A Framework for Simplification of Quantitative Systems Pharmacology Models in Clinical Pharmacology

Abdallah Derbalah[1], Hesham Al-Sallami[1], Chihiro Hasegawa[1], Abhishek Gulati[2], and Stephen Duffull[1]

[1]University of Otago
[2]Astellas Pharma Global Development Inc

May 6, 2020

Abstract

Quantitative systems pharmacology (QSP) is a relatively new discipline within modelling and simulation that has gained wide attention over the past few years. The application of QSP models spans drug-target identification and validation, through all drug development phases as well as clinical applications. Due to their detailed mechanistic nature, QSP models are capable of extrapolating knowledge to predict outcomes in scenarios that have not been tested experimentally making them an important resource in experimental and clinical pharmacology. However, these models are complicated to work with due to their size and inherent complexity. This makes many applications of QSP models for simulation, parameter estimation and trial design computationally intractable. A number of techniques have been developed to simplify QSP models into smaller models that are more amenable to further analyses while retaining their accurate predictive capabilities. Different simplification techniques have different strengths and weaknesses and hence different utilities. Understanding the utilities of different methods is essential for selection of the best method for a particular situation. In this paper, we have created an overall framework for model simplification techniques that allows a natural categorisation of methods based on their utility. We provide a brief description of the concept underpinning the different methods and example applications. A summary of the utilities of methods is intended provide a guide to modellers in their model endeavours to simplify these complicated models.

Background

Mathematical modelling of clinical pharmacology processes is fundamental to understanding the time course of the determinants of patient response and inform key decisions in the discovery, development, and clinical use of drugs. In drug development, modelling and simulation (M&S) applications span a wide spectrum from target identification and validation through analysis of preclinical data, determining the best first-in-human dose, and analysis of phase I, II, and III data to optimise doses for safety and efficacy [1]. M&S also plays an important role in clinical practice, serving purposes such as dose individualisation based on covariates or measured biomarkers, optimising doses for special populations or off-label use, and optimising clinical trial design [2]. A wide variety of models and approaches are available, ranging from empirical and conceptual models that include no specific representation of biologic mechanisms to fully mechanistic models that describe biochemical, (patho)physiological, and pharmacological mechanisms in detail [3] [4].

Quantitative systems pharmacology (QSP) is a discipline within M&S that creates mathematical models that focus on the quantitative interplay between biological and pharmacological mechanisms (see [5-7] for detailed descriptions). QSP has roots in engineering, systems biology, and pharmacology. A key feature of QSP models is the integration of information and knowledge from various disciplines with experimental data into a single quantitative model that describes drug response often at multiple spatial scales (e.g. spanning from cellular to whole body effects), and temporal scales (e.g. from milliseconds to decades). Due to their

mechanistic structure, QSP models are capable of extrapolating knowledge to predict outcomes in scenarios that have not been tested experimentally [4]. The ability to use QSP models to extrapolate findings has many important applications for the development and clinical use of drugs. For example, extrapolating efficacy from preclinical *in-vitro* or *in-vivo* studies to predict efficacy in humans [8] and extrapolating from adults to paediatric patients [9]. Integration of data collected from preclinical and clinical studies [10], and from various drugs that act on the same or similar targets [11], into a single mathematical framework can expedite drug development by cumulating knowledge gained throughout development of both successful and failed drug candidates [12].

Despite the benefits offered by the QSP approach, there are several challenges in their application. The level of detail in which pharmacological and biological mechanisms are represented in QSP models, usually results in models with a large number of states (in pharmacokinetics states represent a concentration in a particular compartment) and parameters. We show a (relatively simple) hypothetical QSP model schematic in figure 1. This schematic is illustrated with 16 circles (called nodes), that may represent different pharmacological species or physiological measures. Each node is connected to another node using arrows or dashed lines with a bar at the end (collectively called edges). We use an arrow to illustrate movement based on mass (or molar) balance and a dashed line with a bar to illustrate mass (or molar) action through positive or negative feedback. The size (dimension) of the model is a function of the number of nodes (circles) and edges (lines). Feedback mechanisms relate to homeostasis through damping (e.g. blood pressure control [13]) or amplification (e.g. formation of a blood clot [14]). These mechanisms make it difficult to understand which component(s) of a system are important (or unimportant) with respect to observing a phenomenon of interest or interpreting a response of the system given a change (e.g. administration of a dose). Both damping and amplification behaviours result in nonlinearity in QSP models which makes them mathematically difficult to work with and solve computationally [15]. Additionally, the large number of parameters and the often-limited measurable responses makes traditional modelling techniques infeasible due to structural identifiability [16]. Thus, even if an infinite amount of observational data are available (e.g. infinite INR values), it may not be possible to reliably estimate all parameters in a QSP model nor determine which parameters can be estimated [17].

