LATVIAN JOURNAL OF PHYSICS AND TECHNICAL SCIENCES
2010, N 6
10.2478/v10047-010-0028-8

EVALUATION OF CO_2 EMISSIONS FROM ENERGY SOURCES IN LATVIA

D. Blumberga, I. Veidenbergs, J. Gusca, M. Rosa

Riga Technical University, Institute of Energy Systems and Environment,
1 Kronvalda Blvd., Riga, LV-1010, LATVIA
e-mail: dagnija.blumberga@rtu.lv

The authors propose an empirical model for evaluation of the CO_2 emissions released from the energy generation sources as a function of the fuel types, energy efficiency of the technologies used, and the emission factors of the fuels. In the research, multifactor linear and nonlinear regression equations are employed. The developed model has been proved with the data of the energy generation sources taking part in the Latvian Emission Trading Scheme.

Key words: *energy generation, CO_2 emissions, regression equations, empirical model, correlation.*

1. INTRODUCTION

Energy production is one of the leading sectors in generation of CO_2 emissions in Latvia. Since the European Union Emission Trading Scheme was started in 2005, the enterprises have become more interested in evaluating the processes in which CO_2 emissions are generated, with the aim to reduce their total amount. The general correlation between the technically-economic parameters of power generation and the CO_2 emissions created during it as well as the impacts on the Emission Trading Scheme participants are evaluated and the mathematical models developed in several research works [1–3].

The influence exerted by implementation of the EC Directive on Carbon Capture and Storage upon the power generation sector of Latvia has not been evaluated before, and therefore might change the validity and interconnections of the dependent and independent parameters of the developed mathematical models. The paper presents a mathematical description for the CO_2 emissions generated by the Latvian energy production sources with the carbon capture and storage scheme implemented.

2. METHODOLOGY

2.1. Data of energy sources on CO_2 emissions

Operational data of the energy sources are collected from direct and indirect measurements. Direct measurements represent an objective method for determination of relevant parameters. Generally, this method gives lesser errors than other approaches [4]. The accuracy of direct measurements depends on the precision of the measuring equipment used and other factors. One of the most significant issues is of how simultaneously and carefully the measurement data are read and whether the situation in which measurements are made characterises the normal operation of the system − in other words, whether the collected data are representative.

The data used in the research are provided by the enterprises taking part in the Emission Trading Scheme. These data are:

- fuel used;
- energy produced from each fuel type (E_i), MWh/year;
- energy efficiency of technologies (η_i);
- CO_2 emissions of each fuel type (CO_{2i}), tCO_2/year.

Using the above mentioned data, the following parameters are defined:

- levelled annual CO_2 emissions of the energy sources, tCO_2/year;
- levelled annual energy produced by the energy sources (E_i), MWh/year;
- levelled annual energy efficiency factor of the energy technologies, η;
- levelled annual emission factors of the energy sources (F), tCO_2/MWh.

2.2. Empirical data processing

Empirical data processing, made to obtain the equation for determination of CO_2 emissions from the energy sources, was performed using the statistical data processing methods – the correlation and regression analysis.

One of the key issues in the data statistical analysis is the size of datasets: what size of the set should be used to be representative and to give the opportunity to develop adequate empirical models. Practice shows that the number of measurements should be 10 to 20 times greater than that of independent variables (factors) in the regression equation. For example, if the equation has three independent variables, the minimal size of dataset is 30 measurements. A greater amount of data allows developing more accurate models; however, this requires much more work, time and resources. Therefore, the choice of the dataset size is a compromise, and the optimal solution as to the time and costs is needed for data collection and the required model accuracy.

Statistical data processing is to be performed before the analysis of data. The objective of the processing is to produce graphic or analytical correlations between the dependent and independent variables − e.g., to find the correlation between dependent variables or factors (CO_2 emissions) and independent ones characterised by the operational parameters of the energy source. The sought – for analytical expression is known as the phenomenon or process of the regression equation (mathematical model) that characterises the relationship between independent and dependent variables (factors).

The specific succession of activities performed for statistical data processing is as follows.

- Determine the regression equation of the process. For this, the method of least squares is the most commonly used.
- Perform the statistical analysis of coefficients for the regression equation. This step involves regression analysis.
- Determine the density of the stochastic correlation between independent and dependent random variables. This step involves correlation analysis.

During the statistical processing of the data, the parameters that characterise best the operation of technologies should be selected and further included into the descriptive regression equation. This can be done in two ways:

- Having included all possible factors into the regression equation, to evaluate their significance with the statistical analysis of equation coefficients. Excluding the statistically insignificant factors, to identify the best regression equation of the process.
- To indicate graphically the relationship between a dependent variable and each factor and evaluate their correlation. Factors with significant correlation are included in the regression analysis.

