Storm Damage Modeling at the University of Connecticut

Brian Hartman, PhD ASA

31 January 2014.0 (but I don't care, it doesn't matter)

Workshop on Insurance Mathematics

Quebec, QC

UCONN DAMAGE PREMO

DAMAGE PREDICTION MODEL

University of Connecticut

Acknowledgements

- The UConn DPM team
  - Manos Anagnostou
  - Marina Astitha
  - Maria Frediani
  - David Wanik
  - Xinxuan Zheng
  - Alex Baldenko
  - Tiran Chen
  - Jichao He
  - and many others . . .
- The Utilities
  - Northeast Utilities
    - CL&P
    - WMECO
    - NSTAR
  - Others soon to join

University of Connecticut

Outline

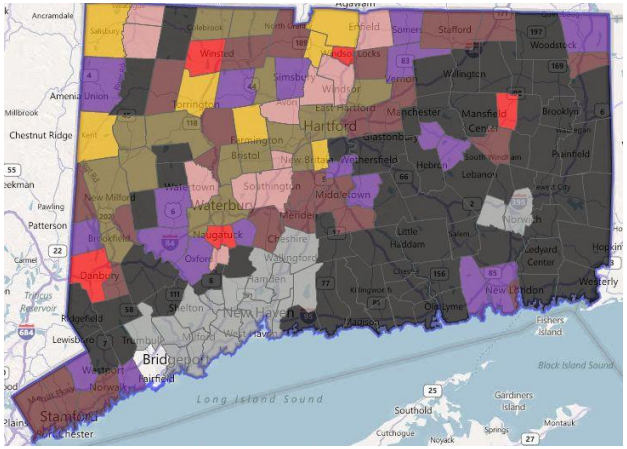
1. Introduction/Motivation of the Project
2. Data Description
3. Model Description
4. Case Studies
5. Current/Future Work

University of Connecticut

Motivation for the Project

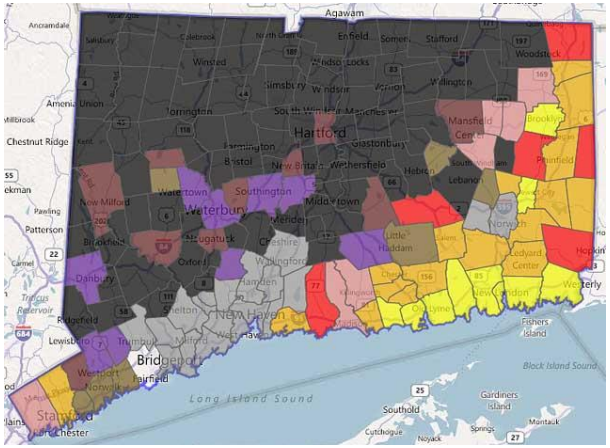
- Major recent CT storm events
  - Tropical Storm Irene (August 2011)
  - October Nor'easter (October 2011)
  - Hurricane Sandy (October 2012)
- Consequences
  - Massive long outages
  - Series of investigative reports (Witt Associates, Two Storm Panel, Davies Consultants)
  - CL&P President Jeff Butler resignation
  - Performance standards and penalties for electric restoration times (CT and MA Attorney General)