An alternative approach is to use simpler models, such as compartmental models, that are built based on available data. These models are readily amenable for use in simulation and estimation analyses. However, since they are built based on data they do not necessarily provide mechanistic insights into how the data arose and may not be appropriate for extrapolation (see for example [18,19]). Therefore, a hybrid approach to building simpler but mechanistically accurate models would be of significant benefit. Model-order reduction techniques provide a set of methods to harness the mechanistic characteristics of large QSP models but render them into simpler models that are amenable for use in simulation, estimation and design. Simpler models using these methods have been used to extrapolate beyond the data used to build them in predicting changes in bone mineral density [18,19], be used to predict response to warfarin therapy [20], individualise treatment for children with acute lymphoblastic leukaemia [21], predict fibrinogen kinetics in snake envenomed patients [22], and design of clinical trials [23].

The aim of this article is to provide a framework to illustrate the utility of various methods that can be used for simplifying QSP models. These methods are globally referred to here as model-order reduction approaches. We will focus on the concepts and potential applications of various model-order reduction methods in the context of QSP with relevant references provided for readers who are interested to delve into more technical details. We denote the original QSP model as full-order and a simpler model (produced by model-order reduction) as a reduced-order model. The order of a model represents its size, i.e. number of compartments and parameters, which is also termed the degrees of freedom, i.e. the number of independent parameters in the model. A reduced-order model will, therefore, by definition, have fewer parameters than the (original) full-order model. We provide an explanation of terms in Appendix 1.

A framework for considering model-order reduction methods

A wide range of model-order reduction methods have been described [18,21,22,24-27] each of which has its own strengths and weaknesses. An ideal model-order reduction method should have the following characteristics:

1. System agnostic, i.e. universally applicable to different biological systems and model use,
2. Automatable, i.e. does not require user input,
3. Mechanistically relevant, i.e. the simplified model contains the necessary features that enable its use for a wide variety of settings
4. Accurate, i.e. able to replicate important data predictions,
5. Computational ease, i.e. identifiable and be of little computational burden,
6. Preserve extrapolation ability of the full-order model,
7. Requires no experimental data to derive the reduced-order model, and
8. Able to be used for parameter estimation.

However, an ideal method does not exist, therefore it is necessary for the investigator to select the method. Understanding how the methods relate to each other and what they contribute is therefore of particular importance.

There are several ways that methods of model-order reduction can be categorised. Snowden et al. [28] in their review adopted four categories to summarise these techniques based on their internal mathematical methods, namely, time-scale exploitation, optimisation and sensitivity analysis, lumping, and singular value decomposition methods. We contend that, the utility of model-order reduction techniques lies not in their internal mathematical properties but in their use-effectiveness of simplifying various problems. In this review, we have adopted a framework that allows a natural categorisation of methods based on their utility. The framework is outlined in Figure 2.

In the proposed framework, we categorise model reduction methods into parametric and nonparametric methods. Parametric methods aim to simplify the structure of the model, which is typically achieved by either reducing the number of nodes (species of interest and compartments) or edges (reactions, fluxes and interactions). Nonparametric methods, on the other hand, simplify the input-output relationship through construction of an empirical (surrogate) model that has a simpler structure but can emulate the behaviour of the full-order model. There are, of course, hybrid methods that use both approaches and approximate some components of the system with a black box while using mechanistic simplification methods on the structure of other components [29].

All methods provide a simplified relationship between the model inputs (i.e. drug, dose, dose-time) with the outputs (i.e. the response variable of interest) that can then be manipulated easily to describe and predict new data.

Description of methods

1 Parametric methods

These methods are premised on the principle that the changes in the response variable(s) of interest (e.g. blood pressure) over time within a complicated model can be approximated by a reduced-order model [30]. To illustrate this concept, consider the 3-dimensional relationship illustrated in figure 3A. If we look at this system from a 2-dimensional X-Z perspective (figure 3B), we see that Z only varies within a relatively narrow range for any given value of X. Therefore Z may be approximated from a known value of X regardless of the value of Y. Parametric model-order reduction methods attempt to achieve the same goal but differ in the way they find the simplified model.

Parametric model-order reduction methods can be divided into two main categories; (i) methods which focus on reducing the number of nodes (e.g. state variables or compartments) in the model, and (ii) methods that focus on reducing the number of edges (i.e. reactions and interactions). The following section will provide an overview of both approaches.

1.1 Node Reduction Methods

1.1.1 Lumping

Lumping is a general concept that we use in our daily lives to simplify the description of objects. For instance, we use the collective term "fruit" to describe the sweet and fleshy seed-containing products of plants. This semantic lumping is useful because many fruits have common properties so can be treated as a single concept (a lump). Similarly, in QSP models, some nodes can be grouped (lumped) into a single node to produce a simpler form of the model which can be more easily solved. This reduces the number of parameters that need to be experimentally determined [31].

