SECNIR: A Multi-Year Electricity Consumption Dataset of 881 French Companies in Islands and Overseas Regions

Pierre-Alexis Chevreuil
Pierre Costini
Eco CO2 Nice, France
e-mail: pierrealexis.chevreuil@gmail.com
e-mail: pierre.costini@centraliens.net

Antoine Tavant
Cédric Simi
Eco CO2, Lyon, France
e-mail: antoine.tavant@m4x.org
e-mail: cedric.simi@proton.me

Joris Costes
Eco CO2 and *Univ. Paris-Saclay*,
ENS Paris-Saclay, CNRS, Centre Borelli,
Gif-sur-Yvette, France
e-mail: joris.costes@ens-paris-saclay.fr

***Abstract*—This paper presents the Survey of Electricity Consumption in Non-Interconnected Regions (SECNIR), a multi-year electricity consumption dataset comprising data from 881 companies and local authorities in French overseas regions. To the best of our knowledge, it is the first large-scale electricity consumption open dataset for professionals in islands and overseas regions at the level of individual electricity meters. The dataset includes electricity consumption, weather, and indoor comfort data from a variety of professionals in six different French islands and overseas regions: Reunion Island, French Guiana, Martinique, Guadeloupe, Corsica, and Mayotte. We analyze how local climate influences electricity consumption, in particular in tropical regions. Additionally, we provide insights into the most power-intensive sectors and examine the variability of energy consumption within specific professional sectors.**

Keywords-Energy; Electricity; Smart meters; Non-Interconnected Regions; Thermosensitivity; Dataset.

I. INTRODUCTION

With the increasing share of renewable energies in the worldwide electricity mix, determining where and when electricity is consumed becomes crucial for the stability of electrical grids. The deployment of smart power meters has enabled large-scale studies on individuals [1]–[3] and companies [4], [5]. However, these studies almost always focus on parts of the world connected to large electrical grids and in mild climates.

To the best of our knowledge, no large-scale open dataset with individual meter consumption data exists for Non-Interconnected Regions (NIR). This is notably true for tropical regions, which are particularly interesting to study because of their different thermosensitivity compared to territories with milder climates.

TABLE I. COMPARISON OF THE MEAN DIRECT CO_2 EMISSIONS PER ELECTRICAL KWH AND MEAN ELECTRICAL PRODUCTION COST IN DIFFERENT FRENCH TERRITORIES [6]–[10].

Region	Mean direct emissions $[gCO_2 \cdot kWh^{-1}]$	Mean production cost $[€ \cdot MWh^{-1}]$
Mainland France	32.4 (2023)	72-90 (2020)
Guadeloupe	703 (2019)	351 (2022)
Martinique	575 (2020)	323 (2022)
Reunion Island	732 (2021)	267 (2022)
Corsica	564 (2020)	316 (2022)
French Guiana	468 (2019)	274 (2022)

Obtaining data in such regions is generally challenging, as grid operators tend to deploy fewer smart meters compared to the mainland. For example, in continental France, over 90% of consumers have smart power meters, such as the Linky meters deployed by Enedis [11], while much fewer are installed in French overseas territories. Except for Corsica, very few electricity consumption data are available for these regions. Moreover, these territories heavily rely on fossil energies for electricity generation, leading to significant financial and environmental costs, as shown in Table I. In mainland France, the first cost figure represents the direct production cost, excluding network, import, and export costs, while the second corresponds to the annualized cost. Therefore, understanding how professionals use electricity in these regions and helping them to reduce their consumption is essential. Raising awareness about climate issues and promoting energy savings is an effective way to achieve this goal. Energy savings certificate programs are a French policy measure designed for this purpose. Under this scheme, energy suppliers must either achieve energy savings themselves or purchase certificates from other parties to meet their targets. These programs, defined by ministerial decrees, focus on providing information, training, funding and innovation to manage energy demand and alleviate household energy insecurity.

