REGRESSION ANALYSIS OF THE ENERGY PRODUCED IN COGENERATION AND SUPPLIED TO DISTRICT HEATING SYSTEMS

Corina Chelmenciuc*, ORCID ID: 0000-0002-5126-8539,
Constantin Borosan, ORCID ID: 0000-0003-1975-4577

Technical University of Moldova, 168 Stefan cel Mare blvd., Chisinau, MD-2004 Republic of Moldova
*Corresponding author: Corina Chelmenciuc, *corina.chelmenciuc@tme.utm.md*

Abstract. Currently, the district heating systems (DHSs) are intensively promoted both nationally and globally. The advantages of using these systems in urban areas compared to individual heating systems are practically indisputable. Still, it is essential that the calculation underlying the assessment of the economic profitability of the projects to connect the new heat consumers to DHS be made correctly by taking into account all the necessary investments and the benefits obtained from the additional amount of energy sale. The article presents the methodology for the optimal regression model selection that can be applied to predict the additional electricity produced in cogeneration mode in case of the new heat consumer connection to the DHS, based on the actual data of the thermal and electrical energy supplied to network from a CHP in the Republic of Moldova. At the same time, it is demonstrated that between the thermal energy supplied to a new consumer connected to DHS and the additional electricity produced in the cogeneration mode, there is a direct and linear correlation.

Introduction

The main purpose of the current projects on the district heating systems development in the Republic of Moldova is to provide for the heat consumers a high quality, reliable and affordable service. The success of a heat supply company depends much on the consumers' satisfaction degree and its profit - on the consumers' number connected to the DHS.

At first glance, it would seem that any new consumer connected to the DHS would brought profit to the company, but this may not always be feasible, given that the new consumer connection involves making investments, such as constructing new networks (sometimes with pumping stations), acquiring and installing the heating agent regulation equipment; thermal energy measuring equipment, etc.

In many DHS, the thermal energy is produced from cogeneration. For example, of the total thermal energy supplied to consumers connected to Chisinau DHS, about 80% is produced in cogeneration mode by the combined heat and power (CHP) plants. In this case,

in calculating of economic profitability of connecting the new heat consumer, it is necessary to consider both gains due to additional sales of heat and those related to the sale of electricity produced in cogeneration when assessing revenues.

The amount of thermal energy supplied to the new consumer can be easily determined from the thermal load for space heating, depending on the rooms/buildings volume to be heated. The calculation temperatures of indoor and outdoor air, while the amount of electricity to be additionally produced with the production of thermal energy for the new consumer, cannot be determined by applying a calculation formula. This problem can be solved by using regression analysis.

The regression analysis theory

A linear regression model is a statistical tool used for the propose to determine a dependency relationship between two variables [1]. The statistical methodology is focused on the regression function with the aim of determining the parameters that express to what extent one or more factors influence a particular process [2].

The regression analysis can be quickly done using the personal computer by applying the *Regression* function in the *Microsoft Excel* tool.

The linear regression analysis showed promising results among the statistical models due to its reasonable accuracy and relatively simple implementation compared to other methods [3].

The linear regression method can be successfully applied in the energetics analysis and is used when for a data set represented by the pairs of values (x_i, y_i), the line must be found that best approximates their placement in a diagram y = f(x).

The advantage of the method is that it generates an equation applicable to the entire range of the variable *x* and even performing extrapolations within a specific limit. The equation will allow making predictions on energy consumption, being a valuable tool in the energetics analysis.

At the same time, the regression analysis method, applied by using the Excel calculation tool, has other advantages such as the speed and simplicity of calculation and reduced possibility of the results influence by the user [4].

The linear regression method can be used in the DHS domain to determine dependencies such as:

1. the fuel consumption for the thermal energy production for space heating according to the heating degrees-days or the consumers' thermal load [5];
2. heat losses at the thermal agent distribution depending on the outdoor air temperature or the heating distribution networks length;
3. the additional electricity supplied once the thermal energy is delivered to the new heat consumer;
4. the needed investments for the new consumers' connection (i.e., for the individual heat substations) depending on the thermal load for space heating and domestic hot water preparation;
5. the thermal agent temperature supplied to the consumers depending on the outdoor air temperature, etc.

The linear regression technique must be used with caution, as it can lead to erroneous conclusions. If the data participating in the analysis are lacking, conclusions drawn

based on an insufficient data set have a low degree of credibility. The more data we have available, the higher the degree of trust.

