Generalised Modelling in Systems Biology

Thilo Gross

> Open any issue of Nature and you will find a diagram illustrating the molecular interactions purported to underlie some behaviour of a living cell. The accompanying text explains how the link between molecules and behaviour is thought to be made. For the simplest connections, such stories may be convincing, but as the mechanisms become more complex, intuitive explanations become more error prone and harder to believe.
> – John J. Tyson, Nature **445**, 823, 2007

1 Introduction

The lines from John Tyson quoted above point to an important challenge biology [35]. What is revealed by experimental measurements is not the biology itself, but only its material basis. For understanding the biology, i.e. the emergence and functioning of life, an interpretation has to be attached to the acquired data. When simple systems are considered the process of interpretation may be so intuitive that it is often not recognized as an independent process at all – meaning seems to emerge directly from measurement. But for complex systems *intuitive explanations become more error-prone and harder to believe* [35].

Intuitive conclusions can be formalised and substantiated by conceptual mathematical models. In contrast to realistic models, which aim to replicate a system in detail, a conceptual model tries to capture a given phenomenon with a *minimal set of ingredients*. If such a model succeeds then it confirms the modellers intuition, showing that the chosen set of ingredients is indeed sufficient to replicate the phenomenon. If the model fails it proves not only that the modeller's intuition is wrong, but by failing in a specific way provides new insights pointing to the nature of the ingredient that was overlooked. Thus, conceptual models lead to new hypothesis

Thilo Gross
Max-Planck-Institute for the Physics of complex sytems, Nöthnitzer Str. 38, 01187 Dresden, Germany e-mail: thilo.gross@physics.org

that can be tested experimentally or theoretically and culminate in a theory of the observed behaviour.

The mathematical modelling and analysis of complex biological systems is complicated by two factors: the large number of components that have to be taken into account and the many uncertainties that exist in the precise interaction of these components. Consider for instance cell metabolism. In all probability future models will keep track of tens or hundreds of different metabolites in different compartments. The kinetic rate laws governing the conversion of metabolites into each other may be known for most reactions from theoretical reasoning or in vitro experiments. However, it is not clear that the same laws hold in the crowded environment of cells and under natural conditions. Ultimately models tracking a large number of components might not be necessary, because most phenomena can possibly be explained on an emergent level of description, where the model needs to follow only a few key metabolites or aggregate variables. However, to gain the insight necessary for constructing such aggregate models, analysing models that resolve the system on a more detailed level may be crucial.

For mathematical modelling, dealing with hundreds of variables does not pose a fundamental problem. Even larger systems are routinely handled in engineering. Neither is working with uncertain interactions is itself difficult. In any analytical computation, the objects under consideration can be treated as unknowns. Nevertheless, large and uncertain systems provide a significant challenge because the numerical methods that are used for investigating large systems cannot deal with unknowns and analytical approaches which can accommodate unknown quantities, often become prohibitively difficult for all but the smallest systems.

Despite the difficulties described above, progress can still be made if we seek answers to specific questions. Consider that numerical simulation, the most common approach to model analysis, reveals whole *trajectories*, i.e. timeseries of the dynamical variables. We can say that simulation answers a very general question: “How does a system starting from some initial conditions evolve in time?”. When analysing conceptual models we are often interested in more specific questions: “Is there a stable equilibrium to which the system can settle in the long run?”, “Does the model explain oscillations that are observed in nature?”, or “Can trajectories diverge, pointing to some flaw in the model?”. These questions concern the *long-term dynamics* of the system, i.e. the behaviour observed after sufficiently long time. If these questions are answered based on numerical simulations one considers only the last steps of long simulation runs, which offer the best approximation to the true long-term dynamics, whereas the *transients* leading up to the long term behaviour are discarded. It is therefore not surprising that methods answering questions about the long-term behaviour of the model directly (without computation of transients) are in general more efficient than simulative approaches to the same problem.

By using methods that are designed to answer specific questions we gain efficiency that can be used either to simplify analytical treatment or to speed up numerical computations. The former may allow extending the applicability of analytical approaches to larger systems. The later can be useful as it may enable running the numerical analysis more often with different parameter values, allowing the re-

searcher to estimate the effect of uncertainties using statistical techniques. Although such sampling analysis are also commonly performed using simulation runs, methods that are geared to specific questions can be orders of magnitude faster than simulation and can therefore acquire a much larger number of samples.

In mathematics a vast array of analytical and numerical techniques for analysing specific aspects of a system's dynamics have been developed [10]. Moreover, in the applied literature different modelling approaches have been proposed that facilitate the application of specific mathematical techniques to biological applications (see [28, 34, 14, 25, 30, 27] for reviews). These include metabolic control theory [1], S-systems [26], power-law models [3] and other approaches discussed in the present volume.

