Statistical Upscaling of Numerical Weather Predictions to Enable Coupled Modelling of Local Weather Impacts

L. Treinish, H. Li and A. Praino
IBM Thomas J. Watson Research Center
Yorktown Heights, NY
{lloydt,liho,apraino}@us.ibm.com

http://www.research.ibm.com/weather/DT.html

Statistical Upscaling of Numerical Weather Predictions to Enable Coupled Modelling of Local Weather Impacts

- Background and motivation
- Approach
  - Meteorology + outages (modelling and analysis)
  - Coupling and calibration
- Example results
- Conclusions and future work

Other Presentations of Related Work

Conference on Weather, Climate, and the New Energy Economy:

- **J7.1** ***Operational utilization and evaluation of a coupled weather and outage prediction service for electric utility operations***
- **P782** ***Application of an operational meso-scale modelling system for industrial plant operations***
- **P765** ***Wind farm layout optimization***
- **7B.5** ***Optimal unit commitment and dispatch for wind farm operations***

Background and Motivation

- The operation of the distribution system of an electric utility, particularly with an overhead infrastructure, can be highly sensitive to local weather conditions
- What is the potential to enable proactive allocation and deployment of resources (people and equipment) to minimize time for restoration?
  - Ability to probabilistically forecast weather conditions and their disruptive impact at an area substation level with useful lead time
  - Can highly localized, NWP-based forecasts be adapted to reduce the uncertainty in decision making?
  - Can the link between weather and impact be quantified to improve preparation and response?

Weather prediction

↓

Damage prediction

↓

Resource requirement prediction

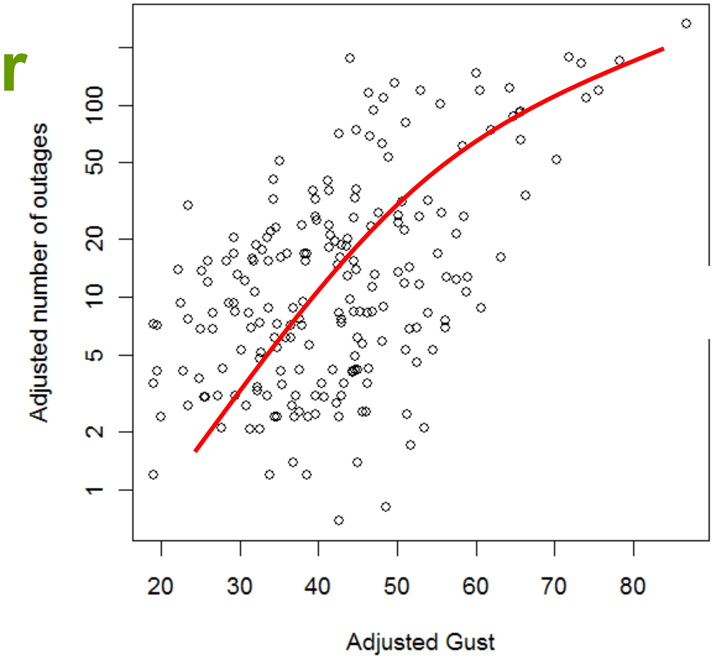
↓

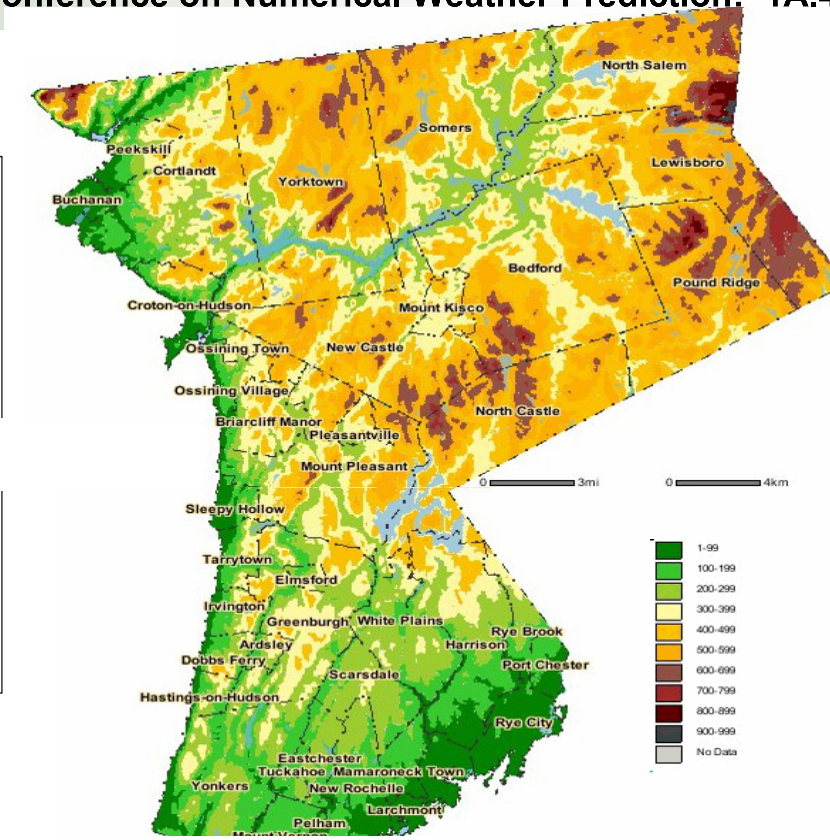
Restoration time prediction

Initial Steps

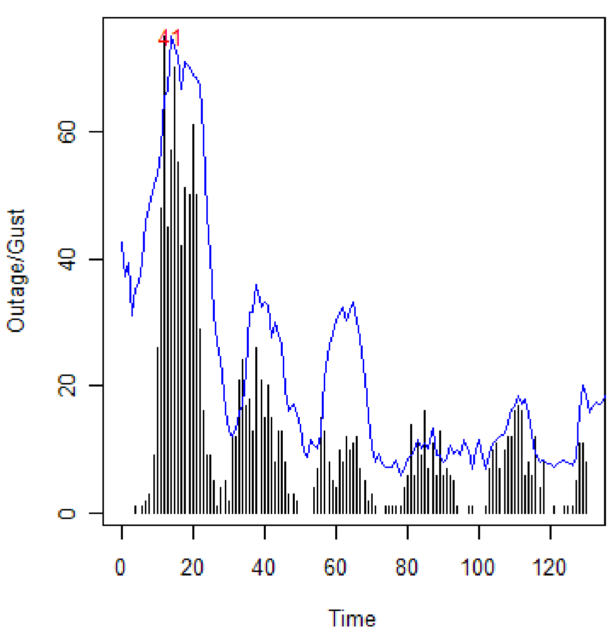
- Determine causes of damage due to storms that lead to outages
- Analyze static and time-varying geo-spatial data, each of which have different uncertainties
  - Damage location, timing, type and response
  - Historical weather observations and NWP data
  - Demographics of affected area
  - Infrastructure impacted in affected area (e.g., poles, wires)
  - Ancillary environmental conditions (e.g., soil, trees)
- Build predictive model from environmental observations, storm impact, weather and related data

Wind Gusts Are a Primary Driver for Outages in this Service Territory

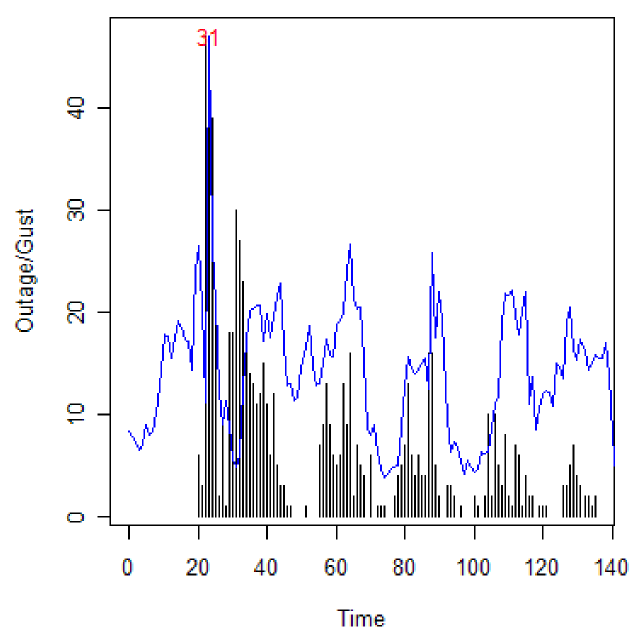




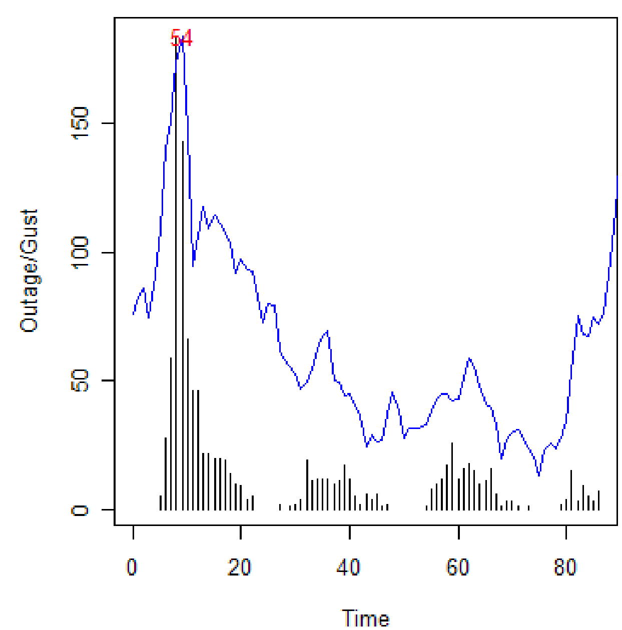
2Sep2006



18Jul2006



18Jan2006

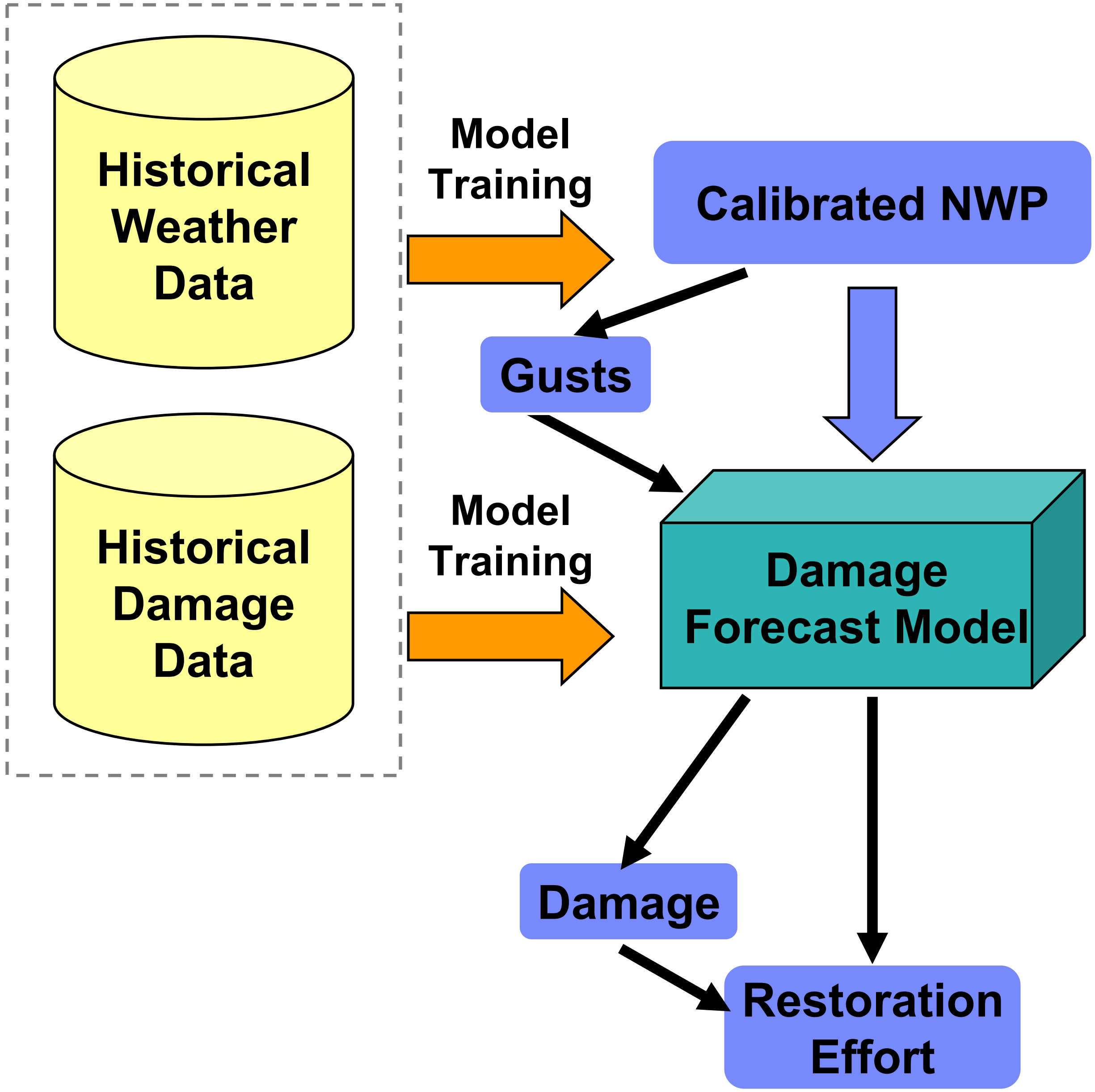


Implication of Gusts as a Driver for Outages

- **Storm-driven disruptions of the overhead electric distribution network (e.g., poles and wires) are caused by physical interaction of the atmosphere directly with that infrastructure or indirectly via nearby trees**
- **Reliable NWP at this turbulence scale with sufficient throughput for operational utilization is neither tractable from a computational perspective nor verifiable from observations**
- **Ensemble NWP cannot capture uncertainty related to impacts**
- **Therefore, outage prediction must be approached stochastically by post-processing NWP**
- **Since the relevant data and applications are inherently temporal and spatial, we use a Bayesian hierarchical model**