The linear regression method application is focused on the heuristic reasoning based on which there is a linear dependence between two variables such as, for example, fuel consumption for building heating and the heating degree-days (HDD) during a month or a heating season. If the data collected over a long period are available, they can be graphically represented in a fuel consumption diagram = f(HDD).

The points thus obtained are arranged approximately linearly and must determine the line that best approximates the sequence of points, i.e., to find the equation of the form y = bx + c, which describes a dependence between discretely represented data. In the considered example, the independent variable *x* is the heating degrees-days during each month of 2020. The dependent one *y*, is the fuel consumption (natural gas) for each month (Figure 1).

By applying Eq. (1), the corresponding natural gas consumption can easily be determined for any value of the heating degree-days.

$$y = -17{,}443x + 35745. \tag{1}$$

Thus, in energetics analysis, the straight-line equation can be used to predict energy consumption for any variable value (driver):

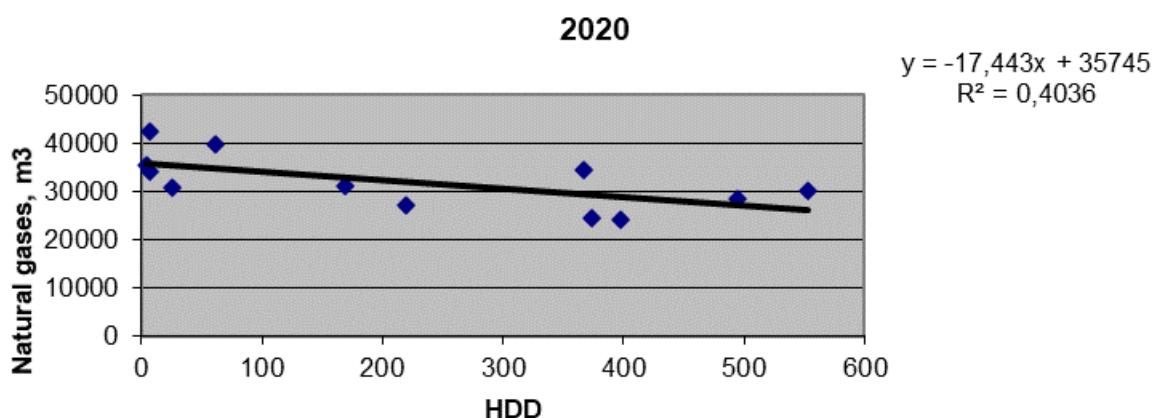$$Energy(y) = Factor \cdot Driver(x) + Constant \tag{2}$$

Figure 1. The natural gas consumption regression analysis based on heating degree-days.

The R^2 value is also significant in determining the dependence intensity of the correlation between the two variable values. The closer to 1 the R^2 value is, the more the variation of the value *x* explains the interpretation of the value *y*. The $R^2 = 0{,}4036$ (Figure 1) analysis results that the heating degree-days influenced the natural gas consumption for space heating in 2020 insignificantly.

Therefore, another driver that influenced the gas consumption in 2020 must be identified (for instance: possible thermal agent leaks in the distribution networks, heat supply with a higher temperature than it is required following the outdoor air temperature, increase the number of thermal energy consumers, etc.).

It should be borne in mind that the regression model obtained from the analysis of fuel consumption data for a given building should not be generalized and applied for the fuel

consumption prediction for any other type of building [6], especially in cases where the structure and dimensions of the building differ significantly from those of the model for which the regression model was made.

The research methodology

The methodology of regression analysis applying for the establishment of the correlation between two values (simple regression analysis), or even between three and more values (multiple regression analysis), both in the field of economics and energetics, is described in detail by authors in the cited papers. For instance, in the article [7], the doctors' number evolution in Romania in the period after accession to the European Union is modeled using the regression analysis depending on three variables: the population in this country, the wage in the health field, and the number of doctors who operates in the private sector.

In the article [8], by applying the regression method, the correlation between the supermarkets' energy consumption in Great Britain and two variables is deduced: the outside air temperature and its relative humidity, to predict the energy consumption evolution in these types of buildings depending on the climate change.

The same regression method can also be successfully applied for the energy consumption prediction for cooling needs, for instance, in fruits cold storages depending on the fruits amount stored and the outdoor air temperature [9].

The article [10] presents the methodology of the regression analysis applying to predict the energy consumption in China depending on some variables less characteristic for energy analyses: the GDP value and the population number in that country.

