SAND2018-1982C

Event Prediction Using Graph-Augmented Temporal Analysis

Sandia National Laboratories

Randy Brost, Vitus Leung, Hamilton Link, Cindy Phillips, Andrea Staid

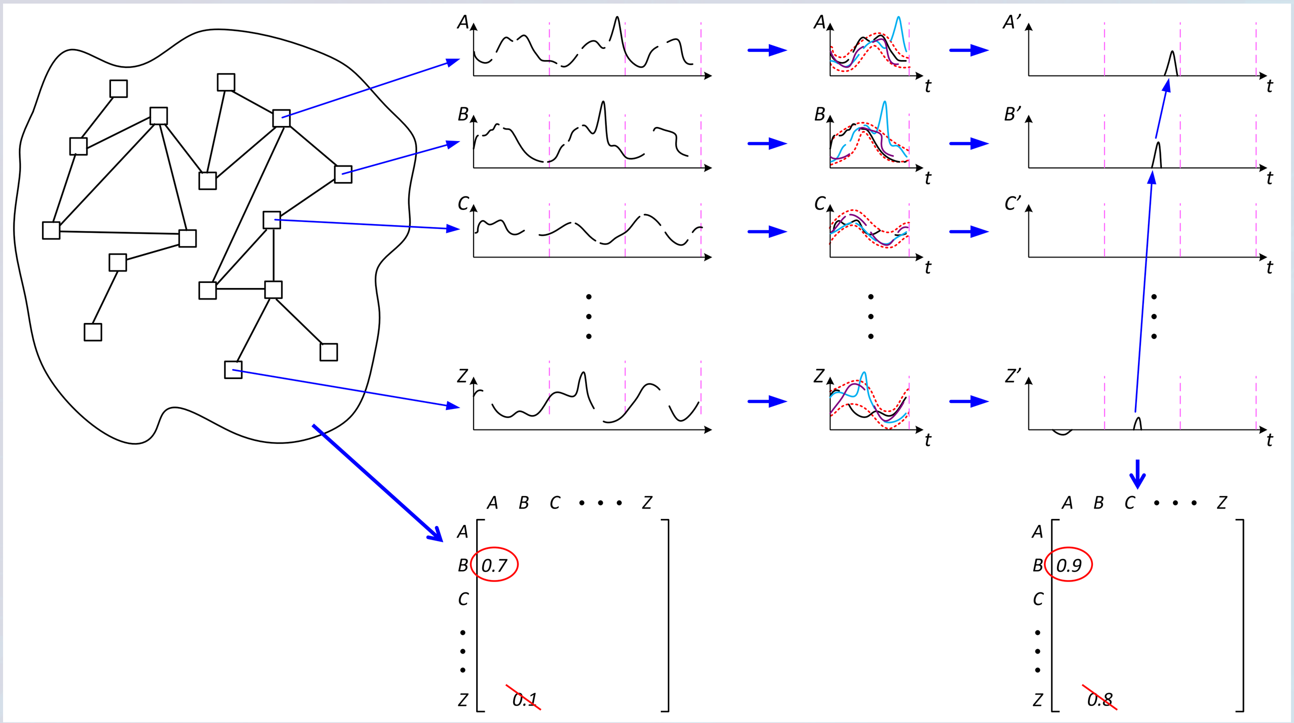
Sandia National Laboratories, Albuquerque, NM 87185

Problem

- Data streams are voluminous and complex.
- Hard to understand event chains leading up to events.
- Prior work: Spatiotemporal search for specified patterns.
- This project seeks to find unknown temporal relationships.

Approach

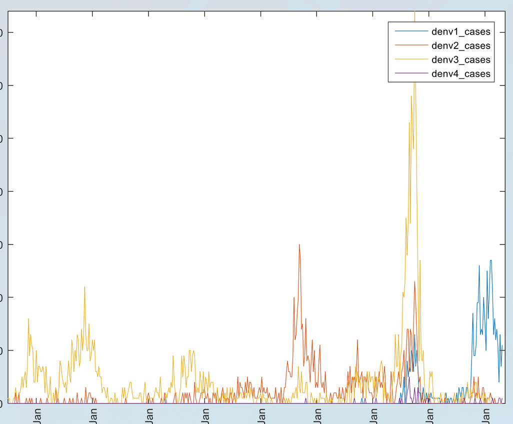
Find precursors via temporal and graph analysis.



A domain has multiple data streams logged at various points. Temporal analysis identifies potential correlations, which are either reinforced or deprecated by graph information.

