

## VERIFIED SOLUTION METHOD FOR POPULATION EPIDEMIOLOGY MODELS WITH UNCERTAINTY

JOSHUA A. ENSZER, MARK A. STADTHERR

Department of Chemical and Biomolecular Engineering  
University of Notre Dame, Notre Dame, IN 46556, USA  
e-mail: markst@nd.edu

Epidemiological models can be used to study the impact of an infection within a population. These models often involve parameters that are not known with certainty. Using a method for verified solution of nonlinear dynamic models, we can bound the disease trajectories that are possible for given bounds on the uncertain parameters. The method is based on the use of an interval Taylor series to represent dependence on time and the use of Taylor models to represent dependence on uncertain parameters and/or initial conditions. The use of this method in epidemiology is demonstrated using the SIRS model, and other variations of Kermack-McKendrick models, including the case of time-dependent transmission.

**Keywords:** nonlinear dynamics, epidemiology, interval analysis, verified computing, ordinary differential equations.

### 1. Introduction

Ordinary Differential Equations (ODEs) are the basis for many mathematical models in the sciences, including population models used in epidemiology. Specifically, the Kermack-McKendrick model was one of the first developed to simulate the spread of infectious diseases such as bubonic plague and cholera. This and other compartmental models in epidemiology partition the population into classes and describe the rate of population change in each class.

Many variations of the original Kermack-McKendrick model have been described, typically using names based on acronyms of the involved classes. The Kermack-McKendrick model is an SIR (Susceptible, Infected, Recovered) model with a simple type of flux between the three classes. Generalizations of this model are sometimes referred to as SEIRS models. These models incorporate a fourth class (Exposed) within the population, accounting for diseases with an incubation period. These models also account for nonpermanent immunity, thus allowing individuals to again become Susceptible (thus the second S in the acronym). Anderson and May (1979) investigated an SIRS model with various mechanisms for transition between the population classes, but assuming a constant total population (either by assuming no deaths within the population or that the number of births of Susceptible persons was equivalent

to the number of deaths of all population classes). They discussed this use of modeling as it applies to a variety of diseases, including measles, smallpox, and tetanus. Hethcote (1976) investigated a variety of mathematical models whose classes and interactions are a subset of the SEIRS model, including SI, SIS, and SIR variations. Several investigators have focused on a single specific model, computing theoretical bifurcation points as well as some transient and steady-state solutions. This includes work on an SEI model (Pugliese, 1990), an SEIR model (Li *et al.*, 1999), and an SEIS model (Fan *et al.*, 2001). Models can be either closed or open. In a closed model, the population is assumed to remain constant, while models with variable total populations are open. In open models, there are fluxes other than those between the compartments, such as the birth and death of individuals.

In the simplest version of the SEIRS model, the mechanisms for transfer between classes lead to transfer rates (fluxes) that are a function of the population of one class (first-order process) or of two classes (second-order process). For example, the rate of exposure is proportional to the product of the susceptible population and the infected population, and the rate of recovery is proportional to the infected population. However, other nonlinear and constant contact rates have also been considered in the literature (Liu *et al.*, 1986; Greenhalgh, 1997; Dushoff *et al.*, 2004). These models have been studied in both continuous (differential equation) and discrete (Markov

chain) form.

Our focus here is on continuous epidemiological models that are systems of ODEs and are formulated as Initial Value Problems (IVPs). Thus, the model is integrated over time, starting with specified initial values for the different population classes. Some population models are simple enough to be solved analytically, at least for steady-state values. However, in most cases, no analytical solution exists and thus numerical schemes are necessary to obtain the population trajectories of each class. Of interest here is verified (i.e., mathematically and computationally guaranteed) solution of such systems of ODEs, especially systems that involve uncertainty in initial conditions or model parameters. Even in the absence of uncertainty, traditional numerical methods, such as Euler's method or the Runge-Kutta schemes, only approximate the solution of an ODE system, since numerical errors from both discretization and machine arithmetic are present. When there is uncertainty in the initial conditions and/or model parameters, normal use of traditional methods cannot account rigorously for the uncertainties.

Accounting for uncertainties is particularly important in the context of epidemiological models, since in many, if not most, cases, initial populations and model parameters (e.g., rate constants) are not known exactly. We will assume that, for such uncertain quantities, only upper and lower bounds are available. That is, uncertain quantities will be represented by intervals. Since this implies that there are infinitely many possible values for the uncertain quantities, it follows that we have infinitely many possible solutions of the underlying ODE system, corresponding to different values of the uncertain quantities. Therefore, we seek rigorous, verified bounds on all the possible trajectories.

For determining rigorous bounds on the solution of an ODE system, with or without uncertainties, the use of interval methods (also called validated or verified methods) is a natural approach, as computations with intervals, as opposed to floating-point numbers, can provide both mathematically and computationally guaranteed enclosures. Excellent reviews of interval methods for IVPs are available in the literature (Nedialkov *et al.*, 1999; Neher *et al.*, 2007). Much work has been done for the case in which the initial values are given by intervals, and there are several available software packages that deal with this case, including AWA (Lohner, 1992), VN-ODE (Nedialkov *et al.*, 2001), and COSY VI (Berz and Makino, 1998). However, relatively little work has been done on the case in which parameters are given by intervals. In the work described here, we will use a recently developed solver for parametric ODEs (Lin and Stadtherr, 2007) called VSPODE (Verifying Solver for Parametric ODEs), which is used to produce guaranteed bounds on the solutions of nonlinear dynamic systems with interval-valued initial states and parameters. Both

COSY VI and VSPODE use Taylor models (Makino and Berz, 1996; 1999; 2003), though in different ways, to deal with the uncertain quantities (parameters and initial values). In this paper, we propose the use of Taylor-model methods, specifically VSPODE, for propagating uncertainties through nonlinear ODE models in population epidemiology. As examples, we will use several variants of the original Kermack-McKendrick model.

This paper is divided as follows: The next section will provide an overview of the general population epidemiology model used, and a general statement of the ODE problem to be addressed. Section 3 gives background on interval analysis and Taylor models. In Section 4, we outline the specific method that is used, and in Section 5, we present examples and highlight results of applying this method to solve population epidemiology problems.

## 2. Problem statement

In this section, we introduce the notation used to describe the epidemiological models of interest and state the general ODE problem to be solved. In order to maintain a consistent set of variables and parameters, we will outline a general epidemiological population model that encompasses all of the specific models to be used as examples. We adopt the notation of Edelstein-Keshet (2005) for this general model and use it as consistently as possible both here and in Section 5.

We assume that all members of a population belong to a class with respect to a disease: either susceptible, exposed, infected, or recovered. The disease is spread, resulting in increased exposed and/or infected populations, when a susceptible member encounters an infected member. We also assume that all members within a class are identical; this means, for instance, that they have the same probabilities of being infected or of recovering. Models based on these assumptions may be developed on a discrete-time scale, in which the populations can be modeled either stochastically or deterministically (e.g., Allen and Burgin, 2000), or on a continuous-time scale, in which the populations are modeled using deterministic ODEs. We consider only the latter approach here.

