], which may explain the accuracy of our method.

Most inference methods in the literature for SRDPs are either based on Brownian dynamics simulations or stochastic simulations of spatially discretized systems, using the RDME. Both approaches are computationally extremely expensive, and quickly become unfeasible for larger systems and in particular for inference purposes. In contrast, our method relies on the solution of (S)PDEs for which a rich literature of efficient numerical methods exists. For the studied example systems, our method turned out to be highly efficient: the computational time for inferring four unknown parameters for the SIRS system, for example, was found to be of the order of 10 s on a 3.1 GHz processor. We therefore expect our method to be applicable to significantly larger systems, containing more species and unknown parameters. Remarkably, simulating a single realization of the SIRS system from Brownian dynamics simulations took about an order of magnitude longer than the whole inference

procedure, using our method, that is, optimizing the likelihood with respect to the parameters, indicating the immense computational costs of inference methods based on such simulations.

Having access to a likelihood function also provides a major advantage in handling model uncertainty: our results on a spatial SIRS model show that likelihood-based criteria can efficiently and accurately discriminate between competing models. This success raises the possibility that our approach could lay the foundations for structure learning of spatio-temporal stochastic systems: leveraging spatially resolved data not only to identify parameters, but to learn directly the mechanisms underlying the data. The availability of a likelihood approximation makes it in principle possible to borrow techniques from fields, where structure learning is more established and where efficient network learning algorithms based on regularized regression or random forests are routinely used, such as learning gene regulatory networks[41,42].

Our approach can also readily handle spatial heterogeneity in the reaction or diffusion rates: the gene regulation example showed that the method can precisely identify simple geometric features of the system, such as the radius of the cell nucleus. While our examples are primarily illustrative of the methodology, and hence simple, it would be in principle straightforward to generalize the approach to SRDPs defined on complex geometries, such as the intracellular landscapes revealed by X-ray tomography[43]. Learning complex geometries directly from the data would potentially be more challenging, however, as it would generally require learning a large number of parameters.

While we believe that the derived representation of SRDPs in terms of Cox process brings clear advantages from a statistical point of view, it is also important to acknowledge the limitations implied by the employed mean-field approximations. Perhaps, the most important step in our approximation is the mean-field treatment of multi-time joint distributions in Result 2. As noticed before, this replaces direct dependencies between particle locations at different time points, with indirect dependences through the intensity field. This implies that self-excitatory behaviours, such as clustering, cannot be accurately captured; at best, these will be mimicked by a local increase in the intensity field within a Cox process framework. More complex point processes that can account for self-excitatory behaviour do exist[44]; in our opinion, it is a question of considerable theoretical interest whether such processes can also arise from a dynamical SRDP representation.

Methods

The chemical master equation. Consider a system of N species X_i, $i = 1,...,N$ that interact stochastically via R reaction channels

$$\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{k_j} \sum_{i=1}^{N} r_{ij} X_i, \quad j = 1, \dots, R, \quad (13)$$
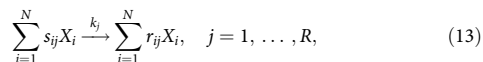
where k_j is the rate constant of the jth reaction, and the s_{ij} and r_{ij} are non-negative integer numbers. Define the stoichiometric matrix $\mathbf{S}$ as $S_{ij} = r_{ij} - s_{ij}$; the jth reaction is of order m if $\sum_{i=1}^{N} s_{ij} = m$. We only consider reactions satisfying $\sum_{i=1}^{N} s_{ij}, \sum_{i=1}^{N} r_{ij} \leq 2$, that is, reactions with a maximum of two reactant and a maximum of two product particles, since higher-order reactions rarely occur in nature. Denote as $\mathbf{n} = (n_1,...,n_N)$ the state of the system, where n_i is the copy number of species X_i. Under well-mixed and dilute conditions, the time evolution of the (single time) marginal probability distribution of the system obeys the chemical master equation[30]

$$\partial_t P(\mathbf{n}, t) = \sum_{r=1}^{R} f_r(\mathbf{n} - \mathbf{S}_r) P(\mathbf{n} - \mathbf{S}_r, t) - \sum_{r=1}^{R} f_r(\mathbf{n}) P(\mathbf{n}, t), \quad (14)$$

where $\mathbf{S}_r$ is the rth column of the stoichiometric matrix $\mathbf{S}$. The propensity function $f_r(\mathbf{n})dt$ gives the probability for the rth reaction to happen in an infinitesimal time

interval dt and is given by

$$f_r(\mathbf{n}) = k_r \Omega \prod_{k=1}^{N} \frac{n_k!}{(n_k - s_{kr})! \Omega^{s_{kr}}}. \quad (15)$$

Here, Ω is the volume of the system.

