Accurate Change Point Detection for Electricity Market Analysis

Will Gu
Lawrence Berkeley National Lab & UC Berkeley

Introduction

- The California Energy Crisis (May 2000-December 2001) cost the state about $40 billion
- Energy companies took advantage of deregulation laws, to reduce California's electricity supply
- Known for being volatile, electricity prices skyrocketed and rose by 800%, as seen in Figure 1

Goal

- To detect points of unnatural price changes within our time frame
  - Has to significantly improve the Change Point Detection algorithm to work with all data records

Gaussian Process Models

- GP assigns its input points to a random (Gaussian) distribution and uses the distribution to predict new points
  - An example is shown in Figure 2.
- GP is specified by a covariance function (K) and a mean function (m):

$$x \sim \mathrm{GP}(\mathrm{m}, K) \quad \text{where } K(x,x') = \sigma_f \left[1 + \sqrt{3}\frac{\| x - x' \|}{3}\right] exp\left(-\sqrt{3}\frac{\| x - x' \|}{3}\right) \quad \text{and} \quad \mathrm{m}(x) = 0.$$

- GP is computationally expensive, requiring $O(n^3)$ time, but much more accurate than parametric models
  - Figure 3 compares the predictive precision of GP with that of the less expensive, parametric ARIMA model; the GP is clearly more accurate.

Bayesian Online Change Point Detection

- BOCPD is a Bayesian change detection algorithm that uses the Gaussian Process recursively
- BOCPD attempts to fit a Gaussian Process to every possible time segment to find the best series of Gaussian Processes to fit the data
- Each such Gaussian Process is called a run (see Figure 4), the beginning of a run is a change point
- BOCPD detects obvious changes such as the reduction of price cap on 2000/07/01 and 2000/08/02, as well as subtler events such as bankruptcy of SDG&E and Enron's heavy use of "ricochet"

Semi-Separable Matrices

- BOCPD is computationally expensive: it invokes GP $O(n^2)$ times, for a total of $O(n^5)$ time
- Challenge: reduce the time needed for BOCPD
- The Matern covariance function for GP on time series data produces semi-separable matrices
  - There exist a diagonal matrix D and matrices P,Q $\in R^{n(p+1)}$ such that:

(covariance matrix)

$$\begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \dots & K(x_1,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n,x_1) & K(x_n,x_2) & \dots & K(x_n,x_n) \end{bmatrix} + \sigma_n^2 I = D + triu(P\,Q^T) + \left(triu(P\,Q^T)\right)^T$$

(triu denotes the upper triangular part of a given matrix)

- Solutions to linear systems with semi-separable matrices can be solved quickly
- Furthermore, the linear systems from the GP are related to each other, which leads us to solve n GP problems in O(n) time
- Overall, BOCPD solves $O(n^2)$ GP problems in $O(n^2)$ time
- Timing measurements in Figure 5 show that the new algorithm is much faster

Conclusion

- By taking advantage of the semi-separable structure of the matrices in GP on 1-dimensional time series, we reduce the computational time of BOCPD from $O(n^5)$ to $O(n^2)$
- The new change point detection algorithm, BOCPD-GPSS, can be applied on much large dataset than previously possible
- BOCPD-GPSS is demonstrated to detect a variety of changes as illustrated in Figure 4

