Jiří Jaromír Klemeš, Petar Sabev Varbanov and Peng Yen Liew (Editors)
Proceedings of the 24th European Symposium on Computer Aided Process Engineering – ESCAPE 24
June 15-18, 2014, Budapest, Hungary.

Clustering Urban Areas for Optimizing the Design and the Operation of District Energy Systems

Samira Fazlollahi[a], Luc Girardin[b], François Maréchal[b*]

[a]*Veolia Environnement Recherche et Innovation (VERI), 291 avenue Dreyfous Ducas, 78520 Limay, France*
[b]*Industrial Process and Energy Systems Engineering Laboratory, Ecole Polytechnique Fédérale de Lausanne,CH-1015 Lausanne, Switzerland*
Francois.marechal@epfl.ch

Abstract

Solving the MILP model for optimizing the design and operating strategy of district energy systems (DES) is a computationally demanding task due to the large number of subsystems (i.e. resources, conversion technologies, buildings and networks) and corresponding decision variables. In order to reduce the number of decision variables and therefore the computational load of the problem, this paper presents a systematic procedure to represent an urban area with a macroscopic view as a set of "integrated zones". The integrated zone is an area where consumers, resources and energy conversion technologies are integrated. This is obtained by developing aggregated district integration models based on GIS data and applying k-means clustering techniques. By using the proposed method, the regional DES is partitioned into limited number of integrated zones. The selected zones allow us to achieve accurate representation of the whole district while significantly reducing the number of decision variables for which more detailed optimization methods can be applied.

Keywords: District energy systems (DES), Geographical information systems (GIS), CO_2 mitigation, Mixed Integer Linear Programming (MILP), *k-means* clustering.

1. Introduction

Higher efficiency, CO_2 mitigation and renewable energy usage in the urban area are achieved by proper system integration. Optimization techniques using mixed integer linear programming (MILP) method have been developed (Fazlollahi and Maréchal, 2013a) to optimize the energy system integration in the urban systems. These methods are however limited by the size of the problem and the number of decision variables (i.e. number of conversion technologies, buildings and networks) that prevents to solve regional scale problems.

Several researches concerned with developing reduction techniques for large-scale optimisation models. For example Holló et al. (2009) reviewed the reduction techniques for the process network synthesis (PNS). Lam et al. (2011) proposed an analytical method for merging several zones to analyse the biomass supply networks. Ng et al. (2013) introduce a functional clustering approach for integrating the industrial facilities of material supply networks.

This paper presents a systematic procedure and an optimization model to represent an urban area with a set of optimal integrated zones. The integrated zone is an area where resources, energy conversion technologies, and energy requirements of consumers are

aggregated. The developed model allows to geo-localise the zones in which a distribution network has a good potential to be implemented, and for which the energy system design using multi-objective, multi-period optimizations approach can be applied. The method is demonstrated and discussed by its application to a test case.

2. Methodology

In order to represent the district area using a limited set of "integrated zones", a method based on GIS data and the k-means clustering technique is developed. The aim of the developed model is to optimize the number of integrated zones (N_k^*). k-means is a greedy optimization algorithm, which goal is to minimize the squared error over clusters (Eq.(1)), for a given value of N_k.

$$k-means:\ min\ [\sum_{k=1}^{N_k}\sum_{i=1}^{N_S}(\sum_{a=1}^{N_a}(\hat{\mu}_{k,a}^{v}-\hat{s}_{i,a})^2\times z_{i,k}^{v})],\ \ \forall v\ ,\forall N_k \quad (1)$$

$S=\{s_{i,a}\}$ is a set of N_S subsystems such as buildings to be grouped into a set of N_k integrated zone. The main attributes ($a\in\{1,\dots,N_a\}$) of subsystems ($s_{i,a}\ \forall i\in\{1,\dots,N_S\}$) are their locations with geographical coordinates, and the temperatures of requirement (i.e. heating and hot water demands). The normalized set of attributes ($\hat{s}_{i,a}$) is calculated by Eq.(2). $\mu_{k,a}^{v}$ refers to the center of each zone, $\hat{\mu}_{k,a}^{v}$ denotes the normalized center, v is an index for starting point, and $z_{i,k}^{v}$ is a binary variable equal to 1 if a subsystem s_i is assigned to the zone k.

$$\hat{s}_{i,a}=\frac{s_{i,a}}{max\ \{s_{i,a}\ \forall i\in\{1,\dots,N_S\}}\ \ \forall a,i \quad (2)$$

