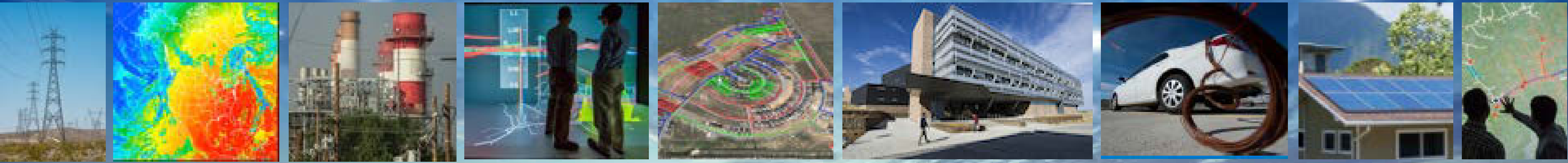
NREL
Transforming ENERGY



Realistic Synthetic Distribution Grids: Summary of Validation Results

Venkat Krishnan, M.S., Ph.D.
Senior Engineer, Power Systems Engineering Center
National Renewable Energy Laboratory, Golden ,CO, USA

Power Systems Engineering Research Center (PSERC) Industry Advisory Board (IAB) Systems Stem Meeting, Dominion Energy, Richmond, VA
December 4, 2019

Validation Team

NREL: Venkat Krishnan, Bugbee Bruce, Tarek Elgindy, and Bryan Palmintier
IIT Comillas (Spain): Carlos Mateo, Fernando Postigo, and Tomás Gómez San Roman
MIT: Pablo Duenas
CyME: Jean-Sebastien Lacroix

SMART-DS*: Synthetic Models for Advanced and Realistic Testing-Distribution Systems and Scenarios

Large-scale, open-source distribution systems in multiple modeling formats

Data set 1: SAF

SAF:
Santa Fe (NM)
84,000 consumers

Urban and suburban regions

Data set 2: GSO

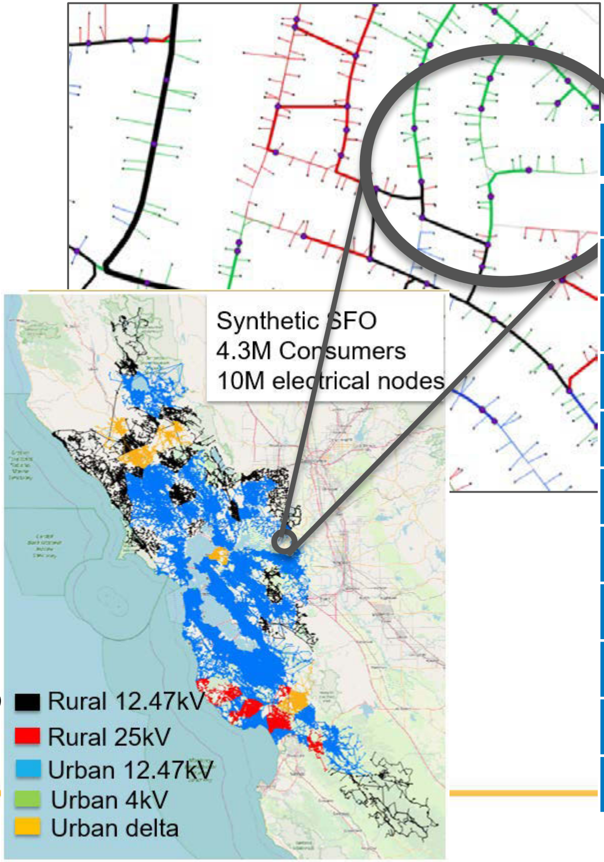
GSO:
Greensboro (NC)
134,000 consumers

Industrial, semi-urban, and rural customers

Data set 3: SFO

San Francisco Bay area (SFO)
4.3 million consumers

Diverse mix of rural, urban, semi-urban, and industrial customers across several counties



Number of buildings	2,265,594
Medium voltage	1,535
Low voltage	2,264,014
Number of consumers	**4,299,805**
Medium voltage	11,503
Low voltage	4,288,297
Number of electrical nodes	**9,868,205**
Number of transmission substations	148
Number of sub-transmission substations	632
Number of distribution transformers	559,151
Length of power lines (km)	116,837
Number of primary feeders	**2,236**

CHANGING WHAT'S POSSIBLE

**An ARPA-E funded project*

Team

National Renewable Energy Laboratory (NREL): Bryan Palmintier (PI), Tarek Elgindy, Venkat Krishnan, Elaine Hale, Bugbee Bruce, Michael Rossol, and Bri-Mathias Hodge

Massachusetts Institute of Technology (MIT): Pablo Dueñas

IIT-Comillas: Carlos Mateo Domingo, Fernando Postigo, Fernando de Cuadra, and Tomás Gómez San Roman.

CyME (Eaton): Jean-Sebastien Lacroix

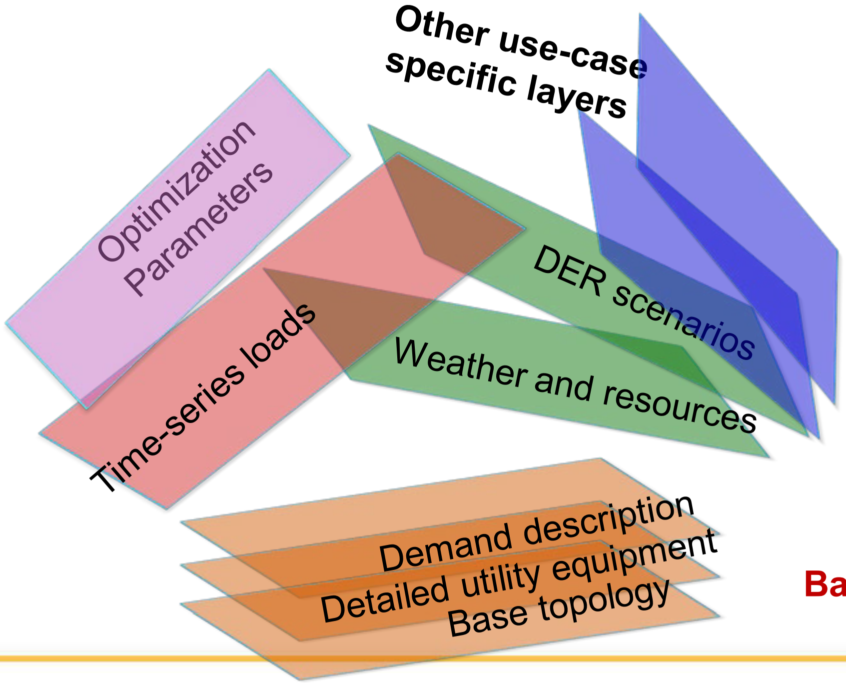
Electrical Distribution Design (EDD): Jason Bank and Robert Broadwater