In the present chapter we focus on the approach of generalised modelling. Generalised models are named for their ability to capture the dynamics of systems in which the interactions between model variables is not restricted to specific functional forms. For instance a generalised model could describe a gene regulatory network in which the functional dependencies between the abundances of proteins and the rate of transcription is unspecified. Thus the generalised model can be regarded as generalisation of the whole class of conventional models describing the same system with specific rate laws. The approach of generalised modelling was originally developed for ecological communities [5, 4] before it was recognized as a general approach to complex dynamical systems [6].

While generalised models were at first studied only analytically, many later works used generalised models in numerical studies [29]. Interestingly, answering certain questions for the whole class of systems described by the generalised model can be easier than answering the same question for specific systems contained in this class [6, 24]. Therefore, the approach of generalised modelling can in some cases be advantageous even if the kinetic rate laws for processes in the model are known [37].

This chapter provides a gentle introduction to generalised modelling and reviews some applications of generalised models in systems biology. The central idea of generalised modelling and its mathematical background is explained in Sec. 2. This idea is then illustrated by a simple example in Sec. 3, before we discuss the application to larger models in Sec. 4. The chapter concludes with a brief overview of past applications in systems biology, in Sec. 5, and a discussion of future perspectives in Sec. 6.

2 Basic Concepts

Considering a whole class of dynamical systems, rather than a model restricted to specific functional forms, narrows the range of tools by which information can be extracted from the model. For instance we cannot simulate the dynamics of a generalised model. However, the basic tools of local dynamical systems theory can still be applied with relative ease. In this section we provide a brief introduction to these

tools and sketch the basic idea behind their application to generalised models. A deeper treatment of dynamical systems theory can be found in several excellent textbooks [10, 18].

For illustration we focus on systems of N ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i = f_i(x_1, \ldots, x_N), \tag{1}$$

where $x_1, \ldots, x_N$ are N dynamical variables and the functions f_i denote the right-hand sides of the equations. We say that a system is in a steady state, if $\mathrm{d}/\mathrm{d}tx_i = 0$ for all i. The dynamics close to a steady state, can be captured by the local linearisation given by the Jacobian matrix

$$J_{ij} = \left.\frac{\partial f_i}{\partial x_j}\right|_* , \tag{2}$$

where $|_*$ indicates that the derivative is evaluated in the steady state. The steady state is stable if all eigenvalues have negative real parts.

If the right-hand side, f_i, is changed, e.g. by changing parameters on which f_i depends, then stability can be be lost if at least one of the eigenvalues crosses the imaginary axis. The critical parameter sets at which such transitions occur are known as bifurcation points.

Conventional models.

In *conventional modelling* one restricts the right-hand-side of the systems of equations to specific functional forms that only depend on a finite number of free parameters. In the specific set of equations that is thus obtained one can then compute the stationary solutions as a function of the parameters. Once the steady states are known, one can investigate their stability and bifurcations by explicit construction of the corresponding Jacobian matrices.

Random-matrix models.

The conventional approach, described above, is routinely applied to a large variety of systems and has revealed important insights in a vast range of applications. However, it is certainly not the only in which real world dynamical systems can be modelled. Assume for instance that too little information is available to formulate a plausible system of dynamical equations. If we unable to restrict the functions f_i to any specific functional forms then we cannot compute the steady states. Nevertheless, even without explicit knowledge of the steady states and the underlying dynamical equations equations, we can still express the Jacobian in the general form of Eq. (2). Although the entries of the matrix cannot be pinned down to specific numerical values, they are formally constants and hence can be treated as unknown parameters. If we are able to come up with plausible distributions for the values

that these parameters assume in the real world then we can study the possible local dynamics in the system by random sampling.

The proposition of the previous paragraph – studying an ensemble of randomly drawn Jacobians, where assumptions on the underlying physical system only enter in the statistical distribution of matrix elements – is used in so-called *random matrix models*, which have been successfully used in important applications [12, 21]. In comparison to conventional models random matrix model require very little information and can therefore yield robust results that are independent of many assumptions made in conventional modelling. The main drawback of random matrix models is that all information that is available on the system has to enter on the level of statistical descriptors of the matrix elements. These descriptors, the parameters of random matrix models, are in general not easily interpretable within the context of the application. It is therefore often difficult to incorporate available information into random matrix models, or, phrased differently, to restrict the ensemble of random matrices under consideration to those that can plausibly appear in the application.

Generalised models.

Generalised modelling is an intermediate approach which requires less information than conventional models but allows easier integration of available information than random-matrix models. Thus generalised models are often as easily interpretable as conventional models but come close to the elegance, robustness, and efficiency of random matrix models.

In generalised modelling we decompose the right-hand-side f_i into individual terms representing specific processes in the system, but do not restrict these terms to specific functional forms. Although we cannot compute the steady states of the system at the chosen level of generality, we can formally compute a Jacobian matrix describing all possible steady states of the system. By applying a normalization, this matrix can be expressed in terms of a number of parameters that have clear interpretations in the context of the application.

