EPFL

IPESE
Industrial Process
and Energy Systems
Engineering

A framework to identify technology solutions at district level

Application of levelized infrastructure-connected regionalisation in energy systems modelling

Semester Project

Author:
Arthur CHUAT

Supervisors:
Prof. François MARÉCHAL
Jonas SCHNIDRIG
Cédric TERRIER

Industrial Processes and Energy Systems Engineering
Ecole Polytechnique Fédérale de Lausanne

Sion, 20th June 2022

Abstract

A proper assessment of technologies' impact on energy consumption and GHG emissions is essential for designing an effective energy transition. In this regard, the modeling of an energy system is a great resource to identify valuable technologies. This paper falls within a larger project that aims to optimize a national scale energy system based on two encapsulated subsystems (building and district scale).

The following paper presents a framework to identify typical configurations of a district energy system. The framework is composed of a two-step GSA. The first step identifies the most influential parameters on the model output using Morris method. The second allows to obtain a representative sampling of the global solution space using the variance-based Sobol method.

The GSA suggests that the sensitivity of the model comes primarily from energy carrier tariffs, while the investment cost and other technology properties weigh little in the model output. Furthermore, the space of optimal district was clustered using multiple techniques. The most coherent results were obtained with a DBSCAN which allowed to identify 10 different typical configurations.

The configurations heat supply is either based on electricity, using HP and electrical heater, or on NG boilers. Regarding the electricity needs, the supply strategy is identical for all configurations. They rely on a combination of PV panels and imported electricity. Finally, the HP and the water tank are coupled in all electric configurations to furnish heat where NG boilers do not require storage unit.

Contents

List of Figures

List of Tables

Acronyms

CAPEX capital expenditure.

CHP combined heat and power.

DBCV density-based cluster validation.

DBSCAN density-based spatial clustering of applications with noise.

DES distributed energy system.

EE elementary effect.

EPFL Ecole Polytechnique Fédérale de Lausanne.

ES Energy Scope.

GHG green-house gas.

GSA global sensitivity analysis.

GWP global warming potential.

HDBSCAN hierarchical density-based spatial clustering of applications with noise.

HP heat pump.

IPCC Intergovernmental Panel on Climate Change.

IPESE Industrial Processes and Energy Systems Engineering.

KPI key performance indicators.

LHS latin hypercube sampling.

LSA local sensitivity analysis.

MPC model predictive control.

NG natural gas.

OPEX operational expenditure.

PCA principal component analysis.

PV photovoltaic.

RE renewable energy.

REHO Renewable Energy Hub Optimizer.

SA sensitivity analysis.

SDG sustainable development goal.

SSE sum of squared error.

TOTEX total expenditure.

Chapter 1

Introduction

As declared in the sixth IPCC's report on the Mitigation of Climate Change "*Meeting the long-term temperature objective in the Paris Agreement implies a rapid turn to an accelerating decline of GHG emissions towards 'net zero', which is implausible without urgent and ambitious action at all scales*" [1]. This statement highlights the urgent need to drastically reduce green-house gases (GHG) emissions. Meanwhile, emissions from the energy sector, which accounts for 73% of global emissions, are expected to increase over the coming decades [2]. Therefore, as mentioned, efforts must be taken at all scales.

The built environment is a major player in the energy sector, contributing to 36% of the final energy consumption in 2018 [3], but it is also seen as a promising pathway towards a sustainable energy mix. Indeed, the sector has the potential to significantly reduce its emissions. It can be achieved by considering a multi-energy system, increasing its inter-connectivity within and outside its borders and naturally by developing its renewable energy (RE) shares.

However, such transition requires careful energy planning from policy makers such as government signatories of the Paris Agreement. Energy system models are excellent tools for understanding such complex systems. For that reason, they support decision-makers outline their strategy for the energy transition. These models can assess the potential impact of deploying new technologies or policies on the current system.

A promising solution to achieve such sustainable system is the development of distributed energy system (DES). Distributed energy systems are composed of energy hubs, interconnected via a multi-energy grid. These energy hubs are composed of energy conversion and storage units, all optimally sized and combined with optimal system operation. Such systems have a great potential when deployed at the district scale as they increase self-consumption and help mitigate grid congestion [4].

This project aims to develop a framework for establishing a panel of optimal district configurations. They are then intended to be fed into a second energy model for optimizing a national energy system,

which will consider typical districts and their different specific configurations.

Chapter 2

Literature review

This chapter assesses the current state-of-the-art regarding energy system modelling (section 2.1) and the application of sensitivity analysis of energy system models (section 2.2). Additionally, diverse clustering techniques are presented in section 2.3. Finally, the goal and scope of the project is presented (section 2.4).

2.1 Energy system modelling

Energy models are mathematical formulations of energy systems. They are one of the main supports to guide decision makers in the ongoing energy transition towards a fossil-free system. The decarbonization of the current system requires a profound restructuring and involves a deeper electrification [5]. Low-carbon technologies, such as solar and wind power, are spreading rapidly. They are expected to become the main source of electricity, as their potential is high [1, 6]. 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 [7]. 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.

There exist many energy system models and all differ in some way. It can either be a simulation or an optimization of the system. A simulation has no degrees of freedom and only provides an evaluation of the model, whereas an optimization finds the optimal solution among the space of possible solutions (several degrees of freedom). The optimality criterion of a problem is defined by its objective function, e.g. minimizing the GHG emissions.

Another large variation between models is the considered energy vectors to fulfill a specific demand. In most cases, only one resource is used to meet the needs. For example, the optimization of a district heating network can be feed by waste recovery without considering the possible local implementation of solar thermal, heat pump or boiler. Likewise, the majority of models focus on one specific energy sector (electricity, heat, mobility, etc.) and do not account for cross-sectoral synergy [8]. Fi-

nally, 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 system, e.g., interconnected buildings to represent a district system

2.1.1 Large-scale energy system modelling

The spatial resolution at which the system is evaluated is a major characteristic of the system. It allows to define the different technologies considered, the energy vectors taken into account or the possible optimization of the system operation [6, 8, 9]. Jalil-vega and al. [10] analyzed the effect of spatial resolution on a large scale energy system. They conclude that the effect of spatial resolution was related to the homogeneity of the region, i.e. a heterogeneous region would be better captured by a thin resolution. Furthermore, their results highlight that finer resolution improves the network design when using optimization models.

The following subsections present three different scopes of energy system. The first is based on the spatial distribution of resources, the second on identical buildings, and the third on the configuration of the district energy system.

2.1.2 Geographic energy hub

The European Commission defines the term Geo-clusters as *"virtual trans-national areas where strong similarities are found in terms of climate, culture and behaviour, construction typologies, economy, energy price and policies, gross domestic product..."*. A geo-cluster can be interpreted as a region in which the selected indicators are homogeneous. The GE2O project, supported by the European Commission, has defined a panel of energy efficient solutions for each identified geo-cluster in Europe. For their implementation, technological and non-technological aspects of the region, such as the wind generation potential and the age distribution within the region, were taken into account [11]. Kuster and al. [12] defined 118 different geo-clusters within Europe, accounting for 16 parameters such as building types, climate and socio-economic indicators. They developed two tools to help users select appropriate technologies for their location by providing case studies with a similar environment.

Germano used political delimitations to discretize Switzerland and provided a regional optimization of the country [13]. An adaptation of its regionalization has been performed by Slaymaker and al. [14]. The paper proposed a discretization of the space using clustering method based on geospatial data. Geographical characteristics were used, such as wind, solar and photovoltaic potential, as they are location dependent. As well as, demographic characteristics such as population density, urban, agricultural and ecological areas, as well as distance to the power grid.

2.1.3 Building archetypes

In 2019, close to a third of the final energy demand (31%) came from the building sector and its CO_2 emissions have increased by 50% in the past 30 years to reach $12GtCO_2$-eq in 2019 (21% of global GHG emissions) [1]. This motivates the will to define large-scale energy system with a building-scale resolution [15]. Kotzur and al. [16] 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 solution. The model was validated with a case study in Germany, and emphasized the importance of PV and HP deployment to reduce GHG emissions. However, those new technologies tend to increase the gap between electricity overproduction in summer and demand in winter.

Stadler and al. [17] assessed the impact of 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 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.

2.1.4 District energy hub

As predicted by the IPCC's report [1], the share of the world population living in urban areas continues to increase. This trend requires the improvement of urban energy systems to mitigate GHG emissions and energy consumption. Distributed energy systems are part of the solution to help improve sustainable development goals (SDGs)[18, 19, 20, 21]. The distributed aspect of such system 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 system by Allgerini and al. [22], there has been a significant improvement of the models and tools used to analyze such system.

In their paper, Morvaj and al. [23] 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 GHG emissions and TOTEX. Maroufmashat and al. [24] 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. And that the operation of interconnected energy hubs can significantly increases the robustness of the power grid, e.g. mitigation of congestion and ensuring reliability.

On the other hand, increasing the size of the considered district has a direct impact on the computational time of the model resolution [25, 26]. Hence, new solving methods are needed to improve

optimization efficiency without reducing spatial and temporal data, which oversimplifies the problem [26, 27].

2.2 Global sensitivity analysis

Sensitivity analysis (SA) is a powerful tool to assess how a set of input parameters can affect the output of a model. There exist two main types of SA: local and global. The most common method is the local sensitivity analysis (LSA), which evaluates the sensitivity by varying one parameter at the time around a specific value. 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 sample space by varying several parameters at once. In doing so, it is able to capture parameter interaction.