It is important to assess whether the significant and independent variable values do not correlate in-between and it is correct to use the factors in the multifactor regression equation in order to draw the relevant equation of the system processes. The statistical processing of data and the development of the multifactor empirical model are conducted with STATGRAPHICS Plus software.

Correlation between the independent and dependent random variables can be estimated with a correlation coefficient. For a single-factor mathematical model, Pearson's expression is applied:

$$r = \frac{\sum_{i=1}^{m}(x_i - x)(y_i - y)}{(m-1)S_x * S_y}, \quad (1)$$

where x_i, y_i – couples of independent and corresponding dependent values;
x, y – averaged arithmetical values of independent and dependent values;
S_x, S_y – dispersions of the sample size.

The multifactor correlation coefficient (R) is used in the multifactor correlation cases. This coefficient is not statistically interpreted; however, it is determined and used as an indirect expedience indicator of the regression equation.

The correlation ratio is employed instead of the correlation coefficient when there is a nonlinear regression. In this case it has the same meaning as the correlation coefficient in the linear regression, i.e. describes the grouping of the results along the nonlinear regression line. The correlation between the datasets in the nonlinear regression analysis is also known as the correlation analysis.

The accuracy of the mathematical models describing the correlation is evaluated with correlation coefficients. It is observed that correlation exists if the values of these coefficients are in the range of 0.8...0.9. It should be noted that statistical processing software normally calculates the square of the correlation coefficient (R^2).

If the correlation coefficient squared (R^2) is multiplied by 100, this value (in percentage) describes the changes in the dependent variables characterised by the regression equation. For example, the value R^2 = 0.9 indicates that the regression equation is characterised with 90% of the dependent random variables.

3. REGRESSION ANALYSIS OF CO_2 EMISSION DATA OF THE ENERGY SOURCES

The research objective stated in this chapter is to obtain the multifactor empirical equations which will describe in a quantitative way the interconnections between the CO_2 emission values, the energy produced, the fuel emission factors, and the technology efficiency.

Regression analysis defines the changes of random factors in a view of accurate quantitative parameters, i.e. it expresses the significance of stochastic connections through functional correlations [5].

The regression analysis in the research has been done according to the following scheme:

- verification of distribution of the dependent variable values;
- definition of the regression equation using the least squares method;
- statistical analysis of the results obtained.

The following regression equations are used in the research:

- multifactor linear equations;
- multifactor nonlinear equations.

Regression analysis is a good analytical tool. This is often used for analysing experimental data; its results are correct if correct analytic rules are used [4]. There are a vast number of rules that not always are fully obeyed in practice. This fact often decreases the accuracy of a regression equation or cripples its evaluation results.

Several main terms of using the regression analysis exist. These are as follows.

1. *Verification test.* The use of regression analysis is correct if the dependent variable (CO_2 emissions of the energy generation sources) complies with the normal allocation rule. This requirement is not valid for independent variables. This means that the analysis begins with defining the allocation of dependent variables, and this can be followed up if the allocation corresponds to the normal allocation rule.

Verification of the allocation rule is conducted for the analysed dataset $m = 72$. The data characterise the CO_2 emissions generated by different fuels and with dissimilar efficiency from the energy generation sources. The allocation of a dependent variable (CO_2 emissions) is shown in Fig. 1.

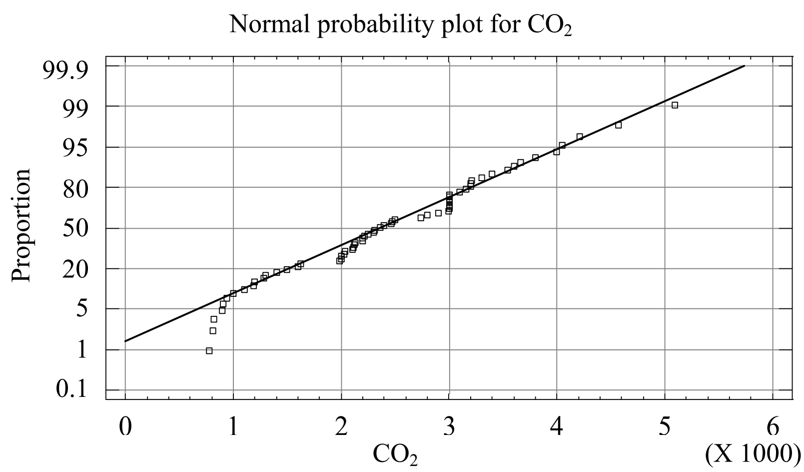

Fig. 1. Allocation of CO_2 emissions of the energy generation sources.

