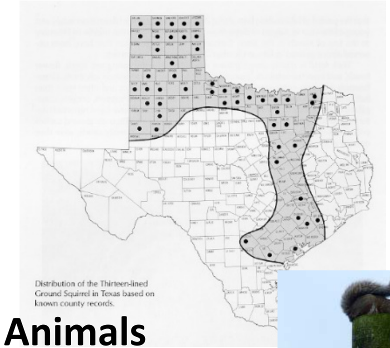
Big Data Framework for Predictive Risk Assessment of Weather Impacts on Electric Power Systems

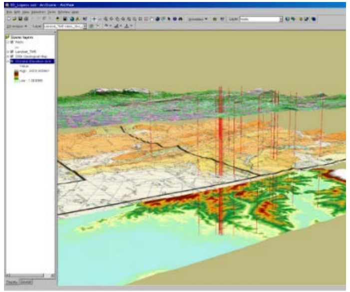
Mladen Kezunovic
Texas A&M University

CIGRE - 2019 Grid of the Future
November 3-6 2019, Atlanta, GA

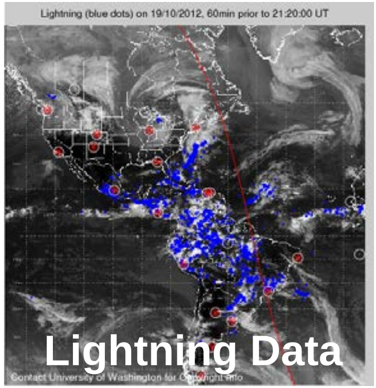
BD Data Aggregation



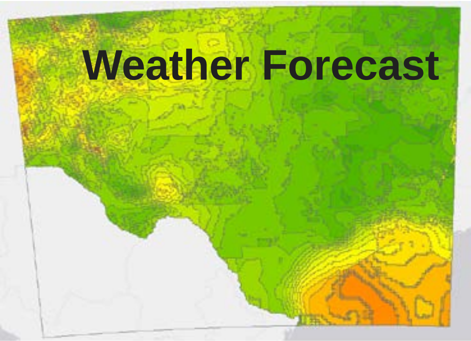
Animals Data

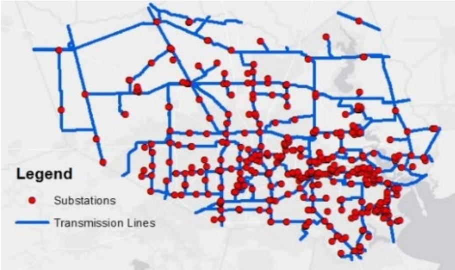


GIS



UAS

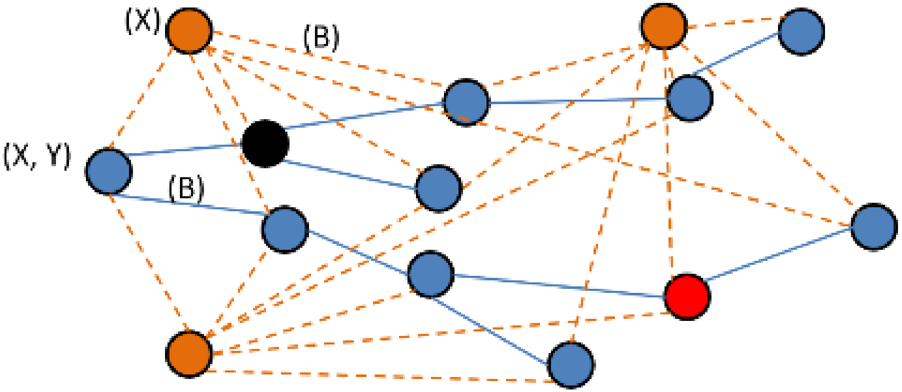




Network Assets Data

Challenges of Big Data

- Spatiotemporal correlation
- Scalability
- Missing data
- Bad data
- High volume, variety, velocity
- Various types of uncertainties



Nodes: a) Input data: X = (Sensor Data, Weather Data, Missing Data, Bad Data)

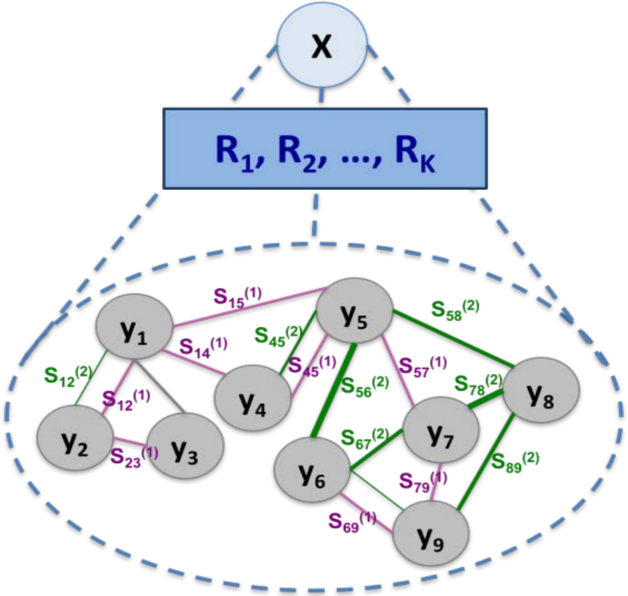
b) Output: Y = Outage Probability

Branches: B = Geographical distance between nodes

Network impedance matrix

Prediction Model

To reduce the computational complexity of learning and inference *A* and *I* can be constructed as quadratic functions of *y* – **Gaussian Conditional Random Fields**

