6

Rigorous Propagation of Imprecise Probabilities in Process Models

Joshua A. Enszer and Mark A. Stadtherr*

University of Notre Dame
Notre Dame, IN 46556

CONTENTS

Introduction 78
Background 78
    Intervals 79
    Probability Boxes (P-boxes) 79
    Taylor Models 81
Problem Statement 82
Solution Procedure 83
    Enclosing the State Variables 83
    Probability Distribution of State Variables 84
Example: Bioreactor Process 85
    Case 1: $X_0 \in [0.794, 0.864]$ and $\mu_{max} \in [1.15, 1.25]$ 86
    Case 2: $X_0 \in [0.81, 0.83]$ and $\mu_{max} \in [1.08, 1.32]$ 89
Concluding Remarks 90
Acknowledgments 91
References 91

ABSTRACT Models of process dynamics often involve uncertain parameters, inputs and/or initial states. Even if probability distributions for the uncertainties are available, they too may be imprecise. An approach for rigorously and tightly bounding the effects of such uncertainty in process models is described here, and it is shown how this can be extended to determine rigorous bounds on the probabilities of achieving desired outcomes.

KEYWORDS *Process dynamics, Ordinary differential equations, Initial value problems, Probability bounds analysis.*

* To whom all correspondence should be addressed

Introduction

The process models used in analysis and design, whether static or dynamic, often involve uncertainties in parameters, inputs, and/or initial states. Determining how these uncertainties propagate through a model to affect its outputs, and doing so rigorously, can be a very challenging problem, especially for nonlinear dynamic systems. 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 (p-box), which provides upper and lower bounds on the cumulative probability distribution function for the uncertain quantity. We will concentrate here on the latter case, in which uncertain quantities in the process model are characterized by imprecise probabilities represented by p-boxes. Furthermore, we will focus on the difficult case of a nonlinear dynamic model, i.e., a nonlinear system of ordinary differential equations (ODEs) for which an initial value problem (IVP) must be solved.

One common approach for dealing with this problem is Monte Carlo analysis. However, in this approach, it is not possible to investigate the complete space of uncertain quantities in a finite number of simulations, and thus Monte Carlo analysis may fail to capture all possible system behaviors, especially in the case of nonlinear systems. We will describe here an approach for rigorously and tightly bounding the effects of uncertainty in process models, and show how this can be extended to determine rigorous bounds on the probabilities that desired outcomes are achieved. This approach is enabled by the use of Taylor models to represent the solution of IVPs with uncertain parameters and/or initial states, as described recently by Lin and Stadtherr (2007b).

The rest of this paper is organized as follows. In the next section, we provide some background on the approaches used here for representing uncertainties, as well as on the use of Taylor models. This is followed by a formal problem statement and then a description of the solution procedure used. Finally a demonstrative example is provided, with comparison to results obtained from Monte Carlo simulation.

Background

There are several ways to treat numerical uncertainty in mathematical models. In this section, we provide background on the specific approaches used here, namely intervals and p-boxes. We also provide background on the use of Taylor models.

Intervals

A real interval X is the set of real numbers between and inclusive of its lower bound $\underline{X}$ and upper bound $\overline{X}$; that is, $X = [\underline{X}, \overline{X}] = \{x \in \Re \mid \underline{X} \le x \le \overline{X}\}$. Thus, an interval can be used to represent an uncertain quantity for which no information is available other than its lower and upper bounds. Intervals are also used to represent computational uncertainties due to machine rounding. That is, a real number that is not exactly machine representable is bounded by an interval determined from the real number's nearest floating-point representations. An interval vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ has n real interval components and can be thought of as an n-dimensional rectangle or box. Interval matrices are similarly defined.

Basic arithmetic operations are defined on intervals according to

$$X \operatorname{op} Y = \{x \operatorname{op} y \mid x \in X, y \in Y\} \tag{1}$$

for $\operatorname{op} \in \{+, -, \times, \div\}$ and, in the case of division, $0 \notin Y$, though division in the case of Y containing zero is allowed in extensions of interval arithmetic (e.g., Hansen and Walster, 2004). Commutativity and associativity hold for addition and multiplication, but these operations are only subdistributive. Interval versions of the elementary functions can also be defined.

For a real function $f(\boldsymbol{x})$, an interval extension $F(\boldsymbol{X})$ encloses the range of $f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$. When $f(\boldsymbol{x})$ can be written as a series of arithmetic operations and elementary functions, substituting $\boldsymbol{X}$ into $f(\boldsymbol{x})$ and evaluating using interval arithmetic gives the "natural" interval extension. However, computing the interval extension in this manner often results in overestimation of the function range due to the "dependency" problem. This issue may arise if there are multiple occurrences of the same variable in the function, since in computing the natural interval extension each such occurrence is treated as being independent, though clearly this is not the case. Another source of overestimation that may arise in the use of interval methods is the "wrapping" effect. This occurs when an interval is used to enclose (wrap) a set of results that is not an interval. If overestimation due to either of these issues is propagated from step to step in an integration procedure for ODEs, it can quickly lead to the loss of a meaningful enclosure. One approach for addressing these issues is the use of Taylor models, as discussed later in this section.