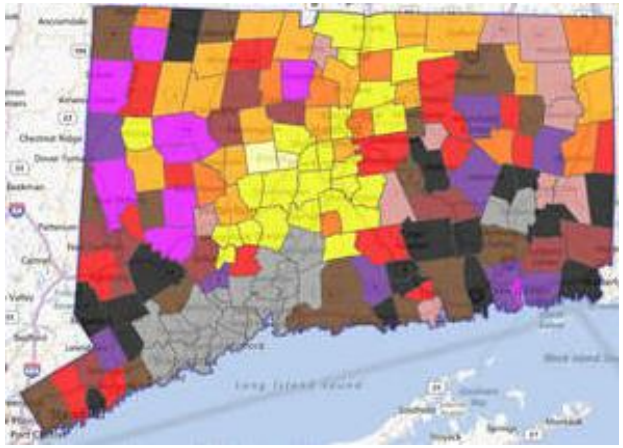
Comparison of Extreme Storm Outages



Storm Irene
671,000 affected
11 days

October Nor'easter
803,000 affected
9 days

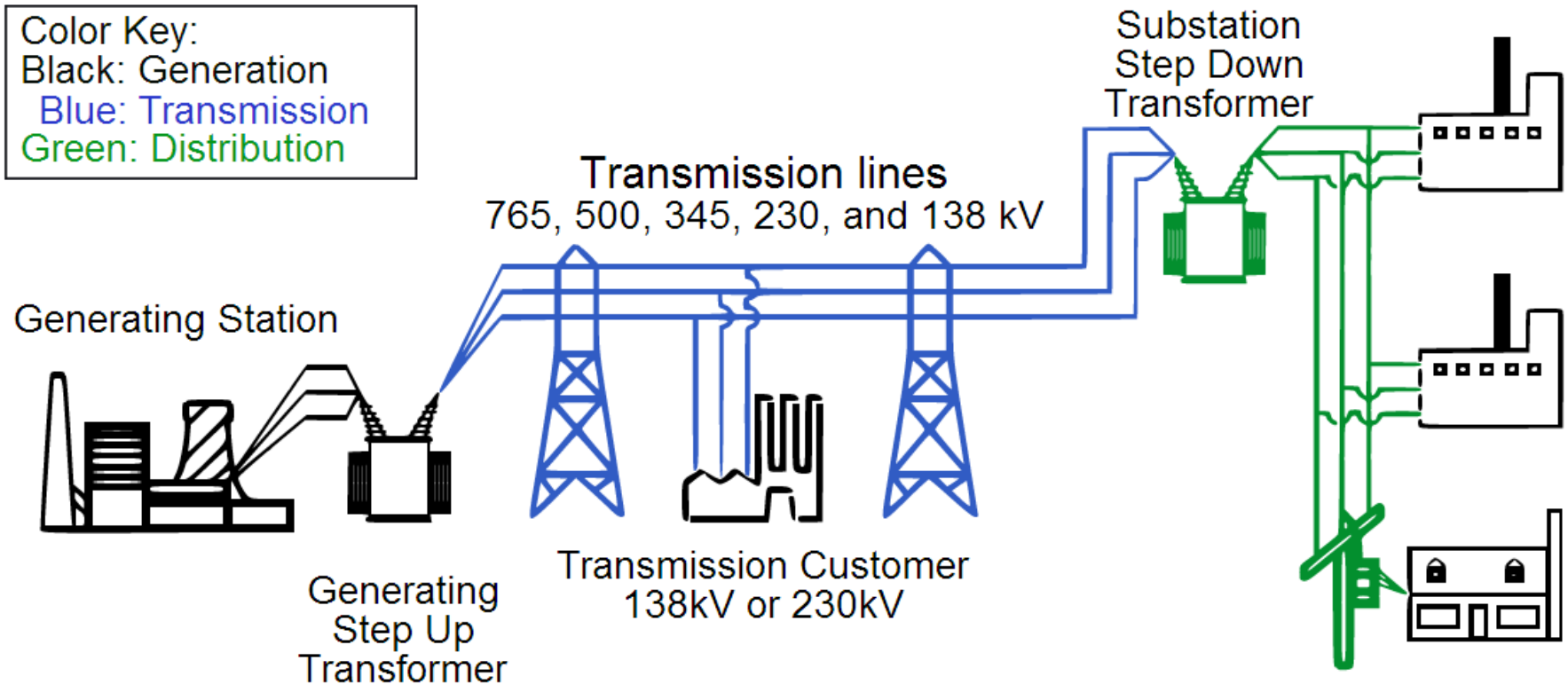




Hurricane Sandy
500,000 affected
10 days

University of Connecticut

Power Generation, Transmission, and Distribution



University of Connecticut

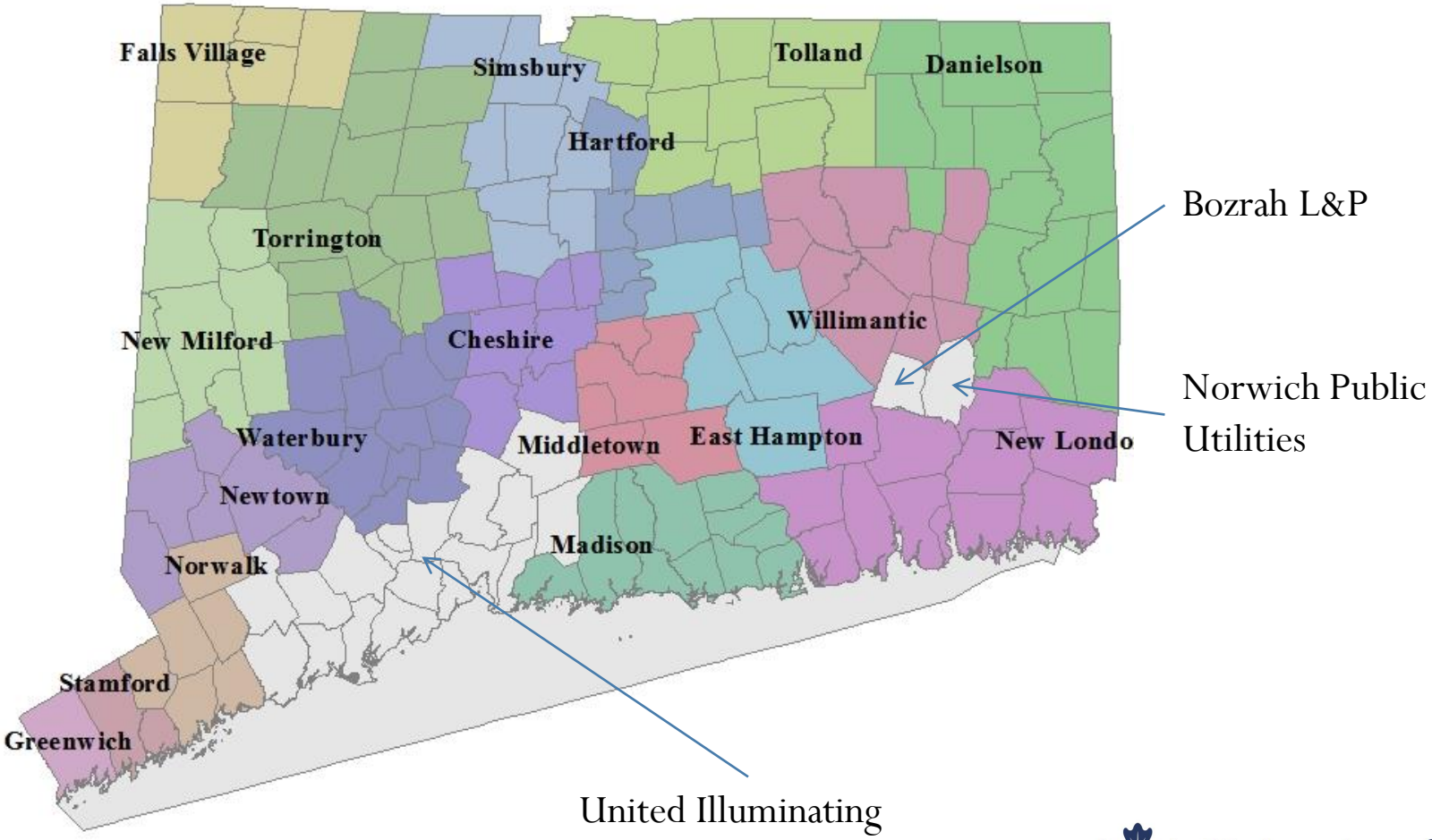
Data Description

University of Connecticut

What is a “trouble spot?”

- A trouble spot is an extended interruption of service (>5 minutes).
- Usually requires human intervention
- Recorded at the nearest isolating device (transformer, fuse, recloser, or switch)

CL&P Service Territories (1.2M customers)



