Scientific Papers. Series E. Land Reclamation, Earth Observation & Surveying, Environmental Engineering. Vol. XII, 2023
Print ISSN 2285-6064, CD-ROM ISSN 2285-6072, Online ISSN 2393-5138, ISSN-L 2285-6064

MACHINE LEARNING-BASED MODELING FRAMEWORK FOR IMPROVING ROMANIAN RESILIENCE STRATEGY TO GREENHOUSE GAS EMISSIONS IN RELATION TO VISEGRAD GROUP

Stefan-Mihai PETREA, Ira-Adeline SIMIONOV, Alina ANTACHE, Aurelia NICA, Cristina ANTOHI, Dragos Sebastian CRISTEA, Maxim ARSENI, Madalina CALMUC, Catalina ITICESCU

REXDAN Research Infrastructure, "Dunarea de Jos" University of Galati, 98 George Cosbuc Street, Galati, Romania

Corresponding author emails: ira.simionov@gmail.com, dragoscristea@yahoo.com

Abstract

The present research reveals the difference between Romania and V4 in terms of the Greenhouse Gas Emissions Strategy and establishes a machine learning (ML) - based modeling framework for improving the ability to reach zero GHG by the mid-21st century. The ML tree-based algorithms, based on dual dimension environmental-economic nexus, revealed that net greenhouse gas emissions (NGHGE) are mostly conditioned by greenhouse gases from agriculture (GHGA), a fact valid both in the case of Romania (feature importance - FI = 0.41) and V4 (FI = 0.86). However, for V4, the 2nd important predictor is identified as greenhouse gases from waste management (FI = 0.26), while in the case of Romania, the national expenditure on environmental protection has a limited impact (FI =0.02) on NGHGE. Both integrated models have good prediction accuracy (Rsq. 0.70, RMSE 0.53 for the model associated with the Romania database and Rsq. 0.76, RMSE 0.47 for the V4 model). It can be concluded that in terms of integrated GHG emissions management strategy, Romania can merge with V4 to increase the environmental efficiency towards achieving the EU environmental goals.

***Key words**: machine learning, environmental modeling, GHG, environmental strategy, tree-based models.*

INTRODUCTION

It is a reality that climate change impacts the world's environmental heritage, a fact revealed by extreme weather conditions such as heat waves of rapidly changing climate. Thus, the European Union Commission responded to this environmental challenge by adopting European Green Deal and Intergovernmental Panel for Climate Change suggested a carbon neutrality target by 2050 (*EUR-Lex - 52021DC0240 - EN - EUR-Lex*, n.d.; Von Der Leyen, n.d.). Greenhouse gas emissions (GHG) are a major concern nowadays and new EU joint countries are making efforts to align with the established desideratum. Thus, regional groups have been created, such as Visegrad Group (V4), to encourage and sustain optimum cooperation between the countries, for the successful accomplishment of common goals, such as GHG reduction.

The European Union has been a leader in addressing climate change and reducing greenhouse gas emissions. In 2019, the EU set a target to achieve net-zero greenhouse gas emissions by 2050 (*Virtual Workshop - UNFC Europe: Ensuring Sustainable Raw Material Management to Support the European Green Deal | UNECE*, n.d.). To achieve this goal, the EU has implemented various policies and initiatives.

