An introduction to time series approaches in biosurveillance

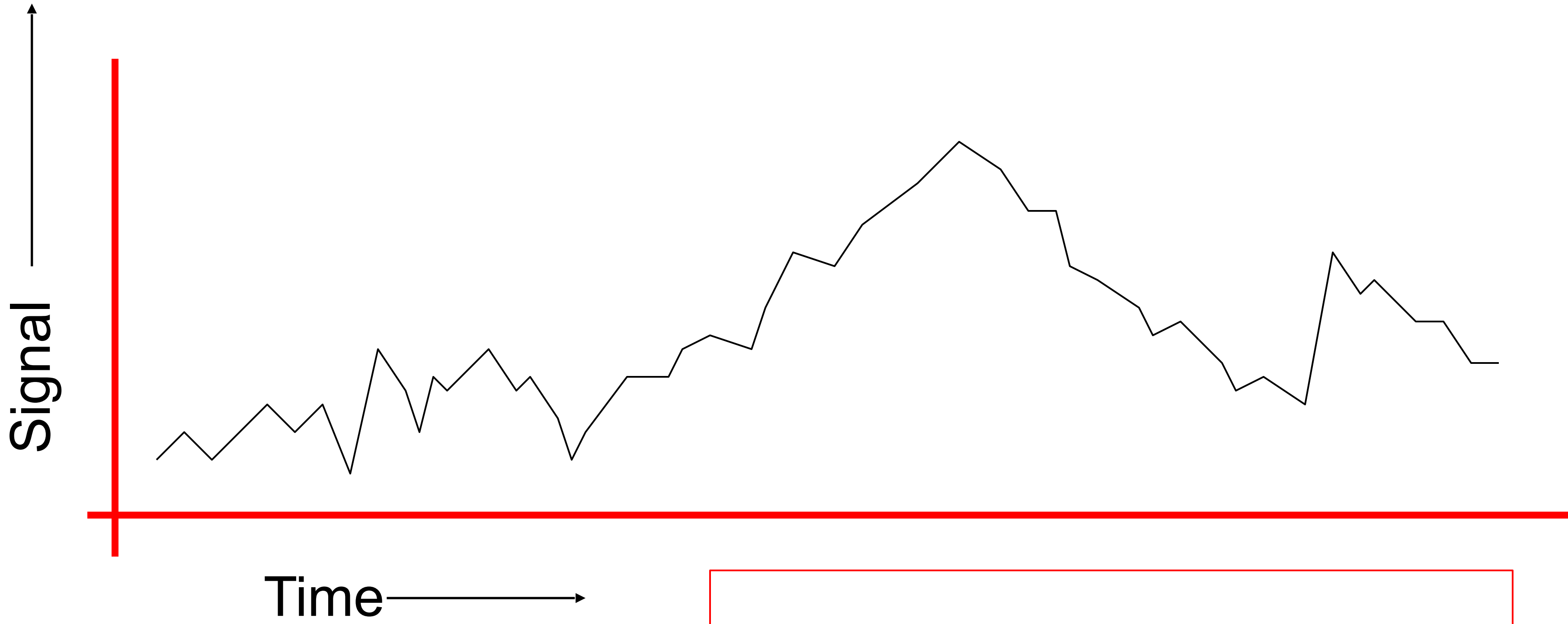
Andrew W. Moore

Professor
The Auton Lab
School of Computer Science
Carnegie Mellon University
http://www.autonlab.org

Associate Member
The RODS Lab
University of Pittburgh
Carnegie Mellon University
http://rods.health.pitt.edu

awm@cs.cmu.edu
412-268-7599

Univariate Time Series



Example Signals:

- Number of ED visits today
- Number of ED visits this hour
- Number of Respiratory Cases Today
- School absenteeism today
- Nyquil Sales today

(When) is there an anomaly?

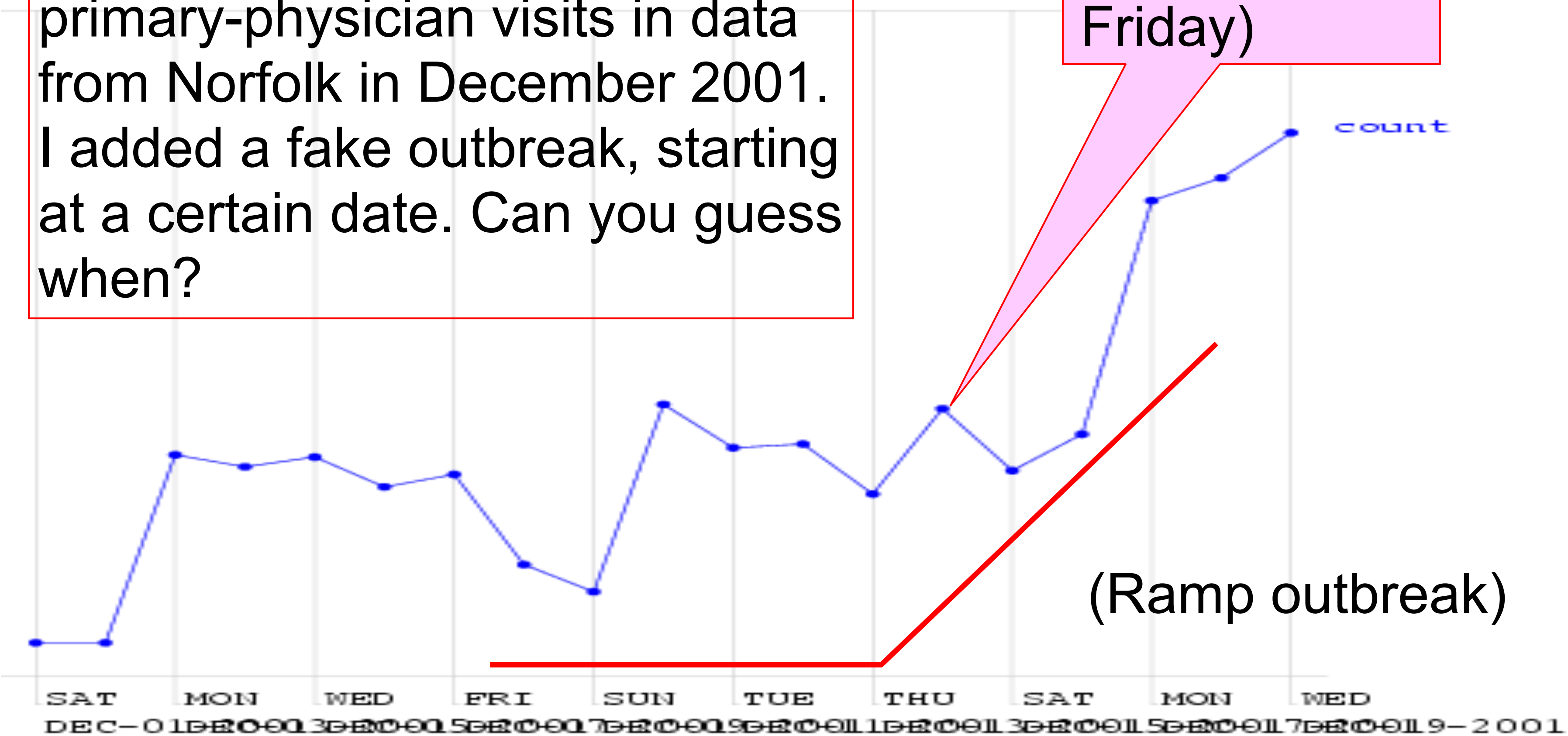
(When) is there an anomaly?

This is a time series of counts of primary-physician visits in data from Norfolk in December 2001. I added a fake outbreak, starting at a certain date. Can you guess when?

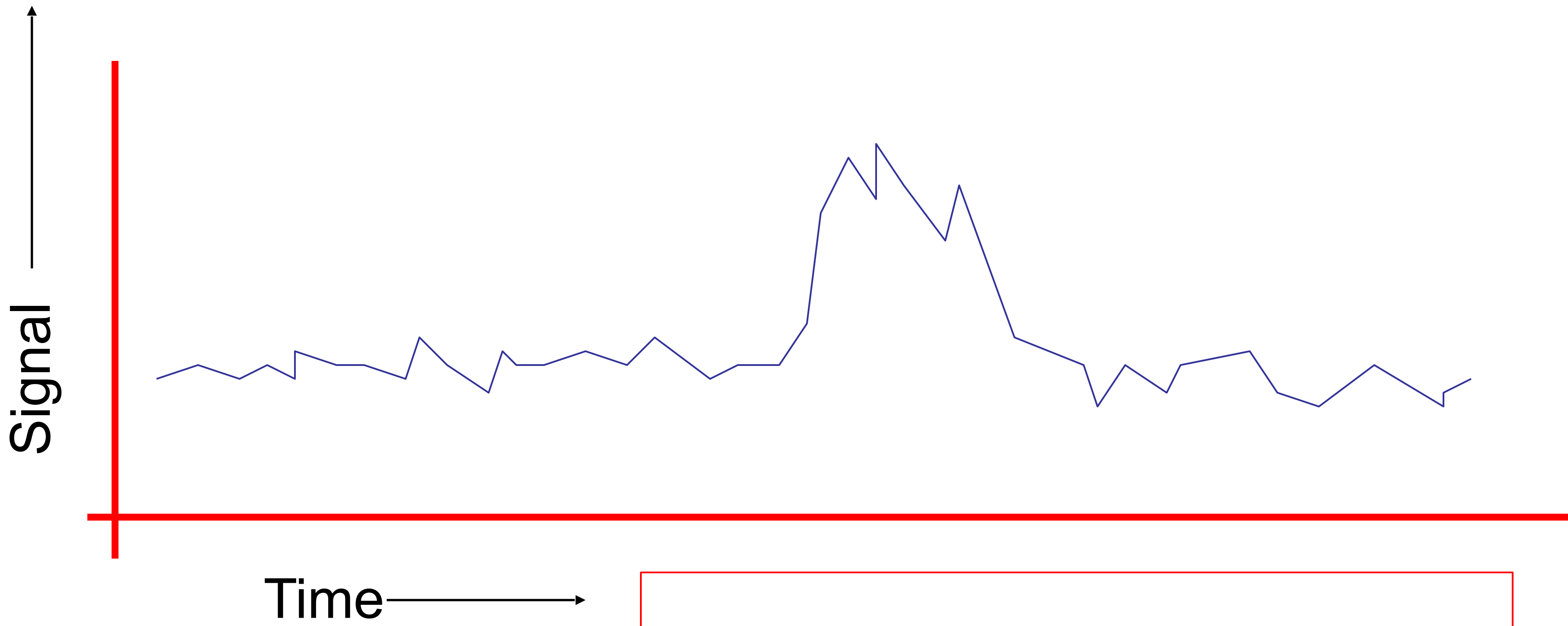
(When) is there an anomaly?

This is a time series of counts of primary-physician visits in data from Norfolk in December 2001. I added a fake outbreak, starting at a certain date. Can you guess when?

Here (much too high for a Friday)



An easy case

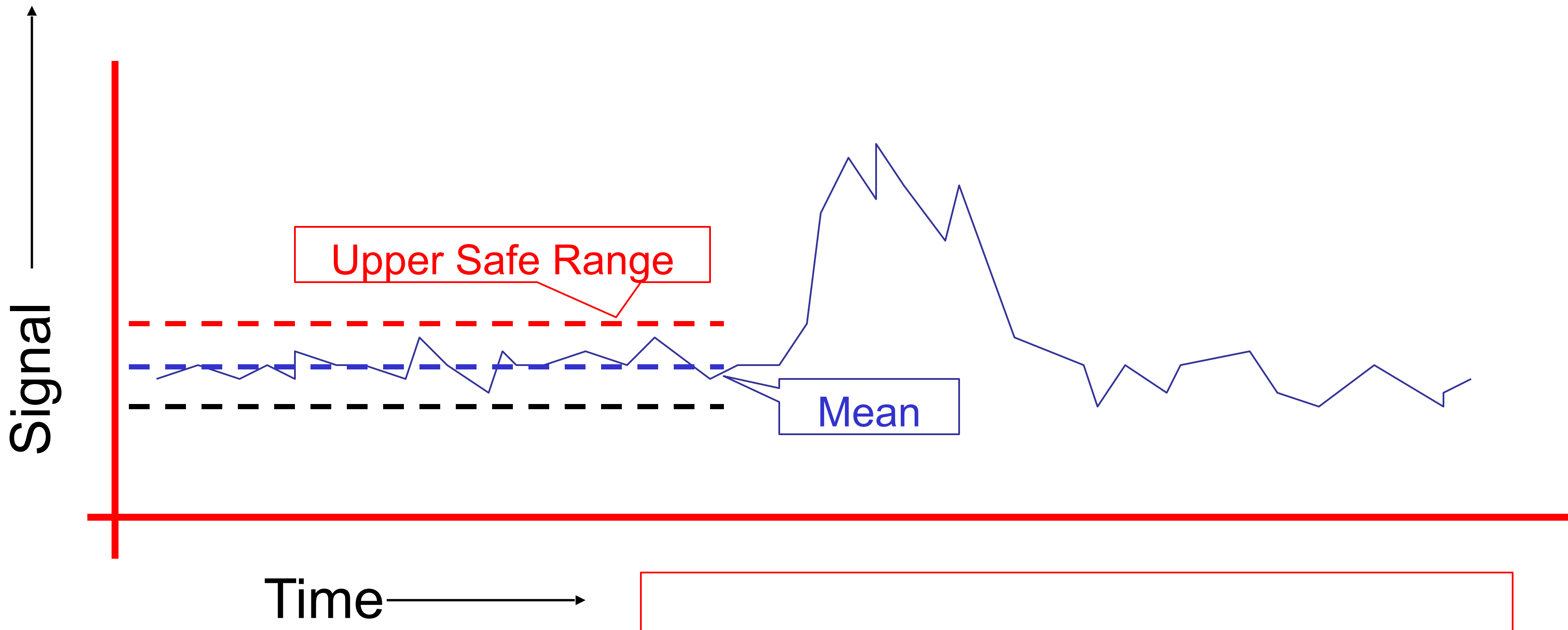


Dealt with by Statistical Quality Control

Record the mean and standard deviation up the the current time.