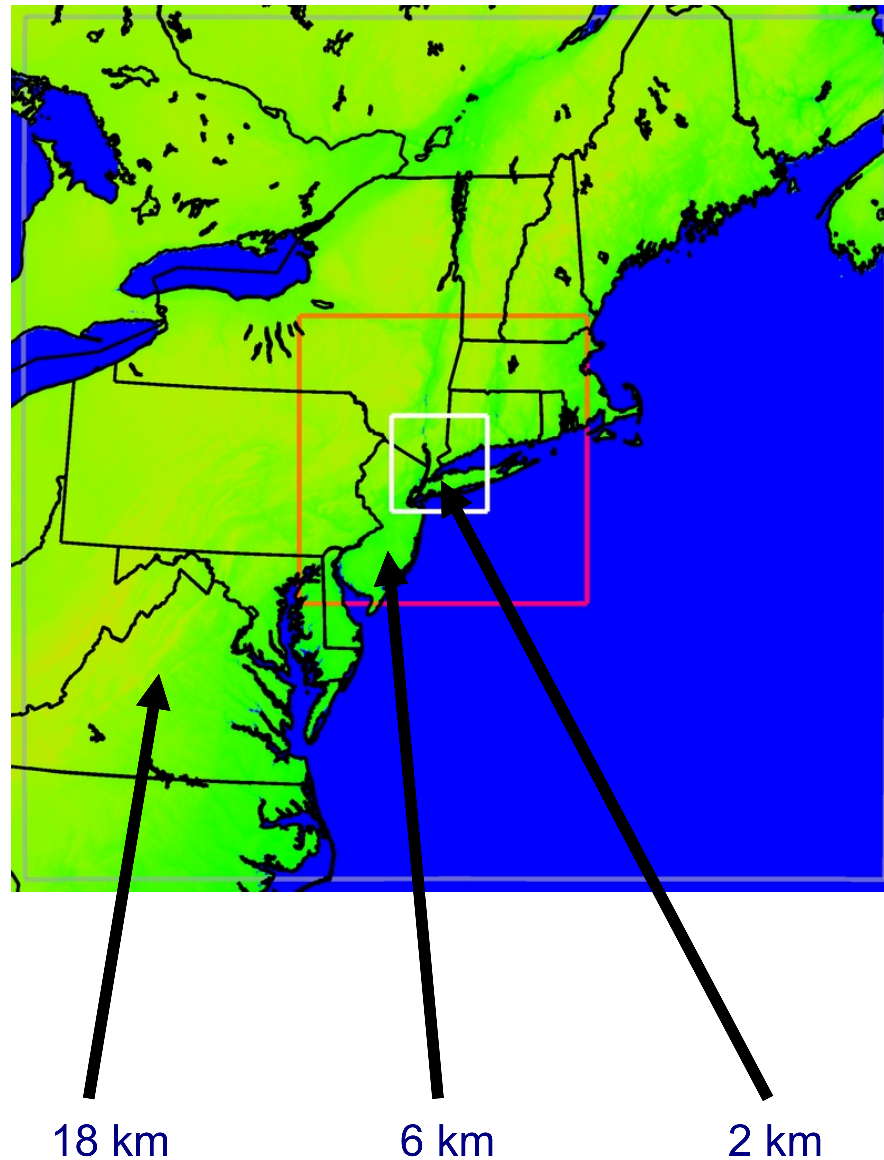
Coupled Weather and Impact Modelling ("*Deep Thunder*")

- Damage: custom modelling for predictions of outages
  - A damage forecast model at the area substation level is developed using historical weather and outage data as well as infrastructure and environmental data by building a hierarchical Poisson regression model
  - Derived gusts are developed statistically using time series analysis based on historical wind forecasts and gust observations
  - Statistical hierarchical modeling integrates various data sources in one model and allows variances or uncertainties analyzed at different levels



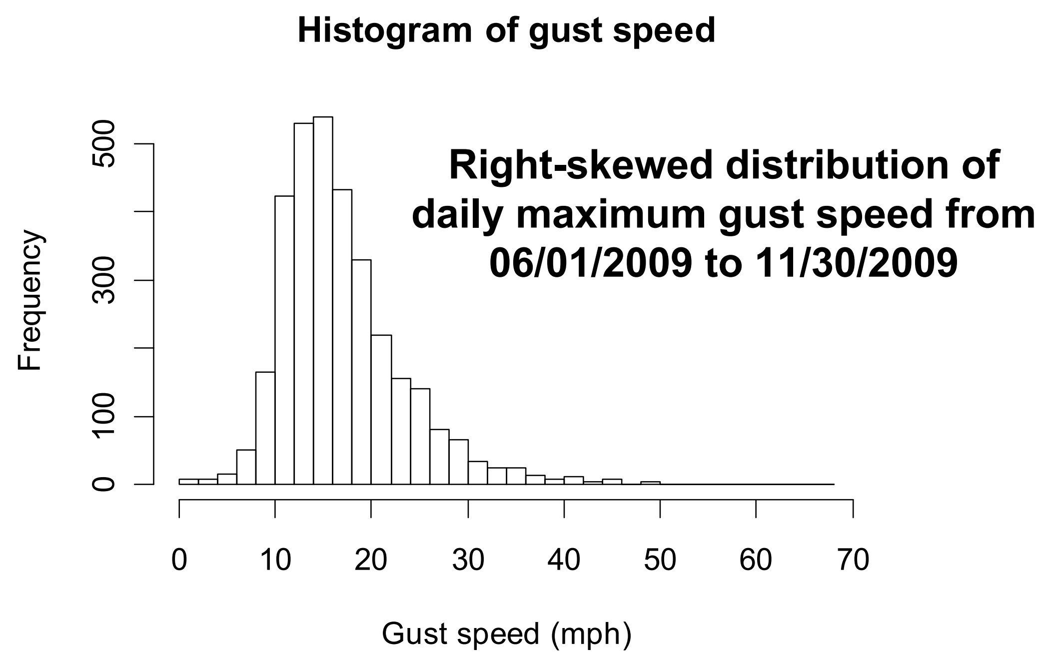
Weather Model Configuration: WRF-ARW (V3.1.1)

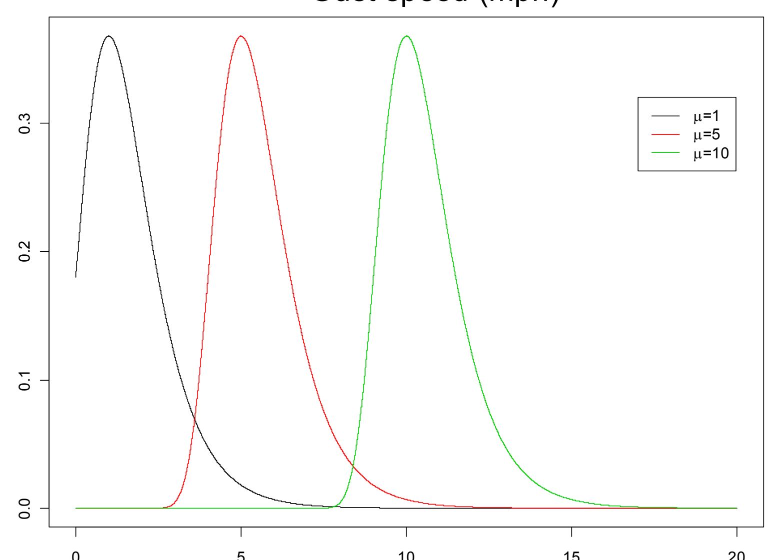
- 18/6/2 km nested (76x76x42) with 2 km resolution across entire extended service area for 84 hours
- Run twice daily (0 and 12 UTC)
- NAM for background and boundary conditions
- Physics configuration for highly urbanized to rural domain
  - WSM 6-class microphysics (includes explicit ice, snow and graupel)
  - Yonsei University non-local-K scheme with explicit entrainment layer and parabolic K profile in the unstable mixed layer for the PBL
  - NOAH land-surface modeling with soil temperature and moisture in four layers, fractional snow cover and frozen soil physics
  - Grell-Devenyi ensemble cumulus parameterization
  - 3-category urban canopy model with surface effects for roofs, walls, and streets



Modelling Extreme Values

- Distribution of daily maximum gust speed shows a highly right skewed tail which indicates a Gaussian distribution assumption does not hold
- Generalized extreme value (GEV) distribution have three parameters, μ, σ, ξ, which controls the location, scale and shape of a distribution, respectively
- Use a GEV distribution to model daily maximum gust speed given a daily maximum wind forecast, while location parameter and scale parameter are spatially correlated