One of the most important initiatives in this area is the EU Emissions Trading System (EU ETS), which is the world's largest carbon market. It covers more than 11000 power stations and industrial plants in the EU and regulates the emissions of carbon dioxide (CO_2) and other GHGs. The EU ETS works by allocating a limited number of permits to each participating company, which allows them to emit a certain amount of CO_2. Companies that emit less than their allocated amount can sell their unused permits to companies that exceed their allocation. Another important directive is the Renewable Energy Directive (RED), which aims to increase the share of renewable energy in the EU's energy mix. It sets binding targets for each EU member state to increase the share of renewable energy in their final energy consumption by 2030. The RED also establishes sustainability criteria for biofuels and bioliquids, ensuring that only those with a high degree of greenhouse gas emission savings can be counted towards the targets. In addition, the Energy Efficiency Directive (EED) sets binding energy efficiency targets for EU member states, with the aim of reducing energy consumption by 20% by 2020 and 32.5% by 2030. The EED requires member states to establish energy-saving schemes for households, businesses, and public buildings, and to carry out energy audits on large companies. The EU has also adopted the Effort Sharing Regulation (ESR), which sets binding national GHG reduction targets for sectors not covered by the EU ETS, such as transport, buildings, and agriculture. The ESR requires each member state to reduce its GHG emissions by a certain percentage compared to 2005 levels. The targets for 2030 range from 0% to 40%, depending on the country's wealth and emissions level. Finally, the EU has recently proposed the Fit for 55 packages, a set of legislative proposals aimed at achieving the EU's climate neutrality goal. The package includes revisions to existing directives such as the EU ETS, RED, and EED, as well as new initiatives such as the Carbon Border Adjustment Mechanism (CBAM), which aims to ensure that the carbon cost of imports into the EU is equivalent to the cost paid by EU companies under the EU ETS. Overall, the EU's environmental policies aim to reduce GHG emissions, promote renewable energy, increase energy efficiency, and achieve climate neutrality by 2050. The directives discussed above are some of the key tools the EU is using to achieve these goals.

As a member state of the EU, Romania is also subject to these policies and initiatives. In 2019, Romania's greenhouse gas emissions were 181 million tons of CO_2 equivalent, which is a decrease from 2005 levels but still above the country's emissions reduction target. Romania has implemented several measures to reduce its greenhouse gas emissions, including promoting energy efficiency and renewable energy sources, improving public transportation, and reducing methane emissions from landfills. On the other hand, Visegrad Group (Poland, Czech Republic, Slovakia, Hungary) was subject to infringement proceedings by EC for failing to implement emissions reduction targets under the Effort Sharing Regulation. The Czech Republic had first considered that the EU Green Deal is a huge threat to the country. It is considered that Bioeconomy GHG emissions make up 11.99% of GHG emissions in Slovakia, 15.15% in Czechia, 21.69% in Poland and 24.63% in Hungary (Lazorcakova et al., 2022). Slovakia, the Czech Republic and Hungary have already achieved the level of emission reductions required by the Paris Agreement and can contribute to the achievement of the EU target (Tucki et al., 2021). In December 2020, Poland announced its intention to achieve net-zero greenhouse gas emissions by 2050, a significant shift in its previous stance on climate change. Overall, EU greenhouse gas policies and climate change initiatives are closely linked, with the aim of achieving net zero emissions by 2050 and mitigating the impact of climate change. Romania and the Visegrad Group are subject to these policies, but there are varying levels of commitment and progress among the member states. The European Union (EU) has set a goal to become climate neutral by 2050, which means that the net greenhouse gas (GHG) emissions will be reduced to zero. To achieve this goal, the EU has implemented several environmental policies, including European directives, aimed at reducing GHG emissions and promoting renewable energy sources.

The aim of the present study is to emphasize the difference between Romania and V4 in terms of GHG and to support decision management in achieving the zero GHG goal by mid-21st century, considering dual dimension environmental-economic nexus, by using a machine learning (ML) tree-based algorithm framework. Therefore, the framework will be used as a support tool for building a strategy which will target to maximize the efficiency of the decision process by using artificial intelligence (AI) methodology. Also, the study will reveal the main trend in GHG research in association with both area of study and AI methodology used to achieving the main goals.

MATERIALS AND METHODS

Bibliographic analysis

The analysis was performed using R code - Bibliometric and Scientometric Python Library, considering the following dimensions, within WoS databse: GHG Emissions Romania, GHG Emissions Visegrad, Environment - economics nexus, GHG machine learning, GHG economic policies, GHG agriculture modeling, GHG waste modeling, GHG industry modeling, GHG energy modeling, GHG modeling, GHG neural networks, GHG deep learning. Thus, during the 2011-2023 timespan, a number of 93 research articles, 3 book chapters, 1 data paper, 45 proceeding papers and 4 review articles were considered and analysed, counting 571 authors with a collaboration index of 29.45. The repartition of all analysed scientific papers, by year publishing year, is presented in Table 1.