The data presented in this paper, collected through the SEIZE program managed by Eco CO2, has been used to create the anonymized, weekly-resampled Survey of Electricity Consumption in Non-Interconnected Regions (SECNIR) dataset. Figure 1 illustrates the locations of the 881 sensors.

This paper is organized as follows: Section II describes the SEIZE program and data collection methodology. Section III details the dataset structure and processing techniques. In Section IV, the analysis and visualization of the dataset is presented, followed by a discussion of its limitations. Finally, Section V concludes the paper and suggests directions for future research.

II. PROGRAM DESCRIPTION

This section describes the deployment and methodology of the SEIZE program, which aims to collect and analyze electricity consumption data from various French overseas regions.

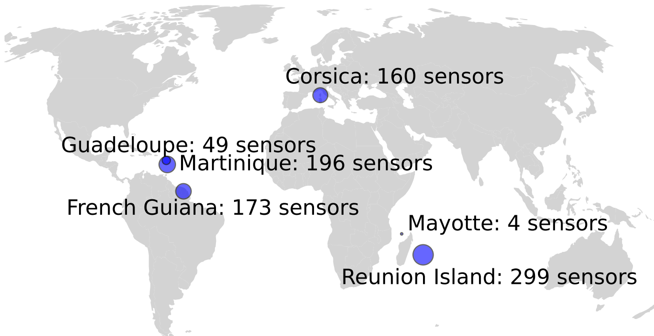

Figure 1. Number of electrical sensors per region.

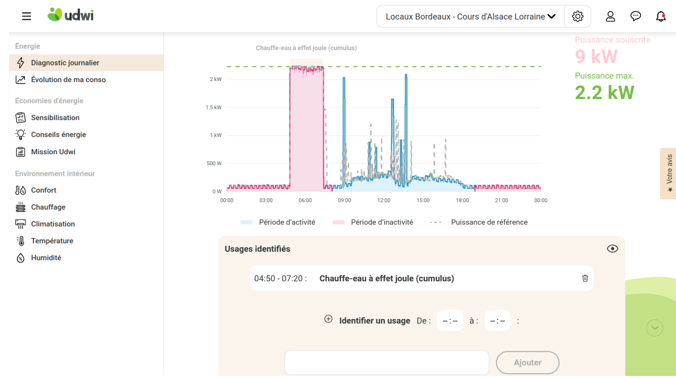

Figure 2. Screenshot of the interactive website showing the daily load curve and identified use periods.

A. Program deployment

The SEIZE program was built around 5 actions, each user being free to carry out one or more actions among:

- Environmental awareness workshops for managers and employees, aiming at raising attention about energy-saving actions in organizations;
- Installation of smart meters and comfort sensors, with access to the consumption data through an online visualization platform;
- Orientation towards financial aid schemes or technical advice and telephone/face-to-face appointment to redirect the professional to solutions adapted to their needs;
- Access to a network of ambassadors, who are professionals committed to promoting the program;
- Technical visits of buildings and delivery of recommendations concerning the management of the building or its energy renovation.

All these actions are offered to beneficiary companies and local authorities on a voluntary basis and free of charge. This paper concentrates on presenting the data collected by the smart meters installed during the SEIZE program. Datasets creation and analyses presented have been carried out by Eco CO2's Research and Development team and funded by Eco CO2.

B. Data visualization platform

To help organizations understand and reduce their electricity usage, users can visualize their consumption data in near real-time through an interactive website. They also receive weekly emails summarizing key trends in their consumption patterns. Figure 2 shows a screenshot of one of the website's pages displaying a daily load curve. Additional information, such as standby power over the last day, indoor and outdoor temperatures, indoor humidity, and the evolution of daily, weekly, and monthly electricity consumption, is also available, along with eco-gestures suggestions.

C. Sensors

Each participant of the study may have multiple sites, typically corresponding to individual buildings, each equipped with a unique electrical sensor. Most sites are also outfitted with one or more indoor comfort sensors measuring temperature and humidity. Outdoor temperature data are obtained using the building location [12].