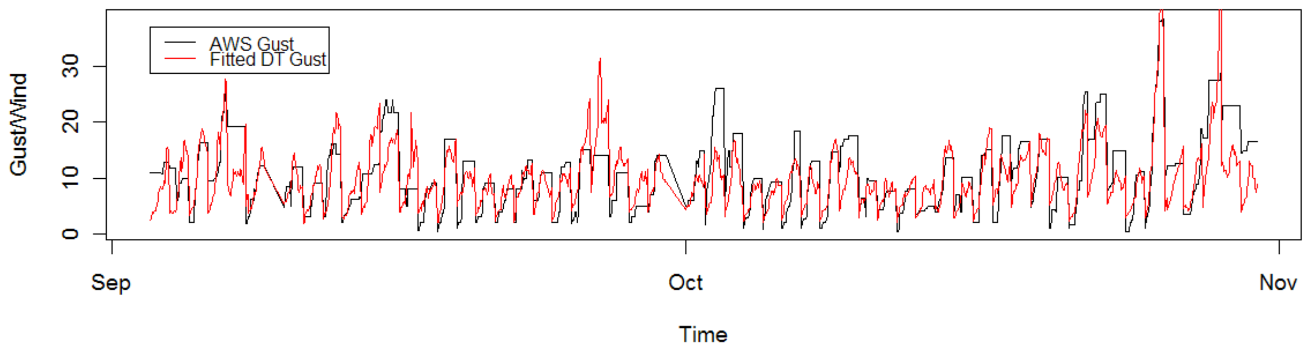


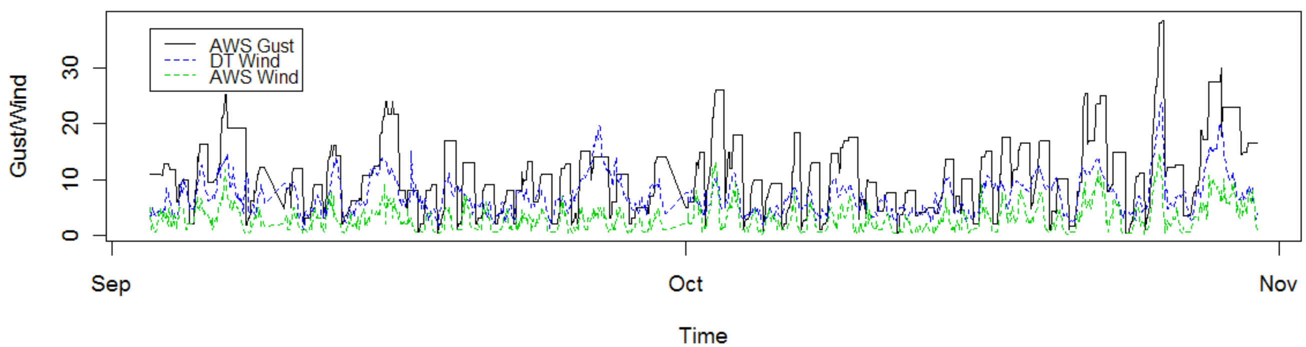


Example of generalized extreme value distribution with location parameter varied

Issues with Wind Data and Example Results

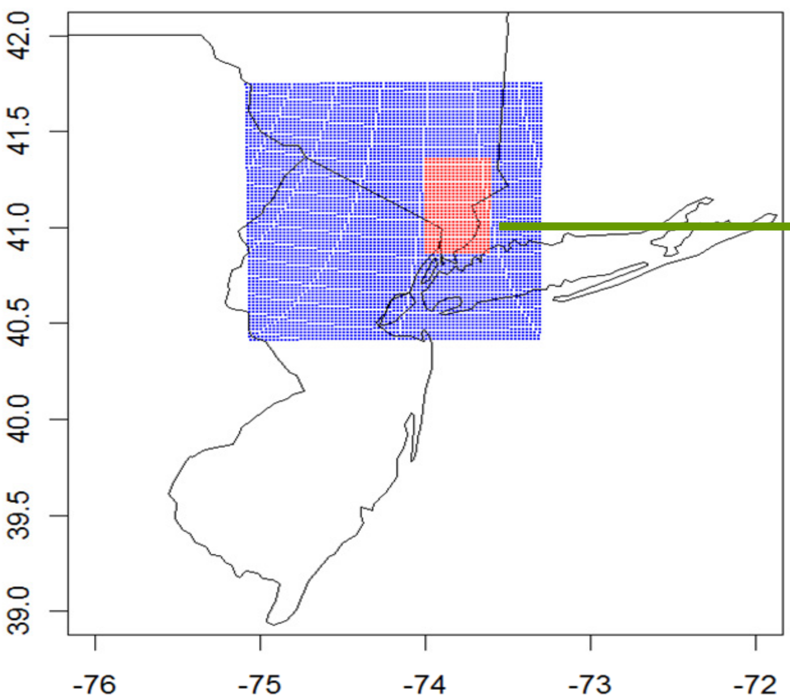
- Gusts are measured as maximum wind speed in 2 seconds
- Sustained wind is averaged wind speed in 2 minutes, recorded as instantaneous
- Gust speed is updated only if the current 5-minute gust speed value exceeds the previous value and will be set as 0 at midnight of each day
- Therefore, gust speed is aggregated to daily maximum for modelling

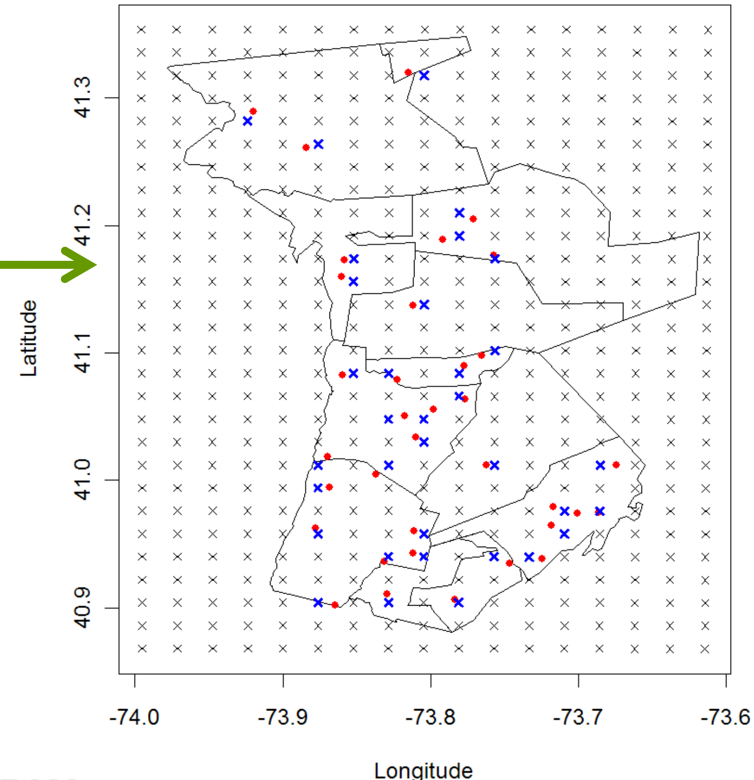




Example Observed and Modelled Wind and Gust Speed (mph) for One Weather Station

Initial Data Preparation





- The blue mesh shows the 2 km resolution WRF-ARW nest
- The red mesh shows the subset, which covers the utility company's service area and is used for the post-processing model

- WRF-ARW output:
  - 28 x 17 grids, covers the entire service area
  - 06/01/2009 to 10/30/2009 for post-processing model
- Weather observations (WeatherBug stations at red dots):
  - 06/01/2009 to 10/30/2009
  - Aggregate weather observations to hourly maximum values

Modelling Process

Model set-up:

Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be the set of locations of AWS weather stations where gust observations are recorded

Let $\mathcal{U} = \{u_1, \ldots, u_m\}$ be the set of DT forecast grids

- AWS daily maximum gust speed: $Y_d(s_i), i = 1, \ldots, n$
- DT daily maximum wind forecast: $X_d(u_j), j = 1, \ldots, m$

The goal is to obtain $\mathbf{Y_d}(\mathcal{U}) | \mathbf{X_d}(\mathcal{S})$

Bayesian Hierarchical Modeling Data Process:

$$f(Y_d(u)) = \frac{1}{\sigma(u)} \left[1 + \xi \left(\frac{Y_d(u) - \mu(u)}{\sigma(u)}\right)\right]^{-1/\xi - 1} \exp\left\{- \left[1 + \xi \left(\frac{Y_d(u) - \mu(u)}{\sigma(u)}\right)\right]^{-1/\xi}\right\}$$

where $1 + \xi \left(\frac{Y_d(u) - \mu(u)}{\sigma(u)}\right) > 0$

Location parameter and scale parameter are location dependent

Bayesian Hierarchical Modeling Latent Process:

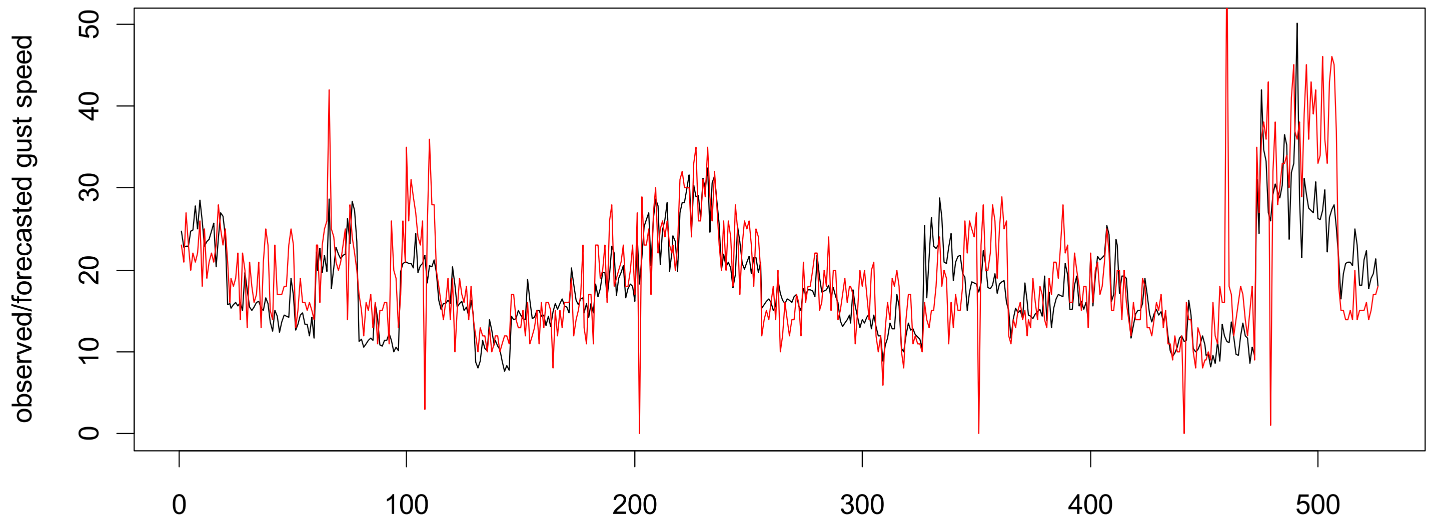
$$\mu(\mathcal{U}) = g_\mu \left(X(\mathcal{S}), \boldsymbol{\alpha}, \omega_\mu\right)$$
$$\sigma(\mathcal{U}) = g_\sigma \left(X(\mathcal{S}), \boldsymbol{\beta}, \omega_\sigma\right)$$

$$\boldsymbol{\alpha} \sim Gau(\mathbf{a}, \boldsymbol{\Sigma}_\alpha(\rho, \tau_\alpha))$$
$$\boldsymbol{\beta} \sim Gau(\mathbf{b}, \boldsymbol{\Sigma}_\beta(\rho, \tau_\beta))$$

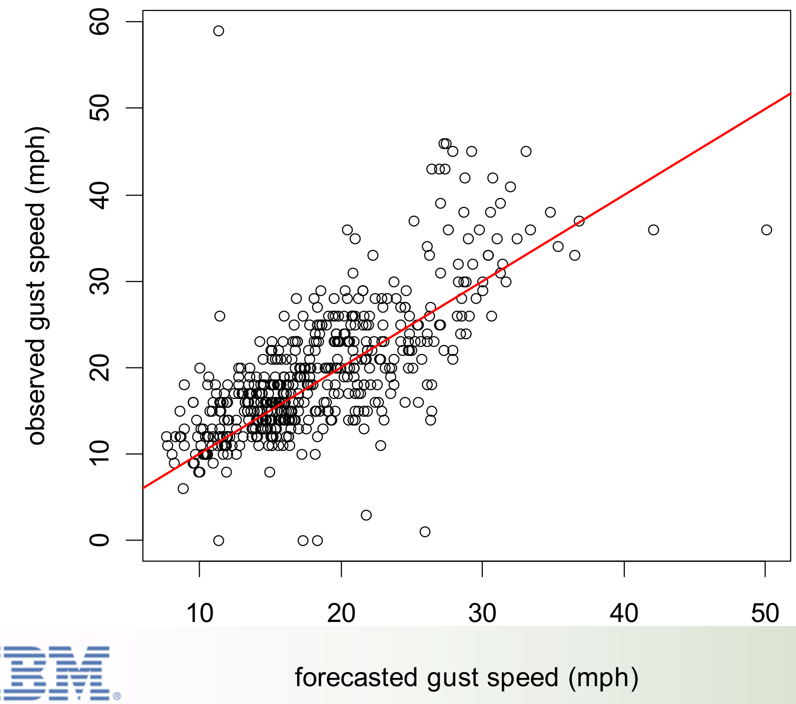
- **Prior set-up and derive posterior distributions of the parameters**
- **Use Markov Chain Monte Carlo (MCMC) to draw posterior samples and stop after convergence is achieved**

Sample Modelling Results

- Initial version uses data from 06/01/2009 to 10/30/2009 to model
- Data from 11/01/2009 to 11/30/2009 to forecast for verification