Table 1. Annual scientific production considering the keywords-based, already established, analytical framework

Annual scientific production	
Year	No. of articles
2011	3
2012	6
2013	8
2014	9
2015	7
2016	15
2017	11
2018	11
2019	23
2020	10
2021	18
2022	19
2023	6

Database descriptiton

Data provided by Eurostat, related to a number of 10 indicators, between the years 2010-2022, were taken into consideration for the analytical framework: National expenditure on environmental protection (mil. euros) *NEEP*; Environmental protection investments of total economy (mil. euros) *EPITE*; Total environmental taxes (mil. euros) *TET*; Taxes on Pollution/Resources (mil. euros) *TPR*; Net greenhouse gas emissions (tonnes per capita) - *NGHG*; Greenhouse gases from agriculture (thousand tons) *GHGA*; Greenhouse gases from waste management (thousand tons) *GHGWM*; Greenhouse gases from industrial processes and product use (thousand tons) *GHGIPP*; Greenhouse gases from energy sector (thousand tons) *GHGES*; Research and development expenditure (*% of GDP*). The indicators were divided, firstly, into 2 groups: dependent variable (NGHG) and predictors (the rest of the indicators). Moreover, the predictors' group was divided into 2 dimensions, as follows: economic dimension (NEEP, EPITE, TET, TPR, RDE) and environmental dimension (GHGA, GHGWM, GHGIPP, GHGES). The dependent variable was firstly analyzed considering each dimension parameters, separately, as the independent variables.

Machine learning algorithms

Two machine learning supervised techniques were used to develop the analytical frameworks within the present study, as follows: random forest tree-based algorithms (RF) and generalized additive models (GAM). The RF represents a machine-learning algorithm, consisting of many decision trees, each being a regressor for the input data. Sampled data is used as input, based on a subset of features randomly selected. Therefore, RF involves bagging and random selection of features, generating a certain number of regression trees. The input data is selected through bootstrap sampling, while the features represent a random subset of the original features.

Boosting and bagging are important concepts for random forest models. Boosting represents the combination of several weak learners into one accurate prediction algorithm, while bagging means that only a random subset of samples is independently drawn from the training sample. The randomly selected samples are used to grow weak learners, dealing well with the overfitting situation, the average prediction value being chosen as the final prediction value. Random forest is a very popular algorithm (Liu et al., 2021; Wang et al., 2016) as it is accurate, robust against noise and outliers, fast, and able to perform feature selection. The parameter feature importance is measured by calculating how the overall score decreases when a feature is not available. Thus, the predictor variables will be ranked in terms of relative significance, by calculating the decrease in node impurity weighted by the probability of reaching that node, probability that is calculated by the number of samples that reach the node, divided by the total number of samples. When the value is high, the feature is more important (eq. 1) (*Random Forest Feature Importance Computed in 3 Ways with Python | MLJAR*, n.d.).

$$n_{ij} = w_j C_j - w_{left(k)} C_{left(j)} - w_{right(j)} C_{right(j)} \quad (1)$$

where:

w_j = weighted number of samples reaching node j;

right (j) = child node from right split on node *j*;
n_{ij} = represents the importance of node *j*;
C_j= the impurity value of node *j*;
left(j) = child node from left split on node *j*;
It is possible to calculate the importance for each feature on a decision tree by using the following formula:

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k} \tag{2}$$

where:
fi_i is the importance of feature i;
ni_j = the importance of node j.
The values obtained from the above equation are normalized (*normfij*), obtaining values between 0 and 1 by dividing by the sum of all feature importance values:

$$normfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T} \tag{3}$$

The feature importance is the average over all the trees. The sum of the feature's importance value on each tree is calculated and divided by the total number of trees:

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T} \tag{4}$$

where:
T = total number of trees;
$normfi_{ij}$ = normalized feature importance for *i* in tree *j*;
$RFfi_i$ = the feature *i* importance, calculated from all the trees in the RF model.
The GAM is represented by the sum of different functions specific to each feature, the relationship between the dependent variable and the predictors being different from the simple weighted sum. The Thus, the linear regression coefficients are replaced by spline functions allowing the description of nonlinear relationships. The GAM's advantage is its interpretability, as each predictor's contribution can be clear.

