Verified Probability Bounds Analysis Around Bifurcations in an Ecosystem Model

Joshua A. Enszer, Kate A. Smith, and Mark A. Stadtherr

Department of Chemical and Biomolecular Engineering
University of Notre Dame, Notre Dame, IN 46556, USA

14th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics, ENS Lyon, France, September 27–30, 2010

Overview

- Overall project — Rigorous propagation of uncertainty in nonlinear models of ecological dynamics
  - Interval uncertainty (upper and lower bounds only)
  - Probability box (p-box) uncertainty (upper and lower bounds on the cumulative distribution function)
- Ecological models
  - May be considerable uncertainty in parameters and/or models
  - Nonlinear dynamics provides a rich variety of possible behaviors
  - Collaboration with Notre Dame Ecology Program — Experimental and modeling studies of new chemicals in aquatic ecosystems

Overview (cont'd)

- Earlier work — Rigorous propagation of interval and p-box uncertainties in ecological models can be done using verified integration and Taylor models (Enszer et al., 2009)
- Today's presentation — Consider cases for which multiple Taylor models (over different parameter subdomains) may be needed
  - State bounds diverge due to accumulation of overestimation errors (dependency, wrapping)
  - Divergence of state bounds is a feature of the model
    - Bifurcation enclosed by parameter interval
    - Separatrix enclosed by initial state interval
  - How to do arithmetic with p-boxes split over multiple subdomains and multiple Taylor models

Example 1: Aquatic Food Web Model

- Consider the aquatic food web model (Kulacki et al., 2008) consisting of the species Chlamydomonas, Daphnia, and Zebra Mussels

$$\frac{dC}{dt} = C\left[r_C\left(1 - \frac{C}{K_C}\right) - \frac{a_{CD}D}{b_D + C} - a_{CZ}Z\right]$$

$$\frac{dD}{dt} = D\left[\frac{a_{DC}C}{b_D + C}\left(1 - \frac{D}{K_D}\right) - d_D\right]$$

$$\frac{dZ}{dt} = Z\left[a_{ZC}C - d_Z\right]$$

- Values of the growth rate r_C and the interaction parameter a_{CD} are uncertain

Example 1: Aquatic Food Web Model

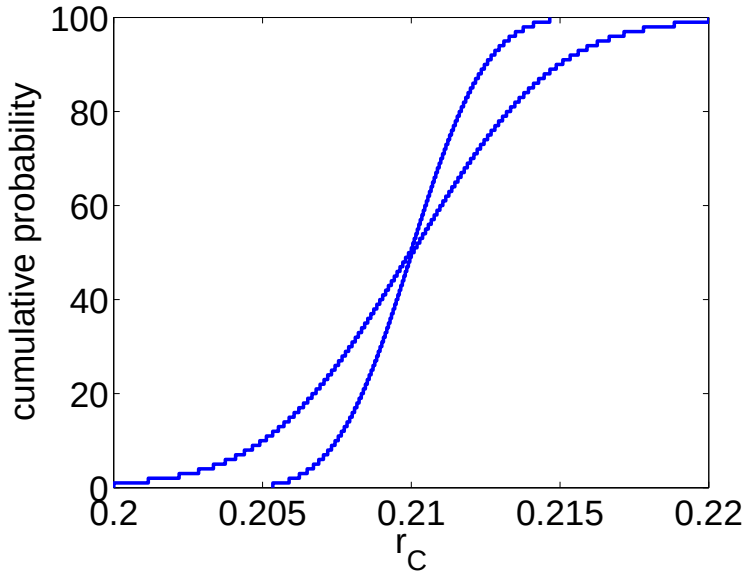
- Parameters and initial conditions:

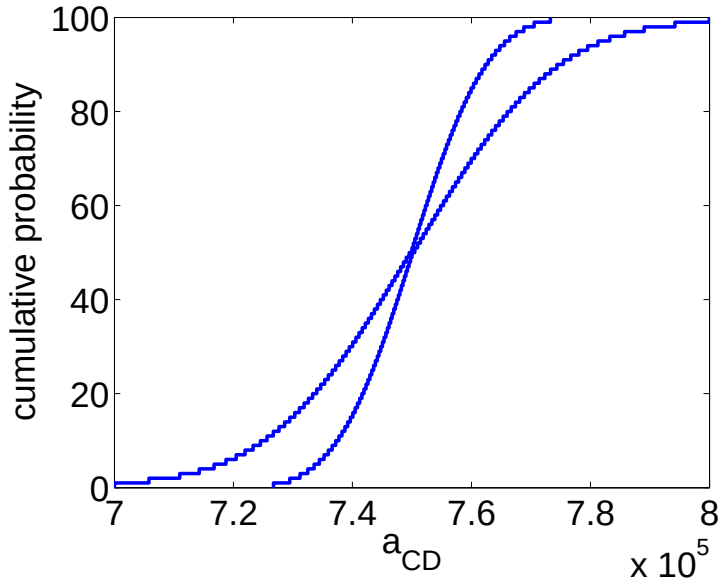
	Value	Units		Value	Units
b_D	8.82×10^6	indv			
K_C	5.15×10^{17}	indv	a_{ZC}	3.78×10^{-8}	day^{-1}
a_{CZ}	7.47×10^{-11}	day^{-1}	a_{DC}	0.125	day^{-1}
d_D	0.016	day^{-1}	d_Z	0.001	day^{-1}
K_D	3.25×10^{10}	indv	$C(0)$	5.15×10^{16}	indv
$D(0)$	1.25×10^9	indv	$Z(0)$	8.9×10^8	indv
r_C	$[0.40, 0.45]$	day^{-1}	a_{CD}	$[4.0, 5.0] \times 10^6$	day^{-1}

- Relatively large uncertainties: 12.5% for r_C and 25% for a_{CD}

Example 1: Aquatic Food Web Model

- P-box representation of uncertain parameters





- Cumulative probability functions for the uncertain parameters are bounded by truncated normal distributions

General Problem Statement

- Consider initial value problem (IVP) for nonlinear ODE problem,

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

  in which at least one of the initial states in $\boldsymbol{y}_0$ or one of the time-invariant parameters in $\boldsymbol{\theta}$ is uncertain (contained in $\boldsymbol{Y}_0$ and/or $\boldsymbol{\Theta}$)

- There is information about the distribution of the uncertainty that can be represented by p-boxes

- Goal 1: Obtain a rigorous, verified enclosure of all possible solutions to this uncertain, parametric IVP to a time horizon t_{f} of interest

- Goal 2: Obtain rigorous, verified bounds (p-boxes) on the probability distribution of the states at times of interest

