Parameter Estimation for Differential Equations: A Generalized Smoothing Approach

Jim Ramsay, McGill University
David Campbell, Simon Fraser University
Jiguo Cao, Simon Fraser University
Giles Hooker, Cornell University

Eurandom Meeting: Eindhoven, 2009

Outline

Overview

- We summarize our approach to fitting data with a model defined by a set of differential equations (ODE's).

- Modeling groundwater level in response to rain in Vancouver contrasts assumptions often made in mathematical treatments of ODE's with the requirements of modeling real world data.

- A controlled experiment involving predator-prey population dynamics shows some challenges researchers can face when they match their model to their data.

- A model for a chemical reaction under stable and unstable regimes.

- We conclude with some reflections on the use of data by dynamical systems modelers, and challenges facing statisticians in working with this community.

Some notation

- Let $\mathbf{x}$ be a vector-valued function of length n varying over time t, and that has first derivative values $D\mathbf{x}(t)$.
- Let $\mathbf{u}$ be a vector containing one or more input or *forcing* functions.
- Let $\boldsymbol{\theta}$ be a vector of parameters defining the ODE.
- A general formulation is

$$D\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}, t | \boldsymbol{\theta}).$$

- Systems involving higher order derivatives $D^m\mathbf{x}$ are reducible to this form by defining new variables,

$$\mathbf{x}_1 = \mathbf{x}, \quad \mathbf{x}_2 = D\mathbf{x}_1, \quad \ldots, \quad \mathbf{x}_{m-1} = D^{m-1}\mathbf{x}.$$

Outline

The Collocation Approach

- For each variable x_i in $\mathbf{x}$, we define a basis function expansion

$$\hat{x}_i(t) = \mathbf{c}_i' \phi_i(t),$$

  where $\mathbf{c}_i$ and ϕ_i are a coefficient vector and a vector of basis functions, respectively.

- Enough basis functions are used to capture any transient localized features. This may require thousands of basis functions.

- Splines are usually the logical choice because of their local support and their capacity to model transient events. Far more knots than data points are often required.

- Discontinuities in a derivative at known locations may require multiple knots.

A Fitting Criterion Conditional on θ

- A criterion $J_i(\mathbf{c}_i | \boldsymbol{\theta}, \lambda_i)$ measures, for variable i,
  - the fidelity of $\hat{x}_i(\mathbf{t}_i | \mathbf{c}_i, \boldsymbol{\theta}, \lambda_i)$ to the data in $\mathbf{y}_i$,
  - how close $\hat{x}_i$ is to being an ODE solution for parameters $\boldsymbol{\theta}$,
  - where smoothing parameter λ_i controls the relative emphasis on these two objectives.

- For example

$$J_i(\mathbf{c}_i) = \|\mathbf{y}_i - \hat{x}_i(\mathbf{t}_i)\|^2 + \lambda_i \int [L_{i,\boldsymbol{\theta}}(\hat{x}_i(t))]^2 dt$$

where

$$L_{i,\boldsymbol{\theta}}(x_i) = Dx_i - f_i(\mathbf{x}, \mathbf{u}, t | \boldsymbol{\theta}).$$

- The first data-fitting term can be -negative log likelihood, or any appropriate function.

The parameter hierarchy

- There are three classes of parameters to estimate:
    - The coefficients c_i defining each basis function expansion.
    - The model parameters θ defining the ODE.
    - The smoothing parameters λ_i.

The roles of the three parameter levels

- Coefficients $\mathbf{c}_i$ are *nuisance* parameters because
  - they are not of direct interest
  - their numbers are apt to vary with the length of the observation interval, density of observation, and other design factors.
  - they can be orders of magnitude greater in number than the number of ODE parameters in $\boldsymbol{\theta}$.
  - each c_{ik} controls the shape of x_i only *locally* over a small neighborhood of t_{ik}.

Structural and complexity parameters

- Parameters θ are *structural* in the sense of being of primary interest. Our clients want to know their values and precisions.
- Each θ_ℓ affects one or x_i's over a wide region of t-values.
- Smoothing parameters λ_i control the overall *complexity* of the model:
    - $\lambda_i \to 0 \Rightarrow$ high complexity in $\hat{x}_i$
    - $\lambda_i \to \infty \Rightarrow$ low complexity in $\hat{x}_i$

The parameter cascade algorithm

- Nuisance parameters are defined as *smooth functions* $\mathbf{c}_i(\boldsymbol{\theta}, \boldsymbol{\lambda})$ of the structural and complexity parameters.
- Structural parameters are defined as *functions* $\boldsymbol{\theta}(\boldsymbol{\lambda})$ of the complexity parameters.
- These functional relationships are defined *implicitly* by specifying a different conditional fitting criterion at each level of the parameter hierarchy.