Data partners: Duke Energy; Arizona Public Service (APS); City of Loveland, Colorado; Southern California Edison (SCE); Exelon Companies (Potomac Electric Power Company [Pepco], Delmarva Power, and Atlantic City Electric)

Technical review committee (additional members): Opus One, Electric Power Research Institute (EPRI), Advanced Microgrid Solutions (AMS), and Tesla

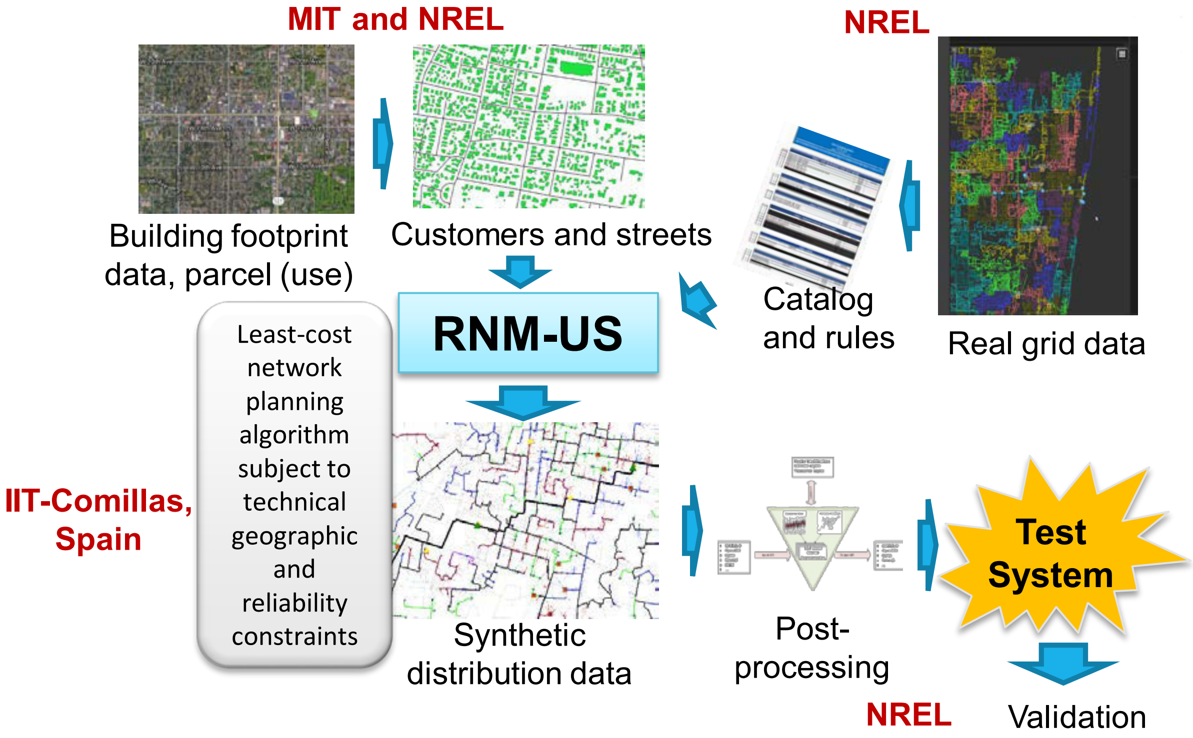
Distribution Data Sets: Development Factory

Full-scale, high-quality synthetic distribution system data set(s) that are realistic (but not real) and useful for testing advanced distribution system algorithms. A **layered** approach:



Base topology creation and validation

Base Topology Creation Process: Summary



https://www.nrel.gov/docs/fy17osti/68764.pdf
https://www.iit.comillas.edu/publicacion/mostrar_publicacion_working_paper.php.es?id=353
https://www.iit.comillas.edu/publicacion/mostrar_publicacion_working_paper.php.es?id=352

