



**HAL**  
open science

## District heating networks: enhancement of the efficiency

Ugis Sarma, Girts Karnitis, Janis Zuters, Edvins Karnitis

► **To cite this version:**

Ugis Sarma, Girts Karnitis, Janis Zuters, Edvins Karnitis. District heating networks: enhancement of the efficiency. *Insights into Regional Development*, 2019, 1 (3), pp.200-213. 10.9770/ird.2019.1.3(2) . hal-02188841

**HAL Id: hal-02188841**

**<https://hal.science/hal-02188841>**

Submitted on 18 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Publisher**

<http://jssidoi.org/esc/home>



## DISTRICT HEATING NETWORKS: ENHANCEMENT OF THE EFFICIENCY\*

Ugis Sarma<sup>1</sup>, Girts Karnitis<sup>2</sup>, Janis Zuters<sup>3</sup>, Edvins Karnitis<sup>4</sup>

<sup>1</sup>*Institute of Energy Systems and Environment, Riga Technical University, Azenes street 12/1, LV1048 Riga, Latvia*

<sup>2,3,4</sup>*Faculty of Computing, University of Latvia, Raina Blvd. 19, LV-1050, Riga, Latvia*

<sup>1</sup>*Latvenergo AS, Pulkveza Brieva Street 12, LV-1230, Riga, Latvia*

E-mails: <sup>2</sup>[girts.karnitis@lu.lv](mailto:girts.karnitis@lu.lv); [janis.zuters@lu.lv](mailto:janis.zuters@lu.lv); [edvins.karnitis@lu.lv](mailto:edvins.karnitis@lu.lv); <sup>3</sup>[ugis.sarma@latvenergo.lv](mailto:ugis.sarma@latvenergo.lv)

Received 10 May 2019; accepted 24 June 2019; published 30 September 2019

**Abstract.** During the decades the district heating's (DH) advantages (more cost-efficient heat generation and reduced air pollution) overcompensated the additional costs of transmission and distribution of the centrally produced thermal energy to consumers. Rapid increase in the efficiency of low-power heaters, development of separated low heat density areas in cities reduce the competitiveness of the large centralized DH systems in comparison with the distributed cluster-size networks and even local heating. Reduction of transmission costs, enhancement of the network efficiency by optimization of the design of the DH networks become a critical issue. The methodology for determination of the key drivers of the cost-efficiency of the DH networks to implement the most efficient (cost-minimal) thermal energy transmission was developed in this study. An inductive benchmarking modelling was applied; the general causal regularity is based on the observations of specific cases, thus determining the relationships between the network's design and thermal indicators as predictors and transmission costs as the target variable. The key drivers of the network efficiency were disclosed – the network length and the largest inner diameter of the pipes. The methodology is applicable for use by municipalities and heat providers for the heating planning of the new housing developments as well as renovation and/or expansion of the existing DH networks.

**Keywords:** district heating, network design, data mining, benchmarking methodologies.

**Reference** to this paper should be made as follows: Sarma, U.; Karnitis, G.; Zuters, J.; Karnitis, E. 2019. District heating networks: enhancement of the efficiency, *Insights into Regional Development* 1(3): 200-213. [https://doi.org/10.9770/ird.2019.1.3\(2\)](https://doi.org/10.9770/ird.2019.1.3(2))

\* The study has been supported by the State Research Program (VPP) project "Development of heat supply and cooling systems in Latvia" (project No. VPP-EM-EE-2018/1-0002), and by the University of Latvia project AAP2016/B032 "Innovative information technologies". The authors would like to thank members of the Latvian Association of Heat Companies for sharing actual data.



**European Research Council**

Established by the European Commission

**JEL Classifications:** C24, C45, O13.

**Additional disciplines:** ecology and environment; energetics and thermoenergetics; informatics.

## 1. Introduction

During the decades, the advantages of the large district heating (DH) systems were accepted almost as an axiom for the densely populated (urban) areas. Even starting the new millennium the recognized development trend was replacement of isolated small DH systems with an adequate heat sources by a powerful central heat station and main transition line that will carry large volumes of heat over long distances (e.g., King, & Shaw, 2010). The advantages were based on the cost-efficient heat generation (due to higher efficiency factor and cheaper fuel as a result of the effect of scale) as well as on reduced air pollution and carbon emissions (due to a small number of controlled sources and extensive use of the renewable energy). These benefits overcompensated the additional costs of transmission and distribution (further transmission) of the centrally produced thermal energy to consumers.

Recent studies (e.g., Delangle, Lambert, Shah, Acha, & Markides, 2017; Song, Li, & Wallin, 2017) and analyses (e.g., DEA, 2016a; DEA, 2016b), however, show the decreasing competitiveness of the large DH systems in the light of the development of new technologies. There has been a rapid increase in the efficiency of low-power heaters in the last decade. The DH systems using natural gas are losing their advantages of the effect of scale; the effect is only partly remained if biomass is used in both central and local plants. Also the modelling results (Fakhri, Ahlgren, & Ekvall, 2016) show that even with an incredibly large difference in heat costs, already at a very small network length (around 3 km), the urban-size centralized network becomes economically disadvantageous in comparison with the distributed cluster-size local networks.

In addition the modern cluster-based urban planning principles (agglomeration, metropolitan area, peri-urbanisation, etc.) and the heat efficient buildings create the low energy density areas that significantly increase specific transmission costs. But advanced green heat generation technologies (e.g., solar thermal, ground-source heat pumps, geothermal, waste) are well suited for the lower thermal power applications; they become acceptable for use even in the density populated areas. All this reduce the advantages of the large centralized DH systems; actually, only the difference in fuel prices is noteworthy due to the large-scale purchases (which can often be implemented in a distributed system too).

Rapid development of the heat generation technologies directly increases the importance of the design of the DH networks. The network-related costs are not an insignificant component of total heat supply costs, they accounts for up to 40% of the total heating tariff, including distribution losses (10.4% of produced heat in EU28, 12.4% in Latvia in 2016; (Eurostat, 2018)). The reduction of the transmission costs, enhancement of the cost-efficiency of the network become the decisive factor for the DH's competitiveness in the heating sector.