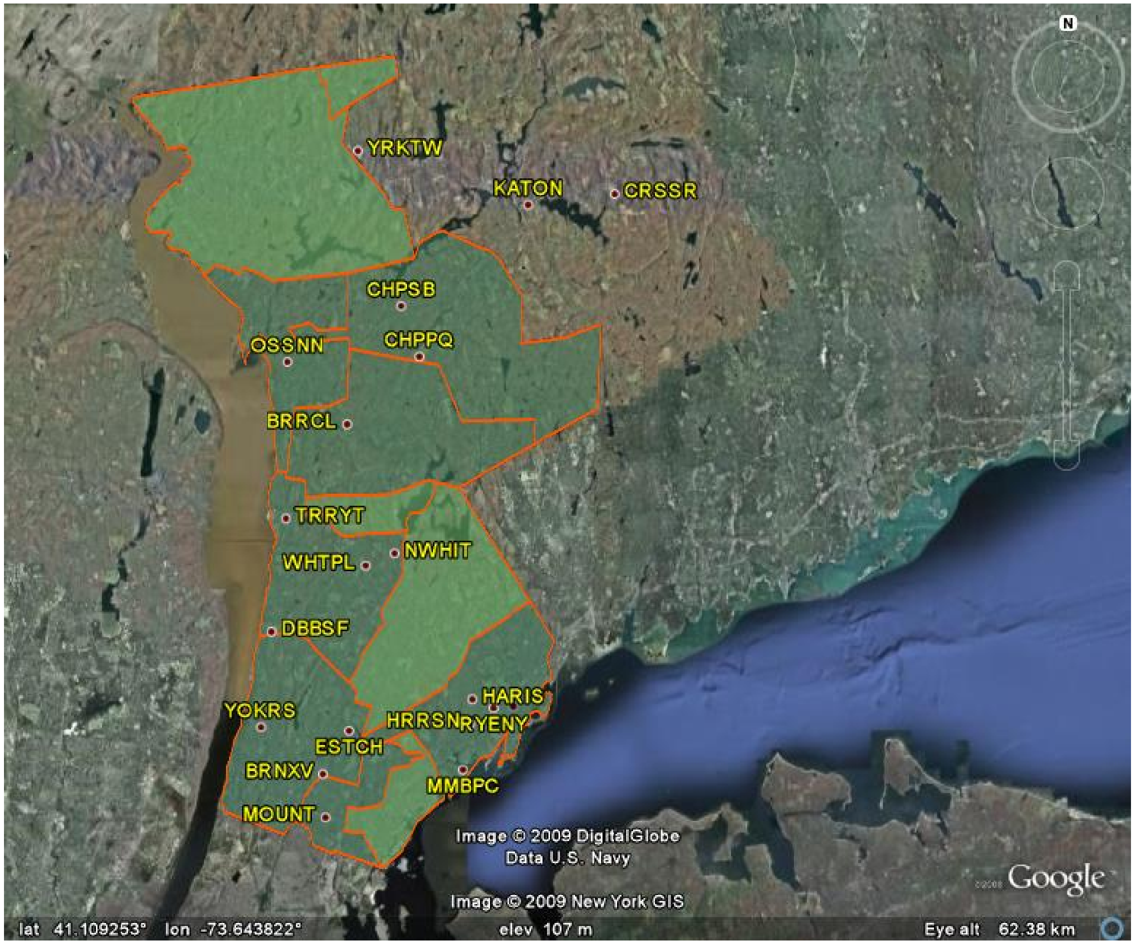
- Comparison of forecasted and observed daily maximum gust speed
- Black curve is the forecasted values and the red curve is the observed values



- Forecasted vs. observed daily maximum gust speed for all of November 2009
- The points are lined up with 45 degree line (red solid line)

Outage/Damage Model Configuration

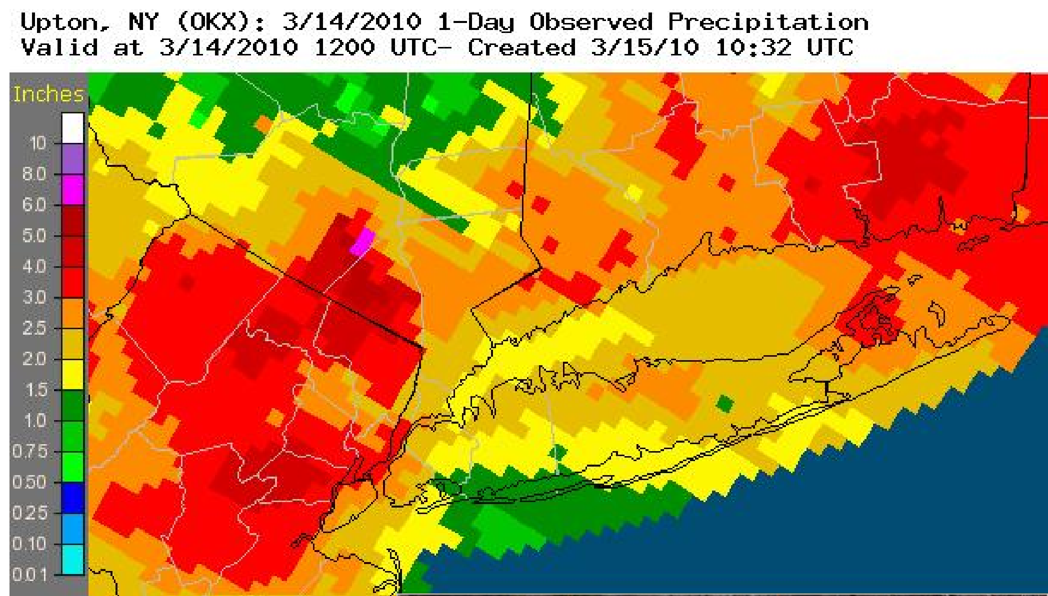
- Post-processor for weather data
  - "Predictive mode" operates on WRF output and generates daily gust maximums
  - "Probable mode" operates on near-real-time WeatherBug data (i.e., nowcasting)
- Produces estimate of number of jobs per substation and gusts
- Trained and calibrated with updated weather and outage data
- Dissemination
  - Tailored weather visualizations available via a web browser, which are automatically updated for each forecast cycle
  - E-mail alert system, triggered by total number of jobs with 75% confidence
- Uncertainty quantification
  - Multiple sources



Location of Consolidated Edison Westchester County Substation Areas and AWS/WeatherBug Stations

Example Event: 13 March 2010 Rain and Wind Storm (NYC)

- Coastal storm with strong winds and heavy rains
- Gusting between 40 and 75 mph observed in the afternoon and evening
- Innumerable downed trees and power lines
- Local flooding and evacuations
- Electricity service lost to over 600,000 residences and businesses in the New York City metropolitan area
- Widespread disruption of transportation systems (e.g., road and bridge closures, airport and rail delays)
- Other forecasts during the late morning on 11 March: "rain may be heavy at times, east winds 20 to 25 mph with gusts up to 40 mph"
- Wind advisories issued (gusts to 45-50 mph) at 1619 EST, 12 March
- High wind warnings issued (gusts to 55-60 mph) at 1343 EST, 13 March