The multi-criterion optimization strategy

- Nuisance parameter functions $\mathbf{c}_i(\boldsymbol{\theta}, \lambda_i)$ are defined by optimizing the regularized fitting criteria $J_i(\mathbf{c}_i)$ *each time* either $\boldsymbol{\theta}$ or $\boldsymbol{\lambda}$ is modified.

$$J_i(\mathbf{c}_i) = \|\mathbf{y}_i - \hat{x}_i(\mathbf{t}_i)\|^2 + \lambda_i \int [L_{i,\boldsymbol{\theta}}(\hat{x}_i(t))]^2 dt$$

- A purely data-fitting criterion $H(\boldsymbol{\theta})$ is then optimized with respect to the structural parameters $\boldsymbol{\theta}$ alone, such as

$$H(\boldsymbol{\theta}) = \sum_i w_i \|\mathbf{y}_i - \hat{x}_i(\mathbf{t}_i | \boldsymbol{\theta}, \lambda_i)\|^2$$

- Finally, at the top level, a complexity criterion, such as generalized cross-validation, $GCV(\boldsymbol{\lambda})$, is optimized with respect to $\boldsymbol{\lambda}$.

Advantages of parameter cascading

- Gradients and hessians at any level can be readily computed using the *Implicit Function Theorem*.
- Parameter estimates have bias and sampling variances as good as those obtained by other methods.
- Confidence interval estimates for parameters are easy to compute, and appear to have excellent bias and coverage properties.
- Compared to marginalizing out the nuisance parameters using MCMC or or other integration schemes, parameter cascading is
  - much faster,
  - much more stable,
  - much easier to program
  - can be deployed to the user community more conveniently.

Two fits for the price of one

- Data fitting function $\hat{x}_i(t)$ is only approximately a solution of the differential equation. It is a *data-regularized* approximation to a solution.

- But parameter estimates $\hat{\boldsymbol{\theta}}$ and initial values $\hat{x}_i(0)$ permit us to numerically approximate an accurate solution $x_i^*(t)$ of the differential equation that fits the data as well as possible.

- Comparing the two fits can be instructive.

- When $\boldsymbol{\lambda}$ are small, the optimization criteria **J** and **H** are smooth functions of their arguments, easily optimized, and global optima are easier to locate.

- By ramping up $\boldsymbol{\lambda}$ slowly we
    - avoid the local minimum problem
    - force $\hat{x}_i(t)$ and $x_i^*(t)$ to coalesce

Outline

The Vancouver mud slide tragedy

- After a mud slide in North Vancouver, BC that killed two people, the firm BGC Engineering was commissioned to monitor groundwater level.
- Rainfall raises groundwater level,
- If groundwater rises to flood a boundary between stable and unstable soil structures, residents will be evacuated.
- Rainfall is measured hourly, and can be viewed as event data, or what statisticians call *marked point process* data.
- BGC Engineering asked us if we could develop an effective six-hour prediction of groundwater level.

Rainfall (mm/hr)

Groundwater level

day
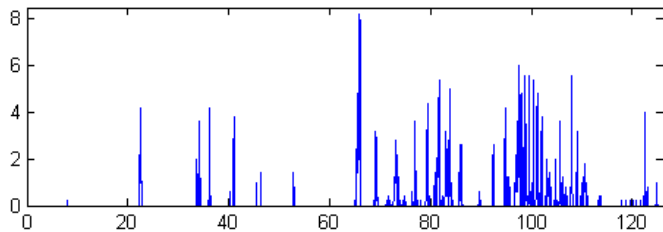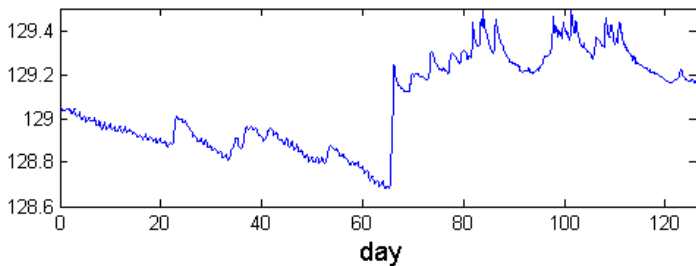
A simple differential equation

