

SCIENTIFIC REPORTS



OPEN

Computationally Efficient Modelling of Stochastic Spatio-Temporal Dynamics in Biomolecular Networks

Jongrae Kim¹, Mathias Foo² & Declan G. Bates²

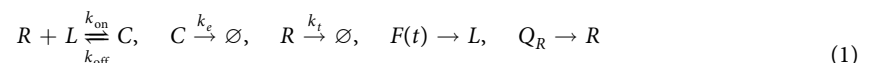
Measurement techniques in biology are now able to provide data on the trajectories of multiple individual molecules simultaneously, motivating the development of techniques for the stochastic spatio-temporal modelling of biomolecular networks. However, standard approaches based on solving stochastic reaction-diffusion equations are computationally intractable for large-scale networks. We present a novel method for modeling stochastic and spatial dynamics in biomolecular networks using a simple form of the Langevin equation with noisy kinetic constants. Spatial heterogeneity in molecular interactions is decoupled into a set of compartments, where the distribution of molecules in each compartment is idealised as being uniform. The reactions in the network are then modelled by Langevin equations with correcting terms, that account for differences between spatially uniform and spatially non-uniform distributions, and that can be readily estimated from available experimental data. The accuracy and extreme computational efficiency of the approach is demonstrated on a model of the epidermal growth factor receptor network in the human mammary epithelial cell.

Measurement techniques in experimental biology have now advanced to a stage where several molecular species can be tracked at the same time^{1,2}. The availability of such data strongly motivates the development of scalable methods for the spatio-temporal modelling of biomolecular networks³. It is well known that some characteristics of biomolecular interaction networks are strongly influenced by stochastic noise effects⁴⁻⁶. Although the key assumption underlying standard stochastic modelling approaches is spatial homogeneity, molecular interactions in cells are in reality highly spatially heterogeneous^{7,8} and also directionally heterogeneous^{9,10}. Current modelling approaches that include both stochastic and spatio-temporal dynamics lead to stochastic reaction-diffusion equations, whose solutions are extremely computationally intensive¹¹. This places strong constraints on the size of biomolecular networks that can be modelled using such approaches.

Efficient modelling approaches to take into account directionalities of biomolecules are considered in⁹ and¹⁰. In this paper, we present a novel method, based on a simple form of the Langevin equation with noisy kinetic constants, to model spatial distribution effects of biomolecular interactions. Spatial heterogeneity in molecular interactions is decoupled into a set of compartments, where the distribution of molecules in each compartment is uniform. Since obtaining such a set of compartments would be challenging in practice, we model the reactions in the network using Langevin equations with correcting terms that account for differences between spatially uniform and spatially non-uniform distributions. We demonstrate the accuracy and computational efficiency of the approach using a model of the epidermal growth factor receptor network in the human mammary epithelial cell.

Results

Langevin equation with spatial heterogeneity. Consider the generic model of a ligand-receptor interaction network in a well-mixed solution given by the following five reactions:



¹School of Mechanical Engineering, University of Leeds, Leeds, LS2 9JT, UK. ²Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, CV4 7AL, UK. Correspondence and requests for materials should be addressed to J.K. (email: menkim@leeds.ac.uk) or D.G.B. (email: D.Bates@warwick.ac.uk)

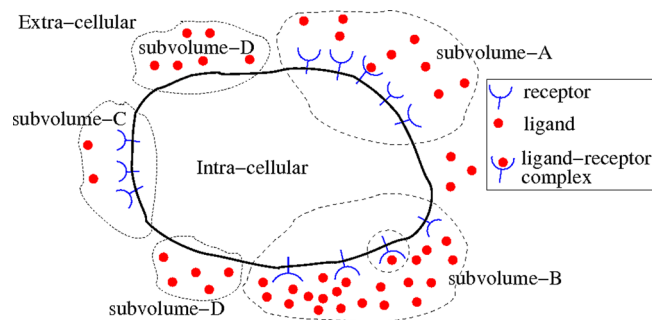


Figure 1. Schematic of receptor-ligand distributions within the cell. The uneven distribution of ligands and receptors can be considered as the sum of multiple subvolumes, where the distributions of each subvolume are approximately uniform. The structure of the subvolumes is not necessarily fixed in time and would change with its own dynamics.