3 A simple example

For illustrating the approach of generalised modelling let us consider the example of a population of cells that increases due to differentiation of precursors and decreases due to further differentiation or cell death. This system can be modelled by a single ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{dt}}X = G(X) - L(X), \tag{3}$$

where X is the population size and G and L describe the gain and loss, respectively. Because we do not restrict these functions specific functional forms, the equation above constitutes a generalised model.

Let us now consider the dynamics close to a steady state X^*. Clearly, the dynamics will depend on some properties of the unknown functions G and L in X^*. The challenge which we address is to capture these properties in a number of parameters that can be interpreted straight forwardly in the context of the model.

For consistency we assume $X^* > 0$, $G(X^*) > 0$, and $L(X^*) > 0$. We can then define a normalized variable

$$x = \frac{X}{X^*} \tag{4}$$

and normalized functions

$$g(x) = \frac{G(xX^*)}{G(X^*)},\ \ l(x) = \frac{L(xX^*)}{L(X^*)}. \tag{5}$$

Rewriting the system yields

$$\frac{\mathrm{d}}{\mathrm{dt}}x = \frac{G(X^*)}{X^*}g(x) - \frac{L(X^*)}{X^*}l(x). \tag{6}$$

It can be shown [17] that the normalised system is smoothly equivalent to the original. In other words, the normalization cannot change the dynamics qualitatively. But what do we gain by normalizing the system? The normalization rescales the unknown steady state X^* such that it is mapped to a known location $x^* = 1$. Further, the unknown rates of processes in the steady state, $G(X^*)$ and $L(X^*)$, are likewise mapped to known values, $g(x^*) = l(x^*) = 1$. The price we have to pay is to introduce two new unknown prefactors, $G(X^*)/X^*$ and $L(X^*)/X^*$. Let us now consider these prefactors more closely.

Because X^* is a steady state it has to satisfy the stationarity condition $G(X^*) - L(X^*) = 0$. Therefore we know that also $G(X^*)/X^* = L(X^*)/X^*$. Defining

$$\alpha = \frac{G(X^*)}{X^*} = \frac{L(X^*)}{X^*} \tag{7}$$

enables us to rewrite Eq. (6) as

$$\frac{\mathrm{d}}{\mathrm{dt}}x = \alpha(g(x) - l(x)). \tag{8}$$

By definition α is a constant real-valued quantity and can thus be interpreted as an unknown parameter of the model. Inspection of Eq. (7) reveals that this parameter has a straight-forward interpretation: It denotes the *per-capita* gain and loss rate in the steady state. In other words α is the inverse of the average time that cells spend in the population. Such parameters, defining characteristic turnover rates of the model variables are called *scale parameters* in the context of generalised modelling.

Having completed the normalization we can write the Jacobian as

$$\mathbf{J} = \alpha\,(g_{\mathrm{x}} - l_{\mathrm{x}})\,, \tag{9}$$

where

$$g_{\mathrm{x}} = \left.\frac{\partial g(x)}{\partial x}\right|_1\,,\ l_{\mathrm{x}} = \left.\frac{\partial g(x)}{\partial x}\right|_1\,. \tag{10}$$

Again, the two quantities, g_{x} and l_{x}, can be interpreted as parameters of the model. Note that g_{x} and l_{x} are logarithmic derivatives of the original functions in the steady state. For instance

$$l_{\mathrm{x}} = \left.\frac{\partial \log L(X)}{\partial \log X}\right|_*\,. \tag{11}$$

Such parameters are sometimes called *elasticities* [1]. They are used in different fields ranging from econometrics to metabolic control theory, because they can be estimated well from empirical data and provide a natural nonlinear measure for the sensitivity of the processes. If the rate of a process follows a power law, e.g. $L(X) = AX^p$ then the corresponding elasticity is the exponent p. For instance, a linear loss rate corresponds to $l_{\mathrm{x}} = 1$, regardless of the slope of the linear dependency. For more complex rate laws the value of the elasticity can depend on the location of the steady state, but remains intuitively interpretable. For example an activating Hill function with exponent n, has an elasticity of n if evaluated close to zero and an elasticity close to 0 if evaluated at saturation. Inhibiting functions such as $1/X$ are charcterised by negative elasticities.

Note that the scale parameters and elasticities are defined directly in the steady state under consideration. By contrast conventional parameters often require reference to different operating points such as the the half-maximum point or the growth rate at saturation, which may be unattainable in the real world system.

Using the three parameters α, g_{x}, and l_{x} we can express the Jacobian in arbitrary steady states. In this simple example the Jacobian is a 1×1-matrix and the only eigenvalue is $\lambda = \alpha(g_{\mathrm{x}} - l_{\mathrm{x}})$. Stability requires that the eigenvalue is negative and therefore

$$g_{\mathrm{x}} < l_{\mathrm{x}} \tag{12}$$

This stability condition applies to all positive steady states in the whole class of models. In particular the stability condition also holds in systems with multiple steady states. In this case the values of the generalised parameters can depend on the steady state under consideration. We remark that the generalised analysis does not prove that a positive steady state exists in a specific model. However, it is generally easy to find a specific model that has a steady state with given values of the generalised parameters [8, 33].

