

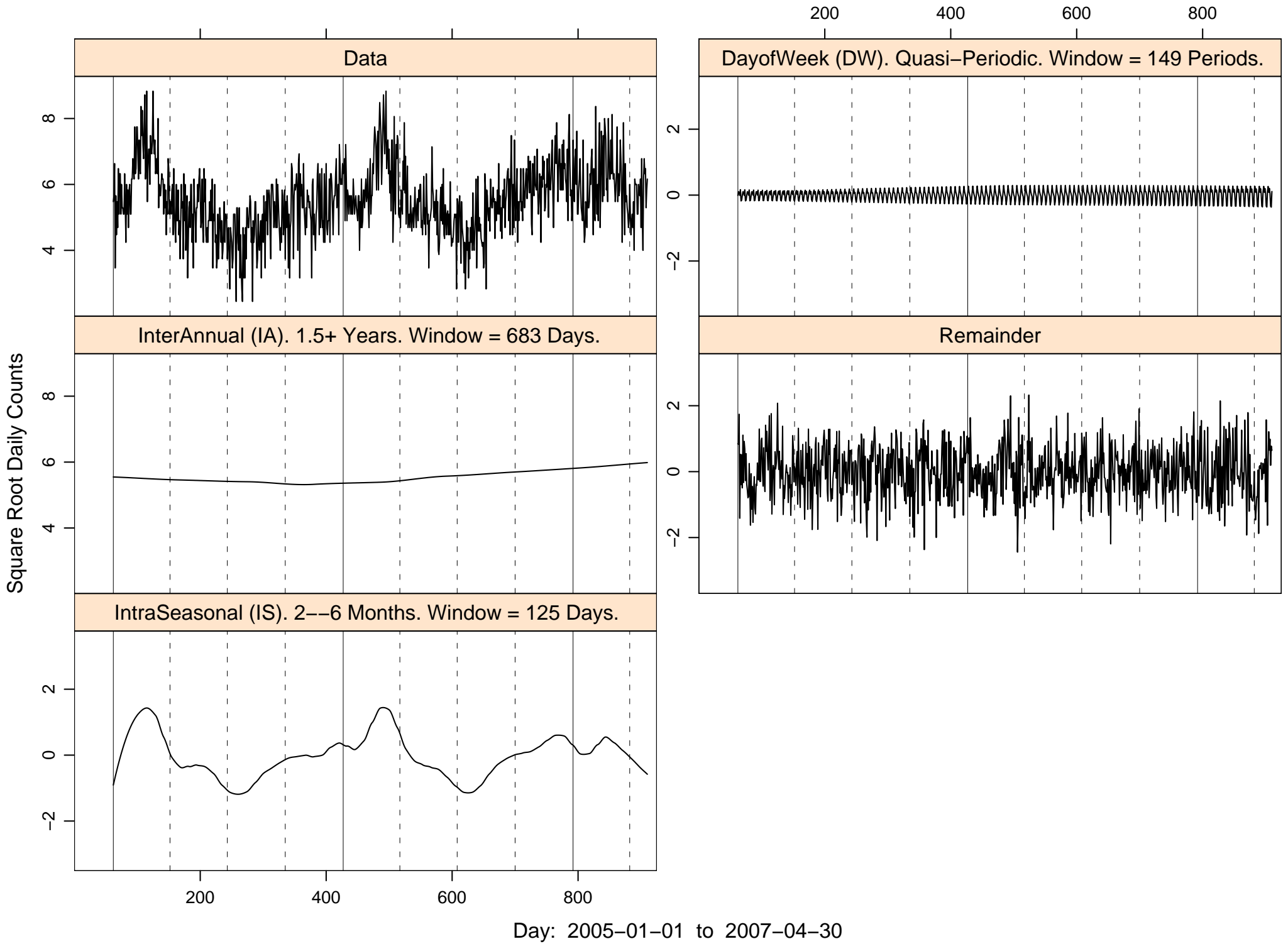
STL and Local Regression for Modeling Disease Surveillance Counts

