



Children's Hospital
Informatics Program



Harvard
Medical School

Quantitative approaches to differentiating signal from noise: Enhancing surveillance for prescription opioid abuse

John Brownstein, PHD
Children's Hospital Boston
Harvard Medical School
Harvard-MIT Health Sciences and Technology

Evolving Systems and Datasets - Challenges

1. Low data levels means that signal to noise levels are typically low.
2. Non-stationary and non-stable data means modeling is difficult.
3. Relatively new systems means that historical data is limited

AEGIS

SELECT DATE



SELECT HOSPITAL/SYNDROME

	BIDMC	CAMB	CHB	MGH	SOMER	WH
RESP	■■■					
GI						
RASH						
NEURO						
HEMOR						
ALL	■	■	■	■	■	■
INI	■	■	■	■	■	■

SELECT ABBERRATION

Aberrations For Current Selection

- BIDMC RESP $\leq 4/2$ [details](#) [export](#)

Other Aberrations

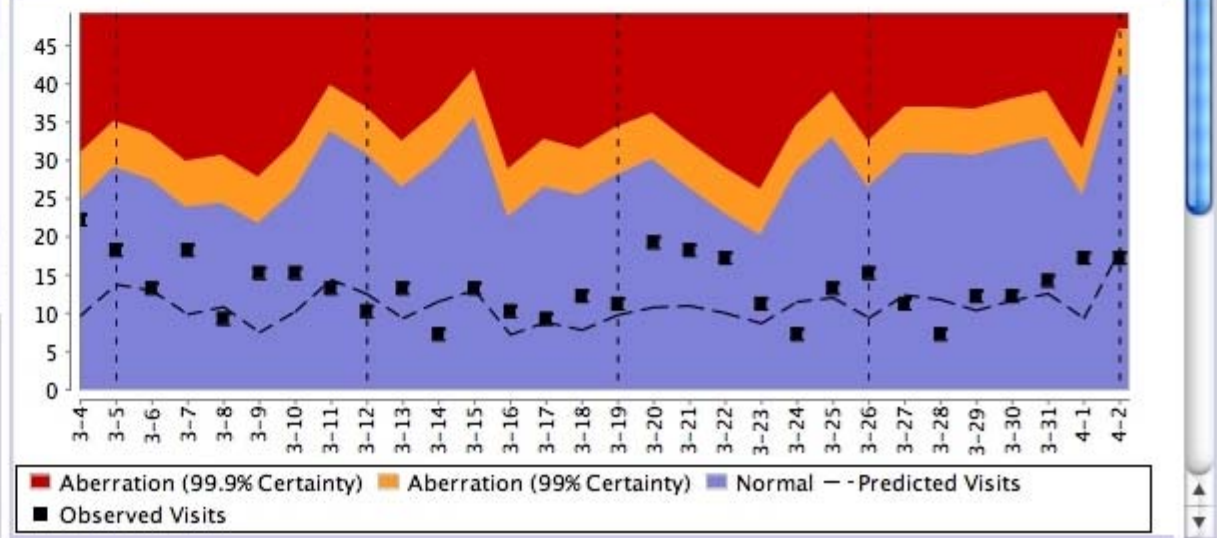
- MGH RESP $\leq 4/5$ [details](#) [export](#)
- BIDMC RESP $\leq 4/2$ [details](#) [export](#)
- CHB GI $\leq 3/11$ [details](#) [export](#)
- CAMB RESP $\leq 3/23$ [details](#) [export](#)

LEGEND

■ Active	■ Inactive	- Normal
■	■	- Aberration (99% Certainty)
■	■	- Aberration (99.9% Certainty)
■	■	- Result not yet available
		S/T - Spatial/Temporal Aberration



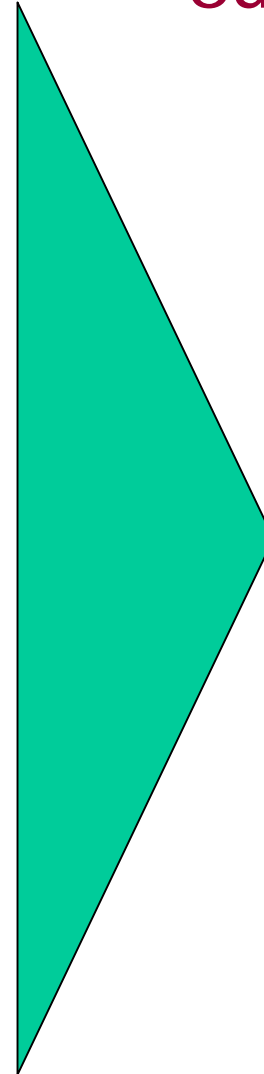
TEMPORAL



Goals of Surveillance

- Timeliness of detection
- Reliability of detection
- Automation and integration
- Robustness under adverse conditions
- Security of patient identifiable data
- Compatibility/adherence to standards

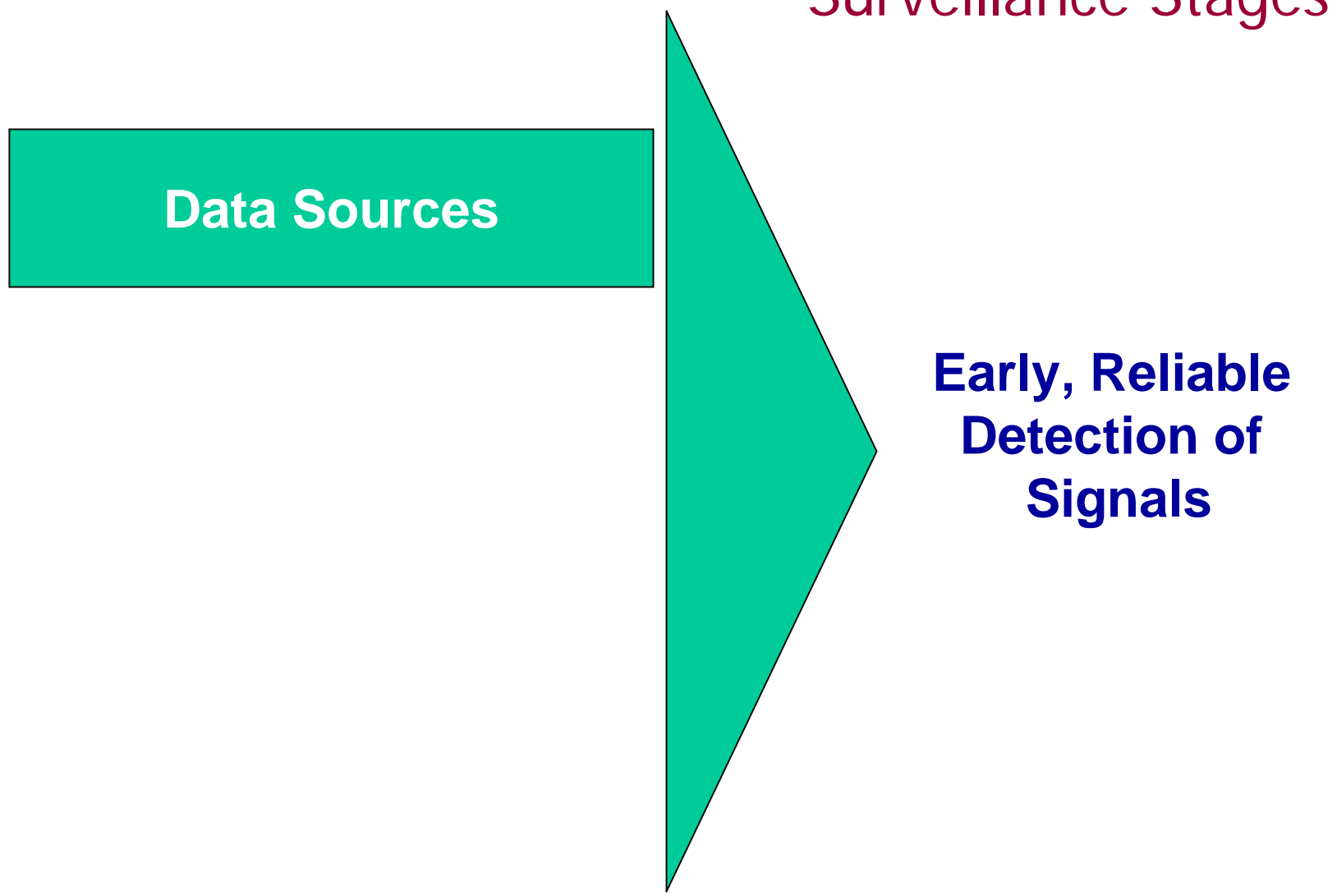
Surveillance Stages



**Early, Reliable
Detection of
Signals**

Surveillance Stages

Data Sources



**Early, Reliable
Detection of
Signals**

Surveillance Stages

Data Sources

Model

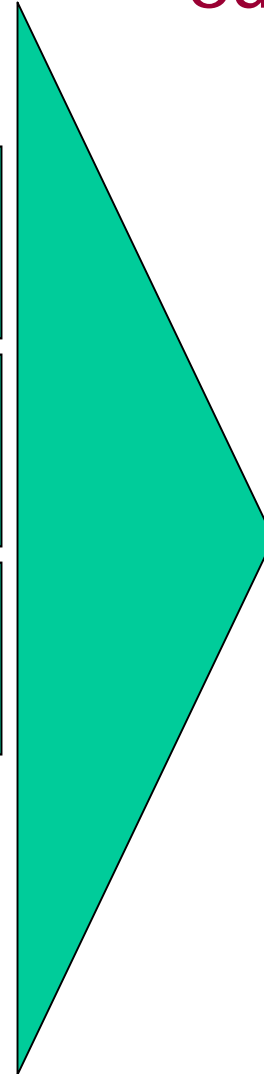
**Early, Reliable
Detection of
Signals**

Surveillance Stages

Data Sources

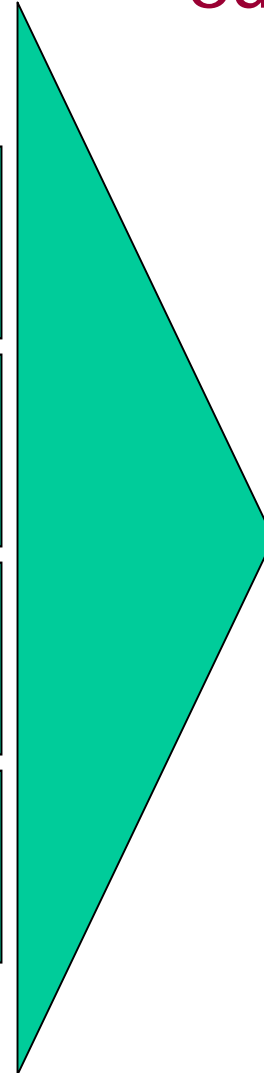
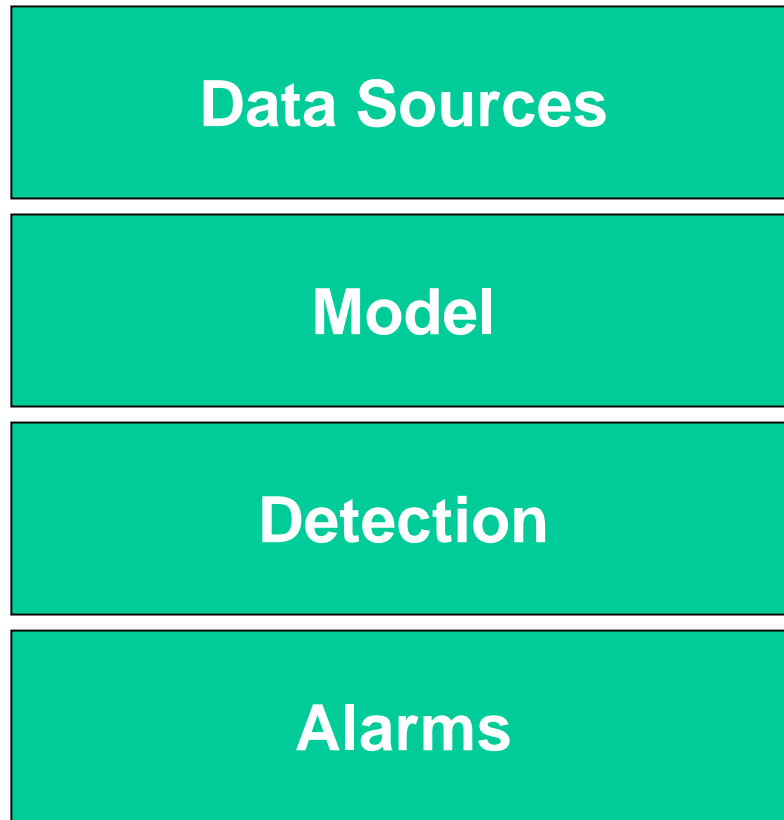
Model

