Biochemical networks with uncertain parameters

W. Liebermeister and E. Klipp

Abstract: The modelling of biochemical networks becomes delicate if kinetic parameters are varying, uncertain or unknown. Facing this situation, we quantify uncertain knowledge or beliefs about parameters by probability distributions. We show how parameter distributions can be used to infer probabilistic statements about dynamic network properties, such as steady-state fluxes and concentrations, signal characteristics or control coefficients. The parameter distributions can also serve as priors in Bayesian statistical analysis. We propose a graphical scheme, the 'dependence graph', to bring out known dependencies between parameters, for instance, due to the equilibrium constants. If a parameter distribution is narrow, the resulting distribution of the variables can be computed by expanding them around a set of mean parameter values. We compute the distributions of concentrations, fluxes and probabilities for qualitative variables such as flux directions. The probabilistic framework allows the study of metabolic correlations, and it provides simple measures of variability and stochastic sensitivity. It also shows clearly how the variability of biological systems is related to the metabolic response coefficients.

1 Introduction

Cell simulations aspired to in systems biology [1] require knowledge of enzyme kinetic parameters. However, owing to a lack of measurements, measurement errors and biological variability, most of these parameters are still unknown or uncertain, which turns out to be a major obstacle in large-scale cell modelling. In this situation, a probabilistic description of the parameters can be helpful: to assess the effects of measurement errors, to find out which model results persist for a wide range of parameters, and to derive probabilities for different model outcomes. Besides this, parameter distributions can also be employed to study the natural variability and robustness of biological systems.

The effects of temporal random fluctuations in gene expression [2, 3] and metabolism [4, 5] have been studied. Here, we focus on models with static yet uncertain parameters: the parameters are described by a probability distribution, and the standard deviation of the resulting variable distributions (their variability) reflects how strongly the variables respond to parameter variations. (In this paper, we use the term 'variable' quite generally for quantitative model results, such as concentrations or fluxes in steady state or as functions of time, or functions of them, such as signal amplitudes or durations [6].) Of course, this influence depends on the system, on the variable of interest and on the parameter: at bifurcation points, a small parameter change can even change the qualitative dynamic behaviour. In other cases, parameters can have a weak influence on the system behaviour. In fact, various biological systems seem to have evolved robustness, that is, low sensitivity and thus low variability, against a typical amount of parameter variation (see [7] and references therein) [8–10].

If the parameter variability is fairly small, the width of variable distributions can be computed by expansion of the variables around the mean parameter values. This expansion has been applied [11, 12] to study the uncertainty of control coefficients due to errors in reaction elasticities. The variability of a variable depends on its parameter sensitivities, the so-called metabolic response coefficients. Variability measures based on the normalised control coefficients have been proposed [13, 14], and, for larger parameter variances, Monte Carlo simulations with randomly generated parameter sets have been used to study the variability of control coefficients and other variables [10, 15]. Brown and Sethna [16] have explored parameter distributions related to the likelihood function, which describes the ability of parameter sets to explain experimental data.

In this paper, we address two practical questions arising with the parameter distributions

(*a*) Sometimes, the kinetic parameters cannot be chosen separately: they are statistically dependent, because of thermodynamic constraints or because they depend on another common parameter. In Section 2, we shall show how to define parameter distributions that account for such thermodynamic and biochemical knowledge.

(*b*) Later, we shall take a closer look at the expansion method: in Section 3, we derive first- and second-order approximations for average values and covariances, compute the distributions of metabolites and fluxes and derive probabilities for qualitative statements, for instance, about flux directions. In Section 4, we show how the expansion method can be used for defining stochastic sensitivity measures, to quantify the control of regulatory parameters on the variability of system variables, and for Bayesian parameter estimation. Three illustrative examples are presented in Section 5.

Mathematical notation: Vectors are given in bold. The symbol ln denotes the natural logarithm. Given a vector $\boldsymbol{x}$, $\ln \boldsymbol{x}$ is short for the vector $(\ln x_1, \ln x_2, \ldots)^T$. A log-normal random variable x is characterised by the mean

© IEE, 2005

IEE Proceedings online no. 20045033

doi:10.1049/ip-syb:20045033

Paper first received 16th November 2004 and in final revised form 12th July 2005

The authors are with the Max Planck Institute for Molecular Genetics, Ihnestraße 73, 14195 Berlin, Germany

E-mail: lieberme@molgen.mpg.de

$\mu_{\ln x} = \langle \ln x \rangle$ and the variance $\sigma^2_{\ln x} = \mathrm{var}(\ln x)$. A list of symbols and the derivations of equations (20)–(22), (25), (29), (31), (45) and (46) can be found in the web supplement.

2 Distributions of the kinetic parameters

2.1 Parameter distributions

Biochemical reaction systems implement both metabolism and signalling in cells and have been widely studied [17, 18]. Their description comprises two parts: the stoichiometric structure and the individual reaction kinetics. Extensive data on network structures have been collected in databases [19–22], whereas quantitative knowledge about reaction kinetics [23] is less detailed, and kinetic parameters can also vary strongly between experiments.

In the following, we shall assume that the stoichiometrical structure of a biochemical network is known and fixed, and the kinetic parameters (or some of them) are constant in time but unknown, uncertain, or varying from situation to situation. Thermodynamic constraints by experimentally [24] or computationally [25] determined equilibrium constants can limit the emerging parameter combinations. Uncertainty can arise for different reasons

(i) a parameter value is known, but with certain measurement errors

(ii) a value is known from a different experiment, and so biological variability will lead to uncertainty: for instance, enzyme activities are actively adapted to the demands of the cell by differential gene expression; in particular, parameters will vary within a population of cells

(iii) a parameter value is unknown, but a rough guess can be made from the range of known parameter values of the same type. For instance, inhibition constants in the Brenda database [23] typically range between 10^{-6} and 100 mM, K_M values lie between 10^{-4} and 100 mM, and turnover numbers vary between 0.01 and 10^6 min^{-1}.

For modelling, we shall quantify our belief in certain parameter values by a probability distribution, characterised by the mean values $\langle \ln \boldsymbol{p} \rangle = \ln \boldsymbol{p}^0$ and the covariance matrix[1] $\mathrm{cov}(\ln \boldsymbol{p})$ of the logarithmic parameters (working with logarithms requires dimensionless positive quantities; for instance, concentrations have to be divided by their unit (moles l^{-1})). We shall call $\boldsymbol{p}^0$ the reference parameter values. Most results derived in this paper remain valid for all parameter distributions of finite variance, but, to be specific, we assume that the parameters follow log-normal distributions.

By definition, a random variable x is called log-normal if its logarithm follows a normal distribution. A joint log-normal distribution of many parameters is fully specified by $\ln \boldsymbol{p}^0$ and $\mathrm{cov}(\ln \boldsymbol{p})$. Mean and variance of the non-logarithmic random variable x can be computed using

$$\mu_x = \exp\left(\mu_{\ln x} + \frac{\sigma^2_{\ln x}}{2}\right),$$

$$\sigma^2_x = (\exp \sigma^2_{\ln x} - 1) \exp(2\mu_{\ln x} + \sigma^2_{\ln x})$$

The diagonal of the covariance matrix contains the variances of the (logarithmic) parameters, and the off-diagonal elements describe the dependencies between them.

[1]The covariance matrix of a random vector $\boldsymbol{x}$ defined as $\mathrm{cov}(\boldsymbol{x}) = \langle (\boldsymbol{x} - \langle \boldsymbol{x} \rangle)(\boldsymbol{x} - \langle \boldsymbol{x} \rangle)^T \rangle$.

2.2 Thermodynamic constraints lead to correlated parameters

If the kinetic parameters are statistically independent, we can characterise each of them by the mean value $\langle \ln p_i \rangle$ and the variance $\mathrm{var}(\ln p_i)$ of its logarithm. In general, however, parameters will depend on each other: for instance, the maximum forward and backward velocities of a reaction are proportional to the enzyme activity and are therefore correlated. Another source of dependencies is the equilibrium constants, which denote the ratio of product and substrate concentrations in thermal equilibrium. They imply a relationship between the kinetic parameters that can be found by equating the kinetics function to zero. From mass-action kinetics with the rate law $v = k_+[S] - k_-[P]$ follows

$$q = [P]/[S] = k_+/k_- \tag{1}$$

whereas, for the reversible Michaelis–Menten kinetics,

$$v = \frac{(V^+_{\max}/K^+_M)[S] - (V^-_{\max}/K^-_M)[P]}{1 + [S]/K^+_M + [P]/K^-_M} \tag{2}$$

q fulfils the Haldane relationship

$$q = \frac{V^+_{\max} K^-_M}{V^-_{\max} K^+_M} \tag{3}$$

To see how a known equilibrium constant leads to off-diagonal elements in the covariance matrix, we consider a single reaction $S \leftrightarrow P$ with mass-action kinetics. If the rate constants k_+ and k_- are independent log-normal, the covariance matrix

$$\mathrm{cov}(\ln \boldsymbol{k}) = \begin{pmatrix} \sigma^2_{\ln k_+} & 0 \\ 0 & \sigma^2_{\ln k_-} \end{pmatrix} \tag{4}$$

of the parameter vector $\ln \boldsymbol{k} = (\ln k_+, \ln k_-)^T$ is diagonal, and the equilibrium constant $q = k_+/k_-$ is a log-normal random variable. Prescribing a value for q, on the other hand, implies a linear relationship between $\ln k_+ - \ln k_- = \ln q$, and the covariance matrix becomes

$$\mathrm{cov}(\ln \boldsymbol{k}) = \begin{pmatrix} \sigma^2_{\ln k} & \sigma^2_{\ln k} \\ \sigma^2_{\ln k} & \sigma^2_{\ln k} \end{pmatrix} \tag{5}$$

where $\sigma^2_{\ln k} = \sigma^2_{\ln k_+} = \sigma^2_{\ln k_-}$. Note that the two parameters $\ln k_+$ and $\ln k_-$ must have the same variance, although, experimentally, their values might be determined with different accuracies.