Most current models assume a perfect knowledge of the input parameters, which induces the absence of uncertainty in the model and makes them deterministic [19]. However, there is some uncertainty in assessing various aspects such as current policy, renewable energy production and economic trends. Therefore, specific models are developed to consider the randomness of input parameters: stochastic models. An alternative to stochastic model is the application of SA to deterministic model as proposed in [28, 29].

The review of the SA method for building energy systems performed by Tian and al. [30] 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 the familiarity with the methods.

Another review was performed by Westermann and Evins [31], 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.

Liu and al. [29] developed a framework to assess the uncertainty and sensitivity of district energy systems. The framework includes a two-stage GSA. First, a screening method filters out non-influential parameters and to reduce computational cost. Secondly, the sensitivity of the influential parameters were evaluated with a SA. The methodology was further validated with a case study of a DES in China.

Mavromatidis and al. [28] applied the same methodology to a DES. The screening method used to reduce the number of parameters was the Morris method. It was coupled with a Sobol sequence to obtain a proper sampling of the space for the final sensitivity analysis. The results showed that the variation from the system optimal cost comes from energy carriers prices and energy demand pat-

terns. Concerning the installed units, a combination of heat pumps, cogeneration units and boiler achieved the lowest cost and a phase-out of CHP was required in order to reduce GHG emissions.

Finally, Østergård and al. [32] used a SA to explore multidimensional design space to help decision making regarding sustainable building design. The identical framework presented earlier was used and provided a portfolio of possible building conception. The author developed a tool to select optimal designs based on specific parameter ranges.

2.3 Clustering

Clustering is used to explore data sets and attempt to find their structure. It reassembles data points with similar attributes into a cluster. There exist two main types of aggregation technique: supervised and unsupervised (machine) learning. The supervised methods are provided with an object which is already labeled, its role is to develop an algorithm to link a new object to a label. Inversely, the unsupervised learning, also known as clustering, is not provided with any label. The algorithm has to identify the groups (clusters) based on the features of each data point. The number of clusters is not necessarily specified by the user. As labels are not identified with unsupervised learning, it is considered more difficult than supervised learning. Unsupervised learning methods can be further divided in two different groups: hierarchical and partitional. The partitional algorithms simply divide the input data space in k clusters. Whereas, the hierarchical clustering method develops a tree-shaped structure of the data, also called dendrogram, to form nested sets of data.

A clustering algorithm can be formulated as a minimization problem where the objective function represents the sum of the distance between each data point within the same cluster [33]. Lloyd and al. [34] proposed the K-means clustering technique. It identifies k clusters among the data points using the algebraic mean as the cluster center. There exist many different methods to define the cluster center such as K-medoids using the most centered points as the cluster center [35] or K-medians using the median as the cluster center [36].

However, these methods require that the data space is convex, which cannot always be verified. Under these conditions, more robust algorithms based on density clustering are essential to aggregate arbitrary shaped data. The data distribution can come from several density functions and with different parameters [37]. Thus, Cheeseman and Sutz [38] developed AUTOCLASS, an algorithm to detect clusters based on the distribution of the data. Ester and al. [39] developed a density-based method named DBSCAN for *Density-based spatial clustering of applications with noise*, which is able to detect non-convex shapes among large data sets. Campello and al. [40] proposed an hierarchical version of DBSCAN. The algorithm is named HDBSCAN and is more robust to outliers.

The application of the aforementioned algorithms in the context of district energy systems is quite rare. One can mention the study published by Felsmann and al. [41], which examined multiple clustering techniques such as Single-Linkage, DBSCAN and OPTICS on a district heating network. The results showed that the DBSCAN robustness outperformed the other algorithms. In order to identify typical buildings Stadler [27] used the K-medoid technique to determine archetypes building among the entire Swiss building stock. The K-medoid was preferred to K-mean as it produced more robust results [42].

Since the choice of the clustering method is a problem in itself, it is a good practice to introduce a cluster validation index to compare several algorithms. Arbelaitz et al. produced an exhaustive comparison of cluster validation indices [43]. 30 indices were compared over several sets of configurations. The researchers created a tool to find the best suited index for a specific application, although the Silhouette index often performed the best. Moulavi and al. [44] introduced a novel index to evaluate the quality of density-based clustering as most of the validation index are developed for globular clusters.

2.4 Goal and scope

As emphasized throughout this chapter, the literature lacks an application of distributed energy system as energy hub in a large-scale energy system optimization. This project attends to provide a framework to identify a set of optimal district configurations. This framework consists of a GSA of a district-wide energy system model to obtain various district configurations. They will then be aggregated to identify robust typical configurations. Those configurations are then intended to feed a larger scale energy system (national scale), now considering various typical districts and their configurations panel as energy hubs. To summarize, the overall project attempts to define an optimal national energy system based on two optimal sub-systems. The finer resolution is the building scale which then forms a district, each building included in the district is optimal. Finally, the national system has access to multiple optimal district configurations defined thanks to this framework.

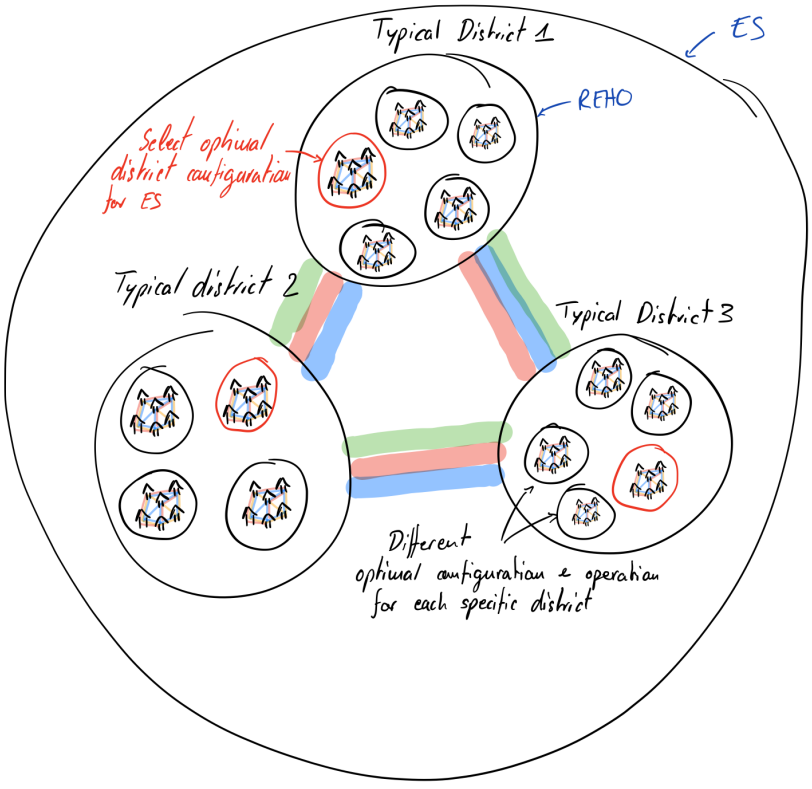

Figure 2.1: Scheme of the overall project. This project provides the tool to define the different configuration for each typical district.

Chapter 3

Methodology

This chapter presents the global sensitivity analysis used to best sample the solution space. First, a screening method is presented to reduce the computational work by considering only the most influential model parameters. Then, the sampling method is introduced to efficiently cover the space of the selected input parameters. Finally, the different clustering techniques are presented to identify the typical configurations.

3.1 Global sensitivity analysis

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. [45] which assesses the state-of-the-art of GSA. A sensitivity analysis can be decomposed in 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. LSA: 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. GSA: a variance-based method is used to quantify the sensitivity of the model

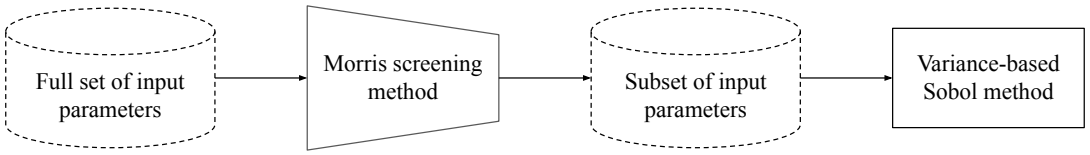

Figure 3.1: Scheme of the two-step GSA composed of first the Morris method as a screening and then the Sobol method is applied on the most influential parameters

3.1.1 Morris screening method

The screening method used is the Morris method, it allows to qualitatively compare the influence on the model output of a large number of parameters with a few evaluations [46]. 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. Then, it performs a one-step-at-the-time method, i.e. it randomly modifies an input parameter by $\pm\Delta$ to generate r trajectories.
Then, it evaluates the elementary effect of the i_{th} input factors (EE_i) as a function of the model output $Y = f(X_1, ..., X_n)$, see Equation 3.1. 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.

$$EE_i = \frac{[Y(X_1, X_2, \ldots, X_{i-1}, X_i \pm \Delta, \ldots X_k) - Y(X_1, X_2, \ldots, X_k)]}{\Delta} \tag{3.1}$$

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 [45]. 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 $r = 10$ [47] whereas Morris used $r = 4$ in [46], which seems to be the minimum usable value.

