Estimating Dynamic Models

Jim Ramsay, McGill University, ramsay@psych.mcgill.ca
Giles Hooker, Cornell University, giles.hooker@cornell.edu
Jiguo Cao, Yale University , jiguo.Cao@yale.edu
David Campbell, McGill University,
campbell@math.mcgill.ca

Outline

Overview

- We want to fit data by a solution to a system of nonlinear differential equations (DIFE's).
- We ignore DIFE's so simple that they can be solved, such as linear constant coefficient systems. These are already well taken care of.
- Our approach is a generalization of smoothing methods combined with a computational approach involving a modification of profiling.
- We will show results for simulated data from two test-bed problems.
- Data from a chemical reactor producing nylon is analyzed to estimate parameters defining equations for reaction kinetics.

What differential equations do

- DIFE's model change.
- The link the behavior of one or more derivative to the behavior of the process itself and, possibly,
- to one or more exogenous inputs.
- Perhaps the grande dame of such dynamic models is $F = Ma$, connecting the rate of change of velocity a to mass M and an exogenous force F.
- Probably more people know about the closely related

$$E = mc^2$$

The 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 forcing functions.
- Let θ be a vector of parameters defining the DIFE.
- A general formulation is $D\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}, t|\theta)$.
- Systems involving higher order derivatives $D^m\mathbf{x}$ are reducible to this form by defining new variables,

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

Nonlinear least squares

- The usual approach is called by textbooks the *nonlinear least squares* or NLS method.
- An initial value numerical method, such as Runge-Kutta, is used to approximate the solution given
  - a trial set of parameter values
  - a trial set of initial conditions.
- The fit value, usually `SSE`, is input into an optimization algorithm to update parameter estimates and the initial conditions.

NLS problems

- NLS is computationally intensive since a numerical approximation to a possibly complex system is required for each update of parameters and initial conditions.
- The size of the parameter set is increased by the set of initial conditions needed to solve the system.
- The inaccuracy of the numerical approximation is added to noise in the data.
- NLS only produces point estimates of parameters, and, where interval estimation is needed, a great deal more computation is required.
- The fitting criterion can have a complex surface geometry, including many local minima.

Other methods

- Simulated annealing can be used if only a few parameters are involved, but can be extremely slow.
- Local linearization combined with methods for linear systems such as the Kalman filter can be used if the nonlinearity is mild.
- Bayesian methods using MCMC are also possible, but require repeated numerical solution and also add initial values to the parameter set.

Outline

The FitzHugh-Nagumo model

- This simple two-component system is widely used to model properties of actual neural networks.
- They describe the reciprocal dependencies of the voltage V across an axon membrane and a recovery variable R reflecting outward currents, and
- the impact of a time-varying external excitation E.
- In the typical experiment only V will be measured, but we will consider both to be available.

The FitzHugh-Nagumo equations

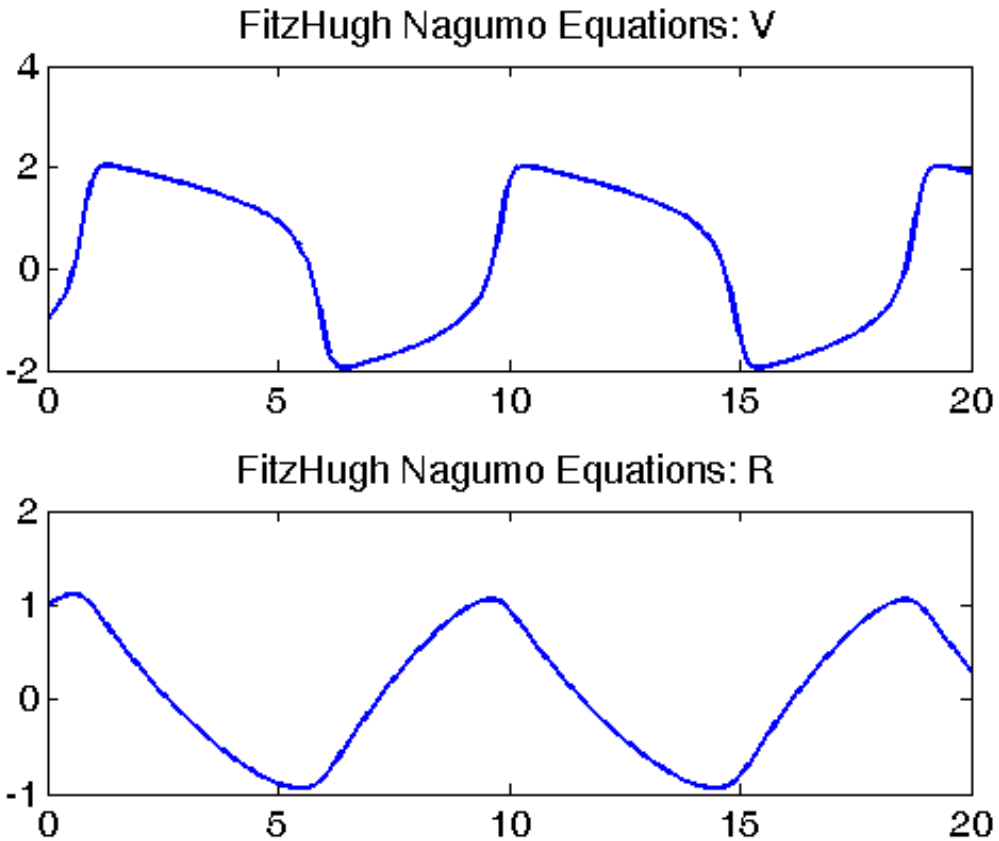
- Here is the system:

