

Analyzing Time Course Microarray Data With Temporal Uncertainty

Stephen C. Billups

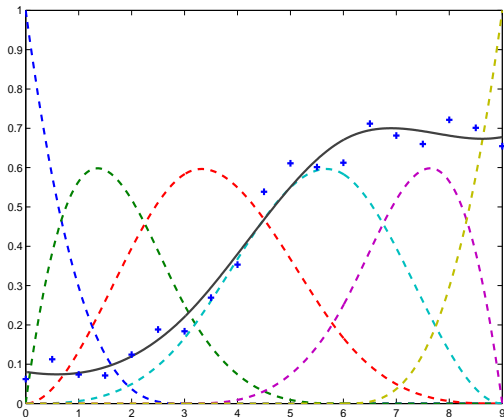
University of Colorado at Denver and Health Sciences Center
Department of Mathematical Sciences

December, 2006

Introduction

- Goal: Understand how gene expression changes over time (relative to a biological process).
- Key Analysis Tool: Spline representations:
$$y_j \approx f(t_j) = c_1 b_1(t_j) + \cdots + c_M b_M(t_j).$$
- Biological Time vs. Clock time.
- Clock time only approximates biological time.
- We can refine this approximation with the gene expression data itself.

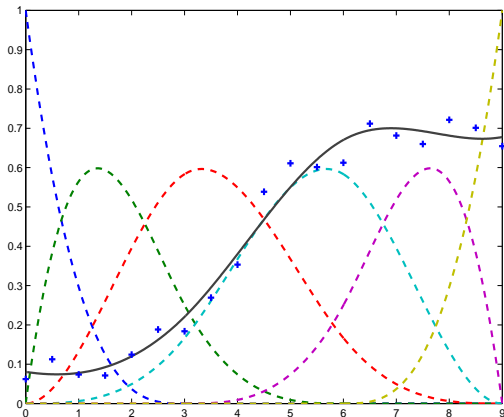
Spline approximations



Choose c_1, c_2, \dots, c_M to minimize the sum of squares error:

$$\sum_j (y_j - f(t_j))^2, \quad \text{where } f(t) = c_1 b_1(t) + \dots + c_M b_M(t).$$

Spline approximations

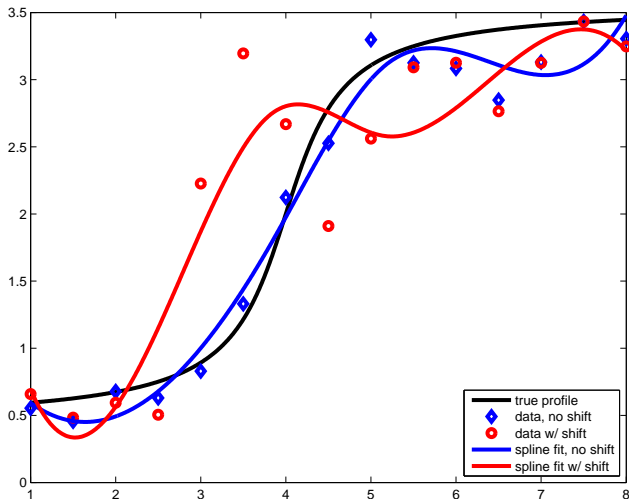


Choose c_1, c_2, \dots, c_M to minimize the sum of squares error:

$$\sum_j (y_j - f(t_j))^2, \quad \text{where } f(t) = c_1 b_1(t) + \dots + c_M b_M(t).$$

$$c = (.08, .09, .1, .92, .6, 3.7)$$

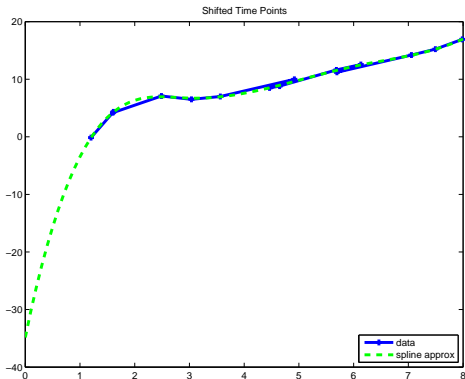
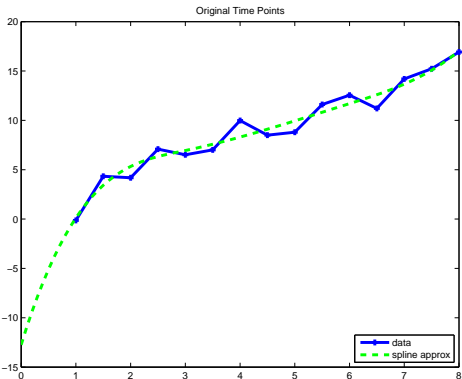
Temporal Inaccuracy Can Make a Big Difference



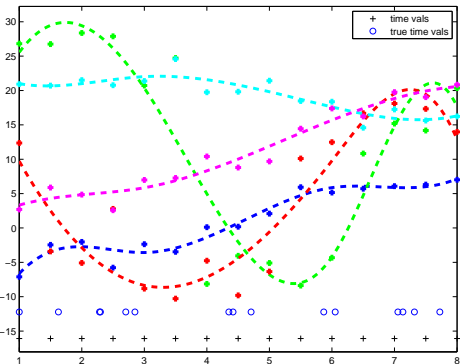
Handling Temporal Uncertainty

Choose c_1, \dots, c_M and s_1, \dots, s_T to minimize

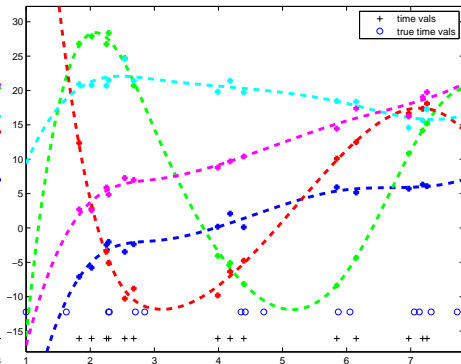
$$\sum (y_j - f(s_j))^2 + \lambda \sum (s_j - t_j)^2.$$



Fitting Multiple Genes



Spline fit-no shift



Spline fit-with shifts