Intellectual analysis and estimation of gross greenhouse gas emissions

Mikhail Gorodnichev¹, Yuri Trofimenko², Timur Potapchenko¹³, Lionella Fedotova³

¹Moscow Technical University of Communications And Informatics, 8a, Aviamotornaya str., 111024, Moscow, Russia
²Moscow Automobile and Road Construction State Technical University (MADI), 64, Leningradskaya avenue, Moscow, Russia
³Federal State Budgetary Institution «Center for Strategic Planning and Management of Medical and Biological Health Risks»(Center for Strategic Planning, Russian Ministry of Health) 10 b.1, Pogodinskaya St., Moscow, Russia, 119121

E-mail: gorodnichev89@yandex.ru

Abstract. Currently, the system of accounting and monitoring of greenhouse gas emissions is being formed in the Russian Federation. It should be noted that within the framework of the system being formed, the transport sector is considered as the main contributor that generates large amounts of greenhouse gas emissions in cities and on busy federal highways. At the same time, it is proposed to take into account emissions directly when burning fuel only by road, city electric, rail, water and air transport. The article discusses methods for estimating gross emissions from road facilities using data mining methods aimed at verifying the results obtained. The article also presents the calculation of greenhouse gas emissions for road facilities in the pilot region with a description of the verification algorithm for the results obtained. The results of testing the methods for determining greenhouse gas emissions by road facilities showed that this calculation method can be used in estimating greenhouse gas emissions at road facilities, however, reliable baseline data for calculating gross greenhouse gas emissions can only be obtained by analyzing the activities of economic entities in transport and in the road sector. It should be borne in mind that business entities are not obliged to provide information constituting commercial secrets from their point of view. Therefore, the results obtained are of an estimated nature and relate only to those objects, data on which could be obtained from official statistics and free sources, while the actual amount of greenhouse gas emissions from road facilities in the pilot region during the period under review may be significantly higher.

1. Introduction

Every year a large number of transport infrastructure facilities are under construction in the Russian Federation, as well as being repaired. At the same time, all stages (preparation of the territory before construction; production of construction materials (for structural layers of pavement and pavement); arrangement of the road; delivery of construction materials; construction works; monitoring of the road condition, repair work, etc., are related to the formation greenhouse gas emissions resulting from the operation of road and construction equipment.
Currently, the system of accounting and monitoring of greenhouse gas emissions is being formed in the Russian Federation. It should be noted that within the framework of the system being formed, the transport sector is considered as the main contributor that generates large amounts of greenhouse gas emissions in cities and on busy federal highways. At the same time, it is proposed to take into account emissions directly when burning fuel only by road, city electric, rail, water and air transport.

However, it should be borne in mind that in modern large cities, the pace of construction and reconstruction of roads and transport infrastructure facilities is constantly increasing and increasing every year, while avoiding inventory and accounting of greenhouse gas emissions associated with the construction and operation of transport infrastructure facilities may result a lack of understanding of the situation related to the volumes of greenhouse gas emissions that are being formed in the sector of road facilities which will affect not only the environment, but also the inability to make an effective policy decisions and measures aimed at reducing greenhouse gas emissions, as well as the transition to low-carbon technologies and materials in the construction.

In accordance with the above, the purpose of this article is to analyze the main methodological approaches associated with estimating greenhouse gas emissions from road facilities, adapting these approaches to conducting an inventory of greenhouse gas emissions, making calculations for the pilot region and evaluating the effectiveness of the results obtained using machine learning methods.

2. Methodical approaches in conducting an inventory of greenhouse gas emissions from road facilities objects

The construction and operation of transport infrastructure is associated with the use of conventional and specialized road equipment, which consumes various types of energy (diesel, gasoline, gas, etc.). As a result of the operation of all this road machines in the process of combustion of fuel, pollutants are released into the atmosphere, which also include greenhouse gases, i.e. In order to estimate greenhouse gas emissions, a tool for calculating should be based on a method for estimating emissions based on compliance with the rules of the fuel and energy balance, since greenhouse gases are generated directly during fuel combustion.

It should be noted that the methodological approaches to estimating greenhouse gas emissions by transport at the international level are based on the requirements of the guidelines and instructions of the Intergovernmental Panel on Climate Change (IPCC) which are based on a method of assessment based on the rules of the fuel and energy balance [1,2,3].