Accounting for (1) is important not only if the equilibrium constants are known, but also to fulfil the Wegscheider condition for metabolic networks: the product of equilibrium constants over each closed path in the network must be equal to one. The reason for this is that the equilibrium constant q is related to the difference of the Gibbs free energies by

$$q = e^{-\beta \Delta g} \tag{6}$$

Δg is the free energy difference between products and substrates (measured in J mole^{-1}), the equilibrium constant q is measured in l mole^{-1}, $\beta = (kT)^{-1}$, T is the absolute temperature, and $k \approx 1.38 \times 10^{-23}$ J K^{-1} is Boltzmann's constant.

An independent choice of the kinetic constants would correspond to statistically independent equilibrium constants and thus violate the Wegscheider condition.

2.3 The dependence graph

We shall now construct parameter distributions that are consistent with thermodynamic and other knowledge: practically, we compute the mean values and covariance matrix of all (logarithmic) kinetic parameters in a biochemical

network. The basic idea is to express the kinetic parameters (e.g. rate constants) by other, underlying parameters (e.g. equilibrium constants) until we reach a set of basic parameters that can be chosen independently. For instance, to fulfil the Wegscheider condition, we may choose independent random values for the Gibbs free energies and use them to compute the enzyme kinetic parameters.

The dependencies and constraints among the different types of parameter can be depicted in an acyclic dependence graph. An example is shown in Fig. 1: parameters with incoming arrows are computed from their 'parent' parameters, whereas the basic parameters (without parents) are either exactly known or chosen independently from a log-normal distribution. Of course, the choice of independence assumptions may have a strong impact on the modelling results. It is important to note the independencies are not meant to be biological facts, but modelling assumptions, and the dependence graph is supposed to make them explicit.

The direct dependencies depicted in the graph can be used to compute the average values and covariance matrix of logarithmic parameters. To illustrate this, we consider a chemical reaction denoted by i, with mass-action kinetics: the forward and backward rate constants can be expressed as

$$\begin{aligned} k_{+i} &= r_i u_i \sqrt{q_i} \\ k_{-i} &= r_i u_i / \sqrt{q_i} \end{aligned} \tag{7}$$

where q_i is the equilibrium constant and u_i is the enzyme activity. The prefactor r_i is related to the free energy barrier of the catalysed reaction, and we can decide to choose it independently for each reaction. The choice (7) fulfils the equilibrium condition $q_i = k_{+i}/k_{-i}$, and both forward and backward rate constants are proportional to the enzyme activity u_i.

The kinetic parameters are related in a multiplicative fashion, which means that their logarithmic values are additive

$$\begin{aligned} \ln k_+ &= \ln r + \ln u + \frac{1}{2}\ln q \\ \ln k_- &= \ln r + \ln u - \frac{1}{2}\ln q \end{aligned} \tag{8}$$

or, in vector notation,

$$\ln \boldsymbol{k} = \ln u \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\ln q \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \ln r \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{9}$$

This linear relationship makes it easy to compute the means and covariance matrix of $\ln \boldsymbol{k} = \begin{pmatrix} \ln k_{+i} \\ \ln k_{-i} \end{pmatrix}$ from the means and variance of $\ln r_i$ and $\ln u_i$. Moreover, if $\ln r_i$ and $\ln u_i$ are normally distributed, then $\ln \boldsymbol{k}$ is also normal.

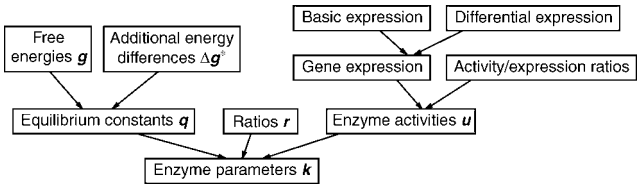

Fig. 1 *Dependence graph for kinetic parameters*

Arrows denote relationships between parameters: for instance, enzyme parameters k depend on equilibrium constants, enzyme activities and additional ratio parameters

Gibbs free energies g_i can be represented by $\exp(g_i)$: then all relationships in dependence graph are multiplicative, and all parameters can be consistently described by log-normal distributions. Together with dependence coefficients characterising individual arrows (see text), this dependence graph defines joint distribution of all parameters

Other relationships in the dependence graph can also be stated in a linear form: the vector $\boldsymbol{g}$ (top left in Fig. 1) contains the Gibbs free energies of all metabolites, and $\Delta\boldsymbol{g}$ contains the energy differences of all reactions. They are related by

$$\ln \boldsymbol{q} = -\beta \Delta \boldsymbol{g} = -\beta \boldsymbol{N}^T \boldsymbol{g} \tag{10}$$

and so the energies can be modelled by a normal distribution. The first equality stems from (6), and the second equality relates the energy differences to the energies and the stoichiometric matrix $\boldsymbol{N}$. For example, the energy balance for a reaction $2A \rightarrow B + C$ can be written $\Delta g = -2g_A + g_B + g_C = \boldsymbol{N}(g_A, g_B, g_C)^T$, where $\boldsymbol{N} = (-2, 1, 1)$ is the stoichiometric matrix.

Sometimes, reactions are known to be irreversible under physiological conditions, that is, the fluxes have to be positive. However, the flux directions depend on an interplay between enzyme kinetic parameters and external metabolite concentrations, so that enforcing them by a proper choice of the parameter distribution is not straightforward: for practical purposes, however, we can bias the flux direction by adding a term $-\beta\Delta g_i^*$ to the logarithmic equilibrium constant $\ln q_i$. In fact, such terms also appear if there are cofactors (such as ATP and ADP) that contribute to the energy balance but do not explicitly appear in the reaction equation.

The logarithmic enzyme activities u_i can be split into a sum of terms describing base level gene expression, differential gene expression and a prefactor, which may, among other things, account for the correlation between mRNA expression and enzyme concentration [26, 27]. Other relationships could be incorporated into the dependence graph: differential gene expression can be described with separate basic parameters accounting for extrinsic and intrinsic noise [2, 3]. Known empirical correlations among external metabolites [28] can also be employed. In addition, model parameters that are not restricted by constraints can depend on a log-normal random parameter describing multiplicative measurement noise.

It is not always possible to obtain linear relationships among the logarithmic parameters: instead of using (7), the rate constants can also be parametrised by the relaxation time $\tau = (k_+ + k_-)^{-1}$. In this case, we can set

$$\begin{aligned} k_{+i} &= \tau_i^{-1} q_i/(q_i + 1) \\ k_{-i} &= \tau_i^{-1}/(q_i + 1) \end{aligned}$$

where τ_i is log-normal, drawn independently for each reaction. However, the logarithm of this formula is not linear in q_i, and thus this parametrisation is only useful if the equilibrium constants q_i are exactly known.

2.4 Computing the parameter distribution

Equation (8) states the relationship between equilibrium constants, enzyme activities and kinetic parameters for mass-action kinetics. For Michaelis–Menten kinetics and other rate laws, we demand that, for each reaction i, the parameter vector $\boldsymbol{k}^{(i)}$ can be written in the form

$$\ln \boldsymbol{k}^{(i)} = \boldsymbol{a}^{(i)} \ln u_i + \boldsymbol{b}^{(i)} \ln q_i + \boldsymbol{c}^{(i)} + \boldsymbol{D}^{(i)} \ln \boldsymbol{z}^{(i)} \tag{11}$$

where $\ln \boldsymbol{z}^{(i)}$ is a vector of independent standard normal random variables. The dependence coefficients in the

vectors $\boldsymbol{a}^{(i)}$, $\boldsymbol{b}^{(i)}$, $\boldsymbol{c}^{(i)}$ and in the matrix $\boldsymbol{D}^{(i)}$ follow from independence assumptions and from the kinetic law. Possible choices for some common rate laws are listed in Appendix 1.

To compute the parameter distribution for the entire network, we merge all $\boldsymbol{k}^{(i)}$ into a large vector $\boldsymbol{k}$. Accordingly, $\boldsymbol{a}^{(i)}$, $\boldsymbol{b}^{(i)}$ and $\boldsymbol{D}^{(i)}$ are merged into block matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$, and the vectors $\boldsymbol{c}^{(i)}$ and $\boldsymbol{z}^{(i)}$ are merged into vectors $\boldsymbol{c}$ and $\boldsymbol{z}$. Altogether, we obtain

$$\ln \boldsymbol{k} = \boldsymbol{A} \ln \boldsymbol{u} + \boldsymbol{B} \ln \boldsymbol{q} + \boldsymbol{c} + \boldsymbol{D} \ln \boldsymbol{z} \tag{12}$$

For instance, the block matrices for a system of two reactions read

$$\begin{pmatrix} \ln \boldsymbol{k}^{(1)} \\ \ln \boldsymbol{k}^{(2)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}^{(1)} & \\ & \boldsymbol{a}^{(2)} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{b}^{(1)} & \\ & \boldsymbol{b}^{(2)} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{c}^{(1)} \\ \boldsymbol{c}^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{D}^{(1)} & \\ & \boldsymbol{D}^{(2)} \end{pmatrix} \begin{pmatrix} \ln \boldsymbol{z}^{(1)} \\ \ln \boldsymbol{z}^{(2)} \end{pmatrix} \tag{13}$$

Note that (9) is a special case of (12). As $\ln \boldsymbol{q}$, $\ln \boldsymbol{u}$ and $\ln \boldsymbol{z}$ are mutually independent, and $\mathrm{cov}(\ln \boldsymbol{z}) = \boldsymbol{I}$ (where $\boldsymbol{I}$ is the identity matrix) the mean values and covariances of all kinetic parameters read

$$\begin{aligned} \langle \ln \boldsymbol{k} \rangle &= \boldsymbol{A} \langle \ln \boldsymbol{u} \rangle + \boldsymbol{B} \langle \ln \boldsymbol{q} \rangle + \boldsymbol{c} \\ \mathrm{cov}(\ln \boldsymbol{k}) &= \boldsymbol{A}\ \mathrm{cov}(\ln \boldsymbol{u}) \boldsymbol{A}^T + \boldsymbol{B}\ \mathrm{cov}(\ln \boldsymbol{q}) \boldsymbol{B}^T + \boldsymbol{D}\boldsymbol{D}^T \end{aligned} \tag{14}$$

Based on (10), $\mathrm{cov}(\ln \boldsymbol{q})$ can be expressed as $\beta^2 \boldsymbol{N}^T\, \mathrm{cov}(\boldsymbol{g}) \boldsymbol{N}$.