Daily data of the heat and electricity supplied into the network from a CHP plant located in the Republic of Moldova will be used to develop the linear regression model related to the correlation between the electricity supplied to the network in addition to the heat supplied to the new consumer connected to the DHS, which is produced at the CHP plant.

Also, the heating degree-days during three consecutive heating seasons (2018 - 2019, 2019 - 2020 and 2020 - 2021), lasting 166 days each, will be used to assess the correlation between these two values.

Table 1 presents the data for the 2018 - 2019 heating season.

Similar data were used for the other two heating seasons, specific to the 2019 - 2020 and 2020 - 2021 heating seasons.

In the first stage, the hypothesis regarding the correlation between the resultant variable - the electricity produced and the variable considered as an influencing factor - the supplied heat, separately for each heating season, will be verified.

Suppose there are different correlation degree results for each heating season. In that case, the regression analysis will be carried out to establish the correlation between the thermal energy produced and HDD, separately for each heating season.

The supplied heat in the DHS should depend on how cold the weather outside is. HDDs were collected from [11].

In the second stage, the most appropriate model will be established that can be recommended for application to predict the volume of additional electricity produced in case of connecting the new consumer to the DHS.

The amount of heat and electricity supplied into the network by CHP and HDD

Day	Heat (Gcal)	Electricity (kWh)	HDD	Day	Heat (Gcal)	Electricity (kWh)	HDD	Day	Heat (Gcal)	Electricity (kWh)	HDD
1	1743,7	1184272	3,8	57	7982,8	4830608	16,9	113	6895,8	3576304	14,3
2	1969,4	1256624	7,9	58	8155	4855312	17,4	114	6937,6	3573376	12,2
3	2494,5	1534432	6	59	8150,9	4838240	19,4	115	6845,4	3547344	12,1
4	2668,6	1818544	5,6	60	7993,9	4752240	18	116	6505,1	3633024	11,2
5	3113,1	2188432	1,3	61	7710,8	4609008	17,5	117	6372,2	3556848	7,5
6	2514	1917392	1,5	62	7685,8	4596256	15,8	118	6285,6	3529408	8,5
7	2486,8	1978592	1,6	63	7714,2	4635408	11	119	6062,6	3489888	10,2
8	2742,7	1983232	0,4	64	7690,7	4605584	14,2	120	6047,7	3485184	9,2
9	3286,7	2064704	2,6	65	7497,8	4444112	14,9	121	6081,1	3595472	9,2
10	3957,9	2406512	2,8	66	7284	4319952	14	122	6531,2	4056496	14,9
11	4490,5	2706272	4	67	7356,7	4428112	11,6	123	6953,6	4199440	20,4
12	4450,4	2680640	4,9	68	7391	4445664	11,1	124	6988,7	4211488	19,1
13	4651,7	2817664	6,5	69	7518,9	4537008	11,5	125	6937	4162640	14,5
14	4953,3	2959664	6,7	70	7423,7	4437520	13,3	126	6623,5	4072320	10,6
15	5224,6	3049840	10	71	7250,6	4318656	14,8	127	6628,3	4104176	11,8
16	5234,2	3088272	11,1	72	7443	4493008	14	128	6462,3	3937392	6,9
17	5482,5	3298640	10,5	73	7828,7	4728992	14,8	129	6299,9	3868368	8
18	5569,1	3349968	9,5	74	7960,8	4746512	16	130	6434,7	4012640	14,6
19	5601,9	3359920	10,3	75	7952,9	4745808	18,3	131	6469	3990624	15,4