In summary we have shown that in all models of the form of Eq. (3) stability of a steady state requires that in the steady state the elasticity of the loss is greater than the elasticity of the gain. For instance, in a system with linear gain and quadratic loss the steady state is stable whereas in the the case of quadratic gain and linear loss it is unstable. For the simple example considered here the reader may have guessed this result straight away. However, the same approach can be applied to large networks of interacting components.

4 Application to large systems

The central concept of generalised modelling is replacing unknown derivatives in the Jacobian by products of scale parameters and elasticities. The application of this concept to complex models is to a large extend analogous to the simple example shown above. Nevertheless, some additional difficulties appear, namely: The dynamical equations can contain more than two terms, we may have to deal with a large number of equations, and we have to explore a potentially huge parameter space.

4.1 Multiple gain and loss terms

A relatively harmless complication is encountered, when the generalised equations contain more than one gain and one loss term. For instance

$$\frac{\mathrm{d}}{\mathrm{dt}}X = G(X) - D(X) - M(X) \tag{13}$$

The reader may wonder why one would need multiple terms at all, given that considering one unspecified function L covers the same range of possibilities than the sum of two unspecified functions $D+M$. The answer is interpretability. By giving the model a more detailed structure we make it easier to relate to the real world and thus facilitate the extraction of insights. If it is known that two processes act independently on a variable then it is generally advantageous to map these processes to independent terms in the model. By contrast if two process are not completely independent it is generally better to describe them by a single term in the generalised model.

Applying the normalization procedure from the previous section to Eq. (13) we obtain the intermediate equation

$$\frac{\mathrm{d}}{\mathrm{dt}}x = \frac{G^*}{X^*}g(x) - \frac{D^*}{X^*}d(x) - \frac{M^*}{X^*}m(x) \tag{14}$$

the challenge is now to absorb the prefactors into meaningful scale parameters. Because there are now three prefactors linked by one stationarity condition $G^*/X^* = D^*/X^* + M^*/X*$ we will obtain two independent scale parameters. It is often advantageous to define first a parameter denoting the per-capita turnover rate of the variable and then introducing further parameters that specify which proportion of the turnover originates from which process. In the present example this yields

$$\frac{\mathrm{d}}{\mathrm{dt}}x = \alpha(g(x) - \beta d(x) - \bar{\beta} m(x)), \tag{15}$$

where $\alpha = G^*/X^* = L^*/X^* + M^*/X^*$ is the per-capita turnover rate, $\beta = \underline{D}^*/(D^* + M^*)$ denotes the fraction of the turnover occurring due to the loss D, and $\bar{\beta} = 1 - \beta$

is not an independent parameter, but the complement of β, i.e. the proportion of the turnover occurring due to the loss M.

The approach of first defining a total per-capita turnover rate and then identifying the proportions contributed by the individual processes can be applied to any number of gain and loss terms. However, care has to be take that the parameters describing the proportions add up to one for both gains and losses to ensure that the state under consideration is stationary. We note that in specific systems other ways of absorbing the prefactors into scale parameters may be advantageous. For instance in models of metabolisms many conservation laws exist that reduce the number of parameters. In these systems it can be advantageous to use the flux modes of the system directly as scale parameters. An example of this approach is the analysis of the mitochondrial TCA cycle in [31].

4.2 Iterative refinement

From Eq. (15) the Jacobian can be computed directly. However, let us compute the Jacobian in a different way that provides the opportunity to introduce a useful refinement procedure. Suppose we had just derived the Jacobian of our initial example system, Eq. (9). Now assume that we realize that the loss L in this model is caused by two independent processes D and M. We can now redo the normalisation for the new model, as we have done in Sec. 4.1, or derive the Jacobian for the extended model by *refining* the Jacobian of the initial model.

The refinement procedure starts with an equation capturing the intended refinement

$$L = D + M. \tag{16}$$

We normalize this equation following the same procedure that we applied for the differential equation, which yields

$$l = \beta d + \bar{\beta} m, \tag{17}$$

where we used $D^*/L^* = D^*/(D^* + M^*) = \beta$ and $M^*/L^* = \bar{\beta}$. By differentiating with respect to x we find

$$l_\mathrm{x} = \beta d_\mathrm{x} + \bar{\beta} m_\mathrm{x} \tag{18}$$

Substituting this relationship into the Jacobian, Eq. (9) yields $\mathbf{J} = \alpha(g_\mathrm{x} - \beta d_\mathrm{x} - \bar{\beta} d_\mathrm{x})$, which is the same expression that we would have found by direct differentiation of Eq. (15).

The iterative refinement procedure allows us to make the model more concrete and therefore more interpretable without repeating the normalization of the complete model. Even when studying large systems of equations, new ideas and insights can thereby often integrated quickly and smoothly into generalised models. A more complex example of the iterative refinement can be found in [36].

