Probability Bounds Analysis for Nonlinear Dynamic Process Models

Joshua A. Enszer and Youdong Lin
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

Scott Ferson
Applied Biomathematics, Setauket, NY 11733

George F. Corliss
Dept. of Electrical and Computer Engineering, Marquette University, Milwaukee, WI 53201

Mark A. Stadtherr
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

Dynamic process models frequently involve uncertain parameters and inputs. Propagating these uncertainties rigorously through a mathematical model to determine their effect on system states and outputs is a challenging problem. In this work, we describe a new approach, based on the use of Taylor model methods, for the rigorous propagation of uncertainties through nonlinear systems of ordinary differential equations (ODEs). We concentrate on uncertainties whose distribution is not known precisely, but can be bounded by a probability box (p-box), and show how to use p-boxes in the context of Taylor models. This allows us to obtain p-box representations of the uncertainties in the state variable outputs of a nonlinear ODE model. Examples having two to three uncertain parameters or initial states and focused on reaction process dynamics are used to demonstrate the potential of this approach. Using this method, rigorous probability bounds can be determined at a computational cost that is significantly less than that required by Monte Carlo analysis. © 2010 American Institute of Chemical Engineers AIChE J, 57: 404–422, 2011
Keywords: design (process simulation), mathematical modeling, numerical solutions, reactor analysis, bioprocess engineering

Introduction

Systems of ordinary differential equations (ODEs) are the basis for many mathematical models in engineering and science. For example, models of reactor dynamics are based on unsteady-state material and energy balances and thus take the form of a system of first-order ODEs, which typically is nonlinear. Generally, the problem of interest is an initial value problem (IVP), in which an initial state is given and the system then integrated numerically until some final time (time horizon) is reached, thus determining numerical approximations of the final state, as well as of the trajectory followed to reach it.

Often these dynamic models involve uncertainties in parameters and/or initial states. Analysis of the impact of such uncertainties is clearly important in models of process dynamics, as used, for example, in state and parameter estimation[1,2]

and process control.[3-5] It is a challenging problem to propagate uncertainties through a nonlinear ODE system to rigorously predict the uncertainty in the model outputs. The problem is further complicated by the fact that the probability distributions describing the uncertainties may not be known precisely, if they are known at all. If there is no known probability distribution for an uncertain quantity, but only bounds, then the uncertainty can be modeled using an interval. If some knowledge of the probability distribution is available, but it is imprecise, then this can be modeled using a probability box[6] (p-box), which provides interval-like upper and lower bounds on the cumulative probability distribution function for the uncertain quantity.[7] We will consider both cases here, but concentrate on the latter case in which p-boxes are used to represent imprecise distributions of uncertainty.

A common approach for studying the effect of uncertainties in this and other contexts is the use of repeated sampling. This is frequently based on Monte Carlo methods, but alternative techniques (e.g., Latin hypercube and quasirandom sequences) are also available and can be more efficient. In general, it is not possible to investigate the complete space of uncertainties in a finite number of samples. Thus, sampling methods such as Monte Carlo cannot produce rigorous bounds, as they may fail to capture all system behaviors, especially when nonlinearities are involved. Monte Carlo methods have other limitations[8] as well, especially in dealing with uncertainties that have unknown dependencies or that cannot be characterized by a precise probability distribution. The latter case can be dealt with using a "second-order" Monte Carlo approach[9,10] in which the space of the probability distributions is sampled in an outer loop, with this sample distribution then used in a "standard" Monte Carlo simulation of the parameter uncertainty in an inner loop. Sampling methods can also become quite expensive computationally, especially if a second-order approach is needed, as the number of samples needed to obtain a meaningful analysis of the uncertainty may be very large.

We will describe here a much different approach, not based on sampling, for the analysis of uncertainty in nonlinear dynamic systems. Given interval bounds on the uncertain quantities, this approach can be used to rigorously bound all possible trajectories of an IVP for an uncertain ODE model. Furthermore, given bounds (p-boxes) on the probability distributions of the uncertain quantities, this approach can be used to bound rigorously the probability distributions of model outputs at specified points in time. This method is enabled by the use of Taylor models to represent the solution of an IVP with uncertain parameters and/or initial states, as described recently by Lin and Stadtherr.[11]

This article is divided as follows. In the next section, we will provide a formal statement of the problems to be addressed. This will be followed by a section in which we provide general background on the tools used here, including intervals, p-boxes, and Taylor models. Then, we introduce the proposed solution methods, and finally, we present a number of examples that demonstrate this method, with comparisons to results obtained from Monte Carlo analysis.

Problem Statement

Consider an IVP for a parametric, autonomous ODE with uncertain parameters and initial states:

$$\frac{d\boldsymbol{y}}{dt} = \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{\theta}), \quad \boldsymbol{y}(t_0) = \boldsymbol{y}_0 \in \boldsymbol{Y}_0, \quad \boldsymbol{\theta} \in \Theta. \qquad (1)$$

Here, $\boldsymbol{y}$ is the n-dimensional vector of state variables with initial value $\boldsymbol{y}_0$, $\boldsymbol{\theta}$ is a p-dimensional vector of time-invariant parameters, and $t \in [t_0, t_f]$ for some $t_f > t_0$. The vectors $\boldsymbol{Y}_0$ and Θ are intervals that enclose uncertainties in the initial states and parameters, respectively. Furthermore, information about the probability distribution of the uncertainty is available for at least one component of $\boldsymbol{Y}_0$ or Θ, and this information is expressed as a p-box, as formally defined in the next section. We also assume that $\boldsymbol{f}$ can be represented by a finite number of standard functions, and that $\boldsymbol{f}$ is $(k - 1)$ times continuously differentiable with respect to $\boldsymbol{y}$ and $(q + 1)$ times continuously differentiable with respect to $\boldsymbol{\theta}$. Here, k is the order of the truncation error in the interval Taylor series (ITS) method used in the solution procedure, and q is the order of the Taylor model used in the solution method to represent dependence on parameters and initial values. ODE models that are nonautonomous, or that involve parameters with known time dependence, can be converted to the form of Eq. 1 by the introduction of additional state variables.

There are two goals. First, we seek to obtain verified (mathematically and computationally guaranteed) enclosures $\boldsymbol{Y}_j$ of the state variables $\boldsymbol{y}_j = \boldsymbol{y}(t_j)$ at the endpoints $t_j \in [t_0, t_f]$ of each time step in the numerical integration procedure used to solve Eq. 1. Second, but of particular interest here, for any such enclosure $\boldsymbol{Y}_j \ni \boldsymbol{y}_j$, we seek to determine verified bounds, in the form of a p-box, on the probability distribution for the values of $\boldsymbol{y}_j$. Thus, we can obtain rigorous bounds on the probabilities that desired outcomes are achieved.

Background

Uncertainties can be represented in a variety of ways. A typical approach is to use real-valued variables characterized by some type of probability distribution. However, in many cases, the true probability distribution is not known precisely. This leads to the concept of a probability distribution variable (PDV), as described by Li and Hyman,[12] which can be characterized using a generalized probabilistic discretization (GPD), such as a p-box. If there is no knowledge at all about the probability distribution, with only upper and lower bounds on the uncertain quantity available, then this can be represented by an interval, which can be regarded as a limiting case of the p-box. Depending on the type of information that is available, other representations of uncertain knowledge may be appropriate, such as the use of fuzzy numbers[13,14] or clouds.[15] We will focus here on the use of intervals and p-boxes, and provide some background on these topics. We also provide background on Taylor models, which provide a key tool for use in the computational method described here.

Interval analysis

A real interval X is the set of real numbers between (and including) a specified lower bound (denoted $\underline{X}$) and upper bound (denoted $\overline{X}$). That is, $X = [\underline{X}, \overline{X}] = \{x \in \Re | \underline{X} \le x \le \overline{X}\}$. A real interval vector $X = (X_1, X_2, \ldots, X_n)^T$ has n real interval

components and can be interpreted as an *n*-dimensional rectangle or box. Interval matrices are similarly defined. Basic arithmetic operations are defined on intervals according to X op $Y = \{x$ op $y \mid x \in X, y \in Y\}$, op $\in \{+, -, \times, \div\}$. Division in the case of Y containing zero is allowed in extensions of interval arithmetic.[16] Addition and multiplication are commutative and associative but only subdistributive. Interval versions of the elementary functions can also be defined. When implemented with outward rounding (lower bound rounded down, upper bound rounded up), interval operations can be used to obtain rigorous bounds on function ranges.

For a real function $f(x)$, the interval extension $F(X)$ encloses the range of $f(x)$ for $x \in X$. That is, $\{f(x) \mid x \in X\} \subseteq F(X)$. When $f(x)$ can be expressed as a series of arithmetic operations and elementary functions, the ''natural'' interval extension can be obtained by substituting the given interval X into the expression for $f(x)$ and evaluating using interval arithmetic. If any variable occurs more than once in the expression for $f(x)$, then computing an interval extension in this way may result in overestimation of the function range because of the ''dependency'' problem. Although a variable may take on any value within its interval, it must take on the *same* value each time it occurs in an expression. However, this type of dependency is not recognized when the natural interval extension is computed. In effect, when the natural interval extension is used, the range computed for the function is the range that would occur if each instance of a particular variable was allowed to take on a different value in its interval range.

Another source of overestimation that may arise in the use of interval methods is the ''wrapping'' effect.[17] This occurs when an interval is used to enclose (wrap) a set of results that is not an interval. If this overestimation is propagated from step to step in an integration procedure for ODEs, it can lead quickly to the loss of a meaningful enclosure. Historically, the problems of wrapping and dependency have caused interval methods to acquire a reputation for producing overly loose and conservative bounds of limited usefulness. However, current interval methods, including the use of techniques such as Taylor models, as discussed below, can often yield rigorous bounds with very little overestimation.

Several good introductions to interval analysis, as well as interval arithmetic and other aspects of computing with intervals, are available.[16,18–21] Implementations of interval arithmetic and elementary functions are also readily available, and recent compilers from Sun Microsystems directly support interval arithmetic and an interval data type.

P-boxes

Intervals give upper and lower bounds only and provide no knowledge about the distribution of uncertainties. If additional knowledge about the distribution is available, then we will capture this through the use of p-boxes and the methods of probability bounds analysis.

Ideas about the use of bounds on probabilities have appeared throughout the development of probability theory. The first methods of probability bounds analysis that could be routinely used in practical applications (e.g., environmental risk assessments) were developed in the 1980s. Yager[22] described the elementary procedures by which it is possible
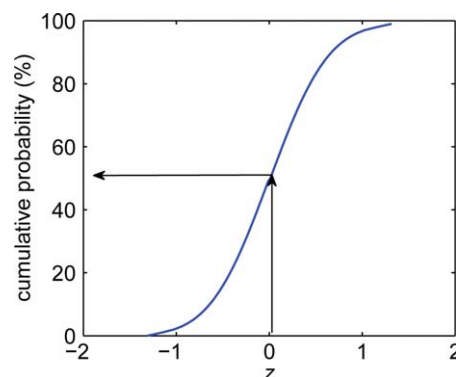

Figure 1. Example of cumulative probability density function (CDF).

Here, $P(x \leq 0) = 50\%$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to determine bounds on convolutions computed from probability distributions for independent random variables. (The operation of finding the probability distribution of a sum—or a product, difference, quotient, etc.—of random variables specified by probability distributions is a convolution.) At about the same time, Frank et al.[23] solved the problem of bounding distributions of sums of random variables when no information about their interdependency is available. Extending this approach, Williamson and Downs[7] developed a semianalytical approach that computes rigorous bounds on the cumulative distribution functions of convolutions without assuming independence between the operands. Ferson et al.[6] coined the name ''probability box,'' or p-box, for the envelope of the bounds on the distribution function. They applied several new and classic results on bounding probabilities to develop a suite of p-box shapes to express a wide array of states of knowledge involving possible information about the range, moments, order statistics, and other shape information about the distributions of random variables.[24,25] Berleant and Zhang[26] described a complementary intermediate approach based on mathematical programming for the case when the correlation but not the dependence function can be specified. In this subsection, we discuss the basic idea of a p-box and describe how operations with p-boxes can be performed.