Several good introductions to interval analysis, as well as interval arithmetic and other aspects of computing with intervals, are available (e.g., Hansen and Walster, 2004; Jaulin et al., 2001; Kearfott, 1996; Neumaier, 1990). 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.

Probability Boxes (P-boxes)

If more is known about an uncertain quantity than simply its upper and lower bounds, then this can be represented in a number of ways. We will

assume here that some information, not necessarily precise, is known about the probability distribution of the uncertainty. In other situations, the use of fuzzy numbers (e.g., Dubois and Prade, 1978), clouds (e.g., Neumaier, 2004), and other representations of uncertain knowledge may be appropriate, depending on the type of information that is available.

For some quantity (variable or parameter) x, the cumulative distribution function (CDF) $F(z)$ gives the probability that $x \leq z$. In practice, knowledge of the probability distribution describing an uncertainty is often itself uncertain. To deal with imprecise probability distributions, we use probability boxes (p-boxes) (e.g., Ferson, 2002; Ferson et al., 2004). A p-box, as defined below, is a way to bound probability distributions, in much the same way that an interval is used to bound a real number. Furthermore, arithmetic operations with p-boxes can be performed, again in much the same way as done with intervals. Computations with p-boxes allow for more information about the uncertainty of a quantity to be utilized in modeling and analysis.

Formally, a p-box is the set of all CDFs enclosed by two bounding CDFs $F(z)$ and $G(z)$; that is,

$$(F, G) = \{H(z) \mid F(z) \geq H(z) \geq G(z)\}. \tag{2}$$

Less formally, a p-box can be thought of as a set of interval bounds on a cumulative distribution function, and thus, in practice, computation with p-boxes and intervals are analogous (Ferson, 2002). The bounding functions $F(z)$ and $G(z)$ are decomposed into interval-mass pairs, and interval arithmetic is then applied. Therefore, computation with p-boxes involves the same issues of dependency and (especially) wrapping that occur in computations with intervals. For a p-box represented as n interval-mass pairs, a single arithmetic operation with another independent p-box provides a result with n^2 interval-mass pairs, and a p-box with n interval-mass pairs must then be used to condense (wrap) this result.

A p-box may be constructed from any available information about an uncertain quantity, including, but not limited to, any combination of its maximum, minimum, mean, median, or standard deviation. An interval is a special case of a p-box where only the maximum and minimum are known. P-boxes may also be created by assuming a particular form of probability distribution for the bounding functions $F(z)$ and $G(z)$. Some sample p-boxes are shown in Figure 1. Figure 1(a) shows a p-box representation of a standard interval. Figure 1(b) gives the p-box for a variable whose minimum, median, and maximum are known. Figure 1(c) is a p-box created from two different uniform distributions as bounding functions. It is important to note that the true probability distribution simply lies between the bounding functions and does not necessarily take the same form as a bounding function; that is, a distribution within a p-box bounded by uniform distributions is not necessarily also uniform.

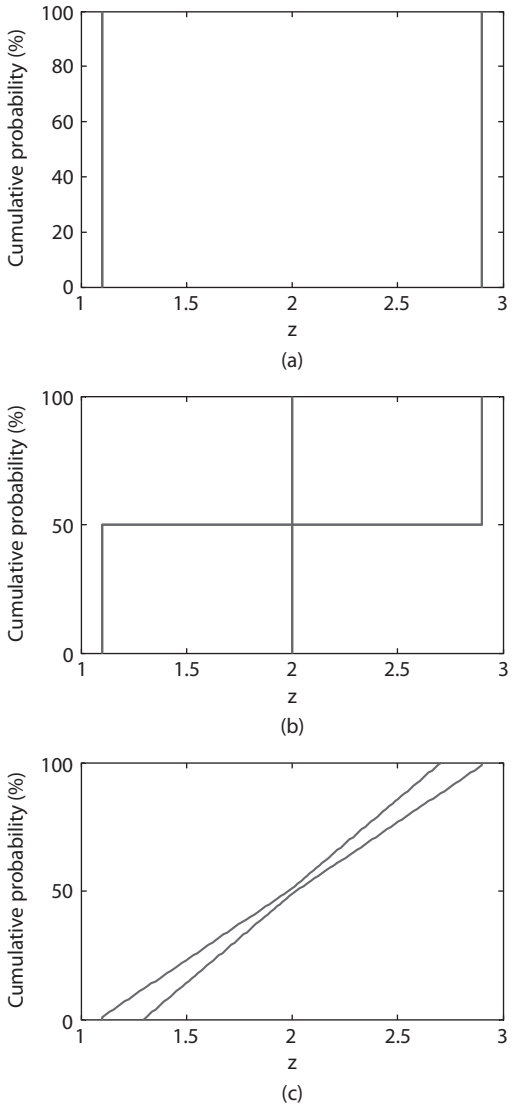

FIGURE 1
Examples of p-boxes for given knowledge of uncertainty. See text for discussion.

