Investigating Process Behaviour Using Functional Data Analysis

Jim McLellan
Department of Chemical Engineering
Queen’s University
Kingston, ON
Canada
mclellnj@chee.queensu.ca
Functional Data Analysis (FDA)

- FDA is an evolving statistical framework in which responses are profiles, not points
Functional Data Analysis (FDA)

- FDA is an evolving statistical framework in which responses are profiles, not points

- FDA has potential for making it easier to estimate dynamic and steady-state models for chemical processes
Functional Data Analysis (FDA)

- FDA is an evolving statistical framework in which responses are profiles, not points.
- FDA has potential for making it easier to estimate dynamic and steady-state models for chemical processes.
- Potential of FDA needs to be assessed. Could FDA help solve your model estimation problems?
Functional Data Analysis

- Motivation
- Functional data analysis

2 Stories –

- Estimating differential equation models using principal differential analysis
- Obtaining reactor settings for desired molecular weight distributions using functional regression
Motivation

When modeling and controlling chemical processes, we frequently encounter responses that are functional – functions of an independent variable such as time or molecular weight

- time traces (time series) – most frequently encountered in process monitoring and control

- species distributions – e.g., polymer molecular weight distributions, particle size distributions

- spectra

The data object is a function of one or more independent variables.

*How do we work with these responses?*
Ways of viewing functional data

Functional data objects

Sampled data series

Multiple responses

$y_1(t), y_2(t)$

$y_1(T), y_1(2T), y_1(3T), \ldots$

$y_2(T), y_2(2T), y_2(3T), \ldots$

$y_{t1} = y(t_1)$

$y_{t2} = y(t_2)$

$y_{t3} = y(t_3)$
Ways of viewing functional data

Sampled data series
- standard approach for dealing with time traces – time series – predominant approach in control modeling and analysis
- typically assume uniform sampling – measurements are available at regular intervals
- models are typically discrete-time – difference or recursion equations
e.g., \( y_{k+1} = a_1 y_k + b_1 u_{k-1} + e_k \)

Multiple responses at discrete points from continuous models
- typical approach in fundamental process modeling – e.g., predicting polymer molecular weight distributions at different chain lengths, chemical concentrations at different times
- responses are measurements at particular times or positions

Functional data objects
- the data object is a continuous curve
Goals of this talk

1. Demonstrate how FDA techniques can be used to model dynamic and steady-state process behaviour

2. Provide an overview of two relevant FDA techniques
   - Principal differential analysis
   - Functional regression

3. Comment on the relationship between FDA approaches and existing approaches
Functional Data Analysis (FDA)

... is a statistical framework in which the data object is a function of one or more independent variables

- Techniques have been developed and used for analyzing handwriting, lip motion, horse gait data, analyzing weather data, eye-hand response times, ...
- FDA toolbox for Matlab available free from Jim Ramsay web site (www.psych.mcgill.ca/faculty/ramsay.html)

Datasets consist of collections of functional observations

- Multiple observations (realizations) of same response function – e.g., temperature profiles for different runs in a batch reactor – \{y_1(t), y_2(t),..., y_N(t)\}
- Observations of different functional responses – e.g., time traces for valve input and temperature – \{u(t), y(t)\}
Functional Data Analysis (FDA)

FDA is a statistical framework for functional data

- Standard summary measures defined - examples
  
  - Sample average
    \[
    \bar{y}(t) = \frac{1}{N} \sum_{i=1}^{N} y_i(t)
    \]
  
  - Sample variance
    \[
    s^2(t) = \frac{1}{N-1} \sum_{i=1}^{N} (y_i(t) - \bar{y}(t))^2
    \]
  
  - Sample covariance, sample cross-covariance, sample covariance matrix

Note that the result is a function of the independent variable – average function, variance function.
Functional Data Analysis

Concept
- data objects are continuous functions of an independent variable

\[ y_2(t) \]
\[ y_1(t) \]