At the same time, the DH currently is well-developed in the Nordic and Eastern European countries as well as in the largest densely populated cities in the EU's countries with mild climate (e.g., ETSAP, 2013; Werner, 2017). Planning and design of the new and renovated networks are long-term projects (for 30-40 years). Evolving over decades, the DH systems have accumulated large amounts of investments; therefore, it is not possible to imagine a rapid dismantling of the DH sector, which would mean colossal sunken investments.

Like in other network industries, heating network operators are the natural monopolists in specific areas. So there is neither direct competition, nor commercially open market; operators have a little motivation to enhance efficiency of the network and business. Therefore the DH transmission segment and its pricing structure are fully regulated. In practice the consumers' choice is limited by administrative and/or fiscal instruments; it is not the

cost-based approach, of course. Nevertheless, consumers would opt for a local heat source, especially in the case of a new building connection. Their benefit (value-in-behaviour) will be based on the rationality and economic self-interest (Butler, Gordon, Roggeveen, Waitt, & Cooper, 2016). Consequently, economic values are becoming the core message in DH marketing (Smaliukiene, 2019); the decreased environmental effect would be an inadequate motivator for choice of the DH.

The European Commission has established a single, detailed and strict legal framework for the electricity and gas sectors, which has been developed for several decades. The DH is very fragmented even within one country, it is a distinct sub-regional issue; there is lack of substantial EU-level regulatory framework devoted to it. The development and functioning of the DH has traditionally been under the responsibility and control of municipalities. Exactly the local government (mostly city municipality) is a planner, normative and financial facilitator and supervisor of the modern heating networks (Saeima, 1994; UNEP, 2015); it is a supplier and large consumer of the thermal energy at the same time. In Latvia as in Estonia and Lithuania, the DH sector is dominated by companies owned by local governments, although there are also some private equity companies. A similar situation can also be seen in the Nordic countries, where municipal enterprises continue to be relatively strong (e.g., Aberg, Falting, & Forssell, 2016).

There is also a government support for further deployment of the heat networks on the national scale and political willingness on European and global scale (e.g., UNEP, 2015; UN, 2017; Euroheat & Power, 2019), taking care on national interests (e.g., energy policy, social policy, climate targets), and recognizing national benefits (modern district energy systems as a key solution to integrate energy efficiency and renewables in cities).

The various factors of increasing the efficiency of the DH systems have been the drivers of further strengthening the competitiveness of the systems over many years. The rapidly growing amount of academic research (e.g., on the search term “district heating” database *ScienceDirect* provided 13 times more issues published in 2017 in comparison with the number of publications in 2008, while the total number of annual records during this period has increased by only 46%) also shows the topicality of the DH related issues. The research has been focused mostly on various aspects of the heat generation (CHP, renewables, thermal storage, heat pumps, etc.) as well as the regulatory issues (e.g., Sarma, & Bazbauers, 2017); the heat transmission temperature is the single practical network issue as well as the hydrodynamic issues as more theoretical topic (see, e.g., Sayegh, Danielewicz, Nannou, Miniewicz, Jadviszczak, Piekarska, & Jouhara, 2017). Currently the studies are linked to development of the fourth generation DH (e.g., Lund, Werner, Wiltshire, Svendsen, Thorsen, Hvelplund, & Mathiesen, 2014).

Each case and each DH network is unique in terms of differences in the amount of the heat transported, operating area, consumers and other factors. The aim of current research is development of the methodology for determination of the key drivers of the cost-efficiency of DH networks to implement the most efficient (cost-minimal) thermal energy transmission. The methodology will be applicable for the heating planning of the new housing developments as well as renovation and/or expansion of the existing DH networks, increasing the degree of heating centralisation by merging smaller and remote DH systems into a larger one, or, on the contrary, routinely fragmenting or even dismantling existing DH systems. The methodology was developed by analysis of various performance aspects and creation of the general causative regularity using the advanced benchmarking modelling procedures on the basis of practical cases (Latvian DH utilities).

## **2. Basic principles of the methodology**

A direct theoretical calculation of transmission costs is an extremely difficult and even practically unachievable task. There are several multivariate analytic methods that are appropriate in principle for the study of cause and effect relationships between the input (network parameters) and output (costs) variables without investigation of

internal aspects of the network (a *black box* principle). The intelligent data mining methods are among them, they are well suited to discover existing regularities in the network’s data set; particularly it relates to the benchmarking data mining algorithms.

On the other hand, the mathematical modelling procedures are generally accepted instruments for the prediction of the scenarios in the wide variety of fields. To achieve the set goal we applied an inductive benchmarking modelling, using the declared design, thermal and financial indicators of networks, which are functioning in homogenous conditions (in particular geographical area with the similar environment, climate and business factors). The benchmarking modelling means search of the mathematical model that is based on observations of specific cases (e.g., Moriarty, 2011), thus determining the relationships between the network’s design and thermal indicators as independent variables (predictors) and some cost indicator as the dependent (target) variable; its reduction should reflect the efficiency progress in the corresponding network. Due to the analogous impact of external factors, the benchmarking provides comparative assessment of transmission costs; this way we will disclose the impact of network indicators on costs and extract the most significant ones (key drivers), reducing a large number of predictors.

The authors already have successfully used several benchmarking modelling tools for the cost assessment in a related industry – water supply (Zuters, Valainis, Karnitis, G., & Karnitis, E., 2016; Karnitis E., Karnitis G., Zuters, & Bobinaite, 2017). The results obtained have shown the suitability of the method to solve the similar tasks.

The necessary performance of the network  $k$  is determined (fig. 1a) by the requested heat capacity  $Ph(k)$ , and the thermal energy  $Qcon(k)$  that really is supplied through  $N(k)$  connection points to consumers. Due to the losses, a higher amount of heat  $Qpr(k)$  should be produced and delivered into the network; the price of the  $Qpr(k)$  is determined by the regulated heat production tariff  $Tpr(k)$ . The network consists of different size pipes (inner diameter changes from  $Dmin(k)$  to  $Dmax(k)$ ;  $Dmed(k)$  is the most frequently used size), their total length is  $L(k)$ . The total costs  $Ctr(k)$  for heat transmission includes all capital expenditures and operational expenses for the network development, maintenance and functioning; the reduction of costs directly characterizes enhancement of network’s efficiency. The regulated transmission tariff  $Ttr(k)$  could be determined as  $Ttr(k) = Ctr(k) / Qcon(k)$ ; actually it don’t provide a new information in our case. All these indicators are clearly and unambiguously defined, they are quantitatively measurable and controllable that is important for modelling quality.