$$DV = c\left(V - \frac{V^3}{3} + R\right) + E(t)$$

$$DR = -\frac{1}{c}(V - a + bR)$$

- V is voltage across axon membrane
- R reflects outward currents
- E reflects external excitation
- The dynamics of the system are controlled by parameters a, b and c.
- The system would be linear except for the V^3 term.

A FitzHugh-Nagumo solution



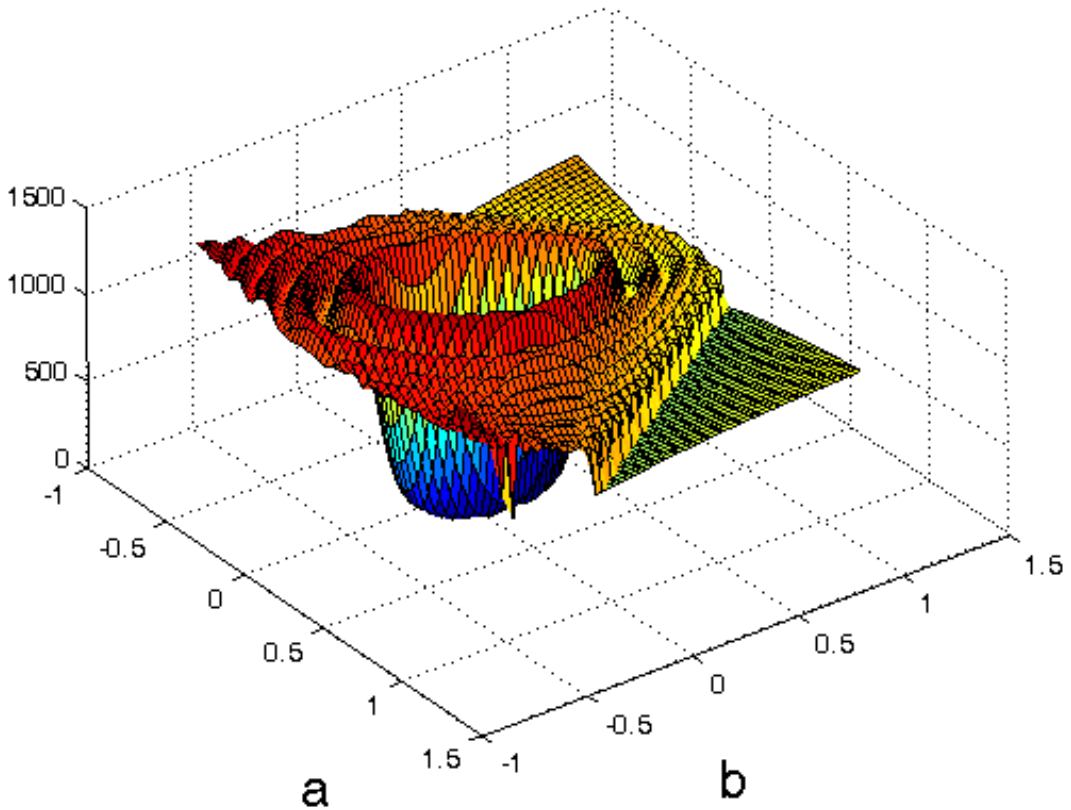
What we see

- The solution quickly reaches a steady state where it is periodic with an asymmetric pattern.
- The parameters control the amplitude and period of the response.
- The second order van der Pol equation is a closely related system.

The response surface can be complex

- Differential equations be simple, and yet define extremely complex behavior.
- This is reflected in the response surface of these equations as a functions of parameters a and b.

A FitzHugh-Nagumo response surface



The tank reactor model

- 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}.