Let us summarise this section: we have described dependencies among model parameters by linear relationships between their logarithms. Thereby we can easily propagate uncertainty from independent basic parameters to all model parameters and, at the same time, also obtain the covariances between them.

3 Distribution of model variables

We have seen how to obtain a log-normal distribution of the kinetic parameters $\boldsymbol{p}$. To infer probabilistic statements about the variables, we have to compute their distributions from the parameter distribution. In this section, we shall first consider continuous variables y, which can describe concentrations, fluxes and control coefficients in steady state, but also time-dependent concentrations or signal characteristics such as signal durations or amplitudes [6]. Later, we examine binary variables, such as the signs of fluxes or the relative order of metabolite concentrations.

3.1 Expansion using response coefficients

If the parameter distribution is narrow, then the variable distributions can be approximated by a linear or quadratic expansion around the reference parameters $\boldsymbol{p}^0$. Fig. 2 illustrates the linear expansion for one parameter p and one variable y. The logarithm $\ln p$ is normally distributed with the standard deviation σ: if $\ln y(\ln p)$ is linearised around $\ln p^0$, then $\ln y$ is also normal, and the standard deviation of $\ln y$ is σR, where R is the slope of the tangent.

This expansion also works if $\boldsymbol{p}$ and $\boldsymbol{y}(\boldsymbol{p})$ are vectors, and the slopes of $\ln y_i$ are the response coefficients defined in metabolic control analysis [17, 29]. Let us briefly recapitulate a few definitions: the non-normalised response

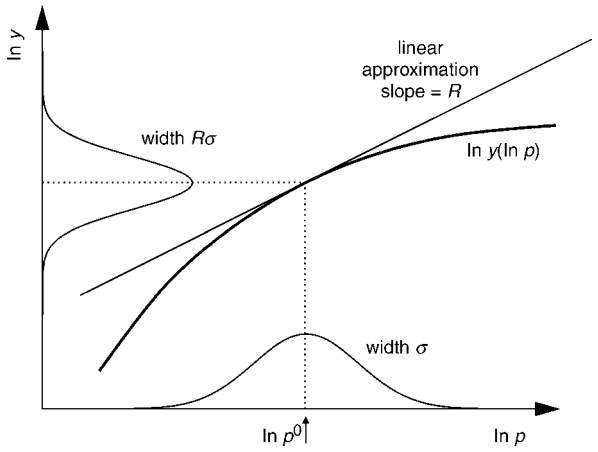

Fig. 2 *Distribution of variables approximated using linear expansion*

Solid curve: variable y as function of parameter p, both in log-scale Probability density of $\ln p$, with mean $\ln p^0$ and width σ, is shown at bottom. Logarithmic variable $\ln y(\ln p)$ is linearly expanded around $\ln p^0$ (straight line), and its slope is called response coefficient R. Linearised $\ln y$ follows normal distribution with width $R\sigma$

coefficients are defined as the derivatives

$$\begin{aligned} \bar{R}_{lm} &:= \frac{\partial y_l(\boldsymbol{p})}{\partial p_m} \\ \bar{R}^{(2)}_{lmn} &:= \frac{\partial^2 y_l(\boldsymbol{p})}{\partial p_m \partial p_n} \end{aligned} \tag{15}$$

where $y_l(\boldsymbol{p})$ is a (stationary or time-dependent) variable of the system. Response coefficients for stationary fluxes and concentrations can be computed analytically from the stoichiometric matrix and the reaction elasticities, which describe the linearised rate laws. The normalised response coefficients read

$$R_{ik} = \frac{\partial \ln y_l}{\partial \ln p_m} = y_i^{-1} \bar{R}_{ik} p_k \tag{16}$$

$$R^{(2)}_{ikl} = \frac{\partial^2 \ln y_l}{\partial \ln p_m\, \partial \ln p_n} = y_i^{-1} \bar{R}^{(2)}_{ikl}\, p_k p_l + R_{ik} \delta_{kl} - R_{ik} R_{il} \tag{17}$$

where Kronecker's δ_{ik} denotes the elements of the identity matrix. The logarithmic variables $\ln y$ can be expanded using normalised response coefficients

$$\ln y_l(\ln \boldsymbol{p}^0 + \Delta \ln \boldsymbol{p}) \approx \ln y_l(\ln \boldsymbol{p}^0) + \sum_m R_{lm} \Delta \ln p_m + \frac{1}{2} \sum_{mn} R^{(2)}_{lmn} \Delta \ln p_m \Delta \ln p_n \tag{18}$$

plus higher-order terms. It should be noted that the expansion of logarithmic fluxes and concentrations does not exactly obey stationarity conditions and conservation relationships. This problem does not occur with a linear expansion of the non-logarithmic variables.

Expansion (18) can be used to compute the distribution of variables. Again, we assume that the parameters have mean $\langle \ln \boldsymbol{p} \rangle = \ln \boldsymbol{p}^0$ and covariance matrix $\mathrm{cov}(\ln \boldsymbol{p})$ and that the variables $\boldsymbol{y}(\boldsymbol{p})$ are continuous, positive and differentiable with respect to the parameters. Also, for large $\ln \boldsymbol{p}$, $|\ln y(\ln \boldsymbol{p})|$ must not increase faster than exponentially. Then, to first order, the logarithms $\ln \boldsymbol{y}$ are normally distributed with

mean and covariances

$$\langle \ln y_l \rangle = \ln(y_l(\boldsymbol{p}^0)) \tag{19}$$

$$\mathrm{cov}(\ln \boldsymbol{y}) = \boldsymbol{R}\,\mathrm{cov}(\ln \boldsymbol{p})\boldsymbol{R}^T \tag{20}$$

Note the similarity to (14): just as within the dependence graph, uncertainty is now propagated towards the model variables. If $\ln \boldsymbol{y}$ is expanded to second order, its mean and covariance read

$$\langle \ln y_l \rangle = \ln(y_l(\boldsymbol{p}^0)) + \frac{1}{2}\sum_{mn} R^{(2)}_{lmn} C_{mn} \tag{21}$$

$$\begin{aligned}\mathrm{cov}(\Delta \ln y_l, \Delta \ln y_k) &= \sum_{mn} R_{lm} C_{mn} R_{kn} \\ &\quad + \frac{1}{4}\sum_{mnrs} R^{(2)}_{lmn} R^{(2)}_{krs}(C_{ms}C_{nr} + C_{mr}C_{ns})\end{aligned} \tag{22}$$

where $\boldsymbol{C} = \mathrm{cov}(\ln \boldsymbol{p})$. The second-order term reflects the curvature of $\ln \boldsymbol{y}(\ln \boldsymbol{p})$ (compare Fig. 2). It implies that, to second order, the average model results cannot be obtained from a single model with average parameters. In the second-order approximation, the distribution of y_l will no longer be log-normal, but, by using (21) and (22), we can approximate it by a normal distribution.

3.2 Change of flux sign

The log-normal distribution describes variables with positive values. For negative variables, we can simply consider the absolute value $|y_i|$, but, if a variable can change its sign, then a log-normal distribution may be a poor approximation. In the important case of metabolic fluxes, this problem can be resolved: we compute the distribution of the total flux from the joint distribution of the partial fluxes in the forward and backward directions. Let us denote the forward and backward fluxes by y_1 and y_2, respectively. We assume that the joint distribution of their logarithms is bivariate normal with means $\langle \ln y_1 \rangle$ and $\langle \ln y_2 \rangle$ and covariance matrix $\boldsymbol{C} = \mathrm{cov}(\ln \boldsymbol{y})$ computed from (19) and (20) (first-order approximation) or (21) and (22) (second-order approximation). Thus the joint probability density of y_1 and y_2 reads

$$\begin{aligned}p(y_1, y_2) = \frac{1}{2\pi|\boldsymbol{C}|y_1 y_2}\exp\Bigg(-\frac{1}{2}(\ln \boldsymbol{y} - \langle \ln \boldsymbol{y} \rangle)^T \\ \times \boldsymbol{C}^{-1}(\ln \boldsymbol{y} - \langle \ln \boldsymbol{y} \rangle)\Bigg)\end{aligned} \tag{23}$$

The probability density $p(J)$ of the total flux J can be computed numerically by integration over all pairs (y_1, y_2), fulfilling $J = y_1 - y_2$. The resulting distribution describes both positive and negative total fluxes (compare the example in Section 5.1). Note that this procedure is not restricted to fluxes, but applies to all variables that can be written as a function of two log-normal variables.

3.3 Qualitative variables: thresholds and order relationships

Let us finally consider binary variables describing signs or order relationships of log-normal variables ('flux i is positive', 'control coefficient a is larger than control coefficient b'). Given a parameter distribution, each of them can be assigned a probability, and, to compute it, we can make use of the means and covariances computed above.