where R , L and C are the numbers of receptors, ligands and ligand-bound-receptor complex molecules, respectively, $F(t)$ is the external source of ligand stimulation, Q_R is the intra-cellular receptor generation rate, and k_i for $i = \{\text{on, off, e, t}\}$ is the kinetic rate of each chemical reaction. k_{on} is the ligand-receptor binding rate, k_{off} is the dissociation rate, and the complex and receptor are internalised with the rates, k_e and k_r , respectively. All of the kinetic rates given above are typically assumed to be constant.

The probability that a ligand-bound complex (C) is produced from the chemical reaction between the receptor (R) and the ligand (L) in some small time interval, dt , is given by $k_{\text{on}}RLdt$. This is called a propensity function and the propensities for the other five reactions are defined similarly. Now, the probability of the number of receptor molecules being reduced by one, $P(dR = -1)$, during the time interval, dt , for the well-mixed condition is

$$P(dR = -1) = k_{\text{on}}RLdt + k_tRdt \quad (2)$$

where dR is the increment in the number of receptors for the time interval, dt . Similarly, the probability of the number of receptor molecules increasing by one, $P(dR = 1)$, for dt is

$$P(dR = 1) = k_{\text{off}}Cdt + Q_Rdt \quad (3)$$

The probability that neither of the above two reactions occurs is

$$P(dR = 0) = [1 - P(dR = -1)] [1 - P(dR = 1)] \approx 1 - k_{\text{on}}RLdt - k_tRdt - k_{\text{off}}Cdt - Q_Rdt \quad (4)$$

where terms involving higher orders of dt are considered negligible.

The well-mixed condition considered above is a strong assumption, since it is known that heterogeneous spatial distributions can have significant impact on the responses of biomolecular networks. As illustrated in Fig. 1, the numbers of both receptor and ligand molecules would vary considerably across different regions of the cell surface, and thus the equations derived in (2), (3) and (4) would potentially have large differences from quantities measured by experiments. Typically, the reaction rates, k_i , would be measured in a well-mixed condition *in vitro*, or calculated from some theoretical derivations, which most likely neglect spatial effects. To model the effects of spatial heterogeneity while keeping mathematical and computational complexity low, we derive a version of the Langevin equation with spatial fluctuations, as follows.

We assume that the heterogeneous space shown in Fig. 1 can be *virtually* divided into several subvolumes within which the well-mixed condition is valid. Then, the probabilities given above are modified as follows:

$$P(dR = -1) = k_{\text{on}}(R_A L_A + R_B L_B + R_C L_C + R_D L_D)dt + k_t Rdt \quad (5)$$

where R_A and L_A are the number of receptors and ligands in the subvolume-A, the numbers in the other subvolumes are defined similarly, and the total numbers are given by

$$R = R_A + R_B + R_C + R_D \quad (6a)$$

$$L = L_A + L_B + L_C + L_D \quad (6b)$$

Note that the kinetic constant, k_{on} , would be measured in a well-mixed condition *in vitro*. To correct (2) to match with (5), we substitute (6) into the RL term of (2), i.e.,

$$\begin{aligned} RL &= (R_A + R_B + R_C + R_D)(L_A + L_B + L_C + L_D) \\ &= (R_A L_A + R_B L_B + R_C L_C + R_D L_D) \\ &\quad + R_A(L_B + L_C + L_D) + R_B(L_A + L_C + L_D) \\ &\quad + R_C(L_A + L_B + L_D) + R_D(L_A + L_B + L_C) \end{aligned} \quad (7)$$