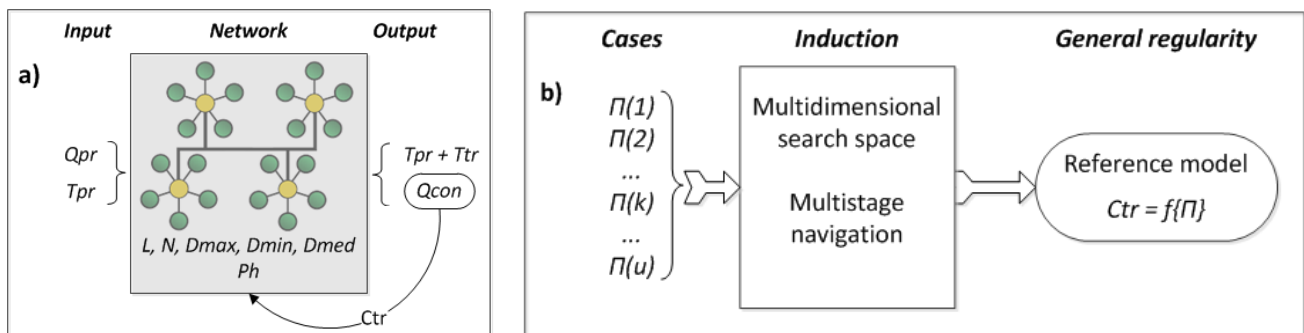


Fig. 1. Indicators of the network (a), and functionality of the benchmark modelling (b)

Developed by authors

For the development of the benchmark model, the DH network  $k$  is formalized as a multiple-input single-output determined search space (fig. 1b); its input data set of predictors  $\Pi(k)$  defines some output (dependent) cost

variable  $C(k)$ . Then the creation of the model means search of the multi-functional regularity (1); the internal structure and operation of the network are irrelevant for performing this task.

$$C(k) = f_k(\Pi(k)) \quad (1)$$

There are  $u$  different functions  $f_1, f_2, \dots, f_k, \dots, f_u$  for  $u$  networks; they form a factual basis for the modelling – set of the specific cases that can be used for search of the general regularity. Then the benchmark modelling means an inductive process – search of the general modelled regularity  $C(m) = f(\Pi)$  on the basis of  $u$  specific cases  $C(k) = f_k(\Pi(k))$  by navigation in the multidimensional search space.

The total transmission costs  $Ctr$  consist of number of components – heat loss costs, labour costs, depreciation costs, payments for electricity and chemicals, etc. An analysis of the structure of declared basic cost components has been made; it identified a great variety in these components. There are several objective (e.g., various business models, wide but uneven investments) and subjective (e.g. accuracy and uniformity of the data on cost components) reasons for it. The  $Ctr$  data can be evaluated as more qualitative (accurate, reliable) ones in comparison with the separate cost components, since the latter very depend on the cost allocation. In addition, exactly total costs are the efficiency criterion as well as the basic regulatory focus as the determinant of transmission tariff  $Ttr$ . Therefore, the  $Ctr$  were chosen as the dependent variable:

$$(C(k) = Ctr(k)).$$

Other mentioned network's design and performance indicators form the set of independent input indicators  $\Pi(k)$ :

$$Ctr(k) = f_k(\Pi(k)) = f_k\{Ph(k), Qpr(k), Tpr(k), Qcon(k), Dmin(k), Dmed(k), Dmax(k), L(k), N(k)\} \quad (2)$$

It can be predicted that it will not be achievable model, which is completely adequate to all real networks. The leading motive for practical purposes is to create the function  $f$  with the best possible quality. To achieve the maximum conformity of the found regularity, correlation of the particular modelled values with corresponding declared specific costs (correl ( $Ctr(m): Ctr$ )) is used as the quality criterion during the creation of the model. We have applied also the Average Normalized Squared Deviation from the Mean value (ANSDM) for assessment of the quality of the models; the normalization was used because of large range of networks size:

$$ANSDM = \frac{1}{u} * \sum_1^u ((Ctr - Ctr(m))/Ctr)^2$$

Number of input data sets  $u$  and consequently the output data is finite (number of networks  $u$ ; in our case data on 23 DH networks were available). Nevertheless, in order to have the possibility to evaluate the new and/or modified undertakings, it is necessary to create the regularity that is continuous and monotonous for any predictor (the economic logic does not indicate any reason for some extremum) against any  $\Pi(k)$  in the determined ranges of input data.

### 3. First stage of modelling: determination of key drivers

The regression analysis, as the method at the crossroads of data mining and modelling, was chosen as the most preferable statistical modelling tool, which is directly focused on the relationships between the dependent variable and several predictors. For modelling we chose the well-developed, powerful and at the same time user-friendly R statistics environment that provides the flexible modelling algorithms; the modelling was started by using the simpler linear regression algorithm. The post-modelling analysis of residuals will show the purposefulness of continuation of modelling by usage of more complicated nonlinear regression algorithms to obtain the stronger cause-effect relationship.

For the linear modelling we used the built-in linear regression model, implemented in R as a function *lm*, which is called by command *lm*. The target was to detail the general regularity (2) and to create the benchmarking model as the linear mathematical expression:

$$Ctr(lin) = \alpha + \beta_1 * i_1 + \beta_2 * i_2 + \dots + \beta_n * i_n \quad (3)$$

where:

- $\alpha$  – constant;
- $\beta_i$  – estimated optimal weight of corresponding predictor *i*.

Identification of outliers (data sets which are too far away from the central data cluster) was made before the modelling by means of standard outlier detecting procedure; the outliers may have a different essence and therefore cannot be used to create the general regularity. After outliers’ rejection from the total data set data samples of remaining 20 DH networks had been used for the modelling ( $u = 20$ ).