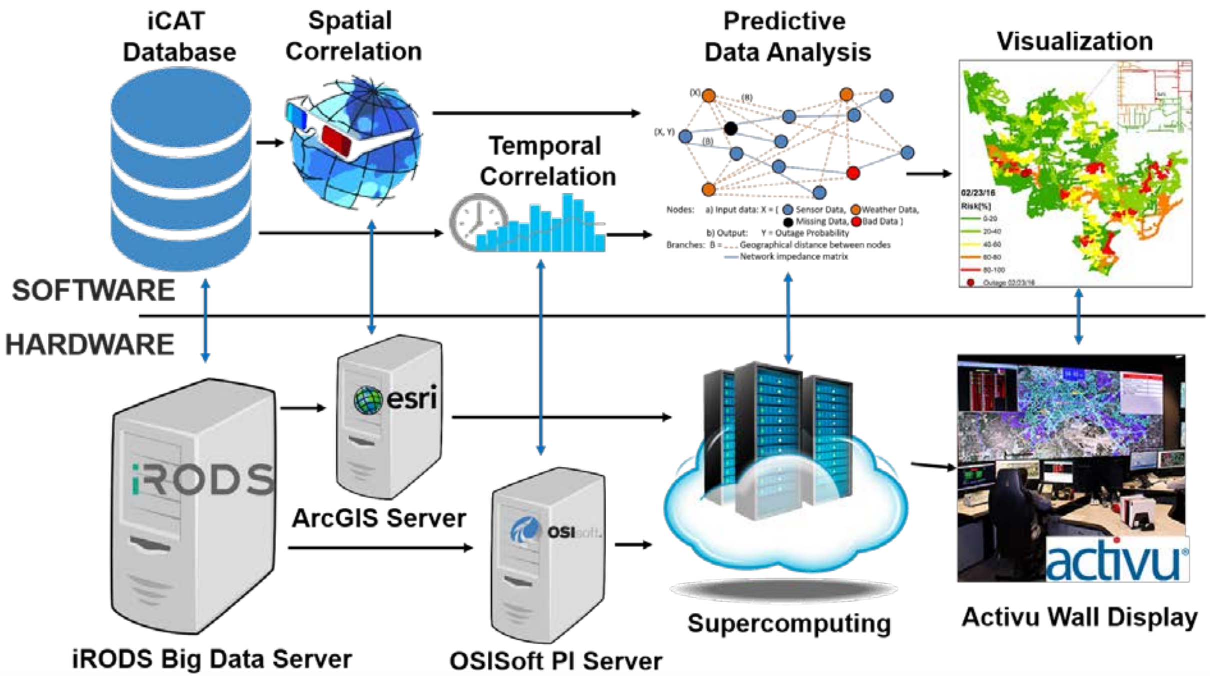


$$P(y|X) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{N}\sum_{k=1}^{K}\alpha_k\left(y_i - R_k(X)\right)^2 - \sum_{i,j}\sum_{l=1}^{L}\beta_l e_{ij}^{(l)} S_{ij}^{(l)}(X)\left(y_i - y_j\right)^2\right)$$

ENGINEERING TEXAS A&M UNIVERSITY

ELECTRICAL & COMPUTER ENGINEERING TEXAS A&M UNIVERSITY

Data Processing

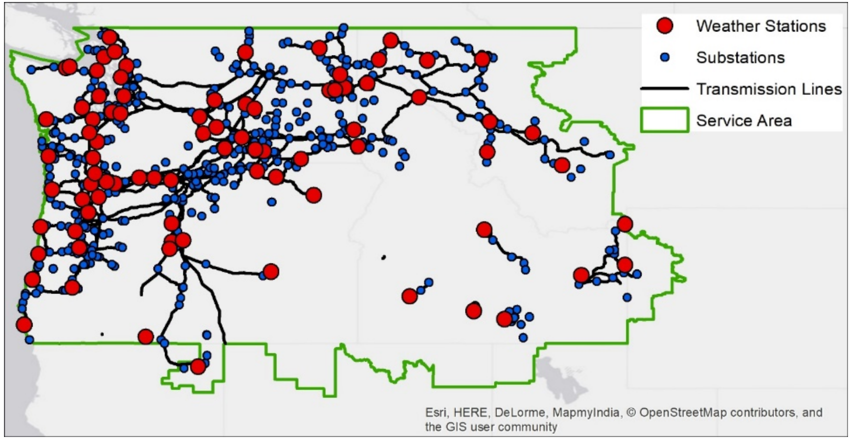


Applications

- Transmission Outage Prediction
- Transmission Insulator Failure Predictions
- Distribution Vegetation Management
- Distribution Transformer Failure Prediction
- Solar and Wind Generation Forecast

Transmission Outage Prediction

Historical Weather Data

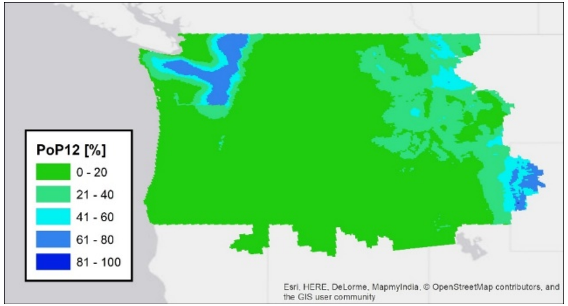


Fractions of missing data from ASOS observations

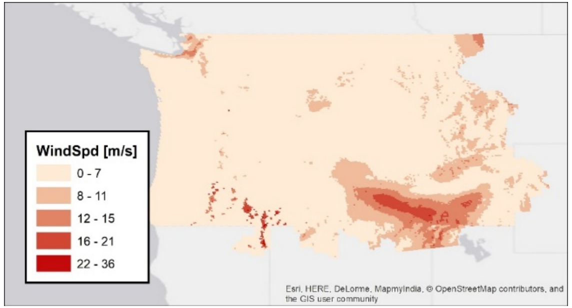
Temperature	Dew Point	Humidity	Wind Direction	Wind Speed	Precipitation	Pressure	Wind Gust	Weather Code
0.146	0.148	0.148	0.145	0.134	0.312	0.265	0.378	0.336

Weather Forecast

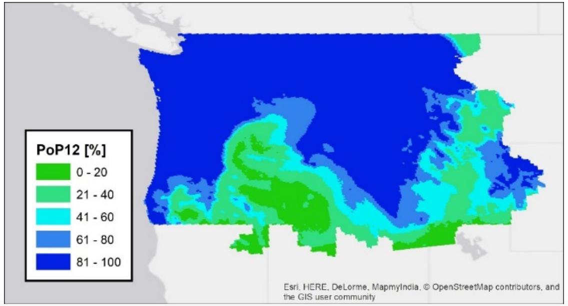
No Outage - Precipitation



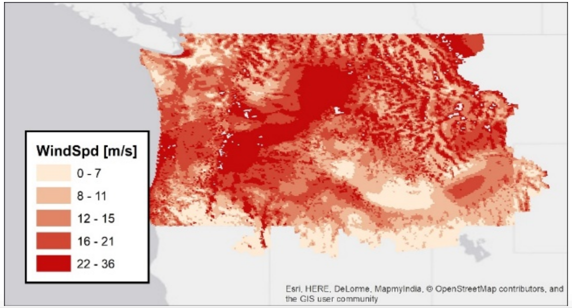
No Outage – Wind Speed



Outage - Precipitation



Outage – Wind Speed



Results – Outage Occurrence Prediction

Experimental Setup

- Training: data from 1999 to 2010
- Prediction horizon: 2010-2018
- Substations were embedded into a $\mathbf{50}$-dimenstional space based on their spatial proximity
- CLEC was run with $M = 30$ components
- $\eta = 30\%$ of the training data were sampled to construct the subset for each LR component

Model	Acc.	AUC	F1	Bias
LR	0.8467	0.9278	0.8097	0.6821
LR (spatial)	0.8624	0.9292	0.8242	0.7075
CLEC	**0.8919**	**0.9313**	**0.8532**	**0.7685**

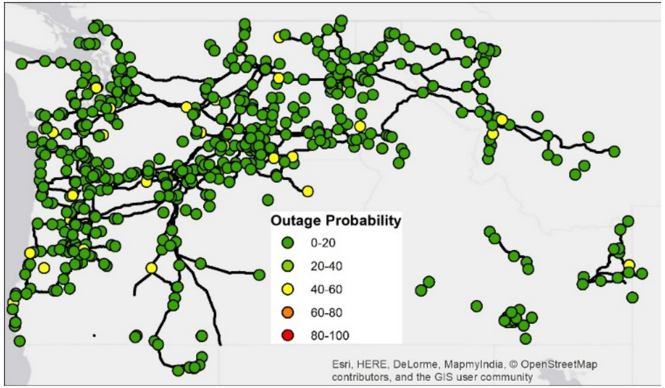
Prediction performance w.r.t. different evaluation metrics.

