Published in final edited form as:

Bull Math Biol. 2012 January ; 74(1): 73–90. doi:10.1007/s11538-011-9664-2.

MARS Approach for Global Sensitivity Analysis of Differential Equation Models with Applications to Dynamics of Influenza Infection

Yeonok Lee and Hulin Wu*

Department of Biostatistics and Computational Biology The University of Rochester, 601 Elmwood Avenue Box 630 Rochester, New York 14642, USA

Abstract

Differential equation models are widely used for the study of natural phenomena in many fields. The study usually involves unknown factors such as initial conditions and/or parameters. It is important to investigate the impact of unknown factors (parameters and initial conditions) on model outputs in order to better understand the system the model represents. Apportioning the uncertainty (variation) of output variables of a model according to the input factors is referred to as sensitivity analysis. In this paper, we focus on the global sensitivity analysis of ordinary differential equation (ODE) models over a time period using the multivariate adaptive regression spline (MARS) as a meta model based on the concept of the variance of conditional expectation (*VCE*). We suggest to evaluate the *VCE* analytically using the MARS model structure of univariate tensor-product functions which is more computationally efficient. Our simulation studies show that the MARS model approach performs very well and helps to significantly reduce the computational cost. We present an application example of sensitivity analysis of ODE models for influenza infection to further illustrate the usefulness of the proposed method.

Keywords

Global sensitivity analysis; Variable selection; Variance decomposition; Multivariate adaptive regression splines; Ordinary differential equation model

1 Introduction

Ordinary differential equation (ODE) models are popularly used in many scientific fields. Most of the studies involve unknown parameters and/or initial values that need to be estimated from experimental data. The solutions to ODE models are usually called the output variables and the undetermined parameters and/or initial values are often referred to as input factors (variables). In practical applications, sensitivity analysis is frequently used to study how the variation of output variables of a model can be apportioned according to the input factors (Saltelli et al., 2000, Chapter1). In other words, the sensitivity analysis is used to assess the relevant importance of each input factor and to identify the factors with significant impacts on model output variables.

In general, the techniques of assessing the importance of input factors on output variables can be classified into local and global sensitivity analysis. The local sensitivity analysis (LSA) addresses the sensitivity problem for a small interval around nominal values of input

*Corresponding author *Phone:* (585)275-6767 *Fax:* (585) 273-1031 *HWu@bst.rochester.edu.*

variables. The LSA typically involves the partial derivatives of the output variables with respect to an input factor while other input factors are fixed. On the other hand, the global sensitivity analysis (GSA) is used if the study interest lies in the sensitivity for the entire range of input factors while simultaneously varying other factors (Saltelli et al., 2000).

In this paper, we focus on GSA, which usually employs sampling-based methods. A standard nonlinear functional model can be written as

$$y=f(x),$$

where f is a real-valued function, $x = (x_1, \dots, x_p)$ is a vector of input factors with dimension p , and y is the corresponding output of the nonlinear model. Input factor samples $x_i = (x_{i1}, \dots, x_{ip})'$, $i = 1, \dots, n$, can be generated from pre-specified distributions that reflect the prior knowledge of input factors. There are many sampling techniques available (Helton and Davis, 2003) and the Latin Hypercube sampling method (McKay et al., 1979) is widely used. The output y is evaluated either numerically or analytically for given input factor samples, and then the association between the output and input factors can be studied.

In specific, we are interested in GSA for ordinary differential equation models (ODE) with a general form

$$y' = g(y, \theta, t), \quad y(t_0) = y_0, \quad (1)$$

where $y' = dy/dt$ is the derivative of a function $y(t)$ with respect to t often indicating the time, θ is a vector of unknown parameters, and $y(t_0) = y_0$ is a vector of initial values for the state variables $y(t)$. Let $x = (\theta, y_0)$ denote the vector of input factors with dimension p . For a general nonlinear ODE model, the closed-form solution may not exist. However, we can numerically solve the ODE model (1) for given input factors and obtain a time trajectory (function) of an output variable (here we only consider a univariate output variable for simplicity). Then we can rewrite the input-output relationship in ODE model (1) in terms of its numerical solution f

$$y_{ij} = f(x_i, t_j), \quad j = 1, \dots, q \quad (2)$$

for a given set of input factors x_i and time t_j for $i = 1, \dots, n$. Then the sampling-based approach can be applied to the ODE model (1) for sensitivity analysis.

Various GSA methods are applicable to evaluate the impact of input factors to the output variables for ODE models at a given time. Blower and Dowlatabadi (1994) pioneered the GSA for ODE models of HIV using the partial rank correlation coefficients (PRCC) (Helton et al., 2006) under the assumption of the nonlinear monotonic input-output relationship. The PRCC is recognized as the most reliable method among the various correlation measures. Recently, Marino et al. (2008) emphasized the advantages of GSA over LSA in accounting for uncertainty of parameter estimates due to natural variation and measurement errors in systems biology. Marino et al. (2008) focused on two popular GSA approaches, PRCC and the extended Fourier Amplitude Sensitivity analysis (eFAST) methods (Saltelli et al., 1999). The eFAST methods, including other variance-based GSA techniques such as correlation ratio (McKay, 1995) and Sobol's total effect indices (Sobol', 1993), are considered as informative and intuitive measures for nonlinear non-monotonic input-output relationships. Note that the impact of each input factor on the output variable was studied for fixed time points in Marino et al. (2008). Other popular GSA approaches are reviewed and developed into software tools by Hoare et al. (2008).

We consider two importance measures based on the variance of conditional expectation (VCE). The VCE plays a key role in formulating the importance indices of input factors. Note that $Var(y|t=t^*)$ can be decomposed into $Var[E(y|x_U, t=t^*)] + E[Var(y|x_U, t=t^*)]$, where x_U is a subset of input factors. Denote x_U^C the complement of x_U (the rest of input factors). The term $Var[E(y|x_U, t=t^*)]$, referred to as $VCE(x_U)$, relates how well x_U controls y while the term $E[Var(y|x_U, t=t^*)]$ is considered as a residual term, explaining the remaining variation in y due to other input factors or unknown sources.