2. Data Description

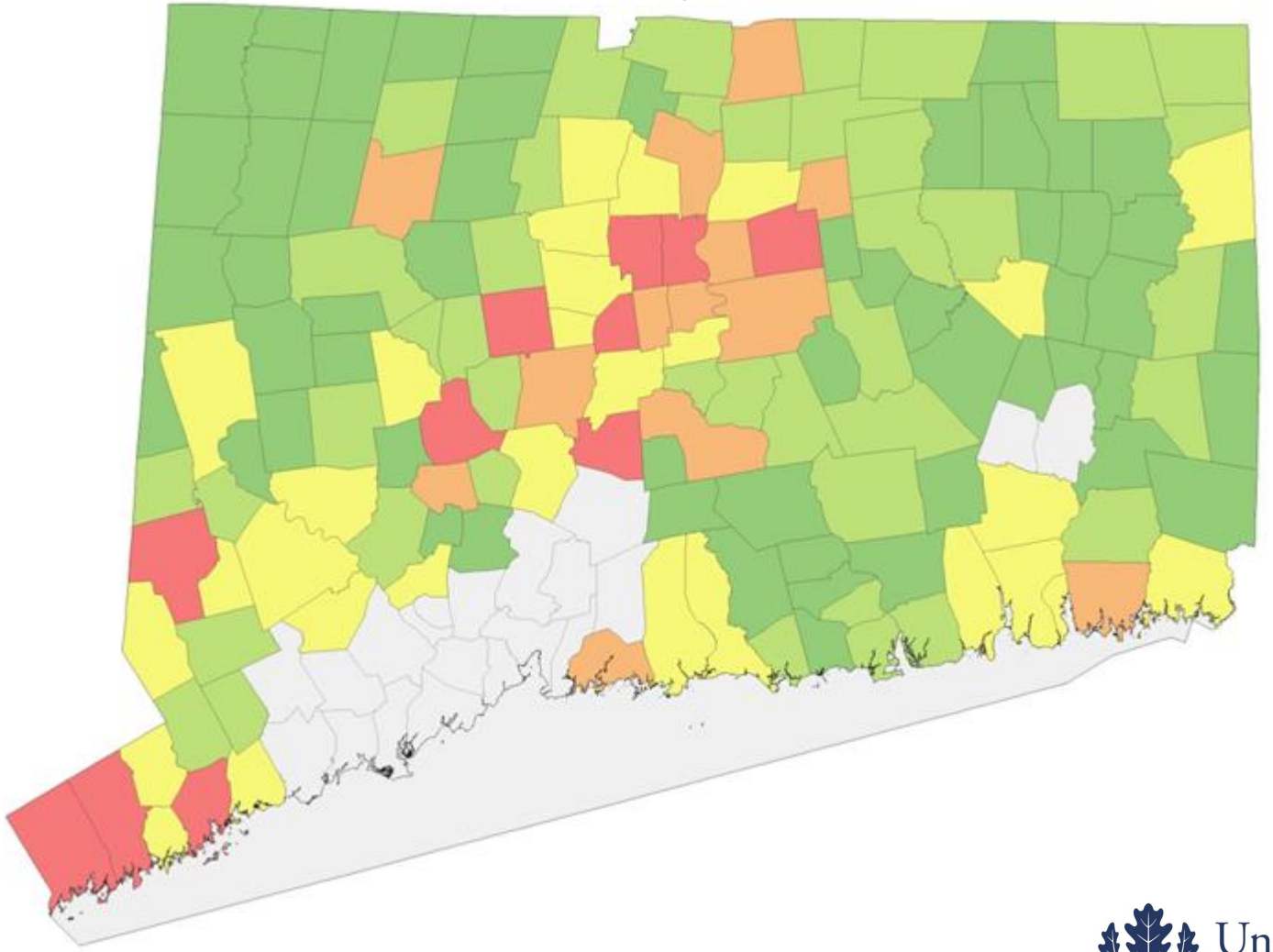
University of Connecticut

Self-reported damage causes CL&P 2004-Present

Cause	Percentage
TREE-Tree Related	89%
LTNG-Lightning	6%
EQUP-Equip Failure	3%
Animals/Birds	1%
VHCL-Veh Accident	0.5%
VAND-Vandalism	<0.5%