In its original work, Morris proposed the computation of the mean μ_i (Equation 3.2) and standard deviations σ_i (Equation 3.4) 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. Thus, the method has been improved by Campolongo and al. [48], by considering the absolute mean elementary effect μ_i^* (Equation 3.3).

$$\mu_i = \frac{1}{r} \sum_{j=1}^{r} EE_i^j \tag{3.2}$$

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} |EE_i^j| \tag{3.3}$$

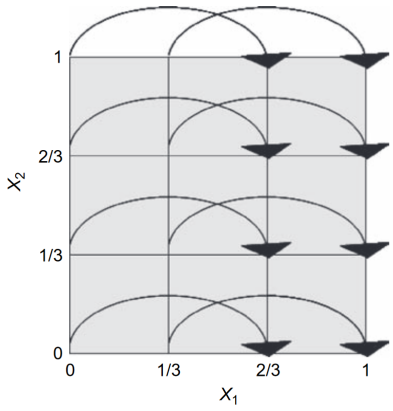

Figure 3.2: Illustration of a k-dimensional hypercube, with $k = 2$ and a four-level grid ($p = 4$) [45]. The arrows represent the 8 points required to estimate the EE of X_1.

$$\sigma_i^2 = \frac{1}{r-1}\sum_{j=1}^{r}\left(EE_i^j - \mu_i\right)^2 \tag{3.4}$$

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 strongly dependent to the sampling point, i.e. to the other values of the input parameters. Conversely, a low value of σ indicates that the elementary effect is not subject to vary with other factors.
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 3.3, 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

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^*.

3.1.2 Sobol sampling sequence

The Sobol method is a variance-based sensitivity analysis named after the mathematician Ilya M. Sobol. The method is developed in [45, 49] and uses Sobol's recommendation on the sequencing of

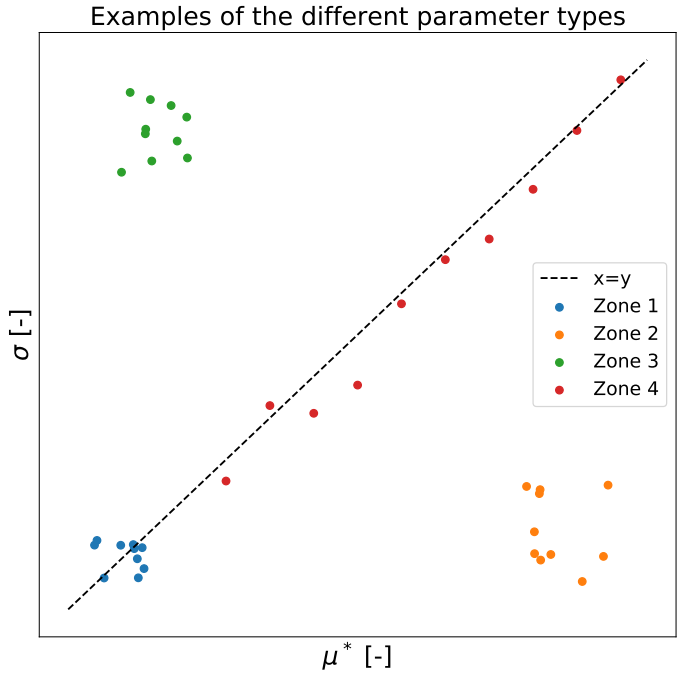

Figure 3.3: Identification of the typical zone on the $\mu^* - \sigma$ plane

quasi-random numbers. However, Saltelli extended her work in [50] to reduce the error rate when computing the sensitivity index. Once the Monte Carlo optimizations of the model are completed the Sobol index can be defined.

Sensitivity indices

The method evaluates two different sensitivity indices, the first one is the first-order sensitivity coefficient (Equation 3.5). It results from the ratio of the variance of the output mean, considering all parameters except the ith, and the variance of the output. The secondary sensitivity coefficient is the total effect index, i.e. first and higher-order sensitivity coefficient (Equation 3.6).
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.

$$S_i = \frac{V_{X_i}\left(E_{\mathbf{X}_{\sim i}}(Y \mid X_i)\right)}{V(Y)} \tag{3.5}$$

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}\left(V_{X_i}(Y \mid \mathbf{X}_{\sim i})\right)}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{\sim i}}\left(E_{X_i}(Y \mid \mathbf{X}_{\sim i})\right)}{V(Y)} \tag{3.6}$$

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 effect 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 [45], it is suggested to choose a power of two.

3.2 Clustering

This section provides the different steps to define the typical district configurations. First, it is essential to standardize the data. Then, the different clustering techniques are presented: K-mean, DBSCAN and HDBSCAN. As the K-mean method requires the number of clusters, two methods are used to identify a range of optimal clusters: the Silhouette score and the Elbow method. Additionally, a cluster validity index is introduced to compare the various clustering technique results.

3.2.1 Standardization and feature 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 3.7. Each features X_i of the data set is standardized as follow:

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \tag{3.7}$$

where μ_i is the mean and σ_i is the standard deviation of the feature X_i. The features selected for the clustering, i.e. the district characteristic, are a mix of economic and technical attributes of the optimization result. The key performance indicators (KPI) chosen are the 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 electricity grid.

3.2.2 K-means

The K-means method minimizes the clusters variance, which is an NP-hard problem. However, the implementation of heuristic algorithms allows to efficiently reach a local optimum. The K-means problem is formulated as follows:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2 \tag{3.8}$$

where, $\mathbf{S}$ represents the k sets of points.

Two different techniques are used to evaluate the required number of clusters: the silhouette score and the Elbow method.

Silhouette score

The Silhouette score evaluates the cohesion and separation of the different clusters, i.e. the points density within a cluster and the distance between clusters. Here, the metric used to evaluate the distance between two points i and j is the Euclidean distance, $d(i,j)$. Two subscores are used to evaluate the Silhouette score. Firstly, a_i measures the quality of the assignment of a data point i to its cluster C_I.

$$a(i) = \frac{1}{|C_I| - 1} \sum_{j \in C_I, i \neq j} d(i,j) \tag{3.9}$$

The second subscore represents the mean dissimilitude between a data point $i \in C_I$ and another cluster C_J.

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i,j) \tag{3.10}$$

The silhouette score for a data point i can be computed as follow:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{3.11}$$

Its value is between -1 and 1, where 1 means that the point is well grouped and conversely -1 indicates that the point would be better represented in a different cluster. A null value means that the point is on the border of two clusters. Finally, the average of the Silhouette scores of all data points can be used to evaluate the effectiveness of the clustering.

Elbow method

The Elbow method, as its name indicates, consists to observe an inflexion on a k-SSE plot, where SSE stands for sum of squared error. The squared error is the sum of the squared distance, here Euclidean, between the points and their cluster center. The optimal number of clusters is found where the reduction of SSE sharply stops decreasing as the number of clusters increases.

3.2.3 DBSCAN

DBSCAN 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 distance of ϵ. 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 as outliers. The choice of *minPts* and ϵ should be based on the data properties [39].

3.2.4 HDBSCAN

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 [40, 51].

3.2.5 Density-based clustering validation

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. [44]. It is based on a parameterless core distance defined on the density of objects and mutual reachability. The index ranges between -1 and 1, which represents a good and a bad score respectively.

3.3 Application

The framework detailed above is applied to a district energy system. The buildings chosen to form the district are located in Rolle, Switzerland, see Figure 3.4. 15 buildings, constituted of 5 single family houses and 10 apartment buildings, have been selected among 30 available as the district synergy seems to appear on tens or more buildings district [26]. The data are either open source or provided by the Swiss government or the canton of Vaud. Dwellings' characteristics such as type and year of construction, reference energy area, height, etc., come from the cantonal and federal Official Building Registry [52, 53] whereas the ground surface area is provided by the cantonal administration [53, 54]. The climate data, solar irradiation and temperature, are extracted from Meteonorm [55]. The characteristics of the building envelope are calculated following SIA norms 380/1 [56]. The PV orientation is optimized based on roofs and façades information extracted from Swisstopo [57, 58]. The grid specifications and electricity demand are furnished by Romande Énergie [59].

The time series are clustered into 10 typical periods formed of 24 hours and 2 extreme periods of 1 hour each, thus 242 different timesteps are considered. Several parameters are fixed during the problem definition, such as units parameters, district parameters, energy carrier price, see chapter 6 for additional information.

The used model for the district level optimization is REHO, which stands for Renewable Energy Hub Optimizer, developed in [27, 61, 62]. Regarding the large-scale energy model, EnergyScope will be used to optimize the collection of typical districts [8].

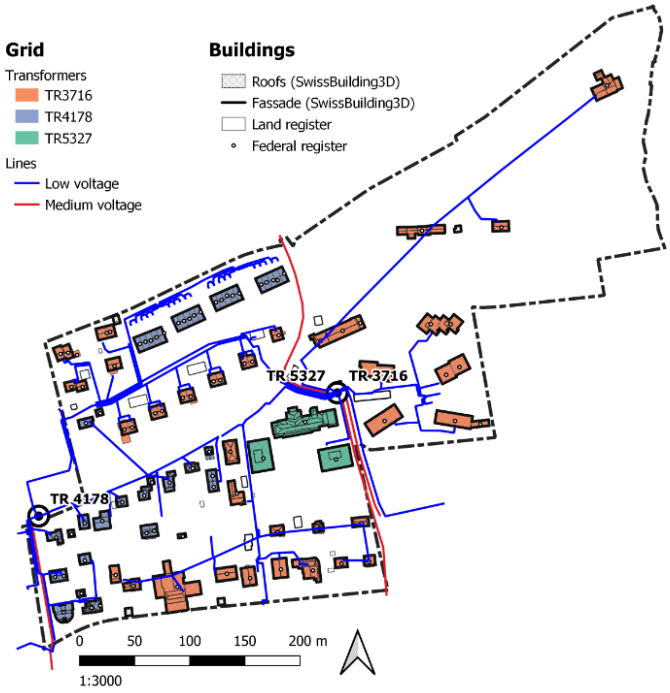

Figure 3.4: Map of the district with the electrical network, typical European residential configuration [60]. The electrical network is connected to a single transformer and the total energy reference area is of 7'755 m^2.

