Supplementary Information

Sensitivity, robustness and identifiability in stochastic chemical kinetics models

Michał Komorowski[1], Maria J. Costa[2], David A. Rand[2], Michael Stumpf[1]

1. Centre for Bioinformatics, Imperial College London, UK
2. Systems Biology Centre, and Mathematics Institute, University of Warwick, UK

This is supplementary information for the paper *Sensitivity, robustness and identifiability in stochastic chemical kinetics models* which is henceforth referred to as **MP**.

1 Derivation of the model equations

We consider a general system of chemical reactions that consists of N chemical species and interacts in a fixed volume through R reactions. Let $x = (x_1, \ldots, x_N)^T$ be the vector representing the numbers of molecules for the N species and $S = \{s_{ij}\}_{i=1,2...N;\ j=1,2...R}$ be the stoichiometry matrix that describes changes in the population sizes due to each of the reactions, so that occurrence of reaction j results in a change

$$(x_1, ..., x_N) \to (x_1 + s_{1j}, ..., x_N + s_{Nj}).$$

We assume that the probability that a reaction of type j occurs in the time interval $[t, t+dt)$ equals $f_j(x, \Theta)dt$, where functions $f_j(x, \Theta)$ are called the reaction rates and $\Theta = (\theta_1, ..., \theta_L)$ is the vector of all model parameters. The probability that more than one event will take place in a small time interval is of higher order (dt^2) with respect to the length of the interval and can thus be ignored. Finally, we assume that events taking place in disjoint time intervals are independent, when conditioned on the events in the previous interval. This specification leads to a Poisson birth and death process; the Chemical Master Equation [1, 2] is widely used to describe the temporal evolution of the probability $P(x, t)$ that the system is in the state x at time t

$$\frac{dP(x,t)}{dt} = \sum_{j=1}^{R} \left(P(x - s._j, t) f_j(x - s._j, \Theta, t) - P(x,t) f_j(x, \Theta, t) \right). \qquad (1)$$

Under the assumption that molecular species are present in sufficiently large copy numbers the model defined above is well described by the following system of equations [1, 3, 4]

$$\begin{aligned} x(t) &= \varphi(t) + \xi(t) && (2)\\ \dot{\varphi} &= S\ F(\varphi, \Theta, t) && (3)\\ d\xi &= A(\varphi, \Theta, t)\xi + E(\varphi, \Theta, t)dW, && (4) \end{aligned}$$

where

$$F(\varphi, \Theta, t) = (f_1(\varphi, \Theta, t), ..., f_R(\varphi, \Theta, t)) \quad (5)$$

$$\{A(\varphi, \Theta, t)\}_{ik} = \sum_{j=1}^{R} S_{ij} \frac{\partial f_j}{\partial \phi_k} \quad (6)$$

$$E(\varphi, \Theta, t) = S\sqrt{diag(F(\varphi, \Theta, t))}. \quad (7)$$

Equation (3) is an ordinary differential equation that in general does not have an explicit solution but can be solved numerically, whereas equation (4) is a linear stochastic differential equation that has a solution of the form [5]:

$$\xi(t) = \Phi(t_0, t)\xi_{t_0} + \int_{t_0}^{t} \Phi(s, t)E(\varphi, \Theta, s)dW(s), \quad (8)$$

where the integral is in the Itô sense and $\Phi(t_0, s)$ is the fundamental matrix of the non-autonomous system of ODEs

$$\frac{d\Phi(t_0, s)}{ds} = A(\varphi, \Theta, s)\Phi(t_0, s), \quad \Phi(t_0, t_0) = I. \quad (9)$$

In order to simplify the further analysis of the system studied, suppose that the initial condition has a multivariate normal distribution (MVN) $x(0) \sim MVN(\varphi(0), V(0))$.

This specification of an initial condition together with equations (2 - 4, 8) implies that $x(t)$ has a multivariate normal distribution [5, 6]

$$x(t) \sim MVN(\varphi(t), V(t)) \quad t > 0, \quad (10)$$

where $\varphi(t)$ is a solution of the macroscopic rate eqaution (MRE), Eqn. (3), with initial condition $\varphi(0)$, and $V(t)$ is a variance at time t. Direct calculations using equations (2 - 4, 8) show that V satisfies

$$\frac{dV(t)}{dt} = A(\varphi, \Theta, t)V(t) + V(t)A(\varphi, \Theta, t)^T + E(\varphi, \Theta, t)E(\varphi, \Theta, t)^T, \quad (11)$$

which is equivalent to the fluctuation dissipation theorem [1]. In the further sections we will also need to specify covariances, $\mathrm{cov}(x(s), x(t))$ $(t > s)$, and therefore we calculate these here. As $\langle x(t)\rangle = \varphi(t)$ we have that $\mathrm{cov}(x(s), x(t)) = \langle \xi(t)\xi(s)^T\rangle$ and therefore equation (8) implies

$$\mathrm{cov}(x(s), x(t)) = V(s)\Phi(s,t)^T. \tag{12}$$

2 Derivation of the likelihood function

In the previous section we have explained that $x(t) \sim MVN(\varphi(t), V(t))$. Now we derive the distribution of experimental data. Three different data types are considered: time series, time-point measurements, and deterministic model data.

2.1 Time series data

We start with the case where a single trajectory is measured at times $t_1, ..., t_n$. Initially suppose that all molecular species x_i are measured. Later we demonstrate that this assumption is easily relaxed. First let $\mathbf{x}_{TS} \equiv (x_{t_1}, \ \ldots, x_{t_n})$ be a nN column vector that contains all measurements and $\tilde{\varphi}(\varphi_0, \Theta, t)$ be a solution of equation (3) such that $\tilde{\varphi}(\varphi_0, \Theta, 0) = \varphi_0$, and let $\tilde{V}(V_0, \Theta, t)$ be a solution of equation (11) such that $\tilde{V}(V_0, \Theta, 0) = V_0$. In order to find the distribution of vector $\mathbf{x}_{TS}$ we write $x_{t_0} = \varphi(t_0) + \varsigma_{t_0}$, where $\varsigma_{t_0} \sim MVN(0, V_0)$ and using equations (3-4) and (8) we have

$$x_{t_1} = \varphi_{t_1} + \Phi(t_0, t_1)\varsigma_{t_0} + \varsigma_{t_1},$$

where $\varsigma_{t_1} \sim (0, \Xi_1)$ and $\Xi_1 = \int_{t_0}^{t_1} \Phi(s, t_1)^T E(s)^T E(s)\Phi(s, t_1)ds$. Using

$$\Phi(t_{j-1}, t_{j+1}) = \Phi(t_j, t_{j+1})\Phi(t_{j-1}, t_j) \tag{13}$$

we can analogously write x_{t_i} as

$$x_{t_i} = \varphi_{t_i} + \sum_{j=0}^{i} \Phi_{t_j}(t_i - t_j)\varsigma_{t_j}, \tag{14}$$

where ς_{t_j} are independently normally distributed random variables with mean 0 and covariance matrix $\Xi_j = \int_{t_{j-1}}^{t_j} \Phi(s, t_j)^T E(s)^T E(s)\Phi(s, t_j)ds$. This representation demonstrates that x_{t_i} is a linear sum of multivariate normal variables and therefore $\mathbf{x}_{TS}$ has a multivariate normal distribution with mean $\mu(\Theta)$ and covariance matrix $\Sigma_{TS}(\Theta)$

$$\mathbf{x}_{TS} \sim MVN(\mu(\Theta), \Sigma_{TS}(\Theta)) \tag{15}$$

where $\mu(\Theta) = (\tilde{\varphi}(t_1), ..., \tilde{\varphi}(t_n))$ and $\Sigma_{TS}(\Theta)$ is a $(n \times N) \times (n \times N)$ block matrix $\Sigma_{TS}(\Theta) = \left\{\Sigma(\Theta)^{(i,j)}\right\}_{i=1,...,N;j=1,...,N}$ such that diagonal elements contain variances $\Sigma(\Theta)^{(i,i)} = \tilde{V}(t_i)$ and non-diagonal elements ($i \neq j$) covariances $\Sigma(\Theta)^{(i,j)} = \text{cov}(x(t_i), x(t_j))$. Diagonal elements are given by the solution $\tilde{V}$. From representation (14) we have

$$\Sigma^{i,j+1} = \Sigma^{i,j}\Phi(t_j, t_{j+1})^T, \tag{16}$$

which demonstrates that non-diagonal elements can be easily computed from diagonal elements given by solutions of equation (11).

