A Simple, Versatile, Data-adaptive Approach for Alerting Based on Temporal Biosurveillance Data

Howard S. Burkom and Sean Patrick Murphy
The Johns Hopkins University Applied Physics Laboratory
National Security Technology Department

2nd Conference on Quantitative Methods on
Defense and National Security

George Mason University
Washington, DC February 7, 2007

Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

Chief Complaint Query

History of ESSENCE | Syndrome Definitions | Detector Algorithms | Data Dictionary | Help

HOME | Alert List | Query | Matrix Portal | Weekly % | Map Portal

Selections Affecting Time Series Analyzed: scale, variability, correlation, cyclic behavior

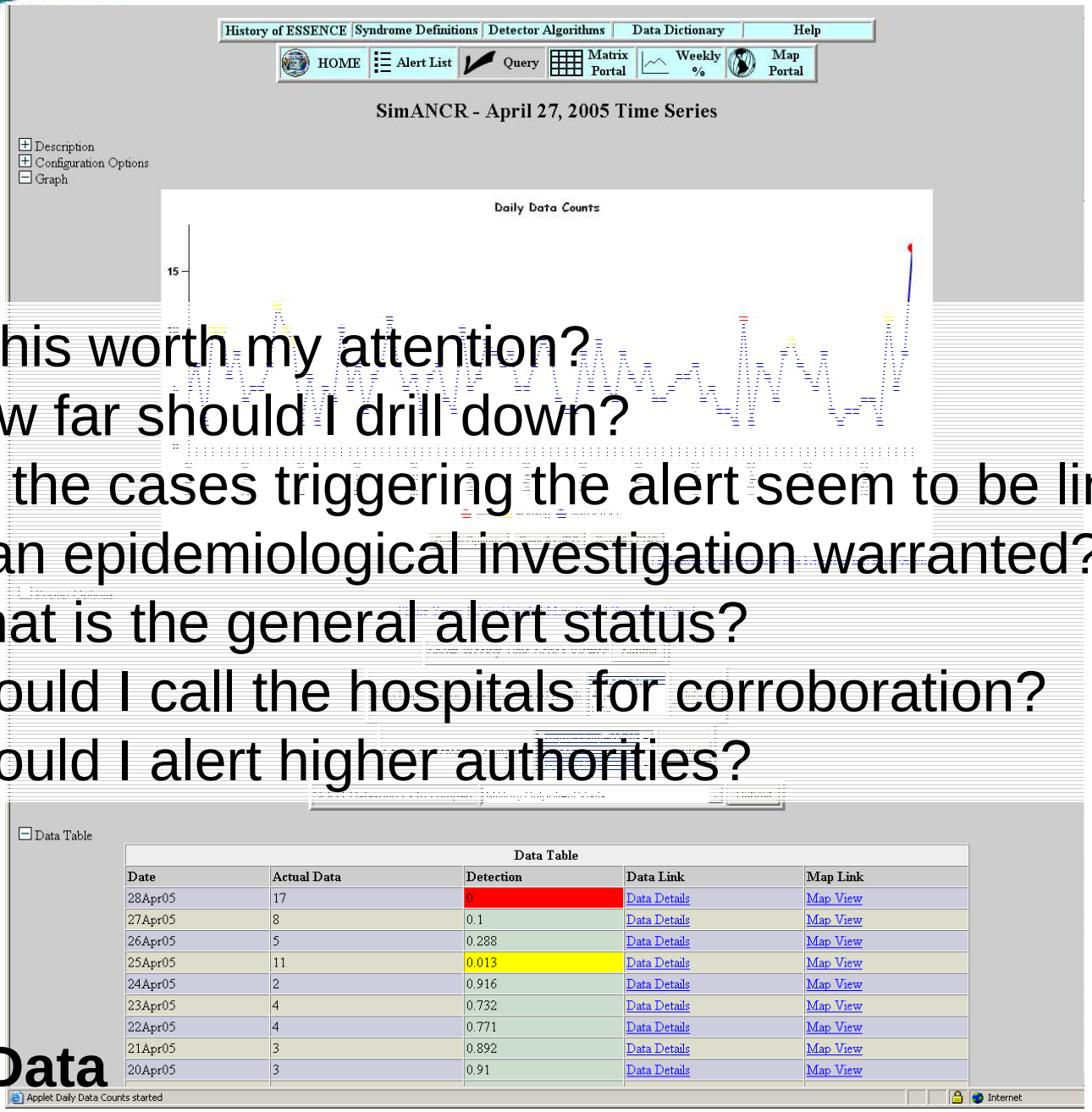
Data Source	Emergency Room Data by Patient Location	**Geography System**	Region
Region	All	**Medical Grouping System**	ChiefComplaints

Next Selections:

Select ChiefComplaints:	^fever^,and,^chills^ Use ^ for wildcards -- Use , for multiple entries Use and/or between entries to make complex queries Example: ^cough^,and,^fever^,or,^cold^ Query History	**Select Detector:**	Regression/EWMA
Select Age Range:	All Age Ranges Unknown 0-4	**Select Sex:**	All Sexs Unknown Male
Select Start Date:	28 Jan 05	**Select End Date:**	28 Apr 05
Submit			

Questions or Problems?