Signal an alarm if we go outside 3 sigmas

An easy case: Control Charts

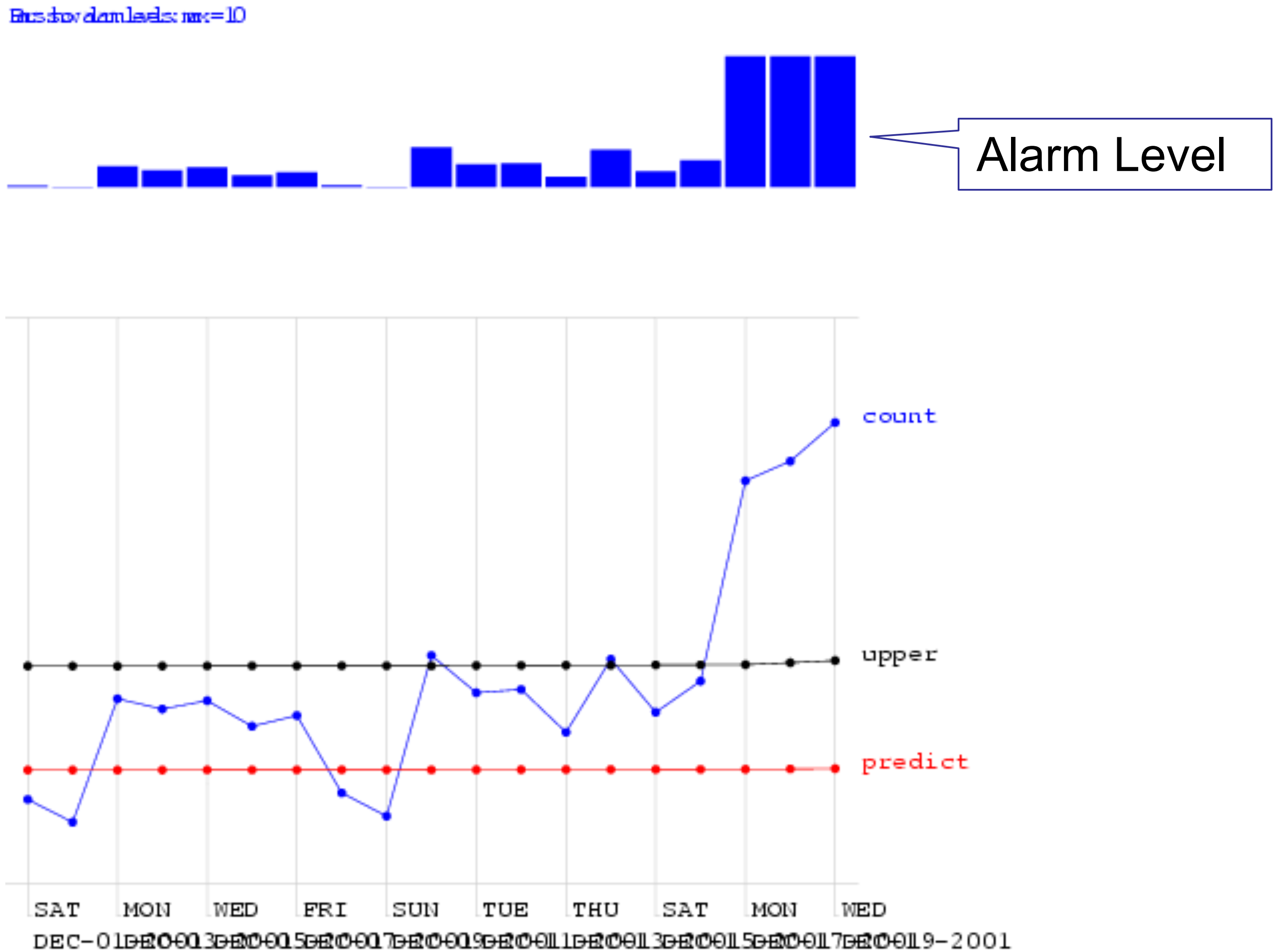


Dealt with by Statistical Quality Control

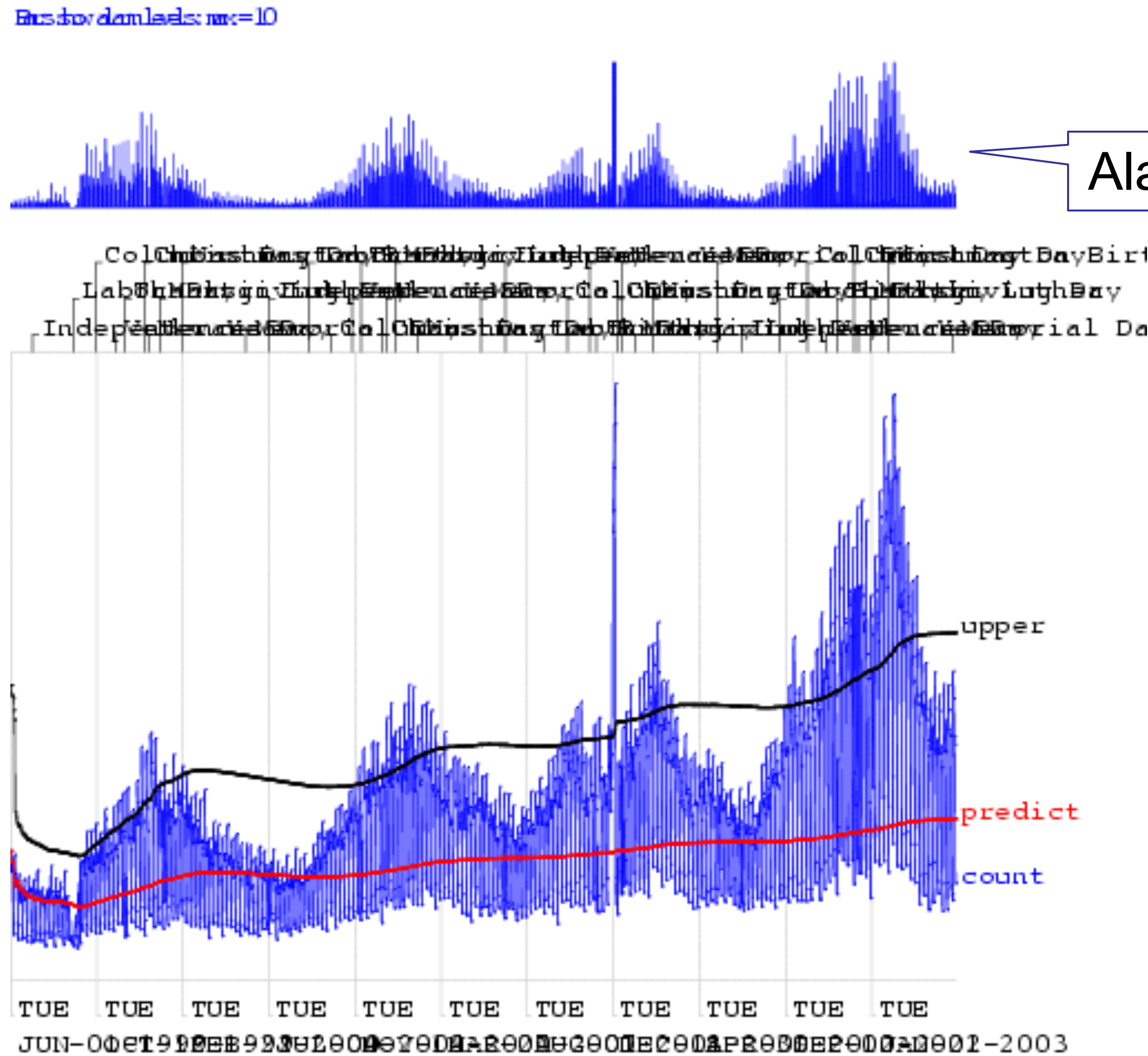
Record the mean and standard deviation up the the current time.

Signal an alarm if we go outside 3 sigmas

Control Charts on the Norfolk Data



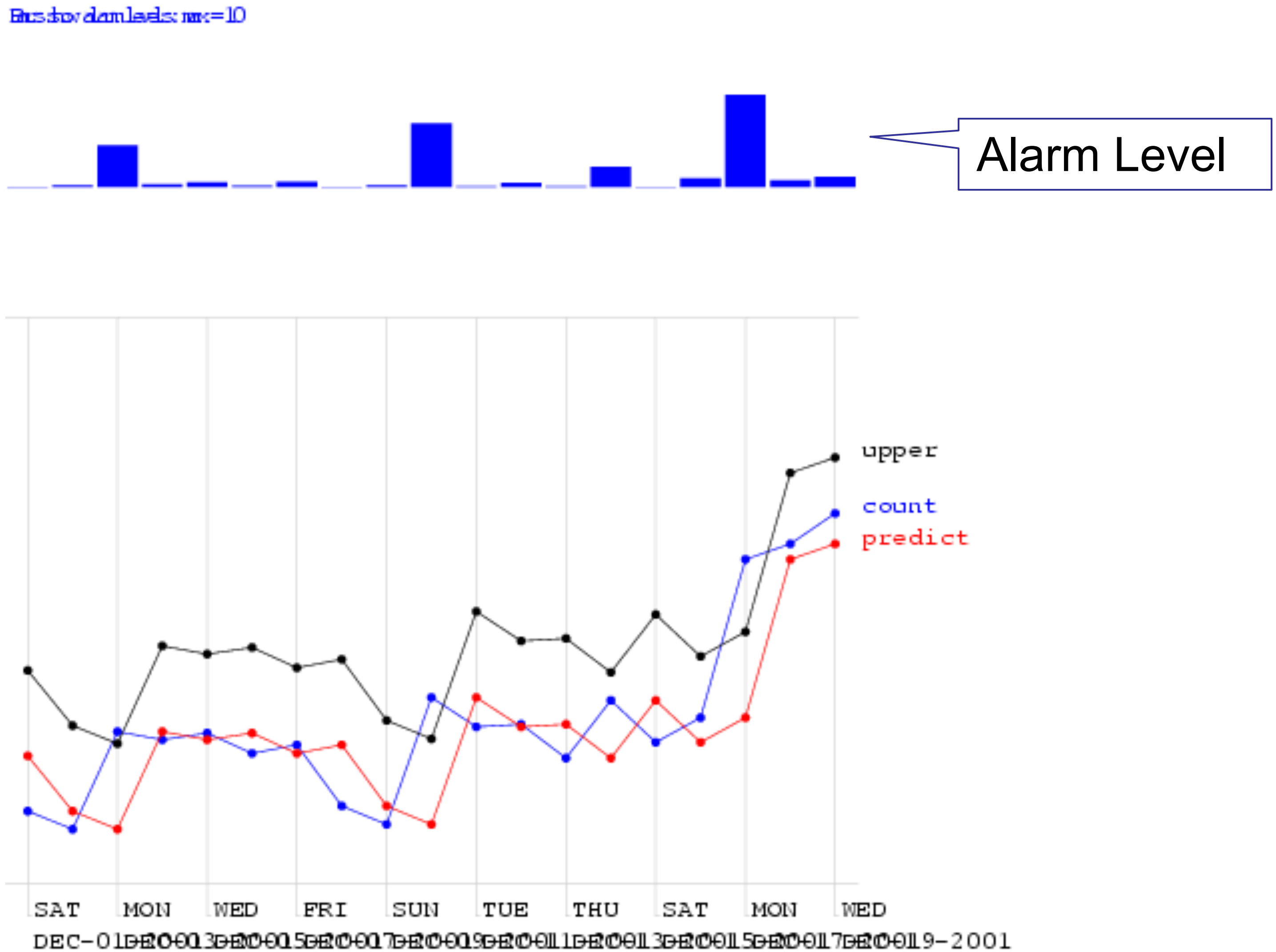
Control Charts on the Norfolk Data



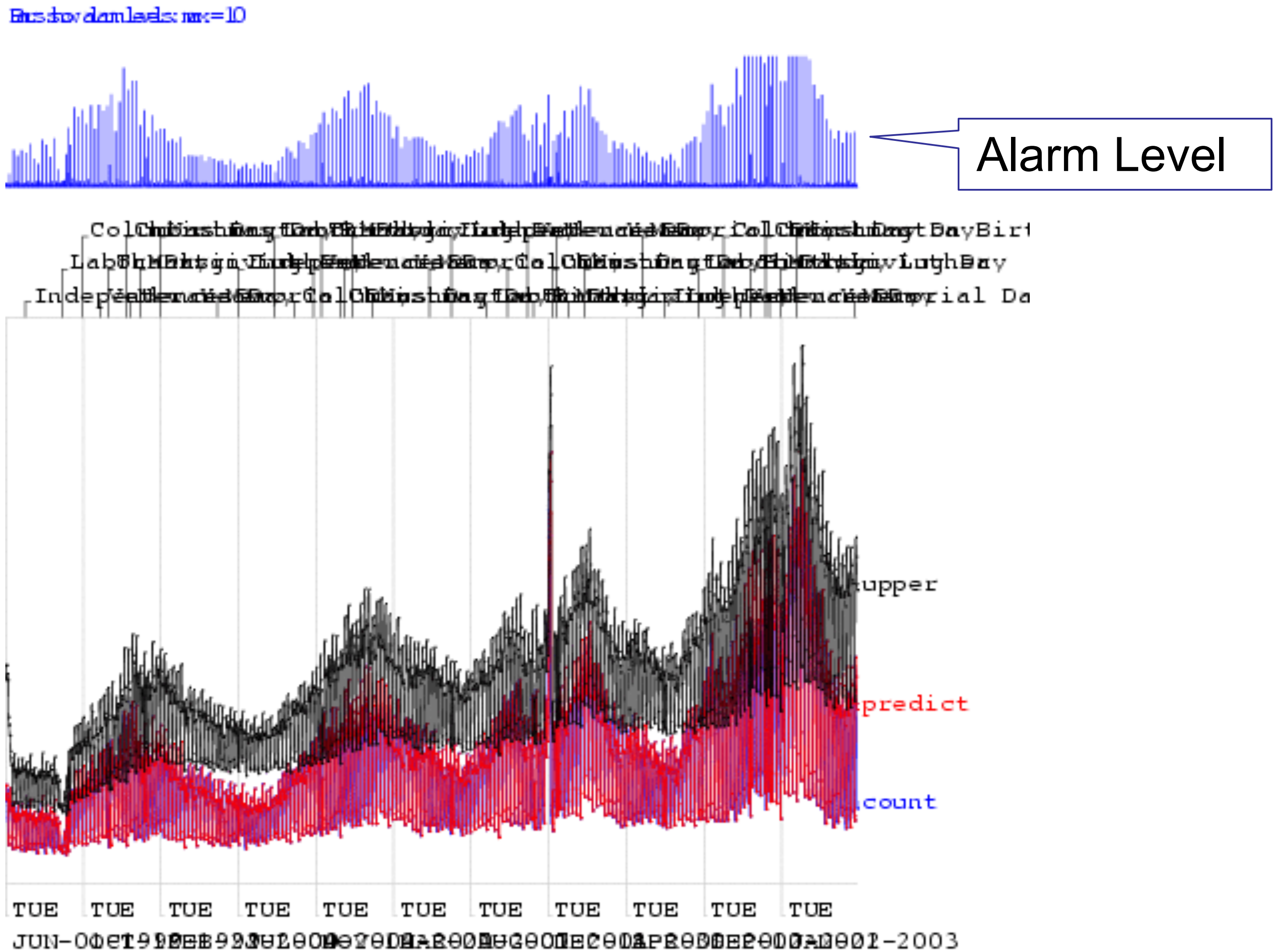
Alarm Level

Looking at changes from yesterday

Looking at changes from yesterday

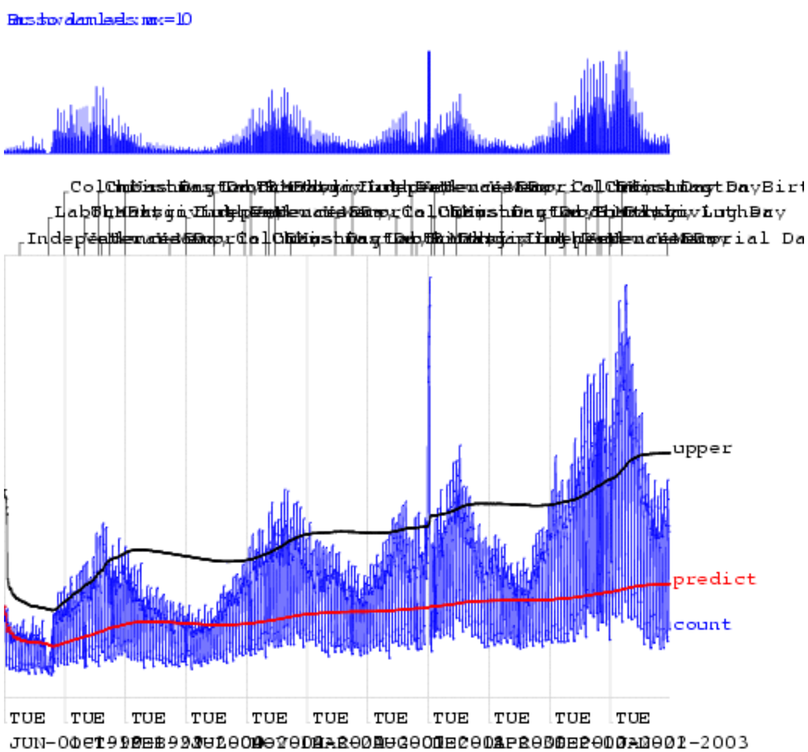


Looking at changes from yesterday

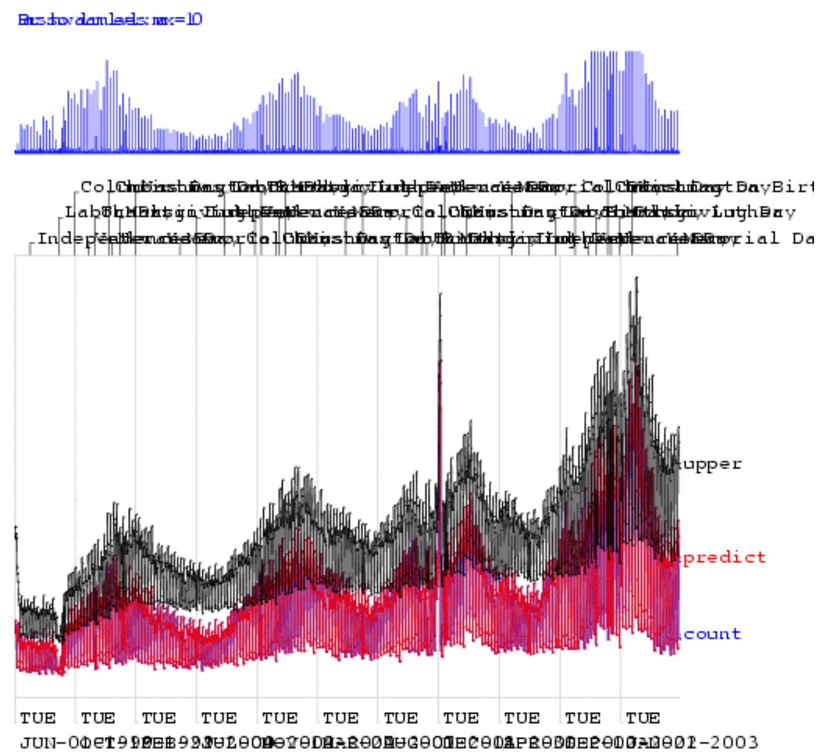


We need a happy medium:

Control Chart: Too insensitive to recent changes

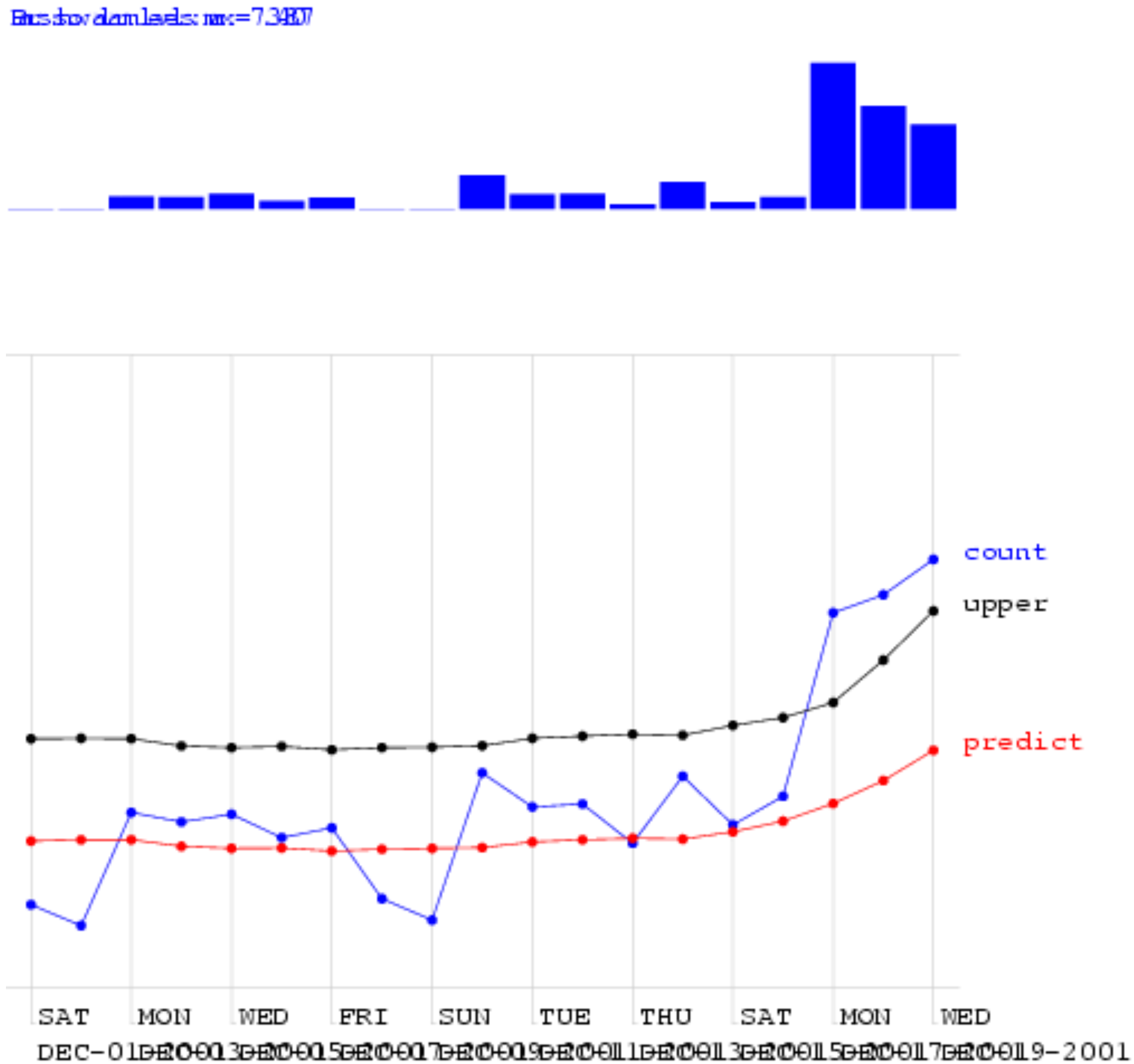


Change from yesterday: Too sensitive to recent changes

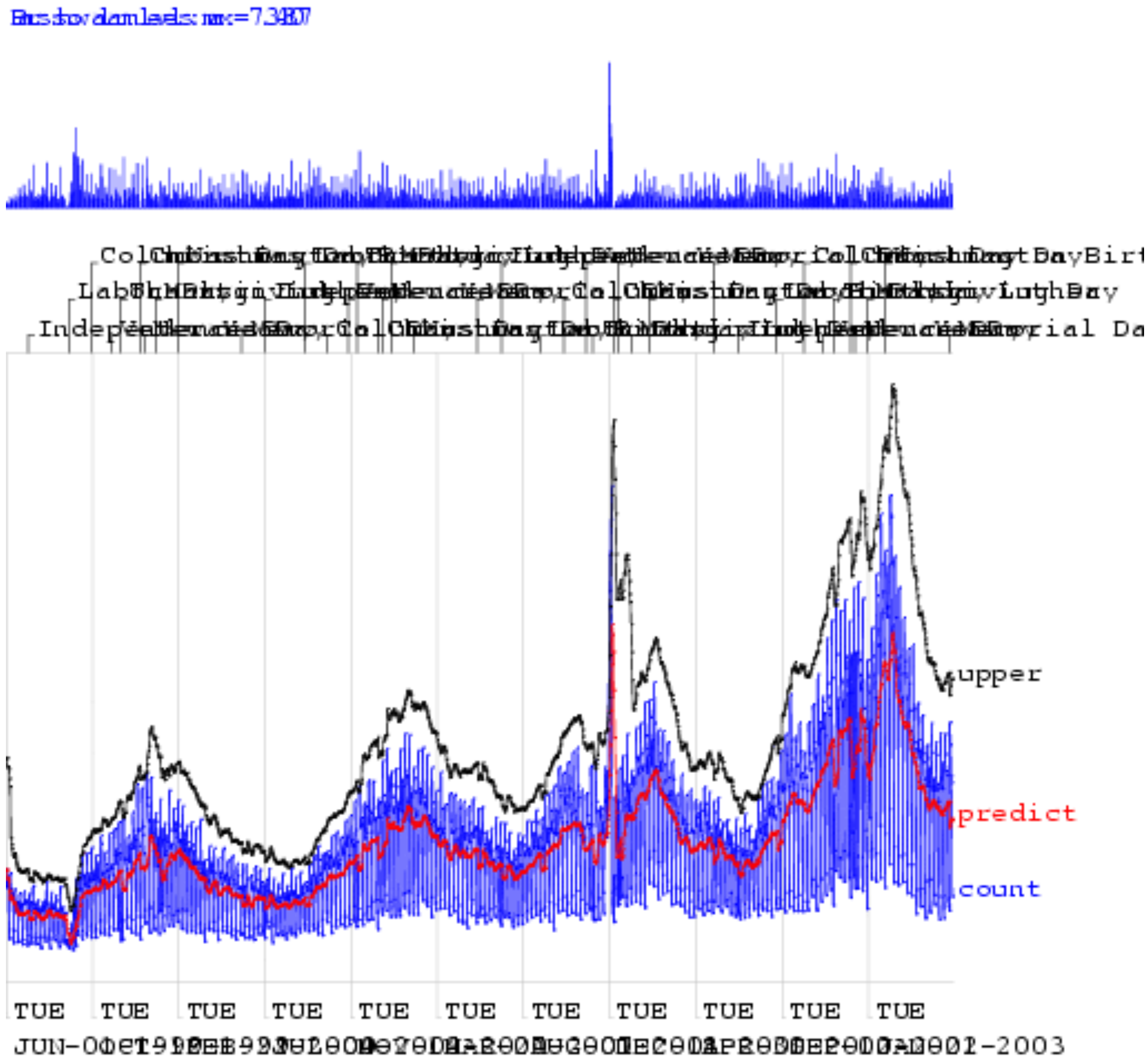


Moving Average

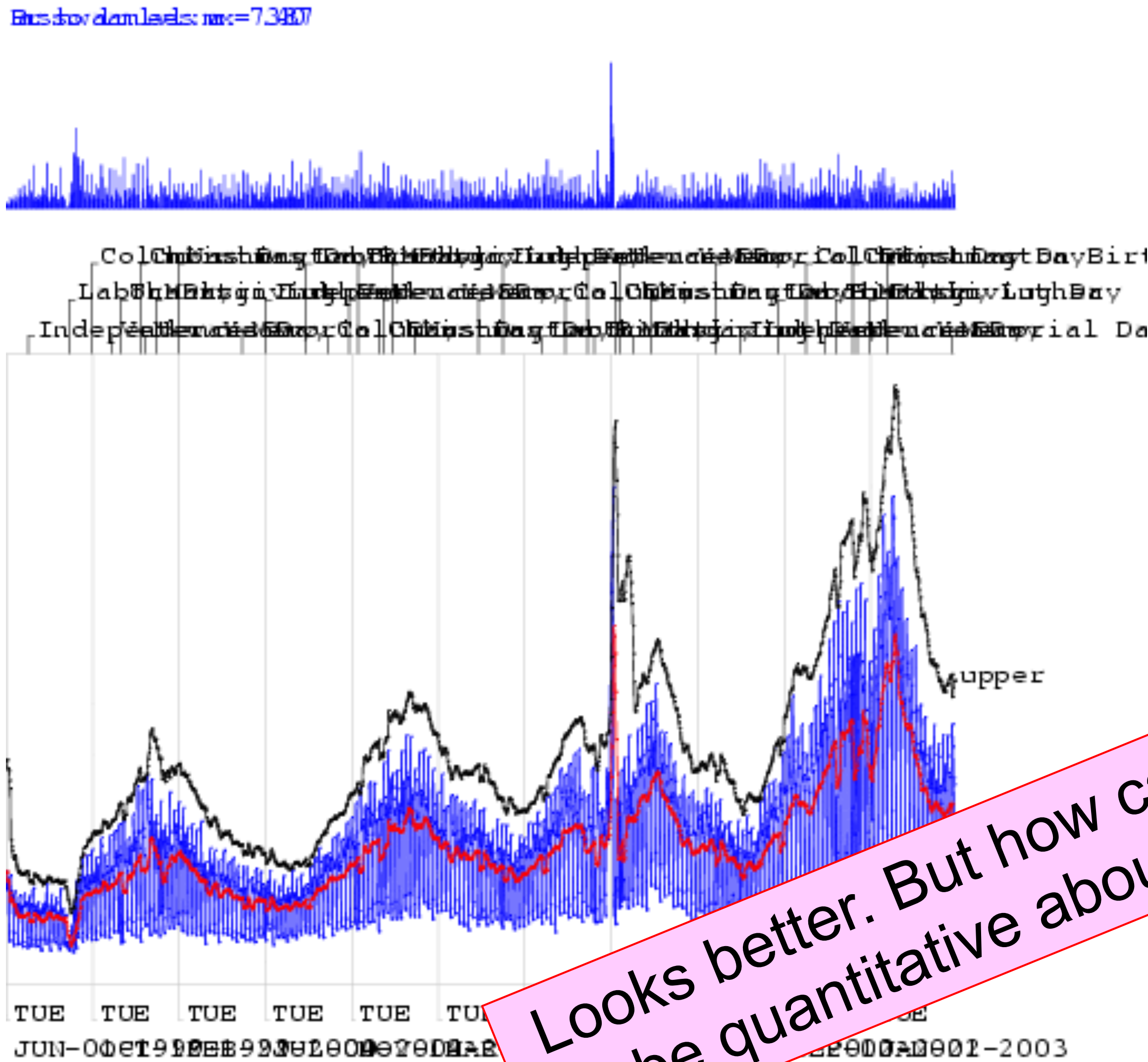
Moving Average



Moving Average



Moving Average



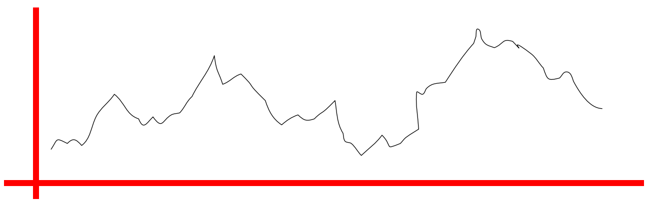
Looks better. But how can we be quantitative about this?

Algorithm Performance

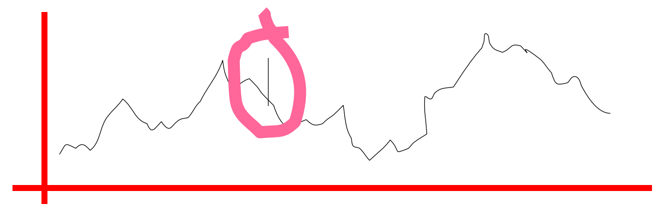
	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7

Semi-synthetic data: spike outbreaks

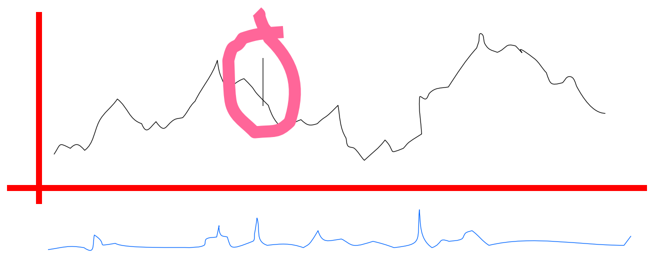
1. Take a real time series



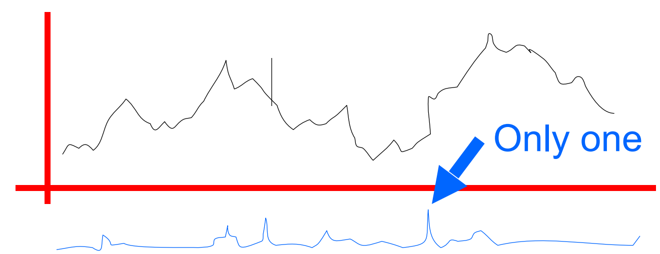
2. Add a spike of random height on a random date