Discussion

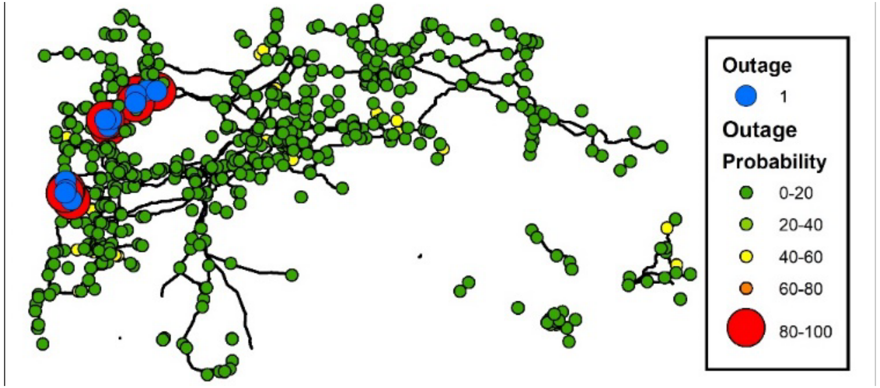
- LR (spatial) obtains greater classification performance compared to LR
  - ⇨ supports the hypothesis **that spatial information is truly relevant for this task**
- CLEC outperforms its alternatives, yielding higher values for accuracy, AUC and F1
- Large lift in Bias
  - ⇨ shows the **benefit of using a subsampling-based ensemble scheme**

Risk Maps

No Outage



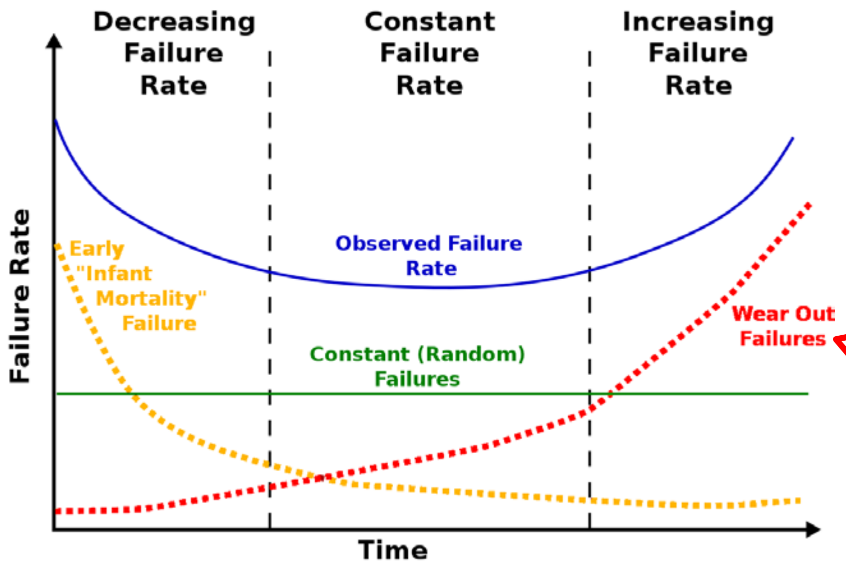
Multiple Lightning Outage



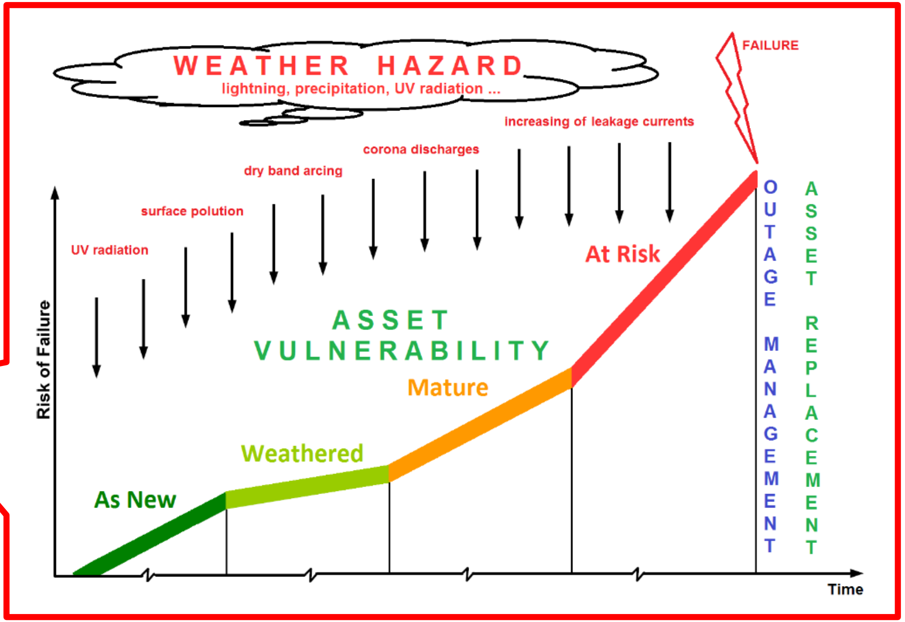
- **No outages occurred** ⇨ outage probabilities are **smaller than 60%** for all substations
- Outages occurred ⇨ the area around the outages has points with probability over 80%
- **In general**: Both LR and CLEC do well at guessing the areas of outage occurrences
- **However**: CLEC does better than LR making prediction **more precise on a spatial level**
  - ⇨ the number of high risk areas far away from the outage locations is much smaller

Transmission Insulation Coordination

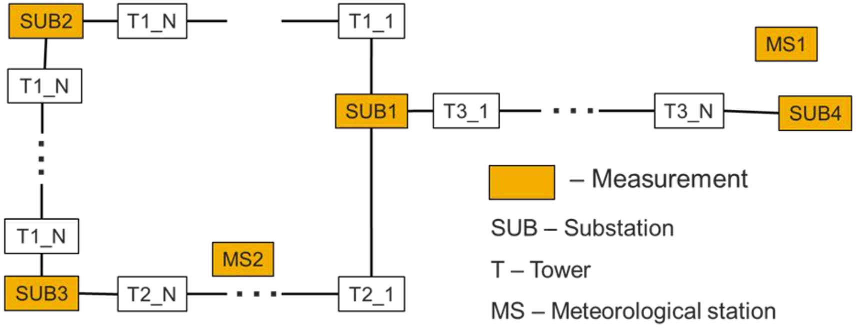
Asset Management for Insulators



*https://commons.wikimedia.org/wiki/File:Bathtub_curve.svg#/media/File:Bathtub_curve.svg



*A. Tzimas, et al. "Asset management frameworks for outdoor composite insulators." IEEE Transactions on Dielectrics and Electrical Insulation 19.6 (2012).



