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
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.
Choose \( c_1, c_2, \ldots, c_M \) to minimize the sum of squares error:

\[
\sum_j (y_j - f(t_j))^2, \quad \text{where } f(t) = c_1 b_1(t) + \cdots c_M b_M(t).
\]
Choose $c_1, c_2, \ldots, c_M$ to minimize the sum of squares error:

$$\sum_{j}(y_j - f(t_j))^2,$$

where $f(t) = c_1 b_1(t) + \cdots c_M b_M(t)$.

$$c = (0.08, 0.09, 0.1, 0.92, 0.6, 3.7)$$
Temporal Inaccuracy Can Make a Big Difference

![Graph showing the impact of temporal inaccuracy on microarray analysis. The graph compares true profiles with data and spline fits with and without shifts.](Image)

**Legend:**
- **Black Line:** True Profile
- **Blue Diamond:** Data, No Shift
- **Red Circle:** Data with Shift
- **Blue Line:** Spline Fit, No Shift
- **Red Line:** Spline Fit with Shift

Stephen C. Billups
Microarray Analysis With Temporal Uncertainty
Choose $c_1, \ldots, c_M$ and $s_1, \ldots, s_T$ 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