Solution Procedure – Goal 1

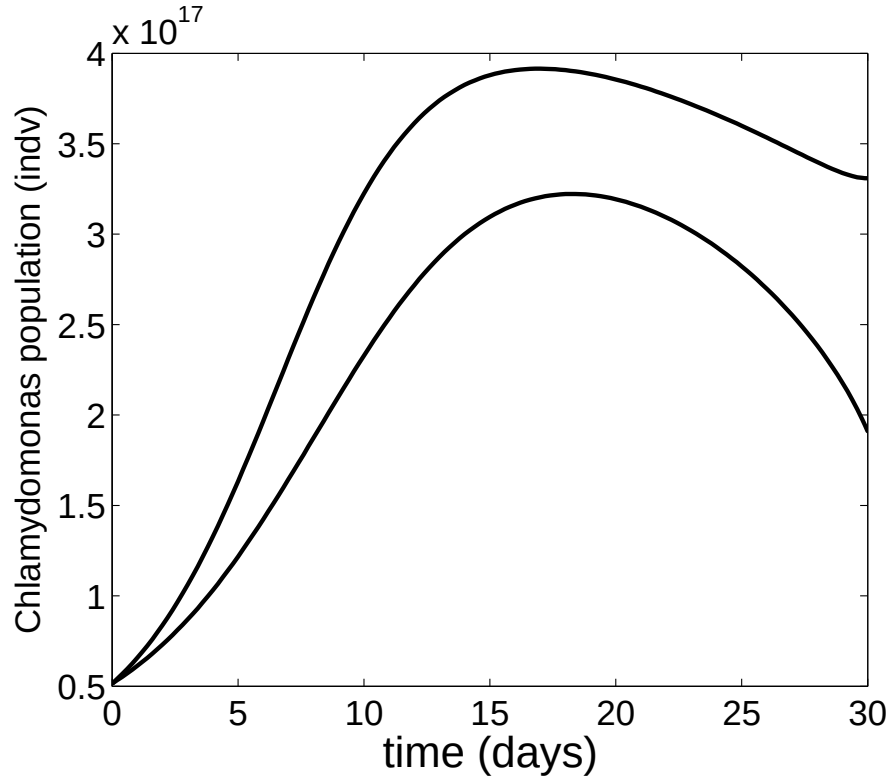
- Goal 1: Obtain a rigorous, verified enclosure of all possible solutions to this uncertain, parameter IVP over a time horizon of interest
- Use a method for verified (validated) solution of IVP
  - Guarantees there exists a unique solution $\boldsymbol{y}(t)$ for interval $t \in [t_0, t_\mathrm{f}]$, for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$
  - At time step j, computes an interval $\boldsymbol{Y}_j$ that encloses all solutions $\boldsymbol{y}_j$ of the ODE system at t_j for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$
- Tools are available – AWA, VNODE, COSY VI, ValEncIA-IVP, VSPODE, etc.

Summary of VSPODE

- Use interval Taylor series to represent dependence on time
- Use Taylor model $\boldsymbol{T}_{\boldsymbol{y}_j} = \boldsymbol{T}_{\boldsymbol{y}_j}(\boldsymbol{y}_0, \boldsymbol{\theta})$ to represent dependence on the initial states $\boldsymbol{y}_0$ and parameters $\boldsymbol{\theta}$; consists of
  - A real-valued polynomial function of $\boldsymbol{y}_0$ and $\boldsymbol{\theta}$
  - A parallelepiped remainder bound
- Assuming $\boldsymbol{Y}_j$ is known, then
  - Phase 1: Compute a coarse enclosure $\widetilde{\boldsymbol{Y}}_j$ and prove existence and uniqueness using fixed point iteration with Picard operator and high-order interval Taylor series (as in VNODE)
  - Phase 2: Refine the coarse enclosure to obtain $\boldsymbol{Y}_{j+1}$ using Taylor models in terms of the uncertain parameters and initial states
- Implemented by Lin and Stadtherr (2007)

Example 1: Aquatic Food Web Model

- VSPODE enclosure of Chlamy population C over $t \in [0, 30]$ based on interval uncertainties in r_C and a_{CD}



- Bounds diverge past $t = 30$
- To extend integration time, can use some bisection strategy, e.g. bisect the parameter intervals

Solution Procedure – Goal 2

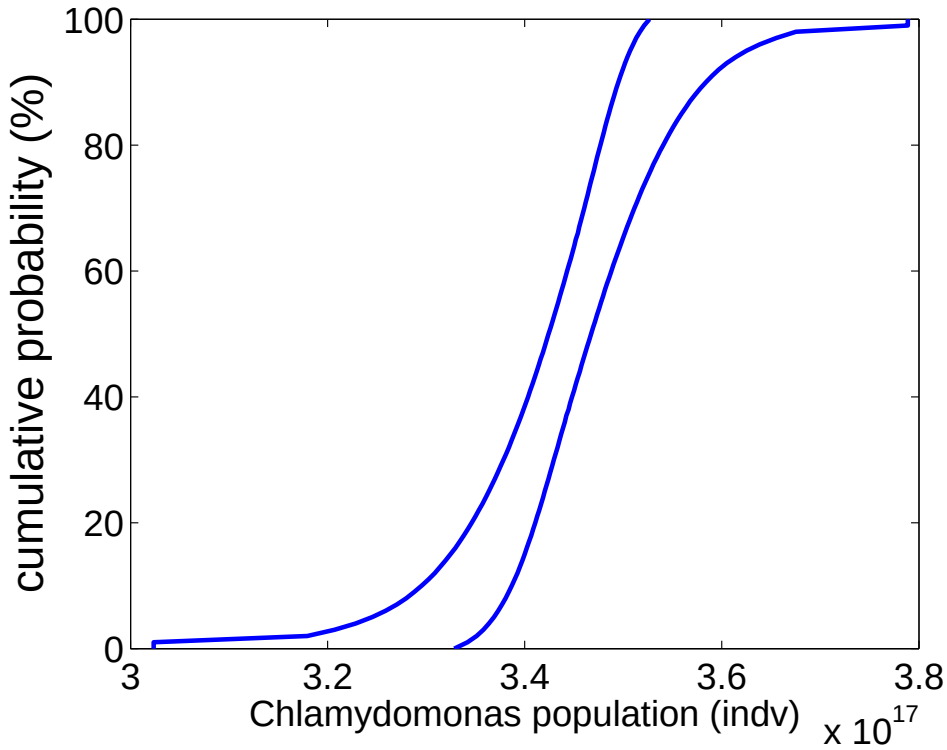
- Goal 2: Obtain rigorous, verified bounds (p-boxes) on the probability distribution of the states at times of interest
- For a time of interest t_j (end of j-th time step), VSPODE has computed a Taylor model representation $\boldsymbol{T}_{\boldsymbol{y}_j} = \boldsymbol{T}_{\boldsymbol{y}_j}(\boldsymbol{y}_0, \boldsymbol{\theta})$ of the state variables as a function of the initial states $\boldsymbol{y}_0$ and parameters $\boldsymbol{\theta}$
- This Taylor model is valid for all $\boldsymbol{y}_0 \in \boldsymbol{Y}_0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- Substitute distributions (p-boxes) for $\boldsymbol{y}_0$ and $\boldsymbol{\theta}$ into $\boldsymbol{T}_{\boldsymbol{y}_j} = \boldsymbol{T}_{\boldsymbol{y}_j}(\boldsymbol{y}_0, \boldsymbol{\theta})$ and use p-box arithmetic to compute p-boxes of state variables $\boldsymbol{y}_j = \boldsymbol{y}(t_j)$

P-box Arithmetic on a Taylor Model (TM)

- For p-box arithmetic, the p-boxes are typically discretized into a set of intervals corresponding to different probability levels
- Each interval is then substituted into the TM $\boldsymbol{T}_{\boldsymbol{y}_j}$ and interval arithmetic is used to evaluate the TM
- The results of each interval TM evaluation are collected and assembled into a final p-box for $\boldsymbol{y}_j$
- Because interval arithmetic is used, there are dependency and wrapping issues that can result in overestimation
- If regular interval arithmetic results in too much overestimation, then for each interval used to represent the p-box, subdivide it, evaluate the TM over each subinterval and take the union of the results (subinterval reconstitution) (Ferson and Hajagos, 2004)
- Then the (tighter) results of the interval TM evaluations are collected into the output p-box

