ELECTRICAL ENGINEERING AND ENERGY

OPEN ACCESS

# Automated Conversion of Unstructured Geospatial Feeder Data into Analytical Models: An 11 kV Case Study

Amon Okemo[1], Christopher Maina Muriithi[1], and John Nderu[2]

[1] Murang'a University of Technology, Murang'a, Kenya
[2] Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Submitted: 15 December 2025
Accepted: 08 January 2026
Online First: 11 January 2026

Corresponding author
Amon Okemo,
amonokemo8@gmail.com

DOI: 10.64470/elene.2026.20



**Abstract:** The modernization of power distribution grids, driven by the integration of DER, necessitates advanced modelling capabilities. A critical challenge is the semantic gap between the extensive geospatial asset data and the detailed electrical models required for engineering analysis. This data is often stored in static, unstructured KMZ formats that contain inherent topological errors. To address this, this paper presents a novel low-cost, Python framework that fully automates the conversion of this raw GIS data into solvable mathematical models for computation. This process generates executable files for industry-standard, script-based simulators such as MATPOWER and OpenDSS. The framework's core technical contributions include an XPath and Regex-based engine for metadata extraction. A graph-theory pipeline then utilizes a Minimum Spanning Tree (MST) algorithm to algorithmically heal topological disconnections. A Dijkstra-based method is then used for model abstraction. The core methodology was first validated using a bidirectional, reverse-engineering process on a 4-bus test case. This confirmed a lossless round-trip conversion to and from a data-rich KMZ format. Subsequently, it was applied to a complex, real-world 11 kV Kenyan distribution feeder. The generated model converged in a Newton-Raphson power flow. This demonstrated its utility as a powerful diagnostic instrument by enabling detailed feeder voltage profiling and loss analysis. Results were cross-validated against DIgSILENT PowerFactory, a graphic-based simulator, showing excellent numerical agreement. This study validates a scalable framework that transforms static, error-prone GIS data into dynamic, multi-platform diagnostic models. This approach provides a feasible pathway to accelerate grid modernization.

**Keywords** Data-driven diagnostics, Distribution networks, KMZ metadata extraction, Power system modelling, Topological healing

## 1. Introduction

The global energy landscape is undergoing a profound transformation. This shift is driven by the urgent need for decarbonization and the increasing integration of Distributed Energy Resources (DERs) such as solar photovoltaics and electric vehicles (Matanov & Nankinsky, 2021). This evolution places unprecedented complexity on electrical distribution networks. These networks were not originally designed for bidirectional power flows or dynamic operational profiles. Consequently, advanced modelling, simulation and analysis has become paramount. Utilities must effectively plan grid reinforcements to ensure

reliability and maintain power quality (Javed et al., 2021; Mohd Azmi et al., 2022).

A critical enabler for such analysis is the availability of accurate network models. However, a fundamental semantic gap persists between the Geographic Information Systems (GIS) and engineering software (Rahman et al., 2020). GIS formats like KML/KMZ are excellent for visualization but inherently lack the structured electrical metadata required for direct simulation (Geth et al., 2023; Tóth et al., 2020). This gap creates a significant operational bottleneck. The process of manually interpreting a KMZ file into analytical format is a well-documented challenge (Abeysinghe et al., 2021; Patel et al., 2025). This manual translation of visual GIS maps into solvable analytical models is slow, labour-intensive and prone to error (AL-Jumaili et al., 2023; Mukherjee et al., 2020).

This challenge is particularly acute for distribution utilities in rapidly developing regions, where operational excellence is a central pillar of strategic planning. For instance, the Kenyan distribution utility aims to reduce system losses from a baseline of 23% to a target of 15.5% (Kenya Power, 2023). A core strategy to achieve this is energy accounting at granular feeder and transformer levels. However, performing such analysis is contingent upon having accurate simulation models. The manual conversion process from the extensive KMZ asset archives severely hinders these data-driven diagnostic efforts.

Comprehensive standards like the Common Information Model (CIM) exist to promote interoperability, but their implementation remains complex and costly for many utilities (Anderson et al., 2022). Recent academic research has proposed various methods to automate model generation, yet limitations persist when dealing with the data realities of many utilities. Studies by Deka et al. (2024) and Subasic et al. (2022) focused on inferring network topology from operational measurements. These data-driven methods leverage time-series voltage correlations from smart meters to deduce connectivity. While effective for real-time grid calibration, these methods utilize a data-driven approach that presupposes a fully deployed Advanced Metering Infrastructure (AMI). This is a capital-intensive resource often unavailable in developing regions. In contrast to these measurement-driven inference methods, our approach focuses on the structural analysis of the static, as-is geospatial asset data, which does not require real-time operational measurements.

Conversely, Montano-Martinez et al. (2024) proposed using external datasets to correct and enhance existing GIS data. This includes using OpenStreetMap and municipal parcel data to correct asset coordinate errors. However, this approach relies on the accuracy of third-party data rather than the utility's own records. Furthermore, existing open-source conversion tools, such as the QGIS2OpenDSS plugin, are typically designed for structured GIS formats like Shapefiles. They fail when confronted with the unstructured, free-text metadata typical of KML/KMZ files used in the field (De-Jesús-Grullón et al., 2024).

These existing methods fail to address the unique challenge presented by KML/KMZ files. This format is inherently designed for visual presentation rather than engineering analysis. It is characterized by unstructured text descriptions and a lack of topological connectivity. The specific research gap is the lack of an integrated pipeline that can autonomously process these error-prone files. Such a pipeline must heal inherent topological disconnections and generate a fully parameterized model, without reliance on operational data streams or external structured databases. This paper presents a complementary approach. It introduces a framework that performs topological healing based on the geographical proximity of assets found within a single, self-contained KML/KMZ file.

To clearly distinguish the contributions of this study from previous literature, this research introduces three specific innovations. First, it eliminates the dependency on dynamic sensor data by reconstructing network topology solely from static asset maps using Graph Theory. This provides a viable solution for utilities that have not yet deployed AMI. Second, it automates metadata extraction through a novel regex engine designed to parse non-standardized HTML descriptions. This directly addresses the rigidity that causes previous conversion tools to fail on unstructured data. Finally, it bridges the technological gap for

developing nations by establishing a low-cost, open-source pathway to digital twins. This is specifically tailored for utilities that rely on KML/KMZ as their primary system of record.

The framework's scalability and practical utility are demonstrated through its application to a large-scale, real-world utility feeder. In this case, it functioned as a powerful diagnostic tool for identifying operational issues like voltage violations. A rigorous validation process established the resulting model's high fidelity and multi-platform compatibility. This showed successful convergence and excellent numerical agreement across open-source environments and commercial-grade simulators (Khanh et al., 2025; Montenegro et al., 2022).

## 2. Methodology

The framework was implemented in Python (v3.9) environment. It leverages a suite of open-source libraries, including Pandas for data manipulation, LXML for KML parsing and NetworkX for graph-based topological analysis. The methodology follows a systematic, multi-stage pipeline designed to autonomously process raw, unstructured KMZ data into a fully parameterized, solvable power system model. This final model is then used to generate executable scripts for multiple industry-standard simulators including MATPOWER and OpenDSS, as illustrated in Figure 1.

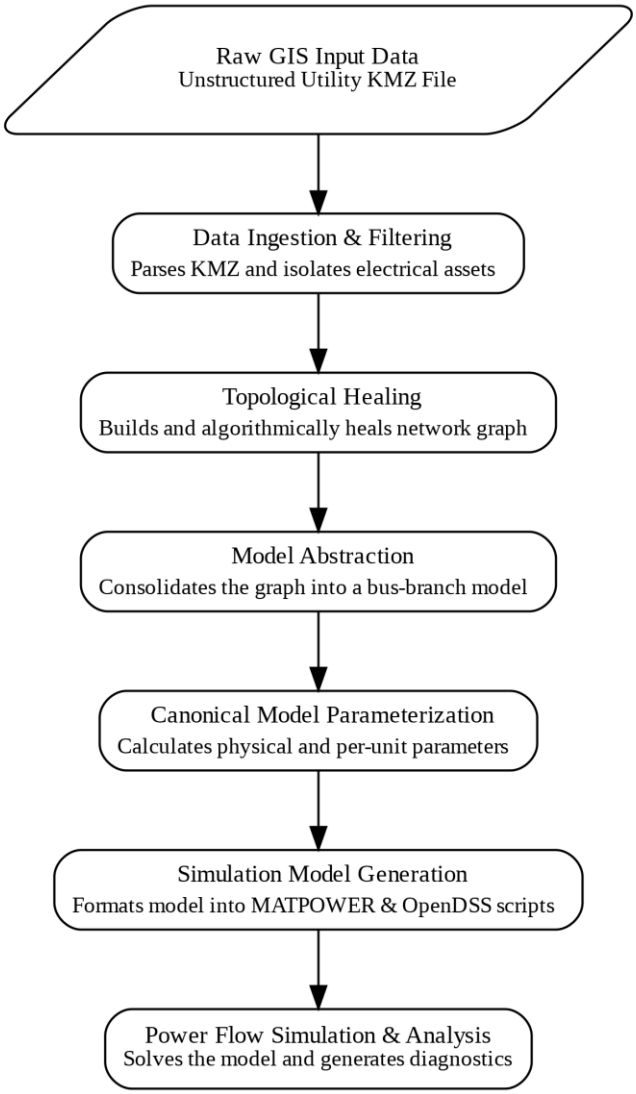


**Figure 1** Conceptual workflow of the automated KMZ to Simulation framework.

### 2.1. Data ingestion and metadata-driven filtering

The process begins by ingesting a standard utility-grade KMZ file, which is a zipped archive containing more KML documents. The framework unzips the archive in memory and parses the primary KML document using the LXML library. An XPath query is used to iterate through all placemark elements, which represent individual geospatial assets. For each element, descriptive metadata and geospatial are extracted from the HTML-formatted description tag using Regular Expressions (Regex).