2.2 Time-point measurements

Here we consider the case where in an experiment at each time point $t_1, ..., t_n$ a different trajectory is measured. Therefore, measurements come from the same process $x(t)$ but from its independent realisations. We define the measurement vector as $\mathbf{x}_{TP} \equiv (x^{(1)}_{t_1}, \ \ldots, x^{(n)}_{t_n})$. Upper indices indicate the number of trajectories from which the measurements were taken in order to emphasis that each measurement is taken from a different trajectory. The distribution of $x^{(i)}_{t_i}$ is given by (10). All measurements are independent so that $\text{cov}(x_{t_i}, x_{t_j}) = 0$ for $i \neq j$, therefore

$$\mathbf{x}_{TP} \sim MVN(\mu(\Theta), \Sigma_{TP}(\Theta)) \tag{17}$$

where $\mu(\Theta) = (\tilde{\varphi}(t_1), ..., \tilde{\varphi}(t_n))$ and $\Sigma_{TP}(\Theta)$ has the same diagonal blocks as $\Sigma_{TS}(\Theta)$ and non-diagaonal blocks are equal to 0

$$\Sigma_{TP}(\Theta)^{(i,j)} = \begin{cases} \tilde{V}(t_i) & \text{for } \ i = j \\ 0 & \text{for } \ i \neq j. \end{cases} \tag{18}$$

2.3 Deterministic model

In order to study differences between stochastic and deterministic regimes we also consider a deterministic model where the system state is described entirely by the MRE (3). In such a model measurements are usually assumed to have the form

$$x(t_i) = \varphi(t_i) + \epsilon_{t_i},$$

, where ϵ_{t_i} is a normally and independently distributed measurement error with mean 0 and constant variance σ^2_ϵ. We denote the measurements for this model by $\mathbf{x}_{DT} \equiv (x_{t_1}, \ \ldots, x_{t_n})$. Finding the data distribution for this case is straightforward,

$$\mathbf{x}_{DT} \sim MVN(\mu(\Theta), \Sigma_{DT}), \tag{19}$$

where $\mu(\Theta)$ is as in the previous cases and Σ_{DT} is a N^2n^2 diagonal matrix with diagonal elements equal to σ^2_ϵ.

2.4 Hidden variables

Usually it is not possible to measure all variables present in the system of interest experimentally. Hence we here demonstrate that the distribution of observed components can be directly extracted from the distributions (15, 17, 19). For simplicity we consider the case of time series data only; analysis for the two other data types proceeds analogously. First, we partition the process $x(t)$ into those components $y(t)$ that are observed and those $z(t)$ that are unobserved. Let $\mathbf{y}_{TS} \equiv (\mathbf{y}_{t_1}, \ldots, \mathbf{y}_{t_n})$ and $\mathbf{z}_{TS} \equiv (\mathbf{z}_{t_1}, \ldots, \mathbf{z}_{t_n})$ denote the time series that of y and z,respectively, at times $t_1, \ldots t_n$.

The distribution of $\mathbf{y}_{TS}$ is a marginal distribution of $\mathbf{x}_{TS}$; we thus have

$$\mathbf{y}_{TS} \sim MVN(\mu_y(\Theta), \hat{\Sigma}(\Theta)), \tag{20}$$

where $\mu_y(\Theta)$ and $\hat{\Sigma}(\Theta)$ are elements of $\mu(\Theta)$ and $\Sigma_{TS}(\Theta)$ that correspond to the observed components of $\mathbf{x}_{TS}$. If for instance first M out of N components of x are observed than $y(t) = (x_1(t), ..., x_M(t))$, and

$$\mu_y(\Theta) = (\tilde{\varphi}_M(t_1), ..., \tilde{\varphi}_M(t_n)) \tag{21}$$

where $\tilde{\varphi}_M(t) = (\tilde{\phi}_1(t), ..., \tilde{\phi}_M(t))$ and $\hat{\Sigma}(\Theta)=$ is a $MN \times MN$ block matrix

$$\hat{\Sigma}(\Theta) = \left\{ \hat{\Sigma}(\Theta)^{(i,j)} \right\}_{i=1,...,N; j=1,...,N} \tag{22}$$

where

$$\hat{\Sigma}^{(i,j)}_{pq}(\Theta) = \Sigma^{(i,j)}_{pq}(\Theta) \quad p = 1, ..., M, q = 1, ..., M. \tag{23}$$

3 Calculation of the Fisher Information Matrix (FIM)

Suppose that a random variable X has an N-variate normal distribution with mean $\mu(\Theta) = (\mu_1(\Theta), ..., \mu_N(\Theta))^T$ and covariance matrix $\Sigma(\Theta)$. We define the FIM for this variable to be $I(\Theta)= \{I(\Theta)_{k,l}\}$ [7]

$$I(\Theta)_{k,l} = E_\Theta \left[\left(\frac{\partial}{\partial \theta_k} \log(\psi(X, \Theta)) \right) \left(\frac{\partial}{\partial \theta_l} \log(\psi(X, \Theta)) \right) \right], \tag{24}$$

where $\psi(.)$ is the density function of a multivariate normal distribution with mean $\mu(\Theta)$ and covariance $\Sigma(\Theta)$. As the random variable X is normally distributed the elements $I(\Theta)_{k,l}$ can be also expressed explicitly as

$$I(\Theta)_{k,l} = \frac{\partial \mu}{\partial \theta_k}^T \Sigma(\theta) \frac{\partial \mu}{\partial \theta_l} + \frac{1}{2} trace(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l}). \tag{25}$$

In this section we demonstrate how to calculate the FIM for the models (2- 4). We consider the model for time series data, as the case of time-point measurements is less general and can be directly extracted from formulas derived below. Previously, we have shown 15 that variable $\mathbf{x}_{TS}$ has a multivariate normal distribution and demonstrated how its mean $\mu(\Theta)$ and covariance matrix $\Sigma(\Theta)$ can be calculated. Formula (25) indicates that two more components $\frac{\partial \mu}{\partial \theta_k}$ and $\frac{\partial \Sigma(\Theta)}{\partial \theta_k}$ need to be known in order to be able to compute FIM. Below we show how these can be obtained.

Let $Y(t)$ be the concatenated vector of $\varphi(t)$ and upper diagonal of the symmetric matrix V

$$Y(t) = (\phi_1(t), ..., \phi_N(t), V_{1,1}(t), ..., V_{N,N}(t), ..., V_{1,2}(t), ..., V_{N-1,N}(t)) \tag{26}$$

and $\tilde{Y}(Y_0, \Theta, t)$ be the concatenation of $\tilde{\varphi}(\varphi_0, \Theta, t)$ and upper diagonal of $\tilde{V}(V_0, \Theta, t)$. Similarly denoting the concatenation of the right hand sides of equations (3) and (11) by W we can write

$$\frac{d}{dt} Y(t) = W(Y(t), \Theta, t). \tag{27}$$

To determine the derivative $Z_k(t) = \frac{\tilde{Y}(t)}{\partial \theta_k}$ we use the fact that it satisfies the following equation (see Appendix)

$$\frac{d}{dt} Z_k(t) = J(\tilde{Y}(t), \Theta, t) Z_k(t) + K_k(t), \tag{28}$$

where $J(\tilde{Y}(t), \Theta, t)$ is the Jacobian $\frac{\partial}{\partial Y(t)} W(Y(t), \Theta, t)$ evaluated at the solution $\tilde{Y}(t)$ and $K_k(t)$ is the vector $\frac{\partial}{\partial \theta_k} W(Y(t), \Theta, t)$ also evaluated at $\tilde{Y}(t)$.
The solution of equation (28), $\tilde{Z}(t)$, provides us with $\frac{\partial \tilde{\phi}(t)}{\partial \theta_k}$ and therefore with $\frac{\partial \mu}{\partial \theta_k}$. Similarly $\tilde{Z}(t)$ contains diagonal elements of $\frac{\partial \Sigma}{\partial \theta_k}$.