The Monitor's Routine Dilemma

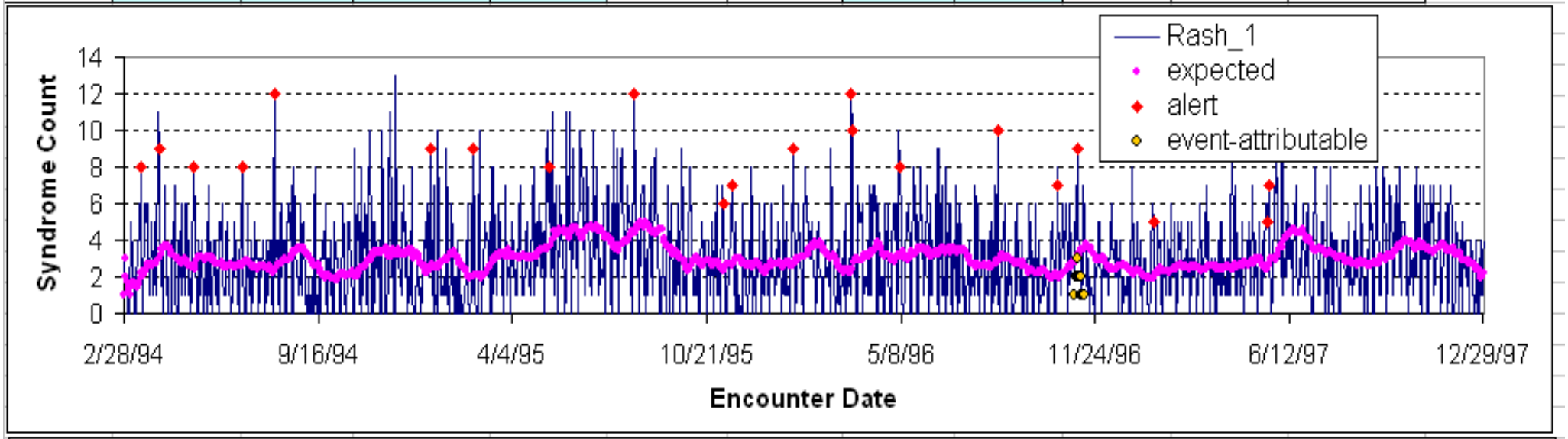


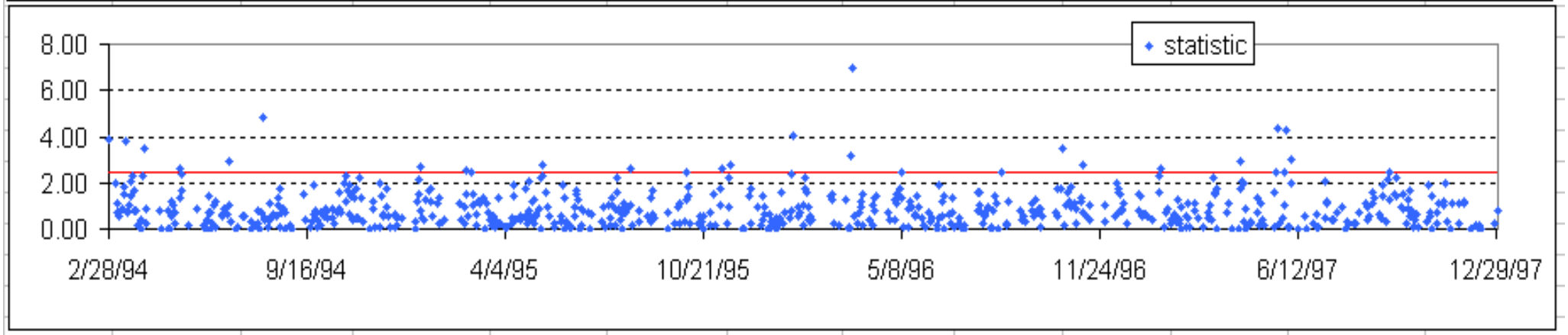
- Is this worth my attention?
- How far should I drill down?
- Do the cases triggering the alert seem to be linked?
- Is an epidemiological investigation warranted?
- What is the general alert status?
- Should I call the hospitals for corroboration?
- Should I alert higher authorities?

Simulated Data

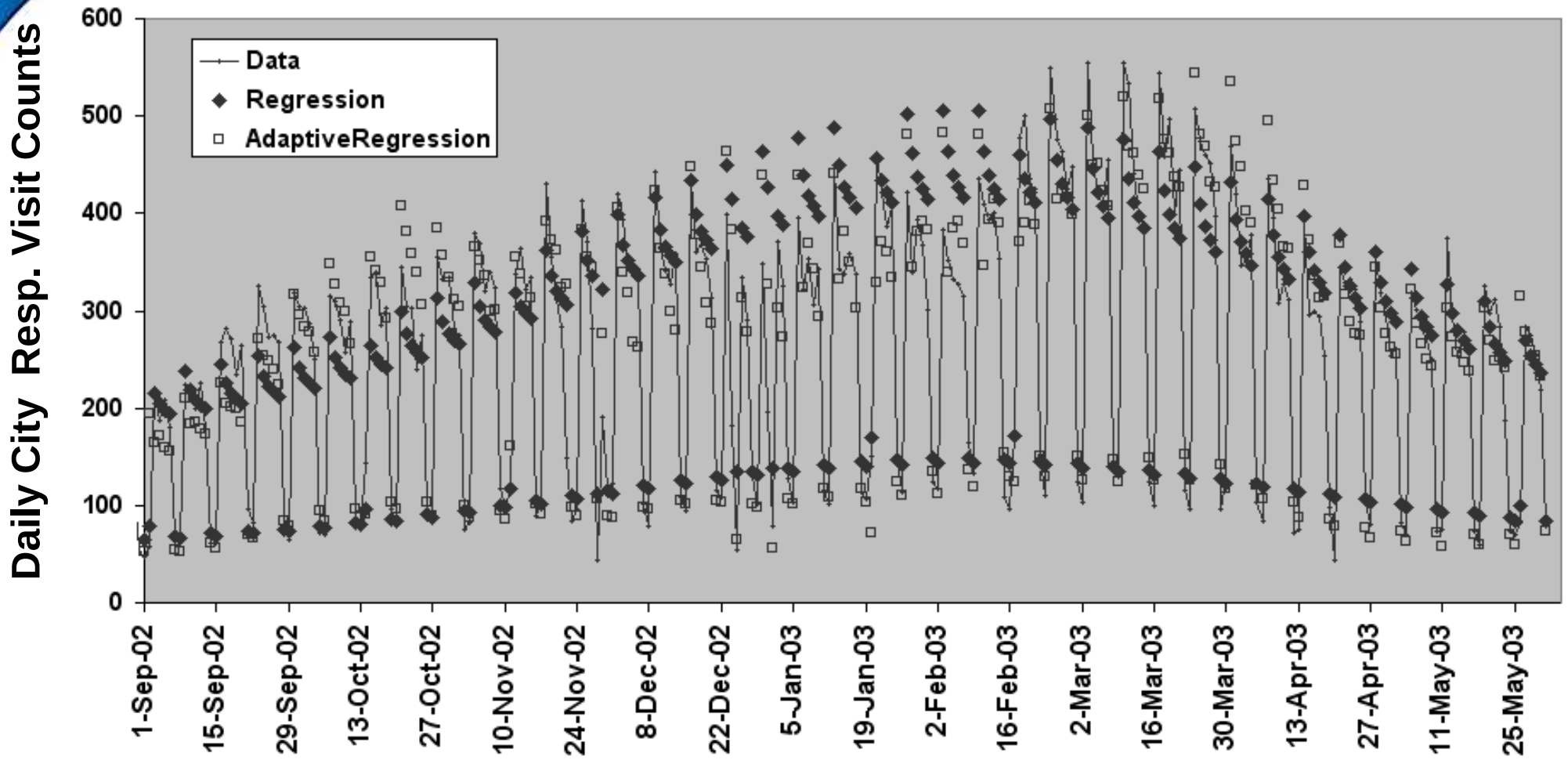
Monitoring for Statistical Anomaly

Number of Data Sets	Data Set Number	Event Start Date	Event Severity (multiplier)	Alerting Threshold	Data start date:	Data end date:	Plot start date:	Days plotted:	Total Alert Count	Alerts Excluding Event	Background recurrence (exp. days)
35	25	11/01/96	0.20	2.473	02/28/94	12/30/97	02/28/94	1402	25	24	57.83





Forecast Comparison: Nonadaptive & Adaptive Regression



Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

EWMA Monitoring

- Exponential Weighted Moving Average
- Average with most weight on recent X_k:

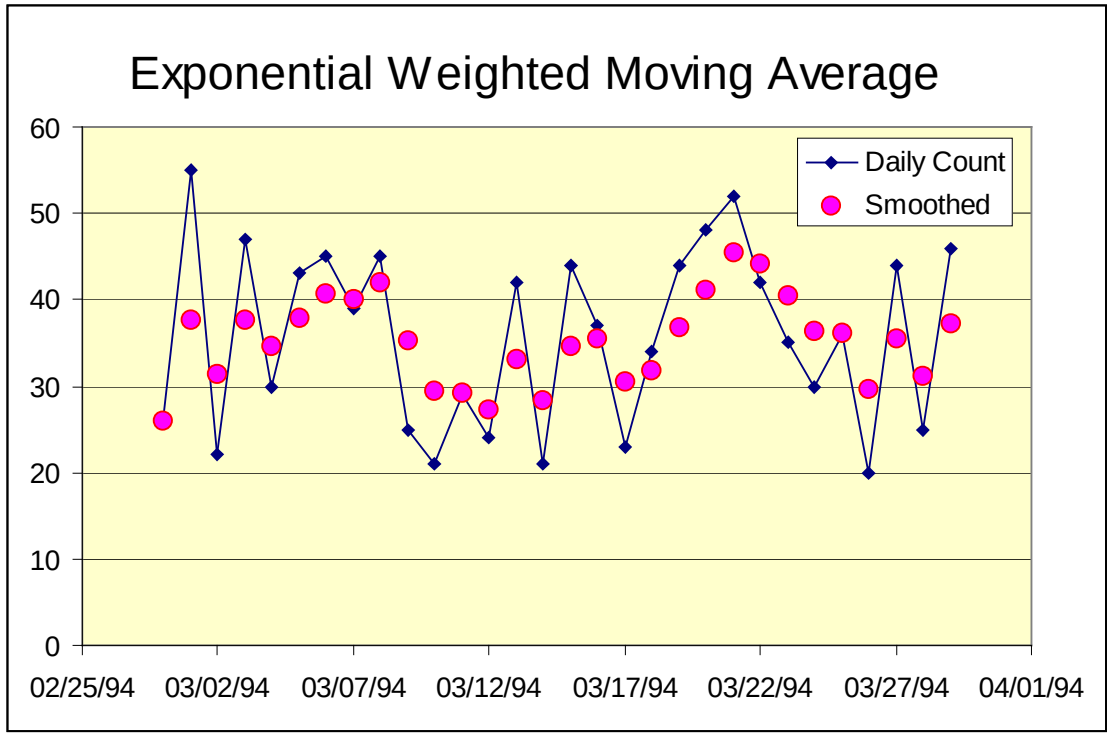
  $S_k = \omega S_{k-1} + (1-\omega)X_k$,

  where $0 < \omega < 1$
- Test statistic:

  S_k compared to expectation from sliding baseline

Basic idea: monitor

$(S_k - \mu_k) / \sigma_k$



- Added sensitivity for gradual events
- Larger ω means less smoothing

Generalized Exponential Smoothing

Holt-Winters Method: modeling level, trend, and seasonality

http://www.statistics.gov.uk/iosmethodology/downloads/Annex_B_The_Holt-Winters_forecasting_method.pdf

Forecast Function:

$$\hat{y}_{n+k|n} = (m_n + k\,b_n)\,(c_{n-s+k})$$

where: m_n = level at time n,
b_n = trend at time n,
c_n = periodic multiplier at time n
s = periodic interval
k = number of steps ahead

and m_n, b_n, c_n are updated by exponential smoothing

Holt-Winters Updating Equations

Updating Equations, multiplicative method:

Level at time t: $$m_t = \alpha \frac{y_t}{c_{t-s}} + (1-\alpha)(m_{t-1} + b_{t-1}), \quad 0<\alpha<1$$

Slope at time t: $$b_t = \beta(m_t - m_{t-1}) + (1-\beta) b_{t-1}, \quad 0<\beta<1$$

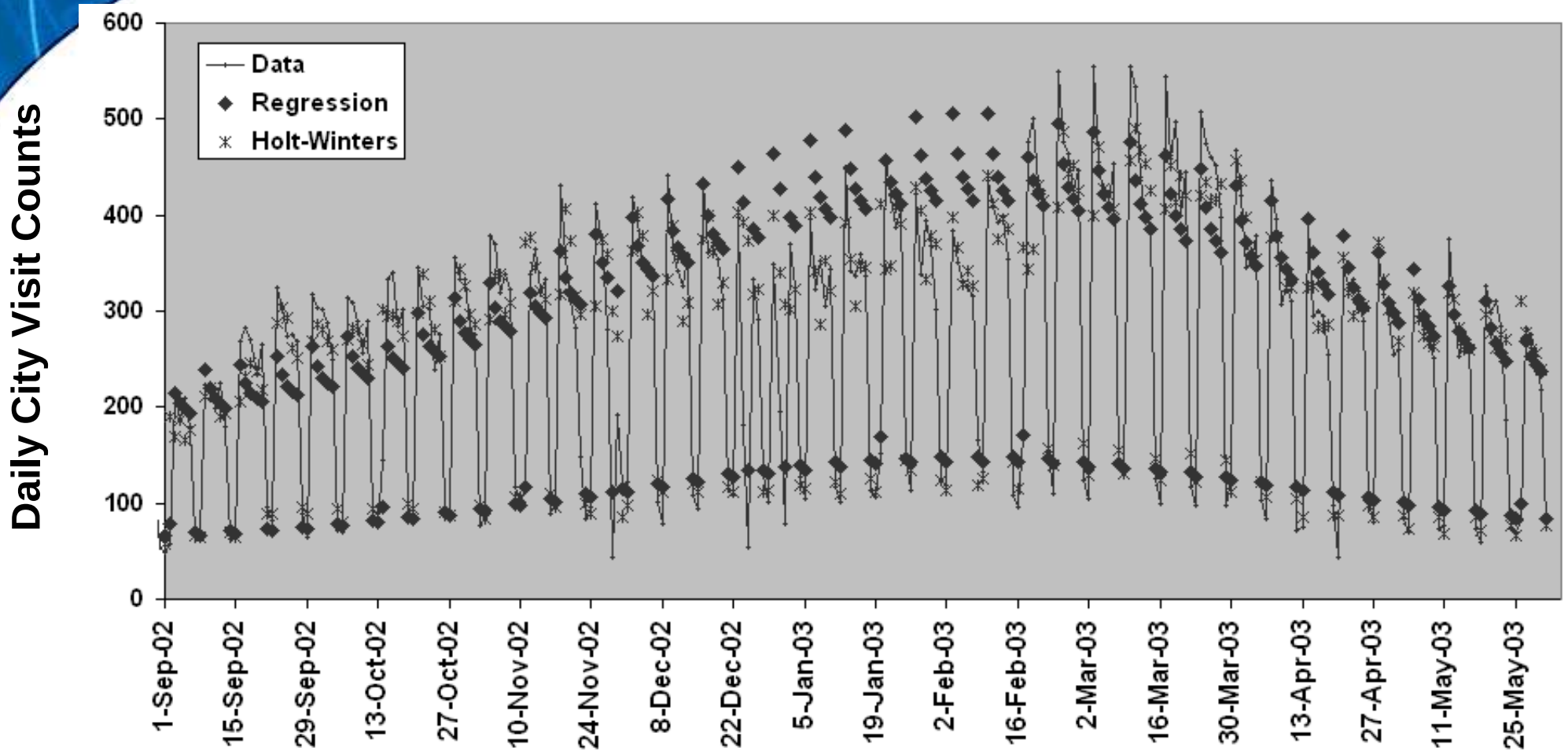
Periodic multiplier at time t: $$c_t = \gamma \frac{y_t}{m_t} + (1-\gamma) c_{t-s}, \quad 0<\gamma<1$$

And choice of initial values m_0, b_0, c_0,…c_{s-1} should be calculated from available data

Holt-Winters Adaptations

- Infer initial values, smoothing constants from (limited) sample data
- 7-day cyclic multipliers for day-of-week effects
- Forecasting on holidays: replace cyclic multiplier with holiday factor
- Suspend updating equations when they involve obvious outliers or holidays
- For sparse data
  - No exp. smoothing for linear, cyclic terms
  - Initial values for cyclic updating may be critical
  - Adjustments to avoid zero divisors

Forecast Comparison: Nonadaptive Regression & Holt-Winters Smoothing



Burkom, H., Murphy, S.P., and Shmueli G, Automated Time Series Forecasting for Biosurveillance, accepted for 2007 publication in Statistics in Medicine.