Prediction Model

Lightning Detection Network

Weather Stations

Nodes: X = (Lightning Current, Temperature, Precipitation, Humidity, Pressure, BIL_old)

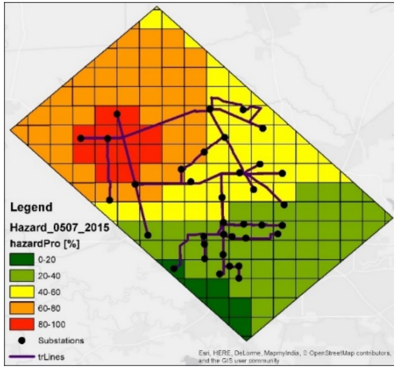
Y = (BIL_new)

Branches: Network Impedance Matrix

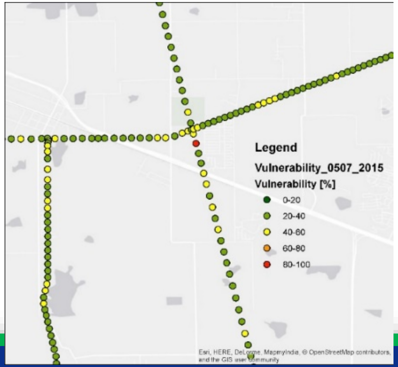
$$P(y|x) = \frac{1}{Z} exp\left(-\sum_{i=1}^{N} \sum_{k=1}^{K} \alpha_k \left(y_i - R_k(x)\right)^2 - \sum_{i,j} \sum_{l=1}^{L} \beta_l e_{ij}^{(l)} \, S_{ij}^{(l)}(x)\left(y_i - y_j\right)^2 \right)$$