Sensors' deployment took place between mid-2021 and end of 2024. At the time this paper is written, most sensors reported data for at least one year, as Figure 3 shows. Only sources with 4 weeks of data with sufficient quality have been kept in the database, which explains the flat lines from 0 to 4 weeks on Figure 3.

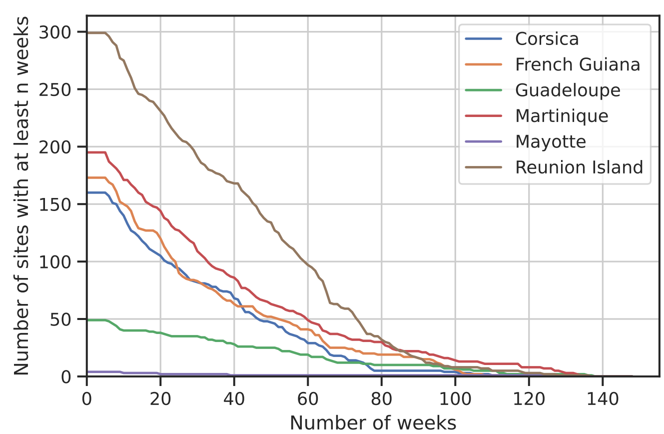

Figure 3. Evolution of the minimum number of weeks with electricity data in the dataset for each region.

D. Types of sensors

To facilitate sensor deployment, only sensors that do not require modifications of the users' electrical installations are used:

- Digital sensors communicate electronically with electricity meters to provide consumption indices.
- Analog sensors measure energy consumption by counting light pulses or wheel rotations from electromechanical meters.

Depending on the technology, the data timestep varies between one minute and one hour. External temperature data [12], indoor temperatures, and humidity levels are recorded every ten minutes.

III. Dataset

This section details the structure, processing, and anonymization of the dataset collected through the SEIZE program.

A. Data anonymization

To protect the users' privacy and follow the European General Data Protection Regulation (GDPR) guidelines [13], we resampled all the data to a weekly basis using `energy_analysis_toolbox`, a Python library developed by Eco CO2 R&D team and released in open-source [14]. Moreover, we provide for each participant its field of activity using the Statistical Classification of Economic Activities in the European Community (NACE) [15] classification.

B. Index data processing

The digital sensors transmit the consumption index, namely the integrated electricity consumption in kWh since the date of installation. Because of technical issues, abnormal index jumps are present in the database. Thus, we preprocess the data to automatically detect these anomalies. Weeks containing such elements are marked as invalid, and an empty entry replaces them in the `weekly_E_kWh` column of the dataset. This implies that the calculated energies can be considered valid on a weekly scale, and that the energy integrated over several weeks may be below reality if no data is provided for a week.

Furthermore, we interpolate between the last index of each week and the first known index of the following week to estimate weekly energy consumption as accurately as possible. If the time between the last known index and the first known index of the following week is too long, an empty entry is also written in the dataset.

C. Power data processing

The analog sensors detecting light pulses on electronic meters or wheel revolutions on electromechanical meters report the average power consumed each minute. The energy consumed is then calculated by taking the integral of the power. Weeks containing aberrant data, such as abnormally low or high power, abnormal load curve shapes, are removed from the analysis and an empty entry is written in the corresponding week of the dataset. Thus, similarly to digital sensors, calculated electricity consumption can be considered valid on a weekly scale, but energy integrated over several weeks may be below reality if no data is provided. In all following figures, we estimate the annual consumption by imputing missing week with the mean value consumed during the weeks kept in the dataset. Table II shows a few statistics about the dataset.

TABLE II. Total number of sites number, cumulated weeks of data, and yearly electricity consumption statistics, per region. IQR stands for interquartile range.