A p-box provides interval-like bounds on the cumulative distribution function (CDF) describing a probability distribution. For some quantity (variable or parameter) x, the CDF $F_x(z)$ gives the probability that $x \leq z$. An example CDF is shown in Figure 1, which is marked to indicate that, for this CDF, the probability that $x \leq 0$ is 50%. In practice, it is possible that the probability distribution describing some uncertain quantity is itself uncertain. In this case, a p-box can be used to bound the uncertain probability distribution. A p-box $PB(x) = (L_x, R_x)$ is the set of all CDFs enclosed by two bounding CDFs $L_x(z)$ and $R_x(z)$ with finite support. That is, $PB(x) = (L_x, R_x) = \{F_x(z) \mid L_x(z) \geq F_x(z) \geq R_x(z)\}$. For a given value of z, the left bound $L_x(z)$ of the p-box gives the upper bound on the probability that $x \leq z$ and the right bound $R_x(z)$ gives the lower bound on this probability. This is shown for an example p-box in Figure 2a, which is marked to indicate that, for this p-box, the probability that $x \leq 1.5$ is bounded by the interval [14.2,23.3]%. Similarly, for a given
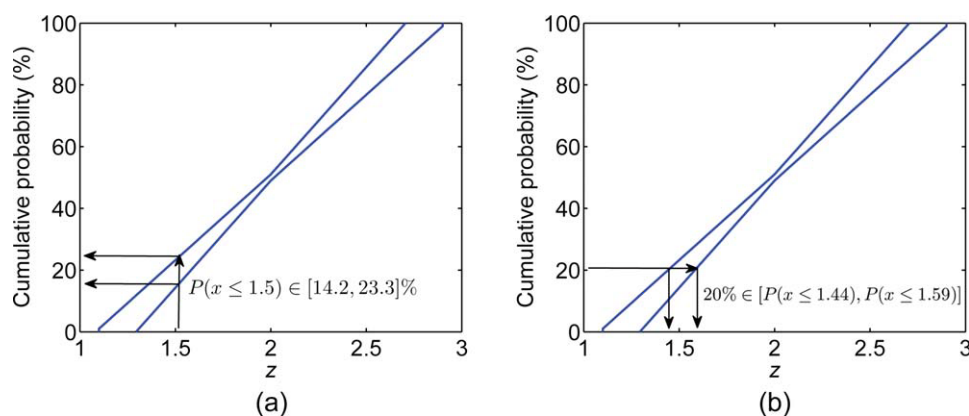
Figure 2. Interpretation of a p-box PB(x).

(a) The probability that $x \leq 1.5$ is bounded by the interval [14.2,23.3]%. (b) The 20th percentile value of x is bounded by the interval [1.44,1.59]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

value of the cumulative probability, $L_x(z)$ and $R_x(z)$ give the lower and upper bounds on the values of x for which this probability is possible. For the example p-box in Figure 2b, this is marked to indicate that the 20th percentile value of x is bounded by the interval [1.44,1.59].

P-boxes may be constructed from any available information about an uncertain quantity or may be generated from distributions of a specified shape but uncertain distribution parameters. Figure 3 shows four p-boxes created from different types of information. Figure 3a is a p-box for the case in which only the minimum and maximum values of an uncer-

tain quantity are known; that is, this is the p-box representation of an interval. If the minimum, maximum, and median are known, this gives the p-box in Figure 3b. For the p-box in Figure 3c, the bounding distributions are obtained by specifying a uniform distribution with known median and uncertain but bounded maximum and minimum. For the p-box in Figure 3d, the bounding distributions are obtained by specifying a normal distribution of known mean and uncertain standard deviation. As the bounding distributions are assumed to have finite support, the normal distributions must be truncated to a finite range (as in Monte Carlo methods),
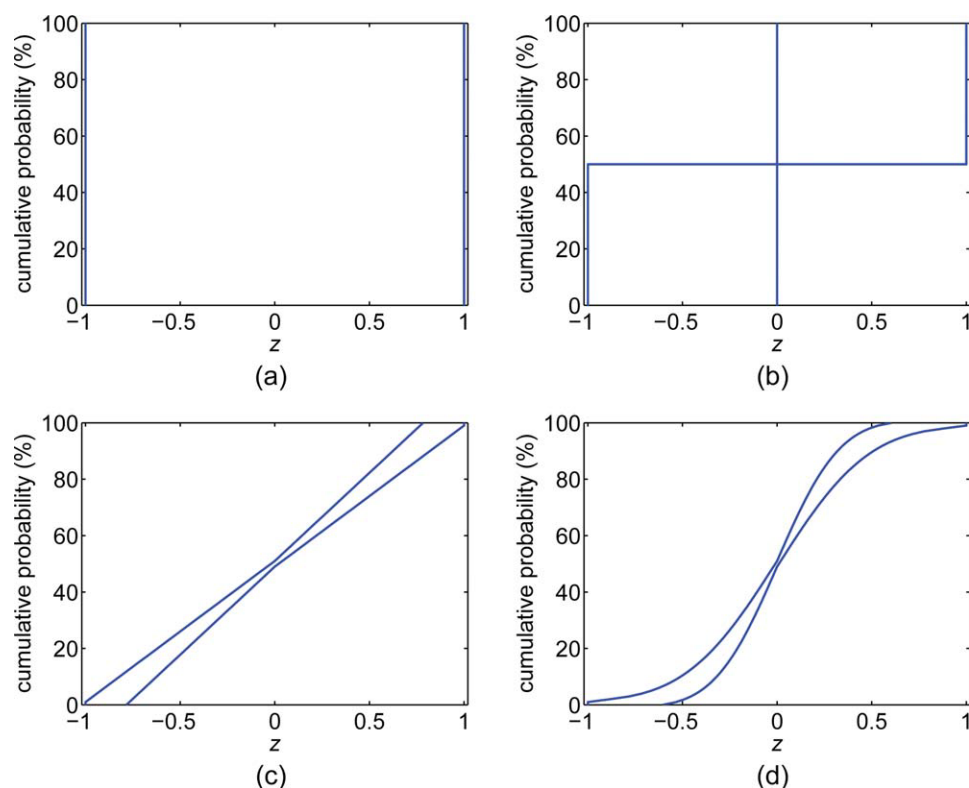

Figure 3. Examples of p-boxes.

(a) P-box of an interval (known minimum and maximum only). (b) P-box for known minimum, maximum, and median. (c) P-box with uniform bounding distributions. (d) P-box with normal bounding distributions. See text for further discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
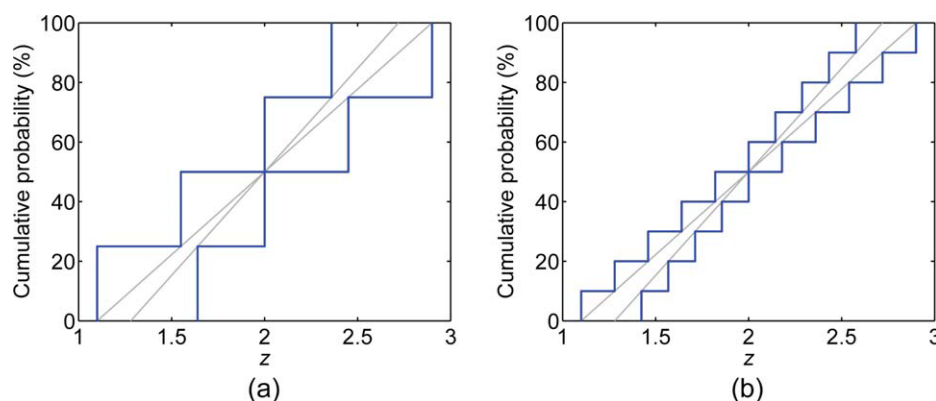
Figure 4. Discretizations of a p-box with uniform bounding distributions.

(a) $N = 4$ and (b) $N = 10$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

which here corresponds to the 99th percentile. When a p-box is formed using bounding distributions of a specified shape, as in Figures 3c, d, the p-box may still enclose distributions of other shapes. For example, the p-box in Figure 3c has uniform bounding distributions but still contains nonuniform distributions.

Williamson and Downs[7] have described methods for rigorously bounding the results of arithmetic operations on random variables when only their bounding distributions are known. This can be done without assuming any knowledge of possible correlation between the operands. It can also be done under the assumption that the operands are independent (or that they are the same, as in a polynomial or other expression with a repeated operand).

In general, these methods are implemented numerically using discretizations of the bounding distributions. To do the discretization, the p-box bounds are enclosed using an ordered set of N intervals, each representing a probability range of equal weight $1/N$. For example, if the p-box discretization is done using $N = 4$ intervals, then each one represents a quartile of probability. An example discretization with $N = 4$ is shown in Figure 4a. Here, a p-box with uniform bounding distributions (median = 2, minimum = [1.1,1.28], and maximum = [2.72,2.9]) is represented by the ordered set of intervals {[1.1,1.64],[1.55,2],[2,2.45],[2.36,2.9]}. Note that this discretization completely encloses the p-box, though with

significant overestimation due to the coarse discretization. Obviously, a tighter enclosure can be obtained using a finer discretization. Figure 4b shows the same p-box discretized with $N = 10$ intervals. For all of the computations done here, we use $N = 100$ intervals to discretize a p-box.

To demonstrate the process of arithmetic operations with p-boxes, we will consider two independent (uncorrelated) variables, x and y, with probability bounds given by the p-boxes, PB(x) and PB(y), and compute the probability bounds for the sum $x + y$, that is, the p-box PB($x + y$). PB(x) and PB(y), discretized with $N = 4$ intervals of weight $1/4$, are shown in Figure 5. For PB(x), the intervals are [1,3], [2,5], [3,5], and [4,6], and for PB(y) the intervals are [0,2], [1,3], [2,5], and [4,5]. To determine PB($x + y$), each interval in PB(x) must be added to each interval in PB(y). This results in $N^2 = 16$ intervals, all of which are equally likely with weight $1/16$, as x and y have independent distributions. To construct PB($x + y$), these 16 intervals are sorted and ordered so that both the left bounding distribution (as indicated by the interval lower bounds) and the right bounding distribution (as indicated by the interval upper bounds) are nondecreasing. The result for this example is shown in gray in Figure 6. The first four intervals in the sorted and ordered set of 16 intervals are [1,3] + [0,2] = [1,5], [1,3] + [1,3] = [2,6], [2,5] + [0,2] = [2,7], and [3,5] + [0,2] = [3,7]. This p-box provides rigorous bounds on the probability distribution for x
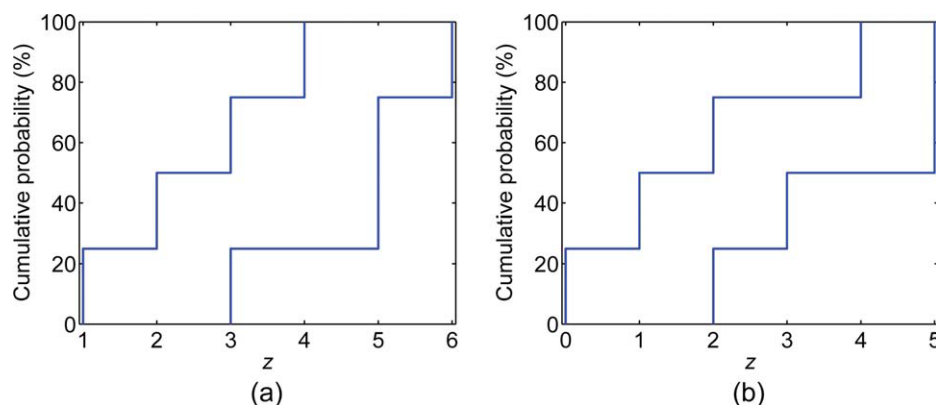

Figure 5. P-boxes used in arithmetic examples.

(a) PB(x) and (b) PB(y). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
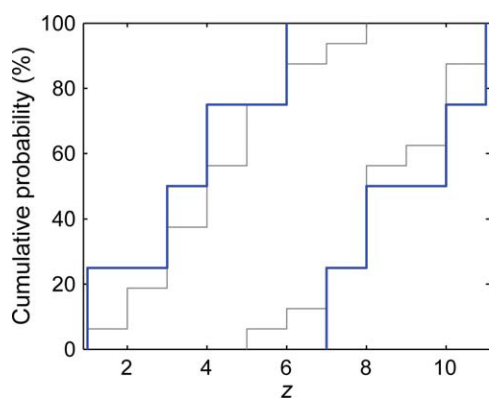
Figure 6. Intermediate (gray) and final (blue) results for PB($x + y$).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

+ y. Analogous procedures can be used to compute probability bounds for the other arithmetic operations as well as for other functions (logarithm, integral powers, etc.). Furthermore, the results of bounding the probability distribution of one operation can then be used as input for other operations. So, for example, if there is a third independent variable u with known PB(u), then PB($x + y + u$) can be computed from PB($x + y$) and PB(u). In general, it is possible to compute p-boxes PB($f(x,y,u)$) for quite complicated functions $f(x,y,u)$ from the p-boxes representing the individual variables.

