Differential Equations & Functional Data Analysis

Parameter Estimations for Differential Equations

Whitney Huang

CLEMSON UNIVERSITY

wkhuang@clemson.edu

Clemson ENVR Group, November 11, 2020

Agenda

Motivation

The Estimation Procedure of Ramsay et al. 2007

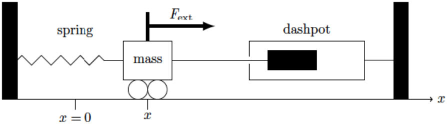
Example: Groundwater Levels

Differential Equations

A differential equation is an equation that relates some function of one or more variables with its derivatives.

Ordinary differential equation:

$$m\ddot{x} + b\dot{x} + kx = F_{\text{ext}}$$



Partial differential equation:

$$\frac{\partial u}{\partial t} - \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0$$

Differential Equations

- Widely used to model dynamic systems
  - Maxwell's equations in electromagnetism
  - Navier-Stokes equations in fluid dynamics
  - The Black-Scholes PDE in Economics
- Forward problem (i.e. solving differential equations) has been studied extensively by mathematicians

Suppose a given data set can be reasonably model by a differential equation but with unknown coefficients. Can we make a statistical inference?

Differential Equations

- Widely used to model dynamic systems
  - Maxwell's equations in electromagnetism
  - Navier-Stokes equations in fluid dynamics
  - The Black-Scholes PDE in Economics
- Forward problem (i.e. solving differential equations) has been studied extensively by mathematicians

> Suppose a given data set can be reasonably model by a differential equation but with unknown coefficients. Can we make a statistical inference?

Differential Equations & Functional Data Analysis

Most dynamic systems defined by the solutions of their differential equations are not fit to data, they intend to capture gross shape features in the specified context. **However**, ...

- Solutions of differential equations are functions
- We can treat the data as an approximated solution of the corresponding differential equation with unknown coefficients

FDA framework can help for solving this inverse problem

Set-up

Differential equations:

$$f(t, x, \dot{x}, \ddot{x}, \cdots; \boldsymbol{\theta}) = 0, \qquad \text{e.g., } f = \dot{x} + \beta x - \mu = 0$$

Observed data:

$$y_i = x(t_i) + \epsilon_i, \qquad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2), \quad i = 1, \cdots, n.$$

Goal: to estimate the unknown $\boldsymbol{\theta}$ in the differential equation from the data and to quantify the uncertainty of the estimates

The Basic Idea of the Estimation Procedure

1. Use basis function expansion to approximate $x(t)$, i.e., $x(t) = \sum_{k=1}^{K} c_k^T \phi_k(t)$

2. Estimate the coefficients $\boldsymbol{c} = \{c_k\}_{k=1}^{K}$ of the chosen basis functions by incorporating *differential equation defined penalty*

3. Estimate the parameters $\boldsymbol{\theta}$ in the differential equation

4. Choosing the amount of smoothing λ

Basic Function Expansion

$$\hat{x}(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

- Choice of basis function:
  splines are usually the logical choice because of the compact support and the capacity to capture transient localized features

- Number of basis function:
  usually large because it requires not only to approximate $x(t)$ but also its derivatives

Data Fitting Criterion

$$J(\boldsymbol{c}|\boldsymbol{\theta}) = \underbrace{\ell\left(\hat{x}(t_i), y(t_i)\right)}_{\text{data fidelity}} + \lambda \underbrace{\int \left[f(\hat{x}(t);\boldsymbol{\theta})\right]^2 \, dt}_{\text{DE defined penalty}}$$

- The first part of the criterion function is the fidelity of basis function approximation to the data
- the second part is the penalty term with respect to differential equation (DE) given $\boldsymbol{\theta}$
- smoothing parameter λ controls the relative emphasis on these two objectives

The Parameter Hierarchy

There are three classes of parameters to estimate:

- The coefficients $\boldsymbol{c}$ in the basis function expansion
- The parameters $\boldsymbol{\theta}$ defining the differential equation
- The smoothing parameter λ

The Roles of the Three Parameter Levels

- $\boldsymbol{c}$ are not of the direct interest $\Rightarrow$ nuisance parameters
- We are primary interest in $\boldsymbol{\theta}$, the parameters that define the differential equation
- Smoothing parameter λ control the overall complexity of the model
  - $\lambda \to 0 \Rightarrow$ high complexity in $\hat{x}(t)$
  - $\lambda \to \infty \Rightarrow$ low complexity in $\hat{x}(t)$

The Parameter Cascade Algorithm

- $\boldsymbol{c}$ are nuisance parameters are defined as a smooth functions $\boldsymbol{c}(\boldsymbol{\theta}, \lambda)$
- Structural parameters $\boldsymbol{\theta}$ are defined as functions $\boldsymbol{\theta}(\lambda)$ of the complexity parameter
- 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 $\boldsymbol{c}(\boldsymbol{\theta}, \lambda)$ are defined by optimizing the regularized fitting criterion