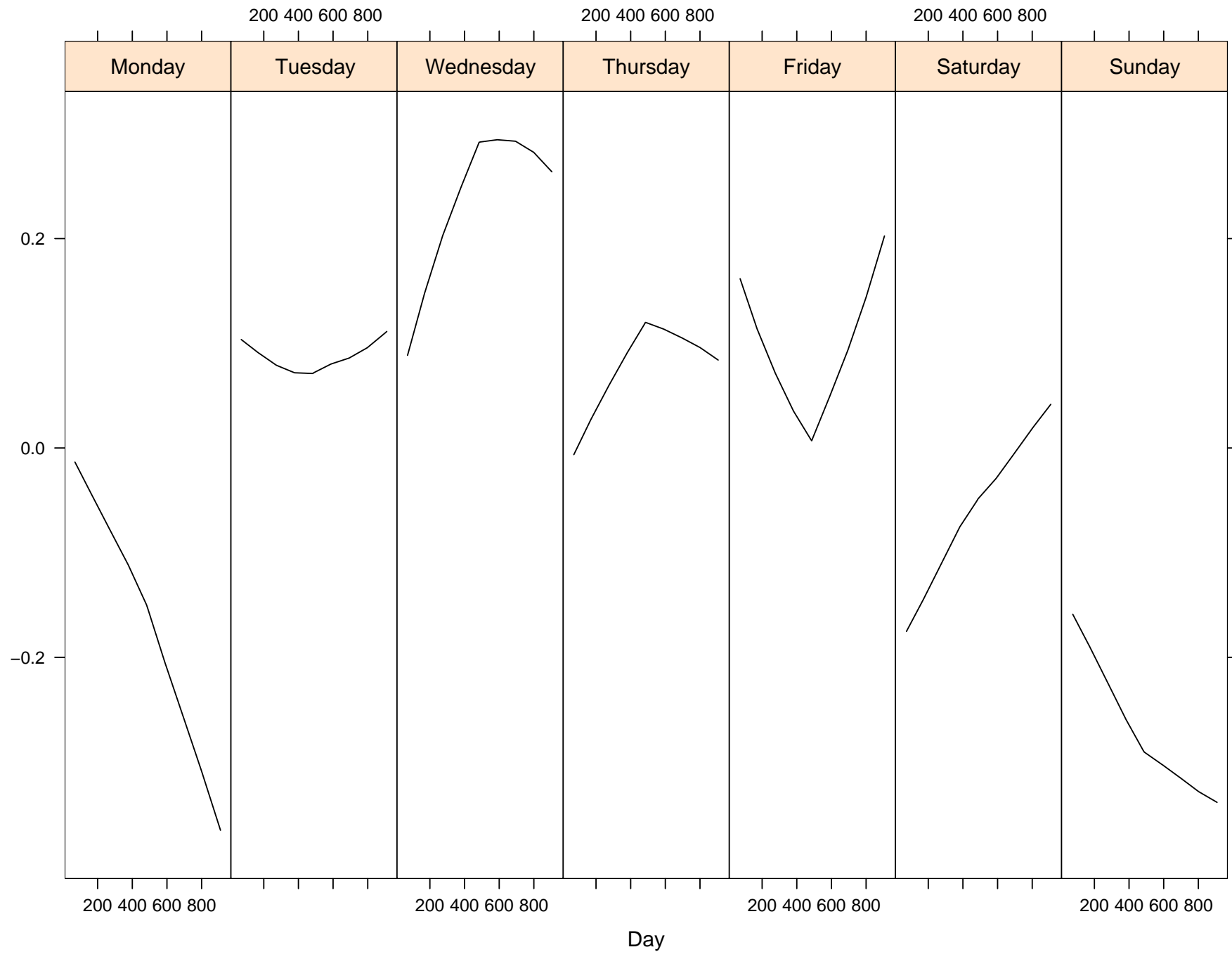
David E. Anderson, Cheng Zheng, Ross Maciejewski, Ryan Hafen, William S. Cleveland, David S. Ebert, Mourad Ouzzani, Purdue Regional Visualization & Analytics Center, W. Lafayette.

Shaun J. Grannis, Regenstrief Inst. & Indiana U. School of Medicine, Indianapolis.

Objectives

- Use STL (Seasonal Decomposition of Time Series by Loess) and local regression to model chief-complaint daily counts.
- Develop visualization tools for model display and checking.





Background

- Removing the influence of interannual and day of week components is essential for accurate outbreak detection.
- STL is a well proven method, based on loess, which can be used to extract out interannual, intraseasonal, and day of week components.
- Dafni et. al. successfully used loess to directly estimate a yearly seasonal effect for “daily syndromic counts in emergency departments of four major hospitals in the Athens area during August 2002–August 2003.”

- Chief-complaint daily counts from Emergency Rooms at 16 Indiana hospitals.
- At least two years of data from 14 of the 16 hospitals. Dates roughly from November 2004 to April 2007.
- Focus on Gastro-Intestinal and Respiratory daily counts.