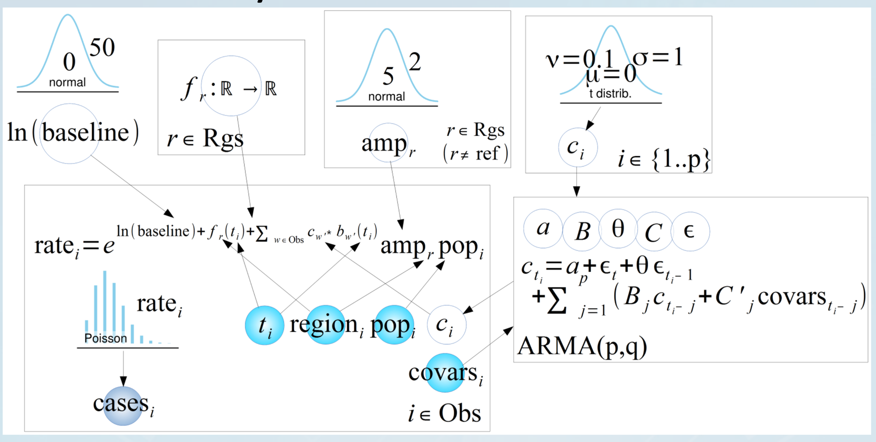
Example Problems:

1. Dengue Fever Outbreaks:



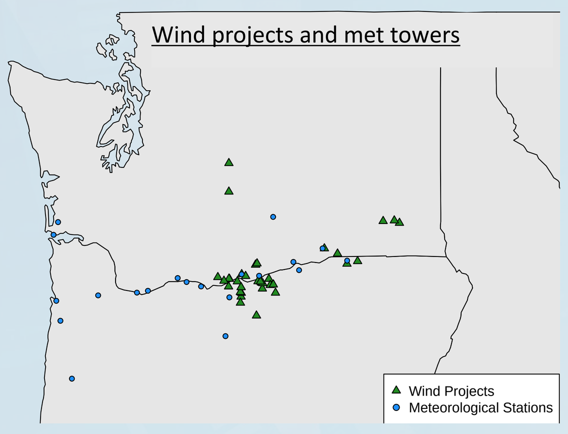
Data: Dengue cases, weather…
Goal: What precedes an outbreak?

Bayesian MCMC Model



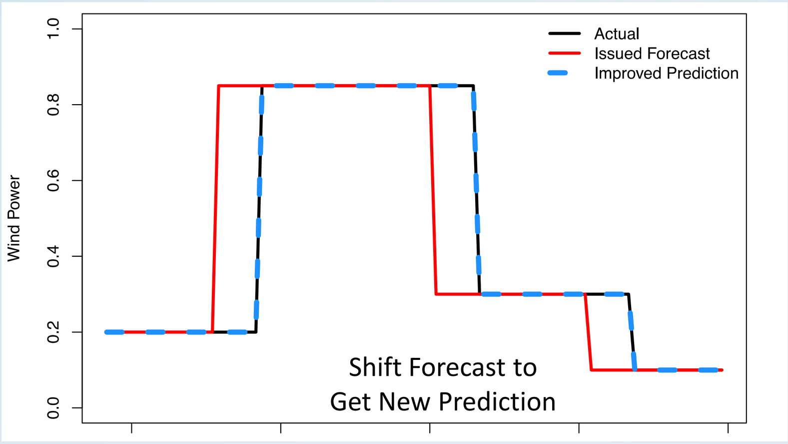
Method: Construct Bayesian model relating context, weather, and Dengue cases. Characterize priors via Monte Carlo analysis. Find parameters connected to outbreaks.

2. Wind Power Analysis:



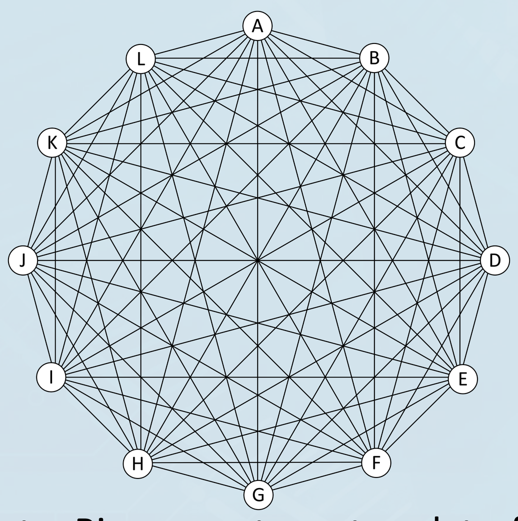
Data: Wind power forecast, actual…
Goal: Can we predict forecast errors?