3. See what alarm levels your algorithm gives on every day of the data

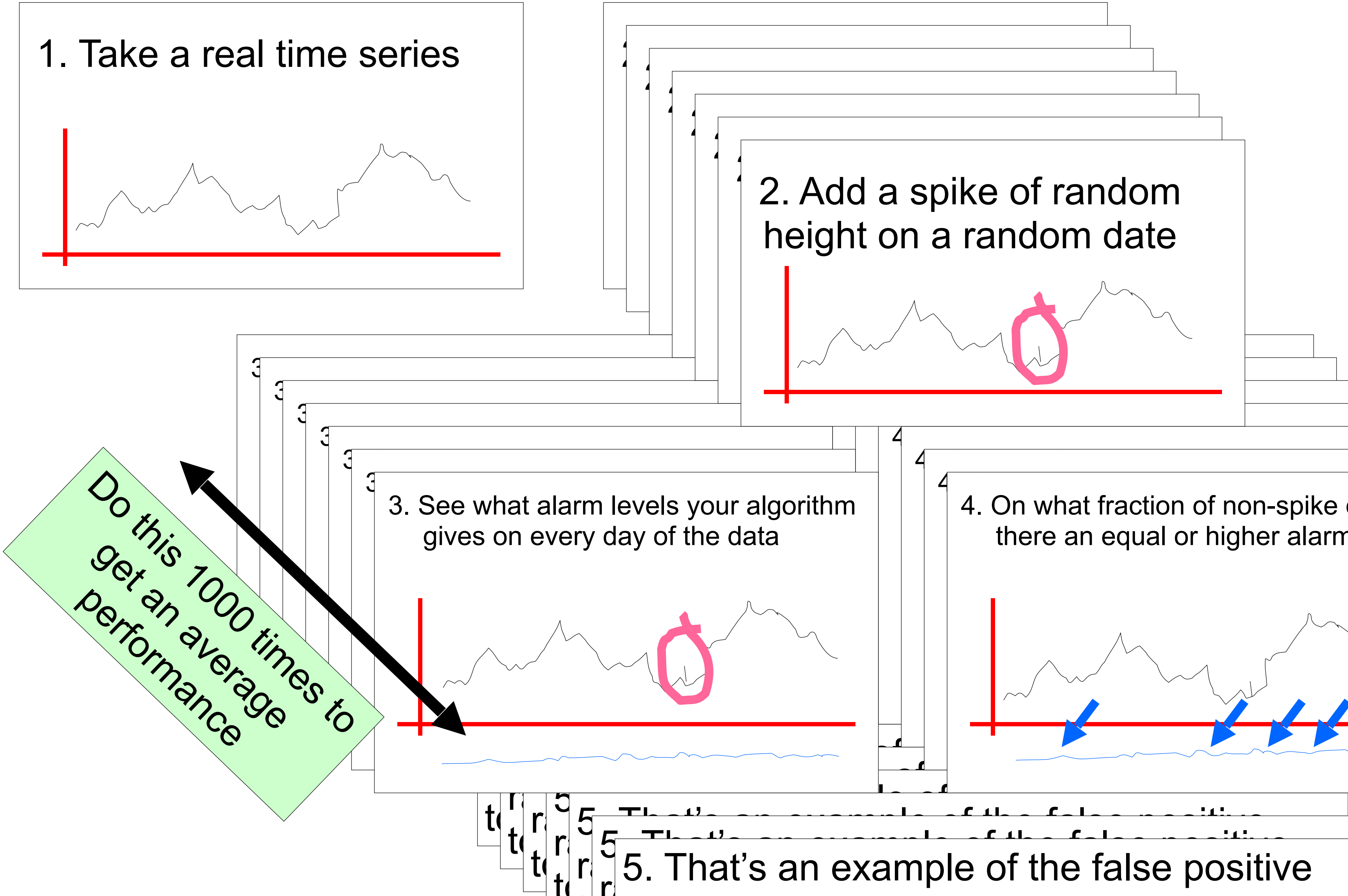


4. On what fraction of non-spike days is there an equal or higher alarm



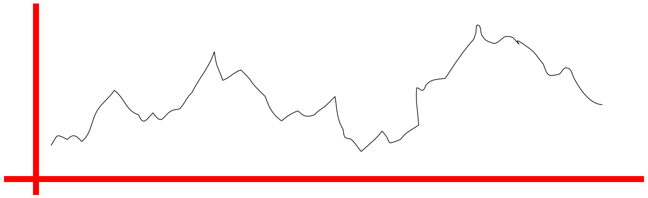
5. That’s an example of the false positive rate this algorithm would need if it was going to detect the actual spike.

Semi-synthetic data: spike outbreaks

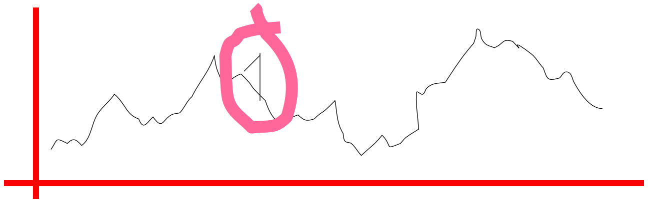


Semi-synthetic data: ramp outbreaks

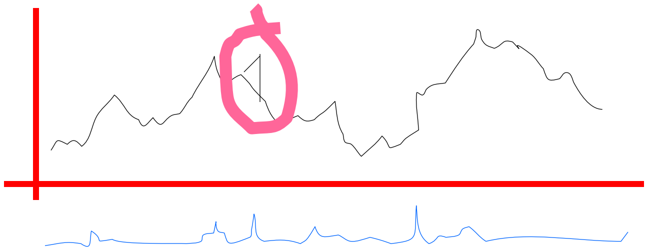
1. Take a real time series



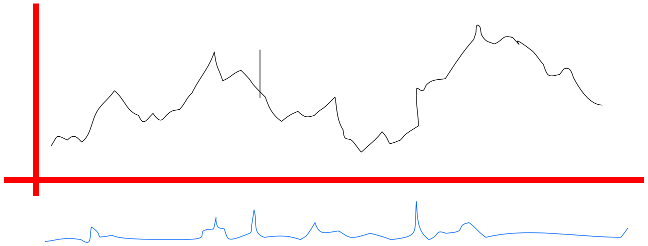
2. Add a ramp of random height on a random date



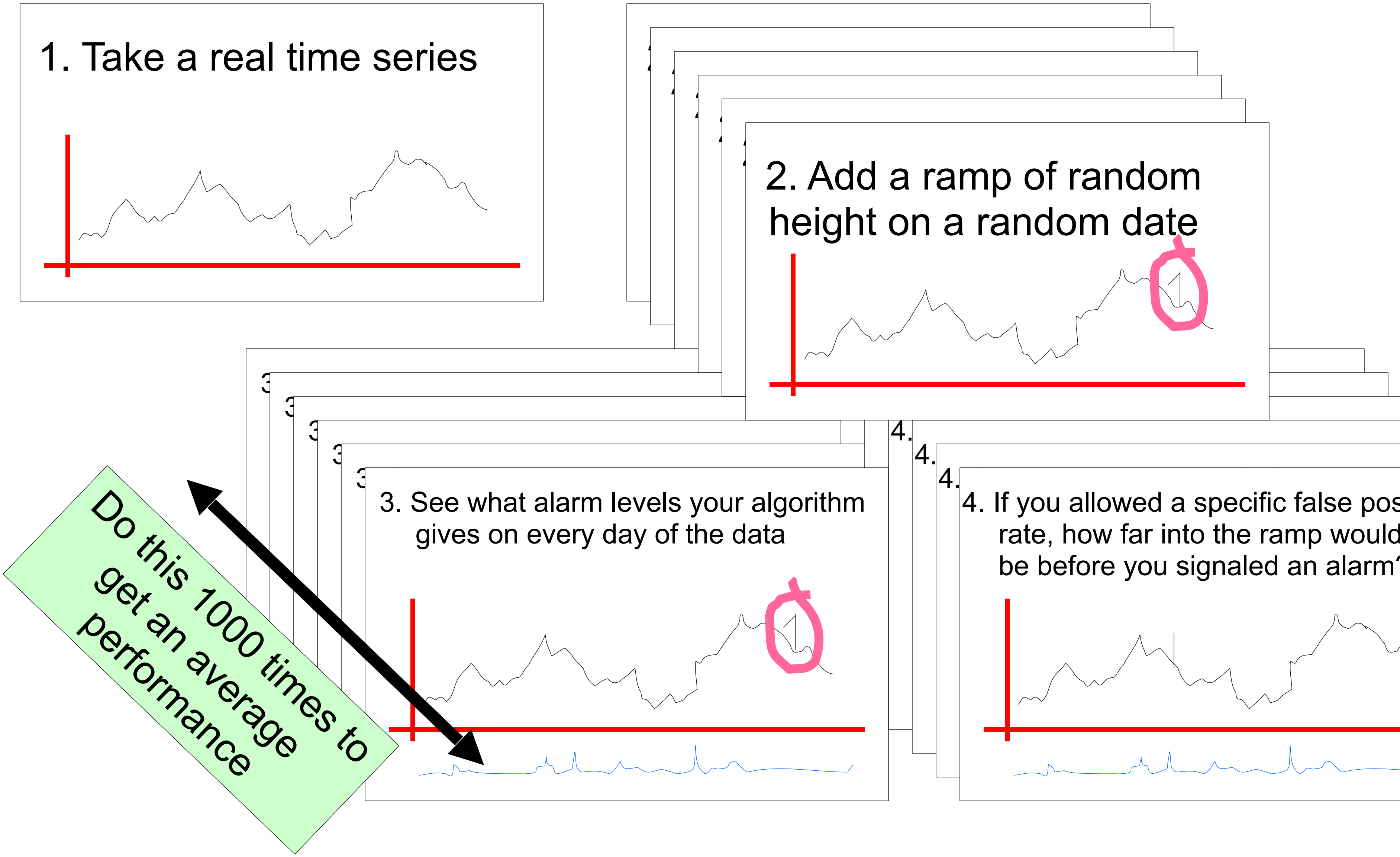
3. See what alarm levels your algorithm gives on every day of the data



4. If you allowed a specific false positive rate, how far into the ramp would you be before you signaled an alarm?



Semi-synthetic data: ramp outbreaks



Evaluation methods

All synthetic

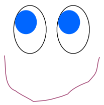
Evaluation methods

All synthetic

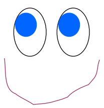
- You can account for variation in the way the baseline will look.
- You can publish evaluation data and share results without data agreement problems
- You can easily generate large numbers of tests
- You know where the outbreaks are

Evaluation methods

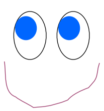
All synthetic



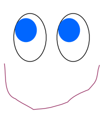
You can account for variation in the way the baseline will look.



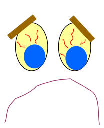
You can publish evaluation data and share results without data agreement problems



You can easily generate large numbers of tests



You know where the outbreaks are



Your baseline data might be unrealistic

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Your baseline data is realistic

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Your baseline data is realistic

Your outbreak data might be unrealistic

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Your baseline data is realistic

Your outbreak data might be unrealistic

All real

You can't get many outbreaks to test

You need experts to decide what is an outbreak

Some kinds of outbreak have no available data

You can't share data

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Your baseline data is realistic

Your outbreak data might be unrealistic

All real

You can't get many outbreaks to test

You need experts to decide what is an outbreak

Some kinds of outbreak have no available data

You can't share data

Your baseline data is realistic

Your outbreak data is realistic

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

You know where the outbreaks are

Your baseline data might be unrealistic

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

Don't know where the outbreaks aren't

Your baseline data is realistic

Your outbreak data might be unrealistic

All real

You can't get many outbreaks to test

You need experts to decide what is an outbreak

Some kinds of outbreak have no available data

You can't share data

Your baseline data is realistic

Your outbreak data is realistic

Is the test typical?

Evaluation methods

All synthetic

You can account for variation in the way the baseline will look.

You can publish evaluation data and share results without data agreement problems

You can easily generate large numbers of tests

Yo
out

You
migl

Semi-Synthetic

Can't account for variation in the baseline.

You can't share data

You can easily generate large numbers of tests

You know where the outbreaks are

ak data
might be unrealistic

All real

You can't get many outbreaks to test

You need experts to decide what is an outbreak

Some kinds of outbreak have no available data

share data

ine data is

ak data

Is the test typical?

None of these options is satisfactory.

Evaluation of Biosurveillance algorithms is really hard. It has got to be. This is a real problem, and we must learn to live with it.

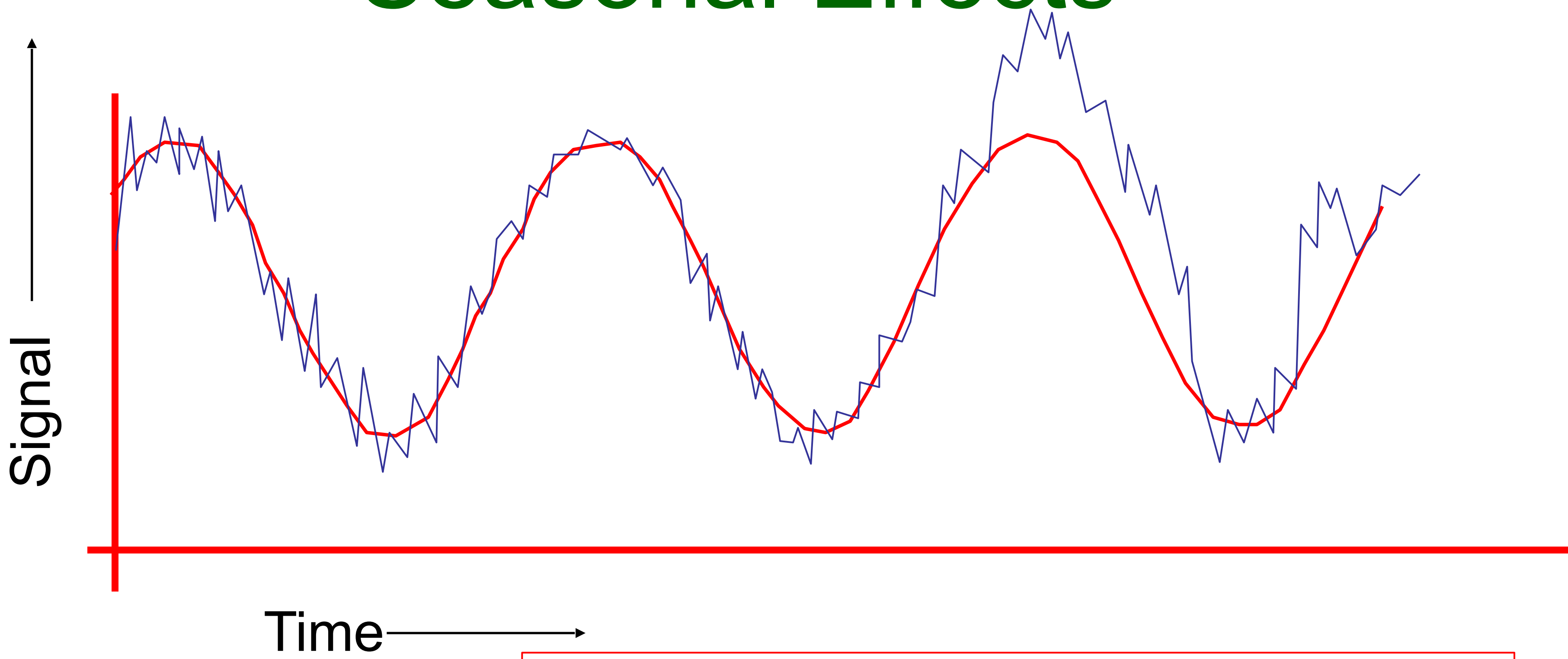
Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 7	0.58	2.79	0.51	3.31

Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54

Seasonal Effects



Fit a periodic function (e.g. sine wave) to previous data. Predict today's signal and 3-sigma confidence intervals. Signal an alarm if we're off.

Reduces False alarms from Natural outbreaks.

Different times of year deserve different thresholds.

Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9

Day-of-week effects

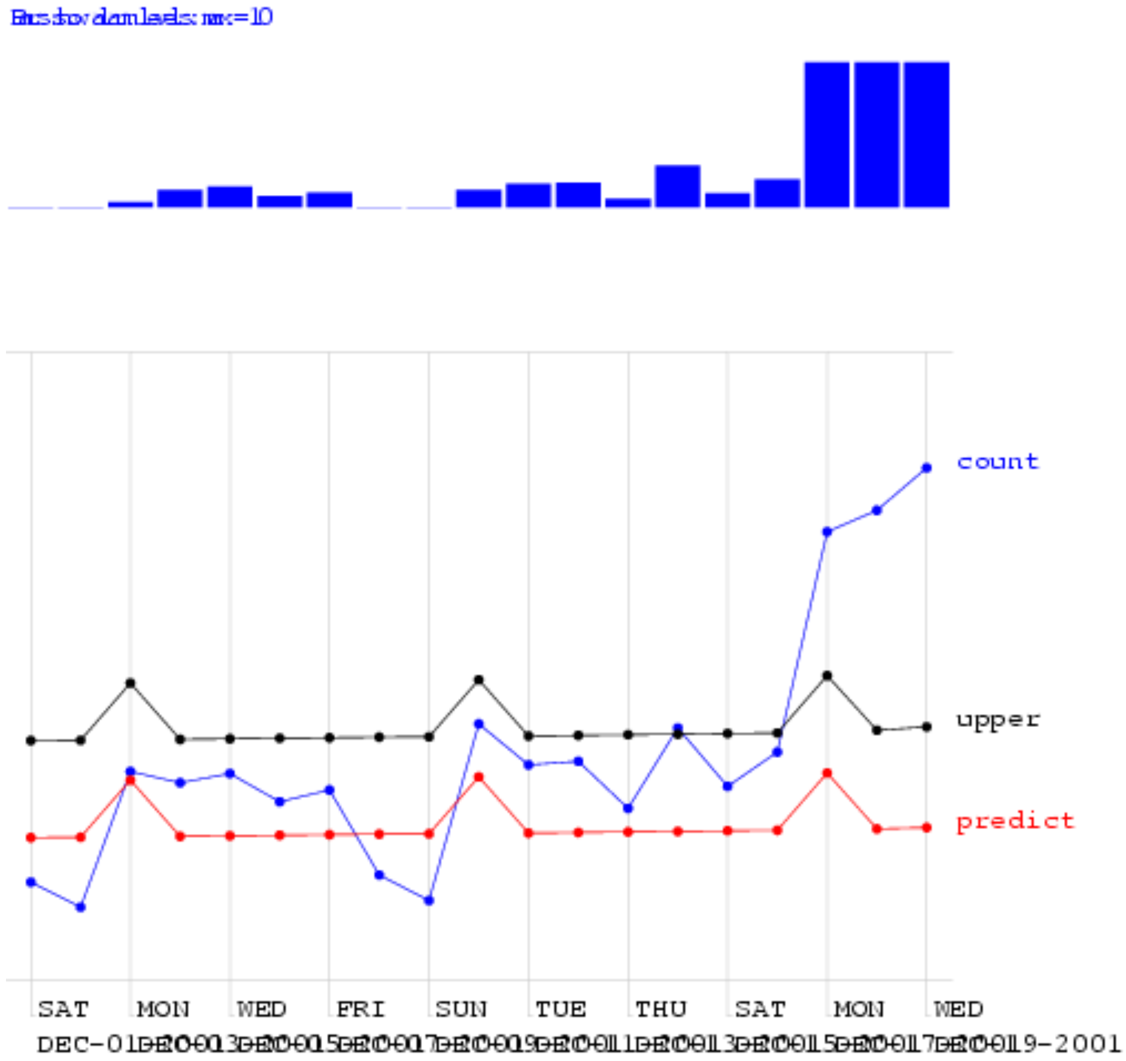
Fit a day-of-week component

$E[\text{Signal}] = a + \text{delta}_{day}$

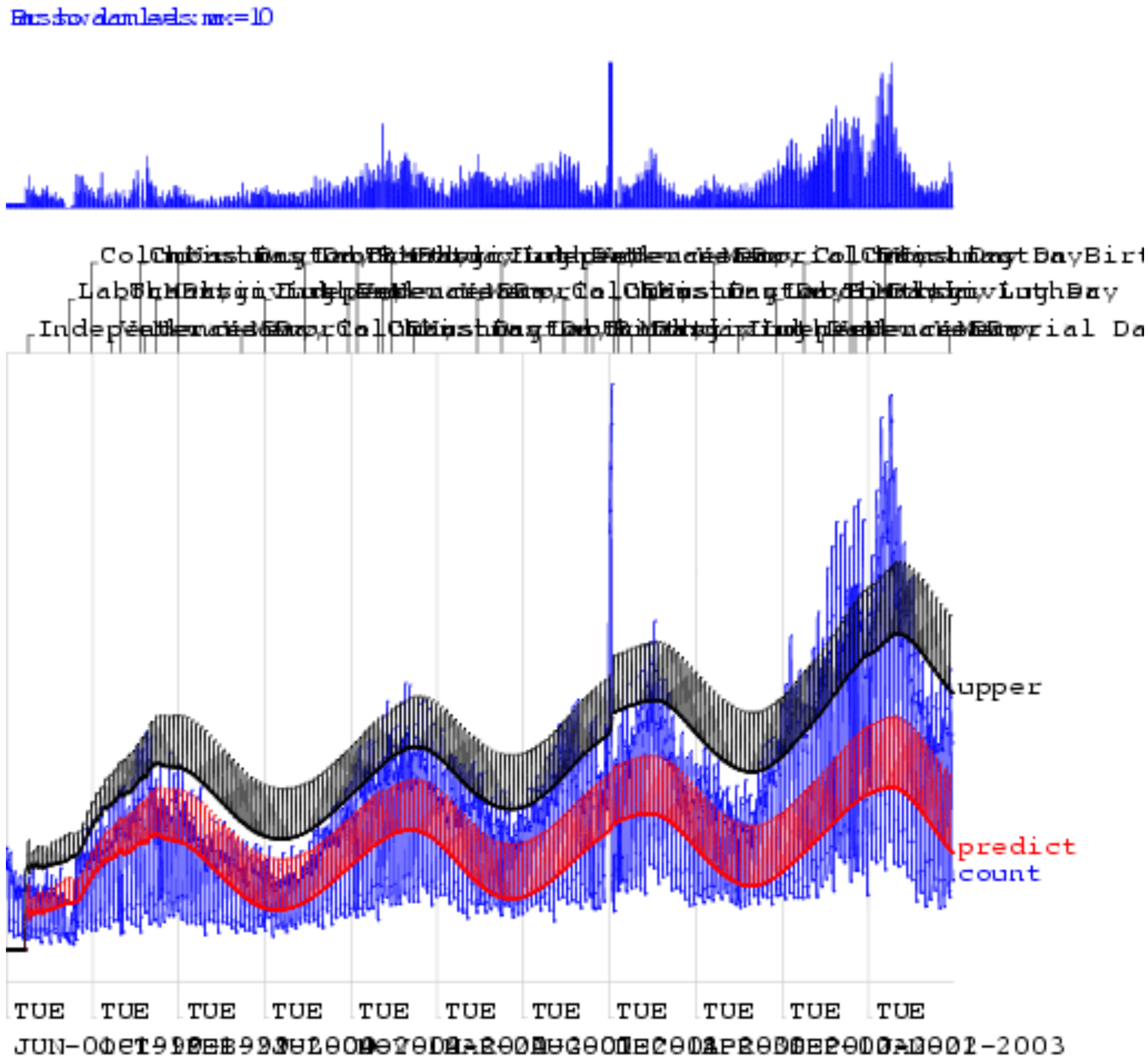
E.G: $\text{delta}_{mon} = +5.42$, $\text{delta}_{tue} = +2.20$, $\text{delta}_{wed} = +3.33$, $\text{delta}_{thu} = +3.10$, $\text{delta}_{fri} = +4.02$, $\text{delta}_{sat} = -12.2$, $\text{delta}_{sun} = -23.42$

A simple form of ANOVA

Regression using Hours-in-day & IsMonday



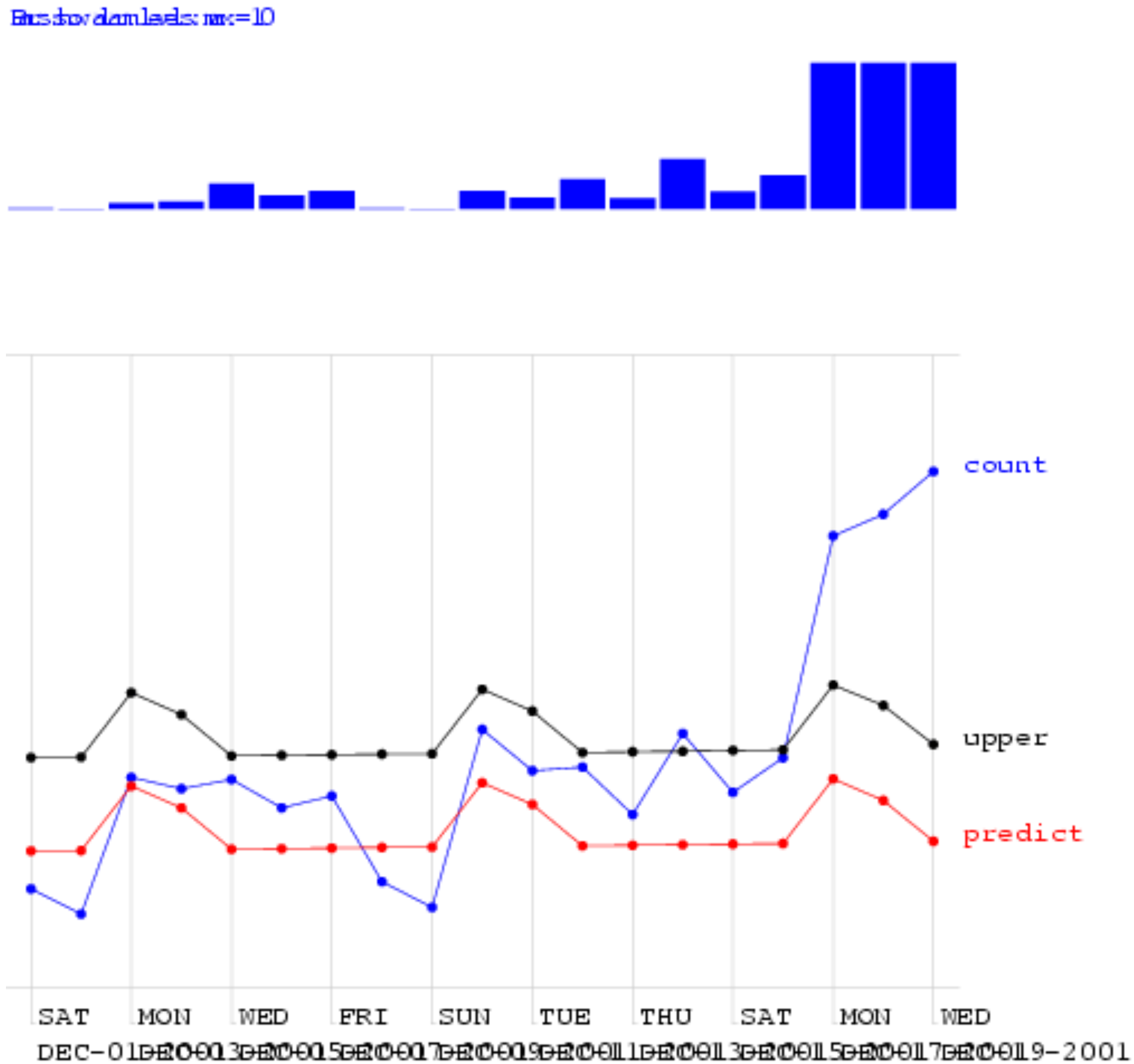
Regression using Hours-in-day & IsMonday



Algorithm Performance

	Allowing one False Alarm per TWO weeks… Fraction of spikes detected	Allowing one False Alarm per TWO weeks… Days to detect a ramp outbreak	Allowing one False Alarm per SIX weeks… Fraction of spikes detected	Allowing one False Alarm per SIX weeks… Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12

Regression using Mon-Tue



Algorithm Performance

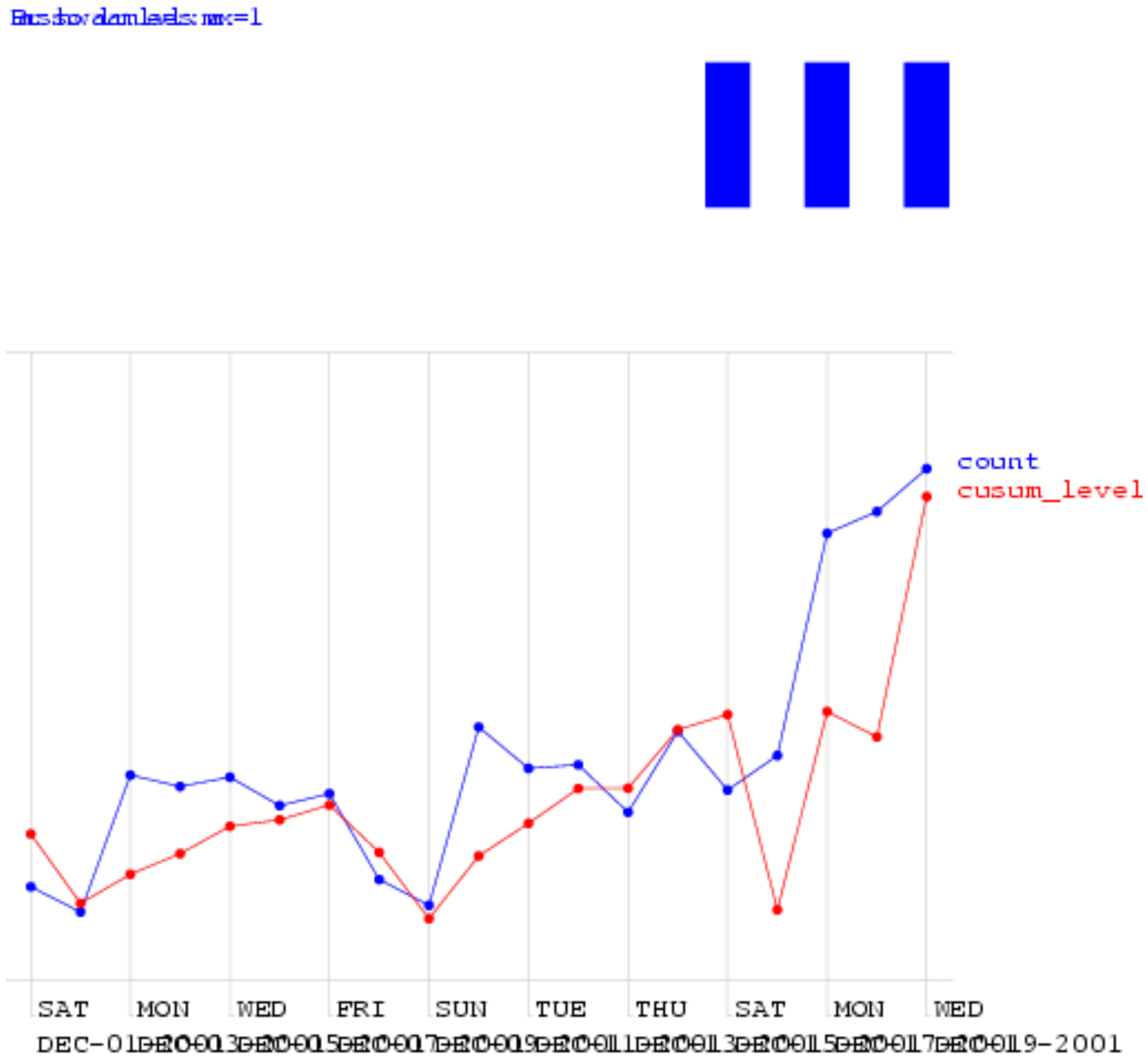
	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12
hours_of_daylight is_mon … is_tue	0.72	1.83	0.57	3.16
hours_of_daylight is_mon … is_sat	0.77	2.11	0.59	3.26