Equation (5) is then rewritten using (7) as follows:

$$P(dR = -1) = k_{\text{on}}(1 - \delta)RL + k_t R dt \quad (8)$$

where the correction term δ is given by

$$\delta = [R_A(L_B + L_C + L_D) + R_B(L_A + L_C + L_D) + R_C(L_A + L_B + L_D) + R_D(L_A + L_B + L_C)] / (RL) \quad (9)$$

By definition, δ is between 0 and 1, where $\delta = 0$ corresponds to the well-mixed condition and $\delta = 1$ corresponds to a complete separation of the molecules, i.e., no reaction occurs. In most biomolecular networks δ would take some value between 0 and 1, and, since the spatial distributions of the molecules might be continuously changing, the value of δ would also change over time.

The proposed modelling approach thus provides a compact parameter, δ , that captures spatial effects on the stochastic dynamics of a biomolecular network, without requiring the introduction of complex reaction diffusion equations. Note that, in principle, unimolecular reactions would not be affected by spatial heterogeneity, and thus Equation (3) remains the same for both the cases of a well mixed and non-uniform distribution. Only the chemical reactions involving two or more molecules will be affected by spatial heterogeneity.

The expected value of dR^2 is calculated as

$$\begin{aligned} E(dR^2) &= (-1)^2 P(dR = -1) + (1)^2 P(dR = 1) \\ &= k_{\text{on}}(1 - \delta)RL dt + k_t R dt + k_{\text{off}} C dt + Q_R dt \end{aligned} \quad (10)$$

and the expected value of dR is given by

$$E(dR) = -k_{\text{on}}(1 - \delta)RL dt - k_t R dt + k_{\text{off}} C dt + Q_R dt \quad (11)$$

where $E(\cdot)$ is the expectation over the sampling space. The variance of dR is obtained by

$$\begin{aligned} E\{[dR - E(dR)]^2\} &= E(dR^2) - E(dR)^2 \approx E(dR^2) \\ &= [k_{\text{on}}(1 - \delta)RL + k_t R + k_{\text{off}} C + Q_R] dt \end{aligned} \quad (12)$$

where all dt^2 terms are negligible. Finally, the Langevin equation is derived by matching the mean and the variance given by (11) and (12), so that

$$dR = [-k_{\text{on}}(1 - \delta)RL - k_t R + k_{\text{off}} C + Q_R] dt + \sqrt{k_{\text{on}}(1 - \delta)RL + k_t R + k_{\text{off}} C + Q_R} dw_R \quad (13)$$

where dw_R is the Brownian motion with variance equal to dt .

The correction term, δ , would have its own stochastic temporal kinetics, for example,

$$d\delta = a_\delta + b_\delta dw_\delta \quad (14)$$

where a_δ and b_δ are unknown functions that determine the temporal evolution of δ , and dw_δ could be the Brownian motion. In some limited cases, the exact or approximate form of the two functions could be modelled. In general, however, we would use experimental data to identify the structures and/or parameters in the functions determining the temporal evolution of δ . For example, if we assume that a_δ and b_δ are constant, then system parameter identification algorithms could be used to determine the constants¹². This process is described in detail in the following example.

Example: Epidermal Growth Factor Receptor (EGFR) network in human cells. A model of the ligand-receptor network kinetics of the human mammary epidermal growth factor receptor (EGFR) in the well-mixed state was developed in¹³, where the following set of ordinary differential equations are derived from the underlying chemical reactions, (1), as follows:

$$\frac{dR}{dt} = -k_{\text{on}}RL + k_{\text{off}}C - k_t R + Q_R \quad (15a)$$

$$\frac{dC}{dt} = k_{\text{on}}RL - k_{\text{off}}C - k_e C \quad (15b)$$

$$\frac{dL}{dt} = -k_{\text{on}}RL + k_{\text{off}}C + F(t) \quad (15c)$$

where $k_{\text{off}} = 0.24$ [1/min], $K_D = 2.47 \times 10^{-9} \times N_{\text{av}} \times V_{\text{cell}}$ [the number of molecules], $N_{\text{av}} = 6.023 \times 10^{23}$, i.e., the Avogadro's number, $V_{\text{cell}} = 4 \times 10^{-10}$ [l], the cell volume, $k_{\text{on}} = k_{\text{off}}/K_D$, $k_t = 0.02$ [1/min], $Q_R = 2 \times 10^5 k_p$ and $k_e = 0.15$ [1/min].