Forecasting Local Linearity: Automatic vs Nonautomatic Methods

Chatfield, C. (1978), "The Holt-Winters Forecasting Procedure," Applied Statistics, 27, 264-279.

Chatfield, C.and Yar, M. (1988), "Holt-Winters Forecasting: Some Practical Issues, " The Statistician, 37, 129-140.

- "Modern thinking favors local linearity rather than global linear regression in time…"
- "Local linearity is also implicit in ARIMA modeling…"
  - **Simple EWMA ~ ARIMA(0,1,1)**
  - **EWMA + trend ~ ARIMA(0,2,2)**
  - **Multiplicative Holt-Winters has no ARIMA equivalent**
- "Practical considerations rule out [*Box-Jenkins*] if there are insufficient observations or …expertise available"
  - "Box-Jenkins… requires the user to identify an appropriate… [*ARIMA*] model"

For "fair" comparison of H-W to B-J, have both automatic or nonautomatic.

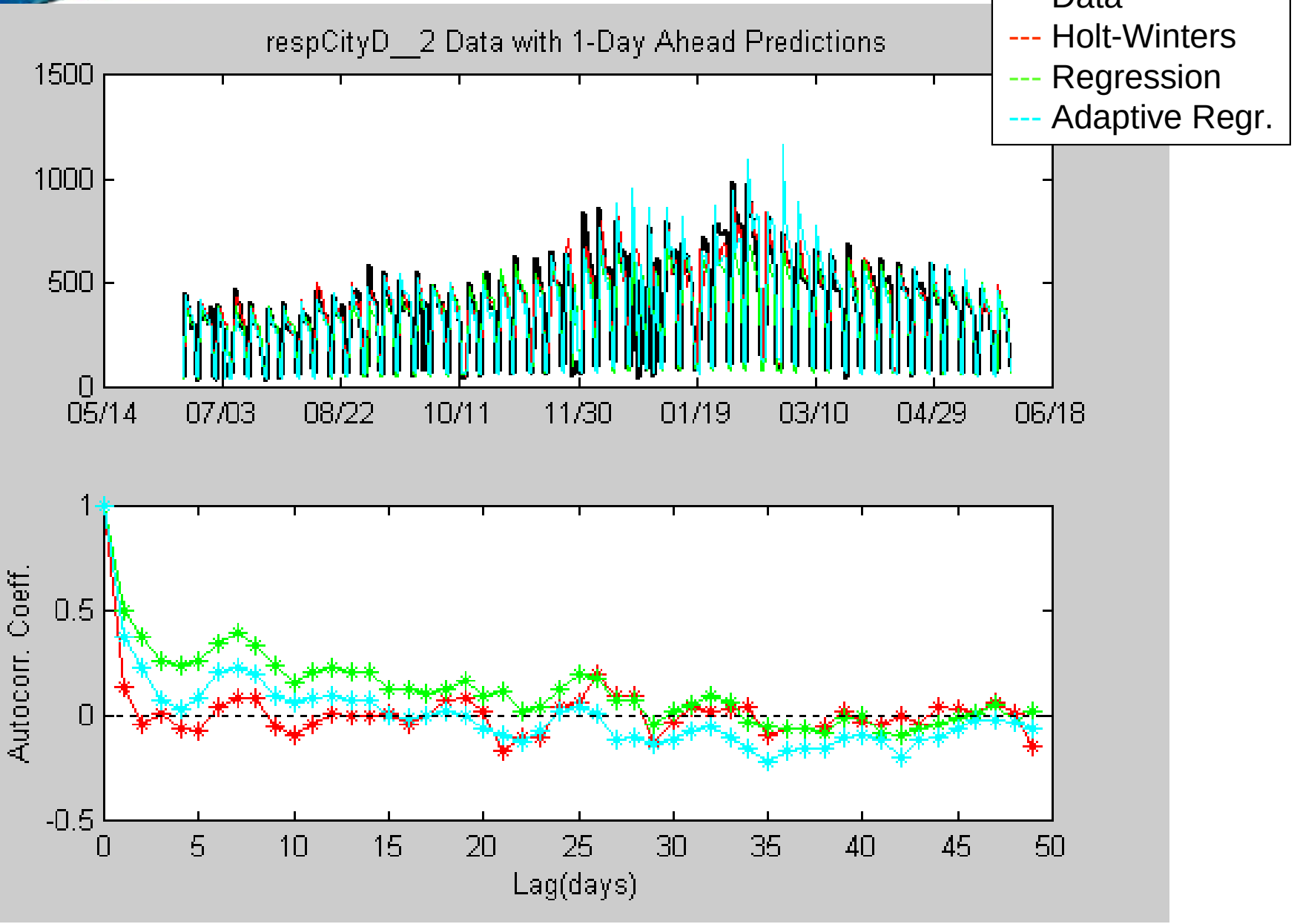
Assertion: The simplicity of H-W permits easier classification, requiring less historic data.

Can an automatic B-J give robust forecasting over a range of input series types?

Median Residual Comparison

		1-day ahead predictions							
		1-day lag				7-day lag			
		Holt-Winters	Regression	Adaptive Regression	HW-Adap	Holt-Winters	Regression	Adaptive Regression	HW-Adap
Resp Data DOW + Seasonal Effects	1	17.54	22.62	19.99	-2.45	21.13	22.62	20.40	0.72
	2	26.67	31.69	29.30	-2.63	28.90	31.69	36.88	-7.98
	3	21.94	83.13	25.77	-3.83	25.99	83.13	28.10	-2.11
	4	13.31	13.25	16.34	-3.03	15.30	13.25	18.03	-2.73
	5	26.28	27.27	27.70	-1.43	28.60	27.27	35.75	-7.16
	6	19.19	35.55	22.90	-3.71	22.51	35.55	23.47	-0.96
	7	24.66	43.88	27.97	-3.32	24.73	43.88	32.01	-7.28
	8	30.59	53.71	47.60	-17.01	39.18	53.71	60.02	-20.84
	9	50.69	76.67	68.14	-17.46	63.68	76.67	81.76	-18.07
	10	33.56	131.06	39.01	-5.45	35.28	131.06	49.32	-14.04
GI Data DOW Effects Only	1	8.66	13.48	9.21	-0.56	8.90	13.48	9.42	-0.52
	2	15.73	35.63	15.89	-0.16	16.71	35.63	16.94	-0.23
	3	8.03	24.27	9.14	-1.12	8.91	24.27	10.53	-1.63
	4	24.62	38.88	25.25	-0.63	23.44	38.88	25.03	-1.59
	5	10.19	24.09	11.63	-1.44	10.67	24.09	13.35	-2.68
	6	12.23	21.14	13.45	-1.23	12.09	21.14	13.31	-1.22