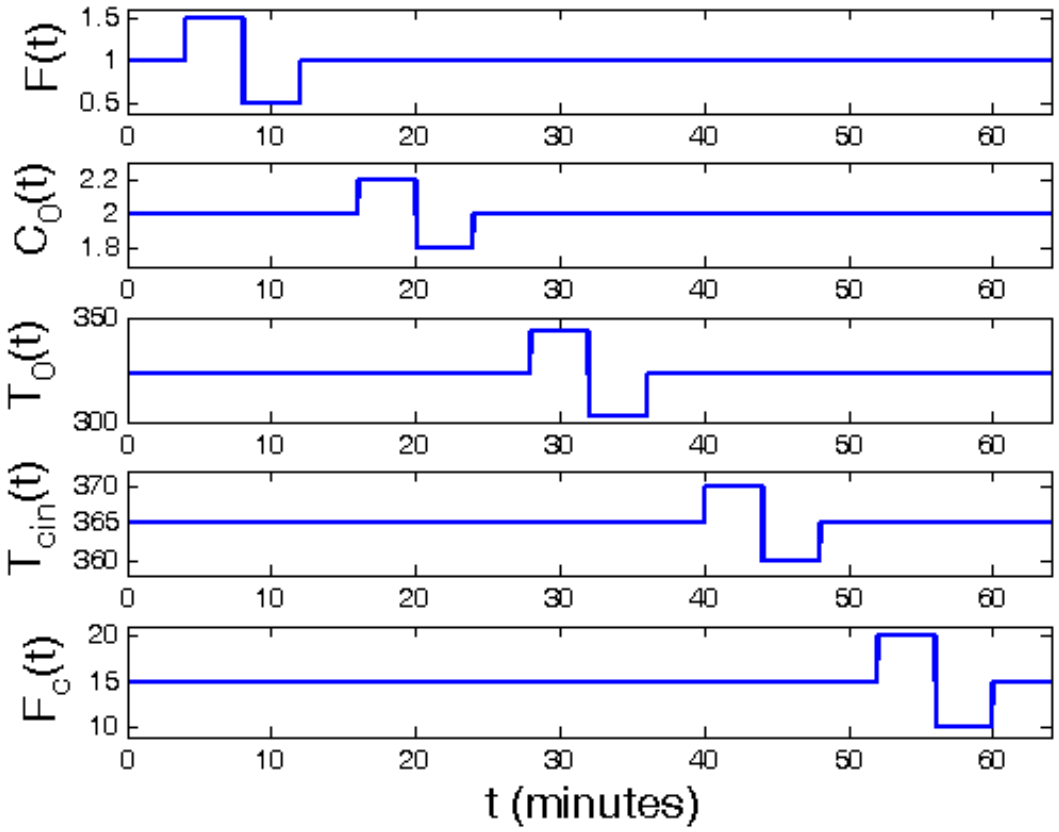
The tank reactor coefficients

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

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

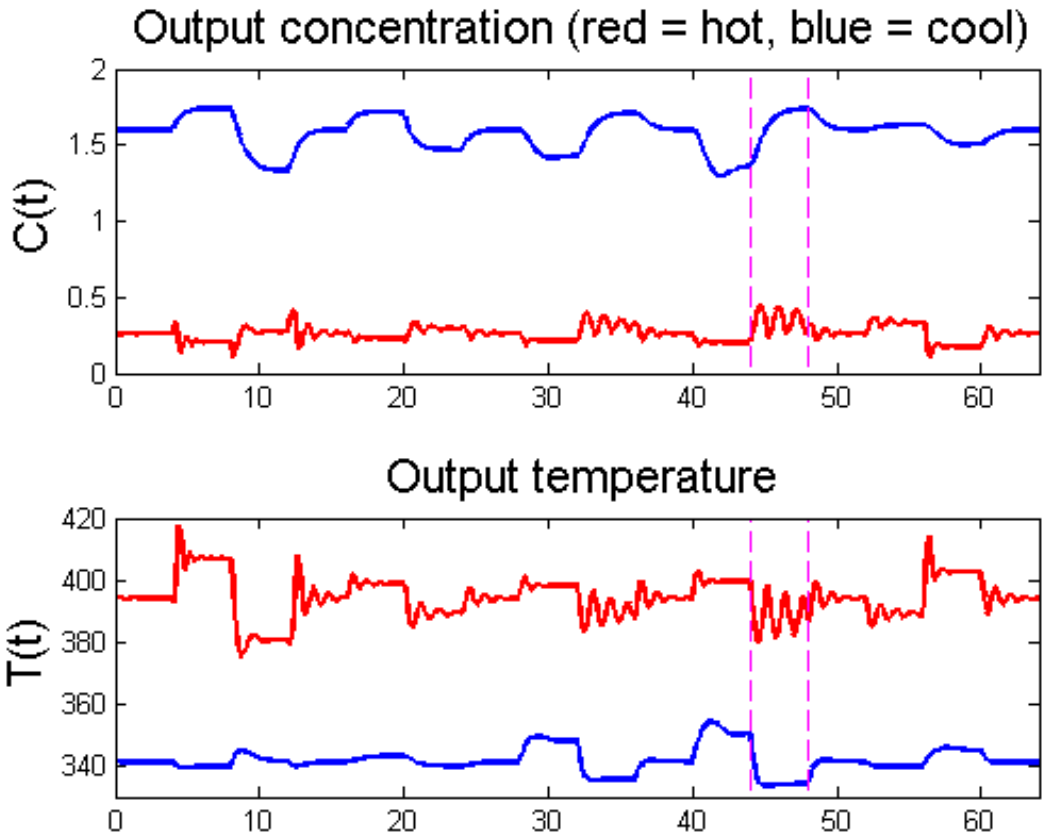
- These functions depend on two paired unknown parameters:
  - κ and τ
  - a and b

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?