The k-means algorithm requires two user-specified parameters, firstly the initial partitioning or starting point (v) and secondly the number of clusters (N_k). The result of the k-means greedy optimization algorithm depends on the starting. In order to overcome this issue, the k-means algorithm is applied with several random starting points ($\forall v\in\{1,\dots,v_{max}\}$, i.e. $v_{max}=1{,}000$). For selecting the best initial partitioning option ($v_{N_k}^*$) of N_k zones, the total costs of distribution networks ($\sum_{k=1}^{N_k}TAC_{\mathrm{v,k}}$ in Eq.(3)) is proposed as an additional indicator. $TAC_{\mathrm{v,k}}$ \$/MWh/y (Eq.(3)) is the specific cost of the heating and the gas distribution networks in zone k of evaluation v. It is estimated based on the following two scenarios.

$$TAC_{v,k}=\ min\{TAC1_{v,k}\ ,TAC2_{v,k}^{d}\ \forall d\},\qquad \forall v\in\{1,\dots,V_{max}\},\forall k\in\{1,\dots,N_k\} \quad (3)$$

$TAC1_{v,k}$: All buildings inside the integrated zone k are connected to the heating distribution network (centralized option). The specific distribution cost ($TAC1_{v,k}$ \$/MWh/y) is equal to the annual investment cost of the heating pipelines (Girardin 2012) divided by the total heat and hot water energy consumptions in zone k.

$TAC2_{v,k}^{d}$: Buildings which are located inside *d % (0 < d < 90)* of distance from the center of the integrated zone k are connected to the heat distribution network as a centralized option, while the rest (last *(100-d)%*) are connected to the gas distribution network for supplying gas to decentralized boilers (combination of decentralized and centralized solutions). The specific distribution cost ($TAC2_{v,k}^{d}$ \$/MWh/y) is equal to the total annual investment cost of the heating pipelines and the gas pipelines divided by the total heat and hot water energy consumptions in zone k. As an assumption $TAC2_{v,k}^{d}$ is estimated for $\forall d$ ($d\in\{0{,}10\ \%,\dots,90\ \%\}$) with step of 10 %.

Three statistical indicators are defined to optimize the number of "integrated zones" (N_k^*) for selected starting point ($v_{N_k}^*$); the average intra-clusters distance, which evaluates the compact character of the clusters ($C(v_{N_k}^*) = \frac{1}{N_k}\sum_{k=1}^{N_k}\sum_{i=1}^{N_S}(z_{i,k}^{v^*} \times \sum_{a=1}^{N_a}(\hat{\mu}_{k,a}^{v^*} - \hat{s}_{i,a})^2), \forall N_k$), the average inter-clusters distance, which evaluates the separation between the clusters ($D(v_{N_k}^*) = \frac{1}{N_k^2}\sum_{k=1}^{N_k}\sum_{j=1}^{N_k}(\sum_{a=1}^{N_a}(\hat{\mu}_{k,a}^{v^*} - \hat{\mu}_{j,a}^{v^*})^2), \forall N_k$), and the statistical measure $ESE(v_{N_k}^*)$ (Pham et al. 2004), which evaluates the ratio of observed to expected squared errors for N_k clusters (Eq.(4) and Eq.(5))

$$ESE(v_{N_k}^*) = \begin{cases} 1 & if\ N_k = 1, \forall v \\ \frac{N_k \times C(v_{N_k}^*)}{\alpha_{N_k} \times (N_k - 1) \times C(v_{N_k-1}^*)} & if\ C(v_{N_k-1}^*) \neq 0, \forall N_k > 1, \forall v \\ 1 & if\ C(v_{N_k-1}^*) = 0, \forall N_k > 1, \forall v \end{cases} \quad (4)$$

$$\alpha_{N_k} = \begin{cases} 1 - \frac{3}{4 \times N_a} & if\ N_k = 2,\ N_a > 1 \\ \alpha_{N_k-1} + \frac{1 - \alpha_{N_k-1}}{6} & if\ N_k > 2,\ N_a > 1 \end{cases} \quad (5)$$

N_a is the number of data set attributes and α_{N_k} is the weight factor.

An optimal value for N_k^* should yield; a low value for the average intra-clusters distance ($C(v_{N_k}^*)$), a high value for the average inter-clusters distance ($D(v_{N_k}^*)$), and a low value for the $ESE(v_{N_k}^*)$ measure. It can be expressed as Eq.(9).