The concept of lumping is relatively simple, conceptually, as all you need to do is to lump together every possible combination of nodes and find the simplest model that retains the desired prediction characteristics of the lumped model when compared to the full-order model. However, the number of possible combinations grows very fast with the size of the model leading to an unfeasibly long list of combinations to be tested. To overcome this issue, Dokoumetzidis et al. [32] proposed a greedy search approach to find the optimal lumping scheme. The utility of the approach was demonstrated by reducing a 62-dimensional model of NF-κB signalling pathway to a 13-dimensional model with acceptable accuracy.

Hasegawa and Duffull proposed an automatable approach to lumping nonlinear QSP models [18] by first linearising the QSP model then lumping using standard techniques. Using this approach, a 28-state bone biology model was reduced to an 8-state model that retained sufficient mechanistic detail to be able to extrapolate to new data for two examples (denosumab and alendronate). In the former case, the model was able to predict long-term responses from short to medium-term data [18] while in the latter, the reduced model for denosumab was reused successfully for alendronate without reference to the original QSP model and without further development of the reduced model [19]. Another approach was applied by Gulati et al. [22] who used an arbitrary series of lumping steps to search through for the optimal lumping scheme. In this approach a 62-state QSP model of the coagulation system was lumped to a 5-state model that was used as the basis for estimating the effect of snake-bite on fibrinogen kinetics.

1.1.2 Conservation Analysis

A conservation relationship occurs when a linear combination of a subset of species within a reaction network model remains constant at all times. For example, consider the following classical example of substrate S conversion to product P catalysed by an enzyme$E$.

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \text{E:S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \text{P + E, (1)}$$

where $E:S$ is a complex formed between the enzyme and substrate. Here, the moles of $(E + E:S)$ is constant and is equal to the initial concentration of the enzyme (E_0). The model can therefore be simplified be removing the constant term. For large models, conservation analysis typically achieves a modest 10–15% reduction in the number of state variables [33]. However, conservation relations in such large models are not obvious and algorithmic approaches are needed to find those relations [27,34,35].

1.1.3 Balanced Truncation

The balanced truncation is based on an observation from control theory that the least observable and least controllable states have no significant contribution to the input-output relationship of interest [36]. Therefore, removing such states from the system results in a reduced-order model that retains most of the input-output behaviour of the full-order model, but, unfortunately, masks the mechanistic basis of the model. Balanced truncation can be viewed as a semiparametric model-order reduction method because it is a function of model parameter although the solution it produces is empirical.

Balanced truncation is typically employed for simplification of linear systems. Of note, Snowden et al. [29] used this approach for reducing a general physiologically based pharmacokinetic (PBPK) model from 16-state to a 5-state system while incurring less than 1% maximal relative error in prediction of venous compartment concentration. A generalisation of this approach to a nonlinear system, called empirical balanced truncation, through numeric approximation of the transformation process, has been proposed [37]. An application of this approach to nonlinear systems biology type models has been recently published [38]. An extension of this approach to a framework where parameters of interest are preserved in the reduced model has also been proposed [25]. These improvements to balanced truncation technique make it highly applicable to QSP models although no such application has been published yet.

1.2 Edge Reduction Methods

1.2.1 Time-Scale Separation

The existence of multiple time-scales is an inherent property of biological systems. Multi-time scale systems contain reactions that occur over several orders of magnitude. A clear example is the effect of denosumab on bone mineral density which is mediated by the high affinity binding to receptor activator of NF-κB ligand (RANKL). Denosumab-RANKL binding kinetics occur over a time scale of seconds-minutes. However, the effect of this binding on bone mineral density evolves over months to years [39]. These separations in time-scales can be exploited to reduce the model by setting the slowly reacting components to be constant relative to the rapidly reacting components or by setting the rapidly reacting species to equilibrate instantaneously. The choice depends on the response variable of interest (in this case binding of RANKL or bone mineral density).

The assumption that rapid reactions are at equilibrium is known as Rapid Equilibrium Approximation (REA) or Quasi-Steady State Approximation (QSSA). The most widely used application of QSSA is the derivation of Michealis-Menten equation of the enzyme kinetics [40,41]. It has also been used to simplify kinetic models for drugs exhibiting target-mediated drug disposition (TMDD) [42].

For large scale QSP models, finding the appropriate partitioning of reactions into fast and slow classes can be challenging. Holland et al. [43] applied the approach to reduce a 25-dimensional model of cardiac β1-adrenergic signalling to a 6-dimensional reduced model with a reasonable predictive performance. Biswal et al. applied a variant of the approach to simplify a 27-state stiff model of calcium homeostasis and bone remodelling and was able to reduce it to "very slow", "slow", and "fast" models that describe different time scales of the full-order model.

1.2.2 Sensitivity Analysis

Sensitivity analysis is a technique that determines how changes in the values of parameters affect a given response variable of interest. Originally, sensitivity analysis was employed to determine the robustness of a system to variability or uncertainty in parameter values. This has been extended to guide model-order reduction by determining the system's state that has the least influence on system outputs. In addition to reducing the number of nodes (compartments), sensitivity analysis can also be used as an edge (reactions) reduction method to determine which reactions contribute least to the output of the system and eliminate those reactions by setting their associated rate parameters to zero [44].