Although PB(x) and PB(y) are both described by four intervals, PB($x + y$) is described by 16 intervals. To prevent continued growth in the number of intervals required for the results of an arithmetic operation, it is convenient to require that all p-box operands and results be expressed in terms of the same number of intervals N. For this example then, the 16-interval result is an intermediate one, which must then be wrapped using a four-interval p-box to obtain the final result. This final result is shown in blue in Figure 6. Obviously, this wrapping process results in overestimation, so it is desirable to use a relatively large N. As noted previously, for our computations with p-boxes we will use $N = 100$.

Because the operands are assumed to be independent, the dependency problem for computations with expressions involving repeated variables also occurs in this context. For example, using the same PB(x) as above, we have computed PB($x + x$) and PB($2x$), with the results compared in Figure 7 on the basis of $N = 4$ intervals per p-box. The probability bounds from PB($2x$) (blue) are tighter than those for PB($x + x$) (gray).

P-box operations are available for the case of independent operands as well as for the case of operands for which no assumption about dependence or independence can be made. P-box operations can be performed using the risk analysis software RAMAS Risk Calc.[24] Basic p-box arithmetic operations may also be performed using Statool.[27] For the p-box computations done in the examples presented below, we used our own basic Matlab implementation of p-box arithmetic.

Taylor Models

To alleviate the problems of dependency and wrapping that occur in using traditional interval methods, and which

lead to the overestimation of bounds, Makino and Berz[28] described a remainder differential algebra (RDA) approach for bounding the ranges of functions.[29] In this method, a function is represented using a "Taylor model," consisting of a Taylor polynomial and an interval remainder bound.

One way of forming a Taylor model of a function is by using the Taylor theorem. Consider a real function $f(x)$ that is $(q + 1)$ times partially differentiable on X and let $x_0 \in X$. The Taylor theorem states that for each $x \in X$, there exists a real ζ with $0 < \zeta < 1$ such that

$$f(x) = p_f(x - x_0) + r_f(x - x_0, \zeta), \qquad (2)$$

where p_f is a q-th order polynomial (truncated Taylor series) in $(x - x_0)$ and r_f is a remainder, which can be quantitatively bounded over $0 < \zeta < 1$ and $x \in X$ using interval arithmetic or other methods to obtain an interval remainder bound R_f. A q-th order Taylor model $T_f = p_f + R_f$ for $f(x)$ over X consists of the polynomial p_f and the interval remainder bound R_f and is denoted by $T_f = (p_f, R_f)$. Note that $f(x) \in T_f$ for $x \in X$, and thus T_f encloses the range of $f(x)$ over X.

In practice, it is more useful to compute Taylor models of functions by performing Taylor model operations. Arithmetic operations with Taylor models can be done using the RDA operations described by Makino and Berz,[28–30] which include addition, multiplication, reciprocal, and intrinsic functions. Using these operations, it is possible to start with simple functions such as the constant function $f(x) = k$, for which $T_f = (k, [0,0])$, and the identity function $f(x_i) = x_i$, for which $T_f = (x_{i0} + (x_i - x_{i0}), [0,0])$, and then to compute Taylor models for very complicated functions. This can be implemented using operator overloading, which makes it easy to compute a Taylor model for any function representable by standard arithmetic operations and elementary functions. It has been shown that, compared with other rigorous bounding methods, the Taylor model often yields sharper bounds for modest to complicated functional dependencies.[28,29,31] A discussion of the uses and limitations of Taylor models has been given by Neumaier.[31]

Solution Procedure

In this section, we describe the methods used to achieve the two goals established in the Problem Statement section.
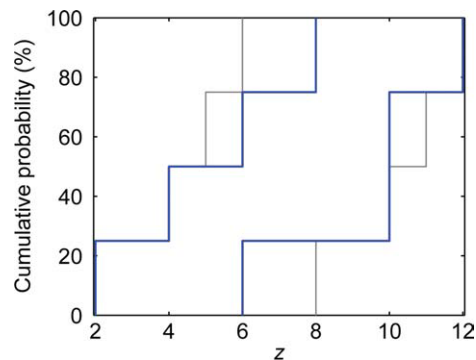

Figure 7. Results of PB($x + x$) (gray) and PB($2x$) (blue) to illustrate dependency issue.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Enclosure of state variables

The first goal is to solve the uncertain ODE system given by Eq. 1, obtaining mathematically and computationally guaranteed enclosures Y_j of the state variables $y_j = y(t_j)$ at the endpoints $t_j \in [t_0, t_f]$ of each time step in the numerical integration.

Interval methods (also called validated methods or verified methods) for ODEs provide a natural approach for computing the desired enclosure of the state variables. Traditional interval methods generally use two processes at each integration step. In the first process, existence and uniqueness of the solution are proven, and a rough enclosure of the solution is computed. In the second process, a tighter enclosure of the solution is computed. In general, both processes are implemented by applying interval Taylor series (ITS) expansions with respect to time and using automatic differentiation to obtain the Taylor coefficients. An excellent review of the traditional interval methods has been given by Nedialkov et al.,[32] and more recent work has been reviewed by Neher et al.[33] For addressing this problem, there are several packages available, including AWA,[34] VNODE,[32,35] COSY VI,[36] and ValEncIA-IVP.[37] In this study, we will use the recently developed solver VSPODE,[11] which is capable of determining guaranteed bounds on the solutions of dynamic systems with interval-valued initial states and parameters, and which offers significant performance improvements over the popular VNODE package. The method makes use, in a novel way, of the Taylor model approach[28–30] to deal with the dependency and wrapping problems involving the uncertain quantities (parameters and initial values).

Assuming an interval enclosure Y_j of the state variables at time t_j, VSPODE determines a time step $h_j = t_{j+1} - t_j$ and an enclosure Y_{j+1} of the state variables at t_{j+1}. In the first phase of the algorithm, a coarse enclosure $\widetilde{Y}_j$ is determined such that a unique solution $y(t) \in \widetilde{Y}_j$ is guaranteed to exist over the time interval $[t_j, t_{j+1}]$ for every $y_j \in Y_j$ and every $\theta \in \Theta$. This is done with a high-order ITS with respect to time, using the Picard-Lindelöf operator and Banach fixed-point theorem. The time step used can be specified, but may be reduced if necessary, or an automatic step size procedure can be used. This represents an extension, to parametric ODEs, of the traditional interval approach used in VNODE.

In the second phase of the algorithm, Taylor models in terms of the uncertain quantities are used. The uncertain initial states and parameters are expressed as Taylor model identity functions T_{y_0} and T_θ. Then, Taylor models $T_{f^{[i]}}$ of the ITS coefficients $f^{[i]}(y_j, \theta)$ are obtained by using RDA operations to compute $T_{f^{[i]}} = f^{[i]}(T_{y_j}, T_\theta)$. Using an ITS expansion for y_{j+1} with coefficients given by $T_{f^{[i]}}$, and incorporating an approach for using the mean value theorem on Taylor models, one can obtain $T_{y_{j+1}}(y_0, \theta)$, the desired Taylor model of y_{j+1} in terms of the parameters θ and initial states y_0. In this process, the wrapping effect of traditional interval methods is reduced by using a new type of Taylor model that uses a parallelepiped (as opposed to interval) remainder bound.[11] The Taylor model $T_{y_{j+1}}(y_0, \theta)$ can then be bounded[11] over $y_0 \in Y_0$ and $\theta \in \Theta$ to obtain Y_{j+1}.

Probability distribution of state variables

The second goal is to determine rigorous bounds, in the form of a p-box, on the probability distributions for the values of $y_j \in Y_j$, given p-box bounds on the probability distributions of the uncertain quantities.

Using the method summarized above, we can obtain, for a time t_j of interest, a Taylor model $T_{y_j}(y_0, \theta)$, which gives the state variables $y_j = y(t_j)$ as a polynomial function of the initial states $y_0 \in Y_0$ and the parameters $\theta \in \Theta$, plus a small remainder bound. If probability distributions (p-boxes) are available for y_0 and for θ, then these can be substituted directly into $T_{y_j}(y_0, \theta)$, and a p-box giving bounds on the probability distribution for y_j can be computed using p-box operations.

Straightforward application of p-box operations to evaluate the Taylor model $T_{y_j}(y_0, \theta)$ may lead to significant overestimation of bounds on the true probability distribution of the state variables, because of the dependency problem and the wrapping effect, as discussed in the Background section. One method to obtain a much tighter enclosure is subinterval reconstitution (SIR). In this procedure,[25] each of the N intervals in the discretized p-box used for arithmetic operations is partitioned into d subintervals. Arithmetic operations are then done on each subinterval separately. The overall results are reconstituted by taking the union of the results from each subinterval within a discretization interval. SIR is available as an option in RAMAS Risk Calc as well as in our own Matlab implementation of p-box arithmetic. SIR is only useful for cases in which the expression to be evaluated involves repeated uncertain quantities, as occurs in the polynomial part of the Taylor model representing the state variables. Unless otherwise noted, in the examples given below, we will use SIR with $d = 100$ subintervals per variable. The first example provides a comparison of results with and without use of SIR.

Examples

To demonstrate the potential of this approach for studying the effect of uncertainties with imprecise probability distributions, we will present here several examples focused on reaction process dynamics. The first two examples involve linear models, and the remainder are nonlinear. In each case, unless otherwise stated, VSPODE was used with an ITS order of $k = 17$ and a Taylor model order of $q = 5$. For the integration procedure, a constant step size $h = 0.01$ was specified for the first two examples and $h = 0.2$ for the last three examples, though the step size is automatically reduced by VSPODE if necessary. Taylor model remainder bounds in VSPODE are obtained using a QR-factorization process.[11] P-boxes are discretized using $N = 100$ intervals, and p-box arithmetic is done using SIR with $d = 100$ subintervals per variable unless otherwise noted. In all cases, we assume that the uncertain quantities are independent (uncorrelated). Computations using VSPODE and p-box arithmetic were done using an Intel Pentium 4 (3.2 GHz) machine running Red Hat Enterprise Linux (RHEL). VSPODE was implemented using C++. P-box arithmetic was implemented using Matlab, with outward rounding of interval operations done using techniques described by Lambov[38] that do not require repeated switching of rounding mode. Second-order Monte Carlo simulations used as comparisons were implemented using Matlab and run using AMD dual-core Opteron 175 (2.2 GHz) hardware running RHEL. To facilitate comparisons of computational effort,
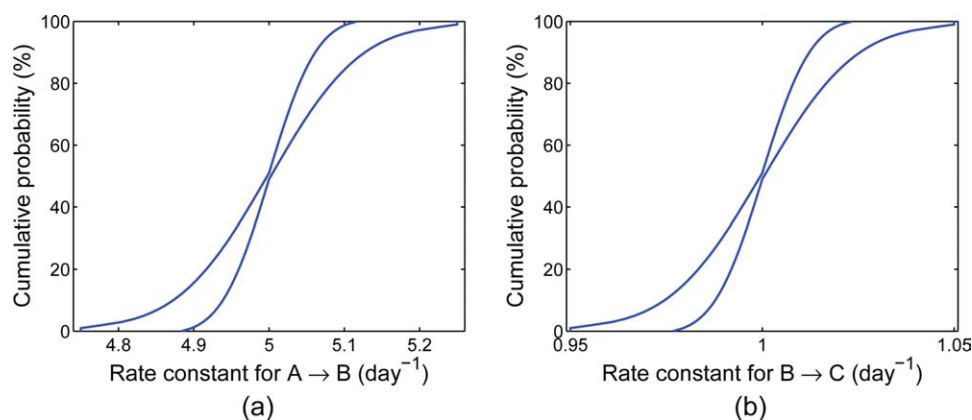
Figure 8. P-boxes used for Case 1 in irreversible series reaction example.