- Loess is a flexible and computationally efficient method of local regression.
- To evaluate the loess fit $g(x)$ at a given x , all data points in the neighborhood of x are assigned tricubic weights, so that the closer a point is to x , the larger its weight.
- Weighted least squares is then used to fit a polynomial through the points, and $g(x)$ equals the value of the polynomial at x .
- Parameters are the size of the neighborhood and the degree of the polynomial (constant, linear, or quadratic).
- STL is an iterative procedure that repeatedly uses different types of loess.

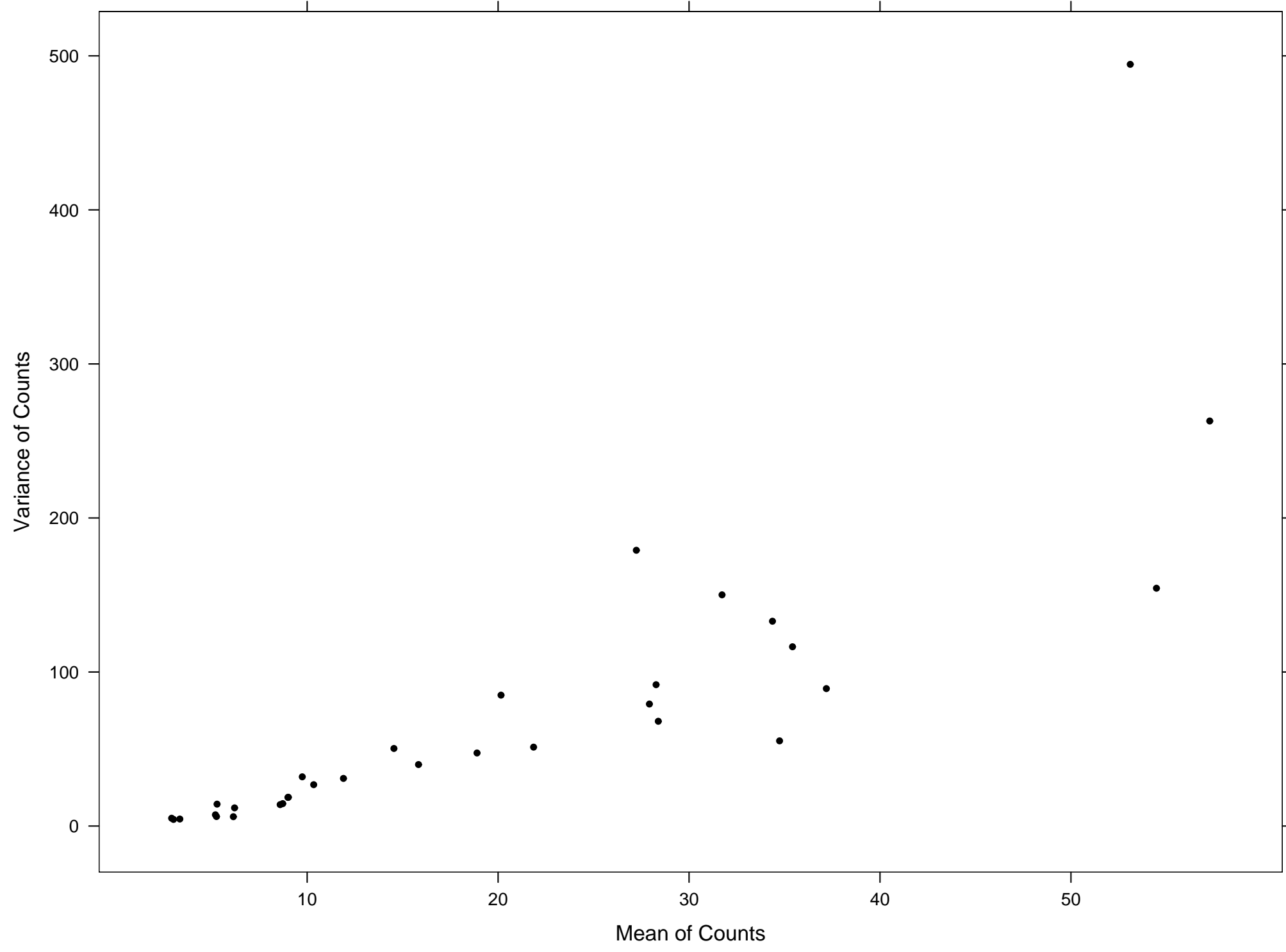
Method

- Use square roots of daily counts to stabilize variances.
- Use linear STL to extract evolving day of week seasonal component (window = 149 periods).
- Use quadratic loess to extract interannual component (window = 683 days).
- Use quadratic loess to extract intraseasonal component (window = 125 days).
- Perform model checking on the remainder, which can be modeled as iid Normal random variables.