A critical challenge is distinguishing the core electrical network from thousands of non-essential elements, such as low voltage poles or purely visual markers. To solve this, a metadata-driven filtering heuristic is applied. It inspects attributes such as origin of the element and feeder of the element to ensure the asset belongs to the target circuit. Assets belonging to adjacent feeders are automatically discarded. The algorithm then classifies the remaining placemarks as valid electrical components based on the presence of specific technical keywords.

For instance, a line segment is classified as a true medium voltage electrical branch only if its description contains a non-null attribute such as type of conductor, size of the conductor and length (km). The framework assumes that these valid line segments are explicitly represented in the KMZ as paths connecting two distinct asset placemarks (such as a line from pole A to pole B). The subsequent topological healing step (Section 2.2) is then applied to connect these defined segments or components where topological gaps exist. The length attribute is particularly critical, as it represents the physical span between individual poles in the raw data. Similarly, a point placemark is identified as a transformer only if it contains attributes such as substation rating (kVA), substation number and number of meters. Non-electrical assets like poles, stay wires or visual pointers typically lack these specific electrical attributes and are therefore discarded. This heuristic operates on a principle of positive identification: only elements containing explicit, recognized electrical keywords or attributes (such as 'kVA', 'conductor', 'feeder') are classified as valid assets. Consequently, non-essential elements such as lighting poles, stay wires, or visual markers, which inherently lack these specific identifiers, are automatically filtered out. This method proved highly effective, successfully isolating the core electrical network for subsequent processing.

### 2.2. Topological healing of network connectivity gaps

Raw, visually digitized GIS data is often topologically fragmented, containing gaps that would cause a simulation to fail. The framework implements an automated topological healing algorithm based on graph theory using the NetworkX library. It utilizes two-dimensional geospatial coordinates (latitude and longitude) for all calculations. While KML supports altitude, distribution feeders in the studied region generally follow the terrain gradient. Consequently, the difference between the 2D geodesic length and the 3D slant length is negligible for steady-state power flow analysis. Therefore, the framework calculates the missing length across topological gaps using the 2D Haversine formula, which determines the shortest great-circle distance between two points on the Earth's surface.

This healing process is implemented using a graph-theory algorithm that involves three key steps:

**Graph construction:** An initial network graph, $G=(V,E)$, is created from the filtered electrical assets. The vertices ($V$) represent the start and end coordinates of these components, and the edges ($E$) represent the explicit line segments themselves.

**Disconnected component identification:** A standard connected components algorithm is run on graph $G$ to identify the set of all disconnected subgraphs, $C = \{C_1, C_2, \ldots, C_k\}$.

**Algorithmic Healing:** To bridge the gaps between these islands, a weighted, complete meta-graph, $G' = (V', E')$, is constructed where each vertex $v_i' \in V'$ represents a component $C_i \in C$. The weight $w(e_{ij}')$ of an edge between any two vertices $v_i'$ and $v_j'$ is the minimum geographical distance between their respective components. This distance $d$ is calculated using the Haversine formula, which proceeds in three steps.

First, the component $a$, representing the square of half the chord length, is calculated using Equation (1). Next, the angular distance $c$ (in radians) is determined using Equation (2). Finally, the physical distance $d$ is computed using Equation (3).

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (1)$$

$$c = 2 \cdot atan2\left(\sqrt{a}, \sqrt{1-a}\right) \quad (2)$$

$$d = R \cdot c \quad (3)$$

where:

- $\phi_1, \phi_2$ are the latitudes of the two points.
- $\Delta\phi$, $\Delta\lambda$ are the differences in latitude and longitude, respectively.
- $a$ is the square of half the chord length between the points.
- $c$ is the angular distance in radians.
- $R$ is the Earth's radius (approx. 6,371 km).

A Minimum Spanning Tree (MST) is then computed for this meta-graph *G'*. The edges of the MST represent the most efficient set of virtual connections required to join all islands into a single, contiguous and topologically correct network with the shortest possible total length of added connections. This approach relies on the principle that distribution networks are designed to minimize conductor usage. Therefore, the MST algorithm inherently associates branching points with the nearest physical pole, validating connectivity without manual intervention. These connections are added back to the original graph *G* to form the final healed graph, $G_{healed}$.

### 2.3. Model consolidation and abstraction

The detailed healed graph is consolidated into an electrically equivalent bus-branch model suitable for simulation. First, the algorithm identifies major nodes ($V_{\text{major}} \subset V(G_{\text{healed}})$) which become the buses in the final model. This set consists of three distinct, programmatically identified categories: the source substation node, all transformer location nodes and all junction nodes. A junction is explicitly defined and identified by the algorithm as any node in the healed graph with a degree greater than two. This ensures that all branching points in the network topology are correctly preserved as buses in the final model.

Next, for every pair of major nodes $(u, v) \in V_{\text{major}}$, Dijkstra's algorithm is used to find the shortest path, $p(u, v)$, in the healed graph. If this path contains no other major nodes, all its constituent segments are consolidated into a single equivalent branch. This consolidation process is strictly geometric; it traces the existing physical path defined by the intermediate pole coordinates in the KMZ file. No new electrical branches are inferred during this stage; the algorithm solely aggregates the series of physical segments between major nodes into a single equivalent electrical branch. The length of this new branch, $L_{\text{branch}(u,v)}$, is the sum of the lengths of all its individual segments, as given by Equation (4).

$$L_{\text{branch}(u,v)} = \sum_{(i,j)\in p(u,v)} l_{\text{segment}(i,j)} \quad (4)$$

Finally, an iterative routine merges any buses that are co-located or connected by a zero-length branch, resulting in a definitive set of buses and branches that form the canonical representation of the feeder.

### 2.4. Canonical model parameterization and load allocation

The abstracted bus-branch model is fully parameterized based on the system's nominal base values of power ($S_{\text{base,MVA}}$) and voltage ($V_{\text{base,kV}}$), from which the base impedance ($Z_{\text{base}}$) is derived using Equation (5):

$$Z_{\text{base}}[\Omega] = \frac{(V_{\text{base,kV}})^2}{S_{\text{base,MVA}}} \quad (5)$$

**Load modelling:** The active power ($P_{\text{d,i}}$) and reactive power ($Q_{d,i}$) for each load bus are calculated from the associated transformer kVA rating ($\text{kVA}_i$) and user-defined parameters. Two alternative scenarios are provided:

i. **Proportional peak load allocation:** The active power is allocated proportionally to transformer ratings using Equation (6), where $P_{\text{peak,total}}$ is the total feeder load and $N_{tx}$ is the number of transformers.

ii. **Uniform Utilization Factor (TUF) allocation:** Alternatively, for a scenario based on a uniform TUF, active power is calculated using Equation (7).

Reactive power is then determined by the system power factor (PF), as shown in Equation (8).

$$P_{d,i} = P_{\text{peak,total}} \times \left( \frac{\text{kVA}_i}{\sum_{k=1}^{N_{tx}} \text{kVA}_k} \right) \quad (6)$$

$$P_{d,i} = \left( \frac{\text{kVA}_i}{1000} \right) \times \text{TUF} \times \text{PF} \quad (7)$$

$$Q_{d,i} = P_{d,i} \times \tan(\cos^{-1}(\text{PF})) \quad (8)$$

**Line impedance parameterization:** For each consolidated medium-voltage line of length $L_{km}$ , its total per-phase resistance ($R'_{\Omega/km}$) and reactance ($X'_{\Omega/km}$) are retrieved from a predefined conductor library. These physical values are then converted to their per-unit equivalents using Equation (9) for resistance and Equation (10) for reactance:

$$r_{\text{p.u.}} = \frac{R'_{\Omega/\text{km}} \times L_{\text{km}}}{Z_{\text{base}}} \quad (9)$$

$$x_{\text{p.u.}} = \frac{X'_{\Omega/\text{km}} \times L_{\text{km}}}{Z_{\text{base}}} \quad (10)$$

**Transformer impedance parameterization:** Transformers are modelled as a branch connecting medium-voltage bus to a newly created low-voltage bus. The impedance is derived from the transformer's kVA rating ($S_{\text{tx,kml}}$), percentage impedance ($\%Z_{\text{kml}}$), and X/R ratio. The impedance is first calculated in per-unit on the transformer's own base using Equation (11) and then converted to the system base using the standard base-change formula (12). This value is then resolved into its resistive ($r_{\text{p.u.}}$) and reactive ($x_{\text{p.u.}}$) components.

$$Z_{\text{p.u., tx-base}} = \frac{\%Z_{\text{kml}}}{100} \quad (11)$$

$$Z_{\text{p.u., system}} = Z_{\text{p.u., tx-base}} \times \frac{S_{\text{base,MVA}}}{S_{\text{tx,kml}}/1000} \quad (12)$$

### 2.5. Simulation script generation

The final step translates the fully parameterized canonical model into executable scripts for target simulation engines. The generated models are designed to be solved using a standard Newton-Raphson algorithm to find a solution for the complex bus voltages ($V$) in the AC power flow Equations (13, 14), which express the balance of active power ($P_i$) and reactive power ($Q_i$) at each bus $i$.

$$P_i = \sum_{k=1}^{N} |V_i||V_k|(G_{ik}\cos(\delta_i - \delta_k) + B_{ik}\sin(\delta_i - \delta_k)) \quad (13)$$

$$Q_i = \sum_{k=1}^{N} |V_i||V_k|(G_{ik}\sin(\delta_i - \delta_k) - B_{ik}\cos(\delta_i - \delta_k)) \quad (14)$$

where:

- N is the total number of buses.
- $|V_i|$ and $|V_k|$ are the voltage magnitudes at bus *i* and bus *k*.
- $\delta_i$ and $\delta_k$ are the voltage angles at bus *i* and bus *k*.
- $G_{ik}$ and $B_{ik}$ are the real and imaginary parts of the *ik*-th element of the bus admittance matrix ($Y_{bus}$).