Checking indicators’ cross-independence level was another preparatory activity; the usage of strongly correlated predictors would lead to the so-called multicollinearity problem, i. e., the possibility of obtaining unreliable results by usage of interdependent predictors. To examine the potential problem the typical tool was used – the Variance Inflation Factors (VIF), which characterise the mutual correlation between any pair of predictors. A strong reliability criterion was chosen:

$$VIF = 1/(1-R^2) < 2.5, \text{ or } R^2 < 0.6$$

where:  $R^2$  – coefficient of determination.

The VIFs were calculated for full matrix of input indicators; the mutual correlations of five variables (*Ph*, *Qpr*, *Qcon*, *L*, *N*) were found as being well above the critical threshold). So only one of the five strongly correlated indicators can be used for modelling, the others should be excluded from the data sets. At the same time it is an approval that the structure and specific load of all networks is similar, that significantly increases trust in the results of benchmarked modelling.

To define the most preferable predictor among the strongly interconnected five indicators and so avoid the potential multicollinearity, we made the prior modelling by using the same linear regression algorithm, using all nine independent variables as predictors. The regularity (3) in this case becomes:

$$Ctr(lin 9) = \alpha + \beta_1 * Ph + \beta_2 * Qcon + \beta_3 * Qpr + \beta_4 * Tpr + \beta_5 * L + \beta_6 * N + \beta_7 * Dmax + \beta_8 * Dmin + \beta_9 * Dmed \quad (4)$$

Although the correlations between any of interconnected five predictors and *Ctr* are very strong ( $> 0.8$ ), their impact on *Ctr(lin9)* is different (their mutual correlation is one of the basic reasons). A perusal of p-values shows that the network length *L* is the most significant predictor among the five mentioned mutually-correlating ones (tab. 1); there is only 3.8% probability on random impact of the *L* on *Ctr(lin9)*. Chances of other indicators to be accidental ones are incomparably higher (27% – 85%). Therefore the network length *L* was chosen as the predictor for the basic modelling together with *Tpr*, *Dmax*, *Dmin* and *Dmed*.

**Table 1.** Statistical characteristics of *Ctr(lin9)* model and its predictors

Variables	Correl with Ctr	<i>Ctr(lin 9)</i> p-value
Ph	0.8739	0.2730
Qcon	0.9063	0.2730
Qpr	0.9285	0.6256
L	0.9200	0.0381
N	0.8138	0.8499

Developed by authors

The modeling process itself was started by creation of linear 5-predictor model in the form:

$$Ctr(lin\ 5) = \alpha + \beta_1 * Tpr + \beta_2 * L + \beta_3 * Dmax + \beta_4 * Dmin + \beta_5 * Dmed \tag{5}$$

The obtained correlation is extremely strong (tab. 2, columns *Ctr(lin5)*); the p-value for the model is very small. At the same time, the excellent correlation is partly misleading because of small amount of data sets in the 30-fold range of the actual *Ctr* values; this is confirmed by the quite large value of ANSDM (7.36%) of the actual costs *Ctr* from those of modelled regularity *Ctr (lin5)*

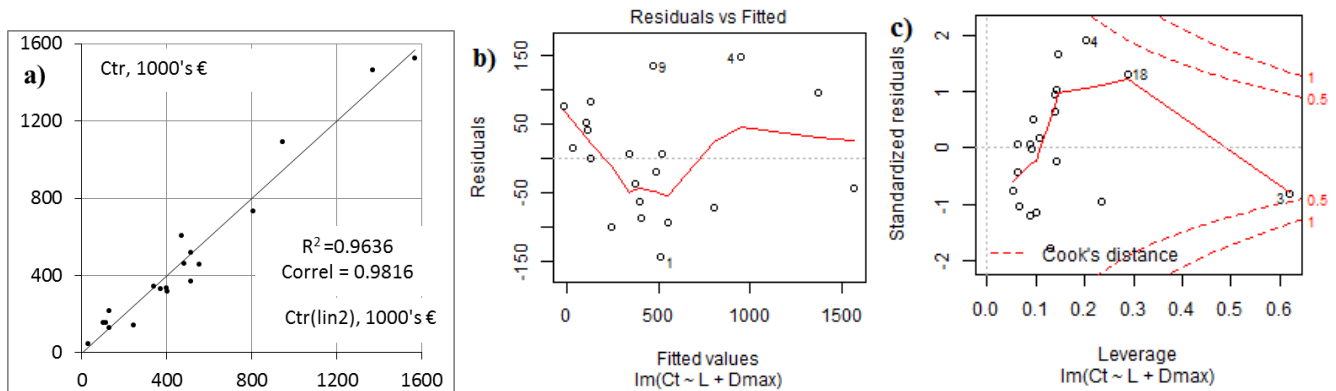
**Table 2.** Characteristics of the models and predictors

Indicators	Characteristics of the predictors					
	Linear models				Nonlinear model	
	Ctr(lin 5)		Ctr(lin 2)		Ctr(nolin)	
	p-value	Model parameters	p-value	Model parameters	p-value	Model parameters
L	5.84E-08		3.05E-10		2.11E-13	
Dmax	0.00428		0.000309		7.25E-06	
Dmin	0.70497					
Dmed	0.83559					
Tpr	0.79119					
	Statistical characteristics of the models					
	1.071E-09		5.89E-13		7.91E-16	
R <sup>2</sup>		0.9651		0.9636		0.9747
ANSDM		0.0736		0.1443		0.0303

Developed by authors

The small p-values of the *Dmax* and *L* show their decisive role in the model's regularity; it clearly means that they are the searched key drivers. The high p-values of other three predictors (*Dmin*, *Dmed* and *Tpr*) show that their impacts on costs are very accidental; these predictors only complicate the model unnecessarily. To evaluate impact of each key driver on costs, in the next modelling stage only variables *L* and *Dmax* were used as the predictors (6) for the creation of the linear 2-predictor model *Ctr(lin 2)*:

$$Ctr(lin\ 2) = \alpha + \beta_1 * L + \beta_2 * Dmax \tag{6}$$



**Fig. 2.** 2-predictor linear model; scatter plot *Ctr* vs *Ctr(lin2)* (a), diagnostic plots (b, c)

Developed by authors



The final result of the linear modelling (tab. 2, columns *Ctr(lin2)*) shows that the correlation parameters are virtually unchanged in comparison with the *Ctr(lin5)*; extremely small p-value confirms the validity of the predictors' choice. The obtained excellent correlation (Fig. 2a) confirms that L and D are the key drivers of transmission costs, i.e., the efficiency of the network; the abandoned three indicators really have not a significant impact on the searched regularity. The results are consistent with the conclusion of the techno-economic model, developed in (Pusat, & Erdem, 2014): the transmission costs increase with the area size, i.e., with the length of the network. On the other hand, costs for lying out pipes as well as the total cost of piping per meter (ETSAP, 2013; Delangle, Lambert, Shah, Acha, & Markides, 2017) directly depends on the pipes' diameter first of all, therefore the significance of the largest diameter is well explained.

