

A methodology for global-sensitivity analysis of time-dependent outputs in systems biology modelling

T. Sumner^{1,2,*}, E. Shephard^{1,2} and I. D. L. Bogle^{1,3}

¹CoMPLEX, ²Structural and Molecular Biology, and ³Chemical Engineering, University College London, Gower Street, London WC1E 6BT, UK

One of the main challenges in the development of mathematical and computational models of biological systems is the precise estimation of parameter values. Understanding the effects of uncertainties in parameter values on model behaviour is crucial to the successful use of these models. Global sensitivity analysis (SA) can be used to quantify the variability in model predictions resulting from the uncertainty in multiple parameters and to shed light on the biological mechanisms driving system behaviour. We present a new methodology for global SA in systems biology which is computationally efficient and can be used to identify the key parameters and their interactions which drive the dynamic behaviour of a complex biological model. The approach combines functional principal component analysis with established global SA techniques. The methodology is applied to a model of the insulin signalling pathway, defects of which are a major cause of type 2 diabetes and a number of key features of the system are identified.

Keywords: sensitivity analysis; computational modelling; insulin signalling

1. INTRODUCTION

Biological systems typically consist of large numbers of interacting components and involve processes operating across a variety of spatial, temporal and biological scales. Systems biology aims to understand such systems by integrating information from all functional levels into a single cohesive model. Mathematical and computational modelling are a key part of the systems biology approach providing a method for formally defining and analysing the structure and the function of a system.

One of the greatest challenges when building mathematical or computational models of biological systems is the precise estimation of parameter values [1]. Values for specific parameters measured *in vivo* are rare [2] and parameters are often estimated from experimental measurements made *in vitro* or by fitting of model simulations to experimental data. Consequently, parameter estimates are often associated with significant uncertainty. Examining the influence of these uncertainties on the model behaviour is crucial to the successful use of mathematical and computational models of biological systems both as predictive tools and as part of an iterative cycle of modelling and experimentation to understand the function of biological systems.

Sensitivity analysis (SA) provides a quantitative approach for investigating the impacts of parameter

uncertainty on model outputs. SA is used in a variety of disciplines from environmental science to software engineering and in many fields is seen as a prerequisite for model building [3]. In systems biology, SA can be used in two key areas: quantifying the variation in model outputs to parameter uncertainty (often referred to as uncertainty analysis (UA)) thereby providing a measure of confidence in the predictive capacity of the model; identifying the parameters that contribute most to the variation in the model outputs allowing us to generate hypotheses about the biological mechanisms that drive the system behaviour which can be tested experimentally. By identifying the most influential parameters we can also improve the predictive capacity of the model by refining our estimates for those parameters.

There are two main classes of SA techniques, local and global. Local SA techniques investigate the effects of small variations in individual parameters around some nominal point and have been applied to a number of signal transduction and metabolic pathway models [4,5]. Where parameter values have large uncertainties or models are nonlinear global SA (GSA) techniques are more appropriate. GSA techniques investigate the effects of simultaneous parameter variations over large but finite ranges and allow the effects of interactions between parameters to be explored [3].

There has been an increased interest in the use of GSA in biological modelling in recent years. A comparative study of local SA and global SA techniques (partial rank correlation coefficients (PRCC) [6], Sobol's method [7] and the extended Fourier amplitude

*Author and address for correspondence: London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK (tom.sumner@LSHTM.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2011.0891> or via <http://rsif.royalsocietypublishing.org>.

sensitivity test (eFAST) [8]) applied to a model of the Erk-MAPK signalling pathway by Zheng & Rundell [2] demonstrated the disparity between the results of the two approaches, and Marino *et al.* [9] developed a methodology for performing global SA of both deterministic and stochastic systems biology models which combined PRCC and eFAST.

In this paper, we consider two aspects of the application of GSA in systems biology: the application of GSA to time-dependent model outputs and the computational efficiency of the commonly used global SA techniques. We describe an alternative methodology for performing global SA of time-dependent model outputs and within that methodology compare existing GSA methods to evaluate their computational performance and utility in systems biology.

In biological systems, we are typically interested in understanding the dynamic behaviour of the system. To investigate how this behaviour depends on the model parameters, we must apply SA to model outputs which are functions of time. The application of SA to time-dependent model outputs is well established. Sensitivity indices are calculated at each time-point to produce a set of time-varying sensitivity indices. These indices provide information on which parameters are influential at particular times and can be integrated over time to identify those parameters which are most important in terms of the entire model output. This approach has been applied to a variety of biological systems [2,5,9,10]. However, by looking at individual time-points, we may miss interesting features in the model output. For example, it would be difficult to infer the impact of a parameter on the period of oscillations or the time at which a peak concentration occurs.

Alternatively, one may define a set of scalar features of the dynamic model output, such as the maximum concentration or the period of oscillations, and calculate sensitivity indices for these derived outputs [4,11]. To use this approach, we need to identify an appropriate set of features for the model of interest and construct algorithms to extract their values from the model outputs.

Here, we present an alternative methodology for the global SA of biological models based on an idea proposed by Campbell *et al.* [12]. The methodology combines existing GSA techniques with functional principal components analysis (fPCA) to investigate the parameter sensitivity of time-dependent model outputs. fPCA is used to transform the model outputs into an alternative format that captures the most important features in the model output. Existing GSA techniques are then used to identify the parameters, which are important in generating these features and hence in driving the associated system behaviour.

To address the second issue, the computational efficiency of GSA methods, we compare the use of two existing GSA techniques within our methodology. When a model contains a large number of uncertain parameters, a common occurrence in biological models, or has a significant run time, the computational cost of many GSA techniques can be significant. The efficiency of GSA techniques becomes increasingly important if they are to be used as part of an iterative cycle of model development and experimentation in

which repeated analysis of the model may be required. Here, we compare the use of the variance-based method of Sobol with a more computationally efficient, but less quantitative design, the Morris method [13], to evaluate the utility of both techniques in a systems biology context.

2. METHODS

This section presents an overview of the methodology including the use of functional principal components (PCs) and introduces the GSA techniques used in the methodology. The methodology is demonstrated by application to a mathematical model of the insulin signalling pathway which is described in §2.2.

2.1. Sensitivity analysis methodology

Consider a model of the form $y(t) = f(u, \mathbf{P}, t)$. The output of the model $\mathbf{y}(t)$ is a set of curves describing the variation in the model variables over time. The output is some function (f) of the external model input (u) and a set of k model parameters, ($\mathbf{P} = (p_1, p_2, \dots, p_k)$). Often this relationship between the model input and parameters and the model output will be described by a set of differential equations that describe the change in the model variables over time.

We assume that the value of each parameter is uncertain and that this uncertainty can be expressed as a probability distribution. The aim of the SA is to identify how the output of the model depends on the uncertainty in the parameters.

We begin by generating N parameter vectors $\mathbf{P}_i = (p_{i1}, p_{i2}, \dots, p_{ik})$ for $i = 1, \dots, N$ from the probability distributions for each parameter. The number of parameter sets and the way in which they are sampled depends on the chosen GSA technique (see §2.2 and figure 1a). The model is then evaluated for each \mathbf{P}_i to generate the corresponding model outputs $y_i(t)$ ($i = 1, \dots, N$) (figure 1b). For convenience, we assume here that the output of interest is a single model variable but the following can be replicated for any number of model variables.

fPCA is then applied to this set of i model outputs to generate the functional PCs and associated scores. Principal component analysis (PCA) is a multivariate statistical procedure, which is commonly used to identify the dominant modes of variation in a set of data consisting of N observations of M variables. PCA works by transforming the original data onto a new set of variables known as PCs. The new variables are such that they describe decreasing amounts of the variation in the data, i.e. the data vary most in the direction of the first PC, the second PC describes the second most important type of variation and so on. A consequence of this is that we can approximate the data in a subset of the new variables thus reducing the dimensionality of the data.

