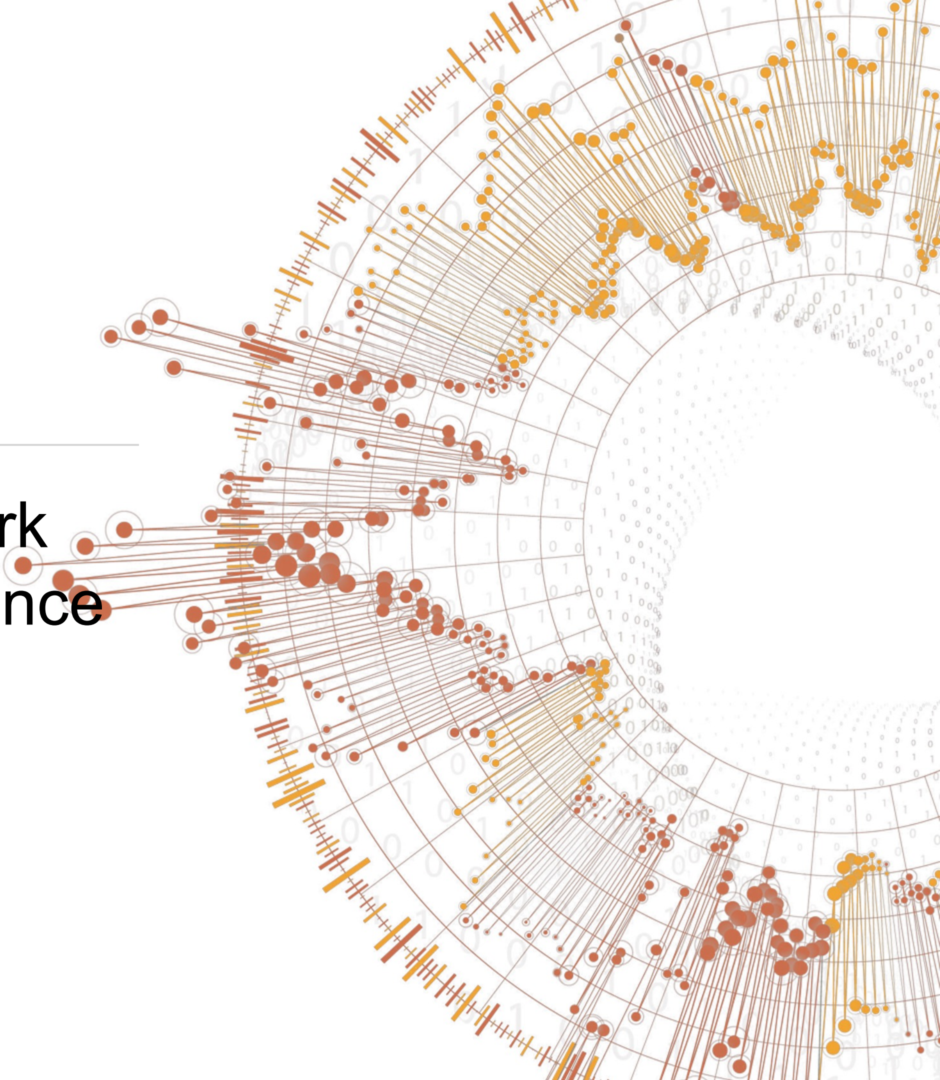
Completing power grid network graph data for property resilience modeling

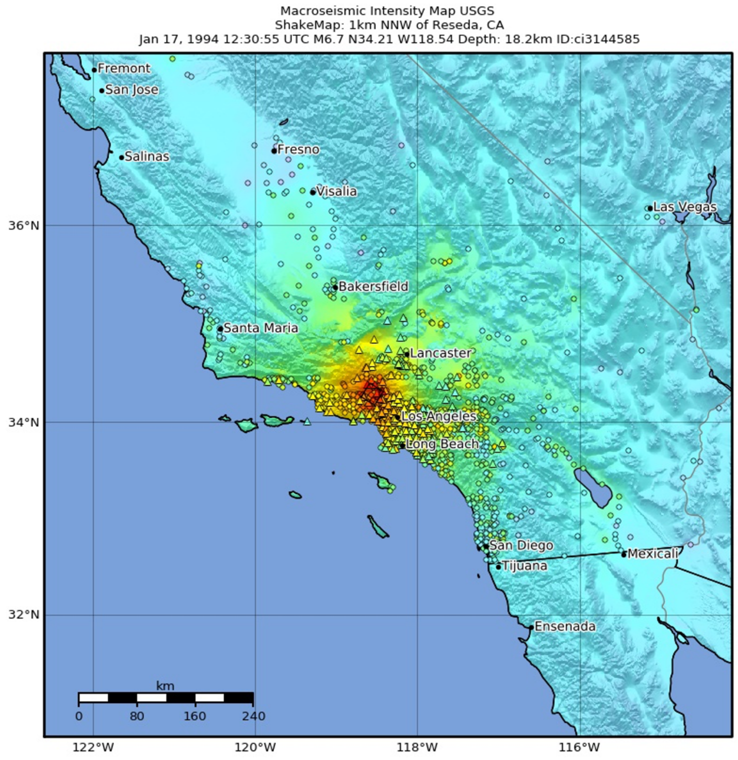
March, 2022



1994 Northridge earthquake

- M6.7 earthquake occurred on Jan. 17, 1994, 20 miles outside Los Angeles, CA, USA

- One of the most damaging earthquakes in recent US history



SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
DAMAGE	None	None	None	Very light	Light	Moderate	Moderate/heavy	Heavy	Very heavy
PGA(%g)	<0.0464	0.297	2.76	6.2	11.5	21.5	40.1	74.7	>139
PGV(cm/s)	<0.0215	0.135	1.41	4.65	9.64	20	41.4	85.8	>178
INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

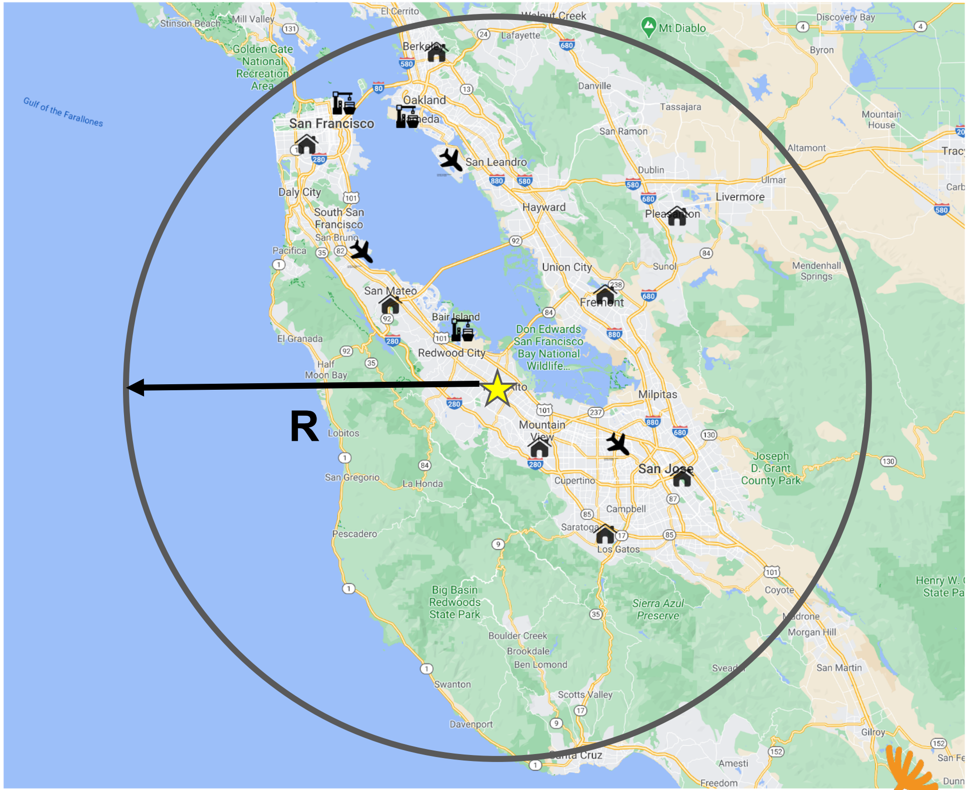
Scale based on Worden et al. (2012) Version 1: Processed 2020-06-03T00:39:44Z

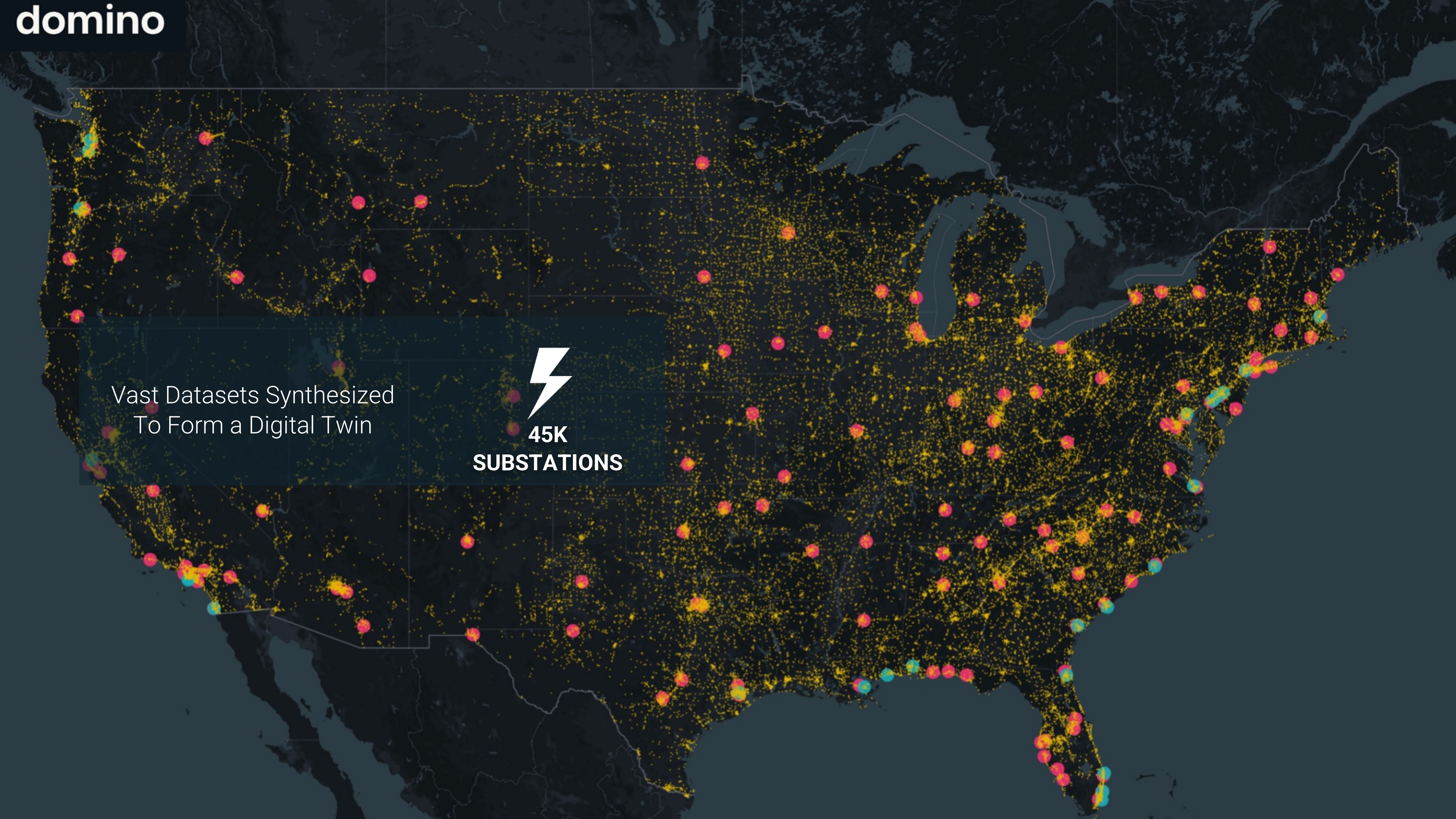
△ Seismic Instrument ○ Reported Intensity ★ Epicenter ▭ Rupture

Incorporate dependent lifeline networks

Calculate downtime for all dependent proximate causes within a given property's radius

1. **Power** – relate substations to buildings
2. **Roads & Bridges** - group by highway name and calculate maximum downtime
3. **Ports & Airports** - use maximum downtime
4. **People** - group by zip code and calculate average downtime




domino
Vast Datasets Synthesized
To Form a Digital Twin
45K
SUBSTATIONS

Resilience is the capability (of individuals, properties, infrastructures, companies, communities, or governments) to withstand & recover from *disasters*, as well as learn from past disasters to strengthen future *response* and *recovery* efforts.

