# Optimizing the logistics operations of distribution network operators from a multinational electric utility company

Diego D. Almeida[1], Mariana Azevedo[1], Victor Vieira[1], Nelson I. Oliveira[1], Anna G. C. D. R. Rodrigues[1], Leonardo C. T. Bezerra[1,2], Lucas Nunes[3], Thaís A. de Mendonça[3], and Rodrigo Manfredini[3]

[1] Federal University of Rio Grande do Norte (UFRN), Natal, RN, Brazil
{diego.dantas.064,mariana.brito.110,victor.vieira.700}@ufrn.edu.br
{nelson.oliveira,anna.ribeiro,leonardo.bezerra}@ufrn.br
[2] University of Stirling, Stirling, Scotland leonardo.bezerra@stir.ac.uk
[3] Neoenergia, Brazil
{lucas.nunes,thais.alves,rodrigo.manfredini}@neoenergia.com

**Abstract.** Distribution network operators (DNOs) share a significant responsibility regarding the assurance of electrical energy supply quality and continuity. In detail, DNOs are legally required to: (i) address electrical emergency occurrences quickly, especially to restore electricity supply, and; (ii) ensure the efficiency of service concerning commercial occurrences. In this work, we propose a solution to optimize the logistics operations of the Spanish multinational electric utility company Iberdrola. Our work scope is Neoenergia, the Brazilian subsidiary controlling five different DNOs. In our work, we follow the CRISP-DM data science framework to address the allocation of operations bases. The solution was developed and successfully deployed in collaboration with the analytics team of Neoenergia. In detail, we model the problem as a knapsack and tackle it with an iterated greedy metaheuristic. Results show a decrease in the distances between bases and occurrences when compared to the current approach adopted by Neoenergia. Our approach also reduces travel times, contributes to the improvement of supply continuity indices, and better meets company business requirements. Importantly, we provide a simulation tool to recommend future base allocation, which comprises valuable input to planning.

**Keywords:** electrical energy· heuristic optimization· data science

## 1 Introduction

Electrical energy is one of the most critical resources for the social and economical development of a country, with clean and affordable energy being one of the United Nations sustainable development goals. In this context, countries and companies have heavily invested in both renewable energy sources and in optimizing costs associated with the generation, transmission, and distribution of energy. Iberdrola, the Spanish multinational electric utility company, has research and innovation initiatives targeting these issues. The Brazilian subsidiary of Iberdrola in Brazil, Neoenergia, controls five distribution network operators (DNOs) accross the country, and also operates in generation, transmission, and commercialization of electrical energy. Altogether, Neoenergia DNOs total over 16 million clients and over 37 million consumers.

Electrical energy distribution in Brazil is regulated by the National Agency for Electrical Energy (ANEEL), which establishes policies for product, service, and commercial quality, including supply continuity indicators that are used to evaluate the performance of distribution companies. Importantly, ANEEL regulates compensation for consumers under certain circumstances, e.g. when indicators fall below pre-determined thresholds or when deadlines are not met. The Neoenergia Cosern DNO, for instance, paid over $1.5 million in consumer compensation in 2022, despite having ranked second nation-wide in terms of energy supply continuity. From 2018 to 2023, Neoenergia DNOs have paid a combined yearly average of $22 million in consumer compensation.

To improve its operational efficiency and mitigate the compensation loss, Neoenergia has been actively collaborating with Brazilian universities for research and innovation. Among the most pressing logistics operations tasks that Neoenergia needs to address is the allocation of operational bases. From a general perspective, managing operational bases is a complex optimization problem that involves (i) determining how many and where the bases will be deployed; (ii) how many teams will be available in each base and their roster, and; (iii) planning daily occurrence routes for each team to address. The analytics team of Neoenergia addresses these problems separately, and in this work we follow the same approach and focus on (i). A further business requirement concerns distribution territorial units (DTUs). In detail, each of Neoenergia DNOs splits the area under its coverage into DTUs, and occurrences within a given DTU must be addressed by an operational base from that DTU. In this context, the underlying optimization problem is to allocate a pre-determined number of operational bases to locations within DTUs such that each DTU has at least one operational base and the total time taken to address occurrences is minimized. Two additional business requirements are that (i) locations selected to host an operational base present a good overall infrastructure to support the base daily activities, and; (ii) occurrences that affect more clients (or clients with higher business priority) be addressed more quickly, as these potentially affect continuity indicators more.

In this work, we follow the CRISP-DM data science framework [15] to address the allocation of operational bases at Neoenergia DNOs. Initially, we identified that the company's most recent model is an unsupervised learning (UL) approach proposed for the allocation of operational bases at the Neoenergia Brasília DNO. Given that all other Neoenergia DNOs operate at a state-level rather than at a city-level, this could heavily affect model assumptions and performance. Next, we propose a general knapsack modeling [12] suitable for all of Neonergia DNOs. By modeling operational bases allocation as a knapsack, we are able to meet all the business requirement set by Neoenergia, including DTU-adherence, location infrastructure, and client priority. Finally, we propose an iterated greedy (IG) algorithm [11] to optimize the knapsack problem. Importantly, because the IG algorithm can use different parameter configurations, we are able to generate a number of alternative allocations that are aggregated as recommendations of locations that would be preferable for hosting an operational base. We empirically demonstrate the benefits of our approach using Neoenergia Cosern DNO data, identifying base allocations that would reduce the median time to address occurrences by 18.53% w.r.t. the UL approach.