Synthetic Data Set: Validation Results Summary

Validation Team

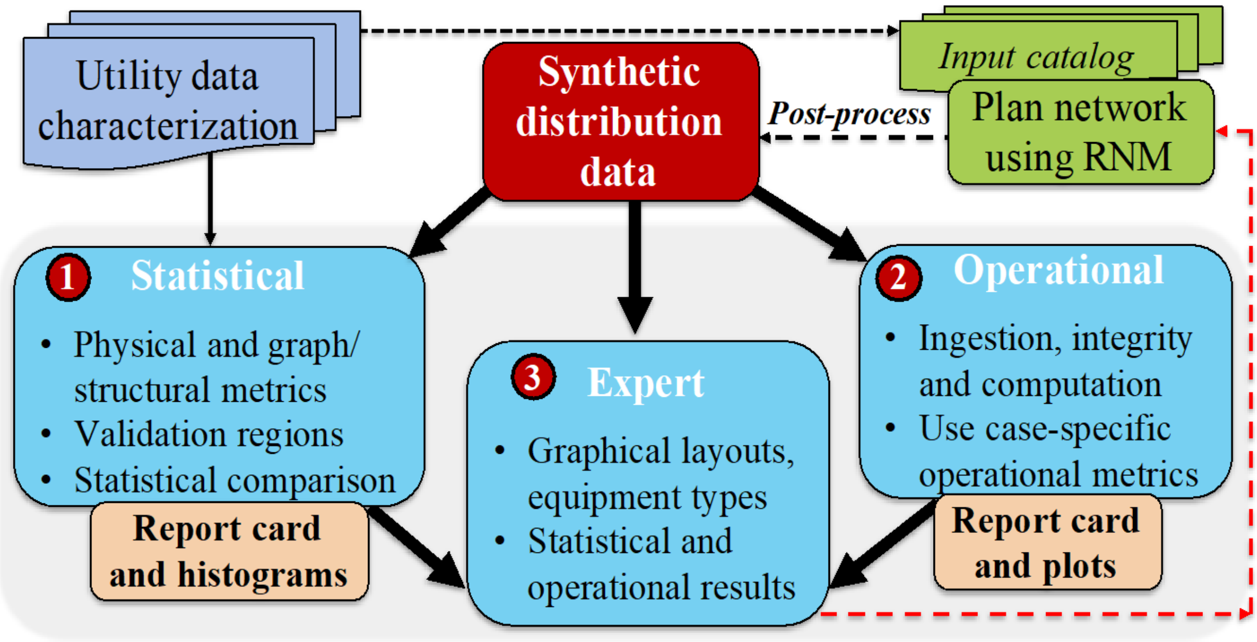
NREL: Venkat Krishnan, Bugbee Bruce, Tarek Elgindy, and Bryan Palmintier
IIT Comillas: Carlos Mateo, Fernando Postigo, and Tomás Gómez San Roman
MIT: Pablo Duenas
CyME: Jean-Sebastien Lacroix

The authors thank the U.S. DOE Advanced Research Projects Agency-Energy (ARPA-E) for their funding support. We are also grateful for the invaluable support from industry experts, vendors and data partners: Duke Energy (Leslie Ponder and Lance Fox); Arizona Public Service (Daniel Haughton); City of Loveland (Brieana Reed-Harmel), Colorado; Southern California Edison (Araya Gebeyehu); Exelon Companies (Steve Steffel) [Potomac Electric Power Company (Pepco) Delmarva Power, and Atlantic City Electric]; and Electrical Distribution Design (Jason Bank). Without their support, the large-scale synthetic data set creation and validation works would not have been possible. Additional thanks to our colleagues for valuable inputs: Mike Coddington and Barry Mather (NREL), Claudio Vergara (Off Grid Electric), Roger Dugan (EPRI), Abilash Thakallapelli (General Electric), and Nicolas Gensollen (LiP6, Sorbonne University).

Validation: Criteria and Parameters

Criteria	Why Important?	Validation Metrics
Realistic physical layout	Stakeholder acceptance, resource/ weather/demographics/communication.	Coordinate data inclusion, Geographic Information System files, visual checks
Realistic system size	Scalability beyond one feeder, reconfiguration options, multi-feeder interactions	Counts: customers, feeders, transformers, per feeder
Realistic topology and components	Critical foundation for all use cases, realistic power flows	Low voltage, medium voltage, high voltage: lengths, counts, customer ratios, impedance
Representative voltage profiles	Key concern for distribution operations, time series, volt/volt ampere reactive (VAR), distributed energy resource (DER) impacts, losses	Voltage distributions, load tap changing (LTC), count of capacitor banks and regulators, set points
Realistic reconfiguration options	Automated reconfiguration (e.g., FLISR) post-reconfiguration operations simulation	Count of switches, reclosers, breakers, fuses, etc.
Comprehensive load specification	Support basic power flow, enable rich scenario layers	ZIP parameters, load types, customer fractions
Computational complexity	Typical power flow solution times PLUS challenging scenarios	Solution times, convergence, violations

Validation Process: Three-Pronged Validation



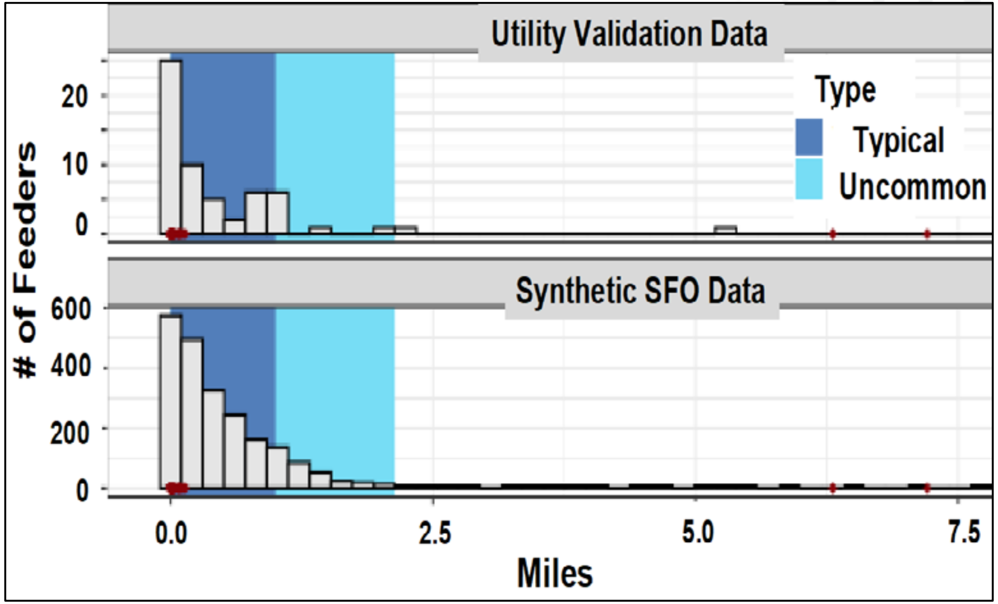
Validation questions:

- Are the synthetic data sets realistic?
- Can the statistical summary of metrics (e.g., line lengths, transformer sizes) come from realistic systems' metric distributions?
- Do they behave operationally (voltage, power flow) as real systems?
- Can they pass expert scrutiny (low-voltage secondary data, phase connections, equipment ratings, and standardization)?