The framework generates executable scripts:

For MATPOWER: The data is assembled into the standard bus, branch and gen NumPy matrices and packaged into a ppc dictionary structure, which can be solved by PYPOWER or exported as an executable .m file.

For OpenDSS: The data is used to programmatically write a text-based .dss script, generating the necessary New Linecode, New Line, New Transformer and New Load commands required by the OpenDSS engine.

This modular approach ensures that the core processing engine is independent of the final simulation software, allowing for flexible output to multiple platforms.

## 3. Validation and Case Study Systems

The methodology was validated and its practical utility demonstrated through a two-part approach. First, a bidirectional round-trip validation was performed on a standard test case to rigorously prove the integrity and fidelity of the core data conversion process. Second, the framework was applied to a large-scale, real-world feeder to demonstrate its scalability and effectiveness as a diagnostic tool.

### 3.1. Bidirectional framework validation: 4-bus proof-of-concept

To establish the fundamental integrity of the data conversion process, a bidirectional or round-trip validation was performed using a standard 4-bus distribution network test case. The system, shown in Figure 2, represents a simplified radial feeder and includes a slack bus (Bus 1), two PQ load buses (Bus 2 and 3), and a PV bus with an underlying load (Bus 4). The initial MATPOWER parameters for the system's buses, generators and branches are detailed in Tables 1-3.

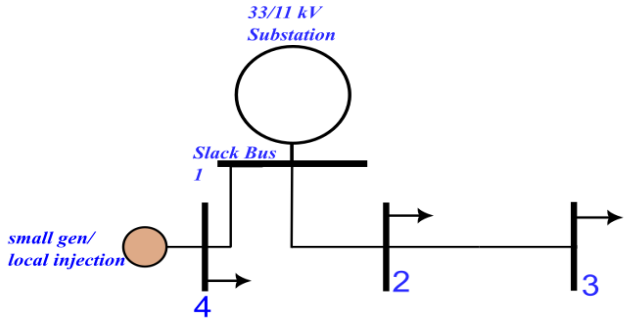


**Figure 2** Single-line diagram of the 4-bus distribution network test case used for methodology validation

**Table 1** Summary of initial bus parameters for the 4-bus test system.

| Bus ID | Type | Pd (kW) | Qd (kVAR) | Vm (p.u.) | Va (°) | Base Voltage (kV) | Vmax | Vmin |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 (Slack Bus) | 0 | 0 | 1.00 | 0 | 12.5 | 1.1 | 0.9 |
| 2 | 1 (PQ Bus) | 0.4 | 0.2 | 1.00 | 0 | 12.5 | 1.1 | 0.9 |
| 3 | 1 (PQ Bus) | 0.4 | 0.2 | 1.00 | 0 | 12.5 | 1.1 | 0.9 |
| 4 | 2 (PV Bus) | 0.4 | 0.2 | 1.00 | 0 | 12.5 | 1.1 | 0.9 |

**Table 1** Summary of initial generator parameters for the 4-bus test system

| Bus ID | Pg (MW) | Qg (MVAR) | Qmax (MVAR) | Qmin (MVAR) | Vg (p.u.) | mBase (MVA) | Pmax (MW) | Pmin (MW) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 10 | -10 | 1.05 | 100 | 10 | 0 |
| 4 | 0 | 0 | 10 | -10 | 1.05 | 100 | 10 | 0 |

**Table 2** Summary of initial branch parameters for the 4-bus test system

| From Bus | To Bus | r (p.u.) | x (p.u.) | angle (°) | status | angmin (°) | angmax (°) |
|---|---|---|---|---|---|---|---|
| 2 | 3 | 0.003 | 0.006 | 0 | 1 | -360 | 360 |
| 1 | 2 | 0.003 | 0.006 | 0 | 1 | -360 | 360 |
| 4 | 1 | 0.003 | 0.006 | 0 | 1 | -360 | 360 |

The validation followed three distinct phases:

**i. Phase I (Forward Conversion: MATPOWER to a Data-Rich KMZ):**

The goal of this phase was to simulate the creation of a utility-grade KMZ file from a known electrical model. First, the initial MATPOWER case data was enriched with descriptive asset metadata such as pole types, substation names, and assigned geographical coordinates which established a plausible physical layout. The framework then executed a parameter transformation algorithm, converting the per-unit electrical data into physical values. A key step was to define the visual representation for each asset, assigning specific shapes, colours and icons to each component. Finally, all of this information including asset names, coordinates, visual styles and the crucial physical and electrical metadata was programmatically written into a single KML file. This was achieved using the standardized metadata schema and packaged into a self-contained KMZ archive. This transformed the abstract MATPOWER model into a tangible, visually rich and machine-readable geospatial file.

**ii. Phase II (Reverse Engineering: KMZ back to MATPOWER):**

The generated KMZ file was used as the sole input for the reverse-engineering pipeline to test its data extraction and model reconstruction capabilities. The framework first parsed the KML structure and extracted all embedded metadata from the description tags for every placemark. A filtering algorithm was then applied to distinguish electrical from non-electrical components. Assets without specific electrical metadata (like intermediate poles) were correctly identified as purely structural and were not included in the final electrical model. The extracted data for the remaining electrical components (the source, loads, lines and transformer) was then used by the parameterization engine. This engine performed the inverse of Phase I, converting the physical values back to the system's per-unit base to reconstruct a new set of bus, generator and branch matrices from scratch, thereby recreating the analytical circuit.

**iii. Phase III (Fidelity Check and Visual Reconstruction):**

The final phase was a two-part verification of the process. First, a direct numerical comparison was performed between the matrices of the original MATPOWER model and the newly reconstructed model.

This fidelity check confirmed a lossless, high-precision round-trip conversion, with the reconstructed matrices being electrically equivalent to the originals. Second, the framework used the reconstructed model to automatically generate visual outputs, including a clean, untangled IEEE-style busbar diagram. The successful generation of this diagram, which correctly reflected the topology of the original 4-bus system, served as a visual validation of the framework's topological interpretation and model reconstruction capabilities. Together, these numerical and visual checks successfully proved the core concept of the methodology.

### 3.2. Real-world application: an 11 kV distribution feeder case study

To demonstrate the framework's scalability and practical utility on a real-world problem, the automated one-way conversion pipeline was applied to an operational 11 kV distribution feeder. The data was provided by a national Kenyan utility as a single, large KMZ file containing 3,765 distinct placemarks representing the feeder's assets, including lines, transformers, poles, and other visual markers.

The framework was configured for a typical planning study scenario based on utility-provided data: a total feeder peak load of 2.7 MW at a power factor of 0.95. This complex, real-world dataset, with its inherent data quality issues such as inconsistent metadata and the topological gaps inherent in visually digitized data, served as the input for the automated conversion process. The key statistics from the automated processing pipeline, which highlight the complexity of the data and the scale of the automated correction, are summarized in Table 4.

**Table 4** Key statistics from the automated model generation process for the 11 kV feeder

| Parameter | Value |
|---|---|
| Total Placemarks Processed | 3,765 |
| Identified Transformers | 94 |
| Identified Electrical Lines | 101 |
| Detected Disconnected Components | 48 |
| Algorithmically Healed Gaps | 47 |
| Final Model Buses (medium-voltage) | 158 |
| Final Model Branches (medium-voltage) | 157 |

The topological healing algorithm proved essential, identifying 48 disconnected network islands. It then created 47 virtual connections to form a single, contiguous graph algorithmically. Following this, the model consolidation and iterative merging process produced the final, topologically correct radial network model. This final model consists of 158 medium-voltage buses and 157 branches, which formed the basis for the power flow analysis.

## 4. Results and Discussion

The methodology was successfully executed in both the validation and real-world application stages, yielding verifiable and insightful results. The following sections detail the outcomes of the 4-bus proof-of-concept, which confirmed the method's numerical fidelity, and the application to the large-scale utility feeder, which demonstrated its scalability and diagnostic power.

### 4.1 Four-Bus validation results

The round-trip conversion process for the 4-bus test case was successful. The forward conversion (Phase I) generated a data-rich KMZ file with all electrical and asset metadata correctly embedded, as illustrated by the geospatial visualization in Figure 3. The reverse-engineering process (Phase II) accurately parsed this KMZ file and reconstructed the bus, gen and branch matrices, confirming a high-fidelity, lossless conversion.

In the final validation phase (Phase III), an AC power flow simulation on the reconstructed 4-bus model

converged successfully. The key steady-state results are summarized in Table 5 and system summary Table 6. These results were numerically identical to four decimal places with a control simulation run on the original case file in MATLAB/MATPOWER, robustly validating the entire bidirectional process.