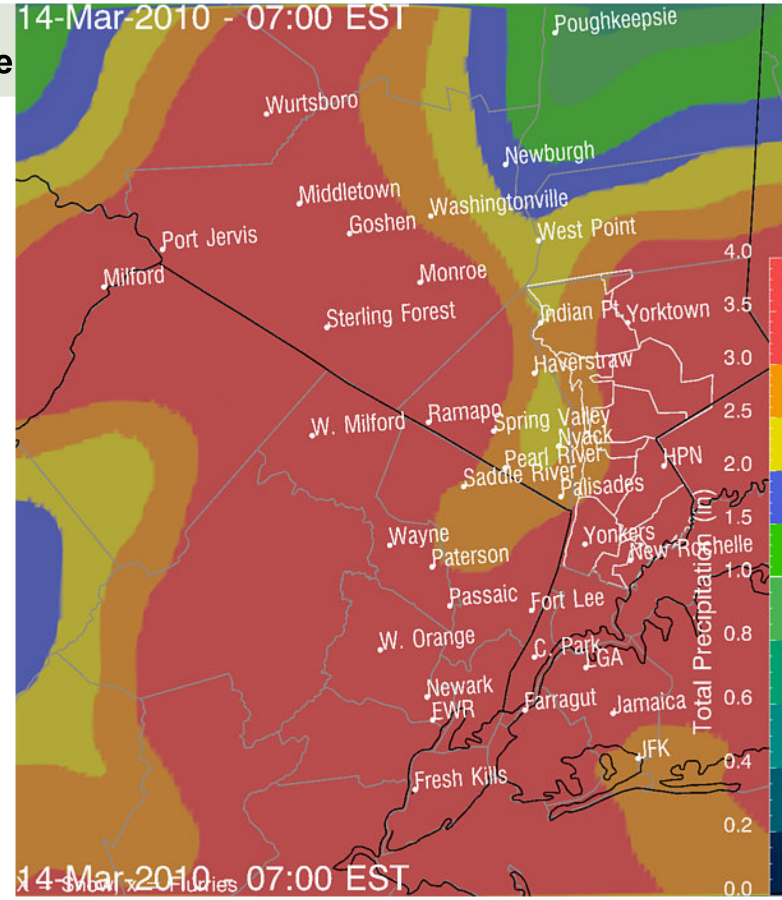


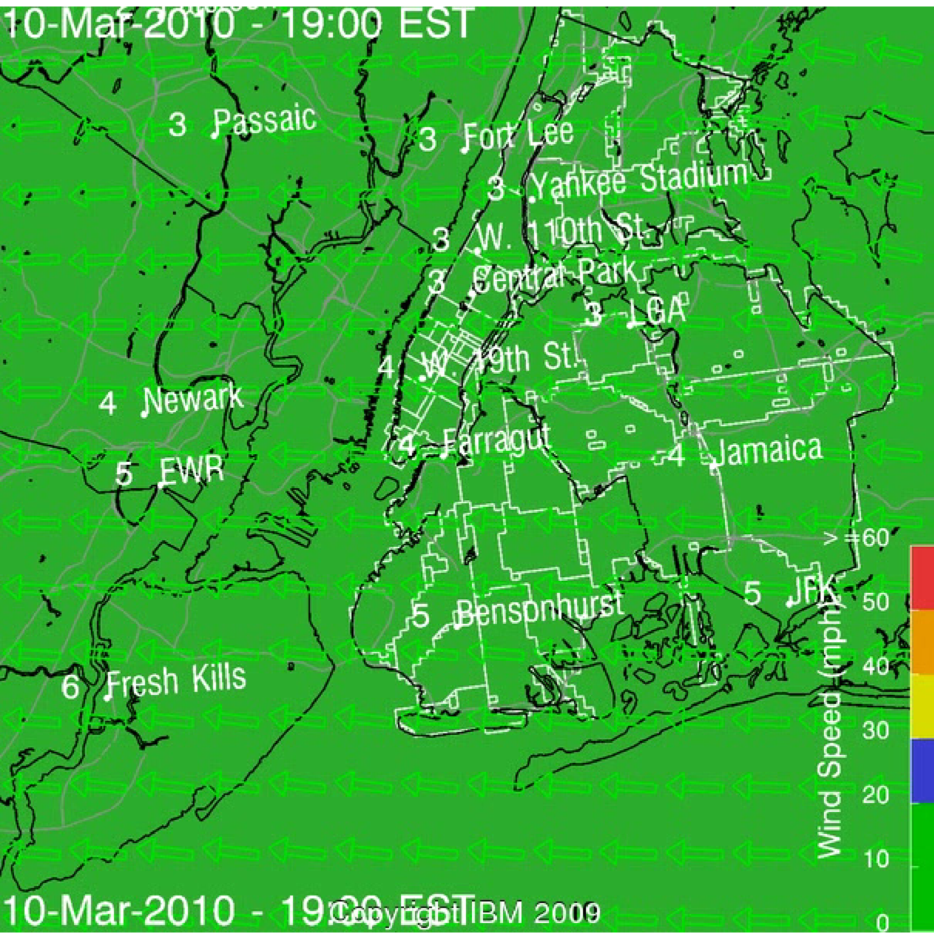
13 March 2010 Nor'easter

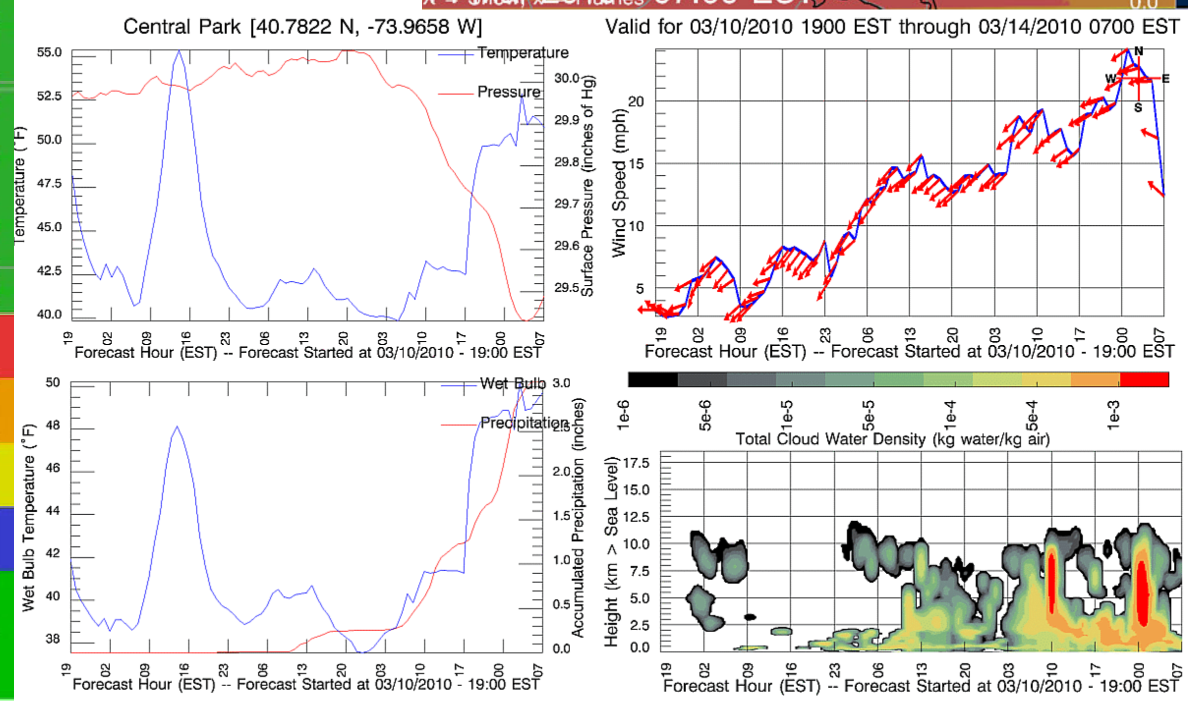
Deep Thunder Weather Forecast

Initiated with data from 00 UTC 11 March with results available ~08 UTC

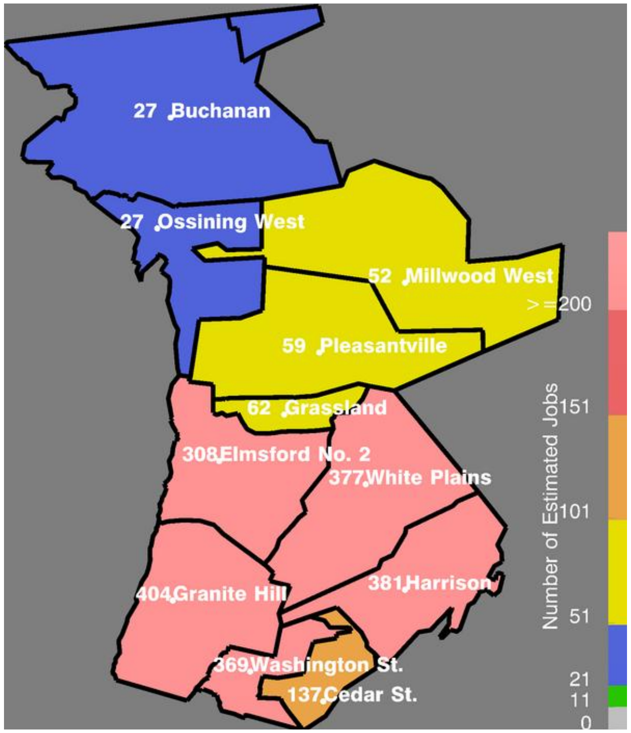
High winds shown in forecast available more than two days before the event and 37 hours before the National Weather Service advisory





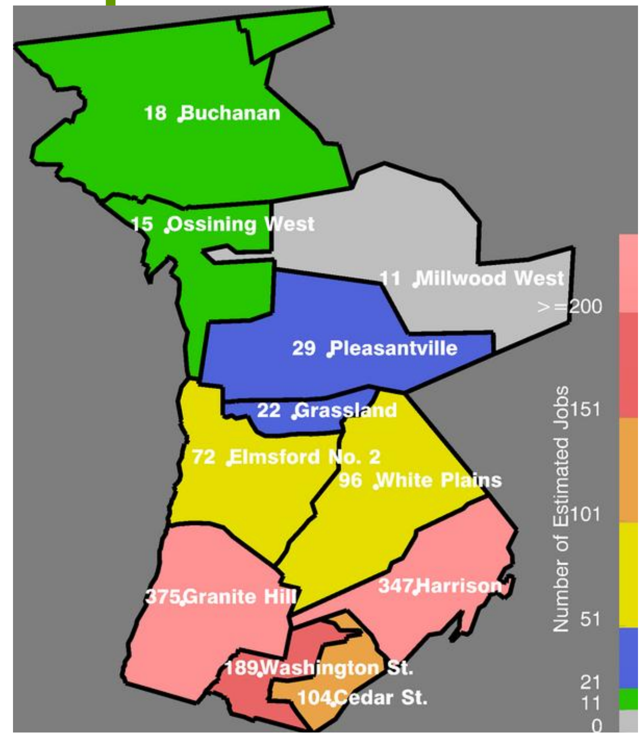


13 March 2010 Nor’easter Deep Thunder Impact Forecast



Actual Outages per Substation (Repair Jobs)

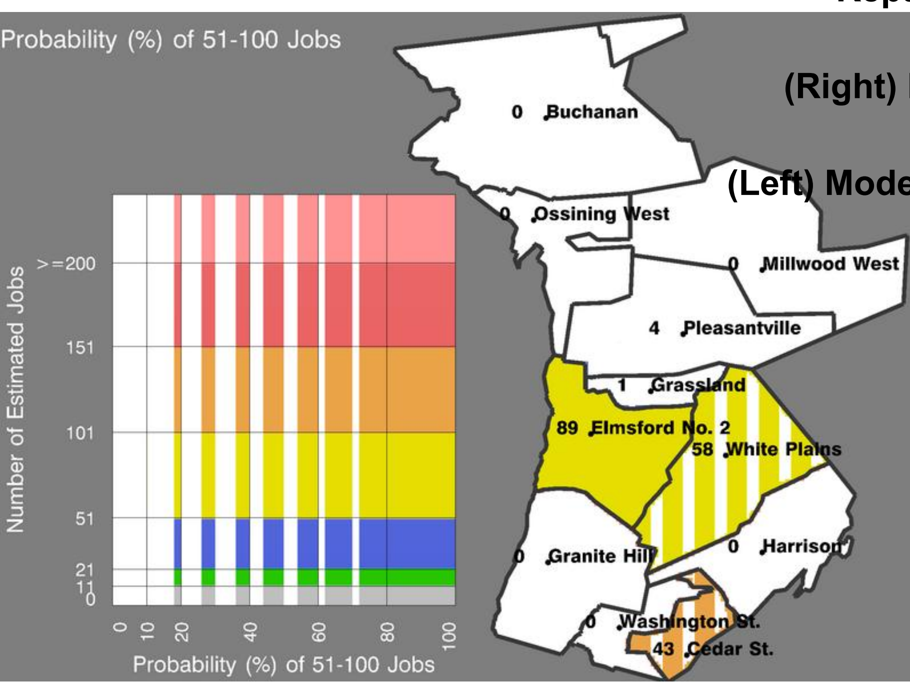
Estimated Outages per Substation (Repair Jobs)

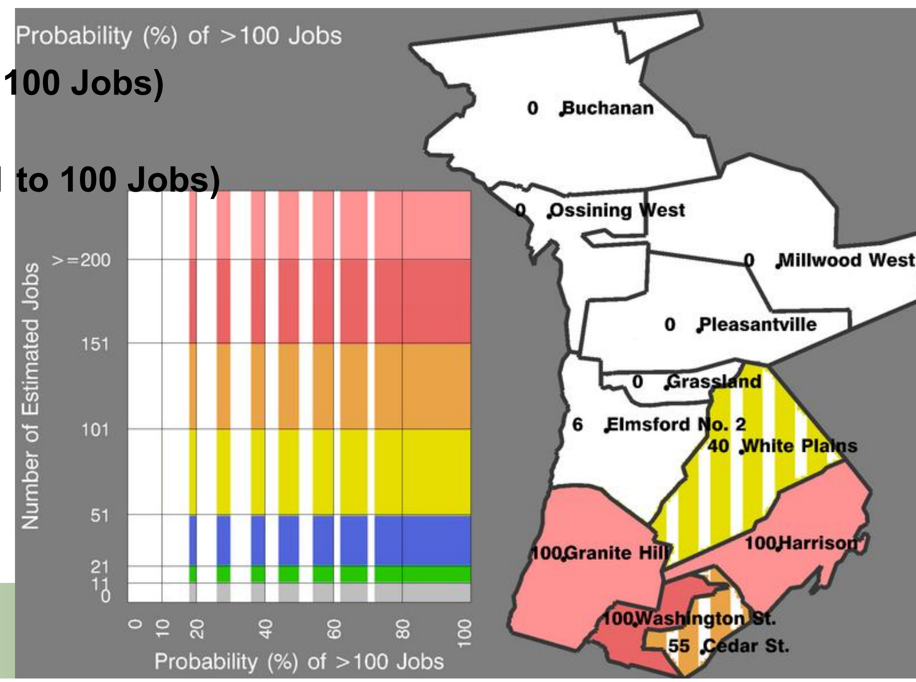


Likelihood (Probability) of a Range of Repair Jobs per Substation

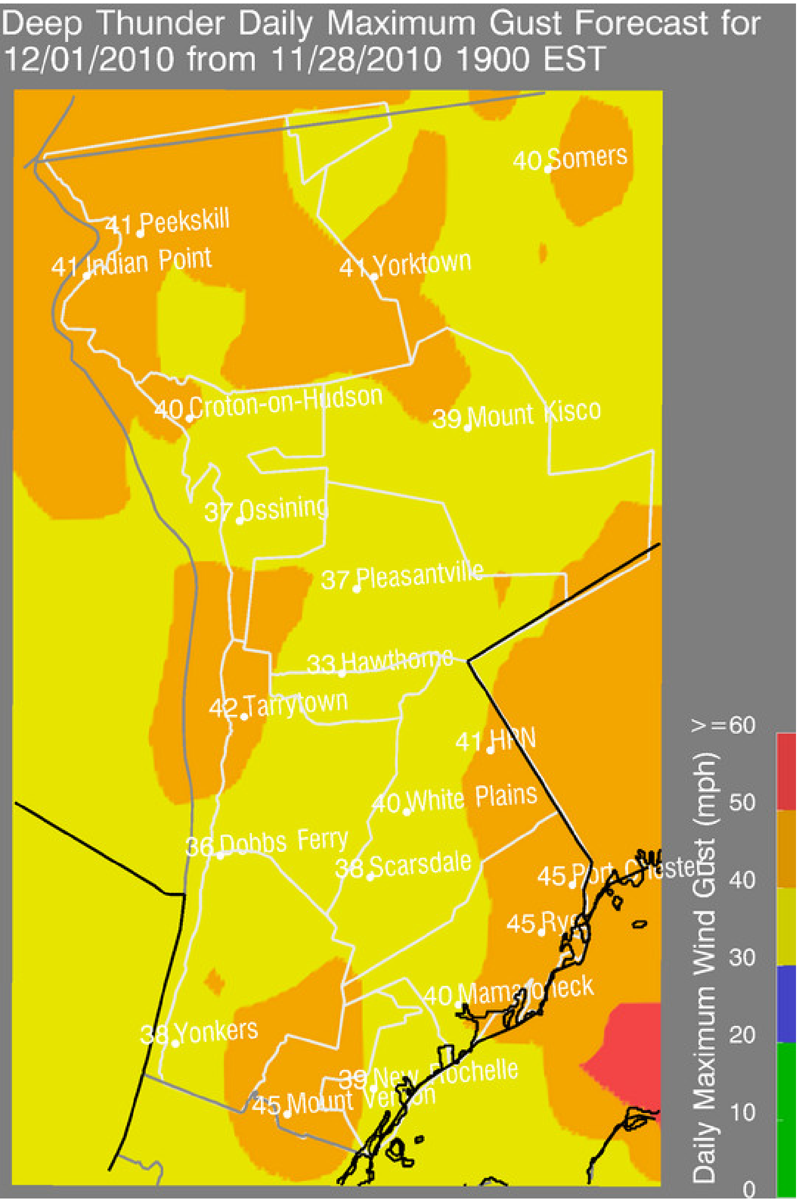
(Right) High Severity (> 100 Jobs)

(Left) Moderate Severity (51 to 100 Jobs)