The remainder of this work is organized according to the CRISP-DM framework, as follows. Section 2 further deepens business understanding and briefly discusses related work funded by ANEEL in Brazil. Next, Section 3 details the data made available by Neoenergia, and provides an exploratory analysis that is instrumental for the modeling proposed in Section 4. The evaluation of the proposed model is given in Section 5 and the development and deployment of the recommendations dashboard is detailed in Section 6. Finally, we conclude and discuss future work in Section 7.

## 2 Business context

Electrical energy distribution is often a monopolized market, demanding indirect incentives for companies to invest in their service. ANEEL adopts a *price-cap* model, setting an initial level for fees that is increased periodically according to a pre-selected official consumer price index. A percentage of that increase is deducted by ANEEL based on whether expected efficiency and productivity gains were reached. As such, distribution companies can only increase their profits by optimizing their operating costs [6, 5]. In addition, ANEEL regulates consumer compensation based on supply continuity indicators, so preventing and solving electrical emergency interruptions becomes critical. In this section, we initially define the main indicators of interest to our work, and then discuss the administrative organization of the Neoenergia Cosern DNO, which is illustrative of the remaining Neoenergia DNOs. Finally, we detail business requirements and overview ANEEL-funded related work in Brazil.

### 2.1 Supply quality indicators

To monitor supply continuity, ANEEL establishes two collective and two individual quality indicators. Our focus is on the former, namely the (i) *equivalent interruption duration per consumer unit* (DEC) and; (ii) *equivalent interruption frequency per consumer unit* (FEC). Each indicator has pre-defined thresholds that companies must not violate over a given period of time for any set of consumer units. If a violation occurs, consumer units are automatically compensated via their utility bill. However, interruptions that last up to three minutes are not included for violation assessment. For this reason, the total time taken to solve an interruption is monitored by ANEEL and companies alike through measures of the individual stages of the electrical emergency occurrences:

**Average preparation time (APT):** average time required for teams to be dispatched to solve electrical emergency occurrences.
**Average travel time (ATT):** average time required between teams dispatch and arrival at electrical emergency occurrence locations.
**Average execution time (AET):** average time required to solve electrical emergency occurrences once teams are on-site.
**Average emergency resolution time (AERT):** average time required for teams to solve electrical emergency occurrences (includes preparation, travel, and execution times).

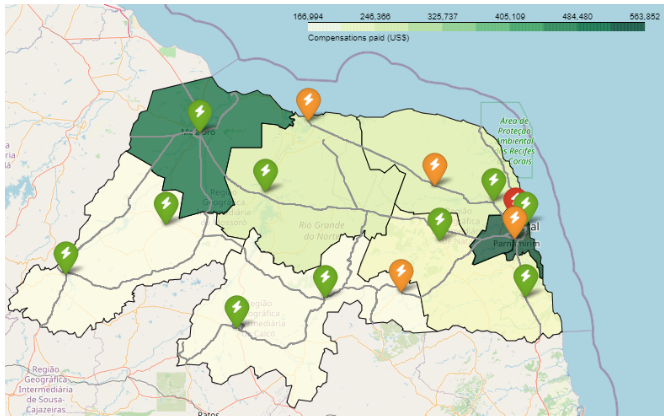


**Fig. 1.** Distribution territorial units (DTUs) spatial discretization adopted by Neoenergia Cosern. Darker DTU colors mean more consumer compensation. The color of an operational base (marker) depicts set up date: up to 2019 (green); between 2020 and 2022 (orange), or; from 2023 on (red).

### 2.2 Administrative organization

As discussed above, ANEEL defines supply continuity indicators based on sets of consumer units. In turn, Neoenergia DNOs further group those sets into distribution territorial units (DTUs). For Neoenergia Cosern, the 56 sets of consumer units are grouped into eight DTUs, illustrated in Figure 1 with DTU geographical limits and markers indicating the locations of operations bases. Details for each DTU are given as supplementary material, for brevity. As a business requirement, occurrences within a given DTU need to be addressed by an operational base from the same DTU. In addition, occurrences can differ as to being (i) *emergency*, meaning they are not planned for and must be addressed as quickly as possible in the case of an interruption, and; (ii) *technical-commercial*, which are planned interventions that are also regulated by ANEEL. Both emergency and technical-commercial occurrences can lead to consumer compensation, as follows. Regarding the former, compensation is due when supply interruptions last longer than three minutes, as previously discussed. Concerning the latter, each type of intervention has a prescribed deadline, and failing to meet the deadline incurs on a corresponding consumer compensation fee.

### 2.3 Business requirements

The solution in use at Neoenergia for the allocation of operational bases was originally developed for the Neoenergia Brasília DNO. As previously discussed, the solution models base allocation as an unsupervised learning (UL) problem. In detail, the target geographic area is considered to be a bi-dimensional Euclidean space where the $x$ and $y$ coordinates correspond to latitude and longitude. In this approach, the analyst inputs the number of DTUs and the occurrence data, and the centroids of the clusters identified by the UL algorithm become the suggested base locations. Currently, Neoenergia uses the $k$-means algorithm, so centroids are computed based on the Euclidean distance between occurrences.