Residual Autocorrelation Comparison



Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

Control Charts Based on Holt-Winters Residuals, I

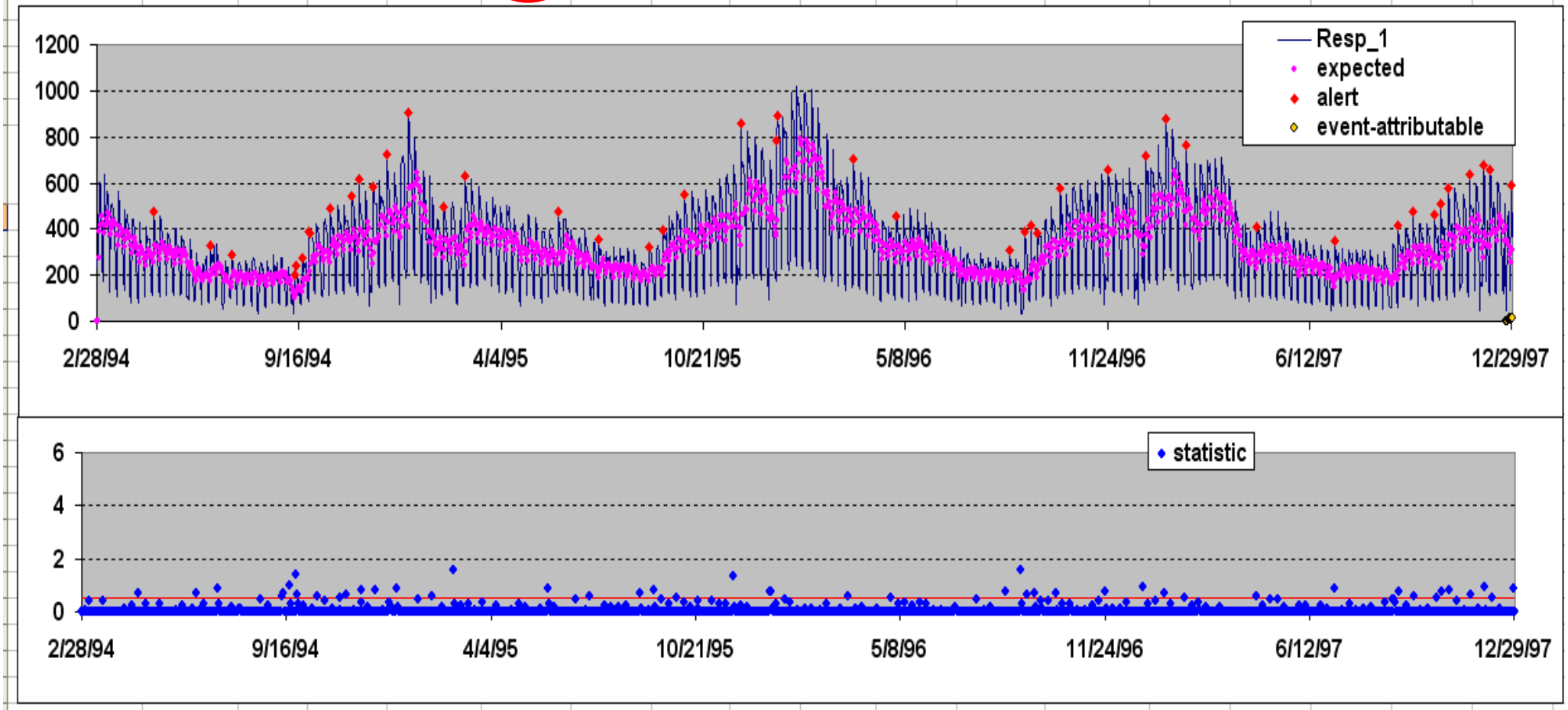
- Select smoothing parameters and initial values based on data sample
- Make data forecasts $pred_t$
- Apply anomaly measure (control chart)
  - For sudden-onset signals, use Xbar-like z-score
  - Test statistic = $(x_t - pred_t) / s^*_t$
  - Normalize with estimate s^*_t of residual standard deviation
  - For clinic-visit-count data, account for residual day-of-week effects in s^*_t

Control Charts Based on Holt-Winters Residuals, II

- For detection of gradual signals:
  - Temporal scan statistic
  - CUSUM (EWMA)
- For sparse data
  - No exp. smoothing for linear, cyclic terms
  - Initial values for cyclic updating may be critical, yielding an EWMA modified by known cyclic effects

Adaptive Xbar Chart ignoring cyclic pattern

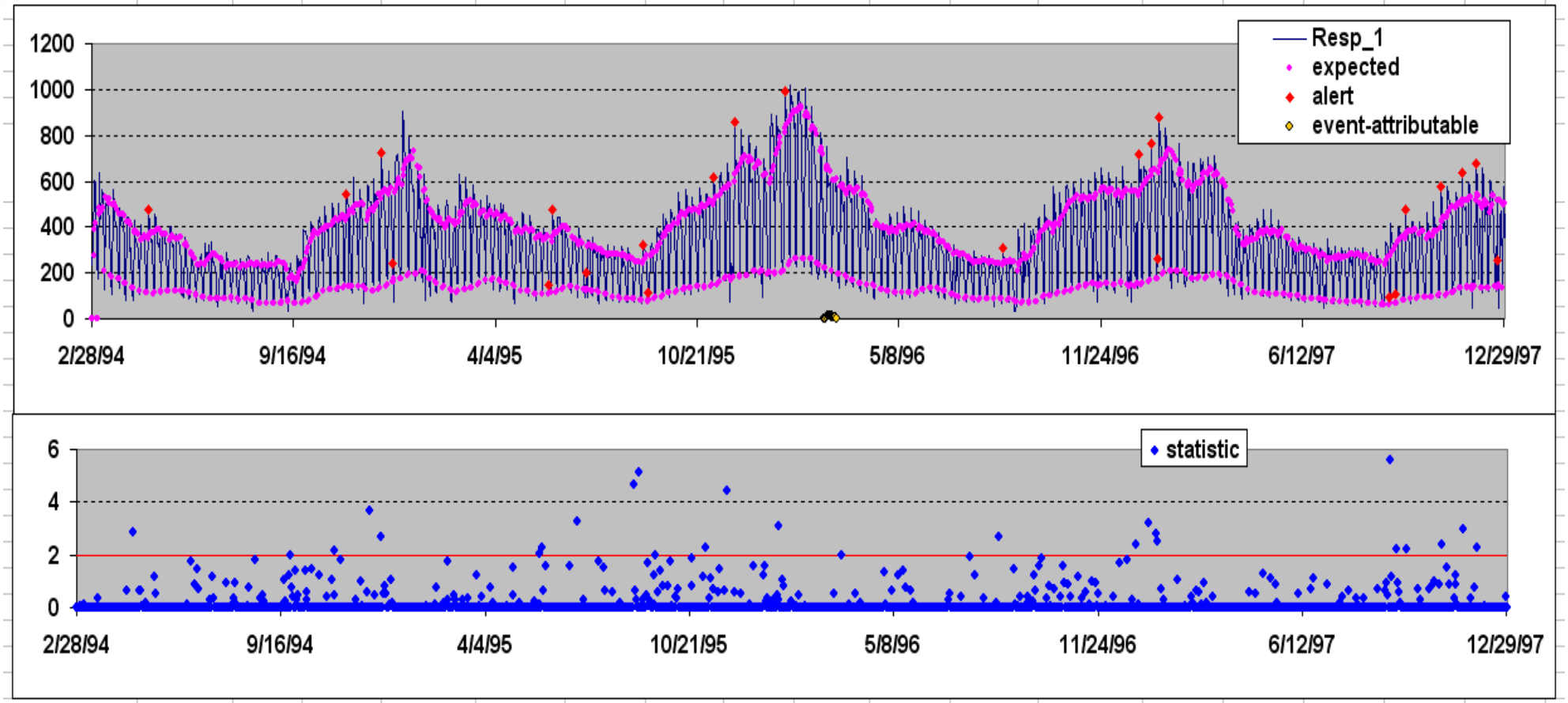
Data Set	Number of Data Sets	Data Set Number	Event Start Date	Event Severity (multiplier)	Alerting Threshold	Data start date:	Data end date:	Plot start date:	Days plotted:	Total Alert Count	Alerts Excluding Event	Background recurrence (exp. days)
SyndromicData	35	28	12/24/97	1.00	0.50	02/28/94	12/30/97	02/28/94	1402	46	45	30.84