Chapter 4

Results and discussion

The results are separated in two distinct sections. The section 4.1 presents the results of the two-steps GSA. First the most influential parameters are identified using the screening method of Morris. Secondly, the Sobol method is used to properly sample the solution space and to determine the sensitivity of the model with respect to the most influential parameters. The section 4.2 identifies and analyzes the district configurations.

4.1 Global sensitivity analysis

The subsection 4.1.1 presents the results of the screening step of the SA and subsection 4.1.2 presents the final results of the SA obtained with the Sobol method.

4.1.1 Parameters screening

This subsection discusses the results obtained during the screening phase of the SA. As discussed in subsection 3.1.1, the Morris method is performed in order to reduce the number of considered input parameters for the GSA. The considered parameters for this step are energy conversion and storage units parameters as well as the energy carriers prices. The parameters are described more explicitly in section 6.3. 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 total expenditure (TOTEX).

The extended results of the Morris method can be found in Table 6.4. However, as discussed in subsection 3.1.1 a good practice is to plot the absolute mean value (μ^*) and standard deviation (σ) of the EE distribution of each input parameter.
Figure 4.1 shows such a plane for the six non-zero results of the Morris method. 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. Its low dependence comes from its key role

in the district model. It serves to directly supply the electric demand when no installed units can fulfill it (PV, CHP) and it can also serve to supply the heat demand via HPs and electrical heaters. The retail price of NG is less influential as only two units consider natural gas as a resource (NG boiler and CHP). 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 boiler. Finally, one can notice that the technology properties have no influence in model output.

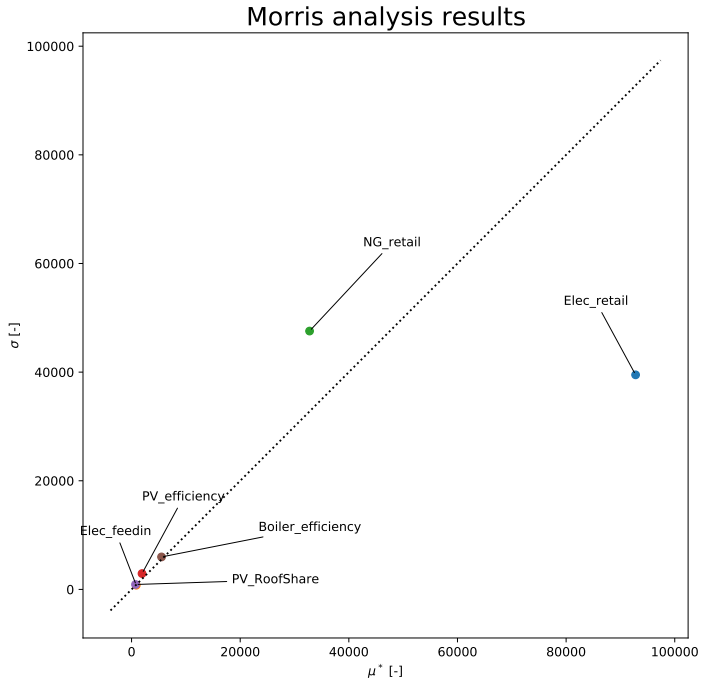

Figure 4.1: Morris analysis results on $\mu^* - \sigma$ plane

Table 4.1 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 boiler efficiency will always reduce TOTEX and, conversely, an increase in electricity cost will always increase TOTEX.

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 chapter 2.

4.1.2 Variance based sensitivity analysis

This subsection focuses on the final sampling of the chosen input parameters using Sobol sequence. The input parameter space is explored with 2560 samples following recommendations presented in subsection 3.1.2. The parameters variation range can be found in Table 6.5. The sensitivity indices of the Sobol can be observed in Figure 4.2 alongside Morris μ^* values for comparison. Detailed val-

Parameter	μ	μ^*	σ
Elec_retail	92793	92793	39500
Elec_feedin	-861	861	776
NG_feedin	32771	32771	47561
PV_efficiency	-1926	1926	2917
PV_RoofShare	-726	726	905
Boiler_efficiency	-5516	5516	5978

Table 4.1: Results of the Morris method for the influential parameters

ues are located in Table 6.6. The Sobol results have a similar trend than the qualitative sensitivity indices from the Morris method. The electricity feed-in price is less influential regarding Sobol's results. Since the first-order Sobol effect and the total effects have a similar value, the interaction between energy carrier prices is mainly of additive in nature. The retail price of electricity is much more influential than NG. One possible explanation is that, as mentioned, there is a greater amount of technology using electricity compared to NG. Thus, a low electricity price can help to significantly reduce TOTEX, as both electricity and heat demand can be met using only electricity. Whereas the maximum installable size of the CHP prevents this behaviour.

Figure 4.2: Comparison of Morris and Sobol sensitivity indices

4.2 Typical configurations identification

This subsection presents the results of the applied clustering techniques as well as the analysis of the various district configurations.

First, the Silhouette score is calculated for various number of cluster to identify an optimal value, see Figure 6.2. The reduction of the score seems to stabilize around $k = 10$. Whereas, for the Elbow

method the inflection occurs in the range of $k = 5 - 7$, see Figure 6.3. Finally, the DBCV index is computed for the K-means clustering with k varying from 5 to 11. The optimal number of clusters appears to be around $k = 10$ (Figure 4.3) with a score of -0.54.

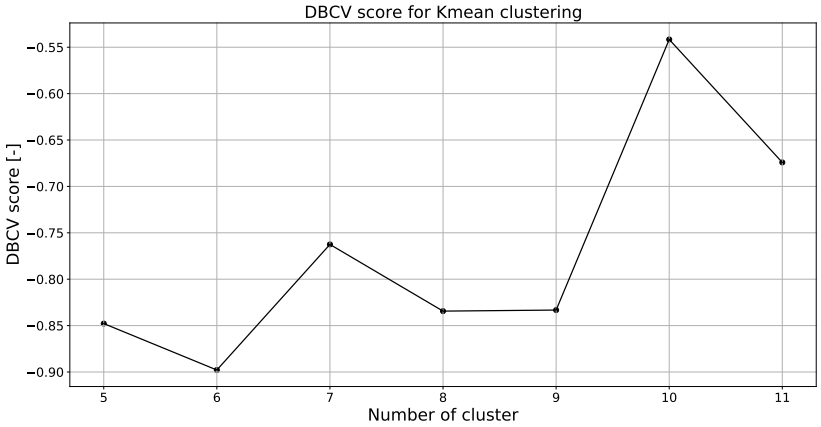

Figure 4.3: Density-based clustering validation indexes evolution with k for K-mean

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 indexes. 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 $k = 10$ obtained previously.
Looking more precisely at the size distribution of the DBSCAN clusters, one can remark that beyond the tenth cluster the size dropped abruptly, see Figure 6.4. 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 of the clusters size. As a consequence, only the data in the then first cluster of the DBSCAN have been considered for further calculation.

4.3 Presentation of the typical district configurations

This section presents the identified configurations. First, the breakdown of the clusters within the sampling space is discussed. Then, the composition of the configurations is investigated, to finally show the correlation among the district indicators.

4.3.1 Distribution of the configuration in the sampling space

Figure 4.4 represents with different colors the distribution of the clusters over the retail tariffs variation range. As explained in subsection 4.1.2, 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 clearly be identified. The space naturally separates itself in half, the separation

is highlighted by a dashed line just above configuration number 1 in Figure 4.4. The configuration below the line, configurations 1, 5, 6, and 7 are based on natural gas and the configurations above, configuration 2, 3, 4, 8, 9 and 10, on electricity. The NG configurations are located on 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 less samples below the separation line, this is due to the data selection done in section 4.2.

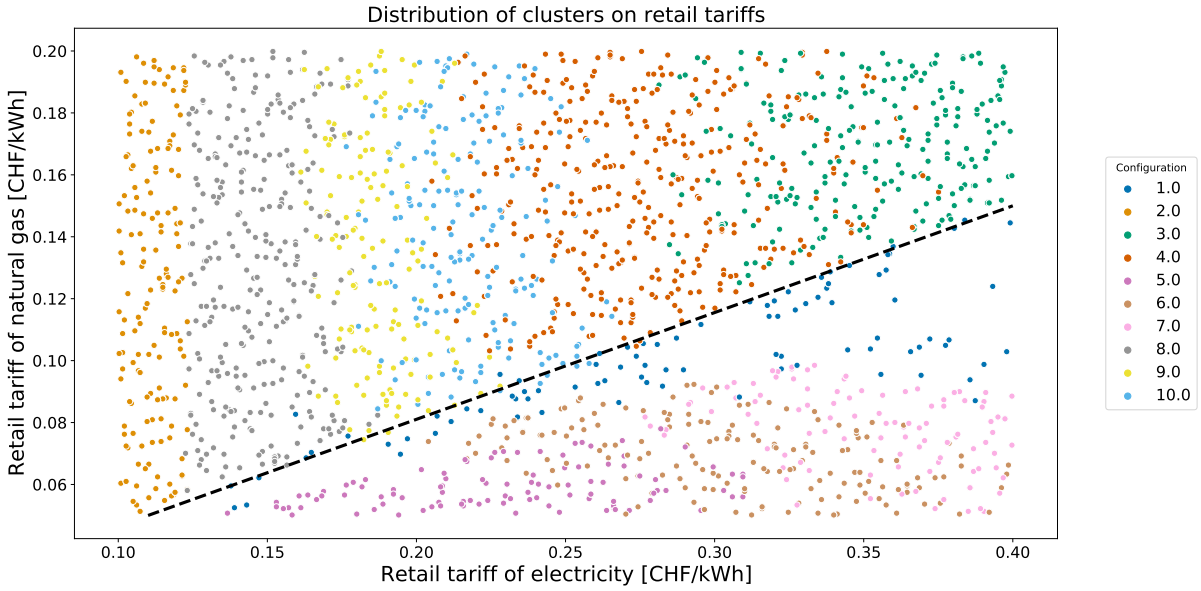

Figure 4.4: Typical configurations distribution in the retail tariffs space

Since the considered parameters for the SA were the prices of energy carriers. It is irrelevant to compare the configurations TOTEX as some of them based mainly their operation on low retail tariff, i.e. configuration 2.