Detection



**Early, Reliable
Detection of
Signals**

Surveillance Stages



**Early, Reliable
Detection of
Signals**

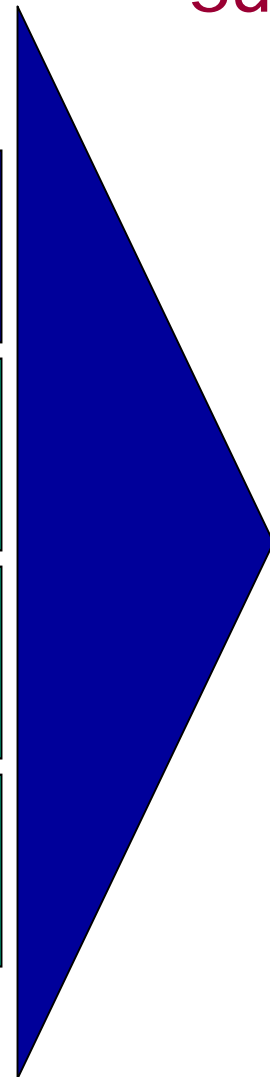
Surveillance Stages

Data Sources

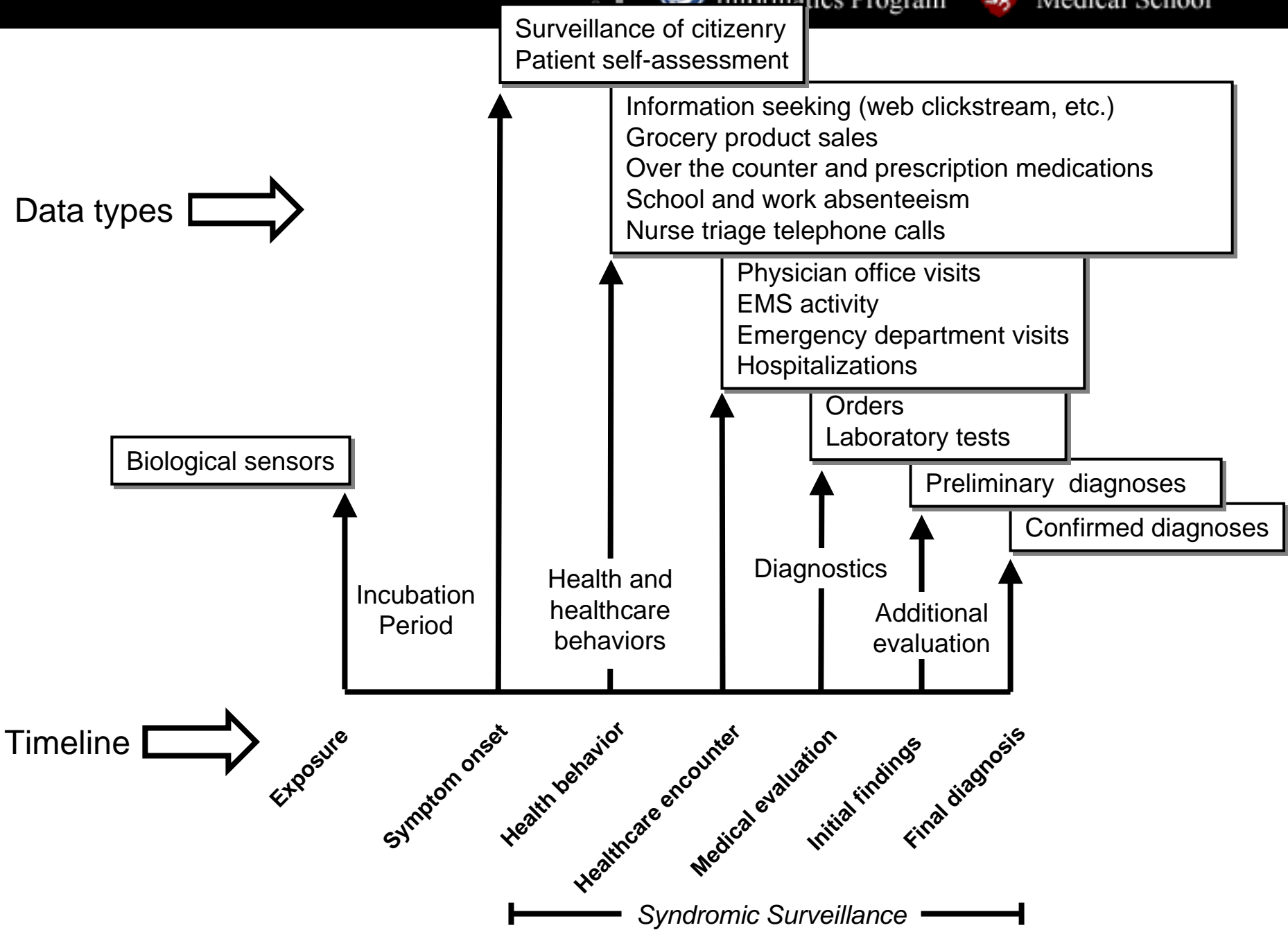
Model

Detection

Alarms



**Early, Reliable
Detection of
Signals**



Health care data sources

- Patient demographic information
- Emergency department chief complaints
- International Classification of Disease (ICD) codes
- Text-based notes
- Laboratory data
- Radiological reports
- Physician reports

“Non traditional” data sources

- Pharmacy data
- Retail sales data
- 911 operators
- Call triage centers
- Poison control centers
- Key informants
- Drug addiction facilities
- Patient self-reporting

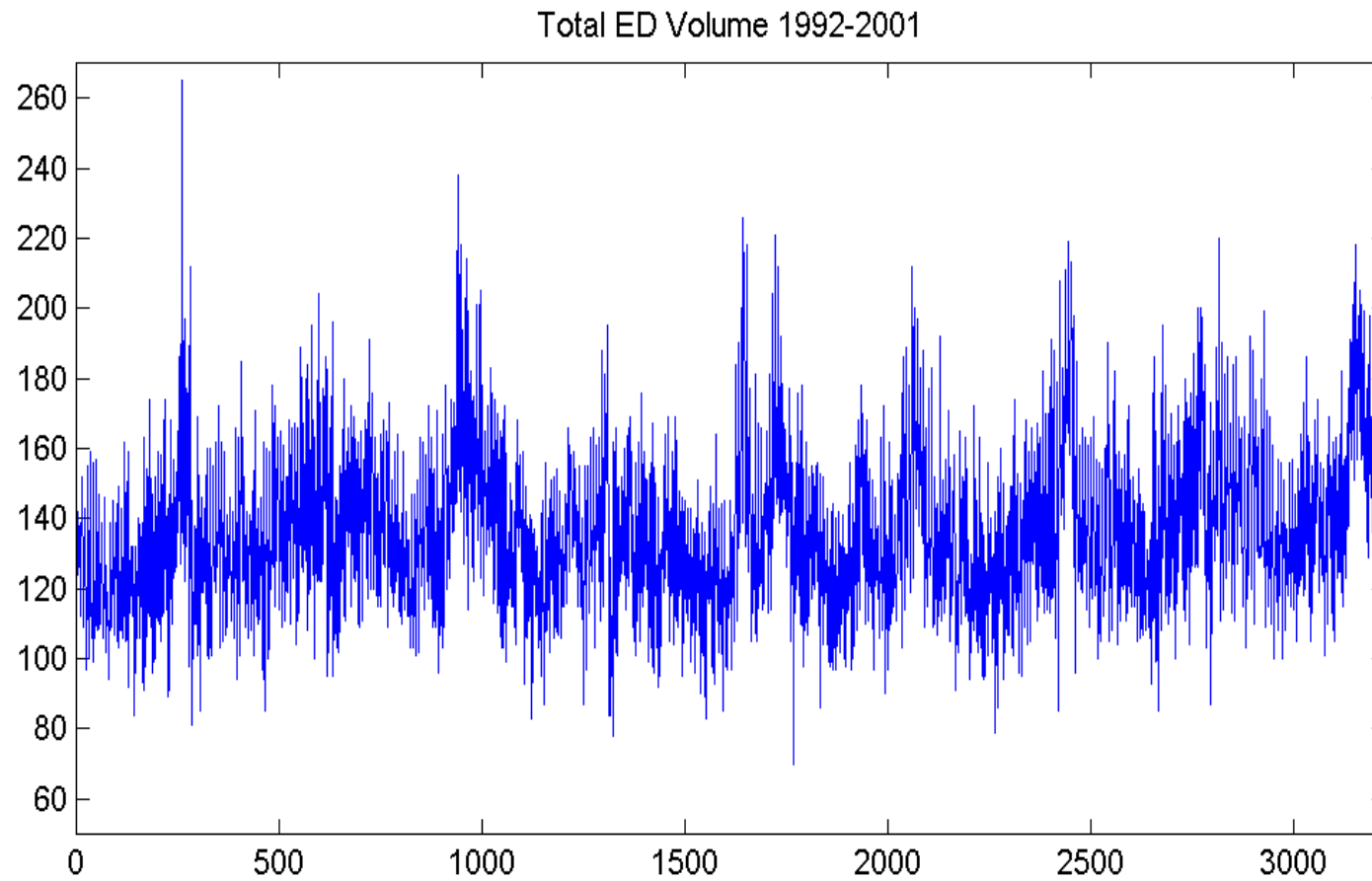
Confounders and Covariates

- Season
- Day of Week
- Meteorological
- Prescription rates
- Drug availability
- Economic Conditions
- Media reports

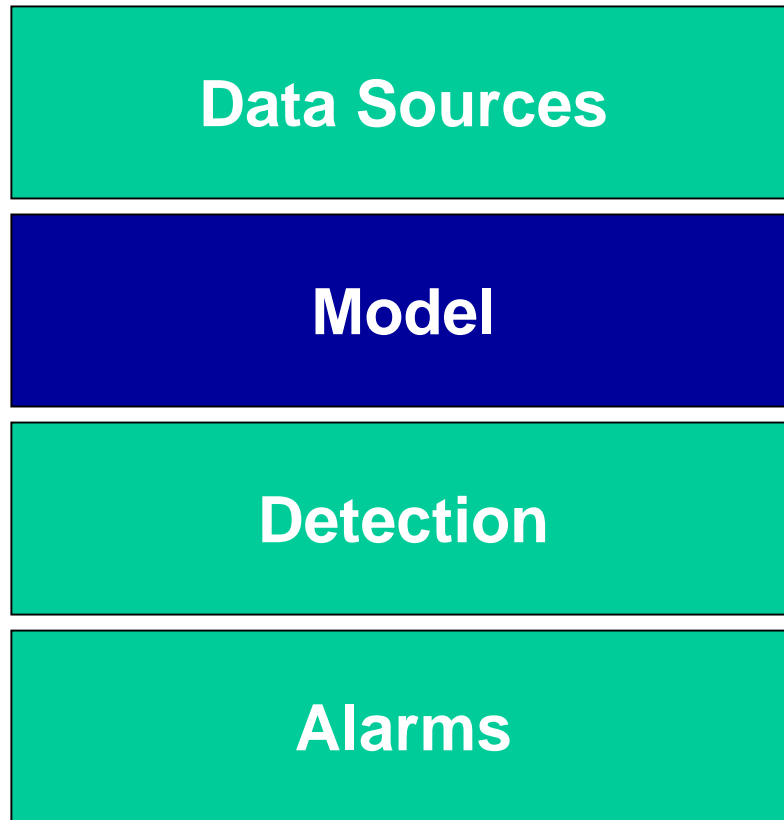
Validation of data sources

- Quality of data is crucial
- Tradeoff of quality for timeliness
- Informant data vs. diagnostic codes
- Comparison with gold standard