Several algorithms for global sensitivity analysis have been developed with Fourier amplitude and Sobol's sensitivity analysis being common [45,46]. Jayachandran et al. employed Sobol's method and achieved a 50% reduction in model parameters for a model of 6-mercaptopurine effects in children with acute lymphoblastic leukaemia. The technique enabled the estimation of model parameters for individual patients and subsequent use in treatment individualisation [21]. Gueorguieva et al. applied an extended Fourier amplitude sensitivity analysis test to reduce a 14-compartment whole body PBPK model of diazepam while preserving the dynamical behaviour of the arterial concentration compartment [47].

2 Nonparametric methods

All nonparametric methods are based on training an empirical function to replicate the input-output behaviour of the QSP system. The training involves generation of a large amount of pseudo-data from the full-order QSP model and then either fitting a user-defined empirical model (termed here Empirical Approximation) to the data or by a machine learning technique such as an artificial neural network (ANN). None of these methods retain mechanistic characteristics of the full-order model and therefore cannot be used for extrapolation to data that might arise from inputs beyond the range of the training inputs (e.g. to a dose that is not within the range of doses used to create the pseudo-data). These approaches are useful when parametric methods are not possible due to high model complexity or even when the structure of the model is unknown, e.g., models provided by third parties. The only requirement of nonparametric approaches is the ability to evaluate the original full-order model at a given set of input variables.

2.1 Empirical Approximation

In this approach an empirical model, chosen by the user, is fitted to the pseudo-data. This approach does not imply or require any particular structure of the full-order model. It is, however, limited to situations in which there are a small number of response variables of interest (e.g. just mean arterial blood pressure and heart rate).

There have been several applications of empirical approximations. Ooi et al. [20] used logistic and polynomial functions to approximate the relationship between warfarin exposure and response in terms of International Normalised Ratio (INR) to predict warfarin dose requirements. Gulati et al. [48] used a similar approach to derive an empirical function that approximated the relationship between the concentration of activating agents and a proposed clotting time test based on a QSP model of the coagulation network. The function was used to optimise the design of a pilot study aimed at evaluating the proposed test as a tool for monitoring of enoxaparin therapy [48]. Similarly, Dumont et al. [49] used PBPK simulated pseudo-data to develop a simplified pharmacokinetic model that was used to design a paediatric clinical trial.

2.2 Artificial neural networks

Artificial neural networks (ANNs) are a set of mathematical algorithms that are able to capture complex relationships between the input variables and response variables. According to the universal approximation theorem [50], an ANN of enough size can approximate any function with any accuracy as defined by the user. This makes ANNs potentially useful for approximating complex multi-dimensional and non-linear input-output relationships of QSP models. They have shown to be an efficient tool for model-order reduction of various types of models in fields such as systems engineering and control [51].

ANNs are particularly useful in nonparametric model-order reduction in situations where it is hard to find an appropriate empirical function to approximate a given input-output relationship. For example, when there is a large number of response variables or highly correlated independent variables to be considered. ANNs have been shown to produce fast but accurate reduced-order models that approximate nonlinear dynamical systems [26]. There are currently no published stduies on the use of ANNs for QSP models. Unpublished work (by the authors) illustrated the use of an ANN surrogate models to approximate the heparin dose-response relationship in children which was 10,000-fold faster for simulation than the original full-order QSP model. Such a speed boost could enable use of complex QSP models in simulation, estimation, and potentially control purposes.

Utility of model-order reduction methods

An all-purpose ideal method for model-order reduction does not exist. Indeed, the context in which model reduction is applied will help determine the best method. The proposed framework allows a natural categorisation of methods based on their utility (illustrated in Table 1).

The model-order reduction techniques discussed in this article have been grouped into parametric and nonparametric methods. Parametric methods are often useful in settings where repeated use of the model particularly in which the parameter values have mechanistic meaning, e.g. in estimation or simulation. In

contrast non-parametric methods are valuable for any setting where repeated simulation is the goal. Since non-parametric models do not retain mechanistic structure their use in estimation may be limited.

Factors that influence the choice of model-order reduction method

The nature of the full-order model

The nature of the model to be reduced is an important factor in choosing the best method for model reduction. Linear models (e.g. most pharmacokinetic models including physiologically based PK models) are the easiest and work well with any method. In contrast, nonlinear models, most pharmacodynamic QSP models, tend to be harder to simplify and many methods need to be adapted to work with them. If the QSP model is based on continuous functions (e.g the Bone model [52]) then most parametric methods would be useful. However models that are a discrete combination of large continuous functions (e.g. the coagulation model [14]) are really only amenable to nonparametric methods. Nonparametric methods are well suited to any type of full-order model and appear to have no requirements about their structure.

The purpose of the reduced model

The two main uses for reduced models are for simulation (to simulate new scenarios and for design) or estimation (to be used in software like NONMEM for estimating parameters for new populations).