Region	Sites	Weeks	Mean [MWh]	Min [MWh]	Max [MWh]	IQR [MWh]
Reunion Island	299	15362	46.9	0.5	517.9	48.4
Martinique	196	9729	38.1	0.6	383.3	42.4
French Guiana	173	7410	36.9	0.5	461.2	31.8
Corsica	160	6731	30.0	0.5	429.9	32.6
Guadeloupe	49	2860	38.5	0.6	436.8	50.7
Mayotte	4	196	12.2	3.1	22.0	5.3

D. Temperature and humidity data

A practical method for using outdoor temperature data alongside electricity consumption data is to calculate Heating Degree-Days (HDD) and Cooling Degree-Days (CDD). In the dataset, we compute these with several reference temperatures T_0 (15°C to 18°C for HDD and from 22°C to 26°C for CDD, both by steps of 1°C) using the integral method:

$$HDD_{T_0} = \int_{\text{Monday 0h}}^{\text{next Monday 0h}} [T_0 - T_{\text{ext}}(t)]^+ \, \mathrm{d}t, \tag{1}$$

$$CDD_{T_0} = \int_{\text{Monday 0h}}^{\text{next Monday 0h}} [T_{\text{ext}}(t) - T_0]^+ \, \mathrm{d}t, \tag{2}$$

where T_{ext} the external temperature, and:

$$x^+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

E. Dataset documentation

The dataset is available for download on data.gouv.fr [16] and Kaggle [17] as a single CSV file. In this file, each line corresponds to one week of data for a given site. Each line in the dataset represents data for a specific week and contains the following columns:

- `year-Wweek`: The ISO week. For example, "2023-W01" corresponds to the week from Monday, January 2 to Sunday, January 8, 2023.
- `user_id`: A unique identification number for each user.
- `site_id`: A unique identification number for each site. Each user can have multiple sites, but each site has precisely one electrical sensor. A site is typically a building, or a group of buildings.
- `department`: The name of the department where the sensor is located. Possibilities are `South Corsica`, `Upper Corsica`, `Guadeloupe`, `Martinique`, `Guyane`, `Reunion Island` and `Mayotte`.
- `nace_code`: The NACE code corresponding to the unique identifier of the company (SIRET) provided by the user (e.g., `47.71` for "Retail sale of clothing in specialized stores").
- `insee_code`: Unique numerical identifier assigned by the French National Institute of Statistics and Economic

Studies (INSEE) to French municipalities. On the January, 1st, January, 2023, France had 34945 municipalities [18].

- `weekly_E_kWh`: Weekly energy consumption in kWh.
- `weekly_dd_heating_<T0>`: Weekly HDD based on the reference temperature T_0, computed with Eq. 1.
- `weekly_dd_cooling_<T0>`: Weekly CDD based on the reference temperature T_0, computed with Eq. 2.
- `<stats_type>_indoor_<data_type>_<nn>`: Various statistics for indoor temperature and humidity data recorded by the comfort sensors. `<stats_type>` can be "min", "max" or "average". `<data_type>` can be "temperature" or "humidity". For temperature data, values are given in degrees Celsius, and for humidity, it is relative humidity in %. `<nn>` is a unique number for each comfort sensor between 00 and 99. The same number is used for temperature and humidity. For example, two columns named `min_indoor_temperature_00` and `min_indoor_humidity_00` correspond to the same physical sensor.

IV. Analysis and Visualization

This section presents the analysis and visualization of the dataset, focusing on electricity consumption patterns across different sectors and regions.

A. Estimated annual consumption by NACE category

To get insights on the annual consumption depending on the activity sector, we grouped companies in 9 categories based on their NACE code, as explained in Table III. As Figure 4 shows, there are significant consumption differences between the various sectors, with differences of median values for two sectors going up to an order of magnitude.

This discrepancy can be attributed to the nature of the activities in these sectors, where office work primarily involves low energy-consuming activities like lighting and computing, whereas shops and transport, as well as accommodation and catering, typically require more energy for operations, such as refrigeration, cooking, cooling, and heating.