$$\frac{dG}{dt} = -\beta G(t) + \alpha R(t - \delta) - \mu$$

where

- β is the speed with which groundwater $G(t)$ reacts to a unit change in rainfall input $R(t)$,
- α defines the impact of a change in $R(t)$; the gain $K = \alpha/\beta$ is the final change in level achieved after a unit increase in $R(t)$, and
- μ is a baseline level, required here because the origin for level $G(t)$ is not meaningful.
- lag δ is the time for rainfall to reach the groundwater level, and is known to be about three hours.
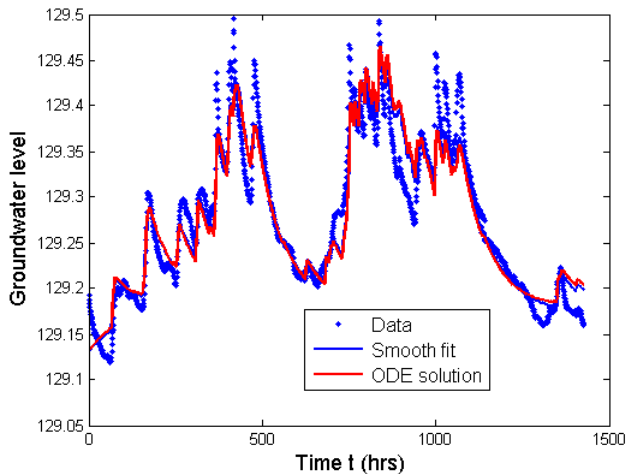
The constant coefficient fit
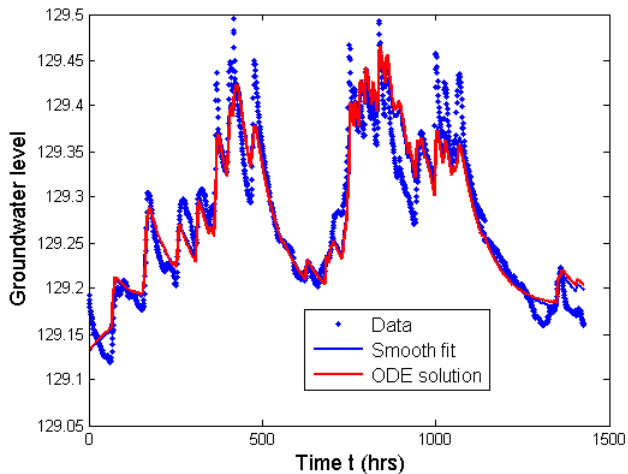
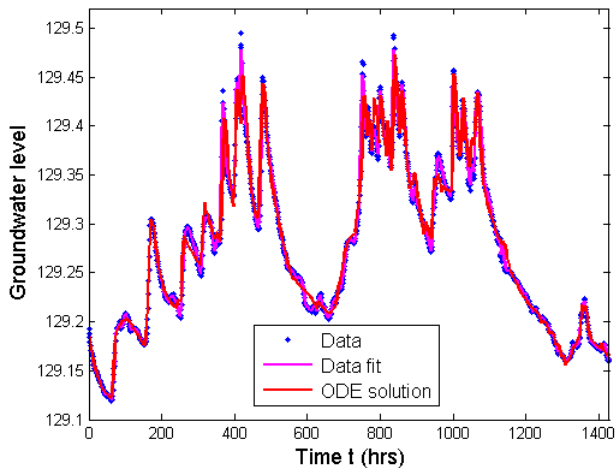# Allowing for long-term trend in ODE parameters

- As groundwater level $G(t)$ changes, the dynamics change, too, because we move through different types of sub-soil structures.
- We weren't given sub-soil transmissioin rates, so we needed to allow $\beta(t), \alpha(t)$ and $\mu(t)$ to vary slowly over time.

$$\frac{dG}{dt} = -\beta(t)G(t) + \alpha(t)R(t - \delta) + \mu(t)]$$

The constant coefficient fit

The variable coefficient fit

Lessons learned along the way

- The point or delta function nature of the rainfall data plays havoc with numerical initial value solvers, and severely violates Lipschitz continuity.
- Of course, this simple differential equation can be solved explicitly.
- But the solution was useless for data fitting because of it's multiple discontinuities and the large amount of data involved.
- The parameter cascading approach was much easier to work with because it used the differential equation itself rather than the solution.
- Its robustness was essential for working with this problem.