The probability that a log-normal variable y exceeds a given threshold $a > 0$ is given by

$$\mathrm{Prob}(y > a) = \Phi\left(\frac{\langle \ln y \rangle - \ln a}{\mathrm{var}(\ln y)^{1/2}}\right) \tag{24}$$

where $\Phi(\cdot)$ is the cumulative density of the standard normal distribution. Likewise, we can compare two variables: for instance, consider again the forward and backward flux velocities y_1 and y_2 of a chemical reaction. If the vector $\ln \boldsymbol{y} = (\ln y_1, \ln y_2)^T$ follows a bivariate normal distribution with mean $\langle \ln \boldsymbol{y} \rangle$ and covariance matrix $\mathrm{cov}(\ln \boldsymbol{y})$, the probability for a positive total flux is given by

$$\begin{aligned}&\mathrm{Prob}(y_1 > y_2) \\ &= \Phi\left(\frac{\langle \ln y_1 \rangle - \langle \ln y_2 \rangle}{(\mathrm{var}(\ln y_1) + \mathrm{var}(\ln y_2) - 2\ \mathrm{cov}(\ln y_1, \ln y_2))^{1/2}}\right)\end{aligned} \tag{25}$$

Likewise, we obtain

$$\begin{aligned}&\mathrm{Prob}(y_1/y_2 > a) \\ &= \Phi\left(\frac{\langle \ln y_1 \rangle - \langle \ln y_2 \rangle - \ln a}{(\mathrm{var}(\ln y_1) + \mathrm{var}(\ln y_2) - 2\ \mathrm{cov}(\ln y_1, \ln y_2))^{1/2}}\right)\end{aligned} \tag{26}$$

which can be seen by replacing y_2 by $a\,y_2$ in (25). Formulae (25) and (26) hold for all kinds of log-normal variables.

4 Further applications of linear expansion

Equation (20) constitutes a simple relationship between parameter distributions, response coefficients and variable distributions. We shall now exploit the linear expansion to derive formulae for the variabilities and sensitivities of model variables. In addition, we shall study, within the same approximation, how observations of the model variables can be used to update a parameter distribution.

4.1 Contributions to variability

The width of a variable distribution can be measured by the variance $\sigma^2_{\ln y}$, which resembles previously proposed variability measures [13, 14] that are based on the normalised control coefficients. For log-normal distributions with variance $\sigma^2_{\ln y}$, the coefficient of variation $\sigma_y/\langle y \rangle$ reads $(\exp(\sigma^2_{\ln y}) - 1)^{1/2}$. Both variability measures depend only on the shape of the distribution and not on the absolute scaling of y. Equation (20) shows how individual parameters contribute to the variability: if they are independent, then $\mathrm{cov}(\ln \boldsymbol{p})$ is diagonal, and (20) can be split into a sum

$$\mathrm{cov}(\ln \boldsymbol{y})_{ik} = \sum_m R^Y_{im}\mathrm{var}(\ln p_m)R^Y_{km} \tag{27}$$

of variabilities caused by the individual parameters. The same holds if certain groups of parameters are mutually independent: for instance, let us recall (8) for the rate constants. If equilibrium constants, enzyme activities and random parameters are chosen independently, the total covariance can be split into

$$\begin{aligned}\mathrm{cov}(\ln \boldsymbol{y}) = \boldsymbol{R}_u\,\mathrm{cov}(\ln \boldsymbol{u})\boldsymbol{R}_u^T + \boldsymbol{R}_q\,\mathrm{cov}(\ln \boldsymbol{q})\boldsymbol{R}_q^T \\ + \boldsymbol{R}_r\,\mathrm{cov}(\ln \boldsymbol{r})\boldsymbol{R}_r^T\end{aligned}$$

which can be computed and analysed separately. The response coefficient matrices $\boldsymbol{R}_u$, $\boldsymbol{R}_q$ and $\boldsymbol{R}_r$ can be easily

computed using the matrices defined in Section 2.4: for instance, the response to changes in the enzmye activities is described by $\boldsymbol{R}_u = \boldsymbol{RA}$.

4.2 Stochastic sensitivity

Sensitivities describe how a variable is affected by individual parameters. In a deterministic setting, only one parameter is varied while all other parameters are kept fixed. For small parameter variation, sensitivity is captured by the response coefficients R. In the probabilistic setting, we suggest two sensitivity measures, and, in both cases, the first-order approximation from Section 3.1 leads to simple formulae. Krewski *et al.* [30] proposed a probabilistic sensitivity measure for pharmacokinetic models: Let $f(\boldsymbol{x})$ denote a random variable that depends on the (possibly correlated) random variables $x_1, \ldots, x_M$, called here the parameters. The influence of a parameter x_m on f, averaged over values of the other parameters, is measured by

$$U(f|x_m) := \mathrm{var}_{x'_m}(\mathcal{E}(f|x_m = x'_m)) \tag{28}$$

$\mathcal{E}(f|x_m = x'_m)$ denotes the expectation value of f where the value of x_m is fixed, and all other parameters are drawn from their distribution conditional on $x_m = x'_m$. This sensitivity measure can be rewritten as $U(f|x_m) = \mathrm{var}(f) - \mathcal{E}_{x_m}(\mathrm{var}(f|x_m = x'_m))$.

The Monte Carlo algorithm (also in [30]) to compute this value is numerically demanding, but we can also approximate it using the linear expansion, yielding

$$U(\ln y_l | \ln p_i) \approx \left(\sum_k R_{lk} \,\mathrm{cov}(\ln p_k, \ln p_i) \right)^2 \Big/ \mathrm{var}(\ln p_i) \tag{29}$$

This sensitivity measure does not distinguish whether a parameter has an increasing or decreasing effect on a variable. To account for the sign of the response, we can use the covariance between the parameter and the variable as a sensitivity measure. Under the above assumptions, it is approximated by

$$\mathrm{cov}(\ln y_l, \ln p_i) \approx \sum_k R_{lk} \,\mathrm{cov}(\ln p_k, \ln p_i) \tag{30}$$

4.3 Posterior distribution

Parameter distributions as defined in Section 2 can also be used as prior distributions for Bayesian parameter estimation [31]. Posterior distributions for biochemical models with poorly determined parameters have been studied [16] by Monte Carlo methods. Here, we give a simple approximative formula for the posterior distribution, based on the linear expansion around the maximum of the prior.

Initially, the parameters are described by a log-normal prior distribution with mean $\ln \boldsymbol{p}^0$ and covariance matrix $\mathrm{cov}(\ln \boldsymbol{p}) = \boldsymbol{C}$. Now suppose that we have collected experimental data for some of the variables, contained in a vector $\boldsymbol{y}^{exp}$. The elements of $\ln \boldsymbol{y}$ and the rows of $\boldsymbol{R}$, which are computed from the mathematical model, correspond to the same variables, in the same order. We assume that the experimental data are given by $\ln \boldsymbol{y}^{exp} = \ln \boldsymbol{y} + \sigma\eta$, where the elements of the vector η are independent standard Gaussian random variables. This error model assumes that the non-logarithmic data carry log-normal multiplicative errors. If the variables are linearly approximated around the reference parameters, we obtain an approximation for the posterior of $\ln \boldsymbol{p}$: it is again a normal distribution, with mean and covariance matrix given by

$$\langle \ln \boldsymbol{p} \rangle_{post} = (\sigma^{-2}\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{C}^{-1})^{-1}(\sigma^{-2}\boldsymbol{R}^T\boldsymbol{w} + \boldsymbol{C}^{-1} \ln \boldsymbol{p}^0) \tag{31}$$

$$\mathrm{cov}_{post}(\ln \boldsymbol{p}) = (\sigma^{-2}\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{C}^{-1})^{-1} \tag{32}$$

where $\boldsymbol{w} := \ln \boldsymbol{y}^{exp} - \ln \boldsymbol{y}(\ln \boldsymbol{p}^0) + \boldsymbol{R} \ln \boldsymbol{p}^0$. For a large measurement error width $\sigma \to \infty$, the experimental data become uninformative, and the posterior equals the prior. On the other hand, if the prior is very broad (all eigenvalues of $\boldsymbol{C} \to \infty$), we obtain

$$\langle \ln \boldsymbol{p} \rangle_{post} \approx (\boldsymbol{R}^T\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{w} \tag{33}$$

$$\mathrm{cov}_{post}(\boldsymbol{p}) \approx \sigma^{-2}\boldsymbol{R}^T\boldsymbol{R} \tag{34}$$

5 Examples

5.1 Steady states of a small network

We study the distribution of concentrations and fluxes in a small metabolic network. The hypothetical metabolic network shown in Fig. 3*a* consists of a closed loop (S_2, S_3, S_5, S_6) connected to external metabolites (X_1, X_4, X_7, X_8). We assume reversible Michaelis–Menten kinetics with a distribution of kinetic constants chosen as described in Section 2.4. The parameter distributions were chosen as follows: with fixed free energies and log-normal enzyme activities with $\langle \ln u \rangle = 0$, $\sigma_{\ln u} = \ln 2$, the distributions of the kinetic constants were computed as described in Appendix 1, with $\mu_K = 1$, $\sigma_K = \ln 2$, $\mu_V = 1$, $\sigma_V = \ln 2$ for all reactions. The reference concentrations of the external metabolites were $(1, 0.7, 0.3, 0.1)^T$: the external concentrations follow independent log normal distributions around these values, with a width of ln 2 of the logarithm.

The distributions computed from the expansion method agree well with the results of Monte Carlo simulation. Fig. 3 shows histograms of parameters, fluxes and concentrations from 1000 simulation runs: solid lines denote the corresponding probability densities, which are exact for the parameters and computed by second-order expansion (see Section 3.1) for fluxes and concentrations. The distributions of fluxes were computed numerically from the log-normal distribution of forward and backward fluxes using (23). The probabilities of flux directions are shown in Fig. 4. The size and direction of the arrowheads in Fig. 4*a* denote the probabilities of flux directions, as computed by (25). Fig. 4*b* shows that the expansion results (abscissa) resemble the results from 1000 Monte Carlo runs: dots and error bars indicate the means and standard deviations due to the finite-sample error.