RESULTS AND DISCUSSIONS

Bibliographic analysis

The top journals revealed by the bibliographic analysis that published papers related to the search dimensions presented in the material and methods section of the present research paper are as follows: Energies - MDPI (17 publications), Environmental Engineering and Management Journal, Journal of Environmental Protection and Ecology, Sustainability - MDPI (each with 6 research papers) and Journal of Cleaner production (4 research papers considered). Most relevant keywords are GHG (30 appearance), followed by renewable energy (13 appearance) and climate change - energy (within 10 appearance each) (Figure 1).

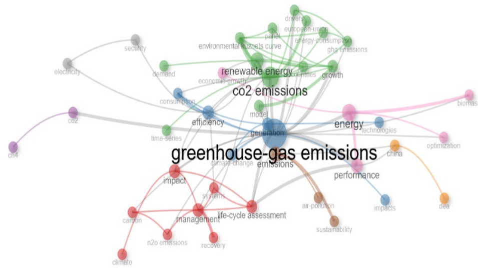

Figure 1. Keywords occurrences network

Considering the keywords conceptual structure (Figure 2) it can be stated that 3 main dimensions are distinguished, as follows: the environmental dimension (blue), the economic dimension (red) and the environmental economics impact dimension (green). However, the red and green dimensions explain almost 40% of the data variance (Figure 2). Therefore, it is necessary to include the 3rd dimension in order to identify a trend within keywords usage in order to elaborate a bibliometric framework which will support the decision process.

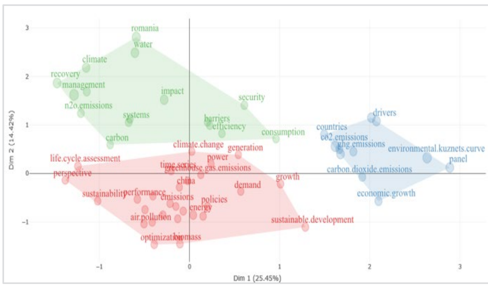

Figure 2. Keyword's conceptual structure map

However, low search results are encountered when AI concepts such as machine learning or autoregression models are included in the search structure. This reveals that there is a gap within the scientific literature related to the use of new, innovative techniques for generating decision-support tools related to GHG emissions for both Romania and V4 group. However, the if analysing the total citations per country, it can

Print ISSN 2285-6064, CD-ROM ISSN 2285-6072, Online ISSN 2393-5138, ISSN-L 2285-6064

be observed that Romania is on the second position, after Italy, with a number of 551 total citations - TC (average article citations - AAC 6.5) while the V4 is mostly represented by Slovakia (4th place) with 80 TC and 16 AAC, followed by Poland (6th place) with 78 TC and AAC 7.71, Hungary (17th place) with TC21 and ACC 7 and the Czech Republic (25th place).

Most cited paper in the area of the study presented in the present article is *EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970-2012* (Janssens-Maenhout et al., 2019) with 250 citations, followed by *Agroforestry creates carbon sinks whilst enhancing the environment in agricultural landscapes in Europe* (Kay et al., 2019) with 79 citations and *Oxy-combustion of coal, lignite and biomass: A techno-economic analysis for a large scale Carbon Capture and Storage (CCS) project in Romania* (Cormos, 2016) with 68 citations.

In terms of countries' collaboration for the elaboration of research articles in the area of interest for the present paper, Romania mostly collaborates with Germany and Austria, while from V4, it can be distinguished 3 clusters as follows: Czech Republic - Slovakia, Hungary (connected with Portugal, Greece and Finland) and Poland (connected with France, Portugal and Germany) (Figure 3).