Consider an initial impulse input for the ligand, i.e., $L(0) = 10,000$, assuming the presence of 200,000 receptors and no complexes. $F(t)$ is set to zero, i.e., no external stimulation. The simulation results are shown in Fig. 2. In Fig. 2A,C, stochastic simulations with the well-mixed condition are performed using Gillespie's direct method, which is an exact stochastic simulation algorithm¹⁴. As expected, for the well-mixed condition, the 10 stochastic

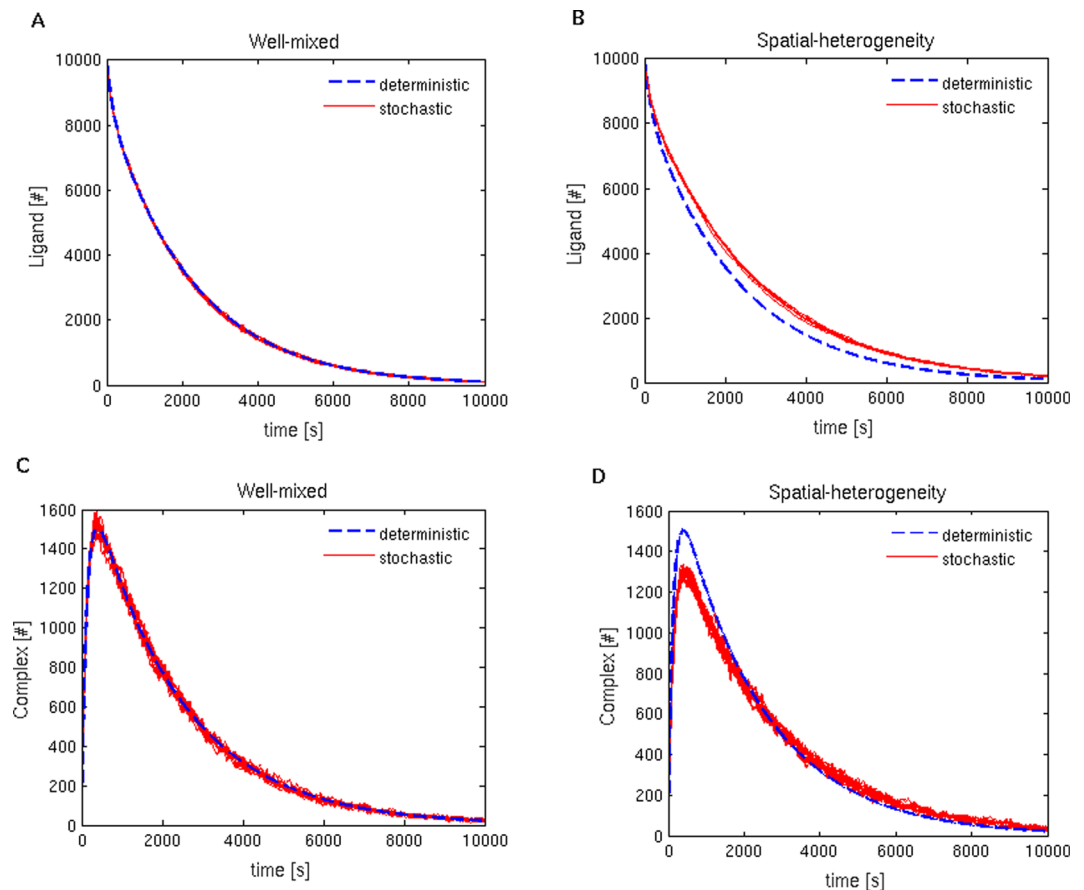


Figure 2. Total number of ligands and complexes in well-mixed and spatially-heterogeneous conditions. For the deterministic case, the time histories of the ligand and complex are obtained by solving the ordinary differential equations corresponding to the molecular interactions. For the well-mixed case, (A,C) 10 realisations of the exact stochastic simulation, using Gillespie's direct method, are shown in red lines. (B,D) 10 realisations of the stochastic reaction-diffusion simulations, shown in red lines, using MesoRD.

simulations of the model produce trajectories that all closely match the deterministic trajectories produced by solving the ordinary differential equations numerically.

However, even for the case where the well-mixed condition is still valid, the simulations including spatial diffusion produce different results from those generated by the ordinary differential equations or Gillespie's direct method, as shown in Fig. 2B,D. Here, the ligand and the receptor are uniformly distributed on the surface of the cell, which is assumed to be a sphere of the radius $100\ \mu\text{m}$. Since Gillespie's algorithm cannot take into account spatial effects, we solve the reaction-diffusion equations using MesoRD (Mesoscopic Reaction-Diffusion Simulator)¹¹, a numerical solver for stochastic reaction diffusion equations using the next subvolume method.