Statistical Validation Process: Step 1—Validation Region

- Regions are defined based on proportion of values of validation data for a particular metric:
  - **Typical**: at least 80% of the observed validation data
  - **Uncommon**: at least 95% of the observed validation data
  - **Rare**: the remaining 5% of the data that occur in the tails.
- Method for region calculation based on data type and availability:
  - **Highest probability density:** finds the collection of possibly disjointed subsets that cover the desired proportion of data
  - **Quantiles:** Defines continuous regions based on empirical quantiles
  - **Group membership:** categorical data are handled by ranking the prevalence of each group and finding appropriate coverage.
- Example: low-voltage three-phase line length:
  - **Typical**: [0+, 1] mile
  - **Uncommon:** [1, 2.14] miles
  - **Rare:** anything outside these regions.

Distribution from aggregated utility data (top)

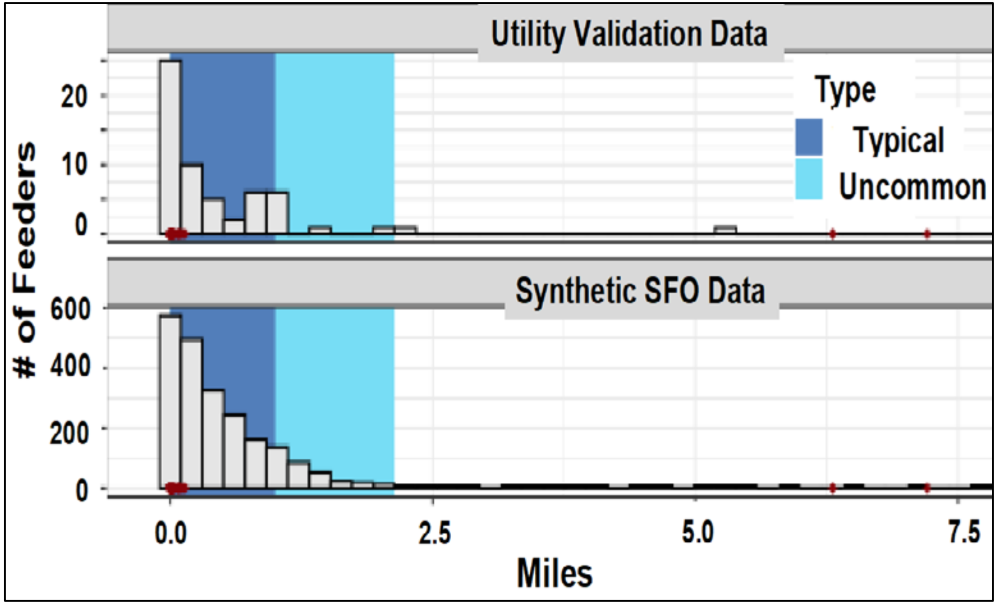


Total low-voltage three-phase line length per feeder: utility (58 feeders) vs. SFO data set

Note: The red markers along x-axis are from open-access test feeders (IEEE, PNNL taxonomy, EPRI, etc. listed on Slide 14)

Statistical Validation Process: Step 2—Comparison and Validation Grade

- Each synthetic feeder is classified as typical, uncommon, or rare based on which region it is in. **The SFO data set is compared in this example.**
- Example: low-voltage three-phase line length:
  - Calculate aggregate totals for each category.
  - **Typical:** 85%
  - **Uncommon**: 12.4%
  - **Rare:** 2.6%.
- Validation grade is based on if we have an appropriate number of "rare" synthetic feeders.
  - **Good:** 0% to 5% (grade for this metric)
  - **OK** (marginally good)**:** 5% to 10%
  - **Check**: >10%
  - 5% threshold is chosen to reflect 95% coverage of typical + uncommon regions. This is a way of asking, "Do the tails of these distributions behave the same?"



Total low-voltage three-phase line length per feeder: utility (58 feeders) vs. SFO data set

Rationality Behind Statistical Validation Approach

- **Ideal goal of statistical validation:** If the metric distributions from utility data are for the same region as the synthetic data set, an exact comparison of the two probability distribution functions may be performed.
- **Issue:** But there are issues in publishing a wide range of statistics from a specific utility (concerns due to critical infrastructure security and respecting the nondisclosure agreement) and at times a lack of a diverse set of metrics from a single utility (e.g., not every utility data set had low-voltage secondary data).
- **Our scenario:** To circumvent this issue and yet disseminate the anonymized real data for research reproducibility, the team proposed using ***aggregated metrics from several regional utilities*** to **"compare the aggregated metric distribution with the synthetic data set metric for a specific region."**
- **Our statistical validation goal:** Therefore, the goal of the statistical validation is not to exactly match real data and synthetic data distributions **but to make sure the synthetic data capture the variety in metrics found in reality and do not have values outside of the aggregated utility distribution (indicating unreal values).** In other words, the goal of the statistical validation process is to assess the validation exit criteria: **"Is it feasible that the metrics from the synthetic data sets could arise from a real system?"** To answer this, we compare the densities of metrics from synthetic feeders against validation regions estimated using aggregated utility data, and we assign grades of **good/OK/check** based on the **rare** region density (Slide 10).

Validation Metrics and Regions

Validation regions obtained from aggregated utility data for each metric.

Utility data:

- Duke Energy
- APS
- City of Loveland
- SCE
- Exelon
- Pepco
- DPL
- ACE.