The normal allocation rule in logarithmic coordinates is graphically shown as a trend line along which the analysed data points are allocated (see Fig. 1). For minor CO_2 emissions a deflection from the line is observed. This means that the allocation is close to the normal allocation rule and that the use of regression analysis is justified.

2. *Checking autocorrelation.* One of the most important requirements at the use of the least squares method (LSM) is to avoid the autocorrelation of residues (the difference between the observed and predicted values of energy consumption). Otherwise the evaluation results of the LSM would be crippled. For checking autocorrelation, the Durbin–Watson test is widely used. In the statistical data processing software a Durbin–Watson criterion is set, and if it is higher than 1.4 the autocorrelation of residues is insignificant.
3. *Multicolinearity test*. The multicolinearity impacts (significant correlation between the independent variables of the regression equation) occur if multifactor regression models are used. In these cases errors in the evaluation of regression coefficients are observed. To analyse the multicolinearity, various methods exist: defining the variance inflation factor (VIF), analysing the correlation matrix of calculated coefficients, etc.
4. *Heteroscedasticity test*. If allocation of the residues is irregular and changeable according to the changes in the factor values, this points to the presence of heteroscedasticity.

The correctness of using the regression analysis method can be verified at different stages of the regression analysis: the variable allocation rule can be proven without any regression equation − only from the input data, while auto-correlation, multicolinearity and heteroscedasticity tests are based on the develop-ment and analysis of the regression equation.

The regression equation is set down with empirical data and can be written as

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + bx = b_0 + \sum_{i=1}^{n} b_i x_i, \tag{2}$$

where y – dependent variable factor;
b_0 – regression factor;
$b_1 ... b_n$ – regression coefficients;
$x_1 … x_n$ – independent variable factors.

The accuracy of the regression equation can be increased by including the double and triple relationship effects of factors and by creating the expanded equation:

$$y = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i,j=0}^{n} b_{i,j} x_i x_j, \tag{3}$$

where $b_{i,j}$ – coefficients of double correlation of factors;
$x_i.x_j$ – effects of double correlation of factors.

It should be mentioned that with increasing number of factors in the equation this becomes more complicated and difficult to use in practice. The empirical analysis shows that the expanded equation does not give a significant increase in the multifactor correlation coefficient, and therefore it was assumed to develop the regression equation (Eq. (2)) in the research.

To evaluate the statistical significance of coefficients b_0.. b_n of Eq. (2), criterion t characterised by the Student allocation with f freedom degrees is used. The number of freedom degrees is determined from the formula:

$$f = m - (n + 1), \tag{4}$$

where m – size of empirical dataset;
n – number of independent variables in the regression equation.

The calculated t criterion is compared with the values from the Student allocation tables (t_{tab}) according to chosen significance level (P) and freedom degrees (f). Significance level $P = 0.1$ is often used for the energy data processing, with the confidence probability thus corresponding to $1 - P = 0.9$. If the inequality $|t| > t_{tab}$ is in force for the evaluated coefficient, it is considered significant and is kept in the regression equation. The developed regression equation should however be evaluated [6].

Dispersion analysis using the Fisher criterion (F) is employed to evaluate the ratio between the dispersion of dependent variables and the dispersion of residues:

$$F(f_1, f_2) = \frac{S_y^2(f_1)}{S_{res}^2(f_2)}, \tag{5}$$

where $S_y^2(f_1)$ – dispersion of dependent variable y;
$S_{res}^2(f_2)$ – dispersion of residue.

The residue is defined as the difference between the value of the dependent variable and that calculated with regression analysis $(y_i - y_i^{calculated})$.

Degrees of freedom (f_1 and f_2) are found from the equalities:

$$f_1 = m - 1;$$
$$f_2 = m - n. \tag{6}$$

If the F criterion exceeds the critical value identified from the F allocation tables taking into account the freedom degrees (f_1 and f_2) as well as the significance level (P), it can be stated that the equation describes the analysed data and can be used in practice.

The regression equation for expression (2) is:

$$CO_2 = b_0 + b_1.E + b_2.F + b_3.\eta, \tag{7}$$

where CO_2 – CO_2 emissions generated by the energy generation sources, tCO_2/year;
b_0, b_1, b_2, b_3 – coefficients;
E – energy consumed, MWh/year;
F – CO_2 emission factors of the fuels used, t CO_2/MWh;
η – energy efficiency of the technologies.

The coefficients of the regression equation and its statistical evaluation are summarised in Table 1.