20	5950,1	3486544	10,5	76	8186,4	4887840	18,5	132	6071,9	3773680	8,1
21	6229	3592480	14,4	77	8279,1	4930880	19	133	5602,3	3412352	5,9
22	6321,7	3688224	16,7	78	8482,8	4997824	19,5	134	5621,5	3437056	10,8
23	6157,8	3581392	12,8	79	8351,7	4922352	22,1	135	5637	3414896	9,8
24	5841,4	3467840	14,8	80	8393,6	4972960	23,5	136	5454,3	3307536	3,7
25	5877,6	3478640	15,1	81	8674,4	5142416	17,9	137	4939,9	3094912	3
26	6402,7	3697376	14,8	82	8532,4	5070800	17,5	138	5210,3	3240736	5,4
27	6932,4	4097600	14,4	83	8420,3	5010880	19,3	139	4915,1	3226976	4
28	7280,3	4409376	14,5	84	8269,3	4820528	18,4	140	5076,9	3243344	9,9
29	7516,8	4584992	15,1	85	8237,1	4813904	18,2	141	5222,1	3262032	12,1
30	7668,9	4674784	18,2	86	8134,7	4822000	16,6	142	5327,5	3259040	10,1
31	7792,5	4754752	19,6	87	7941,6	4680432	17,9	143	5239,6	3224640	8,3
32	7816,8	4768192	18,3	88	7878,3	4649088	16,8	144	5146,3	3131920	8,1
33	7778,3	4719440	14,1	89	7941,3	4710336	14,6	145	5065	3008816	5,4
34	7711,9	4628848	13,1	90	8028,2	4749920	15,5	146	4645,8	2906096	5
35	7711,5	4618944	13,6	91	8076,9	4768080	16,6	147	4528,2	2913760	5,3
36	8048,6	4793904	18,1	92	8039,7	4735728	16,6	148	4803,3	2980480	7,6
37	8287,2	4907344	19,8	93	7880,0	4669168	18,9	149	4890	3017504	7,5
38	8572,1	5111792	21,8	94	8092,4	4804656	15,3	150	4713,7	2935280	3,7
39	8491,8	5039248	21,2	95	8042,5	4743216	17,7	151	4783,6	2926560	7,5
40	8572,1	5086864	22,3	96	7962,7	4698608	21	152	4799,1	2954192	8,2
41	8270,9	4863808	17,4	97	7954,5	4697072	19	153	4691,3	2914496	6
42	7692,7	4631792	12,3	98	7466,6	4426864	20,2	154	4793,3	2967664	9,1
43	7644,5	4624112	13,3	99	7365,1	4387648	15,4	155	5275,2	3324688	11,8
44	7737,7	4686992	14,1	100	7159,8	4279120	12,3	156	5429,6	3248288	11,5
45	7902,2	4800512	17,3	101	7093,2	4285232	11,3	157	5315,2	3093648	9,7
46	7715,3	4624960	14,3	102	7004,4	4287616	12,5	158	4968,5	2934096	6,1
47	7618,5	4540816	13,1	103	6709,4	4213600	12,4	159	4333,7	2861856	5,3
48	7525,6	4519952	14,8	104	6624,2	4286704	11,8	160	3967,9	2869104	4,7
49	7273,6	4363904	13,6	105	6689,3	4291920	7,9	161	4233	2791072	7,3
50	7313	4405312	14	106	6845	4247920	8,9	162	4358,6	2819328	10,3
51	7412,3	4490064	14,2	107	6851,9	4272160	12,1	163	4296,5	2784240	9,1
52	7406,3	4493136	15,3	108	6840,5	4275936	12,3	164	4123,8	2765856	7,8
53	7631,4	4612368	17,3	109	6932,6	4361072	15,9	165	3969,6	2682432	6,8
54	7688,6	4606496	16	110	6935,7	3899920	14,3	166	3520	2631696	4,4
55	7812	4711024	16,9	111	7005,8	3601776	10,5				
56	7868	4761776	17,4	112	6669,3	3618880	11,9				

The results of the regression analysis: produced electricity - supplied heat

The corresponding graphs for each heating season were drawn to verify the correlation between the electricity produced and the thermal energy supplied into the network. The straight-line equation was generated, and the influence degree of the variables over the R^2 value was determined (Figures 2, 3 and 4). The R^2 value must obtain values higher than 0,5 to consider the thermal energy supplied into the network as a determining factor for the produced electricity.