Practice
- observations are typically taken at discrete intervals – not necessarily uniform – and functional observations are constructed using appropriate basis functions - smoothing

\[ y_{2\sim}(t) = \sum_{j=1}^{N_{\text{basis}}} c_{2,j} \varphi_j(t) \]
\[ y_{1\sim}(t) = \sum_{j=1}^{N_{\text{basis}}} c_{1,j} \varphi_j(t) \]

\[ \varphi_j(t) \] are basis functions e.g., splines, polynomials, sinusoids
Two Stories about Functional Data Analysis

**Story #1** – Estimating parameters in differential equation models using **Principal Differential Analysis (PDA)**
- M.Sc. student Andy Poyton

**Story #2** – Obtaining reactor settings for desired molecular weight distributions using **Functional Regression**
What is Principal Differential Analysis (PDA) and Why are We Using it?

• PDA is a technique from Functional Data Analysis (FDA)

• PDA exploits natural smoothness of processes
  – enables analysis using derivatives and rate behaviour

• Does PDA have potential for estimating parameters in dynamic models?
Estimating Differential Equation Models Using Principal Differential Analysis (PDA)

- Parameter estimation in fundamental dynamic models is difficult and iterative using traditional techniques
- Most computational effort arises from repeated numerical solution of ODEs
Parameter estimation in fundamental dynamic models is difficult and iterative using traditional techniques.

Most computational effort arises from repeated numerical solution of ODEs.

**Principal Differential Analysis (PDA)**
- Uses smoothing of dynamic data to provide time derivatives
- Converts ODE estimation to an algebraic problem
- Eliminates need to solve dynamic equations numerically
- Estimates parameters so that differential equation models are satisfied by observed functional responses

\[
\frac{dy}{dt} - \alpha y - \beta u = 0
\]
Smoothing

The first step in any functional data analysis is to obtain a functional representation of the data

- Discrete data are represented as a function of time using basis functions, rather than as a collection of discrete set of points

Smoothing using basis functions
The first step in any functional data analysis is to obtain a functional representation of the data

- Discrete data are represented as a function of time using basis functions, rather than as a collection of discrete set of points

- Functional data are a set of curves with one pair of curves for each run \( \{y(t), u(t)\} \)

- Functional data objects are linear combinations of known basis functions (e.g., B-splines)

\[
y(t) = \sum_{k=1}^{K} c_k \phi_k(t)
\]
Smoothing

The first step in any functional data analysis is to obtain a functional representation of the data

- Discrete data are represented as a function of time using basis functions, rather than as a collection of discrete set of points

- Functional data are a set of curves with one pair of curves for each run \( \{y_{\sim}(t), u_{\sim}(t)\} \)

- Functional data objects are linear combinations of known basis functions (e.g., B-splines)

\[
y_{\sim}(t) = \sum_{k=1}^{K} c_k \phi_k(t)
\]
Principal Differential Analysis

- Assess model fit by how well $y_\sim(t)$ and $u_\sim(t)$ satisfy the ODE
  - E.g., for
  \[
  \frac{dy}{dt} - \alpha y - \beta u = 0
  \]
  substitute in the smoothed curves and their derivatives, and the predicted parameters
  \[
  \frac{dy_\sim(t)}{dt} - \hat{\alpha}y_\sim(t) - \hat{\beta}u_\sim(t) = \varepsilon(t)
  \]
  producing a residual curve $\varepsilon(t)$

- Estimate parameters $\hat{\alpha}$ and $\hat{\beta}$ to minimize the squared residuals
  \[
  \min_{\hat{\alpha},\hat{\beta}} \{ \int \varepsilon(t)^2 \, dt \} \]
PDA Parameter Estimation is a two-step process

1. Data are smoothed and put in functional form

\[
\min_{c_k} \sum_i \left( y(t_i) - y_\sim(t_i) \right)^2
\]

- \(c_k\) values are spline coefficients

\[
y_\sim(t) = \sum_{k=1}^{K} c_k \phi_k(t)
\]

- Roughness penalties on higher-order derivatives of \(y_\sim(t)\) can be used to adjust smoothness