Risk Maps

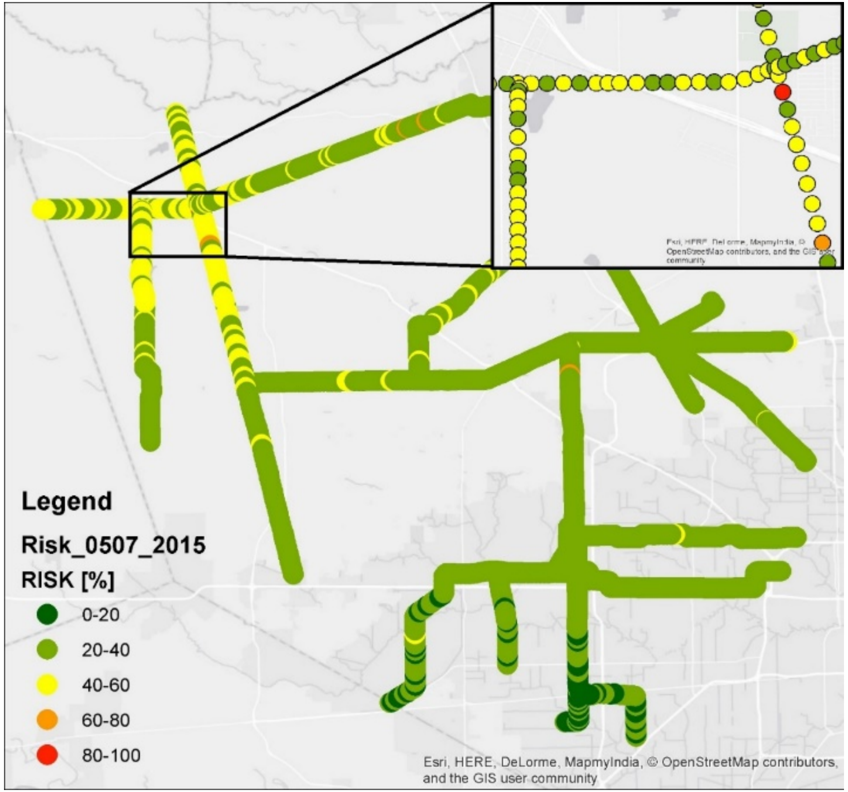
Weather Hazard



Network Vulnerability



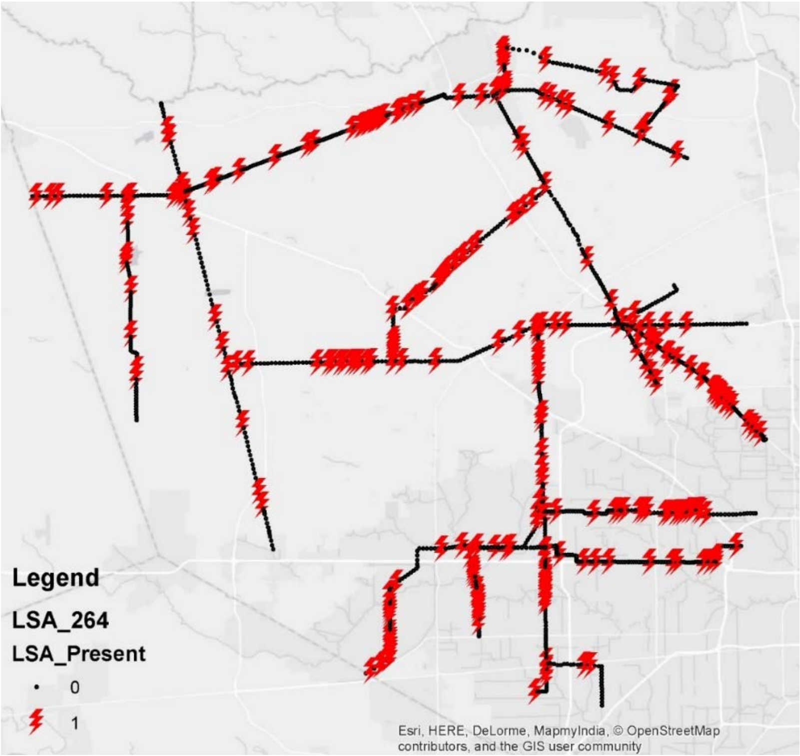
Risk Map



Optimal Location of LSAs

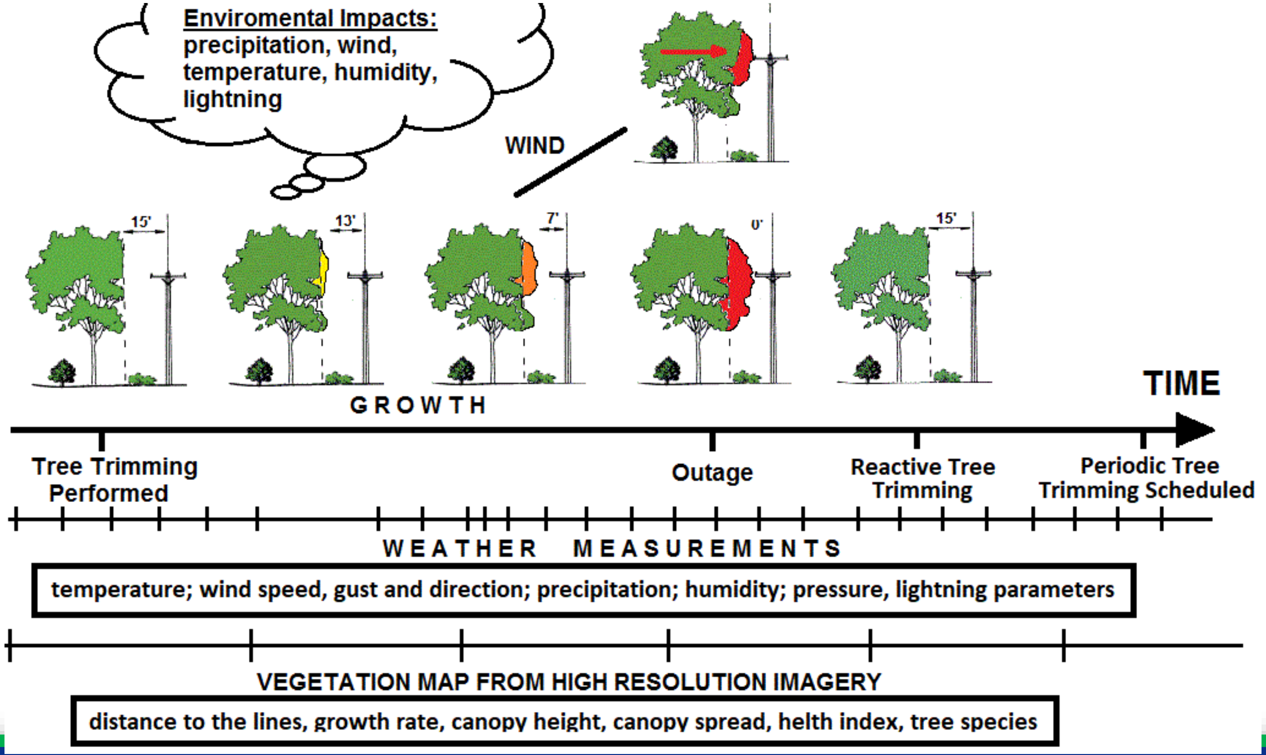
264 LSAs locations

Total Risk Reduction = 72.69%

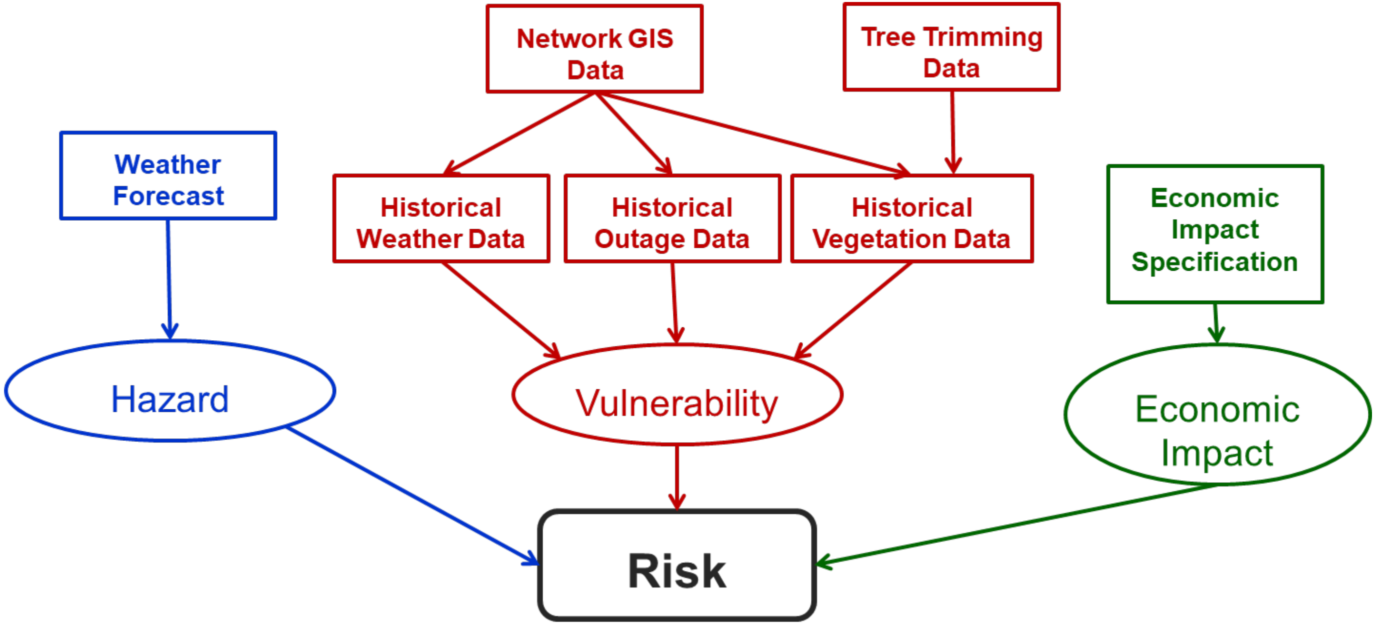


Distribution Vegetation Management

Big Data for Vegetation Management



Risk Model



Network GIS Data
Tree Trimming Data
Weather Forecast
Historical Weather Data
Historical Outage Data
Historical Vegetation Data
Economic Impact Specification
Hazard
Vulnerability
Economic Impact
Risk

Optimal Tree Trimming

For total of N feeder sections maximize the reduction in vegetation risk

$$max\left\{R = \sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}\Delta R_{n,t}\cdot F_{n,t}\right\}$$

$$F_{n,t} = \begin{cases} 0, & \text{section not trimmed} \\ 1, & \text{section trimmed} \end{cases}$$

Difference in component risk before and after action:

$$\Delta R_{n,t} = R_{n,t}^{before} - R_{n,t}^{after}$$

Constrains:

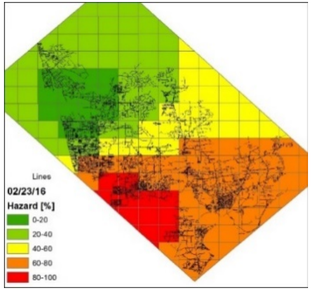
Total cost of tree trimming limit: $\sum_{t=1}^{T}\sum_{n=1}^{N} F_{n,t}\cdot PC_{n,t} \leq PA$

One section trimmed at the time: For t=1,…,T, $\sum_{n=1}^{N} F_{n,t} = 1$

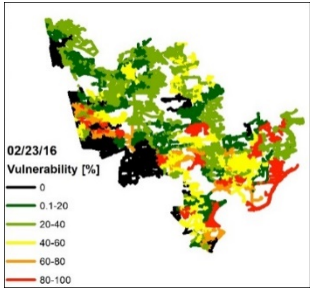
n	feeder section
N	total number of feeder sections
t	time instance
T	number of time instances
$\Delta R_{n.t}$	reduction in risk after trimming
$PC_{n,t}$	cost of trimming on one section
PA	total allocated tree trimming budget