Data Source

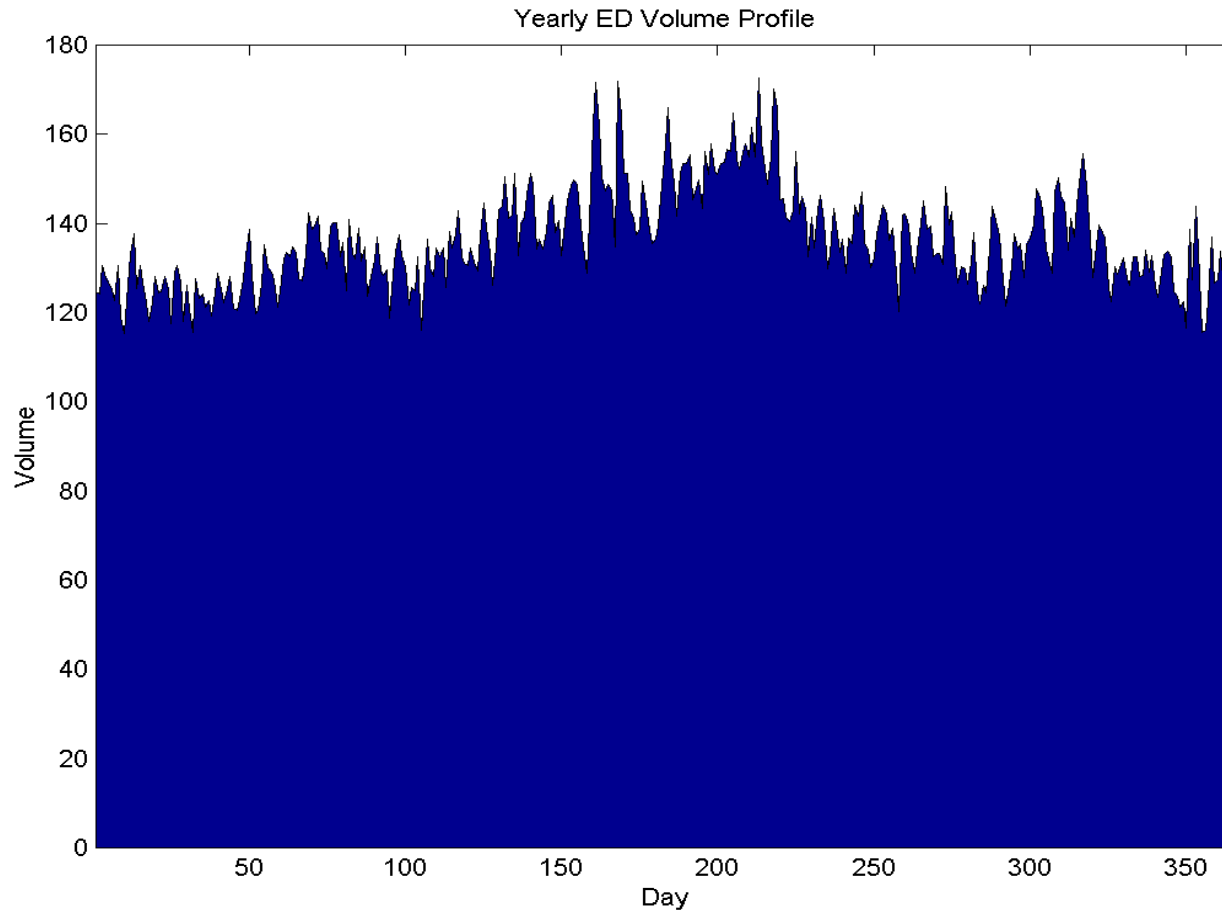


Surveillance Stages

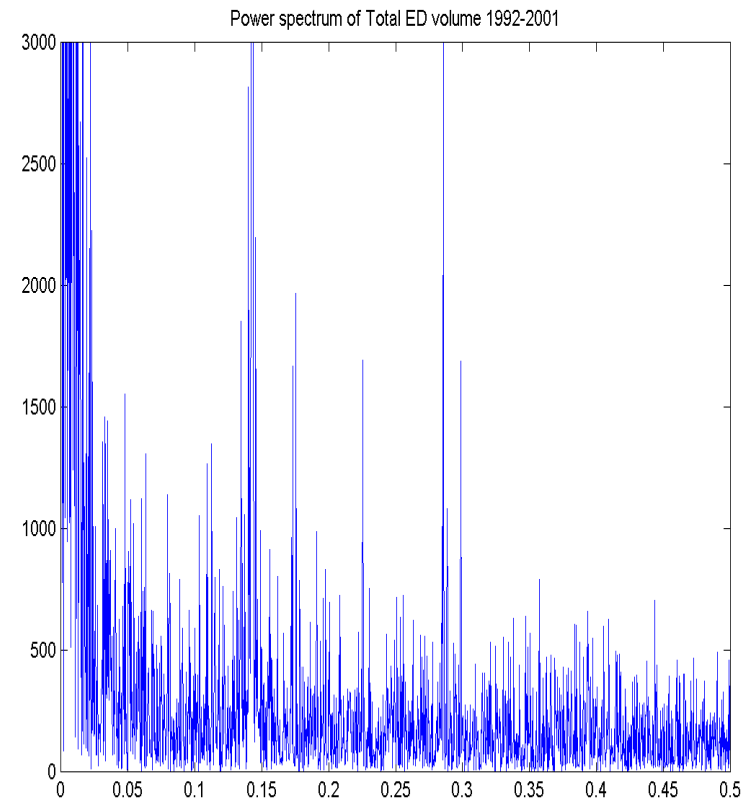
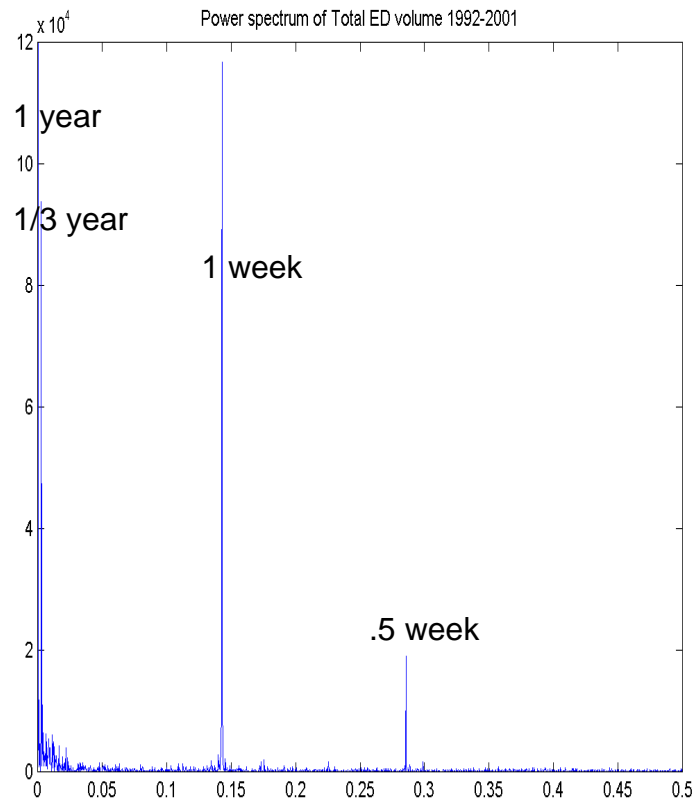