When our data consist of N observations of some function $y(t)$, we can use fPCA to transform the data into a new set of functions (the functional principal components (fPCs)) which capture the most important types of variation in the data. As with multivariate PCA, the functions describe decreasingly small

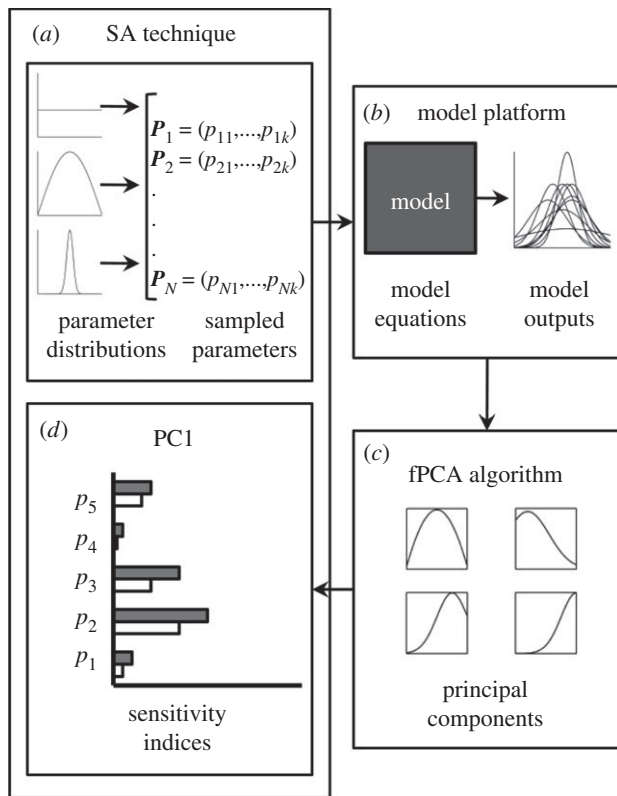


Figure 1. Flow diagram illustrating the main steps in the methodology for performing GSA of time-dependent outputs of systems biology models. Consider a model with k uncertain parameters. (a) An existing GSA technique (Sobol’s method or the Morris method) is used to generate multiple sets of parameter values for the model. Each parameter set consists of a vector $\mathbf{P} = (p_1, \dots, p_k)$ of values for the uncertain model parameter. (b) The model is then evaluated for each set of parameters to produce a set of model outputs. (c) fPCA is applied to the set of model outputs. (d) The PC scores are passed back to the GSA technique and their sensitivity to the model parameters is calculated.

amounts of the variance in the data such that we can typically represent the original data using only the first few fPCs. The fPCs can be viewed as a set of curves $\boldsymbol{\xi}(t) = (\xi_1(t), \xi_2(t), \dots, \xi_q(t))$ such that each model output $y_i(t)$ can be written as a sum of those curves:

$$y_i(t) = \sum_j \omega_{ij} \xi_j(t) \quad \text{for } i = 1, \dots, N, \quad (2.1)$$

where ω_{ij} is the PC score for model output i on component j . The PC scores (ω_{ij}) tell us the amount of component j contained in output i . For example, if the first PC, $\xi_1(t)$, represents an oscillatory time course of period T , model outputs $y_i(t)$ which display this type of dynamics will have high values for the first PC score (ω_{i1}). Details of the computation of the fPCs are given in the electronic supplementary material.

GSA techniques are then used to calculate the sensitivity of the PC scores (ω_{ij}) to the model parameters (figure 1d). If the scores for a given PC are sensitive to a particular parameter, then that parameter is important in producing the type of behaviour in the model output described by the corresponding component.

2.2. Sensitivity analysis techniques

The methodology uses existing SA techniques to generate the uncertain model parameters and calculate the sensitivities of the PC scores to these parameters. Two techniques, the method of Sobol [7] and the Morris method [14] are compared to evaluate their utility in the analysis of biological systems.

2.2.1. The method of Sobol

Variance-based methods are a class of GSA techniques that quantify the importance of a parameter by the reduction in the model output variance which is obtained by ‘fixing’ the parameter to its ‘true’ value [15]. Unlike regression and correlation-based indices which are only suitable when the relationship between parameters and model outputs satisfy certain conditions of linearity or mono-tonicity, variance-based techniques are not dependent on such assumptions and hence are sometimes referred to as ‘model-free’ methods [3]. Marino *et al.* [9] identified variance-based techniques as a key tool in their methodology for applying GSA in systems biology owing to their ability to deal with any type of relationship between model parameters and model outputs. The method of Sobol [7] is one of the most commonly used variance-based methods, in part because it is relatively easy to implement compared with the other approaches [16] and because in its modified form, it is as efficient as any other variance-based technique including the eFAST method [17].

The method of Sobol is based on a decomposition of the model output $y(t) = f(u, \mathbf{P}, t)$ into terms of increasing dimensionality. The function $f(u, \mathbf{P}, t)$ can be written as the sum (note that we drop the dependence on u and t from the following for clarity of presentation):

$$f(p_1, \dots, p_k) = f_0 + \sum_{i=1}^k f_i(p_i) + \sum_{1 \leq i < j \leq k} f_{ij}(p_i, p_j) + \dots + f_{1,2,\dots,k}(p_1, \dots, p_k), \quad (2.2)$$

where

$$f_0 = \int_{\Omega^k} f(\mathbf{P}) d\mathbf{P}, \quad (2.3)$$

where Ω^k is the k -dimensional space of the parameters. The total variance of $f(\mathbf{P})$ can be written as:

$$V = \int_{\Omega^k} f^2(\mathbf{P}) d\mathbf{P} - f_0^2, \quad (2.4)$$

and can be decomposed in the same manner as the function:

$$V = \sum_{i=1}^k V_i + \sum_{1 \leq i < j \leq k} V_{ij} + \dots + V_{1,2,\dots,k}. \quad (2.5)$$

The terms of this decomposition are the contributions to the variance from term $f_{i_1 \dots i_n}$ in (2.2) and

are given by:

$$V_{i_1 \dots i_s} = \int_0^1 \dots \int_0^1 f_{i_1 \dots i_s}^2(x_{i_1}, \dots, x_{i_s}) dx_{i_1} \dots dx_{i_s}. \quad (2.6)$$

The Sobol indices are defined as:

$$S_{i_1 \dots i_s} = \frac{V_{i_1 \dots i_s}}{V}. \quad (2.7)$$

The term $S_{i_1 \dots i_s}$ gives the fraction of the total variance which is due to the combination of parameters $(p_{i_1}, p_{i_2}, \dots, p_{i_s})$. The first-order sensitivity indices, $S_i = V_i/V$, describe the effect of individual parameters on the model output while higher order terms describe the effect owing to interactions between parameters.