Example 1: Aquatic Food Web Model

- P-box enclosure of probability distribution for C at time $t = 15$, based on p-boxes for r_C and a_{CD}



- For beyond $t = 30$, need to bisect the parameter intervals

- Need for p-box arithmetic on multiple Taylor models

Example 2: Tritrophic Rosenzweig-MacArthur (RM) Model

- Model of food chain with three trophic levels - basal (prey), predator, and superpredator
- Logistic prey growth
- Holling type II (hyperbolic) predator (and superpredator) response functions

$$\frac{dx_1}{dt} = x_1 \left[r(1 - \frac{x_1}{K}) - \frac{a_2 x_2}{b_2 + x_1} \right]$$

$$\frac{dx_2}{dt} = x_2 \left[e_2 \frac{a_2 x_1}{b_2 + x_1} - \frac{a_3 x_3}{b_3 + x_2} - d_2 \right]$$

$$\frac{dx_3}{dt} = x_3 \left[e_3 \frac{a_3 x_2}{b_3 + x_2} - d_3 \right]$$

Example 2: Tritrophic RM Model

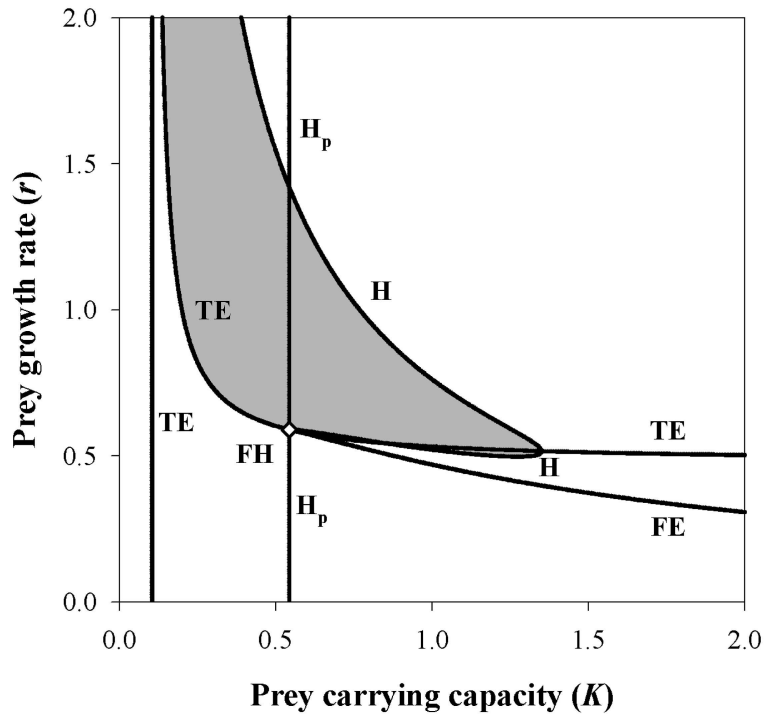
- Fixed parameters (Gragnani et al., 1998)

	Value		Value
a_2	$5/3$	a_3	0.05
b_2	$1/3$	b_3	0.5
e_2	1	e_3	1
d_2	0.4	d_3	0.01

- Initial condition $\boldsymbol{x}_0$ and/or parameters r (prey growth rate) and K (prey carrying capacity) are uncertain

- This system exhibits a rich variety of dynamical behaviors

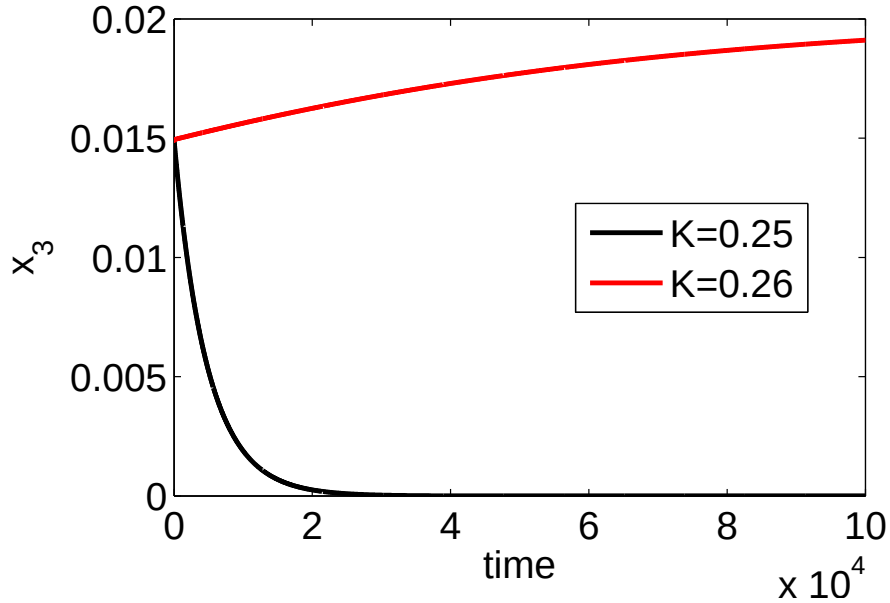
Example 2: Tritrophic RM Model



- This two-parameter bifurcation diagram shows bifurcations of equilibrium (E) only (TE = transcritical; FE = fold; H = Hopf; H_p = planar Hopf)
- There are also bifurcations of cycles
- This diagram was computed rigorously using an interval method (Gwaltney et al., 2007)

Example 2: Tritrophic RM Model

- Consider $r = 0.8$, $\boldsymbol{x}_0 = (0.10, 0.10, 0.015)$, with $K \in [0.25, 0.26]$ containing a transcritical bifurcation.



- Equilibrium (steady-state) value for x_3 is zero for part of the K interval and nonzero for the other part; the states with zero and nonzero x_3 collide at the bifurcation point $\Longrightarrow$ Divergence of state bounds is a feature of the model

- How will VSPODE perform in this situation? Is bisection of parameter interval useful?

Example 3: Model of Two Competitors

- Model of two species who compete for the same resources

$$\frac{dx_1}{dt} = r_1 x_1 \left[1 - \frac{x_1 + \alpha_1 x_2}{K_1} \right]$$

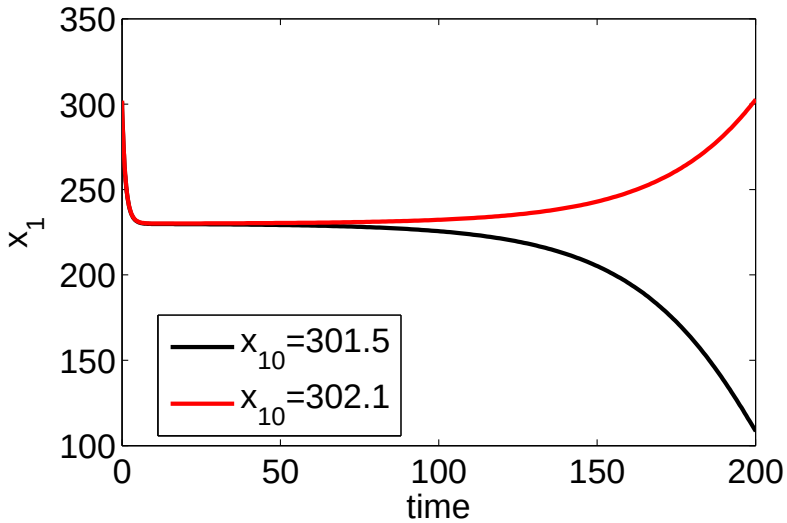
$$\frac{dx_2}{dt} = r_2 x_2 \left[1 - \frac{x_2 + \alpha_2 x_1}{K_2} \right]$$