In reality, it is highly probable that there would be significant differences in the time history of molecular number changes for non-uniform distributions compared to the case of a uniform distribution. To investigate this, a spatially non-uniform distribution is simulated in Fig. 3. Here, the ligands, indicated by the red dots in the figure, are distributed uniformly on the cell surface, while the receptors (not shown in the figure for clarity) are distributed non-uniformly on the cell surface. In this example, we make the distribution of the receptors approximately equal to the distribution of the complexes at $t = 30\ \text{s}$ in the figure, i.e., 99% of the total receptors are placed on the right side of the cell surface and 1% are distributed in the remaining area of the cell surface.

The time histories of the ligand and the complex numbers are compared with those calculated by the deterministic simulation in Fig. 4. The dynamic behaviours for both cases are clearly very different, showing that the explicit consideration of spatial heterogeneity is essential in order to correctly capture the dynamics of the molecules in the ligand-receptor network. However, exact simulation algorithms for stochastic reaction diffusion equations are extremely computationally expensive, and hence cannot be applied to larger networks. Moreover, in general, we do not know how the molecules are distributed - in practice we will only be able to measure total concentration changes.

To get around this problem, we use the proposed model based on the Langevin equation with spatial fluctuations, which gives the following stochastic equations:

$$dR = [-k_{\text{on}}(1 - \delta)RL - k_t R + k_{\text{off}}C + Q_R] dt + \sqrt{k_{\text{on}}(1 - \delta)RL + k_t R + k_{\text{off}}C + Q_R} dw_R \quad (16a)$$

