PROCEEDINGS OF ECOS 2023 - THE 36TH INTERNATIONAL CONFERENCE ON
EFFICIENCY, COST, OPTIMIZATION, SIMULATION AND ENVIRONMENTAL IMPACT OF ENERGY SYSTEMS
25-30 JUNE 2023, LAS PALMAS DE GRAN CANARIA, SPAIN

Identification of typical district configurations: A two-step global sensitivity analysis framework

Arthur Chuat[a] (CA), Jonas Schnidrig[a,b], Cédric Terrier[a] and François Maréchal[a]

[a] Ecole Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland, arthur.chuat@epfl.ch (CA)
[b] University of Applied Sciences Western Switzerland (HES-SO), Sion, Switzerland

Abstract:
The recent geopolitical conflicts in Europe highlighted the sensibility of the current energy system to the volatility of energy carrier prices. In the prospect of defining robust energy system configurations to ensure energy supply stability, it is necessary to understand which parameters modulate the system configuration. This paper presents a framework that identifies a panel of technological solutions at the district level. First, a global sensitivity analysis is performed on a mixed integer linear programming model which optimally size and operate the system. The sensitivity analysis determines the most influential parameters of the model and provides a representative sampling of the solution space. The latter is then clustered using a density-based algorithm to identify typical solutions. Finally, the framework is applied to a suburban and residential Swiss neighborhood. The main outcome of the research is the high sensitivity of the model to energy carrier prices. As a result, the sampling space separates itself into two system types. The ones based on a natural gas boiler, and the ones relying on a combination of electrical heater and heat pump. For both types, the electricity demand is either fulfilled by PV panels or by electricity imports.

Keywords:
Mixed Integer Linear Programming (MILP), global sensitivity analysis (GSA), clustering, district energy system (DES), solution space sampling

1. Introduction

1.1. Background

The energy transition presents itself as one of the main levers to mitigate global warming as the energy sector accounts for 73% of global emissions and is expected to increase over the coming decade [1]. Therefore, developing a sustainable system requires careful energy planning from policymakers. In this prospect, energy system models are essential tools for understanding and assessing the potential impact of new technologies or policies.
The built environment, which accounted for 36% of final energy consumption in 2020 [2], is viewed as a promising pathway to a sustainable energy mix. Moreover, distributed energy systems, such as those composed of interconnected energy hubs, have great potential when deployed at the district scale, increasing self-consumption, and mitigating grid congestion.

1.2. Literature review

1.2.1. Energy system modelling

Energy models play a crucial role in guiding decision-makers towards a fossil-free energy system. The decarbonization of the current system requires a profound restructuring and involves a deeper electrification [3]. Low-carbon technologies, such as solar and wind, are spreading rapidly and are expected to become the main source of electric power [4, 5]. However, those renewable energies are intermittent and decentralized, bringing new challenges such as redistribution of electricity, stabilization of the power grid, intra and inter-day storage [6]. The multiplication of energy vectors, i.e., electricity, heat, oil, waste, biomass, or even hydrogen, increases the complexity of the task to provide an optimal configuration and operation of the system. Consequently, the majority of models focus on one specific energy sector (electricity, heat, mobility, etc.) and do not account for cross-sectoral synergy [7]. Additionally, on a more general aspect, the scale of the system can vary from a single building to an entire country. Whether or not, energy subsystems are considered in the overall scale of the system, e.g., interconnected buildings to represent a district system or the regionalisation of a country-wide system.

https://doi.org/10.52202/069564-0228

Buildings archetypes have been widely used to define large-scale systems [8] as their emissions have doubled in the last 30 years [4]. Kotzur and al. [9] developed a bottom-up model based on residential building stock. They deployed an aggregation algorithm to define archetype buildings. Their configuration and operation, which includes the buildings' interaction, are then optimized to acquire cost-effective solutions. Stadler and al. [10] assessed the impact of model predictive control (MPC) on the Swiss building stock. Three typical buildings were defined, each having 9 variations based on their construction date. The deployment of MPC allowed the reduction of the operational expenditure (OPEX) as well as the increase of the self-consumption. This result emphasizes the importance of the interconnection and operation improvement of building energy system.

In this regard, the design and operation of interconnected distributed energy systems, such as buildings, have proven to be part of the solution to help improve sustainable development in many countries [11–14]. The distributed aspect of such systems comes from the interconnection of multiple energy sources. Those sources can be energy hubs linked by local multi-energy grids. As discussed in the review of district-scale energy systems by Allgerini and al. [15], there has been a significant improvement in the models and tools used to analyze such systems.

In their paper, Morvaj and al. [16] performed an optimization of an urban scale energy system composed of twelve buildings. For each building, an optimal design and operation have been identified and the district heating network associated was optimized to reduce greenhouse gas (GHG) emissions and the total expenditure (TOTEX). Maroufmashat and al. [17] highlighted the importance of considering multiple energy hubs in order to observe significant cost and GHG emissions reduction. Their case study showed that the implementation of distributed combined heat and power (CHP) units was limited while operating an electricity grid with low CO_2 emissions. Additionally, the operation of interconnected energy hubs can significantly increase the robustness of the power grid, e.g. mitigation of congestion and ensuring reliability.