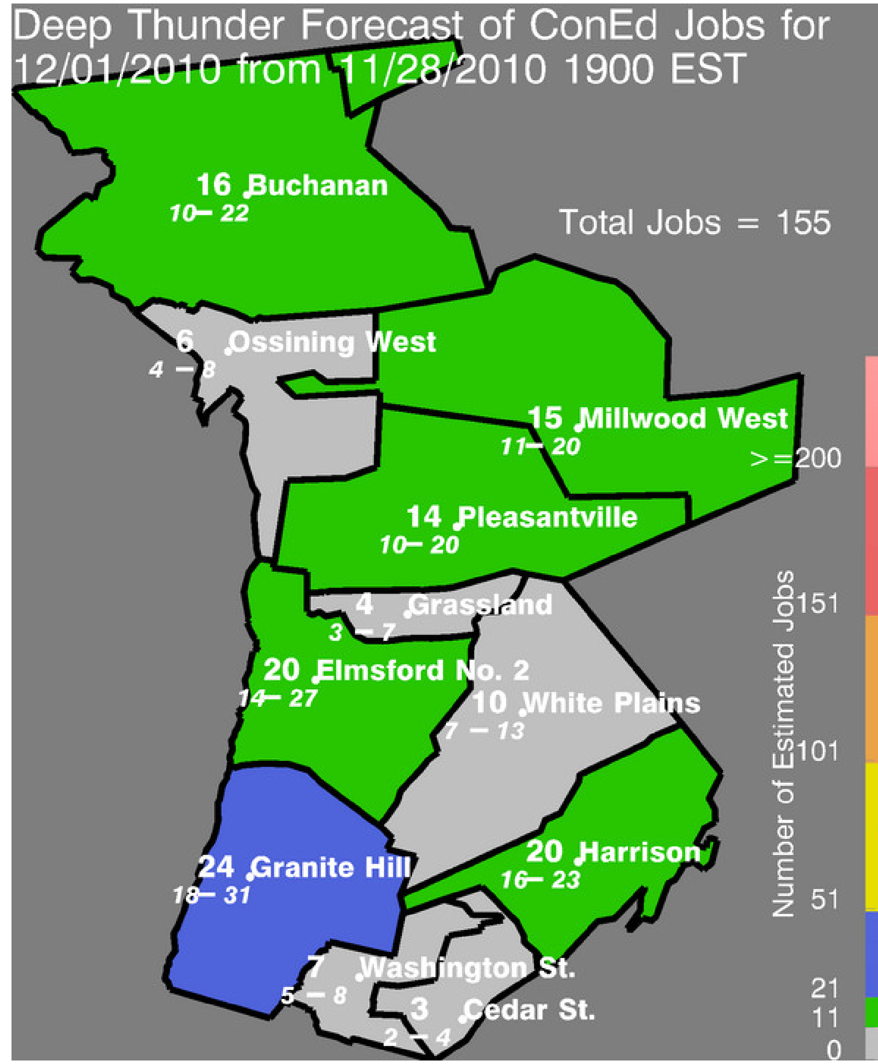


1 December 2010 Deep Thunder Impact Forecast



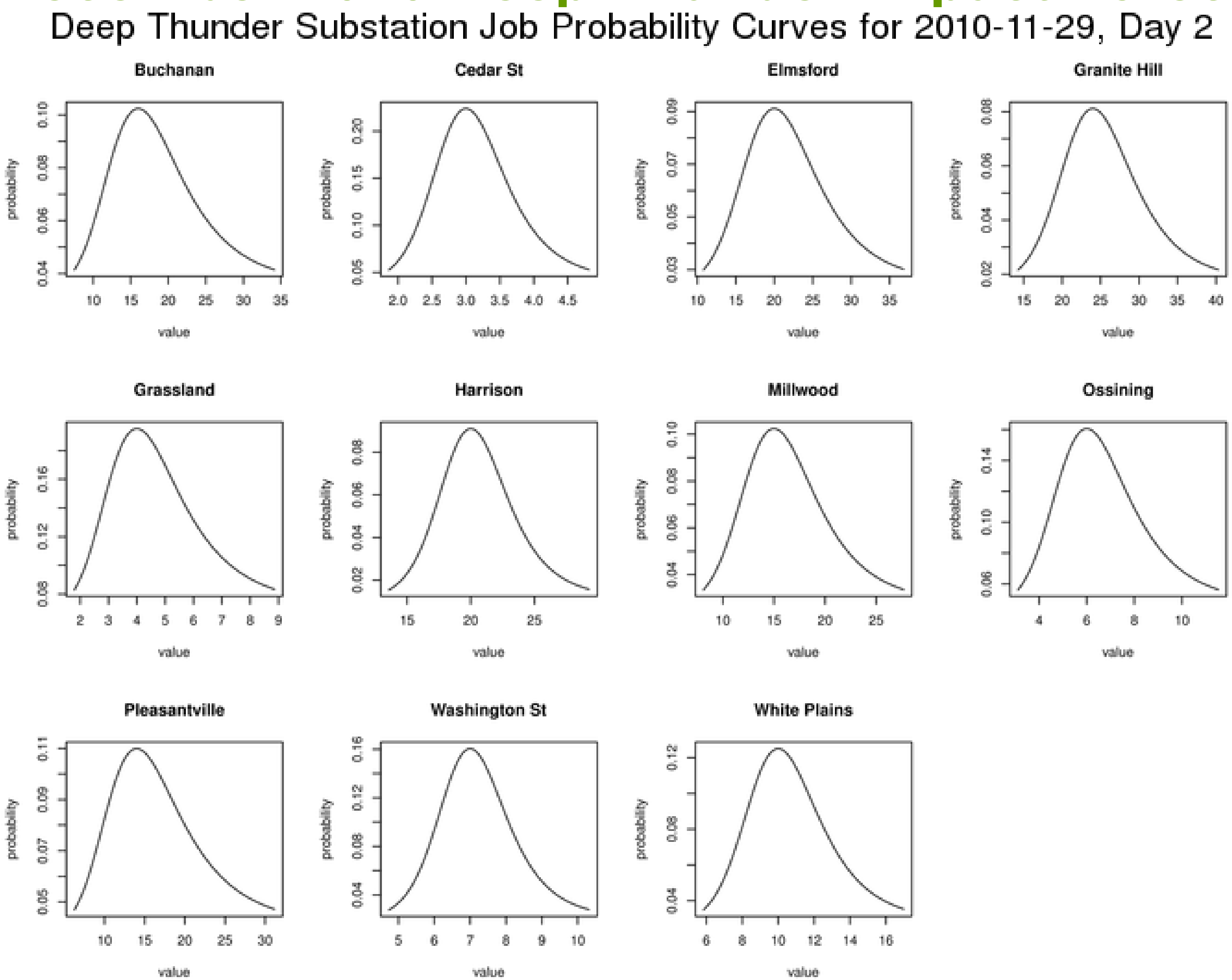
Estimated Daily Wind Gust Speed Maximum

Initiated with data from 00 UTC 29 November with results available ~08 UTC



Estimated Outages per Substation (Repair Jobs)

1 December 2010 Deep Thunder Impact Forecast



Initiated with data from 00 UTC 29 November with results available ~08 UTC

On-Going Work and Challenges

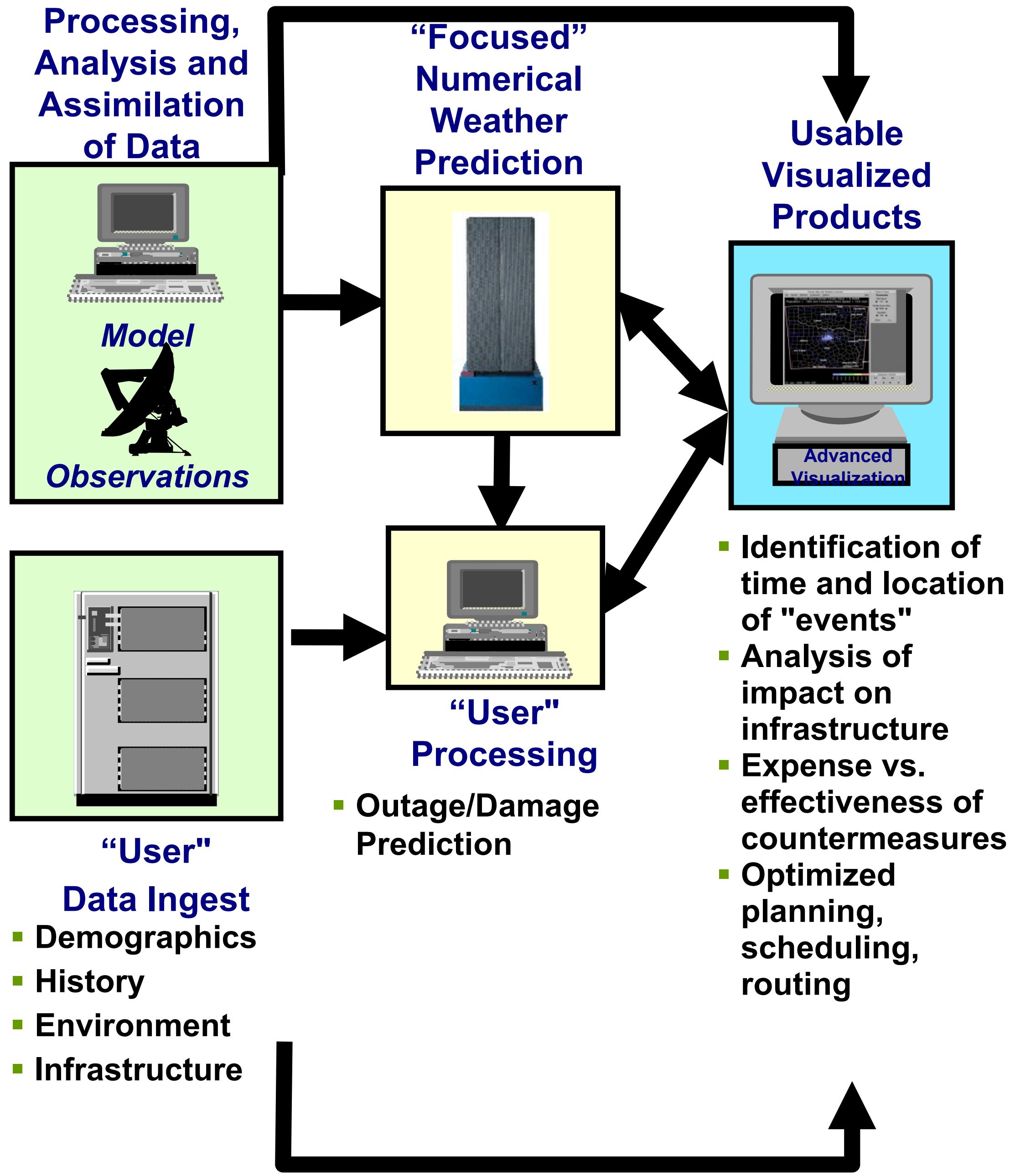
- **Quality of weather observations**
  - Relatively dense network of surface stations from WeatherBug
  - Reporting inconsistent from WeatherBug network
- **Quality of outage (job ticket) data**
  - Must rely on field crews and service representatives
  - Need to filter storm-related damage
- **Continued calibration and training**
  - Determine appropriate inputs (e.g., gusts, soil moisture, foliage, etc.) and their relative correlation for new and old events
  - Damage-driven vs. weather-driven emphasis
  - Changes in infrastructure and environment
  - Incorporation of “Black Swan” events
  - Insufficient events to adequately evaluate
- **Utilization**
  - Need to build trust with diverse users
  - Deliver complex information succinctly (e.g., visualization of probabilistic data)
  - Must be integrated with utility company procedures
  - Development of effective verification methods

Backup

Slides

Challenges of Coupling NWP to the Decision Making Process

- **Damage forecast model inputs**
  - Which weather data really matter (avoid multicollinearity)?
  - For example, gust speed has a stronger relationship to damages vs. wind speed
- **Weather forecast calibration**
  - Forecasted variables (e.g., wind speed) may differ in meaning vs. observations used in the damage-forecast-model training
  - How should physical model outputs be calibrated so that they can be used as the inputs of damage forecast model?
- **How should damage forecasts, multiple spatial resolution interpolations and calibration be integrated in one framework?**