Taylor Models

In order to alleviate the overestimation problems that occur in interval computations, Makino and Berz (1996, 1999) have described a remainder differential algebra (RDA) approach for bounding function ranges. In this method, a function is represented using a model consisting of a real-valued

Taylor polynomial and an interval remainder bound. Such a model is called a Taylor model.

One way of forming a Taylor model of a function is through direct use of the Taylor theorem. Consider a real function $f(\boldsymbol{x})$ that is $(q + 1)$ times partially differentiable on $\boldsymbol{X}$ and let $\boldsymbol{x}_0 \in \boldsymbol{X}$. The Taylor theorem states that for each $\boldsymbol{x} \in \boldsymbol{X}$, there exists a real ζ with $0 < \zeta < 1$ such that

$$f(\boldsymbol{x}) = p_f(\boldsymbol{x} - \boldsymbol{x}_0) + r_f(\boldsymbol{x} - \boldsymbol{x}_0, \zeta) \tag{3}$$

where p_f is a q-th order polynomial (truncated Taylor series) in $(\boldsymbol{x} - \boldsymbol{x}_0)$, and r_f is a remainder term, which can be quantitatively bounded over $0 < \zeta < 1$ and $\boldsymbol{x} \in \boldsymbol{X}$ 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 $f(\boldsymbol{x})$ over X then 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 $f \in T_f$ for $\boldsymbol{x} \in \boldsymbol{X}$; therefore, T_f encloses the range of f over $\boldsymbol{X}$.

In practice, it is more useful to compute Taylor models of functions by performing arithmetic operations on other Taylor models. Arithmetic operations, including addition, multiplication, reciprocal, and intrinsic functions, can be done using the RDA operations described by Makino and Berz (1996; 1999; 2003). Using these, it is possible to start with simple functions such as the constant function $f(\boldsymbol{x}) = k$, for which $T_f = (k, [0, 0])$, and the identity function $f(x_i) = x_i$, for which $T_f = (x_{i0} + (x - x_{i0}), [0, 0])$, and then to compute Taylor models for more complicated functions. Hence, it is possible to compute a Taylor model for any function representable in a computer environment by simple operator overloading through RDA operations. It has been shown that the Taylor model often yields sharper bounds for modest to complicated functional dependencies compared to other rigorous bounding methods (Makino and Berz, 1996, 1999; Neumaier, 2003). A discussion of the uses and limitations of Taylor models has been given by Neumaier (2003).

Problem Statement

Consider the parametric, autonomous IVP

$$\boldsymbol{y}'(t) = \boldsymbol{f}(\boldsymbol{y}, \theta),\ \boldsymbol{y}(t_0) = \boldsymbol{y}_0 \in \boldsymbol{Y}_0,\ \theta \in \Theta \tag{4}$$

over the time interval $t \in [t_0, t_m]$ where $t_m > t_0$. Here, y is the n-dimensional vector of state variables whose initial value is $\boldsymbol{y}_0$, and θ is a p-dimensional vector of time-invariant parameters. The uncertainties in the initial states and parameters are enclosed in the interval vectors $\boldsymbol{Y}_0$ and Θ, respectively. Further, additional probabilistic information is available for at least one component of y_0 or Θ, and this information is expressed in the form of a p-box.

We also assume that $\boldsymbol{f}$ is $(k-1)$ times continuously differentiable with respect to y and $(q+1)$ times continuously differentiable with respect to θ. Here, k is the order of the truncation error in the interval Taylor series (ITS) used in the solution procedure (see below), and q is the order of the Taylor model used to represent dependence on initial values and parameters. We also assume that $\boldsymbol{f}$ is representable by a finite number of standard functions.

Our goal is twofold: 1) to obtain verified (e.g., mathematically and computationally guaranteed) enclosures of the state variables y at specified times t_k of interest from t_0 to t_m, and 2) to obtain a probability distribution, in the form of a p-box, for the values of y within these enclosures.

Solution Procedure

In this section, we summarize the solution methods used to achieve the two goals stated above.