1.2.2. Global Sensitivity Analysis

Most current models assume a perfect knowledge of the input parameters, which induces the absence of uncertainty in the model and characterizes them as deterministic [12]. However, there is some uncertainty in a number of areas, including current policy, renewable energy production, and economic trends. Therefore, specific models are developed to consider the uncertainty of input parameters: stochastic models. However, it might be complicated to transform a deterministic model into a stochastic one without major remodeling. An alternative is the application of sensitivity analysis (SA) to deterministic models as proposed in [18, 19].

The utilization of sensitivity analysis is an effective approach to evaluate the impact of input parameters on a model. There exist two main types of SA: local and global. The most common method is the local sensitivity analysis, which evaluates the sensitivity by computing partial derivatives and gradients. Even though its ease of implementation makes it popular, the sampling scheme does not scan the entire space of input parameters. This gap is filled by the global sensitivity analysis (GSA), which covers the sampling space by varying several parameters at once. In doing so, it can capture parameters interaction [20].

The review of the SA method for building energy systems performed by Tian and al. [21] emphasized the importance of choosing the right method. They concluded that the choice should be based on the following criteria: research purpose, computational cost of energy models, number of input variables, and familiarity with the methods. Another review was performed by Westermann and Evins [22], 57 studies on building design were analyzed, focusing on objective, sampling strategy, and surrogate model type. Among all studies only 16 included a SA of the model, their sampling strategies were primarily based on Latin hypercube sampling (LHS), and only 3 used an optimization model, highlighting the low implementation of SA in optimization models.

A two-stage GSA framework has been used in several studies [18,19,23]. Their goal is diverse, from assessing the uncertainty, and sensitivity of the system, or to explore the multidimensional design space, but the applied framework stays the same. First, a screening method filters out non-influential parameters to reduce the computational cost of further model evaluations. It is then combined with a Sobol sequence to obtain a proper sampling of the solution space for the final sensitivity analysis.

1.3. Gaps and contribution

As emphasized throughout subsection 1.2., the literature lacks an application of distributed energy systems as energy hub in large-scale energy system optimization. The role of districts as renewable energy hubs can be

https://doi.org/10.52202/069564-0228

central in the optimization of a carbon-neutral national energy system as the operation of energy fluxes can help considerably improve the thermo-efficiency of the system. In this prospect, the proposed configurations should represent at best the district's solution space. This paper attends to provide a framework to identify a set of optimal district configurations. The framework consists of a two-step GSA on a district-wide energy system model to explore its solution space. The latter will then be aggregated to identify typical configurations. The purpose of this paper is to contribute to the following research points:

- Identification of the most influential parameters.
- Exploration of the solution space.
- Identification of typical configurations.

2. Methodology

2.1. Global sensitivity analysis

2.1.1. Overview

In order to develop a panel of solutions for the district, it is necessary to explore the whole solution space of the model. The following methodology is inspired by the publication of Saltelli and al. [20] which assesses the state-of-the-art of GSA. A sensitivity analysis can be decomposed into four main steps:

1. Identification: k input parameters of the model are selected
2. Sampling: the input parameters space is discretized by N samples
3. Evaluation: the model outputs are computed for each sample
4. Comparison: some metrics are derived from the N outputs of the model for the k parameters

The methodology used consists of two separate SAs:

1. A screening method is performed to identify the most influential factor of the model, i.e. the parameters inducing the greatest variation of the objective function.
2. A variance-based method is used to efficiently explore the solution space of the model and quantify its sensitivity.

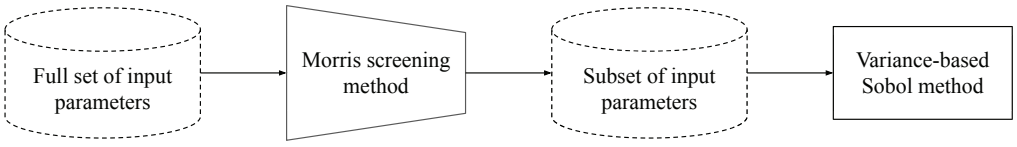

Figure 1: Scheme of the two-step GSA composed of first the Morris method as a screening and then the Sobol method is applied on the most influential parameters previously identified.

2.1.2. Morris screening

The screening method used is the Morris method, which allows to qualitatively compare the influence on the model output of a large number of parameters with a few evaluations [24]. The method discretizes the input parameters space, which is a k-dimensional hypercube, into a p-level grid, where k is the number of independent input parameters and p is a sampling parameter. Afterward, it performs a one-step-at-the-time method (OAT), i.e. it randomly modifies one input parameter by $\pm\Delta$ to generate r trajectories.
Finally, it evaluates the elementary effect of the i_{th} input parameter (EE_i) as a function of the model output $Y = f(X_1, ..., X_n)$, see Equation 1. The EE can be interpreted as a local partial derivative, thereby it represents the sensitivity of the model at a specific point w.r.t the input parameter:

https://doi.org/10.52202/069564-0228

$$EE_i = \frac{\partial Y}{\partial x_i} \simeq \frac{Y(X_1, X_2, \dots, X_{i-1}, X_i \pm \Delta, \dots X_k) - Y(X_1, X_2, \dots, X_k)}{\Delta} \tag{1}$$

Where Δ is defined as a function of p: $\Delta = \frac{p}{2(p-1)}$ and can be considered as the size of the discretization mesh. The total number of model evaluations amounts to $r(k+1)$, where r is suggested between 4-10 [20]. The choice of p and r has to be made jointly to ensure that the k dimensions and their interactions are correctly sampled, Saltelli proposed $p = 4$ and $r = 10$ [25] whereas Morris used $r = 4$ in [24], which seems to be the minimum usable value.