One popular importance measure is to sequentially evaluate and rank the importance of input factors. For a given time $t=t^*$, $\eta^2(x_U|t^*)$, the sequential importance measure of the input factors x_U , is defined in Storlie and Helton (2008b):

$$\eta^2(x_U|t^*) = \frac{VCE(x_U, t=t^*)}{Var(y|t=t^*)}, \quad (3)$$

where $y = f(x, t=t^*)$. This measure can be sequentially applied to a single input factor one by one first so that we can obtain $\eta^2(x_j)$, $j = 1, \dots, p$. The input factor resulting in the largest $\eta^2(x_j)$ is ranked as the most important factor, say x_1^* . The second important input factor is the one that leads to the largest $\eta^2(x_1^*, x_j)$, where $x_j \neq x_1^*$. The j th important factor x_j is determined as the one that has the largest $\eta^2(x_U|t^*)$, where $x_U = \{x_1^*, \dots, x_{j-1}^*, x_j\}$ with $x_j \neq x_1^*, \dots, x_{j-1}^*$. The procedure continues until all input factors are assigned to a rank.

Alternatively, a single factor importance measure at a specified time point $t=t^*$ is given by Homma and Saltelli (1996),

$$T_j(t^*) = \frac{E[Var(y|x_{(-j)}, t=t^*)]}{Var(y|t=t^*)} = 1 - \frac{VCE(x_{(-j)}|t=t^*)}{Var(y|t=t^*)}, \quad (4)$$

where $x_{(-j)} = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p\}$.

The variance-based measures of importance evaluated by sampling-based methods can be computationally intensive. Especially if the evaluation of model output demands heavy computational cost, the sensitivity analysis using sampling-based methods may not be feasible. For example, the VCE needs to be obtained in two stages. First, the conditional expectations are evaluated for given (x_U, t^*) and then the variation of the conditional expectation is evaluated. In addition, this procedure needs to be repeated for $p(p+1)/2 - 1$ times to determine the ranks using the sequential importance measure. Although some efficient sampling techniques (See reviews in Helton and Davis (2003)) have been developed to moderate the computational burden, the computational cost is still substantial. Alternatively, a meta modeling approach can be used to avoid the computational burden. The basic idea of the meta modeling approach is to approximate the complex original model such as ODE models of which the closed-form solution is not available, by simple regression models, i.e.,

$$y_{ij} = \tilde{f}(x_i, t_i) + e_{ij} \quad (5)$$

where $\tilde{f}$ is an approximation of f and e_{ij} are approximation errors with mean zero for $i = 1, \dots, n$ and $j = 1, \dots, q$. Notice that here the independent variables $x = (x_1, \dots, x_p)$ are input factors such as unknown parameters and initial values of the original ODE model. The relationship between the output variable y and these input factors is generally nonlinear with an unknown form. But for any given values of input factors (parameters and initial values of

the ODE model), the output variable can be obtained at different time points t_j on the ODE solution trajectory by numerically solving the ODE. Thus, we can obtain a data set of independent (covariate) variables of sample size n using a sampling-based method as well as the corresponding output (response) variable at q time points by numerical simulations of the original ODE model.

A linear, nonlinear or nonparametric regression models can be used to approximate the relationship between the input factors and the output variable in model (5). Recently, Storlie and Helton (2008a) and Storlie and Helton (2008b) proposed nonparametric smoothing methods that impose a flexible input-output relationship assumption as opposed to a linear or nonlinear regression. Various nonparametric models are examined in Storlie et al. (2009) for their performance on estimating the true importance measure. Among these nonparametric approaches, the multivariate adaptive regression spline (MARS) model (Friedman, 1991) is one of the best performing models.

The MARS model (Friedman, 1991) is a nonparametric approach that allows both a single factor and multiple-factor interaction effects to be flexibly included in the model. In addition, the MARS modeling approach screens out unimportant factors using a variable selection approach, which helps to reduce the number of factors for further sensitivity analysis. This feature is especially useful when there are a large number of insignificant input factors in the model. Also, since the MARS model is of the form of univariate tensor products, the importance measures, η^2 and T_j , can be analytically evaluated (Chen et al., 2005) under the assumption that the input factors are independent of each other. Especially, as the entire ranking procedure using both importance measures only needs one sample set, instead of $p(p+1)/2 - 1 + p$ sample sets in the sampling-based method, the meta-model approach reduces the computational cost of solving ODEs significantly.

The importance measures, η^2 and T_j , are time-varying for ODE models. In Section 2, we provide a brief introduction to the MARS model (Friedman, 1991) and an analytic method to calculate the VCE (Chen et al., 2005). We propose an approach that measures input factors' sensitivity over the time range of interest by integrating important measures over time. In Section 3, a simulation study is performed to evaluate the efficiency and accuracy of the MARS modeling approach and the sampling-based method. We apply the proposed MARS modeling approach to carry out the global sensitivity analysis for two differential equation models for influenza infection in Section 4. We conclude our paper with some discussions in Section 5.

2 MARS Model and Analytic VCE Method

2.1 MARS Algorithm

Define y_{ij} as the output generated from Eq(2) corresponding to x_i (the i th sample of input factors) at time t_j for $i = 1, \dots, n$ and $j = 1, \dots, q$. For notational simplicity, we assume that the observation time points $t_1, \dots, t_q$ are the same for all n samples of input factors, but they are not necessarily equally spaced. Denote the k th input factor of the i th sample by x_{ik} for $k = 1, \dots, p$, where p is the number of input factors. The MARS modeling approach uses the stepwise model selection method to build a nonparametric regression spline model including main effects and interaction terms of both input factors x_i and time t on the output variable y_{ij} , so that we can obtain $\hat{f}$, an estimate of function f in model (5). In this section, we consider time t as another input factor without loss of generality, and that the number of input factors becomes $p+1$. For completeness, we briefly review the MARS approach (Friedman, 1991) and introduce the relevant notation which is also useful in later sections.

The MARS model approach can be completed in three steps (Friedman, 1991; Zhang, 1997). First define a collection of basis functions $\mathfrak{B}$ by

$$\mathfrak{B} = \{b^+ \text{ and } b^- | b^+ = (x_k - \tau)^+ \text{ and } b^- = (x_k - \tau)^-, \tau \in (-\infty, \infty), k=1, \dots, p+1\},$$

where τ is one of the values in x_{jk} , called a knot, and $a^+ = \max(0, a)$ and $a^- = \max(0, -a)$ for any a .

Step 1. Select the initial pair of basis functions—For any given pair of basis functions $(x_k - \tau)^-$ and $(x_k - \tau)^+$ and the selected best knot τ_k , we can fit a linear regression model

$$y_{ij} = \beta_0 + \beta_1(x_k - \tau_k)^- + \beta_2(x_k - \tau_k)^+ + e_{ij} \quad (6)$$

to obtain the estimates of the coefficients β_0 , β_1 and β_2 using the least square method. The covariate will be selected in the initial spline model if its fitted Sum of Square of Residuals

(SSR), $SSR = \sum_i^n \sum_j^q [y_{ij} - \hat{f}(x_i, t_j)]^2$, is smallest among $p+1$ covariates, where $\hat{f}$ is the fitted model. Note that the knot selection methods can be found in Friedman (1991) and Zhang (1997). Denote the selected input factor by x_{k_1} .