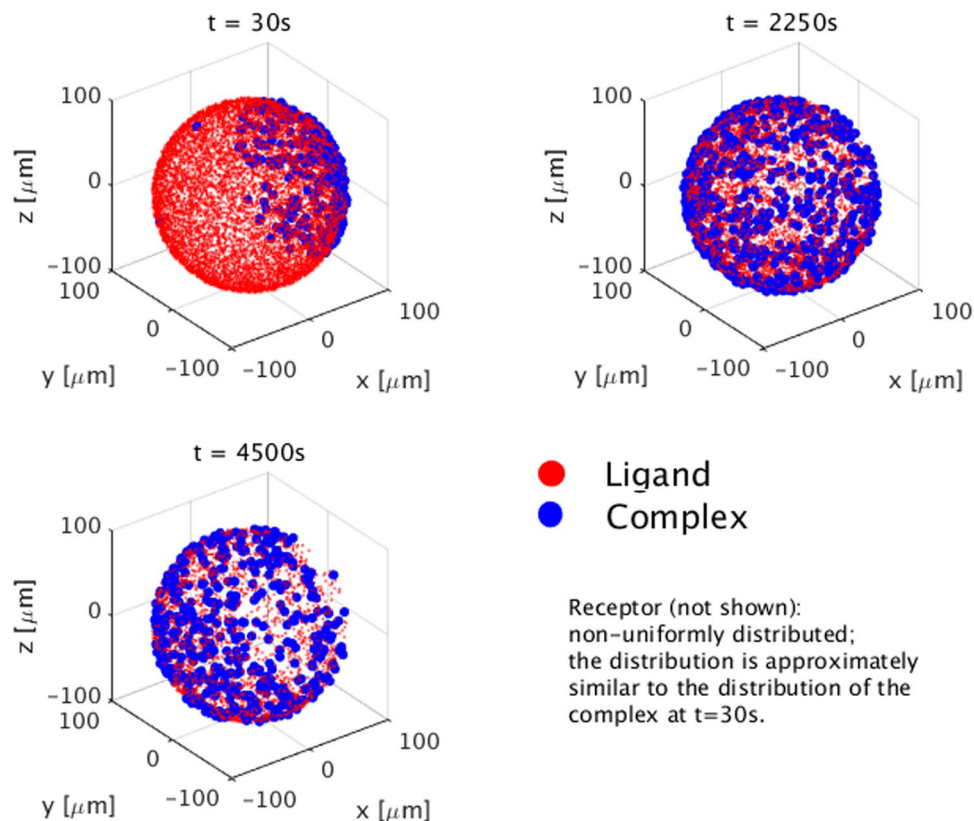


Figure 3. Interaction between non-uniform distribution of receptors and the uniform distribution of ligands.

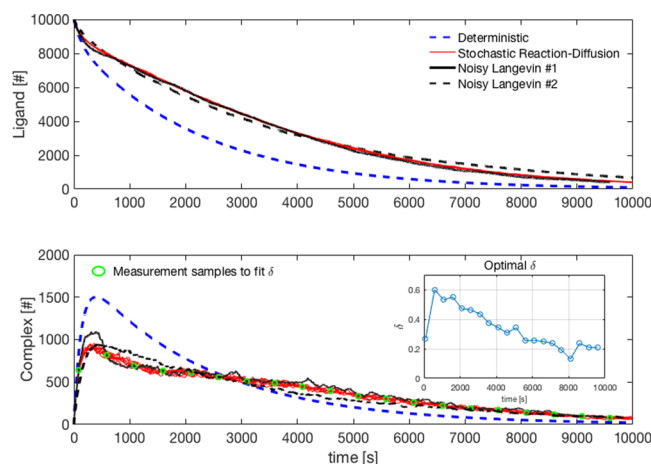


Figure 4. Total number of ligands and complexes in the spatially heterogeneous condition. The time histories of the ligand and the complex obtained from the ordinary differential equations. The 10 stochastic realisations of the reaction-diffusion simulation are carried out using MesoRD and the green circles are the mean values of the MesoRD simulation results (considered as surrogate data points from wet-lab experimental measurements). The new Langevin equation with the optimal δ obtained by solving a minimisation problem provides a highly accurate simulation of the molecular time histories in less than one second (Noisy Langevin #1), compared with 38 minutes to compute one reaction-diffusion simulation using MesoRD. The inset shows the corresponding time history of δ , which is clearly able to capture the effects of spatial heterogeneity. Less accurate but still reasonable matched trajectories with even lower computational overheads are obtained using a random sampling of δ between upper and lower bounds that can be calculated from experimental measurements (Noisy Langevin #2).

$$dC = [k_{\text{on}}(1 - \delta)RL - k_{\text{off}}C - k_e C]dt + \sqrt{k_{\text{on}}(1 - \delta)RL + k_{\text{off}}C + k_e C} dw_C \quad (16b)$$

$$dL = [-k_{\text{on}}(1 - \delta)RL + k_{\text{off}}C + F(t)]dt + \sqrt{k_{\text{on}}(1 - \delta)RL + k_{\text{off}}C + F(t)} dw_L \quad (16c)$$

where the independent parameters dw_R , dw_C and dw_L have zero-mean Brownian motion with variance \sqrt{dt} , and the dynamics of δ depend on the spatial distributions of the molecules involved in the chemical reactions.