From a business perspective, the UL approach presents drawbacks that Neoenergia set as requirements to be addressed.

1. **Existing bases:** do not move existing operational bases given the economical overhead for setting up a base;
2. **Occurrence relevance:** supply quality indicators are more heavily influenced by occurrences that affect a larger number of clients or demand;
3. **DTU adherence:** occurrences from a given DTU need to be addressed by operational bases located within the DTU;
4. **Flexible number of bases:** the number of operational bases should be set independently of the number of existing DTUs;
5. **Optimized DTU allocation:** when more bases than DTUs are given, the allocation of bases to DTUs should optimize supply quality indicators;
6. **Road infrastructure:** allocation should consider real-world routes and distances, and federal routes should be preferred due to their infrastructure;
7. **Location infrastructure:** bases should be allocated to locations that correspond to cities, which need to present a minimally viable infrastructure.

### 2.4 Related work

ANEEL funds projects related to electrical energy in Brazil. These are traditionally industry-academia collaborations, and the data about the projects conducted between 2008 and 2023 is publicly available [1]. Nearly half of the projects funded by ANEEL in the period concern electrical energy distribution, and the two most recurring themes are the (i) supervision, control, and protection of electrical energy systems, and; (ii) alternative sources for electrical energy generation. Most of the projects concern applied research and over recent years have often delivered information technology solutions. However, seldom works involve optimization and/or artificial intelligence.

We then conducted a systematic review of the publications associated with the projects funded by ANEEL, to identify works related to ours that also fit an applied research industry-academia collaboration. Among the projects that employ computational intelligence techniques, the works based on machine, deep, or statistical learning usually focus on modeling continuity indicators [8, 7], forecasting energy demand [4], or fraud detection [10]. By contrast, works that employ metaheuristic algorithms tackle a range of different optimization problems, such as the allocation of remote control switches [9, 13], maintenance planning [14], or consumer set gerrymandering [2]. Regarding metaheuristic techniques, the techniques employed range from variable neighborhood search [9] to artificial immune systems [14], but evolutionary algorithms are the most recurring algorithm [2, 13]. To our knowledge, ANEEL has not yet funded an industry-academia collaboration addressing the allocation of operational bases.

## 3 Data understanding

Logistics operations at Neoenergia uses multiple information systems (IS). In addition, the different companies that comprise the Neoenergia economical group were acquired over the years, and hence use legacy IS that are being gradually replaced. For the context of emergency occurrences, Neoenergia developed a system dubbed GSE, built on ArcGIS, which has already been deployed by

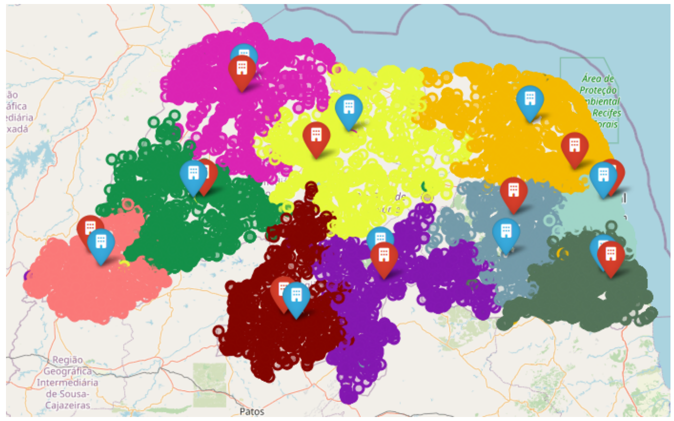


**Fig. 2.** Spatial distribution of the electrical emergency occurrences, colored by the addressing 2019 operational bases (red markers). Blue markers: UL approach allocation.

most of the Neoenergia DNOs. In turn, technical-commercial occurrences are managed by a commercial software that has also been deployed by most DNOs. Given the similarity in IS across DNOs, in this work we use data provided by the Neoenergia Cosern DNO as representative of the remaining DNOs. Importantly, Neoenergia Cosern operates at a state-level, similarly to all other Neoenergia DNOs – the only exception is Neoenergia Brasília, as discussed.

In this section, we discuss insights we obtain from an exploratory analysis of the data using the currently employed unsupervised learning (UL) approach devised by Neoenergia Brasília. Details on data cleaning and enrichment stages are given as supplementary material. Figure 2 illustrates the geographical distribution of electrical emergency occurrences, where color indicates the operational base that addressed a given occurrence. Notice that in very few situations there is a mix in colors, confirming that occurrences are mostly addressed by teams from the nearest operational base. In addition, we observe that some operational bases appear considerably shifted from the centroids of the occurrences they have addressed. This insight is what potentially motivated the analytics team at Neoenergia to propose an UL approach to operational base allocation.

An important remark about the data provided by Neoenergia concerns the effect of the COVID-19 pandemic. In detail, the digital transformation rushed by the pandemic meant that the efforts of the IS sectors were concentrated on transitioning the company from in-place to remote working environments. As a result, some IS such as GSE became inconsistent with developments occurred during the 2020-2021 period, most notably the addition of novel operational bases. In this context, we remark that the bases given in Figure 1 in orange and red were not registered in the occurrences datasets provided. Hence, we differentiate the three time periods depicted when assessing the data below.