Step 2. Enlarge the spline model—Next, the model is extended to one of the three candidates:

$$y_{ij} = \beta_0 + \beta_1 b_1^- + \beta_2 b_2^+ + \beta_3(x_{k'} - \tau_{k'})^- + \beta_4(x_{k'} - \tau_{k'})^+ + e_{ij}, \quad (7)$$

$$y_{ij} = \beta_0 + \beta_1 b_1^- + \beta_2 b_2^+ + \beta_3 b_1^-(x_{k'} - \tau_{k'})^- + \beta_4 b_2^+(x_{k'} - \tau_{k'})^+ + e_{ij}, \quad (8)$$

$$y_{ij} = \beta_0 + \beta_1 b_1^- + \beta_2 b_2^+ + \beta_3 b_2^+(x_{k'} - \tau_{k'})^- + \beta_4 b_1^-(x_{k'} - \tau_{k'})^+ + e_{ij}, \quad (9)$$

The three models are fitted separately for all covariates but $x_{k'} = x_{k_1}$ for (8) and (9). Note that this condition does not apply to (7) since the new terms are not products of basis functions. This condition keeps the MARS model to be in the form of univariate tensor product functions. Then the model with the smallest SSR will be the next extended model. This process repeats until the number of terms reaches the pre-specified maximum number or until the model does not improve by adding more terms in the sense of reduction in SSR. Friedman (1991) suggested to have the pre-specified maximum number of terms to be roughly twice as many as the optimal number of terms of the smallest GCV.

Step 3. Backward deletion—Let M^* be the number of terms included in the model from the forward addition procedure in Step 2. In the backward deletion step, M^*-1 models are sequentially built by deleting a nonconstant term one by one using the generalized cross-validation (GCV) score defined as

$$GCV(m^*) = \frac{1}{n \cdot q} \sum_{i=1}^n \sum_{j=1}^q [y_{ij} - \hat{f}_{m^*}(x_i, t_j)]^2 / \left(1 - \frac{C(m^*)}{n \cdot q}\right)^2, \quad (10)$$

where $\hat{f}_{m^*}$ is the reduced model with m^* terms and $C(m^*) = m^* + d \cdot (m^* - 1)/2$ with d being the penalization parameter for model complexity. The choice of d is upon researchers' decision and the best values are in the range $2 \leq d \leq 4$ (Friedman, 1991).

The final MARS model is the one with the smallest GCV.

2.2 Analytic VCE from the MARS model

Let M be the number of terms in the final MARS model. Note that we consider the time t as another covariate x_{p+1} for the ODE model sensitivity analysis although it is not our interest to evaluate the sensitivity of time on the output variable. Let $B_k(x, t)$ denote the k th term in the final MARS model, which can be expressed as $p+1$ products of univariate functions for $k = 1, \dots, M$. The MARS model can be written as

$$\widehat{f}(x, t) = \sum_{k=1}^M B_k(x, t) = \sum_{k=1}^M a_k \prod_{l=x}^{p+1} b_{kl}(x_l),$$

where a_k are the estimated coefficients in the final MARS model and $b_k(x_j)$ is the univariate basis function of k th factor in k th term. We employ the analytic method as Chen et al. (2005) suggested to evaluate the VCE based on the MARS model in the form of univariate tensor-product functions.

Denote the probability density of x_j by $p(x_j)$, the expectation of $b_{ij}(x_j)$ by $E_{ij} = \int b_{ij}(x_j)p(x_j)dx_j$, and the inner product of $b_{i1,j}(x_j)$ and $b_{i2,j}(x_j)$ by $C_{i1,i2,j} = \int b_{i1,j}(x_j)b_{i2,j}(x_j)p(x_j)dx_j$. Under the assumption that the input factors are independent of each other, $E(\widehat{f})$ and $E(\widehat{f}(x|x_U))$ are $E(\widehat{f}) = \sum_i a_i \prod_j^{p+1} E_{ij}$ and $E(\widehat{f}(x|x_U)) = \sum_i a_i \prod_{j \notin U} E_{ij} \prod_{j \in U} b_{ij}(x_j)$ respectively. Consequently, $\widehat{VCE}(x_U)$ and $Var(\widehat{f})$ can be expressed in terms of E_{ij} and $C_{i1,i2,j}$ as follows:

$$\widehat{VCE}(x_U) = Var(E(\widehat{f}(x|x_U))) = \left(\sum_{i_1}^M \sum_{i_2}^M a_{i_1} a_{i_2} \prod_{j \notin U} E_{i_1 j} E_{i_2 j} \prod_{j \in U} C_{i_1, i_2, j} \right) - E(\widehat{f})^2, \quad (11)$$

and

$$Var(\widehat{f}) = \left(\sum_{i_1}^M \sum_{i_2}^M a_{i_1} a_{i_2} \prod_{j=1}^{p+1} C_{i_1, i_2, j} \right) - E(\widehat{f})^2, \quad (12)$$

where a_{i1} and a_{i2} are estimated coefficients in the final MARS model.

Now we extend the standard VCE to integrate time t for ODE model sensitivity analysis by treating the time t as an additional covariate or input factor. Denote the variance of conditional expectation for given x_U and $t = t^*$ by $VCE(x_U, t = t^*)$. The variance of conditional expectation given x_U over a range of time interval can be expressed as

$$VCE(x_U) = \int_t VCE(x_U, t=t^*) dt = \int \int [E(y|x_U, t=t^*) - E(y)]^2 p(x_U) dx_U dt.$$

We can similarly define the overall sequential importance measure and the single importance measure over a time interval by

$$\eta^2(x_U) = \frac{VCE(x_U)}{Var(y)} \quad (13)$$

and

$$T_j = 1 - \frac{VCE(x_{(-j)})}{Var(y)}, \quad (14)$$

respectively, where y is an observation or output vector at any time point and $x_{(-j)} = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p, t\}$. Note that $VCE(x_U)$ and $Var(y)$ in the above formulas can be estimated by $\widehat{VCE}(x_U)$ and $Var(\widehat{f})$ using Eq(11) and Eq(12), respectively, and that $\widehat{\eta}^2(x_U)$ and $\widehat{T}_j$ are obtained subsequently.