**Early, Reliable
Detection of
Signal**

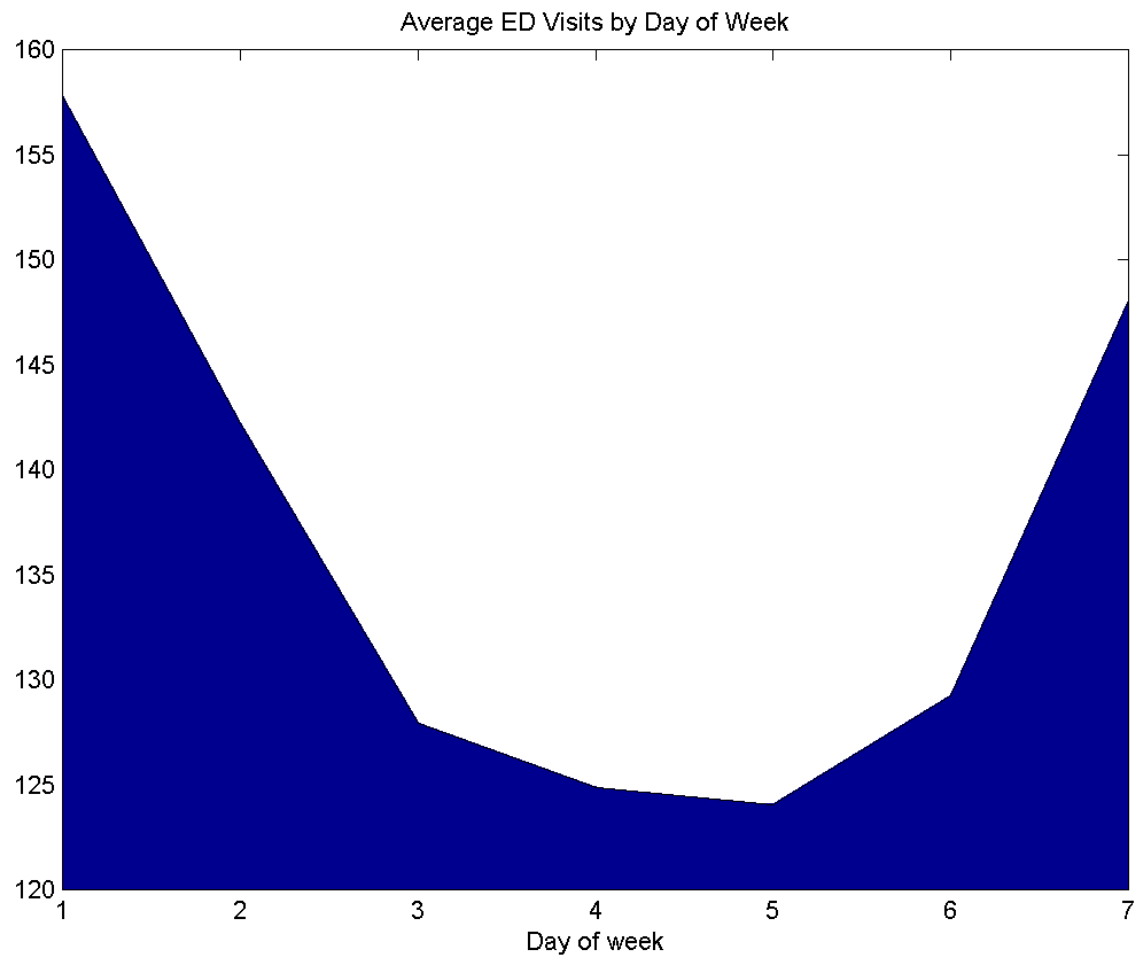
Yearly Trends



Principal Fourier Component Analysis

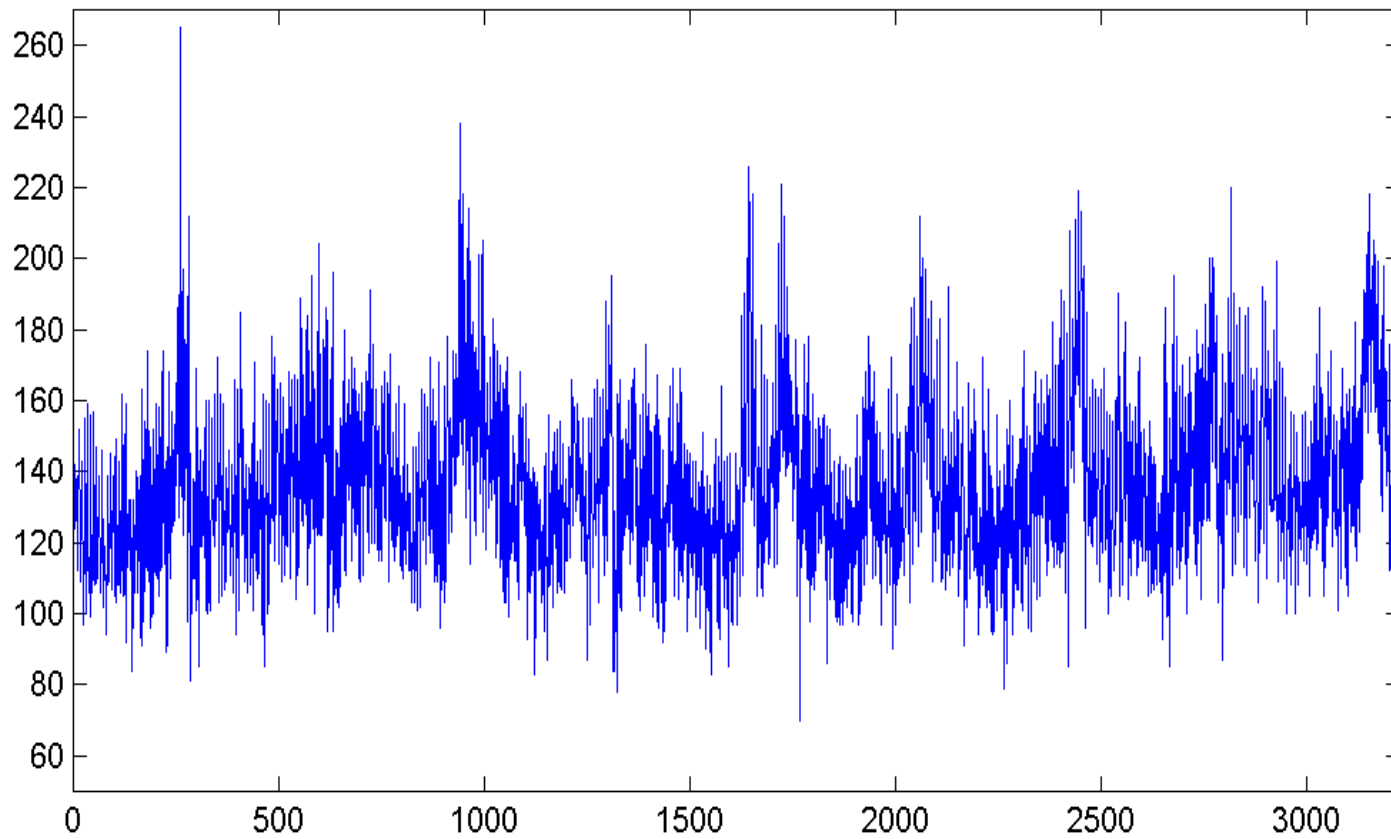


Weekly Trends

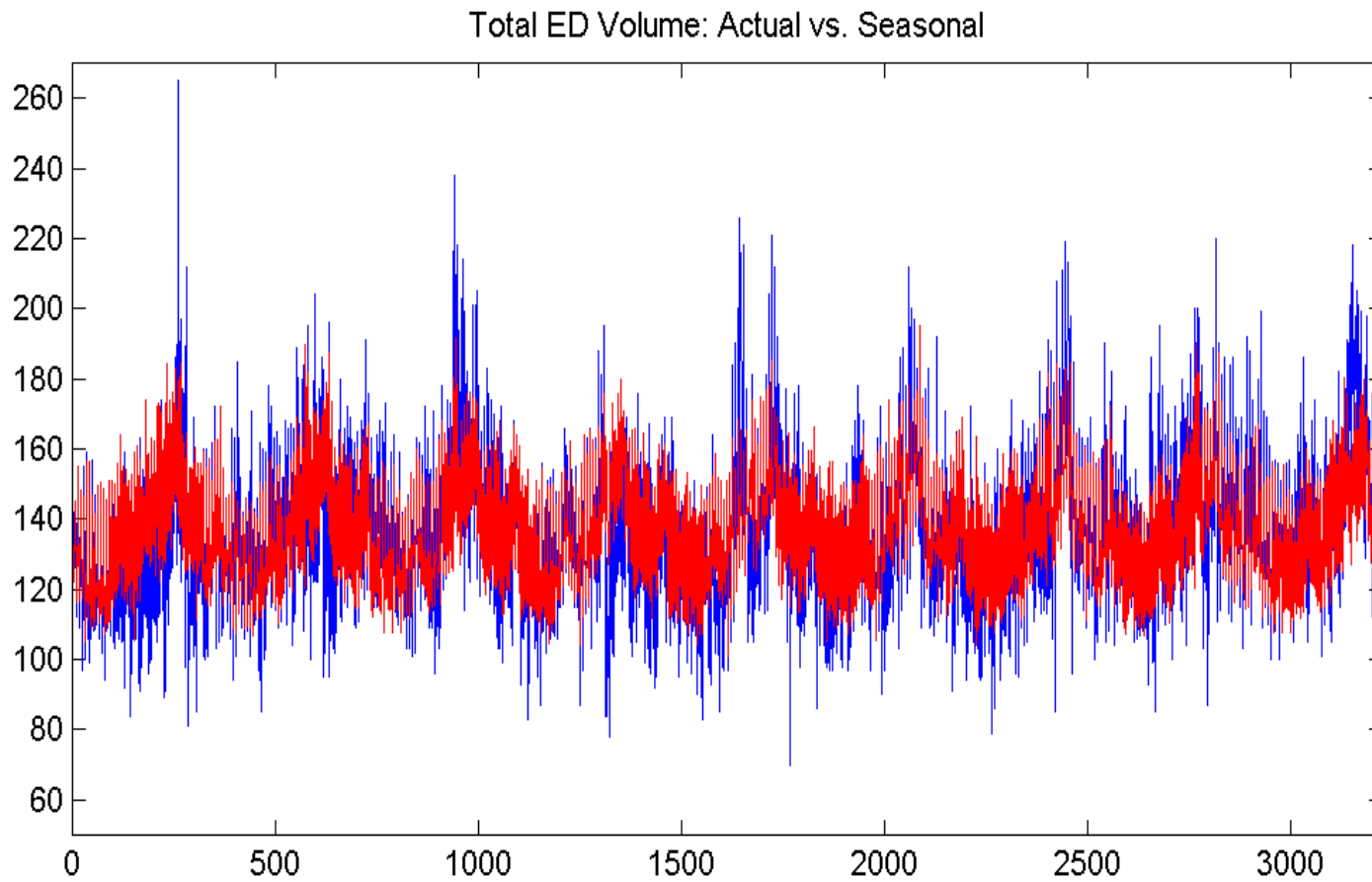


Forecasting

Total ED Volume 1992-2001