Parameters selected through extensive use of data visualization, to prevent overfitting or lack of fit.

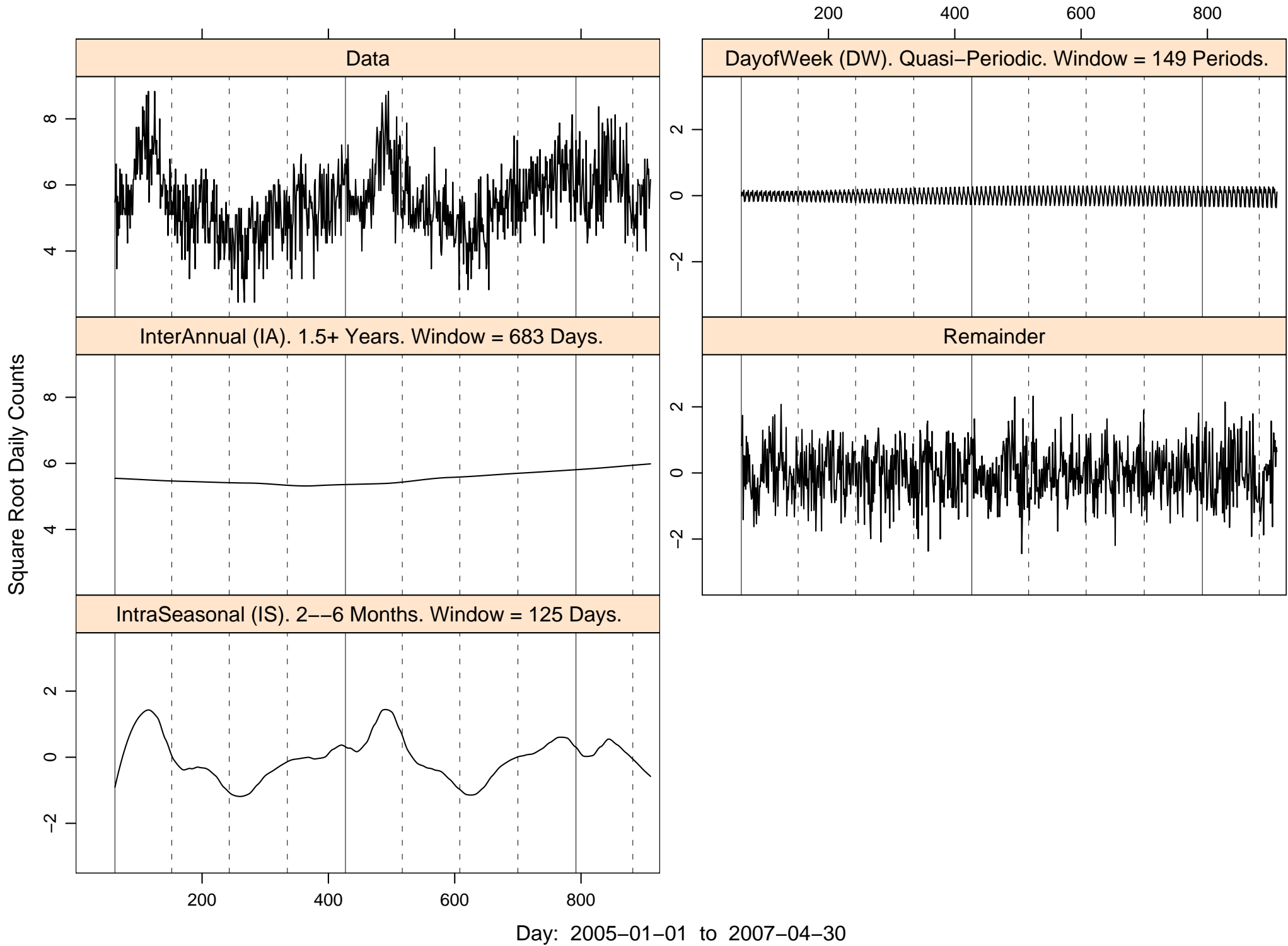
Square Root Transformation