Forecast Shift Analysis



Method: Identify adjustments to forecast based on recent, local actual performance.

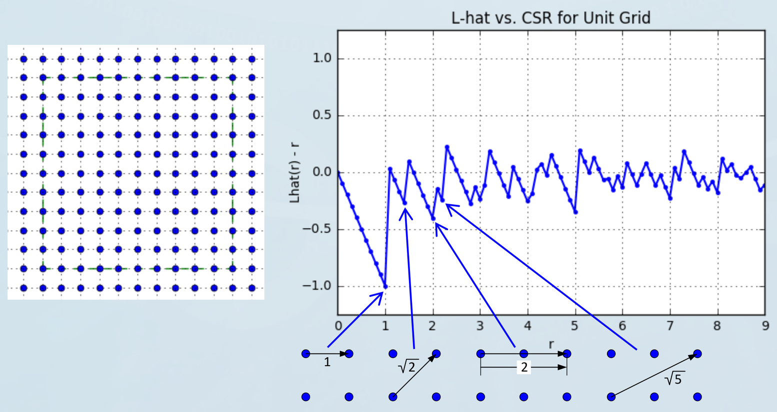
3. Network Analysis:



Data: Pizza events, network traffic.
Goal: Who orders the pizza?

Ripley's K Function

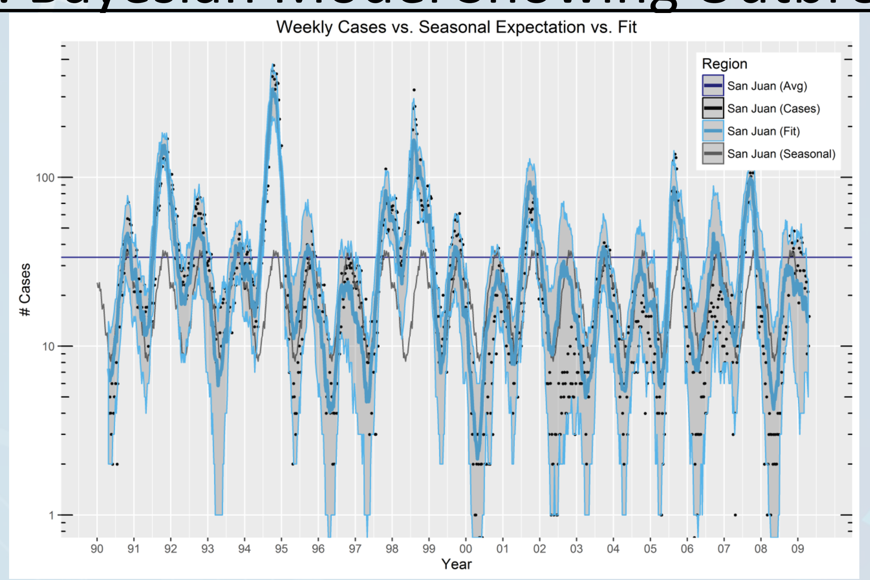
$$\hat{K}(r) = \hat{\lambda}^{-1}\sum_{i}^{N}\sum_{j\neq i}^{N}\frac{I[d_{ij}<r]}{N} \qquad \hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}}$$

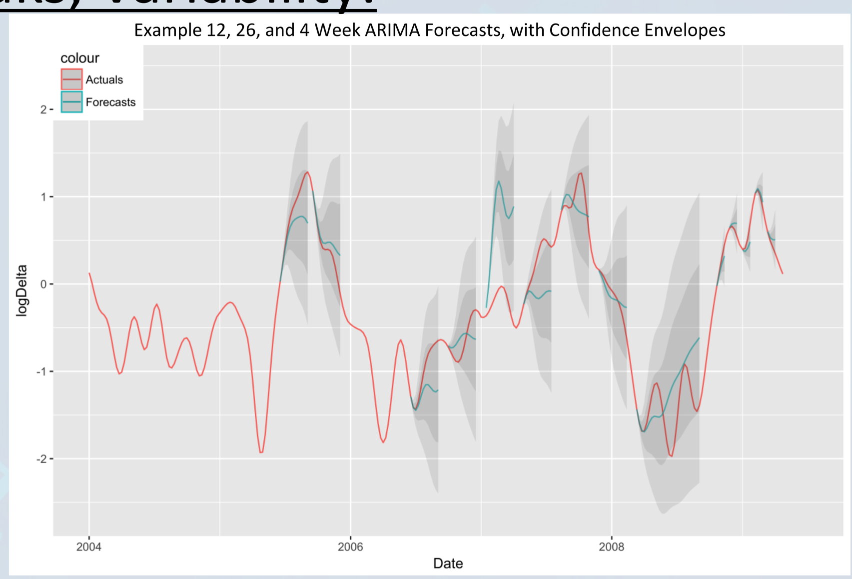


Method: Identify structures in spatial and temporal data that deviate from random.