An individual can move from one class to another via different processes, each of which is population (or, elsewhere in the literature, population density) dependent. These processes include the following:

1. *Exposure.* This is the process in which individuals exit the susceptible class and enter the exposed class. The exposure rate is given by  $\beta si$ , where  $si$ , the product of the susceptible population  $s$  and infected population  $i$ , represents the frequency at which a susceptible member comes in contact with an infected member, and  $\beta$  represents the probability that this contact

spreads the disease agent. The use of the term  $\beta si$  for the exposure rate is often called simple mass-action incidence (Li *et al.*, 1999) or pseudo mass-action incidence (de Jong *et al.*, 1995).

2. *Infection.* In this process, individuals move from the exposed class to the infected class. This occurs at a rate proportional to the exposed population  $e$ , with a proportionality constant of  $\varepsilon$ , the infection rate constant. Thus, the infection rate is  $\varepsilon e$ .
3. *Recovery.* In this process, members move from the infected class to the recovered class, at a rate proportional to the infected population. Thus, the recovery rate is represented by  $\nu i$ , where the proportionality constant  $\nu$  is the recovery rate constant.
4. *Loss of immunity (susceptibility).* In this process, which implies that immunity is only temporary, individuals move from the recovered class to the susceptible class, with a rate proportional to the recovered population  $r$ , and expressed by  $\gamma r$ , with  $\gamma$  as the proportionality constant.

The population of a class also depends on assumptions made regarding births and deaths and their respective rates. Generally, a base probability of death  $d$  is assumed for all classes, but is incremented to  $d + \alpha$  for the infected class. For this case then, the death rates are  $ds$ ,  $de$ , and  $dr$  for the susceptible, exposed, and recovered classes, respectively, and  $(d + \alpha)i$  for the infected class. New births are almost always assumed to add to the Susceptible class (i.e., no individual is born as exposed, infected, or recovered). A base probability of births  $b$  is generally assigned to all classes, but may be decremented for the exposed and/or infected classes. Assuming the same probability of births for all classes, the total birth rate is  $bn$ , where  $n = s + e + i + r$  is the total population of all classes.

Specific variations of the SEIRS framework are developed by deciding which classes are present, what fluxes exist between the classes, how births and deaths are handled, and whether the system is open or closed. Unsteady population balances can then be performed on each class, based on rate expressions of the type described above, for movement between the classes and for births and deaths. This leads directly to a set of ODEs that describes the rate of change of each population. The specific models for each of the examples considered here will be established in Section 5.

Each continuous SEIRS model, or variation thereof, is representable as a system of ODEs, for which an IVP must be solved. As discussed above, the initial values, as well as the parameters in the ODE model, are often uncertain and thus these quantities will be represented by intervals. In a general mathematical form, this problem may be written as

$$y'(t) = f(y, \theta), \quad y(t_0) = y_0 \in [y_0], \quad \theta \in [\theta], \quad (1)$$

where  $t \in [t_0, t_m]$  for some  $t_m > t_0$ . Here  $y$  is the  $n$ -dimensional vector of state variables with the initial value  $y_0$ , and  $\theta$  is a  $p$ -dimensional vector of time-invariant parameters.  $[y_0]$  and  $[\theta]$  are interval vectors (see Section 3.1) that enclose uncertainties in the initial states and parameters, respectively. We assume that  $f$  is  $(k - 1)$  times continuously differentiable with respect to  $y$  and  $(q + 1)$  times continuously differentiable with respect to  $\theta$ . Here,  $k$  is the order of the truncation error in the interval Taylor series (ITS) method used by VSPODE, and  $q$  is the order of the Taylor model in VSPODE used to represent dependence on parameters and initial values.

When its parameters do not depend on time, the SEIRS model  $f$  is polynomial. However, in practical cases, such as nonlinearly time-dependent parameters (see Section 5.2), this model can become more complicated. Therefore, we will treat  $f$  as a general nonlinear function, which we assume can be represented by (i.e., is a composition of) a finite number of standard functions. If the ODE model is nonautonomous (as in Section 5.2), it can be easily converted to the autonomous form of Eqn. (1) by adding a new state variable that is equal to  $t$ . Our specific goal is to obtain a rigorously guaranteed enclosure of the state variables  $y$  at all times of interest from  $t_0$  to  $t_m$ .

### 3. Background

**3.1. Interval analysis.** The real interval vector  $[x] = [\underline{x}; \bar{x}]$  is an enclosure of the real vector  $x = [x_1, \dots, x_n]^T$ ,  $n \geq 1$ . The real vectors  $\underline{x} = [\underline{x}_1, \dots, \underline{x}_n]^T$  and  $\bar{x} = [\bar{x}_1, \dots, \bar{x}_n]^T$  provide the lower and upper bounds, respectively, on the components of  $x$ . That is,  $\underline{x}_i \leq x_i \leq \bar{x}_i$  or  $x_i \in [\underline{x}_i; \bar{x}_i]$ . An  $n$ -dimensional interval vector can be interpreted geometrically as an  $n$ -dimensional rectangle or box.

Basic arithmetic operations are defined on interval scalars according to  $[x] \circ [y] = \{x \circ y \mid x \in [x], y \in [y]\}$ ,  $\circ \in \{+, -, \times, \div\}$ , with division in the case of  $[y]$  containing zero allowed only in extensions of interval arithmetic (Hansen and Walster, 2004). Addition and multiplication are commutative and associative but only sub-distributive. Interval versions of the elementary functions can also be defined.

For a real function  $f(x)$ , an interval extension  $f^I([x])$  encloses the range of  $f(x)$  for  $x \in [x]$ . That is,  $f^I([x]) \supseteq \{f(x) \mid x \in [x]\}$ . When  $f(x)$  can be written as a series of arithmetic operations and elementary functions, an interval extension can be obtained by substituting  $[x]$  into  $f(x)$  and evaluating using interval arithmetic. In this case,  $f^I([x]) = f([x])$ , which is referred to as the natural interval extension. Computing the interval extension in this way may result in the overestimation of the function range due to the “dependency” problem. While 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. If  $f(x)$  is a single-use expression, in which no variable appears more than once, then the natural interval extension corresponds to the exact function range.

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 this overestimation is propagated from step to step in an integration procedure for ODEs, it can quickly lead to the loss of a meaningful enclosure.

Several good introductions to interval analysis, as well as interval arithmetic and other aspects of computing with intervals, are available (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.

**3.2. Taylor models.** Makino and Berz (1996; 1999) have described a Remainder Differential Algebra (RDA) approach for bounding function ranges and the control of the dependency problem of interval arithmetic. In this method, a function is represented using a model consisting of a 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 to use 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  $\zeta$  with  $0 < \zeta < 1$  such that

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

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  $[x]$  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  $x \in [x]$  and so  $T_f$  encloses the range of  $f$  over  $[x]$ . The function  $f$  can thus be bounded by seeking bounds on the Taylor model  $T_f$ , as described below.