To sum up, Eq.(6) can express the developed method.

$$N_k^*: \quad [min\{C(v_{N_k}^*)\ \forall N_k\},\ max\{D(v_{N_k}^*)\ \forall N_k\},\ min\{ESE(v_{N_k}^*)\ \forall N_k\}] \quad (6)$$

Subject to:

$$v_{N_k}^*: \quad min\{(\sum_{k=1}^{N_k} TAC_{v,k})\},\ \forall v \in \{1, \dots, V_{max}\}, \forall N_k \in \{1, \dots, N_{max}\} \quad (7)$$

$$k - means:\ min\ [\sum_{k=1}^{N_k}\sum_{i=1}^{N_s}(\sum_{a=1}^{N_a}(\hat{\mu}_{k,a}^{v} - \hat{s}_{i,a})^2 \times z_{i,k}^{v})],\ \forall v, \forall N_k \quad (8)$$

The solving strategy of the proposed model (Eq.(6)) proceeds as following:

Step 1: Run *k-means* for values of $N_k \in \{1, \dots, N_{max}\}$, with $\forall v \in \{1, \dots, V_{\max}\}$ random starting points (e.g V_{max}=1,000, N_{max}=25).

Step 2: Estimate the total costs of distribution networks ($\sum_{k=1}^{N_k} TAC_{v,k}$) for $N_k \in \{1, \dots, N_{max}\}$, and $\forall v \in \{1, \dots, V_{\max}\}$ (Eq.(3)).

Step 3: Select the best starting point ($v_{N_k}^*$) for $\forall N_k$ (Eq.(7)).

Step 4: Once $v_{N_k}^*$ have been selected, calculate values of average intra-clusters distance ($C(v_{N_k}^*)$), average inter-clusters distance ($D(v_{N_k}^*)$) and $ESE(v_{N_k}^*)$ for $\forall N_k$.

Step 5: Define the ascending order set of $C(v_{N_k}^*)\ \forall N_k$, the descending order set of $D(v_{N_k}^*)\ \forall N_k$, and the ascending order set of $ESE(v_{N_k}^*)\ \forall N_k$.

Step 6: Chose the best integrated zones (N_k^* Eq.(6)) in which;

$$N_k^*: R(N_k^*) = min\ [max\{R_C(v_{N_k}^*), R_D(v_{N_k}^*), R_{ESE}(v_{N_k}^*)\}\ \forall N_k \in \{1, \dots, N_{max}\}] \quad (9)$$

Where $R_C(v^*_{N_k})$ refers to the rank of N_k clusters (zones) in ascending order set of $C(v^*_{N_k})$, $R_D(v^*_{N_k})$ is the rank of N_k clusters in descending order set of $D(v^*_{N_k})$ and $R_{ESE}(v^*_{N_k})$ denotes the rank of N_k clusters in ascending order set of $ESE(v^*_{N_k})$.

3. Illustrative example

The case study presented by Fazlollahi et al. (2013b) is used to illustrate the advantage of the proposed method. The aim is to supply the heating requirements of a city with 475 small zones (Figure 1) with a central plant via heating distribution networks or individually with decentralized equipment. The k-means clustering is applied to split up the area into limited number of integrated zones. For $N_k \in \{1,\ldots,25\}$, and V_{max}=1,000 random starting points, $N_k^* = 13$ has the lowest value for the average intra-clusters distance, the highest value for the average inter-clusters distance and the lowest value for *ESE* measure (Figure 3, Eq.(6)). In addition, the total annual costs of distribution networks for $\forall N_k$, are presented by Figure 3, where the optimal costs observed by $N_k^* = 13$. Therefore, the city can be split up into 13 integrated zones (Figure 2). There are four candidate locations (S1, S2, S3 and S4 in Figure 5) for placing new central plants in the urban area. The design and operation optimizations of the district system are performed with respect to three objectives (Fazlollahi et al. 2013b); maximizing the system efficiency (***EFF***), minimizing the total investment and operating costs (***TAC***), and minimizing the environmental impacts ($\boldsymbol{M_{CO2}}$). Alternative conversion technologies for supplying power and heat services are; solar thermal, natural gas and biomass boilers, natural gas and biomass engines and turbines. These technologies can be placed in locations S1 to S4. There is also a possibility of investing on heat pumps in locations S1 and S2 to recover the waste heat from the wastewater plant. The hourly heating demands, the solar irradiation and electricity price are given by Fazlollahi et al. (2013b).