The MARS model-based sensitivity analysis procedure for ODE models can be summarized as follows: 1) Determine the distributions of input factors, and then sampling input factors from their distributions to obtain n samples. 2) For each sample of input factors, solve the ODE to obtain numerical solutions for output variables at time points $t = t_1, t_2, \dots, t_q$. 3) Perform the MARS model fitting based on the data of input factors and the output variables generated from Steps 1 and 2. 4) Calculate the VCE and find the sequential and single importance measures over a time interval analytically based on the established MARS model as proposed above. 5) Finally, rank the importance or sensitivity of input factors based on one of the importance measures in Eq(13) or Eq(14).

3 Simulation study

In this section, we design a simulation study to evaluate the performance of the MARS model-based sensitivity analysis approach. We examine the goodness-of-fit and predictability of the MARS model for data generated from a differential equation model, and then we evaluate whether the MARS modeling approach can correctly identify the important input factors and compare the results to those of the computationally-intensive sampling-based GSA methods, which will be considered as a benchmark in this article.

First we evaluate the goodness-of-fit of the MARS model using the standard R^2 ,

$$R^2 = 1 - \frac{SSR}{Var(y) \cdot (nq - 1)}, \quad (15)$$

where SSR is the sum of squares of residuals. If R^2 is close to 1, it indicates a good model to represent the data. To evaluate the predictability of the MARS model, we employ a measure of D^2 which is defined as

$$D^2 = 1 - \frac{SSD}{Var(y) \cdot (nq - 1)}, \quad (16)$$

where $SSD = \sum_{i,j} (y_{ij}^{new} - \widehat{f}(x_i^{new}, t_j))^2$ with y_{ij}^{new} and x_i^{new} from a new data set. Having the fitted MARS model, we generate another set of input factors and compare the actual outputs (y_{ij}^{new}) with the fitted values $\widehat{f}(x_i^{new}, t_j)$. When both R^2 and D^2 are close to one, we expect that the established MARS model is a good approximation to the original complex ODE model, which can be reliably used to assess the importance and sensitivity of input factors on the output variables. The small difference between R^2 and D^2 indicates how well the simplified model represents the original model.

In the simulation study, we select a simple univariate nonlinear differential equation model that has an analytic solution (Saltelli et al., 2000, page 44) so that we can minimize the

confounding factors when we evaluate the sensitivity and importance of input factors to output variables:

$$\frac{dY}{dt} = -2x_1 \exp(-x_2/300) Y^2,$$

where x_1 and x_2 are parameters in this ODE model and we set initial value of $Y(t)$ as $Y(0) = x_3$. Let $\mathbf{x} = (x_1, x_2, x_3)'$ be input factors and $Y(t)$ be the output variable. Our objective is to evaluate the importance of the input factors $\mathbf{x} = (x_1, x_2, x_3)'$ on the output variable $Y(t)$. Note that the analytic solution to the above ODE model is

$$Y(\mathbf{x}, t) = (2x_1 \exp(-x_2/300)t + 1/x_3)^{-1}.$$

We assume that the input factors follow Uniform distributions with ranges $x_1 \sim U(8.97 \cdot 10^6, 3.59 \cdot 10^7)$, $x_2 \sim U(0, 1000)$, $x_3 \sim U(1.0, 1.2)$, and the range of time t is $[10^{-10}, 10^{-3}]$. Note that the sensitivity of time t is not our interest. We can consider time t as either deterministic and equally-spaced or random from a Uniform distribution.

A sampling-based method with a fairly large sample size is expected to approximate the true importance measures well and can be served as a comparison basis for the meta-model sensitivity analysis method. We employ the Latin hypercube sampling method (McKay et al., 1979) to generate 500^k samples of $\mathbf{x}_U$, where k is the number of factors in $\mathbf{x}_U$. We assume that time t is deterministic and equally spaced on $[-10, -3]$ on a log-scale. The conditional expectations are evaluated numerically for given $\mathbf{x}_U$ at 500 equally spaced time points, and finally the variance of the conditional expectation is estimated. This entire procedure is repeated for six different subsets of $\mathbf{x}_U$.

Based on the *VCE* estimates from the sampling-based method described above, we can calculate the sequential importance measure over the time interval: $\eta^2(x_1) = 0.9477$, $\eta^2(x_2) = 0.9879$, $\eta^2(x_3) = 0.9455$, $\eta^2(x_2, x_1) = 0.9959$, and $\eta^2(x_2, x_3) = 0.9927$. The single importance measure is calculated as $T_1 = 0.0073$, $T_2 = 0.04769$, and $T_3 = 0.0041$ for x_1 , x_2 and x_3 , respectively. Both the sequential and single importance measures η^2 and T_j produce the same ranking in this simulation study, i.e., x_2 , x_1 and x_3 . The input sample size 500^k for the sampling-based method is quite large for global sensitivity analysis which is necessary in order to achieve reliable results. We expect that the meta-model approach such as the MARS model-based sensitivity analysis approach can reduce the computational cost significantly while it can also achieve consistent importance ranking results compared to that of the sampling-based approaches. We evaluate this claim as follows.

For the MARS meta-model approach, we can use much smaller sample sizes of input factors to establish the MARS model that can be used to approximate the relationship between the input factors and the output variable. We take $n = 25, 50, 150, 300$, and 600 random samples of input factors from the uniform distributions as introduced above by the Latin hypercube sampling method (McKay et al., 1979) and consider the time t as deterministic and equally-spaced time points with the size of $q = 5, 10, 25, 50$, and 100 , respectively. In addition, we also introduce three dummy factors, x_4 , x_5 , and x_6 that do not affect the output variable $Y(\mathbf{x}, t)$, which are generated from the same uniform distributions as those of x_1 , x_2 and x_3 , respectively. These dummy factors are used to test if the MARS model is capable of selecting significant factors using the variable selection procedure, referred to as the screening capability. It is expected that the fitted MARS model does not include the dummy factors.

The MARS model is fitted to each data set with the pre-determined maximum number of terms 100, the maximum four-way interaction terms are allowed, and the penalization parameter of model complexity as $d = 3$. The resulting MARS model for each data set fits the data very well (R^2 ranging from 0.9838 to 0.9959). The predictability of the established MARS model is evaluated using a new data set with 100 samples of input factors at 20 random time points. The D^2 is very close to the corresponding R^2 and the maximum difference of $R^2 - D^2$ is about 0.016. This indicates that the established MARS model approximates the original ODE model very well.