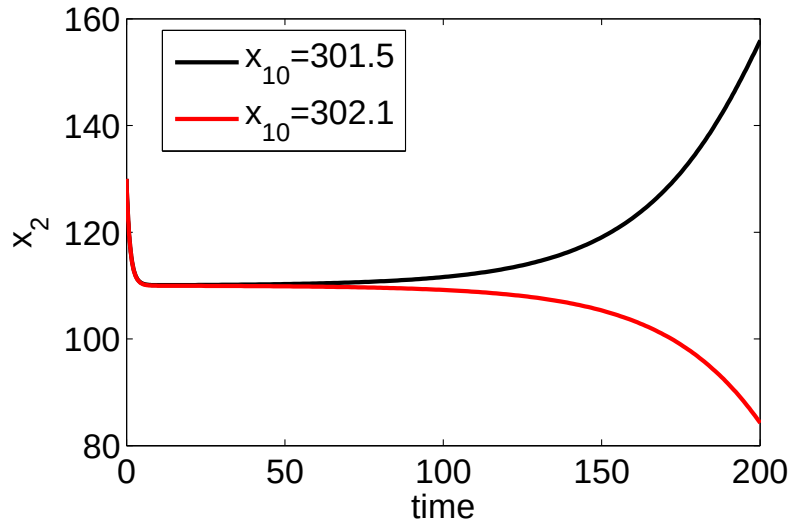
- Parameters and initial conditions

	Value		Value
r_1	1	r_2	0.6
α_1	3	α_2	0.4
K_1	560	K_2	202
x_{10}	$[301.5, 302.1]$	x_{20}	130

Example 3: Model of Two Competitors

- $x_{10} \in [301.5, 302.1]$ contains a separatrix





- For part of the $\boldsymbol{x}_0$ interval, trajectories go to an equilibrium state with $x_1 = 0$ and $x_2 \neq 0$, and in the other part trajectories go to an equilibrium state with $x_1 \neq 0$ and $x_2 = 0 \Longrightarrow$ Divergence of state bounds a feature of the model

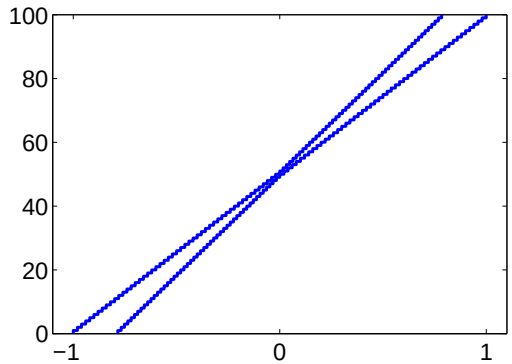
- How will VSPODE perform in this situation? Is bisection of initial state interval useful?

P-box Arithmetic on Multiple Taylor Models

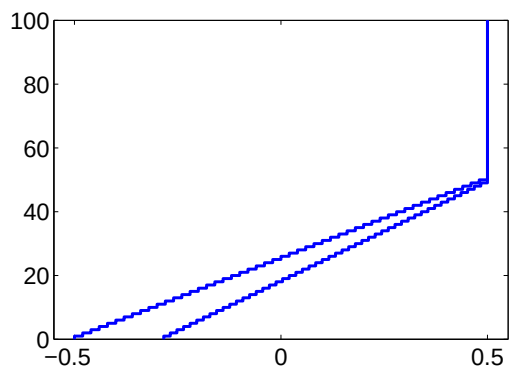
- If parameter and/or initial state interval is bisected, then for the time of interest t_j, VSPODE has computed multiple Taylor model representations $T_{y_j,i}$, one for each subinterval i of the uncertain quantities
- Each Taylor model $T_{y_j,i}$ is only valid over its corresponding subinterval, and each TM has its own expansion point (midpoint of subinterval)
- P-boxes must be reformulated for input to the multiple Taylor models

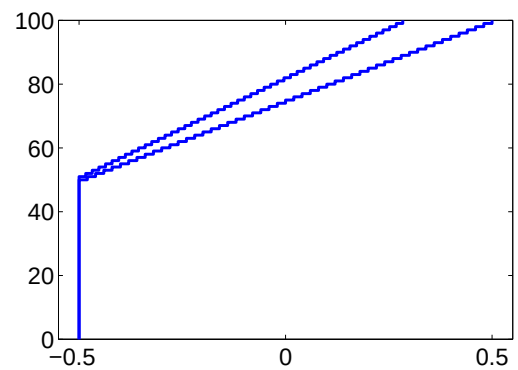
P-box Arithmetic on Multiple Taylor Models

- Preparation of p-box for a bisected parameter interval (original width = 2) for input to Taylor model on each parameter subdomain



$\Downarrow$





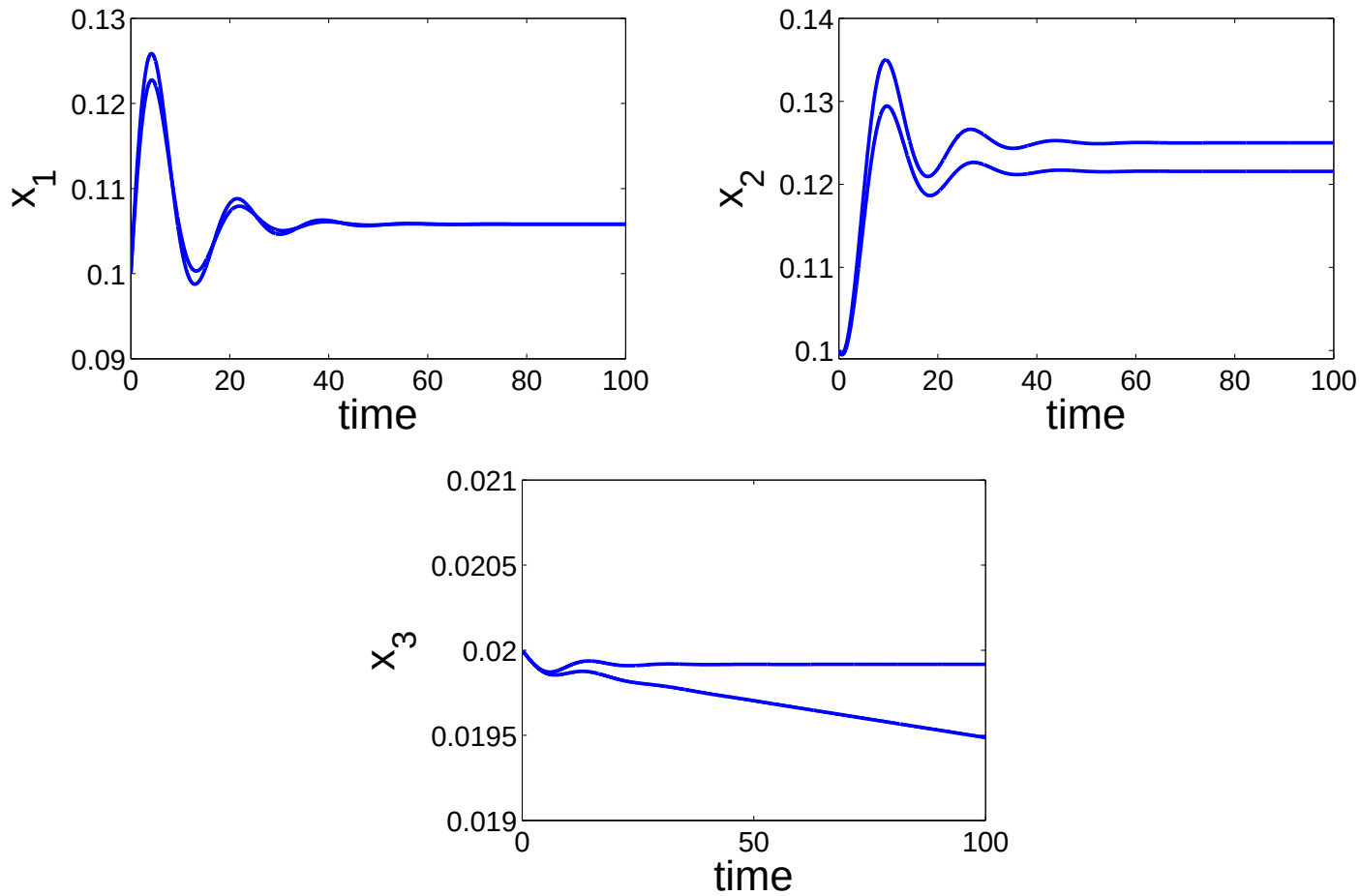
- Parameter values relative to expansion point for TM (here interval midpoint)