Figure 1: Real-time market (ISO) prices of electricity for northern California

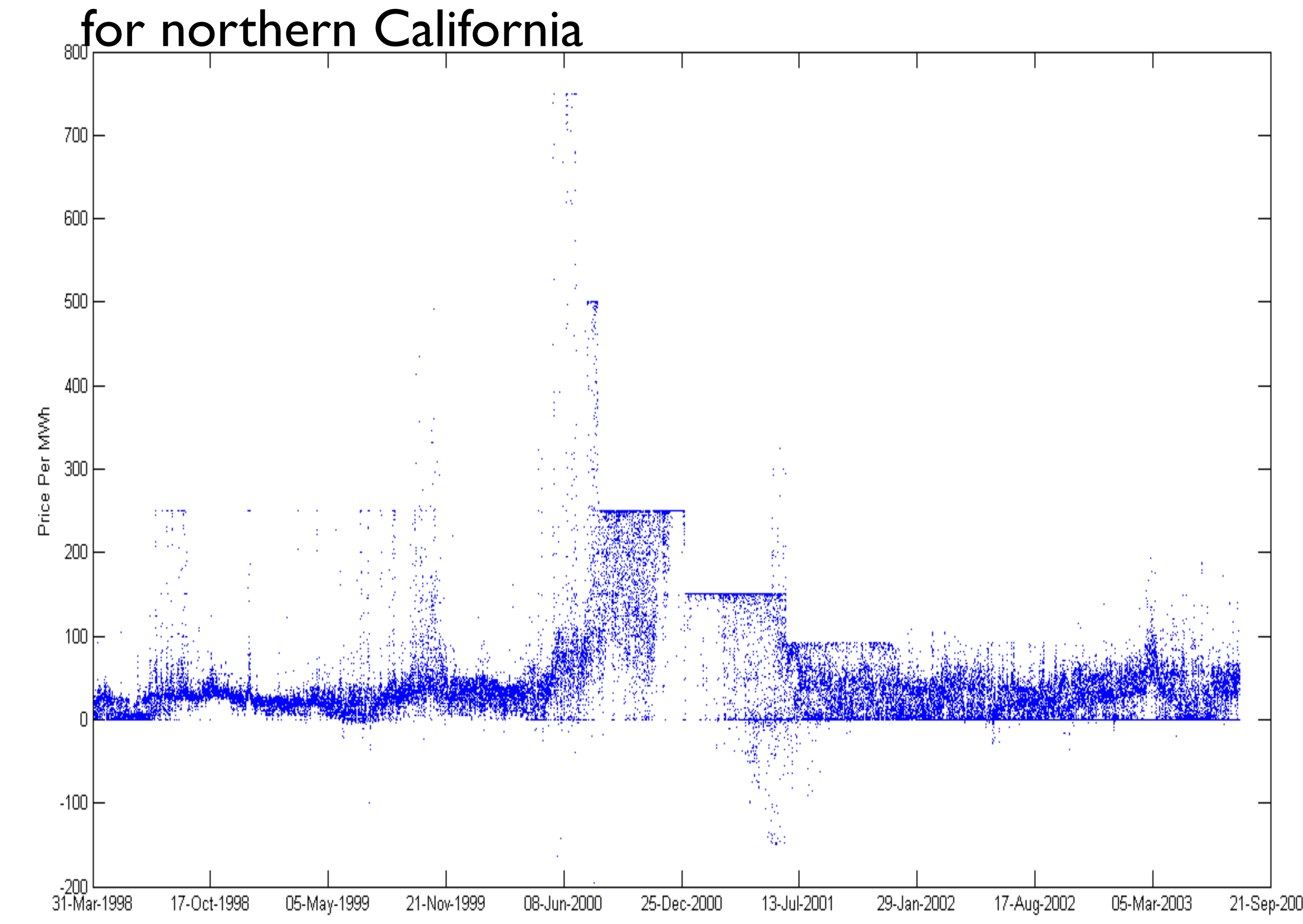
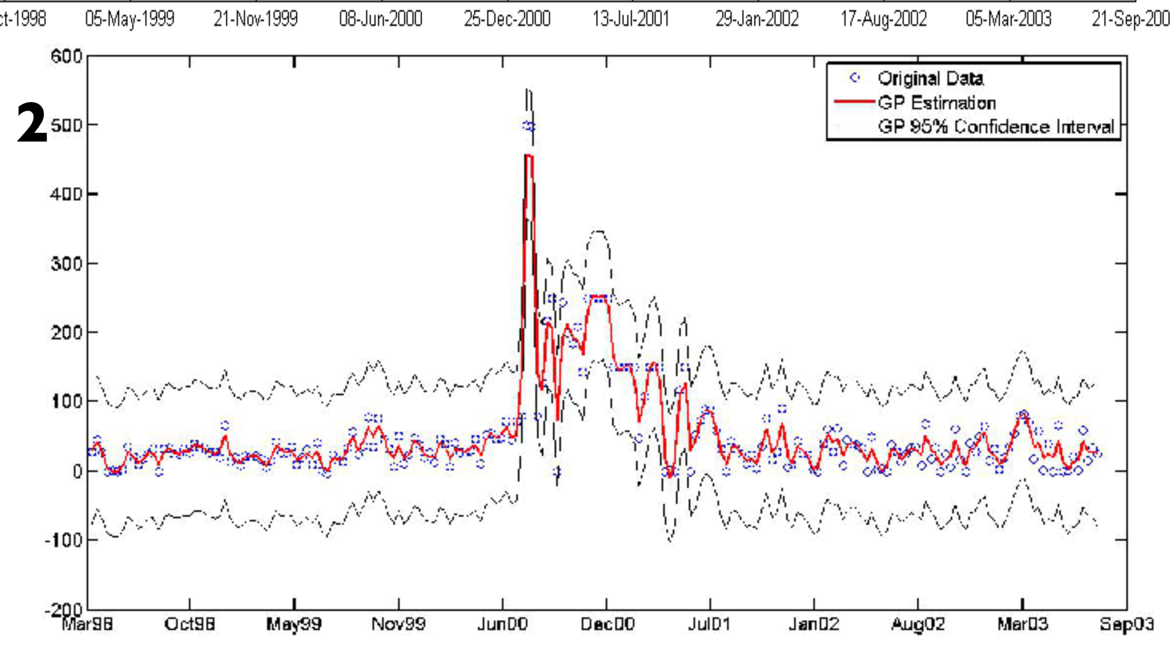