Validation Metric	Utility Data Validation Regions		
	Typical	Uncommon	#Feeder
Dist. Xfrmr Tot. (MVA)	[0+, 1.73], [4.94+, 31]	[1.73+,4.94], [31+,38.629]	5923
Tot. real load (kW)	[4181+, 13793]	[577+,4181], [13793+, 17590]	1330
LV 1ϕ line len. (miles)	[0+, 34.75]	[34.75+, 44.31]	57
LV 3ϕ line len. (miles)	[0+, 1]	[1+, 2.135]	58
MV 1&2 ϕ line ln (mile)	[0+, 35.36]	[35.36+, 124.62]	10632
MV 3ϕ line len. (miles)	[0+, 20.84]	[20.84+, 45.6]	10149
MV OH 1&2 ϕ line ln (mi)	[0+, 19.1]	[19.1+, 84.5]	10099
MV OH 3ϕ line ln (mile)	[0+, 17.7]	[17.7+, 39.7]	9747
% of OH 1&2 ϕ lines	[0+, 0.23], [0.46+, 1]	[0.23+, 0.46]	9350
% of OH 3ϕ lines	[0.4+, 1]	[0.18+, 0.4]	9492
# Cust.	[94+, 2607]	[8+, 11837]	9734
Ratio of MV 1&2 ϕ line len. to Cust.	[0+, 0.12]	[0.12+, 0.24]	9221
Ratio of MV 3ϕ line len. to Cust.	[0+, 0.09]	[0.09+, 0.77]	8556
# Fuses	[4+, 187]	[187+, 281]	6013
# Reclosers	[0+, 5]	[5+, 9]	6013
# Regulators	[0+, 3]	[3+, 8]	11574
Sectionalizer	[0+, 1]	[1+, 3]	5020
# Switches	[3+, 392]	[392+, 635]	5020
# Cap. Banks	[0+, 5]	[5+, 7]	11574
Avg. degree	[1.9+, 2.06]	[1.6+, 1.9], [2.06+, 2.1]	5020
Char. path len. (miles)	[12.4+, 95]	[2+, 12.4], [95+,134.39]	5020
Graph dia. (miles)	[32+, 260]	[4+, 32], [260+, 371]	5020

arpa·e CHANGING WHAT'S POSSIBLE

Report Card

Nearly all metrics pass the test, meaning we can infer that statistically ***the data in the synthetic network could have originated from real-world utility networks****.*

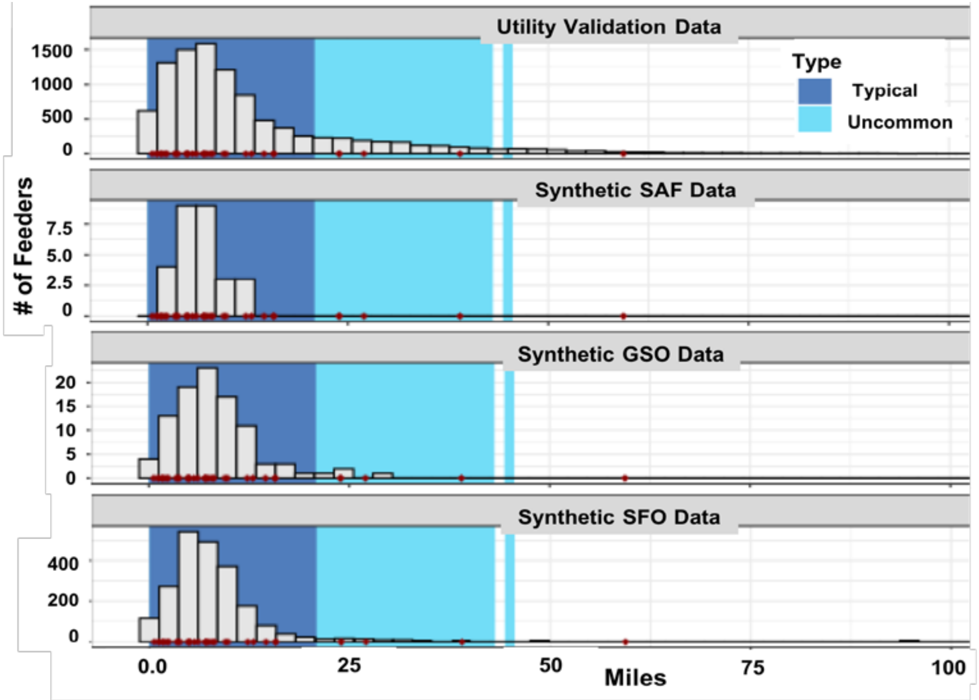
Note: The goal is to attain “good” grades for as many metrics as possible; however, not every metric needs to get “good” grades. Instead, “OK” or “check” are indications to look further (more data set creation iterations) or understand better (some regional peculiarities might be needed). For instance, when the team compared metrics from one actual utility to those from all utilities, we found that although most metrics were “good,” one to two metrics received a “check” grade, indicating different levels of diversity among utilities.