Lessons learned along the way

- The point or delta function nature of the rainfall data plays havoc with numerical initial value solvers, and severely violates Lipschitz continuity.
- Of course, this simple differential equation can be solved explicitly.
- But the solution was useless for data fitting because of it's multiple discontinuities and the large amount of data involved.
- The parameter cascading approach was much easier to work with because it used the differential equation itself rather than the solution.
- Its robustness was essential for working with this problem.

Outline

The data

- Prof. Gregor Fussmann of the McGill Biology Dept. studies the interaction between one-celled algae organisms (Chlorella) and a multicellular rotifer that eats them (Brachionis).
- The environment is a tank of water into which is pumped nitrogen, the nutrient for the algae cells.
- The two populations oscillate in the classic manner, but the tightly controlled experimental nature of the data has attracted considerable attention among ecologists.

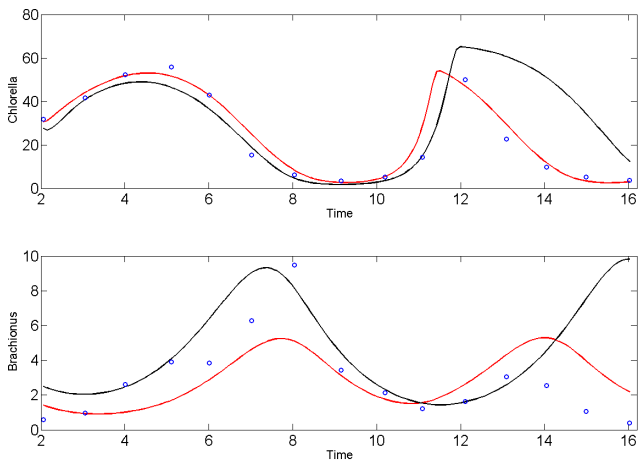# Fussmann's data
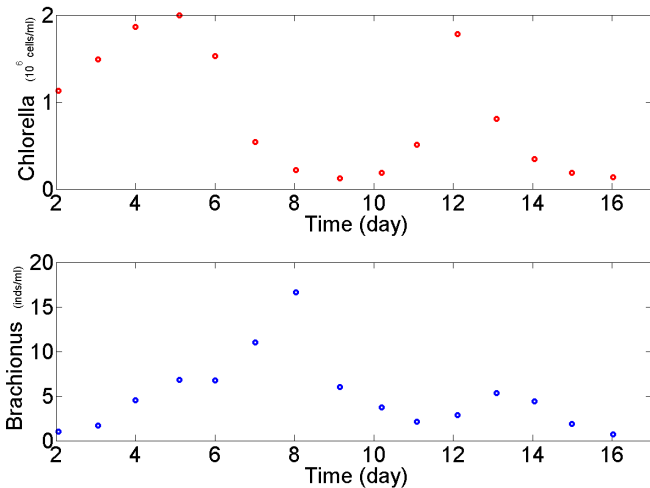
The 1-input / 3 output model

$$
\begin{aligned}
DN(t) &= -\delta_N N(t) + \delta_N N_{in} - \alpha_C \frac{N(t)}{K_C + N(t)} C(t) \\
DC(t) &= -\delta_C C(t) + \alpha_C \frac{N(t)}{K_C + N(t)} C(t) - \alpha_B \frac{C(t)}{K_B + C(t)} B(t)/\epsilon \\
DB(t) &= -\delta_B B(t) + \alpha_B \frac{C(t)}{K_B + C(t)} R(t) \\
DR(t) &= -\delta_R R(t) + \alpha_B \frac{C(t)}{K_B + C(t)} R(t)
\end{aligned}
\tag{1}
$$

- $N(t)$ is nitrogen concentration (A known input variable).
- $C(t)$ is Chlorella population size.
- $B(t)$ is Brachionis population size.
- $R(t)$ is reproducing Brachionis sub-population size.

The results

- The parameters to estimate are:
  - The first order dynamics parameters $(\delta_N, \delta_C, \delta_B, \delta_R)$.
  - The forcing function multipliers $(\alpha_C, \alpha_B, \epsilon)$.
  - The Michaelis-Menton or hill function baselines (K_C, K_B).
- Some of these could be experimentally determined from other experiments.

Fitting Fussmann's data

Chlorella values are multiplied by 28.0 and Brachionis by 0.57.

Lessons learned along the way

- Although dynamical systems are widely used in ecology to model population dynamics, high quality data are hard to find.
- Fussmann's data are among the best, but his model only captures shape features to a limited extent, and fails to reproduce actual data.
- Dr. Jiguo Cao's analyses, as well as Fussmann's published results, showed that model values were wrong by large factors.
- Dynamical systems modelers are often content to reproduce shape features rather than fit data, and their models are perhaps more like mathematical metaphors.