Prior to modeling, we conduct an exploratory analysis of the current operational bases allocation of the Neoenergia Cosern DNO. To do so, we employ the Neoenergia Brasília UL solution, which also evidences the limitations that the business requirements provided by Neoenergia seek to address. For instance, Euclidian space representation violates the (6) road infrastructure requirement by definition. The violations of other requirements are detailed below, comprising four different scenarios. For brevity, most of the allocations are illustrated in the supplementary material.

**(1) Existing bases.** Figure 2 compares the 2019 operational base allocation (red markers) with the UL allocation (blue markers). Most of the 10 then existing operational bases are relatively near the cluster centroids computed from the 2019-2022 data, with the exception of three. Importantly, two of the three alternative centroids match the operational bases that Neoenergia Cosern set up between 2019-2022, and are justified by the absence of operational bases in the north coast up to 2019. Altogether, these insights help validate the planning currently adopted by Neoenergia Cosern for the allocation of their operational bases. Importantly, we use this scenario to evidence the importance of business requirement (1) *existing bases*, as this allocation incurs in base relocation.

**(2) Occurrence relevance.** When weighted UL is employed, many significant changes to the 2019 allocation are observed. Two of the most important concern: (i) an operational base at the north of the state would be replaced by two novel bases, and (ii) two operational bases at the south of the state would be replaced by a single novel base. These insights evidence that the existing base allocation could be improved if occurrence importance were factored, but the clustering approach provides this at the cost of business requirement (1) *existing bases.*

**(3) DTU adherence.** We compare the 2022 allocation with a DTU-adherent UL allocation. To do so, we split the original dataset into DTU-exclusive datasets, and run the UL approach taking each DTU-exclusive dataset and their 2022 existing number of operational bases. The centroids identified match most of the existing operational bases, once again validating the allocation planning of Neoenergia Cosern. However, one of the centroids is located in a region where no cities are eligible to host an operational bases, violating business requirement (7) *location infrastructure.* Furthermore, requiring a predefined number of bases per DTU violates business requirement (5) *optimized DTU allocation.*

**(4) Flexible number of bases.** When the number of clusters exceeds the number of existing operational bases, the centroids differ considerably from the 2022 allocation. Among the most relevant modifications are (i) replacing three operational bases with a single, novel base; (ii) replacing an existing base with two novel bases, which the model recommends in three separate regions. Altogether, this allocation further reinforces the difficulty to meet business requirement (i) *existing bases* when using an UL approach.

## 4 Data preparation and modeling

The operational bases allocation problem addressed in this work can be naturally modeled as an optimization problem. Depending on scale, the problem can be addressed through exact methods or metaheuristics. In this work, we opt for the second option as the different Neoenergia DNOs present different numbers of operational bases and locations that would be eligible to host them. More importantly, the stochastic nature of metaheuristics enable the analysis of alternative scenarios that can be aggregated as recommendations for stakeholders. In this section, we initially discuss how we use data preparation to meet part of the business requirements from Neoenergia. Later, we detail our proposed problem formulation, solution representation, and algorithmic approach.

### 4.1 Data preparation

To complement the datasets provided by Neoenergia, we additionally engineer a real-world distance matrix to meet four business requirements, namely (2) *occurrence relevance*; (3) *DTU adherence*; (6) *road infrastructure*, and; (7) *location infrastructure*. Below, we detail the three-stage enrichment procedure.

**Location infrastructure.** We identify cities with at least 10 000 residents in 2023 according to official projections. This threshold was selected given the demographics of the cities currently hosting Neoenergia Cosern bases. In total, 74 cities are eligible to host an operational bases according to this criterion.

**Road infrastructure.** We preprocess real-world distances between occurrences and eligible cities using the osmnx [3] real-world routing Python library. For feasibility, we spatially-aggregated occurrences prior to computing real-world distances, as follows. First, we associated each occurrence $o_i$ to its nearest city $c_j$. For each city $c_j$, we then computed the centroid of its associated occurrences $o_i$ using their latitude and longitude information. In practice, centroids roughly matched city centers. Finally, we computed real-world distances between the 74 eligible cities $e_k$ and all 167 cities $c_j$ that Neoenergia Cosern is responsible for, resulting on a 167x74 real-world distance matrix. Effectively, this matrix enables us to identify for each city $c_j$ the nearest eligible city $c_k$ in a given allocation. We further enriched this matrix with three columns computed from the occurrences $o_i$ associated with city $c_j$, namely (i) $n_j$, the number of occurrences; (ii) $r_j$, the number of clients affected by the occurrences (a proxy for **occurrence relevance**), and; (iii) $m_j$, the median distance between city $c_j$ and the occurrences.

**DTU adherence.** We set distance matrix entries $d_{jk} = \infty$ for every combination of city $j$ and eligible city $k$ that do not belong to the same DTU, effectively discarding them from the analysis.

### 4.2 Modeling

We model the operational bases allocation problem as a knapsack problem, where each eligible city is considered a unit-weight item and the knapsack capacity equals the total number of bases to be allocated. For a given eligible city $e_k$, item profit is computed as a function of the cities that a base located in $e_k$ would address. As a result, item profit is dynamic, since the addition or removal of a base from the knapsack affect city-base association. Next, we detail the objective function, our iterated greedy algorithmic approach, and the heuristic function used by the greedy constructive procedure.