Challenges of Building a Damage Model

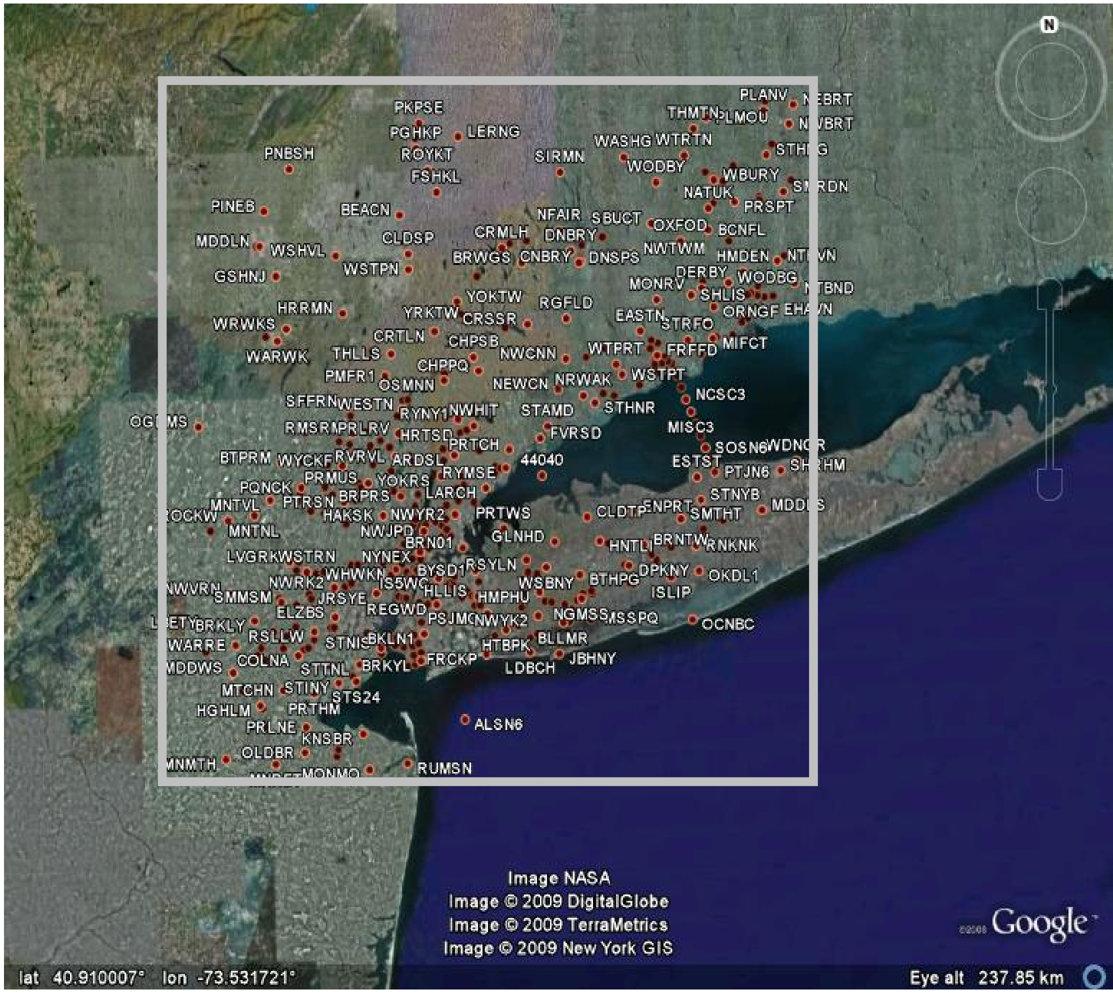
- **Damage forecast model inputs**
  – Which weather inputs are important for damage forecast?
  – Most weather variables are correlated
  – Multicollinearity may cause invalid interpretation of weather predictors
- **Weather forecast calibration**
  – Forecasted variables (e.g., wind speed) may differ in meaning vs. observations used in the damage-forecast-model training
  – How should physical model outputs be calibrated so that they can be used as the inputs of damage forecast model?
- **Gust speed calculation**
  – Exploratory data analysis indicates that gust speed has a stronger relationship to damages vs. wind speed
  – General meteorological models do not provide a direct gust forecast
  – How should gust speed be calculated based on limited weather data?
- **Model integration**
  – How should damage forecasts, multiple spatial resolution interpolations and calibration be integrated in one framework?
- **Uncertainty quantification and visualization**
  – Uncertainties come from various data sources

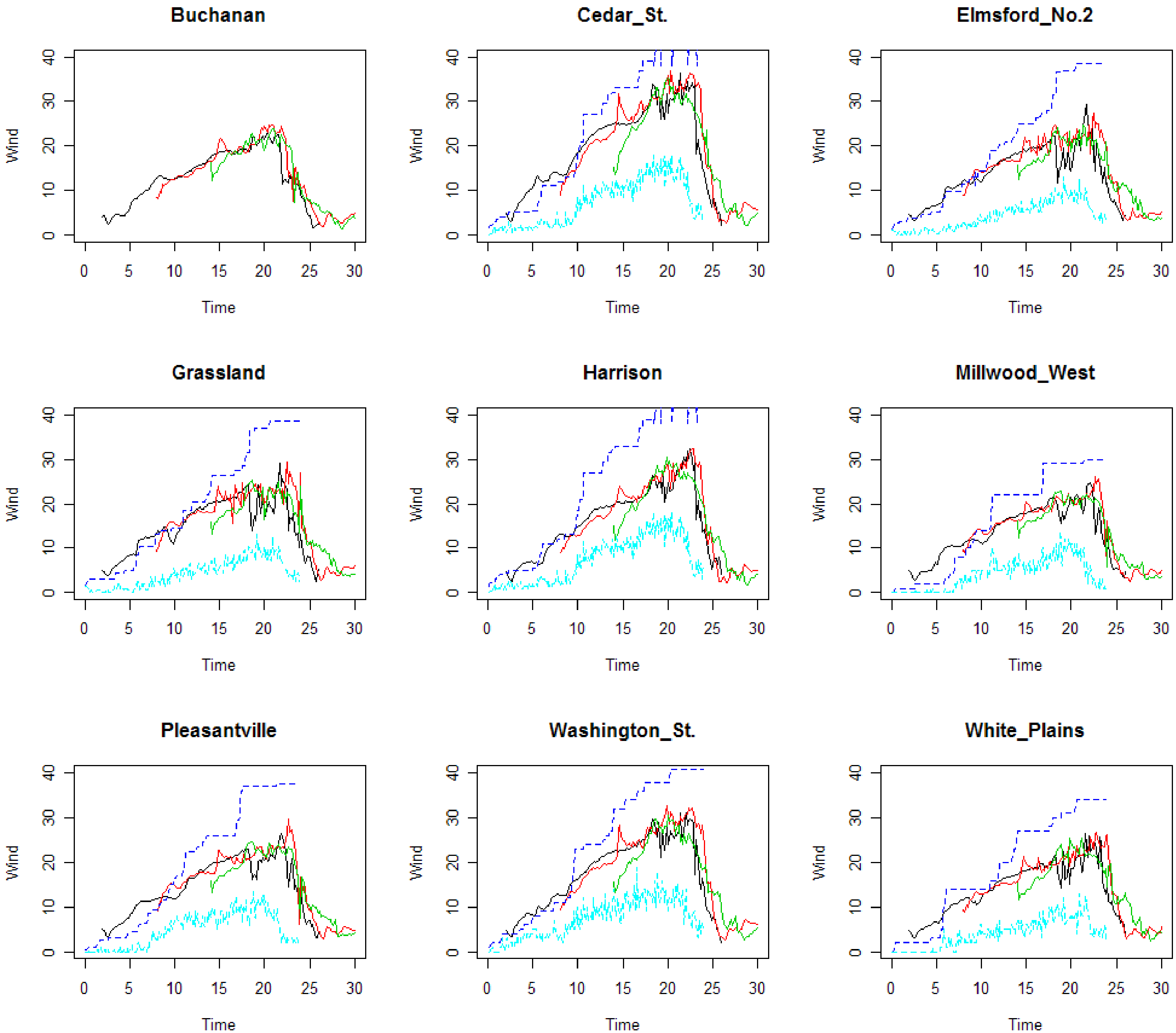
Data Description – Weather Observations

- Mesonet data, 2004~present (AWS/WeatherBug)
- Observations for every 5 minutes
- Around 20 weather stations located in utility company's service area, some of which do not produce regular, reliable data
- Weather variables:
  - Temperature, humidity, pressure, rainfall, wind speed, gust speed
- Gust vs. wind
  - Gust is measured as maximum wind speed in 2 seconds
  - Sustained wind is averaged wind speed in 2 minutes
  - Gust speed is updated only if the current 5 minute gust speed value exceeds the previous value and will be set as 0 at midnight of each day
  - Sustained wind is recorded as instantaneous values

AWS/WeatherBug Surface Observations

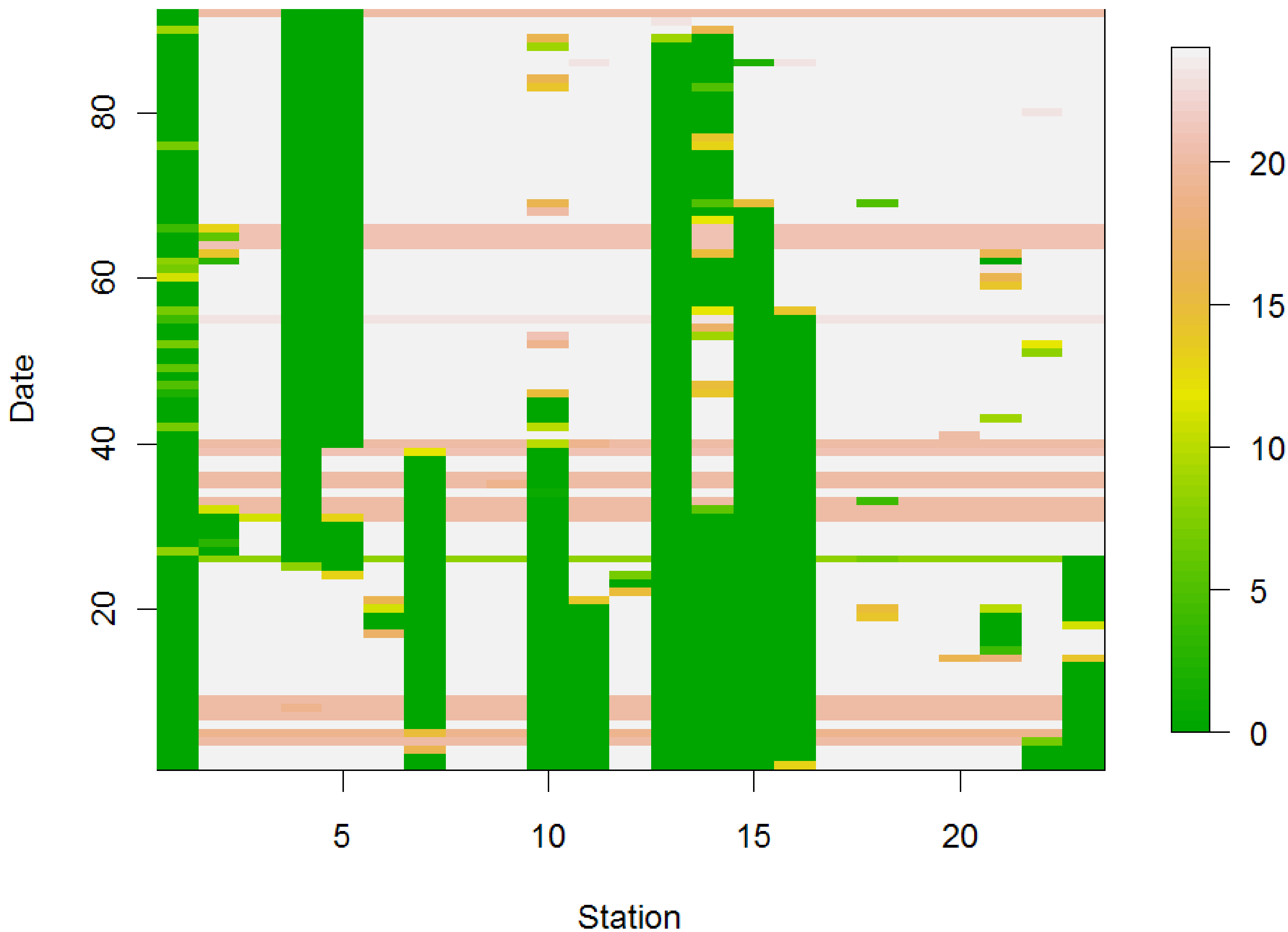
- More than 400 stations in model domain that covers extended ConEd service territory – close sampling to model resolution
- Primary data used include temperature, relative humidity, wind speed and direction, rainfall
- Real-time data used for damage assessment during severe weather
- Improve model initialization via data assimilation in near-real time
- Historical data used for retrospective analysis, forecast verification and tuning
- Data also used to calibrate model and other sensor data