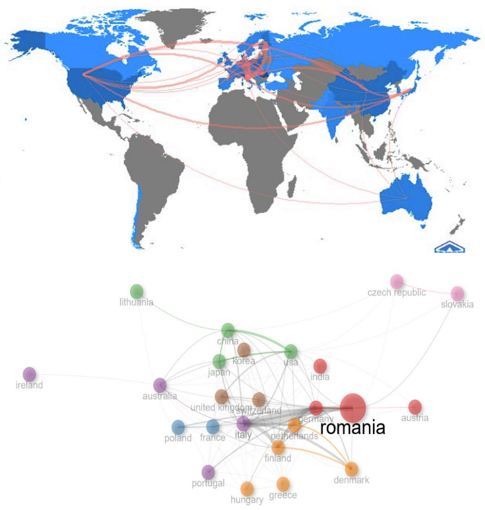

Figure 3. Countries' collaboration in the area of study

GAM modelling framework

The environmental economics dimension generates a prediction of NGHG related to RDE, TET and MEEP (Figure 4A, 4C, 4D) revealing that research and development spendings are directly dependent on GHG emissions, therefore, the increase of emissions generating a similar increase of RDE, TET and MEEP in Romania. Also, in terms of the NGHG-EPITE relation, it can be observed (Figure 4B) that the share of environmental protection investment will increase as the NGHG increases, but with a slower dynamic - however, the EPITE will continue its increase more over the NGHG will decrease to values. The TPR - NGHG relation in Romania has a certain cyclicity as an increase of TPR (between 9 -11 mil. euros) generates positive effects on NGHG, while the variation of TPR outside this interval is promoting high values of NGHG in Romania (Figure 4E).

The Romania environmental dimension emphasizes that the energy sector does not impact the NGHG since an increase of GHGES will generate a continuous decrease of the predicted variable (Figure 5A). However, the situation is different in the case of the waste management sector where an increase in GHGWM generates a direct and proportional increase in NGHG (Figure 5B), a fact that reveals the necessity of applying an optimized waste management strategy in order to increase the sustainability related to GHG emissions in Romania.

The agriculture sector can be considered one of the sectors where *green* measurements can generate an increase of sustainability related to NGHG, revealing its crucial importance, at least in the first stages, in decreasing emissions by acting as a substitute for other industries. However, practising more intensive agriculture (increasing productivity) as well as maximizing production can generate a negative effect on NGHG (Figure 5D).

Industrial processes are also an important source of GHG and can influence NGHG especially if the industry overcomes a maximum sustainable production capacity point (Figure 5C). However, if variated within the optimal range, the industry can be a proper tool for controlling the NGHG in the case of Romania.

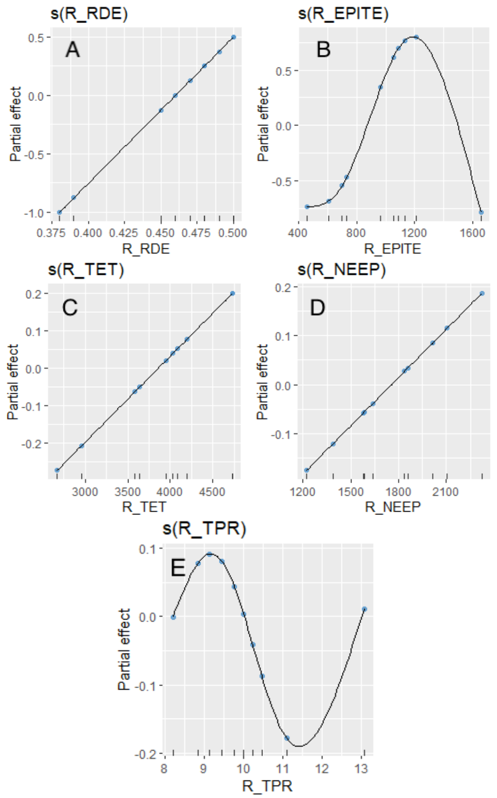

Figure 4. Romania GAM prediction models for the NGHG dependent variable, based on several predictors from environmental economics dimension
(A - RDE; B - EPITE; C - TET; D - NEEP; E - TPR)

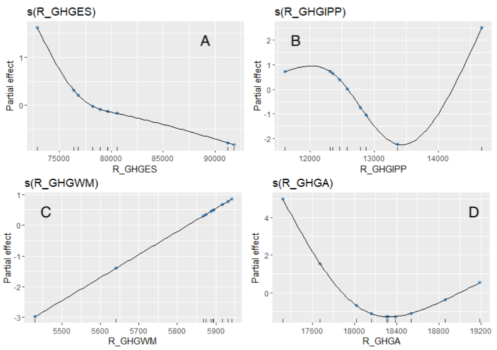

Figure 5. Romania GAM prediction models for the NGHG dependent variable, based on several predictors from the environmental economics dimension
(A - GHGES; B - GHGIPP; C - GHGWM; D - NEEP; E - GHGA)

The environmental economics dimension for V4 generates a prediction of NGHG related to RDE (Figure 6A) and reveals that research and development in the environment can be used in order to decrease NGHG, at least during the first layers of a resilience plan. However, RDE cannot offer a linear relation in association with NGHG and overusing this tool can have a boomerang effect. However, TET and EPITE can be used as efficient tools for decreasing NGHG in V4 (Figure 6B, 6C) since their increase will generate a decrease in the predicted variable.

