Chemometrics and Intelligent Laboratory Systems 120 (2013) 25–41

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Hierarchical multivariate regression-based sensitivity analysis reveals complex parameter interaction patterns in dynamic models

Kristin Tøndel [a,*], Jon Olav Vik [a], Harald Martens [a], Ulf G. Indahl [a], Nicolas Smith [b], Stig W. Omholt [c]

[a] Centre for Integrative Genetics (CIGENE), Dept. of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway
[b] Department of Biomedical Engineering, Kings College London, The Rayne Institute, Lambeth Wing, St Thomas' Hospital, London SE1 7EH, United Kingdom
[c] CIGENE, Dept. of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P. O. Box 5003, N-1432 Ås, Norway

ARTICLE INFO

Article history:
Received 28 March 2012
Received in revised form 18 September 2012
Accepted 26 October 2012
Available online 7 November 2012

Keywords:
Sensitivity analysis
Parameter interactions
Hierarchical multivariate regression analysis
Metamodelling
Nonlinear dynamic models
Cluster analysis

ABSTRACT

Dynamic models of biological systems often possess complex and multivariate mappings between input parameters and output state variables, posing challenges for comprehensive sensitivity analysis across the biologically relevant parameter space. In particular, more efficient and robust ways to obtain a solid understanding of how the sensitivity to each parameter depends on the values of the other parameters are sorely needed.

We report a new methodology for global sensitivity analysis based on Hierarchical Cluster-based Partial Least Squares Regression (HC-PLSR)-based approximations (metamodelling) of the input–output mappings of dynamic models, which we expect to be generic, efficient and robust, even for systems with highly nonlinear input–output relationships. The two-step HC-PLSR metamodelling automatically separates the observations (here corresponding to different combinations of input parameter values) into groups based on the dynamic model behaviour, then analyses each group separately with Partial Least Squares Regression (PLSR). This produces one global regression model comprising all observations, as well as regional regression models within each group, where the regression coefficients can be used as sensitivity measures. Thereby a more accurate description of complex interactions between inputs to the dynamic model can be revealed through analysis of how a certain level of one input parameter affects the model sensitivity to other inputs. We illustrate the usefulness of the HC-PLSR approach on a dynamic model of a mouse heart muscle cell, and demonstrate how it reveals interaction patterns of probable biological significance not easily identifiable by a global regression-based sensitivity analysis alone.

Applied for sensitivity analysis of a complex, high-dimensional dynamic model of the mouse heart muscle cell, several interactions between input parameters were identified by the two-step HC-PLSR analysis that could not be detected in the single-step global analysis. Hence, our approach has the potential to reveal new biological insight through the identification of complex parameter interaction patterns. The HC-PLSR metamodel complexity can be adjusted according to the nonlinear complexity of the input–output mapping of the analysed dynamic model through adjustment of the number of regional regression models included. This facilitates sensitivity analysis of dynamic models of varying complexities.

1. Introduction

Dynamic models describing complex biological systems, processes or traits are normally rich in input parameters, i.e. quantities that are constant over the time-scale of the particular dynamic model being studied but can be varied between simulations to create variation in model output. In cases where a dynamic model is sensitive to changes in a set of parameters, and the effects of change in one parameter are not dependent on the values of the other parameters, the causal structure of the system is simple, although possibly nonlinear, and the associated sensitivity analysis is relatively trivial across the whole parameter range giving rise to biologically meaningful results. However, for the majority of nonlinear complex dynamic models, the effects of changes in a parameter are often highly dependent on the values of other parameters (the parameters interact), precluding a parameter-by-parameter approach. This situation is likely to become even more pronounced with the emergence of ever more high-resolution, multi-scale dynamic models characterised by high-dimensional input parameter- and output state variable spaces due to improved genomics and phenomics technologies [1].

Means to systematically elucidate the sensitivity features of such dynamic models, including ways to reveal complex interaction patterns between input parameters manifested in high-dimensional

* Corresponding author. Tel.: +47 64 96 52 83; fax: +47 64 96 51 01.
E-mail address: kristin.tondel@umb.no (K. Tøndel).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2012.10.006

phenotypes, are instrumental for efficient model construction, validation and application. However, many traditional methods for sensitivity analysis are primarily suitable for systems with relatively few input- and output variables and for analysing the effects on only one output at a time [2–5]. A generic sensitivity analysis methodology must be able to handle even the most complex modelling situations, such as highly nonlinear high-dimensional systems [6–8]. Ideally, it should reveal the sensitivity of all dynamic model outputs to any parameter as a function of all other parameters, within the entire operative domain of the analysed model.

In statistical sensitivity analysis, a major branch of the sensitivity analysis field, a selection of data points is derived by experimental design or (semi-) random sampling, and the input–output relations are analysed by statistical methods such as e.g. regression methodology [3] (see Section 5 for more details). In such regression-based sensitivity analysis, the regression coefficients provide direct measures of the impact of the individual inputs on the output (model sensitivity). A major concern is that most regression-based sensitivity analyses published are based on relatively simple linear regression models fitted by ordinary least squares (OLS) regression. Since the input–output relations may be highly nonlinear, linear regression analysis may lead to suboptimal descriptions of dynamic model behaviour, and subsequent difficulties with revealing important interaction patterns. Simple curvature and interaction effects may be modelled successfully by polynomial regression with cross-terms, but when functionally distinct input parameter space regions with clearly different input–output relations and complex interaction patterns between inputs are present, more flexible multivariate analysis methods are needed for a detailed analysis of dynamic model behaviour. Furthermore, most regression-based sensitivity analysis methods are primarily focused on analysing the effects on a single output variable at a time. In many situations it might be advantageous to explore the effects of simultaneous input variation on the whole set of output variables to reveal intricate covariance patterns within both the input- and the output space, in addition to relationships between inputs and outputs. This motivates the development of multivariate *metamodels* [9]; statistical approximations to the input–output mappings of dynamic models that facilitate accurate analysis of their sensitivity features even if these vary substantially across parameter space. In the following, the term “model” refers to the analysed dynamic simulation model if not otherwise specified. Metamodels and regression models are specified as such when discussed.