(a) $PB(k_1)$ and (b) $PB(k_2)$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

CPU times on the Pentium 4 (3.2 GHz) have been converted to equivalent times on the faster Opteron 175 (2.2 GHz). This was done based on the PassMark CPU Benchmark, on which the performance scores are 1022 for the Opteron 175 (2.2 GHz) and 512 for the Pentium 4 (3.2 GHz). Thus, CPU times on the Pentium 4 (3.2 GHz) have been multiplied by a factor of $512/1022 \approx 0.5$ (we have also confirmed this performance ratio with our own direct comparisons).

First-order irreversible series reaction

In this example, we consider the irreversible series reaction

$$ A \xrightarrow{k_1} B \xrightarrow{k_2} C $$
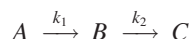
occurring in a batch reaction process. Each individual reaction is first order in the concentration of reactant. However, the rate constants k_1 and k_2 are uncertain. The balance equations describing the concentrations of A and B are

$$ \frac{dC_A}{dt} = -k_1 C_A \qquad (3) $$
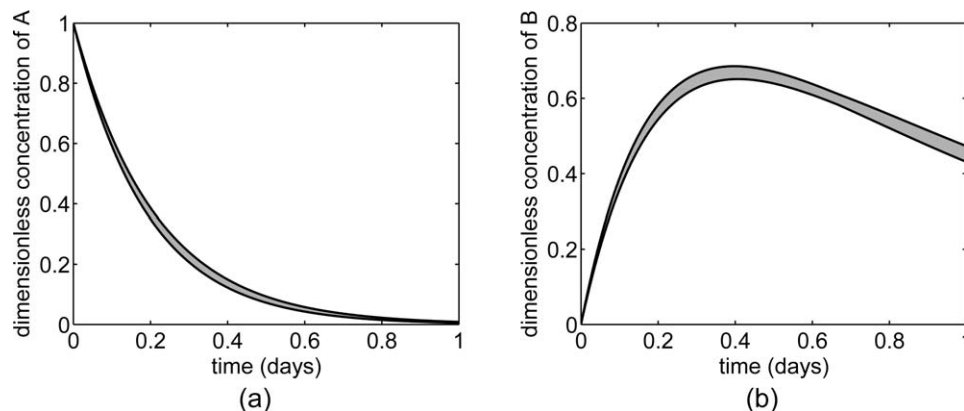
$$ \frac{dC_B}{dt} = k_1 C_A - k_2 C_B. \qquad (4) $$

The concentrations C_A and C_B are treated as dimensionless (relative to the constant total number of moles). Initially, the reactor contains only species A; that is, $C_A = 1$ and $C_B = 0$ at time $t = 0$. The time horizon of interest is $t_f = 1$ day. We will consider three cases for this linear ODE model, with increasing levels of uncertainty in the rate constants, which have mean values of $k_1 = 5$ day^{-1} and $k_2 = 1$ day^{-1}. For each case, the goals are to bound the possible concentration profiles $C_A(t)$ and $C_B(t)$ and to bound the probability distribution of C_B values at $t = t_f = 1$ day.

Case 1 has 5% uncertainty in the rate constants: $k_1 \in [4.75, 5.25]$ day^{-1} and $k_2 \in [0.95, 1.05]$ day^{-1}. The distribution of the uncertainties is not known precisely, but can be enclosed by p-boxes whose bounding distributions are obtained from a truncated (99th percentile) normal distribution of known mean and uncertain standard deviation. For k_1 the standard deviation is in the interval $[0.05, 0.095]$ day^{-1}, and for k_2 the standard deviation is in the interval $[0.01, 0.019]$ day^{-1}. These p-boxes are illustrated in Figure 8.

VSPODE was used to determine rigorous bounds on the trajectories $C_A(t)$ and $C_B(t)$ over the time horizon of interest. These results are shown as the black curves in Figure 9, which represent mathematically and computationally guaranteed bounds on the possible trajectories. As interval methods

Figure 9. Trajectories of C_A and C_B for irreversible series reaction example with Case 1 parameters.

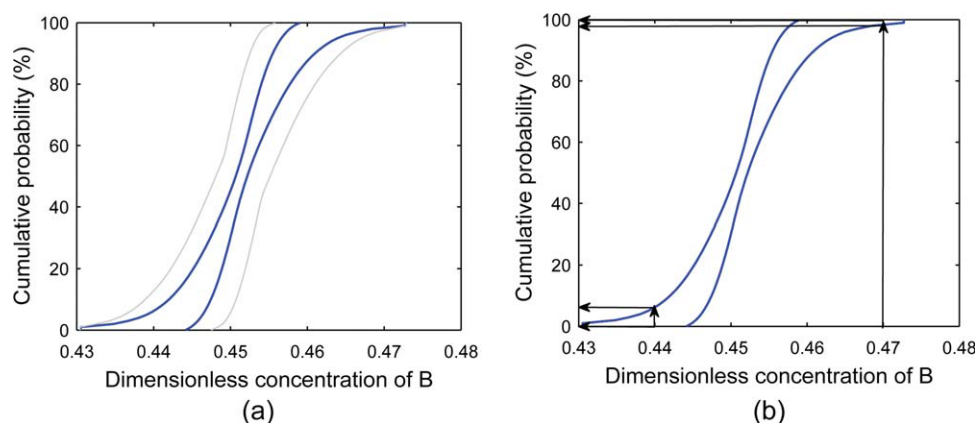VSPODE bounds are in black and Monte Carlo simulation results (200 samples) are in gray.

Figure 10. Probability distribution bounds for $C_B(1)$ for irreversible series reaction example with Case 1 parameters.

(a) Computed PB($C_B(1)$) when using SIR (blue) and when not using SIR (gray). (b) Determining probability bounds for $C_B(1) \leq 0.44$ and $C_B(1) \leq 0.47$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

may produce loose bounds, we checked the tightness of the VSPODE bounds by comparison to the results of a Monte Carlo simulation with 200 trials. For each trial, real values of k_1 and k_2 were selected at random from within their specified interval bounds. Bounds obtained from Monte Carlo analysis are not guaranteed and in general yield an inner estimate of the true bounds (the guaranteed VSPODE bounds represent an outer estimate). The Monte Carlo simulation results are shown by the shaded areas in Figure 9. On the scale of these figures, there is no apparent gap between the VSPODE bounds and the Monte Carlo simulation results, indicating that VSPODE provides very tight bounds on the possible concentration trajectories for this system. Of course, for this simple linear system, an analytic solution can be obtained and used for bounding. In general, however, no analytic solution is available for the problems of interest.

The Taylor model from VSPODE at $t_f = 1$ day then was used to compute bounds on the probability distribution for $C_B(1)$. This was done both with and without the use of the SIR procedure for tightening the bounds. The results are shown as p-boxes in Figure 10a, with the p-box obtained with SIR in blue and the p-box obtained without SIR in gray. Clearly, it is possible to obtain tighter bounds using

SIR, and this procedure is used in all of the remaining examples. The SIR p-box indicates, for example, as shown in Figure 10b, that the probability of $C_B(1) \leq 0.44$ is in the interval $[0,7]\%$ and that the probability of $C_B(1) \leq 0.47$ is in $[98,100]\%$. The CPU time necessary to compute p-box bounds without SIR is about 2 s, and with SIR the time necessary is about 13 s. The computation time does not increase by a factor equal to the number of SIR subintervals (100 $\times$ 100). P-box computations without SIR require use of a sort function with every function call, but when SIR is used, there are some savings in the amount of sorting work that can be achieved.

For Case 2, there is 10% uncertainty in the rate constants: $k_1 \in [4.5,5.5]$ day^{-1} and $k_2 \in [0.9,1.1]$ day^{-1}. As in Case 1, the distribution of the uncertainties can be bounded by p-boxes whose bounding distributions are obtained from a truncated (99th percentile) normal distribution of known mean and uncertain standard deviation. For k_1 the standard deviation is in the interval $[0.1,0.19]$ day^{-1}, and for k_2 the standard deviation is in the interval $[0.02,0.038]$ day^{-1}. VSPODE bounds on the concentration trajectories, along with comparison to Monte Carlo simulation results, are given in Figure 11. Again, the bounds determined using VSPODE
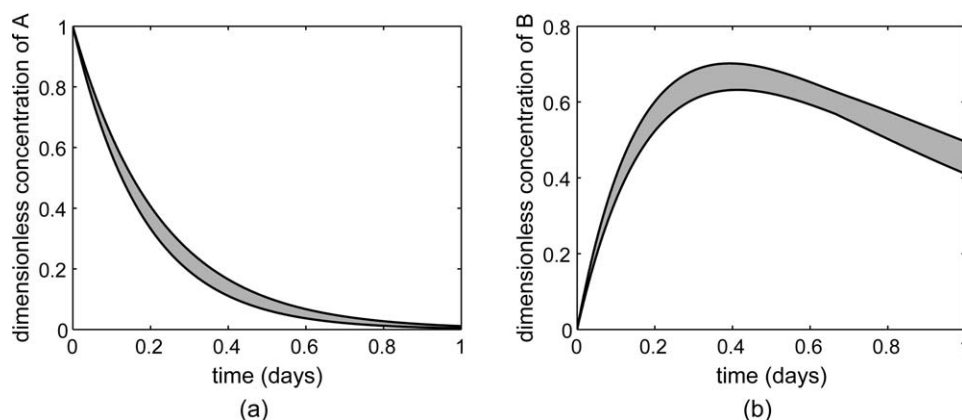

Figure 11. Trajectories of C_A and C_B for irreversible series reaction example with Case 2 parameters.

VSPODE bounds are in black and Monte Carlo simulation results (200 samples) are in gray.
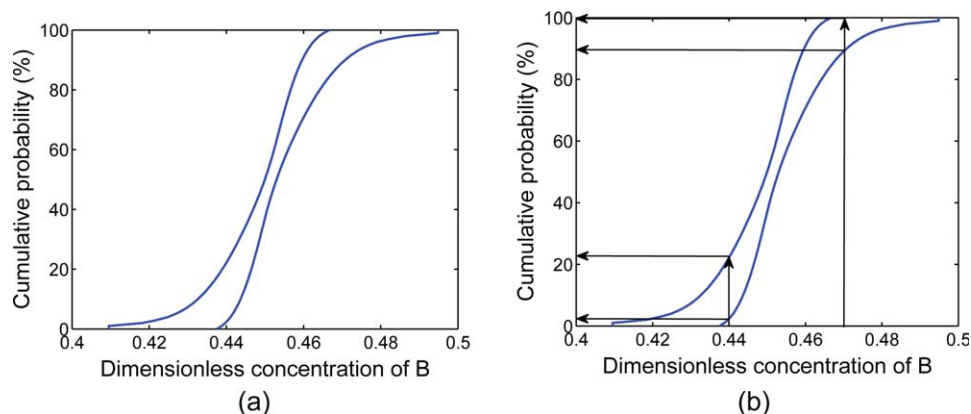
Figure 12. Probability distribution bounds for $C_B(1)$ for irreversible series reaction example with Case 2 parameters.

(a) Computed PB($C_B(1)$). (b) Determining probability bounds for $C_B(1) \leq 0.44$ and $C_B(1) \leq 0.47$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

appear to be very tight. The probability bounds for $C_B(1)$ are shown in Figure 12a. Now, as shown by Figure 12b, the probability that $C_B(1) \leq 0.44$ is in the interval [2,23]%, and the probability that $C_B(1) \leq 0.47$ is in [89,100]%. The range of standard deviations for this case is outside that of the first case, so it is not necessarily expected that these probability bound results enclose those from the first case.

Finally, for Case 3, we consider 80% uncertainty in k_2 so that it lies in the interval [0.2,1.8] day^{-1} with a standard deviation in [0.160,0.304], whereas k_1 remains at 10% uncertainty as described in the previous case. As shown in Figure 13, VSPODE again provides very tight bounds on the concentration trajectories. For this case, as well as for the first two cases, continued integration with VSPODE past $t_f = 1$ day tightly bounds the trajectories as they approach the ultimate steady state, where $C_A = C_B = 0$. The probability bounds shown in Figure 14 now indicate that the probability that $C_B(1) \leq 0.44$ is in the interval [39,48]%, and the probability that $C_B(1) \leq 0.47$ is in [55,65]%.

