NC STATE UNIVERSITY

ADAC

Exploratory Analysis of Massive Data for Distribution Fault Diagnosis in Smart Grids

Yixin Cai, Mo-Yuen Chow

Electrical and Computer Engineering,

North Carolina State University

July 2009

Outline

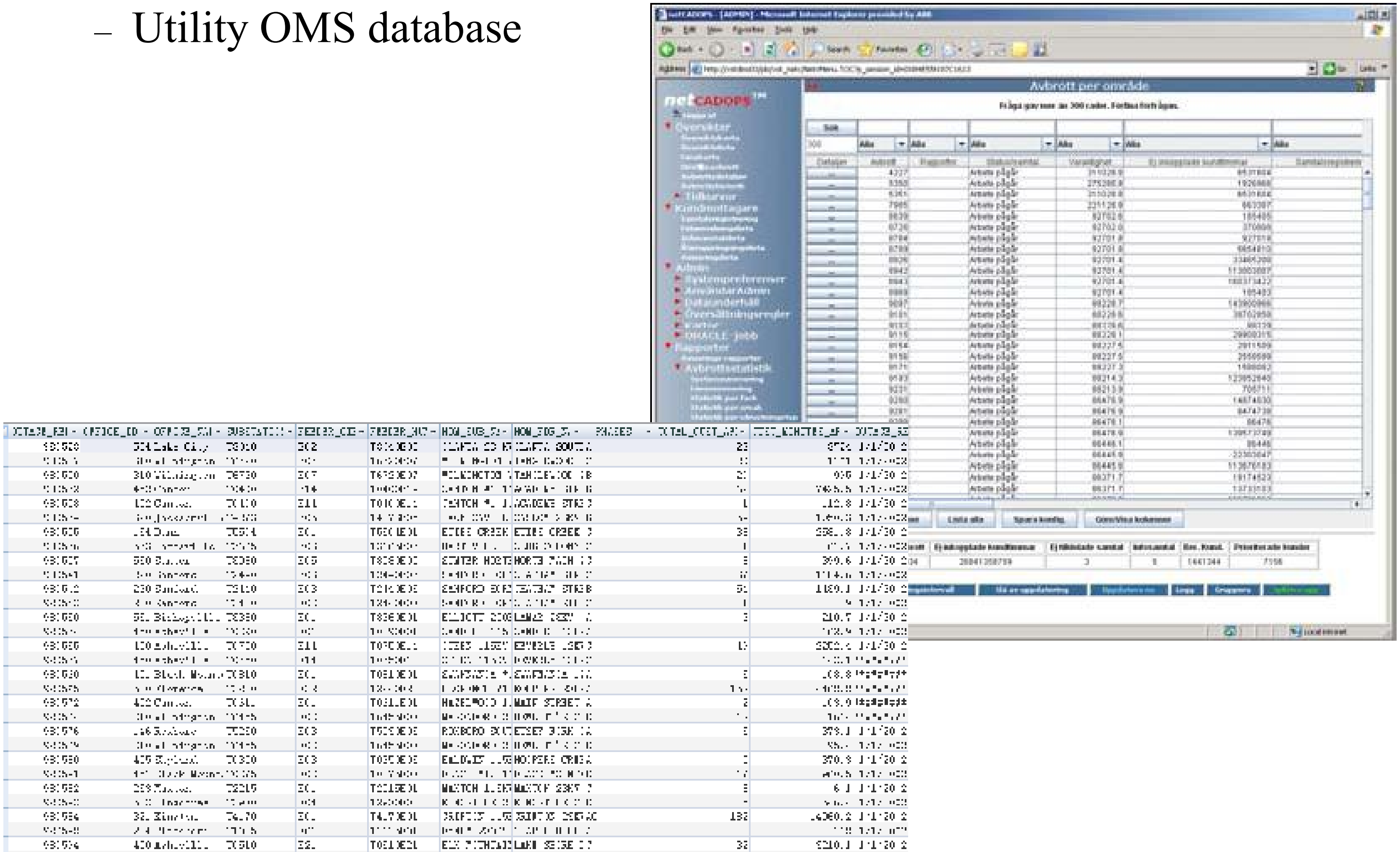
- Introduction
- Integrate data
- Evaluate a single feature
- Build a fault cause classifier
- Summary