Enclosing the State Variables

Interval methods (also called validated methods or verified methods) for ODEs provide a natural approach for computing the desired enclosure of the state variables in the problem stated above. Traditional interval methods usually consist of two processes applied at each integration step. In the first process, existence and uniqueness of the solution are proved using the Picard-Lindelöf operator and the Banach fixed point theorem, 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 realized 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. (1999), and more recent work has been reviewed by Neher et al. (2007). For addressing this problem, there are various packages available, including AWA (Lohner, 1992), VNODE (Nedialkov, 1999; Nedialkov et al., 2001) COSY VI (Berz & Makino, 1998) and ValEncIA-IVP (Rauh et al., 2006). In this study, we will use a new validating solver VSPODE (Lin & Stadtherr, 2007b) for parametric ODEs, 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 (Makino & Berz, 1996, 1999, 2003) to deal with the dependency and wrapping problems on the uncertain quantities (parameters and initial values). We will summarize here the basic ideas of the method used.

As in traditional interval methods, each integration step in VSPODE consists of two phases, as noted above. Assume that at time t_j there is a known enclosure (computed in the previous time step) $\boldsymbol{Y}_j$ of $\boldsymbol{y}_j = \boldsymbol{y}(t_j)$. In the first phase of the next time step, a step size $h_j = t_{j+1} - t_j$ and coarse enclosure $\tilde{\boldsymbol{Y}}_j$ are determined such that a unique solution $\boldsymbol{y}(t) \in \tilde{\boldsymbol{Y}}_j$ is guaranteed to exist for all $t \in [t_j, t_{j+1}]$, all $\boldsymbol{y}_j \in \boldsymbol{Y}_j$ and all $\theta \in \Theta$. This is achieved using a high-order (k) ITS with respect to time and the traditional approach using the Picard-Lindelöf operator and the Banach fixed-point theorem. In the second phase of the method, a tighter enclosure $\boldsymbol{Y}_{j+1} \subseteq \tilde{\boldsymbol{Y}}_j$ is computed, such that $\boldsymbol{y}_{j+1} \in \boldsymbol{Y}_{j+1}$ for all $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$ and all $\theta \in \Theta$. This is done by using an ITS approach to compute $\boldsymbol{T}_{\boldsymbol{y}_{j+1}}(\boldsymbol{y}_0, \theta)$, a Taylor model of $\boldsymbol{y}_{j+1}$ in terms of the initial values $\boldsymbol{y}_0$ and parameters θ. To compute this Taylor model, we begin by representing the interval initial states and parameters by the Taylor models (identity functions) $\boldsymbol{T}_{\boldsymbol{y}_0}$ and $\boldsymbol{T}_\theta$, respectively. Then, we can determine Taylor models $\boldsymbol{T}_{\boldsymbol{f}^{[i]}}$ of the Taylor series coefficients $\boldsymbol{f}^{[i]}(\boldsymbol{y}_0, \theta)$ by using RDA operations to compute $\boldsymbol{T}_{\boldsymbol{f}^{[i]}} = \boldsymbol{f}^{[i]}(\boldsymbol{T}_{\boldsymbol{y}_0}, \boldsymbol{T}_\theta)$. Using an ITS for $\boldsymbol{y}_{j+1}$ with coefficients given by $\boldsymbol{T}_{\boldsymbol{f}^{[i]}}$, and using the mean-value theorem, one can obtain $\boldsymbol{T}_{\boldsymbol{y}_{j+1}}(\boldsymbol{y}_0, \theta)$, the desired Taylor model of $\boldsymbol{y}_{j+1}$ in terms of $\boldsymbol{y}_0$ and θ. To control the wrapping effect, the state enclosures are propagated using a new type of Taylor model consisting of a polynomial and a *parallelepiped* (as opposed to an interval) remainder bound. Complete details of the computation of $\boldsymbol{T}_{\boldsymbol{y}_{j+1}}$ using VSPODE are given by Lin and Stadtherr (2007b), who also describe a procedure for efficient bounding of $\boldsymbol{T}_{\boldsymbol{y}_{j+1}}(\boldsymbol{y}_0, \theta)$ over $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$ and $\theta \in \Theta$ to obtain the final state enclosure $\boldsymbol{Y}_{j+1}$.

Probability Distribution of State Variables

Using the method summarized above, we can obtain, for the specified time of interest t_k, a Taylor model $\boldsymbol{T}_{\boldsymbol{y}_k}(\boldsymbol{y}_0, \theta)$, that gives the state variables $\boldsymbol{y}_k = \boldsymbol{y}(t_k)$ as a polynomial function $\boldsymbol{p}_{\boldsymbol{y}_k}(\boldsymbol{y}_0, \theta)$ of the initial states $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$ and the parameters $\theta \in \Theta$, plus a small remainder bound. If probability distributions (p-boxes) are available for $\boldsymbol{y}_0$ and for θ, then these can be substituted directly into $\boldsymbol{T}_{\boldsymbol{y}_k}(\boldsymbol{y}_0, \theta)$, and a p-box giving bounds on the probability distribution for y_k can be computed using p-box operations.