4.3.2 Installed units capacity

Figure 4.5 shows installed units capacity for each configuration alongside 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. 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 electrical heater and heat pump.

Concerning the units' installed capacity, there is no positive correlation between heat pump and PV installation, as one would have expected from previous results discussed in section 2.1. If no PV panels are installed, electricity imports increase considerably to supply the heat pump with electricity. The heat pumps installation triggers the deployment of water tanks to serve as a buffer. The install-

ation of boilers is accompanied by significant imports of natural gas. Since, boilers are used as the main heat source when they are installed.

The presence of heat pumps 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 increased, requiring a sufficient absorption capacity of the grid. It should be noted that the electrical configurations are more numerous and diversified than the natural gas configurations thanks to a larger panel of electrical units.

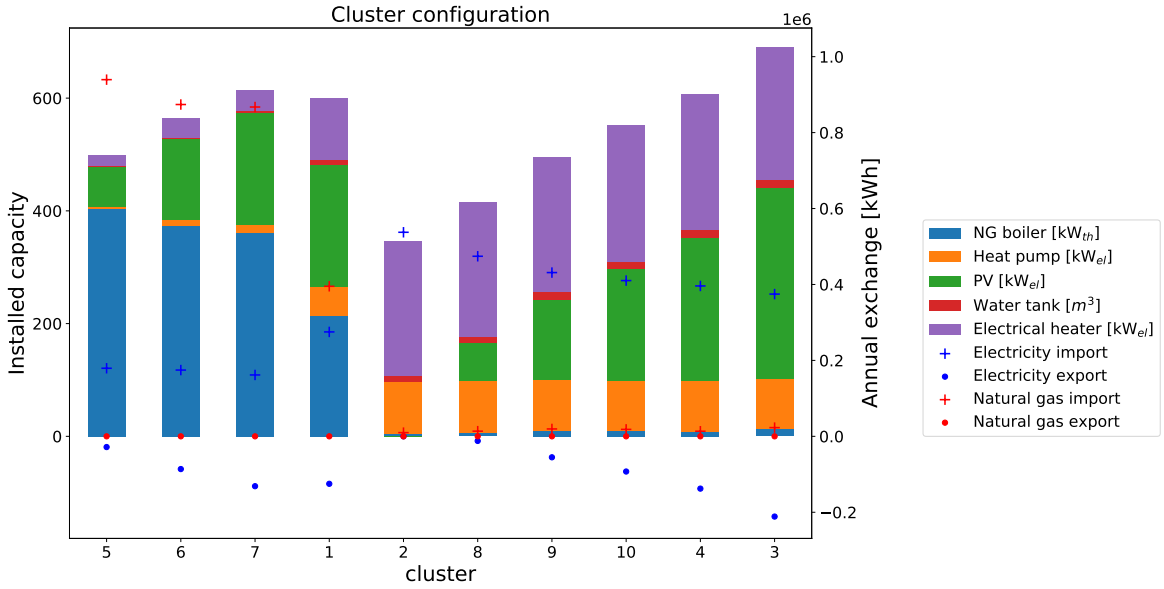

Figure 4.5: Distribution of installed units capacity and network exchange within the identified district configurations

4.3.3 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 component. Usually, only the first two components are used to project the data points, as it can be observed in Figure 4.6.

The first dimension explains 64.8% of the variance and the second 33.2%, thereby the plot allows to explain 98% of the data set variance. The indicators can be regrouped in three main groups which align with the principal components. This is the result of having 98% of the variance explained in only 2 dimensions. The first component can be interpreted as the 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 natural gas import, boiler installed capacity and GWP. Their correlation is natural as the boiler is fueled by natural gas, which has a high CO_2 emission factor. Oppositely, another group is inversely correlated to the first dimension. It contains the installed capacity of the heat pump and electrical heater, which are low emissions technologies. The electricity import is closely related to the group but with a negative correlation to the second dimension. This second dimension is highly correlated with installed PV panel capacity, electricity exports 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 excess into 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.

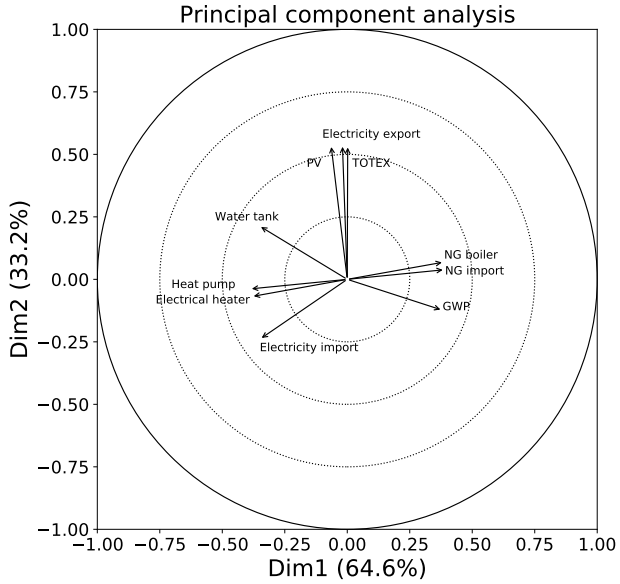

Figure 4.6: Variable correlation plot of the first two principal components. Indicators pointing in the same direction are correlated, those pointing in the opposite direction are inversely correlated, and those that are perpendicular are uncorrelated. The length of the arrow indicates the influence of the indicator.

The capacity of the water tank is inversely correlated with the GWP, highlighting the importance of thermal storage at the building scale to help mitigate climate change. The non-connection between the installed capacity of the heat pump and PV is surprising as the two technologies are complementary from each other to meet electric and heat demand. Similarly, one might have expected a negative correlation between GWP and TOTEX, as in most scenarios, low GWP leads to high TOTEX and vice versa. This is most likely due to optimizations having a low retail electricity price. The grid electricity consumption having a relatively low CO_2 emission factor; this reduces the positive correl-

ation between TOTEX and GWP. Finally, regarding the arrows' length their low variance among the indicators, only the TOTEX, PV and electrical export are slightly more influential.

The 2D distribution of the standardized data for every combination of the district indicator is available in the appendix, see Figure 6.5. It allows to validate the PCA and see in more detailed possible correlations.

4.4 Outlook

This following lines present some possible outlook on the applied GSA and clustering techniques.

First, a more extensive selection of input parameters for the Morris method could bring to the fore additional parameters. Additionally, the 95% confidence interval of the absolute mean of EE distribution is abnormally high compared to the mean value, see Table 6.4. Regarding the Sobol method, more accurate variation ranges could help reduce outliers as each configuration would have a real economical and physical meaning. The reduction of computational time of the model optimization would allow to increase the number of evaluations and therefore increase the number of data points for the clustering.

The application of K-means on the data set is not necessarily correct as the space is probably not convex. However, the utilization of the algorithm to obtain an evaluation of the required number of clusters is acceptable. The curse of dimensionality should be investigated when using a Euclidean distance with DBSCAN. Finally, the use of the peak and total exchange with the grid can be improved by introducing wavelet. This would allow to consider the evolution of the exchange over time in the clustering at minimum computational cost.

Chapter 5

Conclusion

In this research, a framework is proposed to identify typical configurations of a district energy system. It is part of a larger project which attend to define a national scale energy system based on typical district. The framework plays a key role in providing a panel of optimal configurations for each typical district. In the course of this research, two main steps are performed.

First, a global sensitivity analysis of the district energy system model is performed to efficiently sample the space of solutions. The Morris method is used to screen input parameters and the sampling scheme is acquired using Sobol method. The conclusion of the SA emphasized the importance of the retail tariff of energy carriers in such systems.

Finally, 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. They showed that 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.

Furthermore, the identified configurations can be separated into two types of system, 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 have a positive, respectively negative, correlation with the installed PV capacity.

Chapter 6

Appendix

6.1 Reduction of the computational time

The model used is a distributed energy model of a district scale name REHO for Renewable Energy Hub Optimizer. The application of a GSA to this model would require, following [45] recommendations, few hundreds to a thousand of optimizations per parameters. This would end up in an unreasonable computational time of several tens of days. Prior to this project the computation time of a 15 buildings district was of around 10 to 15 minutes per optimization after the following improvement the computational time has been reduced to 120 seconds per optimization. AMPL refers to the optimization language used in REHO and stands for *a mathematical programming language)*.