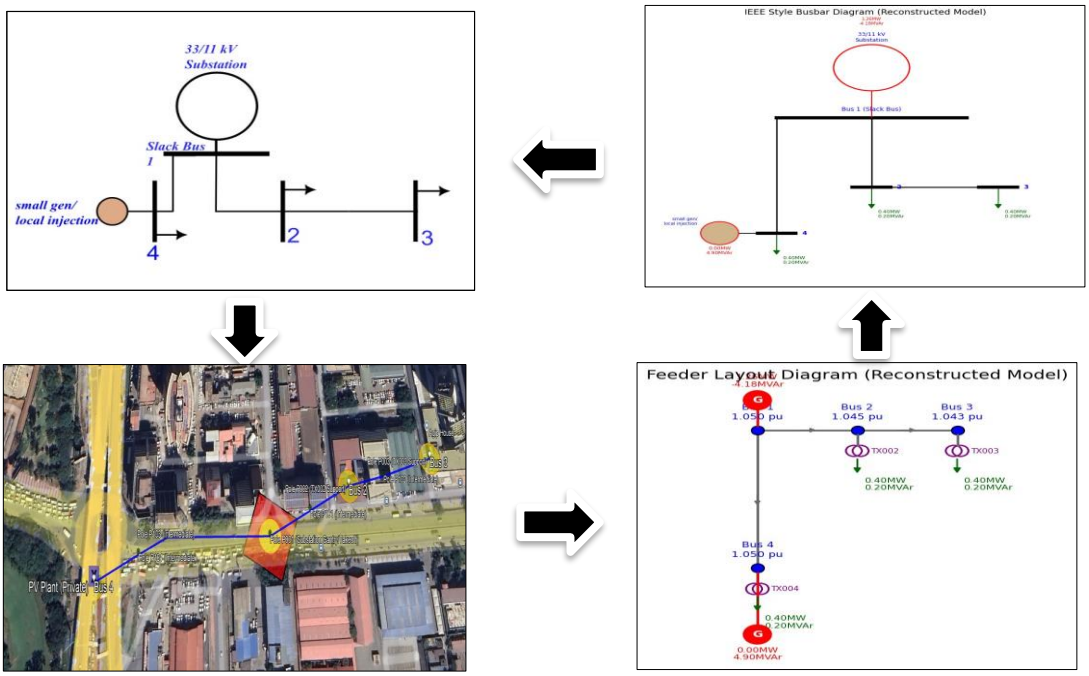


**Figure 3** Visual workflow of the 4-bus validation, from the original MATPOWER model (top-left) to geospatial visualization (bottom-left), and the automatically reconstructed analytical models (right)

**Table 5** Solved bus voltage magnitudes and angles from the power flow solution on the KML-reconstructed 4-bus model, confirming numerical fidelity

| Bus | VM (pu) | VA (deg) | Pd (MW) | Qd (MVAr) |
|---|---|---|---|---|
| 1 | 1.0500 | 0.0000 | 0.000 | 0.000 |
| 2 | 1.0454 | -0.1595 | 0.400 | 0.200 |
| 3 | 1.0430 | -0.2398 | 0.400 | 0.200 |
| 4 | 1.0500 | -0.8794 | 0.400 | 0.200 |

**Table 6** System Summary from the power flow solution on the KML-reconstructed 4-bus model

| Parameter | MW | MVAr |
|---|---|---|
| Total Generation | 1.2637 | 0.7262 |
| Total Load | 1.2000 | 0.6000 |
| Total Losses | 0.0637 | 0.1262 |

### 4.2 Case study: automated model generation and visualization

The application of the framework to the real-world 11 kV feeder demonstrated its effectiveness in handling large, unstructured and topologically flawed utility data. The automated pipeline successfully transformed

the raw geospatial data into a clean, solvable electrical model through a multi-stage process.

The process began with the raw KMZ data, a dense geospatial file containing 3,765 placemarks. In the first step, the framework applied its metadata-driven filter to parse this data, successfully isolating the core electrical assets. This crucial filtering step correctly identified the source substation, 94 transformers and 101 true medium-voltage line segments, while discarding 186 non-electrical marker lines and thousands of irrelevant pole nodes, resulting in the clean extracted network shown in Figure 4.

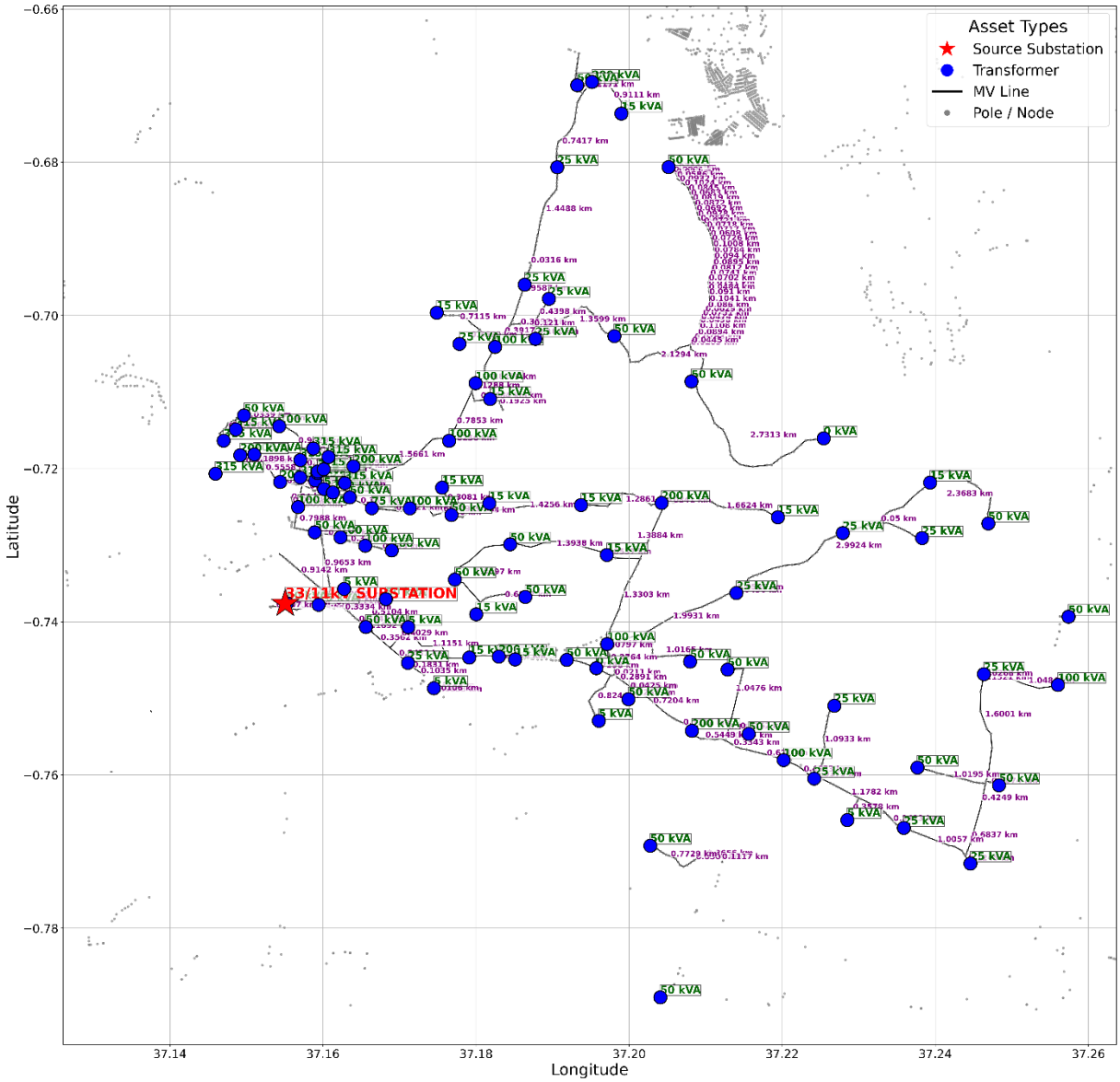


**Figure 4** Extracted network from the raw KMZ data with relevant electrical assets

Analysis of this extracted network revealed that it was topologically fragmented into 48 disconnected components. The topological healing algorithm was then executed, algorithmically adding 47 virtual connections, shown as red dashed lines, to bridge these gaps. Subsequently, the model abstraction process consolidated the detailed, pole-by-pole graph into a simplified, electrically equivalent bus-branch model containing the final 158 medium-voltage buses and 157 branches. Figure 5 and Figure 6 show this final healed and abstracted geographical model, which is now topologically complete and ready for parameterization.