4.3 Extracting information from generalised models

In the simple example systems studied so far, the Jacobian was a 1×1-matrix, so that we were able to read off the only eigenvalue directly. However, the analytical computation of eigenvalues becomes tedious already in systems with more than 2 variables and is generally impossible in systems with more than four variables. This problem can be overcome in two possible ways: First, we can compute the eigenvalues of a given Jacobian numerically. Second, we can compute the bifurcation points at which stability changes by so-called direct methods that do not rely on computation of eigenvalues.

Let us first discuss the direct computation of bifurcation points. Direct methods are based on the construction of test-functions that vanish in bifurcation points. Using the method described in [11, 7] such testfunctions can be formulated straight forwardly. The analytical expressions that are obtained in this way directly reveal the critical parameter values at which the stability of steady states is lost. Further, the testfunctions distinguish between different types of bifurcations corresponding to different types of instabilities, such as the onset of oscillations (*Hopf bifurcation*) or disappearance of the steady state (*saddle-node bifurcation*). Thereby, the computation of bifurcations reveals the dynamics that can be expected immediately after the stability of the steady state has been lost. Finally, the analysis can detect certain types of bifurcations (bifurcations of higher codimension) that provide some insights into the global dynamics of the system. These bifurcation can for instance indicate parameter regions where chaotic dynamics can be observed [18, 8, 37].

Using the method cited above testfunctions for local bifurcations can, at least in principle, be constructed for systems of any size. However, the analytical expressions that are obtained become too complicated to interpret directly for systems of more than approximately six variables. In the larger systems one can still use the testfunctions to visualise bifurcations in bifurcation diagrams. The major drawback of visualisation compared to direct inspection of the equations is that only a very limited number of parameters can be plotted in any one bifurcation diagram. Ideally, one uses a three-parameter diagrams (Fig. 1), which can be created using the algorithm described in [32]. In the three-dimensional space spanned by these diagrams every single point corresponds to a steady state with the respective parameter values. The bifurcation points form surfaces in the space that divide the parameter volume into regions of different stability. They thereby allow the researcher to see which combinations of parameter values lead to stable steady states and which cause instability.

Already in systems with only 6 parameters a three-parameter diagram can only show a small slice of the parameter space. Plotting three-parameter diagrams can still be valuable as they can provide intuition on the shape of stability boundaries and efficiently reveal bifurcations of higher codimension. Nevertheless, an exhaustive analysis of a system with many parameters cannot be based on the inspection of bifurcation diagrams alone, but must be supplemented by other techniques that use information from the whole parameter space accessible by the system.

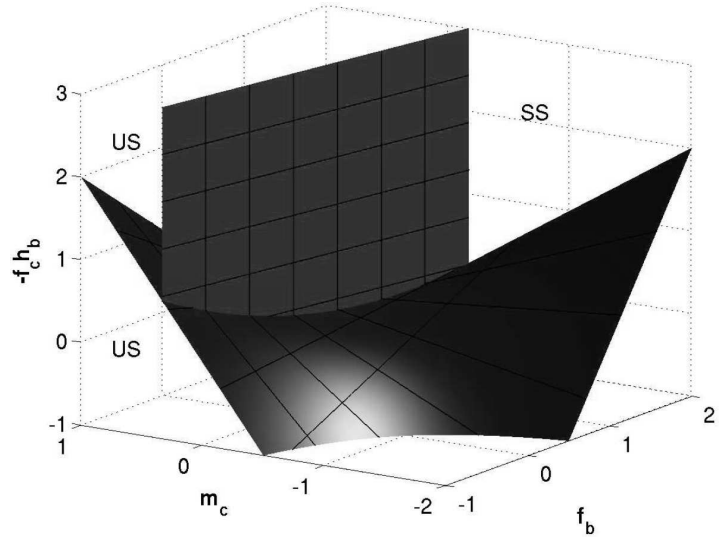

Fig. 1 Three-parameter bifurcation diagram of two-variable models of bone remodelling from [38]. The stability of steady states in this class of models depends only on the elasticity of the net growth rate of osteoclasts with respect to osteoclast abundance, m_c, the elasticity of osteoblast reproduction with respect to osteoblast abundance, f_b, and the product of the elasticities of the growth rate of osteoclasts and osteoblasts with respect to the abundance of the other cell type, f_c and h_b, respectively. Every point in the three dimensional volume corresponds to a family of steady states characterised by the respective parameter values. Regions of stable local dynamics (SS) are divided from unstable states (US) by surfaces of bifurcation points, which correspond to different types of instabilities (shiny-black=saddle-node bifurcation, matte-grey=Hopf bifurcation). For small generalised models such three-parameter diagrams can provide an intuitive picture of the parameter space. Larger models are better explored with statistical methods based on numerical random sampling of the parameter space.