2. Squared residual curve is minimized to obtain parameter estimates \(\hat{\alpha}\) and \(\hat{\beta}\)

\[
\min_{\hat{\alpha}, \hat{\beta}} \left\{ \int \left[ \frac{dy_\sim(t)}{dt} - \hat{\alpha} y_\sim(t) - \hat{\beta} u_\sim(t) \right]^2 dt \right\}
\]

\(\varepsilon(t)\)

★ No numerical solution of the ODE is required. Differentiate \(y_\sim(t)\) to obtain \(\frac{dy_\sim}{dt}\)
PDA Parameter Estimation Example

• Linearized CSTR model to produce example with known analytical solution

• 1 linear ODE, 2 nonlinear parameters to estimate

• Generate simulated responses with white noise

• Compared PDA estimation with conventional nonlinear least squares estimation
  – Computing time
  – Precision and bias of parameter estimates
  – Precision and bias of model predictions
Linearized Non-Isothermal CSTR

Linearized model with nonlinear parameter dependence (Marlin, 2000)

\[
\frac{dC_A}{dt} = -\left(\frac{F_s}{V} + k_0 \frac{E}{R T_s}\right)C_A - \left(k_0 \frac{E C_{As}}{R T_s^2} e^{-E/RT_s}\right)T
\]

- Input – reactor temperature \( T \) (perfect T control)
- Constant inlet concentration and flow
- Parameters are \( k_0 \) and \( E/R \). True values are \( k_0 = 1.0 \times 10^{10} \text{ min}^{-1} \), and \( E/R = 8330.1 \text{ K} \)
Linearized Non-Isothermal CSTR

Linearized model with nonlinear parameter dependence (Marlin, 2000)

\[
\frac{dC_A}{dt} = -\left( \frac{F_s}{V} + k_0 \frac{E}{R T_s} \right) C_A - \left( k_0 \frac{E}{R T_s^2} e^{-E/RT_s} \right) T
\]

- Input – reactor temperature $T$ (perfect T control)
- Constant inlet concentration and flow
- Parameters are $k_0$ and $E/R$. True values are $k_0 = 1.0 \times 10^{10} \text{ min}^{-1}$, and $E/R = 8330.1$ K

First PDA Attempt

<table>
<thead>
<tr>
<th>PDA</th>
<th>1.591</th>
<th>8.580</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLS</td>
<td>0.997</td>
<td>8.329</td>
</tr>
<tr>
<td>True</td>
<td>1.000</td>
<td>8.330</td>
</tr>
</tbody>
</table>

Not so good!
What was wrong?

- Poor spline fit to data
- Cubic B-spline requires first and second-order derivatives to match at intersection of adjacent intervals
- Can’t match sharp corner – problem for all first-order models with step inputs
Fixing the spline fit: knot placement

- Cubic B-splines require continuous 1\textsuperscript{st} and 2\textsuperscript{nd} derivatives at knots, giving poor fits near sharp corners
- Multiple knots at or near sharp corners (3 for cubic B-splines) alleviate this problem and give better spline fit

Examined effects of
- Coincident knots at sharp corners
- Closely spaced (not coincident) knots near sharp corners
- Improved spline fits on parameter estimates in ODE model
Results of PDA with Coincident Knots

- Simulations – step down in T; 30 runs simulated
- Base case: cubic B-splines with uniform knot spacing at 0.1 min intervals
  
<table>
<thead>
<tr>
<th></th>
<th>$k_p/10^{10}$ (min$^{-1}$)</th>
<th>$E/R/10^3$ (K)</th>
<th>Run time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original PDA</td>
<td>1.591</td>
<td>8.580</td>
<td>0.219</td>
</tr>
<tr>
<td>PDA w/ coincident knots</td>
<td>0.998</td>
<td>8.329</td>
<td>0.219</td>
</tr>
<tr>
<td>PDA w/ close knots</td>
<td>1.216</td>
<td>8.438</td>
<td>0.234</td>
</tr>
<tr>
<td>Traditional NLS</td>
<td>0.997</td>
<td>8.329</td>
<td>0.750</td>
</tr>
<tr>
<td>True Parameter Values</td>
<td>1.000</td>
<td>8.330</td>
<td>-</td>
</tr>
</tbody>
</table>