In its original work, Morris proposed the computation of the mean μ_i (Equation 2) and standard deviation σ_i (Equation 4) of the elementary effect distribution for each parameter i. However, by doing so, the positive and negative effects cancel each other out, which would falsely influence the results of the mean value. Thereby, the method has been improved by Campolongo and al. [26], by considering the absolute mean elementary effect μ_i^* (Equation 3).

$$\mu_i = \frac{1}{r} \sum_{j=1}^{r} EE_i^j \tag{2}$$

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} |EE_i^j| \tag{3}$$

$$\sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^{r} \left(EE_i^j - \mu_i \right)^2 \tag{4}$$

Those indicators allow to compare the input parameters between each other. A small absolute mean value means a non-influential parameter. The standard deviation reflects the interaction between parameters: a high value means that the output is highly dependent on the sampling point, i.e. on the other input parameters values. Conversely, a low standard deviation indicates that the elementary effect is not subject to fluctuate with other factors variation.

The representation of the mean absolute value of the EEs and their standard deviation makes it easy to identify to which group the parameter belongs. As represented on Figure 2, one can see the different zones and the line ($x = y$) separating quadrant 1 that defines whether parameters interact together or not. The different zones can be defined as follows:

1. Non-influential parameters.
2. Influential, non-interacting parameters.
3. Influential, interacting parameters.
4. Influential parameters.

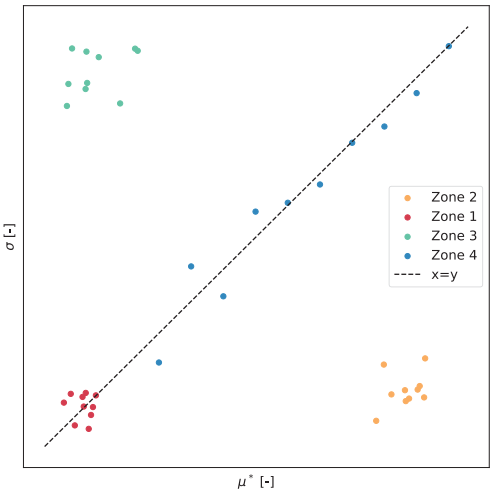

Figure 2: Identification of the typical zone on the $\mu^* - \sigma$ plane.

The comparison of μ and μ^* gives an extra insight in the monotony of the model. If the model output increases with an augmentation of the parameter, the EE will stay positive thus μ^* and μ will have a similar value. Whereas, if the EE changes sign regularly its cumulative will be lower, i.e. μ will be lower than μ^*.

2.1.3. Sobol sampling

The Sobol method is a variance-based sensitivity analysis named after the mathematician Ilya M. Sobol. The method is developed in [20, 27] and uses Sobol's recommendation on the sequencing of quasi-random numbers. However, Saltelli extended her work in [28] to reduce the error rate when computing sensitivity indices.

https://doi.org/10.52202/069564-0228

2.1.3.1 Sensitivity indices

The method evaluates two different sensitivity indices, the first one is the first-order sensitivity coefficient (Equation 5). It results from the ratio of the variance of the mean output, considering all parameters except the i-th, to the variance of the output. The secondary sensitivity coefficient is the total effect index, i.e. first and higher-order sensitivity coefficient (Equation 6).

Assuming: $Y = f(X_1, ..., X_n)$ is the model output, $\mathbf{X}_{\sim i}$ symbolizes all parameters but $\mathbf{X}_i$, $E_{\mathbf{X}_{\sim i}}(Y \mid X_i)$ is the mean of Y for every possible $\mathbf{X}_{\sim i}$ and finally the variance V_{X_i} is calculated for all values of X_i.

$$S_i = \frac{V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y \mid X_i)\right)}{V(Y)} \tag{5}$$

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}\left(V_{X_i}(Y \mid \mathbf{X}_{\sim i})\right)}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{\sim i}}\left(E_{X_i}(Y \mid \mathbf{X}_{\sim i})\right)}{V(Y)} \tag{6}$$

The first-order coefficient only takes into account the effect of itself on the output value, but not the possible effect when considering higher-order interactions with other parameters. The total-order index evaluates the effect of a parameter considering all possible interactions with other parameters. Meaning that a parameter with a value of $S_T = 0$ can be considered as non-influential on the output Y. The total number of model evaluations is $N(k+2)$ with N being typically between 500 and 1000 and it is suggested to choose a power of two [20].

2.2. Clustering

2.2.1. Standardization and features selection

It is necessary to standardize the data when various features are used for clustering. Otherwise, the aggregation algorithm mainly takes into account the large numerical values. The chosen standardization technique is the *z-score*, see Equation 7. Each feature Y_i of the data set is standardized as follows:

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i} \tag{7}$$

where μ_i is the mean and σ_i is the standard deviation of the feature Y_i. The features selected for the clustering, i.e. the district characteristics, are a mix of economic and technical attributes of the optimization result. The key performance indicators (KPI) chosen are capital and operational expenditure (CAPEX and OPEX). Installed capacity of energy conversion and storage units are included in the clustering as they are key properties of the district configuration. Regarding the exchange with the network, the total and peak energy supply and demand are considered for the natural gas (NG) and the electricity grid.