University of Connecticut

Population Density

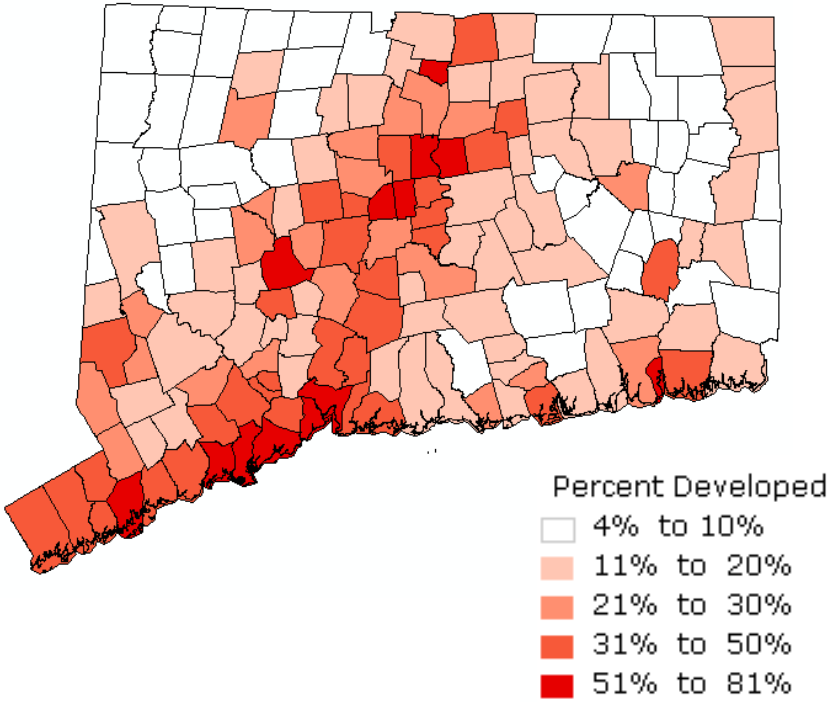


2. Data Description

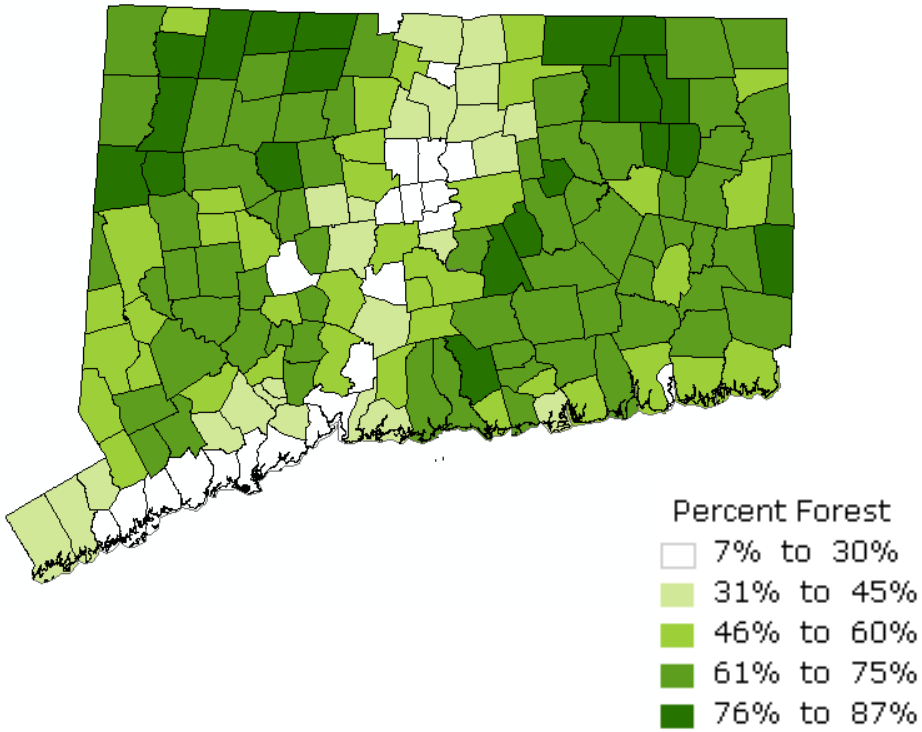
University of Connecticut

Land Cover

Percent developed



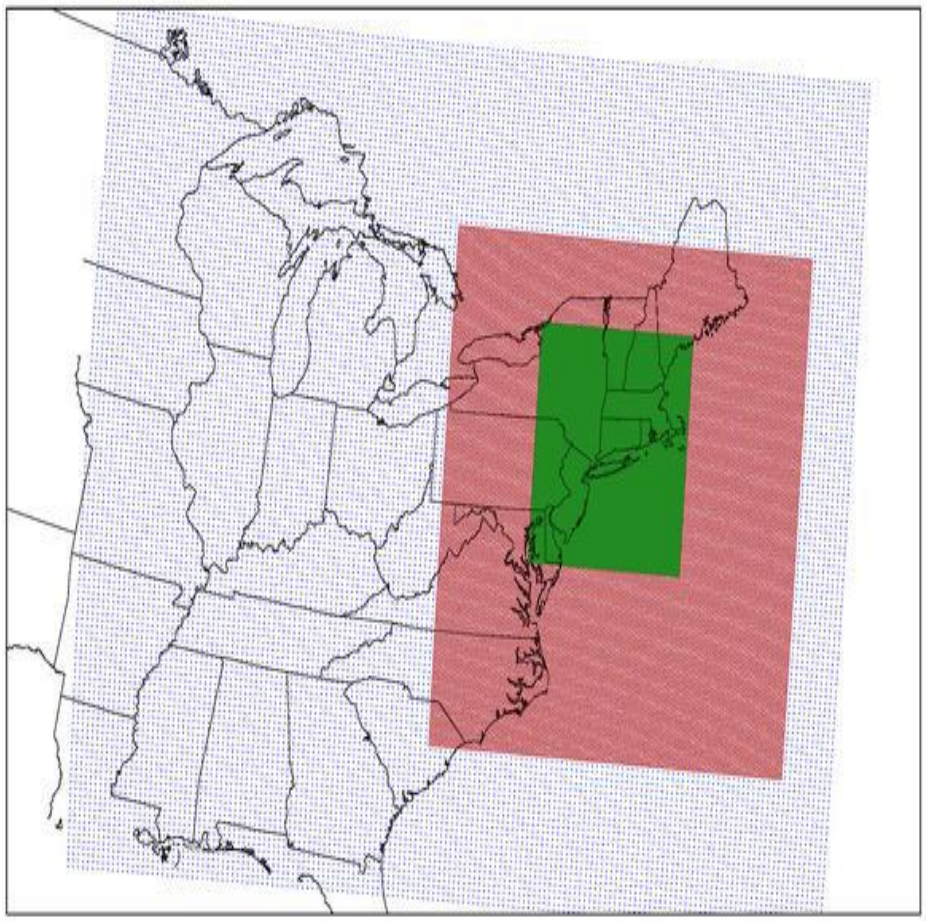
Percent forested



University of Connecticut

2. Data Description

Weather Forecasts



- Simulations initialized with Global Forecast System (GFS) analysis fields at 0.5 degrees, with 6 hour intervals.
- Blue domain (18 km grid), red domain (6 km), green domain (2 km)

University of Connecticut

Some examples of possible covariates

- Physical Weather (GFS)
  - Wind speed (@10m)
  - Wind gust
  - Rainfall
  - # hours wind > 20 mph
- Met Indices (GFS)
  - CAPE (instability)
  - CIN (required energy)
  - BRN (free/forced convection)
- Geographic Characteristics (public sources)
  - Household income
  - Tree density
  - Population density
  - Leaf status
- System Characteristics (NU)
  - Pole count (offset)
  - Tree trimming (std/enhncd)

Processing covariates

- To summarize the time series generated in the weather forecasts we:

1. Use a 4 hour running window and find the maximum mean wind speed.
2. At that same window, calculate both the mean and max of each physical weather characteristic and meteorological index over each town.

University of Connecticut

Model Description

University of Connecticut

Candidate Models

- We are currently looking at a few models for the damage
  - Poisson GLM
  - Negative Binomial GLM
  - Zero-inflated Poisson
  - Zero-inflated Negative Binomial
  - Decision/Regression Trees

Variable Selection

- All told, we have about 90 potential covariates.
- How to choose which to keep?
- Whenever doing variable selection, first ask "Why am I making this model?"
  - For the utilities, their main concern is storm-wise prediction
  - Town-wise accuracy is secondary

Predictive Variable Selection

- Because of that goal, we choose a model based on the hold-out predictive accuracy.

1. Choose one storm to hold out
2. Fit the model using the other storms with the same leaf status.
3. Compute some measure of predictive accuracy (MAPE, MSPE, lift, R^2_{pred}, GINI).
4. Repeat for all storms and all models

Computational Difficulties

- Say each model takes 1 second to predict all storms (70 storms).
- Because each variable can either be in or out of the model, we have $2^{90} = 1.24 \times 10^{27}$ possible models.
- That would take approximately 3.93×10^{19} years, so even with parallel processing you are not going to crack that nut
- This is even assuming that the count and zero models include the same covariates

Variable Selection Steps

1. Prior to hitting it with the heavy computational hammer, remove other variables by talking with experts, for example
   a. Meteorological Parameters
   b. Snow (for non-winter storms)
2. Choose a good set of 15 possibilities from the remaining variables
3. Continue adjusting the possibility set until the model starts to converge.