Nevertheless, R diagnostic plots point the incomplete compliance of the actual data with the linear model *Ctr(lin2)*. The plot 2b shows that residuals are not equally spread around a horizontal line (especially at small fitted values). It is an indication that the linear model doesn't fully capture the existing non-linear relationship between the predictors and the target variable. The plot 2c identifies the data point that lies very near to the threshold, so called Cook's distance. This point is not an outlier; nevertheless it can become influential against a general regularity. In addition the ANSDM even increased, its reduction is necessary for more accurate evaluation of the key drivers' impact on the network's efficiency. In general the model needs to be further developed to assess the impact of the key drivers on the efficiency of transmission network.

#### 4. The second stage of modelling: evaluation of the impact of the key drivers

Using the traditional method of the least squares, it was found that the nonlinear trend line in the scatter plot of the created linear model (fig. 2a) provides a slightly higher coefficient of determination  $R^2$  in comparison with the standard linear one. It approves a necessity to add some nonlinearity in the model. To find the correct way, we checked the scatter plots (*Ctr* vs *L*) and (*Ctr* vs *Dmax*) (fig. 3). The impact of each individual predictor on the result of the total mining is, of course, different from the individual correlation (e.g., due to some mutual impact of predictors). Nevertheless, the qualitative differences between both scatter plots provide some comparative indication. The optimum trend line on the scatter plot (fig. 3a) is linear, while on the scatter plot (fig. 3b) one can observe a moderate non-linearity as well as can indicate that the inclusion of power function is preferable.

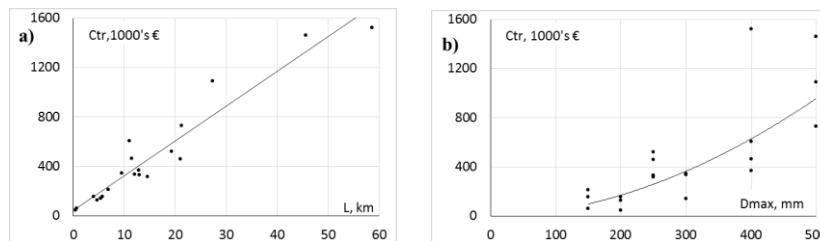


Fig. 3. Scatter plots *Ctr* vs *L* (a) and *Ctr* vs *Dmax* (b)

*Developed by authors*

Two mutually non-related modelling procedures have been applied to reduce the mentioned inconsistencies, to increase the coincidence of the searched regularity with the declared transmission costs, and to ensure the cross-check of results, thus improving the quality of the model.

To define the non-linearity, which best matches the regularities of the actual costs *Ctr*, as the first step of the nonlinear regression process we used the NLS function, which determines the nonlinear (weighted) least-squares estimates of the parameters of the nonlinear model:

$$\text{mod}(1) \leftarrow \text{nls}(Ctr \sim \beta_1 * L + \beta_2 * (Dmax)^\gamma, \text{start} = \text{list}(\beta_1 = 23, \beta_2 = 1)) \quad (7)$$

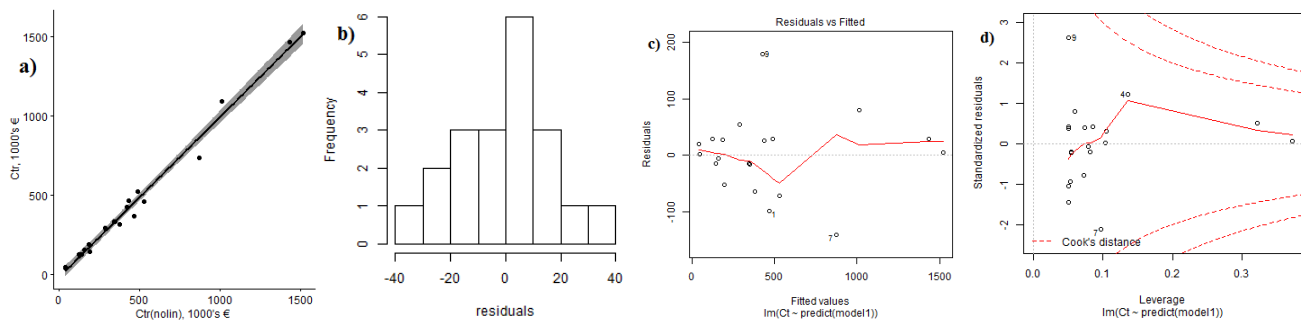
where:

- the start values of the  $\beta_1$  and  $\beta_2$  were determined according to the linear model (5);
- the optimum values of the coefficients  $\beta_1$  and  $\beta_2$  were calculated by NLS function for number of picked  $\gamma$  values using the correlation  $\text{correl}(Ctr: Ctr(\text{nolin}))$  as the quality criterion.

As the second step we used the linear modelling to find the free term of the final model  $Ctr(\text{nolin})$  by processing function (7) that was found in the first step:

$$Ctr(\text{nolin}) \leftarrow -\ln(Ct \sim \text{predict}(\text{mod}(1))) \quad (8)$$

Slightly improved formal correlation and p-value (tab. 2, columns  $Ctr(\text{nolin})$ ) in comparison with the linear model  $Ctr(\text{lin}2)$  is not the main achievement. The ANSDM is significantly reduced (3.03%), the deviation of actual costs  $Ctr$  from the modelled  $Ctr(\text{nolin})$  has decreased for 14 of 20 networks, while it has insignificantly increased only for 4 networks. The highlighted confidence interval (at showed 95% confidence level) is satisfactorily narrow. One can see that the nonlinear model (fig. 4a) is much more coinciding with the actual costs of the utilities in comparison with the linear one.