First-order reversible series reaction

In this example, we consider a batch reaction system similar to that in the previous subsection, but now with reversible reactions:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} B \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C.$$
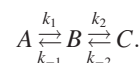
We assume that the second reaction has been well studied and that there is no uncertainty in these rate constants, which are $k_2 = 40$ day^{-1} and $k_{-2} = 20$ day^{-1}. However, for the first reaction, we assume that there are very large uncertainties in the rate constants, which are bounded by $k_1 \in [2,6]$ day^{-1} and $k_{-1} \in [1,3]$ day^{-1}. The distribution of the uncertainties is not known precisely but can be enclosed by p-boxes whose bounding distributions are obtained from a uniform distribution of fixed median and uncertain minimum and maximum. For k_1 the median is fixed at 4 day^{-1}, the minimum is in the interval [2,2.4], and the maximum in [5.6,6]. For k_{-1} the median is fixed at 2 day^{-1}, the minimum is in the interval [1,1.2], and the maximum in [2.8,3]. These p-boxes are illustrated in Figure 15.

The balance equations describing the dimensionless concentrations (relative to the constant total number of moles) of components A and B are

$$\frac{dC_A}{dt} = -k_1 C_A + k_{-1} C_B, \tag{5}$$

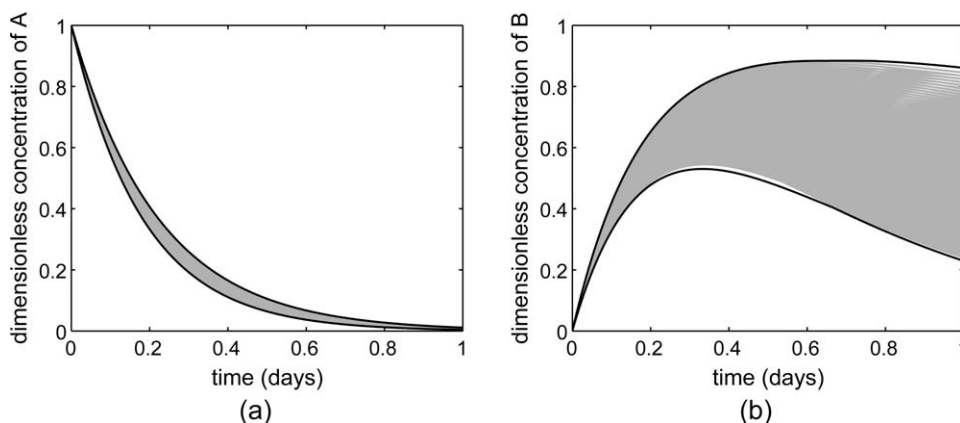$$\frac{dC_B}{dt} = k_1 C_A - (k_{-1} + k_2) C_B + k_{-2}(1 - C_A - C_B). \tag{6}$$

Figure 13. Trajectories of C_A and C_B for irreversible series reaction example with Case 3 parameters.

VSPODE bounds are in black and Monte Carlo simulation results (2000 samples) are in gray.
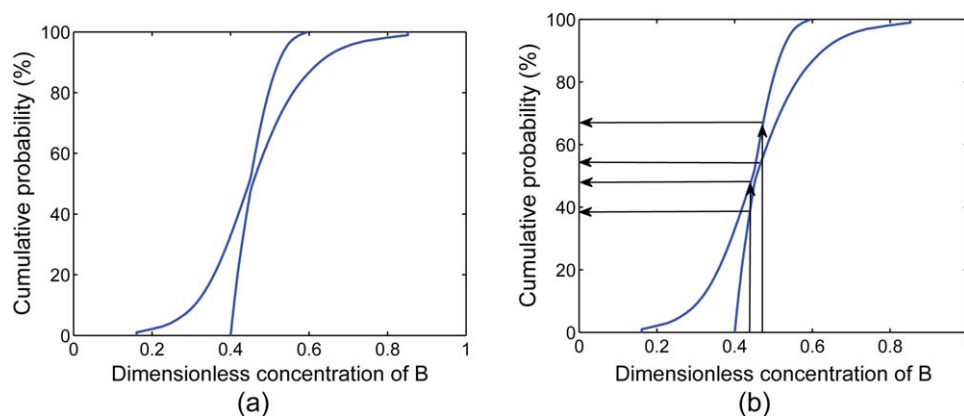
Figure 14. Probability distribution bounds for $C_B(1)$ for irreversible series reaction example with Case 3 parameters.

(a) Computed PB($C_B(1)$). (b) Determining probability bounds for $C_B(1) \leq 0.44$ and $C_B(1) \leq 0.47$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Initially, the reactor contains only species A; that is, $C_A(0) = 1$ and $C_B(0) = 0$. The time horizon of interest is $t_f = 0.5$ day. The goals are to bound the possible concentration trajectories $C_A(t)$ and $C_B(t)$ out to the time horizon and to bound the probability distribution of C_A and C_B values at a point halfway ($t = 0.25$ day) to the time horizon.

VSPODE was applied to these equations to bound the concentration trajectories, with the results as shown by the black curves in Figure 16. These curves represent rigorously guaranteed bounds on the possible trajectories. For comparison, a Monte Carlo analysis with 2000 samples was done, and the results of which are shown by the shaded areas in Figure 16. It can be seen that, up to $t = t_f = 0.5$ day, the bounds from VSPODE are quite good. We also integrated with VSPODE past $t_f = 0.5$ day, and the quality of the computed bounds gradually decreased as time increased, when compared with the known analytic solution. This was not surprising given the very large intervals of uncertainty in this problem. If integration continues beyond $t = 1$ day, the bounds computed by VSPODE will ultimately enclose physically impossible solutions and thus become unusable. To resolve this situation, it would be necessary to divide the large intervals of uncertainty into multiple smaller subintervals, then integrate with VSPODE to determine bounds corresponding to each subinterval, and then combine the results.

The Taylor model from VSPODE at $t = 0.25$ day was used to compute bounds on the probability distributions for values of $C_A(0.25)$ and $C_B(0.25)$. The resulting p-boxes are shown in Figure 17. This computation required a total of about 13 s of CPU time (of which only about 0.1 s was spent in VSPODE).

For comparison, we also did a second-order Monte Carlo analysis to obtain probability distributions for $C_A(0.25)$ and $C_B(0.25)$. In the outer loop of this procedure, uniform distributions were chosen at random from the set of distributions enclosed by PB(k_1) and PB(k_{-1}). Then in the inner loop, the k_1 and k_{-1} intervals were repeatedly sampled based on the probability distribution chosen in the outer loop, and for each sample the ODE system was solved. CDFs for $C_A(0.25)$ and $C_B(0.25)$ were then constructed using the inner loop samples. Such CDFs result for each outer loop used. The concentration p-boxes PB($C_A(0.25)$) and PB($C_B(0.25)$) computed using the VSPODE Taylor model (Figure 17) are rigorous bounds on the CDFs for $C_A(0.25)$ and $C_B(0.25)$, respectively, provided that these CDFs are themselves rigorous, which in principle would require an infinite number of inner loop samples. In practice, this means that CDFs determined from an insufficient number of inner loop samples are not necessarily enclosed by the rigorous p-box bounds.

To determine an appropriate basis for comparison, we performed second-order Monte Carlo simulations using varying
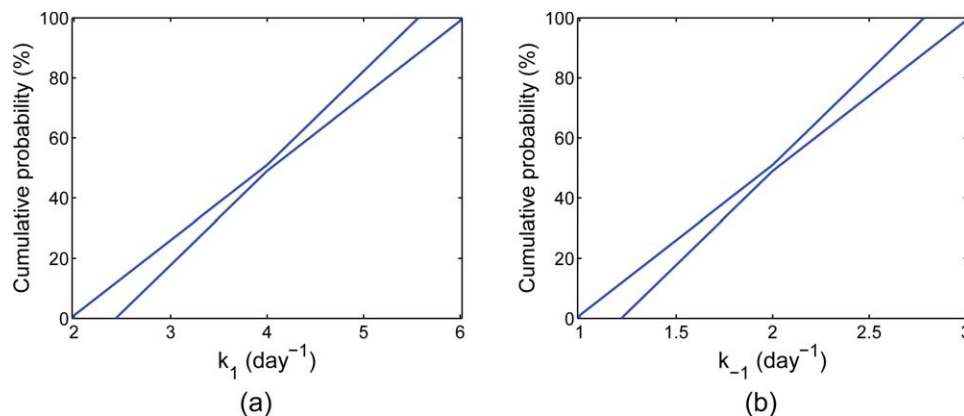

Figure 15. P-boxes used for reversible series reaction example.

(a) PB(k_1) and (b) PB(k_{-1}). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
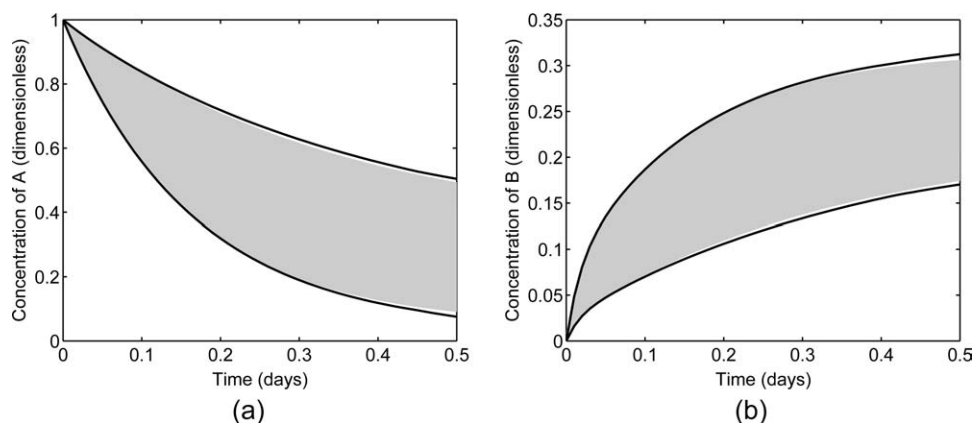
Figure 16. Trajectories of C_A and C_B for reversible series reaction example.

VSPODE bounds are in black and Monte Carlo simulation results (2000 samples) are in gray.

numbers of outer and inner loops, with results characterized by the cases presented in Figure 18. For the case of 1000 outer loops and 1000 inner loops, as shown in Figures 18a, b, a careful comparison to the rigorous probability distribution bounds determined using the VSPODE Taylor models, as shown in Figure 17, indicates that *not all* of the CDFs obtained from the Monte Carlo analysis lie within these rigorous bounds. This implies that 1000 is an insufficient number of inner loop samples. The computational time for this second-order Monte Carlo simulation was about 74.5 min (versus about 13 s to compute the rigorous bounds using the Taylor model method described here). If the number of inner loop samples is increased to 10,000 and the number of outer loops reduced to 100, to keep the computational effort the same as in the previous case, then the results are as shown in Figures 18c, d. Now, all the CDFs lie within the rigorous bounds of Figure 17, and tighter bundles of CDFs are obtained. To determine whether 100 is an adequate number of outer loops, we also considered the case of 1000 outer loops (and again 10,000 inner loops). This computation required about 745 min, with the results as shown in Figures 18e, f. These bundles of CDFs are only very slightly wider than those obtained using 100 outer loops (Figures 18c, d)

and still are within the rigorous bounds of Figure 17. In subsequent examples for which second-order Monte Carlo analysis is done to provide a comparison, we will use 100 outer loops and 10,000 inner loops, as this appears to provide a reasonable balance between accuracy and computational effort.

A shortcoming of the second-order Monte Carlo approach used here for comparison, besides its long execution time, is that only uniform distributions are sampled in the outer loop. The p-box bounds obtained using the Taylor model approach enclose all CDFs from any shape distribution enclosed by the input p-boxes $PB(k_1)$ and $PB(k_{-1})$.