Risk Maps

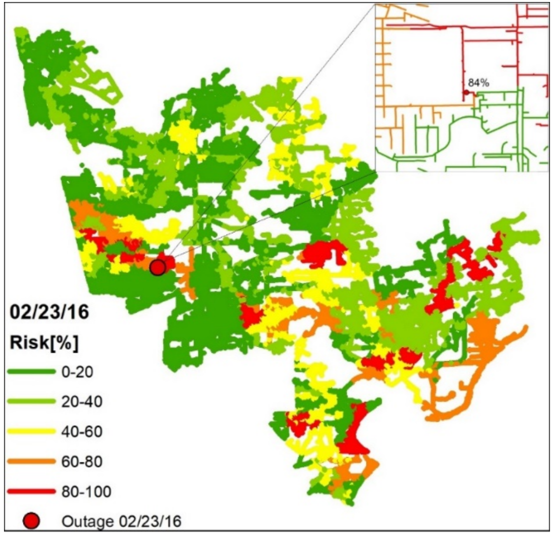
Weather Hazard



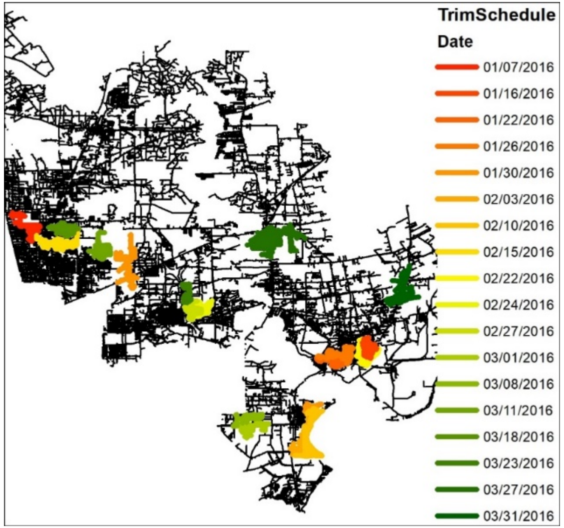
Network Vulnerability



Risk Map



Optimal Tree Trimming Schedule



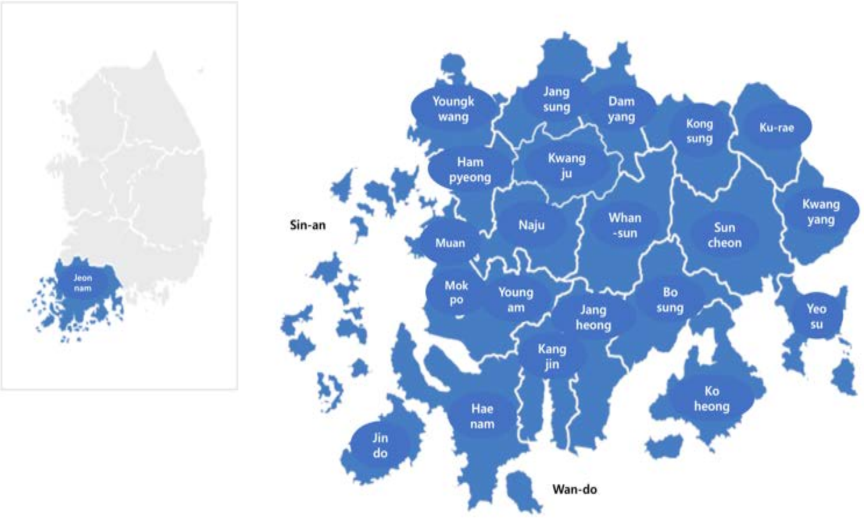
Overall risk reduction 32.85%

Reactive tree trimming cost reduction 27.2%

Distribution Transformer Failure Prediction

The System – South Korea

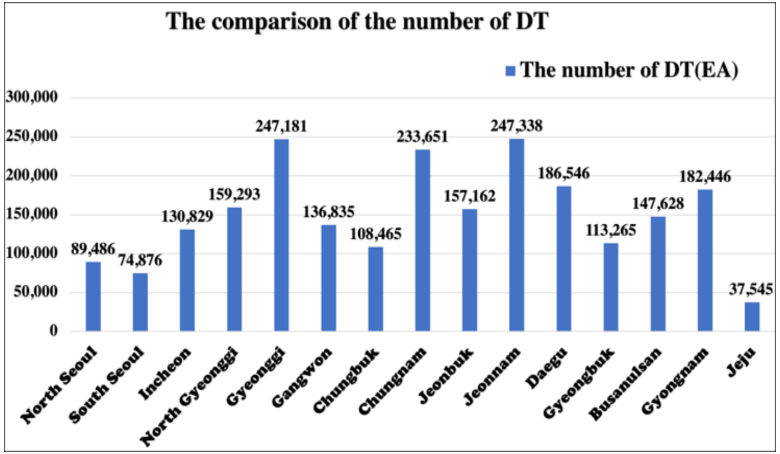
JeonllaNam-do Area



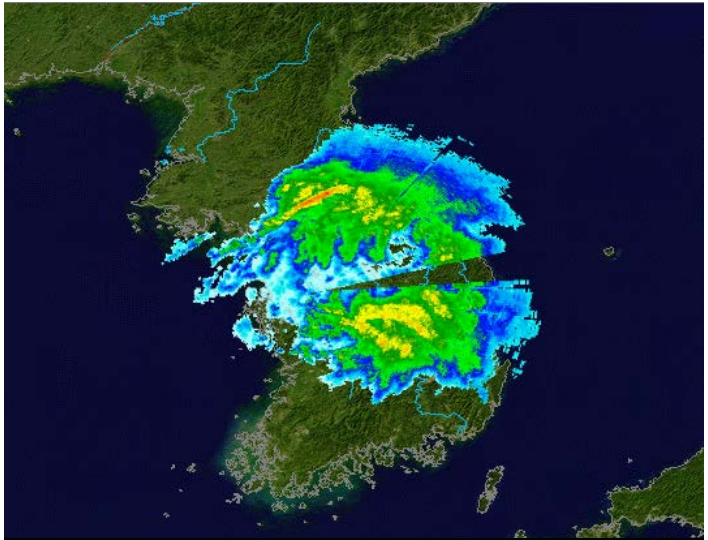
Distribution Facilities in JeonllaNam-do Area

Transformer			Protective Devices		
Bank	Number	Capacity (kVA)	Breaker	Equipment	COS
104	243	9,252	12	1.4	72

The comparison of the number of DT in South Korea



Weather Data



NOAA/U.S. Air Force, Blue Marble map
https://www.ncdc.noaa.gov/nexradinv/chooseday.jsp?id=rksg