Non-diagonal elements of $\frac{\partial \Sigma}{\partial \theta_k}$ can be computed from diagonal elements using the recursive relation

$$\frac{\partial}{\partial \theta_k} \Sigma^{(i,j+1)} = \frac{\partial}{\partial \theta_k} \left(\Sigma^{(i,j)} \Phi(t_j, t_{j+1})^T \right) = \frac{\partial}{\partial \theta_k} \left(\Sigma^{(i,j)} \right) \Phi(t_i, t_{i+1})^T + \Sigma^{(i,j)} \frac{\partial}{\partial \theta_k} \left(\Phi(t_i, t_{i+1})^T \right). \tag{29}$$

from elements $\Phi(t_j, t_{j+1})$, $\Sigma^{(i,i)}$, $\frac{\partial}{\partial \theta_k} \Sigma^{(i,i)}$ that are given by equations (9), (16) and (28) respectively. To simplify notation denote $\Xi_k(s,t) = \frac{\partial \Phi(s,t)}{\partial \theta_k}$. As $\Phi(s,t)$ is a solution of an ODE we use similar techniques as in equations (28) and write $\Xi_k(s,t)$ as a solution of the

differential equation

$$\frac{d\Xi_k}{dt}(s,t) = A(\tilde{\varphi},\Theta,t)\Xi(s,t) + M_k(t), \tag{30}$$

where

$$M_k(t) = \frac{\partial}{\partial\theta_k}(A(\tilde{\varphi},\Theta,t)\Phi(s,t)) = \left(\frac{\partial}{\partial\theta_k}A(\tilde{\varphi},\Theta,t) + (\frac{\partial}{\partial\varphi}A(\varphi,\Theta,t)_{\varphi=\tilde{\varphi}})\frac{\partial\tilde{\varphi}}{\partial\theta_k}\right)\Phi(s,t), \tag{31}$$

and $\Xi(s,s) = 0$ for all s. To summarise, for the experimental data distribution (15) the FIM (25) can be computed using equations (28 - 31):

- the parameter derivative of the mean, $\frac{\partial\mu(\Theta)}{\partial\theta_k}$, can be extracted from a solution of (28)
- diagonal elements of the parameter derivatives of the variance, $\frac{\partial\Sigma_{TS}(\Theta)}{\partial\theta_k}$, can be extracted from a solution of (28)
- non-diagonal elements of parameter derivatives of the variance, $\frac{\partial\Sigma_{TS}(\Theta)}{\partial\theta_k}$, are given by formula (29), which involves (30) and (31).

3.1 Summary of the numerical computation of the FIM

Below we summarise in more details how the FIM can be calculated numerically. We start with the case of time series data as it is most general and the remaining two can be derived from it.

Time series measurmens

1 Read input: Stoichiometry matrix S, reaction rates vector F, initial conditions x_0, V_0

2 Construct equations for φ (eq. (3)) and V (eq. (11)) and for Y (eq. (27))

3 Calculate symbolically the Jacobians A (eq. (6)), J (eq. (28)) and vectors K_k (eq. (28)) , M_k (eq. 30) $(k = 1, ..., L)$

4 Solve equations for φ and V (eq. (27))

5 Compute fundamental matrices $\Phi(t_i, t_{i+1})$ $i = 1, ..., N-1$ (eq. (9))

6 Construct covariance matrix Σ_{TS} from $\tilde{V}(t_i)$ and $\Phi(t_i, t_{i+1})$ $(i = 1, ..., n)$ according to eq. (16)

7 Compute $\frac{\partial \tilde{Y}}{\partial \theta_k}$ (solve eq. (28)) and extract $\frac{\partial \mu}{\partial \theta_k}$ $\frac{\partial \tilde{V}}{\partial \theta_k}$ according to eq. (26)

8 Compute $\frac{\partial}{\partial \theta_k}\Phi(t_i, t_{i+1})$ (eq. (30))

9 Compute $\frac{\partial}{\partial \theta_k}\Sigma^{(i,j)}$ for $j =\geq i+1, ..., n$ and $i = 1, ..., n$ (eq. (29))

10 Construct $\frac{\partial}{\partial \theta_k}\Sigma_{TS}$ from objects computed in steps 7 and 9

11 From $\frac{\partial}{\partial \theta_k}\varphi$, Σ_{TS}, $\frac{\partial}{\partial \theta_k}\Sigma_{TS}$ extract those elements corresponding to observed component according to relations (21), (22) and (23)

12 Compute FIM from elements obtained in the previous steps according to eq. (25)

Time-point measurmens
In order to compute the FIM for time-point measurements the covariance matrix, Σ_{TS}, should be replaced by Σ_{TP}. Additionally non-diagonal blocks of covariance matrix, Σ_{TP}, are equal to 0, therefore steps 5, 8 and 9 are omitted and in step 3 vectors M_j need not be computed.

Deterministic model data
For the deterministic model the covariance matrix does not depend on parameters, therefore the formula for the elements of FIM simplifies to

$$I(\Theta)_{k,l} = \frac{\partial \mu}{\partial \theta_k}^T \Sigma_{DT}(\Theta) \frac{\partial \mu}{\partial \theta_l}, \tag{32}$$

and it requires only calculation of derivatives $\frac{\partial \tilde{\varphi}}{\partial \theta_k}$ for $k = 1, \ldots, L$.

4 Examples

In this section we present details pertaining to the examples of models of single gene expression and the p53 system.

4.1 A model of a single gene expression

The Table 1 contains parameter values used for numerical experiments presented in the main paper.

Param.	Set 1	Set 2	Set 3	Set 4
k_r	100	100	20	20
k_p	2	2	10	10
γ_r	1.2	0.7	1.2	0.7
γ_p	0.7	1.2	0.7	1.2

Table 1: Four parameter sets used in analysis of the single gene expression model. Sets 1 and 3 correspond to slow protein degradation rate γ_p and high and low transcription / translation ratio, respectively. On the other hand Sets 2 and 4 describe fast protein degradation rate and high and low transcription / translation ratio, respectively. All rates are per hour.

4.1.1 Differences in sensitivity and robustness analysis between time-series, time-points and deterministic versions of the model

Considering sensitivity and robustness analysis, there are three main differences between stochastic and deterministic systems. Firstly deterministic models completely neglect variability in the abundances of molecular species. This variability is a function of the kinetic parameters and is therefore also sensitive to them. Secondly, the deterministic model does

Type	# ident. param.	$CR(\log(k_r))$	$CR(\log(k_p)))$	$CR(\log(\gamma_r))$	$CR(\log(\gamma_p))$	det(FIM)
TS	4	0.0017	0.0016	0.0017	0.0017	$4 \cdot 10^3$
TP	2	0.0002	0.0002	0.0002	0.0002	0
DT	1	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	0

Table 2: Identifiability analysis for stationary state data. The table presents the number of non-zero eigenvalues (# ident. param.), Cramer-Rao bounds (CR), determinants of FIM (det(FIM)) for different data types (time series (TS), time point measurements (TP), deterministic model (DT)). The number of non-zero eigenvalues equals to the number of (in principle) identifiable linear combinations of parameters and therefore describes the number of parameters that can be estimated given that others are known. Quantities were calculated for parameter set 3 (see Table 1) and we have set the sampling frequency to $\Delta = 0.3h$, and the number of measurements to $n = 50$. The system was supposed to be in the stationary state. We have assumed that a parameter is identifiable if an eigenvalue of FIM is not lower than 10^{-4} to take account of numerical inaccuracies, and therefore "# ident. param." is calculated as the number of eigenvalues that are greater or equal to 0.1% of the largest eigenvalue. For the same reason the determinant was calculated as the product of eigenvalues that satisfies this condition. As not all parameters were identifiable in all versions of the model we calculated CR for individual estimates (assuming all other parameters to be known). For the deterministic model we have set variance of measurement error $\sigma_\epsilon^2 = 100$ and no measurement error for TS and TP therefore CR-bounds between stochastic and deterministic models cannot be compared.

not include correlations between molecular species. Thirdly, temporal correlations are also neglected.