Mean Abs. Error	Median Abs. Error	Mean Error Ratio	Median Error Ratio
133.58	118.89	0.69	0.32

Adaptive Xbar Chart
Stratified by Weekday / Non-weekday

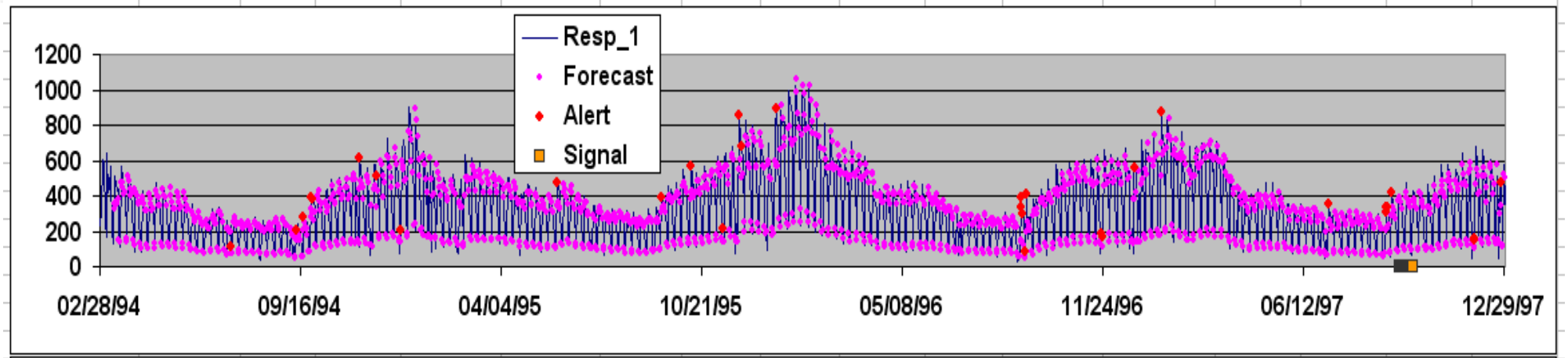
Data Set	Number of Data Sets	Data Set Number	Event Start Date	Event Severity (multiplier)	Alerting Threshold	Data start date:	Data end date:	Plot start date:	Days plotted:	Total Alert Count	Alerts Excluding Event	Background recurrence (exp. days)
SyndromicData	35	28	02/24/96	1.00	2.00	02/28/94	12/30/97	02/28/94	1402	24	24	57.83

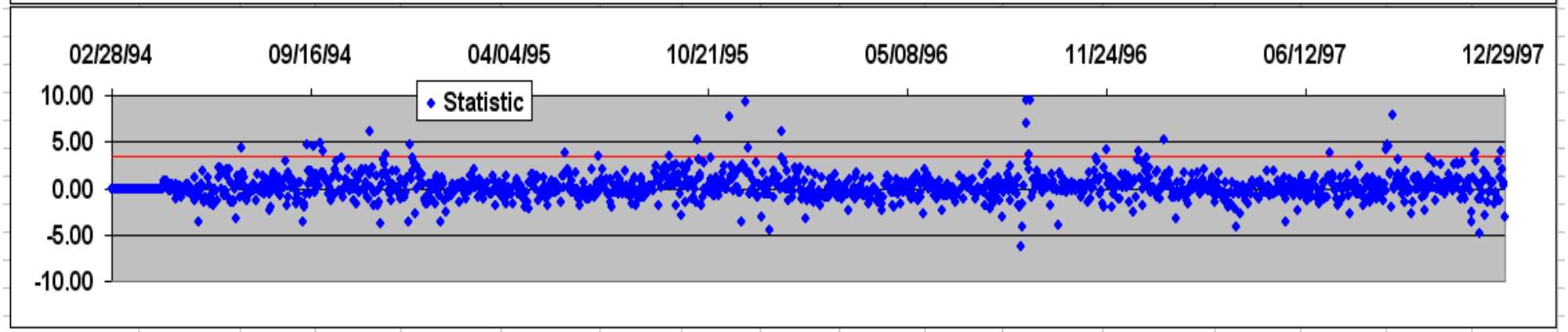


Mean Abs. Error	Median Abs. Error	Mean Error Ratio	Median Error Ratio
45.63	33.00	0.17	0.12

Adaptive Xbar Chart based on Holt-Winters Residuals

Dataset	Number of Data Sets	Data Set Number	Event Start Date	Event Severity (multiplier)	Alerting Threshold	Data start date:	Data end date:	Plot start date:	Days plotted:	Total Alert Count	Alerts Excluding Event	Est. Background Recurrence (days)
syndromicdata	35	28	09/16/97	0.00	3.50	02/28/94	12/30/97	02/28/94	1402	32	32	43.38





Mean Abs. Error	Median Abs. Error	Mean Error Ratio	Median Error Ratio
41.32	25.73	0.140	0.095

Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

Control Chart Comparison: Detecting Stochastic signals in authentic data

- 33 time series of syndromic visit counts at various scales (available for replication)
- Application of 3 control charts to pure counts and H-W forecast residuals
  - Adaptive CUSUM
  - Adaptive Xbar
  - Temporal scan statistic
- Two lognormal signal distributions:
  - rapid-onset (2-3 day mean incubation)
  - slow-rise (10-11 day mean incubation)
  - sampled stochastically for number of cases appropriate to scale of data
- 600 runs with each combination of series, signal type