Among all multi objective optimization results inform of the first Pareto frontier (Figure 4), configuration "A" is selected for more details evaluation. In this solution two central plants, in locations S2 and S4, are chosen (Figure 5). Centralized plant S2 supplies heat via DHN to locations C3, C5, C6 and C8. This center features; 6 MW gas engine, 4 MW gas turbine, 29.5 MW gas boiler and 28 MW backup boiler. Centralized plant S4 features 28 MW biomass boiler and 10 MW backup natural gas boiler for supplying heat via DHN to locations C7, C9 and C10. Due to the transportation cost of biomass fuel the biomass boiler is only placed in S4. It is economically viable that individual gas boilers supply heat directly to locations C1 (with total capacity of 1.1 MW), C2 (2.4 MW), C4 (1 MW), C11 (4 MW), C12 (3.8 MW) and C13 (0.2 MW) as decentralized solutions and without local networks. The extension of pipelines between locations in solution "A" is illustrated in Figure 5.

4. Conclusions

In order to reduce the size of the district energy system design optimization model, a systematic procedure is proposed. The goal is to aggregate the urban area into a limited number of integrated zones for which the distribution cost and the aggregated energy demand can be calculated. The integrated zones are obtained using GIS data and applying k-means clustering techniques. The selected zones allow us to achieve accurate representations of the whole district while significantly reducing the number of decision variables. Table 1 presents the size of the optimization model of illustrative example for different number of integrated zones. It demonstrates the optimization size is decreased significantly by applying the proposed clustering model.

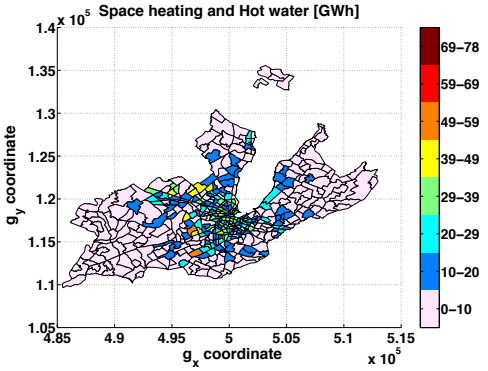

Figure 1. The energy demand [MWh] of a city with 100,000 populations

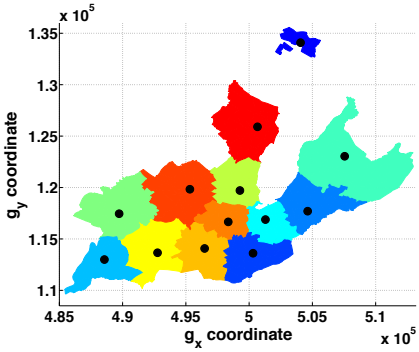

Figure 2. The city with 13 representative "Integrated zones"