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 given by Makino and Berz

(1996; 1999; 2003), which include addition, multiplication, reciprocal, and intrinsic functions. Using these, 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. Therefore, 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, compared to other rigorous bounding methods, the Taylor model often yields sharper bounds for modest to complicated functionals (Makino and Berz, 1996; 1999; Neumaier, 2003). The uses and limitations of Taylor models are discussed in greater detail by Neumaier (2003).

An interval bound on a Taylor model  $T = (p, [r])$  over  $[x]$  is denoted by  $[T]$  and given by  $[T] = [p] + [r]$ , where  $[p]$  is an interval bound on the polynomial part  $p$ . The range bounding of the interval polynomial  $[p] = p([x] - x_0)$  is an important issue, which directly affects the performance of Taylor model methods. The exact range bounding of an interval polynomial is NP-hard, even for a quadratic function, and direct evaluation using interval arithmetic is very inefficient, often providing only loose bounds, which may negate any benefit of choosing Taylor model methods over traditional interval methods. Alternative bounding schemes are mostly focused on the exact bounding of the first- and second-order terms of  $p$  (Neumaier, 2003), but exact bounding of a general interval quadratic is still NP-hard, and so it can still be computationally expensive. Lin and Stadtherr (2007) use a simple compromise approach in which only the first-order and the *diagonal* second-order terms are considered for exact bounding, and other terms are evaluated directly using interval arithmetic. This is the approach used with the problems considered here.

#### 4. Solution procedure

In this section, we outline the method used for solving the problem described in Section 2. Specifically, it is desired to determine a rigorously verified enclosure of all possible solutions to the IVP expressed in Eqn. (1). We denote by  $y(t; t_j, [y_j], [\theta])$  the set of solutions  $\{y(t; t_j, y_j, \theta) \mid y_j \in [y_j], \theta \in [\theta]\}$ , where  $y_j = y(t_j)$  and  $y(t; t_j, y_j, \theta)$  denotes a solution of  $y'(t) = f(y, \theta)$  for the initial condition  $y = y_j$  at  $t = t_j$ . We will summarize a method for determining enclosures  $[y_j]$  of the state variables at each time step  $j = 1, \dots, m$ , such that  $y(t_j; t_0, [y_0], [\theta]) \subseteq [y_j]$ .

Assume that at  $t_j$  we have an enclosure  $[y_j]$  of  $y(t_j; t_0, [y_0], [\theta])$ , and that we want to carry out an integration step to compute the next enclosure  $[y_{j+1}]$ . Then, in the first phase of the method, the goal is to find a step size  $h_j = t_{j+1} - t_j > 0$  and a rough enclosure  $[\tilde{y}_j]$  of the

solution such that a unique solution  $y(t; t_j, y_j, \theta) \in [\tilde{y}_j]$  is guaranteed to exist for all  $t \in [t_j; t_{j+1}]$ , all  $y_j \in [y_j]$ , and all  $\theta \in [\theta]$ . We apply a traditional interval method, with high order enclosure, to the parametric ODEs by using an Interval Taylor Series (ITS) with respect to time. That is, we determine  $h_j$  and  $[\tilde{y}_j]$  such that for  $[y_j] \subseteq [\tilde{y}_j]^0$ ,

$$[\tilde{y}_j] = \sum_{i=0}^{k-1} [0; h_j]^i f^{(i)}([y_j], [\theta]) + [0; h_j]^k f^{(k)}([\tilde{y}_j]^0, [\theta]) \subseteq [\tilde{y}_j]^0. \quad (3)$$

Here  $k$  denotes the order of the Taylor series,  $[\tilde{y}_j]^0$  is an initial estimate of  $[\tilde{y}_j]$ , and the  $f^{(i)}$  are the Taylor coefficients of  $y(t)$  with respect to time, which can be obtained recursively in terms of  $y'(t) = f(y, \theta)$ . When Eqn. (3) is satisfied, it demonstrates (Corliss and Rihm, 1996) that there exists a unique solution  $y(t; t_j, y_j, \theta) \in [\tilde{y}_j]$  for all  $t \in [t_j; t_{j+1}]$ , all  $y_j \in [y_j]$ , and all  $\theta \in [\theta]$ .

In the second phase of the method, we compute a tighter enclosure  $[y_{j+1}] \subseteq [\tilde{y}_j]$ , such that  $y(t_{j+1}; t_0, [y_0], [\theta]) \subseteq [y_{j+1}]$ . This is done by using an ITS approach to compute  $T_{y_{j+1}}(y_0, \theta)$ , a Taylor model of  $y_{j+1}$  in terms of the initial values  $y_0$  and parameters  $\theta$ , and then obtaining the enclosure  $[y_{j+1}] = [T_{y_{j+1}}]$ . For the Taylor model computations, we begin by representing the interval initial states and parameters by the Taylor models (identity functions)  $T_{y_0}$  and  $T_\theta$ , respectively. Then, we can determine Taylor models  $T_{f^{(i)}}$  of the Taylor series coefficients  $f^{(i)}(y_j, \theta)$  by using RDA operations to compute  $T_{f^{(i)}} = f^{(i)}(T_{y_j}, T_\theta)$ . Using an interval Taylor series for  $y_{j+1}$  with coefficients given by  $T_{f^{(i)}}$ , and using the mean value theorem, one can obtain  $T_{y_{j+1}}(y_0, \theta)$ , the desired Taylor model of  $y_{j+1}$  in terms of the parameters  $\theta$  and initial states  $y_0$ . 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  $T_{y_{j+1}}$  are given by Lin and Stadtherr (2007). An implementation of this approach, called VSPODE (Verifying Solver for Parametric ODEs), was developed and tested by Lin and Stadtherr (2007), who compared its performance with results obtained using the popular VNODE package (Nedialkov *et al.*, 1999; 2001). For the test problems used, VSPODE provided tighter enclosures on the state variables than VNODE and required significantly less computation time. Information about the availability of VSPODE can be obtained by contacting the authors.

## 5. Examples

In this section, we explore variations on the general population model outlined in Section 2. These examples have been explored for real-valued initial conditions and parameters elsewhere in the literature. When possible, we

use similar values. Interval values are used where some initial values or parameters are not reported.

As in Section 2, we adopt the notation and terminology of Edelstein-Keshet (2005) where applicable, so that all models are comparable. VSPODE was used with its default ITS order  $k = 17$  and a default Taylor model order  $q = 5$ . When Monte Carlo simulations are run for purposes of comparison, they are done so in Matlab, using the *ode45* routine with default tolerances, unless otherwise noted.