In order to obtain a time history of δ , assume that a web-lab experiment has been performed and multiple measurements have been obtained of the number of complexes, indicated by green circles in Fig. 4 (here, we use a set of sampled average values from the stochastic reaction-diffusion simulations as surrogate data points). An optimal δ for the time period $(t_{k-1}, t_k]$ can then be calculated by solving the following minimisation problem:

$$\text{Minimise } J(\delta) = |\bar{C}_k - \bar{C}_k|, \quad (17)$$

where k is an integer between 1 and N , N is the total number of measurements, t_0 is the initial time, \bar{C}_k is the complex number measurement at t_k , \bar{C}_k is the mean value of the complex number at t_k provided by solving the noisy Langevin Equation (16), and dt is set to 1 s. The minimisation problem is a one-dimensional search, which can be efficiently solved by golden-section search or parabolic interpolation, as implemented in *fminbnd* in the MATLAB optimisation toolbox¹⁵. The results are shown in Fig. 4. As expected from the stochastic spatial-diffusion simulation setting, the spatial heterogeneity is larger at the beginning and slowly diminishes as the molecules interact and are diffused. The Langevin equations are integrated using the 1st-order Euler method. The comparisons with the trajectories from the solutions of the full 3-dimensional reaction diffusion equations are shown in Fig. 4, where both the number of ligands and the number of complexes (Noisy Langevin #1 in the figure) are well matched with the spatial-diffusion simulation. As shown in the figure, the proposed modelling approach provides a much better approximation of the true network dynamics than the ordinary differential equation based model, and in fact closely approximates the exact solution provided by solving the reaction diffusion equations.

Crucially, however, solving the noisy Langevin equations requires only a fraction of the computing time - less than a second to solve one complete time history for this example on a typical desktop computer. On the other hand, it takes about 38 minutes for MesoRD to finish one simulation¹⁶. On the same computer, for the uniform mixed condition, MCell, another Monte-Carlo stochastic reaction-diffusion simulator, takes about 24 minutes to obtain one realisation of the stochastic simulation¹⁷.

Note that a simpler approach of estimating δ instead of solving the optimisation problem would also be possible, and could potentially be useful for application of the approach to large-scale networks. Here, we assume that δ is a uniform distribution over $[\underline{\mu}, \bar{\mu}]$ which changes every t_N sampling time, and $\underline{\mu}$ and $\bar{\mu}$ are estimated using the total concentration changes from the experimental measurements. The number of complex molecules is then calculated for various values of $\underline{\mu}$ and $\bar{\mu}$ to fit the measurements, and t_N is set to 10 s. t_N is set to a value greater than dt as spatial fluctuations would not be faster than the chemical reactions. For our example, the estimated bounds for the best fit are $\underline{\mu} = 0.3$ and $\bar{\mu} = 0.7$. As expected, the resulting simulations are less accurate than those generated by the optimisation based approach, but reasonably well matched trajectories (Noisy Langevin #2 in Fig. 4) are still obtained.

Conclusions

A novel modelling methodology for simultaneously capturing stochastic and spatial effects on the dynamics of biomolecular networks was presented. We showed how, using a relatively simple formalism, i.e. the Langevin equation with noisy kinetic constants, spatial heterogeneity in the interaction space of the network can be included in the model. The effectiveness of the modelling approach is shown through its application to the modelling of the epidermal growth factor receptor network of a human mammary epithelial cell. Using the proposed method, highly accurate simulations of stochastic spatio-temporal dynamics were produced with massively reduced computational overheads. As spatial heterogeneity produces noticeably different dynamics in the network, it is vital that future modelling formalisms are able to take this heterogeneity into account, while minimising the computational complexity of the resulting simulations. The method presented in this paper represents a significant step forward in achieving this goal.