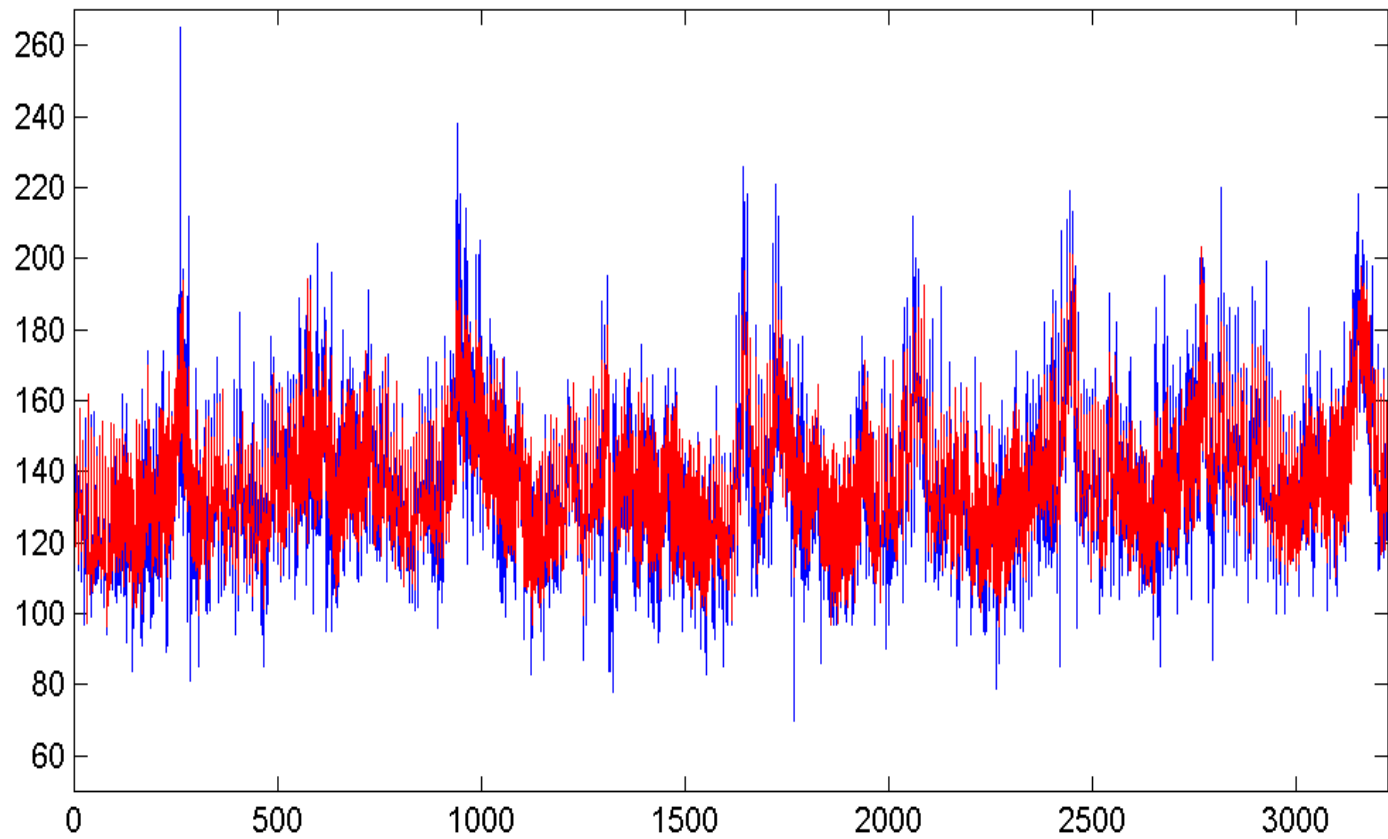


Forecast Model

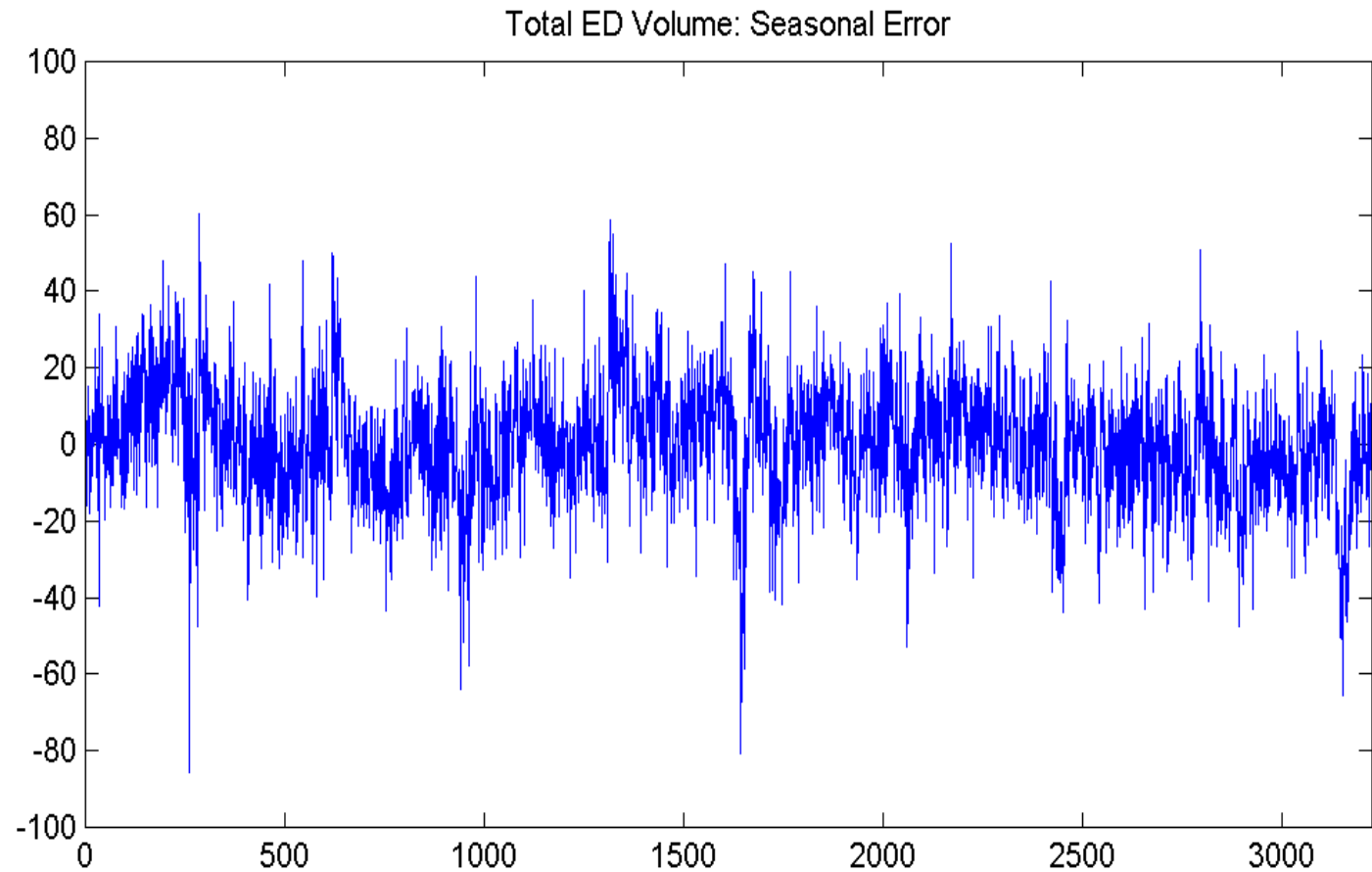


ARIMA Model

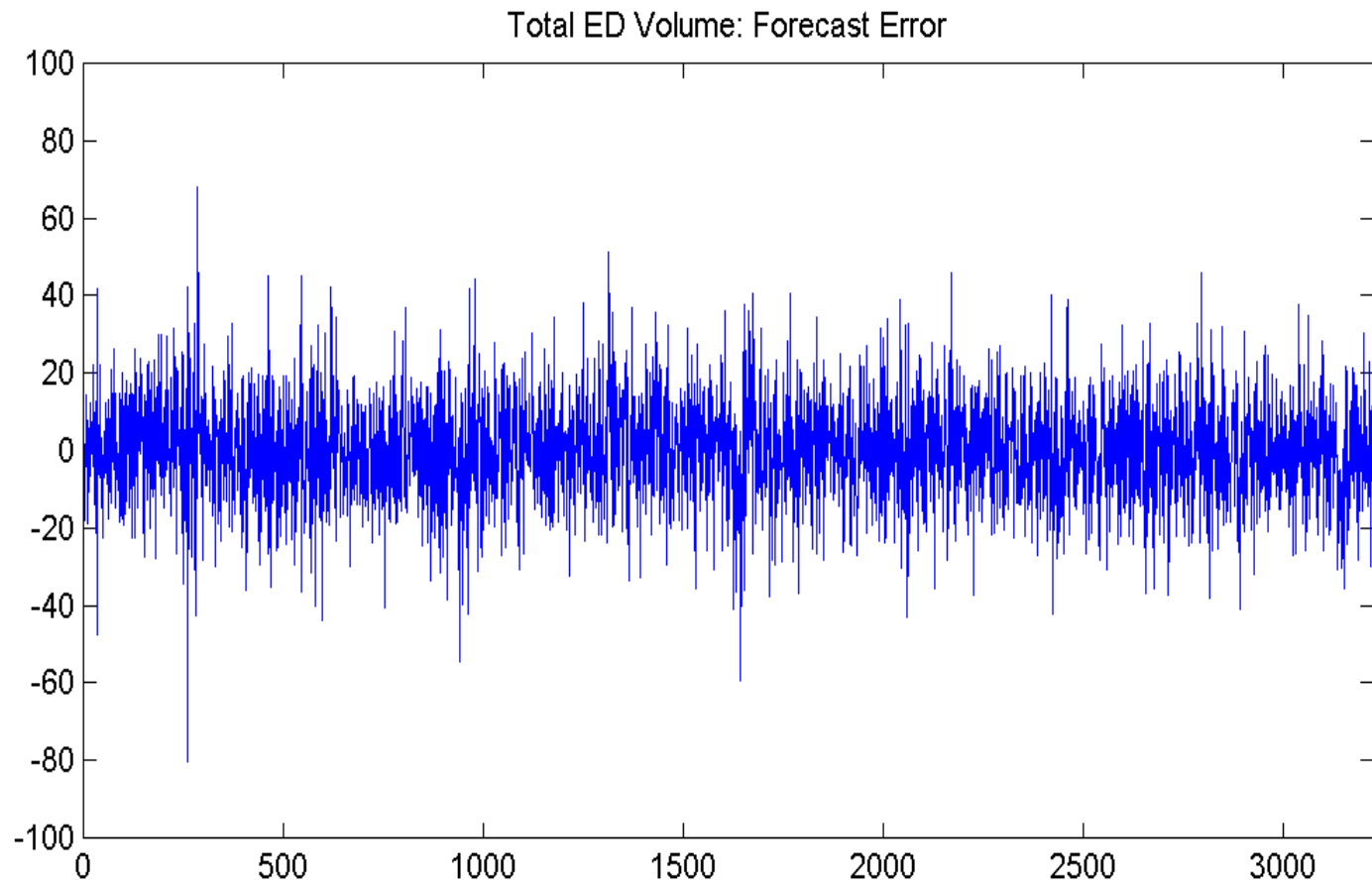
Total ED Volume: Actual vs. Forecast



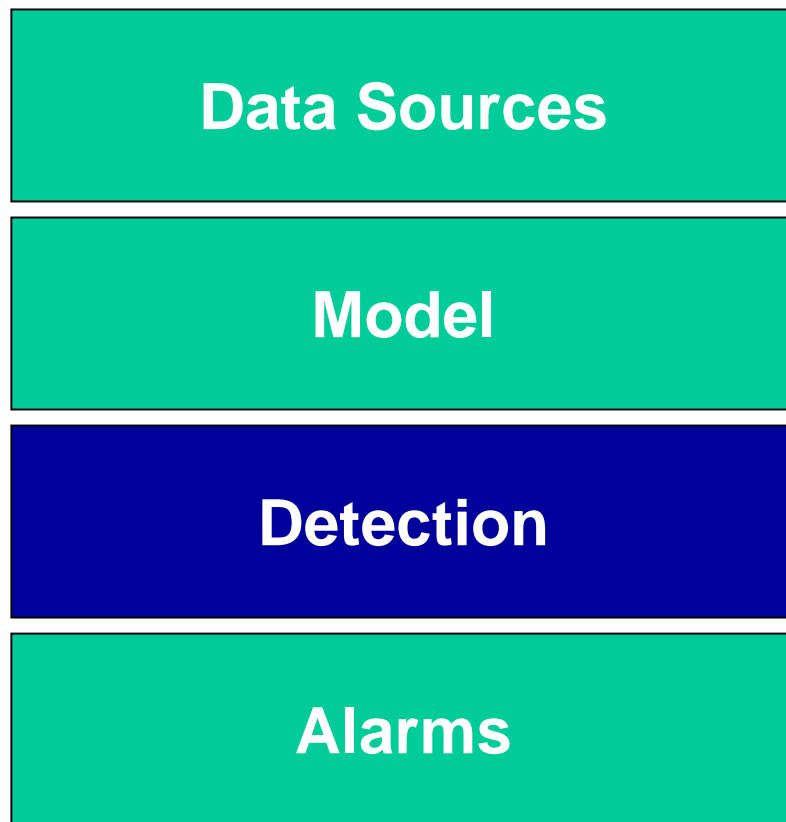
Forecasting (Seasonal Model)



Forecasting (with ARIMA)