**5.1. Basic SIRS model.** The basic SIRS model is most similar to the constant-population model first studied by Kermack and McKendrick (1927). In this model, infection is guaranteed and instantaneous after successful exposure, so there is no exposed class as in the general model. The steady-states of this model can be easily found algebraically (Edelstein-Keshet, 2005) and the transient trajectories are relatively simple, so this serves as a good initial test problem for evaluating the performance of VSPODE.

This SIRS model assumes a constant total population  $n = s + i + r$ . Since we can determine the recovered population from  $r = n - s - i$ , we only need unsteady population balances on the susceptible and infected classes. For the susceptible class, this balance is

$$\frac{ds}{dt} = -\beta si + \gamma r = -\beta si + \gamma(n - s - i). \quad (4)$$

Here the first term on the right-hand side is the loss of susceptibles due to infection (flux from the susceptible to the infected class), and the second term is the gain of susceptibles due to loss of immunity (flux from the recovered to the susceptible class). Similarly, for the infected class, the balance is

$$\frac{di}{dt} = \beta si - \nu i, \quad (5)$$

where the second term on the right-hand side represents the flux from the infected to the recovered class.

For this simple example, we have chosen a total population of  $n = 500,000$  individuals (indv), with an initial infected population of  $i_0 = 2000$  indv and initial susceptible population of  $s_0 = 498,000$  indv. We also set the susceptibility rate to be  $\gamma = 50 \text{ yr}^{-1}$ . Uncertain values are assumed for the recovery rate,  $\nu \in [0.125; 0.250] \text{ yr}^{-1}$ , and for the infection probability,  $\beta \in [2; 2.5] \times 10^{-5} \text{ yr}^{-1} \text{ indv}^{-1}$ .

VSPODE was applied to determine a verified enclosure of all possible solutions to this model for  $t = 0$  to  $t = 10 \text{ yr}$ . The results, out to  $t = 2 \text{ yr}$ , are shown for  $s(t)$  in Fig. 1 and for  $i(t)$  in Fig. 2. The curves shown in these figures are upper and lower bounds, which are mathematically and computationally guaranteed, on the possible trajectories of the susceptible and infected populations.

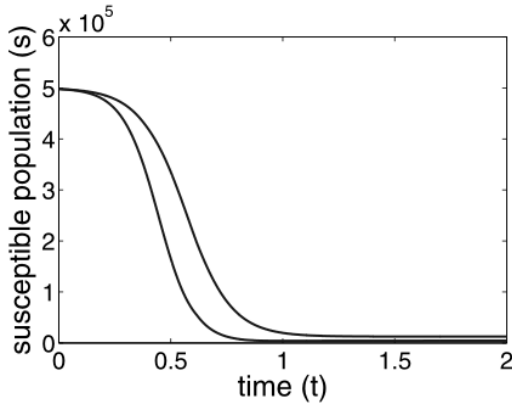


Fig. 1. VSPODE enclosure of the susceptible population trajectory for a simple SIRS model.

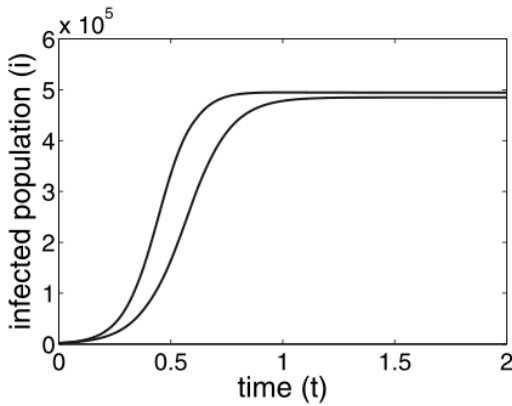


Fig. 2. VSPODE enclosure of the infected population trajectory for a simple SIRS model.

Since interval methods have a reputation of often producing only very loose bounds, we checked the tightness of the VSPODE bounds by comparison to the results of a Monte Carlo simulation with 100,000 trials. For each trial, real values of  $\nu$  and  $\beta$  were selected at random from within their specified interval bounds. Bounds obtained from the Monte Carlo analysis are not guaranteed and in general will 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 Figs. 3 and 4, onto which the VSPODE bounds from Figs. 1 and 2 have been superimposed. 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 population trajectories for this system. This can be seen more quantitatively in Tables 1 and 2, which provide a direct numerical comparison of the bounds obtained from VSPODE and from the Monte Carlo analysis. The true bounds on the trajectories will be between the VSPODE bounds (outer estimate) and the Monte Carlo bounds (inner estimate). The

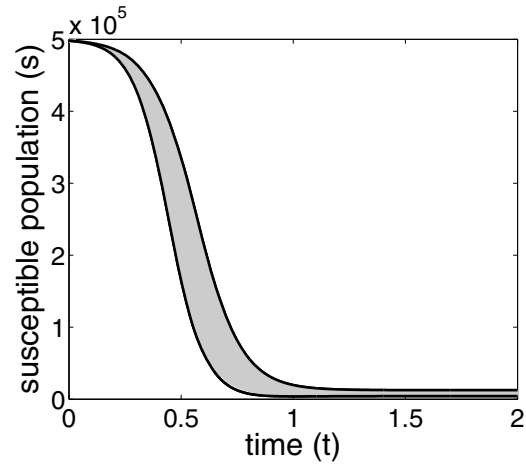


Fig. 3. Monte Carlo simulation (shaded area) and the VSPODE enclosure (solid curve) of susceptible population trajectory for a simple SIRS model.

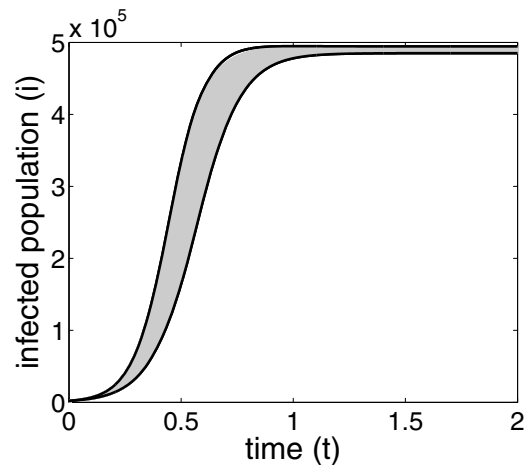


Fig. 4. Monte Carlo simulation (shaded area) and the VSPODE enclosure (solid curve) of the infected population trajectory for a simple SIRS model.

closeness of these two sets of bounds demonstrates that the method used in VSPODE is capable of determining verified bounds that are in fact quite tight. For the final time of  $t = 10$  yr, the VSPODE bounds converge to a solution of  $s \in [4373; 12501]$  and  $i \in [485073; 494389]$ . This numerical result can be compared to exact interval bounds obtained from the analytical steady-state solution,  $s_s = \nu/\beta = [5000; 12500]$ , and  $i_s = (\gamma n - s_s)/(\nu + \gamma) = [485074; 493766]$ . The method employed by VSPODE accurately and tightly bounds the true solution. The Monte Carlo simulation results for  $t = 10$  yr give bounds of  $s \in [5010; 12486]$  and  $i \in [485094; 493757]$ , which are clearly not rigorous bounds on the true solution. For VSPODE, the computation time required was 17.9 s, and for 100,000 Monte Carlo trials with *ode45* in Matlab the computation time was 1820 s (both times on an Intel Pen-

Table 1. Numerical comparison of the VSPODE enclosure and the Monte Carlo simulation (MC) for the susceptible population  $s$  in a simple SIRS model.