5.2 Temporal behaviour of a linear reaction chain

To illustrate how the expansion method can be applied to time-dependent problems, we consider a linear chain of five chemical reactions

$$X_1 \to S_1 \to S_2 \to S_3 \to S_4 \to X_2$$

with reversible mass-action kinetics. The parameter vector $\boldsymbol{p}$ comprises the rate constants, the constant external concentrations x_1 and x_2, and the initial values of the internal concentrations s_1, s_2, s_3 and s_4. As reference parameter values (in arbitrary units), we chose a value of 1 for all rate constants and for x_1, and a value of 0.1 for x_2, $s_1(0)$, $s_2(0)$, $s_3(0)$ and

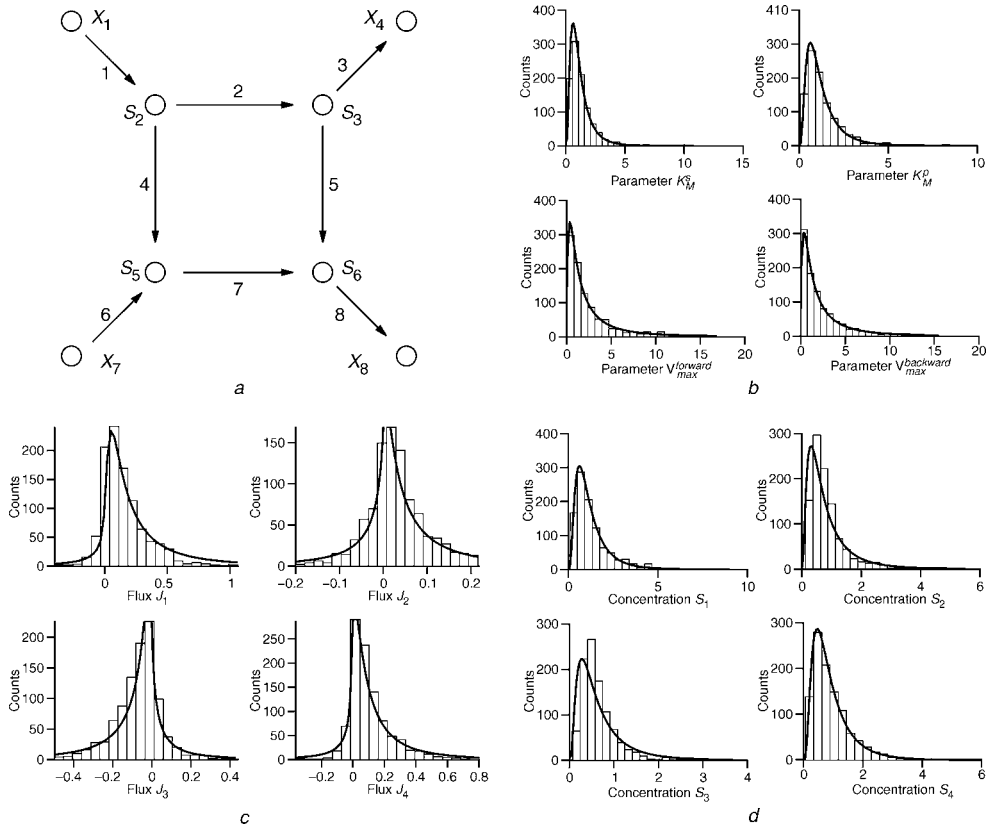

Fig. 3 *Stationary fluxes and concentrations in biochemical network*

a Network consists of loop of four reactions, connected to four external metabolites. Arrowheads show nominal reaction directions. Kinetic parameters (for Michaelis–Menten kinetics) were drawn from log-normal distribution that accounts for thermodynamic constraints
b Random samples of four kinetic parameters of reaction 1 (shown as histograms) follow prescribed probability density (solid line)
c Histograms of four stationary fluxes
d Histograms of internal concentrations. Solid lines show corresponding approximative solutions. Log-normal distributions of metabolites concentrations were computed by expansion method ((21) and (20)). Distributions of fluxes were computed from forward and backward reactions, as described in Section 3.2

$s_4(0)$. The parameters follow independent log-normal distributions around the reference values, with widths $\sigma_{\ln p} = 0.02$.

Fig. 5*a* shows concentration time courses from 100 Monte Carlo simulation runs. Fig. 5*b* shows the corresponding distributions, estimated from a first-order expansion. The mean curves (solid lines) result from the reference parameters. The variability was computed from (20): owing to the independent choice of parameters, the

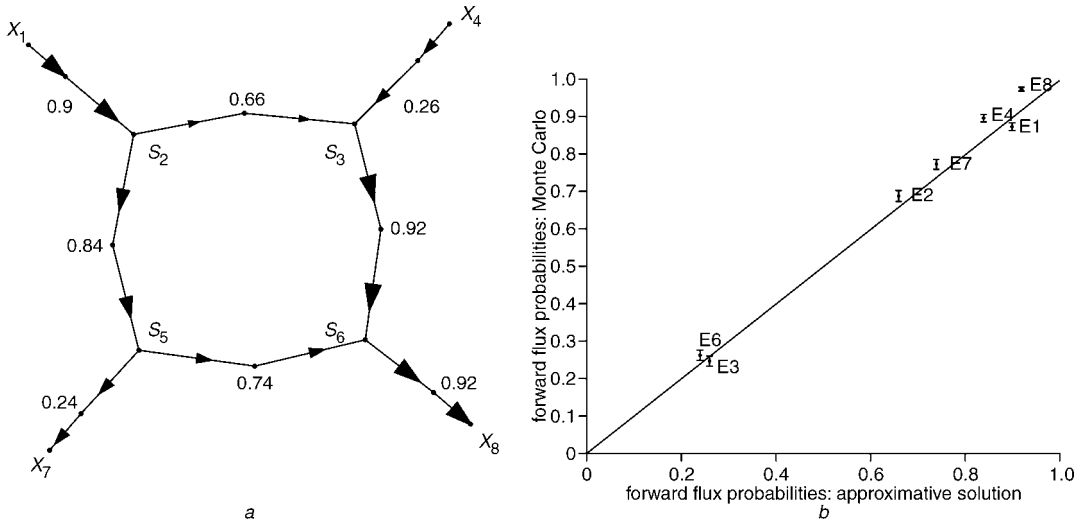

Fig. 4 *Probabilities of flux directions in loop network (compare Fig. 3)*

a Probabilities of flux directions (computed using expansion method, Section 3.3) are given by numbers and represented by sizes of arrowheads
b Results from expansion method (abscissa) and from Monte Carlo calculations (ordinate) are in good agreement. Error bars show estimation error due to finite number of 1000 Monte Carlo runs

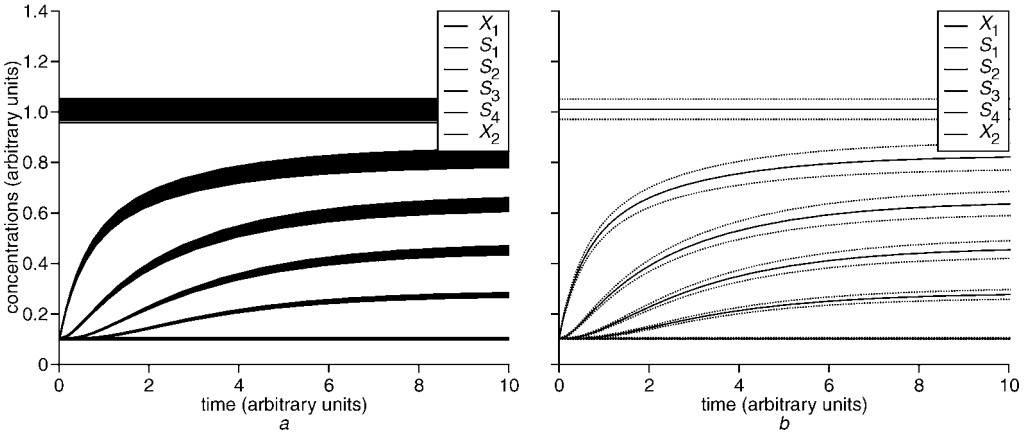

Fig. 5 *Variability of concentration time courses for linear chain of 5 reactions*

Model parameters (initial conditions and kinetic rate constants) were drawn from independent log-normal distribution with widths $\sigma_{\ln p} = 0.02$
a Results of 100 Monte Carlo runs
b Distribution of concentration time courses, computed using linear expansion around reference parameters
—— Reference parameter set $\boldsymbol{p}^0$; . . . . . Percentiles enclosing about 95% of all realisations of time courses

logarithmic concentrations have a standard deviation of

$$\sigma_{\ln s_i(t)} = \sigma_{\ln p}\, s_i(t)^{-1} \sum_k R^S_{ik}(t)\, p_k \tag{35}$$

The response coefficients $R^S_{ik}(t)$ describe how the concentration time courses depend on the parameters. They were computed numerically as described in [32]. With the normality assumption for $\ln s_i(t)$, about 95% of all observations should fall into an interval of two standard deviations $\sigma_{\ln s_i(t)}$ around the reference value, as indicated by the dotted curves.