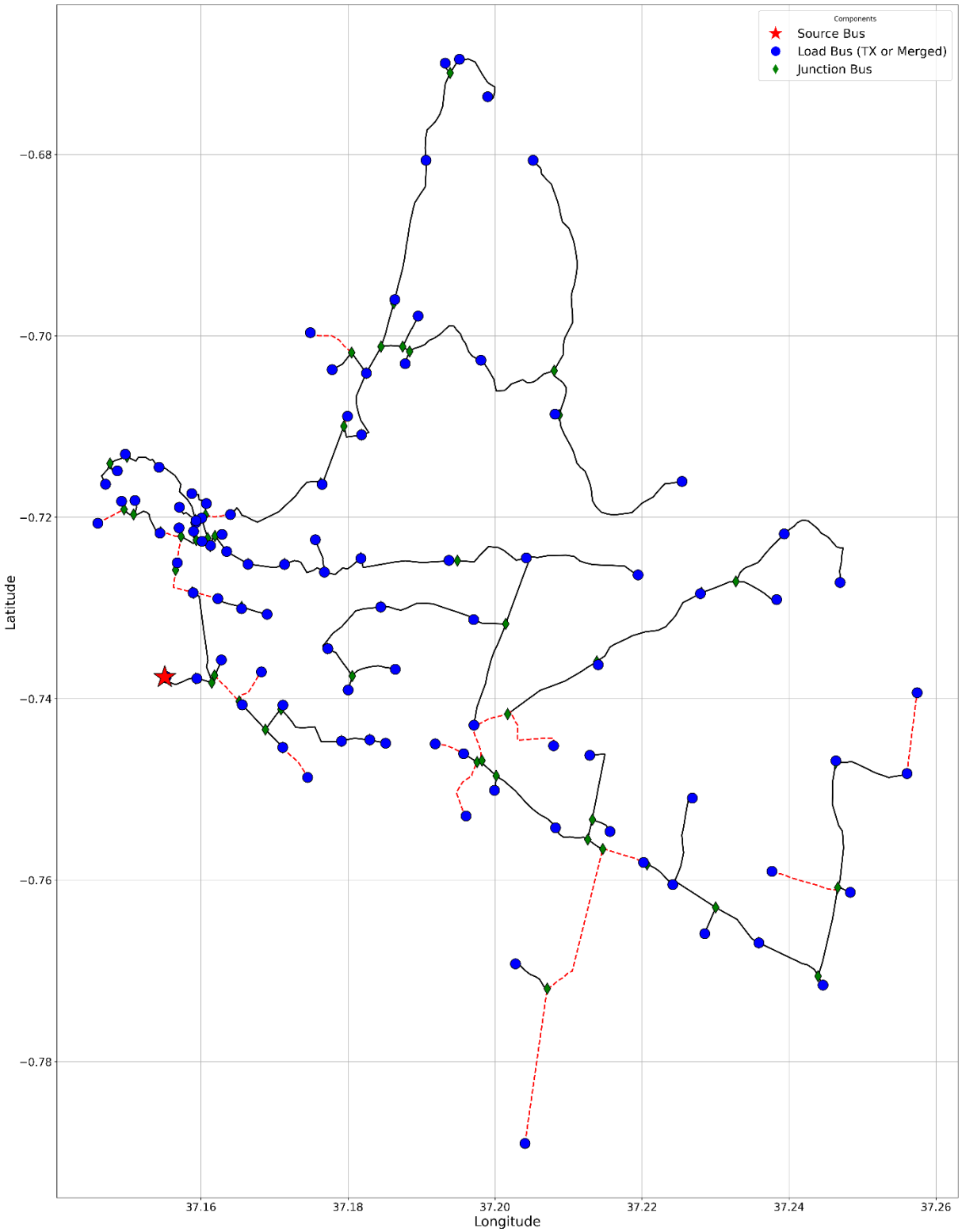


**Figure 5** Healed geographical model, provides a topologically correct spatial view

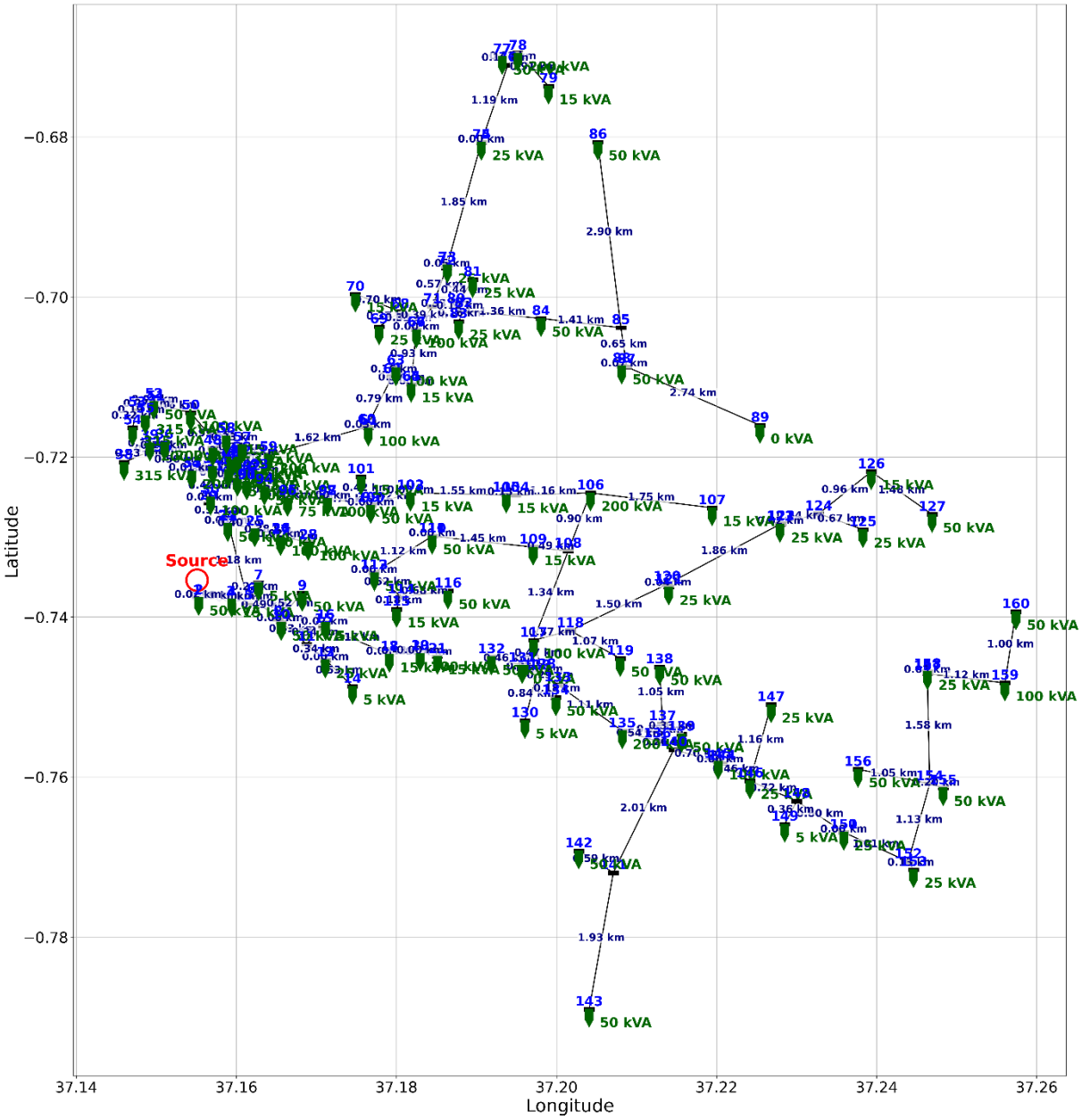


**Figure 6** Final abstracted geographical model, topologically complete and ready for parameterization

From this final model, the framework automatically generated a clean, untangled IEEE-style single-line diagram (Figures 7 and 8). This schematic provides a purely logical representation of the network topology, complete with the final bus numbering, transformer kVA ratings and consolidated medium-voltage branch lengths, serving as the definitive, human-readable blueprint of the simulation model.

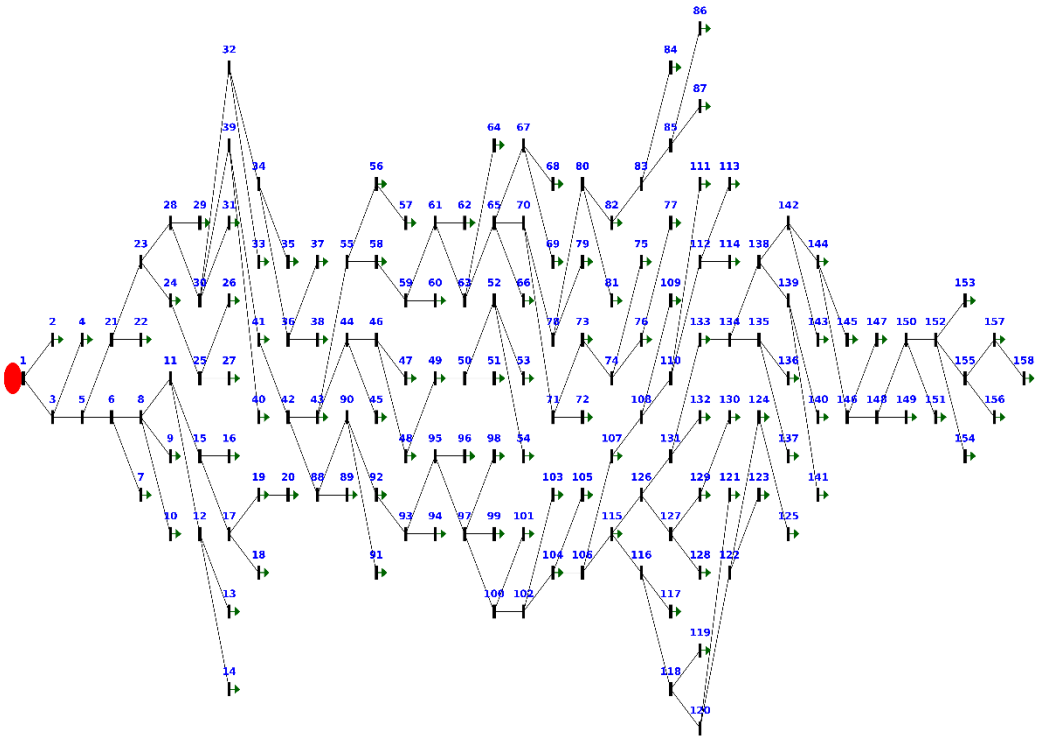

**Figure 7** Clean, untangled IEEE-style single-line diagram with medium-voltage buses

### 4.3 Case study: power flow analysis

Following the successful generation of the 158-bus (medium-voltage) model, an AC power flow was executed to analyse the feeder's performance under the specified 2.7 MW peak load scenario. The simulation, run using the Newton-Raphson solver in PYPOWER, successfully converged in 3 iterations, a strong indicator of a well-posed and valid network model. The high-level performance metrics are summarized in Table 7.