Results

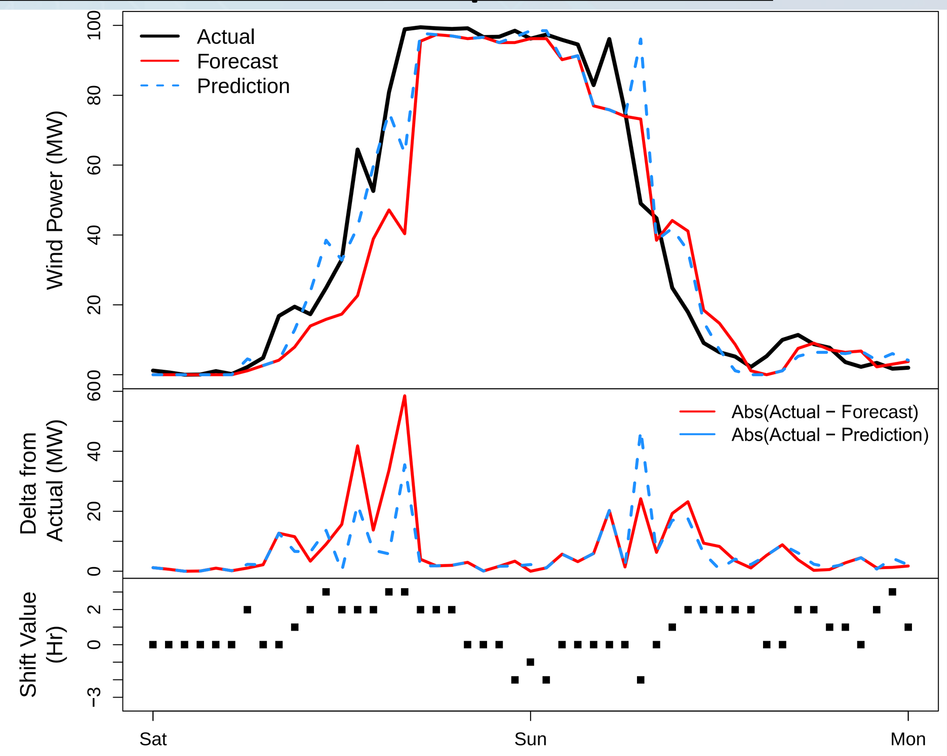
1. Bayesian Model Showing Outbreaks, Variability:





Result: Model fit captures seasonal, out-of-normal. ARIMA prediction demonstration.
Next: Assess environmental influence.

2. Wind Forecast Improvement:



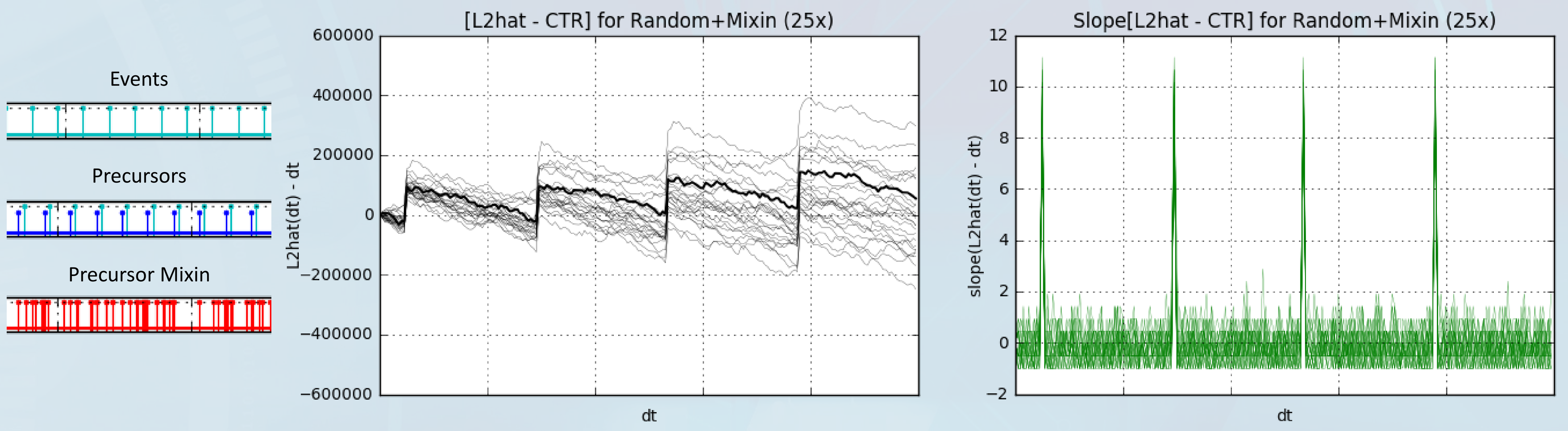
Annual Summary (normalized capacity)