CUSUM

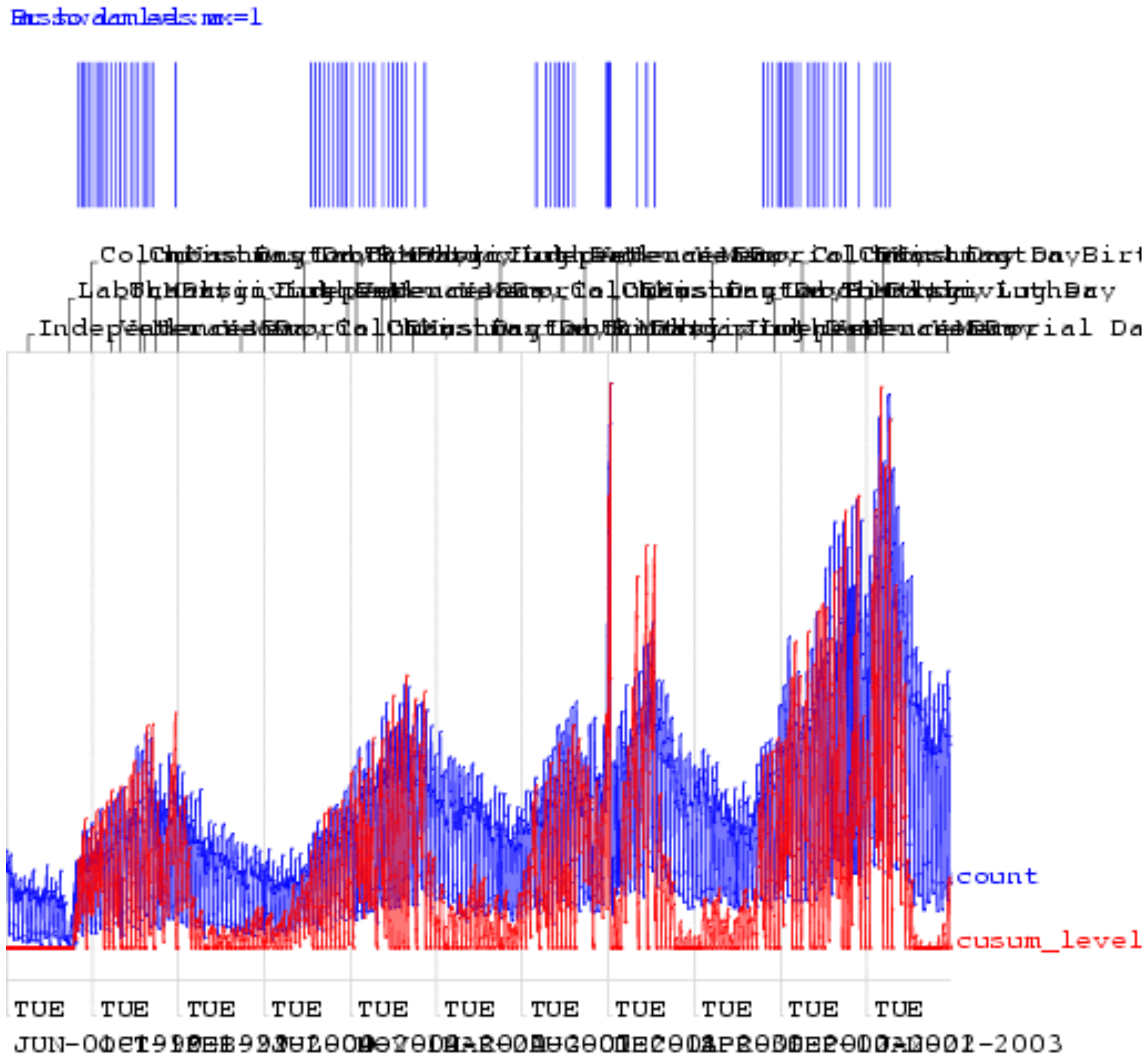
- CUmulative SUM Statistics

- Keep a running sum of “surprises”: a sum of excesses each day over the prediction
- When this sum exceeds threshold, signal alarm and reset sum

CUSUM



CUSUM

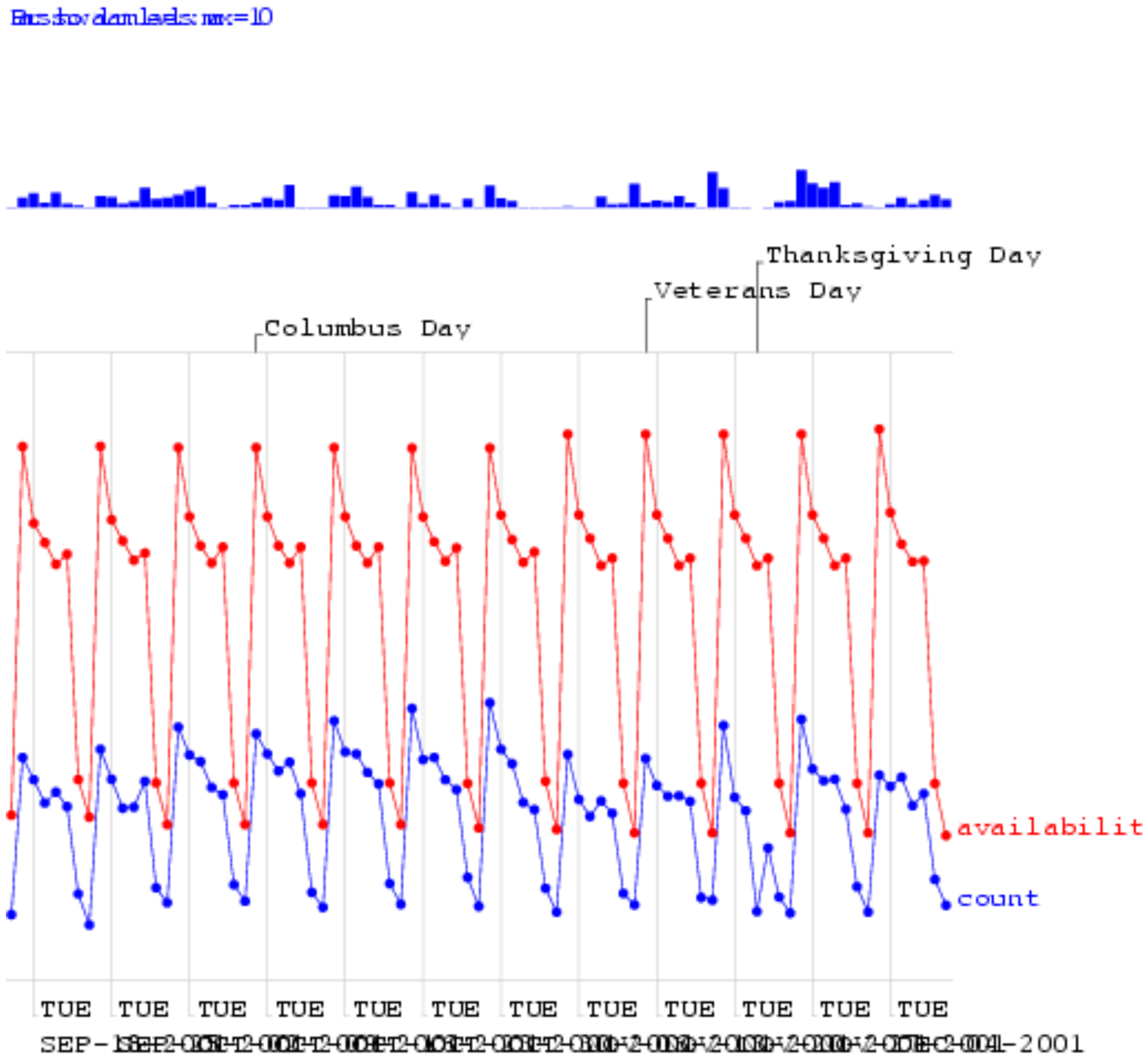


Algorithm Performance

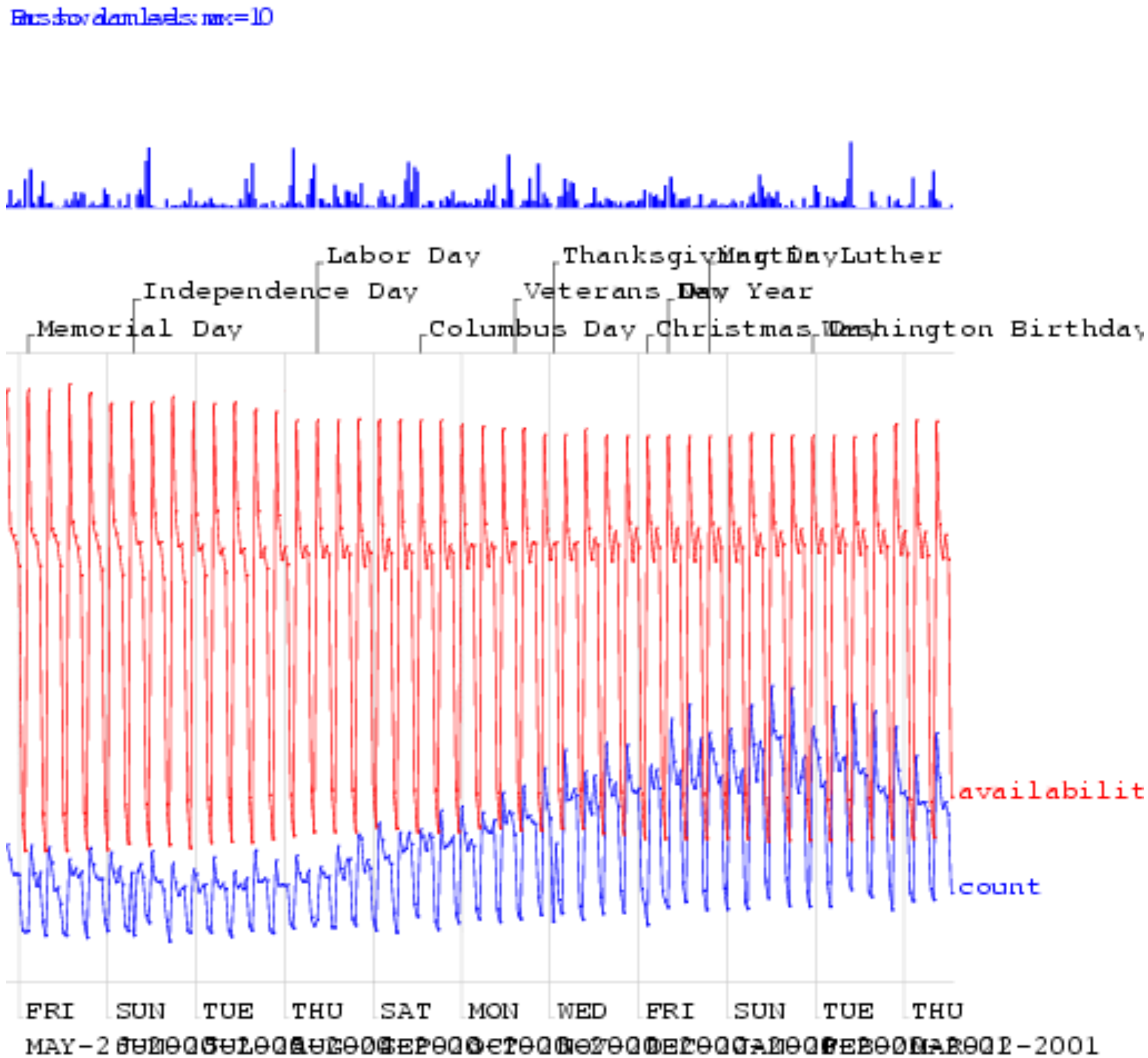
	Allowing one False Alarm per TWO weeks… Fraction of spikes detected	Allowing one False Alarm per TWO weeks… Days to detect a ramp outbreak	Allowing one False Alarm per SIX weeks… Fraction of spikes detected	Allowing one False Alarm per SIX weeks… Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12
hours_of_daylight is_mon ... is_tue	0.72	1.83	0.57	3.16
hours_of_daylight is_mon ... is_sat	0.77	2.11	0.59	3.26
CUSUM	0.45	2.03	0.15	3.55