Straightforward application of p-box operations to evaluate the Taylor model $\boldsymbol{T}_{\boldsymbol{y}_k}(\boldsymbol{y}_0, \theta)$ may lead to significant overestimation of bounds on the true probability distribution of the state variables, due to the dependency problem and the wrapping effect. One method to obtain a much tighter enclosure is subinterval reconstitution (SIR). In this procedure, when p-box operations are done using the Taylor model, the intervals of the decomposed p-box are further partitioned into subintervals, which are then projected through the Taylor model separately. The final bounds are then reconstituted using the union of the subinterval results (Ferson and Hajagos, 2004). Results obtained from use of SIR will, in general, also overestimate the bounds somewhat, but if a reasonably large number of subintervals are used, the p-box bounds can

become quite good. P-box operations and evaluation of the Taylor model, including optional use of SIR, can be performed using the risk analysis software RAMAS Risk Calc (Ferson, 2002). We also employ our own skeletal Matlab implementation of p-box arithmetic and SIR.

Example: Bioreactor Process

We consider here the microbial growth of a single biomass feeding on a substrate in a bioreactor. The process is described by the ODE system

$$X' = (\mu - \alpha D)X \tag{5}$$

$$S' = D(S_f - S) - k\mu X, \tag{6}$$

where X and S represent the biomass and substrate concentrations, respectively, α is the process heterogeneity parameter, D is the dilution rate, S_f is the concentration of substrate in the influent, k is the yield coefficient, and the growth rate μ of biomass follows Monod reaction kinetics (Bastin and Douchain, 1990; Bequette, 2003). For Monod kinetics,

$$\mu = \frac{\mu_{\max} S}{K_S + S}, \tag{7}$$

where $\mu_{\max}$ is the maximum growth rate and K_s is the saturation parameter. The initial states ($t = 0$) are X_0 and S_0.

For this example, we will consider two quantities to be uncertain, namely the maximum growth rate parameter $\mu_{\max}$ and the initial biomass concentration X_0. Two different cases for these uncertain quantities will be considered. The other parameters, including the initial substrate concentration, are taken to be fixed at the values shown in Table 1 (Lin and Stadtherr, 2007a).

For all p-box operations, the p-boxes used were discretized into 100 interval-mass pairs (each interval corresponding to a single percentile). In applying

TABLE 1

Fixed Quantities in Bioreactor Example.

Parameter	Value	Units
S_0	0.8	g S/L
α	0.5	
D	0.36	day^{-1}
S_f	5.7	g S/L
k	10.53	g S/g X
K_s	7.0	g S/L

VSPODE, the order of the interval Taylor series used was $k = 17$, while the order of the Taylor model used was $q = 5$. In the integration procedure, a constant step size of $h = 0.2$ was used, though this step size is automatically reduced if needed. The Taylor model remainder bounds were obtained using a QR-factorization process (Lin and Stadtherr, 2007b). In the SIR procedure for computation with p-boxes, each of the 100 interval-mass pairs of the p-box was bisected. All problems were solved on an Intel Pentium 4 3.2 GHz machine running Red Hat Linux.

Case 1: $X_0 \in [0.794, 0.864]$ and $\mu_{max} \in [1.15, 1.25]$

In this first case, there is an uncertainty of about ±4.2% (relative to the mean) in both μ_{max} and X_0. We shall assume that the uncertainty in both these quantities can be described by p-boxes bounded by uniform distributions, as shown in Figure 2. We further assume that these two uncertainties are independent from one another (i.e., there is no correlation between them).

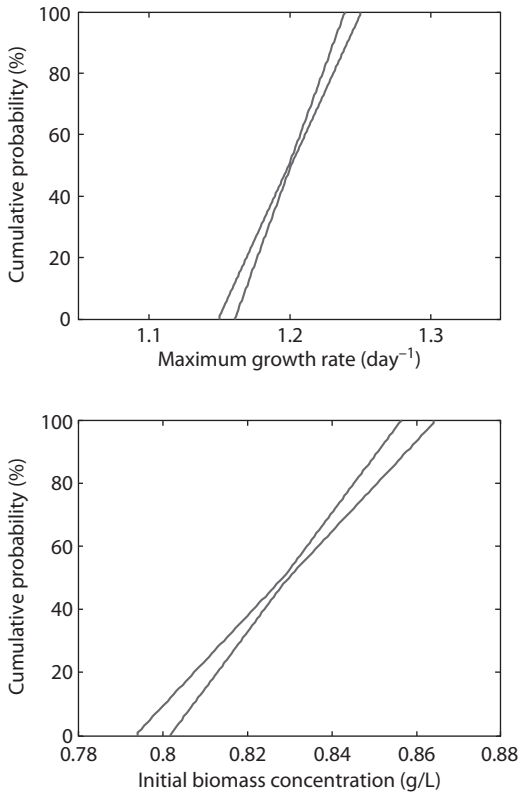

FIGURE 2
P-box representation of uncertainties for case 1 in bioreactor example.