Outline

An overview

- For each variable x_i in $\mathbf{x}$, we define a basis function expansion $\mathbf{c}_i'\phi_i$, where $\mathbf{c}_i$ and ϕ_i are a coefficient vector and a vector of basis functions, respectively.
- Over 400 basis functions per variable are used to capture the sharp variation in outputs.
- A data-fitting criterion $F(\mathbf{y}|\mathbf{x})$ is chosen that measures the fidelity of $\mathbf{x}$ to the data in vector $\mathbf{y}$, and also to the differential equations.
- The extent to which $\mathbf{x}$ is a solution of the differential equation system is assessed by the use of additional penalty terms, and
- the relative balance between these two desiderata is controlled by a set of smoothing parameters.

Structural and nuisance parameters

- There are two classes of parameters to estimate:
  - the parameters θ defining the equation, such as the four reaction kinetics parameters in the CSTR equations
  - the coefficients $\mathbf{c}_i$ defining each basis function expansion.
- The equation parameters are *structural* in the sense of being of primary interest.
- The coefficients $\mathbf{c}_i$ are *nuisance* parameters because they are not of direct interest and
- because their numbers are apt to vary with the length of the observation interval, density of observation, and other factors.
- As a rule, the number of nuisance parameters can be orders of magnitude larger than the number of equation parameters, with a ratio of about 200 applying in the CSTR problem.

Eliminating nuisance parameters

- Nuisance parameters are removed from the problem by defining them as *functions* $\mathbf{c}_i(\theta)$ of the structural parameters using a modified profiling procedure.
- The fitting criterion is then optimized with respect to the structural parameters θ alone.
- An analytic expression for the gradient is developed using the Implicit Function Theorem.
- Compared to marginaling out the nuisance parameters using MCMC, this process is
  - much faster,
  - much more stable, and
  - much easier to program.

The data fitting criterion

$$\mathrm{SSE}(\mathbf{c}|\mathbf{y}) = \sum_{i}^{n} w_i \|\mathbf{y}_i - x_i(\mathbf{t}_i)\|^2.$$

Weights w_i are defined to compensate for differences in scale in the variables.

Assessing fidelity to the equations

- x_i solves the corresponding differential equation if

$$L_i(x_i) = Dx_i - f_i(\mathbf{x}, \mathbf{u}, t|\theta) = 0.$$

- A measure of fidelity to the equation is

$$\texttt{PEN}_i(\mathbf{x}) = w_i \int [L_i(x_i)]^2 dt.$$

- These are combined into the composite penalty term

$$\texttt{PEN}(\mathbf{c}|\theta, \boldsymbol{\lambda}) = \sum_i^n \lambda_i \texttt{PEN}_i(\mathbf{x})$$

- PEN depends on θ through operator **L**.

The inner optimization for estimating $\mathbf{c}(\theta)$

- Each time θ is changed, we optimize

$$G(\mathbf{c}|\theta, \boldsymbol{\lambda}) = \texttt{SSE}(\mathbf{c}|\mathbf{y}) + \texttt{PEN}(\mathbf{c}|\theta, \boldsymbol{\lambda})$$

- This profiling process *implicitly* defines the estimating function $\mathbf{c}(\theta)$.
- As $\lambda_i \to \infty$, variable x_i is forced to satisfy the differential equation more and more exactly.

The outer optimization for estimating θ

- We optimize
$$F(\mathbf{c}(\theta)|\boldsymbol{\lambda}) = \mathtt{SSE}(\mathbf{c}(\theta)|\mathbf{y})$$
- No penalty term is needed here because $\mathbf{c}(\theta)$ is already been regularized in the inner optimization.
- The iterations are greatly accelerated by computing the gradient and Hessian using the *Implicit Function Theorem*.

Outline

An overview of interval estimation

- To a first order of approximation, we can approximate $\theta(\mathbf{y}^*)$ evaluated at an alternative observation $\mathbf{y}^*$ by

$$\begin{aligned}\theta(\mathbf{y}^*) - \theta(\mathbf{y}) &\approx \frac{d\theta}{d\mathbf{y}}(\mathbf{y}^* - \mathbf{y}) \\ &= [D^2_{\theta} F(\hat{\theta}, \hat{\mathbf{c}}|\mathbf{y})]^{-1} D^2_{\theta,\mathbf{y}} F(\hat{\theta}, \hat{\mathbf{c}}|\mathbf{y})(\mathbf{y}^* - \mathbf{y}).\end{aligned}$$