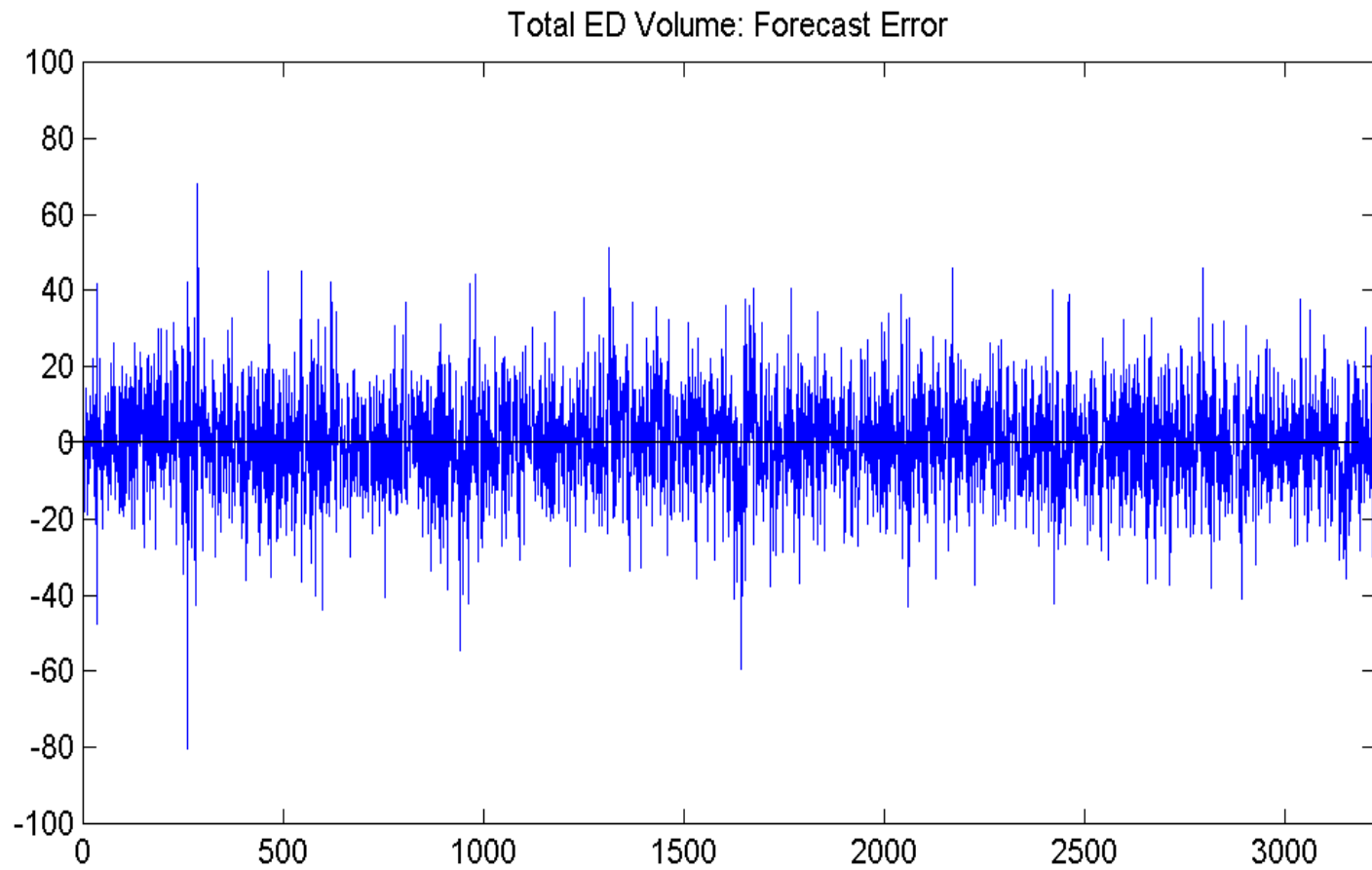


Surveillance Stages

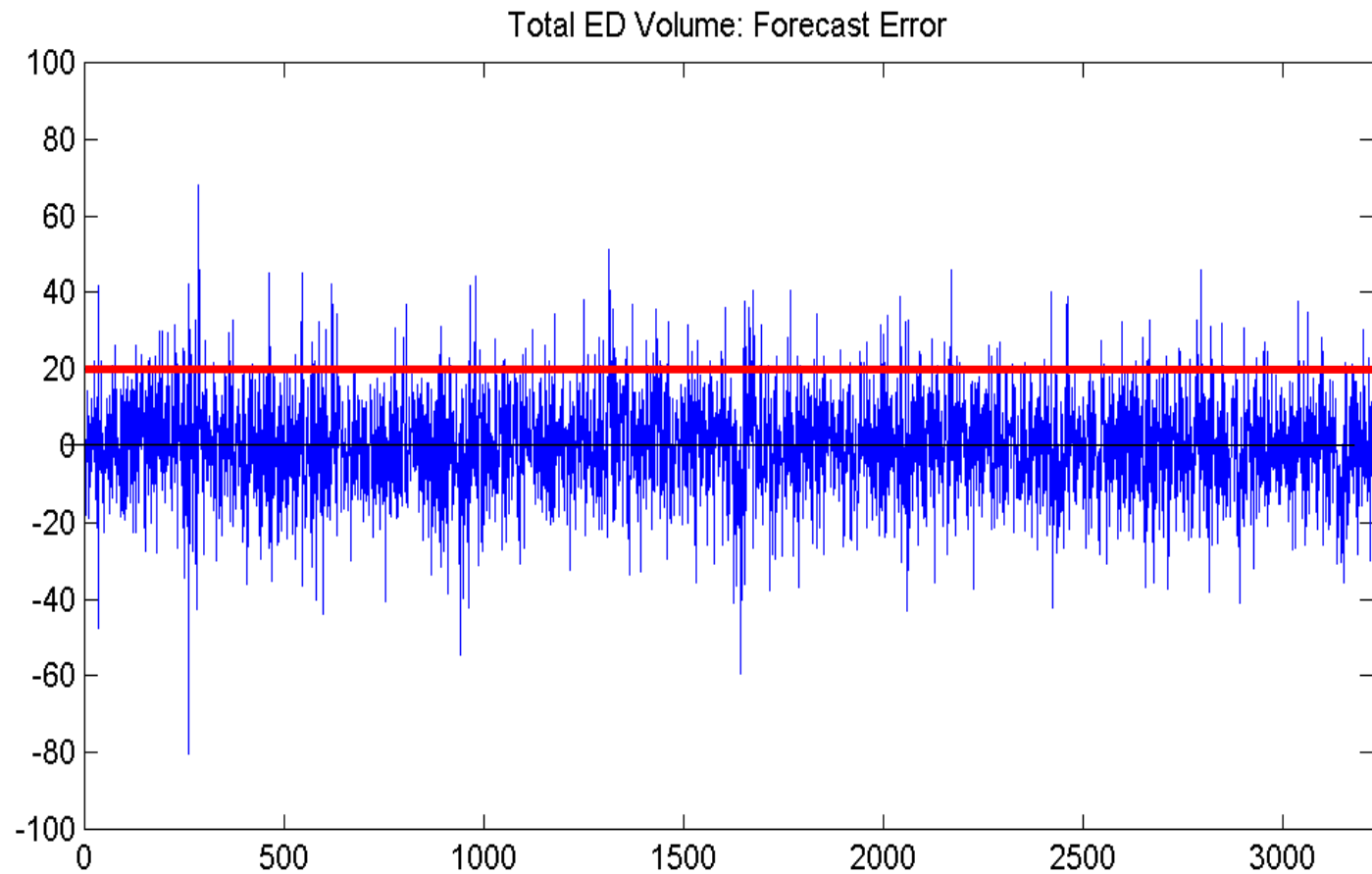


**Early, Reliable
Detection of
Signals**

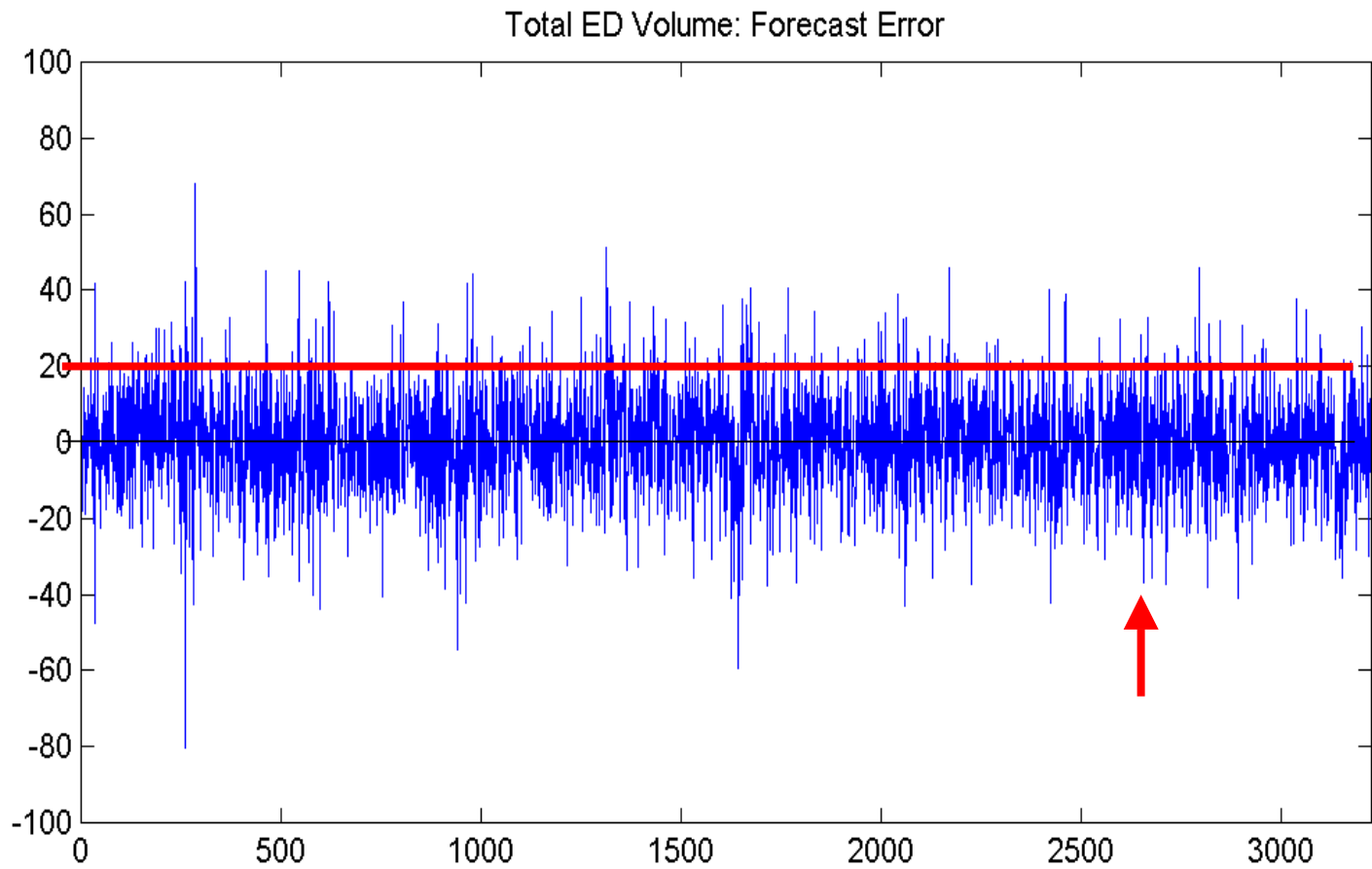
Intrinsic Noise



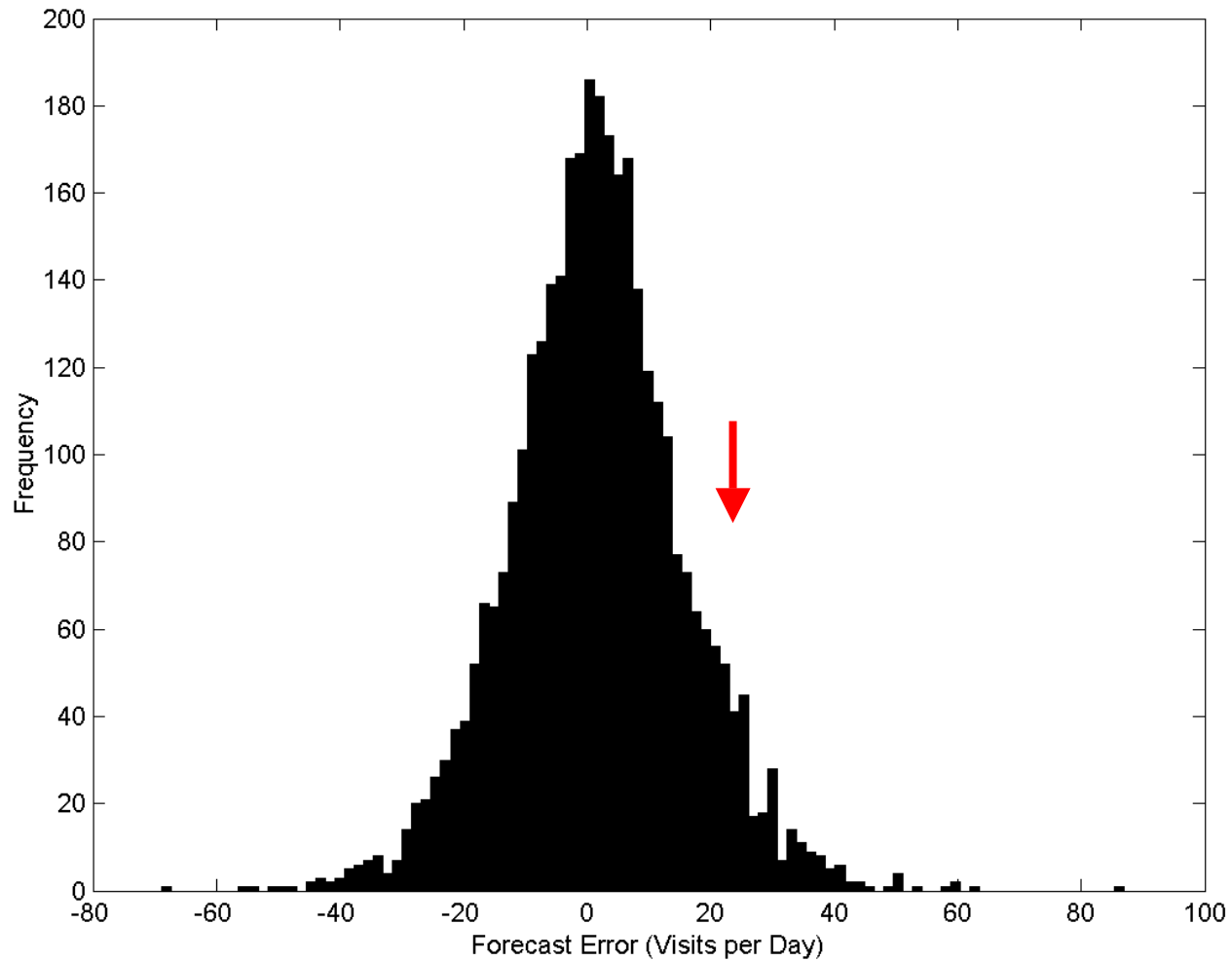
False Positives

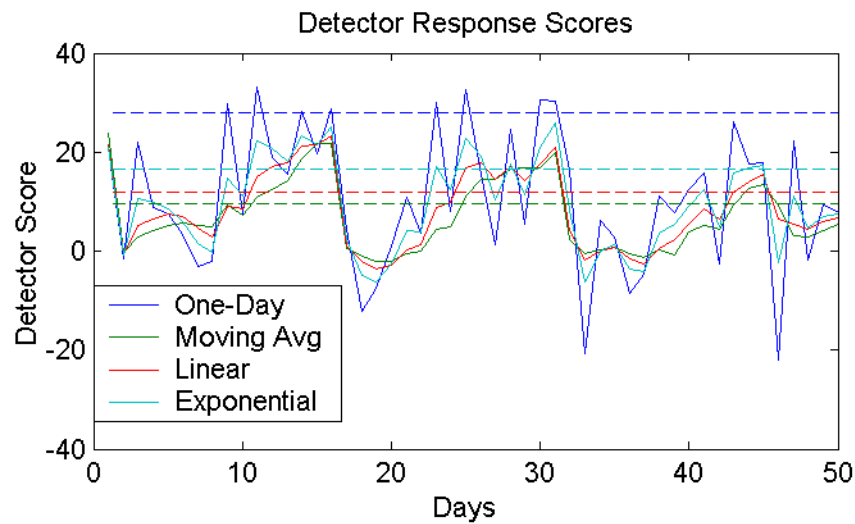
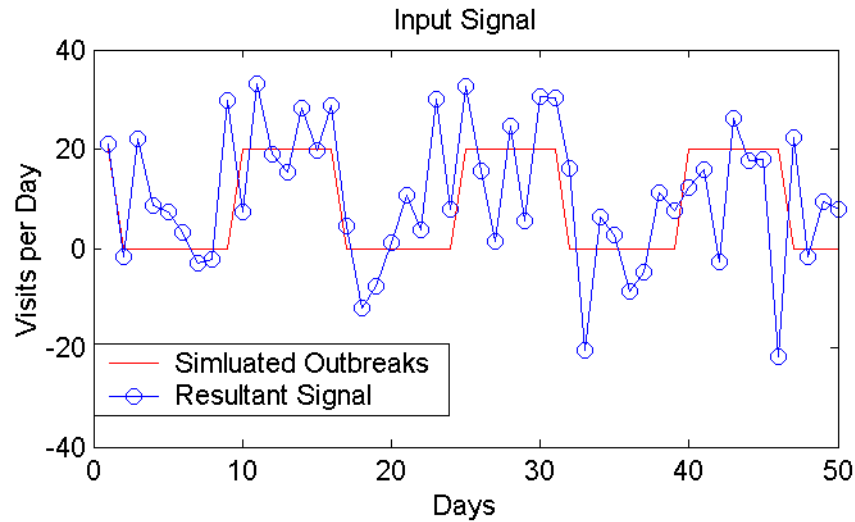


False Negatives

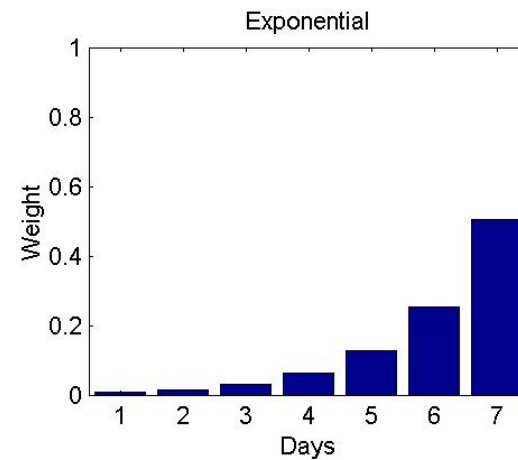
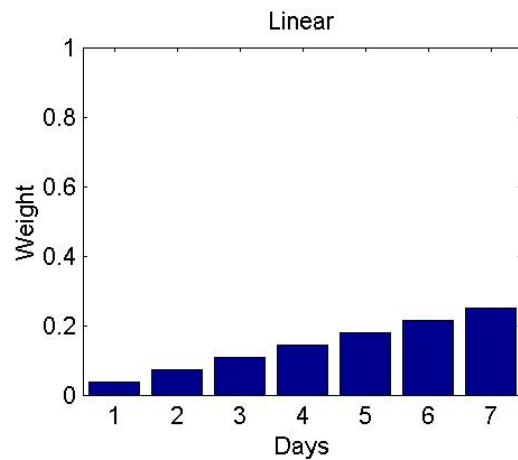
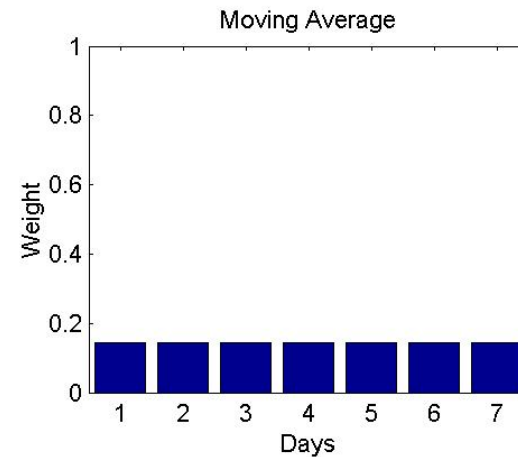
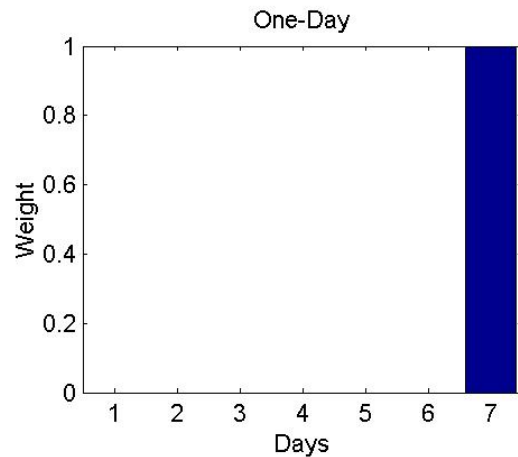