The Sickness/Availability Model

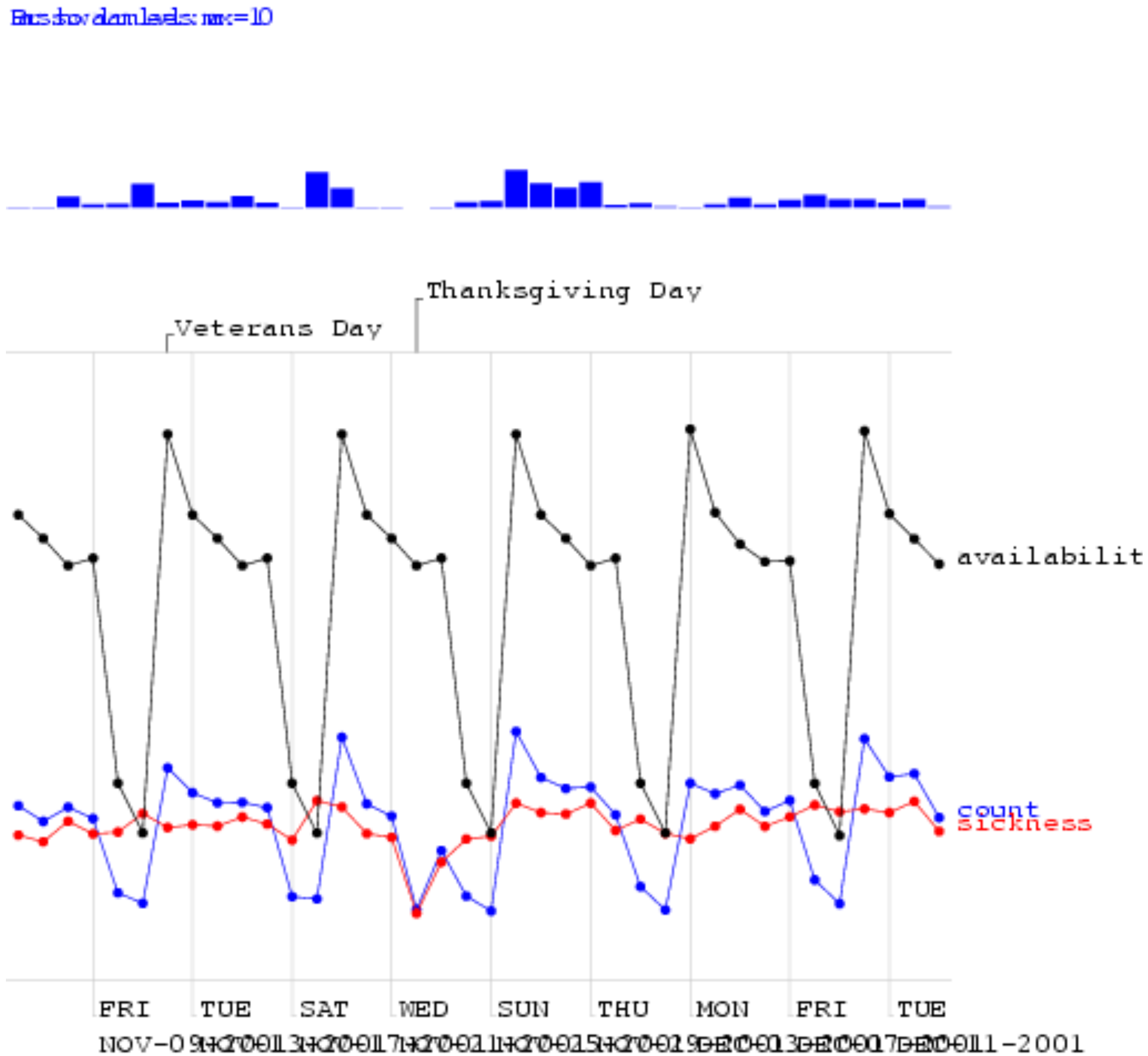
The Sickness/Availability Model



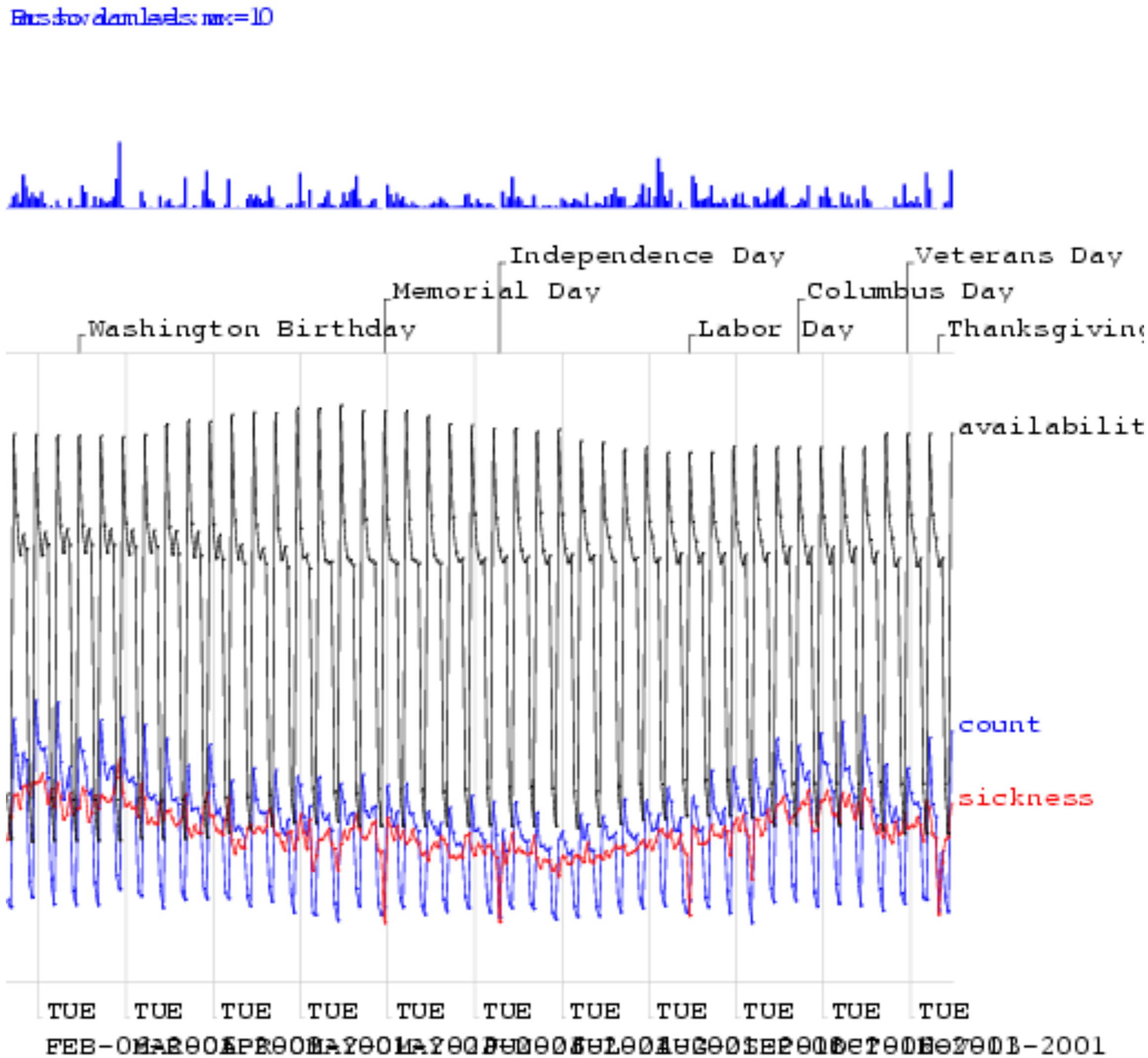
The Sickness/Availability Model



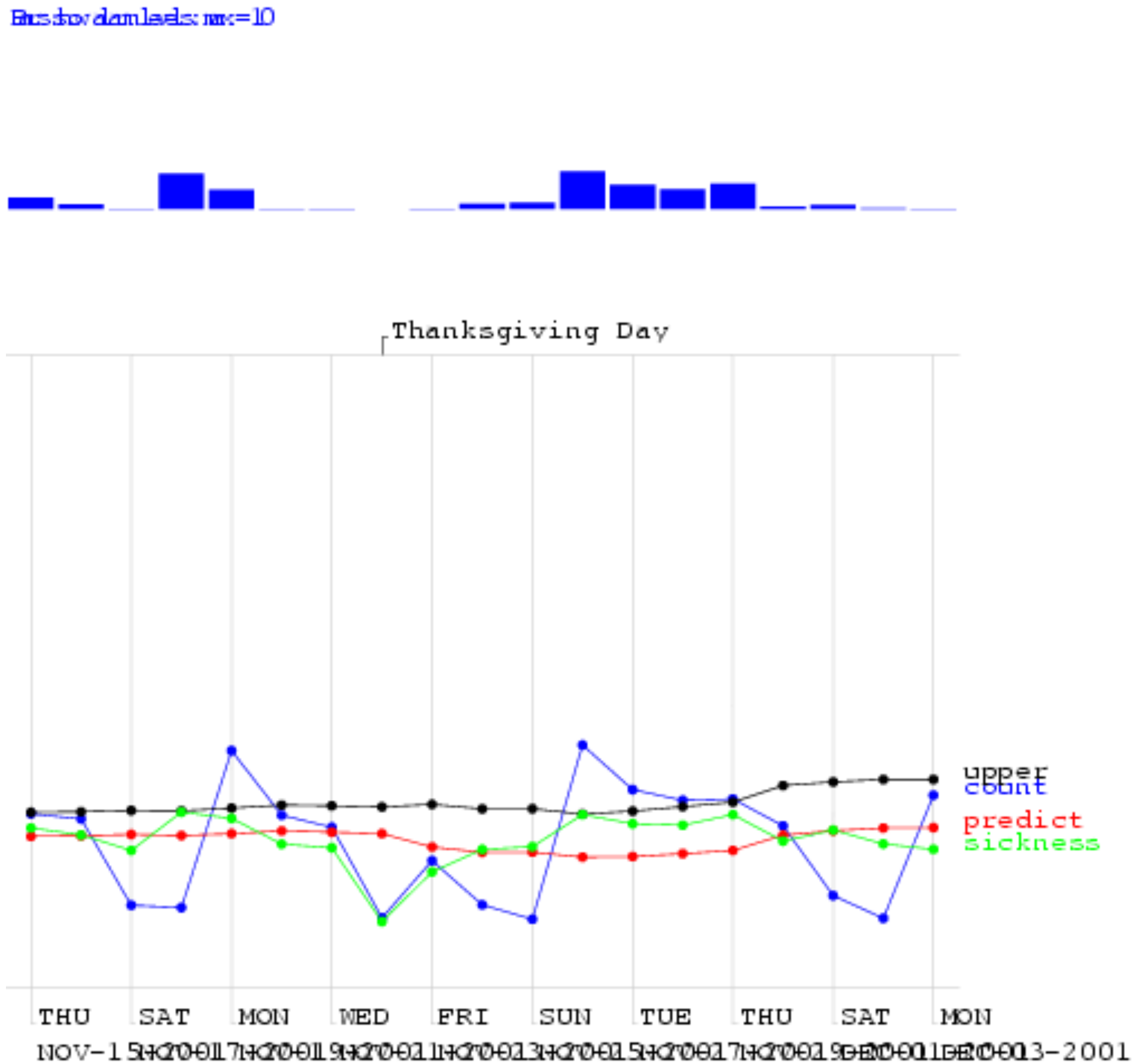
The Sickness/Availability Model



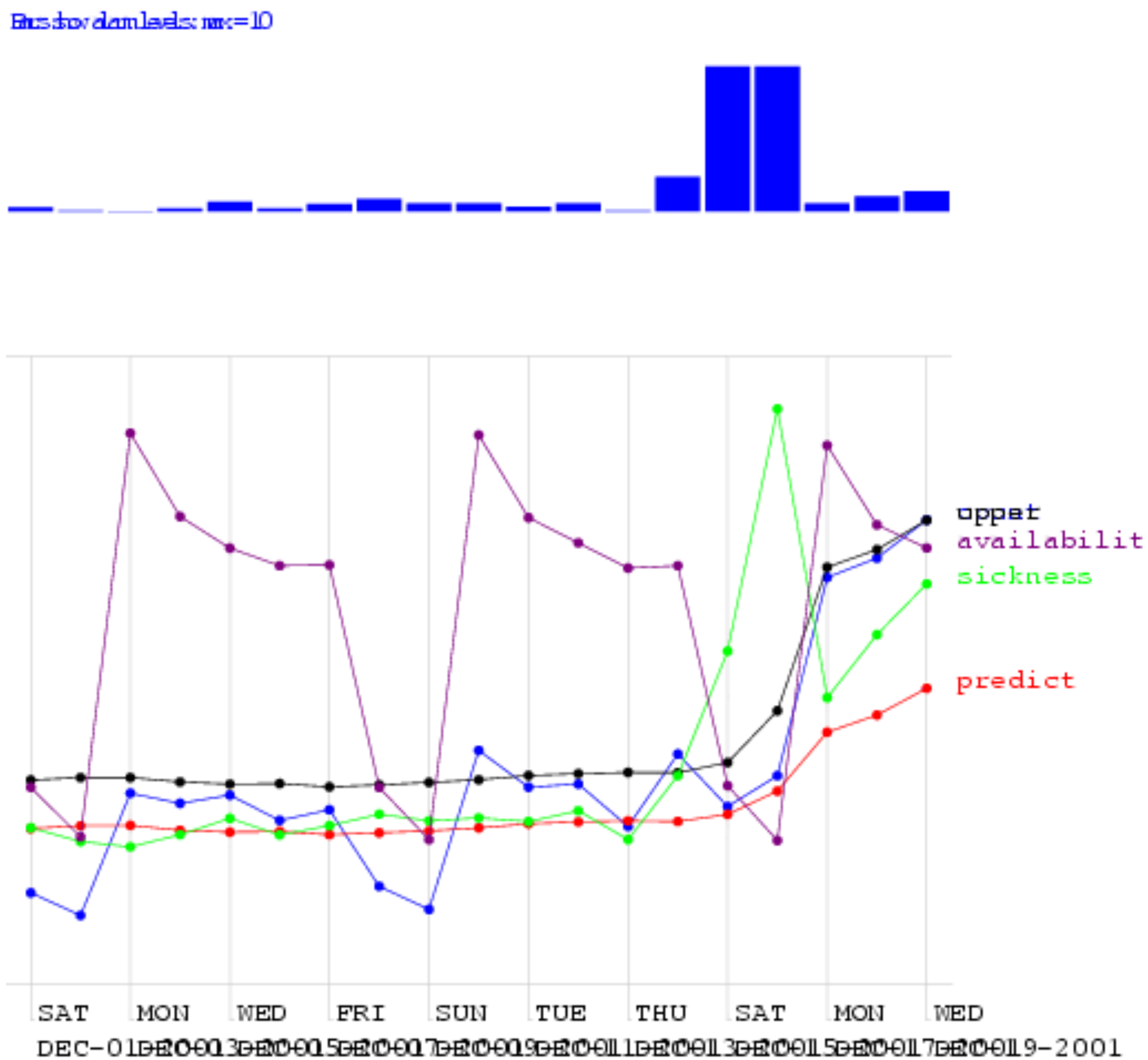
The Sickness/Availability Model



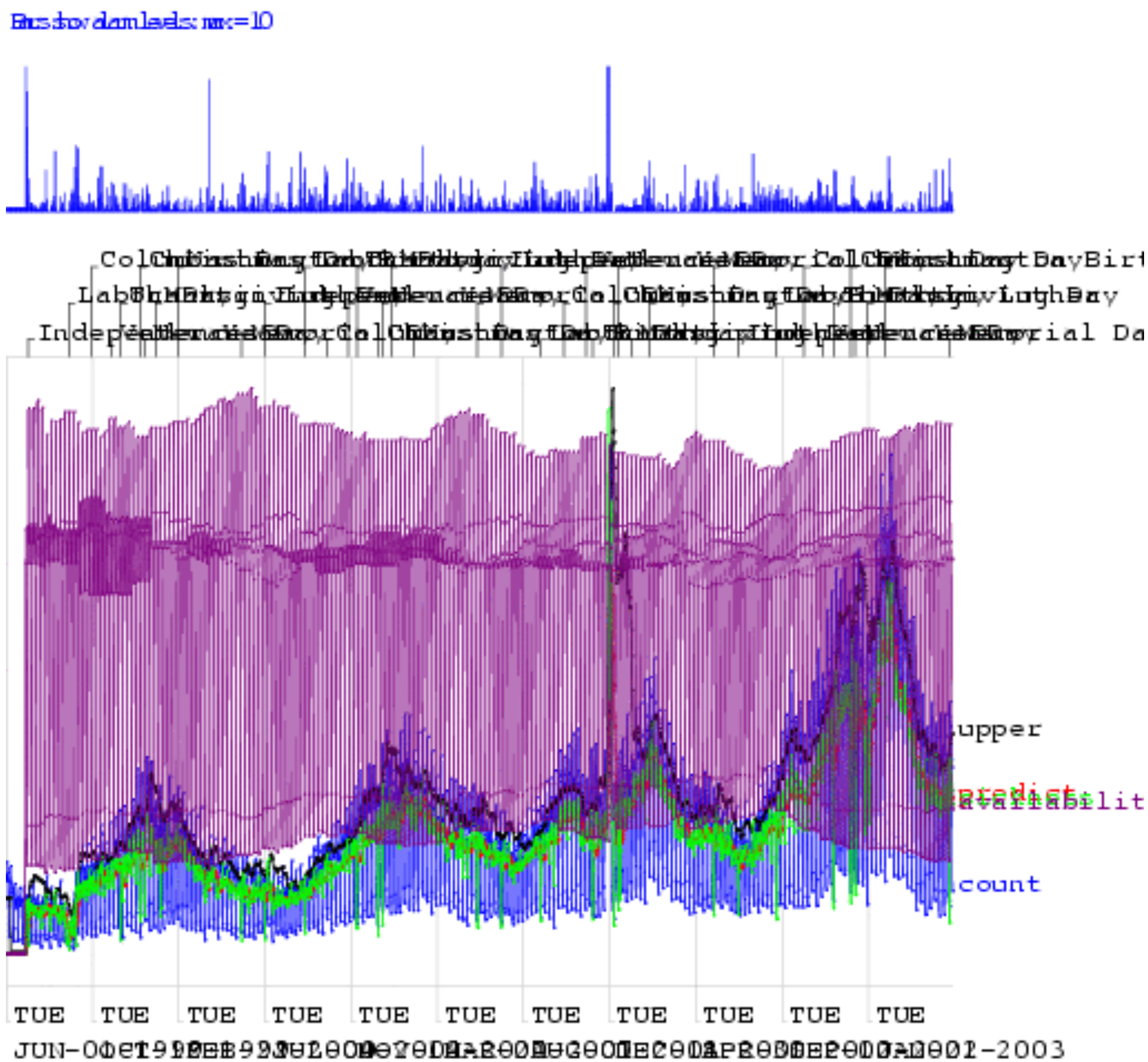
The Sickness/Availability Model



The Sickness/Availability Model



The Sickness/Availability Model



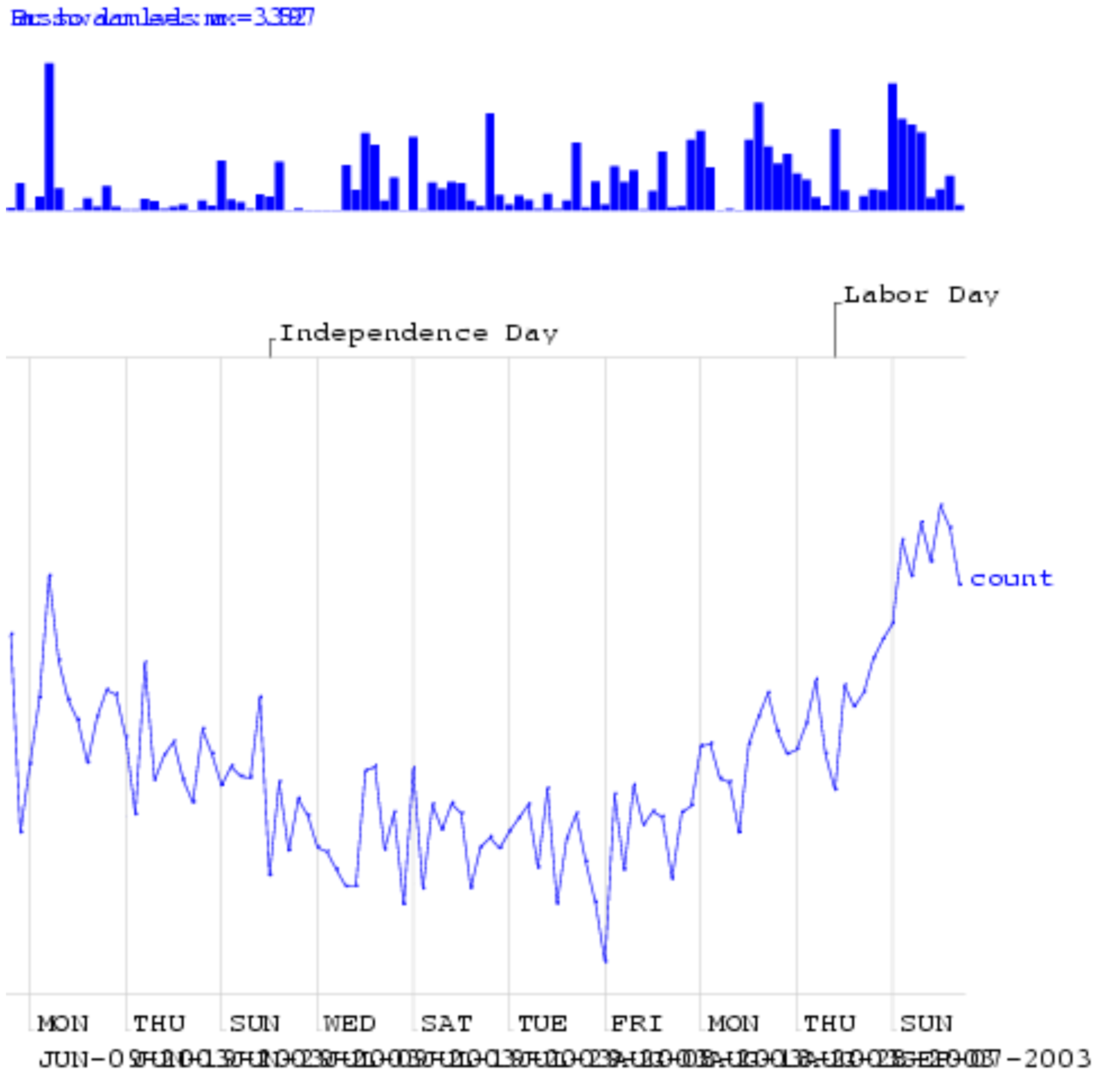
Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12
hours_of_daylight is_mon … is_tue	0.72	1.83	0.57	3.16
hours_of_daylight is_mon … is_sat	0.77	2.11	0.59	3.26
CUSUM	0.45	2.03	0.15	3.55
sa-mav-1	0.86	1.88	0.74	2.73
sa-mav-7	0.87	1.28	0.83	1.87
sa-mav-14	0.86	1.27	0.82	1.62

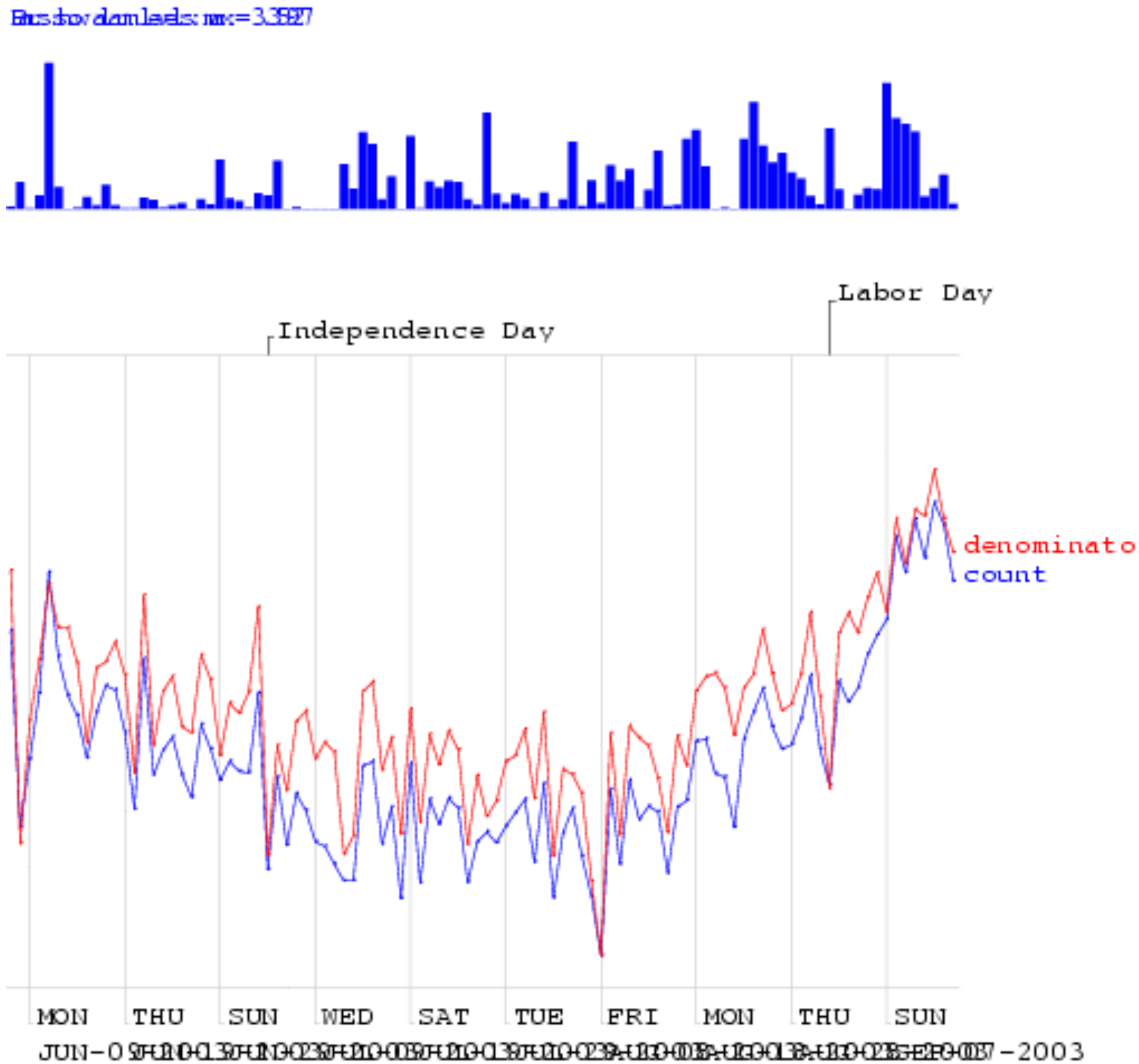
Algorithm Performance

	Allowing one False Alarm per TWO weeks…		Allowing one False Alarm per SIX weeks…	
	Fraction of spikes detected	Days to detect a ramp outbreak	Fraction of spikes detected	Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12
hours_of_daylight is_mon … is_tue	0.72	1.83	0.57	3.16
hours_of_daylight is_mon … is_sat	0.77	2.11	0.59	3.26
CUSUM	0.45	2.03	0.15	3.55
sa-mav-1	0.86	1.88	0.74	2.73
sa-mav-7	0.87	1.28	0.83	1.87