For these simulated data sets, the corresponding MARS models select the significant factors correctly, although not all rankings are consistent with the results from the sampling-based method as presented above. For each combination of input sample sizes and number of time points, the estimated importance measures using the MARS model are reported in Table I. The ranks from the MARS model approach are either (x_2, x_3, x_1) or (x_2, x_1, x_3) . The shaded cells in Table I indicate the data sets from which the MARS model consistently identifies the ranking with that of the computationally-intensive sampling-based method, i.e., (x_2, x_1, x_3) . Note that the other factors x_4 , x_5 , and x_6 are not included in any of the final MARS models, and there is no need to rank them. The MARS model identifies the most important factor successfully regardless of the input sample sizes and the time points, although the second and the last important factors are not successfully identified sometimes.

Table I only presents the estimation result of the two importance measures from a single simulation study. In order to further confirm our conclusions, we repeated the simulations for 1000 times. We summarize the frequency of that the MARS model correctly identifies significant factors in Table II and that the MARS model correctly determines the rankings (consistent with the sampling-based method) in Table III. From Table II, we can see that the MARS model approach can correctly identify the significant input factors with more than 95% of all simulation cases except one with input sample size 25 and time points 5. If the sample size of input factors is 50 or larger, significant factors are correctly identified for more than 99% of times. However, Table III shows that the significant factors can be correctly ranked by the MARS model approach only when the sample sizes of input factors and time points are large enough. Our simulation results show (Table III) that it requires the sample size of input factors to be 150 or larger and the sample size of time points to be 10 or larger in order to achieve about 80% or more correct rankings. We also notice that, for the case of the number of time points $q = 25$, the correct ranking rate is particularly lower than that of $q = 10, 50$ or 100 for the same sample size of input factors. This is caused by the flip ranking of x_1 and x_3 presumably due to some interaction effects of input factor sampling and the time factor.

In summary, our simulation results illustrate that the MARS model works effectively as a meta model for global sensitivity analysis. It can select significant factors correctly with correct rankings with a high probability and it can reduce the computational cost significantly compared to the sampling-based approach.

4 Application: dynamic models for influenza infection

Seasonal and pandemic Influenza A virus (IAV) continues to be a public health threat worldwide. Seasonal influenza claimed about 23, 600 deaths each year from 1976 to 2009 in the United States alone (CDC report, 2010). The pandemic influenza has resulted in multiple health disasters in the history. Vaccination is an effective way to prevent the epidemics of infectious diseases in general. However, as the virus strains such as influenza virus evolve unpredictably, it is difficult to develop a vaccine to be effective for all the strains emerged every year. Thus, it is important to deeply understand how our immune system responds to

different strains of viruses during infection. Recently mathematical models have become an important tool to study immune responses to influenza infection (Bocharov and Romanyukha, 1994; Beauchemin et al., 2005; Mohler et al., 2005; Baccam et al., 2006; Hancioglu et al., 2007; Lee et al., 2009; Miao et al., 2010). In combination with experimental data, we can understand pathogenesis of influenza infection quantitatively using computational biology approaches.

Miao et al. (2010) have proposed two sets of differential equation models to quantify the innate immune and adaptive immune responses to influenza infection. The innate immune response model for the time range between 0 to 5 days postinfection is

$$\frac{d}{dt}E_p = \rho_E E_p - \beta_\alpha E_p V, \quad (17)$$

$$\frac{d}{dt}E_p^* = \beta_\alpha E_p V - \delta_{E^*} E_p^* \quad (18)$$

$$\frac{d}{dt}V = \pi_\alpha E_p^* - c_V V, \quad (19)$$

and the adaptive immune response model for the time range between 5.5 and 14 days postinfection is given as

$$\frac{d}{dt}E_p = \rho_E E_p - \beta_\alpha E_p V \quad (20)$$

$$\frac{d}{dt}E_p^* = \beta_\alpha E_p V - k_E E_p^* T_E(t) - \delta_{E^*} E_p^* \quad (21)$$

$$\frac{d}{dt}V = \pi_\alpha E_p^* - c_V V - k_{VG} V A_G(t) - k_{VM} V A_M(t), \quad (22)$$

where $E_p(t)$ is the number of uninfected epithelial cells in the airway and lung at time t , $E_p^*(t)$ is the number of infected epithelial cells at time t , $V(t)$ is the number of influenza virus at time t , ρ_E is a net growth rate of epithelial cells, β_α is the rate that uninfected cells react with free virus to give rise to infected cells, δ_{E^*} is the death rate of infected epithelial cells, π_α is virus production rate, and c_V is the clearance rate of influenza virus. For additional parameters in the adaptive immune response model, k_E is depletion (killing) rate of infected epithelial cells by CD8⁺ effector cells ($T_E(t)$), k_{VG} and k_{VM} are depletion (neutralization) rates of virus by IgG and IgM ($A_G(t)$ and $A_M(t)$) antibody, respectively. More details of the model description and parameter interpretation can be found in Miao et al. (2010).

Although Miao et al. (2010) have estimated almost all the parameters in the above models based on experimental data in mice, it is necessary to perform formal sensitivity analysis in order to evaluate the reliability of the results and conclusions. The GSA can also help us to assess whether all model structures and assumptions are valid in the sense whether the output (measured) variable is significantly affected by all the terms in the models.

For the innate immune response model (17)–(19), we consider the initial value and parameters, $E(0)$, ρ_E , β_α , δ_{E^*} , and c_V as input factors and viral titer $V(t)$ as the output