2.2.2. Aggregation techniques

Several aggregation techniques, such as K-means, DBSCAN, and HDBSCAN, performance are analyzed for this specific clustering. The K-means method is a clustering technique that minimizes the variance within clusters, which is a difficult problem to solve optimally. However, heuristic algorithms can be used to find a local optimum efficiently. To determine the appropriate number of clusters for K-means, two common techniques are the Silhouette score and the Elbow method. The Silhouette score evaluates how well each data point fits within its assigned cluster, based on the density of points within the cluster and the distance between clusters. The Elbow method involves observing the point of inflection on a plot of the sum of squared errors (SSE) as the number of clusters increases.

Another assessed method is the DBSCAN, which is based on the concept of core points. Points are defined as core points of one specific cluster when they can reach a minimum of *minPts* neighbours within a ϵ distance. Additionally, points within reach, but not satisfying the minimum neighbours criterion still belong to the cluster. However, points non-reachable from a core point are considered outliers. The choice of *minPts* and ϵ should be based on the data properties [29].

The HDBSCAN algorithm is an extension of the DBSCAN method with a hierarchical approach. The algorithm can be decomposed in a few steps to get a broad overview. First, the space is transformed based on its density. This allows to construct a minimum spanning tree used to build the cluster hierarchy. Then, the cluster tree is condensed to finally extract the clusters [30, 31].

To support the clustering task, a score is introduced to quantify the quality of the clusters. The chosen score is the density-based clustering validation (DBCV) index presented by Moulavi and al. [32]. It is based on a parameterless core distance defined by the density of objects and mutual reachability. The index ranges between -1 and 1, which represents a bad and a good score respectively.

2.3. Case study

The application of the presented framework is carried out on a suburban and residential district located in the climatic zone of Geneva, Switzerland. The energy system consists of 15 buildings, including 5 single-family houses and 10 apartment buildings [33], selected from a pool of 30 available options. Various characteristics of the buildings, such as type, year of construction, energy reference area, and height, are obtained from the cantonal and federal Official Building Registry [34,35], while the ground surface area is provided by the cantonal administration [35, 36]. The climate data, solar irradiation, and temperature are extracted from Meteonorm [37], and the characteristics of the building envelope are calculated following SIA norms 380/1 [38]. The physical information gathered on the buildings allows to define the electric and heat demands based on a 1R1C model detailed in [39]. The PV orientation is optimized based on roofs and façades information extracted from Swisstopo [40,41], while the grid specifications and electricity demand are furnished by Romande Énergie [42]. The time series are divided into 10 typical periods formed of 24 hours and 2 extreme periods of 1 hour each, resulting in a total of 242 different timesteps being considered. Various parameters are fixed during the problem definition, including unit parameters, district parameters, and energy carrier price. More information on these parameters can be found in [43].

3. Results and Discussions

3.1. Sensitivity analysis

The considered parameters for the screening step are energy conversion and storage units parameters as well as the energy carriers prices. A total of 60 parameters were considered. Following the sampling indications of Morris, 610 optimizations were required to compute the sensitivity metrics. As the result of each run is a complete system configuration and operation, a specific output value had to be identified to compare each optimization. The chosen indicator is the TOTEX, assuming an economically rational behaviour.
A good practice is to plot the absolute mean value (μ^*) and standard deviation (σ) of the EE distribution of each input parameter, Figure 3 shows such a plane for the six most influential parameters. Two parameters stand out from the rest: the supply cost of electricity and natural gas. The supply cost of electricity has a high influence on the TOTEX, high μ^*, however, its standard deviation is low, indicating that its EE is not correlated to the other input parameters. This low correlation comes from its key role in the district model. It serves to directly supply the electric demand when no installed units can fulfill it and it can also serve to supply the heat demand via HPs and electrical heaters. The retail price of NG is less influential as fewer units consider natural gas as a resource. However, it can be considered as an influential parameter as a small price of NG can help provide heat at a very low cost with a gas boiler. Note that the HP and boiler parameters are also influential with a low standard deviation, highlighting that the technologies installation is not correlated to the energy carriers prices.

Table 1 presents the Morris results of the aforementioned parameters. Comparing the absolute (μ^*) and non-absolute (μ) mean value of EE shows the monotonic correlation between TOTEX and input parameters values, e.g., an increase in electricity supply cost will always increase the TOTEX. Furthermore, the negative sign of the mean EE of the electricity feed-in tariff indicates that it contributes to the reduction of TOTEX.

Table 1: Sobol's indices, mean and absolute mean EE for retail and feed-in tariffs of energy carriers. The sensitivity of the parameters for the Morris and Sobol approaches all trend in the same direction.