In the case of road facilities objects, key categories of emission sources in accordance with the IPCC guidelines can be determined by two methods: the Tier 1 method and the Tier 2 method. The Tier 1 method assumes that the key categories include the emission source categories that contribute the most or the total amount of emissions greenhouse gases, or in the trend of total emissions relative to the base year. The Tier 2 method is similar to the Tier 1 method, but for calculations you need to use only regional values of emission factors based on real carbon content in the fuel.

It should be noted that initially, the approaches laid down in the IPCC methodology were focused on providing information on the national inventory of greenhouse gas emissions, however, an analysis of foreign documents showed that the IPCC methodology also adapts to the needs by large companies, which allows them to estimate greenhouse gas emissions for making management decisions.

In the case of road facilities, the estimation of greenhouse gas emissions using the simplified Tier 1 methodology is based on data on the consumption of various types of fuel and energy resources during construction (reconstruction), maintenance, repair and overhaul of federal, regional, local roads of different technical categories, The IPCC average recommended emission factors for greenhouse gases (CO2, N2O, CH4) from the combustion of various types of fuels, production of heat and electrical energy. The calculation of greenhouse gas emissions by road construction according to the Tier 1 methodology is made using the formula:

\[ E_i = \sum_a \sum_k (FC_{ak} \cdot EF_{aTCE} \cdot EF_{aNCV} \cdot EF_{i,a} ) \]
where \( E_i \) is the greenhouse gas emission \( i \), \( t / \text{year} \); \( FCak \) - total mass of consumed energy of type \( a \) on the roads of the \( k \)-th technical category, \( t / \text{year} \); \( EF \alpha TCE \) - conversion factor to tonnes of reference fuel by type of energy source \( a \), tons of ton / ton; \( EF \alpha NCV \) - conversion factor to energy units by type of energy resource \( a \), TJ / tf; \( EF_i, a \) - greenhouse gas emission factor \( i \) by type of energy resource \( a \), t / TJ.

The Tier 2 method is similar to the Tier 1 method, but regional values of greenhouse gas emission factors based on real carbon content in fuel and other energy resources consumed in the region during the year are used for calculation, and also data on the length of federal, regional, and local automobiles roads are used, roads of different categories, built, repaired, being in operation during the year in the region and the specific consumption of energy resources (per unit length) during construction (reconstruction), maintenance, repair (overhaul) of roads.

The calculation of greenhouse gas emissions by the road sector according to the Tier 2 methodology is based on the formula:

\[
E_i = \sum_a \sum_k \sum_j \left( L_{kj} \cdot m_{kj} \cdot EF_{\alpha TCE} \cdot EF_{\alpha NCV} \cdot EF_{i, a} \right) = \sum_a \sum_k \sum_j \left( \Delta L_{kj} \cdot g_{ijk} \right)
\]

where \( E_i \) is the greenhouse gas emission \( i \), \( t / \text{year} \); \( \Delta L_{kj} \) - increase in the length of highways of the \( k \)-th technical category, \( \text{km} \); \( m_{kj} \) - specific energy consumption of a type at the \( j \)-th stage of the life-cycle (construction, maintenance, repair) of \( k \)-th category roads, \( \text{t} / \text{km length} \); \( g_{ijk} \) - specific emissions of greenhouse gas \( i \) at the \( j \)-th stage of the life cycle (construction, maintenance, repair) of roads of the \( k \)-th category, \( \text{t} / \text{km length} \).

It is necessary to clarify that the increase in the length of the road network \( \Delta L_{kj} \) is used to calculate greenhouse gas emissions only during the construction (reconstruction) phase. In the remaining stages of the life cycle (repair, maintenance) - is calculated using data on the total length of roads.

The variety of mobile and stationary sources of greenhouse gas emissions in the road industry of varying intensity, make it difficult to reliably determine the initial data when quantifying greenhouse gas emissions require a number of assumptions and limitations. These restrictions are typical for the inventory of greenhouse gas emissions for other types of transport and are set for them by regulatory documents prior to the introduction of other requirements in the Russian Federation for road facilities.