**Fig. 4.** Nonlinear model  $Ctr(\text{nolin})$ ; scatter plot  $Ctr$  vs  $Ctr(\text{nolin})$  (a), histogram of residuals (b), diagnostic plots (c, d)

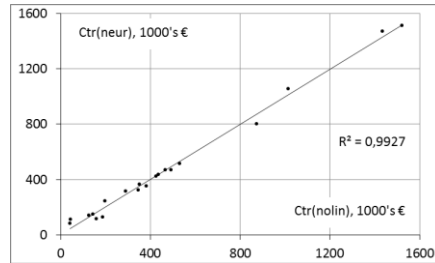
*Developed by authors*

R diagnostic plots show better behaviour of the residuals in comparison with the linear model. The residuals are more equally spread around the horizontal zero line (fig. 4c). There are no patterns between the residuals and fitted values, thus the independence assumption might hold as well. No data point is near the Cook's distance (fig. 4d). The histogram on fig. 4b suggests the normality of the residuals around the zero value that is sign of a good model.

Another modelling procedure that was applied – a classical version of multi-layer perceptron with the error back-propagation training algorithm. The multi-layer perceptrons have proven themselves in a variety of data mining tasks (e.g., Daranda, 2016; Jan, & Ali, 2018; Briedis, & Freivalds, 2018). To obtain the neural model for our case, in addition to standard procedures we introduced techniques to cope with the task-specific conditions: (a) small amount of training data; (b) the monotony requirements for the resulting function with regard to each of the input parameters.

It was realized in the pre-processing phase that it is enough to have just three neurons (computing units) in the *hidden* layer of the network to achieve approximately maximum possible accuracy; input data normalization was made by the S-type function because of very diverse scales of data values. It was also experimentally confirmed that it's enough with just 2 input parameters out of 5 to achieve almost the same quality of the model –  $Dmax$  and

*L*. In addition to correlation between modelled specific costs and declared ones, the accuracy of the modelling was controlled by the stopping criterion – the continuation of the training process until the modelling error falls below the some predetermined threshold  $\epsilon$ ; the best possible correlation obtained was 0.995 with  $\epsilon$  value of 0.00001.

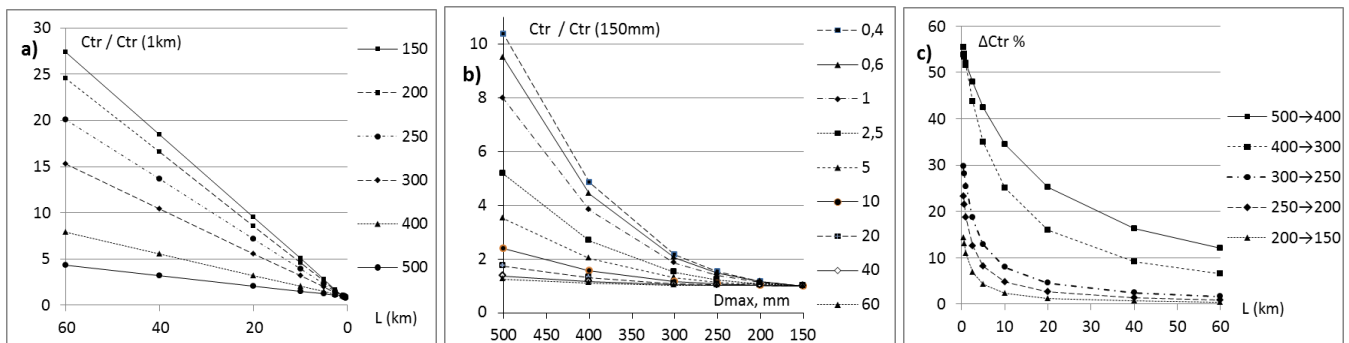


**Fig. 5.** The mutual coincidence of the neuron network model  $Ctr(neur)$  and the nonlinear regression model  $Ctr(nolin)$

*Developed by authors*

Nevertheless, the models obtained with the maximum possible accuracy/correlation too much adapt to the concrete data points used for training and thus they are invalid to model the overall process (so called *overfitting*). Therefore tuning of the model parameters was accomplished to obtain the maximum quality of the searched regularity while avoiding the undesirable overfitting and thus preserving its monotony. To overcome the effect of the overfitting, we used a qualitative method by defining several specific monotony tests with regard to each of the input parameters and applying them against the trained model. After each training episode, the obtained formula was tested to have the monotone cross-relations. E.g., to achieve the monotony, the final value of the threshold  $\epsilon$  was increased until 0.0001, the obtained correlation – 0.987.

The excellent coincidence of the both nonlinear regression and neural network models (fig. 5), created by the two mutually independent modelling tools approves that modelling results are very near to the true causal regularity.



**Fig. 6.** Impact of key drivers on network cost-efficiency;

(a): the relative transmission costs compared to  $Ctr = f(1km; Dmax)$ ;

(b): the relative transmission costs compared to  $Ctr = f(150 mm; L)$ ;

(c): percentage reduction of the transmission costs ( $\Delta Ctr$ ), reducing  $Dmax$  by one standardized size

*Developed by authors*

As it would be predicted, reduction of the length of the network  $L$  reduces costs  $Ctr$  proportionally (fig. 6a), while the reduction of the  $Dmax$  leads to a much faster enhancement of the network efficiency (fig. 6b). The percentage reduction of the transmission costs ( $\Delta Ctr$ ), reducing  $Dmax$  by one standardized size, is shown in the fig. 6c. It is understandable, that reduction of  $Dmed$  and even  $Dmin$  also will provide some effect. With larger pipes, the effect

is greater; also, as the total network length decreases, the effect increases as the  $D_{max}$ -sized pipes make up a larger part of the total network length.