P-box Arithmetic on Multiple Taylor Models

- An additional complication is that when the p-boxes are discretized into intervals for p-box arithmetic, there are two possibilities
  - The interval is completely contained in a single subdomain, so it is evaluated only on its single corresponding TM
  - The interval is not contained in a single subdomain, so it must be evaluated on multiple TMs and the union of the result used
- Ultimately, the results of all the TM evaluations are collected and a single output p-box is constructed

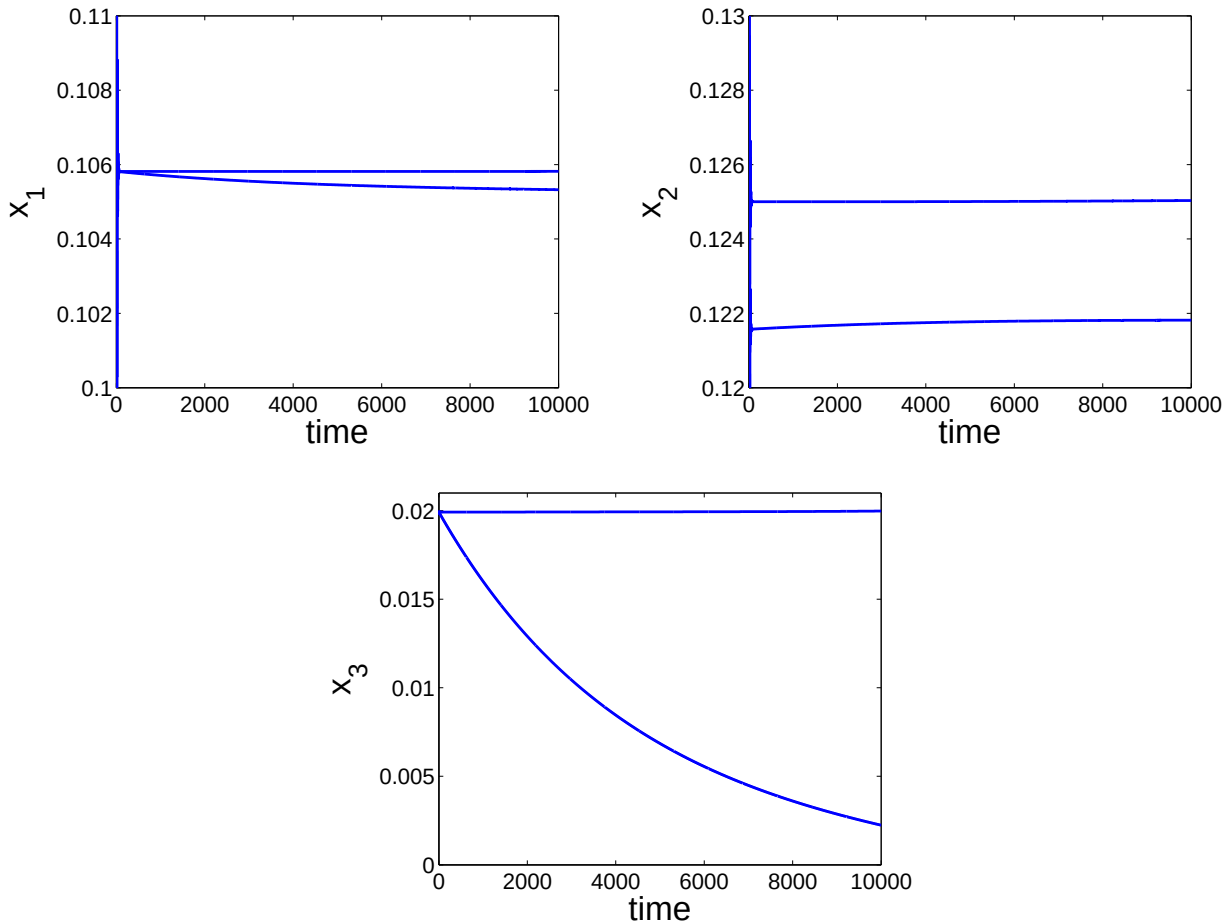
Example 2: Tritrophic RM Model

- Revisit the RM model with transcritical bifurcation: $r = 0.8$, $K \in [0.25, 0.26]$, $\boldsymbol{x}_0 = (0.10, 0.10, 0.02)$
- VSPODE enclosures of $x_1(t)$, $x_2(t)$, and $x_3(t)$ over $t = [0, 100]$



Example 2: Tritrophic RM Model

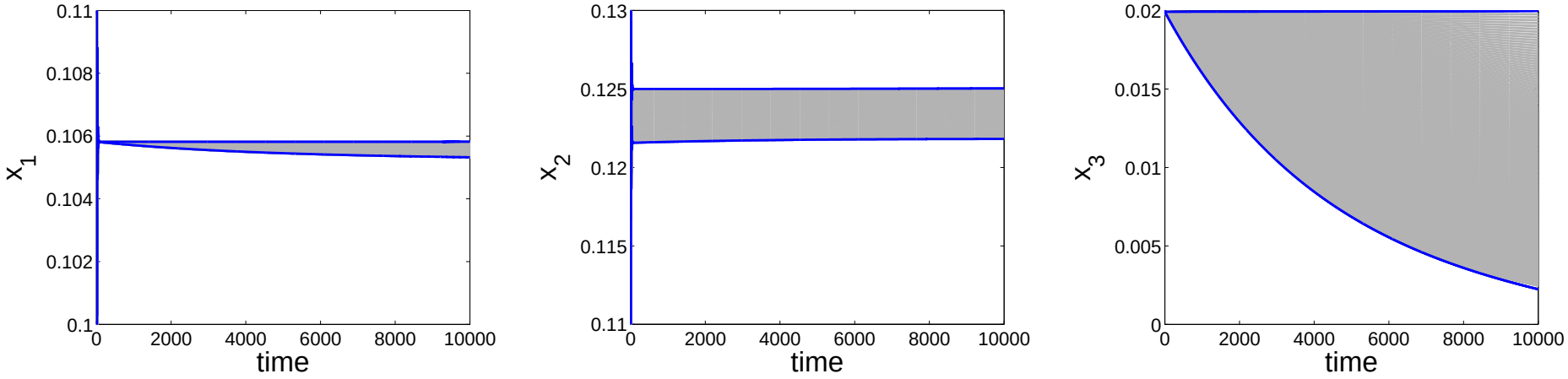
- VSPODE enclosures of $x_1(t)$, $x_2(t)$, and $x_3(t)$ over $t = [0, 10000]$



- How tight are these bounds?