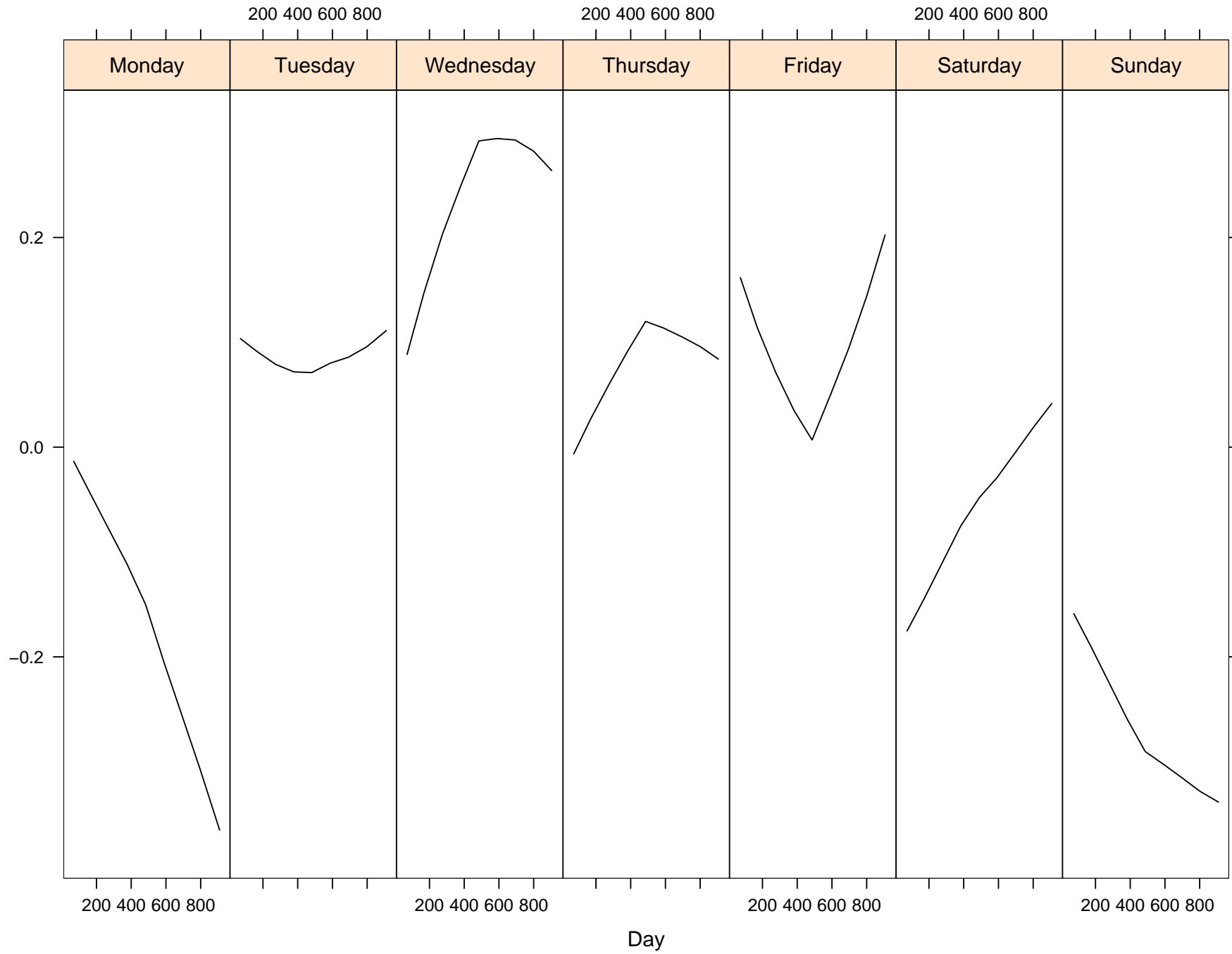
- Makes mean and standard deviation roughly independent.
- Makes distribution of the remainder much closer to normal.
- Allows for a more parsimonious model.