Introduction

- Intelligent fault management
  - detection, recording, location, diagnosis, restoration, …
- Problem of interest
  - Diagnosis: predict what the root cause is based on the available information before the engineers go on-site
  - Help (not replace) the engineers to identify the root cause faster
- Challenges
  - Stochastic nature of faults
  - Noisy data with errors
  - More and more incoming data in Smart Grids

Data Integration

- Data sources
  - Utility OMS database

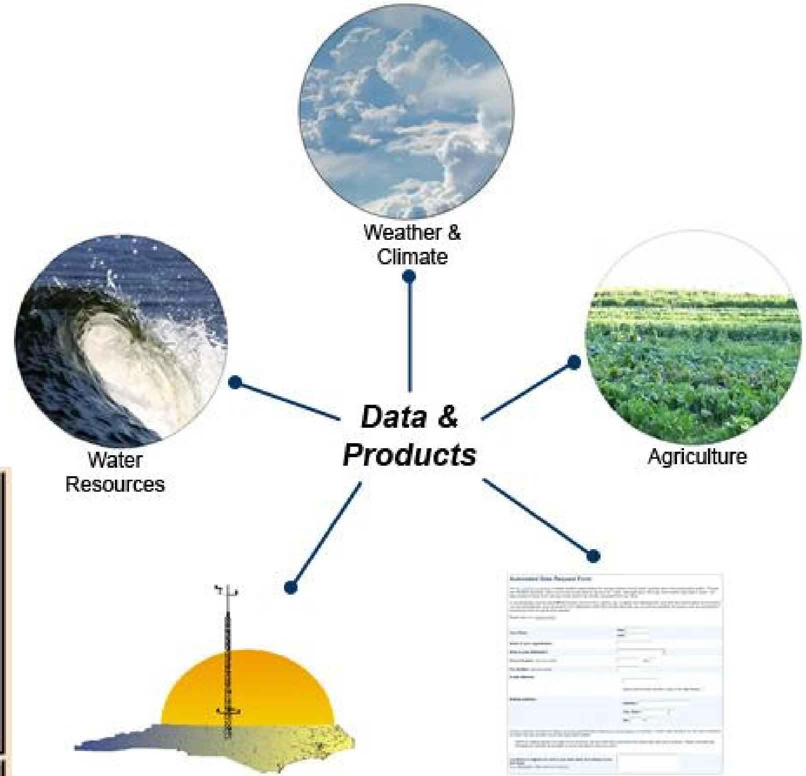


NC STATE UNIVERSITY

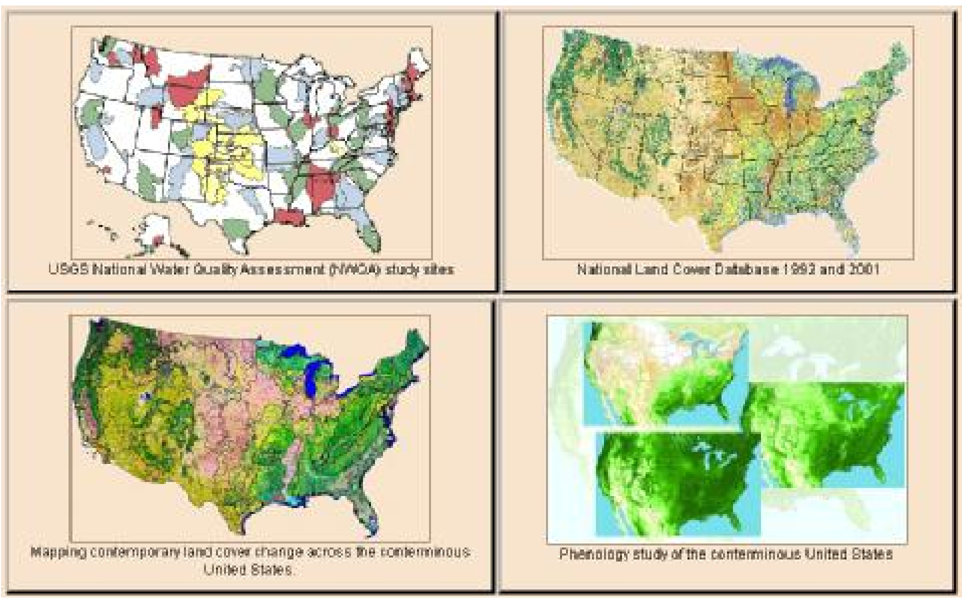
Data Integration

ADAC

- Data sources
  - Utility OMS database
  - Public database on weather, environment, geographic features, etc.