Figure 2



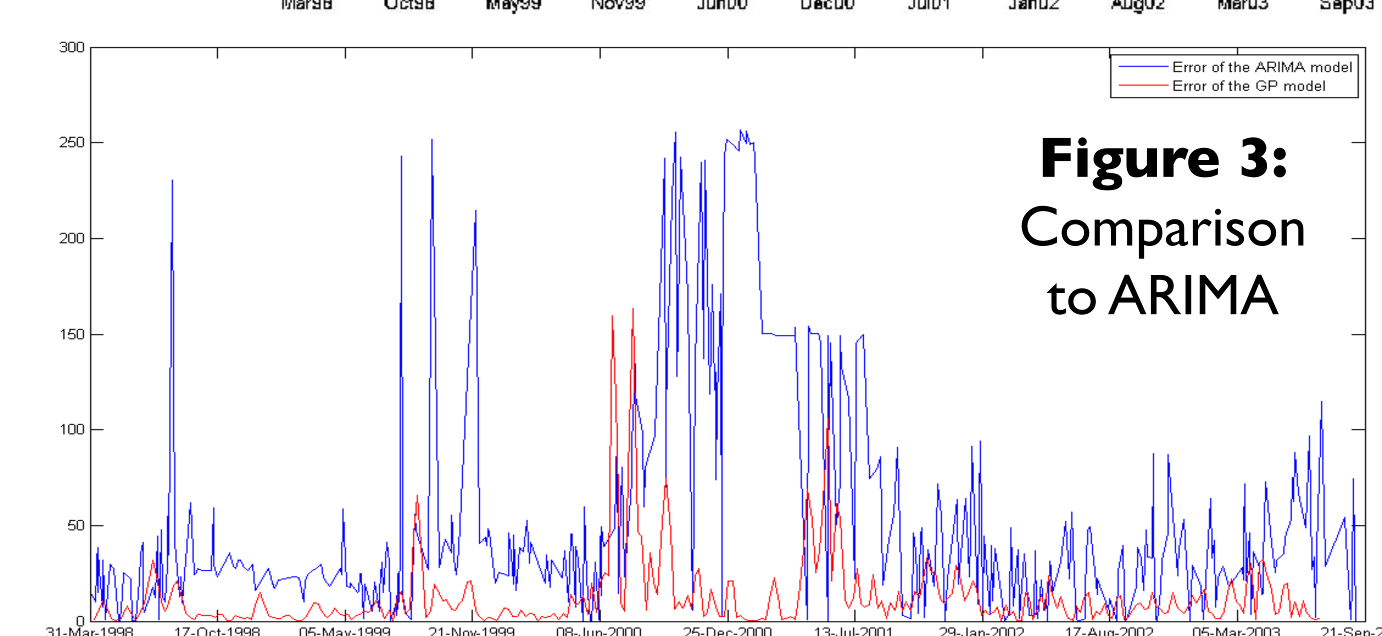


Figure 3: Comparison to ARIMA

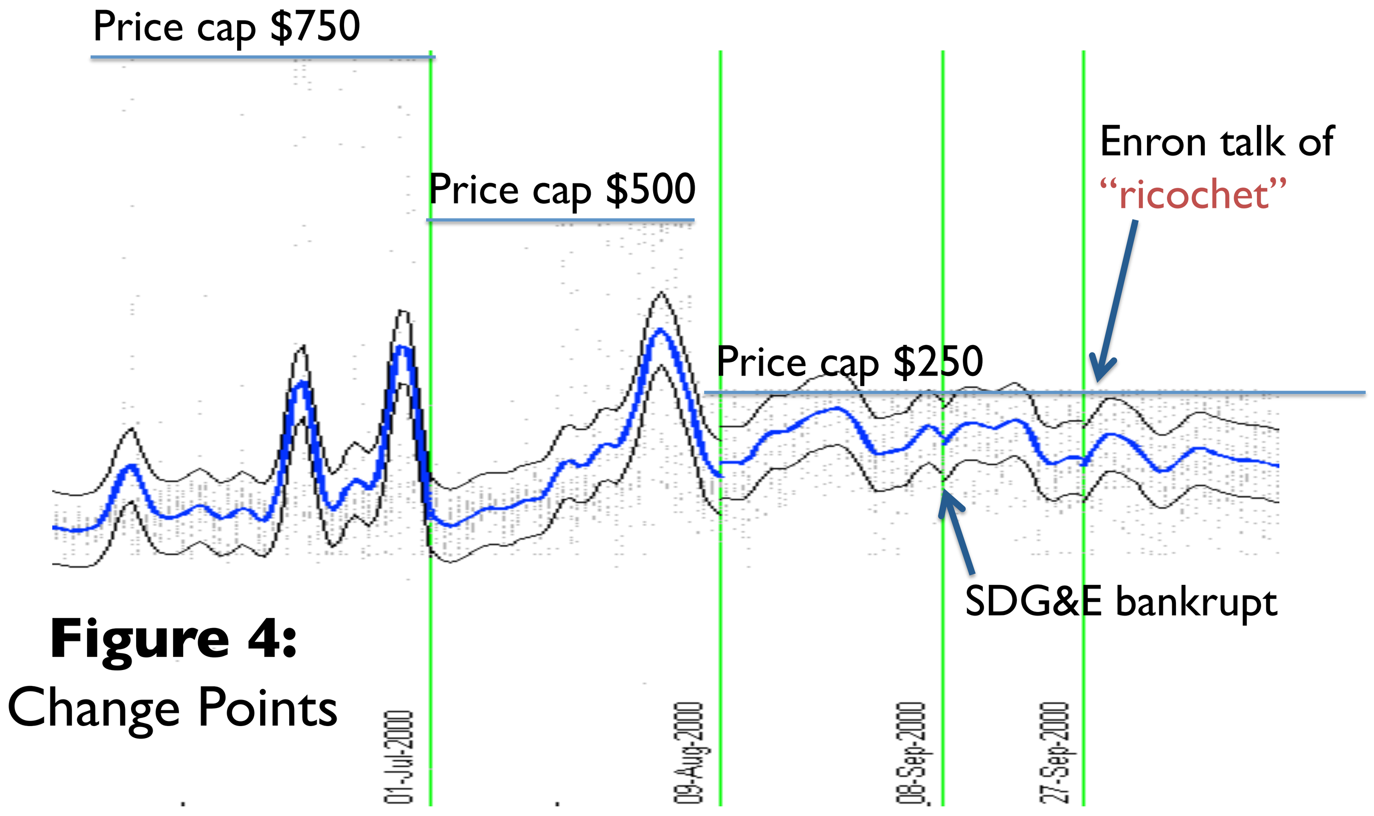