The Poisson representation. The PR makes the ansatz to write $P(\mathbf{n}, t)$ as a Poisson mixture[27]

$$P(\mathbf{n}, t) = \int d\mathbf{u} \, \mathcal{P}(n_1; u_1) \dots \mathcal{P}(n_N; u_N) p(\mathbf{u}, t), \quad u_i \in \mathbb{C}, \quad (16)$$

where $\mathbf{u} = (u_1,...,u_N)$ and $\mathcal{P}(n_i; u_i) = (e^{-u_i} u_i^{n_i})/n_i!$ is a Poisson distribution in n_i with mean u_i, and the u_i are complex in general. The integrals in equation (16) in general run over the whole complex plane for each u_i. Using the ansatz (equation 16) in the generating function equation that can be derived from equation (14), one can derive the following PDE for the distribution $p(\mathbf{u},t)$[27],

$$\partial_t p(\mathbf{u}, t) = \sum_{r=1}^{R} \Omega k_r \left(\prod_{i=1}^{N} \left(1 - \frac{\partial}{\partial u_i} \right)^{r_{ir}} - \prod_{i=1}^{N} \left(1 - \frac{\partial}{\partial u_i} \right)^{s_{ir}} \right)$$
$$\times \prod_{j=1}^{N} \Omega^{-s_{jr}} u_j^{s_{jr}} p(\mathbf{u}, t). \quad (17)$$

Note that this PDE generally involves derivatives of higher order than two, which means that $p(\mathbf{u},t)$ can generally become negative, in which case it does not admit a probabilistic interpretation. However, since we only consider reactions satisfying $\sum_{i=1}^{N} s_{ir}, \sum_{i=1}^{N} r_{ir} \leq 2$, equation (17) simplifies to

$$\partial_t p(\mathbf{u}, t) = - \sum_{i=1}^{N} \frac{\partial}{\partial u_i} [A_i(\mathbf{u}) p(\mathbf{u}, t)] + \frac{1}{2} \sum_{i,j=1}^{N} \frac{\partial}{\partial u_i} \frac{\partial}{\partial u_j} [B_{ij}(\mathbf{u}) p(\mathbf{u}, t)], \quad (18)$$

which is a Fokker–Planck equation with drift vector $\mathbf{A}(\mathbf{u})$ and diffusion matrix $\mathbf{B}(\mathbf{u})$ given by

$$A_i(\mathbf{u}) = \sum_{r=1}^{R} S_{ir} g_r(\mathbf{u}), \quad (19)$$

$$B_{ij}(\mathbf{u}) = \sum_{r=1}^{R} g_r(\mathbf{u}) (r_{ir} r_{jr} - s_{ir} s_{jr} - \delta_{i,j} S_{ir}), \quad (20)$$

$$g_r(\mathbf{u}) = \Omega k_r \prod_{j=1}^{N} \Omega^{-s_{jr}} u_j^{s_{jr}}, \quad (21)$$

where $\delta_{i,j}$ denotes the Kronecker delta. The corresponding Langevin equation reads

$$d\mathbf{u} = \mathbf{A}(\mathbf{u})dt + \mathbf{C}(\mathbf{u})d\mathbf{W}, \quad \mathbf{C}\mathbf{C}^T = \mathbf{B}, \quad (22)$$

where $\mathbf{W}$ is a l-dimensional Wiener process and l is the number of columns of $\mathbf{C}$.

Depending on the reactions in the system, the diffusion matrix may be zero, in which case the Langevin equation (equation 22) reduces to deterministic ordinary differential equations. On the other hand, depending on the reactions, $\mathbf{B}(\mathbf{u})$ is not positive semidefinite and thus $\mathbf{C}\mathbf{C}^T = \mathbf{B}$ cannot be fulfilled for real $\mathbf{C}$, which means that equation (18) is not a proper Fokker–Planck equation in real variables. Rather, it needs to be extended to complex space. Specifically, this is the case whenever the system contains bimolecular reactions or reactions with two non-identical product molecules.