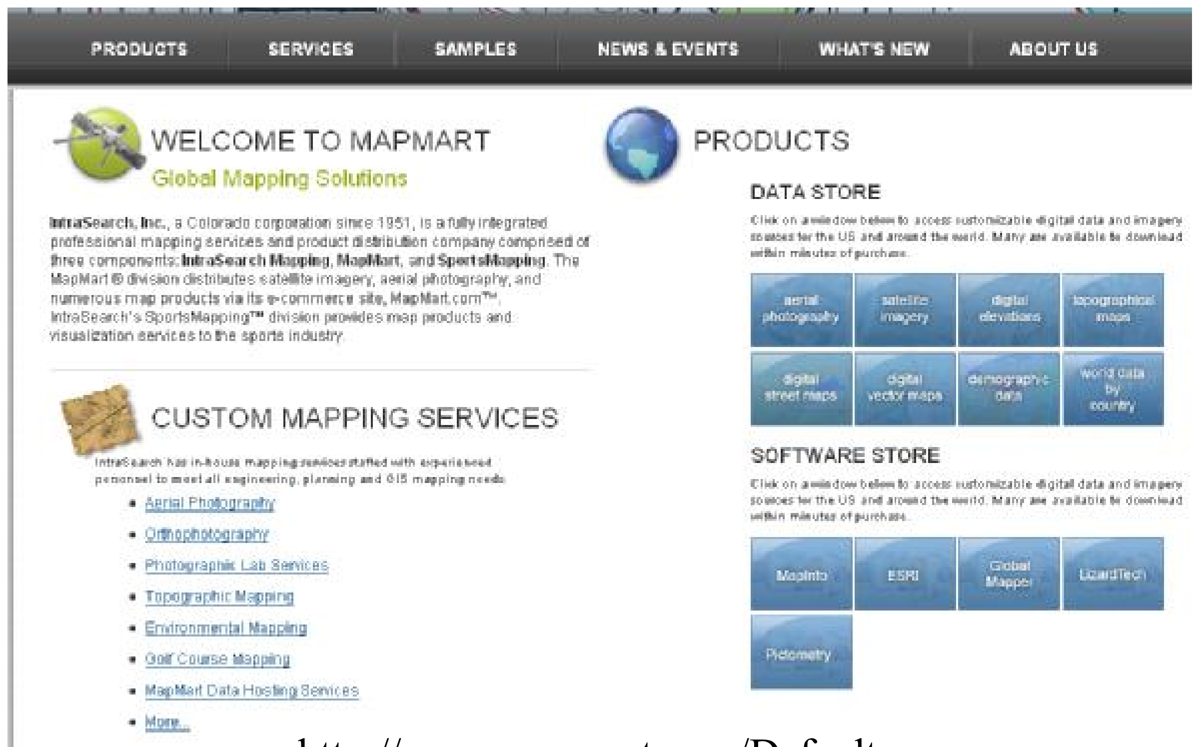
http://www.nc-climate.ncsu.edu/products



http://landcover.usgs.gov/usgslandcover.php

Data Integration