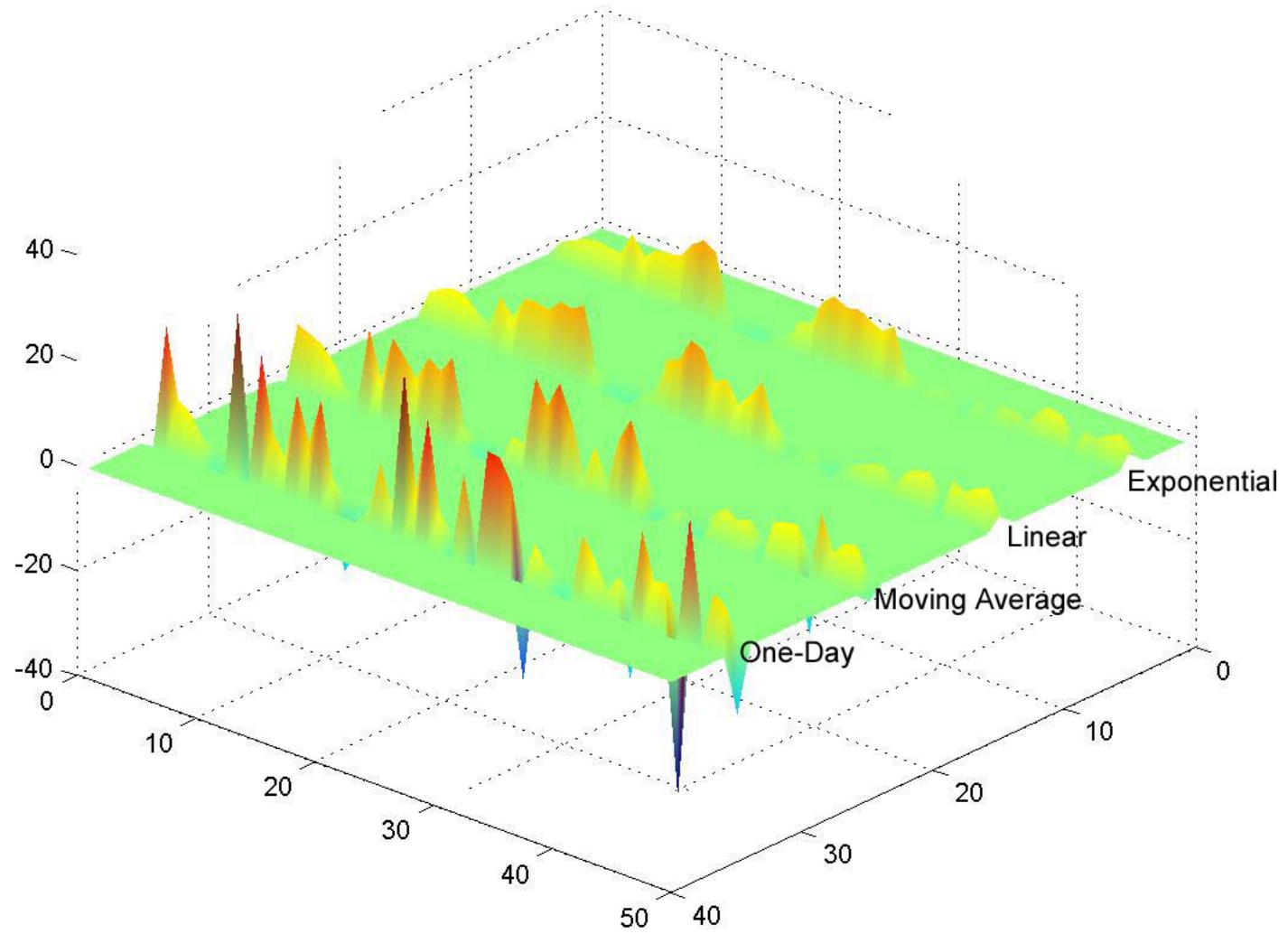
Error Profile





Combating Noise: Multiple Moving Filters





Reis, Pagano, Mandl PNAS 2003

Harvard team suggests route to better bioterror alerts

Jonathan Knight, San Francisco

A simple improvement in the way health data are monitored for signs of a bioterror attack could speed up the process and cut the number of false alarms, says a team of specialists at Harvard Medical School.

The findings, reported online this week (B. Y. Reis, M. Pagano and K. D. Mandl *Proc. Natl Acad. Sci USA* doi:10.1073/pnas.0335026100; 2003), are expected to influence the current US drive to improve early-warning systems for such attacks.

Public-health officials fear that biological attacks may not be recognized until it is too late to prevent casualties. Smallpox and anthrax, for example, start with 'flu-like symptoms, and the first victims are likely just to be sent home to rest.

Biodefence researchers have been looking at everything from patterns of hospital visits to sales of cough syrup. Dozens of systems are now being field-tested by state and local health departments across the United States. In theory, a sudden outbreak of disease, whether natural or deliberate, will register as a spike in the data, alerting health officials.

But the chief difficulty is separating this from day-to-day variation. "There are a lot

of bumps and noise in public healthcare data," says Ben Reis, a biosurveillance specialist with Harvard Medical School at the Children's Hospital in Boston. To prevent false alarms, most systems set the alert threshold so high that they risk missing the first signs of a real outbreak.

The standard approach is to forecast the number of emergency cases that hospitals have to deal with one day at a time, based on historical data. Departures from the forecast send an alert to a regional epidemiologist for further investigation.

Reis designed his system to look at the data a week at a time. He reasoned that the wider window would make it easier to disregard blips that might otherwise register as false positives. It should also spot rising trends earlier than the standard software.

He tested the approach with emergency records from the Children's Hospital, which comprise the main complaint of every patient who checked in from 1992 to 2002 — a total of more than 500,000 visits. Because there were no real outbreaks, Reis added simulated ones calculated to look like small or large releases of anthrax or smallpox.

The week-long average was able to reveal

outbreaks that the one-day system missed, Reis and colleagues report, because its detection threshold could be set much lower without triggering false alarms.

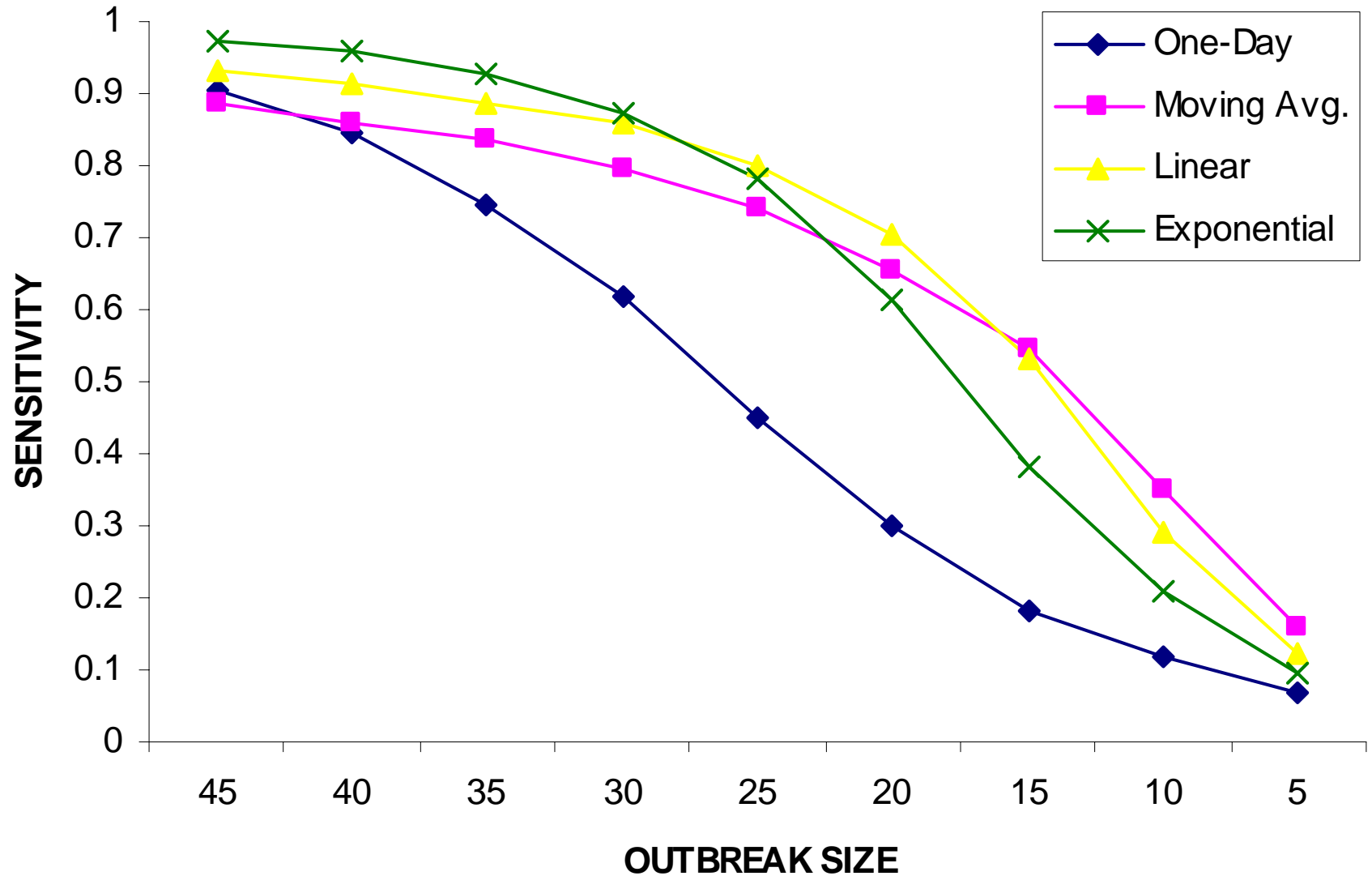
Marc Overhage, an expert in healthcare informatics at the Regenstrief Institute in Indianapolis, says that false alarms need to be eliminated. Investigations into possible outbreaks are expensive, costing an average of \$50,000, he says. Too many could render a system worthless. "We shouldn't build all these surveillance networks until we know they work," he says. "Reis is the only one running through how to do it best." ■