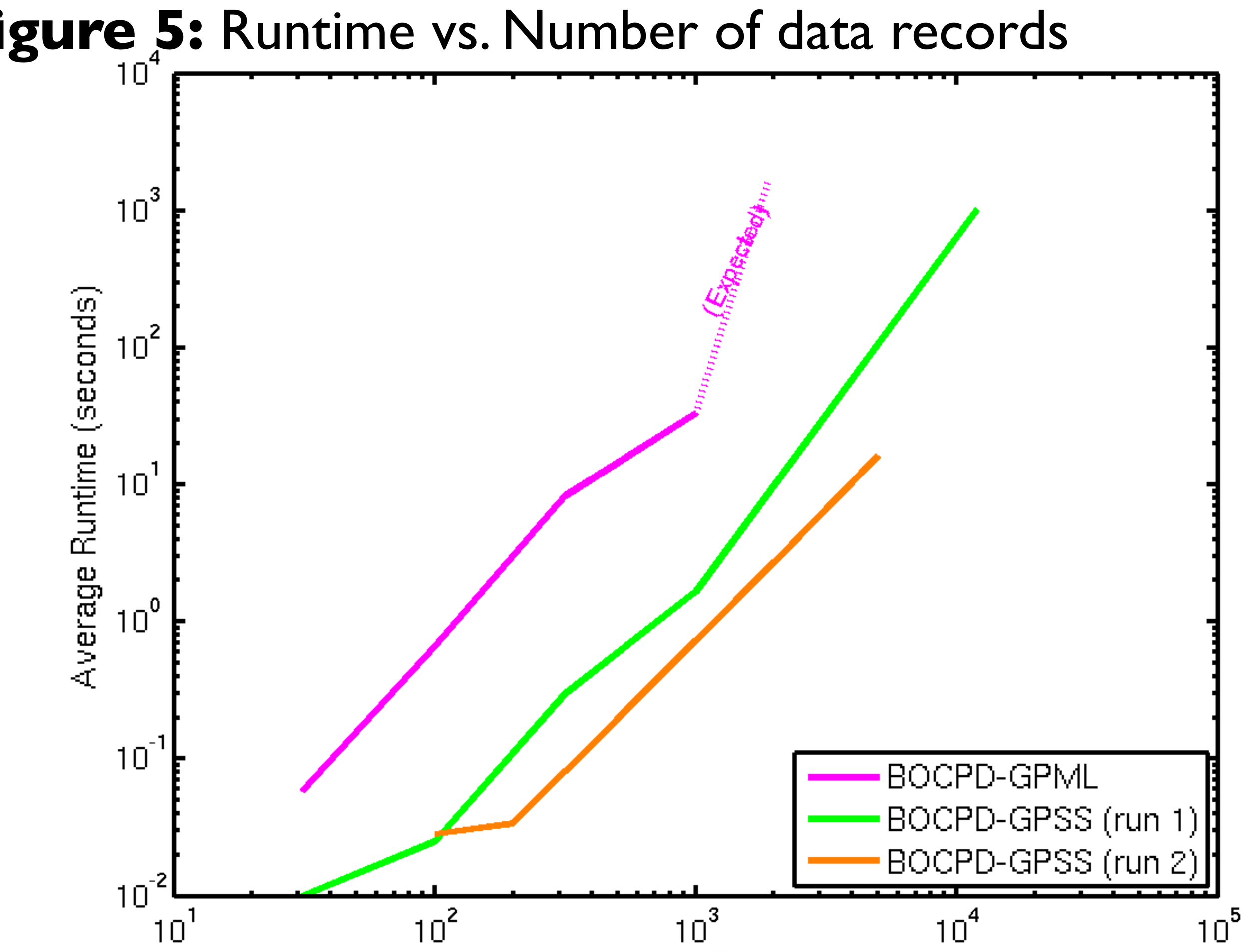


Figure 4: Change Points

Figure 5: Runtime vs. Number of data records



This work is support in part by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231