Heating seson 2018-2019

$$y = 550,41x + 338014$$
$$R^2 = 0,9739$$

Electricity, kWh / 24 h

Heat, Gcal / 24 h
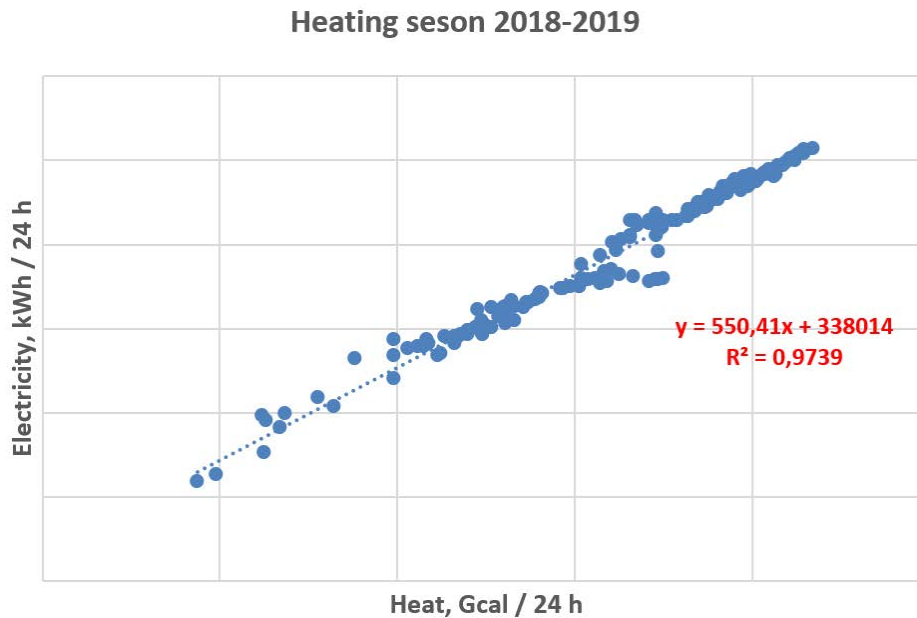
Figure 2. The regression analysis of produced electricity depending on the supplied thermal energy for the 2018 - 2019 heating season.

Heating seson 2019-2020

$$y = 433,65x + 825856$$
$$R^2 = 0,6852$$

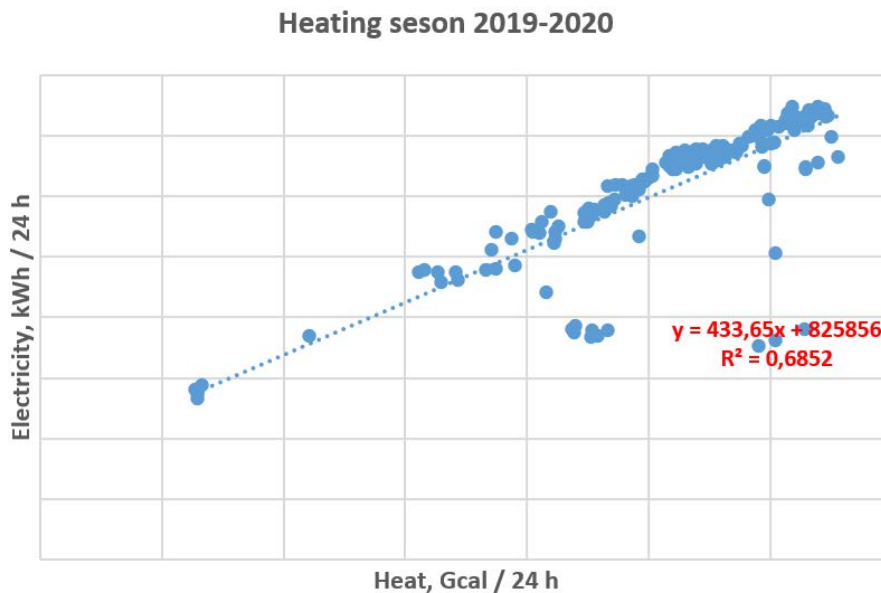Electricity, kWh / 24 h

Heat, Gcal / 24 h

Figure 3. The regression analysis of produced electricity depending on the supplied thermal energy for the 2019-2020 heating season.

When analysing the graphs in Figures 2, 3, and 4, the graphical representation of the two variables is a straight line. Thus, it can be stated that there is a direct correlation between the electricity produced and the heat supplied, including the fact that $R^2 > 0,5$.