For simulation, any method can be used if the simulation is based on sets of input variables (e.g. doses, dose intervals, observation times) that were considered when the model reduction was conducted. For instance, if pseudo-data were simulated from the full-order model to develop a reduced model then the reduced model can only be used to interpolate within the range of pseudo-data inputs. In contrast, any fully parametric method, can be used to simulate under (in theory) any set of input variables. In this context, semiparametric methods such as balanced truncation behave like nonparametric methods.

However, if the purpose of the reduced model is to be used for estimation then any fully parametric method could be used for model-order reduction. The utility of the Empirical Approximation will depend on the circumstance. Importantly, though if it was desirable for the parameters to provide some mechanistic meaning then only the fully parametric methods would be of value.

How easy is it to do?

This is about balancing simplicity and accuracy of the reduced model. Conservation analysis is an especially important parametric method as it does not incur any error and therefore retains all properties of the original model. It is often therefore recommended as a first step for reduction of a large model. However, finding conserved moieties is not straightforward for larger models and it seldom reduces a model by more than 10-15% [33]. Most methods are automatable, meaning that user input is not required for the algorithm to run. This makes complicated mathematical methods, such as linearisation and lumping of nonlinear models [18], more user friendly and accessible. The simplest method that retains acceptable accuracy is using an Empirical Approximation however it has limitations regarding generality.

Most of the model-order reduction methods require the use of a specialised software, except for the Empirical Approximation method. Many of the algorithms are available as pre-coded toolboxes in MATLAB. For example, the Global Optimisation Toolbox for simulated annealing, Deep Learning Toolbox for ANNs, and the SAFE Toolbox for global sensitivity analysis [53].

In all parametric approaches there is a trade-off between model simplicity (model-order) and its predictive performance. Hasegawa et al. [54] proposed a composite criterion-based approach that couples the two opposing factors together to find the optimal reduced model. For all methods (except conservation analysis) it will be necessary for the user to define the trade-off.

Combining methods

Of note, different model-order reduction methods are not mutually exclusive and a combination of one or more methods could be applied to effectively reduce a model while exploiting the potential advantages of

each of the methods used. For example, Snowden et al. [29], used proper lumping and balanced truncation to reduce two independent PBPK and systems biology models, respectively, which were then combined into a single simplified QSP model.

Discussion

Models characterising drug actions and biological systems can have different degrees of complexity. Compartmental models, that also incorporate empirical pharmacodyanmic functions, e.g. the Emax model, for instance, are built chiefly on the observed data with the aim of being utilised for predictions to new scenarios. However, their predictive performance can be limited [18,19,55] when used for extrapolation. Such models when incorporating sufficient population data usually provide good insights for answering many drug development and utilisation questions. However, extrapolating beyond the available data remains problematic. In contrast, QSP models (including PBPK models) are built upon mechanistic knowledge of biological systems and drug actions and therefore allow for predictions of outcomes in previously untested scenarios. Despite the potential benefits of these mechanistic models, the issue of model complexity can make them intractable for many applications. There remains the need for a balance between model usability and complexity. Model-order reduction can act as a bridging approach that provides a means of producing intermediary scale models that bring together the strengths of both simplicity and mechanistic predictive ability. Model-order reduction can also be used to estimate between-subject variability in some parameters using population approach [22] which could then be used to feedback into the original QSP model.

Despite the benefits that model-order reduction provides, careful consideration of the limitations of different methods should be given. Most model-order reduction methods will incur some degree of information loss relative to the full-order model. The greater extent of model reduction means, in most situations, greater loss of information. The optimal balance of the trade-off between model simplicity and accuracy will largely depend on the intended use of the reduced model. It is also important to realise the local nature of many model-order reduction methods, i.e., the reduced model may only be valid for a specific set of parameter values. Of note, most model-order reduction techniques incur a large upfront computational cost for producing a reduced model. This should be balanced against the expected computational speed-up gained by the reduced model in order to determine whether a particular model-order reduction technique is worthwhile.

In this work, we have developed an overall framework for considering various methods of model-order reduction that can be useful in the context of QSP models (figure 2 & table 1). The framework provides a natural categorisation of various methods based on their utility. We hope this can guide modellers to choose the best model-order reduction method that suits their needs.

In summary, model-order reduction can have a significant impact on the future of modelling and simulation in drug development and clinical use [56]. The vast repository of information contained in QSP models can act as source library from which smaller models to describe a specific input-output relationship(s) can be extracted through model-order reduction methods. The existence of automated tools for model-order reduction can streamline the process of model building saving time and effort. It is therefore anticipated that pharmacometricians and clinical pharmacologists would benefit from being familiar with various model-order reduction techniques and there benefits and potential uses.

Acknowledgment

None

Conflict of Interest

The authors declare no conflict of interest related to this article.

References

1. Zhang L, Pfister M, Meibohm B. Concepts and challenges in quantitative pharmacology and model-based drug development. *AAPS J.* 2008;10(4):552-559.