For studying a systems with many parameters we typically have to resort to the numerical computation of eigenvalues. Because numerical eigensolvers can only work on matrices in which all parameters have been fixed to specific values, we explore the parameter space by random sampling. For this purpose we first create a large ensemble of random parameter sets. For each set we then substitute the respective parameter values into the Jacobian and numerically compute the eigenvalue with the largest real part. Depending on the sign of the eigenvalue we assign a stability of 0 (unstable) or 1 (stable) to the parameter set under consideration. Thereby, we obtain a large database of parameter sets and the stability of the steady states they describe. We can then analyse this database by statistical methods aiming to discover the features that characterise stable states.

The simplest approach for analysing the sample database is to correlate the parameter values with the stability value [29, 37]. We extract for example the value of the first parameter and the stability value from all parameter sets. Then we compute the Pearson correlation coefficient between the parameter values and the corre-

sponding stability values. The resulting value is positive if increasing the parameter promotes stability and it is negative if a high value of the parameter hinders stability. Computing the correlation coefficients for every parameter reveals their respective effect on the stability of the system and provides some indication of their relative importance.

Once important parameters have been identified, the database of random parameter sets can be used for more detailed analyses. For instance we can create histograms that show the probability of randomly drawing a stable steady state over the value of one of the parameters. Although such diagrams show only the effect of one parameter, they contain more information than 1-parameter bifurcation diagrams, because the other parameters are not fixed to specific values. In some applications other techniques have been used to extract additional information or answer specific questions. These include the computation of mutual information between parameters and stability, the analysis of eigenvalues and eigenvectors of unstable states [37], two-parameter histograms [9], and bifurcation diagrams generated from computed eigenvalues [29].

For the computation of the largest eigenvalue, highly efficient algorithms are readily available in common software packages. Therefore sufficiently large databases for the correlation and histogram analysis can be created in reasonable computational time (often seconds to minutes) and with very little manual effort.

5 Applications in systems biology

In some applications even the analysis of very simple generalised models can reveal new insights. One such application is bone remodelling. Throughout an individual's life bone is constantly remodelled by the interplay of two cell types. Old bone is resorbed by by specialised cells called *osteoclasts*, while new bone is formed by *osteoblasts*. For remaining in a physiological state it is essential that the number of osteoclasts and osteoblasts remain in balance, but in order to be responsive to changing conditions the abundances of cells must be sensitive to external stimuli.

For understanding the implications of the partially contradictory demands of stability and responsiveness it is useful to study simple non-spatial models of the regulatory interactions controlling the abundances of the two cell types. The simplest such models contain only two variables, which describe the abundance of osteoclasts and osteoblasts, respectively [16]. In some models also responding osteoblasts, a class of precursor cells is modelled explicitly [19, 22].

Generalised models of two and three-variable models for bone remodelling were analysed in [38]. The analysis showed that two-variable models are unlikely to have stable steady states in the experimentally observed parameter ranges, which is a potential shortcoming of such models. Further, computation of bifurcations in the general three variable model showed that the area of parameter space that is most likely realized in vivo is very close to Hopf and saddle-nod bifurcations. Operating in this range enhances responsiveness of the system, but decreases its stability

against perturbations. A system operating in this parameter regime may therefore be destabilized by small variations in certain parameters. Although theoretical analysis alone cannot prove that such transitions are the cause of pathologies in patients, it is apparent that a bifurcation happening in vivo would lead to pathological dynamics. In particular, a Hopf bifurcation may lead to oscillatory rates of remodelling that are observed in certain forms of Paget disease.

The direct computation of bifurcation surfaces, was also recently used in an insightful paper that compared continuous and boolean models of regulatory networks [2]. The comparison revealed that approximating continuous time regulation by boolean functions generally changes the dynamics of the system.

A larger generalised model that was recently proposed focusses on the mitogen-activated-protein-kinase (MAPK) cascade, a basic motif of cell signalling that appears in all eucaryotic cells. A major question concerning the MAPK cascade is why the cascade has evolved, apparently independently, in different organisms and why it is conserved throughout evolution. It has been previously pointed out that the MAPK cascade acts as an ultra-sensitive switch [13, 20], but can also exhibit sustained oscillations [15, 23].

For gaining general insights into these dynamics of the cascade, a generalised model was formulated and analysed in [37]. In this model numerical sampling and bifurcation analysis immediately revealed the Hopf bifurcations which mark the onset of oscillatory dynamics. Furthermore the bistability, which causes the switch-like behaviour can be inferred from the presence of a saddle-node bifurcations in the generalised model. Analysing the eigenvectors close to this bifurcation provides insights into the mechanism of the switch. Perhaps most importantly, the authors found that if adding additional feedback loops known in certain organisms increases rather than decreases the region in which oscillatory dynamics is observed. The latter result suggests that oscillations in the MAPK system may have functional significance, at least in certain stages of development. However, if the MAPK system indeed functions both as a switch and as an oscillator, it is not surprising that it is well conserved throughout evolution, because mutations are likely to interfere with at least one of the two functions. The generalised analysis in [37] further revealed a parameter region in which highly irregular and apparently chaotic dynamics can be observed. Whether this parameter region can be observed in experiments remains to be seen.