To understand fully which investments and portfolios can withstand shocks to the system, investors must look at all relevant facts, not just those with which they are familiar. The missing piece of the puzzle is *resilience*.

---- Judith Rodin & Saadia Madsbjerg

Outline

- Overall methodology for property resilience estimation
- Dependent lifeline network resilience
- Synthetic infrastructure data generation: Power distribution network
  - Methodology
  - Model validation
- Final thoughts

Property resilience methodology

Resilience modeling terminology

- **Hazard**: Natural or man-made phenomenon that can cause potential *loss*. A hazard refers to the potential for a *disaster*. Hazards may or may not cause an actual disaster.
- **Peril**: Specific cause of damage/disruption, as opposed to the hazard that produced it.
- **Resilience-adjusted valuation & risk models:** Modified valuation and risk models that include exceedance probabilities and relevant conditional losses associated with property function disruption resulting in losses.
- **Fragility functions:** Map out probabilities of reaching or exceeding different damage states given a hazard intensity or the parameters that define an intensity measure.
- **Exceedance probability (EP)**: Probability a peril causes damage/disruption to the point a property's normal functioning is impaired.
- **Downtime statistic (DS)**: Duration over which a property's normal functioning is impaired.
- **Conditioning dimensions**
  - Hazards (e.g., earthquakes, hurricanes/ typhoons, cyber attacks, pandemics)
  - Perils (e.g., shaking ground, flood, landslide, wind)
  - Climate scenarios (representative concentration pathways-- RCP in terms of radiative forcing/warming measured by watts per square meter e.g., 4.5, 8.5)
  - Planning horizon measured in years
  - Dependent proximate causes (e.g., direct damage, power, transportation, community)

Objectives

- Estimate **exceedance probabilities (EP)**, **conditional downtime statistics** (**DS**), and **conditional downtime-related losses** (**EDL**) at several planning horizons
- Incorporate **dependencies** into analyses to identify non-damage interruption
- Use EPs and EDLs at relevant tenors to **resilience adjust** valuation & risk models
- Assess **climate-scenario** impact on resilience models
- Facilitate return-on-investment calculations for resilience-focused mitigation efforts
- Build analytical foundation to estimate portfolio effects and construct benchmarks
- Improve basis for pricing resilience risk and business interruption insurance

Resilience Metrics

Damage Ratio

Cost to repair damage expressed as a percentage of the building's total value. For example, a damage ratio of 0.2 means that if the building is worth $100k, the cost to repair it is $20k.

Downtime

Time during which a building, business, or facility is unable to function because of damage.

Properties

- Commercial Buildings

- Power
- Sea Ports
- Highways
- Airports
- Bridges
- People

Hazards & Time Horizons

Earthquake

Ground shaking (e.g., peak ground acceleration)

- 10 return periods
- 5 planning horizons

Flood

Inundation depth

- 2 climate change scenarios
- 15 return periods
- 5 planning horizons

Windstorm

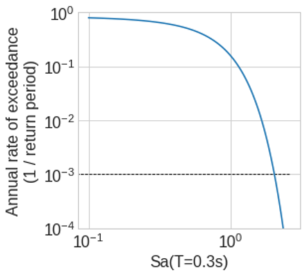
Wind speed

- 2 climate change scenarios
- 15 return periods
- 5 planning horizons

Downtime Calculation Framework

Hazard Intensity

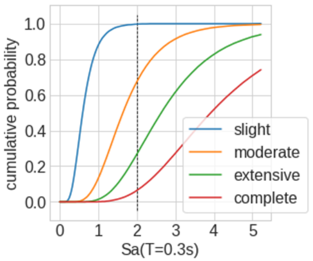
- Ground shaking (seismic)
- Inundation depth (flood)
- Wind speed (windstorm)



Example seismic hazard curve; SA at 1000 yr return period (RP) = 2.0g

Damage Model

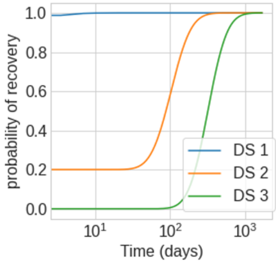
- Seismic algorithmic model (buildings)
- Fragility functions (all others)



Fragility functions predict probability of each damage state

Recovery Model

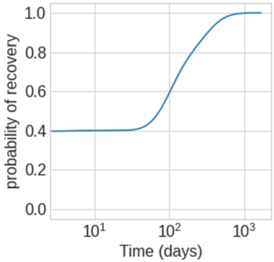
- Recovery curves for each damage state



Recovery curves predict probability of recovery vs. time for each damage state

Recovery Curve

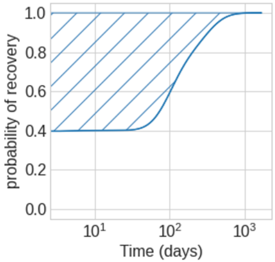
- Combine probability of damage with recovery curves for each damage state



Example recovery curve for a single asset; assumed to be a cumulative density function (CDF) for recovery time

Downtime Statistics

- Mean
- Standard deviation

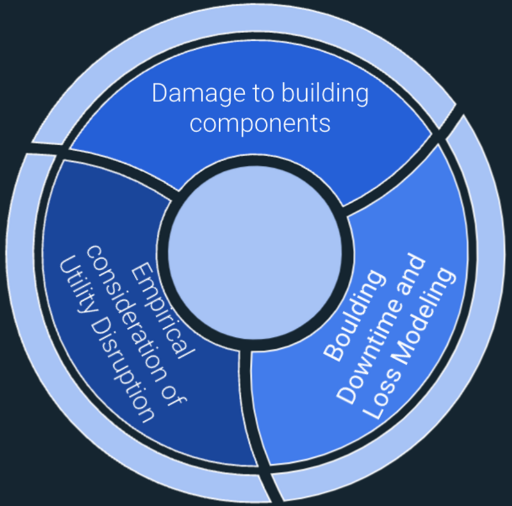


Calculate mean and std of downtime (e.g., mean is the area above the recovery CDF)

*** Calculated for each asset, return/planning period, and peril**

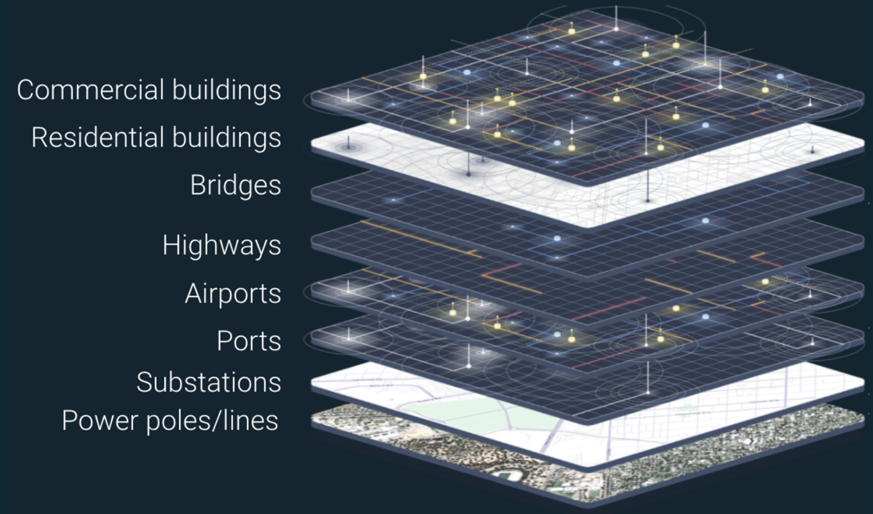
Data Feed/Analysis – *Domino and DNA*

~ 100s Data Points



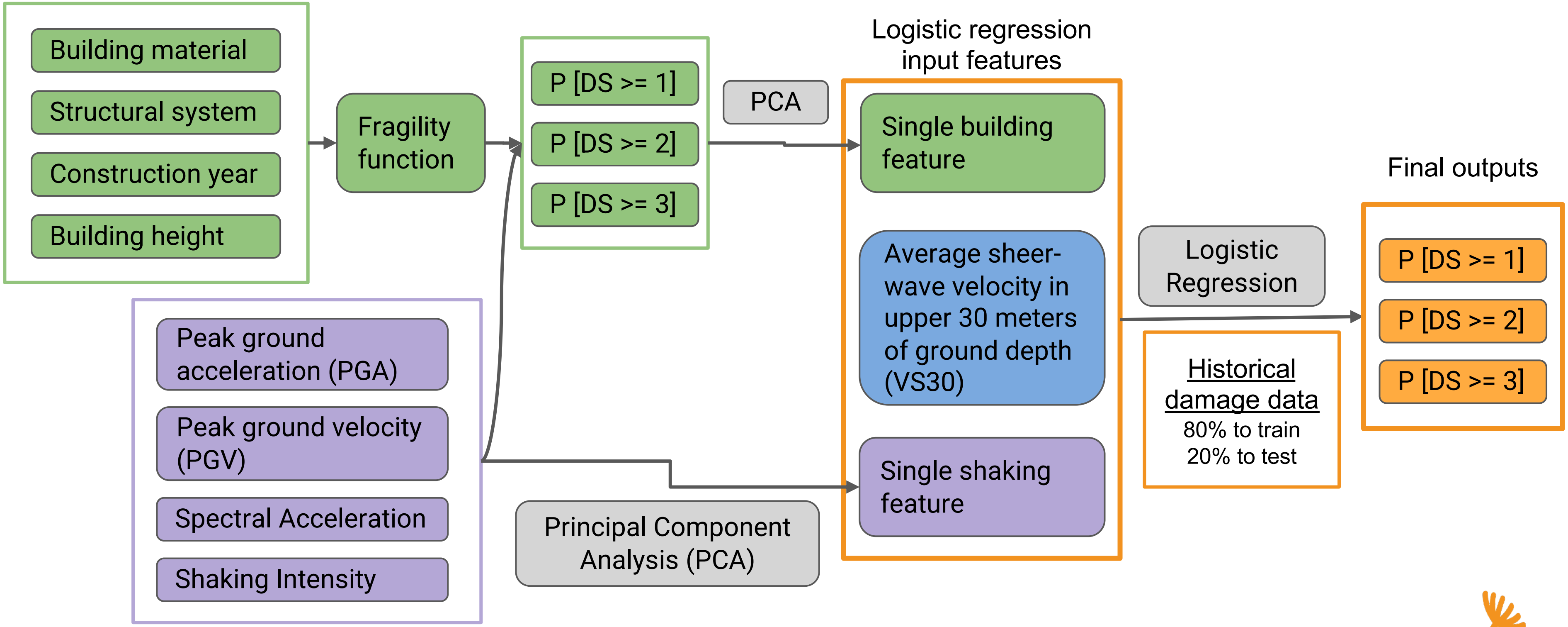
Conventional Models

Millions of Data Points



Domino/DNA Framework

Seismic machine-intelligence-enabled framework



Simplified algorithmic overview

1. Collect & curate data
2. Synthetically generate missing data
3. Estimate hazard-conditional probability of direct damage

   Estimate hazard-conditional probability of impairment due to dependent lifeline disruptions
4. Estimate EP, DS, and EL

   Incorporate climate scenarios into estimation process

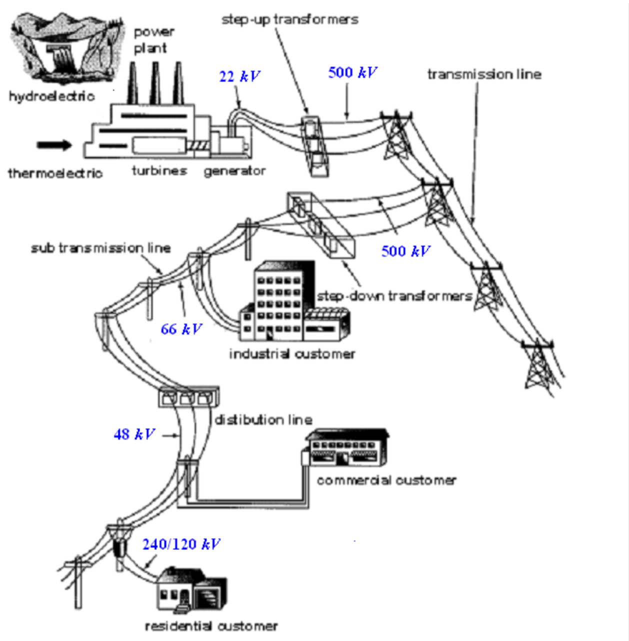
Dependent lifeline networks

Dependent lifeline networks

- Lifeline Infrastructure Systems:
  - Electric Power
  - Water
  - Communication
  - Gas and Liquid Fuels
  - Transportation

- Characteristics
  - Large geographically distributed system
  - Interlinked specialized components
  - Interdependent
  - Consists of a variety of subsystems