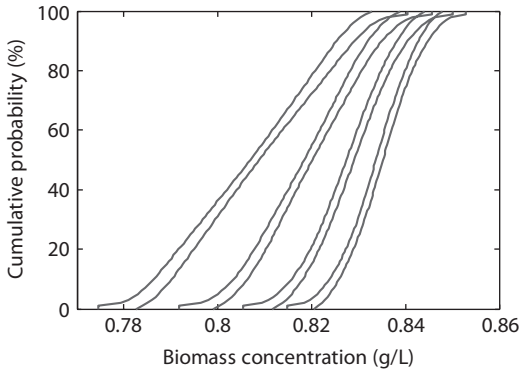

FIGURE 3
Case 1 results from Taylor model method. P-box bounds for biomass concentration $X(t)$ at (left to right) t = 2.5, 5, 7.5 and 10 days.

Using the procedure described above, we computed p-box representations for the probability distribution of the state variables at four different times, t = 2.5, 5, 7.5, and 10 days. These results are plotted in Figure 3 for the biomass concentration $X(t)$, showing (from left to right) the p-box results as time increases. For example, these results show that the probability that $X \le 0.82$ is bounded by the interval $[72.1, 78.3]\%$ at $t = 2.5$ days, by $[49.4, 54.8]\%$ at $t = 5$ days, by $[14.9, 20.7]\%$ at $t = 7.5$ days, and by $[0, 3.5]\%$ at t = 10 days. These bounds are mathematically and computationally rigorous. The computational expense of obtaining these results was quite small. Use of VSPODE to determine the Taylor model for the state variables at t = 10 days required 0.444 seconds. Once the Taylor model was obtained, the p-box operations (with SIR) needed to get the final results for t = 10 days required 188.2 seconds using Matlab.

As a basis for comparison, we also determined probability bounds for $X(t)$ at the same four points in time using Monte Carlo simulation. To do this first requires sampling the space of the probability distributions for $\mu_{\max}$ and X_0. We used 100 samples, each a uniform distribution chosen randomly from the p-boxes for $\mu_{\max}$ and X_0. For each of these 100 distributions, we then ran (using Matlab with ode45) 50,000 simulations to obtain a probability distribution for X. Combining the results for each of the 100 input distributions, we obtain Figure 4, again showing the results for increasing time from left to right. The results (Figure 3) obtained using the Taylor model approach described here are clearly consistent with the MC results. It is important to note: 1) Probability bounds obtained from MC analysis are not rigorous, but those obtained from the Taylor model analysis are. For the number of trials done here, which is relatively large to ensure meaningful results, the computational expense was quite large, about 9 hours (vs. about 3 minutes for the more rigorous Taylor model approach). 2) The probability bounds from MC become quite narrow at the median, less so than obtained from the Taylor

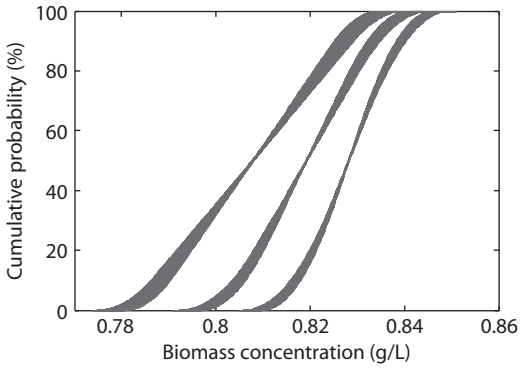

FIGURE 4
Case 1 results from Monte Carlo analysis. Probability distributions for biomass concentration $X(t)$ at (left to right) t = 2.5, 5, 7.5 and 10 days.

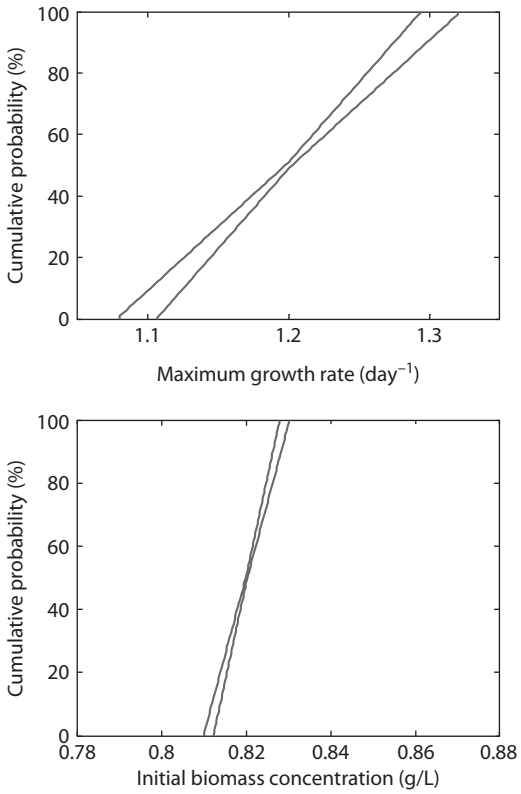

FIGURE 5
P-box representation of uncertainties for Case 2 in bioreactor example.