2. Standing JF. Understanding and applying pharmacometric modelling and simulation in clinical practice and research. *Br J Clin Pharmacol.* 2017;83(2):247-254.

3. van der Graaf PH, Benson N. Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm Res.* 2011;28(7):1460-1464.

4. Benson N. Quantitative Systems Pharmacology and Empirical Models: Friends or Foes? *CPT Pharmacometrics Syst Pharmacol.*2019;8(3):135-137.

5. Birtwistle MR, Hansen J, Gallo JM, et al. Systems Pharmacology: An Overview. In: Mager DE, Kimko HHC, eds. *Systems Pharmacology and Pharmacodynamics.* Cham: Springer International Publishing; 2016:53-80.

6. Peterson MC, Riggs MM. FDA Advisory Meeting Clinical Pharmacology Review Utilizes a Quantitative Systems Pharmacology (QSP) Model: A Watershed Moment? *CPT: pharmacometrics & systems pharmacology.*2015;4(3):e00020-e00020.

7. Sorger PK, Allerheiligen SR, Abernethy DR, et al. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. Paper presented at: An NIH white paper by the QSP workshop group2011.

8. Benson N, Matsuura T, Smirnov S, et al. Systems pharmacology of the nerve growth factor pathway: use of a systems biology model for the identification of key drug targets using sensitivity analysis and the integration of physiology and pharmacology. *Interface Focus.*2013;3(2):20120071-20120071.

9. Barrett JS, Bucci-Rechtweg C, Amy Cheung SY, et al. Pediatric Extrapolation in Type 2 Diabetes: Future Implications of a Workshop. *Clin Pharmacol Ther.* 2020.

10. Lazarou G, Chelliah V, Small BG, Walker M, van der Graaf PH, Kierzek AM. Integration of Omics Data Sources to Inform Mechanistic Modeling of Immune-Oncology Therapies: A Tutorial for Clinical Pharmacologists. *Clin Pharmacol Ther.* 2020;107(4):858-870.

11. Milberg O, Gong C, Jafarnejad M, et al. A QSP Model for Predicting Clinical Responses to Monotherapy, Combination and Sequential Therapy Following CTLA-4, PD-1, and PD-L1 Checkpoint Blockade. *Sci Rep.*2019;9(1):11286.

12. Musante CJ, Ramanujan S, Schmidt BJ, Ghobrial OG, Lu J, Heatherington AC. Quantitative Systems Pharmacology: A Case for Disease Models. *Clin Pharmacol Ther.* 2017;101(1):24-27.

13. Chae D, Son M, Kim Y, Son H, Park K. Mechanistic Model for Blood Pressure and Heart Rate Changes Produced by Telmisartan in Human Beings. *Basic Clin Pharmacol Toxicol.* 2018;122(1):139-148.

14. Wajima T, Isbister GK, Duffull SB. A comprehensive model for the humoral coagulation network in humans. *Clin Pharmacol Ther.*2009;86(3):290-298.

15. Wanner G, Hairer E. *Solving ordinary differential equations II.* Springer Berlin Heidelberg; 1996.

16. Ribba B, Grimm HP, Agoram B, et al. Methodologies for Quantitative Systems Pharmacology (QSP) Models: Design and Estimation. *CPT Pharmacometrics Syst Pharmacol.* 2017;6(8):496-498.

17. Shivva V, Korell J, Tucker IG, Duffull SB. An approach for identifiability of population pharmacokinetic-pharmacodynamic models. *CPT Pharmacometrics Syst Pharmacol.* 2013;2:e49.

18. Hasegawa C, Duffull SB. Automated Scale Reduction of Nonlinear QSP Models With an Illustrative Application to a Bone Biology System. *CPT Pharmacometrics Syst Pharmacol.* 2018;7(9):562-572.

19. Hasegawa C, Duffull SB. Reusing Smaller Versions of Large Models: A Case Example of Reuse of a Simplified Bone Model. *Clin Pharmacol Ther.* 2019;106(6):1184-1186.

20. Ooi Q-X, Wright DFB, Isbister GK, Duffull SB. A factor VII-based method for the prediction of anticoagulant response to warfarin.*Sci Rep.* 2018;8(1):12041.

21. Jayachandran D, Rundell AE, Hannemann RE, Vik TA, Ramkrishna D. Optimal chemotherapy for leukemia: a model-based strategy for individualized treatment. *PLoS One.* 2014;9(10):e109623.

22. Gulati A, Isbister G, Duffull S. Scale Reduction of a Systems Coagulation Model With an Application to Modeling Pharmacokinetic–Pharmacodynamic Data. *CPT: Pharmacometrics & Systems Pharmacology.* 2014;3(1):90.

23. Gulati A, Faed JM, Isbister GK, Duffull SB. Development and evaluation of a prototype of a novel clotting time test to monitor enoxaparin. *Pharm Res.* 2012;29(1):225-235.