A considerable fraction of the generalised models studied so far dwescribe metabolic systems. The ancestor of this line of work is [29]. Apart from a simple model for glycolysis in yeast, the paper analysed a generalised model of the photosynthetic Calvin cycle, which remains one of the most complex models that have been studied with generalised modelling. To facilitate the dynamics of this system the authors proposed a convenient notation. Generalised models using this notation are sometimes called *structural kinetic models* after the title of the paper.

One example of generalised modelling of metabolism is the analysis of the mitochondrial TCA cycle in plants [31]. This cycle plays a central role in energy metabolism and biomass production. Among other results, a sampling analysis revealed that of all processes considered in the model the import of pyruvate into the

mitochondria has the strongest destabilizing effect. This may explain why pyruvate is in some plants only imported into the mitochondria in very low amounts.

Other metabolic systems that have been studied by generalised modelling include a simple model of calcium oscillations [32] and a recent study of the stability of metabolic cycles [24]. The latter study showed analytically that a certain class of one-input one-output metabolic cycles is unconditionally stable, whereas autocatalytic cycles are only stable under certain conditions.

6 Discussion

In the beginning of this chapter we motivated the need for alternative modelling approaches with a quote from John Tyson. This quote is from an essay entitled *Bringing cartoons to life* [35], which argues that modelling approaches are needed that can extract the dynamics of the large networks of dynamical elements that are found in biology. But can generalised modelling really address this challenge? In this discussion I argue that the answer is at least partially, yes.

One major challenge that future models have to cope with is the sheer number of variables in the systems under consideration. However, in generalised modelling the number of variables is unlikely to be a serious obstacle. At least one important technique for analysing generalised models, the numerical computation of leading eigenvalues, can be scaled to systems with thousands of variables.

A potentially more serious challenge is the huge parameter space that has to be explored. The numerically efficiency of generalised modelling means that a very high number of samples can be created. For instance a recent publication [9] considered a model containing more than 5000 unknown parameters. This model was analysed presented by randomly sampling the parameter space in 10^{11} sample points. However, even 10^{11} is relatively small when compared to the size of the parameter space. Consider that, placing a sample point at each corner of the parameter space would have required $2^{5000} \approx 10^{1500}$ samples. Nevertheless, for many analyses performed in this paper already ensembles of 10^8 parameter sets were sufficient to obtain significant statistical results with small error bars. The reason for the fast convergence is possibly that the stability boundary is relatively flat. Its properties can therefore be estimated more efficiently than the size of the embedding space would suggest. These results indicate that also future generalised models in systems biology, containing thousands of parameters, can be understood based on as few as some billion parameters sets. We believe that this number may be reduced further if the parameter space is explored by techniques from machine learning, rather than fully-random sampling.

Another considerable challenge in future models will be the formulation of the model itself. For large systems containing hundreds of variables even generating the equations can be a demanding task. In this area generalised modelling offers a clear advantage. Formulating a generalised model requires less manual work then formulating a conventional model because the functional forms in the model do no

need to be specified. The subsequent construction of the Jacobian matrix can be automated in computer algebra systems and thus requires no manual work.

The price that we have to pay for the advantages of generalised modelling is that the generalised model can only be analysed with a limited set of tools. Generalised models can reveal the parameters that have a strong influence on the stability of steady states. They can be used to determine the threshold parameter values at which the stability of steady states are lost. Furthermore, they can determine the nature of the instability that caused the destabilization and reveal the dynamics that is observed immediately after the destabilization occurs. Some other information, such as the presence of bistability and in special cases chaotic dynamics, is only revealed indirectly in generalised models. Generalised modelling is therefore useful mainly for analysing local dynamics close to steady states. By contrast, if the modeller seeks to explain some transient or highly non-stationary behaviour of the system, generalised models are unlikely to be the right tool.

For studying a complex biological system, generalised models will probably need to be backed up by other modelling approaches including conventional models. In this context generalised modelling will most likely be used as a high-throughput prescreening tool that is applied before conventional modelling is attempted. Used in this way, the generalised model can rapidly narrow down the ranges of parameters and functional forms that can possibly describe the phenomenon at hand. Using the iterative refinement procedure, described above, specific details can be integrated successively into the generalised model. Thereby the ground is prepared for the formulation of a conventional model that in comparison reveals more detailed insights in a narrower range of parameters and functions.