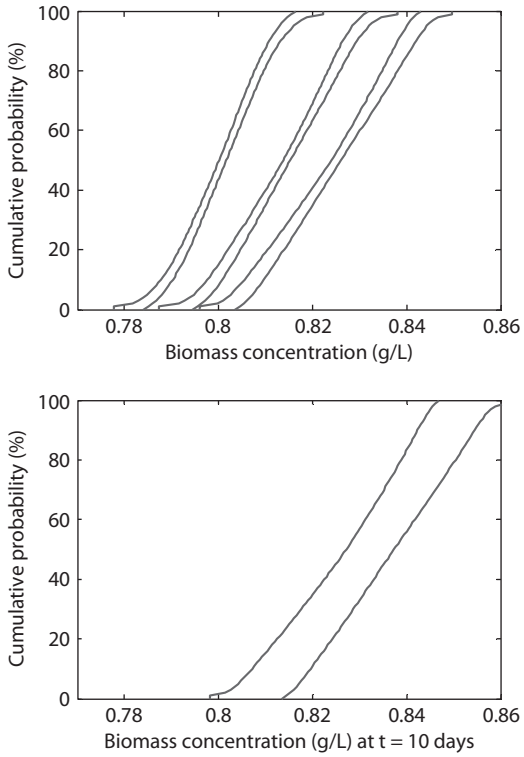

FIGURE 6
Case 2 results from Taylor model method. P-box bounds for biomass concentration $X(t)$ at (top, left to right) t = 2.5, 5, 7.5 and (bottom) 10 days.

model analysis. This reflects the use of only uniform distributions in the MC analysis. A p-box with uniform bounds also contains non-uniform distributions, and this is accounted for in results of Figure 3.

Case 2: $X_0 \in [0.81, 0.83]$ and $\mu_{max} \in [1.08, 1.32]$

In this case, we consider a larger degree of uncertainty in μ_{max}, now ±10%, than in the previous case, and a smaller degree of uncertainty in X_0, now about ±1.2%. Again we assume probability distributions contained in p-boxes with uniform bounds, as shown in Figure 5.

Figure 6 shows the p-box enclosures for the reactor biomass concentration obtained from the Taylor model approach with p-box arithmetic, and Figure 7 shows the probability distributions obtained using Monte Carlo analysis. Though the Taylor model results are quite good for t = 2.5, 5 and 7.5 days, they are not as good for t = 10 days, the p-box for which is noticeably wider than

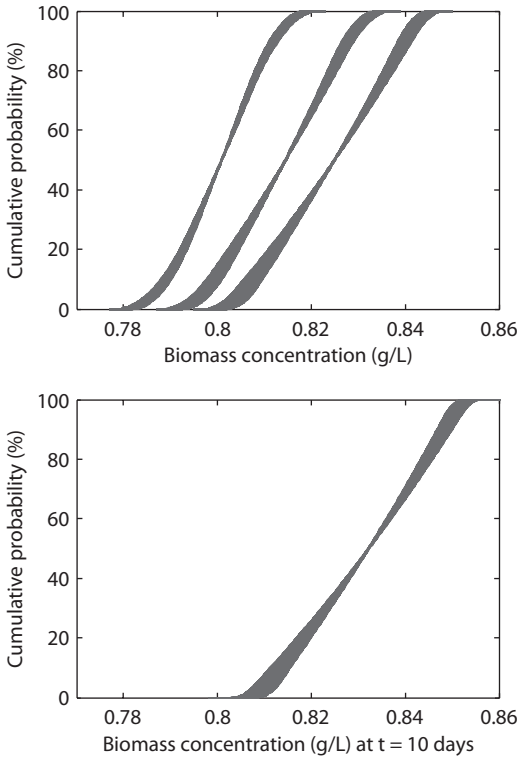

FIGURE 7
Case 2 results from Monte Carlo analysis. Probability distributions for biomass concentration $X(t)$ at (top, left to right) t = 2.5, 5, 7.5 and (bottom) 10 days.

the other p-boxes, and in comparison to the MC results. This is occurring in part because of growth, as time increases, in the width of the remainder bound term in the Taylor model, and is a reflection of the relatively large range of $\mu_{\max}$ considered. Better results could be obtained by bisecting $\mu_{\max}$ and employing a SIR-like procedure on the Taylor model level (i.e., a different Taylor model for each $\mu_{\max}$ subinterval).