Table 1

Coefficients of regression equation and its evaluation

Independent variables	Coefficients	t , statistics	P
Constant b_0	4432.4	1.71	0.0967
Constant b_1	0.2427	164.755	0.0000
Constant b_2	39526.9	1.9379	0.0568
Constant b_3	–16058.0	–1.772	0.0809

The significance level $P = 0.1$ is chosen for data processing and corresponds to the probability level of 0.9 and the freedom degree of 68. For all cases (see Table 1), the relationship $I\ t\ I > t_{tab}$ is in force; therefore, all the parameters are defined as significant and have to be included in the equation.

In the research, the regression equation describing the dependence of CO_2 emissions on the energy efficiency of technologies, fuel type and produced energy (see Eq. (8)) has been developed:

$$CO_2 = 4432.4 + 0.2427\ .E + 39526.9\ .F + (-16058.0.\eta). \quad (8)$$

The defined regression coefficient (in the statistical data processing) of the developed empirical model is 0.94 ($R^2 = 0.94$). This means that the developed model (Eq. (8)) describes 94% of the analysed data changes and the residue (6%) corresponds to the impacts from the independent variables excluded from the equation (or not analysed in the research) or from the interconnections of the effects.

The adequateness of Eq. (8) is evaluated with the Fisher criterion: its calculated value ($F = 9111.3$) is compared with that from the table ($F_{tab.} = 1.47$) using the values of freedom degrees. Equation (8) can be considered adequate and is used correctly as far as $F > F_{tab}$.

Autocorrelation test. Using the Durbin–Watson test and processing the statistical data, the DW criterion is defined (DW=1.67). As far as this value exceeds 1.4, it can be concluded that no significant autocorrelation of the residues is observed and that the evaluation of the achieved values is not crippled in the process of the LSM evaluation.

Multicolinearity test is developed analysing the correlation matrix of the calculated coefficients of regression equation (see Table 2).

Table 2

Correlation matrix of regression equation coefficients

Coefficients	b_1	b_2	b_3
b_1	1.0	0.0473	0.0938
b_2	0.0473	1.0	0.35
b_3	0.0938	0.35	1.0

Analysis of the correlation matrix (low values of the correlation coefficients (<0.5)) shows that insignificant correlation between the coefficients and the independent variables exists. It can therefore be concluded that the definition of the regression coefficients is correct.

Heteroscedasticity test is performed in a graphic way, checking the allocation of residues as a function of the technological energy efficiency. The allocation of residues is shown in Fig. 2.

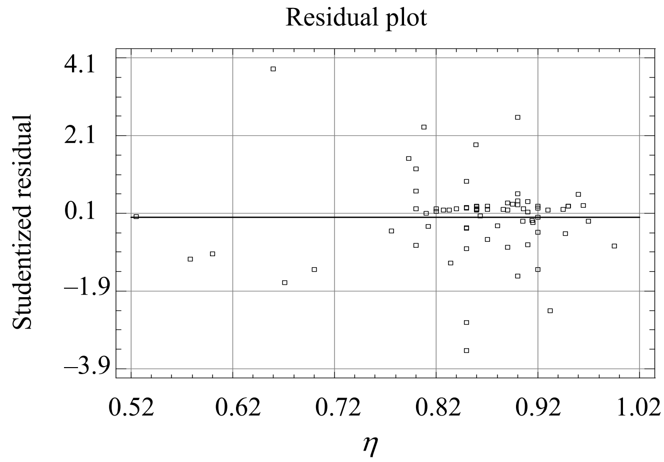

Fig. 2. Allocation of the residues depending on the energy efficiency of energy generation sources.

As shown in Fig. 2, there are no significant changes in the allocation of the residues depending on the energy efficiency; thus, the allocation is observed as regular. In the research, analysis of other factors has been performed, with no heteroscedasticity impacts found. Thus the standard error is defined correctly.

The validity of the physical nature and, therefore, the logical validity of the described processes are evaluated by signs of the factors. The signs of two first parameters of the regression equation (Eq. (8)) used for defining the CO_2 emissions from the energy generation sources are positive: increase in the energy production and the CO_2 emission factors of fuels (F) leads to increase in CO_2 emissions; in turn, increase in the energy efficiency (η) leads to decrease in CO_2 emissions. The considered tendency complies with the real nature of the processes and is logically explainable.