Results

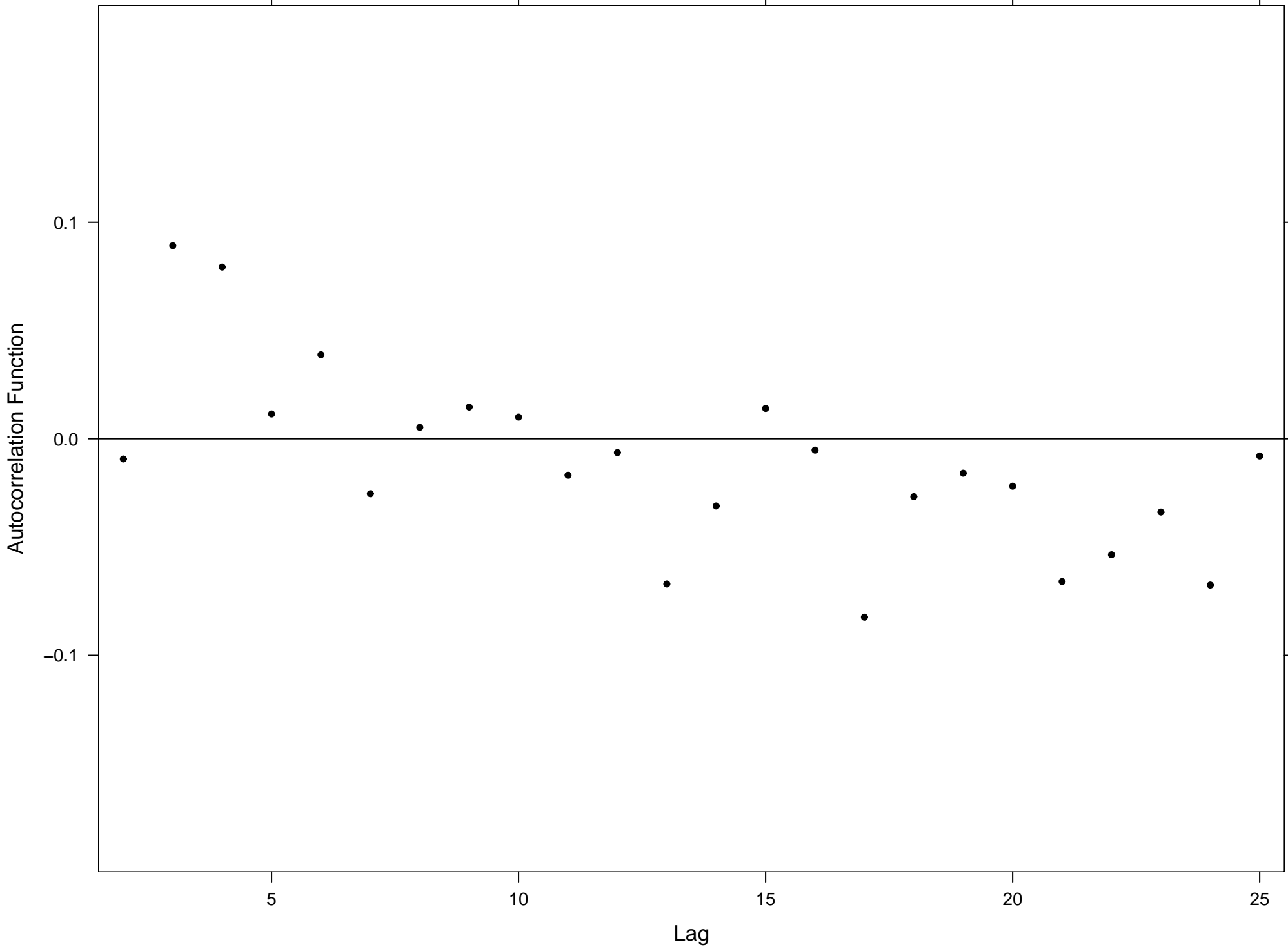
- Intraseasonal components show seasonal influenza.
- Interannual components show growth over time in some hospitals.
- Day of week components for most hospitals peak on Monday, drop through Friday, and then rise over the weekend.
- No significant effects showed up in the remainder terms.