References

1. Fell D A, Sauro H M (1985) Metabolic control and its analysis. European Journal of Biochemistry 148:555–561
   J. Guckenheimer and M. Myers. Computing Hopf Bifurcations II. SIAM J. Sci. Com- put., 17(6):12751301, 1996.
2. Gehrmann E, Drossel B (2010) Boolean versus continuous dynamics on simple two-gene modules. Physical Review E 82:046120–9
3. González J V, Balsa-Canto E, Wellstead P et al (2007) Power-law models of signal transduction pathways. Cellular Signalling 19:1531–1541
4. Gross T (2001) Population dynamics. Der Andere Verlag, Tönning
5. Gross T, Ebenhöh W, Feudel U (2004) Enrichment and foodchain stability. Journal of theoretical biology 227:349–358
6. Gross T, Feudel U (2006) Generalized models as an universal approach to the analysis of nonlinear dynamical systems. Physical Review E 73:016205–14
7. Gross T, Feudel U (2004) Analytical search for bifurcation surfaces in parameter space. Physica D 292–302
8. Gross T, Feudel U (2005) Long food chains are in general chaotic. Oikos 109:135–155
9. Gross T, Rudolf L, Levin S A et al (2009) Generalized models reveal stabilizing factors in food webs. Science 325:747–750
10. Guckenheimer J, Holmes P (1983) Nonlinear oscillations, dynamical systems and bifurcations of vector fields. Springer Verlag, Heidelberg

11. Guckenheimer J, Myers M (1996) Computing Hopf bifurcations II. SIAM Journal of Scientific Computing 17:1275–1301
12. Jamshidi N, Palsson B Ø(2008) Formulating genome-scale kinetic models in the post-genome era. Molecular Systems Biology 4:1–10
13. Huang C, Ferrell J (1996) Ultrasensitivity in the mitogenactivated protein kinase cascade. PNAS 93:10078–10083
14. Guhr T, Müller-Groeling A, Weidenmüller H A(1998) Random-matrix theories in quantum physics. Physics Reports 299:189-425
15. Kholodenko B (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. European Journal of Biochemistry 267:1583–1588
16. Komarova S, Smith R, Dixon S et al (2003) Mathematical model predicts a critical role for osteoclast autocrine regulation in the control of bone remodeling. Bone 33:206–215
17. Kuehn C, Siegmund S, Gross T (2010) On the dynamical analysis of evolution equations via generalized models. arXiv:1012.4340
18. Kuznetsov Yu A (2004) Elements of applied bifurcation theory. Springer Verlag, Heidelberg
19. Lemaire V, Tobin F, Greller L et al (2004) Modeling the interactions between osteoblast and osteoclast activities in bone remodeling. Journal of Theoretical Biology 229:293–309
20. Markevich N, Hoek J, Kholodenko B (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. Journal of Cell Biology 164:353–359
21. May R M (1972) Will a large complex system be stable? Nature 238:413–414
22. Pivonka P, Zimak J, Smith D W et al (2008) Model structure and control of bone remodeling. Bone 43:249–263
23. Qiao L, Nachbar R, Kevrekidis I G et al (2007) Bistability and oscillations in the Huang-Ferrell model of MAPK signaling. PLoS Computational Biology 3:1819–1826
24. Reznik Segrè (2010) On the stability of metabolic cycles. Journal of Theoretical Biology 266:536-549
25. Rodriguez A, Infante D (2009) Network models in the study of metabolism. Electronic Journal of Biotechnology 12:1–12
26. Savageau M A, Voit E O (1987) Recasting nonlinear differential equations as S-systems. Mathematical Biosciences 87:83–115
27. Schallau K, Junker B H (2010) Simulating plant metabolic pathways with enzyme-kinetic models. Plant Physiology 152:1763–1771
28. Steuer R (2007) Computational approaches to the topology, stability and dynamics of metabolic networks. Phytochemistry 68:16–18
29. Steuer R, Gross T, Selbig J et al (2006) Structural kinetic modeling of metabolic networks. PNAS 103:11868–11873
30. Steuer R, Junker B H (2009) Computational models of metabolism. Advances In Chemical Physics 142:105–251
31. Steuer R, Nunes Nesi A, Fernie A R et al (2007) From structure to dynamics of metabolic pathways. Bioinformatics 23:1378–1385
32. Stiefs D, Gross T, Steuer R et al (2008) Computation and visualization of bifurcation surfaces. International Journal Bifurcation and Chaos 18:2191–2206
33. Stiefs D, van Voorn G A K, Kooi B W et al (2010) Food quality in producer-grazer models. The American Naturalist 176:367–380
34. Sweetlove L J, Fell D, Fernie A R (2008) Getting to grips with the plant metabolic network. Biochemical Journal 409:27–41
35. Tyson J J (2007) Bringing cartoons to life. Nature 445:823–823
36. Yeakel J A, Stiefs D, Novak M et al (2011) Generalized modeling of ecological population dynamics. Theoretical Ecology, in press
37. Zumsande M, Gross T (2010) Bifurcations and chaos in the MAPK signalling cascade. Journal of Theoretical Biology 265:481–491
38. Zumsande M, Stiefs D, Siegmund S et al (2011) General analysis for mathematical models in bone remodeling. Bone, in press