TABLE III. Relationship between NACE sections and the categories used in Figure 4.

Category	NACE sections
Natural resource extraction and agriculture	A, B
Manufacturing and construction	C, D, E, F
Shops and transport	G, H
Accommodation and catering	I
Information and finance	J, K, L
Technical and administrative services	M, N
Public and social services	O, P, Q
Arts and recreation	R, S, T
Extra-territorial activities	U

Consumption varies significantly within each category, spanning more than two orders of magnitude between the minimum and maximum values. This may be due to other explanatory parameters not provided in the dataset, such as the number of employees, building surface area, number of equipment, equipment efficiency, or heating energy.

B. Thermosensitivity analysis

French islands and overseas regions studied in this paper exhibit diverse climatic conditions. Reunion Island, Mayotte, French Guiana, Martinique and Guadeloupe all experience tropical climates. However, while Reunion Island and Mayotte are both located in the southern hemisphere, Martinique and Guadeloupe are in the northern hemisphere, and French Guiana is very close to the Equator. Hence, the climate experienced in Reunion Island and Mayotte is relatively similar to that experienced in Martinique and Guadeloupe, but offset by 6 months because of the locations in different hemispheres.

Reunion Island and Mayotte are both characterized by hot and humid weather throughout the year, with distinct wet and dry seasons. The dry season takes place during May to

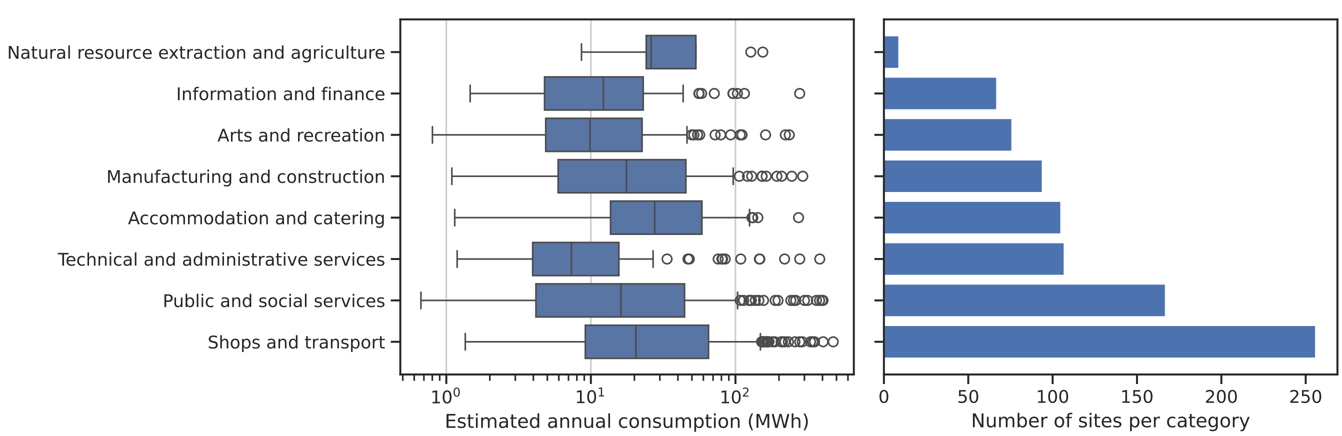

Figure 4. (left) Boxplot of the estimated annual consumption by company category. Whiskers are drawn at ±1.5 interquatile range (IQR), and all values outside this interval are represented with circles. (right) Number of sites per category.