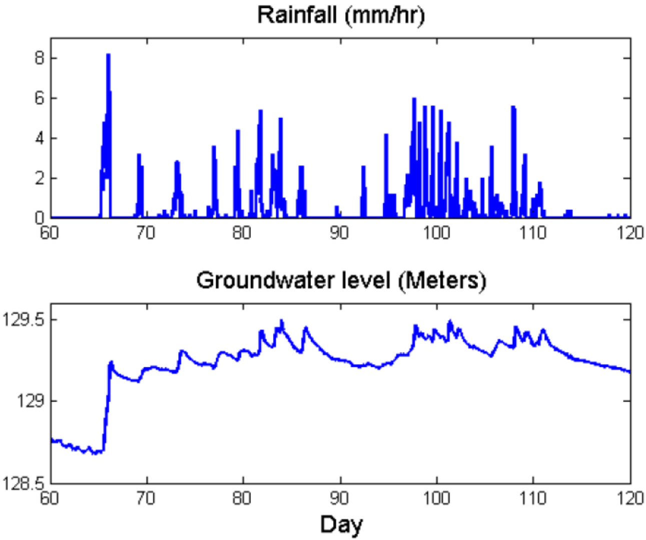
$$J(\boldsymbol{c}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \{y_i - \hat{x}(t_i)\}^2 + \lambda \int [f(\hat{x}(t); \boldsymbol{\theta})]^2 \, dt$$

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

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{y_i - \hat{x}(t_i; \boldsymbol{\theta})\}^2$$

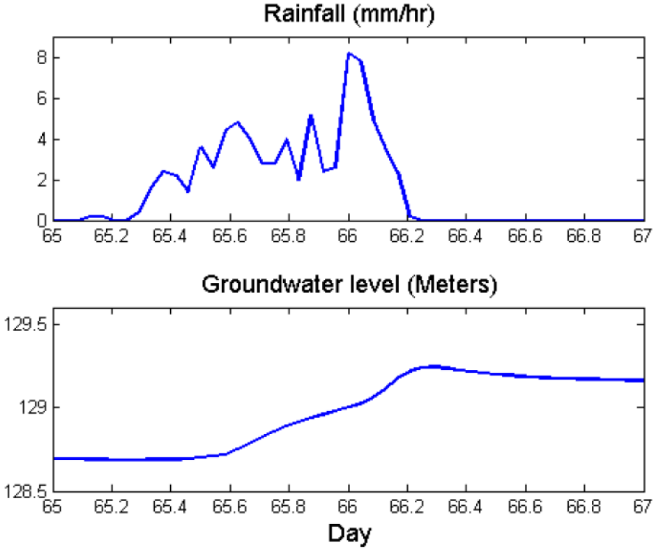
- At the top level, a complexity criterion, is optimized with respect to λ

Rain & Groundwater



Source: Ramsay's slides on "Linear Models for Output-Buffered Systems", 2010 SAMSI

Rain & Groundwater: A Smaller Time Scale



Source: Ramsay's slides on "Linear Models for Output-Buffered Systems", 2010 SAMSI

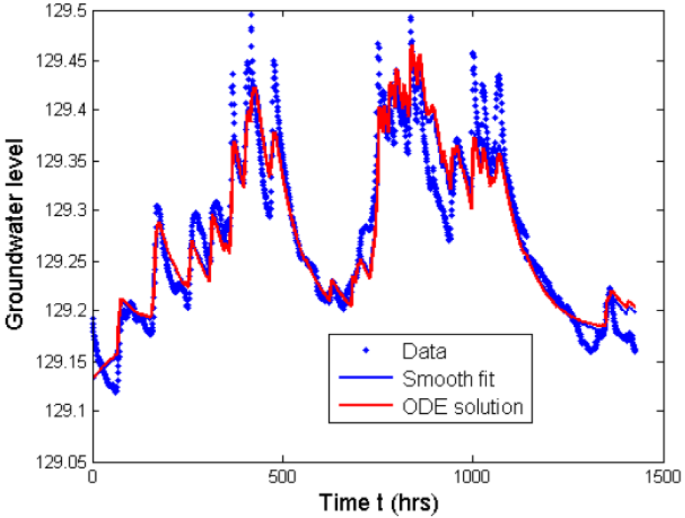
Differential Equation for Groundwater Level

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

where

- β specifies the the rate of change of $G(t)$ with itself
- α defines the impact of $R(t)$
- 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 3 hours

The Constant Coefficient Fit



Source: Ramsay's slides on "Linear Models for Output-Buffered Systems", 2010 SAMSI

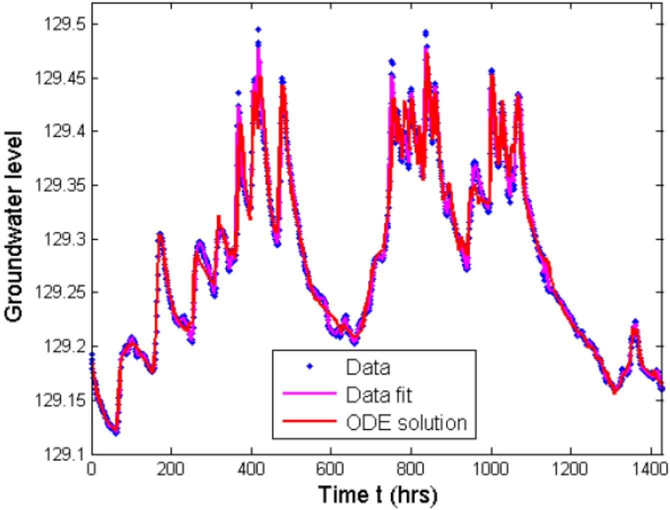
Allowing ODE Coefficients Time Varying

- As groundwater level $G(t)$ changes, the dynamics change, too, because water move through different types of sub-soil structures

- We weren't given sub-soil transmission rates, so we needed to allow $\beta(t)$; $\alpha(t)$ and $\mu(t)$ to vary slowly over time:

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

The Time-Varying Coefficient Fit



Source: Ramsay's slides on "Linear Models for Output-Buffered Systems", 2010 SAMSI