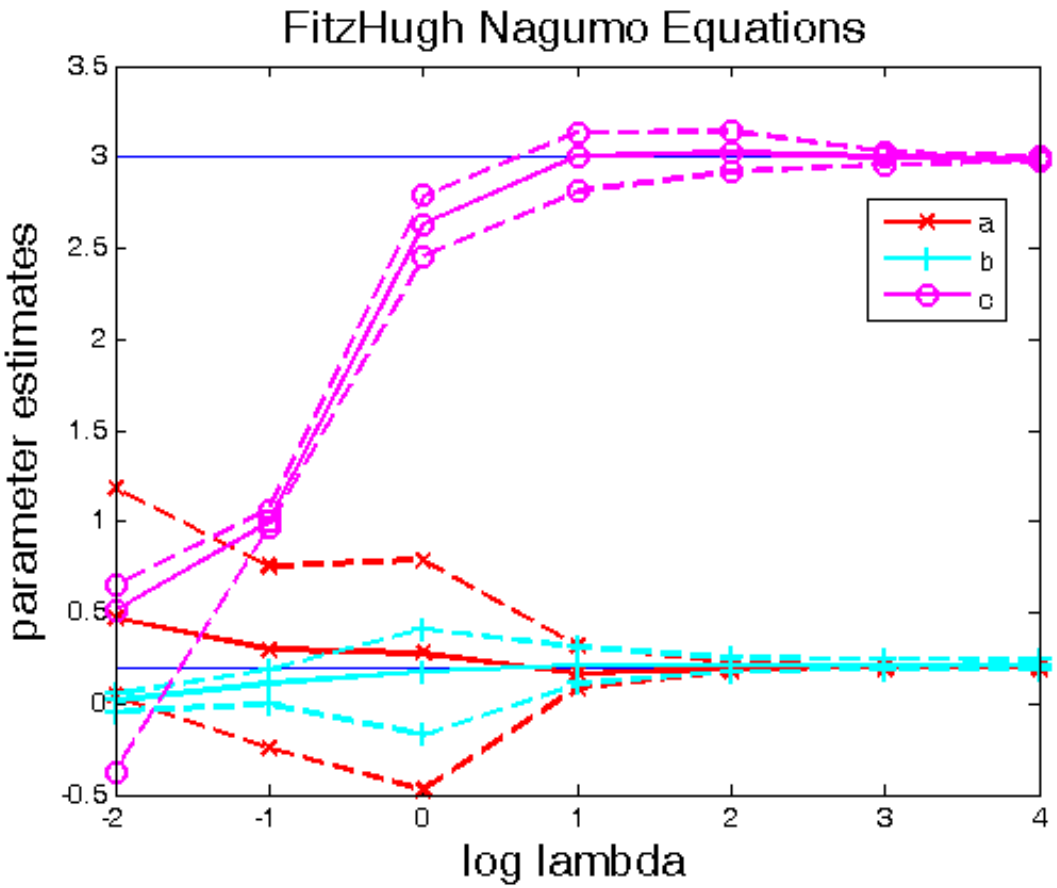
- The derivatives involved can also be computed using the Implicit Function Theorem.
- Sampling variance of θ is then obtained using the Delta method.
- An analogous procedure is used for the variance of $\mathbf{c}(\theta)$.

Outline

Results for the Fitzhugh-Nagumo equations

- The solution to be estimated was determined by $\{a, b, c\} = \{0.2, 0.2, 3\}$ and initial values $\{V(0), R(0)\} = \{-1, 1\}$.
- The paths were measured at 0.05 time units on the interval [0,20].
- Noise was then added to these values with standard deviation 0.5.
- 500 simulated samples were analyzed.

Parameter estimate variation



Both bias and sampling variance decrease as $\lambda_i \to \infty$.

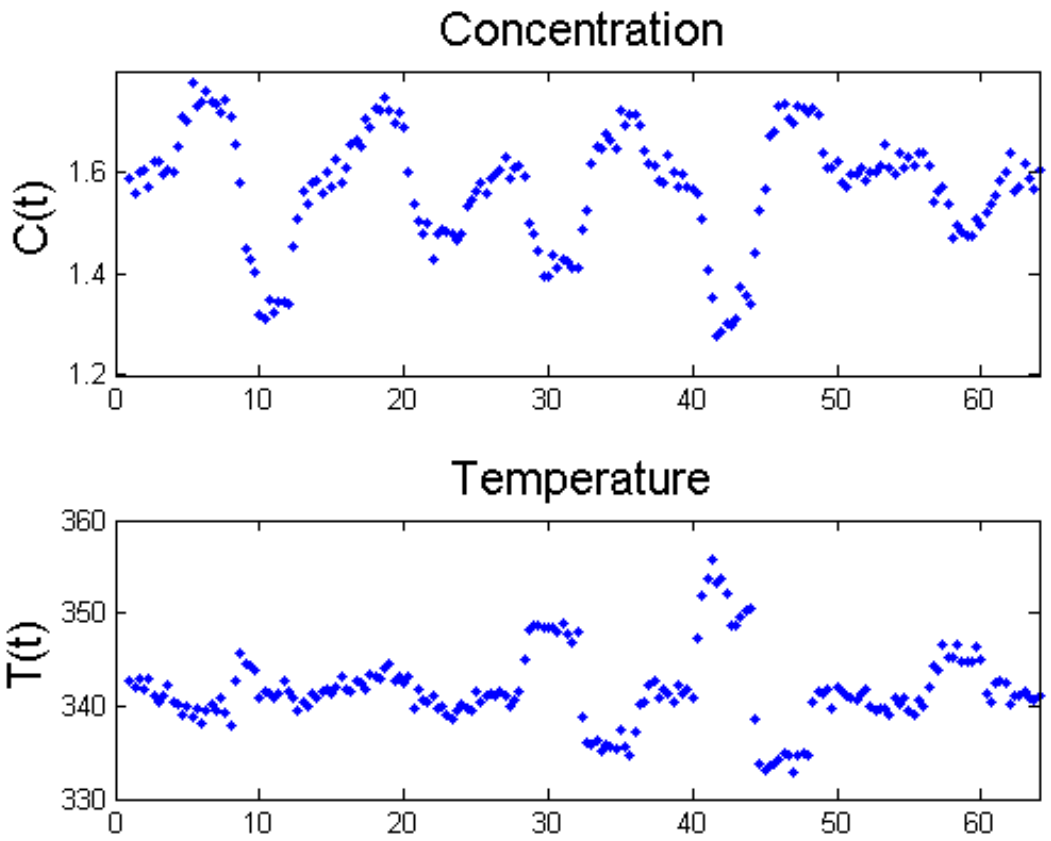
Summary statistics for parameter estimates