In the main paper we argued that these three factors can have a significant impact on how stochastic and deterministic systems respond to perturbations in parameters. Here we provide further explanation using the model of single gene expression. The formulae for mean,

Type	# ident. param.	$CR(\log(k_r))$	$CR(\log(k_p))$	$CR(log(\gamma_r))$	$CR(\log(\gamma_p))$	det(FIM)
TS	4	0.0413	0.0112	0.0098	0.0072	$6.96 \cdot 10^4$
TP	4	0.0185	0.0036	0.0036	0.0020	$6.94 \cdot 10^3$
DT	3	$0.47 \cdot 10^{-4}$	$0.04 \cdot 10^{-4}$	$0.07 \cdot 10^{-4}$	$0.02 \cdot 10^{-4}$	0

Table 3: Identifiability analysis for perturbation experiment. Identical analysis as in Table 2 but with an initial mean increased 5 fold and the initial variance 25 fold.

variances and covariance for this model are

$$\langle r \rangle = \frac{k_r}{\gamma_r} \tag{33}$$

$$\langle p \rangle = \frac{k_r k_p}{\gamma_r \gamma_p} \tag{34}$$

$$\langle \delta r^2 \rangle = \frac{k_r}{\gamma_r} \tag{35}$$

$$\langle \delta p^2 \rangle = \langle p \rangle (1 + \frac{k_p}{\gamma_r + \gamma_p}) \tag{36}$$

$$\langle \delta r \delta p \rangle = \frac{k_p k_r}{(\gamma_r + \gamma_p)\,\gamma_r}. \tag{37}$$

In order to understand the effect of incorporating variability into the sensitivity analysis we are considering changes in parameters, e.g. k_p, γ_p, by a factor δ $(k_p, \gamma_p) \rightarrow (k_p + \delta k_p, \gamma_p + \delta \gamma_p)$. Means of RNA and protein concentrations are not affected by this perturbation, whereas protein variance is.

To understand the effect of correlation between RNA and protein levels we note that formulae (33 - 37) demonstrate that at constant mean, correlation increases with γ_p and compensating decrease in k_p. Figure 7 presents neutral spaces for all parameter pairs for different values of correlation $\rho_{rp} = \frac{\langle \delta r \delta p \rangle}{\sqrt{\langle \delta r^2 \rangle \langle \delta p^2 \rangle}}$. Differences between deterministic and stochastic model increase with correlation.

We also perform similar analyses for different levels of temporal correlation between observations by varying the sampling frequency Δ. Figure 8 presents neutral spaces for all parameter pairs for three different sampling frequencies. The differences between stochastic and deterministic models decrease with Δ as the samples are less correlated for high Δ, and therefore the factor that distinguishes two models becomes less significant.

4.2 P53 system

In the study of the 53 system we have used parameter estimates presented in Table 4. These parameters has been obtained by appropriate scaling of the parameters given in [8]. For all numerical experiments for p53 model we assumed sampling frequency $\Delta = 1h$ and number of measurements $n = 30$.

Param.	Value
β_x	90
α_x	0.002
α_k	1.7
k	0.01
β_y	1.1
α_0	0.8
α_y	0.8

Table 4: Parameters of p53 system.

4.2.1 Sloppiness analysis

Here we compare neutral spaces for all pairs of parameters of the P53 model for tree data types (TS, TP, DT). We use parameter values presented in Table 4 and logarithmic parametrisation. Results are presented in Figues 4, 5 and 6.

4.3 Dependance of analysis on parameter values and qualitative model behaviour

Our method allows to study model sensitivity given the parameter values. Here we show that results depend on parameter values just as the dynamical behaviour of the system does; i.e. the sloppiness of a system also depends on the parameters and is not a fixed property of a mathematical model. Figure 9 demonstrates that p53 undergoes a Hopf bifurcation as parameter α_y is varied from 0.8 to 2, while all other parameters remain unchanged. Parameters thus determine the qualitative dynamical behaviour and therefore varying parameters

influences the structure of the FIM (compare Figures 4 and 10). Change in the FIM in turn has consequence for the magnitude of eigenvalues (Figure 9), sensitivity coefficients (compare Figures 4 in **MP** and 12) and informational content of TS and TP data (Figures 2 and 13).

5 Appendix

5.1 Calculating derivatives of a solution of ODE

In order to derive equations (28) and (30) we differentiated the solution of an ODE with respect to a parameter. Here we provide more details about this procedure. Suppose that the differential equation being considered is

$$\dot{x} = \mathsf{F}(x, \Theta, t),$$

where $x \in R^N$ and the set of parameters are collected together into a parameter vector $\Theta \in R^L$. Suppose that $\tilde{x}(\Theta, x_0, t)$ is the solution of interest with an initial condition $x_0 = \tilde{x}(\Theta, x_0, 0)$. For $Y(t) = \frac{\partial \tilde{x}(\Theta, x_0, t)}{\partial \theta_k}$ it can be shown [9] that

$$\dot{Y} = J(t)Y(t) + K_k(t) \tag{38}$$

where $J(t)$ is the Jacobian of F with respect to x evaluated at $\tilde{x}$, $K_k(t)$ is the n-dimensional vector $\frac{\partial F}{\partial \theta_k}$ and $Y(0) = 0$.

5.2 Logarithmic parametrisation

In the analysis of examples presented in our study we used a logarithmic parametrisation. Below we provide a rationale for this and explain that the FIM for logarithmic parametrisation can be directly obtained from derivatives calculated to obtain the FIM for original parameters.

In biochemical systems, the values of two parameters may differ by orders of magnitudes. Therefore, it is usually not appropriate to consider the absolute changes in the parameters θ_k, but instead to consider the relative changes. A good way to do this is to introduce new parameters $\eta_k = \log(\theta_k)$, because absolute changes in η_k correspond to relative changes in θ_k. Then, for small changes $\delta\theta_k$ to the parameters θ_k, the changes η_k are scaled and non-dimensional. Analyses in terms of absolute and relative changes are closely related and do not require additional computational cost. For any differentiable function $f(\theta)$

$$\frac{\partial f}{\partial \log(\theta)} = \frac{\partial f}{\partial \eta} = \theta \frac{\partial f}{\partial \theta}. \tag{39}$$

and therefore any element of FIM for η can be easily converted into that for θ or *vice versa*

$$I(\eta)_{k,l} = \frac{\partial \mu}{\partial \eta_k}^T \Sigma(\eta) \frac{\partial \mu}{\partial \eta_l} + \frac{1}{2} trace(\Sigma^{-1} \frac{\partial \Sigma}{\partial \eta_k} \Sigma^{-1} \frac{\partial \Sigma}{\partial \eta_l})$$

$$= \theta_k \theta_l \frac{\partial \mu}{\partial \theta_k}^T \Sigma(\theta) \frac{\partial \mu}{\partial \theta_l} + \theta_k \theta_l \frac{1}{2} trace(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l})$$

5.3 The FIM as a measure of system's sensitivity

Here we provide an alternative explanation why the FIM provides a measure of sensitivity for a stochastic system. For notational simplicity we assume that the studied system depends on a single parameter θ as generalisation for multidimensional parameter is straightforward. We start with definitions of classical sensitivity coefficients.

