# Automated Time Series Forecasting for Biosurveillance

**Howard S. Burkom, Sean P. Murphy and Galit Shmueli**

*The Johns Hopkins University Applied Physics Laboratory and the Decision & Information Technologies Department Robert H. Smith School of Business, U. of Maryland College Park*

### Objective

For robust detection performance, alerting algorithms for biosurveillance require input data free of trends, day-of-week effects, and other systematic behavior. Time series forecasting methods may be used to remove this behavior by subtracting forecasts from observations to form residuals for algorithmic input. This abstract examines and compares methods for the automatic preconditioning of health indicator data to enable the timely prospective monitoring required for effective syndromic surveillance.

### Background

The statistical process control (SPC) community has developed a wealth of robust, sensitive monitoring methods in the form of control charts [1]. Although such charts have been implemented for a wide variety of health monitoring purposes [2], some implementations monitor data that violate basic assumptions required by the control charts [3] yielding alerting methods with uncertain detection performance. This problem highlights an inherent obstacle to the use of traditional SPC methods for syndromic surveillance: the nature of the data. Syndromic data streams are based not on physical science, as are manufacturing processes, but on changing population behavior and evolving data acquisition and classification procedures. To overcome this obstacle, either more sophisticated detection algorithms must be developed or the data must be preconditioned so that it is appropriate for traditional monitoring tools.

### Methods

This study evaluated 3 forecasting techniques—a nonadaptive loglinear regression model using a long historical baseline, an adaptive regression model with a shorter, sliding baseline, and the Holt-Winters (HW) exponential smoothing-method for generalized exponential smoothing—using 16 authentic syndromic time series derived from 3 data sources containing day of week (DOW) and seasonal trends. To evaluate each technique's predictive performance, 3 measures were applied to the 1-day ahead and 7-day ahead residuals: the root mean squared error, the median absolute deviation, and the median absolute percentage error (MedAPE). The autocorrelation of the residuals was also compared.

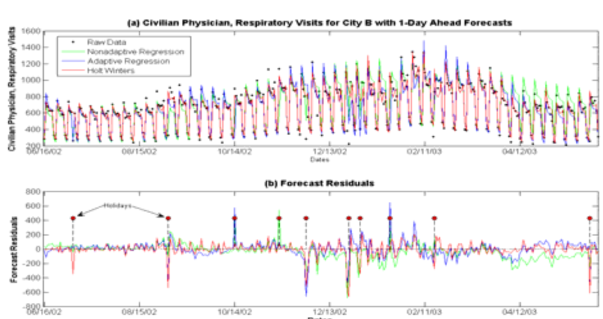


Figure 1—Respiratory syndrome time series with forecasts (a) and residual comparisons (b) with 10 calendar holidays indicated.

### Results

The two median-based criteria showed best overall performance for the HW method. The MedAPE measures over the 16 test series averaged 16.5, 11.6, and 9.7 for the nonadaptive regression, adaptive regression, and HWs methods, respectively. The nonadaptive regression forecasts were degraded by changes from the data behavior in the fixed baseline period used to compute model coefficients. The mean-based criterion was less conclusive due to the effects of poor forecasts on a small number of calendar holidays. The HW method was also most effective at removing serial autocorrelation, with most 1-day-lag residual autocorrelation coefficients below 0.15. The forecast methods were compared without tuning them to the behavior of individual series. We achieved improved predictions by tuning the HW method, but practical use of such improvements for routine surveillance will require reliable data classification methods.

### Conclusions

This examination suggests that HW smoothing is better suited than regression-type methods for removing DOW and seasonal patterns from a range of syndromic time series. This method produces lower forecast errors, can easily be automated, and does not require much data history. Several adaptations are required, most notably accounting for calendar holidays. For the data tested, the HW residuals were much closer to meeting the requirements for input to popular control chart methods such as CuSum and EWMA monitoring.

### References

[1] Ryan TP. Statistical Methods for Quality Improvement. New York: John Wiley & Sons: New York, 1989

[2] Benneyan JC. Part I: Introduction and Basic Theory, Part II: Chart use, statistical properties, and research issues. Statistical Quality Control Methods in Infection Control and Hospital Epidemiology, Infection and Hospital Epidemiology, Vol. 19, (3)194-214.

[3] Woodall, W.H. (2006). "The Use of Control Charts in Health Care Monitoring and Public Health Surveillance", Journal of Quality Technology (in press)