 ISBN: 978-1-68558-242-5

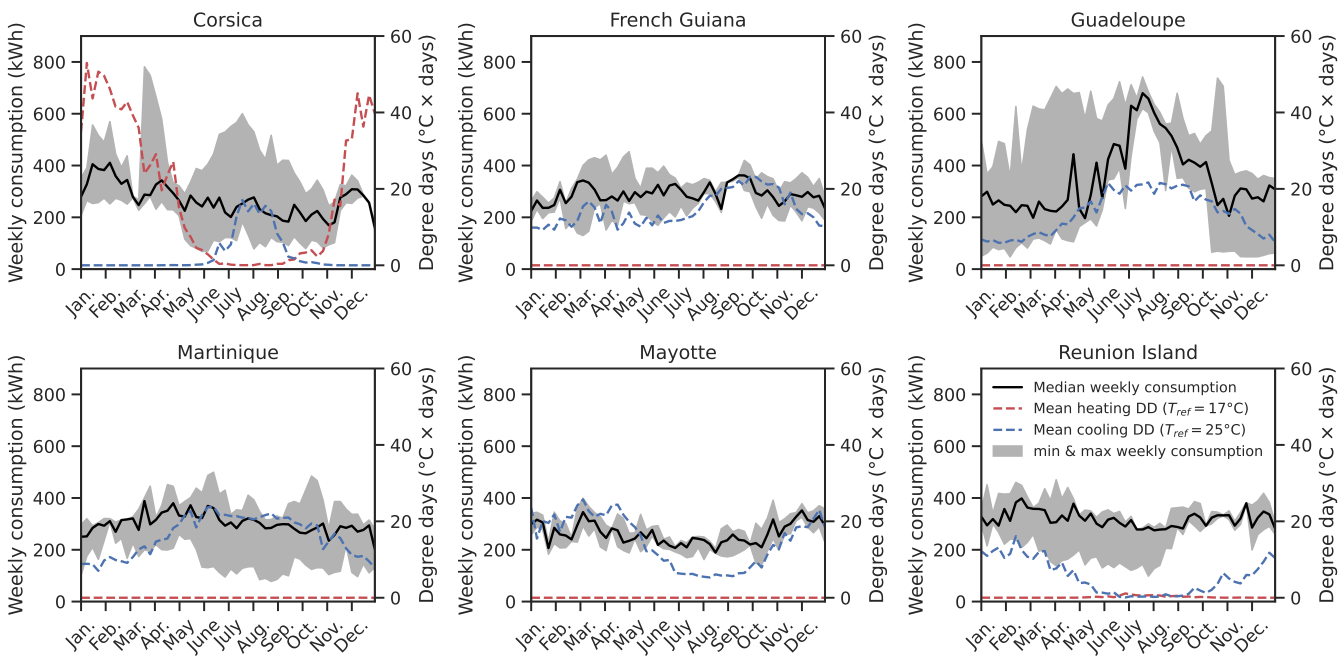

Figure 5. Seasonal evolution of the weekly consumption for each region. Each point represents one week. The black line represents the median weekly consumption across years and sites for each week of the year. For instance, the point for week 10 is the median consumption for all sites in a region across week 10 of 2021, 2022, 2023 and 2024. The gray-shaded area represents the minimum and maximum of the weekly consumption across sites, respectively. The blue and red dashed curves represent the mean HDD and CDD computed using Eq. 1 and Eq. 2 for heating (T_{ref} = 17°C) and cooling (T_{ref} = 25°C), respectively, averaged across years and sites.

November, when the weather is generally mild (typically 21 to 28°C) on the coast. The rainy season begins in December and ends in April. As Figure 5 shows, this translates into no need for heating throughout the year, and some cooling needs between November and April. This is reflected in the consumption, where a small correlation between mean CDD and usage can be seen.

Due to its proximity with the Equator, temperature and humidity in French Guiana varies very little throughout the year. Therefore, even if participants use air conditioning, the corresponding consumption remains roughly constant, which we indeed observe in Figure 5. It is also worth noting that consumption can vary greatly from one year to the other, depending on the territory, as the gray-shaded area in the same figure indicates.