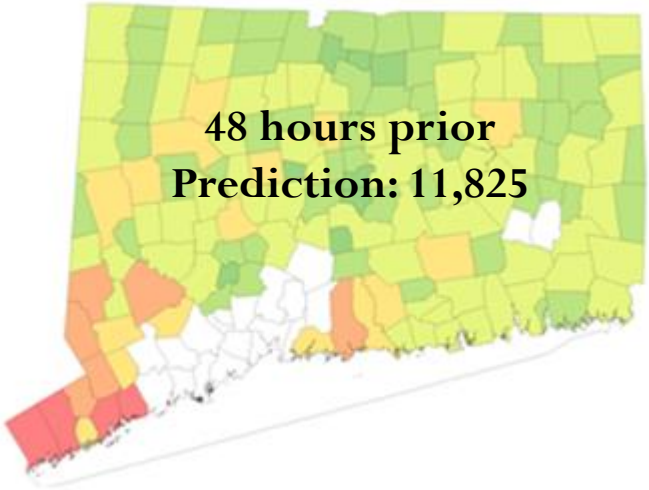
Current Model

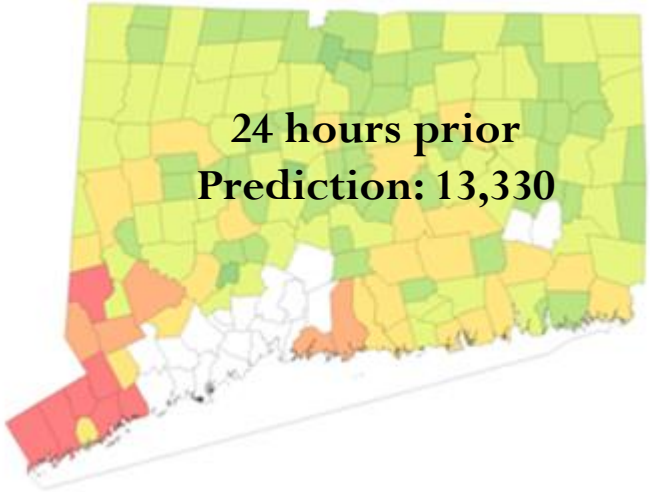
- Current best model: zero-inflated Poisson including
  - Wind speed at 10 meters
  - Wind gust
  - Percent of lines trimmed in the last year
  - Interaction between division and wind speed at 10 meters
  - Interaction between division and wind gust
  - Interaction between precipitation rate and wind speed at 10 meters

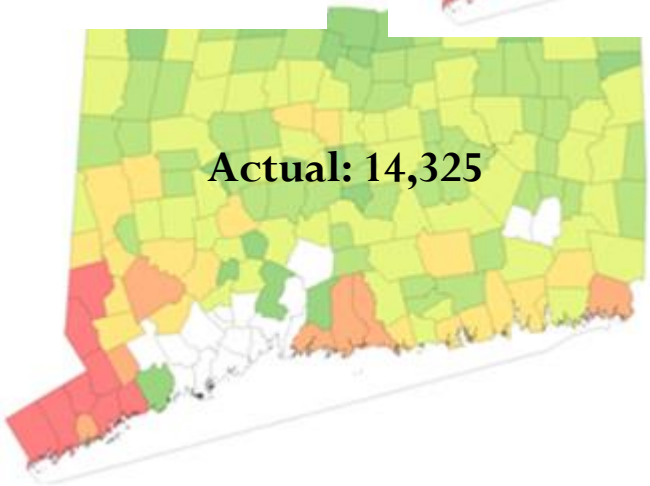
Case Studies

University of Connecticut

Case Study: Hurricane Sandy

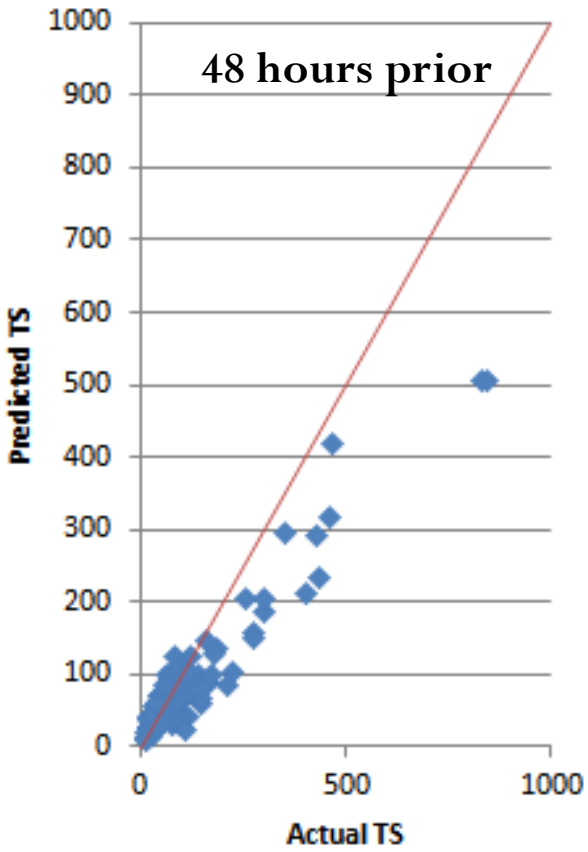


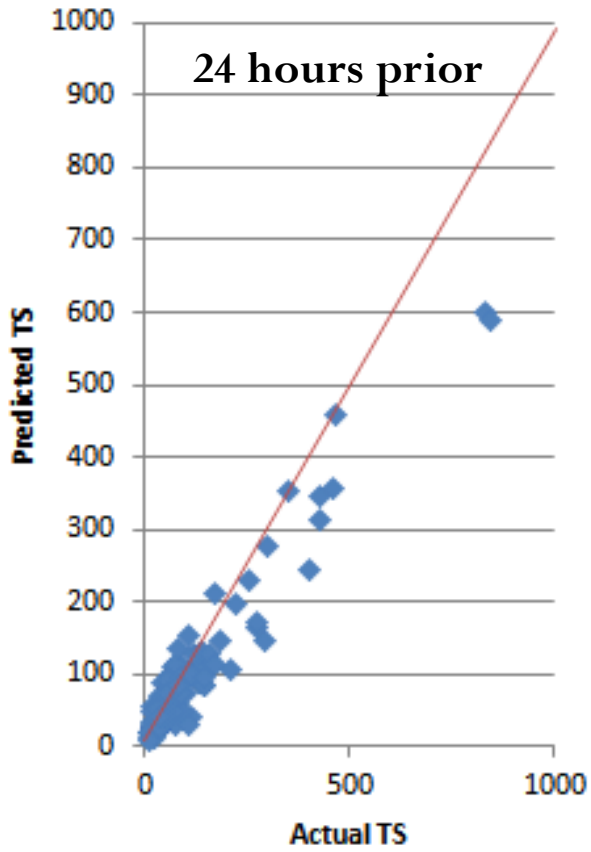




University of Connecticut

Town-wise prediction accuracy





University of Connecticut

4a. Case Study: Hurricane Sandy

What really happened

- We organized a large portion of the data in September/October 2012.
- We had our first model in late October
  - no real variable selection/model testing yet
  - Variables chosen by a single stepwise AIC
- Sandy happened in late October.
- Using that model, we predicted 52,000 trouble spots (basically Armageddon)

What really happened cont.