An important property of the PR is that the mean values of the particle numbers n_i are equal to the mean values of the corresponding PR variables u_i, that is, $\langle n_i \rangle = \langle u_i \rangle$.

The reaction–diffusion master equation. Consider a system as in equation (13), but in an M-dimensional volume discretized into L cubic compartments of edge length h and volume h^M. Denote as $\mathbf{n} = (n_1^1,...,n_N^1,...,n_1^L,...,n_N^L)$ the state of the system, where n_i^l is the copy number of species X_i in the lth compartment. Under well-mixed and dilute conditions in each compartment, the reaction dynamics in each compartment is governed by a corresponding chemical master equation as in equation (14). If we model diffusion of species X_i between neighbouring compartments by linear reactions with rate constant $d_i = D_i/h^2$, where D_i is the microscopic diffusion constant of species X_i, the time evolution of the (single time) marginal probability distribution of the system obeys the RDME[30]:

$$\partial_t P(\mathbf{n}, t) = \sum_{l=1}^{L} \sum_{m \in \mathcal{N}(l)} \sum_{i=1}^{N} d_i \left[(n_i^m + 1) P(\mathbf{n} + \boldsymbol{\delta}_i^m - \boldsymbol{\delta}_i^l, t) - n_i^l P(\mathbf{n}, t) \right] \quad (23)$$

$$+ \sum_{l=1}^{L} \sum_{r=1}^{R} [f_r(\mathbf{n}^l - \mathbf{S}_r) P(\mathbf{n} - \mathbf{S}_r^l, t) - f_r(\mathbf{n}^l) P(\mathbf{n}, t)], \quad (24)$$

where $f_r(\mathbf{n}^l)$ is the propensity function of the rth reaction evaluated at the state vector $\mathbf{n}^l = (n_1^l, \dots, n_N^l)$ of the lth compartment, $\boldsymbol{\delta}_1^l$ is a vector of length $N \times L$

with the entry corresponding to species X_i in the lth compartment equal to 1 and all other entries zero, and $\mathbf{S}_r^l$ is a vector of length $N \times L$ with the entries corresponding to the lth compartment equal to the rth row of the stoichiometric matrix $\mathbf{S}$ and zero otherwise.

Real-valued Poisson representation in space. We next apply the PR to the RDME in equations (23) and (24) after applying the mean-field approximations defined in the Results section to bimolecular reactions and reactions with two non-identical product molecules, and subsequently take the continuum limit. Consider first the diffusion term in equation (23). Since different species do not interact with each other if there are no chemical reactions happening, we can consider a system containing only a single species, say species X_1, for which equation (23) reduces to

$$\partial_t P(\mathbf{n}, t) = \sum_{l=1}^{L} \sum_{m \in \mathcal{N}(l)} d\left[(n^m + 1)P(\mathbf{n} + \boldsymbol{\delta}^m - \boldsymbol{\delta}^l, t) - n^l P(\mathbf{n}, t)\right], \quad (25)$$

where $\mathbf{n} = (n^1, \dots, n^L)$, n^m is the number of X_1 particles in the mth compartment, $\boldsymbol{\delta}^m$ is a vector with a one in the mth entry and zero otherwise, d is the diffusion constant of species X_1 and the sum over m runs over all neighbouring compartments $\mathcal{N}(l)$ of the lth compartment. For this system, the PR is real and deterministic, and we use the PR without any approximations. The corresponding Langevin equation reads

$$du^l = D\frac{2Mu^l - \sum\limits_{m \in \mathcal{N}(l)} u^m}{h^2} dt, \quad l = 1, \dots, L, \quad (26)$$

where M is the spatial dimension of the system and $D = dh^2$ the microscopic diffusion constant. Since the sum over m runs over all adjacent compartments of the lth compartment, the fraction in equation (26) is just the discretized version of the Laplace operator $\Delta = \partial_1^2 + \dots + \partial_M^2$. Introducing a discretized density field $u(x^l) = u^l/h^M$, where x^l is the centre of the lth compartment, and taking the continuum limit of equation (26), we get the PDE

$$du(x, t) = D\Delta u(x, t)dt, \quad (27)$$

which is just the diffusion equation for the field $u(x,t)$.