**Objective function.** We compute the objective function using an auxiliary distance matrix, as follows. Given each city $c_j$, the objective function first computes city cost $t_j = n_j \times (m_j + d_{jk})$, where $k$ is the index of the eligible city $e_k \in S$ nearest to city $c_j$. Effectively, $t_j$ is a proxy for the total distance a team departing from eligible city $e_k$ would travel to address all occurrences from city $c_j$. In addition, we remark that $n_j$ may be replaced by $r_j$ to meet (2) *occurrence relevance*. The next stage of the objective function is to compute the cost $z_k$ of each eligible city $e_k \in S$, which equals $z_k = \sum t_j$, for all cities $c_j$ associated with

**Algorithm 1** Iterated greedy with tabu list pseudocode

**Require:** $B$: existing base set; $E$: eligible city set; $n$: number of bases to be allocated; $d$: ruin degree; $t$: tabu tenure; $\mathcal{R}$: relaxed acceptance criterion

1: $T \leftarrow \emptyset$ ▷ tabu list initialization
2: $S \leftarrow generateSolution(n, B, E, T)$ ▷ incumbent solution creation
3: $S^* \leftarrow S$ ▷ best-so-far solution
4: $E' \leftarrow E \setminus S$ ▷ working set of eligible cities
5: **repeat** ▷ ruin-and-recreate loop
6:     **if** *current iteration* $> t$ **then**
7:         $E' \leftarrow E \cup T.pop(d)$ ▷ oldest tabu city becomes eligible
8:     $S' \leftarrow S$ ▷ candidate solution
9:     $T.append(S'.popRandom(d))$ ▷ solution ruin and tabu list update
10:     $S' \leftarrow generateSolution(d, B, E', T, S')$ ▷ solution recreation
11:     $E' \leftarrow E' \setminus S'$ ▷ update working set of eligible cities
12:     **if** $f(S') < f(S')$ **then**
13:         $S \leftarrow S'$ ▷ update incumbent solution
14:         **if** $f(S') < f(S^*)$ **then**
15:             $S^* \leftarrow S$ ▷ update best-so-far solution
16:     **else if** $\mathcal{R}(S', S)$ **then**
17:         $S \leftarrow S'$ ▷ update incumbent solution
18: **until** *stopping criterion met*

**Ensure:** $S_b$

$e_k$. Finally, the objective function computes the total cost $f(S) = \sum z_k$, for all of eligible cities $e_k \in S$.

**Algorithmic approach.** We adopt an iterated greedy coupled with a tabu list approach including a relaxed acceptance criterion, depicted in Algorithm 1. The algorithm requires the (i) existing base set $B$; (ii) eligible city set $E$; (iii) number of bases to be allocated $n$; (iv) ruin degree $d$; (v) tabu tenure $t$, and; (vi) relaxed acceptance criterion $r$, to accept a tentative solution $S'$ even if it is not better than the incumbent solution $S$. Notice that argument (iii) ensures business requirement (4) *flexible number of bases*, which combined with data preprocessing DTU-adherence also ensures (5) *optimized DTU allocation*.

The algorithm first initializes the tabu list $T$ (Line 1) and generates an incumbent solution by adding the existing base set and filling the best remaining spots with the best eligible cities from $E$ (Line 2), which is then recorded as the best-so-far solution (Line 3). Next, a working eligible city set $E'$ is created by excluding cities in $S$ (Line 4). The ruin-and-recreate loop is given in Lines 5–18. Initially, Lines 6–7 update the working set of eligible cities and tabu list once the number of iterations is greater than the tabu tenure $t$. Next, a candidate solution $S'$ is created from the incumbent solution $S$ (Line 8), and the ruin procedure removes $d$ random eligible cities $e_k$ from $S'$, which become tabu (Line 9). Recreation is given in Line 10, when procedure *generateSolution* adds to $S'$ the best $d$ available eligible cities $e_k \in E' \setminus T$. The cities added to the solution are then removed from the working eligible city set $E'$ (Line 11). Finally, the candidate solution $S'$ may replace incumbent (Lines 12–13) and best-so-far solutions (Lines 14–15). $S'$ may also replace $S$ based on the relaxed acceptance criterion $\mathcal{R}$ (Lines 16–17).

The greedy procedure *generateSolution* either generates a solution from scratch or completes a partial solution, depending on the number of arguments provided. The procedure requires four arguments, namely (i) remaining solution capacity, (ii) existing base set; (iii) a set of eligible cities (excluding existing bases), and; (iv) a tabu list. When no additional argument is provided, the algorithm creates a solution from scratch, initially populating the solution with the existing bases to respect business requirement (1) *existing bases*. Next, eligible cities are greedily selected to fill the remaining solution capacity. Conversely, if a partial solution is provided as fourth argument, procedure *generateSolution* does not use the existing base set and only greedily fills the remaining solution capacity. In both cases, cities present in the tabu list cannot be selected.

**Heuristic function.** The heuristic value of a given eligible city $e_k$ is computed in two steps, as follows. First, we compute for each city $c_j$ its value $v_j = n_j \cdot m_j$, where $n_j$ and $m_j$ are respectively the precomputed number of occurrences associated with city $c_j$ and the median distance between those occurrences and city $c_j$. For a weighted analysis, the number of clients $r_j$ replaces $n_j$. Finally, the heuristic value $h_k$ of a given eligible city $e_k$ equals $h_k = \sum v_j$ for the 10 cities $c_j$ that are nearest to $e_k$.