sa-mav-14	0.86	1.27	0.82	1.62
sa-regress	0.73	1.76	0.67	2.21

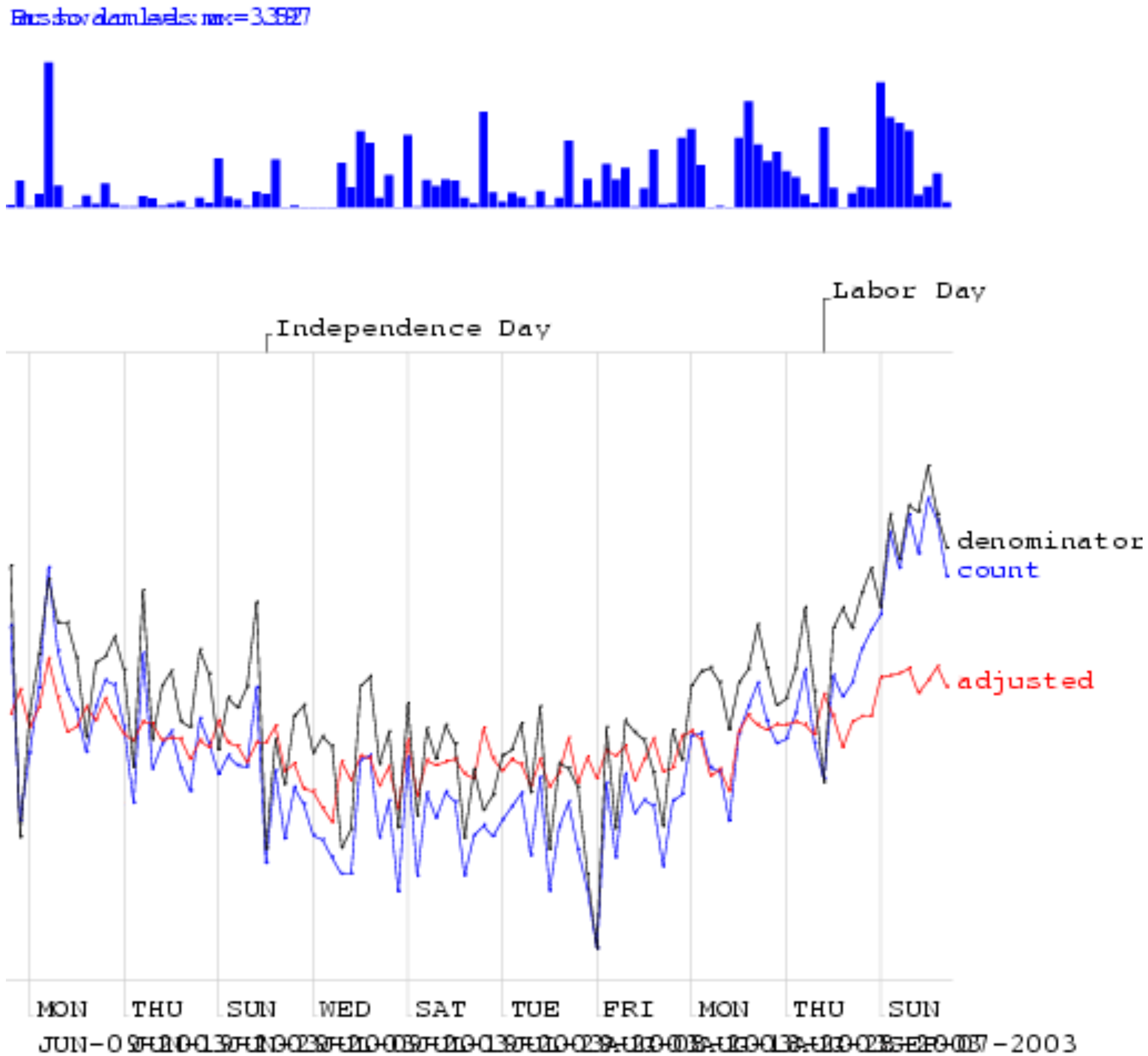
Exploiting Denominator Data



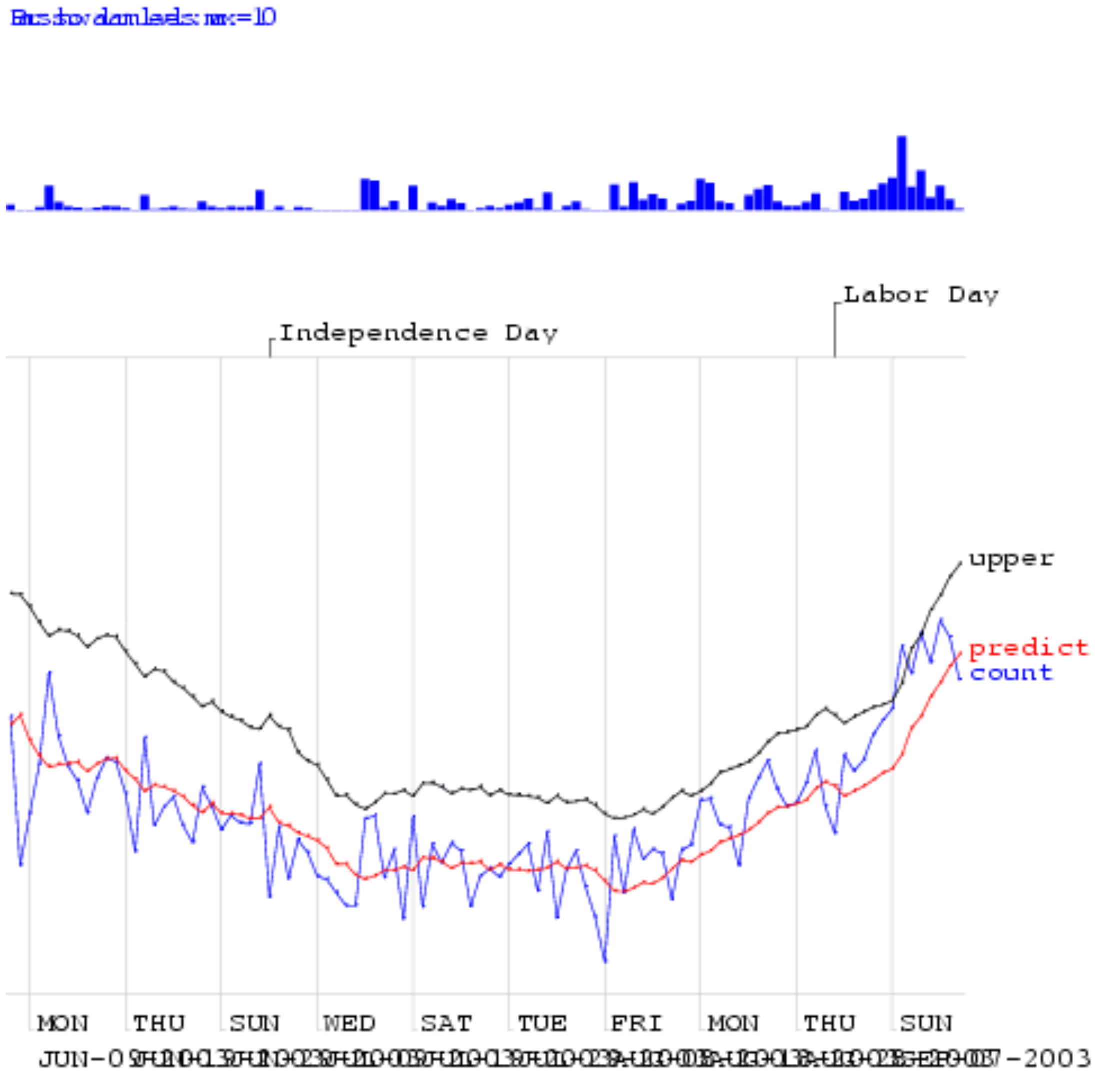
Exploiting Denominator Data



Exploiting Denominator Data



Exploiting Denominator Data



Algorithm Performance

	Allowing one False Alarm per TWO weeks… Fraction of spikes detected	Allowing one False Alarm per TWO weeks… Days to detect a ramp outbreak	Allowing one False Alarm per SIX weeks… Fraction of spikes detected	Allowing one False Alarm per SIX weeks… Days to detect a ramp outbreak
standard control chart	0.39	3.47	0.22	4.13
using yesterday	0.14	3.83	0.1	4.7
Moving Average 3	0.36	3.45	0.33	3.79
Moving Average 7	0.58	2.79	0.51	3.31
Moving Average 56	0.54	2.72	0.44	3.54
hours_of_daylight	0.58	2.73	0.43	3.9
hours_of_daylight is_mon	0.7	2.25	0.57	3.12
hours_of_daylight is_mon … is_tue	0.72	1.83	0.57	3.16
hours_of_daylight is_mon … is_sat	0.77	2.11	0.59	3.26
CUSUM	0.45	2.03	0.15	3.55
sa-mav-1	0.86	1.88	0.74	2.73
sa-mav-7	0.87	1.28	0.83	1.87
sa-mav-14	0.86	1.27	0.82	1.62
sa-regress	0.73	1.76	0.67	2.21
Cough with denominator	0.78	2.15	0.59	2.41
Cough with MA	0.65	2.78	0.57	3.24

Show Walkerton Results

Other state-of-the-art methods

- Wavelets
- Change-point detection
- Kalman filters
- Hidden Markov Models