24. Biswal B, Sen S, Maka S. A structure preserving model order reduction method for calcium homeostatic system. *Math Biosci.*2019;312:8-22.

25. Prajapati AK, Prasad R. Model Order Reduction by Using the Balanced Truncation and Factor Division Methods. *IETE J Res.*2019;65(6):827-842.

26. San O, Maulik R, Ahmed M. An artificial neural network framework for reduced order modeling of transient flows. *Communications in Nonlinear Science and Numerical Simulation.* 2019;77:271-287.

27. Vallabhajosyula RR, Chickarmane V, Sauro HM. Conservation analysis of large biochemical networks. *Bioinformatics.*2006;22(3):346-353.

28. Snowden TJ, van der Graaf PH, Tindall MJ. Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends. *Bull Math Biol.* 2017;79(7):1449-1486.

29. Snowden TJ, van der Graaf PH, Tindall MJ. Model reduction in mathematical pharmacology. *J Pharmacokinet Pharmacodyn.*2018;45(4):537-555.

30. Antoulas AC. Approximation of Large-Scale Dynamical Systems: An Overview. *IFAC Proceedings Volumes.* 2004;37(11):19-28.

31. Wei J, Kuo JCW. Lumping Analysis in Monomolecular Reaction Systems. Analysis of the Exactly Lumpable System. *Industrial & Engineering Chemistry Fundamentals.* 1969;8(1):114-123.

32. Dokoumetzidis A, Aarons L. Proper lumping in systems biology models.*IET Syst Biol.* 2009;3(1):40-51.

33. Vallabhajosyula RR, Sauro HM. Complexity Reduction of Biochemical Networks. Paper presented at: Proceedings of the 2006 Winter Simulation Conference; 3-6 Dec. 2006, 2006.

34. Reder C. Metabolic control theory: A structural approach. *J Theor Biol.* 1988;135(2):175-201.

35. Sauro HM, Ingalls B. Conservation analysis in biochemical networks: computational issues for software writers. *Biophys Chem.*2004;109(1):1-15.

36. Moore B. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control.* 1981;26(1):17-32.

37. Hahn J, Edgar TF. An improved method for nonlinear model reduction using balancing of empirical gramians. *Comput Chem Eng.*2002;26(10):1379-1397.

38. Snowden TJ, van der Graaf PH, Tindall MJ. A combined model reduction algorithm for controlled biochemical systems. *BMC Syst Biol.*2017;11(1):17.

39. Peterson M, Riggs M. Predicting Nonlinear Changes in Bone Mineral Density Over Time Using a Multiscale Systems Pharmacology Model.*CPT: Pharmacometrics & Systems Pharmacology.* 2012;1(11):14.

40. Michaelis L, Menten M. Die kinetik der invertinwirkung Biochem Z 49: 333–369. *Find this article online.* 1913.

41. Briggs GE, Haldane JB. A Note on the Kinetics of Enzyme Action. *The Biochemical journal.* 1925;19(2):338-339.

42. Gibiansky L, Gibiansky E, Kakkar T, Ma P. Approximations of the target-mediated drug disposition model and identifiability of model parameters. *J Pharmacokinet Pharmacodyn.* 2008;35(5):573-591.

43. Holland DO, Krainak NC, Saucerman JJ. Graphical approach to model reduction for nonlinear biochemical networks. *PLoS One.*2011;6(8):e23795.

44. Zhang XY, Trame MN, Lesko LJ, Schmidt S. Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT: pharmacometrics & systems pharmacology.*2015;4(2):69-79.

45. Christopher Frey H, Patil SR. Identification and Review of Sensitivity Analysis Methods. *Risk Anal.* 2002;22(3):553-578.

46. Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation.* 2001;55(1):271-280.

47. Gueorguieva I, Nestorov IA, Rowland M. Reducing whole body physiologically based pharmacokinetic models using global sensitivity analysis: diazepam case study. *J Pharmacokinet Pharmacodyn.*2006;33(1):1-27.

48. Gulati A, Faed JM, Isbister GK, Duffull SB. Application of Adaptive DP-optimality to Design a Pilot Study for a Clotting Time Test for Enoxaparin. *Pharm Res.* 2015;32(10):3391-3402.

49. Dumont C, Mentre F, Gaynor C, Brendel K, Gesson C, Chenel M. Optimal sampling times for a drug and its metabolite using SIMCYP((R)) simulations as prior information. *Clin Pharmacokinet.*2013;52(1):43-57.

50. Hornik K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* 1991;4(2):251-257.

51. Rayas-Sánchez JE. Artificial Neural Networks and Space Mapping for EM-Based Modeling and Design of Microwave Circuits. In: Koziel S, Leifsson L, eds. *Surrogate-Based Modeling and Optimization: Applications in Engineering.* New York, NY: Springer New York; 2013:147-169.

52. Peterson MC, Riggs MM. A physiologically based mathematical model of integrated calcium homeostasis and bone remodeling. *Bone.*2010;46(1):49-63.