**Table 7** High-level power flow analysis summary for the 11 kV feeder at 2.7 MW peak load

| Parameter | Value |
|---|---|
| Total Feeder Load | 2.700 MW |
| Total Generation Required | 2.814 MW |
| Total System Losses (MW) | 0.114 MW |
| System Losses (%) | 4.05% |
| Minimum medium-voltage Voltage | 0.937 p.u. |
| Bus with Minimum Voltage | Bus 158 |
| Maximum Line Current | 156.31 A |

The most significant finding of the analysis was the identification of considerable voltage drop along the feeder's electrically distant sections. As detailed in the voltage profile plot (Figure 9), the voltage drops steadily along the feeder with numerous buses operating close to the lower statutory limit of 0.90 p.u. The most dropped condition was observed at Bus 158, the furthest point from the source, registering a voltage of 0.937 p.u. This result demonstrates the framework's capability as a powerful diagnostic instrument for identifying specific network weaknesses that may require reinforcement to maintain power quality.

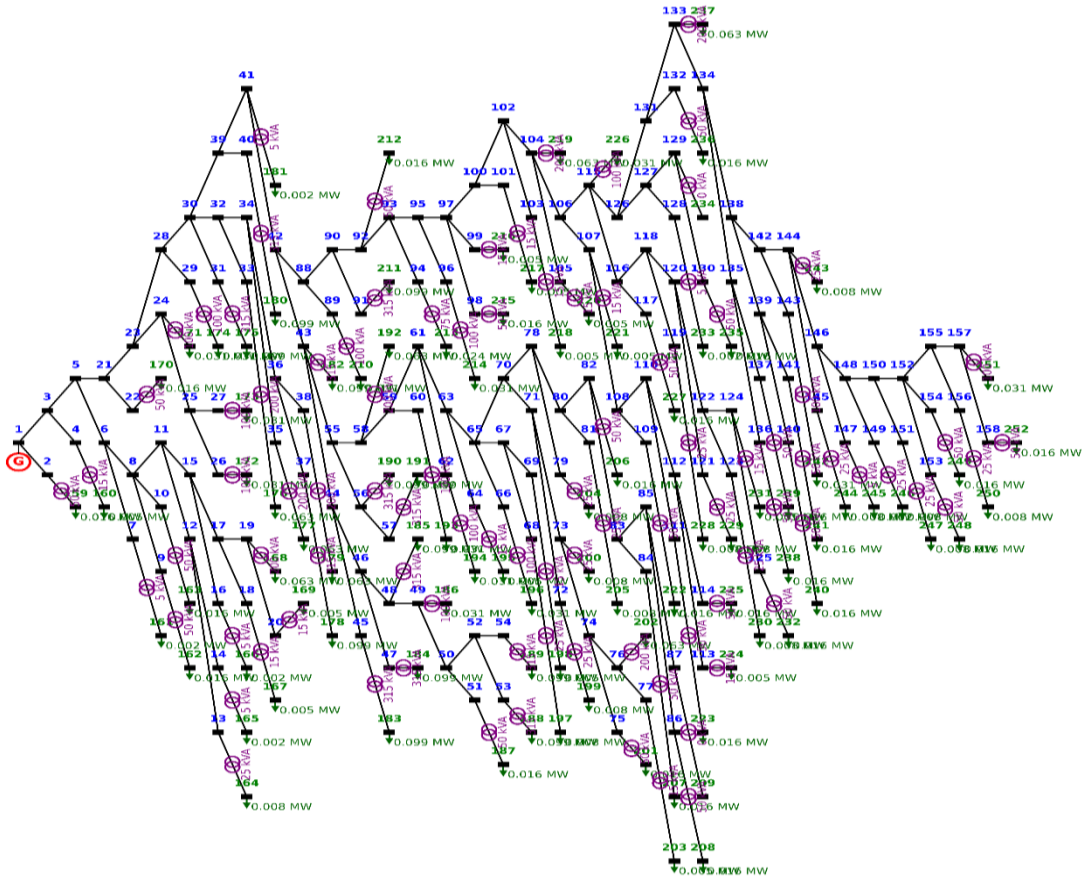

**Figure 8** Untangled IEEE-style single-line diagram with medium-voltage buses, transformers and low-voltage buses

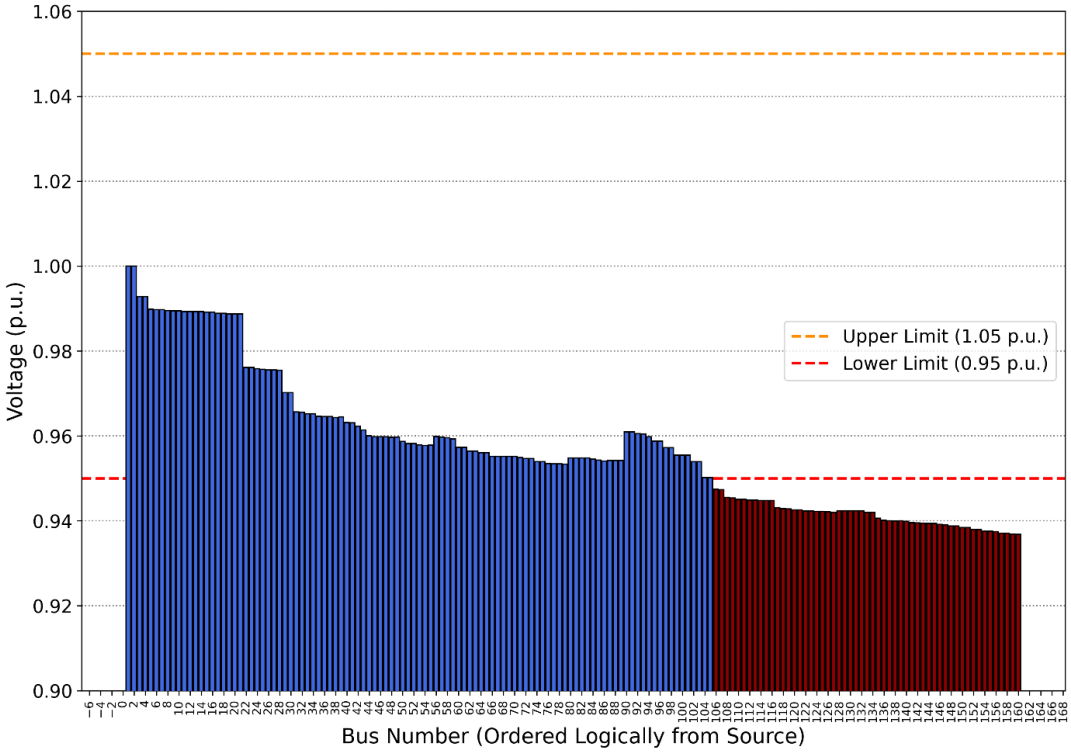


**Figure 9** Medium-voltage profile for the 11 kV distribution feeder. The plot shows the per-unit voltage at each of the 158 medium-voltage buses, ordered logically from the source

The geographical distribution of these results is visualized in the heat map shown in Figure 10. This map clearly illustrates the spatial nature of the power quality issue, with healthy voltages (green) near the source substation and progressive voltage drops (yellow) along the feeder's extremities. This result demonstrates the framework's capability not just as a modelling tool, but as a powerful diagnostic instrument for identifying specific geographical areas with network weaknesses that may require reinforcement.

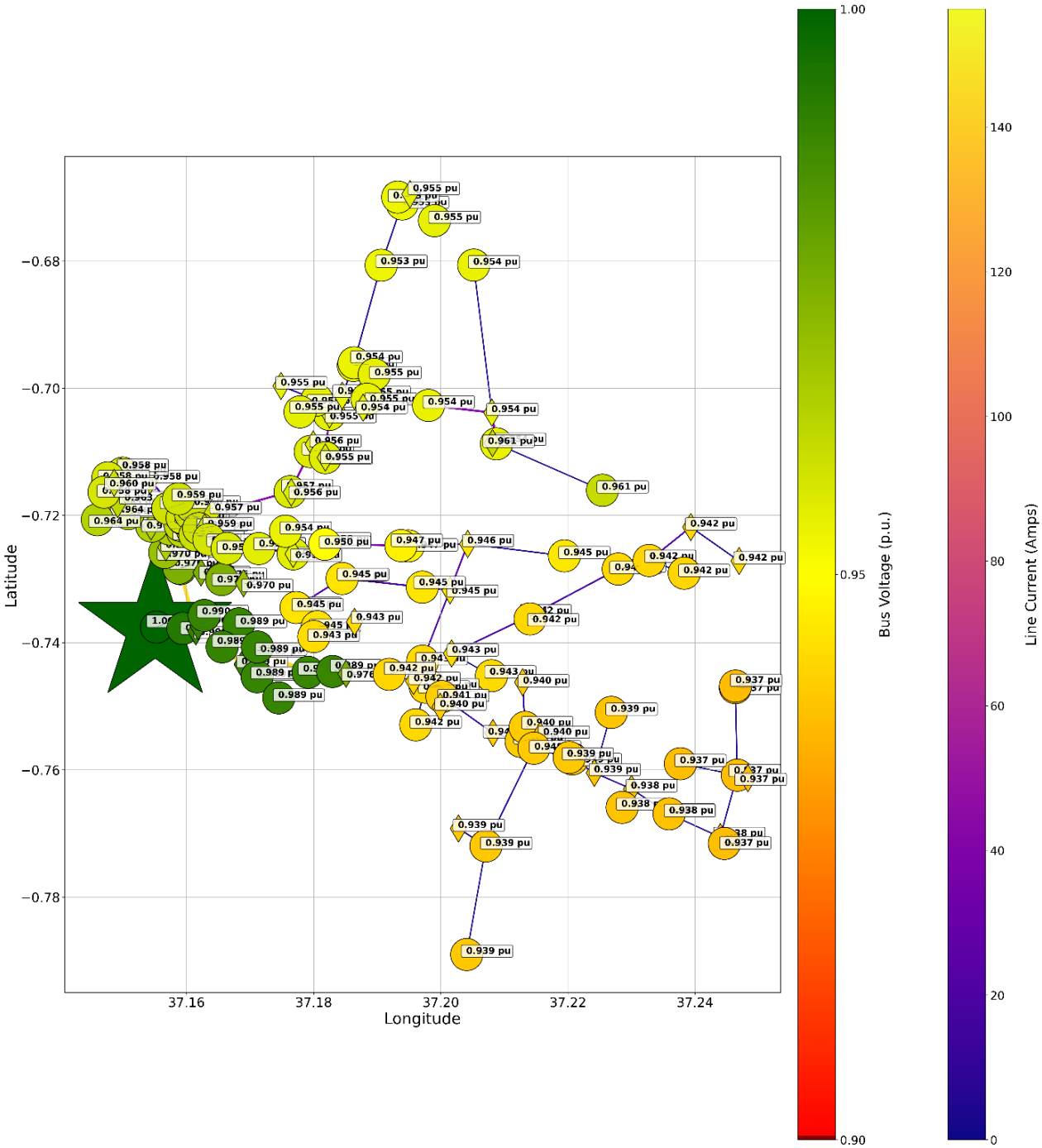


**Figure 10** Geographical visualization of power flow results. Bus colors represent voltage and line color and thickness represent current loading

### 4.4 Cross-platform validation and fidelity

To rigorously validate the accuracy and platform-independence of the automatically generated model, two separate simulation files, a MATPOWER .m file and an OpenDSS .dss script, were generated from the same canonical model. Both were executed in their respective native environments (PYPOWER and dss-python). The successful convergence of both solvers (3 iterations for MATPOWER, 3 for OpenDSS) immediately indicated the robustness of the generated model. Results were further cross-validated against the commercial solver DIgSILENT PowerFactory, showing excellent numerical agreement.

As detailed in Table 8, a direct comparison of the key simulation outputs reveals excellent numerical agreement between the two open-source platforms and the commercial solver. The critical finding is that the location and magnitude of the minimum medium-voltage are nearly identical, with all solvers identifying Bus 158 as the weakest point with a voltage difference of less than 0.01%. Other key metrics, such as total system losses and maximum line current, show less than 3% and 1.1% variance respectively, which is well within expected tolerance for different solver algorithms.

The consistency is further confirmed by the voltage profile plots (Figures. 11 and 12), which show a virtually identical voltage drop characteristic for both the medium-voltage and full medium-voltage/low-voltage systems across both simulations. This successful cross-platform validation robustly demonstrates that the automated framework accurately regenerates an electrically equivalent model whose solution is not dependent on a single simulation engine, confirming the model's high fidelity.

**Table 8** Comparison of key power flow results between the framework's generated MATPOWER, OpenDSS models and cross-platform validation using a commercial solver

| Metric | MATPOWER | OpenDSS | DIgSILENT | Analysis |
|---|---|---|---|---|
| Maximum MV Voltage | 1.0000 p.u. (at Bus 1) | 0.9992 p.u. (at Bus 1) | 1.01 | Both confirm the source bus voltage is correct. |
| Minimum MV Voltage | 0.937 p.u. (at Bus 158) | 0.9376 p.u. (at Bus 158) | 0.936 p.u. | Critical feeder voltage is nearly identical; difference is < 0.1%. |
| Minimum LV Voltage | 0.9271 p.u. (at Bus 252) | 0.9284 p.u. (at Bus 252) | 0.92 | The difference is 0.0013 p.u. (0.13%). |
| Total System Losses | 0.114 MW (4.05% of generation) | 0.1108 MW (3.98% of injection) | 4.05% | < 3% difference, well within expected modelling variance. |
| Total Power Supplied | 2.814 MW (Generation) | 2.7855 MW (Injected) | 2.809 MW | < 1% difference; consistent system-wide power flow modelling. |
| Maximum Line Current | 156.31 Amps | 154.61 Amps | | < 1.1% difference on the most heavily loaded line. |
| Solver Convergence | 3 iterations | 3 iterations | 2 iterations | Both solvers show rapid convergence, indicating model robustness. |