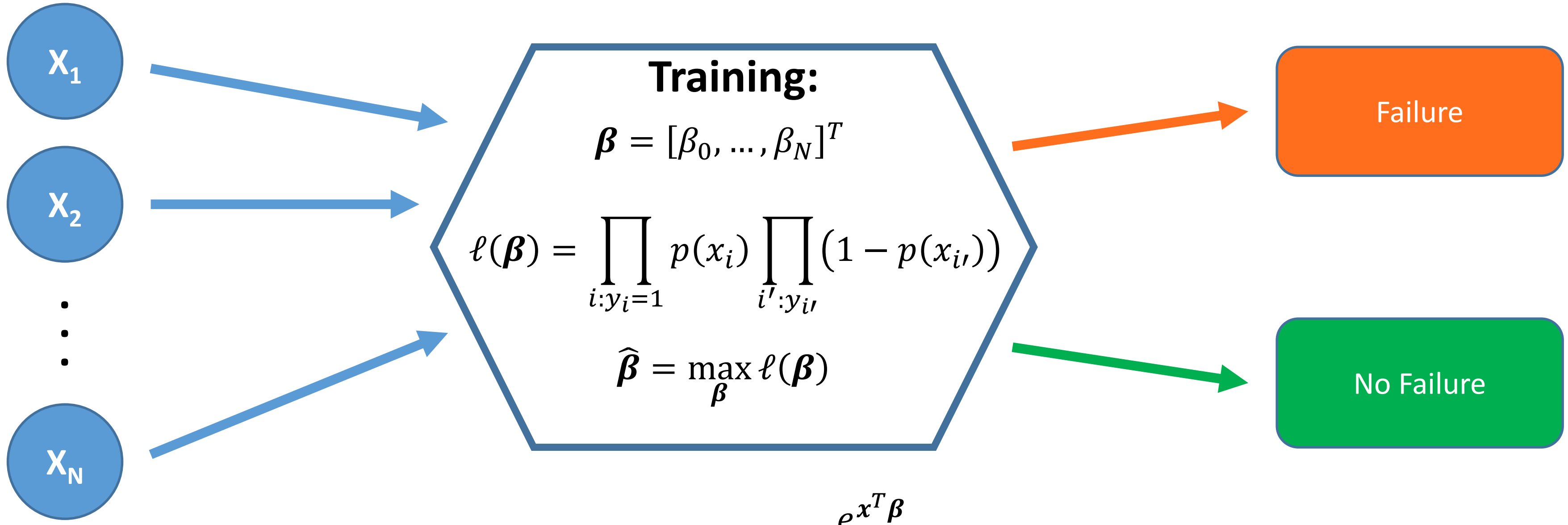
Historical Weather Measurements

Lightning [0/1] (LI)	Average Temperature [˚F] (AT)	Highest Temperature [˚F] (HT)	Relative Humidity [%] (RH)	Maximum Wind Speed [m/s] (MWS)	Wind Gust [m/s] (WG)

ENGINEERING
TEXAS A&M UNIVERSITY

ELECTRICAL & COMPUTER ENGINEERING
TEXAS A&M UNIVERSITY

Logistic Regression



$$\boldsymbol{\beta} = [\beta_0, \dots, \beta_N]^T$$

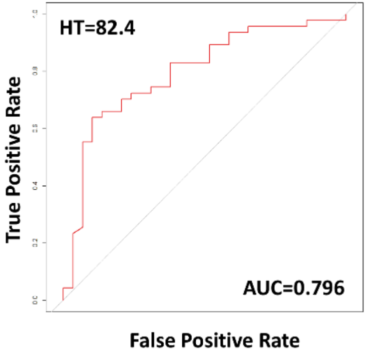
$$\ell(\boldsymbol{\beta}) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}} \left(1 - p(x_{i'})\right)$$

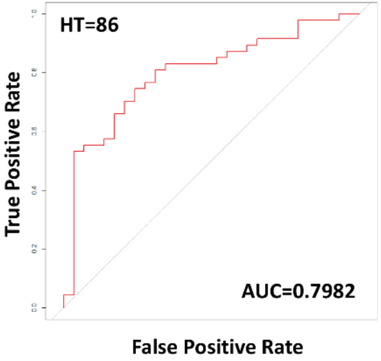
$$\widehat{\boldsymbol{\beta}} = \max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$$

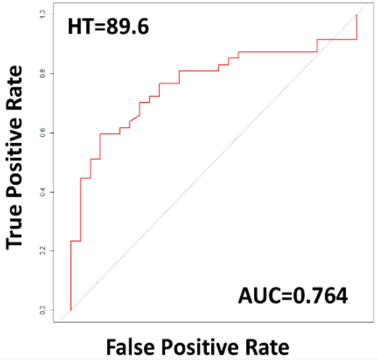
Testing: $p(\boldsymbol{x}) = \dfrac{e^{\boldsymbol{x}^T\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}^T\boldsymbol{\beta}}}$

Prediction Results

Receiver Operating Characteristics Curve (ROC)







Event vs. Prediction of Failure

Event	Failure (Y/N)	Prediction	
		Y=0	Y=1
HT 86°F or below	Y=0	113	47
	Y=1	35	190
HT 86°F - 89.6°F	Y=0	112	47
	Y=1	36	190
HT 89.6°F or above	Y=0	111	54
	Y=1	37	183

Solar Generation Forecast

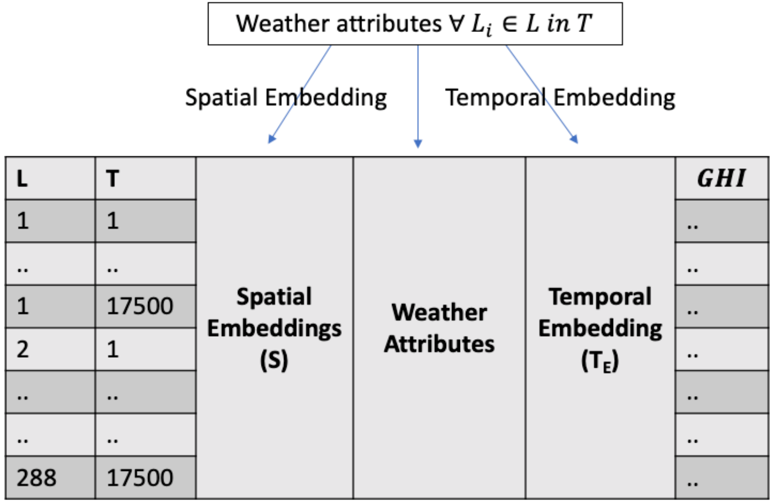
Overview

Goal: Real-time forecast for solar power generation

Challenges:

- **Challenge 1 - Complex data relationships**: spatial and temporal structural dependencies in addition to weather data.
- **Challenge 2 - Prediction horizons**: need accurate and long prediction horizons.
- **Challenge 3 - Missing data**: real datasets have missing data (temporally, spatially).

Proposed Model Intuition: Exploit <u>temporal</u>, <u>spatial</u> and <u>weather data</u> to predict solar power generation (represented by GHI).