5.3.1 Classical sensitivity coefficient

The classical sensitivity coefficient for observable Q and parameter θ is defined as [10]

$$S = \frac{\partial Q}{\partial \theta}.$$

Often sensitivity of relative changes $\frac{\Delta Q}{Q}$ needs to be considered. Given that $\Delta \log(Q) \approx \frac{\Delta Q}{Q}$ the formula for sensitivity of relative changes takes the form

$$S = \frac{\partial \log(Q)}{\partial \theta}.$$

5.3.2 The FIM as a sensitivity measure for a stochastic system

The behaviour of a stochastic system is not defined by an observable Q that can be measured experimentally in a reproducible way. It is instead defined by a distribution form which the measurements are taken. Suppose we want to construct a measure of a sensitivity of a distribution of a random variable X with density ψ. Assume we want to examine relative changes of the distribution ψ to changes in θ. This can be written as

$$\frac{\psi(X, \theta + \partial\theta) - \psi(X, \theta)}{\psi(X, \theta)} \simeq \log(\frac{\psi(X, \theta + \partial\theta)}{\psi(X, \theta)}).$$

Averaging over all possible observations we get

$$\int_X \log(\frac{\psi(X, \theta + \partial\theta)}{\psi(X, \theta)}) \psi(X, \theta) dX.$$

The above is the negative Kullback-Leibler divergence between distributions $\psi(X, \theta + \partial\theta)$ and $\psi(X, \theta)$. In order to study changes in ψ resulting from "small" changes in θ we divide the above equation by $\partial\theta$ and take the limit $\partial\theta \to 0$ and get

$$\int_X \frac{\partial \log(\psi(X, \theta))}{\partial\theta} \psi(X, \theta) dX.$$

The above quantity is the average of the score function and it is basic fact of mathematical statistics that it equals to zero [7]. This observation suggests that it is better to study the squared differences

$$\int_X (\log(\psi(X, \theta) - \log(\psi(X, \theta + \partial\theta))^2 \psi(X, \theta) dX,$$

that lead to

$$\int_X \frac{(\log(\psi(X, \theta) - \log(\psi(X, \theta + \partial\theta))^2}{(\partial\theta)^2} \psi(X, \theta) dx \xrightarrow[\partial\theta \to 0]{} \int_X \left(\frac{\partial}{\partial\theta} \log \left(\frac{(\psi(X, \theta + \partial\theta))}{\psi(X, \theta)} \right) \right)^2 \psi(X, \theta) dx,$$

which is precisely the definition of the FIM.
The above derivation suggests that the FIM is a good measure of sensitivity of a probability distribution and that there is a close link between Kulback-Leibler divergence and the FIM. The KL divergence measures the average relative difference between two distributions whereas FIM measures squared relative difference between a distribution and the same distribution with a perturbed parameter relative to the infinitesimal size of the squared perturbation.

References

1. N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 2006.

2. C. Gardiner. *Handbook of stochastic methods*. Springer, 1985.

3. J. Elf and M. Ehrenberg. Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation. *Genome Res.*, 13(11):2475–2484, 2003.

4. Thomas G. Kurtz. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

5. L. Arnold. *Stochastic differential equations: theory and applications.* Wiley-Interscience, 1974.

6. B. Oksendal. *Stochastic differential equations (3rd ed.): an introduction with applications.* Springer, 1992.

7. S.D. Silvey. *Statistical inference.* Chapman & Hall, 1975.

8. N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, et al. Oscillations and variability in the p53 system. *Molecular Systems Biology*, 2(1), 2006.

9. D. Zwillinger. *Handbook of Differential Equations.* San Diego, 1989.

10. D. A. Rand. Mapping the global sensitivity of cellular network dynamics. *Journal of The Royal Society Interface*, 5:S59, 2008.

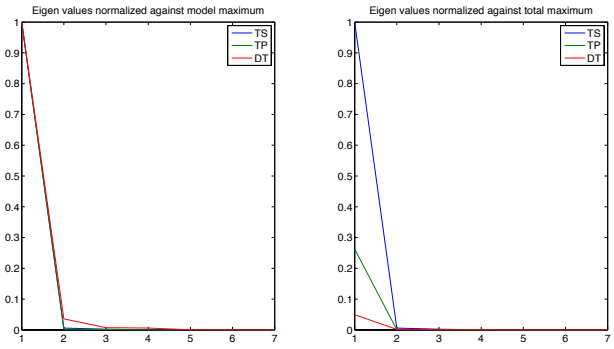

Figure 1: Eigenvalues of FIM for p53 model for three data types: time series (blue), time-points (green) and deterministic model (red). Eigen values were normalised against maximal eigenvalue for each data type (left) and against maximal eigenvalue among all three types (right). FIM was calculated for logs of parameters from Table 4 .

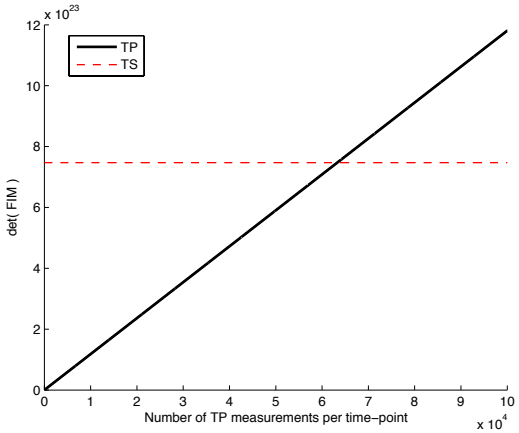

Figure 2: Comparison of informational content of TP and TS samples for p53 model. Determinant of the FIM for TP data is plotted against number of measurements per time point (black line). Due to independence of measurements we observe the linear increase. For comparison determinant of the FIM for a single TS sample is also depicted (red dashed line). Intersection of the two lines indicates how many TP measurements are necessary to obtain the same amount of information in a single trajectory (TS data). In this case around $6.5 \cdot 10^4$ measurements are needed.

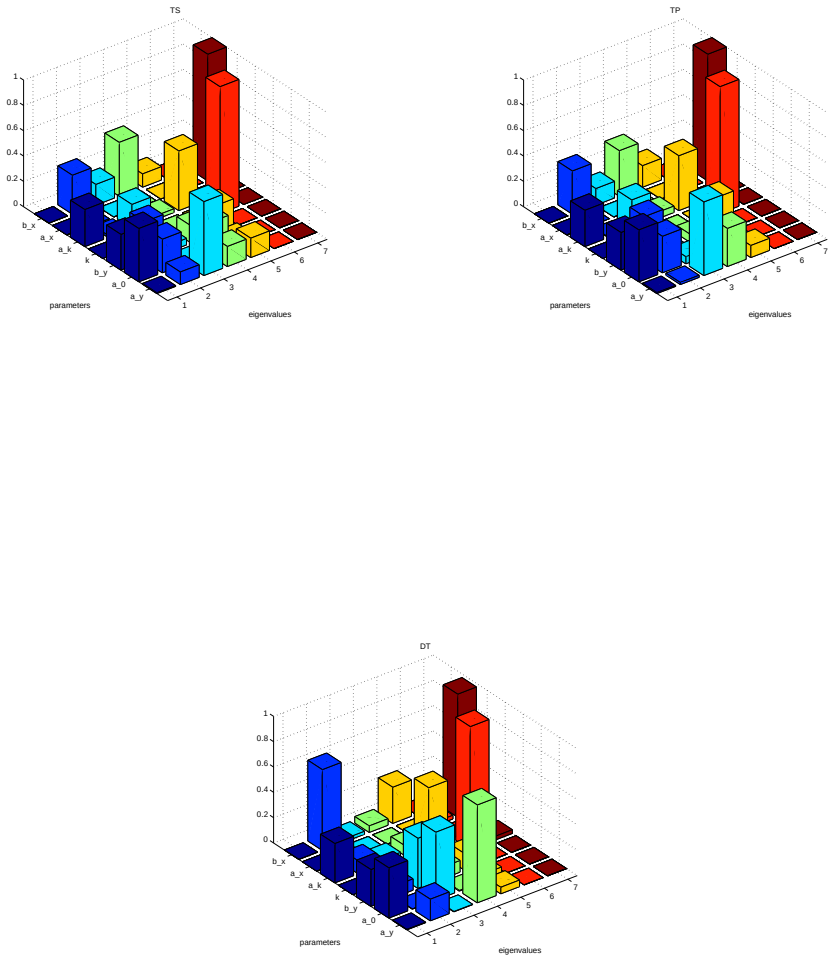

Figure 3: Sensitivity matrices C_{ij}^2 for p53 model for three data types (TS, TP, DT) calculated using parameters presented in Table 4.