Validation Metric	SAF Dataset Results				GSO Dataset Results				SFO Dataset Results			
	Typ	UnC	Rar	Grd	Typ	UnC	Rar	Grd	Typ	UnC	Rar	Grd
Dist. Xfmr Tot. (MVA)					0.74	0.25	0.01	G	0.79	0.19	0.01	G
Tot. real load (kW)	0.07	0.90	0.03	G	0.70	0.29	0.01	G	0.55	0.36	0.09	OK
Tot. reactive load (kVAR)	0.97	0.03	0.00	G	0.05	0.91	0.05	G	0.03	0.74	0.23	Chk
Avg. degree	0.97	0.03	0.00	G	0.86	0.13	0.01	G	0.91	0.08	0.01	G
Char. path len. (miles)	0.97	0.03	0.00	G	0.75	0.23	0.02	G	0.83	0.16	0.01	G
Graph dia. (miles)	0.97	0.03	0.00	G	0.81	0.18	0.01	G	0.86	0.13	0.01	G
LV 1ϕ line len. (miles)	1.00	0.00	0.00	G	0.98	0.02	0.00	G	0.95	0.04	0.01	G
LV 3ϕ line len. (miles)	1.00	0.00	0.00	G	0.66	0.29	0.05	G	0.91	0.07	0.02	G
Ratio of MV 1&2 ϕ line len. to Cust.	0.98	0.02	0.00	G	0.94	0.06	0.00	G	0.89	0.10	0.01	G
Ratio of MV 3ϕ line len. to Cust.	0.17	0.67	0.16	Chk	0.77	0.20	0.03	G	0.71	0.27	0.03	G
MV 1&2 ϕ line ln.(mile)	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G
MV 3ϕ line len. (miles)	1.00	0.00	0.00	G	0.95	0.05	0.00	G	0.97	0.03	0.00	G
LV OH 1ϕ line ln (mile)	0.62	0.38	0.00	G	0.69	0.29	0.02	G	0.69	0.30	0.01	G
LV OH 3ϕ line ln (mile)	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G
MV OH 1&2 ϕ line ln (mi)	1.00	0.00	0.00	G	0.98	0.02	0.00	G	0.92	0.08	0.00	G
MV OH 3ϕ line ln (mile)	1.00	0.00	0.00	G	0.95	0.05	0.00	G	0.97	0.03	0.00	G
% of OH 1&2 ϕ lines	0.05	0.95	0.00	G	0.85	0.15	0.00	G	0.80	0.19	0.01	G
% of OH 3ϕ lines	0.00	1.00	0.00	G	0.85	0.11	0.04	G	0.89	0.10	0.01	G
# Cust.	0.98	0.02	0.00	G	0.72	0.25	0.04	G	0.80	0.17	0.03	G
# Breakers	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G
# Fuses	0.98	0.02	0.00	G	0.82	0.18	0.00	G	0.80	0.20	0.00	G
# Reclosers	1.00	0.00	0.00	G	0.93	0.07	0.00	G	0.92	0.07	0.01	G
# Regulators	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G
Sectionalizer	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G
# Switches	0.71	0.29	0.00	G	0.61	0.27	0.11	Chk	0.64	0.23	0.13	Chk
# Cap. Banks	1.00	0.00	0.00	G	1.00	0.00	0.00	G	1.00	0.00	0.00	G

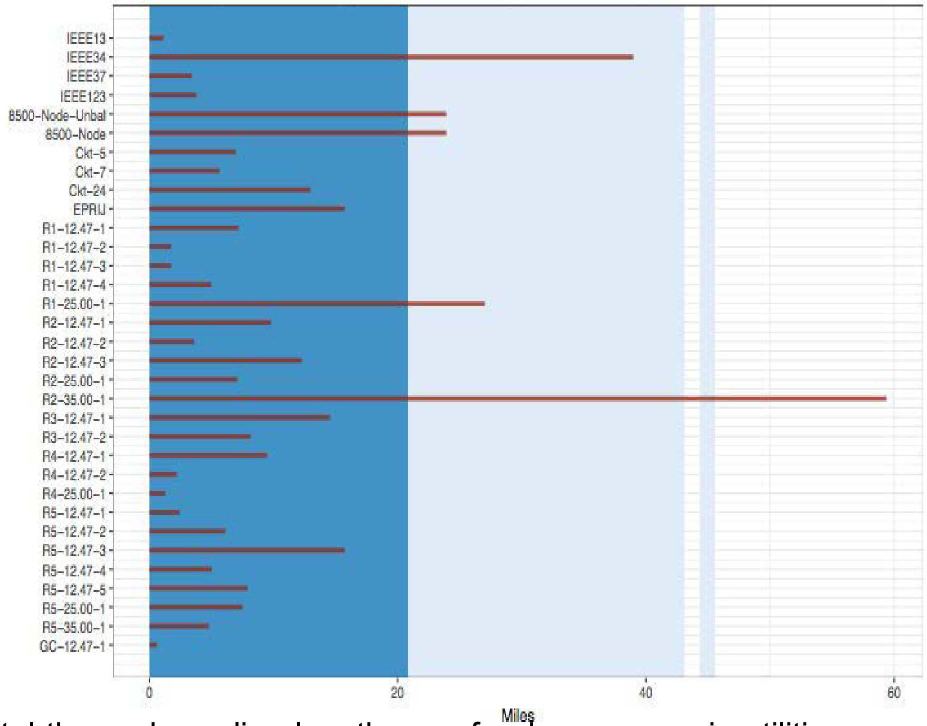
Typ *= Typical;* ***UnC*** *= Uncommon;* ***Rar*** *= Rare;* ***Grd*** *= Grade;* ***G*** *= Good* ***OK*** *= Marginally good;* ***Chk*** *= Check*