	S_1	S_{Ti}	μ^*	μ
Electricity retail tariff	0.86	0.91	93k	93k
Natural gas retail tariff	0.09	0.14	33k	33k
Electricity feed-in tariff	1e-4	6e-4	0.9k	-0.9k

In order to reduce the computational time of the SA a selection of the input parameters was required. The decision to focus on the energy carrier prices was based on obtained results and similar conclusions presented in subsubsection 1.2.2..

https://doi.org/10.52202/069564-0228

The input parameter space is explored with 2560 samples following recommendations presented in subsubsection 1.2.2., $p = 4$, $r = 10$, and $N = 512$. The parameters variation range is $\pm 50\%$ for retail tariffs, as displayed in Figure 4, and $\pm 30\%$ for the feed-in tariff. The sensitivity indices of the Sobol can be observed in Table 1 alongside Morris μ^* values for comparison.
The Sobol results have a similar trend than the qualitative sensitivity indices from the Morris method. The electricity feed-in tariff is less influential regarding Sobol's results. Since the first-order Sobol effect and the total effect have a similar value, the interaction between energy carrier prices is mainly of additive nature, ie. S_1 and S_{Ti} are close. The retail tariff of electricity is much more influential than the NG tariff. One possible explanation is that, as mentioned before, there is a greater amount of technology using electricity compared to the NG. Therefore, a low electricity price can help to significantly reduce TOTEX, as both electricity and heat demand can be met using electricity.

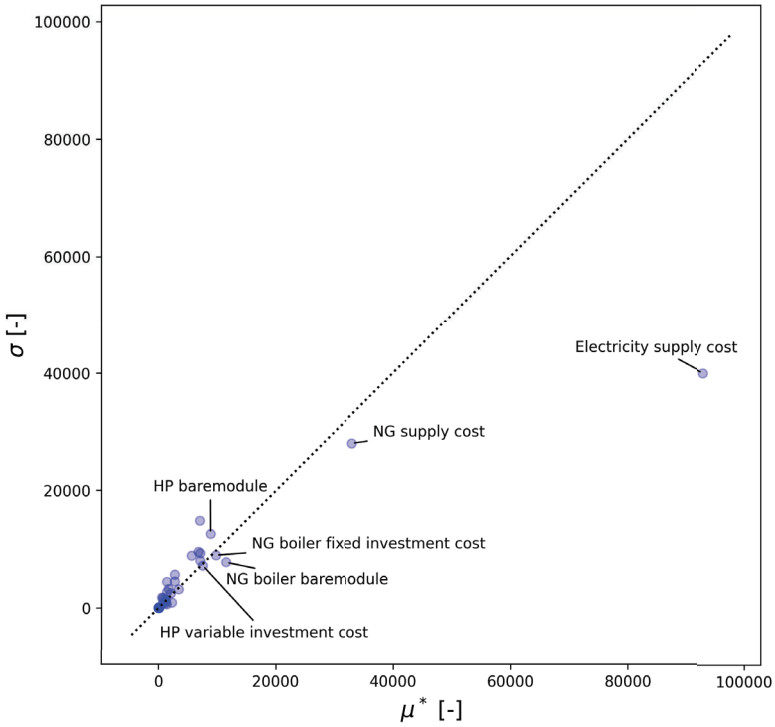

Figure 3: Morris analysis results on $\mu^* - \sigma$ plane.

3.2. Typical configurations identification

First, a Silhouette score is calculated for various numbers of clusters to identify a range of potential optimal values. The score decreases from 1 to 10 and then seems to stabilize, meaning that the optimal number of clusters is $k = 1$. Whereas, for the Elbow method the inflection occurs in the range of $k = 5 - 7$. Finally, the DBCV index is computed for the K-means clustering with k varying from 5 to 11 and the optimal number of clusters appears to be around $k = 10$ with a score of -0.54.

The DBSCAN algorithm recognizes a total of 28 clusters and obtains a DBCV index of -0.12, which is significantly better than the K-means DBCV indices. Regarding the HDBSCAN result, it obtains the best DBCV index with a value of 0.04. However, it identifies 73 clusters, which is far from the initial estimation of $k \simeq 10$.

Looking more precisely at the size distribution of the DBSCAN clusters, one can remark that beyond the tenth cluster, the size dropped abruptly. Indeed, the first ten clusters represent more than 90% of the data points, corresponding to the approximated required number of clusters. Regarding the HDBSCAN distribution, there was no sharp decrease in the cluster size. As a consequence, only data from the first ten DBSCAN clusters were considered for further calculations.

https://doi.org/10.52202/069564-0228

3.3. Presentation of the typical district configurations

3.3.1. Distribution of the configuration in the sampling space

Figure 4 represents with different colors the distribution of the clusters over the retail tariffs variation range. As observed in subsection 3.1., the output of the model, i.e. the district configuration, is strongly correlated to those tariffs. This relation can be observed in the figure below as the different configurations can be identified. Additionally, the sampling space naturally separates itself in half, the separation is highlighted by the amount of NG imports. Configurations 1 to 4 are based on NG and the configurations above, 5 to 10, on electricity. The NG configurations are located in the bottom right corner where the electricity tariff is high. Inversely, the electricity-based ones are in the top left region where the NG price is high. One can note that the space has fewer samples below the separation line, this is due to the data selection done in subsection 3.2..