- So I (on my desktop, unfortunately) made a really simple model (wind speed and tree trimming).
  - No statistical variable selection
  - No meteorological expertise
  - Hardly anything more than dumb luck
- The utility was very excited by these rather accurate predictions

Case Study: Summer 2013

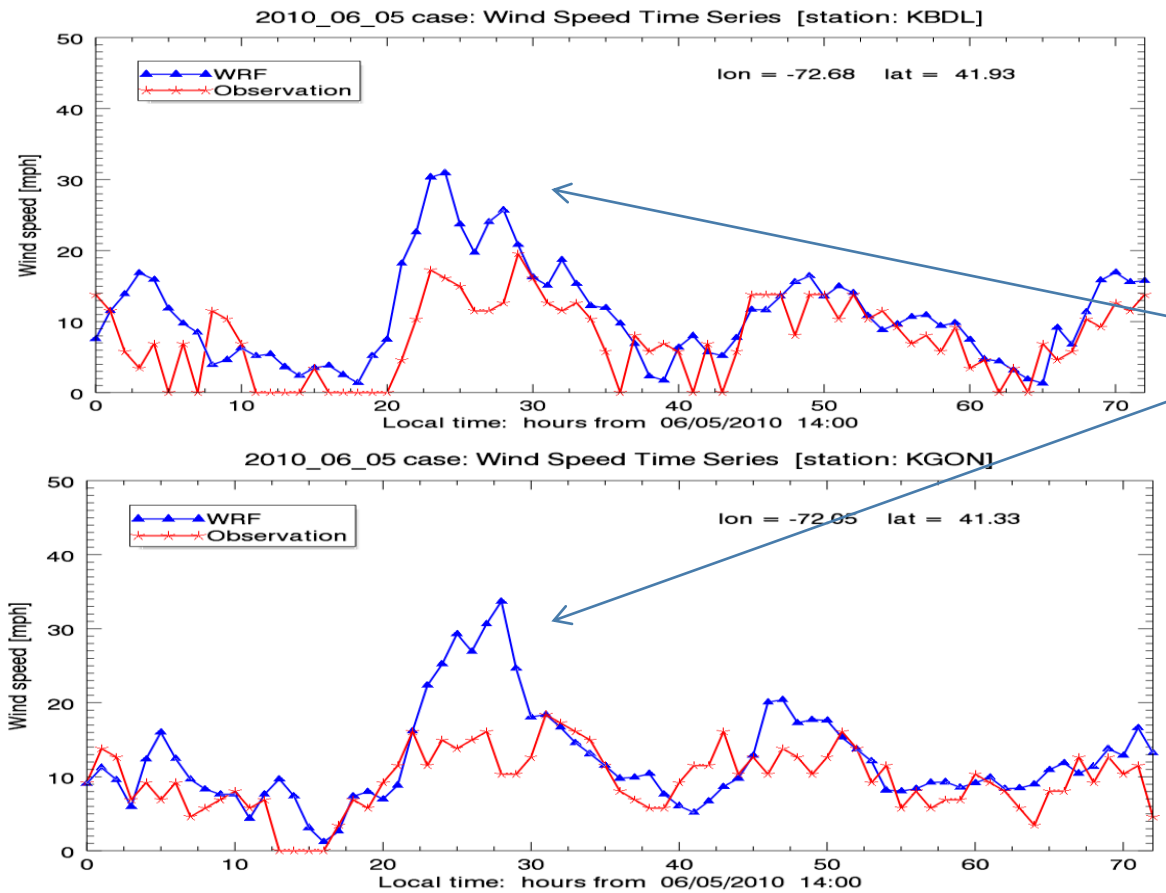
- We had 5 storms in May and June of this year where NU wanted us to run the model.
- Each time, we predicted 300-600 trouble spots and the actuals were 75-150.
- Why were we over-predicting?

Reasons for over-prediction

Two reasons:

1. Our database consisted entirely of "major storms" which are classified as having more than 300 trouble spots.
   - We first showed them this problem 6 months previous.
   - "Why do you want data on damage when there is not a storm, but a few days before you think there might be?"
2. Our weather forecasts were consistently over-predicting the wind speed

Wind speed comparison



Worse in the peaks, where most of the damage occurs.