- Spline smoothing was much improved using coincident knots
- Parameter estimates and model predictions were as good as those obtained using conventional nonlinear least squares
- PDA with coincident knots required less runtime
- Adding coincident knots was significantly more effective than adding closely-spaced knots
Smoothing and PDA

• The smoothing step is critical to the success of the parameter estimates in PDA

• Traditional spline-fitting approach is to penalize higher-order derivatives of the smoothed functions
  – Produces poor results for responses with sharp corners or oscillations

\[ J = \sum_i \left( y(t_i) - y_{\sim}(t_i) \right)^2 + \lambda \int \left( \frac{d^2 y_{\sim}(t)}{dt^2} \right)^2 dt \]

• We can improve the smoothing step by incorporating underlying ODE model structure into the spline fitting
  – Use ODE model with parameter guesses to define roughness penalty for spline smoothing

Profile-based PDA
Profile-Based PDA

- Profile-based PDA objective function for fitting splines

Find spline coefficients that minimize

\[
\sum_i (y(t_i) - y_\sim(t_i))^2 + \lambda \int (\varepsilon(t))^2 dt
\]

Residuals between spline fit and data points

Ensure spline matches data

where

\[
\varepsilon(t) = \frac{dy_\sim(t)}{dt} - \hat{\alpha}y_\sim(t) - \hat{\beta}u_\sim(t)
\]

“Shrink” spline fit to agree with mechanistic model
Profile-Based PDA – Use the ODE to Help Spline Fitting

- Conventional PDA is a 2-step procedure
  1. Fit splines to data
  2. Estimate ODE model parameters using fitted splines

- Profile-based PDA iterates between smoothing and parameter estimation steps
  1. Fit splines using ODE model (with current parameter estimates) to define a roughness penalty
     - Ensures spline fits are smooth and physically reasonable
  2. Update model parameters using fitted splines
     - Adjust ODE parameters to improve ODE model fit to splines
  3. Iterate between steps 1 and 2 until convergence

- The performance of profile-based PDA is currently being investigated using a nonlinear, 2-state CSTR model
PDA Summary

- Functional Data Analysis treats data as curves rather than points.

- Principal Differential Analysis can fit parameters in models without numerical solution of the ODE; effort is in fitting splines to the data.
  - Linearized CSTR model with 2 non-linear parameters.

- Profile PDA uses ODE as spline-fitting penalty.
  - Can be applied to MIMO ODE models with many parameters.
  - Example: nonlinear CSTR model with 2 ODEs and 4 nonlinear parameters.

- Current Limitation: all state variables must be measured.
  - Ideas on how to extend profile PDA to models with unmeasured states.
  - Saeed Varziri – new Ph.D. student.
Two Stories about Functional Data Analysis

Story #1 – Estimating parameters in differential equation models using Principal Differential Analysis (PDA)

Story #2 – Obtaining reactor settings for desired molecular weight distributions using Functional Regression
Obtaining reactor settings for desired MWDs using Functional Regression

• Polymer molecular weight distributions (MWDs) are important because they influence end-use and processing properties

• MWDs are functional observations, in which weight fraction is a function of molecular weight (or log(MW))

• Conventional approaches for modeling and predicting MWDs include
  – characterization using moments
  – detailed mechanistic models to predict fractions for each chain length
  – discretization and treatment as multi-response estimation problems

• Issues
  – loss of information vs. complexity
  – problem conditioning

• Alternative is to treat the MWDs as curves, and use techniques from Functional Data Analysis (FDA)
Functional Regression

- Models in which factors or responses are functional

- Example – functional response $y(r)$ depends on non-functional factors $x_1$ and $x_2$

$$
y(r) = \beta_0(r) + \beta_1(r)x_1 + \beta_2(r)x_2 + \epsilon
$$

In the MWD modeling example, the response is functional and the factors are non-functional, so parameters in the model are functions of the independent variable $r$, which is $\log(MW)$
Functional Regression