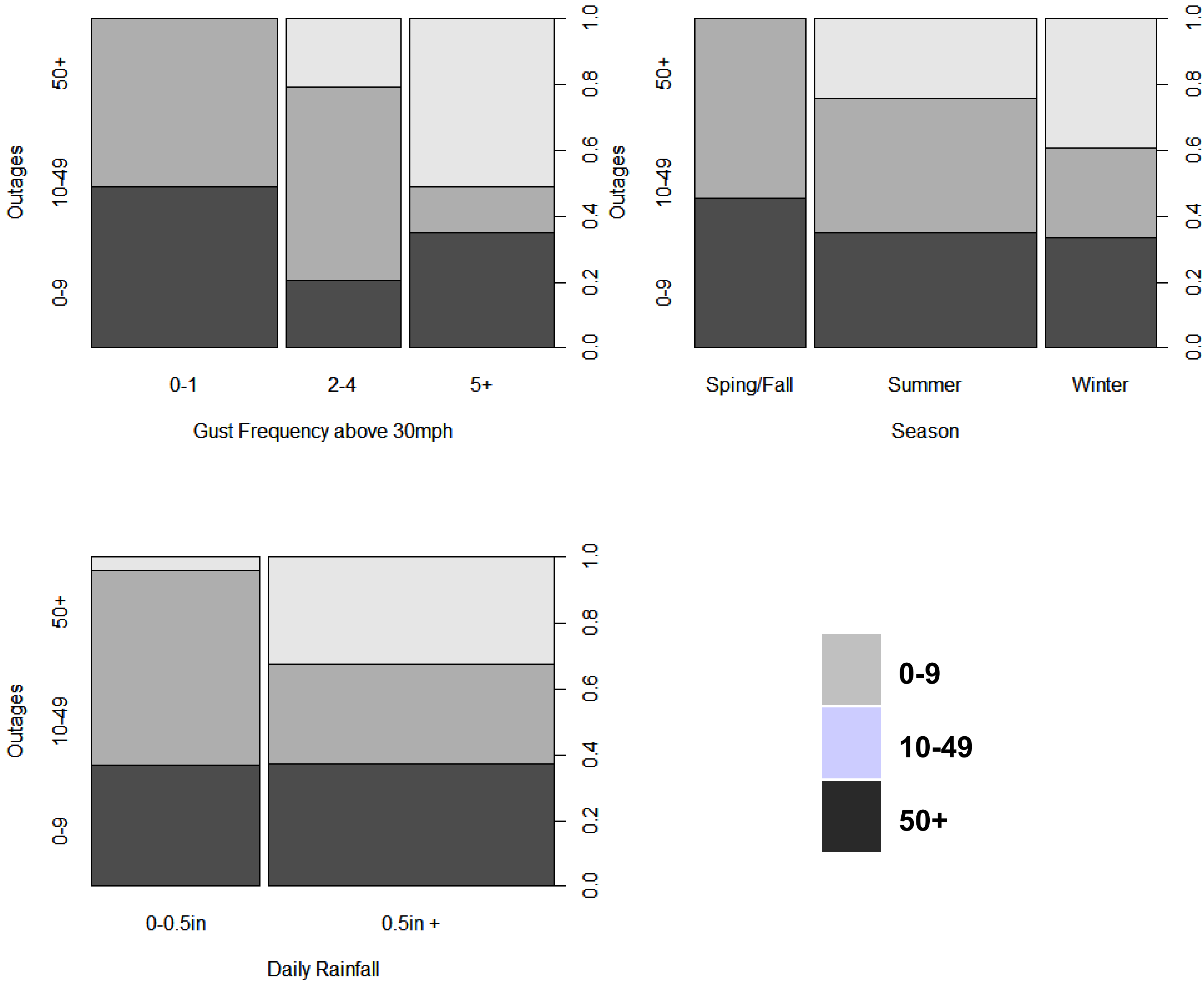
Comparison of Wind and Gust Observations and WRF-ARW Output

Example Data Availability

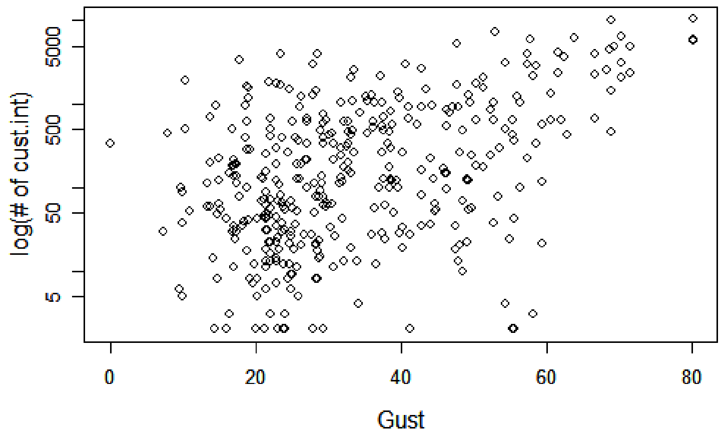


Wind Observations, August ~ October 2009, at 25 Weather Stations

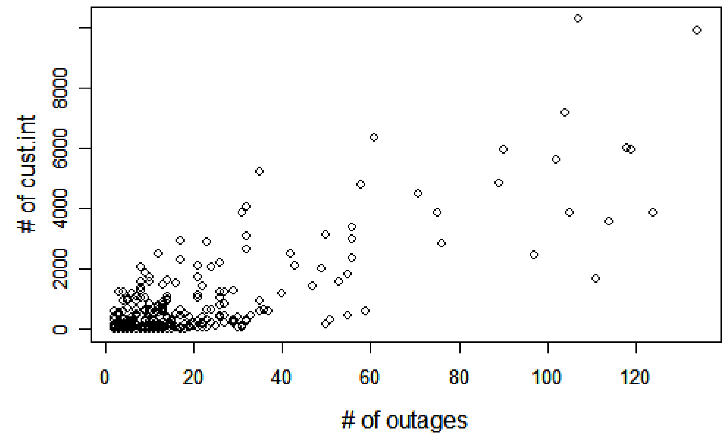
Examples of Correlation Between Historical Outage Data and Weather Observations



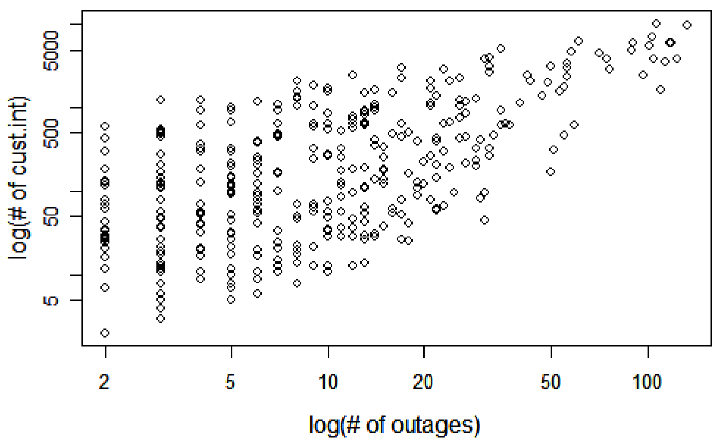
Modelling of Number of Customers Interrupted



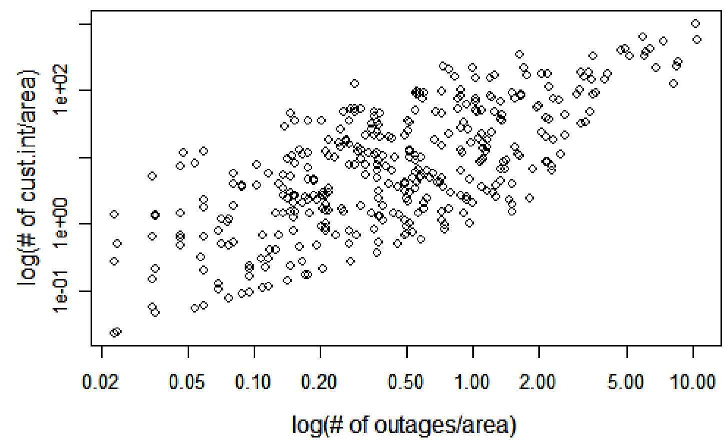
Before transformation



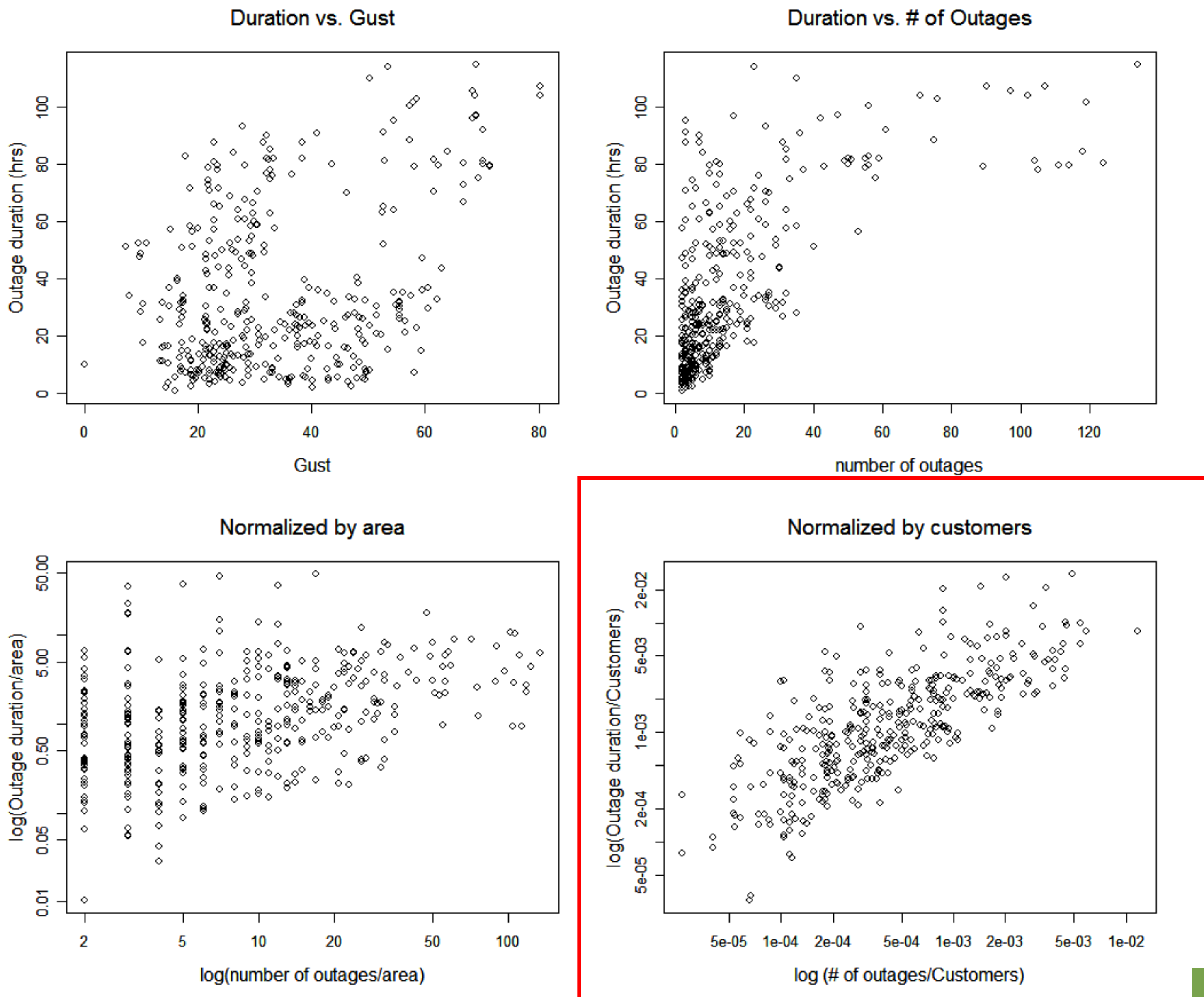
After transformation



After transformation and normalization



Outage Duration:
Number of Hours until 90% of Tickets Are Closed



Correlation of Adjusted Winds with Outages