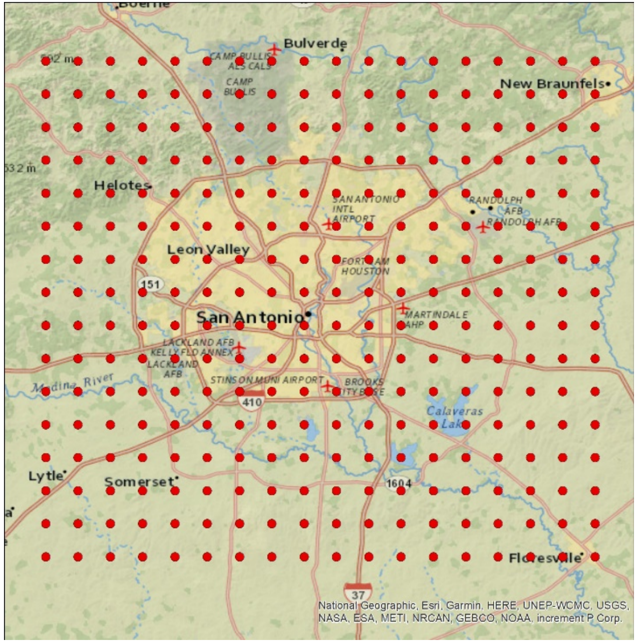


Spatial Data

Spatial relationships: Data is collected for 288 locations (3Km × 3km) around San Antonio, TX for the year 2017.

Using natural representation (longitude and latitude) introduces multiple issues:

- High dimensionality
- Adds complexity to model construction
- Capturing long-range spatial dependencies with local information is harder!



Prediction Results

GHI prediction results for 100 data points

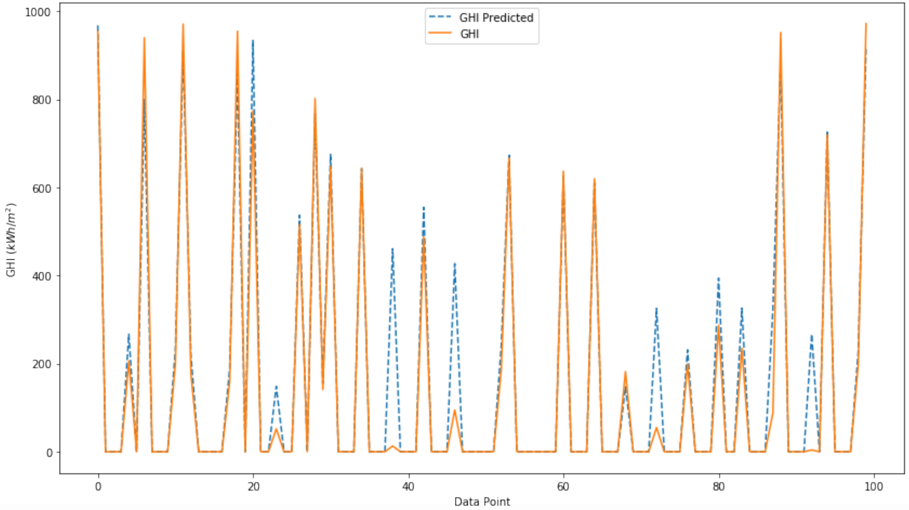


Table 2. Predictions 3 hours ahead by the summer model

Metric	R^2	MAE	RMSE
Value	0.91	42.76	92.8

Table 3. Predictions 3 hours ahead by the winter model

Metric	R^2	MAE	RMSE
Value	0.85	27.3	71.49

Table 4. Predictions 3 hours ahead by the global model

Metric	R^2	MAE	RMSE
Value	0.89	33.4	85.2

Prediction horizons

3 models show good stability over the longer test horizons.

•Summer model has mean values of

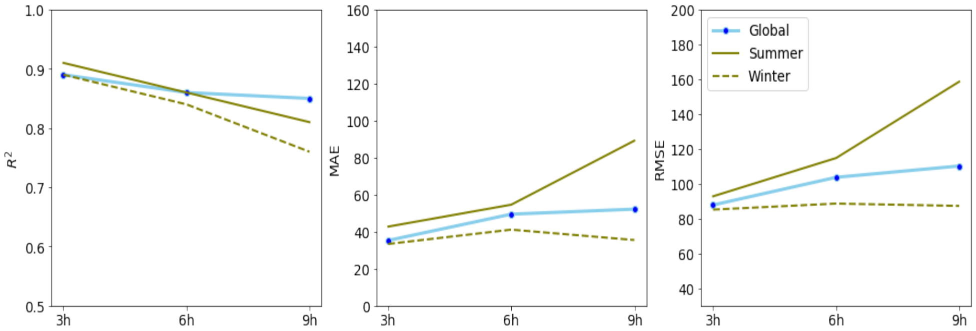
R^2=0.86, MAE=62.2, RMSE=122.1.

•Winter model has mean values of

R^2=0.83, MAE=36.7, RMSE=87.1.

•Global model has mean values of

R^2=0.87, MAE=45.7, RMSE=100.6.



ENGINEERING
TEXAS A&M UNIVERSITY
ELECTRICAL & COMPUTER ENGINEERING
TEXAS A&M UNIVERSITY

Conclusions

- The **weather testbed environment** was designed in order to integrate weather data into utility's control center
- The **spatial granularity of prediction** and **localization of outages** were improved by embeddings and **modeling of spatial interactions.**
- Type of prediction algorithm depends on the application: **Linear Regression, Logistic Regression**
- A **real-time mapping system** was developed to observe **Risk**
- Such a system allows for more **proactive and cost-effective** outage management, asset management, operation.

Publications

- T. Dokic, M. Pavlovski, Dj. Gligorijevic, M. Kezunovic, Z. Obradovic, "Spatially Aware Ensemble-Based Learning to Predict Weather-Related Outages in Transmission," The Hawaii International Conference on System Sciences – HICSS, Maui, Hawaii, January 2019.
- M. Kezunovic, T. Dokic, R. Said, "Optimal Placement of Line Surge Arresters Based on Predictive Risk Framework Using Spatiotemporally Correlated Big Data," at CIGRE General Session, Paris, France, Aug. 2018.
- T. Dokic, M. Kezunovic, "Predictive Risk Management for Dynamic Tree Trimming Scheduling for Distribution Networks," IEEE Transactions on Smart Grid, Vol. 10, No. 5, pp. 4776-4785, September 2018.
- E. Hui Ko, T. Dokic, M. Kezunovic, "Prediction Model for the Distribution Transformer Failure using Correlation of Weather Data," CIGRE 5th International Colloquium Transformer Research and Asset Management, Opatija, Croatia, October 2019.
- M. Alqudah, T. Dokic, M. Kezunovic, Z. Obradovic, "Prediction of Solar Radiation Based on Spatial and Temporal Embeddings for Solar Generation Forecast" HICSS 2020 : Hawaii International Conference on System Sciences, Grand Wailea, Maui, January 2020.

ENGINEERING
TEXAS A&M UNIVERSITY

ELECTRICAL & COMPUTER ENGINEERING
TEXAS A&M UNIVERSITY