## Conclusions

The created methodology and developed benchmarking models well indicate the total network length and the largest inner diameter of pipes as the key drivers of the heat transmission costs, so the network efficiency. Very high coefficient of determination and low ANSDM approve that the modelled causal regularity is very near to the true one in the full range of combination of predictors. The correlation between models, which are created by mutually absolutely independent modelling tools, is excellent. Hence, the trustworthiness in the results is high; it confirms the quality and credibility of the developed methodology as well as its appropriateness for evaluation and optimization of the network design. The methodology is a reasonable compromise between the accuracy and simplicity, which is its strong advantage for the practical applications.

Results of the current study clearly indicates that an optimum planning of the pipe routes, the choice of adequate pipes' sizes, first of all size of large diameter transmission pipes, are the key drivers designing the modern efficient transition and distribution heating networks for the development, expansion and/or renovation of the DH systems. The still used  $D_{max}$  pipes only in 3 of the 20 networks would match the volume of actual  $P_{con}$  (see, e.g., ETSAP 2013), even taking into account the reasonable reserve for the potential future growth of the consumption. In other networks, the  $D_{max}$  exploited for up to 4 standardized sizes exceeds the  $D_{max}$  optimal. Replace of existing  $D_{max}$  size pipes with the matching size pipes during the network renovation would significantly increase the network's efficiency, dropping the  $C_{tr}$  by up to 35% and thus the overall heat supply tariff ( $T_{pr} + T_{tr}$ ) by up to 12%.

In practice, of course, the actual DH systems are not as homogeneous as the theoretical cases discussed above. However, similar principles should be observed when planning their optimal development. An example is the DH system of Riga, which according to its efficiency and other performance indicators ranks among the best city energy systems in the Baltic Sea region. Its structure consists of two large DH systems and more than ten small and very small systems with the boiler houses of adequate capacity (RS, 2018), which supply remote clusters, to which it was not economically justified to build the long poorly loaded transit pipelines, or even such pipes have been liquidated during the reconstruction.

The complete rebuilding of the existing DH systems is, of course, very expensive and therefore practically impossible. Another option would be to intensify the transition to the fourth generation DH networks with a low heat carrier temperature that will require pipes of larger size. But the heating systems of the buildings currently are not yet prepared for this transition. So the analysis of various DH system models, possible re-planning of the network design and selection of adequate pipe's size is a single real option to enhance the efficiency of the network.

Nevertheless, the current study has to be evaluated as the first stage of the research; it demonstrates the correctness of the working hypothesis and perspective of the research. There is a strong basis for continuation of research to further enhance the methodology, to increase its quality and to narrow the standard segment; further improvement of the methodology certainly is possible and necessary.

Currently data on only 23 of 56 functioning networks were available for modelling; in addition three of them were classified as outliers. Increase of the number of data sets would be highly desirable to reflect better the large range of networks' sizes; mentioned differences between the correlation and ANSDM is a direct sign. Our previous experience shows that the availability of data over several consecutive years is a substantial positive factor. It will become possible to identify and to remove from the full data pack several unreliable and/or even contradictory

*bad* data sets, which distort the regularity of the declared data on the networks. Modelling will be made on the basis of the most qualitative and reliable (*good*) data sets; the obtained general regularity will be applicable to the *bad* networks of low confidence too. Fig. 5 clearly shows that some data of several networks seems being too far from the general regularity; it is known that operators of these small networks performs a number of other municipal functions too, obviously the cost accounting is far from perfect in the utility. The consequences of the model's attempts to include such data as much as possible in the common model are the deterioration of the quality of the model. E.g., in our case the exclusion of only two *bad* data sets reduces the ANSDM by one third.

Overall, the results of the study emphasize the municipalities' duty to be proactive in fulfilling their heating supply obligations, ensuring the most efficient functioning of the DH networks and thus minimizing the heat tariffs for their area's population and other consumers. It is time to revise approach of local governments to heat supply; each municipality needs to consider carefully whether to give the preference to a centralized energy model with long, inefficient large-sized pipes, or to several local heat generation stations and an efficient distribution networks. Each individual DH project must be carefully analysed by reviewing the stagnant habits and finding the optimal solution; the work and resources invested will pay off by the most effective network design and implementation for specific area.

## References

- Aberg, M., Falting, L., & Forssell, A. 2016. Is Swedish district heating operating on an integrated market? – Differences in pricing, price convergence, and marketing strategy between public and private district heating companies, *Energy Policy* 90: 222-232. <https://doi.org/10.1016/j.enpol.2015.12.030>
- Briedis, K. M., & Freivalds, K. 2018. On-line Television Stream Classification by Genre, *Baltic Journal of Modern Computing*, 6(3): 235-246. <https://doi.org/10.22364/bjmc.2018.6.3.02>
- Butler, K., Gordon, R., Roggeveen, K., Waitt, G., & Cooper, P. 2016. Social marketing and value in behavior; Perceived value of using energy efficiently among low income older citizens, *Journal of Social Marketing*, 6(2): 144-168. <https://doi.org/10.1108/JSOCM-07-2015-0045>
- Daranda, A. 2016. Neural Network Approach to Predict Marine Traffic, *Baltic Journal of Modern Computing*, 4(3): 483-495.
- DEA 2016a. *Technology Data for Energy Plants for Electricity and District heating generation*. Danish Energy Agency. Retrieved from: [https://ens.dk/sites/ens.dk/files/Analyser/technology\\_data\\_catalogue\\_for\\_el\\_and\\_dh\\_-\\_aug\\_2016\\_upd\\_oct18.pdf](https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_el_and_dh_-_aug_2016_upd_oct18.pdf)
- DEA 2016b. *Technology Data for Individual Heating Installations*. Danish Energy Agency. Retrieved from: [https://ens.dk/sites/ens.dk/files/Analyser/technology\\_data\\_catalogue\\_for\\_heating\\_installations\\_-\\_marts\\_2018.pdf](https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_heating_installations_-_marts_2018.pdf)
- Delangle, A., Lambert, R. S. C., Shah, N., Acha, S., & Markides, C. N. 2017. Modelling and optimising the marginal expansion of an existing district heating network, *Energy*, 140: 209-223. <http://dx.doi.org/10.1016/j.energy.2017.08.066>
- ETSAP 2013. *District heating (2013, January)*. Retrieved from: [https://iea-etsap.org/E-TechDS/PDF/E16\\_DistrHeat\\_EA\\_Final\\_Jan2013\\_GSOK.pdf](https://iea-etsap.org/E-TechDS/PDF/E16_DistrHeat_EA_Final_Jan2013_GSOK.pdf)
- Euroheat & Power 2019. *District Energy in the United Kingdom*. Retrieved from: <https://www.euroheat.org/knowledge-hub/district-energy-united-kingdom/>
- Eurostat 2018. *Energy Balance Sheets, 2018 Edition*. Retrieved from: <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-EN-18-001>
- Fakhri, A. S., Ahlgren, E. O., & Ekvall, T. 2016. *Modelling economically optimal heat supply to low energy building areas – the importance of scales*. Retrieved from: <https://www.slideshare.net/IEA-ETSAP/modelling-economically-optimal-heat-supply-to-low-energy-building-areas-the-importance-of-scales>