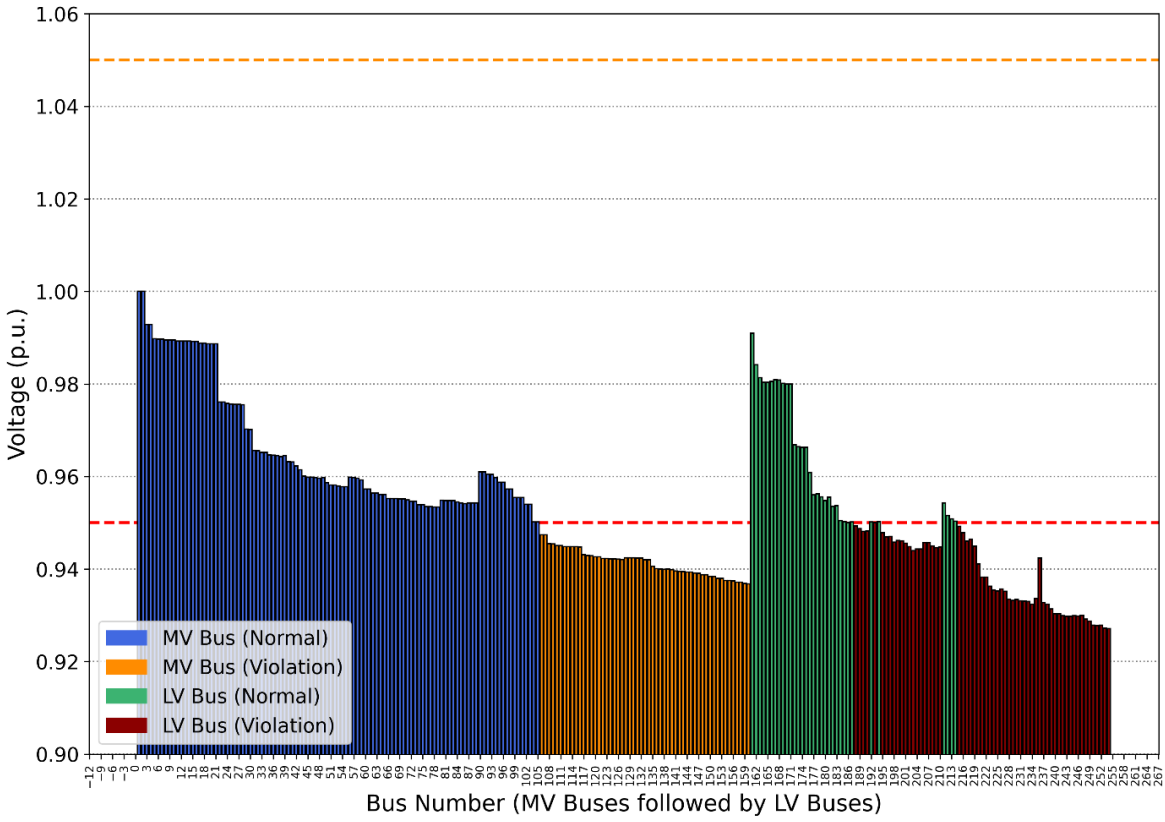


**Figure 11** Medium-voltage and low-voltage voltage profiles from MATPOWER simulations

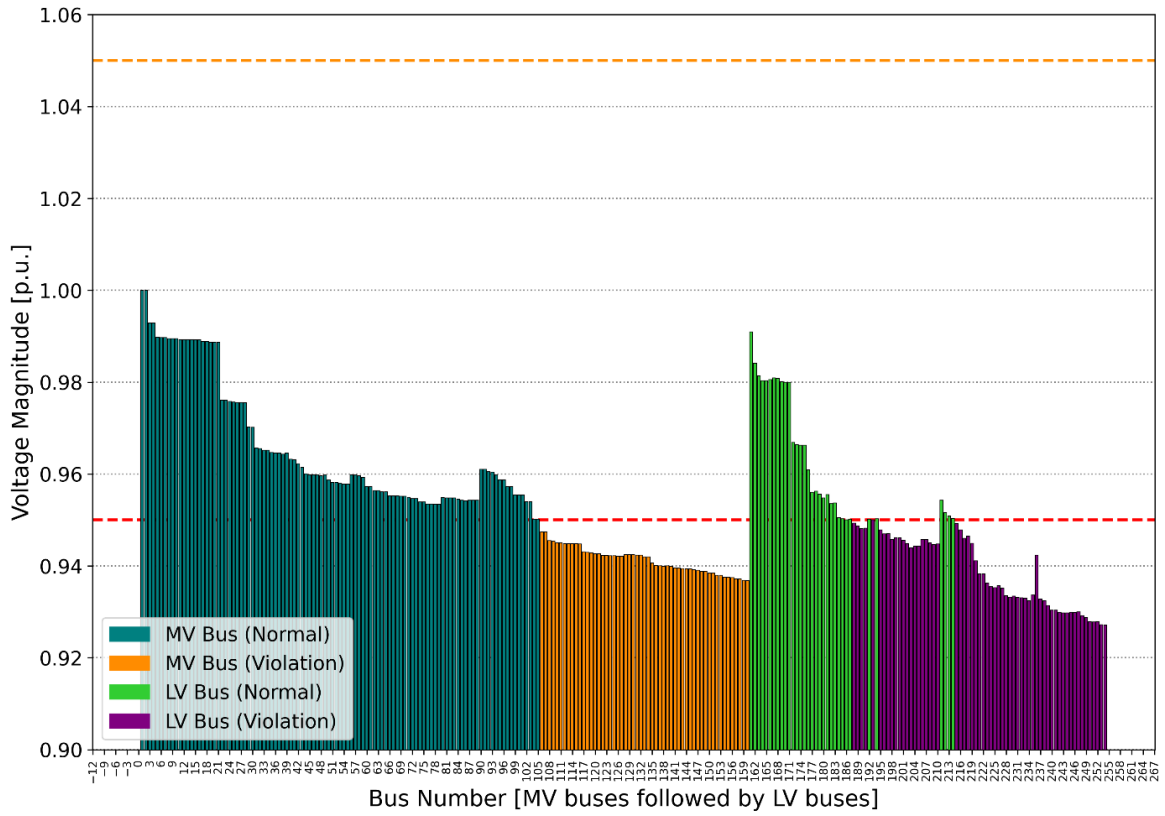


**Figure 12** Medium-voltage and low-voltage profiles from OpenDSS simulations

## 5. Conclusions

This research successfully designed, implemented and validated a novel Python-based framework for power system analysis. The framework automates the conversion of unstructured KMZ geospatial data into high-fidelity, multi-platform mathematical models for computation. It employs a systematic pipeline that integrates metadata-driven filtering, algorithmic topological healing, automated parameterization and intuitive data visualization. This approach effectively bridges the critical semantic gap between static GIS asset data and dynamic engineering simulation. The methodology's robustness and scalability were demonstrated on a complex, real-world 11 kV, 158-bus distribution feeder. The framework autonomously processed the raw data, healing 47 topological disconnections to generate a valid network model. The model's high fidelity was rigorously confirmed through cross-platform validation. Power flow simulations converging successfully across MATPOWER/PYPOWER, OpenDSS and DIgSILENT, all showing excellent numerical agreement. Crucially, the analysis provided a direct, data-driven diagnostic of the network's health, identifying voltage behaviours under peak load conditions. This work demonstrates that an automated framework can serve as both a modelling tool and a powerful diagnostic instrument. It provides a feasible, low-cost and scalable pathway to enhance data-driven network diagnostics and accelerate grid modernization by transforming previously unusable GIS data into actionable insights. The key recommendation is for utilities to adopt such automated techniques to unlock the significant analytical value hidden within their existing data archives.

### Declaration of Ethical Standards

As the authors of this study, we declare that we comply with all ethical standards.

### Credit Authorship Contribution Statement

Amon Okemo: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Visualization, Validation, Writing - Original Draft.

Christopher Maina Muriithi**:** Conceptualization, Software, Supervision, Project Administration, Validation, Writing - Review & Editing.

John Nderu: Methodology, Supervision, Validation, Visualization, Writing - Review & Editing.

### Declaration of Competing Interest

The authors declared that they have no conflict of interest.

### Funding / Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

### Data Availability

The raw geospatial data (KMZ files) for the distribution feeder used in this study were obtained from a utility provider under a confidentiality agreement and are not publicly available. However, the Python automation framework and the derived canonical model parameters are available from the corresponding author upon reasonable request.

## References

Abeysinghe, S., Abeysekera, M., Wu, J., & Sooriyabandara, M. (2021). Electrical properties of medium voltage electricity distribution networks. *CSEE Journal of Power and Energy Systems*, *7*(3), 497–509. https://doi.org/10.17775/CSEEJPES.2020.01640

AL-Jumaili, A. H. A., Muniyandi, R. C., Hasan, M. K., Paw, J. K. S., & Singh, M. J. (2023). Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations. *Sensors*, *23*(6), 2952. https://doi.org/10.3390/s23062952

Anderson, A., Stephan, E., & McDermott, T. (2022). *Enabling Data Exchange and Data Integration with the Common Information Model: An Introduction for Power Systems Engineers and Application Developers*. Pacific Northwest National Laboratory (PNNL). https://doi.org/10.2172/1922947

De-Jesús-Grullón, R. E., Batista Jorge, R. O., Espinal Serrata, A., Bueno Díaz, J. E., Pichardo Estévez, J. J., & Guerrero-Rodríguez, N. F. (2024). Modeling and Simulation of Distribution Networks with High Renewable Penetration in Open-Source Software: QGIS and OpenDSS. *Energies*, *17*(12), 2925. https://doi.org/10.3390/en17122925

Deka, D., Kekatos, V., & Cavraro, G. (2024). Learning Distribution Grid Topologies: A Tutorial. *IEEE Transactions on Smart Grid*, *15*(1), 999–1013. https://doi.org/10.1109/TSG.2023.3271902

Geth, F., Vanin, M., & Van Hertem, D. (2023). *Data quality challenges in existing distribution network datasets* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2308.00487

Javed, A. H., Nguyen, P. H., Morren, J., & Slootweg, J. G. H. (2021). Review of Operational Challenges and Solutions for DER Integration with Distribution Networks. *2021 56th International Universities Power Engineering Conference (UPEC)*, 1–6. https://doi.org/10.1109/upec50034.2021.9548198

*Kenya Power Strategic Plan 2023/24-2027/28*. (n.d.). Retrieved October 12, 2025, from https://kplc.co.ke/storage/01J1BWDEZDE872C70P1CVD6BC2.pdf

Khanh, N. B., Ngo, P. L., Giap, L. N., An, T. N. T., Tien, T. B., Vinh, T. T., Van Nghia, L., & Trong Dat, T. (2025). A review of open-source energy system modeling tools. *International Journal of Advances in Applied Sciences*, *14*(2), 469. https://doi.org/10.11591/ijaas.v14.i2.pp469-480

Matanov, N., & Nankinsky, P. (2021). Modern aspects of the digitalization in distribution systems. *2021 13th Electrical Engineering Faculty Conference (BulEF)*, 1–7. https://doi.org/10.1109/bulef53491.2021.9690811

Mohd Azmi, K. H., Mohamed Radzi, N. A., Azhar, N. A., Samidi, F. S., Thaqifah Zulkifli, I., & Zainal, A. M. (2022). Active Electric Distribution Network: Applications, Challenges, and Opportunities. *IEEE Access*, *10*, 134655–134689. https://doi.org/10.1109/access.2022.3229328

Montano-Martinez, K., Ma, S., Vittal, V., & Rojas, C. (2024). Automated Correction of GIS Data for Loads and Distributed Energy Resources in Secondary Distribution Networks. *IEEE Transactions on Power Systems*, *39*(1), 1628–1636. https://doi.org/10.1109/TPWRS.2023.3243550

Montenegro, D., Dugan, R., Taylor, J., & McGranaghan, M. (2022). Open-source software projects for advancing the power systems analysis. *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*, 1–6. https://doi.org/10.1109/osmses54027.2022.9768968

Mukherjee, M., Lee, E., Bose, A., Gibson, J., & McDermott, T. E. (2020). A CIM Based Data Integration Framework for Distribution Utilities. *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. https://doi.org/10.1109/pesgm41954.2020.9281658

Patel, P. H., Juthani, J. R., Patel, B., & Vahoniya, D. R. (2025). Quantification and Analysis of Human Errors Across the Life Cycle of Electrical Installations. *Archives of Current Research International*, *25*(1), 113–123. https://doi.org/10.9734/acri/2025/v25i11043

Rahman, M. A., Abdul Maulud, K. N., Saiful Bahri, M. A., Hussain, M. S., Ridzuan Oon, A. O., Suhatdi, S., Che Hashim, C. H., & Mohd, F. A. (2020). Development of GIS Database for Infrastructure Management: Power Distribution Network System. *IOP Conference Series: Earth and Environmental Science*, *540*(1), 012067. https://doi.org/10.1088/1755-1315/540/1/012067

Subasic, M., Ave, G. D., Giuntoli, M., Noglik, P., Knezovic, K., Shchetinin, D., Peterson, W., & Li, W. (2022). Distribution Grid Topology Calibration Based on a Data-Driven Approach. *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 1–5. https://doi.org/10.1109/ISGT-Europe54678.2022.9960588

Tóth, R., Balla, D., & Zichar, M. (2020). ANALYSIS AND OPTIMIZATION OF KML FILES. *INTED Proceedings*, *1*, 4209–4216. https://doi.org/10.21125/inted.2020.1175