Martinique and Guadeloupe share a similar tropical maritime climate. These islands experience warm temperatures year-round, with a wet season typically occurring from June to November, and a drier period from December to May. This can indeed be seen in the corresponding blue dashed curves in Figure 5. Interestingly, while the CDD curves are very similar for Martinique and Guadeloupe, we observe a much larger thermosensitivity in Guadeloupe than in Martinique. This suggests that, for the population studied, air conditioning systems are more used in Guadeloupe than in Martinique.

Corsica, in contrast, has a Mediterranean climate, characterized by hot, dry summers and mild, wet winters. This climate creates distinct heating and cooling needs, with a higher sensitivity to cooling during the summer compared to heating in the winter. This may be partially because air conditioning always requires electricity, while some heating systems do not. As a result, heating consumption may not be fully accounted for in Figure 5.

Finally, as Figure 6 shows, the distribution of consumption by region is very similar between all regions, the only exception being Mayotte, due to the fact that only 4 sensors are present in that region.

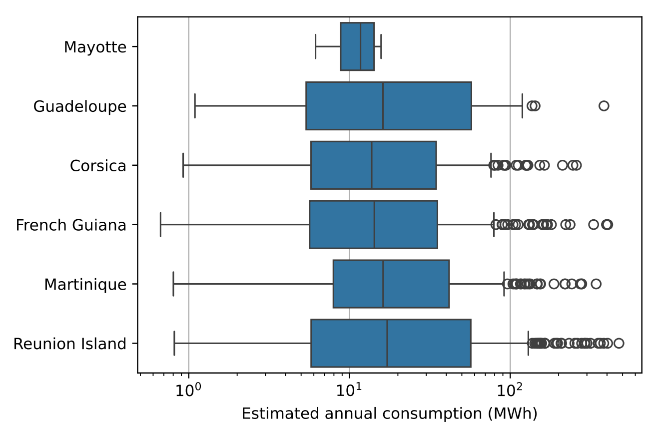

Figure 6. Boxplot of the estimated annual consumption by region. Whiskers are drawn at ±1.5 IQR, and all values outside this interval are represented with circles.

 ISBN: 978-1-68558-242-5

C. Dataset limitations

The SECNIR dataset, while extensive and valuable, has certain limitations that must be acknowledged. Firstly, as explained in sections III-B and III-C, the dataset contains gaps for some weeks: over the 42288 weeks of the dataset, 5358 weeks (12.7%) do not contain electrical consumption data for these reasons.

Secondly, the dataset's territorial coverage is uneven, with significant disparities in the number of sensors deployed across different regions. For example, Mayotte is notably underrepresented with only 4 sensors compared to other regions like Reunion Island and French Guiana, which have 299 and 173 sensors respectively.

The scarcity of sensors in certain regions like Mayotte means that the insights drawn from these regions may not fully capture the variability and unique characteristics of local energy consumption patterns. Consequently, the conclusions derived from the data should be interpreted with caution, especially when comparing different regions.

V. Conclusion and Future Work

This paper presented SECNIR, a dataset on electricity consumption from 881 companies and local authorities across six French overseas regions. This dataset provides a valuable resource for analyzing consumption patterns in regions that are typically underrepresented in large-scale studies due to their isolated and tropical nature. By including electric, weather, and comfort data, this dataset allows for a nuanced understanding of how local climates influence electricity consumption.

The diverse climates of the regions studied—ranging from the tropical climates of Reunion Island, Mayotte, French Guiana, Martinique, and Guadeloupe to the Mediterranean climate of Corsica—highlight the importance of context-specific strategies in managing energy demand.

The findings from this study emphasize the need for tailored energy efficiency programs and policies that take into account the unique consumption patterns and climatic conditions of these regions. As these regions often rely heavily on fossil fuels, there is a pressing need to explore decarbonized energy solutions to mitigate environmental impacts and dependence on fossil fuels.

Future work could explore intra-day consumption variations, in a similar manner to the work done in [5]. This research could help policy-makers and grid managers identify the flexibility measures needed to support the transition to fully decarbonized energy mixes in these regions.