Outline

What is a tank reactor?

- A continuously stirred tank reactor *CSTR* consists of a tank surrounded by cooling jacket and an impeller which stirs the contents.
- It is a basic piece of equipment for a chemical engineer.

The tank reactor variables

- A fluid is pumped into the tank containing a reagent with concentration C_{in} at a flow rate F_{in} and temperature T_{in}.
- Inside the tank a reaction takes place, producing a product that leaves the tank with concentration C_{out} and temperature T_{out}.
- A coolant enters the cooling jacket with temperature T_{cool} and flow rate F_{cool}.
- Temperature T_{out} is can be cheaply measured with little delay and considerable accuracy, but concentration C_{out} requires time and money.

The tank reactor equations

$$
\begin{aligned}
DC_{out} &= -\beta_{CC}(T_{out})C_{out} + F_{in}C_{in} \\
DT_{out} &= -\beta_{TT}(F_{cool}, F_{in})T_{out} + \beta_{TC}(T_{out})C_{out} \\
&\quad + F_{in}T_{in} + \alpha(F_{cool})T_{cool}.
\end{aligned}
$$

- The concentration equation is linear and forced by C_{in}.
- The temperature equation is nonlinear because of the role of T_{out} in coefficient $\beta_{TC}(T_{out})$ multiplying C_{out}.

The tank reactor coefficients

- The dynamics of the system are controlled by these four coefficient functions:

$$
\begin{aligned}
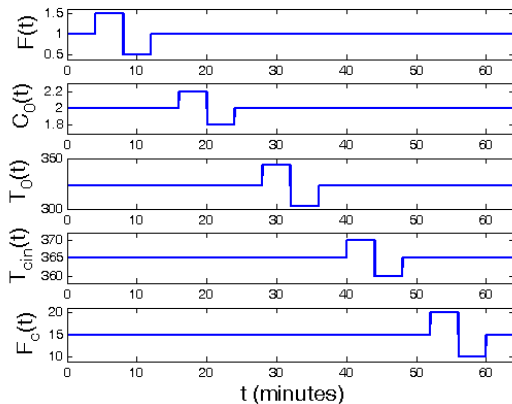\beta_{CC}(T_{out}, F_{in}) &= \kappa \exp[-10^4 \tau (1/T_{out} - 1/T_{ref})] + F_{in} \\
\beta_{TT}(F_{cool}, F_{in}) &= \alpha(F_{cool}) + F_{in} \\
\beta_{TC}(T_{out}) &= 130 \beta_{CC}(T_{out}, F_{in}) \\
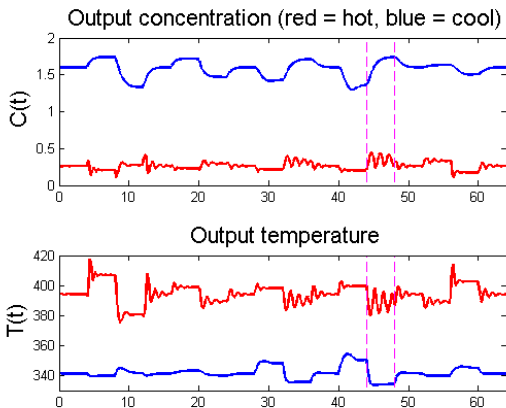\alpha(F_{cool}) &= a F_{cool}^{b+1}/(F_{cool} + a F_{cool}^{b}/2),
\end{aligned}
$$

- These functions depend on two paired unknown parameters:
  - κ and τ
  - a and b

Tank reactor inputs

Each input in turn is stepped up, down and back to baseline.

Tank reactor outputs

The experiment is run at two coolant temperatures: hot and cool.

What we see

- When temperatures are moderate, the reactor responds smoothly to changes in input.
- But when temperatures are higher, sharp high frequency oscillations emerge, and are particularly troublesome for a change in coolant temperature.
- Can we predict reactor response at high temperatures from data collected and parameters estimated under the safer cool regime?
- Can we do this using only temperature measurements?