Exothermic batch reactor

In this example, we model the conversion X and temperature T of a first-order exothermic reaction A $\rightarrow$ B taking place in a batch reactor with cooling jacket. The result is a nonlinear ODE model for which we will assume some relatively large uncertainties. The model is obtained from material and energy balances on the system:

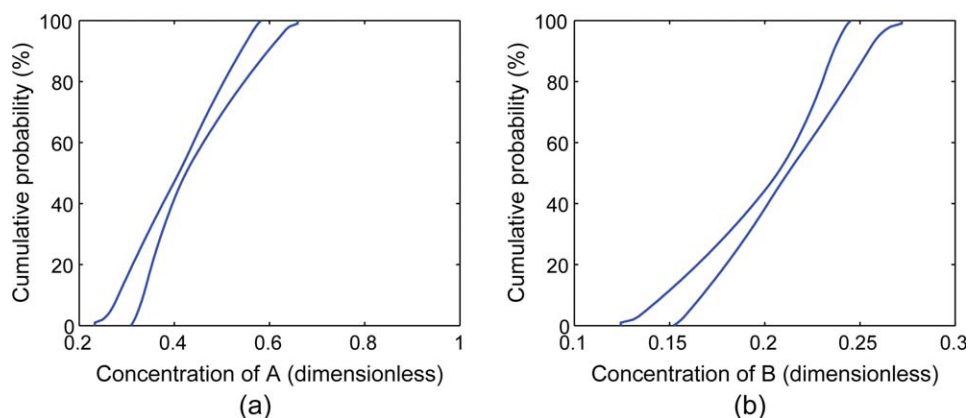$$\frac{dX}{dt} = k_0 \exp\left(-\frac{E_a}{RT}\right)(1 - X) \qquad (7)$$

Figure 17. Probability distribution bounds (based on all distribution shapes for k_1 and k_{-1}) computed using VSPODE Taylor models for dimensionless concentrations of A and B at t = 0.25 day.

(a) $PB(C_A(0.25))$ and (b) $PB(C_B(0.25))$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
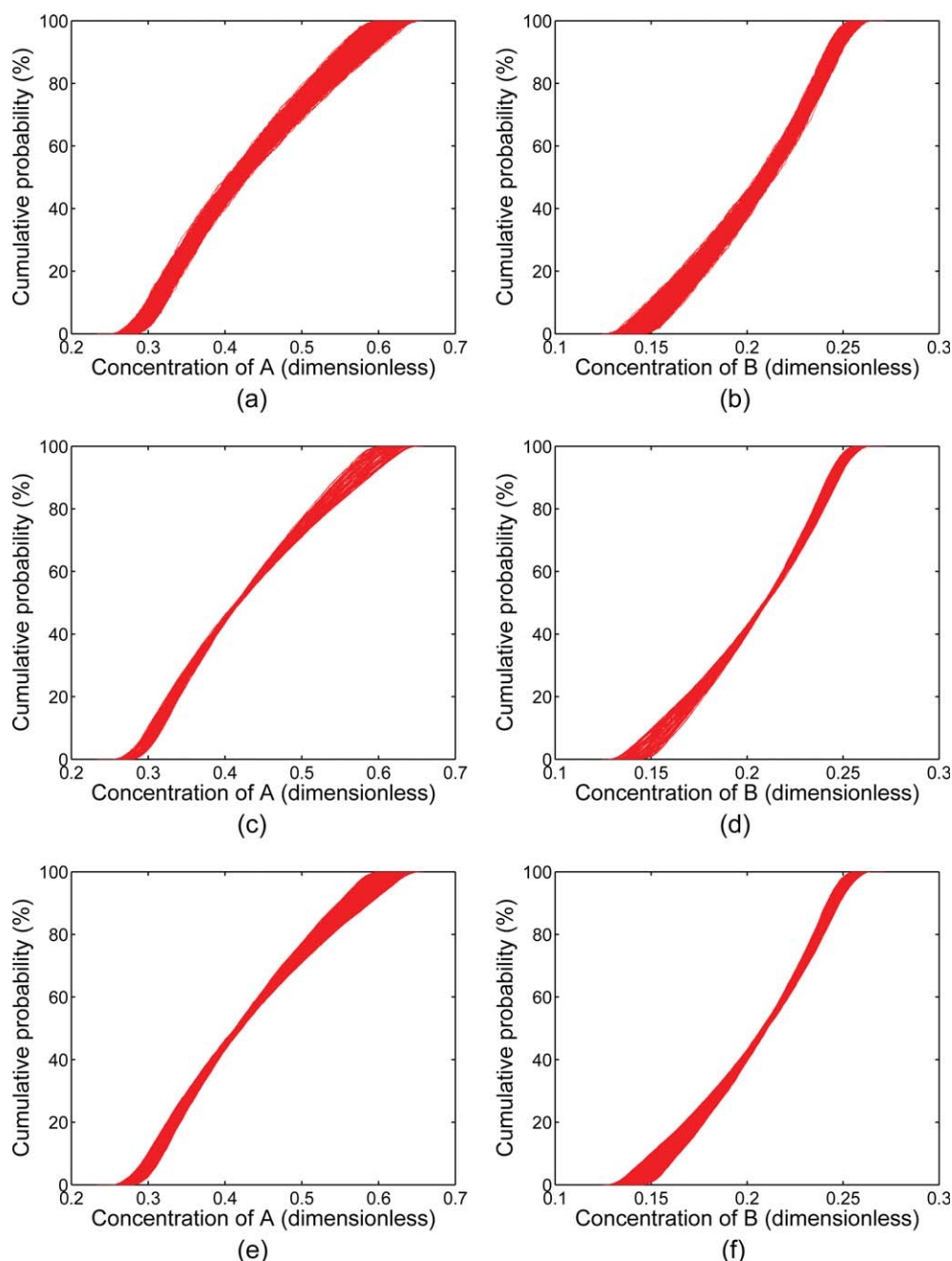
Figure 18. Probability distributions of C_A and C_B at t = 0.25 day for reversible reaction model using second-order Monte Carlo analysis (based only on uniform distributions for k_1 and k_{-1}).

(a, b) 1000 outer loop samples, each with 1000 inner loop samples. (c, d) 100 outer loop samples, each with 10,000 inner loop samples. (e, f) 1000 outer loop samples, each with 10,000 inner loop samples. See text for discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\frac{dT}{dt} = \frac{UA}{C_{A0}VC_p}(T_a - T) - \frac{\Delta H_R k_0}{C_p}\exp\left(-\frac{E_a}{RT}\right)(1 - X). \quad (8)$$

Here, the uncertain quantities are the initial reactor temperature $T(0) = T_0$ and the cooling jacket temperature T_a. Other model parameters have fixed values and are described in Table 1. The distribution of the uncertainties in T_0 and T_a is not known exactly but can by bounded by p-boxes with bounding distributions obtained from a truncated (99th percentile) normal distribution with fixed mean and uncertain

standard deviation. For T_0 the mean is 360 K, and the standard deviation is in the interval [10,19] K. For T_a the mean is 300 K, and the standard deviation is in the interval [2,3.8] K. Based on these p-boxes, $T_0 \in$ [310,410] K, and $T_a \in$ [290,310] K. The time horizon of interest is $t_f = 60$ s. The goals are to bound the trajectories $X(t)$ and $T(t)$ out to this time horizon and then to bound the probability distributions of X and T at this time.

Figure 19 shows the results of modeling this reactor from $t = 0$ to $t = 60$ s using VSPODE; these rigorously guaranteed

Table 1. Parameter Values and Initial States for the Exothermal Batch Reactor Example

Parameter	Description	Value
k_0	Reaction rate constant	0.022 s^{-1}
C_{A0}	Initial concentration of A in reactor	10 mol/m^3
V	Reactor volume	0.1 m^3
C_p	Heat capacity of reactor contents	60 J/mol K
E_a	Activation energy for reaction	6000 J/mol
R	Ideal gas constant	8.314 J/mol K
ΔH_R	Heat of reaction	$-140,000$ J/mol
UA	Product of overall heat transfer coefficient and area	3 W/K
T_a	Cooling jacket temperature	[290,310] K
T_0	Initial reactor temperature	[310,410] K
X_0	Initial conversion	0

bounds on the possible trajectories are given as black curves. Also shown in Figure 19 are the results of a Monte Carlo analysis with 200 samples, shown by the shaded areas. It can be seen that, up to the time horizon of $t_f = 60$ s, the bounds from VSPODE are again quite good. Integration with VSPODE can continue beyond the time horizon to a simulation time of about 270 s before the bounds decrease in quality.

The Taylor model supplied by VSPODE at $t_f = 60$ s was used to compute bounds on the probability distributions for values of $X(60)$ and $T(60)$. The resulting p-boxes are shown by the solid curves (blue) in Figure 20. This computation required a total of about 13 s of CPU time (only about 0.1 s of this was required by VSPODE). This shows, for example, that the probability that the conversion in the reactor is at least 0.22 is in the interval [30,42]%, or that the probability that the temperature in the reactor is at least 470 K is in the interval [32,46]%.

For comparison, we also performed a second-order Monte Carlo analysis to obtain probability distributions for $X(60)$ and $T(60)$. This procedure was the same as described in the previous subsection, except that in the outer loop of the procedure, truncated normal distributions were chosen at random from the set of distributions enclosed by the p-boxes $PB(T_0)$ and $PB(T_a)$. Based on the simulations discussed in the previous subsection, we used 100 outer loops and 10,000 inner loops in the second-order Monte Carlo analysis. The computation time was about 66.5 min. The results are shown

by the shaded areas (red) in Figure 20. It can be seen that these results are consistent with the p-boxes obtained using the Taylor model approach. The Monte Carlo results and Taylor model results are not directly comparable, because in the Monte Carlo analysis only normal distributions were sampled from the input p-boxes, whereas in the Taylor model case distributions of all possible shapes within the input p-boxes are accounted for.

Microbial growth model with Haldane kinetics

In this example, we consider another nonlinear reactor model with uncertain parameters and aim to track the evolution with time of the probability bounds for the state variables. The model equations for microbial growth in a simple bioreactor[1,39] are

$$\frac{dX}{dt} = (\mu - \alpha D)X \tag{9}$$

$$\frac{dS}{dt} = D(S_f - S) - k\mu X, \tag{10}$$

where X represents the concentration in kg/m^3 of biomass in the system, and S represents the concentration of substrate, also in kg/m^3. The growth rate of cells, μ, is dependent on S and may take a variety of forms. In this example, we consider Haldane kinetics,[39] for which

$$\mu = \frac{\mu_{max}S}{K_S + S + K_I S^2}. \tag{11}$$

The uncertain quantities in this example are the initial concentration of cells $X_0 \in [0.8, 0.85]$ kg/m^3 and the maximum growth rate $\mu_{max} \in [1.1, 1.2]$ day^{-1}. These quantities have probability distributions that are enclosed by p-boxes with bounding distributions obtained from a truncated (99th percentile) normal distribution with known mean and uncertain standard deviation. The p-box for X_0 is based on a mean of 0.825 and standard deviation in [0.005,0.0095] kg/m^3, and the p-box for μ_{max} is based on a mean of 1.15 and standard deviation of [0.01,0.019] day^{-1}. The model parameters and initial states are fixed and are given in Table 2. For this
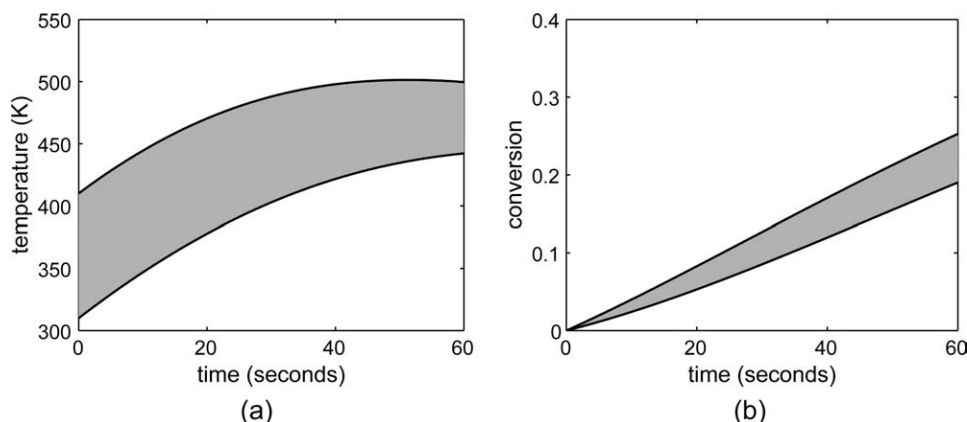

Figure 19. Trajectories of reactor temperature *T* and conversion *X* for exothermic batch reactor example.