As an alternative to calculating the entire set of indices, we can calculate the ‘total effects’ [15] which describe the effect on the model output of a parameter, both individually and in all its possible interactions. If we define one set of parameters to contain only p_i and the other set $\mathbf{P}_{\sim i}$ contains all other parameters except p_i then the total variance can be written as:

$$V = V_i + V_{\sim i} + V_{i, \sim i}, \quad (2.8)$$

the sum of the variance owing to p_i alone (V_i), the variance owing to the set of parameters $\mathbf{P}_{\sim i}$ which excludes p_i ($V_{\sim i}$) and the variance owing to the interaction between p_i and all other parameters ($V_{i, \sim i}$). The total variance owing to p_i (on its own and in interactions with other parameters) is then:

$$V_i^{\text{tot}} = V_i + V_{i, \sim i} = V - V_{\sim i}, \quad (2.9)$$

and the total effect index is defined as:

$$S_{Ti} = \frac{V_i^{\text{tot}}}{V} = 1 - \frac{V_{\sim i}}{V} \quad (2.10)$$

which describes the total variance accounted for by parameter i individually and in all possible interactions with other parameters. The set of S_i s and S_{Ti} s provide an efficient way to quantify the importance of individual parameters and interaction terms. The Sobol indices have a number of useful properties. The first-order effects quantify the amount of variance accounted for by uncertainty in each individual parameter. The sum of the first-order effects $\sum_i S_i$ will be ≤ 1 and the difference between $\sum_i S_i$ and unity provides a measure of the amount of variance which is accounted for by interactions. The ‘total effects’ quantify the variance accounted for by uncertainty in each parameter both on its own and in combination with uncertainty in other parameters. For any individual parameter, i , the difference between S_i and S_{Ti} indicates the extent to which it is involved in interactions with other parameters.

The integrals in (2.3) (2.4) and (2.6) can be evaluated using Monte Carlo integrals of the model outputs generated using random or quasi-random sampling of the model parameters. Using the procedure proposed by Saltelli [17], an estimate of both the first-order and total effects can be obtained at the cost of $N(k + 2)$ model evaluations, where N is of the order of a few thousand. The main weakness of variance-based techniques is their high computational cost. If the model is

expensive to evaluate or the number of parameters, k , is large, the use of variance-based techniques can be impractical. An alternative is the use of more efficient screening techniques such as the Morris method [14].

2.2.2. Morris’ screening design

Screening techniques have a lower computational cost than the variance-based techniques however the trade-off for this economy is that they tend to only provide qualitative measures of sensitivity, that is, they rank the parameters in terms of importance but give no information about how much more important one parameter is than another.

A number of screening designs have been proposed in the literature including the Morris method [14], iterated fractional factorial design (IFFD) [18], sequential bifurcation (SB) [19] and Cotter’s design [20]. IFFD and SB are both group-screening techniques [3] in which the model parameters are combined into groups prior to the analysis. IFFD produces good results only if a very small number of parameters determine the variability in the model output while SB requires the user to know the signs of the effects (i.e. whether a parameter has a positive or negative influence on the output) before the analysis is performed [21]. Cotter’s design may miss important parameters if they have effects on the model output which cancel each other out [15,21]. Morris’ method does not depend on any of these assumptions and is therefore regarded as the most widely applicable of the screening designs [22].

The Morris method is a one-at-a-time (OAT) approach. OAT designs are typically forms of SA in which the parameters are varied individually by small amounts around the nominal point. The results of the analysis only identify the model behaviour in the small region of the parameter space around this point and if the model contains strong nonlinearities selection of a different nominal point can produce vastly different outcomes. In addition, standard OAT approaches do not measure the effect of parameter interactions. These problems can be overcome by using the OAT method proposed by Morris [14].

Consider a model with k parameters where $\mathbf{P} = (p_1, p_2, \dots, p_k)$ is a vector of parameter values and $y(\mathbf{P})$ is the output of the model at parameter point \mathbf{P} . Each parameter may take one of q values. The Morris method defines the elementary effect of the i th parameter at nominal point \mathbf{P} as:

$$d_i = \frac{[y(p_1, \dots, p_{i-1}, p_i + \Delta, p_{i+1}, \dots, p_k) - y(\mathbf{P})]}{\Delta}, \quad (2.11)$$

where Δ is selected such that $\mathbf{P} + \Delta$ is still in the set of allowable values for each parameter k .

Morris’s method removes the dependence of OAT methods on the choice of nominal point, \mathbf{P} , and captures the importance of parameter interactions by calculating r elementary effects for each parameter at different nominal points, $\mathbf{P}_1, \dots, \mathbf{P}_r$. The nominal points are chosen such that each parameter is varied over its entire range. r is typically in the region of 10 [23,24]. The key to the Morris method is an efficient

design that requires $r(k+1)$ model runs to generate the r elementary effects for each of the k parameters [14]. Details of the algorithm can be found in the electronic supplementary material.

The mean and s.d. of the sample of r elementary effects for each parameter provide an approximate global sensitivity measure. They summarize the effects of varying the parameter across its entire range and capture the dependence of its sensitivity on the values of other parameters. A high mean, μ , indicates a parameter with an important overall effect on the output. A high s.d., σ , indicates a parameter with non-linear effects on the output or one which is involved in interactions with other parameters.

If the sample of elementary effects for a given parameter contains both positive and negative elements, that is the relationship between the parameter and the model output is non-monotonic, they may cancel out producing a low value of μ for an important parameter. To overcome this, it has been suggested [22] that the mean of the absolute values of the elementary effects, denoted μ^* should be used. This modified Morris measure has been shown empirically to be a good proxy for the total effect indices, S_{Ti} , of the variance-based measures [24].

It is also important to consider the scaling of the elementary effects. Because the calculation of the elementary effects involves division by the parameter step size Δ (see equation (2.10)) the size of Δ will influence the sensitivity results; parameters with small values will produce larger effects. This can cause the incorrect classification of the importance of parameters. The use of relative sensitivities is well established and has been emphasized in the context of the Morris method by Sin & Germaey [25]. The relative elementary effect for parameter i at point \mathbf{P} is given by:

$$d_i = \frac{[y(p_1, \dots, p_{i-1}, p_i + \Delta, p_{i+1}, \dots, p_k) - y(\mathbf{P})] \sigma_{p_i}}{\Delta \sigma_y}, \quad (2.12)$$

where σ_y and σ_{p_i} are the s.d. of the model output y (in this case the PC scores) and parameter p_i . The scaling of the elementary effects also removes the dependence on the magnitude of the model output. This allows the effects of a parameter on different outputs to be compared.

The use of the Morris and Sobol techniques in the analysis of time-dependent model outputs is compared by applying the methodology presented in §2.1 to a model of the insulin signalling pathway which is described below.

2.3. Insulin signalling model

The SA methodology described in §2.1 is applied to a model of the insulin signalling pathway. The actions of insulin are initiated when the hormone binds to its cell surface receptors triggering a signalling pathway which has pleiotropic effects in virtually all tissues [26]. The pathway is a key part of the blood glucose regulatory system and defects in the pathway are important in the development of insulin resistance, a major cause of type-2 diabetes. Identifying the parameters which drive

the dynamic behaviour of the model may further our understanding of the mechanisms underlying insulin resistance and other related pathologies.