	a	b	c
True value	0.2000	0.2000	3.0000
Mean value	0.2005	0.1984	2.9949
Std. Dev.	0.0149	0.0643	0.0264
Est. Std. Dev.	0.0143	0.0684	0.0278
Bias	0.0005	-0.0016	-0.0051
Std. Err.	0.0007	0.0029	0.0012

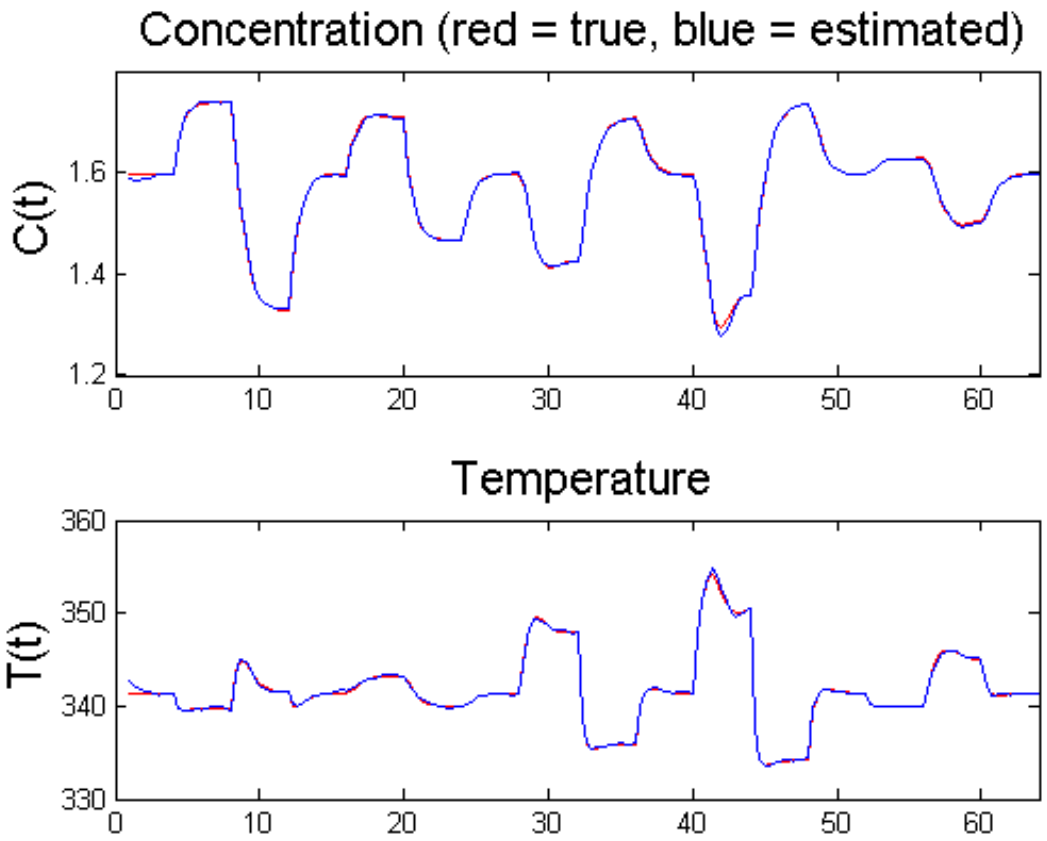
Simulations for the tank reactor equations

- 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.