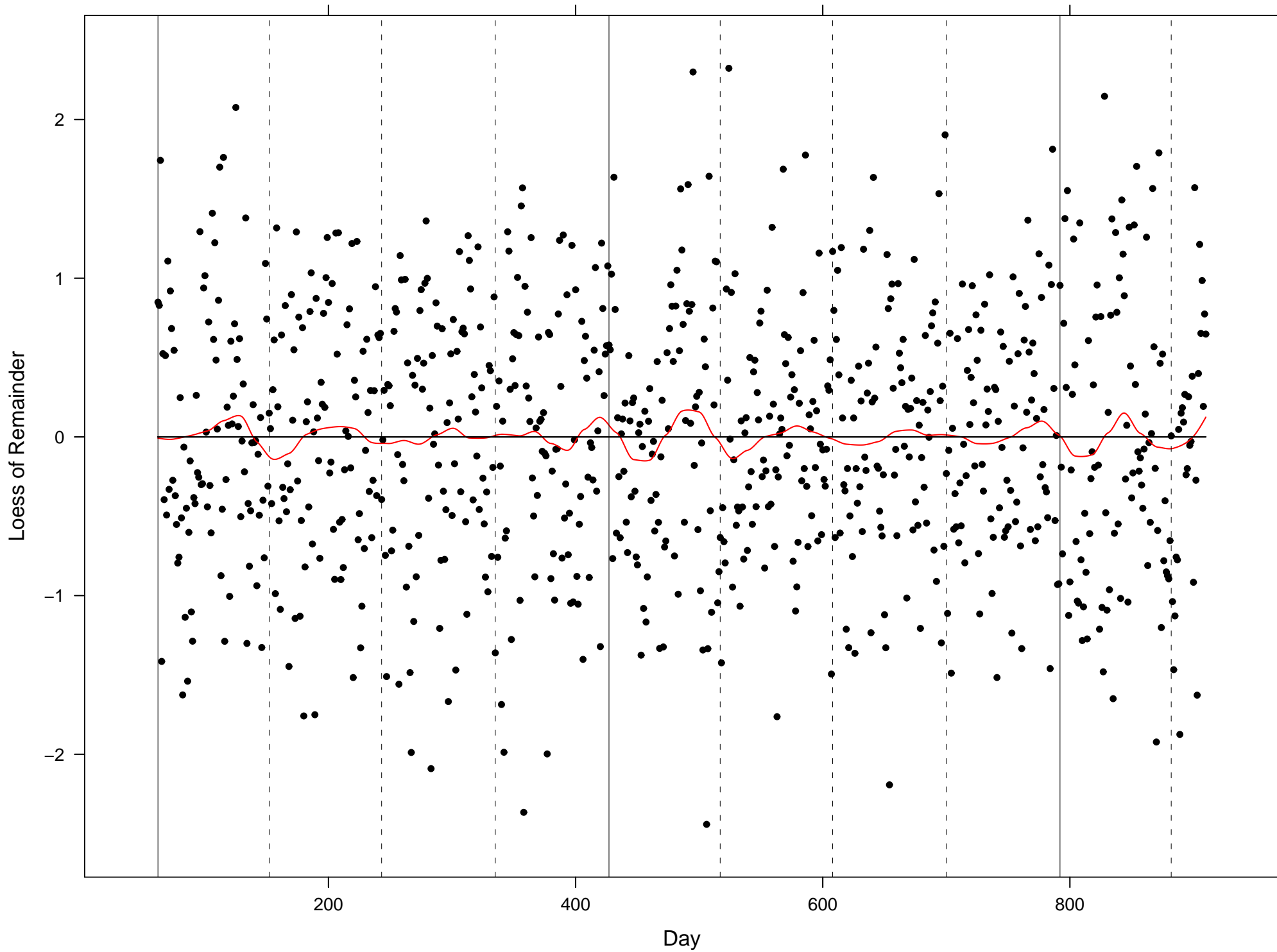




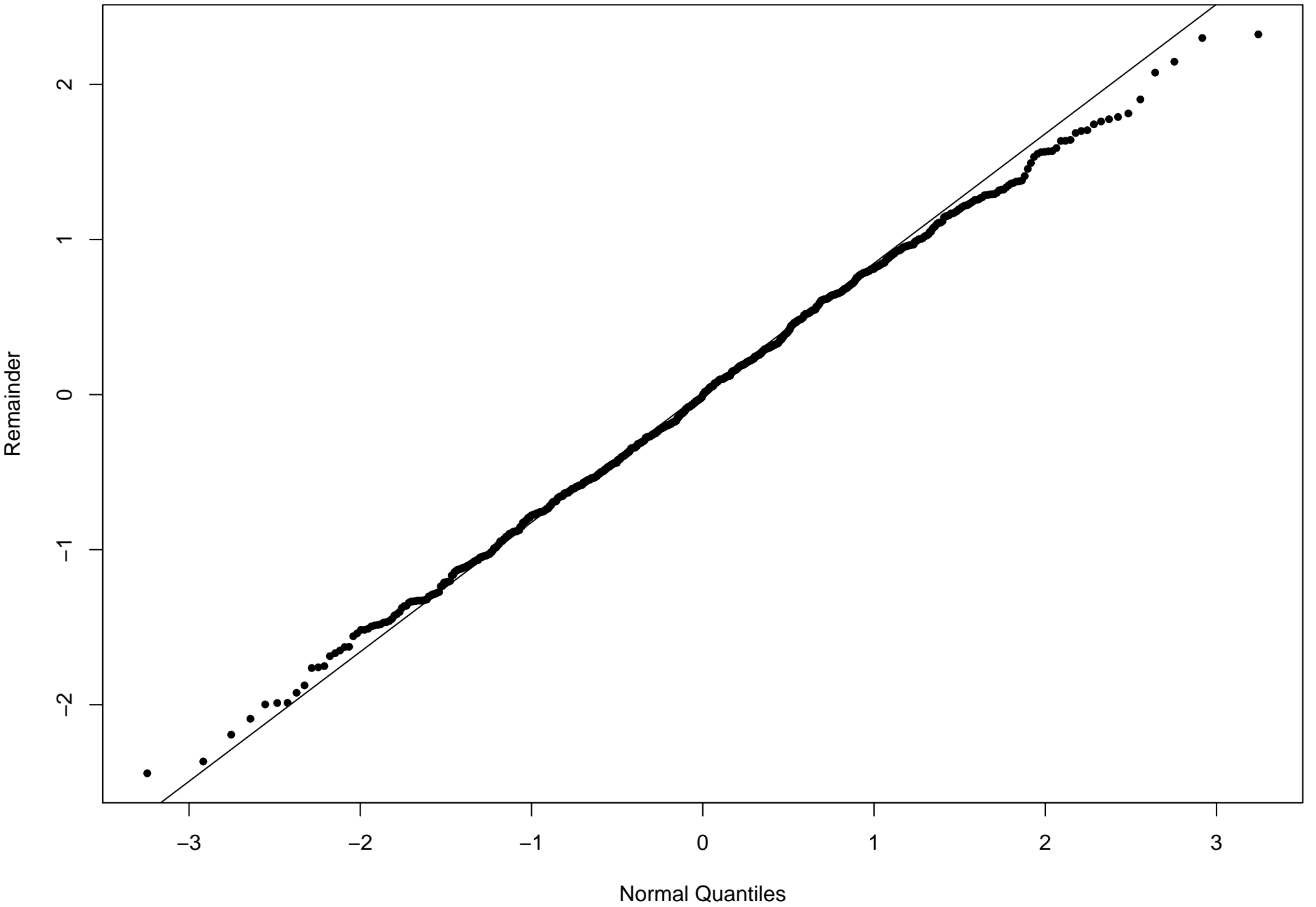
Model Checking

- Autocorrelation function of remainder
- Loess of remainder
- Normal probability plot of remainder
- Test plots using a white noise series





Normal Q-Q Plot



Future Work

- Adding patient addresses and expanding to STL based spatio-temporal procedure.
- Using control charts to detect outbreaks.

Control Charts

- Run STL decomposition each day.
- Build up last day remainder through time. The smoothed distribution provides the control limits.
- Plot up last day remainder plus last day intraseasonal component. Smooth results.
- Can be varied to focus on sudden outbreaks or more gradual problems.

Conclusion

STL and local regression is a promising as a method for modeling chief-complaint daily counts and as a basis for control chart methods for detecting outbreaks.

Special thanks to Department of Homeland Security Center of Excellence Program for funding.

References

- H.S. Burkom. Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance. *Johns Hopkins APL Tech. Digest*, 24, 2003.
- R.B. Cleveland, W.S. Cleveland, J.E. McRae, I. Terpenning. STL: A Seasonal- Trend Decomposition Procedure Based on Loess. *J. Official Statist.*, 6:373, 1990.
- W.. Cleveland, S.J. Devlin. Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J. Amer. Statist. Assoc.*, 83:596610, 1988.
- S.J. Grannis, P.G. Biondich, B.W. Mamlin, G. Wilson, L. Jones, J.M. Overhage. How Disease Surveillance Systems Can Serve as Practical Building Blocks for a Health Information Infrastructure: the Indiana Experience. *AMIA Annual Symp. Proc.*, 286290, 2005.
- U.G. Dafni, S. Tsiodras, D. Panagiotakos, K. Gkolfinopoulou, G. Kouvatses, Z. Tsourti, G. Saroglou. Algorithm for Statistical Detection of Peaks Syndromic Surveillance System for the Athens 2004 Olympic Games. *Morbidity and Mortality Wkly. Rept.*, 53:86-94, 2004.