- Least squares estimation criterion – minimize integral squared error between predicted and observed response functions

\[
\min_{\beta_0(r), \beta_1(r), \beta_2(r)} \int_{r_{\text{min}}}^{r_{\text{max}}} (y(r) - \hat{y}(r, \beta(r)))^2 \, dr
\]

- Solution – can be determined by expressing parameter functions using basis functions
Functional Regression for MWD Analysis

• Estimate an empirical model to predict the effect of isothermal reactor temperature (T) and initial initiator concentration \([I]_0\) on the resulting MWD for bulk polymerization of styrene in a batch reactor

• Response (differential weight fraction, \(y\)) is functional, while factors \(T\) and \([I]_0\) are non-functional

• Synthetic data generated for a 2^2 factorial design in \(T\) and \([I]_0\) using a fundamental model
Functional Regression for MWD Analysis

Steps

– Smooth the raw MWD data using spline basis functions

– Transform MWD responses to ensure model can’t give negative predictions

– Fit main-effects-plus-two-factor-interaction model using transformed responses
  • Obtain parameter estimate curves that are functions of log(MW)

– Transform back to original coordinates
  (differential weight fraction vs. log(MW))

– Use model to determine $T$ and $[\dot{\gamma}]_0$ to produce MWD near target curve
Results

- Simulated MWDs are available for 4 combinations of T and $[I]_0$

- Splines gave good fits to the original data plots using 14 basis functions
  - 4th order B-splines were used with knots at non-uniform intervals

Original data $y(r)$  
Transformed responses $\ln(y(r))$  
Transformed responses $d(\ln(y))/dr$ computed from spline fits
Model

\[
\frac{d(\ln(y))}{dr} = \beta_0(r) + \beta_1(r) T + \beta_2(r) [I]_0 + \beta_{12}(r) T[I]_0 + \varepsilon
\]
Model Predictions

- Regression model fits simulated data well

Observed (blue) and predicted (red) MWDs for $2^2$ factorial design

Main-effects-plus-two-factor-interaction model
Using the model to guide further experimentation

Specified two target distributions, and used model to determine optimal operating conditions

- Conducted as a “blind test” – given target distributions without knowledge of actual operating conditions
- Required operating conditions were correctly identified to within 7% of the true value in both cases

Operating conditions computed to minimize difference between target and predicted MWDs

- solved using Nelder-Mead Simplex algorithm
What we learned

• Influence of Temperature on MWD
  – Increasing T reduces average molecular weight
  – Increasing T increases the breadth of the MWD

• Influence of Initiator on MWD
  – Increasing $[I]_0$ reduces average molecular weight
  – Increasing $[I]_0$ can produce a shoulder on the MWD – asymmetric influence

• Model was useful for selecting operating conditions to produce desired MWDs
Two Stories about Functional Data Analysis

Story #1 – Estimating parameters in differential equation models using Principal Differential Analysis (PDA)

Story #2 – Obtaining reactor settings for desired molecular weight distributions using Functional Regression
Functional Data Analysis (FDA)

- FDA is an evolving statistical framework in which responses are profiles, not points

- FDA has potential for making it easier to estimate dynamic and steady-state models for chemical processes

- FDA is useful for both fundamental and empirical models

- Potential of FDA needs to be assessed. Could FDA help solve your model estimation problems?

- Better understanding is required before FDA will be widely adopted by the chemometrics and chemical engineering community
Acknowledgements

Research collaborations

Jim Ramsay, Psychology Dept., McGill

**Estimating differential equation models using principal differential analysis**
Kim McAuley, Queen’s
Andy Poyton – MSc student, Queen’s
Saeed Varziri – PhD student, Queen’s

**Functional regression for modeling MWDs**
Robin Hutchinson – Queen’s
Hana Sulieman – American University of Sharjah, United Arab Emirates
David Bacon – Queen’s
Acknowledgements

Financial support

Natural Sciences and Engineering Research Council of Canada
MITACS
Queen’s University