Among the assumptions are the following:
1. When quantifying greenhouse gas emissions, only CO2 emissions are taken into account.
2. Greenhouse gas emissions from road machines and mechanisms, as well as during construction and use of other road infrastructure objects (asphalt concrete plants, etc.) are not considered.
3. Greenhouse gas emissions on unpaved roads of all technical categories are not taken into account.
4. When calculating greenhouse gas emissions according to the Tier 1 and 2 methodologies, specific emissions at the construction (reconstruction) and repair (overhaul) stages for certain categories of roads are not considered, but their integral weighted average values for constructed (reconstructed) or repaired federal and regional highways in this year are considered.
5. Greenhouse gas emissions at the stage of maintenance are calculated for the entire length of federal, regional and local roads of all technical categories with the exception of unpaved ones.

Additionally, it should be noted that the IPCC methodology also provides for a third level of calculation of greenhouse gas emissions, however, this method when estimating energy consumption and greenhouse gas emissions uses detailed modeling and involves taking into account the characteristics used in construction (reconstruction), repair (major repairs) and maintenance of automobile roads of different technical categories averaged by the IPCC-recommended greenhouse gas emission factors (CO2, N2O, CH4) from burning various types of fuels, producing heat and electric energy, road construction, and materials used technologies in the life cycle of the roads of different technical categories, climatic and other factors. Therefore, due to the lack of necessary background information and high labor intensity, calculations of greenhouse gas emissions using the Tier 3 method are not performed.
Also, on the basis of formulas 1 and 2, it follows that for estimating greenhouse gas emissions from road facilities, a large amount of input data is required, which may contain erroneous values in the process of collection and processing, which can greatly affect the results obtained, and as a result lead to incorrect assessment of the generated amounts of greenhouse gas emissions and making wrong decisions aimed at reducing emissions. That is why, in carrying out calculations of greenhouse gas emissions from road facilities, it is advisable to apply data mining methods, namely machine learning methods, which allow analyzing a large amount of data with regard to the percentage of correctly classified values and the root-mean-square error.

To solve the problem of verifying the data used, the most suitable is the CART algorithm [4] which forms a model based on the structure of the binary tree, where each internal node is a test on the parameters, each branch is a positive or negative test result, and each node is a label class. A feature of the algorithm is the ability to build not only a classification tree, but also a regression tree.

Associated with each node (including the root node) of the tree is a subset of the training set. The task of the algorithm is to build a hierarchical classification model in the form of a tree of many examples \( T \). The process of building a tree in this algorithm occurs from top to bottom: first the root of the tree is created, then the descendants of the root, then their descendants, etc. Initially, the tree consists of a single root node with which the initial set of examples \( T \) is associated. To split it into 2 subsets, select one of the attributes of the set as a check in the node.

However, the question arises of how to choose from all the attributes of the set necessary for verification and further partitioning. It is worth noting that most often the CART tree to solve this problem uses the “Gini index” as a criterion for evaluating partitioning. At the same time, as an evaluation criterion, it is necessary to specify “information increment”. Below is a description of it [5].

Let \( N(C_j, T) \) be the number of examples from some set \( T \) belonging to the class \( C_j \). Then the probability that a randomly selected example of their set \( T \) will belong to the class \( C_i \):

\[
P_i = \frac{N(C_i, T)}{|T|}
\]

(3)

According to information theory, the amount of information contained in a message depends on its probability. The expression that forms the estimate of the entropy, which is further needed to determine the class of the example from the set \( T \), has the following form:

\[
\text{Entropy}(T) = \sum_{i=1}^{k} P_i \log_2 P_i
\]

(4)

The entropy estimate, but already after splitting \( T \) into subsets by attribute \( X \), is:

\[
\text{Entropy}_x(T) = \sum_{i=1}^{n} \frac{T_i}{|T|} \text{Entropy}(T_i)
\]

(5)

Then the criterion for choosing an attribute will be the following expression of information gain:

\[
\text{Gain}(X) = \text{Entropy}(T) - \text{Entropy}_x(T)
\]

(6)

Criterion (6) is calculated for all attributes, from which one is chosen, the partition with which the 2 subsets will give the greatest gain information.

It should be noted that attributes can be both numeric and categorical. By processing a numeric attribute, the algorithm generates a rule of the form \( x_i \leq c \), where \( c \) is the threshold value. When referring to categorical features, a rule of the form \( x_i \ast F(x_i) \) is formed in a node, where \( F(x_i) \) is a non-empty subset of classes \( x_i \). When generating each node, the algorithm goes through all the attributes and selects the one at which the best result is obtained. For \( n \) numeric attributes, the algorithm makes \( n-1 \) comparisons, and for \( n \) categorical attributes, \((2n-1) – 1\) comparisons. After the selection is made, the data is distributed in descending branches.

3. Results of approbation of the methodology for determining greenhouse gas emissions by road facilities objects using data mining methods
The city of St. Petersburg and the Leningrad Region was chosen as a pilot region for testing the methodology for determining the volume of greenhouse gas emissions from road facilities. This choice is due to the fact that there is a high concentration of vehicles in this region, with a significant amount of transport work carried out in St. Petersburg and the Leningrad Region, leading to the need for continuous improvement and development of the existing road and transport network and infrastructure, as well as volumes works in the region under consideration allow an assessment without the need to take into account the large number of objects and the related difficulties in finding the necessary background information.

The main source data necessary for estimating greenhouse gas emissions in the road sector are the indicators of the length of federal and regional public roads in St. Petersburg and the Leningrad Region by technical categories. In addition, data on the increase in the length of the road network, the volume of construction and reconstruction of public roads is taken into account. Most of the data was obtained from open sources of information, for example, on the website of the Office of the Federal State Statistics Service for St. Petersburg and the Leningrad Region (Petrostat) [6]. The calculation was carried out for the base 2017. Statistics for the specified year is the most complete and reliable.

The results of calculating gross greenhouse gas emissions by enterprises of the pilot region in 2017 are given in Table 1. According to the calculations, the volume of gross greenhouse gas emissions by enterprises of the pilot region in 2017 was 109,965 tons of CO$_2$.

Table 1. Gross greenhouse gas emissions by the enterprises of the road sector in the pilot region in 2017 (Tier 1 method), tons / year.

| Type of fuel and energy resource | CO$_2$ mass emission |
|---------------------------------|----------------------|
| Motor gasoline                  | 7893.2               |
| Diesel fuel                     | 97,387.2             |
| Coal of other deposits          | 4685.1               |
| Total GHG emissions             | **109,965**          |

The main difference between the Tier 2 methodology and the Tier 1 methodology is that the regional values of specific energy consumption and specific emissions of greenhouse gases (per kilometer length) during construction, maintenance and repair of highways should be used to calculate greenhouse gas emissions. carbon in fuel and other energy resources consumed in the region during the year, as well as data on the length of federal, regional, local highways of various categories, Renovated, in operation throughout the year in the region.

According to statistics, in 2017 in St. Petersburg and the Leningrad region an annual increase in the length of the road network amounted to an average of 14.3 km due to the commissioning of the constructed sections of roads.

When calculating greenhouse gas emissions from the pilot area of the pilot region using the Tier 2 methodology in 2017, weighted integral values of specific greenhouse gas emissions were used at the maintenance stage of different technical categories.

The results of the assessment of gross emissions of greenhouse gases by the road sector according to the Tier 2 methodology based on the accounting of specific emissions of greenhouse gases and the length of roads of the respective categories are presented in Figure 1.

According to calculations, gross emissions of greenhouse gases by the road infrastructure of the pilot region according to the Tier 2 methodology in 2017 amounted to 114,353 tons of CO$_2$.

Comparison of the obtained values of gross greenhouse gas emissions from the pilot area of the pilot region according to the Tier 1 and 2 methods showed that the difference in the values of greenhouse gas emissions does not exceed 5%.
Figure 1. Greenhouse gas emissions (million tons of CO2) in the maintenance, repair and construction of roads in St. Petersburg and the Leningrad Region (2017) [7].

Then, the data used in the calculations were studied using the CART algorithm for the percentage of correctly classified data and the root-mean-square error, in order to determine how high the confidence interval the obtained values are, the results of such an assessment are presented in Figure 2 [8,9].

Figure 2. Percentage of correctly classified data using CART algorithm for tier 1 and tier 2 methods.