variable which is measured over time after infection. The point estimates of these input factors in Miao et al. (2010) are $\widehat{E}(0) = 5.84 \cdot 10^5$, $\widehat{\rho_E} = 6.19 \cdot 10^{-8}$, $\widehat{\beta_\alpha} = 2.42 \cdot 10^{-6}$, $\widehat{\delta_{E^*}} = 5.98 \cdot 10^{-1}$, and $\widehat{c_v} = 4.23$. Note that the parameter π_α is not considered as an input factor and is set as 100 since Miao et al. (2010) suggested that π_α is not identifiable only based on the measurements of the output variable $V(t)$. To perform GSA for these input factors on the output variable, we assume that these input factors on a log-scale follow independent uniform distributions centered at their point estimates with ranges of 3-fold and 5-fold changes. We use 300 and 1000 input factor samples at 50 equally spaced time points. The sample size is increased until the fitted MARS model represents the ODE system well based on the predictability measure D^2 . First, we apply the MARS model approach as introduced in Section 2 to the simulated data from the ODE model (17)–(19). We use the five input factors, $E(0)$, ρ_E , β_α , δ_{E^*} , and c_v , the maximum number of terms 100 with upto three-way interactions, and the model complexity penalty $d = 4$ (the largest within the range recommended in Section 2.1). The model evaluation criterion R^2 is obtained as $R^2 = 0.9523$ and $R^2 = 0.9389$ for 3-fold and 5-fold input factor ranges, respectively, which indicates good fits of the MARS models to the simulated data sets. Two separate sets of 100 input factor samples at 50 equally-spaced time points are used to obtain $D^2 = 0.9448$ and 0.9251 for 3-fold and 5-fold ranges, respectively. We can see that D^2 values are quite close to the corresponding R^2 values, which suggests that the MARS model approximates the original ODE model very well. Consequently, we apply the proposed methods in Section 2 to calculate the sensitivity indices $\widehat{\eta}^2$ and $\widehat{T}_j$.

We report the GSA results for the innate immune response model (17)–(19) in Table IV. The values in the column $\widehat{\eta}^2$ are cumulative values corresponding to all the input factors from the top rank. The input factor ρ_E is not included in the final MARS models. This indicates that the term $\rho_E E_p$ in Eq(17) has little effect on the output variable $V(t)$ when the parameter ρ_E has 3-fold and 5-fold changes from its point estimates in Miao et al. (2010), i.e., $\widehat{\rho_E} = 6.19 \cdot 10^{-8}$. From Table IV, we can also see that both importance measures $\widehat{\eta}^2$ and $\widehat{T}_j$ rank the other 4 input factors consistently, but the rankings for the 3-fold and 5-fold range cases are a little different, i.e., the ranks of β_α and c_v are flipped, which indicate that the importance of these two factors is different for different ranges of their values. This also confirms that the ranges of the input factors may have an effect on the sensitivity analysis results, in particular on the ranking similarly important factors.

In summary, among the input factors evaluated for the innate immune response ODE model, the most important factor for the output variable, the viral titer $V(t)$, is the level of target uninfected epithelial cells at baseline, $E(0)$. This is consistent with our understanding that the target cell availability is a limiting factor for influenza virus replication (Miao et al., 2010). The infection rate β_α and the virus clearance rate c_v are ranked as the next important factors for virus replication, which is also quite reasonable since these two parameters are directly related to virus production and virus clearance in the ODE system. The death rate of infected epithelial cells δ_{E^*} is another significant factor for the output variable $V(t)$ since the death rate or life-span of infected cells may affect the number of virions produced per infected cell. Interestingly these ranking results are quite consistent with the AIC-based model selection results in Miao et al. (2010).

For the adaptive immune response model (20)–(22), $T_E(t)$ is the concentration of effector CD8+ T cells (CTL) in lung, and $A_G(t)$ and $A_M(t)$ are concentrations of IgG and IgM at time t , respectively, which are measured from the mice experiments. The cubic spline smoothing is applied to the experimental data to obtain the estimated values for $A_G(t)$, $A_M(t)$, and $T_E(t)$ at any time t and these estimates are substituted in the ODE model. The initial conditions at day 5 for the adaptive immune response model are obtained from the estimated results of the

innate immune response model at day 5, i.e., $E(5) = 0$, $E^*(5) = 49,770$ and $V(5) = 1,370,298$ as suggested by Miao et al. (2010). Here we consider seven parameters ρ_E , β_a , k_E , δ_{E^*} , c_v , k_{VG} , and k_{VM} as input factors and the viral titer $V(t)$ as the output variable in our sensitivity analysis.

Similarly we generate input factor samples from uniform distributions on a log-scale centered at their point estimates from Miao et al. (2010), i.e., $\widehat{\rho}_E = 3.39 \cdot 10^{-1}$, $\widehat{\beta}_a = 7.84 \cdot 10^{-1}$, $\widehat{k}_E = 6.40 \cdot 10^{-5}$, $\widehat{\delta}_{E^*} = 1.50 \cdot 10^{-1}$, $\widehat{c}_v = 3.98 \cdot 10^{-4}$, $\widehat{k}_{VG} = 3.13 \cdot 10^{-5}$, and $\widehat{k}_{VM} = 4.40$. Here we first consider 3-fold and 5-fold range changes of these input factors for GSA. However, we find that the output variable $V(t)$ gets negative values for some of the input factor samples, which are not reasonable for the viral titer and need to be excluded from our analysis. In order to evaluate whether the truncation of input factor samples has an effect on our conclusions, we also consider a 1.5-fold change of input factors for the sensitivity analysis in which no negative value of the viral titer $V(t)$ is found. We collect 300 input factor samples for 50 equally-spaced time points for all three cases. We apply the MARS model approach to the generated data and obtain that $R^2 = 0.999, 0.995$, and 0.993 , respectively, and D^2 are $0.999, 0.995$, and 0.995 for 1.5-fold, 3-fold and 5-fold changes, respectively, which indicate good fits and good approximations of the MARS models to the original ODE model. The sensitivity indices $\widehat{\eta}^2$ and $\widehat{T}_j$ are calculated and reported in Table V.

From Table V, we can see that in all three cases, only three input factors (parameters), k_E , k_{VM} and δ_{E^*} , are selected in the final MARS model. The rankings of the three significant factors for 1.5-fold and 3-fold change cases are the same, i.e., k_E is the most important and followed by k_{VM} , and δ_{E^*} is ranked in the last, and for the 5-fold change case the ranks of k_{VM} and δ_{E^*} flip over. The effect of k_E on the output variable, the viral titer $V(t)$, dominates based on the results on Table V. This conclusion is consistent with the observation that the CTL effect, $k_E E_p^* T_E(t)$, significantly contributes to the killing of infected epithelial cells, thus to the control of viral titer during adaptive immune response. The antibody effect due to IgM is another important adaptive immune response factor to control virus replication. The death rate of infected epithelial cells δ_{E^*} may also affect the viral titer, but its contribution is small in the 1.5-fold and 3-fold range changes. Similar conclusions based on experimental data and AICc model selection criterion are obtained in Miao et al. (2010). Also we can see that the truncation of input factor samples has little effect on the sensitivity analysis results.