5.3 Qualitative effects of differential gene expression

Finally, we outline a potential application of metabolic network modelling. The aim is to assess the effect of differential gene expression on the steady-state concentrations of metabolites, while the kinetic reaction parameters are poorly known. For the calculation, we consider two copies of the network corresponding to the two expression profiles: in network 1, the differential expression values vanish, whereas, in network 2, they are given by the (experimentally determined) differential expression profile $\Delta\boldsymbol{x}$. All other parameters $\boldsymbol{p}$ are uncertain (with reference values $\boldsymbol{p}^0$), but considered equal for both copies of the network. The metabolite concentration of interest is termed $y_1 = y(\ln \boldsymbol{p}, 0)$ under condition 1 (base expression) and $y_2 = (\ln \boldsymbol{p}, \Delta\boldsymbol{x})$ under condition 2 (altered expression). To calculate the joint distribution of y_1 and y_2, we first compute the concentrations $y_1^0 = y(\ln \boldsymbol{p}^0, 0)$ and $y_2^0 = y(\ln \boldsymbol{p}^0, \Delta\boldsymbol{x})$ for both networks at the reference values $\boldsymbol{p}^0$. Likewise, we compute the normalised response coefficients matrix $\boldsymbol{R}$ with two rows related to $\ln y_1$ and $\ln y_2$. The logarithmic ratio $\ln(y_2/y_1)$ is normally distributed with mean and variance

$$\left\langle \ln\frac{y_2}{y_1} \right\rangle = \ln y_2^0 - \ln y_1^0 \tag{36}$$

$$\begin{aligned}\operatorname{cov}\left(\ln\frac{y_2}{y_1}\right) &= \operatorname{var}(\ln y_1) + \operatorname{var}(\ln y_2) + 2\operatorname{cov}(\ln y_1, \ln y_2) \\ &\approx \sum_{ik} (\boldsymbol{R}\operatorname{cov}(\boldsymbol{p})(\boldsymbol{R})^T)_{ik}\end{aligned} \tag{37}$$

The probability that the ratio y_1/y_2 of concentrations exceeds some threshold a can be computed by (26).

6 Discussion

Models with uncertain parameters are useful for predictive simulation with uncertain knowledge, for correlation analysis of metabolic fluctuations, Bayesian parameter estimation, population models and for robustness studies. They fill a gap between kinetic modelling with fixed parameters and the algebraic methods [33–35] that study structural system properties, irrespective of the kinetic parameters. In this article, we addressed two practical issues.

Firstly, the parameter distributions: we studied how to depict and compute statistical dependencies among parameters. We proposed a dependence graph for computing the mean values and covariances of all system parameters. It can help the modeller to make the independence assumptions as transparent as possible, to define parameter distributions that obey the thermodynamic constraints, and to incorporate different sources of uncertainty into the model. Based on multiplicative relationships among the parameters, we derived formulae for the joint distribution of model parameters.

Secondly, we addressed variable distributions: if the parameter distribution is narrow, distributions of variables can be approximated by an expansion around the reference parameter set. The resulting formulae show a connection between the response coefficients, which are related to the network structure, and measures of variability [13, 14] and probabilistic sensitivity. The expansion method applies to both steady-state and non-steady state variables, provided that response coefficients have been defined. For log-normal variables, we also derived probabilities for qualitative statements about their signs, their order and whether they exceed certain threshold values. The second-order expansion yields a correction term that plays a role, e.g. in population models: experimental data from cell populations are often fitted by a single-cell model with the average parameters. The second-order term shows that this will only work if the function $\ln \boldsymbol{y}(\ln \boldsymbol{p})$ is approximately linear for the parameter sets in the cell population.

The expansion method yields a poor approximation for variables of changing sign, which can cause problems in the case of control coefficients [15]. For metabolic fluxes, however, we could remedy this problem by considering individual fluxes in forward and backward directions.

Some of the given formulae, e.g. for the probabilities of flux directions, are based on the log-normality of variables, which was justified here by the linear expansion with log-normal parameters. The log-normal distribution is a convenient choice and consistent with the assumption of multiplicative parameters in the dependence graph. It is also biologically plausible, at least for quantities that depend multiplicatively on many independent random influences, e.g. genetic or environmental conditions. Moreover, if a variable depends on many of the independent parameters in the dependence graph, it will be approximately log-normal, even if the distribution of the individual parameters is not.

Monte Carlo sampling of model results can cope with large parameter variation and arbitrary parameter distributions, it can be used for all kinds of variable and it is easy to implement. A major drawback of Monte Carlo simulation is the estimation error caused by the finite number of simulation runs, which decreases only slowly with sample size ($\mathcal{O}(\sqrt{1/n})$). The expansion method requires metabolic response coefficients, and its error depends on the parameter variance. For steady-state variables, the system has to be solved only once for the set of reference parameters, and the response coefficients can also be computed analytically from the elasticities by simple matrix operations [17]. For other variables, such as signal characteristics, they have be computed numerically from small variations of the parameters, but this still may require much fewer simulations than Monte Carlo simulation. Finally, even the Monte Carlo simulations can be speeded up by use of the metabolic response coefficients: to obtain good initial guesses for the numerical solver, we expanded the metabolite concentrations around the reference parameters.

7 Acknowledgments

The authors would like to thank R. Heinrich and O. Ebenhöh for stimulating discussions. This work was funded by the Berlin Center for Genome Based Bioinformatics and by the European Commission, grant 503269.

8 References

1 Kitano, H.: 'Systems biology: a brief overview', *Science*, 2002, **295**, (5560), pp. 1662–1664
2 Swain, P.S., Elowitz, M.B., and Siggia, E.D.: 'Intrinsic and extrinsic contributions to stochasticity in gene expression', *PNAS*, 2002, **99**, (20), p. 12795
3 Paulsson, J.: 'Summing up the noise in gene networks', *Nature*, 2004, **427**, p. 415
4 Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W.: 'Observing and interpreting correlations in metabolomics networks', *Bioinformatics*, 2003, **19**, (8), pp. 1019–1026
5 Liebermeister, W.: 'Response to temporal parameter fluctuations in biochemical networks', *J. Theor. Biol.*, **234**, (3), pp. 423–438
6 Heinrich, R., Neel, B.G., and Rapoport, T.A.: 'Mathematical models of protein kinase signal transduction', *Mol. Cell.*, 2002, **9**, (5), pp. 957–970
7 Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., and Doyle, J.: 'Robustness of cellular functions', *Cell*, 2004, **118**, pp. 675–685
8 Barkai, N., and Leibler, S.: 'Robustness in simple biochemical networks', *Nature*, 1997, **387**, pp. 913–917
9 Blüthgen, N., and Herzel, H.: 'How robust are switches in intracellular signaling cascades?', *J. Theor. Biol.*, 2003, **225**, (3), pp. 293–300
10 Klipp, E., Liebermeister, W., and Wierling, C.: 'Inferring dynamic properties of biochemical reaction networks from structural knowledge', *Genome Informatics*, 2004, **15**, (1), pp. 125–137
11 Small, J.R., and Fell, D.: 'Metabolic control analysis. sensitivity of control coefficients to elasticities', *Eur. J. Biochem.*, 1990, **191**, pp. 413–420
12 Thomas, S., and Fell, D.A.: 'Metabolic control analysis: sensitivity of control coefficients to experimentally determined variables', *J. Theor. Biol.*, 1994, **167**, pp. 175–200
13 Lee, E., *et al.*: 'The roles of APC and axin derived from experimental and theoretical analysis of the Wnt pathway', *PLoS Biology*, 2003, **1**, (1), pp. 116–132
14 Krüger, R., and Heinrich, R.: 'Model reduction and analysis of robustness for the Wnt/β-Catenin signal transduction pathway', *Genome Informatics*, 2004, **15**, (1), pp. 138–148
15 Ainscow, E.K., and Brand, M.D.: 'Errors associated with metabolic control analysis. Application of Monte-Carlo simulation of experimental data', *J. Theor. Biol.*, 1998, **194**, pp. 223–233
16 Brown, K.S., and Sethna, J.P.: 'Statistical mechanical approaches to models with many poorly known parameters', *Phys. Rev. E*, 2003, p. 021904–1
17 Heinrich, R., and Schuster, S.: 'The regulation of cellular systems' (Chapman & Hall, 1996)
18 Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H.: 'Systems biology in practice. Concepts, implementation, and application' (Wiley-VCH, Weinheim, 2005)
19 Kanehisa, M., Goto, S., *et al.*: 'The KEGG databases at genomenet', *Nucleic Acids Res.*, 2002, **30**, pp. 42–46
20 CSHL Symposium 2003: 'The Genome Knowledgebase: A Resource for Biologists and Bioinformaticists' (CSHL Press, CSH, NY, 2003)
21 Krieger, C.J., *et al.*: 'MetaCyc: a multiorganism database of metabolic pathways and enzymes', *Nucleic Acids Res.*, 2004, **32**, Database issue, pp. D438–442
22 Karp, P.D., *et al.*: 'The EcoCyc database', *Nucleic Acids Res.*, 2002, **30**, (1), pp. 56–58
23 Schomburg, I., *et al.*: 'BRENDA, the enzyme database: updates and major new developments', *Nucleic Acids Res.*, 2004, **32**, Database issue, pp. D431–433
24 Goldberg, R.N.: 'Thermodynamics of enzyme-catalyzed reactions: Part 6—1999 update', *J. Phys. Chem. Ref. Data*, 1999, **28**, p. 931
25 Tanaka, M., *et al.*: 'Extraction of a thermodynamic property for biochemical reactions in the metabolic pathway', *Genome Informatics*, 2003, **14**, pp. 370–371
26 Gygi, S.P., *et al.*: 'Correlation between protein and mRNA abundance in yeast', *Mol. Cell. Biol.*, 1999, **19**, (3), pp. 1720–1730
27 Ideker, T., *et al.*: 'Integrated genomic and proteomic analyses of a systematically perturbed metabolic network', *Science*, 2001, **292**, pp. 929–934
28 Petersen, S., Lieres, E.v., de Graaf, A.A., Sahm, H., and Wiechert, W.: 'Metabolic engineering in the post genomic era' *in* 'A multi-scale approach for the predictive modeling of metabolic regulation' (Horizon Bioscience, 2004)
29 Fell, D.A.: 'Metabolic control analysis: a survey of its theoretical and experimental development', *Biochem. J.*, 1992, **286**, pp. 313–330
30 Krewski, D., Wang, Y., Bartlett, S., and Krishnan, K.: 'Uncertainty, variability, and sensitivity in physiological pharmacokinetic models', *J. Biopharm. Stat.*, 1995, **5**, (3), pp. 245–271
31 Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B.: 'Bayesian data analysis' (Chapman & Hall, 1997)
32 Ingalls, B.P., and Sauro, H.M.: 'Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories', *J. Theor. Biol.*, 2003, **222**, (1), pp. 23–36
33 Varma, A., and Palsson, B.O.: 'Metabolic capabilities of Escherichia Coli: I. Synthesis of biosynthetic precursors and cofactors', *J. Theor. Biol.*, 1993, **165**, pp. 477–502
34 Schuster, S., Fell, D., and Dandekar, T.: 'A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks', *Nature Biotechnol.*, 2000, **18**, pp. 326–332
35 Schuster, S., Klamt, S., *et al.*: 'Use of network analysis of metabolic systems in bioengineering', *Bioproc. Biosyst. Eng.*, 2002, **24**, pp. 363–372
36 Ahrend, P.: 'The multivariate Gaussian distribution'. Technical report, Technical University of Denmark, IMM, 2004

Appendix 1

Dependence coefficients for different kinetic rate laws

To compute the parameter distribution using (14), we have to write the parameters in the form of (11). Here we specify coefficients $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{D}$ (omitting the superscript i) for some common kinetic rate laws. As mentioned above, the choice of these coefficients reflects certain independence assumptions, which need not be justified in every case.