Example 2: Tritrophic RM Model

- Comparison of VSPODE bounds to Monte Carlo simulations

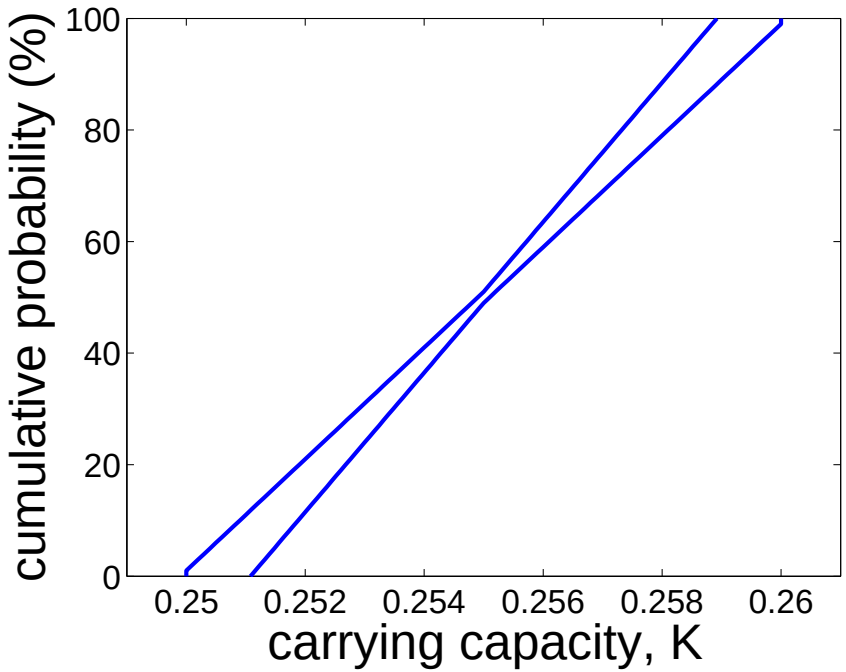


- VSPODE bounds with parameter bisection are the same: No need for bisection

- Knowledge of bifurcation will help avoid unnecessary bisections

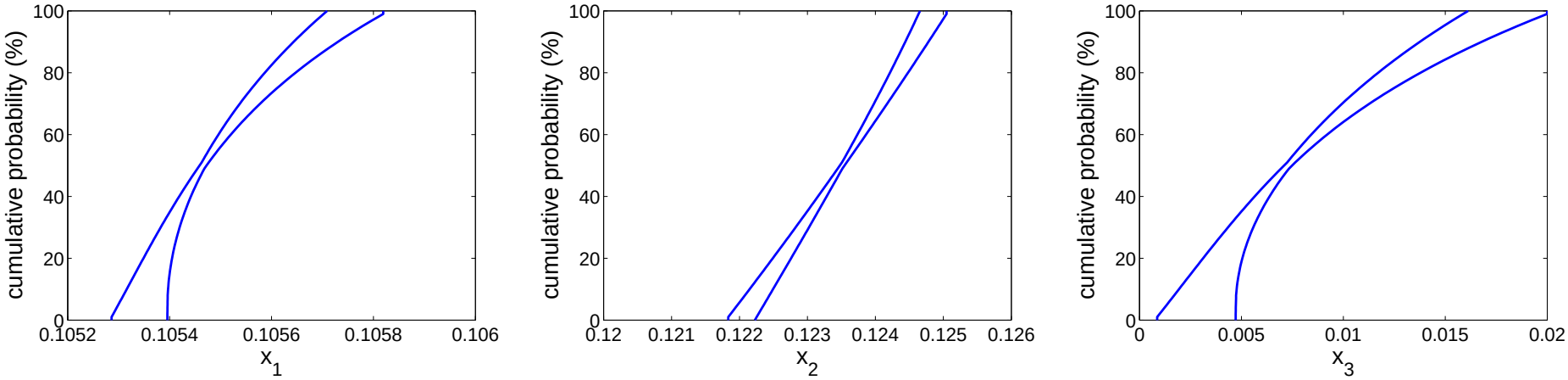
Example 2: Tritrophic RM Model

- P-box representation of uncertain parameter K



Example 2: Tritrophic RM Model

- Probability distributions (p-boxes) for $\boldsymbol{x}$ at $t = 10000$

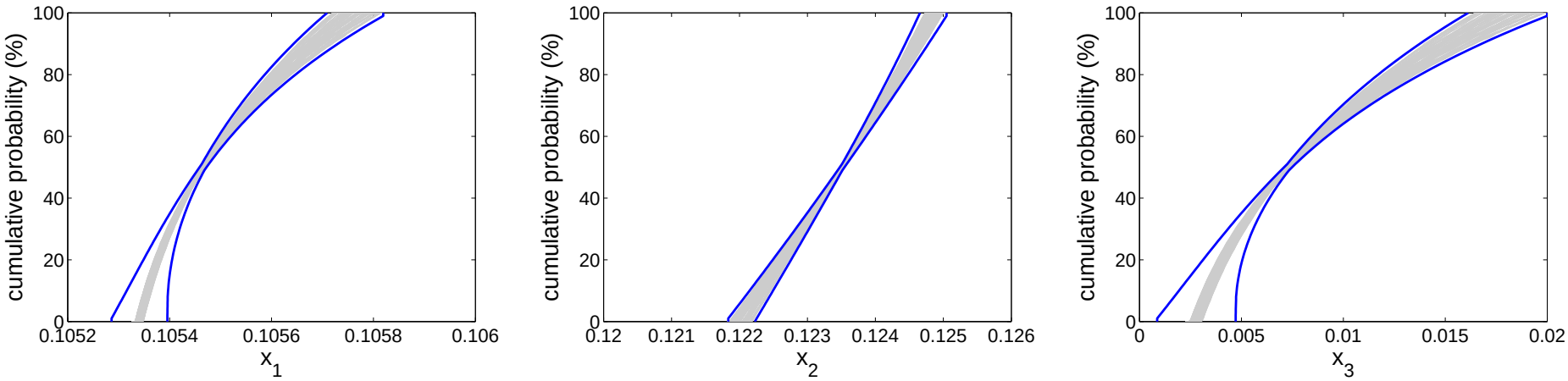


- Same results with or without bisection (at midpoint or at bifurcation)

- Validates new procedure for p-box arithmetic over multiple Taylor models

Example 2: Tritrophic RM Model

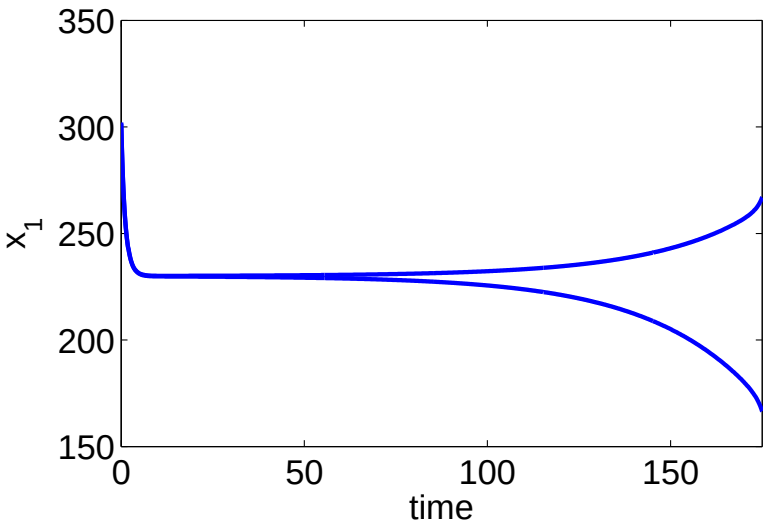
- Comparison with Monte Carlo simulations: 100 simulations, each with a uniform probability distributions for K sampled from the input p-box, with each simulation consisting of 10000 trials