We recently showed that metamodelling based on Hierarchical Cluster-based Partial Least Squares Regression (HC-PLSR) [10], which involves a combination of global and regional regression analysis, was more accurate than ordinary Partial Least Squares Regression (PLSR) [11] (see also [12,13]) and OLS regression for a range of nonlinear dynamic gene regulatory and physiological models. See Supplementary electronic material: Appendix A for a description of these data analysis methods. In general, PLSR is more effective than OLS for handling multiple output variables simultaneously, since it utilises inter-correlations between the response variables for regression model stabilisation, and is therefore used in both the global and the regional regression steps of HC-PLSR. Furthermore, in contrast to OLS, the PLSR does not require linear independency of the input parameters. In multivariate metamodelling of complex dynamic models this is an advantage, since in cases where the number of input variables is large, highly reduced experimental designs or random sampling must often be used to set up the parameter value combinations for the computational experiment, leading to potentially linearly dependent inputs. For some dynamic models the simulations may also fail to converge under certain conditions, leading to non-orthogonal inputs to the metamodelling. The usefulness of global PLSR for sensitivity analysis was recently demonstrated by Sobie [14] and Martens *et al.* [15]. As a nonlinear extension of PLSR, the HC-PLSR separates the input- or output space into local regions based on clustering in an initial global metamodel. Thereafter a regional metamodel is fitted for each cluster. This allows a simpler description of highly nonlinear effects of input parameters, e.g. causing output variations that may apply only in parts of the input space. The HC-PLSR provides a semi-parametric representation of complex interaction patterns that allows e.g. non-monotone parameter-to-phenotype maps to be modelled more accurately.

Here we introduce a flexible and generic methodology for global sensitivity analysis of complex dynamic models. It is based on the two-step HC-PLSR, and can reveal complex, regional interaction patterns between inputs in a multi-dimensional output setting. Both the global and regional regression modelling steps in HC-PLSR provide scores, loadings, regression coefficients and residual matrices that reflect the sensitivity of the dynamic model to variations in the different inputs. Whereas the initial, global regression model provides an overall summary, the subsequent regional regression models can detail input–output relations that pertain only to parts of the dataset. Hence, modifications of the effects of certain parameters on the dynamic model output dependent on the values of other parameters (reflecting complex parameter interactions) can be identified. Furthermore, the regional sensitivity analysis provides the opportunity to test whether a parameter showing little impact on the output in a global sensitivity analysis still has some impact in local regions of the biologically relevant parameter space.

We illustrate our approach using a detailed model of a mouse heart muscle cell (a ventricular myocyte) [16], primarily built to account for the action potential (the time-course of transmembrane voltage, i.e. the cell's electrical signal) and calcium transient (the time-course of the calcium concentration in the cell fluid, which is linked to muscle contraction) of the cell. These are modelled in terms of a large number of constituent ion currents and voltage- and calcium-sensitive ion channels in the cell, represented by a set of coupled ordinary differential equations (ODEs). Our hypothesis was that highly nonlinear dynamic models, like the model analysed here, will exhibit complex interaction effects that are not identifiable using a global regression-based sensitivity analysis alone. The present analysis includes a wide variety of phenotypic measures (outputs from ODE model simulations) related to the action potential (AP), the calcium transient (CT) as well as the dynamics of a range of other state variables (including ion concentrations in the cell fluid and various cellular compartments, and the state distributions of ion channels, whose transition rates between open, closed, and inactivated conformations may depend on transmembrane voltage and calcium concentration). Several auxiliary output variables, such as ion currents, whose magnitude is a function of system state, are also included. Complex interaction patterns between input parameters are revealed through analysis of the sensitivity of output variables to changes in the individual model inputs, conditional on the levels of the other input parameters.

Illustrating how our methodology can lead to new biological insights, we provide a detailed biological interpretation of how the model parameters interact with respect to four of the outputs; the AP time-to-peak, the AP duration to 25% repolarisation, the CT time-to-peak and the CT decay rate. We chose to focus on these four outputs since we know that the AP and the CT are key cell-level phenotypes of consequence for tissue and organ function. We compare our results to those obtained by a global PLSR-based sensitivity analysis, and show that additional parameter interactions can be identified by supplementing the global sensitivity analysis with a regional analysis.

A dynamic model from computational biology is used here to illustrate the two-step metamodelling methodology for sensitivity analysis. However, we believe that the method is generic, and that the HC-PLSR is a promising approach as part of a semi-automatic methodological framework. The reason for this is its possibility for automatic adjustment of the number of regional regression models according to the nonlinear complexity of the response surface of the analysed dynamic model.