Irreversible mass-action kinetics: We assume that the rate constant $k = ru$ is proportional to the enzyme activity u, with a log-normal prefactor r. Setting $\ln r = \mu_r + \sigma_r \ln z$, we obtain

$$\ln k = \ln u + \mu_r + \sigma_r \ln z \tag{38}$$

By comparing this to (11), we obtain the coefficients $a = 1$, $b = 0$, $c = \mu_r$, $d = \sigma_r$.

Reversible mass-action kinetics: We assume that both k_+ and k_- are proportional to the enzyme activity u. Hence, we simply set the geometric mean $\sqrt{(k_+ k_-)} = ru$. With normal $\ln r = \mu_r + \sigma_r \ln z$ and the equilibrium constant $q = k_+/k_-$, we obtain (compare (8))

$$\begin{aligned} \ln k_+ &= \ln u + 1/2 \ln q + \mu_r + \sigma_r \ln z \\ \ln k_+ &= \ln u - 1/2 \ln q + \mu_r + \sigma_r \ln z \end{aligned} \tag{39}$$

For the vector $\boldsymbol{k} = \begin{pmatrix} k_+ \\ k_- \end{pmatrix}$, we obtain

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$$

$$\boldsymbol{c} = \mu_r \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{d} = \sigma_r \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

General sequential reaction mechanism: We employ an approximation for bimolecular or multimolecular reactions: we assume that the overall reaction consists of a sequence of n elementary binding or dissociation steps, each being described by reversible mass-action kinetics with rate constants k_{+i} and k_{-i}. In analogy to the mass-action kinetics, we set

$$\begin{aligned} \ln k_{+i} &= \frac{1}{n}\left(\ln u + \frac{1}{2}\ln q + \mu_r + \sigma_r \ln z_i\right) \\ \ln k_{-i} &= \frac{1}{n}\left(\ln u - \frac{1}{2}\ln q + \mu_r + \sigma_r \ln z_i\right) \end{aligned} \tag{40}$$

and obtain

$$\boldsymbol{a} = \frac{1}{n}\begin{pmatrix} \hat{1} \\ \hat{1} \end{pmatrix}, \qquad \boldsymbol{b} = \frac{1}{2n}\begin{pmatrix} \hat{1} \\ -\hat{1} \end{pmatrix}$$

$$\boldsymbol{c} = \frac{\mu_r}{n}\begin{pmatrix} \hat{1} \\ \hat{1} \end{pmatrix}, \qquad \boldsymbol{D} = \frac{\sigma_r}{n}\begin{pmatrix} \mathrm{diag}\ \hat{1} \\ \mathrm{diag}\ \hat{1} \end{pmatrix}$$

where $\hat{1}$ is a vector of size n with all elements equal to 1.

Irreversible Michaelis–Menten kinetics: The maximum velocity is set $V_{max} = ru$, proportional to the enzyme activity u. Further, we assume that the Michaelis constant K_M is log-normal and independent of V_{max}. With normal $\ln K_M = \mu_K + \sigma_K \ln z_1$ and $\ln r = \mu_r + \sigma_r \ln z_2$, we obtain

$$\begin{aligned} \ln K_M &= \mu_K + \sigma_K \ln z_1 \\ \ln V_{\max} &= \ln u + \mu_r + \sigma_r \ln z_2 \end{aligned} \tag{41}$$

For $\boldsymbol{k} = \begin{pmatrix} K_M \\ V_{\max} \end{pmatrix}$, we obtain

$$\boldsymbol{a} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{c} = \begin{pmatrix} \mu_K \\ \mu_r \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} \sigma_K & 0 \\ 0 & \sigma_r \end{pmatrix}.$$

Reversible Michaelis–Menten kinetics: We choose the K_M values for the forward and backward directions independently from the same log-normal distribution. Further, we assume that the maximum velocities in both directions are proportional to the enzyme activity, with a log-normal prefactor r. Setting $\sqrt{(V^+_{max} V^-_{max})} = ru$ with independent normal $\ln K^+_M = \mu_K + \sigma_K \ln z_1$, $\ln K^-_M = \mu_K + \sigma_K \ln z_2$, and $\ln r = \mu_r + \sigma_r \ln z_3$, we obtain

$$\begin{aligned} \ln K^+_M &= \mu_K + \sigma_K \ln z_1 \\ \ln K^-_M &= \mu_K + \sigma_K \ln z_2 \\ \ln V^+_{\max} &= \ln u + \mu_r + \sigma_r \ln z_3 + \frac{1}{2}(q + \sigma_K \ln z_1 - \sigma_K \ln z_2) \\ \ln V^-_{\max} &= \ln u + \mu_r + \sigma_r \ln z_3 - \frac{1}{2}(q + \sigma_K \ln z_1 - \sigma_K \ln z_2) \end{aligned} \tag{42}$$

For $\boldsymbol{k} = (K^+_M, K^-_M, V^+_{max}, V^-_{max},)^T$, we obtain $\boldsymbol{a} = (0, 0, 1, 1)^T$, $\boldsymbol{b} = (0, 0, 1/2, -1/2)^T$, $\boldsymbol{c} = (\mu_K, \mu_K, \mu_r, \mu_r)^T$ and

$$\boldsymbol{D} = \begin{pmatrix} \sigma_K & 0 & 0 \\ 0 & \sigma_K & 0 \\ \frac{1}{2}\sigma_K & -\frac{1}{2}\sigma_K & \sigma_r \\ -\frac{1}{2}\sigma_K & \frac{1}{2}\sigma_K & \sigma_r \end{pmatrix}$$

By writing $\ln \boldsymbol{k}$ according to (11), with this choice of $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{D}$ and $\boldsymbol{c}$, we can verify the Haldane relationship (compare (3))

$$\begin{aligned} \ln q &= -\ln K^+_M + \ln K^-_M + \ln V^+_{max} - V^-_{max} \\ &= (-1, 1, 1, -1) \ln \boldsymbol{k} \end{aligned} \tag{43}$$

Inhibition: As inhibitors do not change the equilibrium constant, inhibition constants can be chosen independently of the other parameters.

Appendix 2

Variability of control, control of variability

Equations (19)–(22) do not only hold for concentrations and fluxes, but also for other positive, differentiable functions $y(p)$ of the parameters, provided that the respective response coefficients are known. For instance, we may be interested in the ratio of two metabolite concentrations. The distribution of a ratio $z_{ik} = y_i/y_k$ of two variables can

be computed using the response coefficients

$$R^Z_{ikl} = R_{il} - R_{kl} \tag{44}$$

Metabolic control coefficients are defined as $C_{ik} = R_{ik}/\epsilon^P_{kk}$, where ϵ^P_{ik} is the parameter elasticity $\partial v_i/\partial p_k$ and where we assume that each parameter p_k affects only the kinetics of the kth reaction.

To compute the distribution of response and control coefficients themselves, we have to calculate their response coefficients first: the response coefficients of the normalised response coefficients are given by the second-order response coefficients. Response coefficients of the unnormalised control coefficients read

$$R^{C^Y}_{ikl} := \frac{\partial}{\partial p_l} C^Y_{ik}(\boldsymbol{p}) = \bar{R}^{(2)}_{ilk}(\epsilon^P_{kk})^{-1} - \bar{R}_{ik}(\epsilon^P_{kk})^{-2}\epsilon^{PP}_{klk} \tag{45}$$

However, as response and control coefficients can change their signs, their approximation by log-normal distributions is possibly restricted to small parameter variation.

Biological variability itself can be controlled by the cell through regulatory parameters. Let us consider a metabolic system with two types of parameters: the noise parameter vector $\boldsymbol{p}$ follows a probability distribution and creates variability in variables such as metabolic concentrations or fluxes. In contrast to that, the vector $\boldsymbol{q}$ of regulatory parameters can be chosen to control this variability.