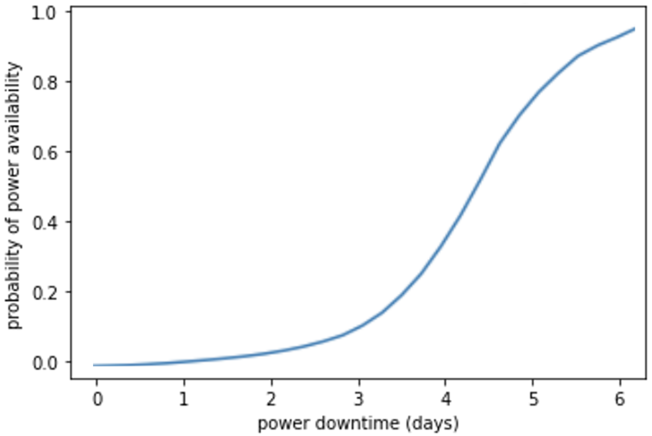
A power grid consists of generation, transmission, and distribution systems



Dependent lifeline network resiliency for buildings and how operation/interruption affects business value

- Minimize business interruption caused by catastrophic events
  - Understand probability of losing services
  - Understand duration of the interruption
  - Understand optimal mitigation strategy

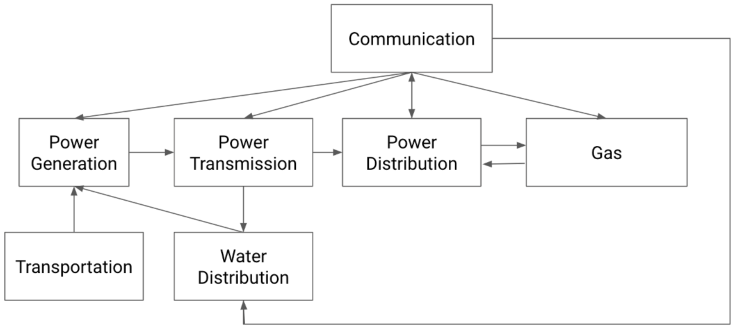
Example: power recovery curve



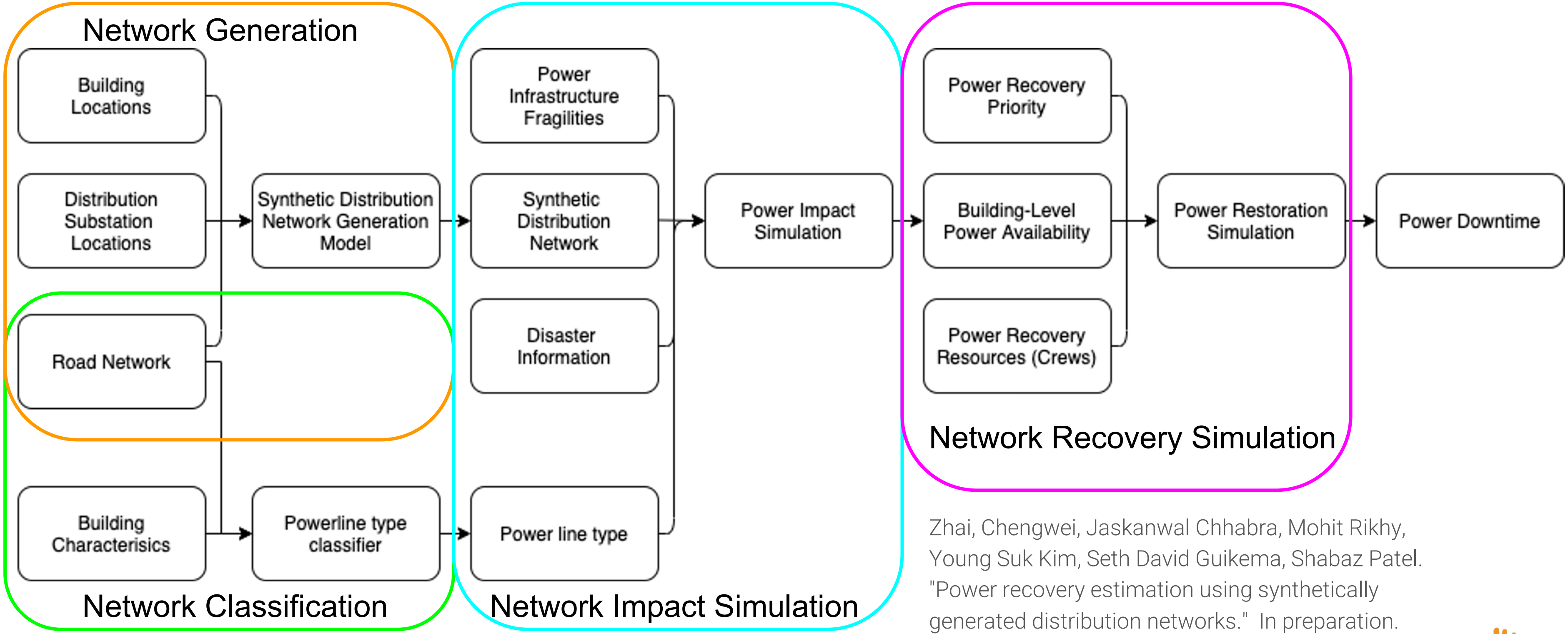
Lifeline networks risk/resilience analysis challenges

- Limited publicly available information
- Sensitive utility information
- Interconnected networks
- Limited research on specialized components
- Interdependency among different systems

Interdependency of Lifeline Systems



Synthetic power-distribution network model



Zhai, Chengwei, Jaskanwal Chhabra, Mohit Rikhy, Young Suk Kim, Seth David Guikema, Shabaz Patel. "Power recovery estimation using synthetically generated distribution networks." In preparation.

Zhai, Chengwei, Thomas Ying-jeh Chen, Anna Grace White, and Seth David Guikema. "Power Outage Prediction for Natural Hazards Using Synthetic Power Distribution Systems." *Reliability Engineering & System Safety* (2021): 107348.

Synthetic infrastructure data generation

Power grid model objectives and challenges

Objectives	Challenges
• Estimate at a high resolution (building level): Probability of outage, estimated downtime (recovery) • Generalizable and scalable	• Network data availability • Component fragility and recovery • Validation data • Compute expense

Power system components

Key components for a distribution circuit

- Station transformers
- Circuit breakers
- Power poles
- Power conductors
    - Overhead
    - Underground
    - Primary
    - Secondary
- Pole-top transformers
- Tie Switches

	Most common value	Other common values
Substation characteristics		
Voltage	12.47 kV	4.16, 4.8, 13.2, 13.8, 24.94, 34.5 kV
Number of station transformers	2	1 – 6
Substation transformer size	21 MVA	5 – 60 MVA
Number of feeders per bus	4	1 – 8
Feeder characteristics		
Peak current	400 A	199 – 630A
Peak load	7 MVA	1 – 15 MVA
Power factor	0.98 lagging	0.8 lagging / 0.95 leading
Number of customers	400	50 – 5000
Length of feeder mains	4 mi	2 – 15 mi
Length including laterals	8 mi	4 – 25 mi
Area covered	25 mi^2	0.5 – 500 mi^2
Mains wire size	500 kcmil	4/0 – 795 kcmil
Lateral tap wire size	1/0	#4-2/0
Lateral tap peak current	25 A	5 – 50 A
Lateral tap length	0.5 mi	0.2 – 5 mi
Distribution transformer size (1ph)	25 kVA	10 – 150 kVA

Typical distribution circuits parameters

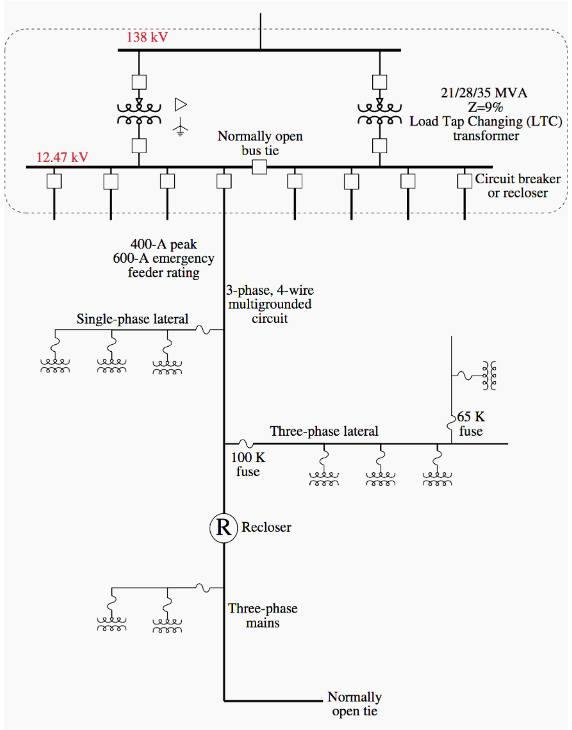
Power: Distribution-network basics

- Radial System
  - Lower cost
  - Easier prediction and control
  - Easier protection

- Arrangement depends on street layouts

- Distribution network
  - Substations (and their components)
  - Power poles/lines

Typical distribution network



Other approaches to modeling power outage/ restoration

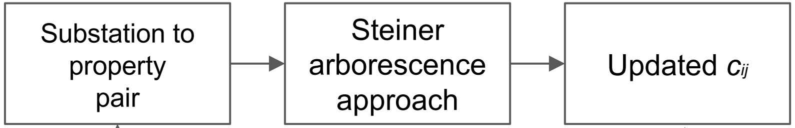
- **Statistical-based approach**
  - Use statistical-learning-based approaches to predict the severity of power outages given a specific hazard and spatial scale.
  - Pros: well-studied, fast in production, applied by utility companies
  - Cons: hazard/region specific, low resolution, insufficient data for rare events

- **Simulation-based approach**
  - Use network data to analyze component damage and impact on system
  - Pros: well-studied, high resolution
  - Cons: data availability, lack of component fragility analysis, slow in production

- **Optimization-based models**
  - Optimize planning for resources, crew teams, and restoration strategies
  - Pros: support decision making
  - Cons: typically not scalable in production

Linking substations and properties

- Properties are allocated to their closest substations [Ouyang, Min, and Leonardo Duenas-Osorio, 2014]
- Simple approach suffers from the following:
  - Does not account for redundancy provided by circuits from nearby substations
  - Uncertain as to the actual connected substation
- Modify Network Flow Optimization if the cost to connect a building to one substation is known. To identify connection cost, require solving the Steiner Arborescence Problem.



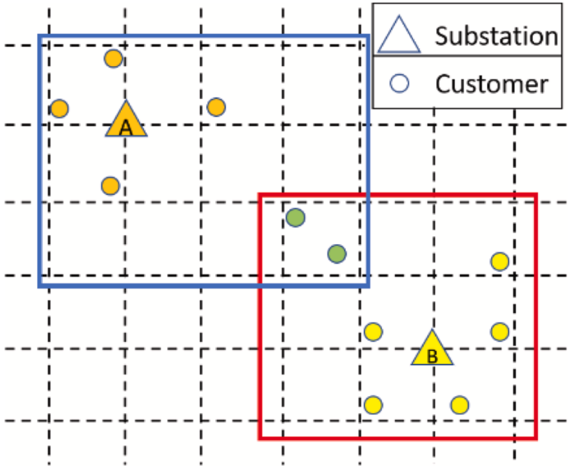
$$\text{minimize} \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij} \quad \text{(A)}$$

$$\text{s.t.} \sum x_{ij} \leq S_i \quad \text{For All } i \in S \quad \text{(B)}$$

$$\sum x_{ij} \geq D_j \quad \text{For All } j \in D \quad \text{(C)}$$

$$x_{ij} \geq 0 \quad \text{For All } i \in S, i \in D \quad \text{(D)}$$

- Simplified approach: Allow for a wide-range of possible substation capacity despite high uncertainty.
- Allow possibility for multiple substations to serve a property

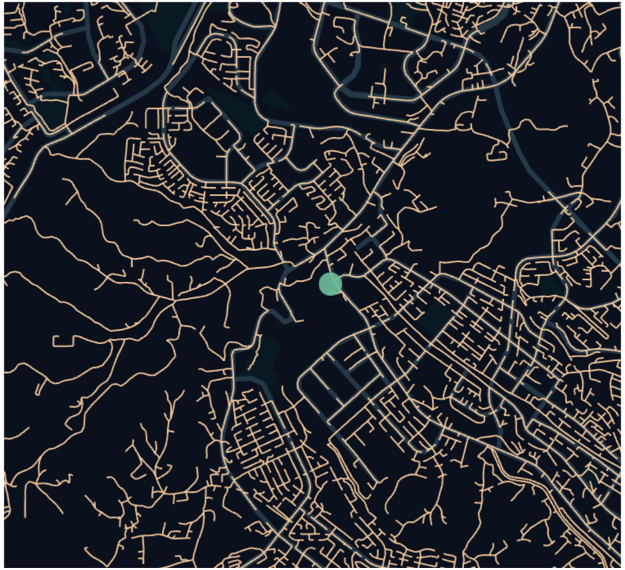


Synthetic infrastructure data generation to address sparsity

- Generate features for specific estimation requirements, which are otherwise not available or hard to acquire