Methods

Stochastic Simulations. For the well-mixed cases, the Gillespie's direct method is implemented and the simulations are performed using MATLAB¹⁸. The reaction-diffusion equations are solved using MesoRD Ver. 1.1¹¹. The uniform mixed condition was also solved using MCell for the comparison of the computation time¹⁷. The noisy Langevin equation is implemented in MATLAB and solved using the 1st-order Euler method with the integration interval equal to 1 seconds.

Computer Hardware. All simulations are performed in a desktop computer with the CentOS Release 6.9, Intel Xeon Quadprocessor 3.20 GHz with memory of 15.6 GB.

Data availability. No datasets were generated or analysed during the current study.

References

1. Roeffaers, M. B. J. *et al.* Single-molecules fluorescence spectroscopy in (bio)catalysis. *Proc. Natl. Acad. Sci.* **104**, 12603–12609 (2007).
2. Wen, J.-D. *et al.* Following translation by single ribosomes one codon at a time. *Nature.* **452**, 598–604 (2008).

3. van Zon, J. S., Morelli, M. J., Tanase-Nicola, S. & ten Wolde, P. R. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.* **91**, 4350–4367 (2006).
4. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science*. **297**, 1183–1186 (2002).
5. Vilar, J. M. G., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci.* **99**, 5988–5992 (2002).
6. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Nature*. **307**, 1965–1969 (2005).
7. Ventura, B. D., Lemerle, C., Michalodimitrakis, K. & Serrano, L. From *in vivo* to *in silico* biology and back. *Nature*. **443**, 527–533 (2006).
8. Fange, D. & Elf, J. Noise-induced Min phenotypes in *E. coli*. *PLoS Comput. Biol.* **2**, e80 (2006).
9. Yang, J. *et al.* Kinetic Monte Carlo Method for Rule-Based Modeling of Biochemical Networks. *Phys. Rev. E* **78**, 031910 (2008).
10. Ruiz-Herrero, T., Estrada, J., Guantes, R. & Míguez, D. G. A Tunable Coarse-Grained Model for Ligand-Receptor Interaction. *PLoS Comput. Biol.* **9**, e1003274 (2013).
11. Hattne, J., Fange, D. & Elf, J. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics*. **21**, 2923–2924 (2005).
12. Kim, J., Bates, D. G., Postlethwaite, I., Heslop-Harrison, P. & Cho, K.-H. Least squares methods for identifying biochemical regulatory networks from noisy measurements. *BMC Bioinformatics*. **8**, 8 (2007).
13. Shankaran, H., Resat, H. & Wiley, H. S. Cell surface receptors for signal transduction and ligand transport: A design principles study. *PLoS Comput. Biol.* **3**, e101 (2007).
14. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
15. MATLAB Optimization Toolbox version 7.6 (R2017a) The MathWorks Inc., Natick, Massachusetts (2017).
16. Fange, D., Mahmutovic, A. & Elf, J. MesoRD 1.0: Stochastic reaction-diffusion simulations in the microscopic limit. *Bioinformatics*. **28**, 3155–3157 (2005).
17. Bartol, T. M., Dittrich, M. & Faeder, J. R. MCell. *Encyclopedia of Computational Neuroscience*. (Springer, New York, 2005).
18. MATLAB version 9.2.0. (R2017a) The MathWorks Inc., Natick, Massachusetts (2017).

Acknowledgements

This work was funded by a BBSRC/EPSRC grant to WISB (BB/M017982/1) and partially funded from the School of Engineering of the University of Warwick.

Author Contributions

J.K. conceived and conducted the experiments, J.K., M.F. and D.G.B. analysed the results. All authors wrote and reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

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



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

© The Author(s) 2018