Consider next the reaction term of the RDME given in equation (24). Since reactions only occur within compartments, we can treat the compartments independently of each other. For a single compartment, equation (24) then reduces to the chemical master equation given in equation (14). Here, however, we first apply the approximations defined in the Results section to bimolecular reactions and reactions with two non-identical product molecules (see Supplementary Methods for more details). These approximations lead to a real-valued PR and only reactions with two identical product molecules lead to stochastic terms. The PR Langevin equation hence simplifies to

$$du_i = \sum_{r=1}^{R} S_{ir} g_r(\mathbf{u})dt + \sum_{r'} \sqrt{2g_{r'}(\mathbf{u})}dW_{r'}, \quad (28)$$

where $\mathbf{u} = (u_1, \dots, u_N)$ and the sum over r' runs over all reactions with two product particles of species X_i. The propensities $g_r(\mathbf{u})$ are obtained by replacing the n_i variables with u_i variables and Ω with h^M in the expressions for the f_r propensities of the approximated reactions. The factor of two in the square root in equation (28) comes from the fact that two identical molecules become produced in these reactions. Reintroducing the label l denoting the compartment number in equation (28), and the species label i in equation (27), we can add the two contributions to obtain

$$du_i^l = D_i \frac{2Mu_i^l - \sum\limits_{m \in \mathcal{N}(l)} u_i^m}{h^2} dt + \sum_{r=1}^{R} S_{ir} g_r(\mathbf{u}^l)dt$$
$$+ \sum_{r'} \sqrt{2g_{r'}(\mathbf{u}^l)}dW_{r'}^l, \quad (29)$$

where $\mathbf{u} = (u_1^1, \dots, u_N^1, \dots, u_1^L, \dots, u_N^L)$ and u_i^l is the PR variable of species X_i in the lth compartment. If we again define discretized density fields $u_i(x^l) = u_i^l/h^M$, where x^l is the centre of the lth compartment, and $dW_r(x^l) = dW_r^l/\sqrt{h^M}$, we can take the continuum limit of equation (29) that leads to the real-valued SPDE for the intensity fields given in equation (4). The $g_r(\mathbf{u}(x,t))$ therein are not functions of single PR variables anymore, but rather functionals of the space-dependent intensity field vector $\mathbf{u}(x,t) = (u_1(x,t), \dots, u_N(x,t))$. They are obtained by taking the corresponding propensity functions $f_r(\mathbf{n})$ of the approximate reactions in real space, replacing $n_i \to u_i(x,t)$ and $\langle n_i \rangle \to \langle u_i(x,t) \rangle$, and omitting Ω factors. The latter can be identified with h^M here, and hence get absorbed in the definition of the intensity fields given below equation (29).

As an example, consider the reaction $A + B \to \varnothing$. The non-spatial propensity in real space for this reaction is $f(n_A, n_B) = k n_A n_B / \Omega$. However, since this is a bimolecular reaction and hence leads to a complex-valued PR, we replace it by two reactions $A \to \varnothing$ and $B \to \varnothing$, with propensities $f_1(n_A, n_B) = k\langle n_B \rangle n_A / \Omega$ and $f_2(n_A, n_B) = k\langle n_A \rangle n_B / \Omega$, respectively. By replacing $n_i \to u_i(x,t)$ and $\langle n_i \rangle \to \langle u_i(x,t) \rangle$, and omitting Ω terms, we thus obtain the corresponding propensity

functions in spatial PR space as $g_1(u_A(x,t), u_B(x,t)) = k\langle u_B(x,t) \rangle u_A(x,t)$ and $g_2(u_A(x,t), u_B(x,t)) = k\langle u_A(x,t) \rangle u_B(x,t)$, respectively. See Supplementary Methods for more details and examples.

Poisson and Cox processes. A (spatial) Poisson process on a spatial region $\mathcal{D}$ of arbitrary dimension defines a measure on countable unions of zero-dimensional subsets (points) of $\mathcal{D}$. A Poisson process is often characterized by an intensity function $u : \mathcal{D} \to \mathbb{R}^+$, giving the probability density of finding a point in an infinitesimal region around x. Now, let $N(A)$ denote the number of points in a subregion $A \subset \mathcal{D}$. Then $N(A)$ is a Poisson random variable with mean given by the integral of $u(\cdot)$ over A:

$$p(N(A) = n) = \mathcal{P}(n; u_A), \quad u_A = \int_A dx\, u(x), \quad (30)$$

where $\mathcal{P}(n; u_A)$ is a Poisson distribution in n with mean u_A.