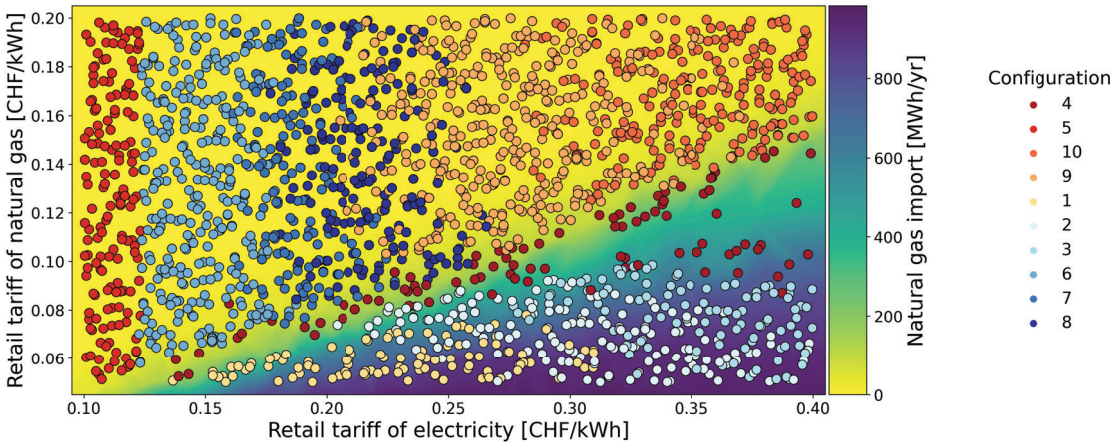

Figure 4: Typical configurations distribution in the retail tariffs space.

3.3.2. Installed units capacity

Figure 5 shows the installed units capacity for each configuration alongside their grid exchanges. The main variation between configurations is the total installed capacity, almost ranging from single to double. This discrepancy is due to the extreme period (high demands and rash environment) included in the model, as the demand is either fulfilled by imports or by an installed unit. The model installs a minimum heating capacity to supply heat in any condition. This minimum heating capacity appears in all configurations and is either composed of NG boiler or a combination of an electrical heater and HP. The additional capacity is formed by various PV capacity deployments.

Concerning the units' installed capacity, there is no positive correlation between HP and PV installation, as one would have expected from previous results discussed in subsubsection 1.2.1.. If no PV panels are installed, electricity imports increase considerably to supply the HP or the electrical heater. The HP installation triggers the deployment of water tanks to serve as a buffer. Large installation of boilers is accompanied by significant imports of NG since boilers are used as the main heat source when they are installed. The presence of HP and PV in most configurations underlines their high potential in district energy systems. Although, the electric grid is more strained with PV implementation as imports are reduced, but exports are increased, requiring a sufficient absorption capacity of the grid.

https://doi.org/10.52202/069564-0228

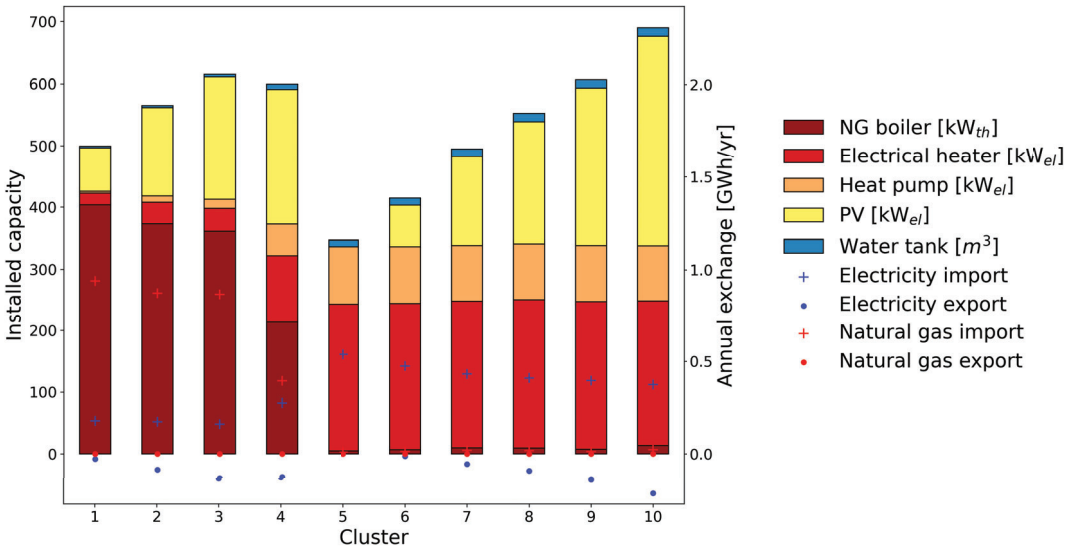

Figure 5: Identified configurations of the district and their grids exchanges.

3.3.3. Dimensionality reduction

Principal component analysis (PCA) is a dimensionality reduction technique used to facilitate the exploration of data sets. The algorithm identifies the eigenvectors of the covariance matrix. The dimensions are sorted from most to least explaining components. Usually, only the first two components are used to project the data points, as can be observed in Figure 6.

The first dimension explains 64.8% of the variance and the second 33.2%, thereby the plot illustrates 98% of the data set variance. The indicators can be regrouped into three main groups which align with the principal components, this is a direct result of having 98% of the variance explained with only 2 dimensions. The first component can be interpreted as the global warming potential (GWP) and the second as the TOTEX. The first deviates slightly from the x-axis, however, the second is perfectly aligned.