Ratio of Medium-Voltage Three-Phase Line Length*

Synthetic data sets



Existing open data sets

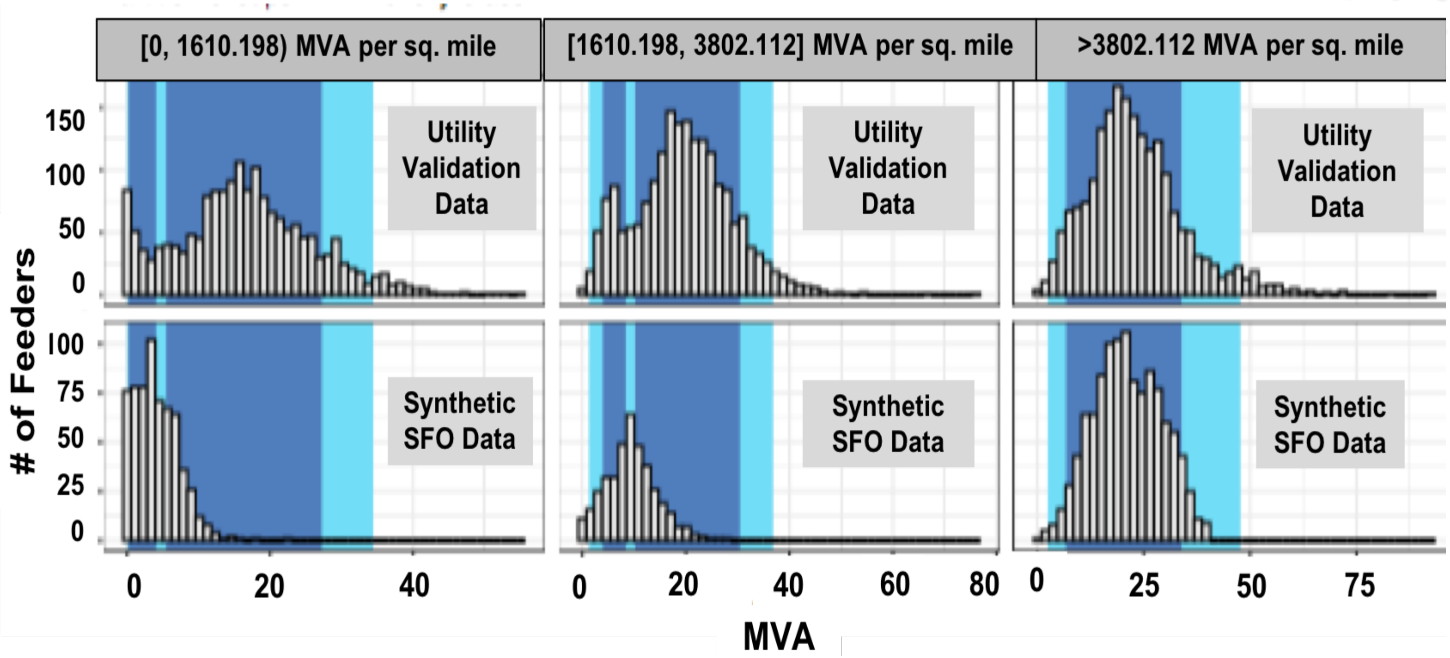


Larger data set (SFO) has more diversity, including increasing total three-phase line lengths per feeder, as seen in utilities.

arpa-e CHANGING WHAT'S POSSIBLE

*Note: Such histogram comparisons exist for all metrics shown on the previous slide, but for brevity are not shown here. A detailed publication will be released later.

SFO: Distribution Transformer Capacity (MVA), with Load Density Partitions



SFO's higher load density region (rightmost plot, urban region) has similar distribution of distribution transformer capacity as observed in utilities (aggregated data). The lower load density region (leftmost, rural) in SFO has fewer customers and smaller load sizes compared to the aggregated utilities; therefore, the distributions are not closely matching. Yet the distribution plot shows that the SFO data set is a subset of what is found in reality.

arpa·e CHANGING WHAT'S POSSIBLE

Operational Validation: Metrics Targets Met with OpenDSS and CyME

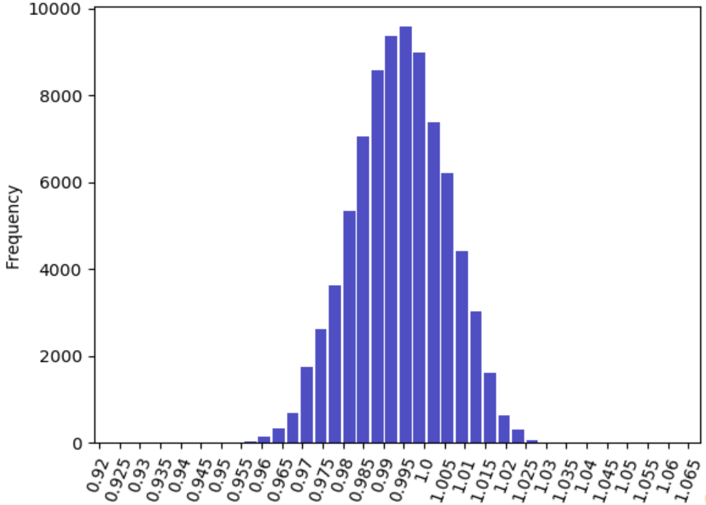
Analysis	Parameters (unit)	Exit target
Power flow convergence	Number of iterations	< 20
	Time to solve (s)	< 30
Power flow performance	Minimum Service Voltage (V)	> 114
	Maximum Service Voltage (V)	< 126
	Average Voltage (V)	114-126
	System Losses (%)	< 10
	Overloads Count	<0
	Under-Voltage Count	<0
	Over-Voltage Count	<0
	Voltage regulation range/bandwidth	Range
	LTCs within range/bandwidth	Range
	Transformer Loading (%)	0-100
	Voltage Unbalanced Factor (%)	<3
Short-Circuit (SC)	Transmission SC Level (kA) (>138 kVLL)	20-40
	Sub-Transmission SC Level (kA) (69 kVLL < V < 138 kVLL)	20-40
	Medium Voltage SC Level (kA) (1 kVLL < V < 69 kVLL)	0.3-40
	Low Voltage SC Level (kA) (<1 kVLL)	0.5-100
Reliability	System Average Interruption Duration Index (SAIDI) (hr/cust-yr)	0.5-4.5
	System Average Interruption Frequency Index (SAIFI) (event/cust-yr)	0.1-3.0
	Customer Average Interruption Duration Index (CAIDI) (hr/cust-event)	0.5-6.5

[[i]] ANSI C84.1-American National Standard for Electric Power Systems and Equipment – Voltage Ratings (60 Hz), 2011
[[ii]] R.C. Dugan, M. F. McGranaghan, and H. W. Beaty, Electrical Power Systems Quality, New York: McGraw-Hill, 1996
[[iii]] IEEE Industry Applications Society, *IEEE Recommended Practice for Calculating Short-Circuit Currents in Industrial and Commercial Power Systems*, IEEE Std. 551TM-2006
[[iv]] J. J. Burke, Power Distribution Engineering: Fundamentals and Applications, 1994
[[v]] R. E. Brown, Electric Power Distribution Reliability, Second Edition, 2009