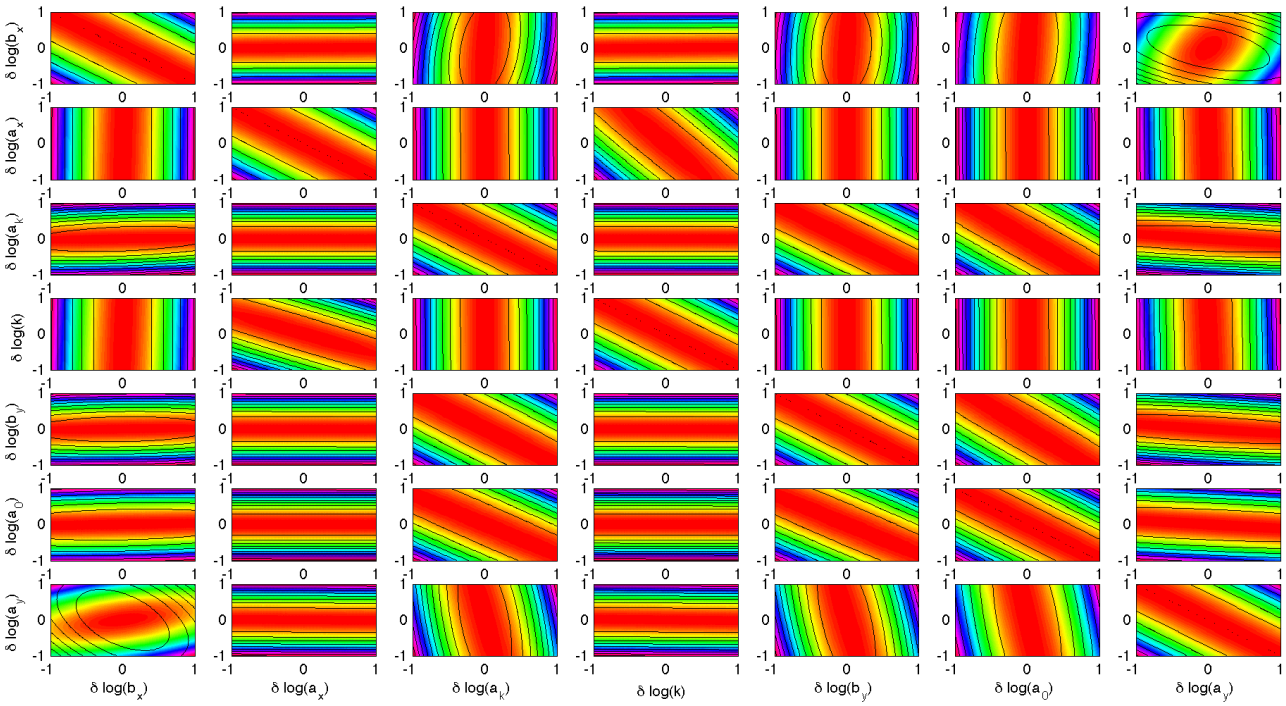

Figure 4: Neutral space for time-series (heatmap) and deterministic (contour plot) versions of the p53 model. The FIM was calculated for the logarithms of parameters in Table 4.

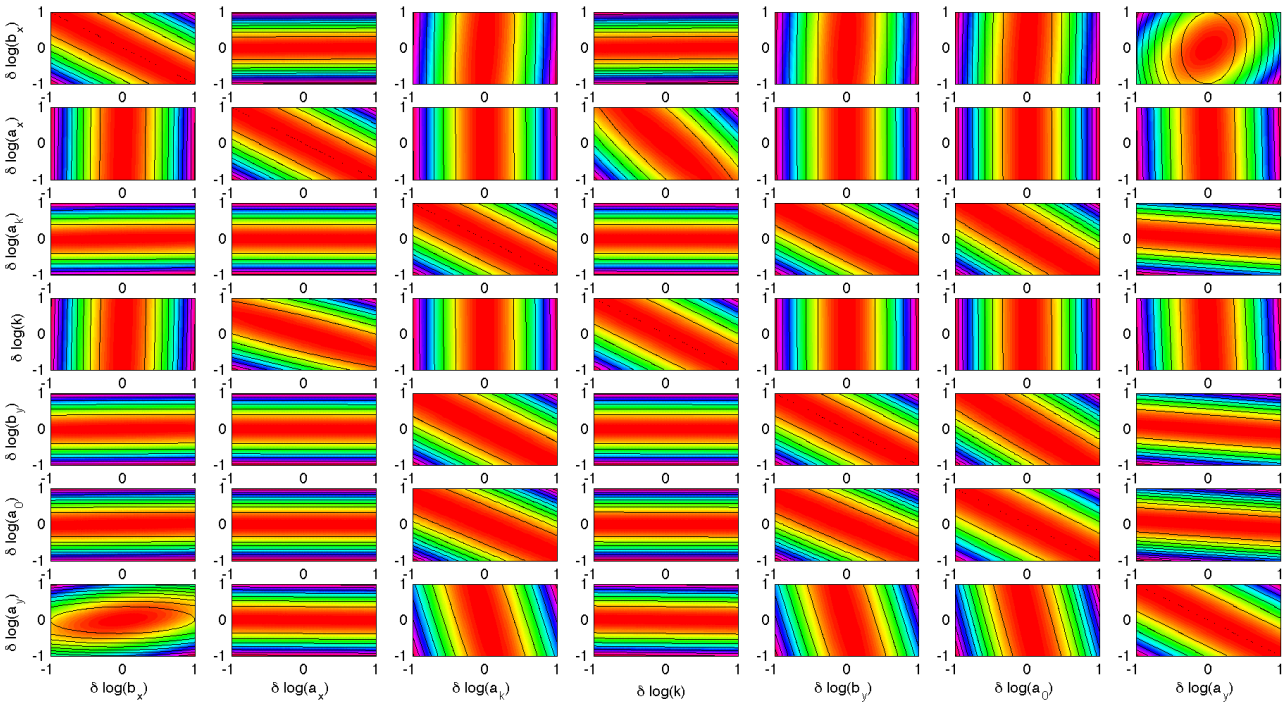

Figure 5: Neutral space for time-series (heatmap) and time-point (contour plot) versions of the p53 model. The FIM was calculated for logarithms of parameters in Table 4.

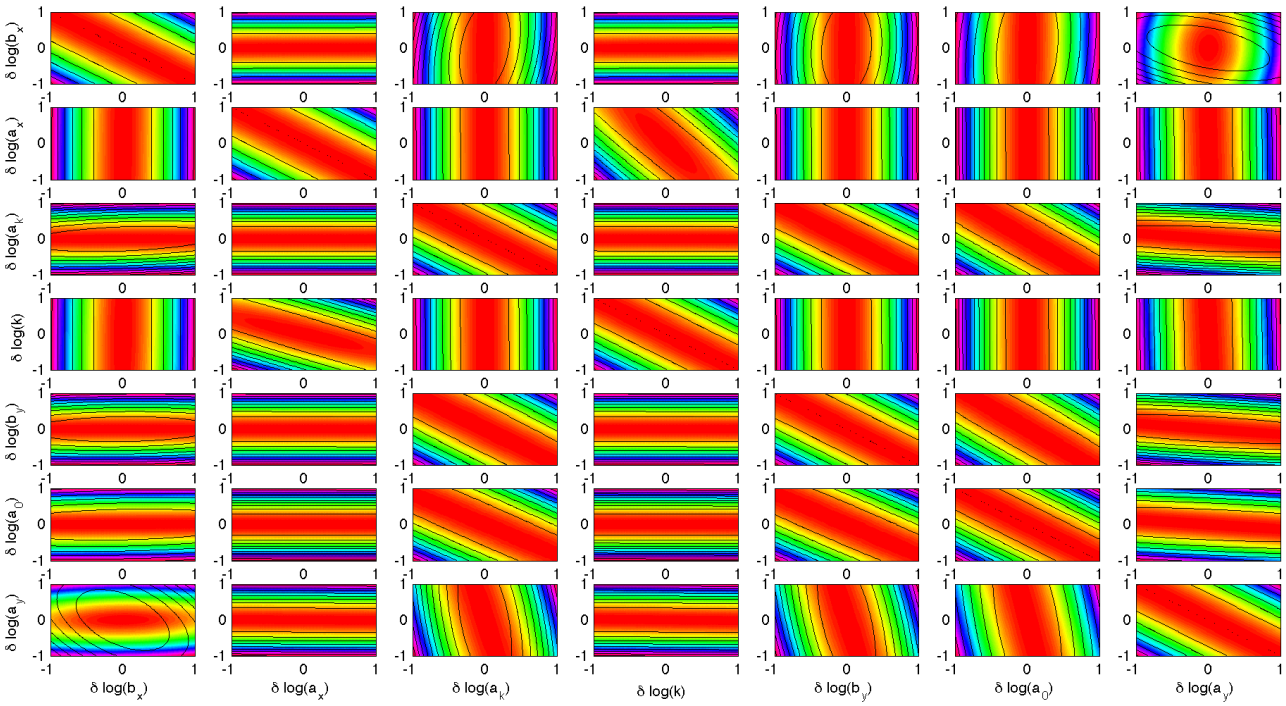

Figure 6: Neutral space for time-points (heatmap) and deterministic (contour plot) versions of the p53 model. The FIM was calculated for logarithms of parameters in Table 4.