A group is positively correlated to the first dimension and contains the NG import, boiler installed capacity, and GWP. Their correlation is natural as the boiler is fueled by NG, which has a high CO_2 emission factor. Inversely, another group is negatively correlated to the first dimension. It contains the installed capacity of the heat pump and electric heater, which are low emissions technologies when coupled with a low-carbon electricity grid like the Swiss one. The electricity import is closely related to the group but helps reduce the TOTEX, negative correlation to the second dimension. This second dimension is highly correlated with the installed PV capacity, electricity export, and TOTEX. The relationship between PV capacity and electricity export stems from the high production potential of the technology in summer, which exceeds the demand. This results in a redistribution of the electricity surplus in the grid. This electricity export is sold at a low price, so the gain from the electricity buyback may not offset the operational cost of the microgrid and the PV installation cost. Note that the influence of the indicators is somewhat similar.

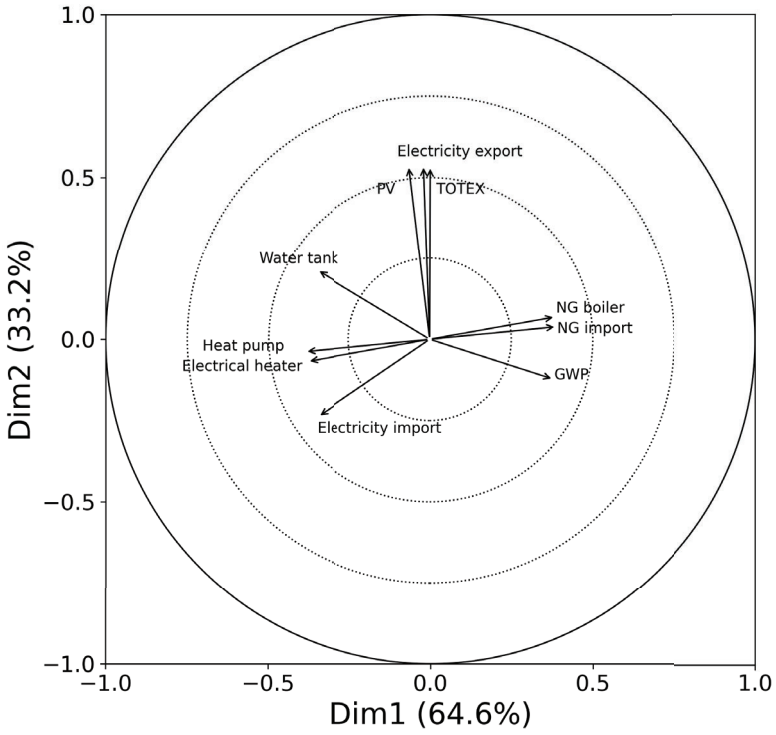

Figure 6: Variable correlation plot of the first two principal components. Indicators pointing in the same direction are correlated, those pointing in the opposite direction are inversely correlated, and those that are perpendicular are uncorrelated. The length of the arrow indicates the influence of the indicator.

4. Conclusion

The objective of this paper was to develop a framework able to identify typical district configurations. The methodology is composed of a two-fold GSA to efficiently sample the space of solutions, and a clustering step to determine typical technical solutions. The Morris method is used to screen input parameters and the sampling scheme is acquired using the Sobol method. The space of solutions obtained from the Sobol sequence was clustered using multiple algorithms. The range of optimal number of clusters was identified using two cluster validation indices and the Elbow method.

- The conclusion of the GSA emphasized the importance of the retail tariff of energy carriers in such systems.
- From 8 to 10 configurations were necessary to represent at best the solution space. The results from each technique were compared using the DBCV index. The tenth first clusters from the DBSCAN were selected as they represent 90% of the data and obtain one of the best score.
- The identified configurations can be separated into two types of systems, those based on NG and the others on electricity. Each type has a pretty basic configuration, i.e. the configurations based on NG install a boiler and the electric ones combine HP and electrical heater.
- In terms of electricity consumption, the import and export of electricity correlate negatively, respectively positively, with the installed photovoltaic capacity.

The presented results indicate that the presented framework allows to efficiently identify districts configurations. The next logical step would be to implement this framework to define a panel of configurations for district representatives of a large-scale energy system and study the impact of optimal built environment in the infrastructure.

https://doi.org/10.52202/069564-0228

5. Fundings

The research published in this report was carried out with the support of the Swiss Federal Office of Energy SFOE as part of the SWEET project acronym. The authors bear sole responsibility for the conclusions and the results of the presented publication.

References

[1] Hannah Ritchie, Max Roser, and Pablo Rosado. CO and greenhouse gas emissions.

[2] GlobalABC, IEA, and UNEP. GlobalABC roadmap for buildings and construction: Towards a zero-emission, efficient and resilient buildings and construction sector.

[3] Daniel M. Kammen and Deborah A. Sunter. City-integrated renewable energy for urban sustainability. 352(6288):922–928.

[4] P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. \{McCollum\}, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley. Climate change 2022: Mitigation of climate change. contribution of working group III to the sixth assessment report of the inter-governmental panel on climate change.

[5] Hans-Kristian Ringkjøb, Peter M. Haugan, and Ida Marie Solbrekke. A review of modelling tools for energy and electricity systems with large shares of variable renewables. 96:440–459.