## 5 Evaluation

The proof-of-concept evaluation conducted in this section concerns two major perspectives. First, we evaluate the heuristic optimization (HO) approach for travel distance, as a proxy to benefits to supply quality indicators. Since the dataset provided by Neoenergia does not include all the existing operational bases for the period considered, we take the unsupervised learning (UL) approach from the Neoenergia Brasília DNO as baseline. Later, we discuss proposed allocations in light of the business requirements set by Neoenergia.

### 5.1 Travel distance

We assess illustrative 15-base allocations adopting the parameter configuration given in the supplementary material for brevity. In detail, the allocations considers the 2019 existing operational bases given in Figure 1, and five new bases must be allocated. Results are given in Figure 3, which also gives baseline results from: (i) the 2019 existing bases, and; (ii) the different UL scenarios discussed in Section 3. However, UL scenarios with $n = 10$ use instead $n = 14$ to increase comparability with HO allocations. The real-world distances for the different allocations are given as boxplots (clipped at 150km).

Overall, no statistical significant difference is observed between approaches, given the overlap in boxplots. However, compared to the 15-base UL allocation, the HO unweighted allocation presents lesser first and second quartile, equivalent average, greater third quartile, and a longer tail. Two insights indicate that this is a consequence of having to respect DTUs, namely (i) a similar difference in distributions between the two 14-base unweighted UL allocations, and; (ii) the similarity between the boxplot of the HO allocation and the boxplot of the 14-base DTU-adherent UL allocation. Regarding (ii), Table 1 shows that the median and mean distances for the HO approach are, respectively, 15.4% and 2.24% lower than for the UL approach, likely a result of having an additional base. Nonetheless, when both 15-base allocations are compared, the median distance for the HO allocation is 18.53% lower than for the UL allocation. Finally, we remark the the addition of five bases by the HO allocation respectively improves the mean and median distances of the 2019 existing bases by 16.27% and 28.76%.

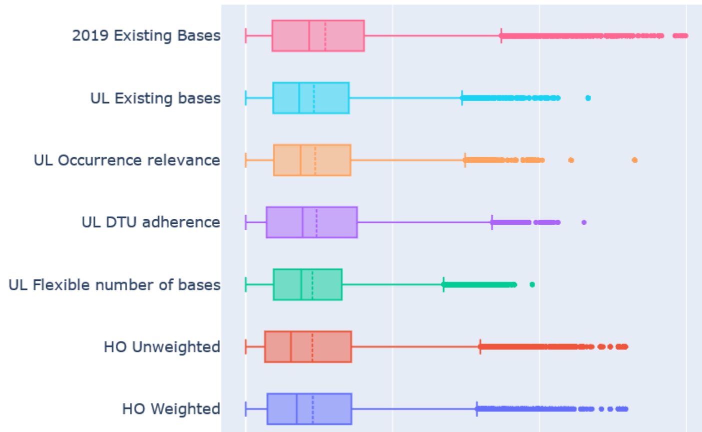


**Fig. 3.** Boxplot of real-world travel distances (in km) using different operational base allocations. Dashed lines depict averages; only data points lesser than 150km are given.

**Table 1.** Real-world distances (km) for different allocations.

| Allocation | Bases | Mean | Median |
|---|---|---|---|
| 2019 Existing bases | 10 | 27.05 | 21.59 |
| UL Existing bases | 14 | 23.17 | **18.18** |
| UL Occurrence relevance | 14 | 23.62 | 18.67 |
| UL DTU adhrence | 14 | 24.09 | 19.36 |
| UL Flexible number of bases | 15 | **22.68** | 18.88 |
| HO Unweighted | 15 | **22.65** | **15.38** |
| HO Weighted | 15 | 22.82 | 17.36 |

When we compare the weighted and unweighted HO allocations, we notice from Figure 3 that the former presents quartiles that are slightly greater than the latter. Indeed, Table 1 shows that means are very much identical between the HO allocations, whereas the median distance for the weighted allocation is 12.87% worse. These results follow a similar pattern observed between UL 14-base weighted and unweighted allocations, with quartiles for the former between slightly greater than for the latter.

### 5.2 Discussion

We conclude the evaluation discussing the illustrative operational bases allocations previously proposed. Figure 4 depicts the allocations for the unweighted (top) and weighted (bottom) analyses.

**Unweighted analysis.** The proposed allocation partially validates the planning of Neoenergia Cosern, as three out of the five recommended bases either match or strongly approximate bases that were deployed after 2019. By contrast, the remaining two bases greatly differ from the ones that were set up by the company. Interestingly, both suggestions favor the coast over the countryside, either becoming closer to the northeastern or the southeastern coast. Importantly, the former suggestion matches the deployment planned by Neoenergia.