1. Improvement of the pools management (multiprocessing)
2. Selection of "important" results to extract from AMPL
3. Launch of the program from the command prompt
4. Regular reset of the AMPL license to avoid unjustified crashes (good practice every 50 to 100 optimizations)

6.2 District characteristic

	Multi family	Multi family	Single family	
Building category	existing	standard	existing	
Number of buildings	11	2	18	
Total net surface	9200	1100	5600	m^2
Total energy ref. area	11500	1400	7000	m^2
Total roof area	4200	560	4400	m^2
Annual electricity demand	37 ± 17	50 ± 21	60 ± 60	kWh/m^2_{net}
Annual hot water demand	25 ± 0	25 ± 0	19 ± 0	kWh/m^2_{net}
Annual internal gains	30 ± 2	32 ± 0	29 ± 2	kWh/m^2_{net}
Solar heat gains	22 ± 6	20 ± 3	31 ± 10	kWh/m^2_{era}
Heat transfer factor	1.74 ± 0.24	0.83 ± 0	1.84 ± 0.21	$W/m^2_{era}/K$
Heat capacity factor	118 ± 5	120 ± 0	120 ± 0	$W/m^2_{era}/K$

Table 6.1: Characteristics of the district building

6.3 Morris

Parameters	Default 50-200%	Increasing 100-200%	Decreasing 50-100%	Constant 85-115%
Cost_inv_1			x	
Cost_inv_2			x	
CO2Emi_unit1			x	
CO2Emi_unit2			x	
lifetime		x		
Units_Fmin				x
Units_Fmax				x
baremodule				x
buy_elec	x			
sell_elec				x
buy_NG	x			
sell_NG	x			
HouseRoofUse				x
BOI_efficiency_max				x
PV_eff		x		

Table 6.2: Parameters variation range of the Morris method

	NG_Boiler	HeatPumpAW	PV	Battery	WaterTankSH	WaterTankDHW	ElectricalHeater_SH	ElectricalHeater_DWH
Cost_inv_1	x	x	x	x	x	x	x	x
Cost_inv_2	x	x	x	x	x	x	x	x
CO2Emi_unit1	x						x	x
CO2Emi_unit2	x	x	x	x	x	x	x	x
lifetime	x	x	x	x	x	x	x	x
Units_Fmin	x	x			x			
Units_Fmax	x	x	x		x	x	x	x
baremodule	x	x	x	x	x	x	x	x

Table 6.3: Units modified parameters for the Morris method

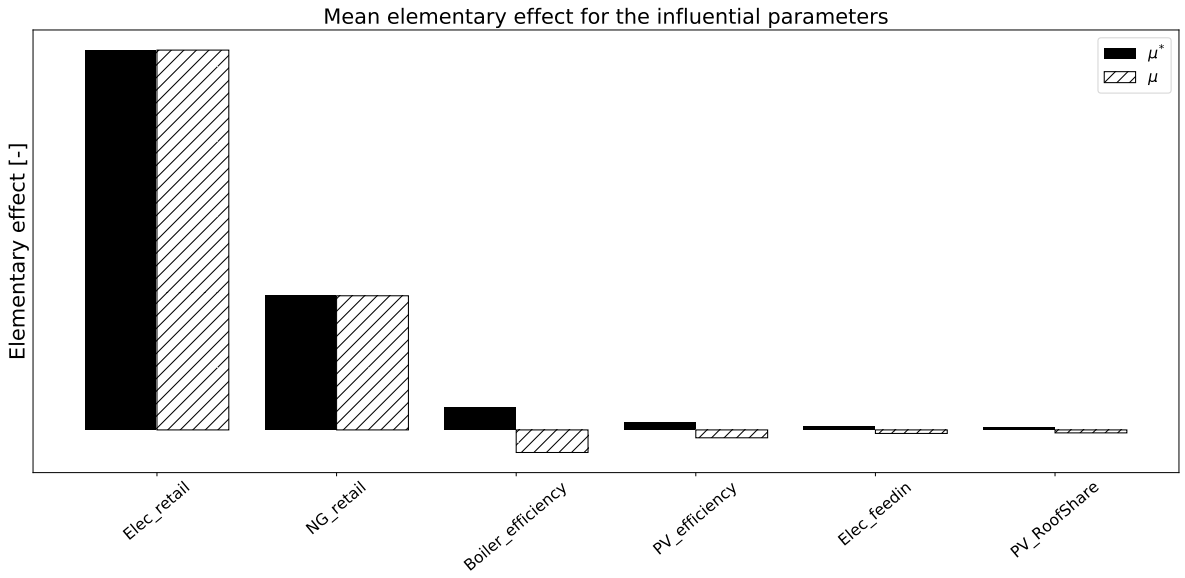

Figure 6.1: Mean and absolute mean of the distribution of the elementary effect of the influential input parameters

	Parameters	mu	mu_star	sigma	mu_star_conf
0	NG_Boiler___Units_Fmin	0.000000	0.000000	0.000000	0.000000
1	NG_Boiler___Units_Fmax	0.000000	0.000000	0.000000	0.000000
2	NG_Boiler___Cost_inv1	0.000000	0.000000	0.000000	0.000000
3	NG_Boiler___Cost_inv2	0.000000	0.000000	0.000000	0.000000
4	NG_Boiler___baremodule	0.000000	0.000000	0.000000	0.000000
5	NG_Boiler___lifetime	0.000000	0.000000	0.000000	0.000000
6	NG_Boiler___CO2Emi_unit1	0.000000	0.000000	0.000000	0.000000
7	NG_Boiler___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
8	HeatPumpAW___Units_Fmin	0.000000	0.000000	0.000000	0.000000
9	HeatPumpAW___Units_Fmax	0.000000	0.000000	0.000000	0.000000
10	HeatPumpAW___Cost_inv1	0.000000	0.000000	0.000000	0.000000
11	HeatPumpAW___Cost_inv2	0.000000	0.000000	0.000000	0.000000
12	HeatPumpAW___baremodule	0.000000	0.000000	0.000000	0.000000
13	HeatPumpAW___lifetime	0.000000	0.000000	0.000000	0.000000
14	HeatPumpAW___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
15	PV___Units_Fmax	0.000000	0.000000	0.000000	0.000000
16	PV___Cost_inv1	0.000000	0.000000	0.000000	0.000000
17	PV___Cost_inv2	0.000000	0.000000	0.000000	0.000000
18	PV___baremodule	0.000000	0.000000	0.000000	0.000000
19	PV___lifetime	0.000000	0.000000	0.000000	0.000000
20	PV___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
21	Battery___Cost_inv1	0.000000	0.000000	0.000000	0.000000
22	Battery___Cost_inv2	0.000000	0.000000	0.000000	0.000000
23	Battery___baremodule	0.000000	0.000000	0.000000	0.000000
24	Battery___lifetime	0.000000	0.000000	0.000000	0.000000
25	Battery___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
26	WaterTankSH___Units_Fmin	0.000000	0.000000	0.000000	0.000000
27	WaterTankSH___Units_Fmax	0.000000	0.000000	0.000000	0.000000
28	WaterTankSH___Cost_inv1	0.000000	0.000000	0.000000	0.000000
29	WaterTankSH___Cost_inv2	0.000000	0.000000	0.000000	0.000000
30	WaterTankSH___baremodule	0.000000	0.000000	0.000000	0.000000
31	WaterTankSH___lifetime	0.000000	0.000000	0.000000	0.000000
32	WaterTankSH___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
33	WaterTankDHW___Units_Fmax	0.000000	0.000000	0.000000	0.000000
34	WaterTankDHW___Cost_inv1	0.000000	0.000000	0.000000	0.000000

35	WaterTankDHW___Cost_inv2	0.000000	0.000000	0.000000	0.000000
36	WaterTankDHW___baremodule	0.000000	0.000000	0.000000	0.000000
37	WaterTankDHW___lifetime	0.000000	0.000000	0.000000	0.000000
38	WaterTankDHW___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
39	ElectricalHeater_SH___Units_Fmax	0.000000	0.000000	0.000000	0.000000
40	ElectricalHeater_SH___Cost_inv1	0.000000	0.000000	0.000000	0.000000
41	ElectricalHeater_SH___Cost_inv2	0.000000	0.000000	0.000000	0.000000
42	ElectricalHeater_SH___baremodule	0.000000	0.000000	0.000000	0.000000
43	ElectricalHeater_SH___lifetime	0.000000	0.000000	0.000000	0.000000
44	ElectricalHeater_SH___CO2Emi_unit1	0.000000	0.000000	0.000000	0.000000
45	ElectricalHeater_SH___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
46	ElectricalHeater_DWH___Units_Fmax	0.000000	0.000000	0.000000	0.000000
47	ElectricalHeater_DWH___Cost_inv1	0.000000	0.000000	0.000000	0.000000
48	ElectricalHeater_DWH___Cost_inv2	0.000000	0.000000	0.000000	0.000000
49	ElectricalHeater_DWH___baremodule	0.000000	0.000000	0.000000	0.000000
50	ElectricalHeater_DWH___lifetime	0.000000	0.000000	0.000000	0.000000
51	ElectricalHeater_DWH___CO2Emi_unit1	0.000000	0.000000	0.000000	0.000000
52	ElectricalHeater_DWH___CO2Emi_unit2	0.000000	0.000000	0.000000	0.000000
53	buy_elec	92793.081813	92793.081813	39500.428732	23738.209307
54	sell_elec	-861.115229	861.115229	775.525332	513.272340
55	buy_NG	32771.053204	32771.053204	47560.528471	28895.379315
56	sell_NG	0.000000	0.000000	0.000000	0.000000
57	PV_eff	-1926.373441	1926.373441	2916.932469	1939.368660
58	HouseRoofUse	-726.115292	726.115292	905.373103	524.583684
59	BOI_efficiency_max	-5516.327114	5516.327114	5978.467172	2980.485230

Table 6.4: Morris results of input parameters

6.4 Sobol

Definition	Parameter	Default value	Range of variation	Units
Retail tariff of electricity	Elec_retail	0.20	50-150%	CHF/kWh
Retail tariff of natural gas	NG_retail	0.10	50-150%	CHF/kWh
Feed-in tariff of electricity	Elec_feedin	0.08	85-115%	CHF/kWh

Table 6.5: Parameters used for Sobol sampling

Parameter	S	S_T
Elec_retail	0.859773	0.914416
Elec_feedin	0.000111	0.000602
NG_retail	0.090910	0.137944

