Final Scientific/Technical Report

Sustainable Data Evolution Technology (SDET) for Power Grid Optimization

Contract Number: 16/CJ000/09/02

Award:	DE-AR0000710
Lead Recipient:	Pacific Northwest National Laboratory
Project Title:	Sustainable Data Evolution Technology (SDET) for Power Grid Optimization
Program Director:	Dr. Kory Hedman
Principal Investigator:	Dr. Zhenyu Huang
Contract Administrator:	Sheena Kanyid
Date of Report:	December 17th, 2018
Reporting Period:	[September 1st, 2016 – December 17th, 2018]

Public Executive Summary

This project develops an innovative sustainable data evolution technology (SDET) to create open-access power grid datasets and facilitate updates to these datasets by the power grid community. The lack of open-access, realistic large-scale datasets significantly limits the ability of researchers, developers and ultimately end users to develop, benchmark and compare new methods and tools for optimizing the operation and planning of the grid, which leads to slow adoption by end users. Available open datasets include the IEEE sponsored 14-300 bus transmission test cases[1] and distribution test feeders[2], the Edinburgh Power Systems Test Case Archive (39-2224 buses)[3], and the Polish cases (2383-3375 buses) available with MATPOWER[4]. These datasets serve certain purposes, but for power grid optimization, they are insufficient in one or more of these aspects: 1) they are too small; 2) they are incomplete in both models and scenarios for optimization purposes; and 3) they are static and are not keeping up with grid needs.

The SDET technology delivers large-scale realistic datasets and data-creation tools capable of generating new datasets. We work closely with the Category-2 data repository project awarded by the same GRID DATA program. The generated datasets and data creation tools are compatible with and available through the data repository. The objective is to make this a sustained effort within and beyond the ARPA-E GRID DATA program so that the datasets can evolve over time and meet the current and future needs for power grid optimization and potentially other applications in power grid operation and planning.

The novel and disruptive features of SDET differentiate itself from prior work in the following aspects:

1) SDET derives features and metrics for both transmission and distribution (T&D) systems by analyzing many public and private datasets provided by our industry partners National Rural Electric Cooperative Association (NRECA) and PJM Interconnection (PJM).
2) SDET develops data-creation tools and use these tools to generate large-scale open-access realistic datasets that comply with the metrics for both T&D systems.
3) SDET validates the created datasets using industry tools provided by our vendor partner GE Energy Solution. The data creation tools, as well as the created datasets, are integrated into the data repository (DR. POWER) funded by ARPA-E GRID DATA program in Category 2. Users can use the datasets and access these data creation tools to create and update datasets, thus datasets can *evolve* per user requirements and power grid development. The SDET is a novel and disruptive technology compared to current ad hoc and static dataset generation.

It is expected the SDET technology will fundamentally enhance the development, assessment, and adoption of new methods and tools for power grid optimization and other applications, which in turn

---

[1] University of Washington, “Power Systems Test Case Archive”, available at http://www.ee.washington.edu/research/pstca/.

[2] IEEE PES Distribution System Analysis Subcommittee’s Distribution Test Feeder Working Group, “Distribution Test Feeders”, available at http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html.

[3] University of Edinburgh, “Power Systems Test Case Archive”, available at http://www.maths.ed.ac.uk/optenergy/NetworkData/.

[4] Cornell University, “MATPOWER files”, available at http://www.pserc.cornell.edu/matpower/docs/ref/matpower5.0/menu5.0.html.

would significantly improve the reliability, resiliency, and efficiency of the power grid – a major ARPA-E mission objective.

Acknowledgements

The project team would like to thank Dr. Kory Hedman, Dr. David Guarrera, Dr. Tim Heidel, and Dr. Ashley Arigoni for their constructive comments and support as the ARPA-E program management, under this project, in developing the innovative software prototype for generating synthetic and realistic large-scale power grid data sets.

Table of Contents

Figures/Tables

Accomplishments and Objectives

The key technical milestones and deliverables proposed in this project with the actual performance is summarized in Table 1 below:

Table 1. List of Key Milestones/Deliverables and Actual Performance

Project Element	**Detailed Description and Targets**	**Actual Performance**
Project Element 1: Metrics of real-world datasets		
Task 1.1 Collect real-world datasets	**Collect industry datasets to be used for extracting key metrics**	
M1.1.1: Datasets requirements Document	Report delivered to ARPA-E and approved by the Program Director listing the team's required data attributes (3-5 transmission, 20 distribution feeders). The data attributes will include topology, network parameters, generation parameters, economical parameters and any other characteristics needed for successful project completion. Transmission datasets need to support OPF problems, and distribution datasets need to support VVO. This document will also identify the specific partners from which data will be obtained.	(12/31/2016) The team completed and delivered the datasets requirements report that includes all the required data attributes details to ARPA-E program director, which was approved.
M 1.1.2 Collection of industry datasets from project partners and other available sources	Receipt of datasets from industry partners and verification and certification that data satisfies all the requirements of M1.1.1. The data attributes and a summary of collected datasets (not the actual datasets) will be reported in the quarterly report.	(3/31/2017) The team received the dataset from industry partners and performed verification and certification of the dataset, which indicated the data satisfied all the requirements of M1.1.1. The data attributes and a summary of collected datasets were reported in the quarterly report.
Task 1.2 Develop initial metrics of real-world datasets	**Develop initial mathematical metrics and features by analyzing the collected real-world datasets**	
M 1.2.1 Initial dataset metrics	Identification of preliminary metrics and features (based on analysis and extraction of common metrics of real-world datasets) that can be used to measure the realism of a dataset. The metrics to be analyzed include at least the following (though not all metrics may prove useful for establishing the realism of datasets): graph clustering coefficients, statistical distributions of parameters such as line impedance, and solvability of targeted applications such power-flow, OPF and VVO. The ranges of these metrics corresponding to "realism" (if applicable) will be specified.	(3/31/2017) The team Identified the preliminary metrics and features (based on analysis and extraction of common metrics of real-world datasets) that could be used to measure the realism of a dataset.
Task 1.3 Finalize dataset metrics via workshop	**Host a workshop with industry experts to review and finalize dataset metrics**	
M 1.3.1 Final dataset metrics	Final sets of metrics for assessing dataset realism delivered to ARPA-E and approved by the Program Director in the form of Word document or slides. The final list of metrics will be informed by expert input gather during a team-organized workshop with industry stakeholders. This report will also summarize exactly what was learned from the workshop and industry experts and how it has informed the path forward.	(3/31/2017) The team completed the final sets of metrics for assessing dataset realism, a technical report detailing the final sets of metrics has been delivered to the ARPA-E program director, which was approved.
Project Element 2: Topology generation		
Task 2.1 Adapt graph algorithms for creating realistic power grid topologies	**Review and develop graph methods for generating realistic power transmission topology.**	

M 2.1.1 Delivery of graph methods for realistic power grid topology generation.	Selection and verification of at least one method (based on existing graph methods or the development of new hierarchical (multi-step stochastic graph models) capable of generating a network-of-networks topology meeting all metrics generated in Project Element 1. Graph method candidates that will be studied include variants of random geometric graph (RGG) models, preferential attachment (PA) models, and Block Two-level Erdos-Renyi (BTER) models. We will study at least five instances of generated topologies in order to meet the selection criteria based on graph theoretic metrics and their ranges (defined in Project Element 1). Findings will be reported in the quarterly report.	(6/30/2017) The methodology of generating the topology of the synthetic transmission power grid has been developed and verified, to satisfy all the topology metrics defined in M1.3.1. The findings was included in the quarterly report.
Task 2.2 Develop a prototype graph tool for topology generation	**Develop an open-source prototype tool based on the proposed algorithm(s) and evaluation of the performance using the developed metrics in Task 1**	
M2.2.1 Topology tool requirements and specification	Identification of user input and outputs for topology generation tool. Description of tool requirements (for example, the allowed ranges and variability of the output given user specifications).	(6/30/2017) The user input and outputs for topology generation tool has been identified and the tool requirements were described in the quarterly report.
M 2.2.2 Completion of a prototype tool for topology generation	Open-source prototype topology generation tool developed with associated documentation. The tool will have input/output clearly defined. The inputs will include the number of buses, the number of voltage levels, degrees of buses, and other graph metrics. The output will be a data file that describes the generated topology (connectivity) in recognized formats such as CIM. A summary report in the form of Word document or slides will be prepared for ARPA-E Program Director. Verification that tool meets specifications in M2.2.1.	(9/30/2017) The team completed the development of the open-source prototype topology generation tool and its associated documentation. The tool had input/output clearly defined. The inputs includes the number of buses, the number of voltage levels, degrees of buses, and other graph metrics defined in M1.3.1. The output is a data file with PTI RAW format. A summary report in the form of Word document has been delivered to ARPA-E Program Director.
Project Element 3: Parameter population		
Task 3.1 Develop deterministic and probabilistic algorithms for parameter population	**Develop both deterministic and probabilistic methods for populating parameters**	
M 3.1.2 Completion of algorithms for populating power grid parameters	Development and verification of algorithms for populating realistic parameters meeting the required metrics derived in Program Element 1. Two methods are expected to be used for this purpose: a deterministic method based on a fragmentation of real systems, and a probabilistic method to populate parameters such as production cost/market bid information, information on variable generation, transmission line parameters connecting the fragments. Four small datasets and two large datasets will be developed. Findings will be reported in the quarterly report.	(6/30/2017) The team completed the development and verification of algorithms for populating realistic parameters meeting the required metrics derived in Program Element 1. A deterministic method based on the fragments of real systems, and a probabilistic method to populate parameters have been developed and tested, and the findings were reported in the quarterly report.
Task 3.2 Develop a prototype tool on parameter population	**Develop an open-source prototype tool to populate model parameters based on the methods in Task 3.1**	
M3.2.1 Parameter population tool requirements and specifications	Identification of user input and outputs for parameter generation tool. Description of tool requirements (for example, the allowed ranges and variability of the output given user specifications) Specifications for numerical thresholds/ranges and other metrics for “realism” will be defined, informed by Program Element 1 activities.	(6/30/2017) The team completed the identification of user input and outputs for parameter generation tool. Description of tool requirements were reported in the quarterly report.

M3.2.2 Completion of an initial prototype tool for parameter population capable for small-scale systems	Initial prototype tool developed for parameter population, targeting for small-scale systems (500 buses). The team will coordinate and compare results with Program Element 2 approach(es). The resulting synthetic system with the parameters from this task and topology from Task 2.2 will meet realism criteria defined in Project Element 1 and refined in M3.2.1.	(9/30/2017) The team completed the initial prototype development of the algorithms and tools for parameter population. The developed algorithms and tools were based on the topology generation tool developed on Program Element 2. The resulting synthetic system was validated with the metrics defined in M.1.3.1.
M 3.2.3 Complete developing the tool for generating topologies with documentation	Open-source prototype parameter population tool further developed and refined with experience from Project Elements 6 and 7. The underlying algorithms may be refined as well. Associated documentation will be developed. The resulting synthetic system with the parameters from this task and topology from Task 2.2 will meet realism criteria defined in Project Element 1 and refined in M3.2.1.The planned documentation and summary slides will be delivered to ARPA-E Program Director.	(12/15/2018) The team completed the refine of the open-source prototype parameter population tool with additional requirement from PJM (no more than 50% of the buses can come from the generic PJM model). Several new synthetic systems (500-bus, 700-bus, 2000-bus, 3000-bus and 5000-bus) was created and validated through the realism criteria defined in Project Element 1 and refined in M3.2.1. The documentation of the prototype tool has been delivered to ARPA-E Program Director.
Project Element 4: Data anonymization		
Task 4.1 Develop methods for data anonymization	**Develop methods for anonymizing power grid ancillary datasets**	
M 4.1.1 Completion of algorithm(s) to anonymize power grid ancillary datasets	Selection, development, and verification of methodologies and supporting software to anonymize distribution datasets and transmission ancillary datasets. This will incorporate different levels of data anonymization, depending upon the needs of the user. Methods will be identified to anonymize at least the following types of information: locations, connectivity, and parameters. Measure of success will include verification against the metrics in Project Element 1 and the experience from previous OMF development.	(6/30/2017) The team completed the selection, development and verification of methodologies and supporting software to anonymize the distribution datasets and transmission ancillary datasets.
Task 4.2 Develop a prototype tool for data anonymization	**Develop an open-source prototype tool to anonymize dataset parameters**	
M4.2.1 Anonymization tool requirements	Identification of user input and outputs for anonymization tool. Also outline of what constitutes "sufficient anonymization" for project goals and technical description of how this will be achieved.	(6/30/2017) The team completed identification of user input and outputs for anonymization tool.
M 4.2.2 Completion of the tool development for anonymizing power grid data, based on users' needs	Data anonymization tool developed with associated documentation based on the algorithm(s) from Task 4.1. Verification that tool meets specification of M4.2.1.	(9/30/2017) The team completed the development of the data anonymization tool with associated documentation. Verification results indicates that tool meets specification of M4.2.1.
M 4.2.3 Release of anonymized distribution systems	At least four distribution systems anonymized by the tool, approved by the utility owners for public use, and released. Delivery of summary report to ARPA-E Program Director describing the differences in powerflow results between the original and anonymized data sets.	(12/15/2018) The team completed the five distribution systems anonymized by the tool, approved by the utility owners for public use, and released. Summary report delivered to ARPA-E Program Director.
M 4.2.4 Updated version of the anonymization tool	Updated version of the anonymization tool released, addressing enhancements and defects discovered during user testing.	(9/30/2018) The team completed the release of the developed anonymization tool.
M 4.2.5 Release of anonymized transmission systems	At least one transmission system anonymized by the tool, approved by the utility owner for public use, and released. Delivery of summary report to ARPA-E Program Director describing the differences in powerflow results between the original and anonymized data sets.	(9/30/2018) The team completed the release of the anonymization of the transmission system. Summary report was delivered to ARPA-E Program Director.

Project Element 5: Validation tools		
Task 5.1 Develop dataset validation methodology	**Develop a methodology for validating generated datasets**	
M 5.1.1 Complete the development of dataset validation methodology	Report delivered to ARPA-E and approved by the Program Director detailing the teams three level component, system, and functionality) validation process. Based on the three-level process, required tools will be identified for validating generated datasets. The report will detail quantitative validation success criteria for all three levels of validations (which may be different for small scale and large scale datasets). In addition, the report will detail the list of validation tools to be used along with any edits/additions to these tools required for successful validation in this project.	(3/31/2017) The team completed the development of dataset validation methodology. Detailed report was delivered to ARPA-E and approved by the Program Director.
Task 5.2 Configure validation tools	**Configure tools in the EMS and DMS products by GE Grid Solutions (previously Alstom Grid) for validation purposes**	
M 5.2.1 Configured tools that support CIM-compatible and CIM-like datasets to validate the datasets created in Task 5	New validation interface to GE EMS and DMS tools developed. The tools will be updated and configured to support CIM-compatible and CIM-like datasets. Other necessary enhancements may be performed (as required by the validation methodology developed in Task 5.1). Confirmation that the modified/developed tools adhere to specifications in M5.1.1.	(6/30/2017) The team completed the development of the new validation interface to GE EMS and DMS tools.
Task 5.3 Install the properly configured tools at PNNL	**Install tools at PNNL and provide training**	
M 5.3.1 Validation tools ready at PNNL	Validation tools installed at PNNL. Relevant PNNL staff trained using these tools. Consulting activities will continue throughout Year 2 to assist the team in performing validation.	(10/3/2017) The team completed the installation of the validation tools at PNNL. Relevant PNNL staff trained using these tools.
Project Element 6: Transmission and distribution basecases		
Task 6.1 Generate and validate a small-scale transmission OPF basebase	**Generate and validate a small transmission basecase using the tools in Tasks 2, 3, 4, and 5**	
M 6.1.1 Creation of a validated small-scale transmission OPF base case with ~500 buses	Delivery of a validated small-scale (500-bus) transmission OPF base case with documentation. Successful completion of three level validation process documented in M5.1.1 on the small-scale base case created using the identified tools. (Additional tool enhancements may be identified and performed in the process of data creation and validation.) Confirmation that the validation meets success criteria specified in M5.1.1. A report will be prepared for ARPA-E Program Director to address this go/no-go milestone.	(03/31/2018) The team delivered a 563-bus transmission OPF base case with documentation. The base case was validated successfully through the three-level validation process in M.5.1.1. A report has been delivered to ARPA-E Program Director.
Task 6.2 Generate and validate T+D basecases with documentation	**Generate and validate small and large T+D basecases using the tools in Tasks 2, 3, 4, and 5**	
M 6.2.1 Complete the creation and validation of a small-scale T+D basecase	Delivery of a validated small-scale T+D basecase (500 transmission buses and 1000 distribution buses) with documentation. Successful completion three level validation process documented in M5.1.1 on this basecase using the identified tools. (Tools may be further refined.) Confirmation that the tools adhere to specifications in M5.1.1. Results and findings will be reported in the quarterly report.	(03/30/2018) The team delivered a validated small-scale T+D base case (563-bus transmission model, in which 104 load buses are connected with detailed feeder models, total 162, 827 nodes for distribution feeder models.) The T+D base case was validated through three level validation and the results and findings were reported in the quarterly report.

M 6.2.2 Complete the creation and validation of a large-scale T+D basecase	Delivery of a validated large-scale T+D basecase (100,000 buses) with documentation. Successful completion three level validation process documented in M5.1.1 on this basecase using the identified tools. (Tools may be further refined.) Confirmation that the tools adhere to specifications in M5.1.1 A summary report will be prepared to ARPA-E Program Director.	(06/30/2018) The team delivered a validated large-scale T+D base case: 4918-bus transmission system, in which 193 load buses are connected with detailed feeder models, total 1,296,446 nodes for distribution feeder models. The 4918-bus synthetic transmission system was created and validated through the realism criteria defined in Project Element 1 and refined in M3.2.1, for both topology and parameter metrics. The distribution data was anonymized from the taxonomy feeders with parameters populated to represent detailed commercial and residential load models. A summary report was prepared for ARPA-E Program Director.
Project Element 7: Scenario generation		
Task 7.1 Enhance OMF capability to specify configuration of the desired scenarios	**Enhance Open Model Framework (OMF) to be compatible with the created basecases in Task 6**	
M 7.1.1 Completion of OMF GUI and configuration for compatibility	GUI developed to allow researchers to generate, view and download scenarios. The GUI is capable of handling at least 1000 scenarios. The GUI will be accessible on the web as part of the Open Modeling Framework (www.omf.coop). Configuration will be completed to facilitate scenario generation. Progress will be reported in the quarterly report.	(03/30/2018) The team completed the GUI development to allow researchers to generate, view and download scenarios. The GUI is accessible on the web as part of the Open Modeling Framework (www.omf.coop). Progress was reported in the quarterly report.
Task 7.2 Generate and validate massive scenarios using the created basecases	**Generate scenarios with 5 min resolution over a year for the T+D basecases created in Task 6**	
M 7.2.1 Delivery of massive scenarios for the created small-scale T+D basecase in 6.2.1.	Delivery of at least 1,000 five-minute scenarios generated for the small-scale T&D basecase. Each scenario will include the data attributes of the basecases and the specified time-series features such as load profiles. Each scenario will be validated in the same way as the basecases and meet all requirements documented in M5.1.1. Complete validation and review of the generated scenarios for their realism (against metrics in Project Element 1), completeness (including all data attributes and time-series features), and open-access nature. Progress will be reported in the quarterly report.	((06/30/2018)) The team completed the creation of the 1000 different scenarios for the 563-bus transmission system, in which 104 load buses are connected with detailed feeder models, total 162, 827 nodes for distribution feeder models. Each scenario includes different detailed commercial and residential load conditions in distribution system, as well as different load profile and unit commitment (generator on/off) and generation dispatches in the transmission system. Each scenario was validated in the same way as the base-cases. Progress was reported in the quarterly report.
M 7.2.2 Delivery of massive scenarios for the created large-scale T+D basecase in 6.2.2.	Delivery of at least 105,120 five-minute scenarios generated for the large-scale T&D basecase. Each scenario will include the data attributes of the basecases and the specified time-series features such as load profiles. Each scenario will be validated in the same way as the basecases and meet all requirements documented in M5.1.1. Complete validation and review of the generated datasets for their realism (against metrics in Project Element 1), completeness (including all data attributes and time-series features), and open-access nature. A summary report will be prepared for ARPA-E Program Director.	(09/30/2018) The team completed the creation of the 105,120 different scenarios for the 4918-bus T+D system, in which 193 load buses are connected with detailed feeder models, total 1,296,446 nodes for distribution feeder models. Each scenario includes different detailed commercial and residential load conditions in distribution system, as well as different load profile and unit commitment (generator on/off) and generation dispatches in the transmission system. Progress was reported in the quarterly report.
Project Element 8: Compatibility with data repository		

Task 8.1 Finalize data schema and formats to facilitate the design of data repository	**Finalize data schema and formats to facilitate the design of data repository. Frequent interaction with data repository development team to ensure a smooth transition of the data creation tools and datasets.**	
M 8.1.1 Common data schema and formats used by the data repository	Dataset schema and formats (in documents) delivered to the data repository team to help design the data repository. Compatibility will be determined by confirmation and data ingestion by at least one data repository team. Progress will be reported in the quarterly report.	(09/30/2017) The team completed the design of the dataset schema and formats (in documents) delivered to the data repository team (DR. POWER) to help design the data repository. Progress was reported in the quarterly report.
Task 8.2 Transfer SDET outcomes into the repository	**Transfer SDET tools, datasets, and scenarios into the data repository**	
M 8.2.1 Completion of integration of the created tools, basecases and scenarios into the data repository	Completion of the integration of the developed data-creation tools and generated datasets in the data repository. A summary report will be prepared for ARPA-E Program Director describing the completion of transfer.	(09/30/2018) The team completed the integration of the developed data-creation tools and generated datasets in the data repository with data repository team DR. POWER. Progress was reported in the quarterly report.
Project Element 9: Project management		
Task 9.1 Monitor and report project progress and facilitate team activities	**Project management to ensure successful execution of the project**	
M 9.1.1 Manage project progress, coordinate team efforts, and interact with ARPA-E program managers and project advisors	Throughout this project, the team will actively interact among the team members as well as with project advisors. The team will ensure NDAs are signed quickly and timely, which is critical to the success of this data-heavy project. Quarterly reporting to ARPA-E is a minimum.	(12/15/2018) The team actively interacted among the team members as well as with project advisors. The team ensured NDAs were signed quickly and timely. Progress was reported in the quarterly reports.
M 9.1.2 NDA among team members	Sign NDAs with team members to exchange and protect sensitive information	(03/31/2017) The team signed NDAs with team members to exchange and protect sensitive information.
M 9.1.3 Contracts for team members	Sign contracts/subcontracts for all team members	(04/05/2017) The team signed contracts/subcontracts for all team members.
Project Element 10: Technology transfer and outreach (TTO)		
Task 10.1 TTO - Strategic engagement for adoption & impact		
M 10.1.1 Stakeholder analysis	Updated T2M plan delivered to ARPA-E and approved by the Program Director. The T2M plan will: a. Map out broad landscape of strengths and weaknesses of existing models and new use cases enabled by the team's new model creation techniques. b. Identify early model adopters (by project end), possible late adopters (once proven and validated). Determine critical decision makers and outline what is not known about model user requirements and constraints. c. Describe process for adoption from start to finish, focusing on barriers and important actions that must be taken by the team to promote adoption.	(03/31/2017) The team delivered the T2M plan to ARPA-E and approved by the Program Director. The T2M includes: 1) new capabilities of the new models and new use cases enabled by the team's new techniques. 2) identification of early model adopters as the ARPA-E GO Competition team. 3) schedule of data delivered to the ARPA-E GO Competition team.

M 10.1.2 Targeted engagement	Update delivered to ARPA-E on targeted engagement activities to drive model adoption by specific stakeholders identified in M 10.1.1. The update should include a summary of the team's efforts to: a. Present update on engagement with specific stakeholders to understand individual needs and initiate the adoption process. Detail strategy for timing of future engagement during the remaining project period. b. Provide tailored documentation and user support for setup and use of the models and any associated software (e.g. format conversion tools) by individual early adopters.	(12/31/2018) The SDET team interacted with the ARPA-E GO Competition frequently to understand the additional needs and requirements from the GO Competition team for preparing the data set. The SDET team also presented the updates frequently to the GO Competition team with software tools and documentations about the scenarios specially created for the GO Competition.
Task 10.2 TTO - User validation and feedback		
M 10.2.1 User validation	Updated T2M plan delivered to ARPA-E and approved by the Program Director. The T2M plan should summarize the results of conducting interviews with stakeholders and potential users outside of core project advisors to validate that: a. Power system model creation methodologies meet stakeholder requirements. Validate that models created using the team's proposed methodology will be viewed by key stakeholders/users as sufficiently realistic and suitable for public release. b. Model data formats are compatible with requirements for widespread community adoption and reasonably future-proof.	(12/31/2018) The SDET team interacted with the ARPA-E GO Competition team frequently to make sure the power system model creation methodologies met the GO Competition team's need, as well as the data format of the models are compatible with GO Competition team's need.
M 10.2.2 Preliminary results	Update delivered to ARPA-E on new power system model characteristics relative to community needs and hands-on feedback from early users: a. Obtain detailed hands-on feedback from early users. Validate that model characteristics and any associated software meet community needs and user goals (e.g. realism, level detail, computational performance, flexibility, visualizations, user interface). b. Quantify adoption of the preliminary models (e.g. number of downloads, conference presentations, citations, industry inquiries, etc.)	(12/31/2018) SDET team worked closely with the GO Competition team and supported the GO Competition. 700-bus, 2000-bus, 3000-bus, and 5000-bus scenarios were delivered to the GO Competition. A methodology and software tool were specially developed for the GO Competition team, which could derive and validate a feasible solution for the security constraint optimal power flow problem for each of the provided scenario. The software tool was sent to the GO Competition team for evaluation and feedback.
Task 10.3 TTO - Community building and transition		
M 10.3.1 Community building	Update delivered to ARPA-E on the team's specific actions taken to involve a broader cross-section of relevant stakeholders. The team should summarize efforts to: a. Raise awareness of the project's technical approach through publications and presentations at major industry events. Promote the goals of the GRID DATA program and associated OPF competitions. b. Attract additional partners (ISOs, utilities, software vendors, etc.) to share data for model development and/or inclusion in the models. c. According to PD's discretion and guidance, conduct workshop(s) to engage collaborators from the technical community and educate potential users.	(12/31/2018) SDET team worked closely with the GO Competition team and supported the GO Competition. 700-bus, 2000-bus, 3000-bus, and 5000-bus scenarios were delivered to the GO Competition. A methodology and software tool were specially developed for the GO Competition team, which could derive and validate a feasible solution for the security constraint optimal power flow problem for each of the provide scenario. The software tool was sent to the GO Competition team for evaluation and feedback.
M 10.3.2 Post ARPA-E transition	Update delivered to ARPA-E on end-of-project transition activities to ensure the continued development and adoption of models and methodologies. The update should detail the team's efforts to: a. Finalize user documentation for models and any associated software (e.g. setup and maintenance, file formats, data processing tools, user interfaces, etc.) b. According to PD's discretion and guidance, develop curriculum materials to educate users and/or technical	(12/31/2018) SDET team worked closely with the GRID DATA DR. POWER team for the end-of-project transition activities to ensure the continued development and adoption of models and methodologies. The user documentation for models and associated software were delivered to the DR. POWER team. The team also secured funding for further model development and maintenance

	contributors. c. Secure funding for further model development and/or maintenance.	from both the GO Competition team and the DR. POWER team.

Project Activities

In this project, the research team derived metrics that define the realism of power grid datasets from industry power grid datasets provided by its industry partners. The research team developed algorithms and software prototypes that could generate topology, populate parameters, and create abundant scenarios for the synthetic power grid models that meet the realistic metrics. Furthermore, by deploying the software prototypes developed in this project, the research team generated and released large-scale Transmission and Distribution base cases and a large number of scenarios based on the generated base cases. The key innovations and achievements are summarized below:

- Based on the industry power grid datasets provided by its industry partners, the research team defined realistic graph-theoretical metrics for the power grid model topology and realistic power system parameters metrics for power grid model network parameters, generation and load parameters, as well as economic parameters for AC OPF.
- The research team designed and developed the SDET architecture that integrates the realism metrics definition, topology generation, and parameter population algorithms and software prototypes. The whole SDET architecture is shown in Figure 1.

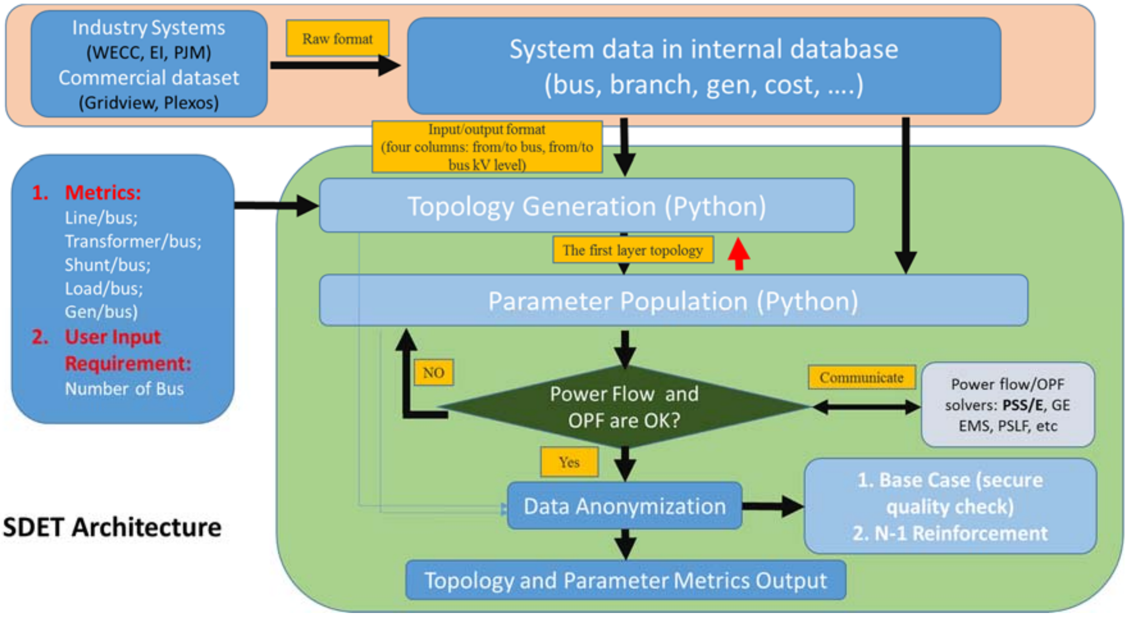

Figure 1. SDET architecture

- The research team developed innovative power grid topology generation algorithms and software prototypes that could derive synthetic and realistic power grid transmission system topology that satisfies the derived realism metrics and user's needs. Figure 2 gives an example of the generated topology for a 5000-bus power grid model

Figure 2. Topology structure of the 5000-bus power grid models

- The research team developed innovative power grid parameter population algorithms and software prototypes that could populate parameters onto the topology generated using graph methods and derive synthetic and realistic power grid models that satisfies the derived realism metrics and user's needs.
- The research team developed innovative algorithms and software prototypes that could generate different scenarios for Transmission & Distribution base cases.
- The research team developed a set of 500-bus, 2000-bus, 3000-bus, and 5000-bus synthetic power grid transmission models that satisfies the derived realism metrics. All the synthetic models were released for public access through the DR. POWER website.
- The research team developed 1000 different scenarios for a 563-bus transmission system, in which 104 load buses are connected with detailed feeder models, total 162, 827 nodes for distribution feeder models. Each scenario includes different detailed commercial and residential load conditions in distribution system, as well as different load profile and unit commitment (generator on/off) and generation dispatches in the transmission system. Each scenario was validated against the realism metrics. All the scenarios were released for public access through the DR. POWER website.
- The team developed 105,120 different scenarios for a 4918-bus T+D system, in which 193 load buses are connected with detailed feeder models, total 1,296,446 nodes for distribution feeder models. Each scenario includes different detailed commercial and residential load conditions in distribution system, as well as different load profile and unit commitment (generator on/off) and generation dispatches in the transmission system. Progress was reported in the quarterly report. Each scenario was validated against the realism metrics. All the scenarios were ready to release for public access through the DR. POWER website.

Project Outputs

A. Journal Articles

B. Papers

Stephen J. Young, Yuri Makarov, Ruisheng Diao, Rui Fan, Renke Huang, James O'Brien, Mahantesh Halappanavar, Mallikarjuna Vallem, and Zhenyu Huang, "Synthetic Power Grids from Real World Models," in the Proceedings of the 2018 IEEE PES General Meeting, Portland, U.S.A, Aug. 2018.

Stephen J. Young, Yuri Makarov, Ruisheng Diao, James O'Brien, Renke Huang, Rui Fan, Mallikarjuna Vallem, Mahantesh Halappanavar, and Zhenyu Huang, "Topological Power Grid Statistics from a Network-of-Networks Perspective," in the Proceedings of the 2018 IEEE PES General Meeting, Portland, U.S.A, Aug. 2018.

C. Status Reports

8 quarterly reports summarizing the project progress over the past 2 years were submitted to ARPA-E program directors.

D. Media Reports

E. Invention Disclosures

An invention disclosure was filed at PNNL, titled "Sustainable Data Evolution Technology (SDET) for Power Grid Optimization".

F. Patent Applications

G. Licensed Technologies

An invention disclosure has been issued for the developed copy right software "Sustainable Data Evolution Technology (SDET) for Power Grid Optimization".

H. Networks/Collaborations Fostered

During the course of this project, the team established strong connections with power industry, including PJM and GE Energy. The team is currently supporting to provide data set for the ARPA-E Grid Optimization Competition.

I. Websites Featuring Project Work Results

https://egriddata.org/group/sustainable-data-evolution-technology-sdet

J. Other Products (e.g. Databases, Physical Collections, Audio/Video, Software, Models, Educational Aids or Curricula, Equipment or Instruments)

The source code of the develop software prototype tool will be sent to ARPA-E.

K. Awards, Prizes, and Recognition

The paper "Synthetic Power Grids from Real World Models" was awarded as one of the four "Conference Prize Paper" among more than 1400 high quality conference papers in the 2018 IEEE PES General Meeting.

Developed Algorithms and Technologies

A. Architecture of the Sustainable Data Evolution Technology

The SDET differentiates itself from other approaches in its novel concept of data evolution. The power grid and its challenges are evolving; so should the datasets that support the development of power grid tools. Our dataset creation is not a one-time effort but a set of reusable data creation tools that interested parties can use to create new and up-date existing datasets. The innovation is to derive features in terms of topologies, parameters, and scenarios from industry-provided private datasets and develop tools that can duplicate these features in publicly available datasets without disclosing the original private datasets. The process of creating a converged base case using SDET method is illustrated in Figure 3.

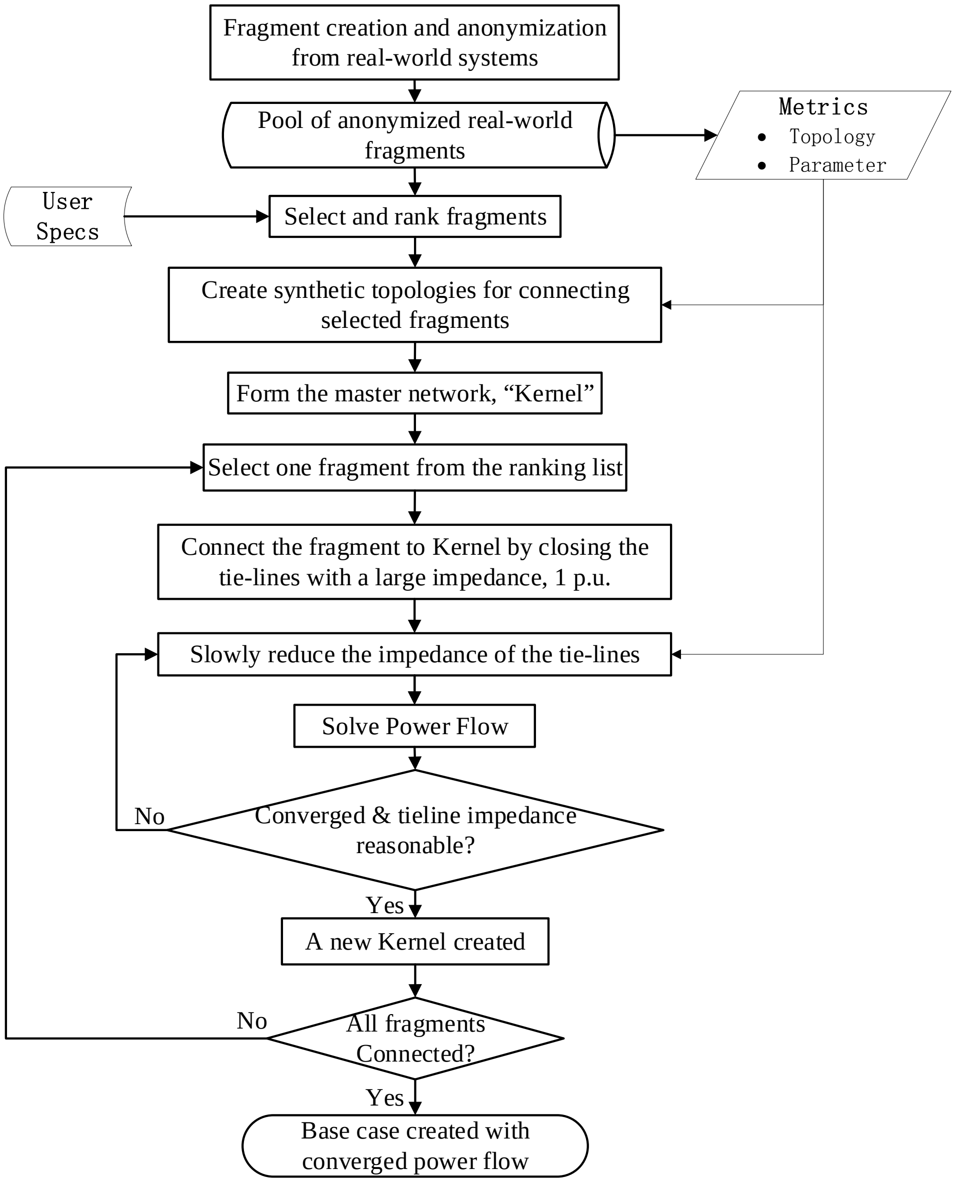

Figure 3. Flow chart of power grid model creation using SDET

The SDET consists of several key steps to create a converged power flow base case:

Fragmentation – real-world power system models, e.g., base cases for planning studies, are used to create fragments based on the actual zone definition. To create a fragment for a desired zone, the tie lines and boundary buses linking this zone to neighboring zones are identified and then removed. Equivalent generators and loads are then created to ensure power balance on the boundaries. A new swing bus is then assigned to ensure power flow convergence. In this way, all the detailed modeling information inside each zone is preserved.

Fragment anonymization – once a fragment is created, all the sensitive information including original bus names, bus numbers, geographical information, and others are replaced by randomly generated values. Generator limits, branch limits, and parameters can be randomly perturbed as well.

Topology generation – a graph-theory–based approach is developed to reconnect selected fragments together to form a large-scale power network model that satisfies the user's specifications and the graph metrics derived from large-scale, real-world models.

Parameters Population and Network Reinforcement – an innovative procedure is developed to ensure easy convergence of power flow for the generated synthetic power grid topology and to enforce the N-1 security for the power flow case.

B. Fragmentation and Anonymization

Fragments are created from real-world or realistic power system models and function as the fundamental building blocks for creating a synthetic system. Several techniques are used to define the buses that belong to a particular fragment, including existing zone information, modularity-based clustering, and local clustering techniques based on the heat kernel page rank. The combination of these techniques in even relatively small power grids (~3000 buses) resulting in hundreds of fragments. Once a collection of buses is identified, a fragment is created by replacing the tie lines between the collection and the rest of the system with a line connected to an equivalent load or generator. The buses with equivalent generators and loads are referred to as boundary buses. Each fragment is "cleaned and polished" to ensure good quality before being used to build a synthetic power system.

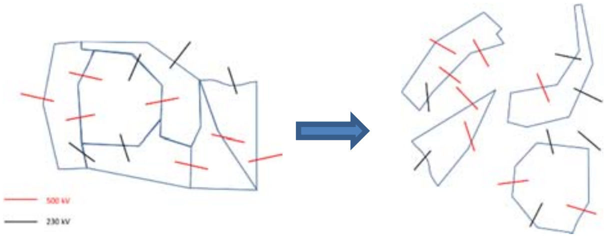

Figure 4. Conceptual illustration of fragmentation: (left) parental system divided in fragments (zones); (right) fragmented parental system.

C. Topology Generation Considering Realistic Properties

The goal of topology generation for the synthetic power grid is to reassemble a selected sets of fragments into a synthetic and realistic power grid transmission system that satisfies the derived realism metrics and user's needs by the addition of new tie lines between equivalent loads and generators. In or-der to distinguish between these new tie lines and existing ones, we refer these tie lines added between fragments as “connectors.” Additionally, the reassembled system should have a meaningful zone structure that is reflective of the underlying geography. Further, as we wish to associate a converged power flow with the reconnected system, it is desirable to minimize the voltage mismatch between connected boundary buses. In a computationally unbounded environment, these four goals would be achieved by exploring the Pareto frontier of these four goals. However, for many of these goals determining the optimality of a particular metric is either computationally expensive (ASPL) or intractable (e.g., number of buses, zone geography, and equivalent bus voltage imbalances). Thus, in order for the developed SDET methodology to be flexible enough to create a wide variety of different types of power systems, including variation in the number of buses, lines, trans-formers, loads, generators, switched shunts, voltage levels, etc., the reconstruction process has a multistage planning phase to ensure computational tractability as well as system realism, followed by a connector integration which produces the final power system.

In ***Planning Phase One***, fragments are selected from the pool to match the user specified generic power system properties, such as the number of buses, lines, transformers, loads, generators, switched shuts, etc. as well as the voltage levels present. These properties typically vary based on the scale and type of the system that is being modeled and can vary independently of other parameters.

In this phase, the fragments are selected from the col-lection of available fragments in an iterative manner. The first fragment is selected to minimize the L2 deviation from the desired user properties (scaled for the number of buses). In order to select subsequent fragments, we maintain an error environment which represents the necessary relative deviation from the user requirements given the currently selected fragments. We determine this error environment by considering the linear programming relaxation of the fragment selection process. The remaining fragments are then selected randomly from those that are consistent with the current error environment and which have an equivalent generator (respectively, load) that matches the nominal voltage of an equivalent load (respectively, generator) of the previously selected fragments. If there are no such fragments, the error environment is increased relative to one of the tight constraints in the linear programming relaxation. This iterative process terminates when the selected fragments satisfy the current error environment.

Planning Phase Two imposes a geographic structure on the fragments by limiting the connectors that could potentially be added in subsequent planning phases. This geo-graphic structure focuses on ensuring two global properties of the power system: robustness to islanding under (n-1) contingency and minimal crossing of lines. As the first of these is not an absolute prescription, we will only consider islanding by a small selection of lines, namely, the connectors between fragments. This approach limits the over-all occurrence of islanding while allowing for some realis-tic islanding to occur. For the second global property, we note that it can be efficiently determined whether there is an instantiation of the power system where no two lines cross, while anything beyond this significantly more challenging. Specifically, determining the number of times power lines must cross in a power system is equivalent to finding what is known as the crossing number of a graph. There is a long history of study the crossing number including the introduction (intentionally and unintentionally) of numerous variants. For all but the smallest power system even estimating the crossing number is computationally impractical. Thus, to mitigate this issue, we require that the created power system model is flat at the fragment level. This requirement can be understood by viewing the fragments as zones in the synthetic power system, and noting that since zones in the power system correspond to (typically contiguous) geographic regions, the structure of the connectors between them should be close to planar. It is worth mentioning that even if all the fragments are planar and the connector structure is planar, this does not imply that the overall structure is planar nor that in the representation with the fewest crossings that none of the connectors cross.

To achieve both these global structure goals, we build an auxiliary network F on the fragments from three collections of edges. In subsequent planning phases, a connector will be allowed only between fragments that are joined by an edge in F.

The first collection of edges is formed by building a weight graph on the fragments, where the weights are the minimum squared difference in voltage between a boundary generator and a boundary load bus at the same nominal voltage level. Within this graph, a minimum spanning tree is then chosen and those edges are added to F. The weights in this phase are chosen to maximize the chance that the overall power system will be connected, as the SDET procedure for adding connectors prefers those with lower voltage difference.

To select the second collection of edges, a graph is built on the fragments where the weight is the number of potential tie lines between two fragments. If the edge was included in phase one, its weight is fixed at 0. From this network, the maximum weight two-factor is chosen in order to in-crease the likelihood that the resulting power system is robust to the deletion of any connector. Since the union of a spanning tree and a two-factor can be non-planar, some of the edges in the two-factor may not be added to F.

The final collection of edge is determined by sequentially adding all other possible edges to F in order of decreasing number of tie lines and excluding those edges that result in F being non-planar. The purpose of this collection of edges is to be as parsimonious as possible in restricting the choice of connectors.

The goal of ***Planning Phase Three*** is the selection of connects which achieve the realism bounds for the NoN properties. As a reminder, the NoN properties are statistical topological properties, such as clustering coefficient, diameter, and average shortest path length (ASPL), associated with maximal subnetworks in the system with constant nominal voltage. These topological properties have been shown to have similar behaviors independent of the power system, voltage level, and size of the system. For instance, both the diameter and ASPL scale (to first order approximation) with the square root of the number of buses in the subnetwork. Thus, we will ensure that all of the generated networks match these statistical properties up to some specified noise tolerance. It is worth noting that for each voltage level the number of potential sets of connectors is exponential in the number of boundary buses and thus it is impractical to consider all possible collections for even one voltage level, let alone all the voltage levels in the system. As a consequence, the proposed SDET procedure in this section is devoted to proceeding through the possible choices of connectors in a structured and intelligent manner.

The first step is to observe that the addition of a connector between fragments has minimal effect on the degree distribution (as it only changes a relatively few number of degree one vertices to degree two) and essentially no effect on the clustering coefficient (as no triangles can be created by the process) and thus, al-most by definition, the degree distribution and clustering coefficient of the resulting network will be realistic. The second step is to break each of the fragments into connected networks with the same nominal voltage and discard those networks which do not contain a boundary bus. Again, the addition of connectors will not change the diameter or average shortest path length for these networks, and they are, by definition realistic, so they can be safely ignored.

For the remaining subnetworks, a bipartite graph can be built for each nominal voltage level. One side of the bipartite graph is the boundary loads and the other contains the boundary generators for the given voltage level. A load and a generator are connected with an edge if their associated fragments are adjacent in F. The weight of this edge is fixed to be the squared difference in actual voltages be-tween the endpoints. The minimum cost assignment can then be found between the two load and generation boundary buses by applying the Hungarian algorithm for weighted matching [30]-[31]. By selectively removing edges and keeping a priority queue of the results, one can iterate through all of the possible sets of connectors in order of decreasing size and increasing total squared voltage difference. This process (returning intermediary steps as appropriate) is continued until a set of connectors which satisfy all the desired metrics has been found or the computational budget has been exhausted.

The final planning phase, ***Planning Phase Four***, focuses on further reinforcing the system against (n-1) contingencies. Specifically, it adds a minimal set of connectors to ensure that each fragment (zone) has connectors to two different fragments, and that no one fragment separates the overall power system. This phase relies on the current “fragment graph” associated with the connectors, that is, the graph whose vertices are fragments and where there is an edge between two fragments if there is a connector be-tween two fragments. The goal of this phase may be alter-natively expressed as adding connectors to ensure that the fragment graph has a minimum degree of two and is bi-connected (that is, every pair of fragments is in a cycle). To achieve this goal, we proceed in two stages. In the first stage an auxiliary graph is constructed from the connectors which have one end in a degree one fragment in the fragment graph. A subset of the connectors is then selected and added to network by finding a maximum matching in the auxiliary graph.

In the second stage, the block-cut tree for the bi-connected components of the fragment graph is built as an auxiliary graph. Each potential connector is then weighted by the distance between its endpoints in the block-cut tree and the connector with the largest distance is added to power sys-tem. This process is repeated until the fragment graph is two-connected.

D. Parameters Population and Network Reinforcement

Following the procedures introduced in the above Section of Topology Generation, a realistic power system model can be obtained, at least from a structural (topology) point of view. However, from an electrical point of view, the system has several significant problems, most notably, the connectors between fragments have very large impedances to prevent each connected fragment exchange power between each other. In the parameter population and network reinforcement phase, more realistic electrical parameters are introduced to the tie lines to generate a converged base-case power flow and ensure security following N−1 contingencies.

Connector Integration and Base-Case Power Flow

The process starts with reconnecting the selected fragments and developing a base power flow by introducing the connectors identified in the topology generation phase with a very large impedance. It is worth noting that because every fragment comes equipped with a converged power flow, this process typically provides a converged power flow for an electrically islanded system without issues.

In order to close all the connectors, reduce their impedances to reasonable levels, and preserve a converged power flow, an iterative process is developed as illustrated in Figure 3. The fragments with the smallest total generation are connected first, and the main network being built is referred to as the “kernel.” For all subsequent stages, the fragment outside the kernel that has the smallest total generation among all fragments is selected to connect to the kernel. We then close the connectors between the kernel and the chosen fragments and re-solve the power flow. The impedance on the connectors added in this step is gradually reduced until they reach a realistic value. This process terminates after all fragments have joined the kernel.

Power Flow Reinforcement for the Base Case and un-der N−1 Contingencies

In this stage, several steps are taken to reinforce the created base case to better comply with North American Electric Reliability Corporation (NERC) reliability performance standards, as illustrated by Figure 5. Specially, the power flow of the system is reinforced to ensure that there are no bus voltage or line flow violations for both the base case and the case under N−1 contingencies; see Table 2.

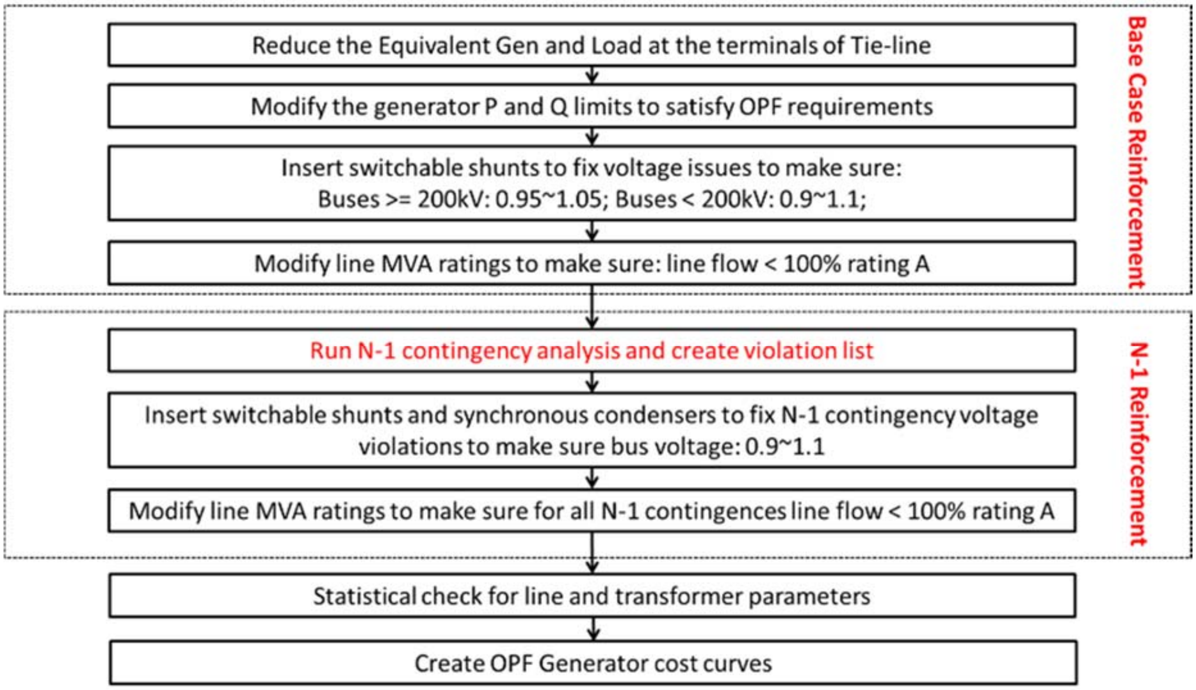

Figure 5. Flowchart of model reinforcement for the base case and under N−1 contingencies.

Table 2. Reinforcement ranges

	Base Case	*N*–1 contingency
Bus Voltage	Above 200 kV: 0.95 – 1.05 p.u.	0.9 – 1.1 p.u.
	Below 200 kV: 0.9 – 1.1 p.u.	

Line Flow	90% of Rate A	100% of Rate C

A few key steps for system reinforcement are summarized below:

1) Base-Case Reinforcement
    a) Reduce generation/load at terminals of connectors.
    b) Modify P and Q limits to satisfy OPF requirements.
    c) Insert switchable shunts to fix bus voltage is-sues.
    d) Modify line MVA ratings to fix line flow issues.
2) N −1 Contingency Reinforcement
    a) Insert switchable shunts and synchronous condensers to fix bus voltage issues.
    b) Modify line MVA ratings to fix line flow issues.

It is worth mentioning that Step (1a) plays an important part in increasing the realism of the resulting power system. In particular, the equivalent loads and generation produced by the fragmentation process are (unsurprisingly) typically larger than one would expect internally to the system because they represent the original inter-zone power flows. Thus, by “canceling out” the equivalent generation and load at these equivalent generators, the resulting base case will be more realistic.

At this point, a reinforced and realistic synthetic power system is generated, which can either be used as-is or further extended to help validate algorithmic performance. For instance, one could add generator cost curves to this system by taking random samples from a large set of cost curves and rescaling them to match the generator ranges specified. The resulting system could then be used to help validate OPF algorithms and perform N−1 contingency analysis.