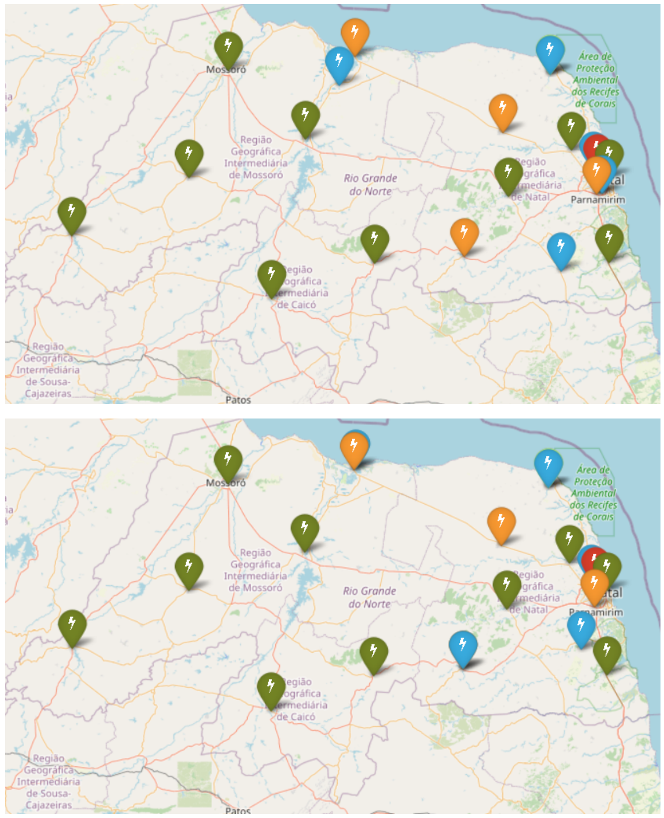


**Fig. 4.** Unweighted (top) and weighted (bottom) allocations proposed by the HO approach. Markers are colored as a function status. Green: existing at 2019; orange: deployed between 2020-2022; red: deployed in 2023, and; blue: recommended.

**Weighted analysis.** When occurrence relevance is factored in, the planning of Neoenergia Cosern is further validated. In detail, three of the suggested bases exactly match the bases that have been set up since 2019. Furthermore, the northeastern coast base under consideration by Neoenergia is once again recommended by the HO approach. The most significant difference between the weighted and unweighted assessment comprises the southeastern base suggestion, which becomes closer to the capital city (located at the middle of the eastern coast) to account for occurrence relevance.

## 6 Deployment

The analytics team at Neoenergia develops dashboards and elaborates reports for the different DNOs with recommendations about alternative scenarios they may consider when planning their operational bases allocations. To meet this need, we implemented and deployed simulation and recommendation dashboards not disclosed for confidentiality. In this section, we give a high-level description of the dashboards, discuss the insights from recommendation, and detail the benefits to Neoenergia DNOs since deployment.

### 6.1 Dashboards

The simulation dashboard depicts plots of an HO approach run taking as input: (i) an occurrence dataset; (ii) an existing bases dataset; (iii) algorithmic parameters, and; (iv) whether to consider occurrence relevance. In turn, the recommendation dashboard ranks eligible cities based on the frequency with which they appear in different allocations. These allocations are the results obtained from running the 108 different parameter configurations of the HO approach given as supplementary material, for brevity.

Five of the eight most recommended eligible cities in the dashboard are cities where Neoenergia Cosern had set up operational bases after 2019. An additional recommended base was also under consideration for deployment, namely the northeastern coast base given in Figure 4. These results serves as further validation that the planning department of Neoenergia Cosern has conducted sound assessments prior to allocating operational bases. More importantly, they help validate the HO approach modeling that we propose. In complement, the remaining most recommended eligible cities are located in the central and mid-southern regions, which the company had not yet considered, providing the analytics teams with novel insights for their planning.

### 6.2 Monitoring

Since deployment at the end of 2022, Neoenergia has experienced a number of benefits resulting from the dashboards we developed. Below, we discuss the benefits observed for Neoenergia DNOs, starting with Neoenergia Cosern.

**Benefits to Neoenergia Cosern.** In addition to validating the planning of the existing bases adopted by the company, the insights produced in this work provided (i) further evidence for Neoenergia Cosern to set up operational bases that were under analysis, and; (ii) indications of the regions where the addition of novel operation bases would bring most benefits. In 2023, the company set up a novel operational base alongside the southern coast recommended in Figure 4 (right). The supply quality indicator thresholds for those regions were respected in 2023, despite adverse climatic conditions that traditionally trigger an increased number of occurrences.

**Benefits to Neoenergia DNOs.** The dashboards developed in this work are being instrumental for holistic assessments of all five DNOs in the Neoenergia economical group. Operational bases locations are revised periodically based on travel distances and team productivity. In this context, the proposed dashboards are being coupled with existing solutions at Neoenergia, thus extending the benefits to DNOs other than Neoenergia Cosern. In 2023, for instance, the analytics team used the suite of solutions to recommend a novel operational base for the Neoenergia Brasília DNO. Along 2024, the Neoenergia Coelba DNO expanded their operational bases, which address around 6 million clients and an area of circa 570,000km$^2$. Importantly, the same demographic profile adopted for identifying the 74 eligibile cities for Neoenergia Cosern results in 341 eligible cities for Neoenergia Coelba, reinforcing the need for scalable solutions. In this context, data science projects such as the one conducted in this work improve the analytics capacity of Neoenergia and lead to more accurate and optimized planning of the required investment for base allocation.