Simulations for the tank reactor equations

- Parameters and initial values for paths were set to those provided by a well known text on control engineering, T. E. Marlin (2000) *Process Control*. New York: McGraw-Hill.
- Parameter b is impossible to estimate because of its correlation with a, and therefore was fixed 0.5.
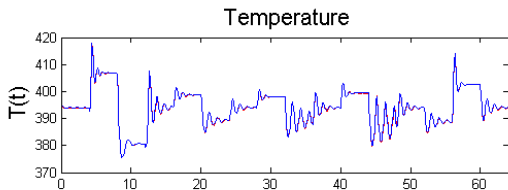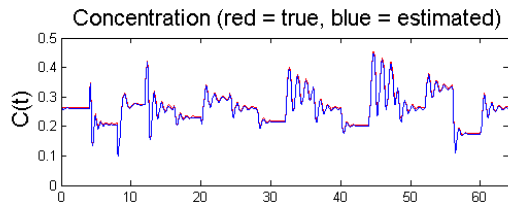- 1000 simulated samples were analyzed.

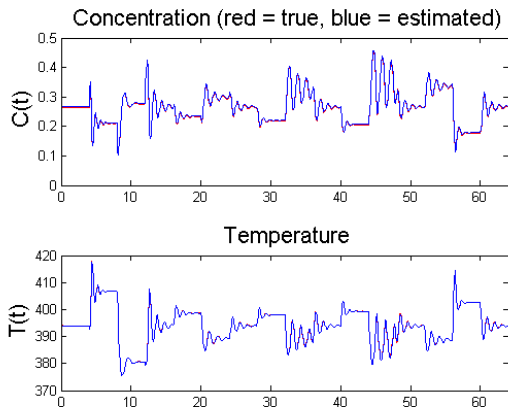# A typical set of tank reactor data
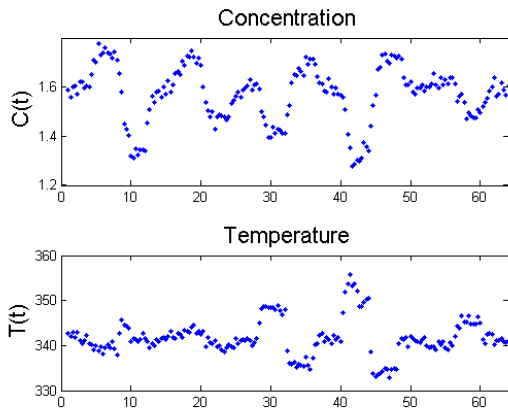
Path estimations, cool mode

Concentration (red = true, blue = estimated)
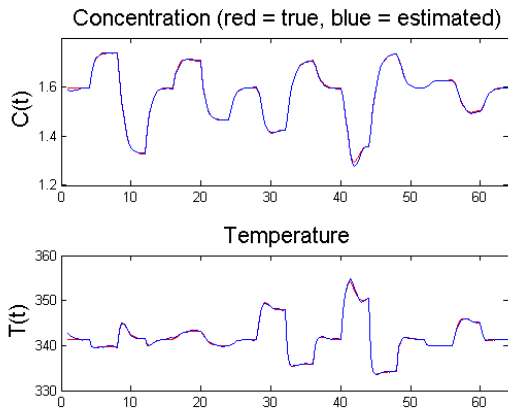
Temperature

Path estimations, hot mode

Path estimations, hot mode

Data for only temperature collected in the cool mode were used.

Summary statistics for parameter estimates from 1000 simulated samples

	κ	τ	a
True value	0.4610	0.8330	1.6780
Mean value	0.4610	0.8349	1.6745
Std. Dev.	0.0034	0.0057	0.0188
Est. Std. Dev.	0.0035	0.0056	0.0190
Bias	0.0000	0.0000	-0.0001
Std. Err.	0.0002	0.0004	0.0012

Lessons learned along the way

- In this severely nonlinear problem with the potential of localized high frequency variation,
    - We can get excellent parameter and confidence interval estimates with measurements on only one of the two output variables,
    - Results obtained on the system run under a benign and smooth regime can be used to predict results for an experiment with much less stable results.

- Although it can be hard to find dynamical systems accompanied by good data, much of great valuecan be learned by working with simulated data and other methods for evaluating what can and cannot be estimated from a specific data design.

Outline

Software

- All the results were computed in Matlab.
- General purpose functions in both R and Matlab have been developed by Giles Hooker, and can be obtained at http://www.bscb.cornell/ hooker/.
- Matlab functional data analysis software was also used. These and a set of software routines that may be applied to any differential equation is available from http://www.functionaldata.org.

References

- Ramsay, J. O., Hooker, G., Cao, J. and Campbell, D. (2007) Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Series B*. **69**, 741-796.

- J. O. Ramsay and B. W.Silverman (2005) *Functional Data Analysis*, Second Edition. New York: Springer.