Let z_i denote the variance of a logarithmic variable (concentration or flux) y_i due to the variability of the parameters $\boldsymbol{p}$. The response coefficients $R^{z_i}_{q_j}$ between the parameters q_j and the variabilities z_i indicate which of the parameters are the most effective regulators of fluctuations. They read

$$R^{z_i}_{q_j} := \frac{\partial z_i}{\partial q_j} = \sum_{mn} R^{(2)}_{imj}\,\mathrm{cov}(\ln \boldsymbol{p})_{mn} R_{in} + \sum_{mn} R_{im}\,\mathrm{cov}(\ln \boldsymbol{p})_{mn} R^{(2)}_{inj} \tag{46}$$

C Web supplement: Mathematical proofs

C.1 List of symbols

q	Equilibrium constant
k_+	Mass action forward rate
k_-	Mass action backward rate
$V^+_{\max}$	Maximal forward velocity
$V^-_{\max}$	Maximal backward velocity
K^+_{M}	Forward Michaelis constant
g	Gibbs free energy
r	prefactor in reaction velocities
u	enzyme activity
$\mathbf{N}$	stoichiometric matrix
τ	relaxation time
$\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{D}$	dependence coefficients between parameters
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$	dependence matrices between all parameters
$\mathbf{k}, \mathbf{p}$	parameter vector
$\mathbf{p}^0$	reference parameter vector
$\mathbf{y}$	vector of metabolic variables, e.g., concentrations, fluxes etc.
$R, R^{(2)}$	response coefficients, first and second order
$\bar{R}, \bar{R}^{(2)}$	non-normalised response coefficients, first and second order
$\bar{R}(t)$	time dependent response coefficients
$\mathbf{C}$	covariance matrix of $\ln \mathbf{p}$
$U(f\|x)$	stochastic sensitivity

C.2 Derivation of equation (17)

$$\begin{aligned} R^{(2)}_{ikl} &:= \frac{\partial}{\partial \ln p_k}\frac{\partial \ln y_i}{\partial \ln p_l} = p_k \frac{\partial}{\partial p_k}(p_l \bar{R}_{il} y_i^{-1}) \\ &= p_k \delta_{kl} \bar{R}_{il} y_i^{-1} + p_k p_l \bar{R}^{(2)}_{ikl} y_i^{-1} - p_k p_l \bar{R}_{il} \bar{R}_{ik} y_i^{-2} \\ &= y_i^{-1} \bar{R}^{(2)}_{ikl} p_k p_l + R_{ik}\delta_{kl} - R_{ik} R_{il} \end{aligned} \tag{47}$$

C.3 Derivation of equations (19)-(22)

For small deviations $\Delta \ln \mathbf{p} = \ln \mathbf{p} - \ln \mathbf{p}^0$ from the reference parameters, the shift $\Delta \ln \mathbf{y}$ of the variables is linearly approximated by

$$\Delta \ln \mathbf{y} \approx \mathbf{R} \Delta \ln \mathbf{p} \tag{48}$$

where $\mathbf{R}$ is the matrix of normalised response coefficients computed at $\ln \mathbf{p}^0$. The derivative $\Delta \ln \mathbf{p}$ is normal with mean $\langle \Delta \ln \mathbf{p} \rangle = 0$ and covariance matrix $\mathbf{C} = \mathrm{cov}(\Delta \ln \mathbf{p})$, hence within the approximation (48), $\Delta \ln \mathbf{y}$ is also normal with $\langle \Delta \ln \mathbf{y} \rangle = 0$ and covariance matrix

$$\mathrm{cov}(\Delta \ln \mathbf{y}) = \langle (\Delta \ln \mathbf{y})(\Delta \ln \mathbf{y})^{\mathrm{T}} \rangle = \langle \mathbf{R}(\Delta \ln \mathbf{p})(\Delta \ln \mathbf{p})^{\mathrm{T}} \mathbf{R}^{\mathrm{T}} \rangle = \mathbf{R}\mathbf{C}\mathbf{R}^{\mathrm{T}}. \tag{49}$$

With the second-order expansion

$$\Delta \ln y_l \approx \sum_m R_{lm} \Delta \ln p_m + \frac{1}{2} \sum_{mn} R^{(2)}_{lmn} \Delta \ln p_m \Delta \ln p_n, \tag{50}$$

we obtain the mean deviation

$$\begin{aligned} \langle \Delta \ln y_l \rangle &= \sum_m R_{lm} \langle \Delta \ln p_m \rangle + \frac{1}{2} \sum_{mn} R^{(2)}_{lmn} \langle \Delta \ln p_m \Delta \ln p_n \rangle \\ &= \frac{1}{2} \sum_{mn} R^{(2)}_{lmn} C_{mn}. \end{aligned} \tag{51}$$

The covariance reads

$$\mathrm{cov}(\Delta\ln y_l, \Delta\ln y_k) \quad = \quad \langle(\Delta\ln y_l - \langle\Delta\ln y_l\rangle)(\Delta\ln y_k - \langle\Delta\ln y_k\rangle)\rangle \tag{52}$$

After inserting (50) and (51) into (52), and using the third- and fourth-order central moments of the normal distribution (see [36]), we obtain eqn. (22)

C.4 Derivation of equation (45)

$$\begin{aligned} R_{ikl}^{C^Y} &:= \frac{\partial}{\partial p_l} C_{ik}^{\mathrm{Y}} \\ &= \frac{\partial}{\partial p_l}\left(\frac{\partial y_i}{\partial p_k}\left(\frac{\partial v_k}{\partial p_k}\right)^{-1}\right) \\ &= \bar{R}_{ilk}^{Y(2)}(\epsilon_{kk}^{\mathrm{P}})^{-1} + \bar{R}_{ik}^{Y}\left(\frac{\partial}{\partial p_l}\left(\frac{\partial v_k}{\partial p_k}\right)^{-1}\right) \\ &= \bar{R}_{ilk}^{(2)}(\epsilon_{kk}^{\mathrm{P}})^{-1} - \bar{R}_{ik}(\epsilon_{kk}^{\mathrm{P}})^{-2}\epsilon_{klk}^{\mathrm{PP}} \end{aligned} \tag{53}$$

C.5 Derivation of equation (25)

The probability a flux in forward direction is the probability density of $\ln\mathbf{y}$, integrated over all values that fulfil $\ln y_1 > \ln y_2$, that is, $(1,-1)\cdot\ln\mathbf{y} > 0$. To compute this integral, we introduce $\mathbf{z} = \mathbf{C}^{-1/2}(\ln\mathbf{y} - \langle\ln\mathbf{y}\rangle)$ (where $\mathbf{C} := \mathrm{cov}(\ln\mathbf{y})$), which follows a standard bivariate normal distribution. Then the probability for forward flux reads

$$\frac{1}{2\pi}\int \mathrm{e}^{-\frac{1}{2}||\mathbf{z}||^2}\mathrm{d}\mathbf{z}, \tag{54}$$

where the integral runs over all values of $\mathbf{z}$ that fulfil $(1,-1)\ \mathbf{C}^{1/2}\mathbf{z} + (1,-1)\langle\ln\mathbf{y}\rangle > 0$. We project $\mathbf{z}$ onto the vector $\mathbf{z}_\parallel$ parallel to $((1,-1)\ \mathbf{C}^{1/2})^{\mathrm{T}}$ and obtain

$$\begin{aligned} \mathrm{Prob}(y_1 > y_2) &= \mathrm{Prob}(||(1,-1)\ \mathbf{C}^{1/2}||\ \mathbf{z}_\parallel + (1,-1)\ \langle\ln\mathbf{y}\rangle > 0) \\ &= \Phi\left(\frac{(1,-1)\ \langle\ln\mathbf{y}\rangle}{||(1,-1)\ \mathbf{C}^{1/2}||}\right) \\ &= \Phi\left(\frac{\langle\ln y_1\rangle - \langle\ln y_2\rangle}{(\mathrm{var}(\ln y_1) + \mathrm{var}(\ln y_2) - 2\ \mathrm{cov}(\ln y_1, \ln y_2))^{1/2}}\right). \end{aligned} \tag{55}$$

where $\Phi(\cdot)$ denotes the distribution function (cumulative density) of the normal distribution.

C.6 Derivation of equation (29)

We assume multivariate normal random variables x_i forming a vector $\mathbf{x}$ with mean $\langle \mathbf{x} \rangle = 0$ and covariance matrix $\mathbf{C} = \mathrm{cov}(\mathbf{x})$. A random vector $\mathbf{f}$ is computed from $\mathbf{x}$ via a linear function $\mathbf{f} = \mathbf{A}\mathbf{x}$. First, we compute the conditional expectation $\mathcal{E}(f_l | x_i = x_i')$. The vector $\mathbf{x}$, conditional on a value $x_i = x_i'$ is multivariate normal, with mean $\mathbf{x}^*$ maximising $\ln \mathrm{Prob}(\mathbf{x})$ while $x_i = x_i'$.

$$\begin{aligned} \mathbf{x}^* &= \mathrm{argmax}_x(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}) \quad \text{while} \quad x_i^* = x_i' && (56) \\ \Rightarrow x_k^* &= C_{ki}(C_{ii})^{-1}x_i' && (57) \end{aligned}$$

$\mathbf{f}(\mathbf{x})$, conditional on $x_i = x_i'$ is also normal, with mean $f(\mathbf{x}^*)$. Thus

$$\mathcal{E}(f_l | x_i = x_i') = \sum_k A_{lk} C_{ki} (C_{ii})^{-1} x_i' \tag{58}$$

The sensitivity measure $U(f_l | x_i)$ reads

$$\mathrm{var}_{x_i'}(\mathcal{E}(f_l | x_i = x_i')) = \langle \mathcal{E}^2(f_l | x_i = x_i') \rangle - \langle \mathcal{E}(f_l | x_i = x_i') \rangle^2 \tag{59}$$

where the averaging is done over all values of x_i. The second term vanishes as $\mathcal{E}(f_l | x_i = x_i')$ is an odd function of x_i, and with (58), the first term yields

$$\langle \mathcal{E}^2(f_l | x_i = x_i') \rangle = (A_{lk} C_{ki} (C_{ii})^{-1})^2 \langle x_i^2 \rangle = \frac{(\sum_k A_{lk}\ \mathrm{cov}(x_k, x_i))^2}{\mathrm{var}(x_i)} \tag{60}$$

Thus

$$U(f_l | x_i) = \frac{\left(\sum_k A_{lk}\ \mathrm{cov}(x_k, x_i) \right)^2}{\mathrm{var}(x_i)} \tag{61}$$

By setting $\mathbf{x} = \Delta \ln \mathbf{p} = \ln \mathbf{p} - \langle \ln \mathbf{p} \rangle$, $\mathbf{f} = \ln \mathbf{y}$, $A = \mathbf{R}$, we obtain eqn. (29).