5 Conclusions

In this paper, we evaluate applications of the multivariate adaptive spline regression (MARS) as a meta model to the global sensitivity analysis of ordinary differential equation models. Based on the concept of the variance of conditional expectation (VCE), a sequential and a single factor importance measures, η^2 and T_j , over a time interval of interest are used to rank the importance of input factors on the output variables. Our simulation studies show that the MARS model is capable of selecting significant factors using the variable selection procedure. If a factor is not included in the final MARS model, it can be considered as insignificant for the output variables over the entire time range. As a result, the MARS model screening reduces the number of factors to be included for further sensitivity analysis. Overall, the MARS model approach performs reasonably good and is computationally efficient for global sensitivity analyses (GSA) for ODE models over a range of time period.

The difficulty conducting global sensitivity analyses on ODE systems is often due to the computational demands to solve the ODEs. The proposed MARS approach only requires to numerically solve ODEs for one set of samples in order to rank the factors while the sampling-based approach requires to solve ODEs for $p(p+1)/2 - 1$ different sets of samples

for the sequential importance measure and p sets of samples for the single importance measure, where p is the number of input factors. Thus, our proposed MARS method can reduce the computational cost significantly when the computational cost for numerically solving ODEs is high. Meanwhile the MARS model and the proposed analytic approach for calculating the importance measures are easy to implement computationally, and the sampling-based approach often encounters computational problems for some samples. In general, provided that the ODE system is considerably expensive to solve, the computation time is approximately proportional to ODE solving time. In addition, the sampling-based approach such as the Latin Hypercube Sampling (LHS) method also requires to solve ODEs to evaluate the variance. Thus, the proposed MARS approach can reduce the total computation time by a factor of $(p + 1)$ for the single importance measure and by a factor of $p(p + 1)/2$ for the sequential importance measure compared to the LHS method. However, if numerically solving the ODEs is not computationally expensive, the total computational cost of the two approaches (MARS and LHS methods) would not be very different, i.e., the MARS method may not have the advantage of computational cost. But it is still more convenient to use the MARS approach in this case since we only need to sample one set of input factors instead of many sets with varying conditions for the LHS method.

From our application examples in Section 4, we can see that the GSA can be used to identify important factors and terms in ODE models. Interestingly the results are consistent with those from statistical model selection methods such as AICc based on experimental data. Note that the proposed GSA can be done without experimental data and we only need to know rough ranges of parameters (or initial values). This is quite useful in helping with design of new experiments and development of dynamic models for the biological processes.

The GSA procedure introduced in this article begins with sampling input factors (parameters) from a specific distribution, and then one can obtain the corresponding outputs of a dynamic system at many time points. Both input factors and the time factor are considered as covariates and the output as a response variable in MARS model (2.1). We use the earth package (Stephen Milborrow derived from mda:mars by Trevor Hastie and Rob Tibshirani, 2009) in the statistical software R (R Development Core Team, 2008), which resulted in the fitted MARS model and will be used to evaluate the two important measures (Eq(13) and Eq(14)) by the analytic *VCE* method introduced in Section 2.2. The algorithm of the analytic *VCE* is written in R and is posted on the website, https://cbim.urmc.rochester.edu/software/importance_measures/. We suggest to evaluate the *VCE* analytically using the MARS model structure of univariate tensor-product functions, instead of Monte-Carlo methods, which is more computationally efficient. The available input factor distributions are currently limited to variants of uniform distributions, but more distribution options will be available in the future and the updates will be posted on the website.

Like other global sensitivity analysis approaches, the ranks using the MARS model is subject to the range and distribution of input factors. Detailed exploration on this problem is beyond the scope of this paper. In addition, in this paper we consider the assumption that the input factors are independent to each other. However, if the covariance structure of input factors can be estimated using available data, the sensitivity analysis without the assumption of independence can be carried out. The bootstrap or Bayesian methods are available to obtain input samples with the estimated covariance structures, but these approaches will be limited due to high computational cost.

Acknowledgments

The work is supported by National Institute of Allergy and Infectious Diseases of the National Institutes of Health of USA under award number N01-AI50020 and RO1-AI087135. The authors thank Dr. Hongyu Miao and other investigators from the Center for Biodefense Immune Modeling (CBIM) for their help. We also appreciate the referee's constructive and insightful comments and suggestions.

References

1. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. Kinetics of influenza A virus infection in humans. *J. Virol.* 2006; 80:7590–7599. [PubMed: 16840338]
2. Beauchemin C, Samuel J, Tuszynski J. A simple cellular automaton model for influenza A viral infections. *J. Theor. Biol.* 2005; 232:223–234. [PubMed: 15530492]
3. Blower S, Dowlatabadi H. Sensitivity and Uncertainty Analysis of complex models of disease transmission: an HIV model, as an example. *Int. stat.Rev.* 1994; 64:229–243.
4. Bocharov GA, Romanyukha AA. Mathematical model of antiviral immune response III. Influenza A virus infection. *J. Theor. Biol.* 1994; 167:323–360. [PubMed: 7516024]
5. CDC report. Estimates of deaths associated with seasonal influenza — united states. 2010:1976–2007.
6. Chen W, Jin R, Sudjianto A. Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. *J. Mech. Design.* 2005; 127:875–886.
7. Friedman JH. Multivariate adaptive regression splines. *Ann. Stat.* 1991; 19:1–67.
8. Hancioglu B, Swigon D, Clermont G. A dynamical model of human immune response to influenza A virus infection. *J. Theor. Biol.* 2007; 246:70–86. [PubMed: 17266989]
9. Helton J, Johnson J, Sallaberry C, Storlie C. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Syst. Safe.* 2006; 91:1175–1209.
10. Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Safe.* 2003; 81:23–69.
11. Hoare A, Regan D, Wilson D. Sampling and sensitivity analyses tools (SaSAT) for computational modelling. *Theor. Biol. Med. Model.* 2008; 5:4. [PubMed: 18304361]
12. Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Safe.* 1996; 52:1–17.
13. Lee HY, Topham DJ, Park SY, Hollenbaugh J, Treanor J, Mosmann TR, Jin X, Ward BM, Miao H, Holden-Wiltse J, Perelson AS, Zand M, Wu H. Simulation and prediction of the adaptive immune response to influenza A virus infection. *J. Virol.* 2009; 83:7151–7165. [PubMed: 19439465]
14. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global uncertainty and sensitivity analysis in system biology. *J. Theor. Biol.* 2008; 254:178–196. [PubMed: 18572196]
15. McKay, M. Technical Report. Los Alamos National Lab.; NM (United States): 1995. Evaluating prediction uncertainty.
16. McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics.* 1979; 21:239–245.
17. Miao H, Hollenbaugh J, Zand M, Holden-Wiltse J, Mosmann T, Perelson A, Wu H, Topham D. Quantifying the early immune response and adaptive immune response kinetics in mice infected by influenza A virus. *J. Virol.* 2010; 84:6687–6698. [PubMed: 20410284]
18. Mohler L, Flockerzi D, Sann H, Reichl U. Mathematical model of influenza A virus production in large-scale microcarrier culture. *Biotechnol. Bioeng.* 2005; 90:46–58. [PubMed: 15736163]
19. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. ISBN 3-900051-07-0
20. Saltelli, A.; Chan, K.; Scott, E. Wiley Series in Probability and Statistics. 2000. Sensitivity Analysis.
21. Saltelli A, Tarantola S, Chan KPS. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics.* 1999; 41:39–56.