A (spatial) Cox process is a generalization of a Poisson process and also called 'doubly stochastic process', in the sense that the intensity function u is itself a random process. Conditioned on the intensity u, the Cox process reduces to a Poisson process. The distribution of the number of points in a subregion $A \subset \mathcal{D}$ is hence a mixture of Poisson distributions,

$$p(N(A) = n) = \int du_A \mathcal{P}(n; u_A) p(u_A). \quad (31)$$

Since we are interested in dynamical systems, we will assume time-dependent intensities $u : \mathcal{D} \times \mathcal{T} \to \mathbb{R}^+$, where $\mathcal{T}$ is a finite real interval denoting time. We then require that for any fixed time point $t \in \mathcal{T}$, the process is a spatial Poisson (Cox) process with intensity $u(\cdot, t)$. In the case of a Poisson (Cox) process, the intensity u may, for example, be defined as the solution of a PDE (SPDE).

References

1. Bullara, D. & De Decker, Y. Pigment cell movement is not required for generation of Turing patterns in zebrafish skin. *Nat. Commun.* **6,** 6971 (2015).
2. Metzler, R. The future is noisy: the role of spatial fluctuations in genetic switching. *Phys. Rev. Lett.* **87,** 068103 (2001).
3. Elf, J. & Ehrenberg, M. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst. Biol.* **1,** 230–236 (2004).
4. Takahashi, K., Tanase-Nicola, S. & Ten Wolde, P. R. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proc. Natl Acad. Sci. USA* **107,** 2473–2478 (2010).
5. Short, M. B., Brantingham, P. J., Bertozzi, A. L. & Tita, G. E. Dissipation and displacement of hotspots in reaction-diffusion models of crime. *Proc. Natl Acad. Sci. USA* **107,** 3961–3965 (2010).
6. Mahmutovic, A., Fange, D., Berg, O. G. & Elf, J. Lost in presumption: stochastic reactions in spatial models. *Nat. Methods* **9,** 1163–1166 (2012).
7. Yates, C. A. *et al.* Inherent noise can facilitate coherence in collective swarm motion. *Proc. Natl Acad. Sci. USA* **106,** 5464–5469 (2009).
8. Cottrell, D., Swain, P. S. & Tupper, P. F. Stochastic branching-diffusion models for gene expression. *Proc. Natl Acad. Sci. USA* **109,** 9699–9704 (2012).
9. Doi, M. Second quantization representation for classical many-particle system. *J. Phys. A* **9,** 1465 (1976).
10. Doi, M. Stochastic theory of diffusion-controlled reaction. *J. Phys. A* **9,** 1479 (1976).
11. Bicknell, B. A., Dayan, P. & Goodhill, G. J. The limits of chemosensation vary across dimensions. *Nat. Commun.* **6,** 7468 (2015).
12. Grima, R. & Schnell, S. A systematic investigation of the rate laws valid in intracellular environments. *Biophys. Chem.* **124,** 1–10 (2006).
13. Holmes, G. R. *et al.* Repelled from the wound, or randomly dispersed? Reverse migration behaviour of neutrophils characterized by dynamic modelling. *J. R. Soc. Interface* **9,** 3229–3239 (2012).
14. Davies, T. P., Fry, H. M., Wilson, A. G. & Bishop, S. R. A mathematical model of the London riots and their policing. *Sci. Rep.* **3,** 1303 (2013).
15. Gardiner, C. W., McNeil, K. J., Walls, D. F. & Matheson, I. S. Correlations in stochastic theories of chemical reactions. *J. Stat. Phys.* **14,** 307–331 (1976).
16. Fange, D., Berg, O. G., Sjöberg, P. & Elf, J. Stochastic reaction-diffusion kinetics in the microscopic limit. *Proc. Natl Acad. Sci. USA* **107,** 19820–19825 (2010).
17. Isaacson, S. A. Relationship between the reaction-diffusion master equation and particle tracking models. *J. Phys. A* **41,** 065003 (2008).
18. van Zon, J. S. & Ten Wolde, P. R. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.* **94,** 128103 (2005).