The model is based on a previously published model [27] which has been used in a number of other studies [28–31] and was originally developed to study the regulation of glucose uptake by GLUT4, the major insulin-dependent glucose transport in fat and muscle tissue [32]. Here, the model has been modified as part of efforts to model the liver–pancreas blood glucose regulatory system using a modular approach to model construction [33]. While there is some evidence that insulin may regulate glucose transport in hepatocytes via increased internalization of GLUT2 [34], the primary role of insulin in the liver is to regulate the synthesis of glycogen. The model was therefore extended to consider the inactivation of GSK3, a key regulator of glycogen synthesis.

The model consists of three subsystems representing receptor binding, receptor recycling and the post-receptor signalling pathway. The processes included in the model are shown schematically in figure 2 and described briefly below. The mathematical details of the model can be found in the electronic supplementary material together with a more detailed description of the processes included in the model.

Binding of insulin to its receptor results in increased tyrosine kinase activity of the receptor towards its substrates. This activity is negatively regulated by dephosphorylation of the receptor by protein tyrosine phosphatases (PTPs) [35]. There are at least nine substrates of the insulin receptor four of which are varieties of the insulin-receptor substrate (IRS) protein [36]. IRS proteins are phosphorylated by the kinase activity of the receptor and act as docking sites for molecules which contain specific sequences known as SH2 domains including phosphoinositide 3-kinase (PI3K). The recruitment of PI3K to the plasma membrane places it in the vicinity of its physiological substrate, phosphatidylinositol (4,5) bisphosphate (PI(4,5)P2) which it phosphorylates to produce PI(3,4,5)P3 (PIP3) [37]. PIP3 binds a variety of signalling molecules including phosphoinositide-dependent kinase 1 (PDK1) and protein kinase B (PKB), also known as Akt, modifying their activity and intracellular location. The co-localization of these molecules allows PDK1 to phosphorylate and activate Akt which in turn phosphorylates and inactivates glycogen synthase kinase (GSK3). GSK3 downregulates the activity of glycogen synthase hence the inactivation of GSK3 results in an increased rate of glycogen synthesis.

3. RESULTS

This section presents the results of applying the new SA methodology described in §2.1 to the insulin signalling model to analyse the sensitivity of the GSK3 dynamics to uncertainty in the model parameters. The methodology produces a set of fPCs describing the key types of variation in the GSK3 output and, for each component, a set of sensitivity indices which measure the importance of the model parameters in producing the variation in GSK3 described by the corresponding component.

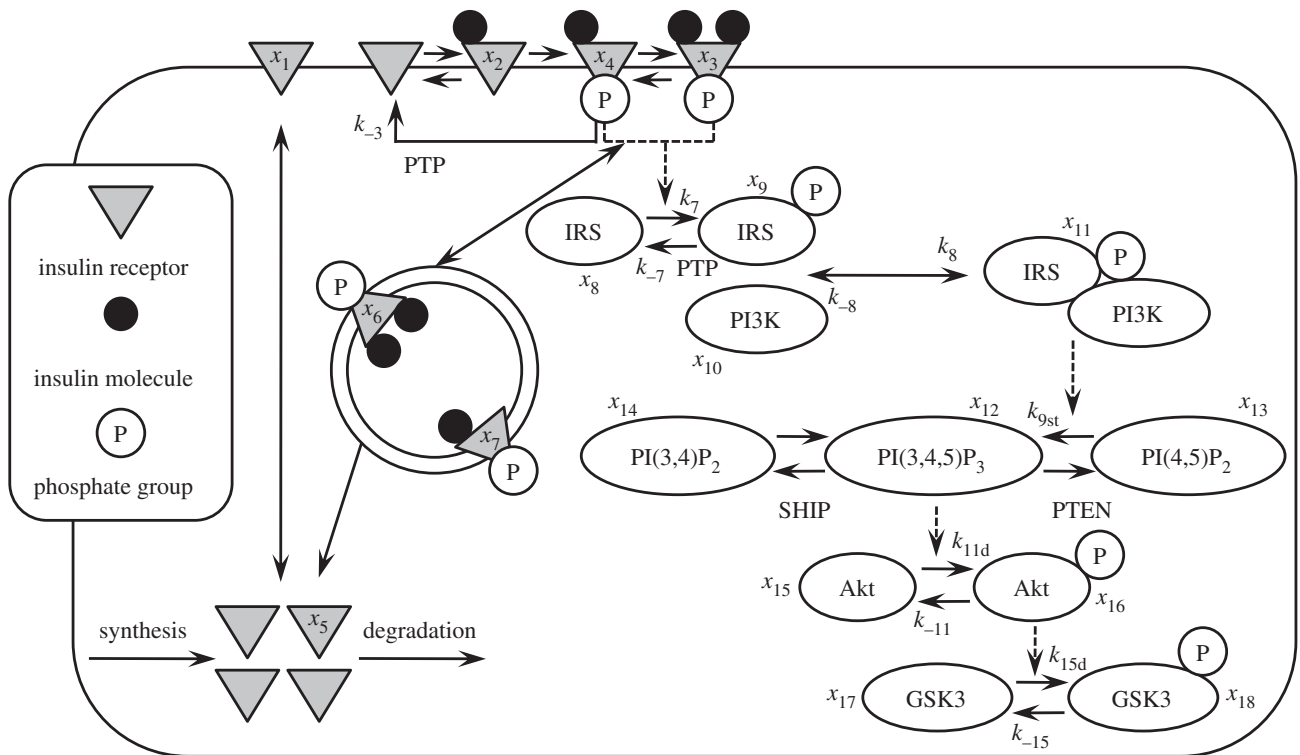


Figure 2. Schematic of the processes included in the insulin model. Triangles represent the various states of the insulin receptor, x_1 – x_7 . The ovals represent the species in the post-receptor signalling pathway that are included as state variables in the model. The parameters identified as important by the sensitivity analysis are shown in bold. Phosphorylated (P) and unphosphorylated forms are modelled as separate species. IRS, insulin receptor substrate; PI3K, phosphoinositide 3-kinase, PI(3,4), PI(3,4,5) and PI(4,5) are phosphoinositid lipids, Akt, protein kinase B, GSK3, glycogen synthase kinase.

The methodology is applied using both the Sobol ($n = 2000$) and Morris ($r = 20$; $q = 4$) methods and the results compared. For each method, 21 parameters were allowed to vary in a uniform range of $\pm 50\%$ of their nominal values (see supplementary material). The external input to the model, u , is the concentration of insulin and is modelled as a step function of magnitude $1 \times 10^{-6} M$ from $t = 0$ to $t = 30$ min. The model is evaluated to $t = 60$ min, sufficient time for dephosphorylation of GSK3 following removal of the insulin stimulus.

3.1. Principal components

The left-hand panels of figure 3 show the first three fPCs of the GSK3 dynamics generated via the Sobol method. The fPCs produced from the Morris method are qualitatively the same, that is, they describe the same types of variation in the GSK3 output (plots are included in the electronic supplementary material). It should be noted that because the parameter sampling is different for the two GSA methods, the PCs generated by the two methods may differ particularly if insufficient sample sizes are used. Together the first three PCs account for 99.7 per cent of the variation in the model output therefore we only present results for these three components. Other components describe minimal amounts of the variation in the model outputs.

The right-hand panel of figure 3 shows the mean GSK3 output (averaged over all model runs) plus and minus some multiple of each of the fPCs. Such plots are useful in visualizing the variation in the GSK3 time-course described by the PCs [38]. Here, they

clearly show that: the first PC, which captures 90.6 per cent of the variation, describes a vertical shift in the time-course of inactive GSK3 with the greatest increase in the steady-state concentration; the second PC, which accounts for a much smaller amount of variation (8.4%), captures variation in the dynamics after the stimulus is removed ($t > 30$) and GSK3 is reactivated owing to dephosphorylation; the third PC (0.7% of the total variation) describes variation in the inactivation phase of GSK3 by phosphorylation.