22. Sobol' I. Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling and Computational Experiment*. 1993; 1:407–414.
23. earth: Multivariate Adaptive Regression Spline Models. R package version 2.3–5. Stephen Milborrow derived from mda:mars by Trevor Hastie and Rob Tibshirani, 2009.
24. Storlie CB, Helton JC. Multiple predictor smoothing methods for sensitivity analysis: Description of techniques. *Reliab. Eng. Syst. Safe*. 2008a; 93:28–54.
25. Storlie CB, Helton JC. Multiple predictor smoothing methods for sensitivity analysis: Example results. *Reliab. Eng. Syst. Safe*. 2008b; 93:55–77.
26. Storlie CB, Swiler LP, Helton JC, Sallaberry CJ. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliab. Eng. Syst. Safe*. 2009; 94:1735–1763.
27. Zhang H. Multivariate adaptive splines for analysis of longitudinal data. *J. Comput. Graph. Stat*. 1997; 6:74–91.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Simulation study for the MARS model approach: The importance measures $\hat{\eta}^2$ and $\hat{T}_j$ from a single simulation study (the shaded cases are those with consistent ranking as the sampling-based method)

Sample size n	$\hat{\eta}^2$					$\hat{T}_j$				
	Number of time points (q)					Factor	Number of time points (q)			
	5	10	25	50	100		5	10	25	100
25	0.9889	0.9896	0.9896	0.9887	0.9897	x_1	0.0050	0.0054	0.0060	0.0056
	0.9950	0.9950	0.9955	0.9946	0.9954	x_2	0.0394	0.0381	0.0390	0.0424
	1.0000	1.0000	1.0000	1.0000	1.0000	x_3	0.0061	0.0050	0.0045	0.0046
50	0.9930	0.9894	0.9899	0.9900	0.9894	x_1	0.0019	0.0060	0.0052	0.0052
	0.9981	0.9954	0.9951	0.9952	0.9948	x_2	0.0428	0.0429	0.0455	0.0447
	1.0000	1.0000	1.0000	1.0000	1.0000	x_3	0.0050	0.0046	0.0049	0.0054
150	0.9933	0.9908	0.9929	0.9902	0.9899	x_1	0.0021	0.0049	0.0027	0.0057
	0.9979	0.9957	0.9973	0.9957	0.9956	x_2	0.0413	0.0421	0.0382	0.0464
	1.0000	1.0000	1.0000	1.0000	1.0000	x_3	0.0046	0.0043	0.0044	0.0044
300	0.9931	0.9929	0.9892	0.9892	0.9890	x_1	0.0023	0.0026	0.0063	0.0065
	0.9977	0.9974	0.9955	0.9955	0.9954	x_2	0.0374	0.0354	0.0431	0.0447
	1.0000	1.0000	1.0000	1.0000	1.0000	x_3	0.0046	0.0044	0.0045	0.0046
600	0.9934	0.9907	0.9897	0.9899	0.9898	x_1	0.0022	0.0049	0.0060	0.0056
	0.9978	0.9959	0.9957	0.9955	0.9955	x_2	0.0397	0.0426	0.0436	0.0440
	1.0000	1.0000	1.0000	1.0000	1.0000	x_3	0.0045	0.0043	0.0044	0.0045

\$watermark-text

\$watermark-text

\$watermark-text

Table II

Simulation study for the MARS model approach: The rate that the MARS model correctly identifies the true factors (among 1000 simulation studies)

Sample size (n)	Number of time points (q)				
	5	10	25	50	100
25	0.884	0.969	0.969	0.982	0.970
50	0.996	0.998	0.999	1.000	0.999
150	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000	1.000

\$watermark-text

\$watermark-text

\$watermark-text

Table III

Simulation study for the MARS model approach: The rate that the ranking of input factors from the MARS model approach is consistent with that of the sampling-based method (among 1000 simulation studies)

Sample size (n)	Number of time points (q)				
	5	10	25	50	100
25	0.090	0.538	0.527	0.600	0.624
50	0.031	0.723	0.677	0.801	0.789
150	0.026	0.954	0.780	0.922	0.918
300	0.124	0.996	0.791	0.943	0.964
600	0.064	1.000	0.810	0.989	0.990

\$watermark-text

\$watermark-text

\$watermark-text

Table IV

Application example for innate immune response to influenza infection: Importance measures of input factors within 3-fold and 5-fold changes from their point estimates. Columns for $\hat{\eta}^2$ are cumulative values from the top rank

3-fold				5-fold			
Rank	Factor	$\hat{\eta}^2$	$\hat{T}_j$	Factor	$\hat{\eta}^2$	$\hat{T}_j$	
1	E_0	0.6688	0.1923	E_0	0.6971	0.1811	
2	β_α	0.8119	0.1342	c_ν	0.8012	0.1126	
3	c_ν	0.9911	0.0996	β_α	0.9003	0.1012	
4	δ_{E^0}	1	0.0885	δ_{E^0}	1	0.0998	

\$watermark-text

\$watermark-text

\$watermark-text

Table V

Application example for adaptive immune response to influenza infection: Importance measures of input factors within 1.5-fold, 3-fold and 5-fold changes from their point estimates. Columns for $\widehat{\eta}^2$ are cumulative values from the top rank

1.5-fold			3-fold			5-fold		
Rank	Factor	$\widehat{\eta}^2$	$\widehat{T}_j$	Factor	$\widehat{\eta}^2$	$\widehat{T}_2$	Factor	$\widehat{\eta}^2$
1	k_E	0.9959	1214	k_E	0.9788	0.3968	k_E	0.9432
2	k_{VM}	0.9984	0.0025	k_{VM}	0.9903	0.0114	δ_{E^*}	0.9731
3	δ_{E^*}	1	0.0016	δ_{E^*}	1	0.0097	k_{VM}	1