19. Isaacson, S. A. & Peskin, C. S. Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. *SIAM J. Sci. Comput.* **28,** 47–74 (2006).

20. Erban, R. & Chapman, S. J. Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.* **6,** 046001 (2009).

21. Drawert, B., Lawson, M. J., Petzold, L. & Khammash, M. The diffusive finite state projection algorithm for efficient simulation of the stochastic reaction-diffusion master equation. *J. Chem. Phys.* **132,** 074101 (2010).

22. Ferm, L., Hellander, A. & Lötstedt, P. An adaptive algorithm for simulation of stochastic reaction-diffusion processes. *J. Comput. Phys.* **229,** 343–360 (2010).

23. Franz, B., Flegg, M. B., Chapman, S. J. & Erban, R. Multiscale reaction-diffusion algorithms: PDE-assisted Brownian dynamics. *SIAM J. Appl. Math.* **73,** 1224–1247 (2013).

24. Fu, J., Wu, S., Li, H. & Petzold, L. R. The time dependent propensity function for acceleration of spatial stochastic simulation of reaction-diffusion systems. *J. Comput. Phys.* **274,** 524–549 (2014).

25. Dewar, M. A., Kadirkamanathan, V., Opper, M. & Sanguinetti, G. Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in D. melanogaster. *BMC Syst. Biol.* **4,** 21 (2010).

26. Ruttor, A. & Opper, M. Approximate parameter inference in a stochastic reaction-diffusion model. *AISTATS* **9,** 669–676 (2010).

27. Gardiner, C. W. & Chaturvedi, S. The Poisson representation. I. A new technique for chemical master equations. *J. Stat. Phys.* **17,** 429–468 (1977).

28. Cressie, N. A. C. & Wikle, C. K. *Statistics for Spatio-temporal Data* (Wiley, 2011).

29. Cseke, B., Zammit Mangion, A., Heskes, T. & Sanguinetti, G. Sparse approximate inference for spatio-temporal point process models. *JASA* (2015) (in the press).

30. Gillespie, D. T., Hellander, A. & Petzold, L. R. Perspective: stochastic algorithms for chemical kinetics. *J. Chem. Phys.* **138,** 170901 (2013).

31. Thomas, P., Straube, A. V. & Grima, R. Stochastic theory of large-scale enzyme-reaction networks: Finite copy number corrections to rate equation models. *J. Chem. Phys.* **133,** 195101 (2010).

32. Kingman, J. F. C. *Poisson Processes* (Oxford Univ. Press, 1992).

33. Gardiner, C. W. *Handbook of Stochastic Methods* (Springer, 1985).

34. Peruani, F. & Lee, C. F. Fluctuations and the role of collision duration in reaction-diffusion systems. *EPL* **102,** 58001 (2013).

35. Abdullah, M., Cooper, C. & Draief, M. in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011* 351–364 (Springer, 2011).

36. Pisarev, A., Poustelnikova, E., Samsonova, M. & Reinitz, J. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Res.* **37,** D560–D566 (2009).

37. Schwarz, G. E. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).

38. Grell, K. et al. A three-dimensional point process model for the spatial distribution of disease occurrence in relation to an exposure source. *Stat. Med.* **34,** 3170–3180 (2015).

39. Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V. & Sanguinetti, G. Point process modelling of the Afghan War Diary. *Proc. Natl Acad. Sci. USA* **109,** 12414–12419 (2012).

40. Ramaswamy, R., Gonzalez-Segredo, N., Sbalzarini, I. F. & Grima, R. Discreteness-induced concentration inversion in mesoscopic chemical systems. *Nat. Commun.* **3,** 779 (2012).

41. Bonneau, R. et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* **7,** R36 (2006).

42. Huynh-Thu, V.-A. & Sanguinetti, G. Combining tree-based and dynamical systems for the inference of gene regulatory networks. *Bioinformatics* **31,** 1614–1622 (2015).

43. Do, M., Isaacson, S. A., McDermott, G., Le Gros, M. A. & Larabell, C. A. Imaging and Characterizing Cells using Tomography. *Arch. Biochem. Biophys.* **581,** 111–121 (2015).

44. Hawkes, A. G. & Oakes, D. A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11,** 493–503 (1974).

Acknowledgements

Author contributions

All authors designed the research. D.S. implemented the research. D.S. and G.S. wrote the paper with input from R.G.

Additional information