Performance Comparison: Rapid-Onset Signals

Average Detection Probabilities with Authentic Data and Attributes for Injected Spike Signals									
Data Attributes				HW-forecast methods			Pure Count Methods		
name	Mean	AC1	AC7	CUSUM	Scan	Xbar	CUSUM	Scan	Xbar
Hemr_ill_1	0.60	0.16	0.17	0.914	0.949	0.947	0.912	0.907	0.941
Gi_3	0.60	0.05	0.03	0.654	0.573	0.781	0.630	0.559	0.766
Lesion_1	1.40	0.28	0.30	0.965	0.836	0.964	0.964	0.951	0.962
Neuro_1	2.02	0.36	0.30	0.550	0.405	0.708	0.562	0.391	0.674
Rash_1	3.00	0.17	0.11	0.328	0.341	0.520	0.257	0.299	0.374
Bot_Like_1	3.15	0.26	0.24	0.260	0.191	0.367	0.234	0.227	0.309
Shk_Coma_2	5.38	0.26	0.19	0.298	0.262	0.521	0.272	0.230	0.379
Lymph_1	6.78	0.24	0.21	0.446	0.568	0.668	0.315	0.272	0.455
UGI_2	7.19	0.34	0.29	0.425	0.173	0.612	0.319	0.357	0.525
Bot_Like_2	13.51	0.40	0.38	0.440	0.648	0.805	0.326	0.378	0.528
Hemr_ill_2	14.20	0.32	0.28	0.334	0.452	0.663	0.296	0.316	0.498
UGI_1	14.94	0.37	0.33	0.369	0.516	0.684	0.306	0.262	0.490
Lesion_2	22.43	0.27	0.23	0.304	0.505	0.735	0.230	0.294	0.517
Neuro_2	35.36	0.37	0.32	0.706	0.769	0.870	0.366	0.443	0.606
LGI_1	38.08	0.42	0.39	0.464	0.634	0.805	0.329	0.300	0.527
Gi_1	53.02	0.42	0.40	0.458	0.695	0.823	0.264	0.300	0.540
LGI_2	54.92	0.36	0.33	0.530	0.773	0.904	0.368	0.424	0.654
Gi_2	62.11	0.36	0.33	0.588	0.801	0.912	0.383	0.437	0.668
Rash_2	77.91	0.31	0.27	0.641	0.796	0.908	0.384	0.447	0.648
Fever_1	78.98	0.51	0.45	0.416	0.529	0.691	0.255	0.204	0.445
Resp_2	161.99	0.35	0.32	0.686	0.835	0.928	0.414	0.376	0.648
Resp_1	334.61	0.52	0.50	0.700	0.823	0.899	0.359	0.237	0.661

Performance Comparison: Gradual Signals

Average Detection Probabilities with Authentic Data and Attributes for Injected Gradual Signals									
Data Attributes				HW-forecast methods			Pure Count Methods		
name	Mean	AC1	AC7	CUSUM	Scan	Xbar	CUSUM	Scan	Xbar
Hemr_ill_1	0.60	0.16	0.17	0.821	0.717	0.787	0.804	0.702	0.780
Gi_3	0.60	0.05	0.03	0.682	0.614	0.658	0.665	0.586	0.646
Lesion_1	1.40	0.28	0.30	0.937	0.760	0.893	0.933	0.847	0.890
Neuro_1	2.02	0.36	0.30	0.833	0.656	0.803	0.821	0.650	0.793
Rash_1	3.00	0.17	0.11	0.740	0.844	0.799	0.686	0.791	0.711
Bot_Like_1	3.15	0.26	0.24	0.618	0.507	0.649	0.652	0.631	0.654
Shk_Coma_2	5.38	0.26	0.19	0.637	0.635	0.710	0.588	0.564	0.622
Lymph_1	6.78	0.24	0.21	0.765	0.945	0.781	0.706	0.697	0.745
UGI_2	7.19	0.34	0.29	0.772	0.678	0.755	0.712	0.810	0.762
Bot_Like_2	13.51	0.40	0.38	0.897	0.975	0.879	0.745	0.824	0.771
Hemr_ill_2	14.20	0.32	0.28	0.783	0.855	0.789	0.682	0.750	0.735
UGI_1	14.94	0.37	0.33	0.775	0.908	0.751	0.708	0.597	0.714
Lesion_2	22.43	0.27	0.23	0.738	0.858	0.808	0.578	0.727	0.743
Neuro_2	35.36	0.37	0.32	0.916	0.979	0.854	0.800	0.892	0.805
LGI_1	38.08	0.42	0.39	0.821	0.876	0.763	0.699	0.648	0.703
Gi_1	53.02	0.42	0.40	0.845	0.904	0.762	0.689	0.651	0.704
LGI_2	54.92	0.36	0.33	0.895	0.977	0.873	0.825	0.897	0.865
Gi_2	62.11	0.36	0.33	0.915	0.985	0.889	0.840	0.912	0.888
Rash_2	77.91	0.31	0.27	0.896	0.978	0.836	0.809	0.905	0.840
Fever_1	78.98	0.51	0.45	0.770	0.691	0.656	0.662	0.473	0.664
Resp_2	161.99	0.35	0.32	0.893	0.969	0.840	0.834	0.805	0.850
Resp_1	334.61	0.52	0.50	0.875	0.932	0.834	0.799	0.458	0.825

Outline of Talk

- Problem: monitoring multiple, evolving data streams for anomalies
- Generalized exponential smoothing with Holt-Winters forecasts
- Control charts for H-W residuals with ad hoc adaptations for biosurveillance
- Sample results
- Research directions