## 7 Conclusion

Neoenergia, the Brazilian subsidiary of the multinational Iberdrola electrical utility company, is an economical group active in the distribution, generation, transmision, and commercialization energy markets. In total, Neoenergia companies operate in 18 different Brazilian states, as well as in the federal district. Altogether, the five distribution network operators (DNO) of Neoenergia serve around 16.5 million consumer units, equivalent to a population of 40 million people approximately. Given its scale, Neoenergia has paid an yearly average $22 million dollars in consumer compensation in the 2018-2022 period, even with DNOs like Neoenergia Cosern ranking among the best in the country.

In this work, we have addressed operational bases allocation, one of the most pressing logistic operations problems that the analytics and planning teams of Neoenergia and their DNOs attempt to optimize. In detail, for this work we have followed the approach adopted by Neoenergia of addressing allocation independent of roster scheduling and route planning. Nonetheless, the heuristic optimization (HO) modeling we proposed demonstrated its benefits in comparison to unsupervised learning model currently adopted by the company. Besides significant reductions of nearly 20% in median travel times for the Neoenergia Cosern DNO, the proposed modeling also meets all the business requirements set by Neoenergia. Importantly, the solution is general and scalable, being applicable to the context of all five Neoenergia DNOs.

The contributions of this work extend beyond modeling. Concretely, we have developed and deployed simulation and recommendation dashboards to assess planning operational bases allocation. Simulation allows analysts to evaluate custom runs of the HO solution, whereas recommendation aggregates results from several runs using alternative parameter configurations. The ranking of eligible cities provided by the recommendation algorithm is instrumental for Neoenergia DNOs, helping to validate planning regarding (i) existing and (ii) under analysis operational bases. More importantly, recommended eligible cities often provide novel insights that planning and analytics teams had not yet antecipated, as observed in the context of Neoenergia Cosern.

The practical benefits to Neoenergia DNOs already observed since dashboard deployment are encouraging. In detail, two of Neoenergia DNOs have already deployed novel operational bases with the aid of the proposed dashboards, as well as a significant expansion in 2024 for the largest DNO. Importantly, the supply quality indicators for Neoenergia Cosern were positively affected by the implementation of the recommendations provided in this work, even in an year of adverse climatic conditions. Future work will address the different aspects of operational bases management, with the goal of developing a solution that integrates allocation, scheduling, and routing. Critically, the work envisioned will be eligible for research funding from the ANEEL Brazilian regulatory agency, which will enable the development of yet more robust solutions. In addition, this work is also applicable to the context of DNOs from other subsidiaries of Iberdrola, which often promote international knowledge exchange workshops. We then expect to expand the current academia-industry collaboration to other countries, e.g. Scottish Power DNOs in the United Kingdom.

## References


1. Agência Nacional de Energia Elétrica: Projetos de P&D em energia elétrica (2023), https://dadosabertos.aneel.gov.br/dataset/projetos-de-p-d-em-energia-eletrica
2. Araújo, R.J.P.d.: Otimização de desempenho de indicadores de continuidade do serviço em concessionárias de distribuição utilizando algoritmos evolutivos. Ph.D. thesis, Universidade de São Paulo (2011)
3. Boeing, G.: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems **65**, 126–139 (2017). https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2017.05.004
4. De Medeiros, R.Á.O., et al.: Previsão de demanda no médio prazo utilizando redes neurais artificiais em sistemas de distribuição de energia elétrica. Master's thesis, Universidade Federal da Paraíba (2016)
5. Filho, J.B.S.A., Bacic, M.J.: Modelos de custos na regulação da indústria de distribuição de energia elétrica. In: XV Congresso Brasileiro de Custos. Curitiba, PR (Nov 2008)
6. Littlechild, S.C.: Privatisation, competition and regulation. The 29th Wincott Lecture delivered on 14th October 1999. Forward by Sir Geoffrey Owen., United Kingdom: The Wincott Foundation (1999)
7. Magalhães, E.F.A.: Modelagem e simulação de indicadores de continuidade: Ferramenta auxiliar para a manutenção em redes de distribuição de energia elétrica. Master's thesis, Universidade Federal da Bahia (2017)
8. Mattos, J.B.: Aplicação de redes neurais artificiais na estimação de indicadores de continuidade do fornecimento de energia elétrica. Master's thesis, Departamento de Engenharia Elétrica, Centro Tecnológico, UFES, Vitória, ES (2018)
9. Nunes, F.S.: Alocação ótima de chaves em sistemas de distribuição de energia elétrica utilizando a metaheurística VNS. Undergraduate dissertation. Universidade Federal do Ouro Preto (2018)
10. Pulz, J.: Abordagem alternativa para cálculo regulatório de perdas não-técnicas do sistema de distribuição e técnicas de machine learning para detecção de fraude na baixa tensão. Ph.D. thesis, Universidade de São Paulo (2022)
11. Ruiz, R., Stützle, T.: A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. European journal of operational research (2007)
12. Salkin, H.M., De Kluyver, C.A.: The knapsack problem: a survey. Naval Research Logistics Quarterly **22**(1), 127–144 (1975)
13. Tenfen, D.: Alocação ótima de chaves telecomandadas em redes de distribuição com multi-objetivo via algoritmos genéticos de Pareto. Master's thesis, Universidade Federal de Santa Catarina (2011)
14. Trentini, C.: Planejamento da manutenção de sistemas de distribuição com foco em custos e confiabilidade. Master's thesis, Universidade Federal de Juiz de Fora (2019)
15. Wirth, R., Hipp, J.: CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining. vol. 1, pp. 29–39. Manchester (2000)