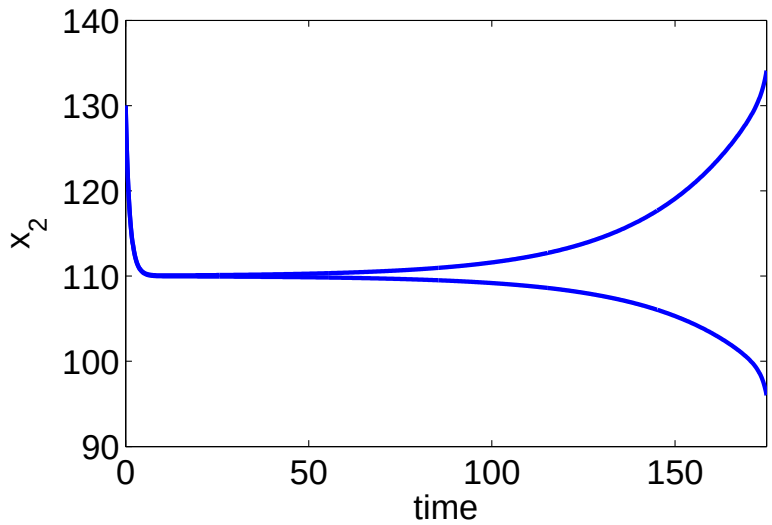


- Additional resolution in p-box arithmetic at lower probability levels is needed
- The second-order (nested) Monte Carlo simulation is computationally expensive

Example 3: Model of Two Competitors

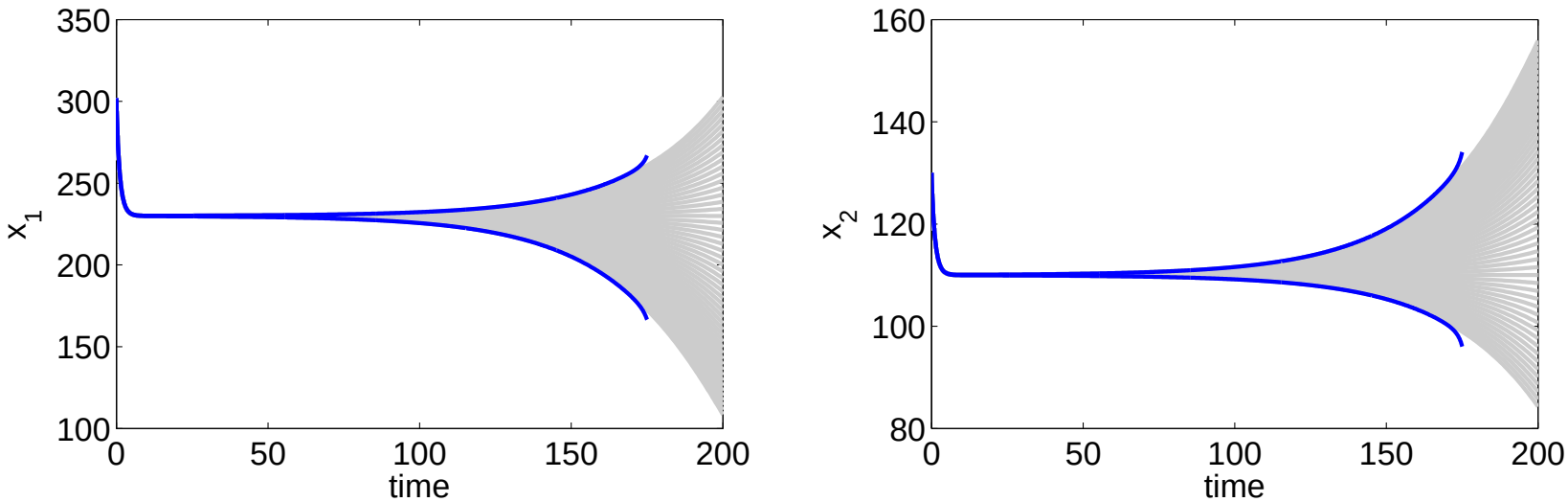
- Revisit the two-competitors model with separatrix: $x_{10} \in [301.5, 302.1]$
- VSPODE enclosures of $x_1(t)$ and $x_2(t)$ over $t = [0, 175]$