At the same time, it can be observed that for the three heating seasons, different values of the variable R^2 were obtained, the best value - $R^2 = 0,9739$ was obtained for the 2018 - 2019 heating season, and the weakest value - $R^2 = 0,6852$ - for the one from 2019 - 2020.
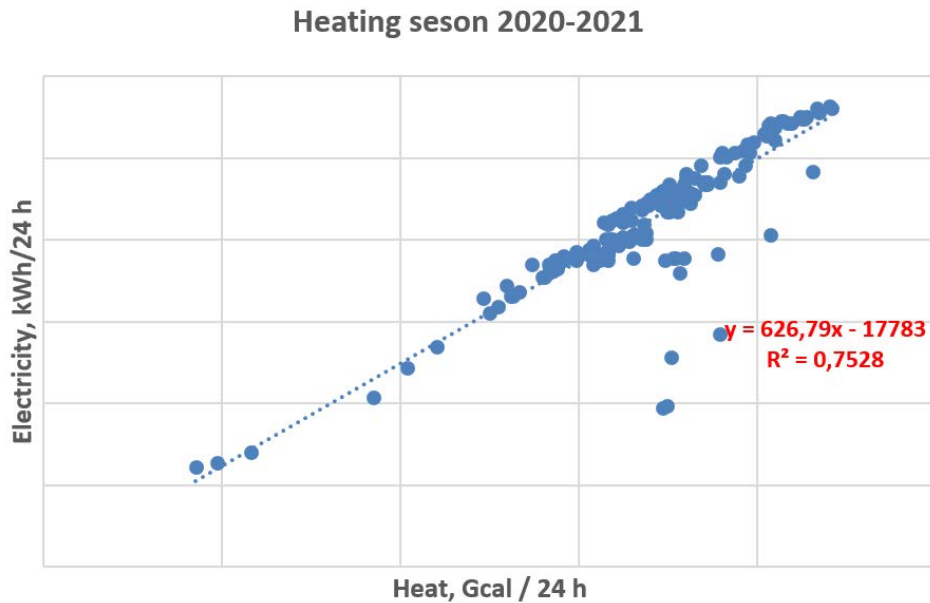
Heating seson 2020-2021

$$y = 626,79x - 17783$$
$$R^2 = 0,7528$$

Figure 4. The regression analysis of produced electricity depending on the supplied thermal energy for the 2020-2021 heating season.

The results of the regression analysis: supplied thermal energy - HDD

Different correlation degrees between the produced electricity and the supplied heat was obtained in this article for different heating seasons. Therefore, it is necessary to perform the regression analysis for the thermal energy produced depending on the HDD to recommend the linear regression model for application to predict the additional produced electricity in case of the new consumer connection to the DHS.
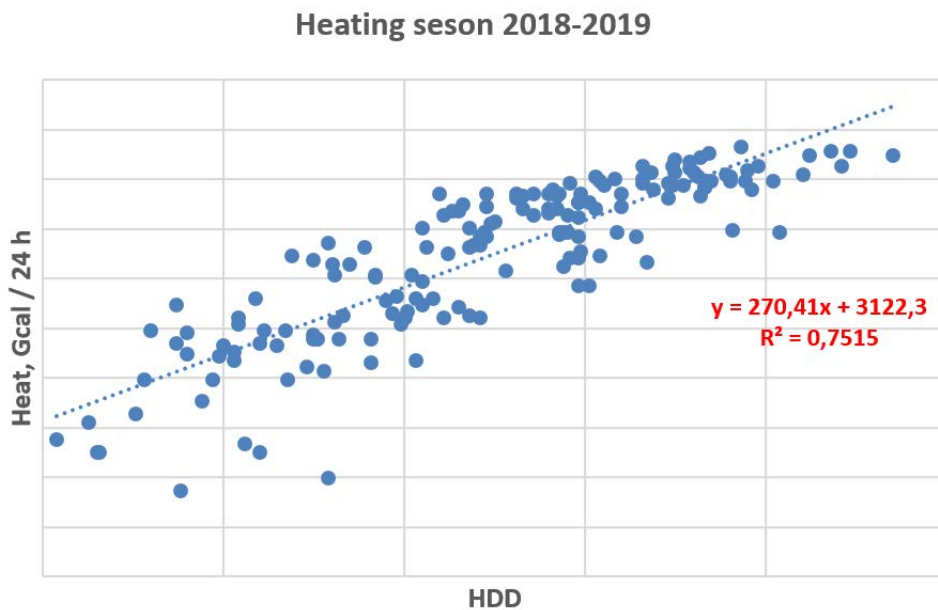
Heating seson 2018-2019

$$y = 270,41x + 3122,3$$
$$R^2 = 0,7515$$

Figure 5. The regression analysis of supplied heat depending on HDD for the 2018-2019 heating season.