- Use case:
  - Create synthetic lifeline network
    - **Power Distribution network**
    - Drinking water system
    - Communication system
  - Simulate lifeline resilience using synthetic network

Example of a synthetic distribution network for one substation



Power: Distribution-network generation

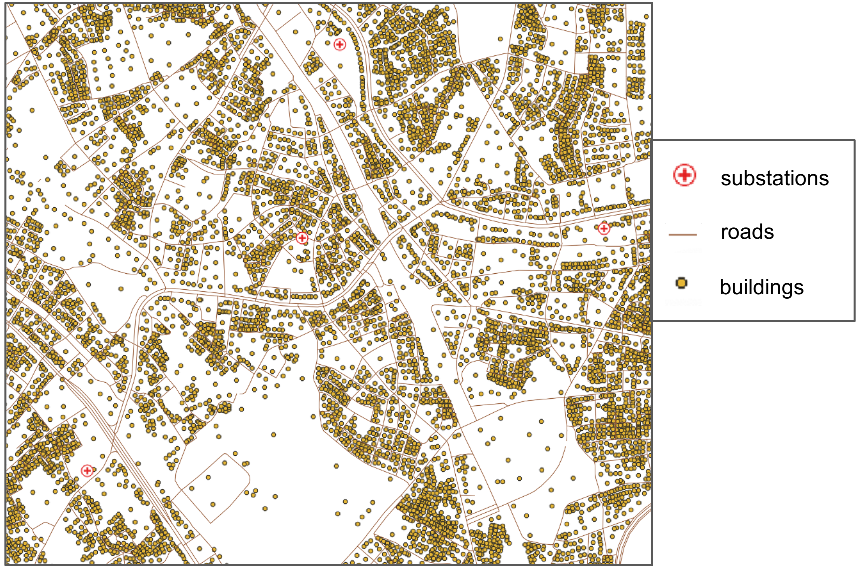
Model Objective

- Find most realistic distribution networks that enable power to flow from substations to buildings

- Challenges
  - Scale
  - Validation

- Solution
  - Break down original problem into sub-problems linked to other features e.g., buildings
  - Validate against ground truth utility networks

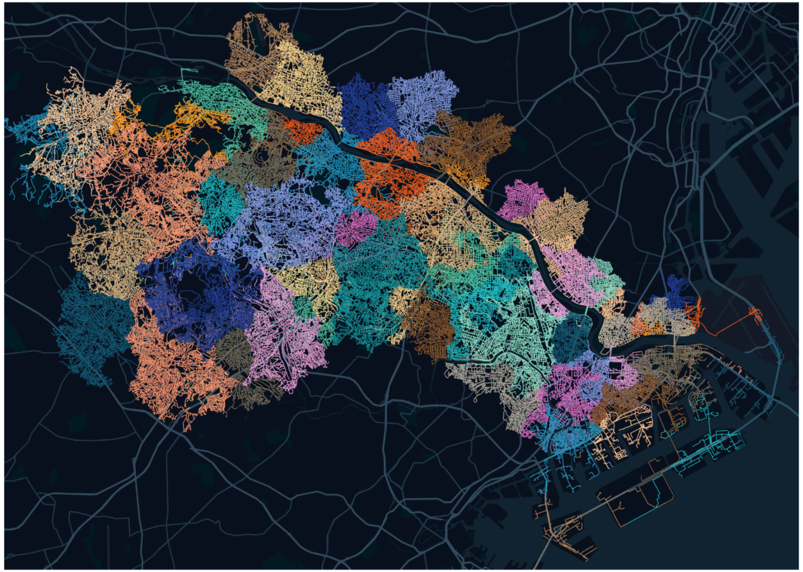
Model Inputs Example



Power network model

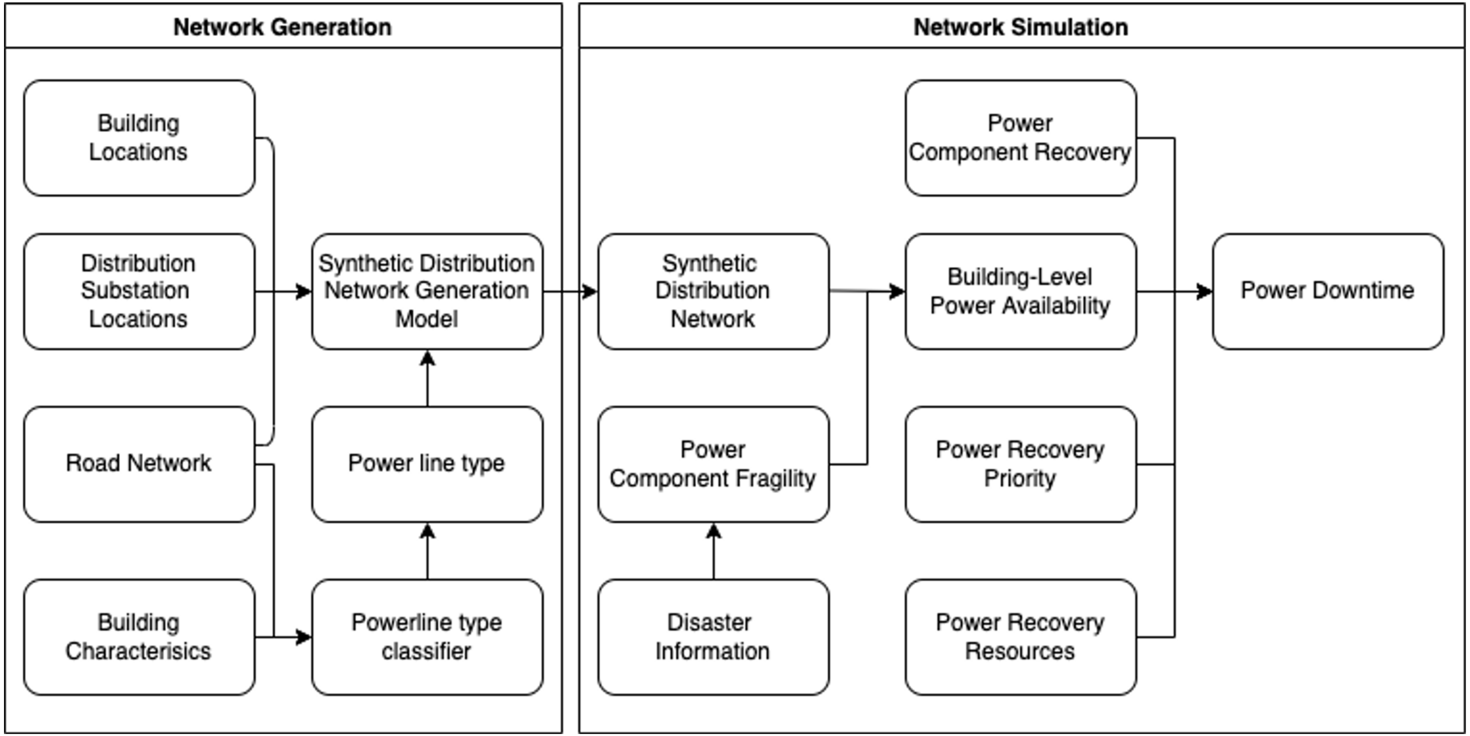
- Physics-based + machine learning to generate **synthetic network**
- Estimate building-level power recovery based on synthetic distribution network
- Develop component fragility and recovery functions
- Simulate efficiently at a high resolution
- Support multiple perils

Synthetic network generated for a city in Japan
Each color represents the service territory for one substation



Synthetic power distribution network model

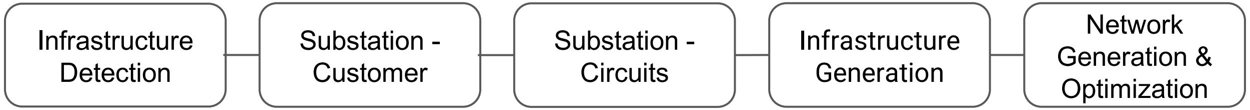
Overview Model Flow



Power: Distribution network generation model

Model Breakdown

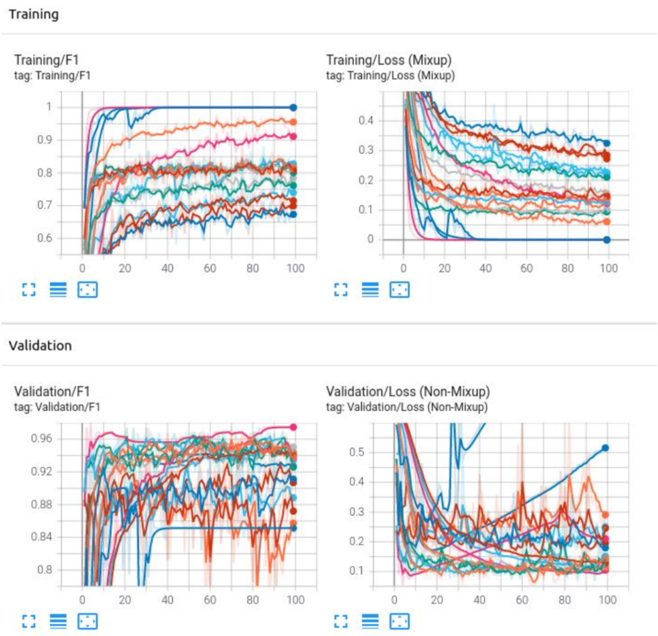
- Infrastructure detection
  - Identify key infrastructure such as power substations
- Substation - Customers
  - Identify service territory of each substation
- Substation - Circuits
  - Identify number of circuits for each substation
- Infrastructure Generation
  - Model infrastructure
- Network Generation & Optimization
  - Generate network delivering power to customers



Infrastructure: Collect data

- Identify power infrastructure locations from satellite images to improve publicly available datasets, e.g.,
  - **Substations**
  - Transformers
  - Transmission Towers
  - Distribution Poles
- Substation Dataset
  - US Homeland Infrastructure Foundation-Level Data (HIFLD)
  - Japan
- Satellite Image Dataset
  - USGS 150 / USGS 300 / Google Maps 150 / Google Maps 300
- Model
  - Transformers

Example - model hyperparameter tuning

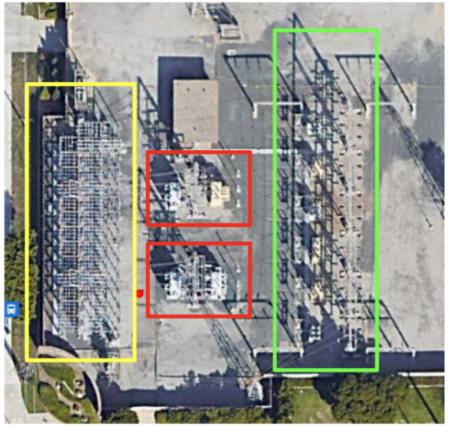


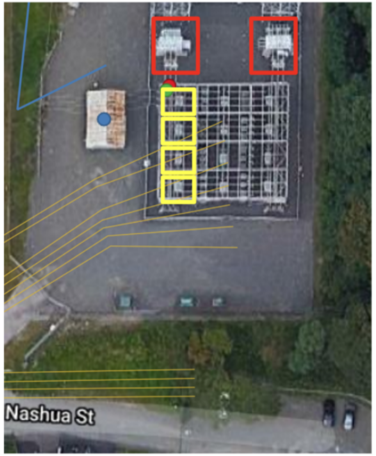
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

Infrastructure Detection — Substation - Customer — Substation - Circuits — Infrastructure Generation — Network Generation & Optimization

Infrastructure: Label data

Example Labeled dataset for substations



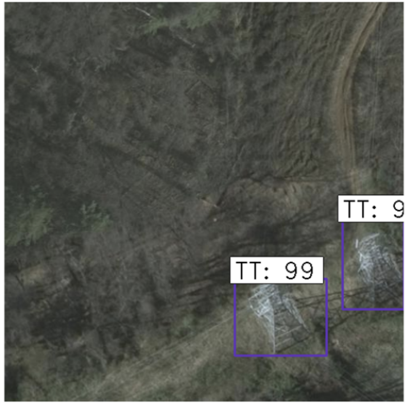


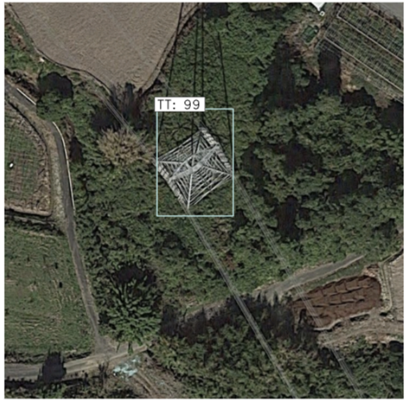
Transformer

Transmission Input

Circuits

Fast Regional-CNN to identify transmission towers





Training

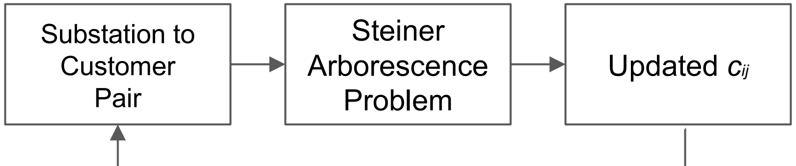
Testing

Infrastructure Detection — Substation - Customer — Substation - Circuits — Infrastructure Generation — Network Generation & Optimization