4b. Case Study: Summer 2013

University of Connecticut

Current and Future Work

University of Connecticut

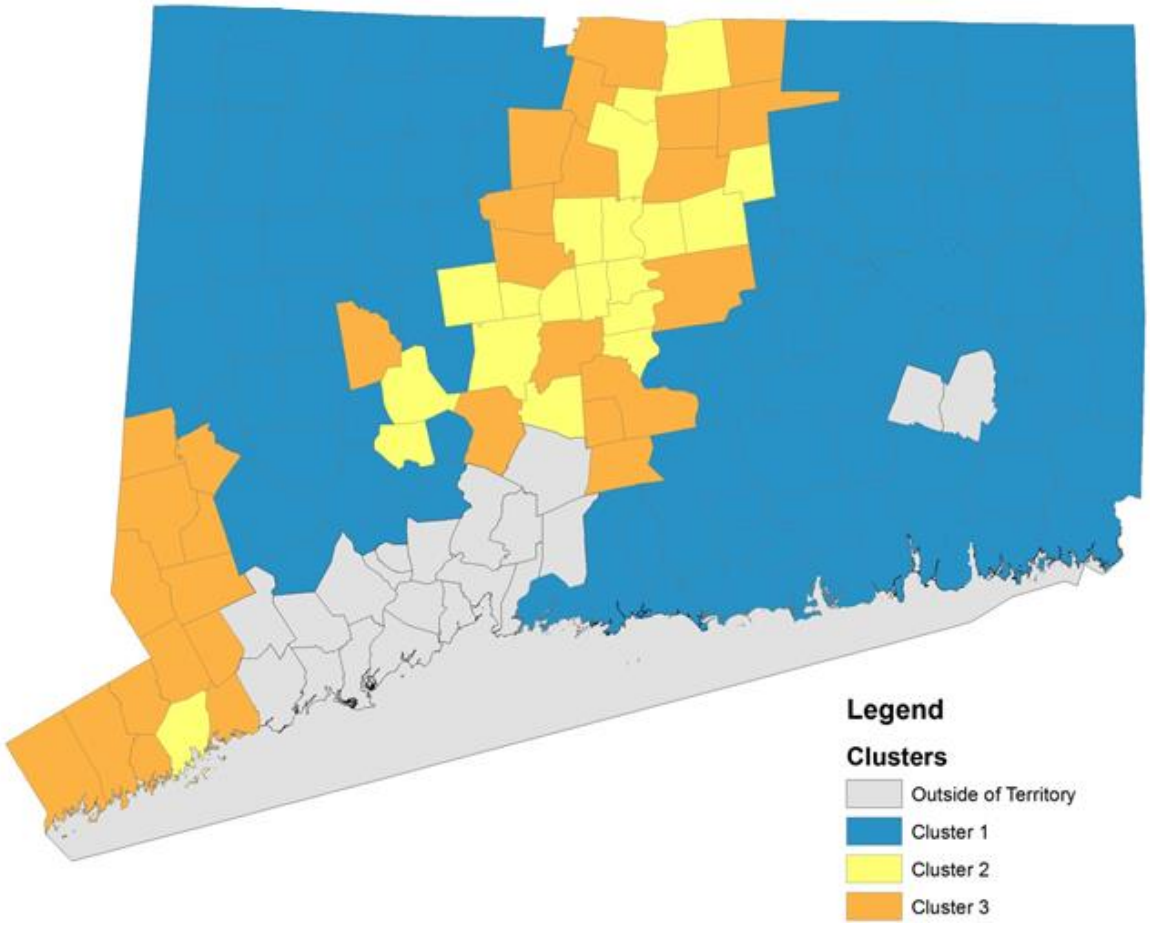
Current Work

1. Clustering the towns
   a. Separate
   b. Hierarchical
   c. Nonparametric Bayes
2. High density (Lidar) data (trimming, HazPix)

Clustering

- Different towns are going to respond in different ways to the storms.
  - Areas with underground power lines will be less affected by the wind
  - Areas on the coast vs. areas inland
- We group the towns into clusters and then fit a separate model on each cluster.

Town-wise clusters



University of Connecticut

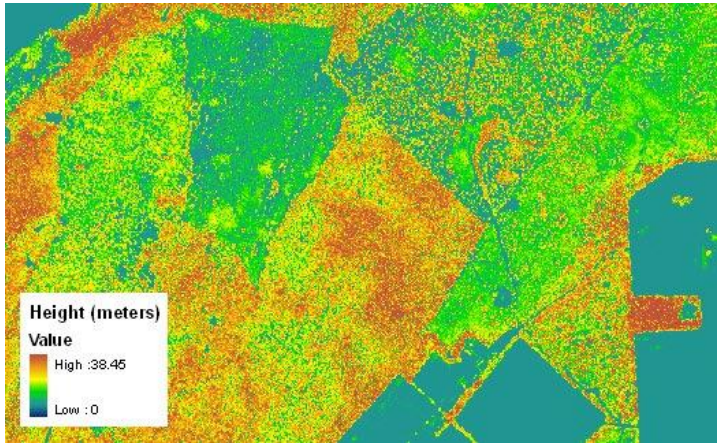
Models on clustered data

Once we have the towns clustered, we have a few options

1. Model each cluster individually
2. Borrow strength using a hierarchical structure.
3. We could combine the clustering and model-fitting into a single step using Bayesian nonparametrics.

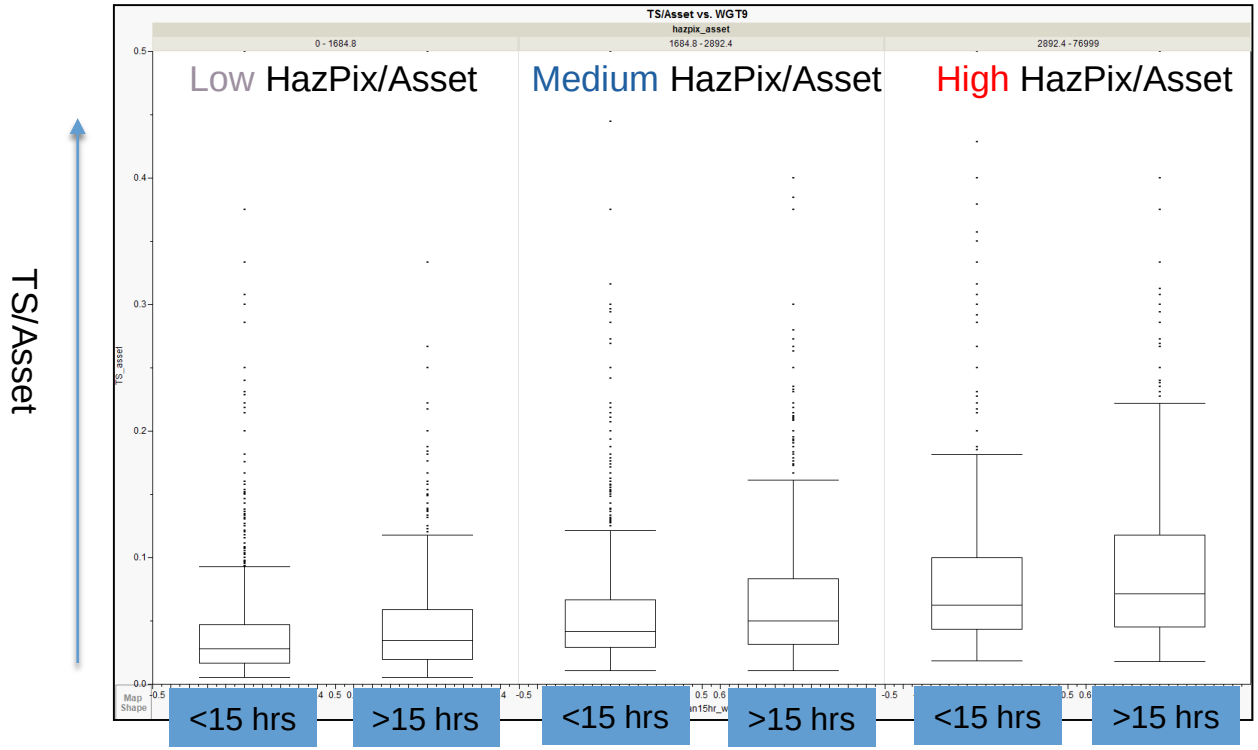
Lidar (Light Detection and Ranging)