When analysing the graphs in Figures 5, 6, and 7, the graphical representation of the two variables is a straight line. Thus, it can be stated that there is a strong correlation between heat supplied into the network and HDD, including the fact that $R^2 > 0,5$.
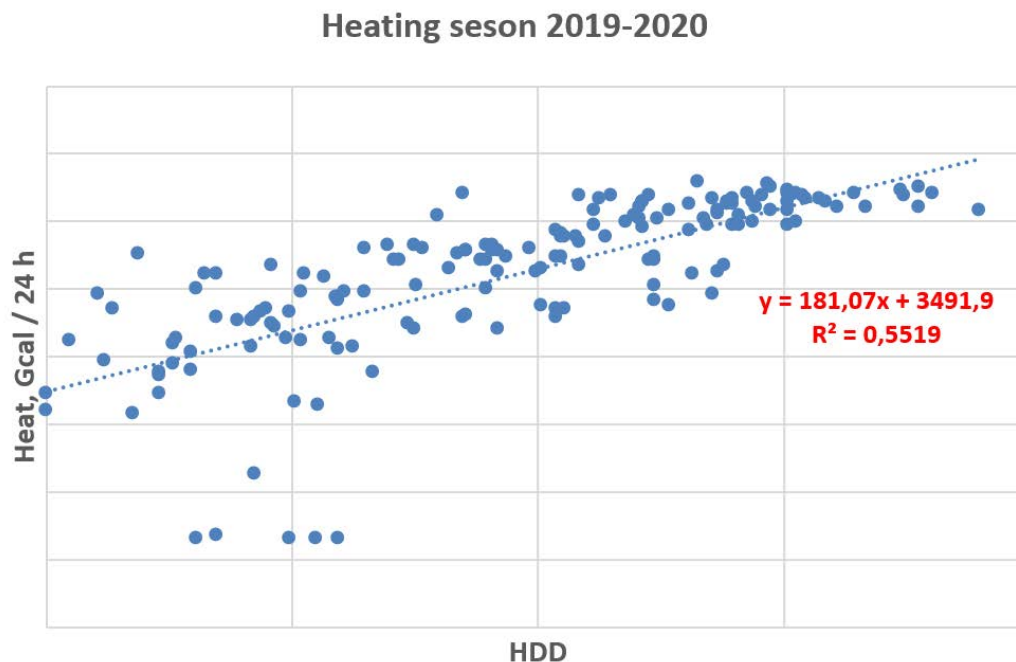
Heating seson 2019-2020

Figure 6. The regression analysis of supplied heat depending on HDD for the 2019 - 2020 heating season.
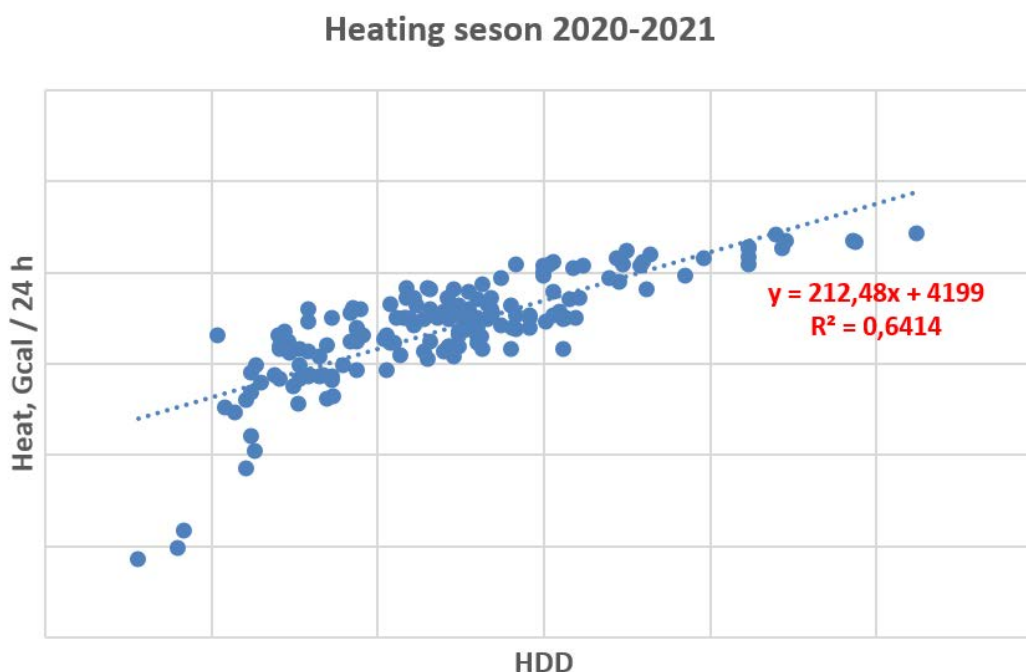
Heating seson 2020-2021

Figure 7. The regression analysis of supplied heat depending on HDD for the 2020 - 2021 heating season.