Concluding Remarks

The parametric nonlinear ODEs that arise in process models for design and analysis often include uncertainty in parameters and initial states, and the distribution of this uncertainty is often not known precisely. We have presented here a new approach, based on Taylor models and probability boxes (p-boxes) for propagating such imprecise probability distributions into the state variable trajectories, enabling the computation of rigorous bounds

on the probabilities that desired outcomes can be achieved. In comparison to Monte Carlo analysis, this new approach provides not only guaranteed bounds, but also a reasonable computational cost. Though we can demonstrate this for a variety of process models, we have focused here on the case of a bioreactor process, computing bounds on the probability distribution of the biomass trajectory.

Acknowledgments

This work was supported in part by the U. S. Department of Energy grant DE-FG36-08GO88020-A0, and by a Lilly Foundation graduate fellowship (JAE).

References

Bastin, G., Douchain, D. (1990). *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, New York, NY.

Bequette, B. W. (2003). *Process Control: Modeling, Design, and Simulation*. Prentice-Hall, Upper Saddle River, NJ.

Berz, M., Makino, K. (1998). Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliable Computing, 4*, 361.

Dubois, D., Prade, H. (1978). Operations on fuzzy numbers. *International Journal of Systems Science, 9*, 613–626.

Ferson, S. (2002). *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*. Lewis Press, Boca Raton, FL.

Ferson, S., Hajagos, J. G. (2004). Arithmetic with uncertain numbers: Rigorous and (often) best possible answers. *Reliability Engineering and System Safety*, *85*, 135–152.

Ferson, S., Nelson, R. B., Hajagos, J., Berleant, D. J., Zhang, J., Tucker, W. T., Ginzburg, L. R., Oberkampf, W. L. (2004). *Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis*. Technical Report, Sandia National Laboratories.

Hansen, E. R., Walster, G. W. (2004). *Global Optimization Using Interval Analysis*. Marcel Dekker, New York, NY.

Jaulin, L., Kieffer, M., Didrit, O., Walter, É. (2001). *Applied Interval Analysis*. Springer-Verlag, London, UK.

Kearfott, R. B. (1996). *Rigorous Global Search: Continuous Problems*. Kluwer, Dordrecht, The Netherlands.

Lin, Y., Stadtherr, M. A. (2007a). Guaranteed state and parameter estimation for nonlinear continuous-time systems with bounded-error measurements. *Industrial & Engineering Chemistry Research*, *46*, 7198.

Lin, Y., Stadtherr, M. A. (2007b). Validated solutions of initial value problems for parametric ODEs. *Applied Numerical Mathematics*, *57*, 1145.

Lohner, R. J. (1992). Computations of guaranteed enclosures for the solutions of ordinary initial and boundary value problems. In: Cash, J., Gladwell, I. (Eds.), *Computational Ordinary Differential Equations*. Clarendon Press, Oxford, UK, 425.

Makino, K., Berz, M. (1996). Remainder differential algebras and their applications. In: Berz, M., Bishof, C., Corliss, G., Griewank, A. (Eds.), *Computational Differentiation: Techniques, Applications, and Tools. SIAM, Philadelphia*, 63.

Makino, K., Berz, M. (1999). Efficient control of the dependency problem based on Taylor model methods. *Reliable Computing*, *5*, 3.

Makino, K., Berz, M. (2003). Taylor models and other validated functional inclusion methods. *International Journal of Pure and Applied Mathematics*, *4*, 379.

Nedialkov, N. S. (1999). *Computing rigorous bounds on the solution of an initial value problems for an ordinary differential equation*. Ph.D. Thesis, University of Toronto, Toronto, Canada.

Nedialkov, N. S., Jackson, K. R., Corliss, G. F. (1999). Validated solutions of initial value problems for ordinary differential equations. *Applied Mathematics and Computation*, *105*, 21.

Nedialkov, N. S., Jackson, K. R., Pryce, J. D. (2001). An effective high-order interval method for validating existence and uniqueness of the solution of an IVP for an ODE. *Reliable Computing*, *7*, 449.

Neher, M., Jackson, K. R., Nedialkov, N. S. (2007). On Taylor model based integration of ODEs. *SIAM Journal on Numerical Analysis*, *45*, 236.

Neumaier, A. (1990). *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge, UK.

Neumaier, A. (2003). Taylor forms—Use and limits. *Reliable Computing*, *9*, 43.

Neumaier, A. (2004). Clouds, Fuzzy Sets, and Probability Intervals. *Reliable Computing*, *10*, 249.

Rauh, A., Hofer, E. P., Auer, E. (2006). VALENCIA-IVP: A Comparison with Other Initial Value Solvers. In: *Proceedings 12th GAMM—IMACS International Symposium on Scientific Computing, Computer Arithmetic, and Validated Numerics (SCAN 2006)*, Duisburg, Germany, 36.