- Jan, H., & Ali, A. 2018. Classification of Latent Fingerprints Using Neural Networks, *Baltic Journal of Modern Computing*, 6(1): 31-44. <https://doi.org/10.22364/bjmc.2018.6.1.03>
- Karnitis, E., Karnitis, G., Zuturs, J., & Bobinaite, V. 2017. Modelling of water supply costs, *Procedia Computer Science*, 104: 3-11. <https://doi.org/10.1016/j.procs.2017.01.040>.
- King, M., & Shaw, R. 2010. *Community Energy: planning, Development and Delivery*. Retrieved from: <https://www.theade.co.uk/assets/docs/resources/CommunityEnergy-PlanningDevelopmentDelivery.pdf>
- Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. 2014. 4th Generation District Heating (4GDH); integrating smart thermal grids into future sustainable energy systems, *Energy*, 68: 1-11. <https://doi.org/10.1016/j.energy.2014.02.089>
- Moriarty, J. P. 2011. A Theory of Benchmarking, *Benchmarking: An International Journal*, 18(4): 588-611. <https://doi.org/10.1108/14635771111147650>
- Persson, U., & Werner, S. 2011. Heat distribution and the future competitiveness of district heating, *Applied Energy*, 88(3): 568-576. <https://doi.org/10.1016/j.apenergy.2010.09.020>
- Pusat, S., & Erdem, H. E. 2014. Techno-economic model for district heating systems, *Energy and Buildings*, 72: 177-185. <https://doi.org/10.1016/j.enbuild.2013.12.051>
- RS 2018. *Annual Reports 2017*, Rigas Siltums AS. Retrieved from: <http://www.rs.lv/en/content/annual-report>
- Saeima, 1994. Law "On Local Governments". Retrieved from: <https://likumi.lv/ta/en/en/id/57255-on-local-governments>
- Sarma, U., & Bazbauers, G. 2017. District heating tariff component analysis for tariff benchmarking model. *Energy Procedia*, 113: 104-110. <https://doi.org/10.1016/j.egypro.2017.04.029>
- Sayegh, M. A., Danielewicz, J., Nannou, T., Miniewicz, M., Jadviszczak, P., Piekarska, K., & Jouhara, H. 2017. Trends of European research and development in district heating technologies, *Renewable and Sustainable Energy Reviews*, 68: 1183-1192. <https://doi.org/10.1016/j.rser.2016.02.023>
- Smaliukiene, R., Monni, S. 2019. A Step-by-step Approach to Social marketing in Energy Transition. *Insights into Regional Development*, 1(1): 19-32. [https://doi.org/10.9770/ird.2019.1.1\(2\)](https://doi.org/10.9770/ird.2019.1.1(2))
- Song, J., Li, H., & Wallin, F. 2017. Cost comparison between district heating and alternatives during the price model restructuring process. *Energy Procedia*, 105: 3922 – 3927. <https://doi.org/10.1016/j.egypro.2017.03.813>
- UN 2017. *New Urban Agenda*. Retrieved from: <http://habitat3.org/wp-content/uploads/NUA-English.pdf>
- UNEP 2015. *District Energy in Cities; Unlocking the Potential of Energy Efficiency and renewable Energy*. Retrieved from: <http://wedocs.unep.org/handle/20.500.11822/9317>
- Werner, S. 2017. International review of district heating and cooling, *Energy*, 137: 617-631. <https://doi.org/10.1016/j.energy.2017.04.045>
- Zuturs, J., Valainis, J., Karnitis, G., & Karnitis, E. 2016. Modelling of adequate costs of utilities services. In: G. Dregvaite, R. Damasevicius (Eds.), *Communications in Computer and Information Science*, 639, (pp. 3-17). Basel, Springer. [https://doi.org/10.1007/978-3-319-46254-7\\_1](https://doi.org/10.1007/978-3-319-46254-7_1)

### *Acknowledgements*

*The study has been supported by the State Research Program (VPP) project "Development of heat supply and cooling systems in Latvia" (project No. VPP-EM-EE-2018/1-0002), and by the University of Latvia project AAP2016/B032 "Innovative information technologies". The authors would like to thank members of the Latvian Association of Heat Companies for sharing actual data.*



**Short biographical note about the contributors at the end of the article (name, surname, academic title and scientific degree, duties, research interests):**

**Ugis SARMA**, PhD student at Riga Technical University, is the Corporate Strategy Director of the Latvenergo AS. He has been employed in the energy sector for more than 20 years. His research interests focus on DH optimisation and regulation issues. Author of more than 20 publications.

<https://orcid.org/0000-0003-4654-4647>

**Prof. Girts KARNITIS**, Dr.sc.comp., is a Professor at the Faculty of Computing, University of Latvia. He is interested in data analysis and big data. During last years he has modeled number of issues – innovative and digital development, water supply, district heating.

<https://orcid.org/0000-0002-7563-6383>

**Prof. Janis ZUTERS**, Dr.sc.comp., is a Professor at the Faculty of Computing, University of Latvia. His research interests include machine learning and natural language processing.

<https://orcid.org/0000-0002-3194-9142>

**Prof. Edvins KARNITIS**, Dr.sc.eng., is the Leading Researcher at the Faculty of Computing, University of Latvia. His research interests focus on data analysis and strategic planning. Author of more than 200 research and analytical publications.

<https://orcid.org/0000-0001-6784-7327>

---

Copyright © 2019 by author(s) and VSI Entrepreneurship and Sustainability Center

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>