Analysis of the data presented in Figure 2 showed that there is a high degree of correctly classified data, with a low mean-square error, the average value of correctly classified data is 97%, ie only 3% of the data is classified incorrectly, which indicates that the data used in calculating greenhouse gas emissions have a high degree of accuracy.
The results of testing the methods for determining greenhouse gas emissions by road facilities showed that this calculation method can be used in estimating greenhouse gas emissions at road facilities, however, reliable baseline data for calculating gross greenhouse gas emissions can only be obtained by analyzing the activities of economic entities in transport and in the road sector. It should be borne in mind that business entities are not obliged to provide information constituting commercial secrets from their point of view. Therefore, the results obtained are of an estimated nature and relate only to those objects, data on which could be obtained from official statistics and free sources, while the actual amount of greenhouse gas emissions from road facilities in the pilot region during the period under review may be significantly higher [10].

It should be noted that reliable information on greenhouse gas emissions by individual transport and road enterprises can appear only after making changes in the regulatory framework, namely after the adoption of the Federal Law on Greenhouse Gases with the introduction of an audit procedure. Also, as shown by the analysis of existing forms of statistics, in order to assess greenhouse gas emissions from stationary and mobile sources of road facilities, the data contained in the statistical reporting forms is not enough, and therefore, these forms of statistical reporting need to be modified would allow a more accurate estimate of greenhouse gas emissions. Accounting for this circumstance is very important for compiling inventories of greenhouse gas emissions for all subjects of the Russian Federation.

4. Discussion

Thus, the assessment of greenhouse gas emissions showed the possibility of using a two-tier system for estimating greenhouse gas emissions for road facilities while confirming the adequacy of the results obtained, it is advisable to carry out using data mining methods, however, it should be noted that machine learning is a large section that includes in itself a variety of different algorithms and approaches, the choice of an algorithm depends on the objectives of the study. It should also be noted that with the help of machine learning it is possible to solve problems not only in relation to large data in order to find and determine patterns that allow one to establish the necessary dependence, but also from the point of view of predictive analytics. For example, the use of such machine learning tools as classification and regression analysis will not only increase the effectiveness of the assessment, but also predict how data may change, which can be crucial in making management decisions.

However, the existing statistical accounting system does not allow for a comprehensive inventory of greenhouse gas emissions, and therefore, to make management decisions aimed at reducing greenhouse gas emissions, it is necessary to improve the statistical observation forms in the area of providing data on road gains and on the number of vehicles used and consumed fuel, making the appropriate changes will significantly improve the accuracy of determining the amount of greenhouse gas emissions.

References

[1] Caviedes A, Figliozzi M 2018 Transportation Research Part F: Traffic Psychology and Behaviour 58 pp 488-499 doi: 10.1016/j.trf.2018.06.032
[2] Kim S, Lee Y, Moon H 2018 Renewable and Sustainable Energy Reviews 81 pp 3061-3069 doi: 10.1016/j.rser.2017.08.067
[3] Marshall W, Ferenchak N 2019 Journal of Transport & Health 13 100539 doi: 10.1016/j.jth.2019.03.004
[4] Assis J, Giacomini H, Ribeiro M 2019 Ecological Indicators 133 pp 303-316 doi: 10.1016/j.ecolind.2018.12.012
[5] Pei J, Zhou B, Lyu L 2019 Applied Energy 241 pp 174-183 doi: 10.1016/j.apenergy.2019.03.033
[6] Erokhin S, Petukhov A, Pilyugin P 2019 Conference of Open Innovations Association FRUCT 24 pp 594-598
[7] Atayero A, Sheluhin O, Ivanov Y 2013 Lecture notes in electrical 170 pp 15-28 doi: 10.1007/978-94-007-4786-9_2
[8] Gorodnichev M, Nigmatulin A 2013 Advances in intelligent systems and computing 224 pp 195-204 doi: 10.1007/978-3-319-00945-2_17

[9] Buldin I, Gorodnichev M, Makhrov S, Denisova E 2018 Wave Electronics and its Application in Information and Telecommunication Systems WECNF 2018 8604408 doi: 10.1109/WECNF.2018.8604408

[10] Gorodnichev M, Polyantseva K, Yashina M 2019 Systems of signals generating and processing in the field of on board communications SOGS 8706721 doi: 10.1109/SOSG.2019.8706721