Table 6.6: First-order and total effect of the chosen energy carrier prices

Cluster	TOTEX [CHF/yr]	OPEX [CHF/yr]	CAPEX [CHF/yr]	GWP [tCO_2/yr]	NG_Boiler [kW_{th}]	HeatPumpAW [kW_{el}]	PV [kW_{el}]	WaterTank [m^3]	ElectricalHeater [kWh/yr]	Elec_total_supply [kWh/yr]	NG_total_supply [kWh/yr]	Elec_total_demand [kWh/yr]
1	156705.44	105215.34	51490.10	116919.20	213.42	50.98	217.75	8.86	108.62	275018.03	395198.54	125036.97
2	85127.01	60408.36	24718.65	96615.80	4.89	92.49	0.00	10.74	238.85	537464.05	9766.88	0.00
3	187189.79	116851.70	70338.08	42100.29	13.91	88.96	338.29	13.72	235.01	374650.08	22969.58	211325.31
4	158631.76	99836.55	58795.22	54687.36	7.67	90.44	254.82	13.81	240.10	396097.10	13413.83	137707.89
5	116146.52	93968.92	22177.60	230348.75	403.42	3.20	70.80	2.60	19.12	179331.67	939460.26	28241.96
6	137916.70	104925.38	32991.32	207066.99	373.26	10.14	143.50	3.34	34.52	174545.69	873966.98	86208.26
7	153241.28	112201.85	41039.44	196838.80	360.94	14.80	198.07	4.21	36.93	161676.96	867373.05	131320.56
8	103056.43	69620.07	33436.35	85874.49	7.01	91.46	66.96	11.22	237.97	474272.31	13598.86	12071.60
9	121668.66	77899.48	43769.19	73924.92	9.94	89.74	143.46	12.79	238.74	431317.83	19466.30	55205.45
10	134286.85	83209.49	51077.36	64730.42	9.43	89.70	198.06	13.51	241.74	410212.06	18289.33	92691.17

Table 6.7: Average values of the identified configurations

6.5 Clustering

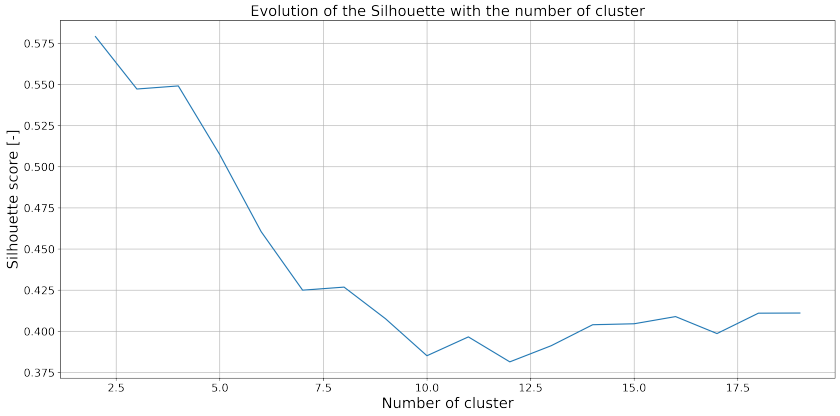

Figure 6.2: Evolution of the Silhouette score with the number of cluster k

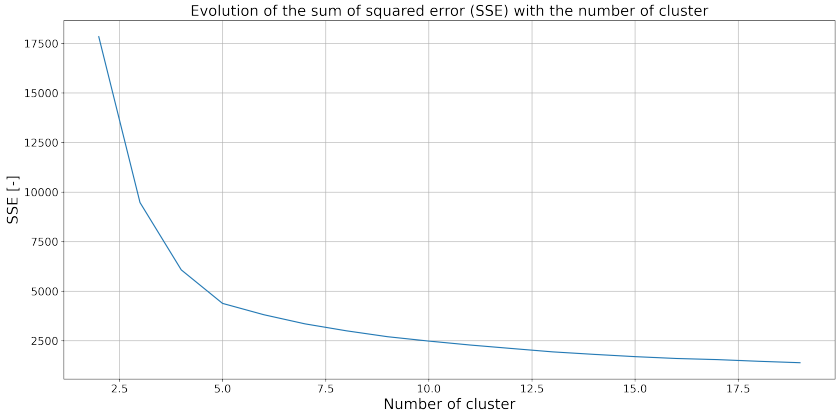

Figure 6.3: Evolution of the SSE score with the number of cluster k

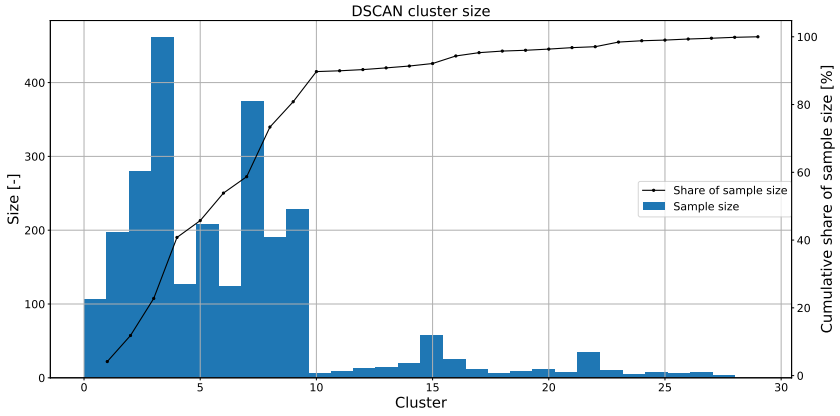

Figure 6.4: Distribution of the clusters size

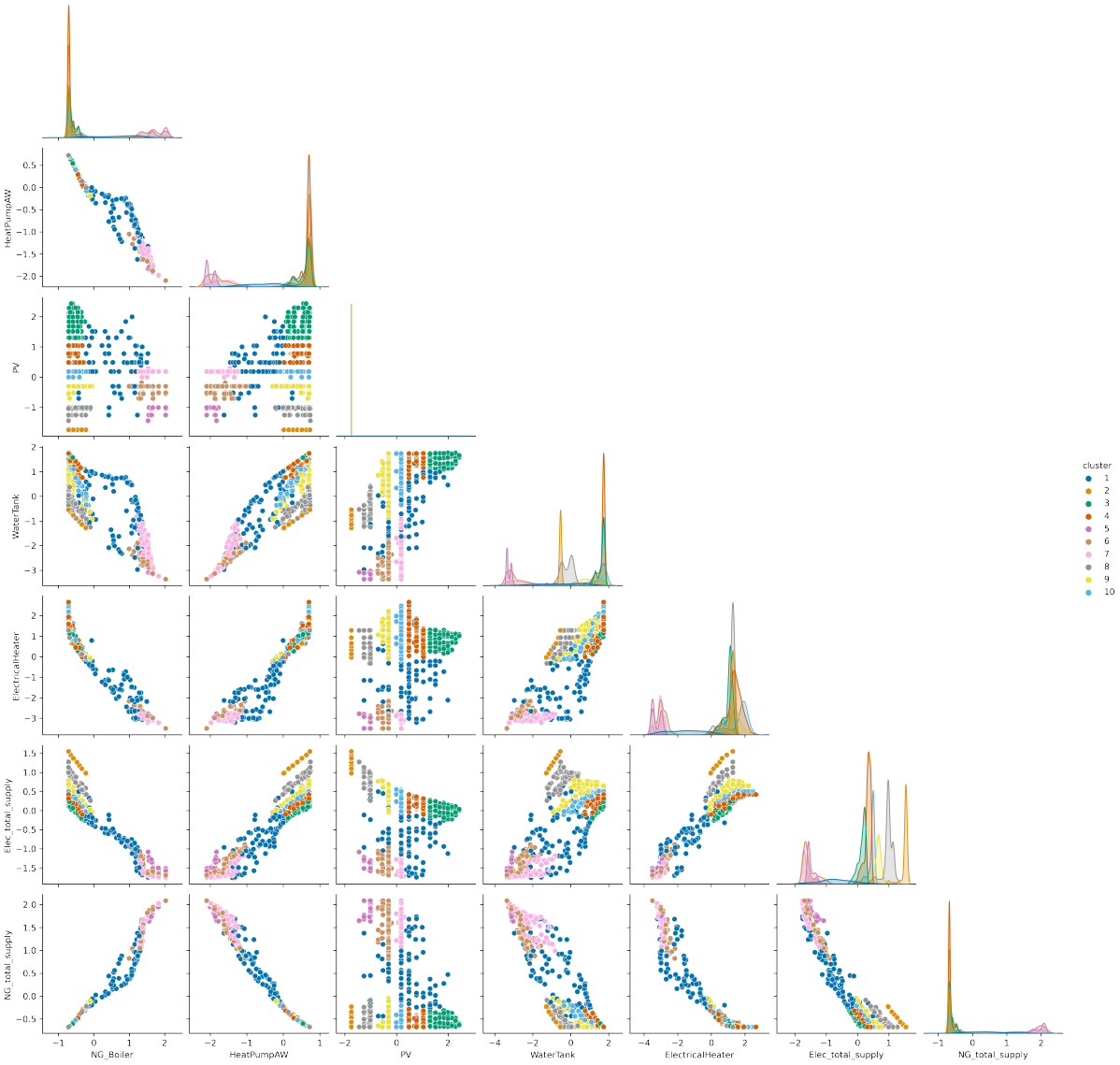

Figure 6.5: Pair plot of the standardized district main indicators. It provided a raw visualisation of the possible correlation between parameters. The color represents the different configuration.