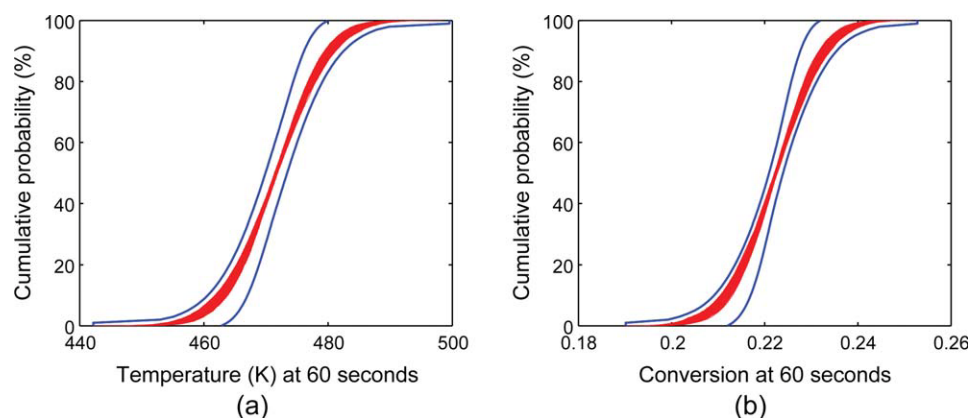VSPODE bounds in black and Monte Carlo simulation (200 samples) results in gray.

Figure 20. Solid curves (blue) show probability distribution bounds (based on all distribution shapes for T_0 and T_a) computed using VSPODE Taylor models for reactor temperature and conversion at $t = 60$ s.

(a) PB($T(60)$) and (b) PB($X(60)$). Shaded areas (red) show probability distributions for $T(60)$ and $X(60)$ as determined using second-order Monte Carlo analysis (based only on normal distributions for T_0 and T_a; 100 outer loop samples, each with 10,000 inner loop samples). See text for discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

example, we want to determine bounds on the cell concentration $X(t)$ up to a final time of $t_f = 10$ days and bounds on the probability distribution for $X(t)$ every 2.5 days.

VSPODE was applied to bound the cell and substrate concentration trajectories from $t = 0$ to $t = 10$ days, with the results as shown by the black curves in Figure 21. These curves represent rigorously guaranteed bounds on the possible trajectories. For comparison, a Monte Carlo analysis with 200 samples was done, and the results of which are shown by the shaded areas in Figure 21. It can be seen that, up to the time horizon of $t_f = 10$ days, the bounds from VSPODE are again quite good. Integration can be continued in VSPODE beyond the time horizon to a simulation time of 33 days before the bounds decrease in quality.

The Taylor model supplied by VSPODE at each time of interest was used to compute bounds on the probability distributions for values of $X(2.5)$, $X(5)$, $X(7.5)$, and $X(10)$. The resulting p-boxes are shown by the solid curves (blue) in Figure 22. This computation required a total of about 51 s of CPU time, representing about 12.75 s each for the SIR computations for the four times of interest (again only a fraction of a second was required by VSPODE). We can observe, for example, that the probability that the concentration of cells in the reactor is greater than 0.81 g/L is in the interval [77,90]% at $t = 2.5$ days, in [50,63]% at $t = 5$ days, in [14,24]% at $t = 7.5$ days, and in [0,7]% at $t = 10$ days.

For comparison, we also performed a second-order Monte Carlo analysis to obtain probability distributions for these

concentrations, using the same nested-loop procedure described in the previous subsection, again with 100 outer loops and 10,000 inner loops. The computation time for the second-order Monte Carlo simulation was about 62 min. The results obtained are shown by the shaded areas (red) in Figure 22 and are consistent with the p-boxes computed from the Taylor model method. However, the Monte Carlo and Taylor model results are not directly comparable. The Taylor model approach accounts for all possible distribution shapes within the input p-boxes, whereas the Monte Carlo approach used accounts for only a subset of the distribution shapes, namely truncated normal distributions.

Three-state bioreactor model

The final example involves another nonlinear bioreactor model, but one for which we will consider a larger number of uncertain parameters than in the previous example. We consider a bioreactor[1] in which the consumption of substrate (concentration x_2) promotes the growth of biomass (concentration x_1) and the formation of a product (concentration x_3). This process can be modeled by

$$\frac{dx_1}{dt} = (\mu - D)x_1 \tag{12}$$

$$\frac{dx_2}{dt} = D(x_{2f} - x_2) - \frac{\mu x_1}{Y} \tag{13}$$

$$\frac{dx_3}{dt} = -Dx_3 + (\alpha\mu + \beta)x_1 \tag{14}$$

with the growth rate μ given by

$$\mu = \frac{\mu_{max}[1 - (x_3/x_{3m})]x_2}{k_s + x_2}. \tag{15}$$

In this example, the uncertain quantities are the initial concentration of cells $x_{10} \in [6.45, 6.55]$ kg/m^3, the maximum growth rate $\mu_{max} \in [0.46, 0.47]$ h^{-1}, and the saturation parameter $k_s \in [1.05, 1.1]$ kg/m^3. All other parameter and initial state values are assumed to be known precisely; these are listed in Table 3. The distribution of the uncertainties is not

Table 2. Parameter Values for the Microbial Growth Model

| Parameter	Description	Value
α	Biomass resistance to washout	0.5
D	Dilution rate	0.36 day^{-1}
S_f	Feed concentration of substrate	5.7 kg/m^3
k	Yield coefficient	10.53 kg substrate/kg cells
K_S	Saturation constant	7.0 kg/m^3
K_I	Inhibition parameter	0.005 m^3/kg
μ_{max}	Maximum growth rate	[1.1,1.2] day^{-1}
X_0	Initial concentration of cells	[0.8,0.85] kg/m^3
S_0	Initial concentration of substrate	0.8 kg/m^3
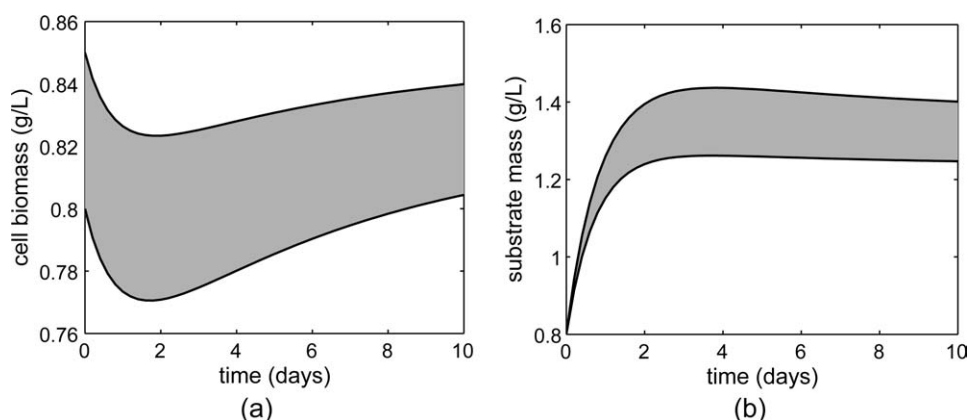
Figure 21. Trajectories of cell concentration *X* and substrate concentration *S* for bioreactor example with Haldane kinetics.

VSPODE bounds in black and Monte Carlo simulation (200 samples) results in gray.

known precisely, but can be enclosed by p-boxes whose bounding distributions are obtained from a uniform distribution of fixed median and uncertain minimum and maximum. For x_{10} the median is fixed at 6.50 kg/m^3, the minimum is in the interval [6.45,6.46] kg/m^3, and the maximum in [6.54,6.55] kg/m^3. For μ_{max} the median is fixed at 0.465 h^{-1}, the minimum is in the interval [0.460,0.461] h^{-1}, and the maximum in [0.469,0.470] h^{-1}. For k_s, the median is fixed at 1.075 kg/m^3, the minimum is in the interval [1.05,1.055] kg/m^3, and the maximum in [1.095,1.1] kg/m^3.

This problem has three uncertain quantities. For a Taylor model of order $q = 5$, this means that the number of polynomial terms increases from 21 (as in the previous examples with two uncertain quantities) to 56. (The number of terms in a polynomial of order q in m variables is $(q + m)!/$ $(q!m!)$.) Although VSPODE can still efficiently provide fifth-order Taylor models of the state variables in terms of this larger number of uncertain quantities, the increased number of terms leads to an increase in the amount of work (p-box arithmetic) needed to compute the p-boxes of the state variables. Furthermore, the increased number of uncertain quantities will also increase the work required to implement SIR, which goes as d^m, where d is the number of SIR subintervals used for each variable, and m is the number of uncertain quantities. Thus, for this problem, we have considered the effect of using either a smaller value of q or a smaller value of d. Specifically, we have considered the case of $d = 100$ with $q = 3$ (this reduces the number of Taylor model polynomial terms to 20) and the case of $d = 50$ with $q = 5$ (the previous examples were done using $d = 100$ with $q = 5$).

Figure 23 shows the results of modeling this reactor from 0 to 10 h using VSPODE with $q = 3$; these rigorously guaranteed bounds on the possible trajectories are given as black lines. Also shown in Figure 23 are the results of a Monte Carlo analysis of 200 samples, as shown by the shaded areas. It can be seen that as the time horizon of 10 h is approached, the bounds determined from VSPODE appear to become a bit loose. This is seen especially in Figure 23a, the cell biomass trajectory $x_1(t)$. When $q = 5$ was used (not shown), tightness of bounds was restored.
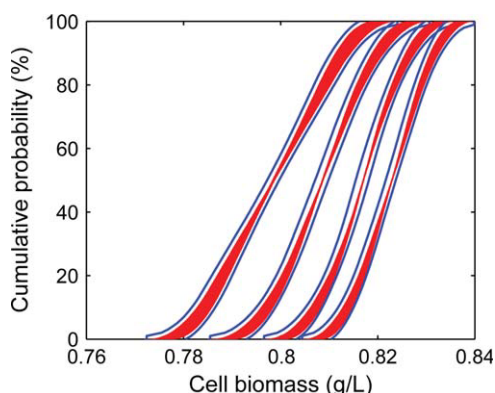

Figure 22. Solid curves (blue) show probability distribution bounds (based on all distribution shapes for X_0 and μ_{max}) for *X* computed using VSPODE Taylor models at times (from left to right) of $t = 2.5$, $t = 5$, $t = 7.5$, and $t = 10$ days.

Shaded areas (red) show probability distribution for X at the same times as computed using second-order Monte Carlo analysis (based only on normal distributions for X_0 and μ_{max}; 100 outer loop samples, each with 10,000 inner loop samples). See text for discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 3. Parameter Values for the Three-State Bioreactor Model

| Parameter	Description	Value
D	Dilution rate	0.202 h^{-1}
x_{2f}	Feed concentration of substrate	20 kg/m^3
Y	A yield parameter	0.4 kg cells/kg substrate
α	A yield parameter	0.5 kg product/kg cells
β	A yield parameter	0.2 h^{-1}
x_{3m}	Maximum product concentration	50 kg/m^3
μ_{max}	Maximum growth rate	[0.46,0.47] h^{-1}
k_s	Saturation parameter	[1.05,1.1] kg/m^3
x_{10}	Initial concentration of cells	[6.45,6.55] kg/m^3
x_{20}	Initial concentration of substrate	5 kg/m^3
x_{30}	Initial concentration of product	15 kg/m^3
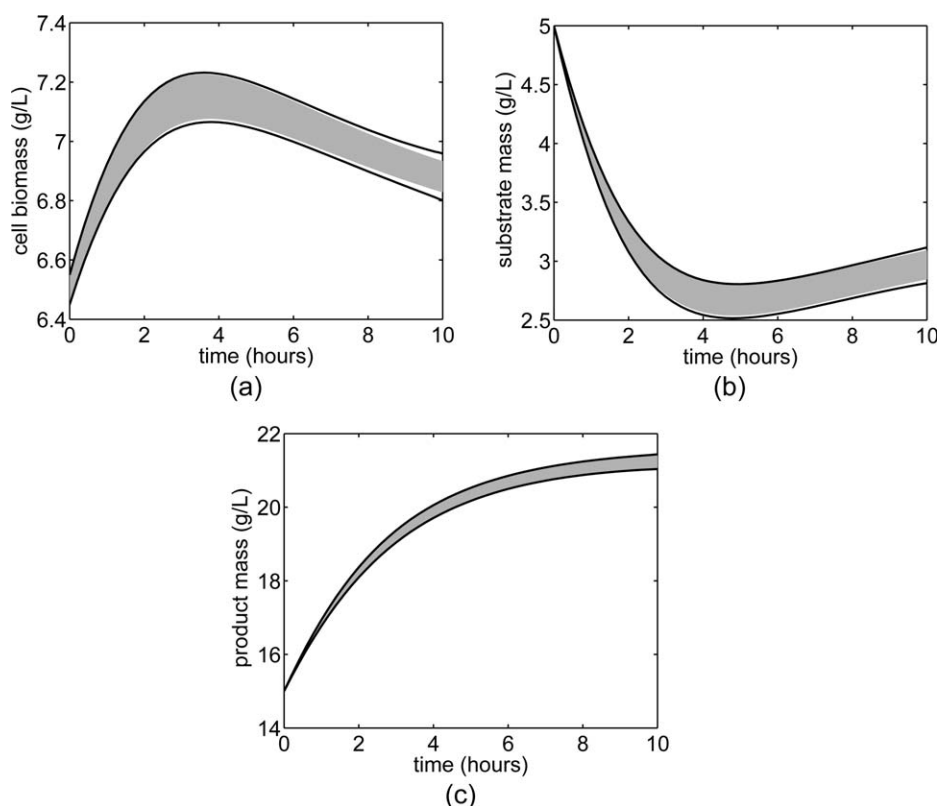
Figure 23. Trajectories of cell biomass concentration x_1, substrate concentration x_2, and product concentration x_3 for three-state bioreactor.