53. Pianosi F, Sarrazin F, Wagener T. A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software.*2015;70:80-85.

54. Hasegawa C, Duffull SB. Selection and Qualification of Simplified QSP Models When Using Model Order Reduction Techniques. *AAPS J.*2017;20(1):2.

55. Duffull SB, Wright DFB. What do we learn from repeated population analyses? *Br J Clin Pharmacol.* 2015;79(1):40-47.

56. Mentré F, Friberg LE, Duffull S, et al. Pharmacometrics and Systems Pharmacology 2030. *Clin Pharmacol Ther.* 2020;107(1):76-78.

Tables

Table 1: Comparison of utilities of different methods of model-order reduction

	Parametric methods
	Lumping
System agnosticism	Utility depends on the nature of the full-orde
Automatability	Yes

	Parametric methods
Mechanistic relevance	Yes
Accuracy	Determined by the user as a trade-off of com
Computational ease	Computationally demanding for large nonline
Predictive ability of the reduced model (interpolation and/or extrapolation)	Both
Requirement of experimental data for reduced model parameter estimation	No
Suitable for parameter estimation	Yes

Figure legends

Figure 1: A hypothetical 16-state nonlinear QSP-style model. Each node (circle) represents a system state (variable). Edges (solid black arrows) represent mass (or molar) balance reactions and dashed lines ending with bar are mass (or molar) action. (+) represents a positive reaction, e.g. stimulation of a reaction, and (-) represents a negative reaction, e.g., inhibition of a reaction.

Figure 2: A general framework for understanding different approaches to model-order reduction. The original full-order model is shown as the starting point. The model can then be reduced using either parametric methods (shown as blue arrows) or nonparametric methods (shown in orange). Parametric approaches produced a reduced-order model based on mathematical techniques that stem from either reducing the number of compartments (nodes) and/or reactions (edges). The nonparametric methods approximate the full-order model input-output relationship (e.g. dose to INR relationship) with a simpler empirical model, which might be a black-box model input-output such as an artificial neural network or a user defined empirical function.

Figure 3: An illustration of the concept of dimension reduction. (A) 3-dimensional surface showing a hypothetical relationship between variables X,Y, and Z; (B) The same surface rotated clock-wise around the Z-axis so that only X-Z perspective is visible. With a small margin of error, Z can be approximated as a function of X, rather than as an independent parameter, thereby reducing the order of the model.

Appendix 1 Glossary of Terms

Amplification: a process in which the response of the system to a stimulus increases over time via internal processes

Automatable : An algorithmic method that requires no user intervention.

Compartment : a chamber (real or hypothetical) in which the variable is homogenous throughout and assumed to be instantaneously mixed, for example the concentration in a compartment.

Controllable state : A state that is changed when the input variables are changed (e.g. the concentration in the body is changed by dose).

Damping: a process in which the oscillations in a system, caused by some stimulus are reduced, restricted or inhibited over time**.**

Edge : a graphical representation of movement or an action. This is usually shown as a line with a bar at the end or an arrow.

Empirical function : A user defined function to describe a set of input-output data with no (or very few) assumption about the underlying processes that generated the data.

Extrapolation : The prediction of an outcome or response by a model under conditions or values of input variables other than those (typically outside of the range) upon which the model was built.

Full-order model: a model that is its original (full) size, e.g. an original QSP model

Input-output (I/O): pertaining to the input variables (e.g. dose, dose time, route, etc) and the output variables represented by a set of response variables (e.g. haemodynamic variables such as blood pressure and heart rate etc)

Model dimension : The number of state variables & parameters in a model.

Model-order reduction : a method to reduce a large model into a simpler model

Model-order : A representation of the size of a model which is related to the number of parameters.

Node : a graphical representation of a compartment in a model. This is usually shown as a box or a circle (depending on preference)

Nonlinear system : A system whose outputs are not proportional to inputs. In a nonlinear system, small changes in the inputs may lead to substantial changes in the outputs.

Observable state : A state (node) that when changed results in a change in the observed response (e.g. changes in the concentration of a drug results in changes in blood pressure).

Pseudo-data : A set of input-output data simulated (e.g. from full-order QSP model) to reflecting its behaviour and is used in training a surrogate model.

Quasi-steady state approximation : An approach to approximating equilibrium systems with different time scales by assuming that rapidly equilibrating states reach equilibrium instantaneously.

Reduced-order model : a model that is smaller in size (number of states & parameters) than the full-order model from which it was derived.

State : a variable in a model whose value evolves through time. This is the same as a concentration in a pharmacokinetic model. It differs from a compartment as the compartment is the location (hypothetical or real) of the concentration. In a pharmacokinetic example, a state is the concentration of a particular drug in a particular compartment. A state is normally shown graphically as a node.

Structural Identifiability : A property of a model whereby no two (or more) sets of parameters give identical model predictions. A model that is identifiable can be used for parameter estimation (and vice versa).

System agnostic : The method applied to the system is not affected by/dependent on the nature of the system.