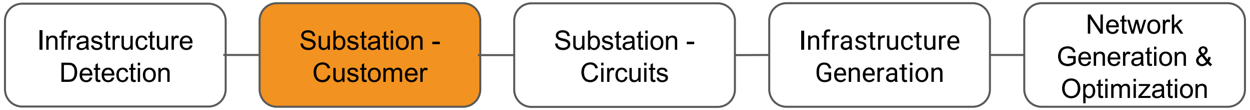
Substations: Conventional approach to assigning buildings

- In some studies, buildings are allocated to closest substations [Ouyang, Min, and Leonardo Duenas-Osorio, 2014]
- Issues
  - Redundancy
  - Uncertainty

Conventionally, solved recursively

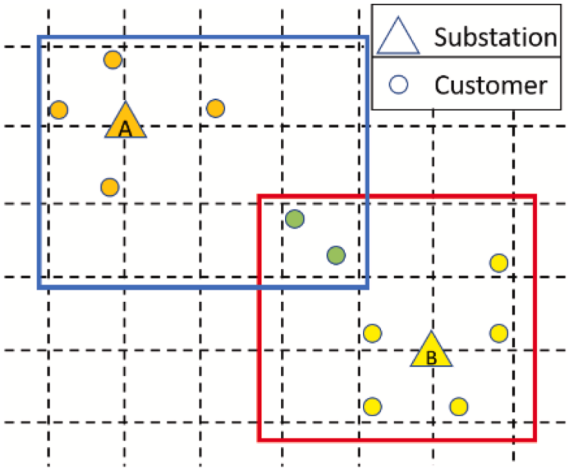


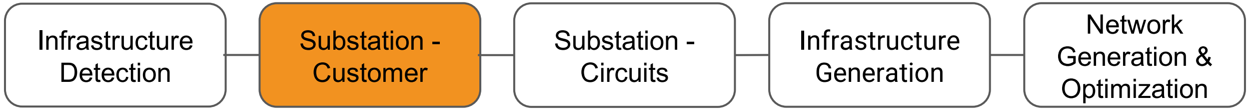
$$
\begin{aligned}
\text{minimize} \quad & \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij} \quad \text{(A)} \\
\text{s.t.} \quad & \sum x_{ij} \le S_i \quad \text{For All } i \in S \quad \text{(B)} \\
& \sum x_{ij} \ge D_j \quad \text{For All } j \in D \quad \text{(C)} \\
& x_{ij} \ge 0 \quad \text{For All } i \in S, i \in D \quad \text{(D)}
\end{aligned}
$$



Substations: Connect buildings

- High uncertainty and wide-range of substation capacity leads to relaxing capacity constraint
- Introduced possibility a building can be served by multiple substations given potential feeder switches
- Assign buildings to closest network distance substation and nearby substations determined algorithmically



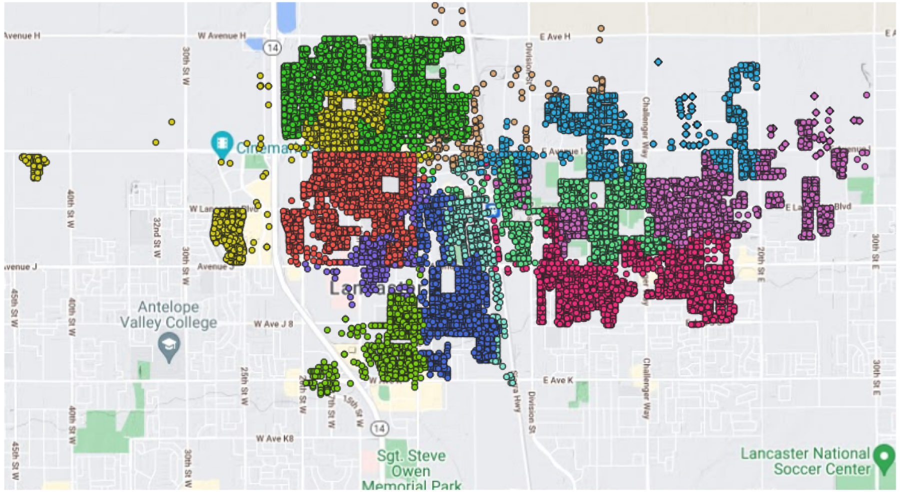


Substation: Circuits

- Estimate number of circuits, and determine buildings served by each circuit

- Unsupervised learning model addresses buildings served by each substation using the following features to cluster customers into groups. Each group is assumed to be served by one circuit
  - Geolocation
  - Year built
  - Occupancy Type
  - Building Size

Example of expected outputs
Each color represents a cluster of buildings served by one circuit



Infrastructure Detection — Substation - Customer — **Substation - Circuits** — Infrastructure Generation — Network Generation & Optimization

Infrastructure: Synthetic data generation

Overhead and underground system characterization essential for resilience modeling:

- Overhead system
  - Power lines
  - Power poles
  - Pole top transformers
- Underground system
  - Power lines
- Network generation algorithm generates data

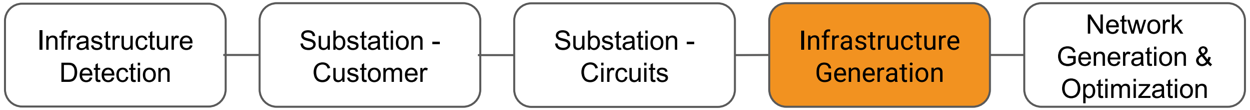
Overhead vs Underground power distribution

Infrastructure Detection — Substation - Customer — Substation - Circuits — **Infrastructure Generation** — Network Generation & Optimization

Infrastructure: Overhead (OH)/ Underground (UG) system classification, 1 of 2

- Computer vision (CV) can help; however, CV-only approach is not scalable
- Approach
  - Build training dataset
  - Collect nearby building features
  - Use classification-and-regression-trees (CART) to classify line type near a building
  - Majority vote on the line type of a road segment when multiple buildings are in training data for that segment

OH/UG system prediction

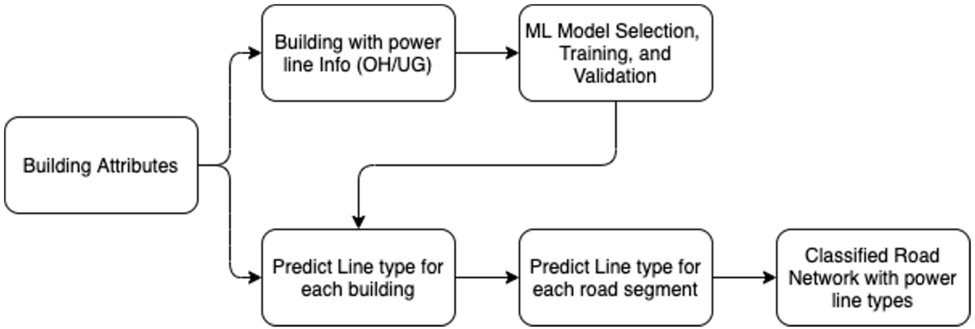


Infrastructure: Overhead (OH)/ Underground (UG) power-line classification, 2 of 2

- Response variable:
  - Power-line type of a buildings’ nearest road segment.
  - Values: Overhead, Underground
- Covariates:
  - Year built
  - Square footage
  - Value
  - Land use type

- Model Selection: Binary Classification Problem
- Holdout tests accuracy: 80-90%

OH/UG prediction flow chart



Network generation & optimization

Steiner Arborescence problem

- Steiner Arborescence Problem
  - Starting from a root node (substation), find the minimum-cost Steiner tree that connects all terminal nodes (customers).
  - NP-hard
- Heuristics
  - Steiner tree approximation (stem from the substation)
  - K-mean + Steiner Tree Heuristic
  - Steiner Arborescence Heuristic (shortest path) -> Chosen

Example formulation of Steiner Arborescence Problem

$$\textbf{Minimize} \sum_{a \in A} c_a w_a$$

$$(IP_{wf}) \qquad \textbf{subject to:}$$

$$(w, f) \in (Q_{wf} \cap (\mathbf{Z}^{|A|} \times \mathbf{R}^{|A \times T_r|}))$$

where

$$Q_{wf} = \{(w,f): f^k(\delta^+(i)) - f^k(\delta^-(i)) = \begin{Bmatrix} 1 & i = r \\ -1 & i = k \\ 0 & i \in V \setminus \{k, r\} \end{Bmatrix} \text{ and } k \in T_r$$

$$f_a^k \le w_a \qquad a \in A \text{ and } k \in T_r$$

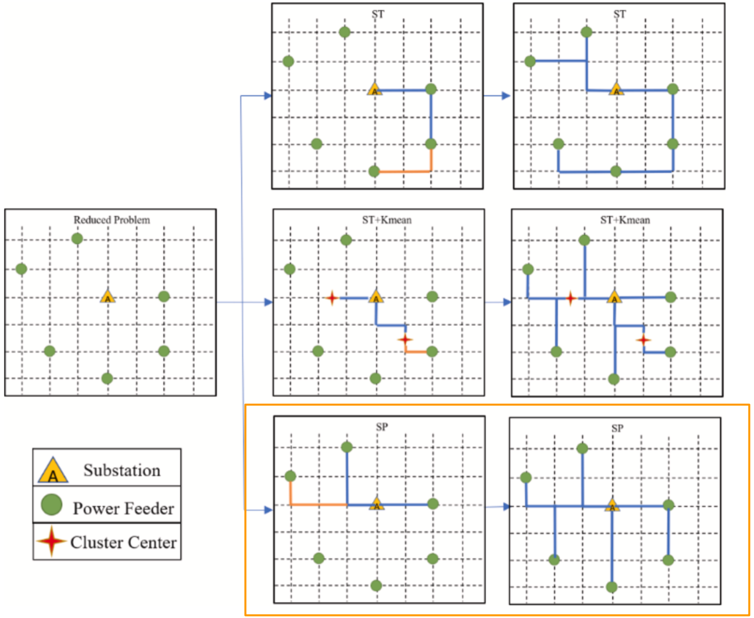
$$f_a^k \ge 0 \qquad a \in A \text{ and } k \in T_r\}.$$

Infrastructure Detection — Substation - Customer — Substation - Circuits — Infrastructure Generation — **Network Generation & Optimization**

Power: Distribution network generation
Model selection

Metrics (compare with ground truth network data)

- Computational Efficiency
- Total Length
- Nodal Degree Distribution
- Betweenness Centrality
- Network Distances from Buildings to Substation
- Pearson Correlation Coefficient
- Simulation Performance



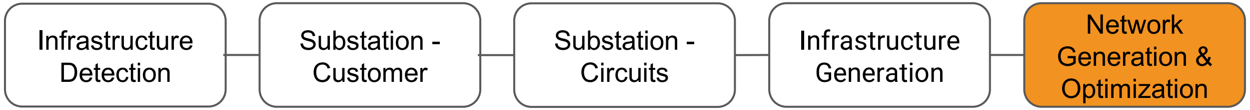
Infrastructure Detection — Substation - Customer — Substation - Circuits — Infrastructure Generation — **Network Generation & Optimization**

Network generation & optimization

Algorithm

Define largest component with all potential edges (power lines) and nodes (e.g. power poles) generated from the road network to be G = (N,E). G defines potential power lines and power poles that deliver power to a building for one substation. Define substation nodes as n_{sub}.

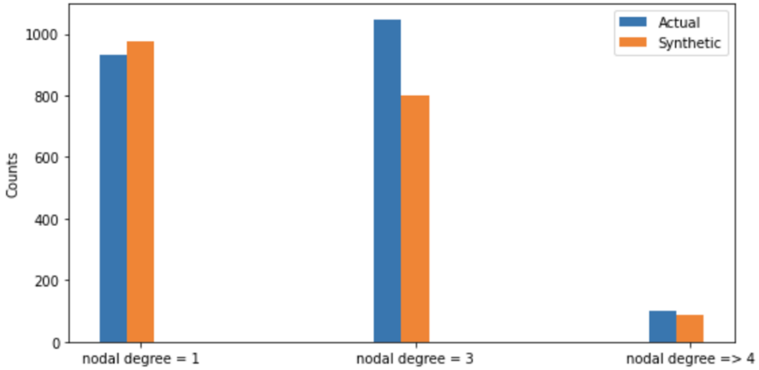
- Step 1: Split customer buildings into residential and non-residential. Associate each building with the closest node on the network as terminal nodes $\{N_{res}, N_{nres}\}$.
- Step 2: For each node in $\{N_{nres}\}$, find weighted shortest path $e \in E$ to substation node n_{sub}. Repeat process until all nodes in $\{N_{nres}\}$ are connected to n_{sub}. New graph defined as G_{main}.
- Step 3: For each node in $\{N_{res}\}$, find weighted shortest path (secondary branch) $e \in E$ to G_{main}. Repeat process until all nodes in $\{N_{res}\}$ are connected to G_{main}.
- Complete network defined as G_{opt}.