3.2. Sensitivity indices

Figure 4 shows the SA results. The results of both methods are consistent, identifying the same subset of 10 parameters as important. All the important parameters are involved in the post-receptor signalling pathway with the exception of k_{-3} , the rate of insulin dissociation from the receptor. This is consistent with the view that post-insulin receptor defects represent the primary sites leading to insulin resistance [39]. Studies have found that insulin binding is normal in diabetic individuals and no difference has been found in receptor numbers in over half of type 2 diabetic patients [40]. The lack of importance of the receptor recycling subsystem parameters supports the recent findings of Giudice *et al.* [41] that internalization of insulin receptors is not important in the activation of the Akt pathway.

The first PC primarily affects the maximal phosphorylation of GSK3. The production and breakdown of PI(3,4,5)P3 (k_{9st}) and the de/phosphorylation of

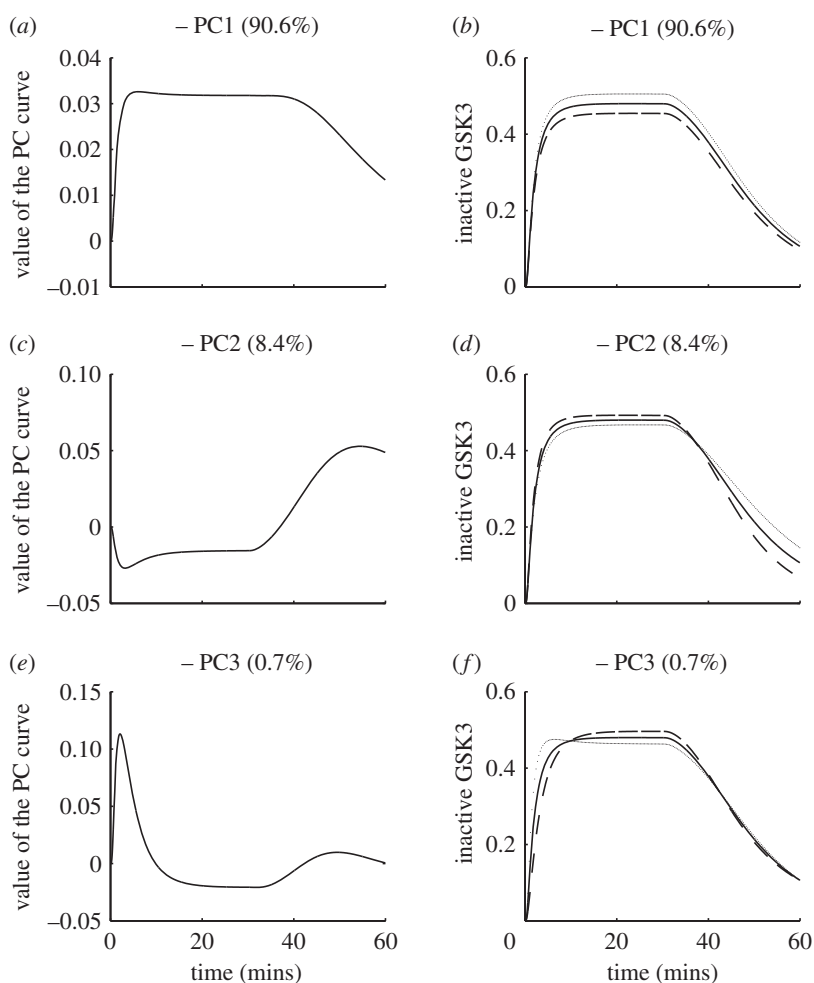


Figure 3. The first three functional principal components (PCs) of the GSK3 output from the insulin model. (a,c,e) The PC curves. (b,d,f) Show the mean GSK3 output averaged over all model runs (solid line) plus (dotted line) and minus (dashed line) a multiple of the PCs. The percentages show the amount of the total variation described by each component.

GSK3 (k_{-15} and k_{15}) are shown to be particularly important in producing variation in the scores on this component. The lipid phosphatases PTEN and SHIP2 which hydrolyse PI(3,4,5)P3 to PI(4,5)P2 and PI(3,4)P2, respectively, have been shown to be negative regulators of insulin signalling and have been proposed as possible therapeutic targets for type 2 diabetes [42,43]. The lesser importance of PI3K regulation (k_8 and k_{-8}) supports the experimental observations that reduced insulin-stimulated activation of PI3K does not affect the downstream activation of Akt [44,45]. There is little interaction between parameters as indicated by the minimal difference between the first-order and total effects for individual parameters and the difference between the sum of the first-order indices, $\sum_i S_i = 0.97$ and unity. Similarly, the low values of σ obtained via the Morris method imply that interactions between parameters are not significant.

The uncertainty in the second PC score is dominated by k_{-3} which accounts for approximately 65 per cent of the variation in this component according to the Sobol method. k_{-3} describes the deactivation of insulin receptors resulting from the dissociation of insulin and the dephosphorylation of the receptors by PTPs. This explains its importance in controlling the reactivation

of GSK3 following removal of the external insulin input (the behaviour described by the second PC). This result is consistent with experimental evidence that insulin signalling can be enhanced by reducing the activity of PTPs [46]. As a consequence of this, there has been a focus on the study of PTPs as potential therapeutic targets for the management of insulin resistance [35]. One particular study has shown that inhibition of PTP-1B expression in obese, insulin-resistant non-human primates can reduce fasting concentrations of glucose and insulin [47]. Figure 4 also shows that there is an increased role of interactions in the second PC ($\sum_i S_i = 0.74$, higher values for σ).

The third PC (which describes the initial phosphorylation of GSK3) is largely controlled by k_{-15} , the dephosphorylation rate of the kinase. This is in line with the view that processes downstream of Akt are crucial in propagating the insulin signal [48]. Hoehn *et al.* [49] have shown that only relatively small amounts of Akt phosphorylation are required to produce a maximal downstream response suggesting that defects in the regulation of Akt may be of less importance. As with the second PC, there is a significant interaction effect, especially for the parameters k_8 , k_{9st} , k_{11d} and k_{15d} (which describe association of the