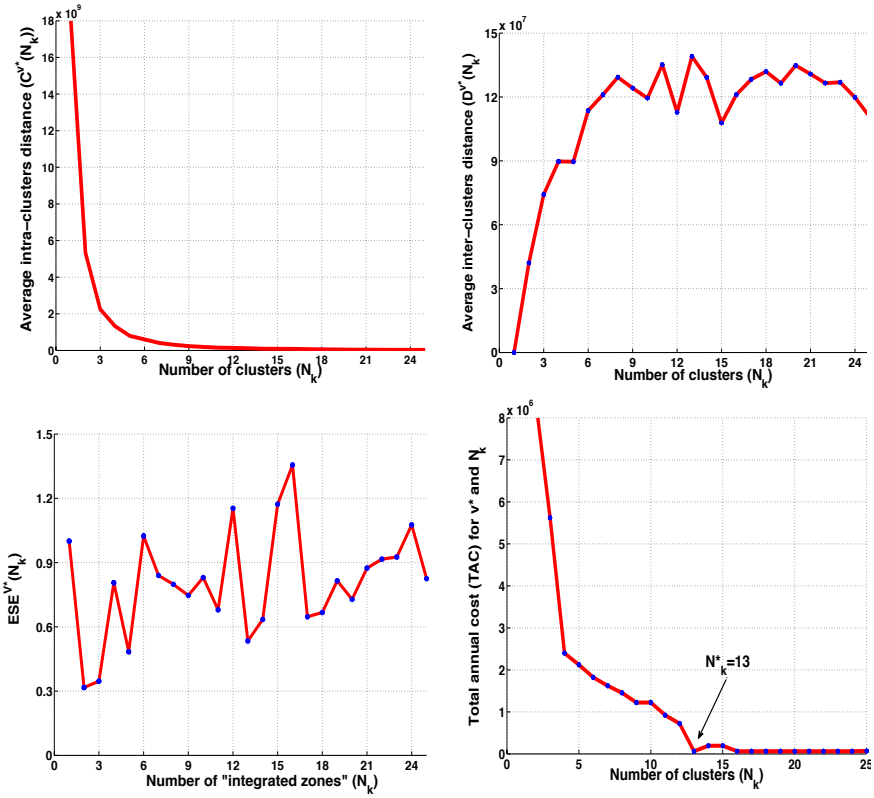

Figure 3. Intra and inter-clusters distances, $ESE(v^*_{N_k})$, and the total annual costs of DHN

Table 1. Comparison between the sizes of the optimization for various number of integrated zones

Number of "integrated zones"	5	7	13	475
Constraints	63,225	88,225	209,499	$\approx 9*10^6$
Variables	191,029	226,500	576,337	$\approx 27*10^6$

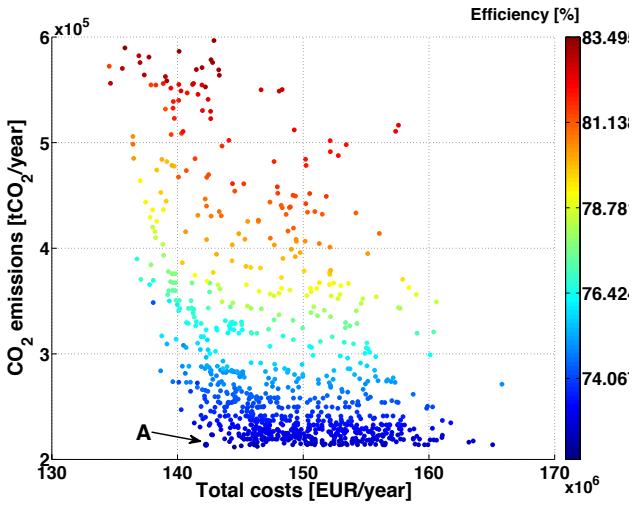

Figure 4. Multi-objective optimization results- first Pareto frontier

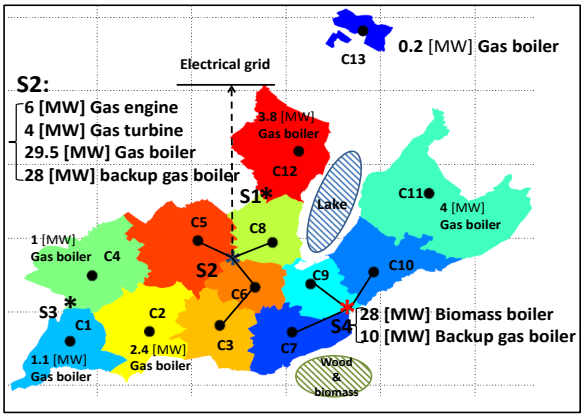

Figure 5. Illustrative example: Solution "A"

References

C. Holló, B.Imreh, C.Imreh, 2009, Reduction techniques for the PNS problems: a novel technique and a review, Optimisation Engineering, 10, 351–361.

D.T.Pham, S.S.Dimov, C.D.Nguyen, 2004, Selection of k in k-means clustering, Mechanical Engineering Science, 219,103–119.

H.L.Lam, J.J.Klemeš, Z.Kravanja, 2011, Model-size reduction techniques for large-scale biomass production and supply networks, Energy, 36, 8, 4599-4608.

L.Girardin, 2012, A GIS-based Methodology for the Evaluation of Integrated Energy Systems in Urban Area, PhD thesis, Ecole Polytechnique Federale de Lausanne, Switzerland.

S.Fazlollahi, F.Maréchal, 2013a, Multi-objective, multi-period optimization of biomass conversion technologies using evolutionary algorithms and mixed integer linear programming (MILP), Applied Thermal Engineering, 50, 2, 1504–1513.

S.Fazlollahi, G.Becker, F.Maréchal, 2013b, Multi-objectives, multi-period optimization of district energy systems: II-Daily thermal storage, Computers & Chemical Engineering, DOI: 10.1016/j.compchemeng.2013.10.016.

W.P.Q.Ng, H.L.Lam, 2013, A supply network optimisation with functional clustering of industrial resources, Journal of Cleaner Production, DOI: 10.1016/j.jclepro.2013.11.052.