Validation

Power: Distribution network generation
Network characteristics for an example substation

- Length: Relative difference of Length
- Average Nodal Degree: Node degree is number of edges adjacent to node
- Nodal Degree Distribution: distribution of nodal degree for nodal degree = 1,3,4+.
- Pearson Correlation (strongly correlated with power outage risks)

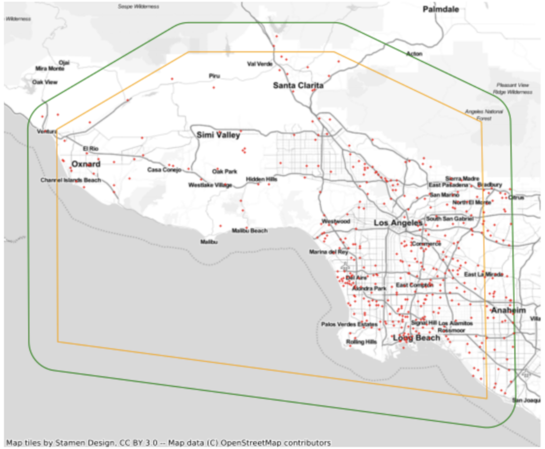
Network Characteristics	Actual	Synthetic
Length	380 km	300 km
Average Nodal Degree	2.04	2.00
Nodal Degree Distribution		
Pearson Correlation Coefficient of Network Distance	0.94	

Power validation based on Northridge earthquake

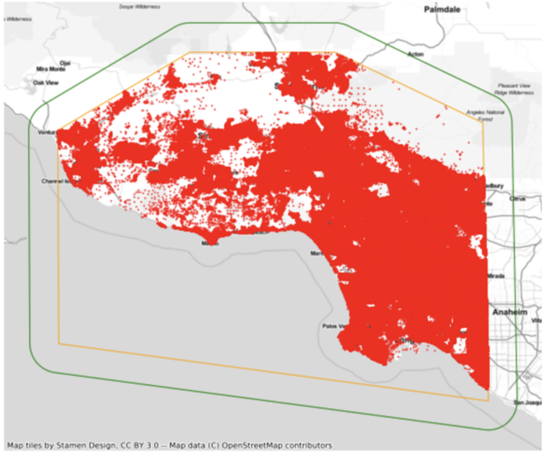
Compare power recovery estimates with observed and publicly available empirical data and update appropriately.

Power network generated using the data shown below:

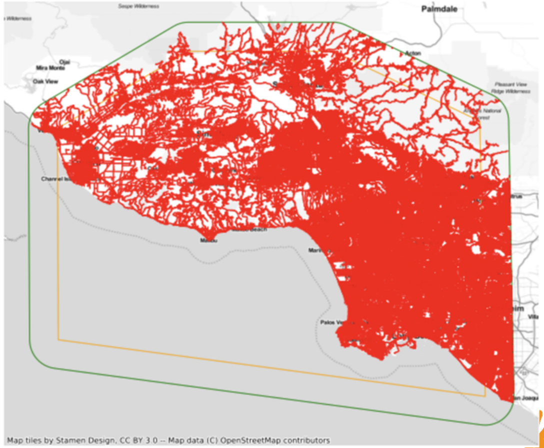
#Substations: 432



#Buildings: 2,210,408



#Roads: 351,131



Power: Distribution network generation

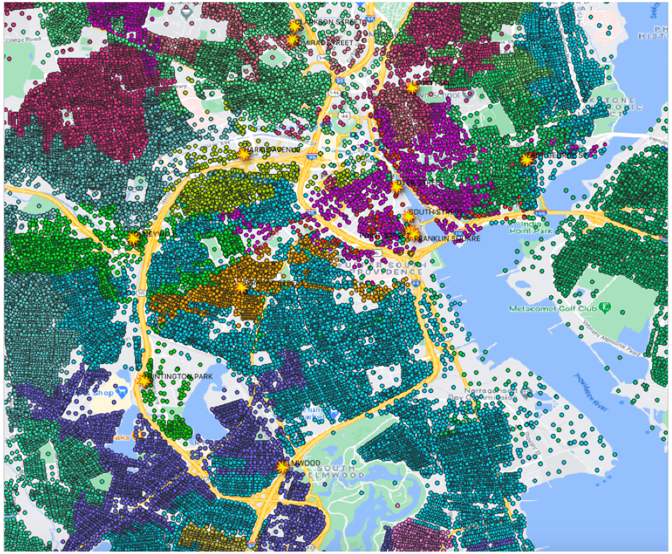
Substation - Targeted validation

- Southern California Edison
    - primary electricity supply company for much of Southern California
    - Serve more than 15M customers
    - 700+ substations

	Southern California Edison
Substation - Customer Accuracy	92.5%

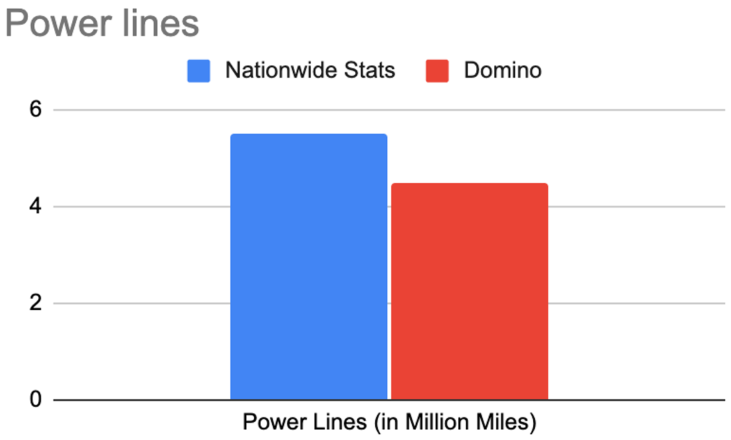
Accuracy measured as percentage of customers served by primary distribution substation per the model.

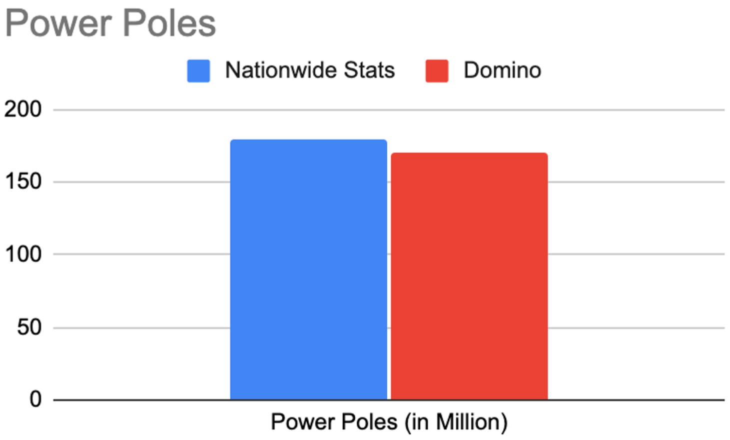
Primary substation of customers



https://drpep.sce.com/drpep/

Power: Distribution network generation
Nationwide validation





* https://www.scientificamerican.com/article/what-is-the-smart-grid/
** https://www.electrocuted.com/2016/09/08/utility-pole-facts/

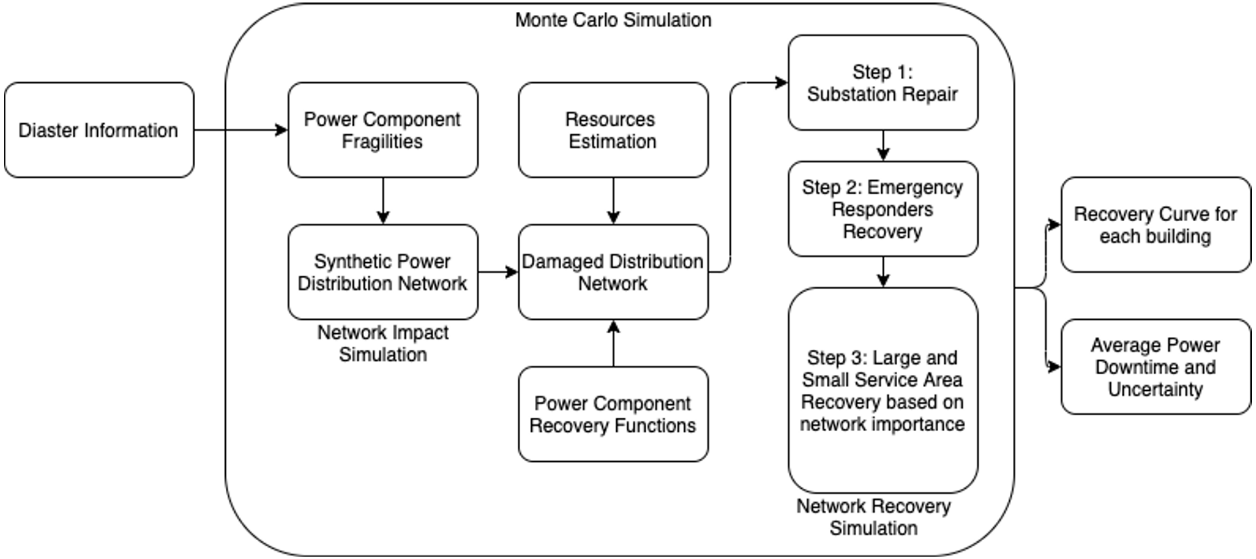
Final remarks

References

- Zhai, Chengwei, Thomas Ying-jeh Chen, Anna Grace White, and Seth David Guikema. "Power Outage Prediction for Natural Hazards Using Synthetic Power Distribution Systems." Reliability Engineering & System Safety (2020): 107348.
- One Concern, "Synthetic Power Distribution Network Model Technical Documentation." (2021)

Appendix

Power-distribution network simulation



Network Generation: Algorithm 1

Steiner Tree heuristic

Evaluated **three** different algorithms to solve the problem with different assumptions.

Original road network defined as an undirected graph G = (N,E), consisting the set N of all the potential power nodes we created from 3.2.3 and the set E of all the road segments that connect the power nodes. We define another undirected graph G' = (N' , E') to be the Steiner tree we are looking for, N' ∈ N, E' ∈ E. The weight of each edge is the length of the power line. We define set NI ∈ N of all the power feeders. (H. Takahashi and A. Matsuyama)

Algorithm 1

Steiner tree to generate distribution layout.

Step 1: Select a random vertex $s \in N_I$, and find a vertex $t \in N'$, $s \neq t$ that gives the shortest weighted path e_{st} to s. Add e to E' and all the vertices in e_{st} to N'. This gives a starting tree that connects s and t. Remove s and t from N_I.

Step 2: Search from all the vertices in N_I and find a vertex u such that the weighted path e_u from u to G' is the shortest. We then add e_u and all the vertices in e_u to N'. We then remove u from N_I.

Step 3: Repeat Step 2 until all the vertices in N_I have been connected to G' and N_I becomes empty.

Network Generation: Algorithm 2

K-mean + Steiner tree heuristic

Second method (Algorithm 2) we examine is called K-mean clustering Steiner tree. The intuition of this method is trying to imitate the development progress of communities. We assume that the development of each substation cluster's distribution system begins with building major power lines from the substation to each neighborhood (i. e., grouping of buildings). Then the power lines within each group of buildings are added, connecting to the original main power line. The algorithm changes as following.

Algorithm 2

Steiner tree with K-mean clustering to generate distribution layout.

Step 1: Spatially cluster customers within each substation cluster with a K-mean algorithm and determine the best number of clusters based on silhouette score. We use the closest power node to each cluster center in G as our centers for communities, i.e. n_1, n_2, n_3.