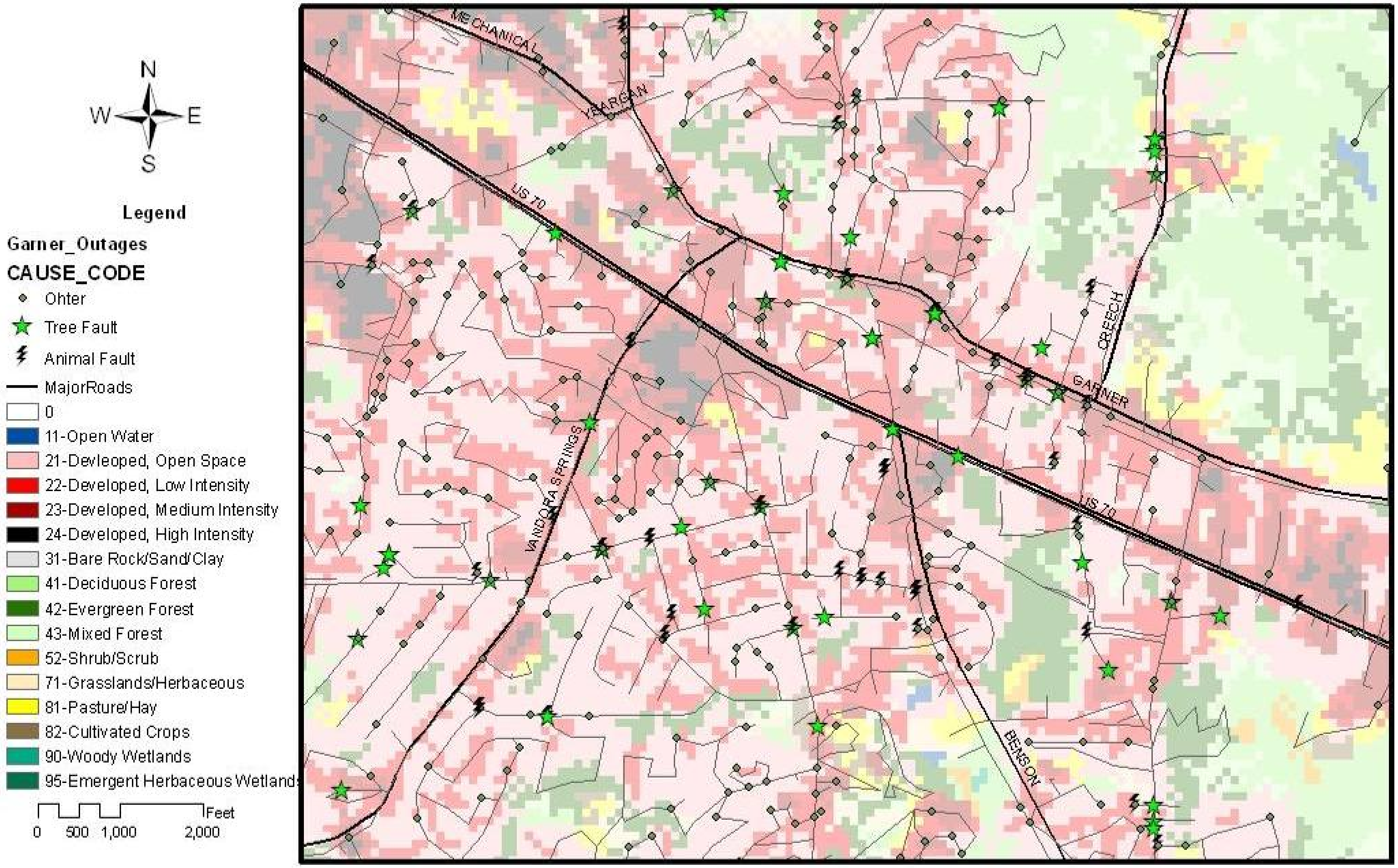
- Data sources
  - Utility OMS database
  - Public database on weather, environment, geographic features, etc.
  - Private venders