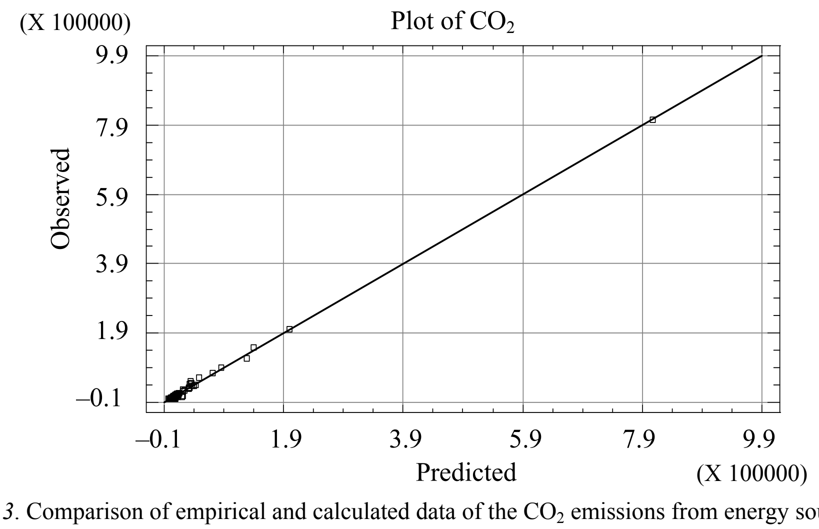

Fig. 3. Comparison of empirical and calculated data of the CO_2 emissions from energy sources.

One of the most significant issues when using empirical equations is: whether and how well the results calculated with empirical models correlate with the empirical data. The data comparison of the adequacy test of the empirical equation is given in Fig. 3.

As seen from Fig. 3, there is a good correlation between both datasets.

4. CONCLUSIONS

Processing of the statistical data for CO_2 emissions from the energy generation sources and analysis with regression methods have been performed. The relationship between the CO_2 emissions values, the energy produced, the fuel emission factors and the energy efficiency of technologies is developed in the form of a regression equation (the empirical model). Testing for correctness of the chosen stages of the regression analysis and possibilities to move to the next stage of analysis was conducted regularly during the research.

The performed analysis leads to the following conclusions:

- the use of the regression analysis is correct as far as the values of dependent variables comply with the normal allocation rule;
- the use of the least squares method is reasonable for definition of the factors whose values are not crippled because the defined DW criterion exceeds the allowed limit value (1.67>1.4);
- evaluation of the coefficients of the regression equation is correct as far as no correlation between them is noticed;
- the standard error of the data assessment is evaluated correctly as far as the allocations of residues are even for dependent and independent variables.

REFERENCES

1. Boemare, C., Quirion, P., & Sorrell, S. (2003). The evolution of emissions trading in the EU: tensions between national trading schemes and the proposed EU directive. *Climate Policy*, *3,* Suppl. 2, S105-S124.
2. Laurikka, H. (2002). Absolute or relative baselines for JI/CDM projects in the energy sector. *Climate Policy*, *2* (1), 19–33.
3. Dorer, V., & Weber, A. (2009). Energy and CO_2 emissions performance assessment of residential micro-cogeneration systems with dynamic whole-building simulation programs. *Energy Conversion and Management*, *50* (3), 648–657.
4. Rošā, M. (2006). CO_2 emisiju samazinājuma mehānismu analīze sadedzināšanas iekārtām Latvijā. *Promocijas darbs,* Rīga: RTU. 124 lpp.
5. Blasnik, M. (1995). The need for statistical analysis and reporting requirements: some suggestions for regression models. *Energy Program Evaluation Conf.*, 15–17 October, p.45, Chicago (USA).
6. Revina, I. (2002). *Ekonometrija*. Rīga: EuroFaculty.
7. Veidenbergs, I., & Blumberga, A. (2001). Ēku energopatēriņa regresijas modelis. *Proceedings of International Scientific Conf.* "Power and Electrical Engineering", 11–13 October, 125–130. Riga: RTU.

ENERGOAVOTU CO_2 EMISIJU NOVĒRTĒJUMS LATVIJĀ

D. Blumberga, I. Veidenbergs,
J. Gusca, M. Rosa

Kopsavilkums

Darbā veikta energoavotu CO_2 emisiju datu statistiska apstrāde un analīze, izmantojot regresijas analīzes metodes. Modelis izstrādāts energosistēmām, kurās enerģijas ražošana ir apvienota ar oglekļa dioksīda uztveršanas un uzglabāšanas procesu izmantošanu. Sakarība starp CO_2 emisiju vērtību, saražoto enerģiju, kurināmā emisiju faktoriem un iekārtu energoefektivitāti iegūta regresijas vienādojuma (empīriska modeļa) veidā. Regresijas analīzes gaitā katrā no tās posmiem izdarītas pārbaudes par veiktā soļa pareizību un iespēju pāriet pie analīzes nākamā posma.

28.10.2010.