The NEEP can be an indicator that can be used for monitoring the NGHG in the first steps of the expansion of emissions since it has high sensitivity (Figure 6D), a fact valid also is considering TPR (Figure 6E). However, if the increase in NEEP can be associated with a slower increase in NGHG (Figure 6D), the situation of TPR differs - a decrease in the predicted variable is associated with the increase in TPR.

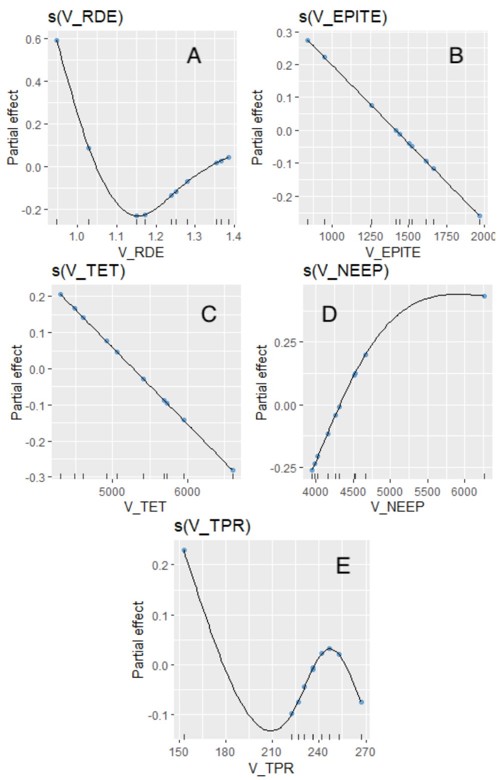

Figure 6. V4 GAM prediction models for the NGHG dependent variable, based on several predictors from environmental economics dimension
(A - RDE; B - EPITE; C - TET; D - NEEP; E - TPR)

The V4 environmental dimension emphasizes that the energy sector and agricultural sector can be used as tools for decreasing the NGHG (Figure 7A, 7D) since an increase of GHG within these two economic sectors can lead to a decrease of NGHG. However, the situation is different in the case of GHGIPP (Figure 7B) since this parameter can be used as a monitoring tool in NGHG dynamics, being directly correlated with the predicted parameter. The waste management sector can be used as a tool for decreasing the NGHG only if GHGWM variates within an optimum interval (5600-5900 thousand tons) (Figure 7C).

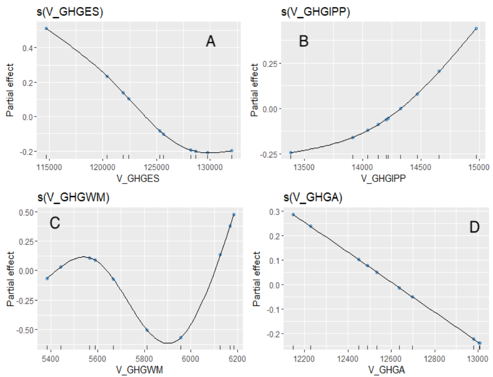

Figure 7. V4 GAM prediction models for the NGHG dependent variable, based on several predictors from the environmental economics dimension (A - GHGES; B - GHGIPP; C - GHGWM; D - NEEP; E - GHGA)

Similar to other studies, also this study is subject to several limitations. The main limitation is related to the dataset time period since it covered a period of 10 years. Thus, although the accuracy of the prediction was high, increasing the time associated with dataset collection will explain better the complex relations between the analytical framework parameters.

CONCLUSIONS

Summarizing the findings, it can be concluded that in terms of integrated GHG emissions management strategy, Romania can merge with V4 to increase the environmental efficiency towards achieving the EU environmental goals. This can be due to similarities between several relations which characterize NGHG within the environmental economics dimension (NGHG - TPR, NGHG - NEEP), but especially within the environmental dimension, where Romania is closely align with the V4 strategy.