[6] Rahul Gupta, Fabrizio Sossan, and Mario Paolone. Countrywide PV hosting capacity and energy storage requirements for distribution networks: The case of switzerland. 281:116010.

[7] Gauthier Limpens, Stefano Moret, Hervé Jeanmart, and Francois Maréchal. EnergyScope TD: A novel open-source model for regional energy systems. 255:113729.

[8] Kody M. Powell, Akshay Sriprasad, Wesley J. Cole, and Thomas F. Edgar. Heating, cooling, and electrical load forecasting for a large-scale district energy system. 74:877–885.

[9] Leander Kotzur, Peter Markewitz, Martin Robinius, Gonçalo Cardoso, Peter Stenzel, Miguel Heleno, and Detlef Stolten. Bottom-up energy supply optimization of a national building stock. 209:109667.

[10] P. Stadler, Luc Girardin, and F. Maréchal. The swiss potential of model predictive control for building energy systems.

[11] M. Di Somma, B. Yan, N. Bianco, G. Graditi, P. B. Luh, L. Mongibello, and V. Naso. Operation optimization of a distributed energy system considering energy costs and exergy efficiency. 103:739–751.

[12] M. Di Somma, B. Yan, N. Bianco, G. Graditi, P. B. Luh, L. Mongibello, and V. Naso. Multi-objective design optimization of distributed energy systems through cost and exergy assessments. 204:1299–1316.

[13] Hongbo Ren, Weisheng Zhou, and Weijun Gao. Optimal option of distributed energy systems for building complexes in different climate zones in china. 91(1):156–165.

[14] David A. Copp, Tu A. Nguyen, Raymond H. Byrne, and Babu R. Chalamala. Optimal sizing of distributed energy resources for planning 100% renewable electric power systems. 239:122436.

[15] Jonas Allegrini, Kristina Orehounig, Georgios Mavromatidis, Florian Ruesch, Viktor Dorer, and Ralph Evins. A review of modelling approaches and tools for the simulation of district-scale energy systems. 52:1391–1404.

[16] Boran Morvaj, Ralph Evins, and Jan Carmeliet. Optimising urban energy systems: Simultaneous system sizing, operation and district heating network layout. 116:619–636.

[17] Azadeh Maroufmashat, Ali Elkamel, Michael Fowler, Sourena Sattari, Ramin Roshandel, Amir Hajimiragha, Sean Walker, and Evgueniy Entchev. Modeling and optimization of a network of energy hubs to improve economic and emission considerations. 93:2546–2558.

[18] Georgios Mavromatidis, Kristina Orehounig, and Jan Carmeliet. Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. 214:219–238.

[19] Tianjie Liu, Wenling Jiao, and Xinghao Tian. A framework for uncertainty and sensitivity analysis of district energy systems considering different parameter types. 7:6908–6920.

[20] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer*. Wiley. Edition: 1.

[21] Wei Tian. A review of sensitivity analysis methods in building energy analysis. 20:411–419.

[22] Paul Westermann and Ralph Evins. Surrogate modelling for sustainable building design – a review. 198:170–186.

[23] Torben Østergård, Rasmus L. Jensen, and Steffen E. Maagaard. Early building design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. 142:8–22.

[24] Max D. Morris. Factorial sampling plans for preliminary computational experiments. 33(2):161–174. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1991.10484804.

[25] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. 145(2):280–297.

[26] J {and} Saltelli A Campolongo, F {and} Cariboni. Sensitivity analysis: the morris method versus the variance based measures.

[27] Will Usher, Jon Herman, Calvin Whealton, David Hadka, Xantares, Fernando Rios, Bernardoct, Chris Mutel, and Joeri Van Engelen. Salib/salib: Launch!

[28] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. 181(2):259–270.

[29] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pages 226–231. AAAI Press.

[30] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. Density-based clustering based on hierarchical density estimates. In Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science, pages 160–172. Springer.

[31] Leland McInnes and John Healy. Accelerated hierarchical density based clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 33–42. ISSN: 2375-9259.

[32] Davoud Moulavi, Pablo A. Jaskowiak, Ricardo J. G. B. Campello, Arthur Zimek, and Jörg Sander. Density-based clustering validation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 839–847. Society for Industrial and Applied Mathematics.

[33] Luise Middelhauve, Cedric Terrier, and Francois Marechal. Decomposition strategy for districts as renewable energy hubs. page 10.

[34] Federal Statistical Office. Géodonnées etat de vaud, registre cantonal des bâtiments (RCB).

[35] Federal Statistical Office. Federal register of buildings and dwellings.

[36] OpenStreetMap contributors. Planet dump.

[37] J. Remund et al. METEONORM - global meteorological database for engineers, planners and education.

[38] SIA. vernehmlassung sia 380/1 heizwärmebedarf.

[39] Luc Girardin. A GIS-based methodology for the evaluation of integrated energy systems in urban area.

[40] Federal Office of Topography. swissBUILDINGS3d 2.0.

[41] Swiss Federal Office of Energy Geoinformation. Switzerland in 3d.

[42] Administration cantonale vaudois. Distribution networks database.

[43] Middelhauve, Luise. On the role of districts as renewable energy hubs.

 https://doi.org/10.52202/069564-0228