$t$	$\underline{s}$ (VSPODE)	$\underline{s}$ (MC)	$\bar{s}$ (MC)	$\bar{s}$ (VSPODE)
0.2	477193	477197	485712	486323
0.4	318703	319584	417148	419180
0.6	63839	68319	212357	216113
0.8	7769	11062	56262	57489
1.0	3988	5528	19204	19673
1.2	4194	5053	13459	13590
1.4	4324	5013	12631	12670
1.6	4361	5010	12515	12528
1.8	4370	5010	12494	12505
2.0	4372	5009	12496	12502

Table 2. Numerical comparison of the VSPODE enclosure and the Monte Carlo simulation (MC) for the infected population  $i$  in a simple SIRS model.

$t$	$\underline{i}$ (VSPODE)	$\underline{i}$ (MC)	$\bar{i}$ (MC)	$\bar{i}$ (VSPODE)
0.2	13612	13728	22751	22761
0.4	80385	82448	180048	180898
0.6	282559	286360	430625	434961
0.8	440344	441604	487722	490957
1.0	477942	478411	493240	494764
1.2	483989	484123	493713	494565
1.4	484904	484945	493753	494436
1.6	485045	485062	493756	494400
1.8	485068	485079	493757	494391
2.0	485072	485080	493757	494389

tium 4 3.2GHz workstation running Red Hat Linux).

Physically interpreting the results of this model, it is clear that an epidemic is sustained in this population. The large value of  $i$  at long times is due to the small probability of recovery compared to the fast rate of susceptibility. At large  $i$ , the term  $\beta si$  dominates the dynamics of this system, maintaining a large value of  $i$  while suppressing the populations of the  $s$  and  $r$  classes.

**5.2. SIRS model with time-dependent parameter.**

Another use of the SIRS model assumes that the probability of infection  $\beta$  is time-variant, according to an expression such as

$$\beta(t) = \beta_0(1 + \frac{\beta_1}{n} \cos(2\pi t)). \tag{6}$$

This model has been used to simulate an illness like influenza, which is known to exhibit such “seasonal forcing” (Dushoff *et al.*, 2004).

For this example, we adopt parameter values consistent with the study of Dushoff *et al.* (2004). These values are  $\nu = 50 \text{ yr}^{-1}$ ,  $\gamma = 0.125 \text{ yr}^{-1}$ ,  $n = 500,000$

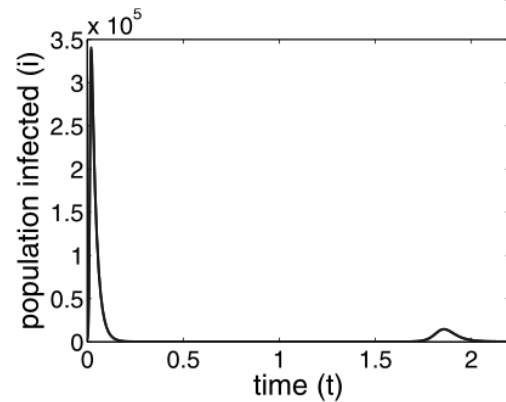


Fig. 5. VSPODE enclosure of the infected population trajectory for an SIRS model with seasonal forcing, showing initial transients.

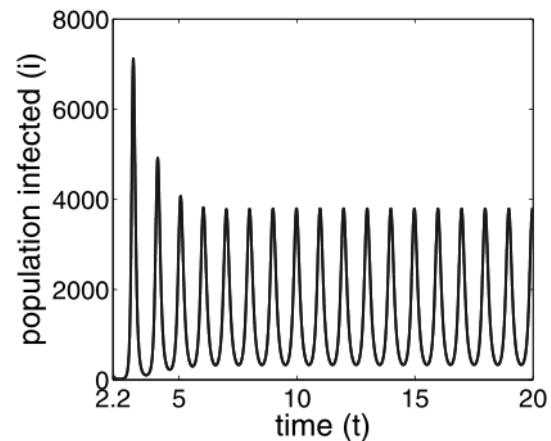


Fig. 6. VSPODE enclosure of the infected population trajectory for an SIRS model with seasonal forcing, for  $t \geq 2.2$  yr.

indv,  $\beta_0 = 400 \text{ yr}^{-1}$ , and  $\beta_1 = 0.04$  (Dushoff *et al.* (2004) mistakenly report using  $\beta_1 = 0.02$ ). For initial conditions, we assume an initial outbreak in the interval  $i_0 \in [1000; 4000]$  indv and  $[s_0] = n - [i_0]$ . The results of Dushoff *et al.* (2004) suggest that any initial condition in this range should converge to the same limit cycle. Thus, we test here whether the VSPODE bounds contract from the initial bounds to a tight bound on the limit cycle.

When the time-dependent expression for  $\beta$ , Eqn. (6), is substituted into the basic SIRS model, Eqns. (4) and (5), the result is a nonautonomous model. Prior to the use of VSPODE, this is converted into the autonomous form of Eqn. (1), by defining  $t$  as a new state variable with a derivative of one. VSPODE was then used to compute verified bounds on the population trajectories for this model from  $t = 0$  to  $t = 20$  yr.

The initial transient results for  $i(t)$  out to  $t = 2.2$  yr are shown in Fig. 5. On the scale of this figure, the upper and lower trajectory bounds cannot be distinguished. There is a sharp peak in the number of infections almost

instantly in this simulation, with a maximum value of around  $i = 340,000$  indiv at time  $t = 0.02$  yr. At this point, the diameter of the interval bounding  $i$  is about 20,000 indiv, which is about a six-fold increase over the interval uncertainty on the initial condition  $i_0$ . However, by a time of about  $t = 0.33$  yr, the diameter of the interval bounding  $i$  has shrunk to 1 or less, where it remains for larger values of time. Since  $i$  is in units of individuals, this means that we have essentially zero-width bounds for  $t$  larger than about 0.33 yr. Thus, while the uncertainty in the initial conditions affects the magnitude of the initial peak in infections, the trajectories very quickly converge to a solution that is independent of the initial states, and this behavior has been captured by VSPODE. For  $t$  between 2.2 and 20 yr, the results are shown (with a different scale for  $i$ ) in Fig. 6. Since the bounds on  $i$  are essentially zero-width, they are not distinguishable in the figure. These results are consistent with the results of Dushoff *et al.* (2004), who show only the limit cycle behavior on the time interval  $t \in [10; 20]$  yr and fairly accurately track the data associated with a strain of influenza.

**5.3. SEI model with a variable total population.** The SEI model (Pugliese, 1990) in theoretical epidemiology assumes that recovery from illness is impossible, so there is no recovered class to consider. However, the model remains dynamic because the total population  $n = s + e + i$  is not assumed to be constant. In this model, the total population may increase through births of new individuals, with different birth rates for each of the population classes. The different birth rates are represented by using a base birth rate constant  $b$ , and then decrementing this by “penalty” parameters  $\delta_1$  for the exposed class and  $\delta_2$  for the infected class. It is also assumed that the infected class has a death rate that is higher than the death rate of the other classes, with this additional rate represented by an increment  $\alpha$  in the base death rate constant  $d$ .

In these terms, the SEI model is given by

$$\frac{ds}{dt} = (b - d)s + b(1 - \delta_1)e + b(1 - \delta_2)i - \beta si, \quad (7)$$

$$\frac{de}{dt} = \beta si - (d + \varepsilon)e, \quad (8)$$

$$\frac{di}{dt} = \varepsilon e - (d + \alpha)i. \quad (9)$$

Following Pugliese (1990), and taking days as the time unit, we set the model parameters as  $b = 0.15 \text{ day}^{-1}$ ,  $d = 0.1 \text{ day}^{-1}$ ,  $\delta_1 = 0$ ,  $\delta_2 = 0.9 \text{ day}^{-1}$ ,  $\beta = 0.025 \text{ day}^{-1} \text{ indiv}^{-1}$ ,  $\varepsilon = 1 \text{ day}^{-1}$ , and  $\alpha = 0.1 \text{ day}^{-1}$ . Note that, since  $\delta_1 = 0$ , there is a birth rate penalty only for infected individuals. The initial states are taken to be  $e_0 = 0$ ,  $i_0 = 10,000$ ,  $s_0 \in [480000; 490000]$ .

The trajectory bounds of the infected class for the first 30 days, as determined by VSPODE, are shown in

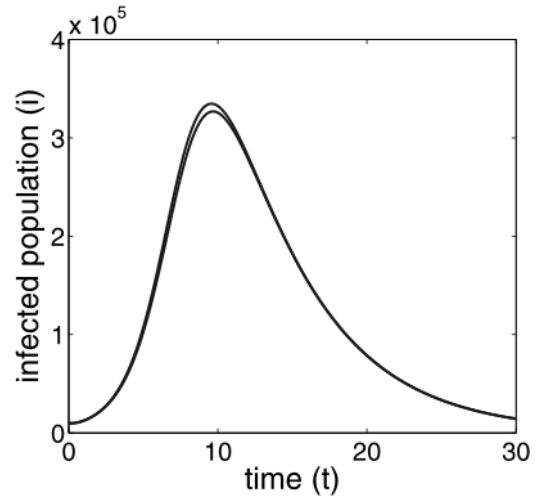


Fig. 7. VSPODE enclosure of the susceptible population trajectory of an SEI model.

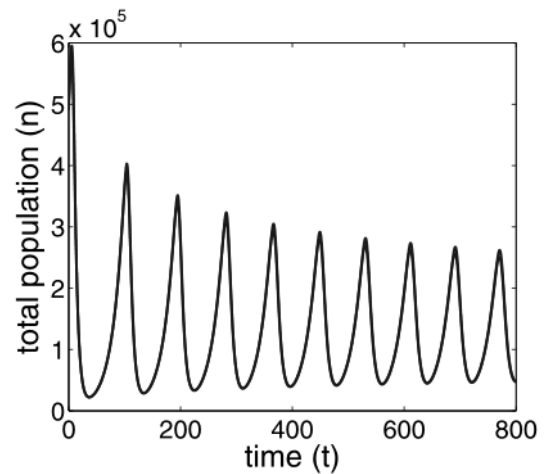


Fig. 8. VSPODE enclosure of the total population trajectory of an SEI model.

Fig. 7. This shows that the uncertainty in the initial population of susceptibles is propagated into uncertainty in the infected population, and that this uncertainty has been bounded by VSPODE. The longer term behavior of this system is cyclic, as shown in Fig. 8, out to  $t = 800$  days, which is consistent with the results of Pugliese (1990). For this problem, the phase shift of the cycles depends on the initial state, so zero-width bounds, as observed in the previous example, cannot be expected here. On the scale of Fig. 8, the VSPODE bounds are not distinguishable and appear to be quite tight. However, comparison with a Monte Carlo analysis shows that, for larger values of time, the VSPODE bounds get increasingly worse at the extremes in the trajectories. This behavior is shown quantitatively in Table 3. Eventually, beyond about  $t = 950$  days, VSPODE fails to obtain meaningful bounds. To obtain bounds for larger values of  $t$ , the initial interval on



Table 3. Numerical comparison of the VSPODE enclosure and the Monte Carlo simulation (MC) of the total population for an SEI model near peaks of each oscillation.

$t$	$\underline{n}$ (VSPODE)	$\underline{n}$ (MC)	$\bar{n}$ (MC)	$\bar{n}$ (VSPODE)
104.88	386332	392070	394594	400326
196.00	332511	339078	341463	348021
283.10	304527	310961	313300	319723
367.74	286613	292756	295064	301193
450.66	274015	279870	282127	287968
532.33	264497	270088	272308	277884
613.03	257015	262387	264575	269931
692.96	250895	256136	258294	263521
772.29	245420	250804	252971	258339
851.22	238813	245638	247947	254756
932.13	196818	221081	225851	250080

$s_0$  could be subdivided, with VSPODE, then run on each subinterval, and the results combined. Of course, this “subinterval reconstitution” procedure will significantly increase the computational expense.

**5.4. SEIR model with variable population size.** This theoretical model (Li *et al.*, 1999) is similar to the previous example, except that recovery is possible, with recovery providing permanent immunity. Thus, there is now a recovered population class, and there is no flux from the recovered class to the susceptible class. The only growth in the susceptible class is from new births, with the birth rate assumed to be the same for all classes. In this particular SEIR model, another important feature is that the transmission probability  $\beta$  is taken to be inversely proportional to the total population. Thus,  $\beta = \sigma/n$  with  $n = s + e + i + r$  and  $\sigma$  a proportionality constant.

For this situation, the model equations (Li *et al.*, 1999) are

$$\frac{ds}{dt} = bn - ds - \sigma si/n, \tag{10}$$

$$\frac{de}{dt} = \sigma si/n - (\varepsilon + d)e, \tag{11}$$

$$\frac{di}{dt} = \varepsilon e - (\nu + \alpha + d)i, \tag{12}$$

$$\frac{dn}{dt} = (b - d)n - \alpha i. \tag{13}$$

Note that the final equation is a balance on the total population. Alternatively, a balance on the recovered population could be used. We use the parameter values given by Li *et al.* (1999) and assume years as the time unit. These values are  $b = 0.5 \text{ yr}^{-1}$ ,  $d = 0.18 \text{ yr}^{-1}$ ,  $\sigma = 20 \text{ yr}^{-1}$ ,  $\varepsilon = 4 \text{ yr}^{-1}$ ,  $\nu = 1.5 \text{ yr}^{-1}$ , and  $\alpha = 6 \text{ yr}^{-1}$ . For the initial state, we choose a total population of  $n_0 = 500,000$  indiv with 10% already infected, so  $i_0 = 50,000$  indiv.

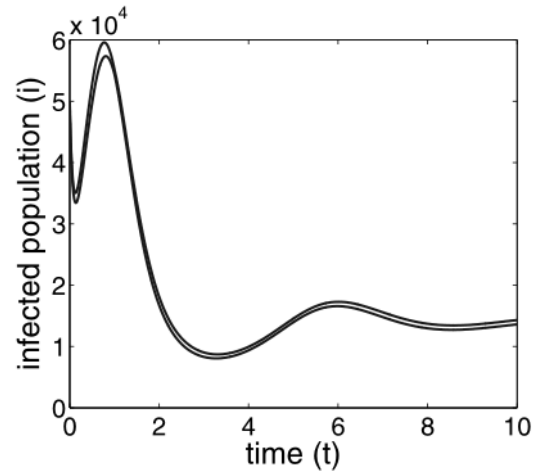


Fig. 9. VSPODE enclosure of the infected population trajectory for an SEIR model.

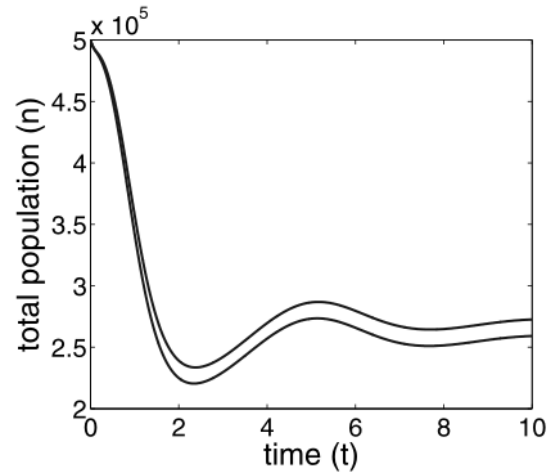


Fig. 10. VSPODE enclosure of the total population trajectory for an SEIR model.

The initial populations of the other classes are uncertain and assumed to be  $s_0 \in [400000; 405000]$  indiv and  $e_0 \in [10000; 15000]$  indiv.

VSPODE was applied to determine a verified enclosure of all possible population trajectories for this model for  $t = 0$  to  $t = 10$  yr. The results for  $i(t)$  and  $n(t)$  are shown in Figs. 9 and 10, respectively. The curves shown are mathematically and computationally guaranteed upper and lower bounds on the infected and total populations as a function of time.

Again, we checked the tightness of the VSPODE bounds by comparison to the results of a Monte Carlo analysis (100,000 trials). For this problem, we could obtain accurate trajectories from the Matlab *ode45* routine only by using fairly strict tolerances (relative tolerance of  $1 \times 10^{-5}$  and absolute tolerance of  $1 \times 10^{-8}$ ). The Monte Carlo simulation results are shown in Figs. 11 and 12, onto which the VSPODE results from Figs. 9 and 10 have

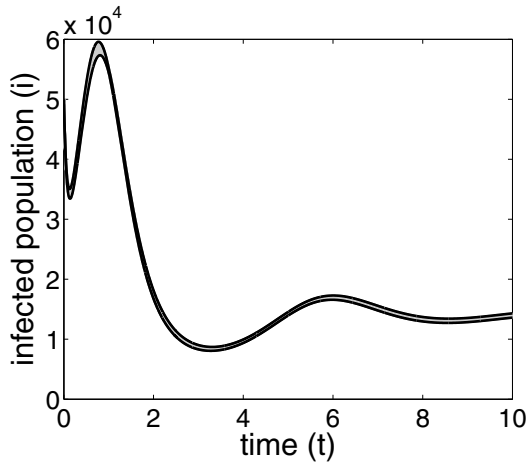


Fig. 11. Monte Carlo simulation (shaded area) and VSPODE enclosure (solid curve) of the infected population trajectory for an SEIR model.

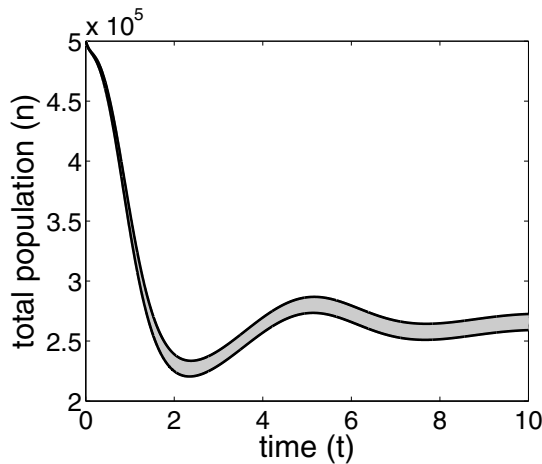


Fig. 12. Monte Carlo simulation (shaded area) and VSPODE enclosure (solid curve) of the total population trajectory for an SEIR model.

been superimposed. It is clear that VSPODE provides very tight bounds on the possible population trajectories for this system. A more quantitative comparison is provided by Tables 4 and 5, in which there is a direct numerical comparison of the bounds obtained from VSPODE and from the Monte Carlo analysis. It should be emphasized that the Monte Carlo analysis does not provide true bounds, only an inner estimate of the bounds. VSPODE provides rigorous bounds that tightly bound the true solution in this case.

The results shown in Figs. 9 and 10 are consistent with those of Li *et al.* (1999), who discuss the numerical conditions necessary to sustain an epidemic in detail. In this simulation, the epidemic is sustained within a small fraction of the total population. As the simulation continues past  $t = 10$  yr, the population bounds approach their expected numerical steady-state

Table 4. Numerical comparison of the VSPODE enclosure and the Monte Carlo simulation (MC) of the infected population of an SEIR model.

$t$	$\underline{i}$ (VSPODE)	$\underline{i}$ (MC)	$\bar{i}$ (MC)	$\bar{i}$ (VSPODE)
1.0	54447	54448	55188	55203
2.0	16743	16753	18112	18115
3.0	8306	8312	9026	9028
4.0	9374	9375	9950	9951
5.0	13949	13950	14541	14545
6.0	16573	16574	17281	17285
7.0	14789	14790	15571	15573
8.0	12967	12970	13707	13709
9.0	12846	12850	13545	13549
10.0	13599	13608	14300	14310

Table 5. Numerical comparison of the VSPODE enclosure and the Monte Carlo simulation (MC) of the total population of an SEIR model.

$t$	$\underline{n}$ (VSPODE)	$\underline{n}$ (MC)	$\bar{n}$ (MC)	$\bar{n}$ (VSPODE)
1.0	341797	341833	357282	357316
2.0	224886	224949	239007	239033
3.0	229684	229703	242026	242046
4.0	256999	257008	269565	269599
5.0	273219	273229	286461	286496
6.0	265838	265854	279566	279590
7.0	253450	253474	267028	267054
8.0	251486	251509	264779	264807
9.0	255898	255928	269158	269198
10.0	259120	259178	272517	272586

values for this theoretical model.

## 6. Concluding remarks

Nonlinear ODE models in population epidemiology often involve uncertainty in the parameters related to disease transmission or in the initial states of the populations. We have demonstrated here the use of a recently developed interval method (Lin and Stadtherr, 2007) for determining mathematically and computationally guaranteed bounds on the population trajectories that are possible for given bounds on the uncertain quantities. Using a Monte Carlo analysis, it was also shown that it is possible for these bounds to be quite tight.

It has been assumed here that the uncertainties in parameters and initial conditions are represented by intervals. This means that there is no information provided about the probability distribution of the uncertain values, only their upper and lower bounds. In subsequent work, we will consider the case of epidemiological models in which there are probability distributions for the uncertainty and show how this can be used to obtain probabilistic bounds on the population trajectories.

## Acknowledgments

This work was supported in part by a Lilly Foundation graduate fellowship (JAE).

## References

- Allen, L. J. S. and Burgin, A. M. (2000). Comparison of deterministic and stochastic SIS and SIR models in discrete time, *Mathematical Biosciences* **163**(1): 1–33.
- Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: Part 1, *Nature* **280**(5721): 361–367.
- Berz, M. and Makino, K. (1998). Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models, *Reliable Computing* **4**(4): 361–369.
- Corliss, G. F. and Rihm, R. (1996). Validating an a priori enclosure using high-order Taylor series, in G. Alefeld, A. Frommer and B. Lang (Eds.), *Scientific Computing and Validated Numerics*, Akademie Verlag, Berlin, pp. 228–238.
- de Jong, M. C. M., Diekmann, O. and Heesterbeek, H. (1995). How does transmission of infection depend on population size?, in D. Mollison (Ed.), *Epidemic Models: Their Structure and Relation to Data*, Cambridge University Press, Cambridge, pp. 84–94.
- Dushoff, J., Plotkin, J. B., Levin, S. A. and Earn, D. J. D. (2004). Dynamical resonance can account for seasonality of influenza epidemics, *Proceedings of the National Academy of Sciences* **101**(48): 16915–16916.
- Edelstein-Keshet, L. (2005). *Mathematical Models in Biology*, SIAM, Philadelphia, PA.
- Fan, M., Li, M. Y. and Wang, K. (2001). Global stability of an SEIS epidemic model with recruitment and a varying total population size, *Mathematical Biosciences* **170**(2): 199–208.
- Greenhalgh, D. (1997). Hopf bifurcation in epidemic models with a latent period and nonpermanent immunity, *Mathematical and Computer Modelling* **25**(2): 85–107.
- Hansen, E. R. and Walster, G. W. (2004). *Global Optimization Using Interval Analysis*, Marcel Dekker, New York, NY.
- Hethcote, H. W. (1976). Qualitative analysis of communicable disease models, *Mathematical Biosciences* **28**(4): 335–356.
- Jaulin, L., Kieffer, M., Didrit, O. and Walter, É. (2001). *Applied Interval Analysis*, Springer-Verlag, London.
- Kearfott, R. B. (1996). *Rigorous Global Search: Continuous Problems*, Kluwer, Dordrecht.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London, Part A* **115**(772): 700–721.
- Li, M. Y., Graef, J. R., Wand, L. and Karsai, J. (1999). Global dynamics of a SEIR model with varying total population size, *Mathematical Biosciences* **160**(2): 191–215.
- Lin, Y. and Stadtherr, M. A. (2007). Validated solutions of initial value problems for parametric ODEs, *Applied Numerical Mathematics* **57**(10): 1145–1162.
- Liu, W., Levin, S. A. and Iwasa, Y. (1986). Influence of non-linear incidence rates upon the behavior of SIRS epidemiological models, *Journal of Mathematical Biology* **23**(2): 187–204.
- Lohner, R. J. (1992). Computations of guaranteed enclosures for the solutions of ordinary initial and boundary value problems, in J. Cash and I. Gladwell (Eds.), *Computational Ordinary Differential Equations*, Clarendon Press, Oxford, pp. 425–435.
- Makino, K. and Berz, M. (1996). Remainder differential algebras and their applications, in M. Berz, C. Bischof, G. Corliss and A. Griewank (Eds.), *Computational Differentiation: Techniques, Applications, and Tools*, SIAM, Philadelphia, PA, pp. 63–74.
- Makino, K. and Berz, M. (1999). Efficient control of the dependency problem based on Taylor model methods, *Reliable Computing* **5**(1): 3–12.
- Makino, K. and Berz, M. (2003). Taylor models and other validated functional inclusion methods, *International Journal of Pure and Applied Mathematics* **4**(4): 379–456.
- Nedialkov, N. S., Jackson, K. R. and Corliss, G. F. (1999). Validated solutions of initial value problems for ordinary differential equations, *Applied Mathematics and Computation* **105**(1): 21–68.
- Nedialkov, N. S., Jackson, K. R. and 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**(6): 449–465.
- Neher, M., Jackson, K. R. and Nedialkov, N. S. (2007). On Taylor model based integration of ODEs, *SIAM Journal on Numerical Analysis* **45**(1): 236–262.
- Neumaier, A. (1990). *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge.
- Neumaier, A. (2003). Taylor forms—Use and limits, *Reliable Computing* **9**(1): 43–79.
- Pugliese, A. (1990). An SEI epidemic model with varying population size, in S. Busenberg and M. Martelli (Eds.), *Differential Equations Models in Biology, Epidemiology and Ecology*, Lecture Notes in Computer Science, Vol. 92, Springer, Berlin, pp. 121–138.



**Joshua A. Enszer** is a Ph.D. candidate at the University of Notre Dame, at the Department of Chemical and Biomolecular Engineering. He holds a B.Sc. in chemical engineering from Michigan Technological University and an M.Sc. in chemical engineering from the University of Notre Dame. His current research includes verified solution of uncertain dynamic systems, with applications in chemical engineering and life sciences.



**Mark A. Stadtherr** is Keating-Crawford Professor of Chemical and Biomolecular Engineering at the University of Notre Dame. He holds a B.Ch.E. degree from the University of Minnesota and a Ph.D. degree from the University of Wisconsin. He was awarded the 1998 Computing in Chemical Engineering Award by the American Institute of Chemical Engineers. The focus of his research is primarily on the development and application of strategies for reliable

engineering computing, with particular interests in global optimization, verified solution of dynamic systems, the computation of phase behavior, and ecological modeling.

Received: 18 September 2008

Revised: 9 February 2009