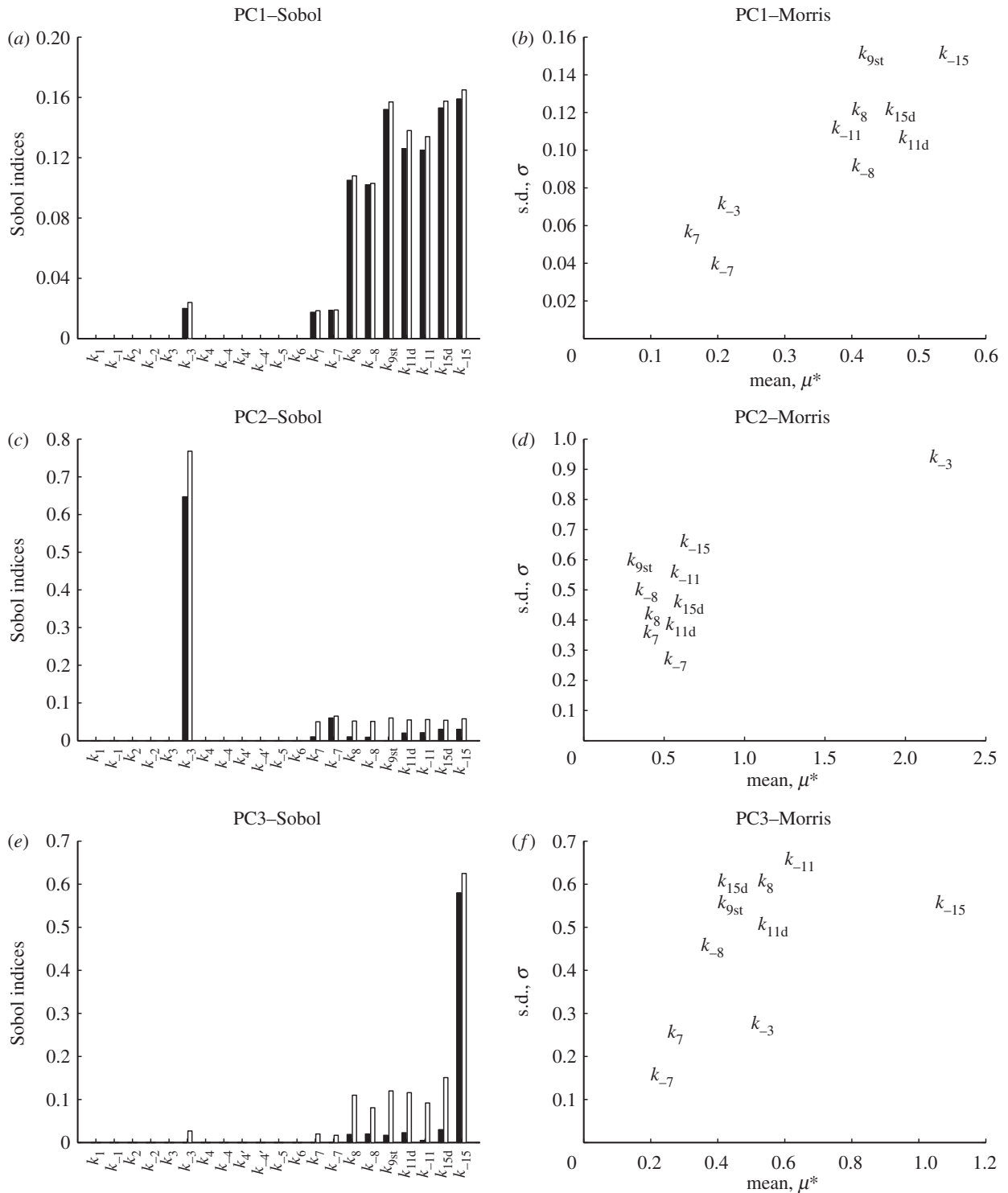


Figure 4. The sensitivities of the first three PCs of the insulin model output. (a,c,e) Show the first-order and total effect Sobol indices. The first-order indices quantify the amount of variance in the model output accounted for by each individual parameter. The total effects quantify the dependence on parameters both individually and in combination with the uncertainty in other parameters. This provides a measure of the extent of interaction between the parameters. (b,d,f) Show the mean (μ^*) and s.d. (σ) of the Morris measures. The mean captures the overall impact of uncertainty in a parameter on the model output (high values indicate important parameters). The s.d. indicates the extent to which the sensitivity to changes in a parameter is nonlinear or depends on the values of other parameters. (a,c,e) Black bars, first-order; grey bars, total.

IRS protein with PI3K, the conversion of PI(4,5)P2 to PI(3,4,5)P3, the phosphorylation (activation) of Akt and the rate of phosphorylation (inactivation) of GSK3, respectively) The importance of interactions highlights the need to use GSA methods to understand the behaviour of biological systems. Local methods, in

which parameters are varied one at a time, do not allow the effects of interactions to be explored.

3.3. Computation times

The Sobol method requires $N(k + 2)$ model evaluations. With $n = 2000$, $k = 21$ and a model evaluation time

on the order of 2 s, this equates to approximately 1.06 days of computation on a desktop computer. This is compared with a computational time of approximately 15 min for the Morris method ($r(k+1)$ model evaluations, $r=20$).

4. DISCUSSION

We have presented a new approach for the SA of time-dependent model outputs which combines fPCA with established global SA techniques. The application of the approach to a model of insulin-dependent GSK3 inactivation identified a number of interesting features of the system. The important model parameters, with the exception of the rate of dissociation of insulin from the receptor, were restricted to the post-receptor signalling sub-model supporting the view that defects in the intracellular pathway represent the primary sites leading to insulin resistance. The analysis also identified an important role for insulin receptor deactivation suggesting that the PTPs which dephosphorylate the receptor may be potential targets for therapeutic interventions. These results support many of the existing theories about the function of the insulin signalling network which have been uncovered through experimental studies.

Liu *et al.* [31] have also analysed a version of the Seda-ghat insulin pathway model [27] embedded in a larger model of glucose regulation. They performed a local SA (varying one parameter at a time) to calculate sensitivity indices at multiple time points in the simulated concentration of glucose. These indices were averaged to provide a single sensitivity measure for each parameter which describes its total effect on the glucose output. Consistent with our results they found that parameters governing PI(3,4,5)P3 conversion were particularly important. However, by combining fPCA with SA, we were able to identify different sets of parameters which affected different aspects of the model output. This is the key advantage of our methodology over the standard application of SA to time-dependent model outputs from which it is not so straightforward to infer the qualitative effect that parameters have on the model output. However, given that the main computational cost of most SA techniques is in the evaluation of the model and not the calculation of sensitivity indices, it is possible to calculate and compare both time-varying and PCA-based sensitivity indices for a model at minimal additional cost.

If we are interested in the sensitivity of a particular scalar feature of a dynamic model output (for example, peak concentration), it may be possible to evaluate that feature from each model run and calculate indices for it. This approach has been used in biological modelling but has its limitations. In their local SA of the oscillatory behaviour of the NF- κ B signalling pathway, Ihekwaba *et al.* [50] found that for certain parameter values, the chosen features (amplitude and period of oscillations) could not be evaluated. This would cause problems in GSA techniques such as the Sobol method, where we require the complete set of model outputs to calculate the sensitivity indices. By using fPCA, we can avoid this problem as the features that we perform the SA on are identified from the data itself. The limitation

of the fPCA-based approach is that it may not identify a type of output behaviour which is of biological interest if it does not contribute sufficiently to the variation in the model outputs. Similarly features of biological interest may be combined into single PCs.

A number of other studies have used mathematical modelling approaches to investigate insulin signalling. Giri *et al.* [28] used steady-state analysis of an insulin signalling model to investigate the translocation of GLUT4. They looked at the effect of changing the concentrations of various components in the model to replicate the variation in protein expression levels observed in pathological states. While we have focused on parameter sensitivities, GSA techniques can be applied to the initial conditions of a model in the same way to investigate similar questions. Cedersund *et al.* [51] and Brännmark *et al.* [52] used a model-based hypothesis testing approach, in which they compared different classes of models to experimental data, to identify the mechanisms which must be included to generate the changes observed in the insulin receptor substrate (IRS1) following exposure of cells to insulin. The GSA techniques presented in this paper are not commonly used to study the impact of qualitative inputs (such as model structure) on model outputs and instead typically focus on quantitative factors such as parameter values or initial conditions.