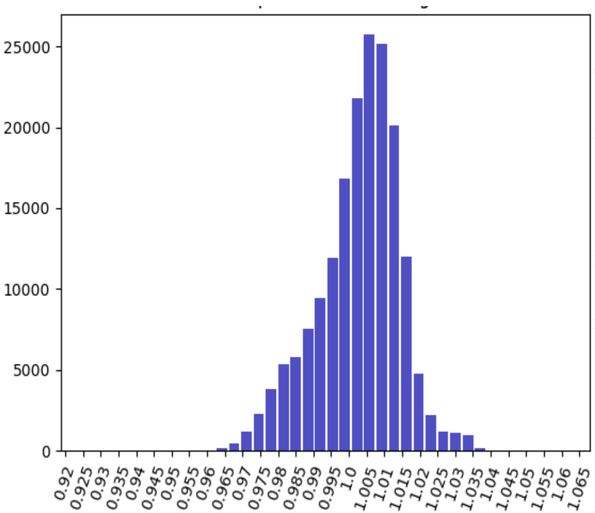
Operational Validation Results and Report Card

Network validation process: correctly ingested by OpenDSS and CyME software?

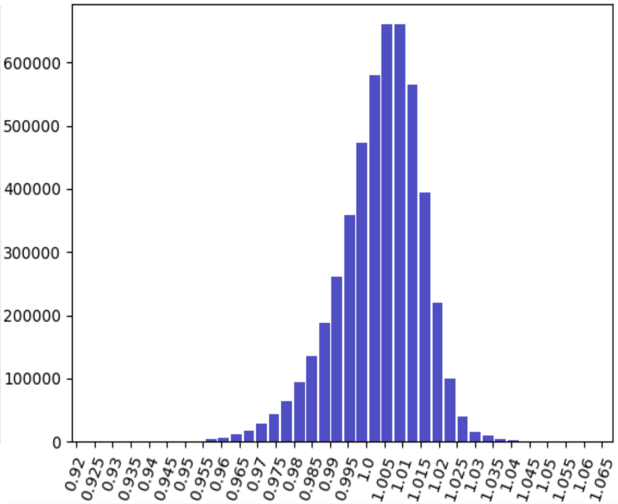
1) Network segmentation and assigning coordinates (for better visualization, section identification, and faster distributed simulation)
2) Line and transformer model corrections (impedances and susceptance values as well as configurations checked against typical line types)
3) Transformer types and low-voltage (120/240-V) configuration (U.S.-style, center-tapped)
4) Appropriate protection equipment (breaker and fuse) ratings to match line ratings.



SAF



GSO



SFO

All feeder voltages respect ANSI limits for average loading scenario

arpa·e CHANGING WHAT'S POSSIBLE

Expert Comments/Input (Selected)

- Fewer medium-voltage consumers per feeder (about 0–2 per feeder)
- Number of feeders per substation: ~ multiples of 4
- A tree-like low-voltage configuration
- Equipment standardization (less rating variety)
- Maximum ampacity of power lines: approximately 600A
- Capacitors instead of regulators in urban-suburban
- Install capacitors only in three-phase buses
- More switches needed for reconfiguration and reliability
- Fuses should be on laterals
- Underground lines should have lots of switches
- Some key per-feeder design parameters: medium-voltage/low-voltage line lengths, consumers per distribution transformers, distribution transformer capacity, capacitors/regulators, and breakdown of distribution transformer size by phase (e.g., fewer single-phase 75-kVA types)
- Qualify switches to be manual (elbow) and automatic
- Distribution transformers—center-tapped shell form
- Typical transformer impedance: approximately 2%–30%
- LTC settings updated to make sure voltage at medium-voltage high side can be set to 1.05, thereby avoiding too low a voltage that leads to consequent low-voltage limit violations in the low-voltage network
- Number of loops were large initially because of a large number of normally open (NO) switches being closed (in the medium-voltage network). These were corrected, and the only loops left were at the subtransmission level (ring configuration for reliability) and at the industrial networks (secondary meshed networks for reliability).
- Spots loads were diversified (ranging from 5 MW to 0.28 kW), whereas originally only few selected values were used repeatedly without diversity.
- Spacing for single- and three-phase lines were updated based on IEEE four-node spacing data
- Substation configuration updated to resemble realistic details, including placement of circuit breakers and reclosers closer to the substation.

Summary: "Validated" Synthetic Large-Scale "Realistic" Distribution System Data Sets

Basic data set (validation results): power flow, quasi-static time-series simulations

Future work: enable advanced use cases.

1. Distribution voltage impacts of DERs:
   - e.g., solar, electric vehicles, advanced inverters
   - Optimize control settings (including grid-edge devices).
2. Single-period Distribution-Optimal Power Flow (D-OPF):
   - Including DLMP or LMP+D calculation.
3. Distribution system reconfiguration:
   - Fault location, isolation, and service restoration (FLISR)
   - Topology changes for enhanced operations.
4. Volt/VAR optimization:
   - Adds dispatchability layer to DER voltage impacts.
5. Advanced DER time-series simulations:
   - Storage
   - Deferrable loads, thermal loads.
6. Multi-period D-OPF:
   - Distribution system operator markets, tariffs, transactive energy.

Thank you

www.nrel.gov

Contact: Venkat.Krishnan@nrel.gov, Bryan.Palmintier@nrel.gov

NREL/PR-5D00-75723

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by Advanced Research Projects Agency-Energy (ARPA E) under the GRID DATA program. A portion of this research was performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) and located at the National Renewable Energy Laboratory. The views expressed in the article do not necessarily represent the views of the U.S. Government. The publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

CHANGING WHAT'S POSSIBLE

NREL Transforming ENERGY