VSPODE bounds are in black, and Monte Carlo simulation (200 samples) results are in gray.

The third-order Taylor models were used with $d = 100$ to compute bounds on the probability distributions for values of $x_1(5)$, $x_1(7.5)$, and $x_1(10)$. This computation required a total
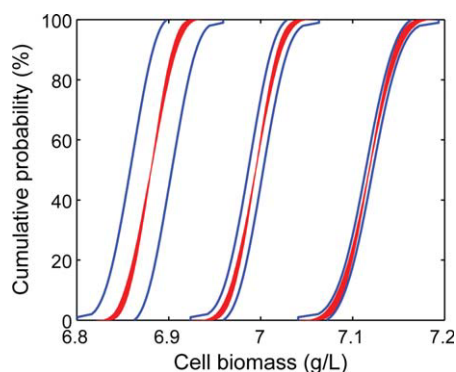

Figure 24. Solid curves show probability distribution (based on all distribution shapes for x_{10}, μ_{max}, and k_s) of x_1 for three-species bioreactor at times (from left to right) of $t = 10$, $t = 7.5$, and $t = 5$ h, as computed with p-boxes using a third-order Taylor model and SIR with 100 discretizations.

Shaded areas (red) show probability distribution of x_1 at the same times as computed with Monte Carlo simulations (based only on uniform distributions for x_{10}, μ_{max}, and k_s). See text for discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of about 13.6 min of CPU time. The resulting p-boxes are shown by the solid curves (blue) in Figure 24. The loosening of bounds seen in Figure 23a is clearly apparent here as well, especially for the result at $t = 10$ h. We repeated the computation of probability bounds using fifth-order Taylor models with $d = 50$. This required about 4.7 min of CPU time (with $q = 5$ and $d = 100$, the CPU time required is about 39.5 min). These results are shown by the solid curves (blue) in Figure 25. At $t = 5$ h, the probability bounds are not quite as tight as those of the $q = 3$ and $d = 100$ case (Figure 24), as at this time the effect of using more SIR subintervals has a greater tightening effect than the higher order Taylor model. However, for the larger times, especially $t = 10$ h, the probability bounds are tighter with the use of the fifth-order Taylor model, even with fewer SIR subintervals. Clearly, what is most important is using a polynomial order large enough to obtain a high-quality Taylor model in the first place, one that is capable of providing tight trajectory bounds. No amount of additional SIR subintervals can tighten the probability bounds if the trajectory bounds provided by the Taylor model are loose. Computation times for determining the probability bounds using p-box arithmetic can be reduced by using the various acceleration techniques noted by Ferson and Hajagos.[25]

For comparison, we also performed a second-order Monte Carlo analysis to obtain probability distributions for $x_1(5)$, $x_1(7.5)$, and $x_1(10)$. The procedure used was the same as described previously, here with uniform distributions contained in the p-boxes PB(x_{10}), PB(k_s), and PB(μ_{max}) chosen at random in an outer loop, and then used in an inner loop
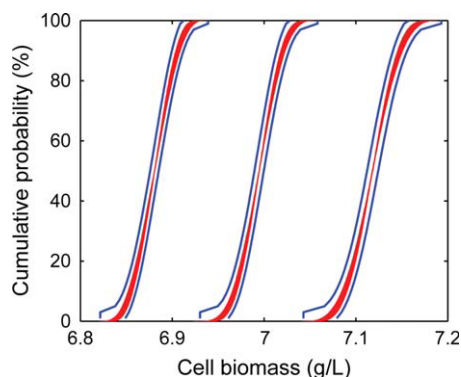
Figure 25. Solid curves show probability distribution (based on all distribution shapes for x_{10}, μ_{max}, and k_s) of x_1 for three-species bioreactor at times (from left to right) of $t = 10$, $t = 7.5$, and $t = 5$ h, as computed with p-boxes using a fifth-order Taylor model and SIR with 50 discretizations.

Shaded areas (red) show probability distribution of x_1 at the same times as computed with Monte Carlo simulations (based only on uniform distributions for x_{10}, μ_{max}, and k_s). See text for discussion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

for sampling values of x_{10}, k_s, and μ_{max}. As in the previous example, we used 100 outer loops and 10,000 inner loops. The computation time was 58.5 min. The results are shown by the shaded areas (red) in both Figures 24 and 25 and are consistent with results obtained from the Taylor model approach. We note again that the Monte Carlo and Taylor model results are not directly comparable, because the former accounts for only a subset (uniform distributions) of the input distributions accounted for by the latter.

Concluding Remarks

We have described here a method for the direct computation (without need for sampling) of probability bounds for the outputs of an uncertain nonlinear dynamic system. This method uses Taylor models that represent the state variables of interest in terms of the uncertain quantities (parameters and/or initial states). The Taylor models are computed using the verified ODE solver VSPODE,[11] which provides a powerful tool for bounding the solutions of nonlinear parametric ODEs. Given p-box representations of the uncertainties in initial states and parameters, the Taylor models can be used to directly compute probability bounds (p-boxes) for the state variables, as demonstrated in several example problems. This technique is best suited to problems in which the effect of relatively few uncertain quantities is to be investigated. Assuming that efficient matrix factorization and multiplication methods are used in connection with VSPODE, it should be possible to handle problems with a relatively large number of state variables, though to date we have little experience with such problems.

Literature Cited

1. Lin Y, Stadtherr M. Guaranteed state and parameter estimation for nonlinear continuous-time systems with bounded-error measurements. *Ind Eng Chem Res*. 2007;46:71–98.
2. Rawlings JB, Bakshi BR. Particle filtering and moving horizon estimation. *Comput Chem Eng*. 2006;30:1529–1541.
3. Bansal V, Perkins JD, Pistikopoulos EN, Ross R, van Schijndel JMG. Simultaneous design and control optimisation under uncertainty. *Comput Chem Eng*. 2000;24:261–266.
4. Faanes A, Skogestad S. Feedforward control under the presence of uncertainty. *Eur J Control*. 2004;10:30–46.
5. Pajula E, Ritala R. Measurement uncertainty in integrated control and process design — a case study. *Chem Eng Proc*. 2006;45:312–322.
6. Ferson S, Ginzburg L, Akçakaya R. Whereof one cannot speak: when input distributions are unknown. Technical Report, Applied Biomathematics, Setauket, NY, 1996 (available at www.ramas.com/whereof.pdf).
7. Williamson RC, Downs T. Probabilistic arithmetic I. Numerical methods for calculating convolutions and dependency bounds. *Int J Approx Reason*. 1990;4:89–158.
8. Ferson S. What Monte Carlo methods cannot do. *Hum Ecol Risk Assess*. 1996;2:990–1007.
9. Iman RL, Helton JC. The repeatability of uncertainty and sensitivity analyses for complex probabilistic risk assessments. *Risk Anal*. 1991;11:591–606.
10. Helton J. Treatment of uncertainty in performance assessments for complex systems. *Risk Anal*. 1994;14:483–511.
11. Lin Y, Stadtherr MA. Validated solutions of initial value problems for parametric ODEs. *Appl Numer Math*. 2007;57:1145–1162.
12. Li W, Hyman JM. Computer arithmetic for probability distribution variables. *Reliab Eng Syst Safe*. 2004;85:191–209.
13. Dubois D, Prade H. Operations on fuzzy numbers. *Int J Syst Sci*. 1978;9:613–626.
14. Zadeh LA. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Set Syst*. 1978;1:3–28.
15. Neumaier A. Clouds, fuzzy sets, and probability intervals. *Reliab Comput*. 2004;10:249–272.
16. Hansen ER, Walster GW. *Global Optimization Using Interval Analysis*. New York: Marcel Dekker, 2004.
17. Moore RE. *Interval Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1966.
18. Jaulin L, Kieffer M, Didrit O, É Walter. *Applied Interval Analysis*. London: Springer-Verlag, 2001.
19. Kearfott RB. *Rigorous Global Search: Continuous Problems*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.
20. Neumaier A. *Interval Methods for Systems of Equations*. Cambridge, UK: Cambridge University Press, 1990.
21. Moore RE, Kearfott RB, Cloud MJ. *Introduction to Interval Analysis*. Philadelphia, PA: SIAM Press, 2009.
22. Yager RR. Arithmetic and other operations on Dempster-Shafer structures. *Int J Man Mach Stud*. 1986;25:357–366.
23. Frank MJ, Nelsen RB, Schweizer B. Best-possible bounds for the distribution of a sum — a problem of Kolmogorov. *Probab Theory Rel*. 1987;74:199–211.
24. Ferson S. *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*. Boca Raton, FL: Lewis Press, 2002.
25. Ferson S, Hajagos JG. Arithmetic with uncertain numbers: Rigorous and (often) best possible answers. *Reliab Eng Syst Safe*. 2004;85:135–152.
26. Berleant D, Zhang J. Using Pearson correlation to improve envelopes around the distributions of functions. *Reliab Comput*. 2004;10:139–161.
27. Berleant D, Xie L, Zhang J. Statool: a tool for Distribution Envelope determination (DEnv), an interval-based algorithm for arithmetic on random variables. *Reliab Comput*. 2003;9:91–108.
28. Makino K, Berz M. *Remainder differential algebras and their applications*. In: Berz M, Bishof C, Corliss G, Griewank A, editors. *Computational Differentiation: Techniques, Applications, and Tools*. Philadelphia: SIAM, 1996:63–74.
29. Makino K, Berz M. Efficient control of the dependency problem based on Taylor model methods. *Reliab Comput*. 1999;5:3–12.

30. Makino K, Berz M. Taylor models and other validated functional inclusion methods. *Int J Pure Appl Math*. 2003;4:379–456.
31. Neumaier A. Taylor forms — use and limits. *Reliab Comput*. 2003; 9:43–79.
32. Nedialkov NS, Jackson KR, Corliss GF. Validated solutions of initial value problems for ordinary differential equations. *Appl Math Comput*. 1999;105:21–68.
33. Neher M, Jackson KR, Nedialkov NS. On Taylor model based integration of ODEs. *SIAM J Numer Anal*. 2007;45:236–262.
34. Lohner RJ. *Computations of guaranteed enclosures for the solutions of ordinary initial and boundary value problems*. In: Cash J, Gladwell I, editors. *Computational Ordinary Differential Equations*. Oxford, UK: Clarendon Press, 1992:425–435.
35. Nedialkov NS, Jackson KR, Pryce JD. An effective high-order interval method for validating existence and uniqueness of the solution of an IVP for an ODE. *Reliab Comput*. 2001;7:449–465.
36. Berz M, Makino K. Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliab Comput*. 1998;4:361–369.
37. Rauh A, Hofer EP, Auer E. ValEncIA-IVP: a comparison with other initial value solvers. In: Proceedings of the 12th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic, and Validated Numerics (SCAN 2006), Duisburg, Germany. 2006:36.
38. Lambov B. *Interval arithmetic using SSE-2*. In: Hertling P, Hoffman CM, Luther W, Revol N, editors. *Reliable Implementation of Real Number Algorithms: Theory and Practice*. Berlin: Springer, 2008: 102–113.
39. Bastin G, Douchain D. *On-line Estimation and Adaptive Control of Bioreactors*. New York: Elsevier, 1990.