The selection of appropriate parameter distributions is an important part of global SA methods because the results of the analysis may be dependent on the choices [53]. This dependence on the choice of parameter distributions should be remembered when interpreting the results of the analysis of the insulin signalling model presented above. It can also be the most difficult and time-consuming stage of performing the analysis [3]. The choice of distribution is often governed by the availability of data but should also be guided by the purpose of the analysis. In the case of a biological or physiological system, a SA may have a number of different aims. If we are interested in understanding the behaviour of the system under normal conditions, we need to select ranges that represent the variation in the parameters observed in normal subjects. Alternatively, we may wish to investigate the important parameters in a particular disease state or condition. In this case, we should extend the ranges to include plausible values associated with the pathology of interest. More generally, we may be interested in investigating the parameters that the model output is sensitive to, for example, to identify potential targets for therapeutic interventions. In this type of analysis, the uncertainty distributions need not be based on the experimentally observed uncertainty in the model parameters. Instead, we may include parameters whose value is not regarded as uncertain to investigate the potential effects of artificially perturbing those parts of the system. A convenient form of input distribution in these cases is to adopt uniform ranges based on a percentage of the nominal parameter values.

The comparison of two SA methods, the Sobol method and Morris' screening design, has demonstrated that the Morris method is capable of providing qualitative sensitivity information for biological models which is

consistent with the Sobol method at a greatly reduced computational cost. The Morris method, therefore, provides a practical approach for the analysis of models with large numbers of parameters or long computational run times. It also offers a useful approach for studying parameter sensitivities under multiple scenarios (for example, different external model inputs) in a realistic time frame. When more quantitative sensitivities are required, for example, to assess the relative reduction in output variance which could be achieved by interventions that target different steps in a signal transduction pathway, variance-based techniques such as the method of Sobol must be used. The computational cost of a quantitative analysis can potentially be reduced by first performing parameter screening using the Morris method to identify non-influential parameters and thus reducing the dimension of the parameter space to be studied in further analysis.

Systematically exploring the impact of parameter uncertainty on model outputs is an important part of the modelling cycle. Global SA provides a powerful tool, which allows us to identify the model parameters, and hence biological mechanisms that drive the system outputs. This information may support the existing understanding of the system or be used to suggest new hypothesis which can be tested experimentally. The methodology presented here provides a new way to apply global SA to investigate the function of dynamic models of biological systems in a computationally efficient manner.

REFERENCES

- van Riel, N. A. 2006 Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform.* **7**, 364–374. (doi:10.1093/bib/bbl040)
- Zheng, Y. & Rundell, A. 2006 Comparative study of sensitivity analyses of the TCR-activated ERK-MAPK signalling pathway. *Syst. Biol.* **153**, 201–211.
- Saltelli, A., Chan, K. & Marian Scott, E. 2000 *Sensitivity analysis*. Chichester, UK: Wiley.
- Ihekwa, A. E., Broomhead, D. S., Grimley, R. L., Benson, N. & Kell, D. B. 2004 Sensitivity analysis of parameters controlling oscillatory signalling in the NF- κ B pathway: the roles of IKK and I κ B α . *Syst. Biol.* **1**, 93–103. (doi:10.1049/sb:20045009)
- Hu, D. & Yuan, J. 2006 Time-dependent sensitivity analysis of biological networks: coupled MAPK and PI3K signal transduction pathways. *J. Phys. Chem. A* **110**, 5361–5370. (doi:10.1021/jp0561975)
- Draper, N. & Smith, H. 1981 *Applied regression analysis*. New York, NY: Wiley.
- Sobol, I. M. 1993 Sensitivity estimates for non-linear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414.
- Saltelli, A., Tarantola, S. & Chan, P. S. 1999 A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **4**, 39–56. (doi:10.2307/1270993)
- Marino, S., Hogue, I. B., Ray, C. J. & Kirschner, D. E. 2008 A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J. Theor. Biol.* **254**, 178–196. (doi:10.1016/j.jtbi.2008.04.011)
- Liu, G., Swihart, M. T. & Neelamegham, S. 2005 Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signalling. *Bioinformatics* **21**, 1194–1202. (doi:10.1093/bioinformatics/bti118)
- Hetherington, J. P., Warner, A. & Seymour, R. M. 2006 Simplification and its consequences in biological modelling: conclusions from a study of calcium oscillations in hepatocytes. *J. R. Soc. Interface* **3**, 319–331. (doi:10.1098/rsif.2005.0101)
- Campbell, K., McKay, M. D. & Williams, B. J. 2006 Sensitivity analysis when model outputs are functions. *Reliab. Eng. Syst. Safety* **91**, 1468–1472. (doi:10.1016/j.res.2005.11.049)
- Morris, C. D., Larson, V. L. & Lounibos, L. P. 1991 Measuring mosquito dispersal for control programs. *J. Am. Mosq. Control. Assoc.* **7**, 608–615.
- Morris, M. D. 1991 Factorial sampling plans for preliminary computational experiments. *Technometrics* **33**, 161–174. (doi:10.2307/1269043)
- Homma, T. & Saltelli, A. 1996 Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Safety* **52**, 1–17. (doi:10.1016/0951-8320(96)00002-6)
- Saltelli, A., Ratto, M., Tarantola, S. & Campolongo, F. 2004 Sensitivity analysis for chemical models. *Chem. Rev.* **105**, 2811–2828. (doi:10.1021/cr040659d)
- Saltelli, A. 2002 Making the best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **145**, 280–297. (doi:10.1016/S0010-4655(02)00280-1)
- Andres, T. H. & Hajas, W. C. 1993 Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model. In *Joint Int. Conf. on Mathematical Methods and Supercomputing in Nuclear Applications* (eds H. Kusters, E. Stein & W. Werner), pp. 328–337. Karlsruhe, Germany: Kernforschungszentrum.
- Bettonvil, B. & Kleijnen, J. P. C. 1997 Searching for important factors in simulation models with many factors: sequential bifurcation. *Eur. J. Oper. Res.* **96**, 180–194. (doi:10.1016/S0377-2217(96)00156-7)
- Cotter, S. 1979 A screening design for factorial experiments with interactions. *Biometrika* **66**, 317–320. (doi:10.1093/biomet/66.2.317)
- Cacuci, D., Ionescu-Bujor, M. & Navon, I. 2003 *Sensitivity and uncertainty analysis: applications to large-scale systems*, vol. 2. Boca Raton, FL: CRC Press.
- Campolongo, F., Cariboni, J. & Saltelli, A. 2007 An effective screening design for sensitivity analysis of large models. *Environ. Model. Softw.* **22**, 1509–1518. (doi:10.1016/j.envsoft.2006.10.004)
- Saltelli, A. et al. 2004 *Sensitivity analysis in practice*. Chichester, UK: Wiley.
- Campolongo, F. & Saltelli, A. 1997 Sensitivity analysis of an environmental model: application of different analysis methods. *Reliab. Eng. Syst. Safety* **57**, 49–69. (doi:10.1016/S0951-8320(97)00021-5)
- Sin, G. & Gernaey, K. V. 2009 Improving the Morris method for sensitivity analysis by scaling the elementary effects. In *19th European Symp. on Computer Aided Process Engineering—ESCAPE19* (eds J. Jezowski & J. Thullie), pp. 925–930. Oxford, UK: Elsevier.
- Plum, L., Belgardt, B. & Bruning, J. 2006 Central insulin action in energy and glucose homeostasis. *Eur. J. Clin. Invest.* **116**, 1761–1766. (doi:10.1172/JC129063)
- Sedaghat, A. R., Sherman, A. & Quon, M. J. 2002 A mathematical model of metabolic insulin signalling pathways. *Am. J. Physiol. Endocrinol. Metab.* **283**, 1084–1101.