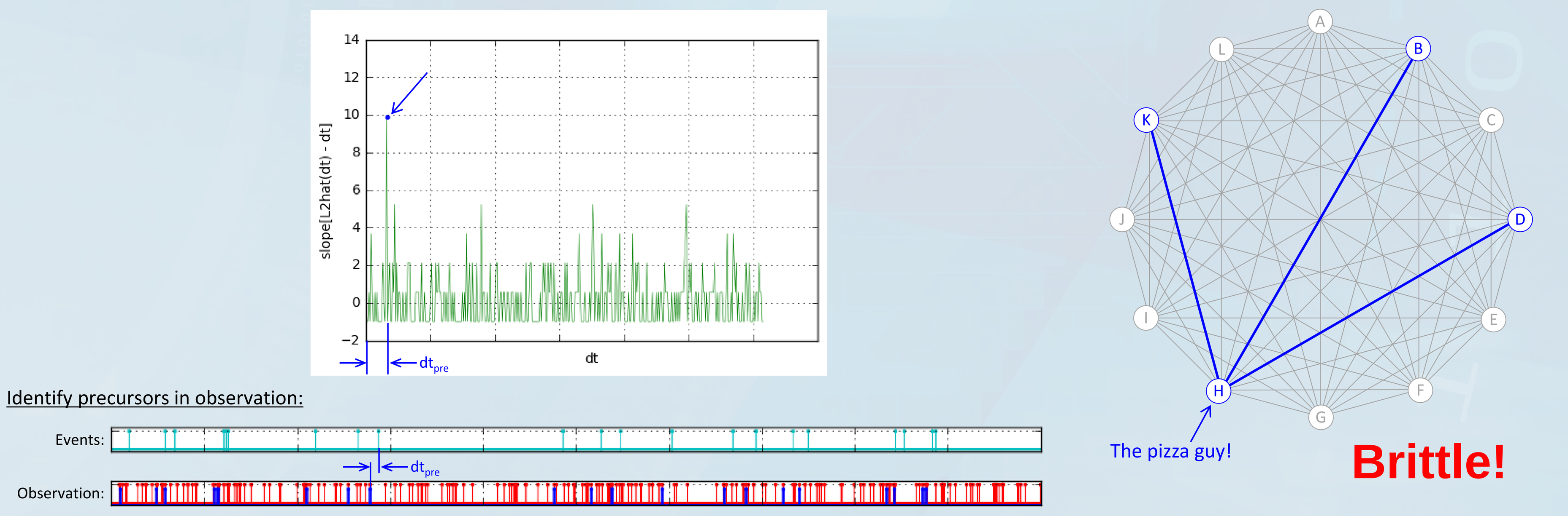
Wind Project	RMSE: Forecast	RMSE: Prediction	Annual Production (MWh)	Annual Savings (MWh)	Annual Savings (%)
A	18.3	17.6	236,021	6,022	2.6%
B	21.2	19.2	204,509	7,136	3.5%
C	18.6	17.6	244,459	4,597	1.9%
D	17.3	16.8	185,649	4,344	2.3%
E	22.9	19.8	201,838	12,018	6.0%
F	20.0	17.8	196,161	6,728	3.4%
G	22.7	20.0	203,610	9,548	4.7%

1-hour lead time for graph (project E) and table.
Data courtesy of Bonneville Power Administration.

Result: Improved forecast accuracy for some wind projects, for short 1-hour lead times.
Savings vary by project, but up to 6% annually for 1-hour lead time.
Next: Improved algorithms, geographic coupling.

3. Network Precursor Analysis (synthetic data):





Result: Modified K-function analysis found precursor network traffic within pure synthetic noise, given prefect precursor lead time.
Next: Add variable lead time, noise, real data. Seek robust method.

Significance

Potential impact:

- New algorithms for solving spatiotemporal analysis problems.
- Supports forensics, warning, relationship identification.
- Multiple potential national security applications.

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories