Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

Assessment of Greenhouse Gas Emissions Using Hotspot Analysis and Machine Learning Methods

Master's Final Degree Project

Mantas Ginkus

Project author

Assoc. prof. dr. Kristina Šutienė

Supervisor

Kaunas, 2026

Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

Assessment of Greenhouse Gas Emissions Using Hotspot Analysis and Machine Learning Methods

Master's Final Degree Project

Data Science and Artificial Intelligence (6211AX013)

Mantas Ginkus

Project author

Assoc. prof. dr. Kristina Šutienė

Supervisor

Assoc. prof. dr. Mindaugas Kavaliauskas

Reviewer

Kaunas, 2026

Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

Mantas Ginkus

Assessment of Greenhouse Gas Emissions Using Hotspot Analysis and Machine Learning Methods

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Mantas Ginkus

Confirmed electronically

Ginkus, Mantas. Assessment of Greenhouse Gas Emissions Using Hotspot Analysis and Machine Learning Methods. Master's Final Degree Project / supervisor dr. Kristina Šutienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Data Science and Artificial Intelligence (Mathematical Sciences).

Keywords: Greenhouse gas emissions, spatial analysis, panel data, Gaussian Process Boosting, machine learning.

Kaunas, 2026. 57 p.

Summary

Greenhouse gas (GHG) emissions vary across countries due to differences in economic activity, energy use, technology, and other persistent country-specific factors. Empirical research on these differences typically relies on country-level panel data. While traditional panel-data regressions are widely used to control for unobserved heterogeneity, they impose restrictive functional forms. More flexible machine learning methods have also been applied to emissions data, but often in pooled settings that do not explicitly account for persistent country-level heterogeneity. At the same time, the emissions literature documents spatial clustering across countries, motivating the analysis of spatial dependence both in the raw data and in model residuals. This thesis analyses cross-country differences in production-based per-capita GHG emissions in Europe using a combination of exploratory spatial analysis and machine-learning methods for panel data. The analysis is based on a balanced panel of 25 European countries observed annually from 2000 to 2023. It addresses two main questions: whether GHG emissions exhibit spatial clustering, and how explicitly modelling persistent country-specific heterogeneity affects predictive performance, residual behaviour, and model interpretation in flexible machine-learning models. Spatial patterns in emissions are analysed using local hot spot analysis and spatio-temporal clustering methods, which identify statistically significant clusters of high and low per-capita emissions and characterise their evolution over time. Predictive modelling is conducted using Gaussian Process Boosting (GPBoost), estimated both in pooled form and with explicit country-specific intercepts and time trends. Models are trained and evaluated using time-respecting expanding-window cross-validation. Beyond average predictive accuracy, residuals are analysed across countries, time, and space to assess whether models capture persistent heterogeneity and spatial structure. Model interpretation is examined using SHAP values, with feature attributions compared across model specifications. The results show that European GHG emissions exhibit spatial clustering that is persistent in some regions and evolving in others. Explicitly accounting for country-level heterogeneity in GPBoost leads to only modest improvements in predictive accuracy, but clear improvements in residual behaviour, including reduced systematic country-level bias, lower within-country residual dispersion, and the removal of spatial structure in prediction errors. Differences in SHAP-based feature attributions demonstrate that the estimated contribution of observed variables depends on whether persistent country-specific heterogeneity is modelled, even when overall predictive performance changes little. These results indicate that unobserved heterogeneity should be explicitly considered when applying machine-learning methods to panel data, particularly for interpretation.

Ginkus, Mantas. Šiltnamio efektą sukeliančių dujų emisijų vertinimas taikant karštųjų taškų analizę ir mašininio mokymosi metodus. Magistro studijų baigiamasis projektas / vadovė dr. Kristina Šutienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Duomenų mokslas ir dirbtinis intelektas (Matematikos mokslai).

Reikšminiai žodžiai: Šiltnamio efektą sukeliančių dujų emisijos, erdvinė analizė, paneliniai duomenys, Gauso procesų stiprinimas (ang. *Gaussian Process Boosting*), mašininis mokymasis.

Kaunas, 2026. 57 p.

Santrauka

Šiltnamio efektą sukeliančių dujų (ŠESD) emisijos tarp šalių skiriasi dėl ekonominės veiklos, energijos vartojimo, technologijų ir kitų ilgalaikių, toms šalims būdingų veiksnių. Šių skirtumų ir juos lemiančių priežasčių analizei dažniausiai naudojami šalių lygmens paneliniai duomenys. Tradiciškai darbui su paneliniais duomenimis taikomi regresijos modeliai, leidžiantys kontroliuoti nepastebėtą heterogeniškumą (angl. *unobserved heterogeneity*), tačiau šie modeliai paprastai remiasi tiesinėmis arba ribotų funkcinių formų priklausomybėmis. Mašininio mokymosi metodai taip pat taikomi emisijų tyrimuose, tačiau dažniausiai jie naudojami kaip jungtiniai (angl. *pooled*) modeliai, kuriuose stebėjimai traktuojami kaip nepriklausomi, o nepastebėtas šalių heterogeniškumas nėra modeliuojamas. Be to, ankstesni tyrimai rodo erdvinį emisijų klasterizavimąsi tarp šalių, o tai leidžia teigti, kad emisijos nėra atsitiktinai pasiskirstę geografinėje erdvėje. Dėl šios priežasties emisijas tikslinga vertinti tiek prieš modelio taikymą, kaip tiriamąjį analizės etapą, tiek ir atlikus modelio apmokymą, siekiant įvertinti, ar prognozių liekanose išlieka erdvinė priklausomybė. Šiame darbe analizuojamos ŠESD emisijos Europoje taikant erdvinę analizę bei mašininio mokymosi metodus, pritaikytus paneliniams duomenims, kurie apima 25 Europos šalių duomenis nuo 2000 iki 2023 metų. Darbe nagrinėjami du pagrindiniai klausimai: ar ŠESD emisijos pasižymi erdviniu grupavimusi ir kaip nepastebėto heterogeniškumo modeliavimas veikia mašininio mokymosi modelio prognozavimo tikslumą, liekanų elgseną ir modelio paaiškinamumą. Erdvinė emisijų klasterių analizė atliekama taikant lokalių karštųjų taškų metodą bei laike kintančių karštųjų taškų metodą. Šie metodai leidžia identifikuoti statistiškai reikšmingus aukštų ir žemų emisijų klasterius bei apibūdinti jų kitimą laike. Prognozavimas atliekamas naudojant kelias GPBoost modelio specifikacijas: be atsitiktinių efektų (angl. *random effects*) ir su atsitiktiniais efektais, skirtais nepastebėtam heterogeniškumui kontroliuoti. Modeliai apmokomi ir vertinami taikant laiko struktūrą išsaugančią kryžminę validaciją. Be bendro prognozavimo tikslumo, atliekama prognozių liekanų analizė siekiant įvertinti, ar nepastebėto heterogeniškumo kontroliavimas daro įtaką prognozavimo kokybei. Modelių paaiškinamumui analizuoti taikoma SHAP analizė. Tyrimo rezultatai rodo, kad ŠESD emisijos Europoje pasižymi erdviniu grupavimusi į klasterius. GPBoost modelių specifikacijų palyginimas rodo, kad nors bendras prognozavimo tikslumas tarp modelių skiriasi nedaug, specifikacijos su įtrauktais atsitiktiniais efektais pagerina prognozių liekanų elgseną: sumažėja liekanų dispersija, prognozės tampa tolygesnės ir tikslesnės tarp šalių, be to, tarp liekanų nėra erdvinės priklausomybės. SHAP analizės rezultatai taip pat reikšmingai skiriasi tarp GPBoost modelio specifikacijų ir priklauso nuo to, ar modelyje kontroliuojamas nepastebėtas heterogeniškumas. Šie rezultatai rodo, kad taikant mašininio mokymosi metodus paneliniams duomenims yra svarbu atsižvelgti į nepastebėtą heterogeniškumą, ypač kai modeliai naudojami interpretacijai.

Table of contents

List of figures **8**
List of tables **9**
Introduction **10**
1. Literature review **11**
1.1. Drivers and frameworks of GHG emissions 11
1.1.1. Greenhouse gases 11
1.1.2. Drivers of GHG emissions 11
1.1.3. Conceptual frameworks for analysing emissions drivers 12
1.1.4. Summary of frameworks 14
1.2. Spatial analysis of emissions 14
1.2.1. Global spatial autocorrelation 15
1.2.2. Local hot spot analysis 15
1.2.3. Spatial weights and neighbourhood definition 16
1.2.4. Emerging hot spot analysis 16
1.2.5. Spatial clustering of emissions in the literature 17
1.3. Panel-data modelling of emissions 18
1.3.1. Traditional panel-data approaches in emissions research 18
1.3.2. Machine learning: prediction versus causal inference 19
1.3.3. Machine learning panel-data approaches in emissions research 19
1.3.4. Gaussian Process Boosting (GPBoost) 20
1.3.5. SHAP values 20
1.4. Justification of the research topic and tasks of the final project 22
2. Methodology **23**
2.1. Data 23
2.1.1. Energy mix variables 25
2.1.2. Economic structure 25
2.2. Spatial hot spot and emerging hot spot analysis 26
2.2.1. Software and implementation 26
2.2.2. Spatial weights matrix 26
2.2.3. Local Getis-Ord Statistics 26
2.2.4. Permutation-based inference 26
2.2.5. Hot and cold spot identification 27
2.2.6. Relationship between $\boldsymbol{Gi}$ and $\boldsymbol{Gi*}$ when calculating permutation-based $\boldsymbol{z}$ and $\boldsymbol{p}$ values 27
2.2.7. Emerging Hot Spot Analysis (EHSA) 28
2.3. GPBoost 28
2.3.1. Software and implementation 28
2.3.2. General model structure 28
2.3.3. GPBoost specifications 29
2.3.4. Linear benchmark models 29
2.3.5. Model training and assessment 30
2.4. Residual analysis 31
2.4.1. Residual definition 31
2.4.2. Residual global spatial autocorrelation 31

2.4.3. Residual local spatial clustering 32
2.4.4. Time-specific median residuals 32
2.4.5. Cross-country residual dispersion 32
2.4.6. Country-level bias 32
2.4.7. Within-country residual variance 32
2.5. Variance decomposition of features 33
3. Results 34
3.1. GHG emission trends across European countries 34
3.2. Local hot spot and emerging hot spot analysis 35
3.2.1. Local hot and cold spots of GHG emissions 35
3.2.2. Temporal trends in permutation-based ***z*** values 36
3.2.3. Temporal patterns of GHG emissions hot and cold spots 37
3.3. GPBoost 38
3.3.1. Performance accuracy (RMSE) 38
3.3.2. Residual analysis 40
3.3.3. Explainability and feature attribution 45
List of references 54
Appendices 58

List of figures

Fig. 1. GHG emissions by country (2000-2023) 34
Fig. 2. Local hot and cold spots of GHG emissions in Europe 35
Fig. 3. Statistically significant trends in permutation-based Getis-Ord $Gi * z$ values across Europe 37
Fig. 4. Emerging Hot Spot Analysis (EHSA) results for GHG emissions across Europe 38
Fig. 5. Residuals for model M1 (left) and model M2 (right) are shown for the validation years 41
Fig. 6. Median residual across countries for each year in the validation sample 43
Fig. 7. Dispersion of residuals across countries for each year in the validation sample 43
Fig. 8. Distribution of within-country residual variance by model specification 45
Fig. 9. SHAP-based feature importance for GPBoost models M1 (left) and M2 (right) 47
Fig. 10. Local SHAP explanations for Germany (top) and Spain (bottom) in 2022 48
Fig. 11. SHAP bee swarm summary plot for GPBoost model M2 49
Fig. 12. SHAP dependence plots for energy consumption per capita (left) and energy intensity (right) in GPBoost model M2, with observations coloured by GDP per capita 50
Fig. 13. SHAP dependence plots 50

List of tables

Table 1. Main drivers of GHG emissions identified in the literature 11
Table 2. Variables used in the analysis 23
Table 3. Cross-validation RMSE across three folds for each model 39
Table 4. Test sample RMSE for all model specifications 40
Table 5. Moran's I values and permutation-based p values for residual spatial autocorrelation 42
Table 6. Most over-predicted and under-predicted countries based on average validation residuals 44
Table 7. Median absolute country-level residuals across the validation period 44
Table 8. Between-country variance share of observed features 46

Introduction

Reducing greenhouse gas (GHG) emissions is a key goal of climate policy, as emissions from human activities are the main cause of global warming [1]. GHG emissions vary across spatial units and over time due to differences in economic activity, energy use, technology, and other factors. To study this variation and the factors associated with it, researchers commonly rely on panel data.

Much of the empirical literature uses regression-based panel data methods, which are used to account for unobserved, time-invariant unit-specific factors. More flexible approaches, including machine learning methods, have also been applied to model complex and potentially nonlinear relationships between emissions and observable variables. However, when such models are estimated in a pooled manner without explicitly accounting for unit-specific heterogeneity, estimated relationships and predictive performance may be distorted if unit effects are correlated with regressors [2]. In cross-country emissions data, this heterogeneity may reflect long-lasting differences between countries that are not captured by the observed variables, which can affect both model predictions and interpretation.

The goal of this thesis is to investigate cross-country differences in European GHG emissions using flexible machine learning models applied to panel data and exploratory spatial and spatio-temporal cluster analysis, and to assess how accounting for country-level heterogeneity affects predictive performance and SHAP-based interpretation.

This goal is addressed through the following objectives:

1. Describe spatial and temporal patterns in emissions to identify geographic clustering and its evolution over time.
2. Estimate and compare the predictive performance of linear and GPBoost model specifications, each estimated with and without explicit country-specific effects, using time-respecting (expanding window) validation.
3. Assess residual behaviour across space, countries, and time to evaluate whether fitted models capture spatial patterns and persistent country-specific heterogeneity beyond average predictive accuracy.
4. Analyse SHAP-based feature attribution in different GPBoost specifications to assess its sensitivity to the inclusion of explicit country-specific effects.

The contribution of this thesis is both methodological and applied. Methodologically, it addresses a common limitation in the use of machine learning methods for panel data, where observations are often treated in a pooled manner without explicitly modelling unit-specific heterogeneity. It does so by evaluating how accounting for country-specific effects influences predictive performance, residual behaviour, and SHAP-based feature attribution. The applied contribution lies in demonstrating these differences in a cross-country analysis of European GHG emissions.

1. Literature review

This literature review covers three areas relevant to the analysis of production-based per-capita greenhouse gas (GHG) emissions across countries. The first section reviews how emissions and their main drivers are conceptualised in the literature. The second section reviews spatial methods used to describe geographic clustering and its evolution over time. The third section reviews panel-data modelling approaches applied to emissions data, including machine learning methods.

1.1. Drivers and frameworks of GHG emissions

This section reviews how GHG emissions are defined in the literature and summarises the main drivers commonly associated with emissions. It also introduces the conceptual and analytical frameworks used to organise these drivers in empirical research.

1.1.1. Greenhouse gases

The main greenhouse gases (GHGs) such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) are closely linked to climate change. GHGs absorb and re-emit infrared radiation, which slows the escape of heat from the Earth's surface into outer space and leads to warming of the climate [3]. Since the industrial era, human activities have strongly increased atmospheric concentrations of GHGs. Limiting climate change therefore requires deep reductions in GHG emissions [4].

In this thesis, production-based GHG emissions per capita are used. These emissions are measured in tonnes of carbon dioxide equivalent (CO_2-eq) per capita. Following Eurostat's definition, total GHG emissions include CO_2, CH_4, N_2O, and fluorinated gases (hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), nitrogen trifluoride (NF_3), and sulphur hexafluoride (SF_6)). Total GHG emissions can be reported both including and excluding emissions and removals from land use, land-use change, and forestry (LULUCF). In this thesis, GHG emissions exclude LULUCF, because its inclusion can mask production-based emissions by accounting for carbon sinks (e.g., forests) which can lower total reported emissions [5].

1.1.2. Drivers of GHG emissions

The drivers of total GHG emissions have been discussed in the literature for several decades. Most studies view emissions as the result of multiple interacting human factors. Rosa and Dietz [23] identify five broad candidate drivers that recur throughout the literature: population, affluence and consumption, technology, institutional arrangements, and culture. These drivers are not independent. They act together and cannot be meaningfully analysed in isolation. Table 1 summarises the main drivers of GHG emissions and briefly describes their role.

Table 1. Main drivers of GHG emissions identified in the literature

Driver	Description
Population	Population size determines the overall scale of human activity. A larger population is generally associated with higher total emissions. However, population size alone does not explain emissions levels. Environmental impacts depend on how people consume and how goods and services are produced. In addition, there are other population-related characteristics that matter for emissions beyond population size alone. Factors such as the number of households, age structure, and urbanisation can influence energy use and consumption patterns in ways not captured by population size alone.

Driver	Description
Affluence and consumption	Higher income levels are associated with higher levels of consumption and production, which typically result in greater energy use and higher emissions. Rosa and Dietz emphasise that consumption is mainly driven by affluence, although cultural factors may influence the composition of consumption.
Technology	Technology determines how consumption and production translate into emissions. Emissions depend on the technologies used in energy generation and in the production of goods and services. Improvements in energy efficiency or shifts towards lower-carbon energy sources can reduce emissions for a given level of consumption.
Institutions and culture	Institutional and cultural factors can influence emissions by shaping consumption patterns, policy and technology choices. However, Rosa and Dietz note that these factors have been studied less systematically.

The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report similarly highlights the importance of these drivers. It notes that population growth and increases in GDP per capita have been major contributors to emissions growth, while improvements in energy intensity and changes in energy systems have partly offset these effects [1].

Because these drivers are closely related and often change together, many studies use simplified analytical frameworks that group them into a small number of aggregate components. The Kaya identity and the related IPAT and STIRPAT frameworks provide such representations and are discussed in the following subsections.

1.1.3. Conceptual frameworks for analysing emissions drivers

1.1.3.1. Kaya Identity

A commonly used framework for describing the main components associated with energy related CO_2 emissions is the Kaya identity, originally introduced by Yoichi Kaya in 1990. The Kaya identity expresses energy-related CO_2 emissions as the product of four factors: population, economic output per capita, energy intensity, and carbon intensity [24]. It can be written as:

$$CO_2 = P \cdot \frac{G}{P} \cdot \frac{E}{G} \cdot \frac{CO_2}{E} \tag{1.1}$$

where P is population, G is gross domestic product (GDP), and E is the total energy consumption. The ratios G/P, E/G, and CO_2/E correspond to GDP per capita, energy intensity (energy per unit GDP), and carbon intensity (emissions per unit energy) [24].

For example, the IPCC frequently uses the Kaya identity to decompose past CO_2 emissions growth into contributions from population, economic activity, energy intensity, and carbon intensity. However, these contributions represent accounting attributions rather than causal effects [25]. The Kaya identity is therefore useful for organizing and summarizing emissions trends, but it is limited in explaining why emissions change or in assessing the effects of policy interventions.

1.1.3.2. IPAT

The IPAT identity, introduced by Ehrlich and Holdren, is a widely recognized formula for analysing the effects of human activities on the environment. It expresses impact (I) as the product of population (P), affluence (A) and technology (T) [26, 27].

$$I = P \cdot A \cdot T \tag{1.2}$$

In this identity, affluence is defined as per capita consumption or production, typically modelled as GDP per capita, while technology is defined as the environmental impact per unit of consumption or GDP [26, 27]. The technology term is not directly observable and captures other factors that affect environmental impact (e.g., sectoral composition, institutions, or production technologies) [26].

Ehrlich and Holdren used IPAT to emphasise that environmental impact arises from the joint, multiplicative interaction of all three factors [27]. They argued that environmentally harmful technologies (T) generate far greater absolute impacts in large (P), affluent (A) societies than in smaller, low-consumption ones. Importantly, Ehrlich and Holdren explicitly warned that IPAT is a highly simplified identity and emphasised that, in reality, the variables P, A, and T are not independent but instead interact and influence one another (e.g., changes in P may cause changes in A or T, and vice versa) [26, 27].