- 28 Giri, L., Mutalik, V. K. & Venkatesh, K. V. 2004 A steady state analysis indicates that negative feedback regulation of PTP1B by Akt elicits bistability in insulin-stimulated GLUT4 translocation. *Theor. Biol. Med. Model.* **1**, 2. (doi:10.1186/1742-4682-1-2)
- 29 Hori, S. S., Kurland, I. J. & DiStefano, J. J. 2006 Role of endosomal trafficking dynamics on the regulation of hepatic insulin receptor activity: models for Fao cells. *Ann. Biomed. Eng.* **34**, 879–892. (doi:10.1007/s10439-005-9065-5)
- 30 Chew, Y. H., Shia, Y. L., Lee, C. T., Majid, F. A., Chua, L. S., Sarmidi, M. R. & Aziz, R. A. 2009 Modelling of glucose regulation and insulin-signalling pathways. *Mol. Cell. Endocrinol.* **303**, 13–24.
- 31 Liu, W., Hsin, C. & Tang, F. 2009 A molecular mathematical model of glucose mobilization and uptake. *Math. Biosci.* **221**, 121–129. (doi:10.1016/j.mbs.2009.07.005)
- 32 Mueckler, M. 1994 Facilitative glucose transporters. *Eur. J. Biochem.* **219**, 713–725. (doi:10.1111/j.1432-1033.1994.tb18550.x)
- 33 Hetherington, J. *et al.* 2011 A composite model of liver glucose homeostasis. I. Building the composite model. *J. R. Soc. Interface* **9**, 689–700. (doi:10.1098/rsif.2011.0141)
- 34 Leturque, A., Brot-Laroche, E. & Le Gall, M. 2009 GLUT2 mutations, translocation and receptor function in diet and sugar managing. *Am. J. Physiol. Endocrinol. Metab.* **296**, E985–E992. (doi:10.1152/ajpendo.00004.2009)
- 35 Drake, P. G. & Posner, B. I. 1998 Insulin receptor-associated protein tyrosine phosphatase(s): role in insulin action. *Mol. Cell. Biochem.* **182**, 79–89. (doi:10.1023/A:1006808100755)
- 36 White, M. F. 1998 The IRS-signalling system: a network of docking proteins that mediate insulin action. *Mol. Cell. Biochem.* **182**, 3–11. (doi:10.1023/A:1006806722619)
- 37 Lizcano, J. M. & Alessi, D. R. 2002 The insulin signalling pathway. *Curr. Biol.* **12**, 236–238. (doi:10.1016/S0960-9822(02)00777-7)
- 38 Ramsay, J. O. & Silverman, B. W. 1997 *Functional data analysis*. New York, NY: Springer.
- 39 Pessin, J. E. & Saltiel, A. R. 2000 Signalling pathways in insulin action: molecular targets of insulin resistance. *J. Clin. Invest.* **106**, 165–169. (doi:10.1172/JCI10582)
- 40 DeFronzo, R. 2004 Pathogenesis of type 2 diabetes mellitus. *Med. Clin. North Am.* **88**, 787–835. (doi:10.1016/j.mcna.2004.04.013)
- 41 Giudice, J., Coluccio Leskow, F. C., Arndt-Jovin, D. J., Jovin, T. M. & Jares-Erijman, E. A. 2011 Differential endocytosis and signaling dynamics of insulin receptor variants IR-A and IR-B. *J. Cell. Sci.* **124**, 801–811. (doi:10.1242/jcs.076869)
- 42 Gupta, A. & Dey, C. 2009 PTEN and SHIP2 regulates PI3K/Akt pathway through focal adhesion kinase. *Mol. Cell. Endocrinol.* **309**, 55–62. (doi:10.1016/j.mce.2009.05.018)
- 43 Sasaoka, T., Wada, T. & Tsuneki, H. 2006 Lipid phosphatases as a possible therapeutic target in cases of type 2 diabetes and obesity. *Pharmacol. Therapeutics* **112**, 799–809. (doi:10.1016/j.pharmthera.2006.06.001)
- 44 Kim, Y. B., Nikoulina, S. E., Ciaraldi, T. P., Henry, R. R. & Kahn, B. B. 1999 Normal insulin-dependent activation of Akt/protein kinase B, with diminished activation of phosphoinositide 3-kinase in type 2 diabetes. *J. Clin. Invest.* **104**, 733–741. (doi:10.1172/JCI6928)
- 45 Choi, K. & Kim, Y. B. 2010 Molecular mechanism of insulin resistance in obesity and type 2 diabetes. *Korean J. Intern. Med.* **25**, 119–129. (doi:10.3904/kjim.2010.25.2.119)
- 46 Goldstein, B. J., Ahmed, F., Ding, W., Li, P. M. & Zhang, W. R. 1998 Regulation of the insulin signalling pathway by cellular protein-tyrosine phosphatases. *Mol. Cell. Endocrinol.* **182**, 91–99. (doi:10.1023/A:1006812218502)
- 47 Swarbrick, M. *et al.* 2009 Inhibition of protein tyrosine phosphatase-1B with antisense oligonucleotides improves insulin sensitivity and increase adiponectin concentrations in monkeys. *Endocrinology* **150**, 1670–1679. (doi:10.1210/en.2008-0885)
- 48 Brady, M. J. & Saltiel, A. R. 1999 Closing in on the cause of insulin resistance and type 2 diabetes. *J. Clin. Invest.* **104**, 675–676. (doi:10.1172/JCI8216)
- 49 Hoehn, K., Hohnen-Behrens, C., Cederberg, A., Wu, L. E., Turner, N., Yuasa, T., Ebina, Y. & James, D. E. 2008 IRS1-independent defects define major nodes of insulin resistance. *Cell Metab.* **7**, 421–433. (doi:10.1016/j.cmet.2008.04.005)
- 50 Ihekweaba, A. E., Broomhead, D. S., Grimley, R., Benson, N., White, M. R. & Kell, D. B. 2005 Synergistic control of oscillations in the NF-kappaB signalling pathway. *IEEE Proc. Syst. Biol.* **152**, 153–160. (doi:10.1049/ip-syb:20050050)
- 51 Cedersund, G., Roll, J., Ulfhielm, E., Danielsson, A., Tidefelt, H. & Strålfors, P. 2008 Model-based hypothesis testing of key mechanisms in initial phase of insulin signalling. *PLoS Comput. Biol.* **4**, e1000096.
- 52 Brannmark, C., Palmér, R., Torkel Glad, S., Cedersund, G. & Strålfors, P. 2010 Mass and information feedbacks through receptor endocytosis govern insulin signalling as revealed using a parameter-free modeling framework. *J. Biol. Chem.* **285**, 20 171–20 179. (doi:10.1074/jbc.M110.106849)
- 53 Lipton, J., Shaw, W. D., Holmes, J. & Patterson, A. 1995 Short communication: selecting input distributions for use in monte carlo simulations. *Regul. Toxicol. Pharmacol.* **21**, 192–198.