

- [20] A. J. Conejo, L. Baringo, S. J. Kazempour, and A. S. Siddiqui, *Investment in Electricity Generation and Transmission*. Berlin, Germany: Springer, 2016.
- [21] K. Willis and G. Garrod, "Electricity supply reliability: Estimating the value of lost load," *Energy Policy*, vol. 25, no. 1, pp. 97–103, 1997.
- [22] J. R. Fletcher, T. Fernando, H. H.-C. Iu, M. Reynolds, and S. Fani, "Spatial optimization for the planning of sparse power distribution networks," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6686–6695, Nov. 2018.
- [23] H. Sangrody, N. Zhou, and X. Qiao, "Probabilistic models for daily peak loads at distribution feeders," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2017, pp. 1–5.
- [24] W. Lai, "Fitting power law distributions to data." [Online]. Available: https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Willy_Lai.pdf. Accessed on: Dec. 11, 2018.
- [25] Y. Dvorkin, Y. Wang, H. Pandzic, and D. Kirschen, "Comparison of scenario reduction techniques for the stochastic unit commitment," in *Proc. IEEE PES Gen. Meeting Conf. Expo.*, 2014, pp. 1–5.
- [26] H. Zhang, *Transmission Expansion Planning for Large Power Systems*. Tempe, AZ, USA: Arizona State University, 2013.
- [27] C. Lin, C. Fang, Y. Chen, S. Liu, and Z. Bie, "Scenario generation and reduction methods for power flow examination of transmission expansion planning," in *Proc. IEEE 7th Int. Conf. Power Energy Syst.*, 2017, pp. 90–95.
- [28] D. Arthur and S. Vassilvitskii, "k-Means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
- [29] M. Syakur, B. Khotimah, E. Rochman, and B. Satoto, "Integration K-means clustering method and elbow method for identification of the best customer profile cluster," in *Proc. IOP Conf. Ser. Mater. Sci. Eng.*, 2018, vol. 336, no. 1, Art. no. 012017.
- [30] R. Lleti, M. C. Ortiz, L. A. Sarabia, and M. S. Sánchez, "Selecting variables for K-Means cluster analysis by using a genetic algorithm that optimises the silhouettes," *Analytica Chimica Acta*, vol. 515, no. 1, pp. 87–100, 2004.
- [31] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. Roy. Statist. Soc., Ser. B (Statist. Methodol.)*, vol. 63, no. 2, pp. 411–423, 2001.
- [32] E. Rosenthal, "GAMS-A user's guide," GAMS Development Corporation, Washington, DC, USA, 2008.
- [33] "Solar home electricity data." [Online]. Available: <https://www.ausgrid.com.au/Industry/Innovation-and-research/Data-to-share/Solar-home-electricity-data>, Accessed on: July 3, 2018.

Mahdi Mehrtash (S'16–M'20) received the B.Sc. degree (First-Rank Hons.) in electrical engineering from the Shahid Chamran University of Ahvaz, Ahvaz, Iran, in 2012, the M.Sc. degree (First-Rank Hons.) in electrical engineering from Shiraz University, Shiraz, Iran, in 2014, and the Ph.D. degree in electrical engineering from Louisiana State University, Baton Rouge, LA, USA, in 2019. His research interests include power system optimization, stochastic optimization, and distributed optimization.

Amin Kargarian (M'14) received the B.S. degree in electrical and computer engineering from the University of Isfahan, Isfahan, Iran, in 2007, the M.S. degree in electrical and computer engineering from Shiraz University, Shiraz, Iran, in 2010, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2014. In 2014, he joined the ECE Department, Carnegie Mellon University as a Postdoctoral Research Associate. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA, USA. His research interests include power system optimization and machine learning.

Antonio J. Conejo (F'04) received the M.S. degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1987, and the Ph.D. degree from the Royal Institute of Technology, Stockholm, Sweden, in 1990. He is currently a Professor with the Department of Integrated Systems Engineering and the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA. His research interests include control, operations, planning, economics and regulation of electric energy systems, as well as statistics and optimization theory and its applications.

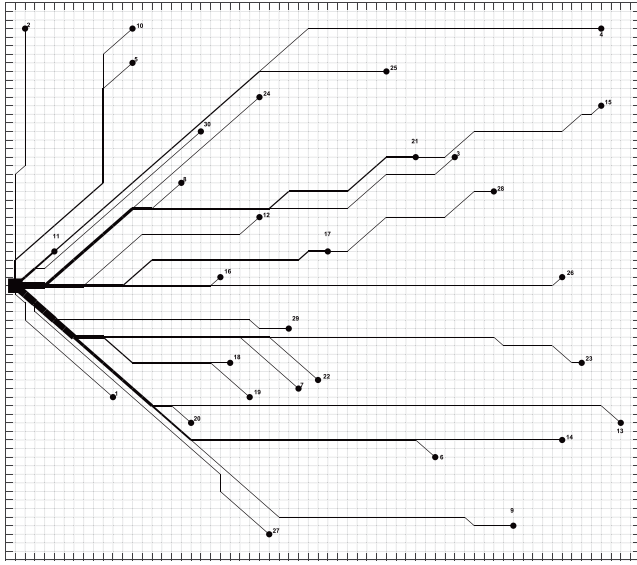


Fig. 12. Solution of stochastic MISOCP GIS-based FR for the large system.

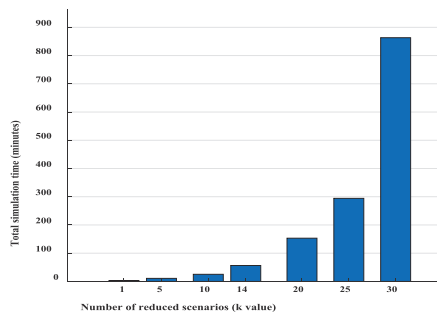


Fig. 13. Total simulation time for different numbers of clustered scenarios.

optimal objective function value is \$13.156M, and the optimal FR plan is shown in Fig. 12.

The simulation time depends on the number of clustered scenarios. The simulation time with respect to different numbers of clustered scenarios is shown in Fig. 13. For a number of clustered scenarios larger than 50, the solver did not provide a feasible solution after 100,000 seconds of computing time. However, as it can be observed in Fig. 11, a number of clustered scenarios (K value) that provides sufficient accuracy is 14. The total simulation time for 14 clustered scenarios is 56.5 minutes. Therefore, the proposed GIS-based FR has reasonable runtime for real-size systems.

VII. CONCLUSION

This paper proposes a novel MISOCP model for feeder routing using GIS data. Economic objectives, technical constraints, and geographical restrictions of FR are considered in the proposed model. Additionally, a model for the cost of distribution feeder resiliency using available GIS data is presented. The uncertainty of rooftop solar generations and demand forecasting errors are considered. A stochastic programming-based solution algorithm is developed to solve this problem. We have shown,

and illustrated by numerical results, that incorporating GIS data leads to a better (less expensive) FR solution. While the cost of the FR without GIS nodes in the case study is \$0.4275M, it is \$0.4114M for the GIS-based model. We numerically show that linearizing economic characteristics of individual conductors leads to a reasonable simulation time even for real-size systems (i.e., around 10 min for the case study). Furthermore, the presence of geographical obstacles leads to extra routing cost (by 1.06% in the case study) while it reduces the required simulation time by 68% (in the case study). In addition, considering the cost of resiliency, the proposed algorithm chooses to install feeders that are closer to the RMC node to take into account the resiliency target. Finally, numerical studies on a large test system illustrate that the proposed GIS-based FR has reasonable runtime for real-size systems.

REFERENCES

- [1] R. Brown, S. Gupta, R. Christie, S. Venkata, and R. Fletcher, "Automated primary distribution system design: Reliability and cost optimization," *IEEE Trans. Power Del.*, vol. 12, no. 2, pp. 1017–1022, Apr. 1997.
- [2] T. Fawzi, K. Ali, and S. El-Sobki, "A new planning model for distribution systems," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 9, pp. 3010–3017, Sep. 1983.
- [3] S. Jonnavithula and R. Billinton, "Minimum cost analysis of feeder routing in distribution system planning," *IEEE Trans. Power Del.*, vol. 11, no. 4, pp. 1935–1940, Oct. 1996.
- [4] E. Vugrin, A. Castillo, and C. Silva-Monroy, *Resilience Metrics for the Electric Power System: A Performance-Based Approach*. Albuquerque, NM, USA: Sandia National Laboratories, 2017.
- [5] "ArcGIS application in electric utilities." [Online]. Available: <https://www.esri.com/en-us/industries/electric-gas-utilities/segments/electric>. Accessed on: Apr. 10, 2018.
- [6] N. G. Boulaxis and M. P. Papadopoulos, "Optimal feeder routing in distribution system planning using dynamic programming technique and gis facilities," *IEEE Trans. Power Del.*, vol. 17, no. 1, pp. 242–247, Jan. 2002.
- [7] J. Shu, L. Wu, Z. Li, M. Shahidepour, L. Zhang, and B. Han, "A new method for spatial power network planning in complicated environments," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 381–389, Feb. 2012.
- [8] R. M. de Lima, R. Osis, A. R. de Queiroz, and A. H. M. Santos, "Least-Cost Path analysis and multi-criteria assessment for routing electricity transmission lines," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 16, pp. 4222–4230, 2016.
- [9] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1458–1459, Aug. 2006.
- [10] A. Lopez, "Seminar in theoretical computer science." [Online]. Available: <http://math.mit.edu/~goemans/18434S06/steiner-adriana.pdf>. Accessed on: Oct. 15, 2018.
- [11] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*. Berlin, Germany: Springer, 1972, pp. 85–103.
- [12] S. Mandal and A. Pahwa, "Optimal selection of conductors for distribution feeders," *IEEE Trans. Power Syst.*, vol. 17, no. 1, pp. 192–197, Feb. 2002.
- [13] A. Wu and B. Ni, *Line Loss Analysis and Calculation of Electric Power Systems*. Hoboken, NJ, USA: Wiley, 2016.
- [14] C. S. Park, G. Kim, and S. Choi, *Engineering Economics*. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.
- [15] H. L. Willis, *Power Distribution Planning Reference Book Second Edition, Revised and Expanded*. New York, NY, USA: Marcel Dekker, 2004.
- [16] N. Li, L. Chen, and S. H. Low, "Exact convex relaxation of OPF for radial network branch flow model," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun.*, 2012, pp. 7–12.
- [17] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2009.
- [18] A. Samui, S. Samantaray, and G. Panda, "Distribution system planning considering reliable feeder routing," *IET Gener., Transmiss. Distrib.*, vol. 6, no. 6, pp. 503–514, 2012.
- [19] R. H. Myers, *Classical and Modern Regression With Applications*. Belmont, CA, USA: Duxbury Press, 1990.

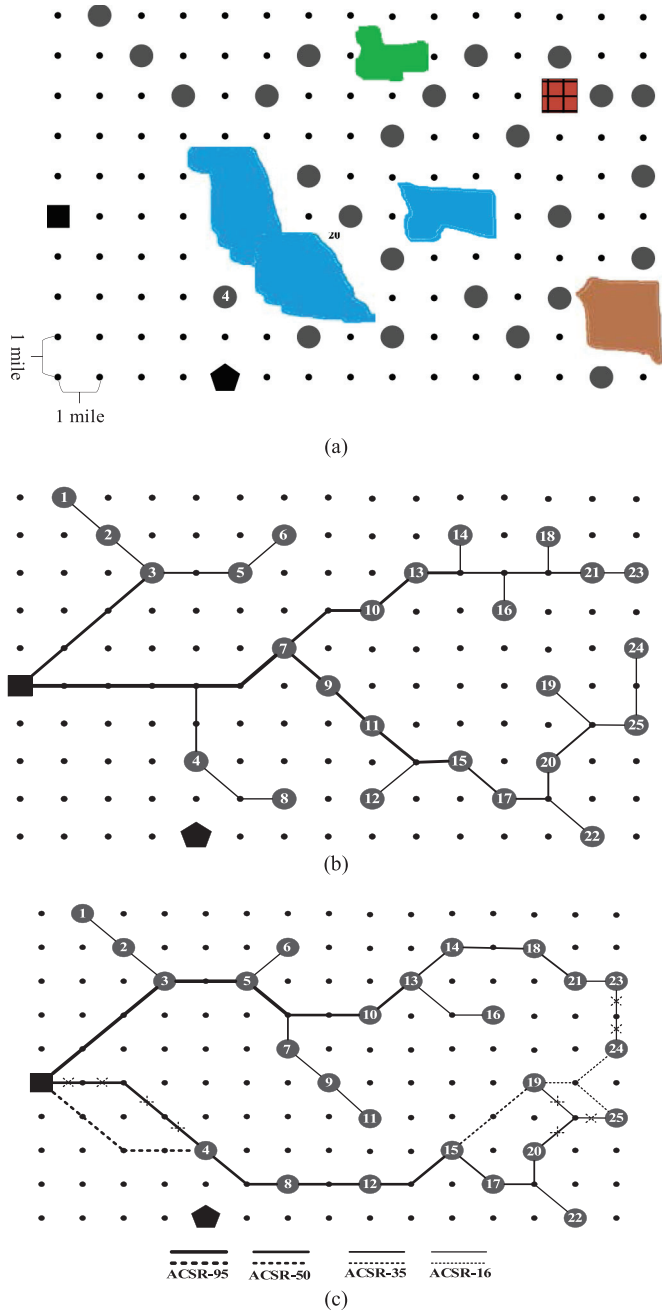


Fig. 9. Realistic test system: (a) GIS map, (b) Case 1, (c) Case 2 (solid lines)/ Case 3 (dashed lines).

10 \$/kWh, respectively. As illustrated in Fig. 9(c), the optimal planning configuration is different from that in case 2. To enhance the network resiliency, feeders marked with dashed lines are installed instead of feeders marked with crosses. The optimal objective function value is \$2.049M, from which \$1.524M is the investment cost and \$0.525M is the cost of resiliency. The proposed algorithm chooses to install feeders that are closer to the RMC node to take into account the resiliency target. This changes the system configuration and increases the investment cost. Figure 10 shows the expected voltages and the voltage ranges for the ten final scenarios.

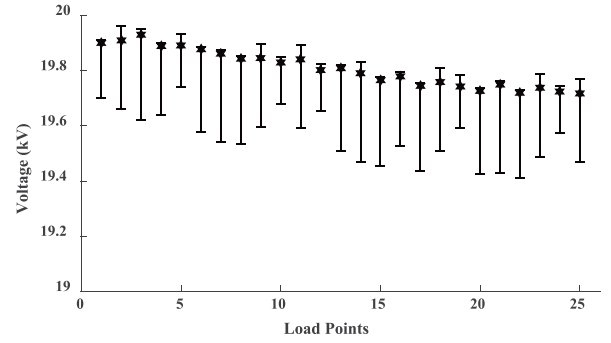


Fig. 10. Expected voltages with the tolerance of ten final scenarios (Case 3).

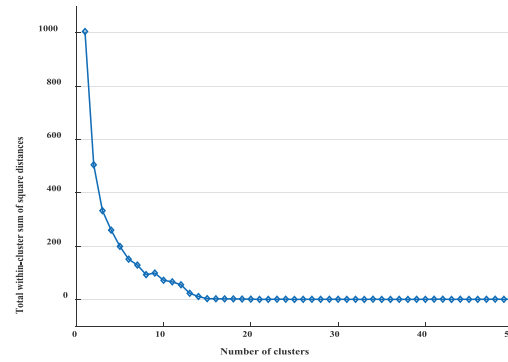


Fig. 11. Total within-cluster sum of squared distances for different numbers of clusters of K-means++.

C. Large Test System

A large test system is constructed and used to investigate the scalability of the proposed solution algorithm. Thirty demand points are randomly located on a GIS map including $64 \times 64 = 4096$ GIS nodes. Considering the typical electric pole distance in rural areas of 300 feet (distance between two adjacent GIS nodes), this test system covers an area of 13.23 square miles. Hence, the system is sufficiently large to be considered as a real-size system. We assume that each load point includes a solar photovoltaic system with a capacity (kWp) equal to 10% of the annual peak load. We consider the same data used in the previous test systems. The linearized economic characteristics of components and the procedure for finding the final clusters of scenarios discussed in case 4 of the small test system are used for this case study. Given 10,000 initial scenarios, the K-means++ method is applied to cluster these scenarios [28]. Considering the total within-cluster sum of squared distances in the elbow method [29], the final number of clusters (K value) is set to 14. The total within-cluster sum of squared distances in the elbow method measures the compactness of the clustering, and it is desired to be as small as possible. The total within-cluster sum of squared distances with respect to different K values are depicted in Fig. 11. It can be observed that for K values greater than 14, the total within-cluster sum of squared distances does not change significantly.

Considering 14 scenarios, the proposed stochastic MISOC P GIS-based FR with a linear objective function is solved. The

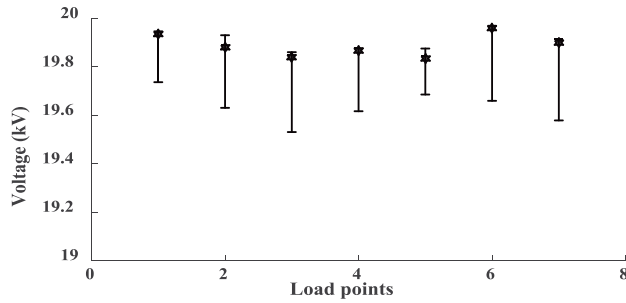


Fig. 8. Expected voltages with the tolerance of ten final scenarios (Case 4).

loss of generality, the elevation of all GIS nodes is assumed to be the same. Four cases are considered.

Case 1: FR model without GIS nodes. Ignoring GIS nodes, only eight electrical nodes remain, seven load points plus the substation node. We perform a complete enumeration to find the global optimal solution of this FR problem. The simulation time for performing such complete enumeration is high and not comparable with the solution time required by standard solvers. However, carrying out a complete enumeration ensures the global optimality of the solution. The total number of possible trees (radial configurations) on $|\mathcal{V}|$ vertices is calculated by Cayley's formula as $|\mathcal{V}|^{(|\mathcal{V}|-2)}$, which result in $8^6 = 262,144$ possible radial configurations. Given the system structure, the FR problem is simplified to a minimum flow problem with a quadratic cost for each possible configuration. The optimal planning configuration shown in Fig. 7(b) is obtained with an objective function of \$0.4275M.

Case 2: The proposed MISOCP model is solved using Gurobi. The objective function is not linearized in this case. The optimum planning configuration is shown in Fig. 7(c), and the objective function is \$0.4114M.

Case 3: The proposed MISOCP model is solved. Economic characteristics for each conductor are linearized into five equal segments. The optimum planning configuration is the same as that of case 2.

Case 4: The stochastic MISOCP model with linearized objective function is solved. The power law distribution given by (39) is used for probabilistic net load forecasting.

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \quad \text{for } x \geq x_{\min} \quad (39)$$

MLE and Kolmogorov-Smirnov tests are used for parameter estimation ($x_{\min} = 1564$, $\alpha = 37.11$). After obtaining the distribution function, 10,000 scenarios are generated. The K-means++ method is applied to cluster these scenarios into ten final scenarios [28]. The optimal planning configuration is the same as that in case 3. However, the objective function increases to \$0.4201M due to the presence of uncertainties. The expected voltages of load nodes for the ten final scenarios are shown in Fig. 8. All voltages are within the permissible limit.

1) *Comparison of Results:* The objective functions of all four cases and the total simulation times are presented in Table I. The

TABLE I
SMALL-SCALE TEST SYSTEM RESULTS

Case No.	Optimal objective function (M\$)	Simulation time (min)
Case 1: FR without GIS nodes (enumeration)	0.4275	402.5
Case 2: MISOCP GIS-based FR (quadratic obj.)	0.4114	55.3
Case 3: MISOCP GIS-based FR (linear obj.)	0.4114	4.1
Case 4: Stochastic MISOCP GIS-based FR (linear obj.)	0.4201	10.1

TABLE II
LARGE-SCALE TEST SYSTEM RESULTS

Stochastic MIP models	Optimal objective function (M\$)	Simulation time (min)
Case 1: GIS-based FR without obstacles	1.499	32
Case 2: GIS-based FR with obstacles	1.515	10.1
Case 3: Resilient GIS-based FR with obstacles	1.524+0.525=2.049	9.5

objective function of case 2 is smaller than that of case 1. That is, considering GIS nodes reduces the FR cost. Furthermore, linearizing the cost functions improves the solution time significantly, which apparent comparing cases 2 and 3. Modeling load forecasting errors and uncertainty of rooftop solar generations in case 4 leads to a higher variable cost (cost of losses), while the investment cost remains unchanged due to the small size of the system.

B. Realistic Test System

The realistic case study presented in [6] is considered here to illustrate the effect of modeling resiliency and the presence of geographical obstacles in the FR problem. The forecast annual peak demand for each load node is given in Fig. 9(a). We assume that each load point includes a solar photovoltaic system with a size (kWp) of 10% of the annual peak load and with the same data used in the previous test system. Three cases are studied considering the same linearized economic characteristics and the same procedure for finding the final clusters of scenarios discussed in case 4 of the small test system. According to simulation results, to keep the system reliable, the minimum number of feeders starting from the substation node needs to be two (i.e., $F = 2$) for all three cases.

While the presence of obstacles is ignored is case 1, these obstacles are considered in case 2. The optimal objective function and total simulation times are provided in Table II. The optimal objective function values for cases 1 and 2 are \$1.499M and \$1.515M, respectively, and the optimal planning configurations are shown in Figs. 9(b) and (c) with solid lines. The investment cost in the presence of obstacles is higher because the geographical alternatives are more restricted. However, the simulation time of case 2 is smaller since this case includes 25% fewer binary variables as compared to case 1.

In case 3, the cost of resiliency and the presence of obstacles are considered. The values of average repair time, average speed to repair, and $VoLL$ are 5 hours, 20 miles per hours, and

and the null hypothesis is rejected [24]. We apply the approach presented in [23], and after estimating the power law distribution for each demand, a large enough number of scenarios is generated. Then, only a limited number of non-redundant scenarios are considered. A multivariate K-means algorithm is applied to reduce the number of scenarios generated into multiple groups. This reduces the computational burden of the problem while keeping an acceptable level of accuracy. K-means is a model-free method for data clustering and partitioning. K-means is compared with alternative techniques for scenario reduction in [25]. Additionally, [26] and [27] use K-means to reduce scenarios in transmission expansion planning problems. Here, to alleviate the dependency of the K-means algorithm to the initial centroids, we apply an initialization approach, called K-means++, that is proven to work better than K-means [28]. To find the optimal number of clusters (K value), different approaches are presented in the literature, such as the elbow method [29], the Silhouette method [30], and the gap statistic method [31]. We have used the elbow method for numerical studies.

Step 5 (solving MISOCP model): The proposed stochastic MISOCP model is solved: minimize (38) subject to (15), (16), (19)–(22), (24)–(32), (34), and (35).

Steps 6–7 (reliability evaluation): Using the obtained FR solution, reliability indices are calculated. If the indices are not good enough, the reliability violation is corrected by increasing the minimum degree of the substation node (i.e., parameter F).

B. On Computational Complexity

One might be concerned with the fact that the proposed model is impractical as it includes a large number of binary variables. We provide the following discussion on the computational complexity of the proposed method to address this concern:

- The GIS-based FR problem is NP-hard. That is, there is no guarantee that it can be solved within a reasonable solution time if the size of the problem grows up significantly. However, as illustrated in the Numerical Results section, the solution time is acceptable for real-world systems.
- Parallel processing might be a practical solution for intractable instances.
- For realistic cases, a considerable number of binary variables are known in advance due to existing edges and geographical barriers. This is particularly so in urban areas.
- Some parameters of the model can be adjusted to provide an acceptable tradeoff between solution time and accuracy level, e.g., the optimality gap of the solver, the number of segments of the piecewise linear approximations, and the number of GIS points (GIS resolution). These parameters can be set to appropriate values to reduce the solution time for large-scale problems.
- One additional criterion that could be considered for determining the GIS resolution is the distance between poles. It does not generally make sense to have a resolution in which the distance between GIS nodes is less than the standard distance between two neighboring poles.

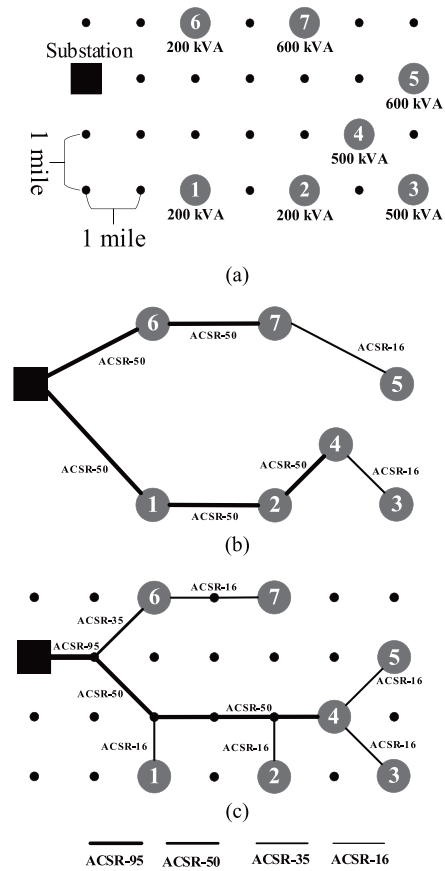


Fig. 7. Small-scale test system: (a) GIS map, (b) Case 1, and (c) Case 2–4.

VI. NUMERICAL RESULTS

The proposed algorithm is tested on three case studies: a small test system, a realistic one, and a synthetic large test system. A nominal voltage of $V = 20$ kV is assumed, and the permissible voltage range is 0.95 to 1.05 per unit. Four types of candidate conductors are selected with economic characteristics shown in Fig. 4. Admittances of the conductors, ACSR-16, ACSR-35, ACSR-50, and ACSR-95, are $0.317-j0.065$, $0.63-j0.268$, $0.81-j0.486$, and $0.989-j1.038$ siemens, respectively. All problems are solved using Gurobi 8.1 under GAMS [32]. A computer with an Intel(R) Xeon(R) CPU @2.6 GHz, including eight cores and 16 GB of RAM, is used for simulations.

A. Small Test System

This test system is used to illustrate the impact of including GIS nodes and to evaluate the presence of uncertainties in the FR problem. The test system is shown in Fig. 7(a). The annual peak load of each load point is depicted in the figure, and the cost of resiliency is ignored [6]. It is assumed that each load point includes a solar photovoltaic system with a capacity (kWp) of 10% of the annual peak load. The data of solar generations are obtained from the Ausgrid study on 300 solar homes from July 2010 to June 2013 [33]. For the sake of simplicity and without

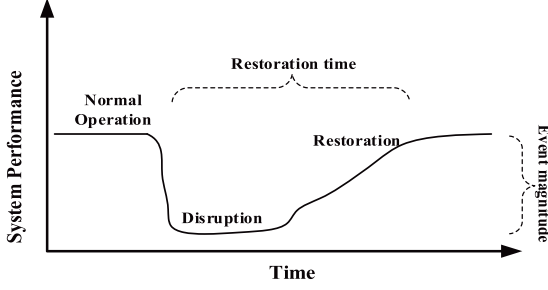


Fig. 5. Illustration of the resiliency concept used in this paper.

a disruption event is provided in Fig. 5. The grid resiliency can be improved by reducing the magnitude of the event and the restoration time.

We introduce below an economic resiliency metric that can be included in the proposed GIS-based FR model. Our resiliency metric directly relies on the resiliency concept as shown in Fig. 5. That is, no specific disruption is considered, but its consequences. Note that it is common to add costs pertaining to future disruptions (e.g., a penalty cost for violating the $N - 1$ security criterion and a cost for lack of resiliency) to the objective function of a planning problem to ensure obtaining reliable network proposals [20]. The consistency of the exact geographical locations of power system components and the GIS nodes is taken into account in the metric definition. Knowing which GIS nodes correspond to the geographical location of the repair and maintenance centers (RMCs), the restoration time for edge (i, j) is given by (36). The restoration time depends on the distance of edge (i, j) to the closest RMC, the average speed of the repair team, the required time to repair edge (i, j) and availability of nearby components. The total resiliency cost of the system is calculated by (37). The magnitude of the outage is modeled by the MW flow carried by the conductor installed ($P_{ij,c,\gamma}^s$). Note that the restoration time for each edge is considered known, and that the value of lost load ($VoLL$), in dollars per MWh, is assumed to be known [21]. The constant coefficient $VoLL$, i.e., the cost of unserved energy, is used to express the resiliency cost in dollars (i.e., $\tau_{ij}(h) \times VoLL(\frac{\$}{MWh}) \times P_{ij,c,\gamma}^s(MW) \rightarrow C_{res}(\$)$).

$$\tau_{ij} = \frac{\text{Distance from the closest RMC}}{\text{Average speed of repair team}} + \text{Repair time of edge } (ij) \quad (36)$$

$$C_{res} = \sum_{(i,j,c) \in \mathcal{E}} \sum_{s \in S} \sum_{\forall \gamma} \rho_s \cdot \tau_{ij} \cdot VoLL \cdot P_{ij,c,\gamma}^s \quad (37)$$

To minimize the resiliency cost of the planned network, (37) is added to the objective function of the FR problem. The objective function of the resilient GIS-based FR problem is given by (38).

$$\min \sum_{(i,j,c) \in \mathcal{E}} \{d_{ij} \cdot a_{ij,c} \cdot I_{ij,c} + \sum_{s \in S} \sum_{\forall \gamma} \rho_s \cdot (d_{ij} \cdot m_{ij,c,\gamma} + \tau_{ij} \cdot VoLL) \cdot P_{ij,c,\gamma}^s\} \quad (38)$$

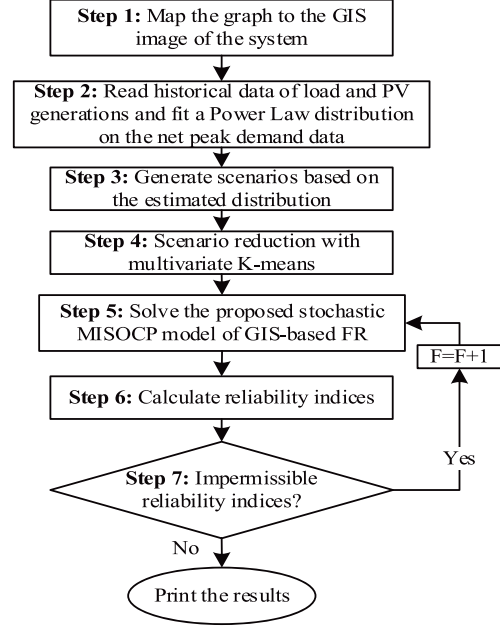


Fig. 6. Flowchart of the solution algorithm.

Finally, to solve the resilient GIS-based FR, objective function (38) is minimized subject to (15), (16), (19)–(22), (24)–(32), (34), and (35).

V. SOLUTION ALGORITHM

A. Solution Steps

The flowchart of the proposed solution algorithm for the GIS-based FR problem is shown in Fig. 6.

Step 1: The first step is to map the graph to the system's GIS image with the desired resolution level, i.e., number of points in the representing graph. Geographical mapping and map rasterizing techniques have been previously studied using sophisticated techniques [7] and [22]. However, for the sake of simplicity and without loss of generality, the representing graph of Fig. 3 is mapped to the GIS image so that most electrical nodes are covered by GIS nodes. Each electrical node that is not covered by a GIS node is associated with its closest GIS node.

Steps 2–4 (uncertainty modeling): Historical data of load and PV generation for each load point is obtained. A probabilistic load forecasting technique is applied to find a pdf that is the best fit for the net load. As experimentally shown in [23], the daily peak load of distribution feeders follows a power law distribution. Parameters of the power law distribution are obtained using a maximum likelihood estimator (MLE) and a Kolmogorov-Smirnov test, while the uncertainty of the model is quantified by a bootstrapping method. Having the null hypothesis of the test as “the observed data follow the power law distribution”, the p -value is greater than the significance level of the test. This means that the null hypothesis cannot be rejected. However, for other distributions, e.g., exponential, gamma, and lognormal, the p -value will be smaller than the significance level,

$$\text{s.t. } \sum_{j|(j,i,c) \in \mathcal{E}} P_{ji,c}^s - \sum_{j|(i,j,c) \in \mathcal{E}} P_{ij,c}^s = PD_i^s; \quad i \in \mathcal{V}, \quad s \in S \quad (15)$$

$$\sum_{j|(j,i,c) \in \mathcal{E}} Q_{ji,c}^s - \sum_{j|(i,j,c) \in \mathcal{E}} Q_{ij,c}^s = QD_i^s; \quad i \in \mathcal{V}, \quad s \in S \quad (16)$$

$$P_{ij,c}^s = \left(\sqrt{2} G_{ij,c} u_i^s - G_{ij,c} R_{ij,c}^s + B_{ij,c} L_{ij,c}^s \right) \cdot I_{ij,c}; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (17)$$

$$Q_{ij,c}^s = \left(\sqrt{2} B_{ij,c} u_i^s - B_{ij,c} R_{ij,c}^s - G_{ij,c} L_{ij,c}^s \right) \cdot I_{ij,c}; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (18)$$

$$2 u_i^s u_j^s \geq (R_{ij,c}^s)^2 + (L_{ij,c}^s)^2; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (19)$$

$$R_{ij,c}^s = R_{ji,c}^s \geq 0; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (20)$$

$$L_{ij,c}^s = -L_{ji,c}^s; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (21)$$

$$\frac{(V_i^{\min})^2}{\sqrt{2}} \leq u_i^s \leq \frac{(V_i^{\max})^2}{\sqrt{2}}; \quad i \in \mathcal{V}, \quad s \in S \quad (22)$$

$$0 \leq P_{ij,c}^s \leq P_c^{\max} \cdot I_{ij,c}; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (23)$$

$$(P_{ij,c}^s)^2 + (Q_{ij,c}^s)^2 \leq (S_{ij,c}^{\max})^2 \cdot I_{ij,c}; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (24)$$

$$\sum_{\forall c} I_{ij,c} \leq 1; \quad \forall (i, j) \in \mathcal{E} \quad (25)$$

$$\sum_{(i,j,c) \in \mathcal{E}} I_{ij,c} \geq F; \quad i = \text{Substation vertex} \quad (26)$$

$$I_{ij,c} \in \{0, 1\}; \quad (i, j, c) \in \mathcal{E} \quad (27)$$

$$\sum_{i \in \mathcal{V}} PD_i^s \leq 0; \quad s \in S \quad (28)$$

It should be noted that the constraints of the proposed model must be satisfied for all possible realizations. The nodal power balance for each GIS node (or vertex) is enforced by (15) and (16). Parameter PD_i^s / QD_i^s is the forecast peak value of the net active/reactive power demand connected to vertex i in scenario s . Conic expressions of the power flow constraints are given by (17)–(19). The constraints of conic variables must be satisfied as shown in (20) and (21) [9]. Permissible voltage constraints, maximum thermal limit of conductors, and capacity limit for each conductor are given by (22), (23) and (24), respectively. Constraint (25) enforces that only one conductor from a set of parallel conductors can be selected. The minimum number of feeders starting from the substation node is restricted by (26). Since the reliability of the planned distribution network increases by increasing the number of feeders leaving the substation node [18], parameter F is used to achieve the required reliability level. Parameter F is initialized to one and then iteratively increased to reach the required reliability level. Since the typical number of feeders leaving a substation node (the actual value of parameter F) is limited, e.g., less than eight as presented in [15], few

iterations would be enough to reach the required reliability level. If it is not possible to reach the required reliability level, the planner might need to add new substations. Definitions of binary variables are stated in (27). The value of PD_i^s is positive for load points, and negative for the substation node as indicated by (28).

The product of binary and continuous variables in (17) and (18) make the model nonlinear. We replace (17) and (18) with their equivalent linear constraints (29)–(32) using a disjunctive technique and Big-M values ($M_{ij,c}$):

$$P_{ij,c}^s - \left(\sqrt{2} G_{ij,c} u_i^s - G_{ij,c} R_{ij,c}^s + B_{ij,c} L_{ij,c}^s \right) + M_{ij,c} \cdot (1 - I_{ij,c}) \geq 0; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (29)$$

$$P_{ij,c}^s - \left(\sqrt{2} G_{ij,c} u_i^s - G_{ij,c} R_{ij,c}^s + B_{ij,c} L_{ij,c}^s \right) - M_{ij,c} \cdot (1 - I_{ij,c}) \leq 0; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (30)$$

$$Q_{ij,c}^s - \left(\sqrt{2} B_{ij,c} u_i^s - B_{ij,c} R_{ij,c}^s - G_{ij,c} L_{ij,c}^s \right) + M_{ij,c} \cdot (1 - I_{ij,c}) \geq 0; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (31)$$

$$Q_{ij,c}^s - \left(\sqrt{2} B_{ij,c} u_i^s - B_{ij,c} R_{ij,c}^s - G_{ij,c} L_{ij,c}^s \right) - M_{ij,c} \cdot (1 - I_{ij,c}) \leq 0; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (32)$$

The proposed model includes a large number of binary variables, which is equal to the number of edges ($|\mathcal{E}|$) of the representing graph. Solving models including a large number of binary variables using standard solvers might be time consuming or infeasible. Therefore, deriving a more tractable version of the GIS-based FR problem is desirable. With this purpose, we apply a piecewise linear approximation of the economic cost curves of conductors using an equal segmentation technique. Other linearization techniques are also possible [19]. The proposed linearized objective function is:

$$\min \sum_{(i,j,c) \in \mathcal{E}} \{d_{ij} \cdot a_{ij,c} \cdot I_{ij,c} + \sum_{s \in S} \sum_{\forall \gamma} \rho_s \cdot d_{ij} \cdot m_{ij,c,\gamma} \cdot P_{ij,c,\gamma}^s\} \quad (33)$$

$$P_{ij,c}^s = \sum_{\forall \gamma} P_{ij,c,\gamma}^s; \quad (i, j, c) \in \mathcal{E}, \quad s \in S \quad (34)$$

$$0 \leq P_{ij,c,\gamma}^s \leq P_{c,\gamma}^{\max} \cdot I_{ij,c}; \quad (i, j, c) \in \mathcal{E}, \quad s \in S, \quad \forall \gamma \quad (35)$$

where $m_{ij,c,\gamma}$ is the slope of segment γ of the cost curve, and $P_{ij,c,\gamma}^s$ is a continuous variable for active power flow in edge (i, j, c) corresponding to segment γ in scenario s . We note that (34) is added to the model, and (23) is replaced by (35).

D. Resiliency Modelling for GIS-Based FR

The main objective of grid resiliency is to reduce the magnitude of undesirable events and alleviate its consequences that result in disruptions. These consequences are closely related to power delivery and grid operation. Therefore, most of the performance metrics defined for resiliency modeling are consequence-based metrics [4]. A simple illustration of the three stages (i.e., normal, disruption, and restoration) of system operation during

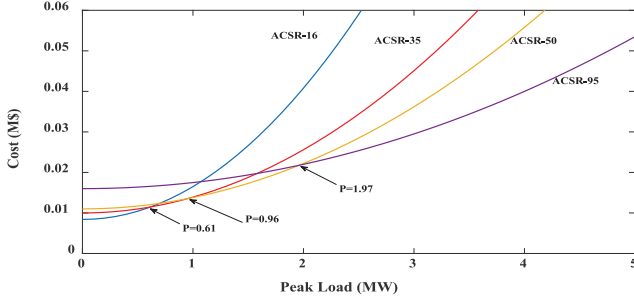


Fig. 4. Economic characteristics of four conductors.

$$\omega_c = a_1 + b \cdot P^2 \quad (\$/\text{mile}) \quad (7)$$

The total cost for each conductor in terms of peak load is a quadratic function. As shown in Fig. 4, four quadratic curves, corresponding to different conductors, intersect with each other. That is, for a specific peak load, there is only one type of conductor that results in minimum cost [15]. In other words, the economic loading limit of a conductor, which is always less than its thermal loading limit, is at the intersecting point of its curve and the curve of the next larger conductor. For instance, ACSR-16 is the optimal conductor size for peak loads less than 0.61 MW, and ACSR-95 is the optimal conductor size for peak loads higher than 1.97 MW.

B. SOCP Model for Power Flow

Considering two adjacent nodes i and j connected through one conductor, the active and reactive power flow through the conductor can be calculated by (8) and (9).

$$P_{ij} = G_{ij}V_i^2 - G_{ij}V_iV_j \cos \theta_{ij} + B_{ij}V_iV_j \sin \theta_{ij} \quad (8)$$

$$Q_{ij} = B_{ij}V_i^2 - B_{ij}V_iV_j \cos \theta_{ij} - G_{ij}V_iV_j \sin \theta_{ij} \quad (9)$$

where V_i and θ_i are voltage magnitude and angle at node i , $Y_{ij} = G_{ij} - jB_{ij}$ is series admittance of conductor ij , and $\theta_{ij} = \theta_i - \theta_j$. The sufficient conditions under which the SOCP relaxation is exact for branch flow models of radial distribution systems are discussed in [16]. Since these conditions hold for most practical power distribution systems, (8) and (9) can be expressed as conic constraints (10)–(12) [9]. By defining $u_i = V_i^2/\sqrt{2}$, $R_{ij} = V_i V_j \cos \theta_{ij}$, and $L_{ij} = V_i V_j \sin \theta_{ij}$, (8) and (9) are rewritten as (10)–(12).

$$P_{ij} = \sqrt{2} G_{ij}u_i - G_{ij}R_{ij} + B_{ij}L_{ij} \quad (10)$$

$$Q_{ij} = \sqrt{2} B_{ij}u_i - B_{ij}R_{ij} - G_{ij}L_{ij} \quad (11)$$

$$2u_iu_j \geq (R_{ij})^2 + (L_{ij})^2 \quad (12)$$

Note that the linear constraints (10) and (11) along with the rotated quadratic cone (12) are a set of convex constraints, which are computationally efficient.

Incorporating the representing graph of Fig. 3 in the FR model ensures radiality of the network. Hence, there is no need to add any radiality constraints in the model and therefore SOCP is an exact relaxation of the full AC power flow.

C. Stochastic MISOCP Model for GIS-Based FR

Having economic curves of different conductors and the proposed graph shown in Fig. 3, the stochastic MISOCP model of the GIS-based FR problem is described below. The objective function is modeled as (13).

$$\min \sum_{(i,j,c) \in \mathcal{E}} \{d_{ij} \cdot a_{ij,c} \cdot I_{ij,c} + d_{ij} \cdot b_{ij,c} \cdot (P_{ij,c}^s)^2\} \quad (13)$$

where indices i and j represent GIS nodes. Individual conductors within a set of parallel conductors are indicated by c . We note that an edge (i, j, c) belongs to the set of all possible edges $\mathcal{E}$. Parameter d_{ij} is the distance between GIS nodes i and j . Note that the distances between all pairs of adjacent GIS nodes are not the same since the geographical elevations of GIS nodes are different. Parameter $b_{ij,c}$ is the variable cost of edge (i, j, c) obtained by (6). Parameter $a_{ij,c}$ is the fixed cost of edge (i, j, c) calculated by (2) plus additional penalty (positive) or incentive (negative) costs (i.e., $a = a_1 + a_2 + a_3 + \dots$). These penalty/incentive costs are defined according to the environmental condition of each edge. For instance, an edge that is close to a road should be incentivized since it is comparatively more accessible for operation and maintenance. On the other hand, an edge close to a commonly flooded area should be penalized since it deteriorates the reliability of the FR solution [8]. For edges that are geographically impossible for installing a line, the parameter $a_{ij,c}$ should be set to a large enough value. For all other edges, both $a_{ij,c}$ and $b_{ij,c}$ should be normalized to improve the efficiency of the solution algorithm. Continuous decision variable $P_{ij,c}$ is the flow of the line installed at edge (i, j, c) . Binary variable $I_{ij,c}$ represents the installation status of a candidate conductor at edge (i, j, c) . It is equal to one if the conductor is installed, and zero otherwise. In addition, for upgrading an existing system, variables $I_{ij,c}$ corresponding to existing feeders must be set to one.

FR is a long-term expansion planning problem that needs long-term load and solar generation forecasts. A stochastic framework is necessary to cope with deviations from forecast values. There are two types of decisions in the FR problem: *here and now* decisions ($I_{ij,c}$), and *wait and see* decisions ($P_{ij,c}$) that depends on realization of the stochastic parameters. A two-stage stochastic optimization model is applied in this situation. We combine load and solar generation to derive the net load (PD_i) and transform the two-stage stochastic optimization model into the equivalent deterministic formulation presented in (14)–(32) (see [17] for more details). The second term of the objective function is replaced by the expected value of the variable cost, which is the sum over all possible net load realizations. A finite number of scenarios (s) with associated weights (ρ_s) is considered. Note that the summation of weights over all scenarios must be equal to one (i.e., $\sum_{s \in S} \rho_s = 1$).

$$\min \sum_{(i,j,c) \in \mathcal{E}} \{d_{ij} \cdot a_{ij,c} \cdot I_{ij,c} + E_s [Q(P_{ij,c}, s)]\} \quad (14)$$

$$E_s [Q(P_{ij,c}, s)] = \sum_{(i,j,c) \in \mathcal{E}} \sum_{s \in S} \rho_s \cdot d_{ij} \cdot b_{ij,c} \cdot (P_{ij,c}^s)^2$$

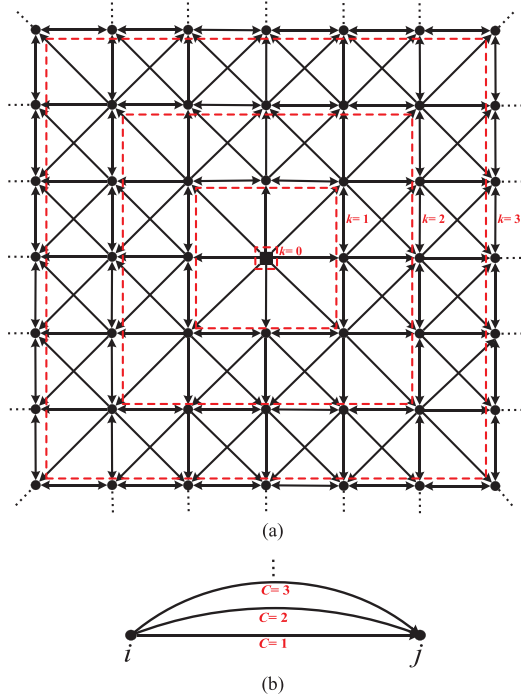


Fig. 3. Representing graph for the GIS-based FR problem: (a) network graph and (b) possible edges (different conductor size) between two adjacent nodes.

time algorithm to find the solution of the problem. Since the dependency of the cost of each edge on its energy loss increases the computational complexity of the problem, it holds that the GIS-based FR problem is NP-hard. ■

Because of the NP-hard nature of the GIS-based FR problem and the recent progress in integer programming, we propose a mixed-integer second-order cone programming (MISOCP) model for the GIS-based FR problem. The proposed MISOCP model accurately represents power system constraints and the accuracy of the solution can be measured by the duality gap.

Dynamic programming is not a good option for the GIS-based FR problem due to two main reasons. First, the principle of optimality does not hold for the GIS-based FR (Steiner tree) problem [11]. Therefore, the solution obtained from dynamic programming for the GIS-based FR problem is just an approximation. Second, as presented in [6], the accuracy of this approximation cannot be characterized.

To construct a graph-based model for the GIS-based FR problem, we propose a representing graph in the Euclidean plane as shown in Fig. 3(a). Although the representing graph of Fig. 3(a) is mapped on a two-dimensional Euclidean plane, the elevation of the vertices is considered in the mathematical model. Considering this representing graph, the continuous geographical space can be discretized to nodes (vertices of the representing graph) and lines (edges of the representing graph) as shown in Fig. 3(a). This is useful since graph-based models can be constructed for the FR problem after discretization. The substation node is located in the center of the graph and is named point zero ($k = 0$) (to avoid any confusion with stochastic optimization, we have used the term “point zero” instead of “stage zero”). Dashed red lines distinguish other points ($k = 1, 2, \dots$). The specific property of this directed graph is that there is no path

from farther points to closer ones. Having this feature and knowing that the optimal solution is a subgraph of the initial graph, radially constraint of the solution is guaranteed. This means that there will be no cycle in the final solution. Each edge in the representing graph includes parallel candidate paths, each of which represents a candidate conductor size as depicted in Fig. 3(b). This is an important feature for modeling the FR problem in the context of MISOCP.

IV. MATHEMATICAL MISOCP FORMULATION

A. Cost of a Feeder

The total cost of a conductor is divided into three components: initial installation cost, annual operation and maintenance cost, and losses. The initial installation cost c_i is a constant cost to be paid at the time of building the feeder. This cost differs for different conductor sizes [12]. Annual operation and maintenance cost $c_{O\&M}$ is also higher for larger conductors. The present worth of this cost with a discount rate d over a period of n years is calculated by the present worth factor ω_1 given in (1). Thus, the fixed part of the cost per unit length (\$/mile) is modeled as (2).

$$\omega_1 = \frac{(1+d)^n - 1}{d(1+d)^n} \quad (1)$$

$$a_1 = c_i + (\omega_1 \times c_{O\&M}) \quad (2)$$

The variable cost depends on energy losses, which are a function of the load. This cost increases yearly due to load growth. For a specific demand for which the peak load in the first year is P MW, the power factor PF , and the line-to-line voltage V kV, the required current is calculated by (3). Hence, for a three-phase line with a resistance of r ohm and an annual loss factor L_f , the total energy loss per mile for the first year is calculated by (4). Note that having a fixed power factor, the annual loss factor, which is the ratio of average loss to actual loss at the peak load, can be obtained directly from the conductor resistance and load profile [13].

$$I = \frac{10^3 \times P}{\sqrt{3} \times V \cdot PF} \quad (A) \quad (3)$$

$$\ell = 3 \times r \cdot I^2 = 8760 \times \frac{10^3 \times L_f \cdot r \cdot P^2}{(V \cdot PF)^2} \quad (\text{kWh/mile}) \quad (4)$$

To obtain the cost of energy loss in \$/mile for the first year, (4) is multiplied by the cost of energy e_c (\$/kWh). Considering an annual increasing rate σ for the peak load, the energy loss will increase at a rate of $j = \sigma^2 + 2\sigma$. Therefore, to derive the present worth of losses with a discount rate of d over a period of n years, the value of energy loss for the first year is multiplied by a present worth factor ω_2 [14]. Finally, the total present worth cost per mile for a specific conductor is calculated by (1)–(7) considering installation cost and the present worth of operation, maintenance, and losses.

$$\omega_2 = \frac{1 - \left(\frac{1+j}{1+d}\right)^n}{d - j} \quad (5)$$

$$b = 8760 \times \frac{10^3 \times L_f \cdot r \cdot e_c \cdot \omega_2}{(V \cdot PF)^2} \quad (6)$$

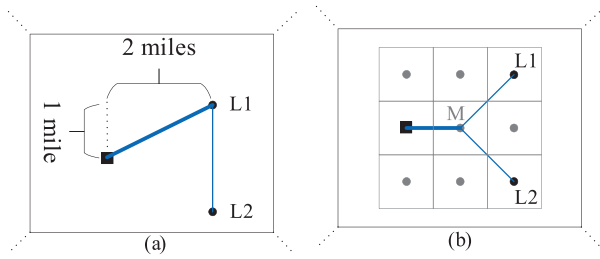


Fig. 1. A small segment of a distribution system: illustration of lateral nodes; (a) without GIS nodes, (b) with GIS nodes.

three electrical nodes. Assume that there is a load at each electrical node identified by a circle, and that a substation is located at the electrical node identified by a square. Note that no other electrical nodes exist in this small segment of the distribution system. Hence, the solution of the FR problem is a connection between the load points L1 and L2 and the substation node. In the case with no GIS nodes (the space is discretized in a way that only electrical nodes exist), the optimal solution is shown in Fig. 1(a). However, with GIS nodes (the space is discretized in a way that several GIS nodes identified by gray dots exist), the optimal solution is obtained if the lateral node “M” is located at a GIS node as depicted in Fig. 1(b). Note that GIS nodes are assumed to be at the center of GIS pixels, as shown in Fig. 1(b). Assuming that the solution cost depends only on the lengths of feeders, it can be concluded that the solution cost in case (b) with GIS nodes ($1 + 2\sqrt{2} \approx 3.83$) is better than that in case (a) without GIS nodes ($2 + \sqrt{5} \approx 4.24$). Thus, it can be concluded that discretizing the continuous space without having GIS-nodes reduces the quality of the solution.

- Multiple right-of-ways between two adjacent electrical nodes can be defined if GIS nodes are in place. That is, there are multiple alternatives to connect two adjacent electrical nodes, each of these alternatives having a specific cost related to its geographical situation (see Fig. 1). However, if GIS nodes are not in place, there is only one candidate path (right-of-way) between two adjacent electrical nodes, and the feasible design space of the optimization problem is reduced. By using GIS information, geographical obstacles, high-cost routes, and land ownership issues can be avoided. Moreover, having alternative paths to connect two electrical nodes is more realistic and gives more flexibility to optimization.
- If GIS nodes are not considered, the cost of FR can be as high as twice the GIS-based FR optimal cost. This is shown in Proposition 1.

Definition: Given an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a cost function for each edge, and a partition of $\mathcal{V}$ into two sets $\mathcal{O}$ and $\mathcal{S}$, the problem of finding a minimum cost tree that contains all vertices in $\mathcal{O}$ and any subset of the vertices in $\mathcal{S}$ is called Steiner tree problem. If the set of vertices $\mathcal{S}$ is empty, and $\mathcal{O} = \mathcal{V}$, it is called minimum spanning tree problem.

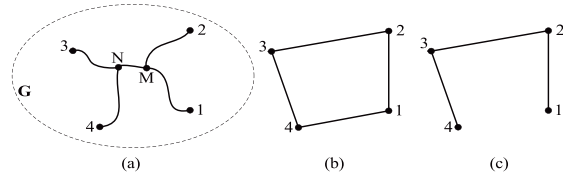


Fig. 2. Illustrative proof for optimality of the GIS-based FR solution.

Proposition 1: If the cost of each feeder depends only on its length, the FR solution without GIS nodes is within a factor of 2 from the GIS-based FR optimal solution.

Proof: If radiality constraints are enforced, the GIS-based FR problem is equivalent to the Steiner tree problem [10], while the FR problem without GIS nodes is equivalent to the minimum spanning tree problem of the corresponding Steiner tree problem. Therefore, to prove proposition 1, it is sufficient to show that the cost of the minimum spanning tree problem is at most twice the cost of its corresponding Steiner tree problem. To do so, we proceed as follows.

Let us consider the graph in Fig. 2(a) as the optimal Steiner tree (GIS-based FR) solution, where $\mathcal{O} = \{1, 2, 3, 4\}$, and $\mathcal{S} = \{M, N\}$ is the set of GIS nodes. If we double each edge of this graph, we get an Eulerian graph. An Eulerian graph is a connected graph including even degree vertices. Using Eulerian graph properties, we follow that there exists an Eulerian tour whose cost is twice the optimum Steiner tree cost. The reason is that in the Eulerian tour ($1 \rightarrow M \rightarrow 2 \rightarrow M \rightarrow N \rightarrow 3 \rightarrow N \rightarrow 4 \rightarrow N \rightarrow M \rightarrow 1$), each edge of the graph shown in Fig. 2(a) (the optimum Steiner tree) is passed twice. The Eulerian tour, which includes the vertices in $\mathcal{O}$ and $\mathcal{S}$, is a Hamiltonian cycle on $\mathcal{O}$. A Hamiltonian cycle is a closed-loop on a graph where every node is visited exactly once. The corresponding Hamiltonian cycle is redrawn in Fig. 2(b) in a way that the cost of path ($1 \rightarrow 2$) in Fig. 2(b) is equal to the cost of path ($1 \rightarrow M \rightarrow 2$) in Fig. 2(a). Removing one optional edge from the Hamiltonian cycle as shown in Fig. 2(b) results in a minimum spanning tree of $\mathcal{O}$ as shown in Fig. 2(c). Since one edge is removed, the cost of the minimum spanning tree is at most twice the optimal Steiner tree cost. Thus, the cost of a minimum spanning tree on $\mathcal{O}$ (i.e., FR solution without GIS nodes) is between once and twice the cost of the Steiner tree (i.e., GIS-based FR solution). This concludes the proof of Proposition 1. ■

III. COMPUTATIONAL COMPLEXITY AND DIRECTED GRAPH REPRESENTATION

In this section, the computational complexity of the GIS-based FR problem is discussed. Then, a representing directed graph, including the location of GIS nodes and candidate edges, is presented.

Proposition 2: If the cost of each feeder depends on its length and energy loss, the GIS-based FR problem is NP-hard.

Proof: Assume that the cost of each feeder (edge) depends only on its length. In this case, the GIS-based FR problem is the Steiner tree problem. It was proven by Karp that the Steiner tree problem is NP-hard [11]. This means that there is no polynomial

Multiple factors such as system planning cost, losses, reliability, voltage quality, and grid resiliency are directly related to the configuration of the system. The financial justification of a new expansion plan depends on all these factors, and therefore, they should be considered in an integral evaluation. Various FR formulations have been presented in the literature. A static concave nonlinear cost function was used in [2] for radial distribution system planning. A nonlinear non-differentiable optimization model for minimizing the FR reliability cost was proposed in [3]. Due to the complexity of the proposed model, a heuristic algorithm was used to find a feasible FR solution. In [1], the total cost of reliability for feeder routing design was considered. Criteria were developed for improving customer and utility reliability costs.

One important aspect of FR is the resiliency of the planned grid that is defined as its ability to continue operating and delivering power even in low probability events that produce high disruptions, such as hurricanes, floods, earthquakes, and cyber-attacks [4]. A grid with low resiliency is vulnerable and difficult-to-recover when an event occurs. In other words, the cost of a system with low resiliency is high in the case of an extreme event. On the other hand, designing a highly resilient grid might be very expensive. Hence, a tradeoff exists between grid resiliency and investment costs.

Recently, applications of geographical information system (GIS) have been extended to distribution system operation and maintenance [5]. Power distribution systems can be represented in more detail by taking advantage of GIS. GIS tools can be further extended to cover distribution systems expansion planning and feeder routing. The use of GIS tools results in more effective resiliency analysis since they, for instance, help to incorporate the exact geographical location of each component into the expansion planning model. As another advantage, GIS tools can be used to model the level of accessibility and repair time of each feeder after an extreme event. This is useful, in particular, for resiliency modeling.

Despite traditional FR models without GIS data, few works have been reported on GIS-based feeder routing. To the best of our knowledge, a GIS-based FR problem was firstly presented in [6], and a dynamic programming technique was proposed to solve it. However, since the principle of optimality does not hold for the GIS-based FR problem, the solution obtained is an approximation with no accuracy guarantee. Afterward, a mixed-integer nonlinear programming model was reported in [7] for spatial power system planning using GIS tools. Unlike transmission systems, the energy loss in distribution systems cannot be ignored. Using the DC power flow model reported in [7] is not appropriate for the FR problem since the variable cost of feeders (cost of energy losses) cannot be modeled. A GIS-based methodology was proposed in [8] for transmission line routing using an analytic hierarchical process. It is important to note that in all of these references, uncertainties of demand and renewable energy resources, energy losses, voltage limits, and resiliency of the planned network are ignored.

In this paper, we propose a stochastic programming model for the GIS-based resilient feeder routing problem in power distribution networks. By introducing a representing graph, the

radiality of the planned network is guaranteed. A second-order cone programming (SOCP) formulation is used to represent feeder power flows. Since the goal of a feeder routing problem is to design a radial distribution system, the developed SOCP model is exact and provides the same results as those of a full-AC power flow model [9]. Hence, the proposed SOCP-based feeder routing model is expected to provide more accurate results as those provided by a DC power flow-based model or other approximated linearized mixed-integer programming models. Using a power law distribution function for generating scenarios and a K-means algorithm for scenario reduction, we model the uncertainty pertaining to rooftop solar generations and demand forecasting errors. The proposed GIS-based resilient FR model represents investment costs, power losses and resiliency, while maintaining feasibility regarding voltage limits and reliability.

In addition to the proposed systematic approach for the GIS-based FR problem using available techniques such as SOCP, maximum likelihood estimator, bootstrapping method, power law distribution, and K-means, this paper contributes as follows:

- A proof is provided to show the cost reduction improvement of a GIS-based FR solution with respect to that of an FR model without GIS data.
- A representing graph including candidate edges for the GIS-based FR model is proposed. The specific feature of the proposed graph is that any of its subgraphs is a tree. Using this representing graph and considering the fact that the FR solution is a subgraph of the representing graph, radiality of the planned network is guaranteed. This means that no radiality constraint is required in the GIS-based FR model. Moreover, this also means that the SOCP model is an exact AC power flow model.
- The cost of resiliency of the planned distribution network is quantified using GIS data. The FR objective function is augmented with additional cost terms to incorporate the resiliency component.

The remainder of the paper is organized as follows. The GIS-based FR model and its advantage are discussed in Section II. Computational complexity of the GIS-based FR problem and the proposed representing graph are discussed in Section III. The mathematical MISOCP formulation and the solution algorithm are presented in Sections IV and V, respectively. Numerical results are discussed in Section VI. Concluding remarks are provided in Section VII.

II. GIS-BASED FEEDER ROUTING

A GIS-based FR model has several advantages over conventional models:

- A GIS-based FR model is comparatively more realistic due to its capability for representing barriers and obstacles that are mapped with geographical settings. This benefit is compatible with representing resiliency aspect.
- If only electrical nodes are considered, laterals (branches with lower thermal limits that are separated from the main-line by fuses) are restricted to start from electrical nodes. This might degrade the FR solution quality. Consider Fig. 1 that shows a segment of a distribution system with

Graph-Based Second-Order Cone Programming Model for Resilient Feeder Routing Using GIS Data

Mahdi Mehrtash , *Member, IEEE*, Amin Kargarian , *Member, IEEE*, and Antonio J. Conejo , *Fellow, IEEE*

Abstract—One important task in power distribution system expansion planning is feeder routing (FR), which is to determine the optimal route from a medium voltage substation to load points and the optimal size of the conductors to be installed. This paper proposes a novel graph-based model for the resilient feeder routing problem using geographical information system (GIS). We show that incorporating GIS data enhances the FR solution optimality. Moreover, by introducing a representing graph, radiality of the planned network is guaranteed. A second-order cone programming (SOCP) model is used to model power flows through feeders. Since the representing graph ensures radiality, the SOCP model is exact. The uncertainty of rooftop solar generations and demand forecasting errors are taken into account, and a stochastic programming-based solution algorithm is developed. The proposed model represents practical aspects such as economic objectives (installation cost, power losses, resiliency), technical constraints (voltage limits, radiality constraint, reliability), and geographical constraints. The efficiency of the algorithm is demonstrated using three case studies: a small test system, a realistic one, and a synthetic large test system.

Index Terms—Distribution network expansion planning, feeder routing, geographical information system, resiliency, second-order cone programming.

NOMENCLATURE

A. Indices and Sets

c	Index for conductors.
i, j	Index of sending and receiving nodes of edge (i, j) .
s	Index for scenarios.
γ	Index for segments of cost curves.
$\mathcal{E}$	Set of all possible edges of the representing graph.
$\mathcal{V}$	Set of all nodes of the representing graph.

B. Parameters

$a_{ij,c}$	Fixed cost of conductor c at edge (i, j) .
------------	--

B_{ij}	Series susceptance of edge (i, j) .
c_i	Installation cost.
$c_{O\&M}$	Annual operation and maintenance cost.
C_{res}	Total resiliency cost of the system.
d_{ij}	distance between nodes i and j .
e_c	Cost of energy [\$/kWh].
F	Minimum number of feeders starting from the substation node
G_{ij}	Series conductance of edge (i, j) .
L_f	Annual loss factor.
$m_{ij,c,\gamma}$	Slope of segment γ of the cost curve of conductor c .
$M_{ij,c}$	Large enough numbers, called big-M.
PF	Power Factor.
PD_i^s	Net active load at node i in scenario s .
QD_i^s	Net reactive load at node i in scenario s .
r	Resistance of conductors.
$S_{ij,c}^{\max}$	Maximum thermal limit of conductor c .
$VoLL$	Value of lost load.
τ_{ij}	Restoration time for edge (i, j) .
ω_1, ω_2	Present worth factors.
ρ_s	Weight of scenario s .

C. Variables

$b_{ij,c}$	Variable cost of conductor c at edge (i, j) .
$I_{ij,c}$	Binary decision variable represents the installation decision of conductor c at edge (i, j) .
ℓ	Total energy loss per mile [kWh/mile].
$P_{ij,c}$	Active power of conductor c at edge (i, j) .
$Q_{ij,c}$	Reactive power of conductor c at edge (i, j) .
R_{ij}, L_{ij}	Auxiliary variables for conic AC power flow model.
u_i	Auxiliary variable of voltage magnitude in conic model.
V_i	Voltage magnitude at node i .
θ_i	Voltage angle at node i .

Manuscript received May 26, 2019; revised September 2, 2019 and October 25, 2019; accepted December 7, 2019. Date of publication December 11, 2019; date of current version July 23, 2020. This work was supported in part by National Science Foundation under Grant ECCS-1711850, in part by Louisiana Board of Regents under Grant LEQSF (2016-19)-RD-A-10, and in part by National Science Foundation under Grant 1808169. Paper no. TPWRD-00570-2019. (Corresponding author: Amin Kargarian.)

M. Mehrtash and A. Kargarian are with the Electrical and Computer Engineering Department, Louisiana State University, Baton Rouge, LA 70803 USA (e-mail: mmehrt3@lsu.edu; kargarian@lsu.edu).

A. J. Conejo is with the Electrical and Computer Engineering Department and the Integrated Systems Engineering Department, The Ohio State University, Columbus, OH 43210 USA (e-mail: conejo.1@osu.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TPWRD.2019.2959229

I. INTRODUCTION

POWER distribution networks are key components of the electricity infrastructure. A distribution network should have adequate capacity and appropriate design for reliable, secure, and high-quality electricity delivery to consumers. Designing new distribution systems and upgrading existing ones are required to support load growth.

An important step in distribution expansion planning is feeder routing (FR) [1]. By definition, FR refers to finding the optimum radial routes and conductors' sizes from a medium voltage substation to residential, commercial, or industrial load points.