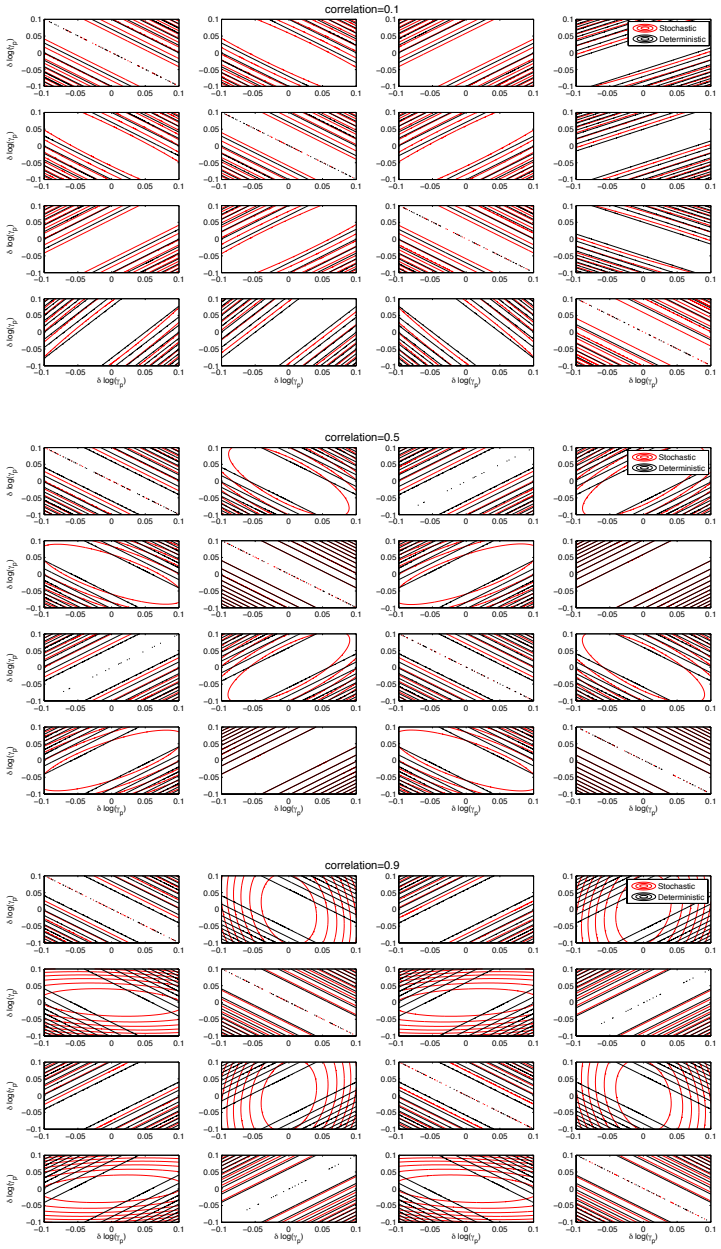

Figure 7: FIM for a single measurement from stationary distribution of the model of single gene expression. For $\rho_{rp} = 0.1$ (top), $\rho_{rp} = 0.5$ (middle), $\rho_{rp} = 0.9$ (bottom). Correlation 0.5 was obtained using parameter set 3 from Table 1. Correlation was varied by equal-scaling of parameters k_p, γ_p.

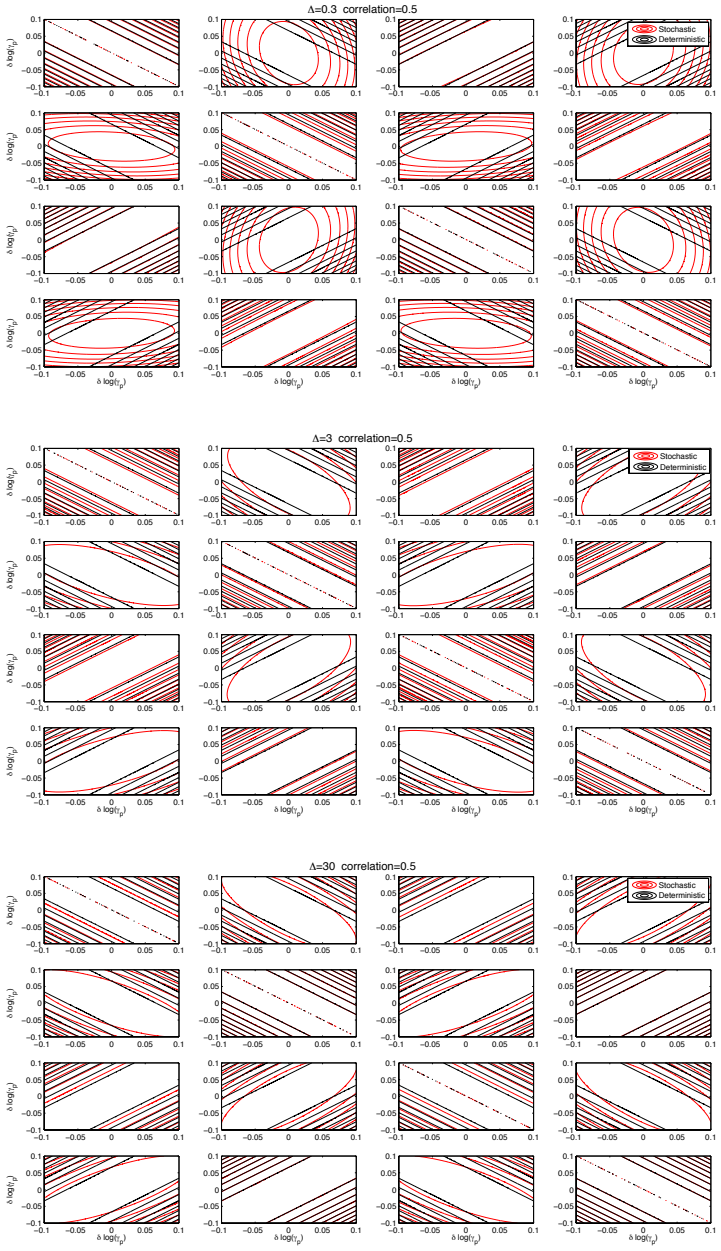

Figure 8: FIM for a 20 times-series type measurements from the stationary distribution of the model of single gene expression for three different sampling frequencies: $\Delta = 0.3$ (top) $\Delta = 3$ (middle) $\Delta = 30$ (bottom). Parameter set 3 from Table 1 was used.

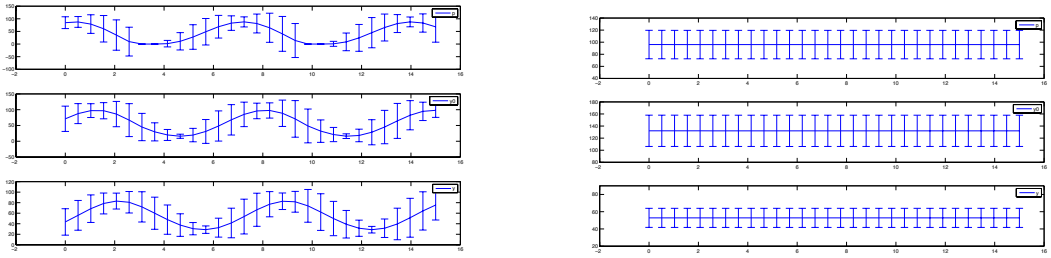

Figure 9: Trajectories of the p53 system plotted together with the standard deviations bars. In the left panel parameters from Table 4 were used and in the right panel the same parameters except $\alpha_y = 2$. The system undergoes a Hopf bifurcation and moves from oscillatory behaviour to dynamics with a stable stationary state.