Future studies are recommended to be performed in order to widen the number of dimensions involved in the present analytical framework and to perfect the decision support tool based on in-depth observations.

ACKNOWLEDGEMENTS

The present research was supported by the project: "An Integrated System for the Complex Environmental Research and Monitoring in the Danube River Area", REXDAN, SMIS code 127065, co-financed by the European Regional Development Fund through the Competitiveness Operational Programme 2014-2020, contract no. 309/10.07.2021.

The present research was supported by the project "Integrated research and sustainable solutions to protect and restore Lower Danube Basin and coastal Black Sea ecosystems, (ResPonSE) ", 760010/30.12.2022, component C9. PRIVATE SECTOR SUPPORT, RESEARCH, DEVELOPMENT AND INNOVATION, Investment "I5. Establishment and operationalization of Competence Centers".

REFERENCES

Cormos, C.C. (2016). Oxy-combustion of coal, lignite and biomass: A techno-economic analysis for a large scale Carbon Capture and Storage (CCS) project in Romania. *Fuel*, *169*, 50–57. https://doi.org/10.1016/J.FUEL.2015.12.005

EUR-Lex - 52021DC0240 - EN - EUR-Lex. (n.d.). Retrieved October 8, 2022, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2021:240:FIN

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., Van Aardenne, J. A., Monni, S., Doering, U., Roxana Petrescu, A. M., Solazzo, E., & Oreggioni, G. D. (2019). EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970-2012. *Earth System Science Data*, *11*(3), 959–1002. https://doi.org/10.5194/ESSD-11-959-2019

Kay, S., Rega, C., Moreno, G., den Herder, M., Palma, J. H. N., Borek, R., Crous-Duran, J., Freese, D., Giannitsopoulos, M., Graves, A., Jäger, M., Lamersdorf, N., Memedemin, D., Mosquera-Losada, R., Pantera, A., Paracchini, M. L., Paris, P., Roces-Díaz, J. V., Rolo, V., … Herzog, F. (2019). Agroforestry creates carbon sinks whilst enhancing the environment in agricultural landscapes in Europe.

Land Use Policy, *83*, 581–593. https://doi.org/10.1016/J.LANDUSEPOL.2019.02.025

Lazorcakova, E., Dries, L., Peerlings, J., & Pokrivcak, J. (2022). Potential of the bioeconomy in Visegrad countries: An input-output approach. *Biomass and Bioenergy*, *158*, 106366. https://doi.org/10.1016/J.BIOMBIOE.2022.106366

Liu, J., Murshed, M., Chen, F., Shahbaz, M., Kirikkaleli, D., & Khan, Z. (2021). An empirical analysis of the household consumption-induced carbon emissions in China. *Sustainable Production and Consumption*, *26*, 943–957. https://doi.org/10.1016/J.SPC.2021.01.006

Random Forest Feature Importance Computed in 3 Ways with Python | MLJAR. (n.d.). Retrieved July 12, 2023, from https://mljar.com/blog/feature-importance-in-random-forest/

Tucki, K., Krzywonos, M., Orynycz, O., Kupczyk, A., Bączyk, A., & Wielewska, I. (2021). Analysis of the Possibility of Fulfilling the Paris Agreement by the Visegrad Group Countries. *Sustainability 2021, Vol. 13, Page 8826*, *13*(16), 8826. https://doi.org/10.3390/SU13168826

Virtual Workshop - UNFC Europe: Ensuring sustainable raw material management to support the European Green Deal | UNECE. (n.d.). Retrieved July 12, 2023, from https://unece.org/sustainable-energy/events/virtual-workshop-unfc-europe-ensuring-sustainable-raw-material-management

Von Der Leyen, U. (n.d.). *A Union that strives for more My agenda for Europe*.

Wang, Y., Ge, X. ling, Liu, J. liang, & Ding, Z. (2016). Study and analysis of energy consumption and energy-related carbon emission of industrial in Tianjin, China. *Energy Strategy Reviews*, *10*, 18–28. https://doi.org/10.1016/J.ESR.2016.04.002