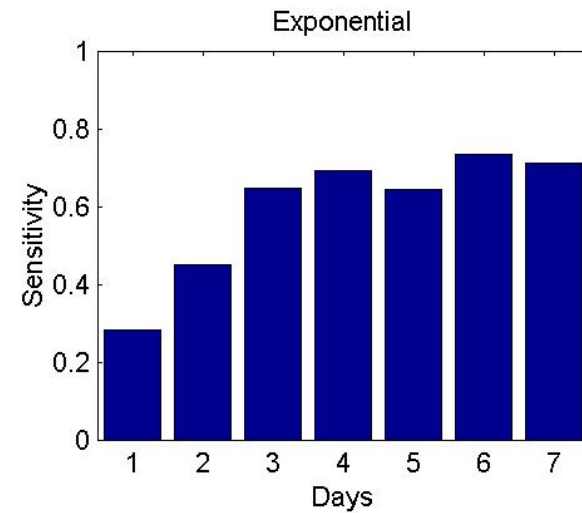
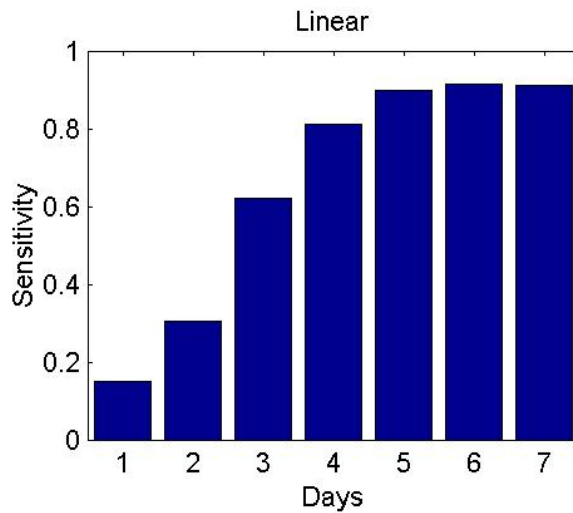
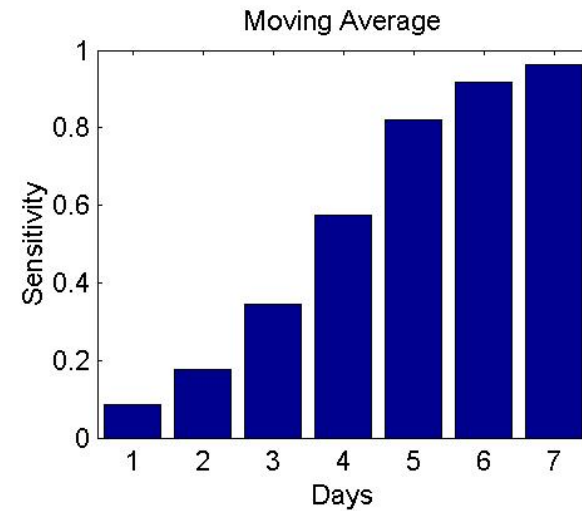
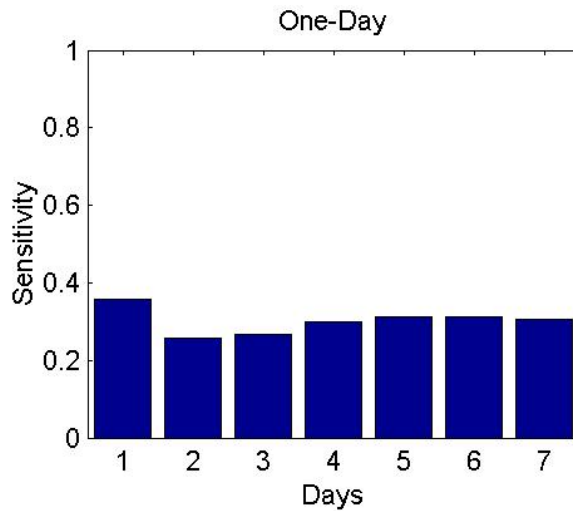
Anthrax investigation: false alarms are costly.

K. LAMBERT/AP

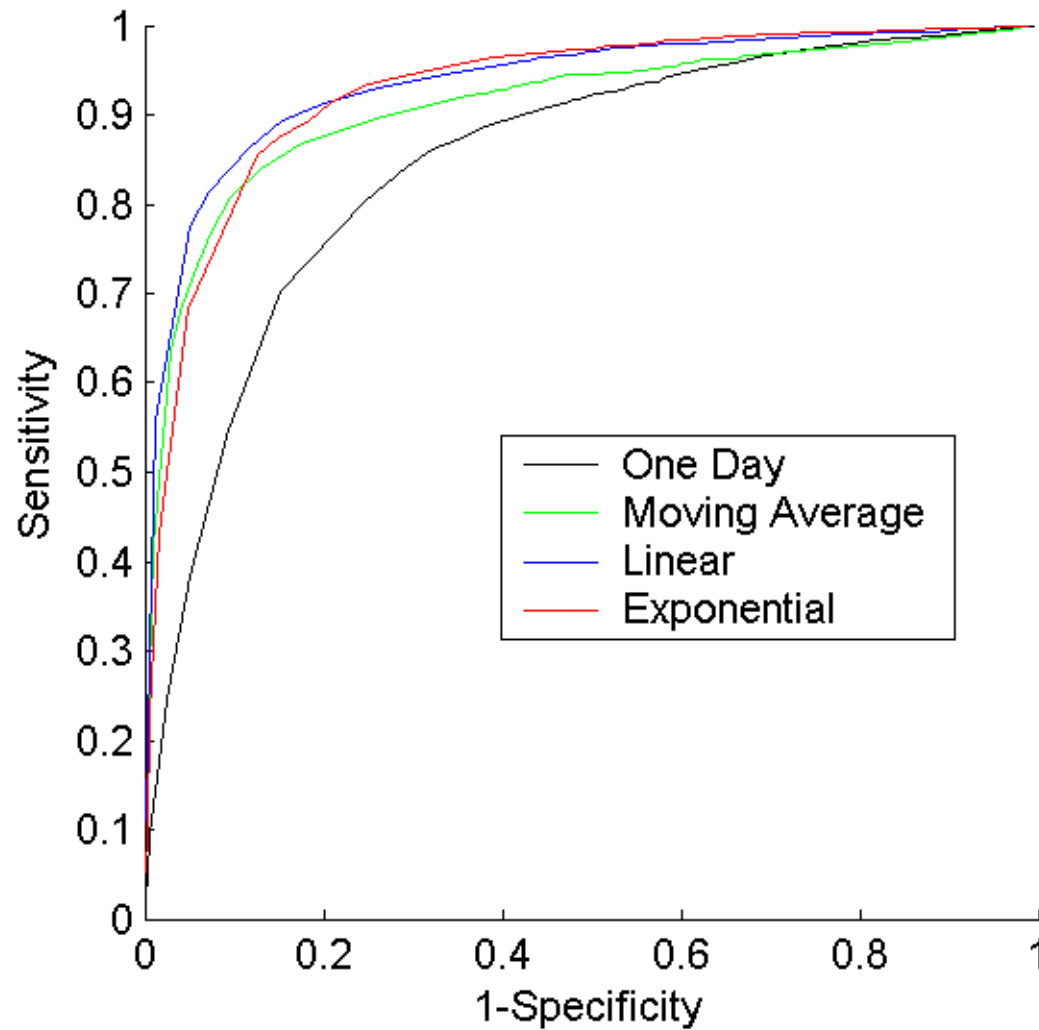
More Reliable Detection



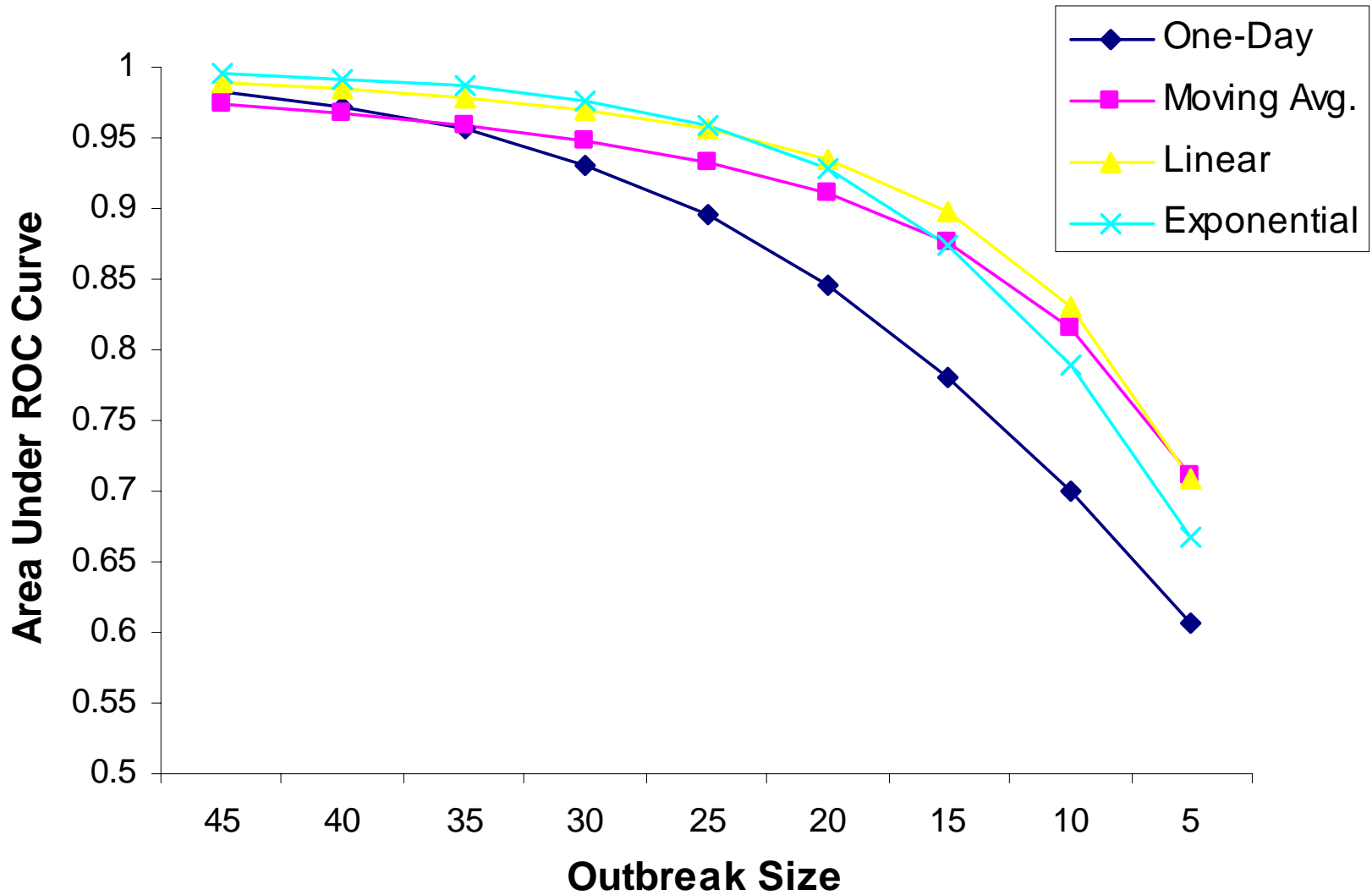
Earlier Detection



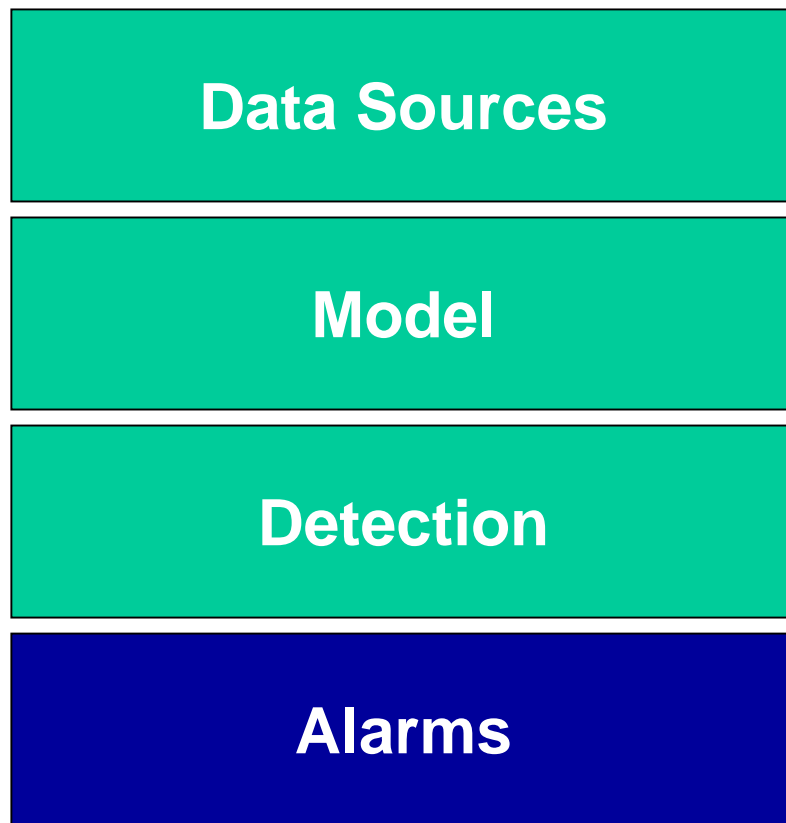
ROC Curve Analysis



Overall Detection Quality



Surveillance Stages



**Early, Reliable
Detection of
Signals**

Notification and Alarms

- Maximize overall value of system
 - Diminishing returns from false alarms
 - Identify the audience(s) of system
 - Explain tradeoff between sensitivity and specificity
 - Let target audience determine value judgment
 - e.g. number of false alarms per month

Overall Conclusions

- 1) Data sources:
 - Evaluate tradeoff between quality and timeliness
- 2) Model:
 - Consider cyclical patterns, confounding factors, and autocorrelation
- 3) Detection:
 - Enhance sensitivity and timeliness by looking at multiple time points, data sources and algorithms
- 4) Alarms:
 - Understand target audience