It can be observed that for the 2018-2019 heating season, the best correlation was obtained between the supplied thermal energy and the heating degree-days, namely R^2 = 0,7515. For the other two heating seasons, weaker correlation degrees were obtained.

Thus, taking into account the fact that the volumes of supplied thermal energy for the consumer's buildings heating should be dependent on the outdoor air temperature, it can be concluded that in the 2018-2019 heating season, to a greater extent, this condition was met.

The regression model

As a result of the performed analysis, the best correlation degree was obtained between the electricity produced and the heat supplied into the network ($R^2 = 0,9739$) and between the heat supplied into the network and HDD ($R^2 = 0,7515$) in the 2018-2019 heating season. As a result, it is recommended to use the obtained linear regression model for the dependence of produced electricity - heat supplied into the network in the 2018-2019 heating season, to predict the additional electricity produced in cogeneration mode in case of the new heat consumer connection to the DHS, used as a model in this article, namely:

$$y = 550,41 \cdot x + 338014,$$ (3)

where: y - the additional annual electricity produced, x - the thermal energy yearly supplied to the new consumer connected to the DHS.

Conclusions

The regression model use, having as variables the thermal energy supplied to the new consumer connected to the DHS and the additional electricity produced in cogeneration mode, demonstrated a direct and linear correlation between these two variables.

Thus, it is recommended to use the described method in this article to predict the additional quantity of the produced electricity, along with the heat supplied to the new consumer connected to any DHS in which these two forms of energy are produced in cogeneration mode. The regression model choice will be made carefully, selecting the one corresponding to the heating season for which the highest correlation degree was obtained both between the pairs of variables electricity produced - thermal energy supplied into the network and the thermal energy supplied into the network - HDD.

The obtained results will be used in the pre-feasibility studies to connect the new thermal energy consumer to the DHS by considering the income due to the production and supply of the additional electricity.

References
1. Sfetcu M., Marinescu I.: Using the linear regression model in order to analyse the correlation between the gross domestic product and the household effective individual final consumption. In: Romanian Statistical Review, 2016, Supplement nr.12, pp. 69 - 73.
2. Anghelache C, Radu I.: Construction of the regression model for the economic risk analysis. In: Romanian Statistical Review, 2020, Supplement nr.7, pp.4 - 22.
3. Fumo N., Biswas M.A. *Regression analysis for prediction of residential energy consumption,* [online]. Renewable and Sustainable Energy Rewiews, 2015, Volume 47, pp.332 - 343.
4. Vesterberg J. *A regression approach for assessment of building energy performance.* Umea: Print&Media, 2014.
5. Moletsane P. P., Motlhamme T. J., Malekian R., Bogatinoska D. C. Linear regression analysis of energy consumption data for smart homes. In: International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)**,** 2018, pp. 433 - 437.
6. Leigh S-B., Won J-S., Bae J-I.: An energy management process and prediction of energy use in an office Building. In: Journal of Asian Architecture and Building Engineering, 2005, pp. 501 - 507.
7. Mărăcine L., Poșircă I. Using the regression method in the factorial analysis of the evolution of the number of medicines in Romania in the post-accession period to the EU. In: ECOSTUDENT – Revistă de cercetare științifică a studenților economiști, 2017, nr. 10, pp. 4 - 11[in Romanian].

8. Braun M.R., Altan H., Beck S.B.M. *Using regression analysis to predict the future energy consumption of a supermarket in the UK*, [online]. 2014, Volume 130, pp. 305 - 313. [accesed 20.05.2021]. Available: https://www.sciencedirect.com/science/article/pii/S0306261914005674
9. Safa M., KC B., Safa M.: Linear model to predict energy consumption using historical data from cold stores. In: International Journal of Advances in Science Engineering and Technology, 2015, Spl. Issue 3, pp. 146 - 150.
10. Li Y. Prediction of energy consumption: variable regression or time series? A case in China. In: Energy Science & Engineering, 2019, pp. 2510 - 2518.
11. Regression Analysis of Energy Consumption and Degree Days in Excel. https://www.degreedays.net/regression-analysis [accesed 28.05.2021].