- High resolution height measurements (1-9 feet^2)
- Comparison studies between
  - Lidar
  - GIS system map
  - Observation
- Can model on the isolating device level knowing the proportion of each line where trees are threatening



University of Connecticut

Trees, Weather and Infrastructure



TS/Asset for various durations of hazardous winds (20 mph) by HazPix/asset

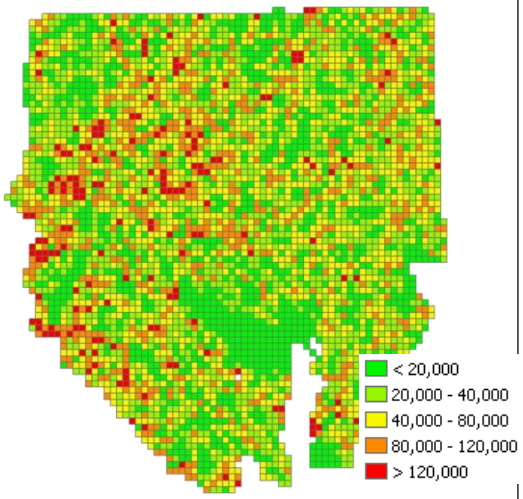


Figure 3: HazPix across Eastern CT

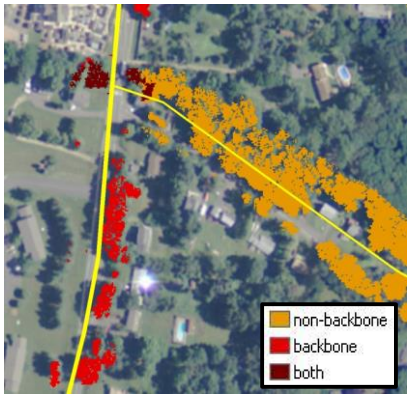
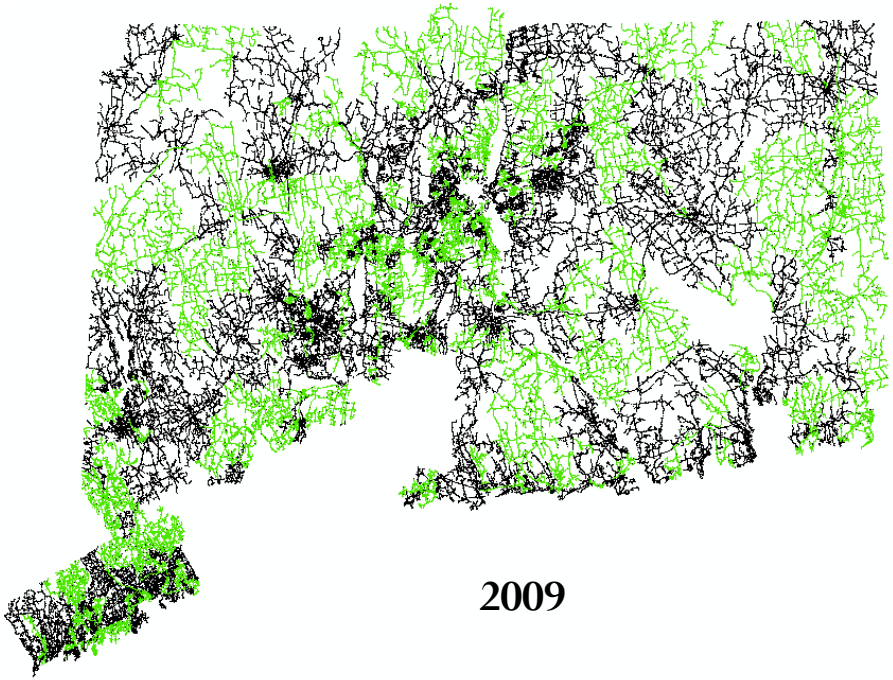
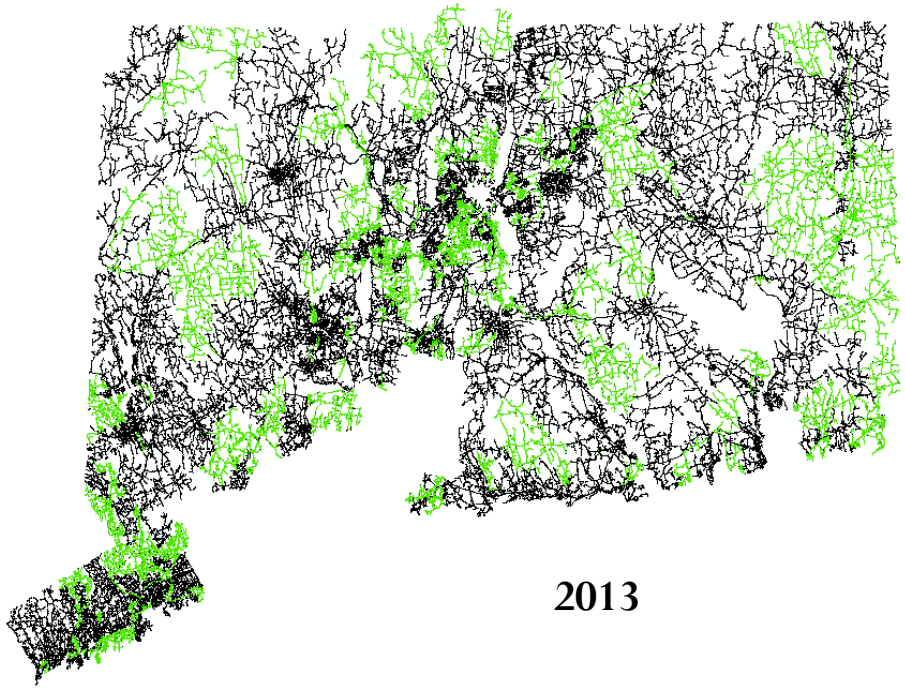


Figure 4: HazPix on distribution ckt

Tree Trimming Data



2009



2013

Storm Damage Modeling at the University of Connecticut

Brian Hartman, PhD ASA

31 January 2014

Workshop on Insurance Mathematics

Quebec, QC

UCONN DAMAGE PREMO

DAMAGE PREDICTION MODEL

University of Connecticut