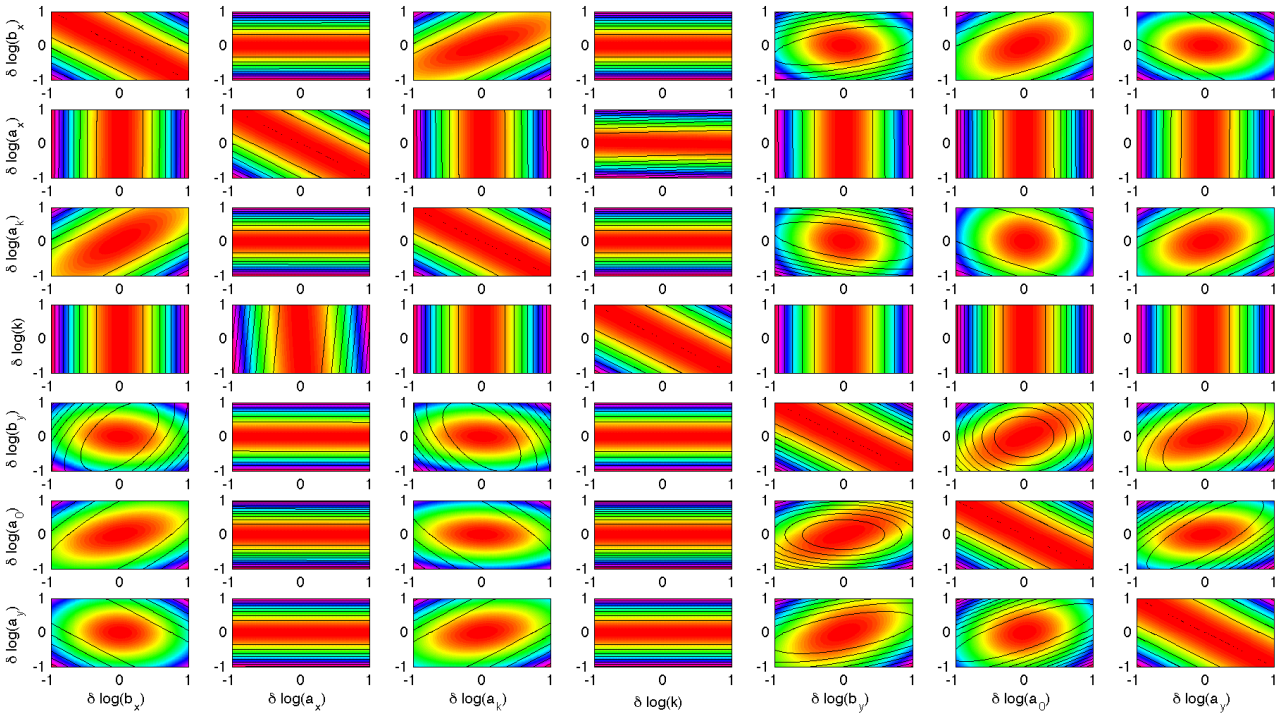

Figure 10: Neutral spaces for time-points (heatmap) and deterministic (contour plot) versions of the p53 model. FIM was calculated for logarithmss of parameters in Table 4 except $\alpha_y = 2$. Differences compared with Figure 6 demonstrate the dependance of neutral spaces on parameter values and the qualitative dynamics of the system.

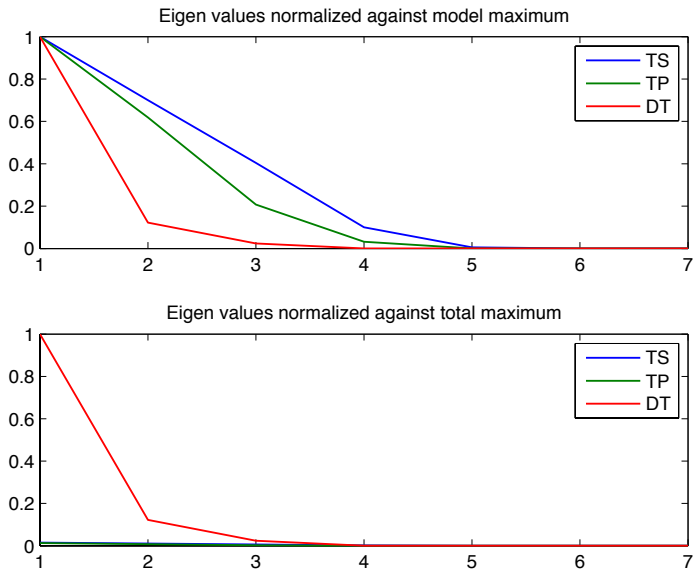

Figure 11: Eigenvalues of FIM for p53 model for three data types: time series (blue), time-points (green) and deterministic model (red). Eigenvalues were normalised against maximal eigenvalue for each data type (top) and against maximal eigenvalue among all three types (bottom). FIM was calculated for logs of parameters from Table 4 except $\alpha_y = 2$. Figure demonstrates that the behaviour of the eigenvalues depends on parameter values (compare with Figure 3 in the **MP**).

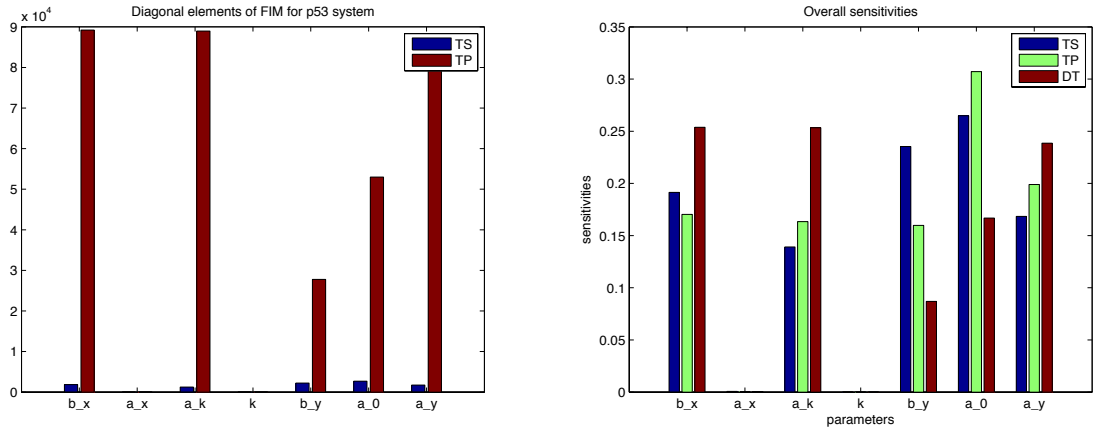

Figure 12: **Left:** Diagonal elements of FIM for TS and TP versions of p53 model. **Right:** Sensitivity coefficients $\mathcal{T}_i$ for TS, TP, DT version of p53 model. FIMs were calculated for parameters presented in Table 4 except $\alpha_y = 2$.

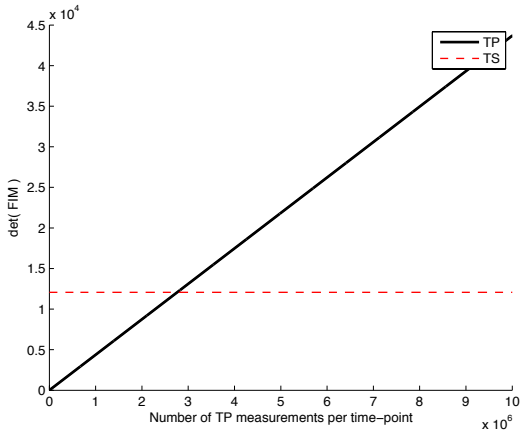

Figure 13: Comparison of informational content of TP and TS samples for p53 model similarly as Figure 2 but with $\alpha_y = 2$ instead of $\alpha_y = 0.8$.