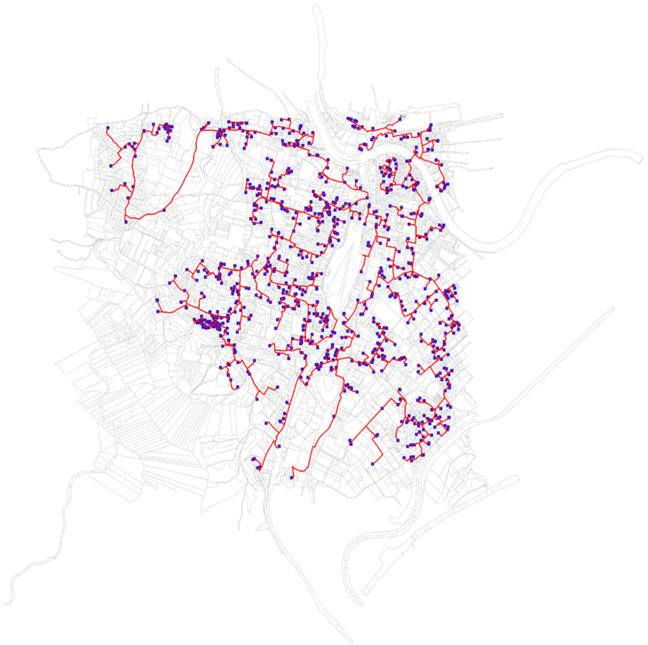
Step 2: Then connect the substation's closest power node n_s to each substation cluster's power nodes n_1, n_2, n_3 with the shortest weighted path on G. Then add these vertices and arcs in G'. Remove substation power nodes and community cluster power nodes from N_I.

Step 3: Apply the Steiner tree algorithm to include all the important nodes and paths into G'.

Network Generation: Algorithm 3 (chosen)

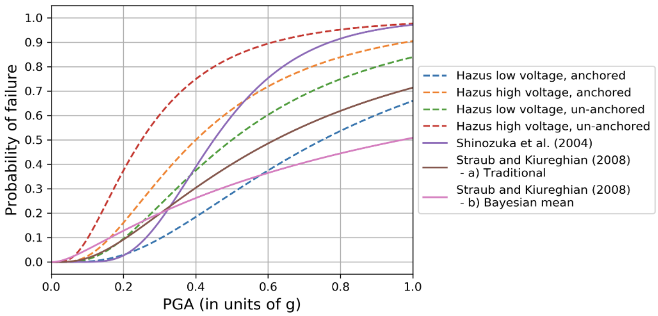
Steiner Arborescence heuristic (shortest path)

- Shortest weighted path approach chosen
- Reflects a power distribution's network typical topology
- Consider power feeders (n1, n2 , n3 ...) and the substation power node n_s , then find shortest path from each power feeder node to a substation power node and include path in G'.
- Method can be viewed as special case of the second method when number of clusters is equal to the number of customers.

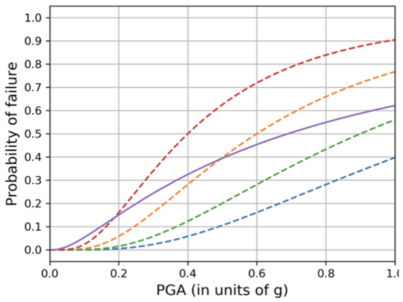


Seismic Fragility Development of Substations

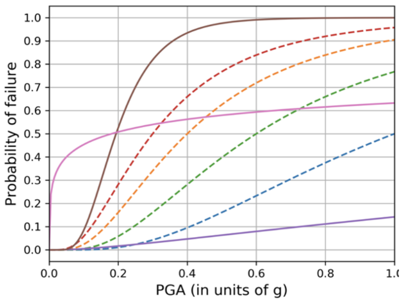
Transformers



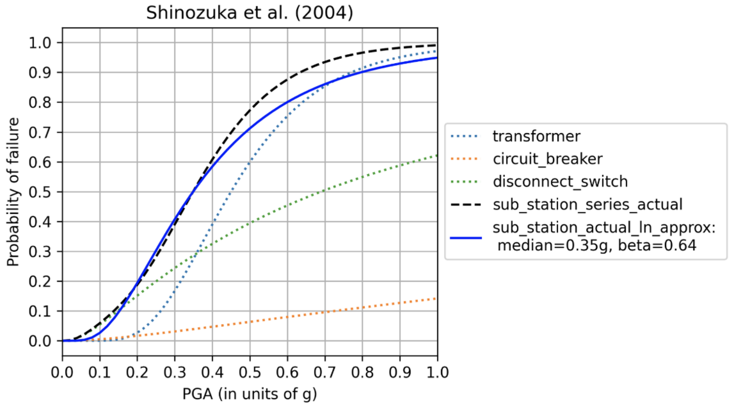
Disconnect switch



Circuit breaker

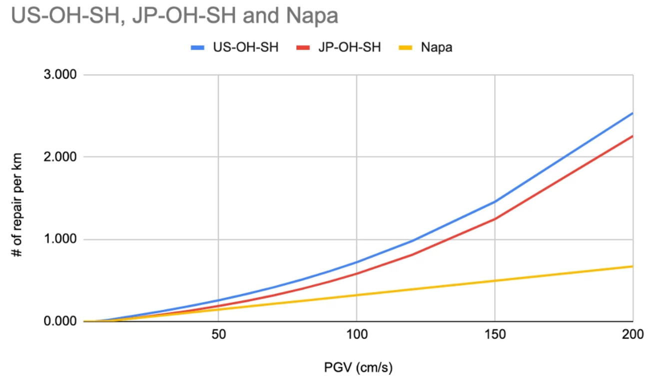


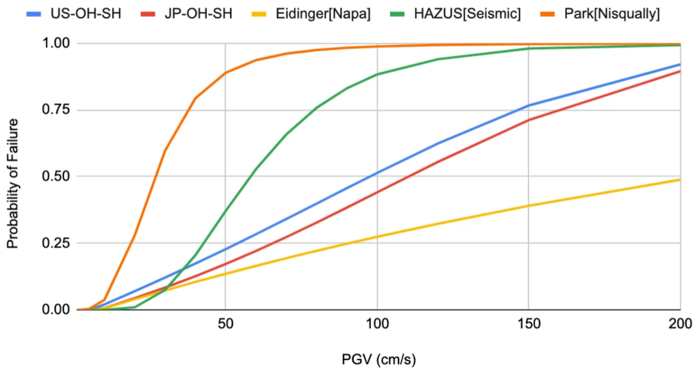
- Goal: Develop a substation fragility function to determine its post-earthquake functionality
- Development based on component fragility functions in literature.



Seismic Fragility Development of Power Lines

- Estimate power line repair rate (# repairs per km) based on:
    - Ground shaking (peak ground velocity) model by Eidinger et al. (2016) augmented to account for damage to adjacent buildings
    - Ground displacement (permanent ground displacement) by Eidinger et al. (2016)





Step 1: Substation Recovery

In a damaged distribution system, substation is having the highest priority to be repaired. Under extreme disaster events, substations might be shut down to prevent the power grid. Therefore if we estimate the substation is having a relatively high chance of being damaged (>X%), we assume the substation is closed due to emergency. Substation repair time is different for each distribution substation and consists three components: the transmission downtime, the inspection time and the actual repair time.

The first component to consider is the availability of upstream of power delivery through the transmission system. Typically the transmission system is built with reliability protocols (e.g.,N-1) and should be recovered very fast. Ttrans is the time it takes for the transmission system to be cleared with damages and can deliver power to the distribution substation.

The second part of time to consider is the inspection time to these high risk substations, Ti. This time is to inspect the damage situation of a substation to evaluate if it can be turned back online for operation. We assume all the inspection starts at the same time in parallel for all substations. After the inspection, if there is no damage to the substation, then the substation will be back to operating. Customers within the distribution substation's service area will get power back if they do not have any damage in their distribution network.

The online time for a substation is also dependent on its damage situation. If there is a damage in the substation, there will be an extra repair time Tr for the substation. Tr is determined as a function of, a) crew availability, b) inventory stock, c) damage severity of the equipment.

$$Tsub = Ttrans + Ti + Tr$$

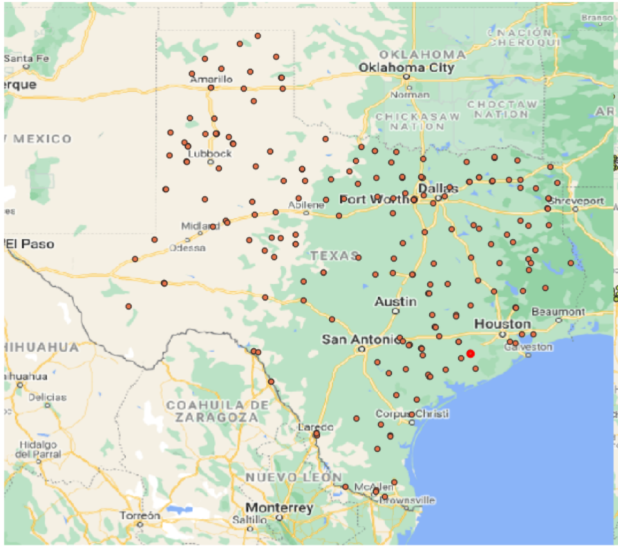
Step 2: Emergency Responders Recovery

Emergency responders are having the highest priority in power restoration as the functionality of these facilities are critical amid disasters. Power recovery process often starts with these facilities after the substation is repaired. In our simulation model, we assume that all the crew teams will focus on repairing the power infrastructure or components that are necessary to bring power back to these facilities first.

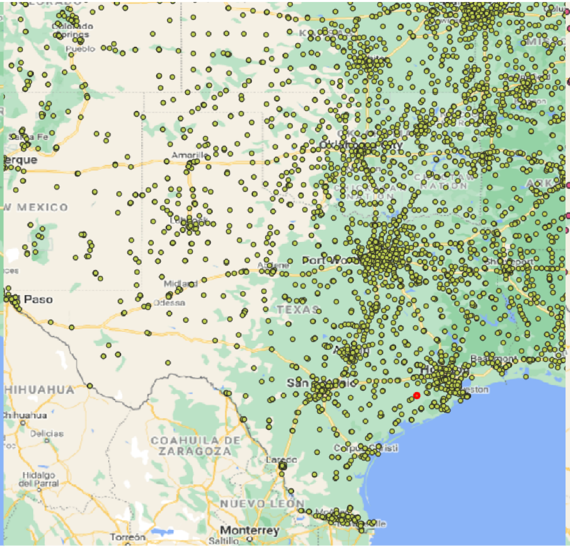
- Definition of important buildings
  - hospitals
  - police, fire and emergency facilities
  - water and sanitary authorities
  - nursing homes
  - assisted living facilities
- Model implementation
  - Pre-identified location of important buildings
  - Rank these priority buildings given their importance (depending on data availability)
  - Identify damaged components in simulation and rank them given building importance
  - Repair the damaged components and restore power for these buildings

Step 2: Emergency Responders Recovery

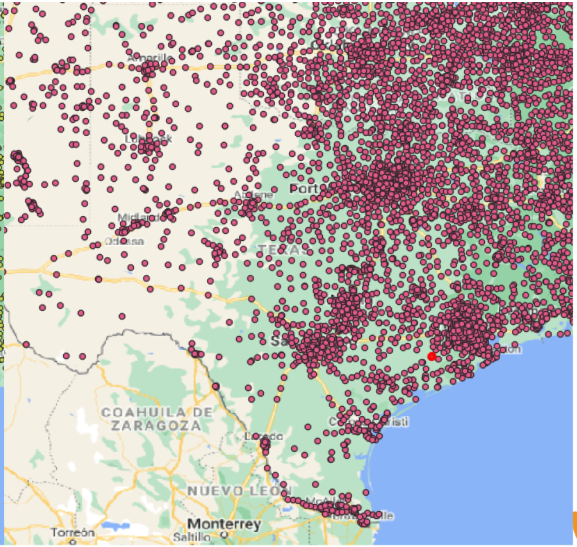
Location of hospital and nursing homes



Police Stations

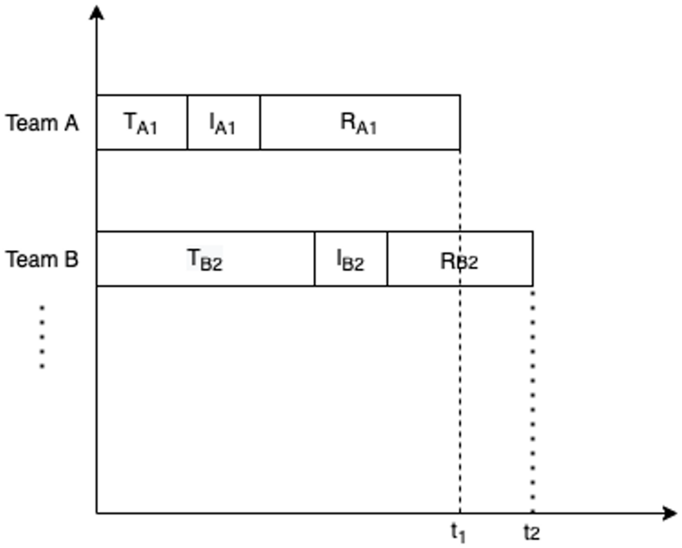


Fire Stations



Power Restoration Simulation for distribution network

After substations are repaired to work, we assume each substation's service area is independent and the number of crew teams are assigned given damage situation. We formulate this process as a scheduling problem, where each crew team will be assigned with the next highest priority job within a certain distance of where they are. Number of crew is determined by damage, crew availability, repair equipments/vehicles, general damage to other dependent infrastructures.