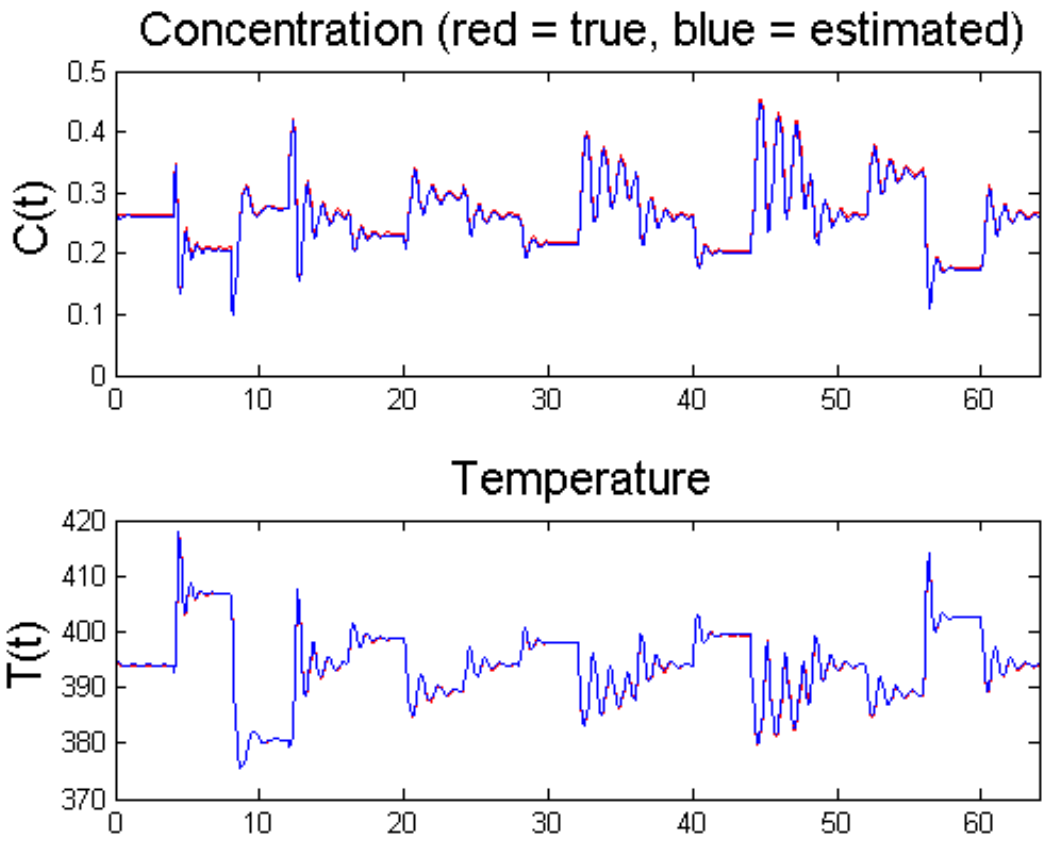
A typical set of tank reactor data



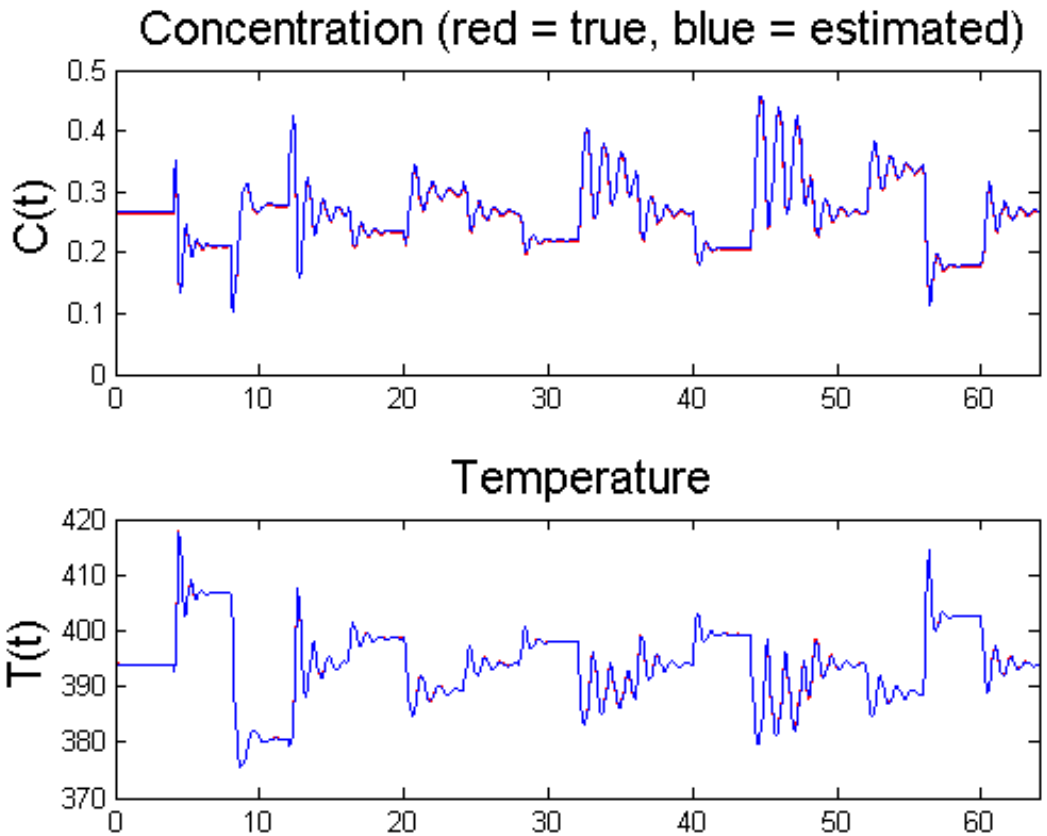
Path estimations, cool mode



Path estimations, hot mode



Path estimations, hot mode



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

Summary statistics for parameter estimates

	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

Outline

The nylon experiment

- Nylon and other polymers are created by a chemical reaction in which molecules with two special types of endings chain together to form long molecules.
- The reaction requires water to form the molecules.
- The long molecules can also be broken up, releasing water.
- Temperature and water are critical control variables.
- There were five runs of the experiment at different temperature settings.
- These data were collected in the laboratory of Prof. K. MacAuley of the Dept. of Chemical Engineering at Queen's University, Kingston, Canada.
- The concentration measurements for variables A and C cost about \$30,000 to obtain.

The variables in the nylon equations

- A: molecules with an amine group end (measured)
- C: molecules with a carboxyl group end (measured)