http://www.mapmart.com/Default.aspx

Data Integration

- Data integration under GIS framework
  - Spatial relation

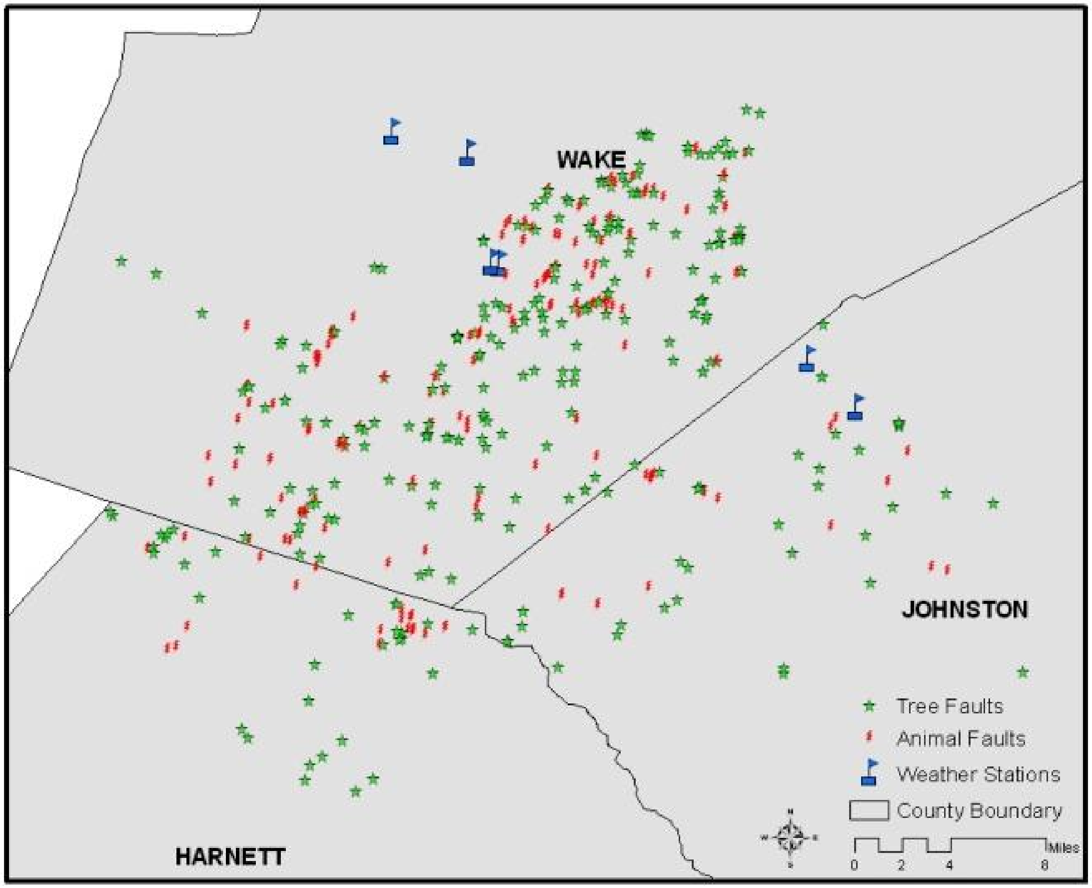


NC STATE UNIVERSITY

Data Integration

ADAC

- Data integration under GIS framework
  - Spatial relation
  - Spatial-temporal relation

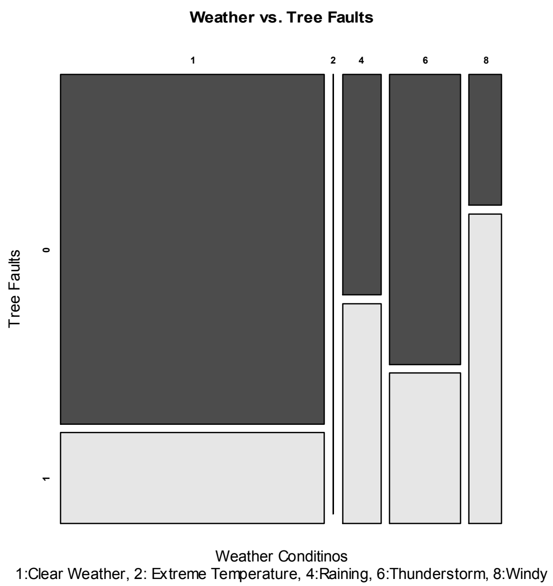


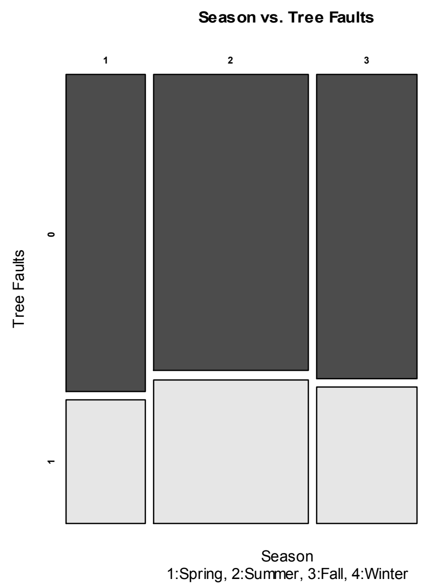
NC STATE UNIVERSITY

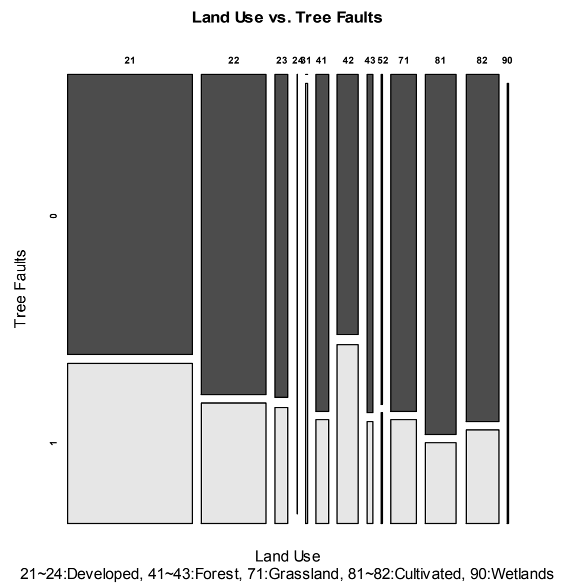
Evaluate a Single Feature

ADAC

- Data preprocess
  - Define the root cause of interest
  - Clean errors and noises
  - Extract features
- Categorical features
  - Likelihood measure $L_{i,j} = P(o_i \mid X = x_j) = \frac{N_{i,j}}{N_j}$
  - (Mosaic) plot



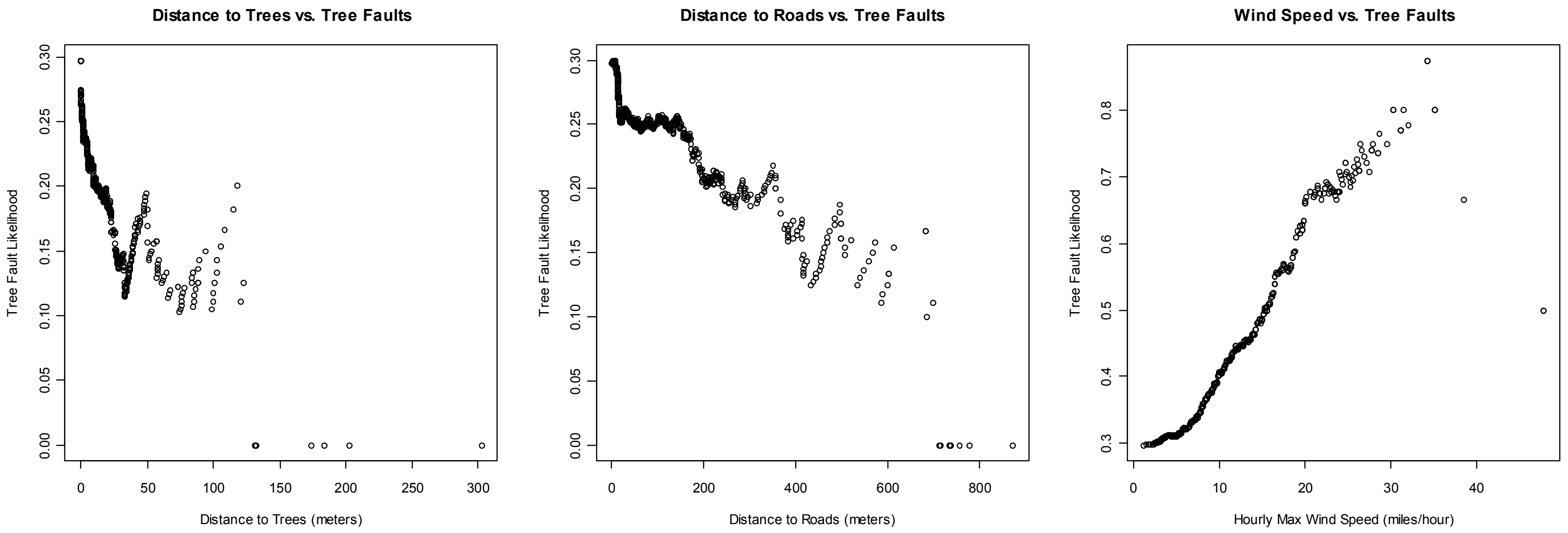




Evaluate a Single Feature

- Continuous features
  - Likelihood measure
  - plot

$$L_{i,j} = P(o_i \mid X \geq x_j) = \frac{N_{i,j}}{N_j}$$



Build a Fault Cause Classifier

- Linear discriminant analysis (LDA)

$$D = \mathbf{w}^T \mathbf{f} = \sum_{i=1}^{N} w_i f_i \qquad D = [\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)]^T \mathbf{f}$$

- Logistic regression (LR)

$$\text{logit}(c=1) = \ln \frac{P(c=1)}{P(c=0)} = \alpha + \boldsymbol{\beta}^T \mathbf{f} \qquad P(c=1) = \frac{1}{1+e^{-\alpha - \boldsymbol{\beta}^T \mathbf{f}}}$$

- Comparison

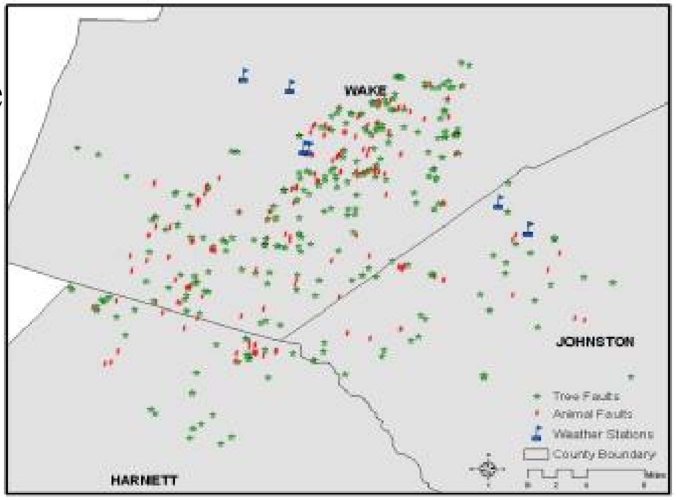
	LDA	**LR**
Model	linear classifier	non-linear classifier
Data assumption	normal distributed with equal variance	none
Computation	matrix manipulation	maximum likelihood

NC STATE UNIVERSITY

Case Study

ADAC

- Data sources
  - Progress Energy Carolinas outage database
  - NC Climate Office
  - NC State Univ. GIS data service
- Fault causes of interest
  - Tree-caused
  - Animal-caused
  - Other
- Features
  - 7 categorical
  - 5 continuous
- Classifiers
  - LDA
  - LR



Classification Performance Using LDA on Sample Dataset

		6 Features		12 Features	
		training	testing	training	testing
Tree fault	ACC	0.75(0.01)	0.76(0.01)	0.77(0.02)	0.76(0.02)
	POD	0.32(0.03)	0.34(0.03)	0.41(0.03)	0.39(0.03)
	FAR	0.34(0.03)	0.32(0.03)	0.32(0.04)	0.33(0.04)
Animal fault	ACC	0.84(0.02)	0.83(0.01)	0.84(0.02)	0.84(0.01)
	POD	0.31(0.04)	0.29(0.04)	0.35(0.03)	0.35(0.03)
	FAR	0.42(0.05)	0.43(0.05)	0.39(0.05)	0.41(0.05)

Classification Performance Using LR on Sample Dataset

		6 Features		12 Features	
		training	testing	training	testing
Tree fault	ACC	0.76(0.02)	0.76 (0.02)	0.77 (0.01)	0.77(0.02)
	POD	0.32(0.03)	0.32(0.03)	0.44 (0.03)	0.44(0.03)
	FAR	0.30(0.04)	0.30(0.04)	0.32 (0.03)	0.34(0.03)
Animal fault	ACC	0.83(0.02)	0.83 (0.02)	0.84(0.01)	0.84(0.01)
	POD	0.30(0.03)	0.31(0.03)	0.37 (0.04)	0.35(0.03)
	FAR	0.42(0.04)	0.41(0.06)	0.41 (0.06)	0.41(0.06)

Summary

- Methods for exploratory data analysis
  - Integrate data from multiple sources under GIS framework
  - Use likelihood measure to evaluate both categorical and continuous features
  - Apply LDA and LR as fault cause classifiers

- Findings
  - LDA and LR performs similar
  - Adding new features helps fault diagnosis

- Future work
  - Systematic feature selection methods
  - Advanced fault diagnosis algorithms
  - Novel sampling strategy