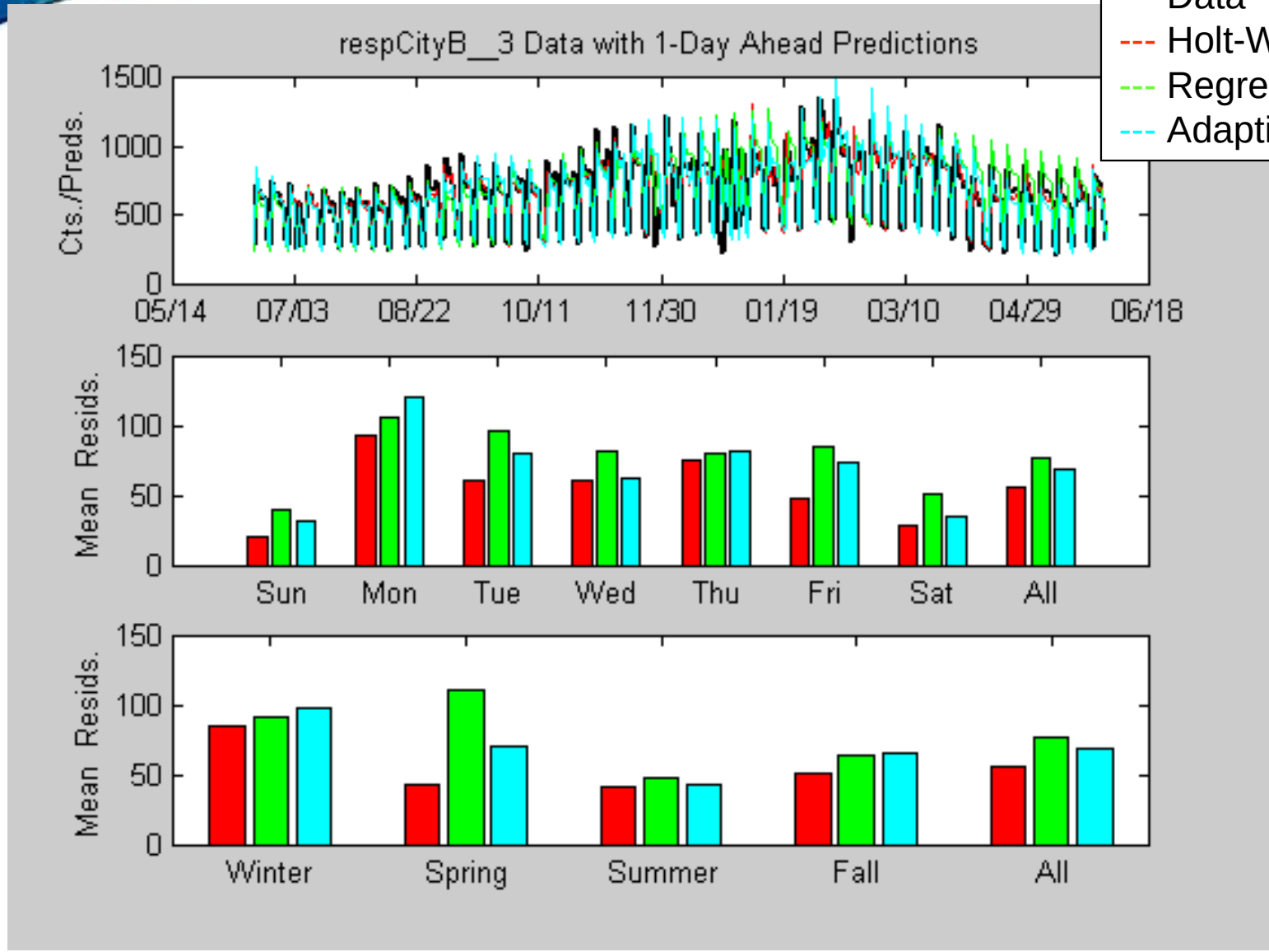
Research Directions

- Classification of time series for improved selection of Holt-Winters coefficients
- System design problems:
  - frequency of inspection for chart modifications
  - improved chart initialization, coefficient choices
- How to consolidate to monitor at manageable alert rates:
  - multiple data streams and syndromes
  - multiple signal types

Multivariate vs Multiple univariate strategies

BACKUPS

Stratified Residual Comparisons



Residual Autocorrelation Comparison
1-Day Ahead Predictions

		1-day ahead predictions							
		1-day lag				7-day lag			
		Holt-Winters	Regression	Adaptive Regression	HW-Adap	Holt-Winters	Regression	Adaptive Regression	HW-Adap
Resp	1	0.11	0.59	0.36	-0.26	0.12	0.62	0.26	-0.14
Data	2	0.14	0.58	0.28	-0.14	0.10	0.60	0.17	-0.06
	3	0.10	0.76	0.21	-0.11	0.06	0.79	0.03	0.03
DOW +	4	0.03	0.37	0.32	-0.28	0.09	0.31	0.24	-0.15
Seasonal	5	0.13	0.50	0.37	-0.24	0.09	0.40	0.23	-0.14
Effects	6	0.24	0.67	0.34	-0.10	0.13	0.66	0.06	0.07
	7	0.08	0.53	0.28	-0.20	0.07	0.57	0.13	-0.07
	8	0.13	0.58	0.34	-0.21	0.20	0.56	0.19	0.02
	9	0.22	0.76	0.59	-0.37	0.16	0.65	0.43	-0.27
	10	0.16	0.79	0.38	-0.22	0.17	0.91	0.26	-0.09
GI Data	1	0.13	0.49	0.13	0.00	-0.02	0.55	0.05	-0.07
	2	0.05	0.67	0.09	-0.04	0.00	0.67	0.05	-0.04
DOW	3	0.14	0.75	0.24	-0.10	0.06	0.75	0.07	-0.02
Effects	4	0.11	0.50	0.20	-0.09	0.01	0.42	-0.05	0.05
Only	5	0.18	0.63	0.22	-0.04	0.09	0.65	0.02	0.08
	6	0.24	0.58	0.16	0.08	0.02	0.62	0.04	-0.02

Residual Autocorrelation Comparison
7-Day Ahead Predictions

	7-days ahead							
	1-day lag				7-day lag			
	Holt-Winters	Regression	Adaptive Regression	HW-Adap	Holt-Winters	Regression	Adaptive Regression	HW-Adap
Resp	0.25	0.59	0.51	-0.26	0.00	0.62	0.40	-0.40
Data	0.27	0.58	0.46	-0.19	-0.06	0.60	0.34	-0.40
	0.18	0.76	0.37	-0.19	-0.01	0.79	0.14	-0.15
DOW +	0.19	0.37	0.46	-0.27	-0.01	0.31	0.36	-0.37
Seasonal	0.29	0.50	0.51	-0.22	0.05	0.40	0.35	-0.30
Effects	0.41	0.67	0.48	-0.06	0.06	0.66	0.11	-0.05
	0.18	0.53	0.40	-0.22	-0.10	0.57	0.27	-0.36
	0.36	0.58	0.49	-0.13	0.10	0.56	0.29	-0.20
	0.40	0.76	0.71	-0.31	0.05	0.65	0.56	-0.50
	0.39	0.79	0.52	-0.12	0.09	0.91	0.41	-0.32
GI Data	0.17	0.49	0.22	-0.05	-0.04	0.55	0.08	-0.12
	0.08	0.67	0.23	-0.15	-0.04	0.67	0.07	-0.10
DOW	0.21	0.75	0.35	-0.14	0.03	0.75	0.15	-0.12
Effects	0.16	0.50	0.34	-0.19	-0.04	0.42	-0.02	-0.02
Only	0.18	0.63	0.33	-0.15	0.06	0.65	0.02	0.04
	0.27	0.58	0.21	0.05	-0.04	0.62	0.01	-0.05

EWMA Concept & Smoothing Constant

Brown, R.G. and Meyer, R.F. (1961), "The Fundamental Theorem of Exponential Smoothing," Operations Research, 9, 673-685.

- Exponential smoothing represents “an elementary model of how a person learns”:

  $\bar{x}_k = \bar{x}_{k-1} + \omega (x_k - \bar{x}_{k-1})$ where $0 < \omega < 1$

- For the smoothed value S_k,

  $S_k = \omega S_{k-1} + (1-\omega)X_k$,

  The variance of S_k is $\sigma_S = [\omega / (2 - \omega)] \sigma_X$

- So a smaller ω is preferred because it gives a more stable S_k; values between 0.1 and 0.3 often used
- But Chatfield: changes in global behavior will result in a larger optimal ω