References

[1] M. Pullinger *et al.*, "The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes," en, *Scientific Data*, vol. 8, no. 1, p. 146, May 2021, ISSN: 2052-4463. DOI: 10.1038/s41597-021-00921-y.

[2] R. Medico *et al.*, "A voltage and current measurement dataset for plug load appliance identification in households," en, *Scientific Data*, vol. 7, no. 1, p. 49, Feb. 2020, ISSN: 2052-4463. DOI: 10.1038/s41597-020-0389-7.

[3] O. Parson *et al.*, "Dataport and NILMTK: A building data set designed for non-intrusive load monitoring," en, in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA: IEEE, Dec. 2015, pp. 210–214, ISBN: 978-1-4799-7591-4. DOI: 10.1109/GlobalSIP.2015.7418187.

[4] G. Liu *et al.*, "EWELD: A Large-Scale Industrial and Commercial Load Dataset in Extreme Weather Events," en, *Scientific Data*, vol. 10, no. 1, p. 615, Sep. 2023, ISSN: 2052-4463. DOI: 10.1038/s41597-023-02503-6.

[5] K. Bellinguer, R. Girard, A. Bocquet, and A. Chevalier, "ELMAS: A one-year dataset of hourly electrical load profiles from 424 French industrial and tertiary sectors," en, *Scientific Data*, vol. 10, no. 1, p. 686, Oct. 2023, ISSN: 2052-4463. DOI: 10.1038/s41597-023-02542-z.

[6] G. Gilboire *et al.*, "Energy balance Reunion Island 2021 + Key Figures 2021 (2022 edition)," France, Tech. Rep., 2022, INIS-FR–23-0083, p. 130.

[7] RTE France, "Electrical Balance 2023," Tech. Rep., Jul. 2024, p. 30.

[8] Court of Audit, "Support for non-interconnected zones," Court of Audit, Final observations S2023-1177, Apr. 2023, p. 5.

[9] French Energy Regulatory Commission, *Energy transition in the non-interconnected zones*, https://www.cre.fr/electricite/transition-energetique-dans-les-zni.html, , retrieved January, 2025.

[10] Court of Audit, *Complete cost analysis of the French power generation system*, Dec. 2021.

[11] Deliberation 2022-64 of the Energy Regulatory Commission on the draft decision on the incentive regulation framework for the Enedis advanced-metering system in the LV voltage field lower or equal to 36 kVA (Linky) for the period 2022-2024 and modifying deliberation no. 2021-13 related to the decision on the tariff for the use of public electricity distribution networks (TURPE 6 HTA-BT), Feb. 2022.

[12] Weather data provided by OpenWeather, https://openweathermap.org, retrieved January, 2025.

[13] Official Journal of the European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)., 2016.

[14] EcoCO2, *Energy_analysis_toolbox*, https://github.com/Eco-CO2/energy_analysis_toolbox, retrieved January, 2025., 2024.

[15] EUROSTAT, "NACE Rev. 2," Tech. Rep., 2008.

[16] P.-A. Chevreuil, A. Tavant, P. Costini, C. Simi, and J. Costes, *SECNIR: Electricity consumption of businesses and local authorities in non-interconnected French regions*, https://www.data.gouv.fr/fr/datasets/consommation-delectricite-des-entreprises-et-collectivites-locales-dans-les-regions-francaises-non-interconnectees-1/, retrieved January, 2025., Sep. 2024.

[17] P.-A. Chevreuil, A. Tavant, P. Costini, C. Simi, and J. Costes, *SECNIR: Electricity consumption of businesses and local authorities in non-interconnected French regions*, https://www.kaggle.com/datasets/ecoco2/secnir-consumption-of-french-overseas-companies, retrieved January, 2025., Sep. 2024.

[18] L. Sebbane, *Statistical information bulletin 172 of the General Directorate of Local Authorities*, Mar. 2023.