We have each team labeled with letters, A, B, …
We have each item needs to be repaired labeled with numbers, 1, 2, ...
T is the travel time from the original location of team A (substation, or utility site) to the damaged site. T is a function of

a) Distance of the two locations, *d*
b) Road damage situation, routing and speed limit, *r*

I is the inspection time to assess the damage.
R is the repair time. R is a function of

a) Distance to inventory site, *d*
b) Road damage situation, routing and speed limit, *r*
c) Inventory in stock or not, *i*
d) Type of repair work, *t*

Model Output

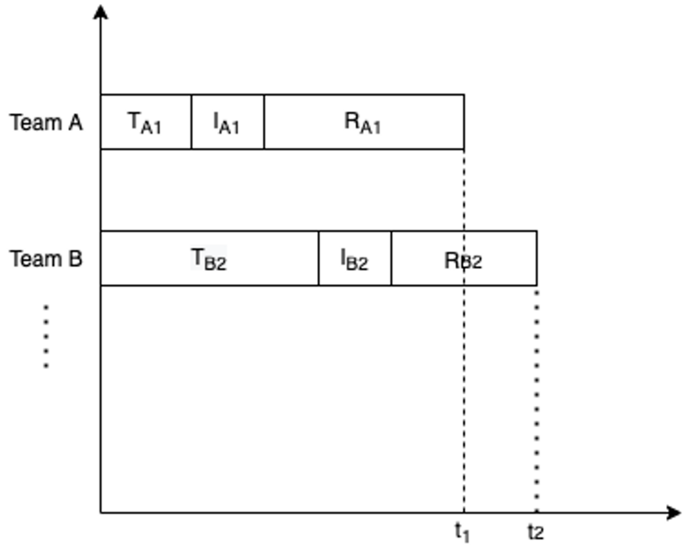
Output of Monte Carlo simulation is the power downtime for each building. We run a large number of trials to achieve convergence for the estimated random variable. Then we create the recovery curve based on the power downtime samples for each building. The curve is defined as the probability of having power recovered given time - P(power on|T<t). This is also called the empirical distribution function. As an example (the number does not reflect any actual simulation, a lot more samples are required to get convergence).

Convergence is determined by coefficient of variation given it is a non-parametric distribution.

$$c_V = \frac{\sigma}{\mu} \, (1)$$

$$n_{\min} = argmax_n |c_V^n - c_V^m| < E, \forall m > n \, (2)$$

Power Restoration Simulation for distribution network



After each component is repaired, e.g., at t1, we will reevaluate the connectivity of the network, and determine which nodes are now having access to the substation. In another word, we know at t1, who will now have access to power (we call these newly powered building {t1}). The power recovery time for these buildings {t1} will be tsub + t1. After the repair is done the team will be assigned with the next job that has the highest priority within a certain region of the location the crew team is at.

The time when a building get power back is defined as DTi for building i.

$$DTi = Tsub + Tcomp$$

where Tsub is the substation that supplies power to the building. Tcomp is the repair time of all damaged components that are used to connect the building to the distribution grid.

Step 3:
Large and Small Service Area Recovery based on network importance

- Job Ranking Algorithms to provide sub-optimal recovery process
  - Given the nature of the synthetic distribution network, assign an importance for each node in the network - higher importance means the node is more important in bringing power back for more customers.
- This step will guarantee larger service area to be recovered first and individual homes/rural area to be recovered later.
- Each repair team will be repairing a component in the damaged network until all nodes are getting back to be connected to the distribution substation.
- In the future model, we can introduce an classification of the line type given the loads to differentiate primary vs secondary distribution lines. And for each type of power lines we can have different parameter in fragilities and recovery.

Job Ranking Algorithms

These network importance can be used to rank the nodes in the distribution network, and each edge's importance is defined by its two nodes. The repairing process will follow this priority list to first recover power for large service region followed by small service region.

Algorithm 1 for pure radial network (high efficiency)

1. Find nodes directly connect to the substation, {N0}. Remove the substation from the network.
2. For each node in {N0}, find nodes directly connect to it as {N1}. Then calculate the number of nodes connect to each node in N0. After this, remove N0 from the graph.
3. Continue the process, until all the nodes are processed and removed from the graph.
4. As a result, we know the number of nodes each node is connected to in their downstream. The more nodes in the downstream, the more important that node is.
5. Then add importance for each power line, the importance is determined by summing up the importance of the nodes of the power line.

Job Ranking Algorithms

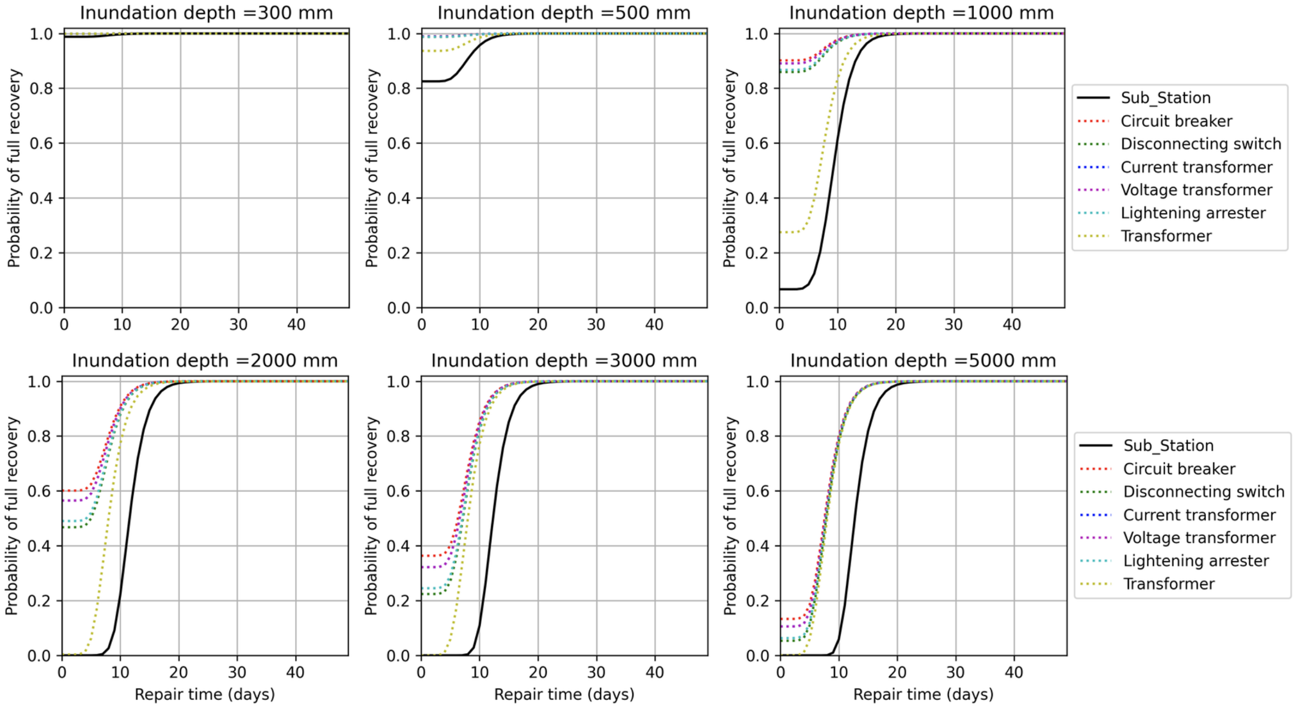
Algorithm 2 for any network 1. For every node in the graph, remove the node 2. Calculate the size of the number of nodes lost connection to the substation due to the removal. 3. Continue until all nodes has been calculated. The importance of each node will be the size of the number of nodes lost connection to the substation 4. Then add importance for routes → lines → nodes connected to important building/facilities, so they will always get repaired first.
Algorithm 3 for any network 1. Given the network, calculate degree centrality, closeness centrality, and betweenness centrality for each node 2. The importance of each node is calculate as the average of these three network metrics (between 0 - 1). 3. Explanation of these metrics are here: https://www.hindawi.com/journals/mpe/2019/9728742/

Recovery function development for power components

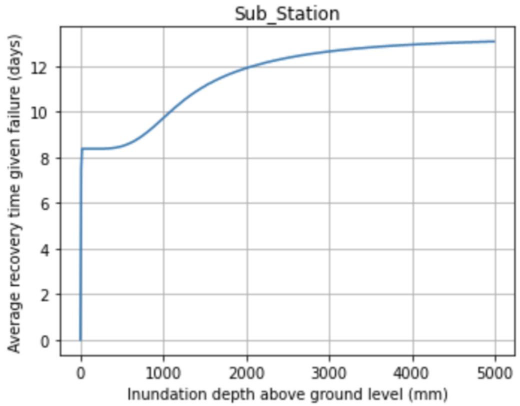
- Developed by One Concern for US and Japan based on historical events and literature.
- Recovery function for vulnerable power components is developed to describe the most accurate estimation of recovery time under the current framework.
- Hazard types
  - Wind
  - Flood
  - Earthquake
- Key components
  - Distribution substation and its components
  - Overhead power lines
  - Underground power lines

Power Component Recovery Parameters - Flood, substation

Recovery function for each component



Average recovery time given depth



Power Component Recovery Parameters - Earthquake

Table 2. Distribution line recovery per data from 2014 Napa earthquake Eidinger (2017)

Repair item	Total Manhours	Number of Repair items	Average Manhours per repair item
Conductor	1,147.0	68.0	16.9
Connector	42.0	4.0	10.5
Cross arm	247.0	12.0	20.6
Cutout	41.0	3.0	13.7
Enclosure, Lid, Frame	24.0	1.0	24.0
Guy wire hardware	45.0	6.0	7.5
Hardware / framing	34.0	3.0	11.3
Insulator	42.0	3.0	14.0
Jumper	81.5	8.0	10.2
Switch / Junction Box	21.0	1.0	21.0
Tie Wire	25.0	2.0	12.5
Transformer, Regulator Booster (OH)	630.0	8.0	78.8
Transformer Pad mount (UG)	28.0	2.0	14.0
Transformer Subsurface (UG)	71.0	2.0	35.5
Logistics	2,000.0	4.0	500.0
Grand Total	4,478.5	127.0	35.3

Repair item	Total Manhours	Number of Repair items	Average Manhours per repair item
Conductor	1,147.0	68.0	16.9
Connector	42.0	4.0	10.5
Cross arm	247.0	12.0	20.6
Cutout	41.0	3.0	13.7
Enclosure, Lid, Frame	24.0	1.0	24.0
Guy wire hardware	45.0	6.0	7.5
Hardware / framing	34.0	3.0	11.3
Insulator	42.0	3.0	14.0
Jumper	81.5	8.0	10.2
Switch / Junction Box	21.0	1.0	21.0
Tie Wire	25.0	2.0	12.5
Transformer, Regulator Booster (OH)	630.0	8.0	78.8
Grand Total	2,379.5	119.0	20.0

Power - Distribution Network Generation

Model Selection

Results of comparing network topology measures and network distance between the real and synthetic distribution systems.

	Average Nodal Degree	Average Betweenness Centrality	Average Number of Circle in Overhead	Total Length of Lines (Meters)	Mean Absolute Difference (m)	Pearson correlation
Actual System	2.066	0.0429	2.44	1.04×10^8	–	–
Steiner Tree Heuristic	1.999	0.0400	0	1.08×10^8	1426	0.699
Steiner Tree + Kmean	1.999	0.0389	0	1.09×10^8	882	0.745
Shortest Path	1.999	0.0363	0	1.07×10^8	693	0.842

Infrastructure Detection

Substation - Customer

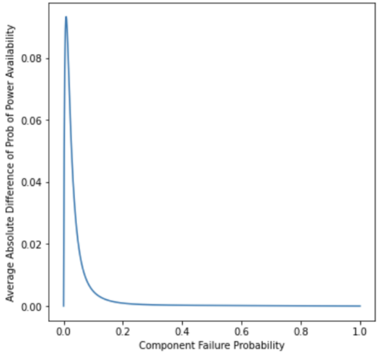
Substation - Circuits

Infrastructure Generation

Network Generation & Optimization

Power - Distribution Network Generation
Network Characteristics Validation

- Betweenness Centrality
- Network Distance Metrics
- Pearson correlation Coefficient
- Simulation Performance

Betweenness Centrality	0.011	0.019
Average Distances	3614 m	2800 m
Average Error Distances	867m	
Average Absolute Error Distances	880m	
Average Relative Error Distances	20%	
Pearson Correlation Coefficient	0.94	
Simulation Performance		

https://drpep.sce.com/drpep/

Power - Distribution Network Generation
Nationwide Validation

	Nationwide	Domino
Power lines	~5.5M Miles*	~4.51M Miles
Power poles	180M**	170,595,385

* https://www.scientificamerican.com/article/what-is-the-smart-grid/
** https://www.electrocuted.com/2016/09/08/utility-pole-facts/