- L: Nylon, a long chain of molecules (a polymer) (not measured)
- W: Water, indirectly adjusted in the experiment
- The variables are related by the mass balance equation

$$A + C \rightleftharpoons L + W$$

Nylon equations

$$DA = DC = -k_p(T)\left(CA - \frac{LW}{K_a(T)}\right)$$

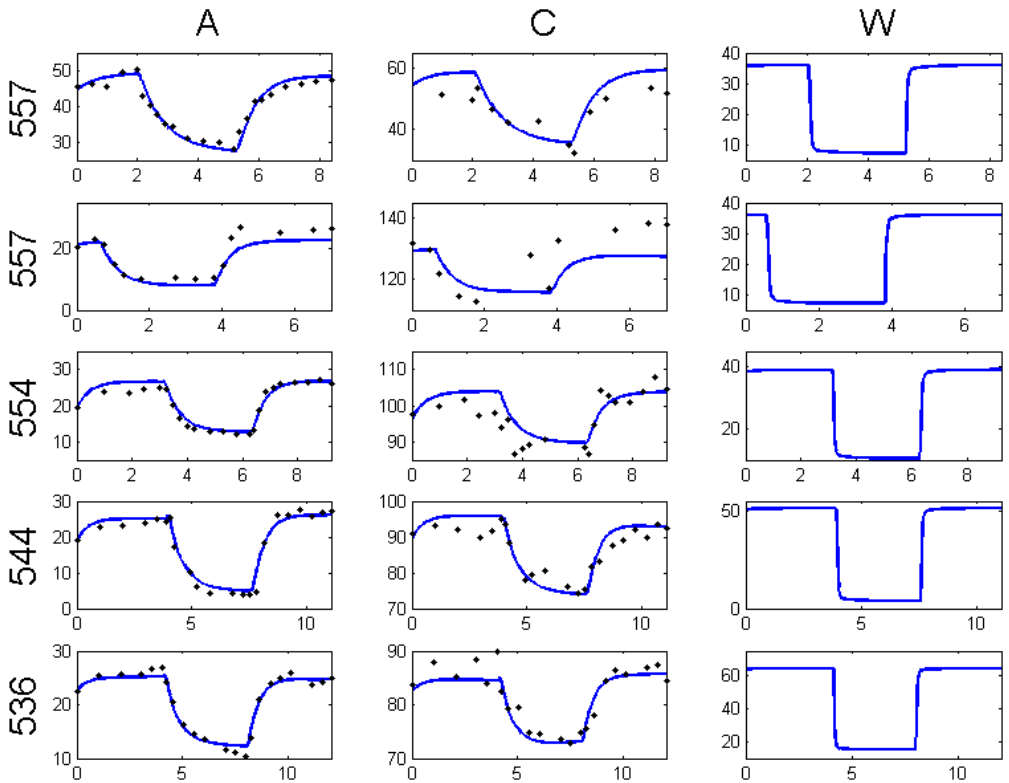
$$DW = k_p(T)\left(CA - \frac{LW}{K_a(T)}\right) - k_m\left(W - W_{eq}\right)$$

$$k_p(T) = k_{p0} \exp\left[-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

$$K_a(T) = \left[\frac{1 + \alpha W_{eq}}{\gamma_w/\gamma_{w0}}\right] K_{a0} \exp\left[-\frac{\triangle H}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

- variables and known constants are black
- parameters to be estimated are in red
- experimentally manipulated and measured constants and variables are in blue

Fits to the data



Outline

Software

- All the results were computed in Matlab.
- 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 the URL: http://www.functionaldata.org.

References

- A paper is available from the URL: http://www.functionaldata.org.
- J. O. Ramsay and B. W.Silverman (2005) *Functional Data Analysis*, Second Edition. New York: Springer.