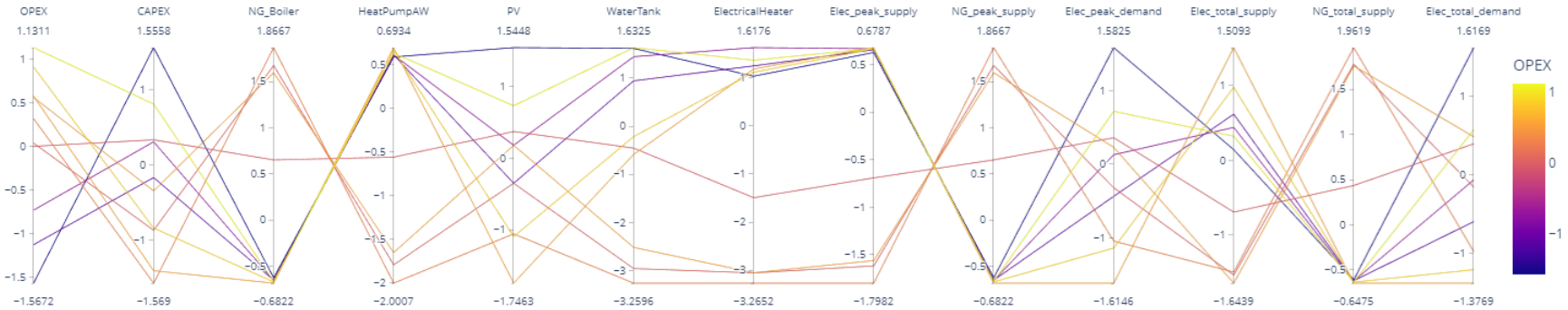

Figure 6.6: Parallel coordinates plot of the typical configurations for standardized features

Bibliography

[1] IPCC. Climate Change 2022: Mitigation of Climate Change.

[2] Hannah Ritchie, Max Roser, and Pablo Rosado. CO and Greenhouse Gas Emissions. *Our World in Data*, May 2020.

[3] IEA. Global Alliance for Building and Construction – Promoting energy efficiency.

[4] Kari Alanne and Arto Saari. Distributed energy generation and sustainable development. *Renewable and Sustainable Energy Reviews*, 10(6):539–558, December 2006.

[5] Daniel M. Kammen and Deborah A. Sunter. City-integrated renewable energy for urban sustainability. *Science (New York, N.Y.)*, 352(6288):922–928, May 2016.

[6] 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. *Renewable and Sustainable Energy Reviews*, 96:440–459, November 2018.

[7] Rahul Gupta, Fabrizio Sossan, and Mario Paolone. Countrywide PV hosting capacity and energy storage requirements for distribution networks: The case of Switzerland. *Applied Energy*, 281:116010, January 2021.

[8] Gauthier Limpens, Stefano Moret, Hervé Jeanmart, and Francois Maréchal. EnergyScope TD: A novel open-source model for regional energy systems. *Applied Energy*, 255:113729, December 2019.

[9] Miguel Chang, Jakob Zink Thellufsen, Behnam Zakeri, Bryn Pickering, Stefan Pfenninger, Henrik Lund, and Poul Alberg Østergaard. Trends in tools and approaches for modelling the energy transition. *Applied Energy*, 290:116731, May 2021.

[10] Francisca Jalil-Vega and Adam D. Hawkes. The effect of spatial resolution on outcomes from energy systems modelling of heat decarbonisation. *Energy*, 155:339–350, July 2018.

[11] CENTRE SCIENTIFIQUE ET TECHNIQUE DU BATIMENT. Geo-clustering to deploy the potential of Energy efficient Buildings across EU, 2013.

[12] Corentin Kuster, Jean-Laurent Hippolyte, Yacine Rezgui, and Monjur Mourshed. A simplified geo-cluster definition for energy system planning in Europe. *Energy Procedia*, 158:3222–3227, February 2019.

[13] Gaël Germano, editor. *Regionalization in energy system planning: Application in Gros-de-Vaud, canton de Vaud*. 2019.

[14] Amara Slaymaker, editor. *Demographic and Geographic Region Definition in Energy System Modelling. A case study of Canada's path to net zero greenhouse gas emissions by 2050 and the role of hydrogen*. 2021.

[15] 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. *Energy*, 74:877–885, September 2014.

[16] 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. *Energy and Buildings*, 209:109667, February 2020.

[17] P. Stadler, Luc Girardin, and F. Maréchal. The swiss potential of model predictive control for building energy systems. *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2017.

[18] 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. *Energy Conversion and Management*, 103:739–751, October 2015.

[19] 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. *Applied Energy*, 204:1299–1316, October 2017.

[20] Hongbo Ren, Weisheng Zhou, and Weijun Gao. Optimal option of distributed energy systems for building complexes in different climate zones in China. *Applied Energy*, 91(1):156–165, March 2012.

[21] 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. *Energy*, 239:122436, January 2022.

[22] 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. *Renewable and Sustainable Energy Reviews*, 52:1391–1404, December 2015.

[23] Boran Morvaj, Ralph Evins, and Jan Carmeliet. Optimising urban energy systems: Simultaneous system sizing, operation and district heating network layout. *Energy*, 116:619–636, December 2016.

[24] 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. *Energy*, 93:2546–2558, December 2015.

[25] Thomas Schütz, Xiaolin Hu, Marcus Fuchs, and Dirk Müller. Optimal design of decentralized energy conversion systems for smart microgrids using decomposition methods. *Energy*, 156:250–263, August 2018.

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

[27] Paul Michael Stadler. *Model-based sizing of building energy systems with renewable sources*. PhD thesis.

[28] Georgios Mavromatidis, Kristina Orehounig, and Jan Carmeliet. Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. *Applied Energy*, 214:219–238, March 2018.

[29] Tianjie Liu, Wenling Jiao, and Xinghao Tian. A framework for uncertainty and sensitivity analysis of district energy systems considering different parameter types. *Energy Reports*, 7:6908–6920, November 2021.

[30] Wei Tian. A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20:411–419, April 2013.

[31] Paul Westermann and Ralph Evins. Surrogate modelling for sustainable building design – A review. *Energy and Buildings*, 198:170–186, September 2019.

[32] Torben Østergård, Rasmus L. Jensen, and Steffen E. Maagaard. Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy and Buildings*, 142:8–22, May 2017.

[33] Hrishikesh D. Vinod. Integer Programming and the Theory of Grouping. *Journal of the American Statistical Association*, 64(326):506–519, June 1969. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1969.10500990.

[34] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. Conference Name: IEEE Transactions on Information Theory.

[35] Partitioning Around Medoids (Program PAM). In *Finding Groups in Data*, pages 68–125. John Wiley & Sons, Ltd, 1990. Section: 2 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316801.ch2.

[36] David Dohan, Stefani Karp, and Brian Matejek. K-median Algorithms: Theory in Practice. page 21.

[37] Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin-Teng Lin. A review of clustering techniques and developments. *Neurocomputing*, 267:664–681, December 2017.

[38] P. Cheeseman and J. Stutz. Bayesian Classification (AutoClass): Theory and Results. *undefined*, 1996.

[39] 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, Portland, Oregon, August 1996. AAAI Press.

[40] 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, Berlin, Heidelberg, 2013. Springer.

[41] Clemens Felsmann, Luise Mann, and Vera Boß. Identification of urban cellular structures for flexible heat and temperature distribution in district heating networks. *Energy Reports*, 7:9–17, October 2021.

[42] Thomas Schütz, Markus Hans Schraven, Marcus Fuchs, Peter Remmen, and Dirk Müller. Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. *Renewable Energy*, 129:570–582, December 2018.

[43] Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M. Pérez, and Iñigo Perona. An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1):243–256, January 2013.

[44] 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, April 2014.

[45] A. Saltelli, editor. *Global sensitivity analysis: the primer*. John Wiley, Chichester, England ; Hoboken, NJ, 2008. OCLC: ocn180852094.

[46] Max D. Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, May 1991. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1991.10484804.

[47] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, May 2002.

[48] J and Saltelli A Campolongo, F and Cariboni. Sensitivity analysis: the Morris method versus the variance based measures. *Technometrics*, 2003.

[49] Will Usher, Jon Herman, Calvin Whealton, David Hadka, Xantares, Fernando Rios, Bernardoct, Chris Mutel, and Joeri Van Engelen. Salib/Salib: Launch!, October 2016.

[50] 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. *Computer Physics Communications*, 181(2):259–270, February 2010.

[51] Leland McInnes and John Healy. Accelerated Hierarchical Density Based Clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 33–42, November 2017. ISSN: 2375-9259.

[52] Federal Statistical Office. Géodonnées Etat de Vaud, Registre cantonal des bâtiments (RCB), 2019.

[53] Federal Statistical Office. Federal Register of Buildings and Dwellings.

[54] OpenStreetMap contributors. Planet dump, 2017.

[55] J. Remund et al. METEONORM - Global Meteorological Database for Engineers, Planners and Education, 2003.

[56] SIA. vernehmlassung sia 380/1 heizwärmebedarf, 2016.

[57] Federal Office of Topography. swissBUILDINGS3D 2.0, 2019.

[58] Swiss Federal Office of Energy Geoinformation. Switzerland in 3D, 2018.

[59] Administration cantonale vaudois. Distribution networks database, 2019.

[60] Middelhauve, Luise. *On the role of districts as renewable energy hubs.* PhD thesis, EPFL, 2022.

[61] Luise Middelhauve, Francesco Baldi, Paul Stadler, and François Maréchal. Grid-Aware Layout of Photovoltaic Panels in Sustainable Building Energy Systems. *Frontiers in Energy Research*, 8:573290, February 2021.

[62] Luc Girardin. *A GIS-based Methodology for the Evaluation of Integrated Energy Systems in Urban Area.* PhD thesis, EPFL, 2012.