Example 3: Model of Two Competitors

- Comparison with Monte Carlo simulation over $t = [0, 175]$

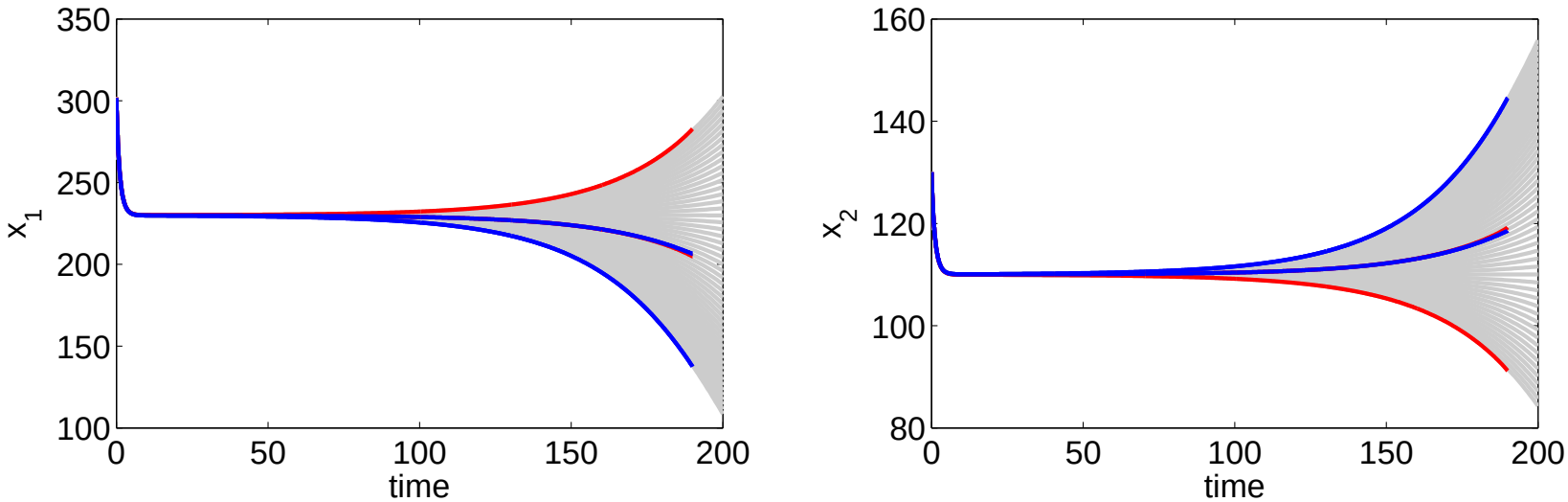


- Begin to see overestimation errors near $t = 175$

- Try bisecting initial state interval

Example 3: Model of Two Competitors

- Comparison with Monte Carlo simulation over $t = [0, 190]$

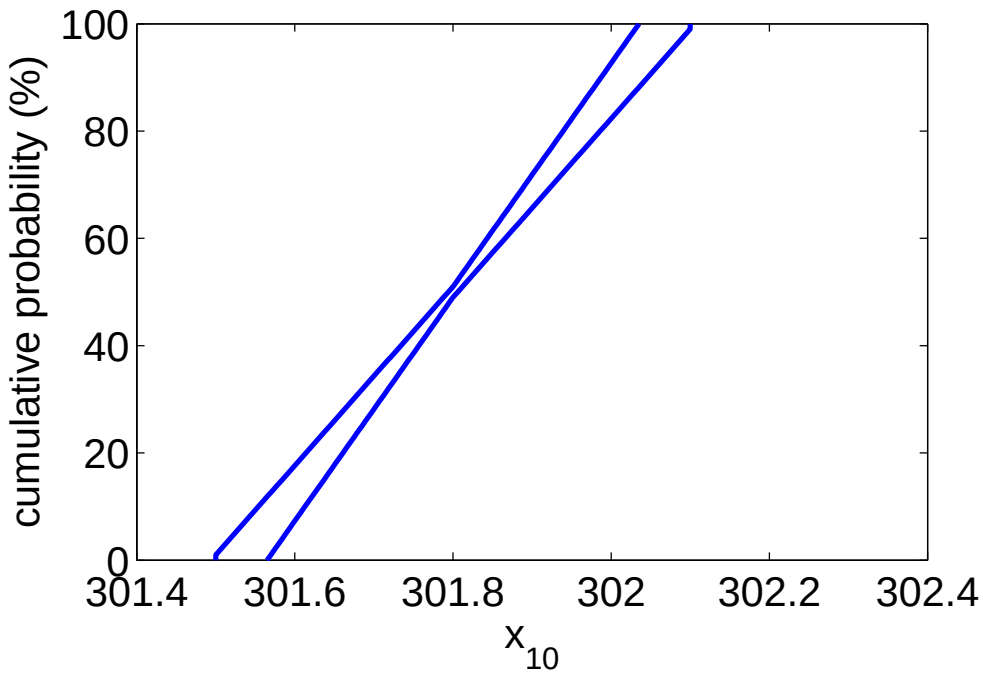


- Subinterval (red) containing the separatrix is beginning to show overestimation errors

- As expected, bisection helps address overestimation errors; additional bisections are needed

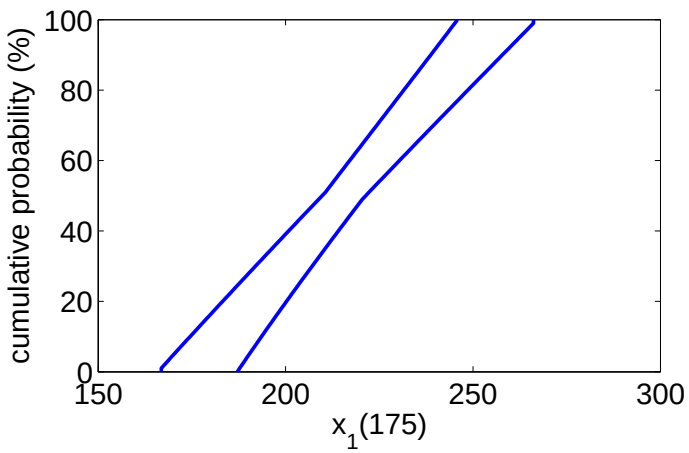
Example 3: Model of Two Competitors

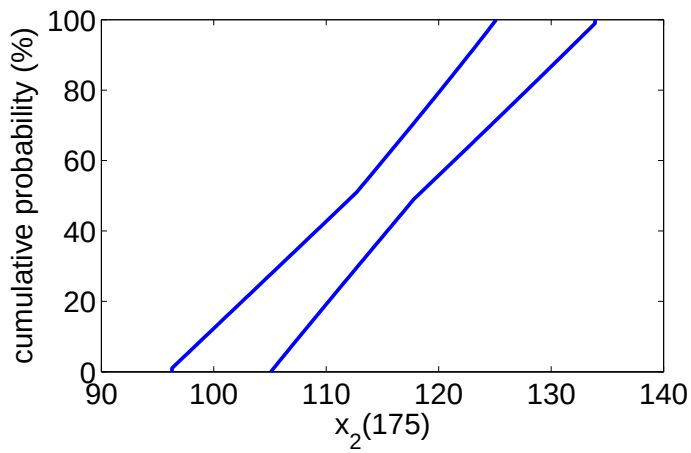
- P-box representation of uncertain initial condition x_{10}



Example 3: Model of Two Competitors

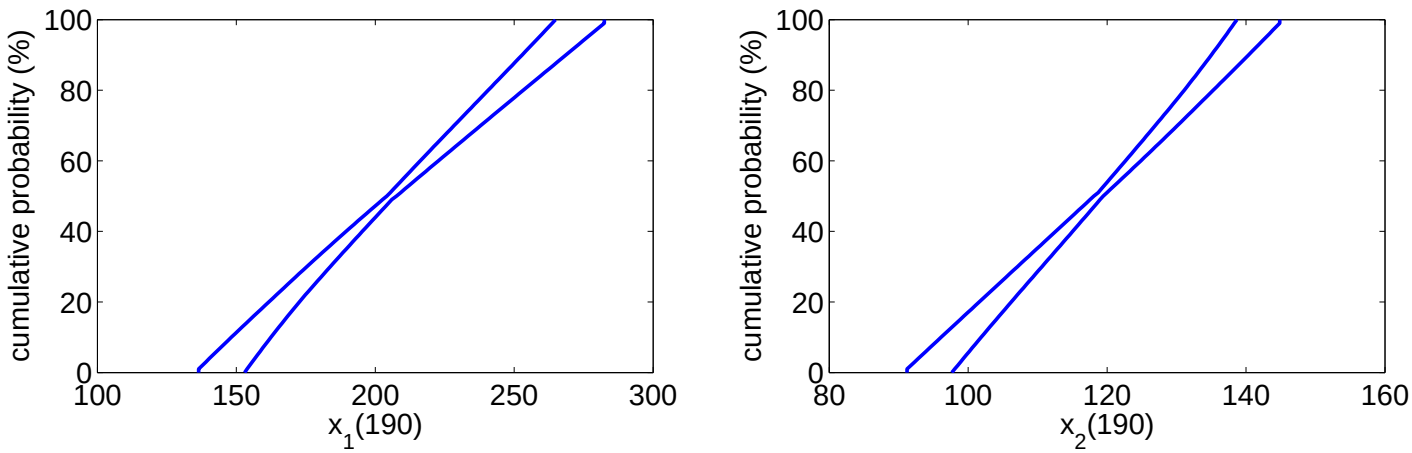
- Probability distributions (p-boxes) for $\boldsymbol{x}$ at $t = 175$ (no bisection)





Example 3: Model of Two Competitors

- Probability distributions (p-boxes) for $\boldsymbol{x}$ at $t = 190$ (with bisection)



- Overestimation errors were reduced

Concluding Remarks

- Behavior of a verified integrator around a bifurcation or separatrix may look like overestimation error
  - Knowledge of bifurcation or separatrix location is helpful
  - Bisection strategies may not help, unless there is overestimation error to reduce
- A procedure was developed for arithmetic with p-boxes split over multiple subdomains and multiple Taylor models
  - Useful in cases where bisection strategies are needed to reduce overestimation error