1.1.3.3. STIRPAT

The primary limitation of the IPAT (beyond the assumption of independence) stems from its nature as a mathematical identity rather than a statistical model. Since the value of any one term is fully determined by the values of the others (e.g., given I, P and A, the value of T is fixed), IPAT does not allow hypothesis testing or statistical inference. As a result, it cannot be used to empirically assess the relative importance of population, affluence, or technology as drivers of environmental impact [28].

To address this limitation, Dietz and Rosa [29] reformulated IPAT as a stochastic model known as STIRPAT (Stochastic Impacts by Regression on Population, Affluence, and Technology). Unlike IPAT, STIRPAT is not an accounting equation but an empirically estimable model. Its general form is:

$$I = aP^b A^c T^d e \tag{1.3}$$

where a is a scaling constant, b, c, and d are parameters to be estimated, and e is a stochastic error term capturing unobserved influences. Taking logarithms yields a linear regression model:

$$\ln(I) = \ln(a) + bln(P) + cln(A) + dln(T) + \ln(e) \tag{1.4}$$

This model allows the use of standard statistical techniques such as regression analysis to test hypotheses about the magnitude and significance of associations between population, affluence, technology, and environmental impact [29].

1.1.3.4. Decomposition Methods

The purpose of decomposition analysis is to quantify how changes in pre-defined factors contribute to changes in a selected aggregate outcome, such as total energy consumption or emissions. By

separating observed changes into interpretable components, decomposition analysis provides insight into the underlying sources of change [30].

Index decomposition analysis (IDA) was introduced in the late 1970s to study how changes in production mix contributed to industrial energy demand [30]. With rising concerns about global warming and environmental sustainability, IDA was increasingly applied to the decomposition of carbon dioxide and other pollutant emissions [31]. IDA represents a family of decomposition methods rather than a single technique. These methods share a common objective of attributing changes in an aggregate indicator to changes in its contributing factors, but they differ in how this attribution is performed [32].

Early or traditional IDA methods suffer from several limitations. First, many methods result in imperfect decomposition, leaving behind an unexplained residual term. Second, some IDA methods have difficulty handling zero values in the data. Third, results from traditional IDA methods may be sensitive to the choice of base year [30]. To address these limitations, the Logarithmic Mean Divisia Index (LMDI) method was developed as an improved implementation within the IDA framework. LMDI provides perfect decomposition, which means that the total change in aggregate variables is fully explained without residuals. In addition, LMDI can handle zero values in the data and is not sensitive to the choice of base year [30]. Because of these advantages, LMDI has become the dominant and preferred IDA method in empirical studies of energy consumption and environmental emissions [31].

1.1.4. Summary of frameworks

The frameworks reviewed in this section play different roles in the emissions literature. The IPAT and Kaya identities are accounting-based frameworks that express emissions as a function of a small number of components [24, 26]. They are commonly used to describe and organise what are considered the main drivers or components of emissions.

Decomposition methods such as the LMDI are closely related to these identities. They use the same breakdown of emissions to show how changes over time can be attributed to individual components [28, 31].

The STIRPAT framework is linked to these approaches in that it is based on the same representation of emissions drivers, but it expresses this relationship in a form that can be estimated from observed data [28].

Reviewing these frameworks helps clarify how the emissions literature groups and discusses the main drivers of emissions. It also provides a shared background for understanding the empirical modelling approaches discussed later.

1.2. Spatial analysis of emissions

This section reviews spatial methods used to describe the geographic and spatiotemporal distribution of GHG emissions. The focus is on exploratory techniques that identify spatial clustering and its evolution over time.

1.2.1. Global spatial autocorrelation

When analysing GHG emissions across European countries, it is informative to examine whether countries with similar emission levels tend to be located near each other in geographic space, rather than considering countries only in isolation. If neighbouring countries exhibit similar levels of GHG emissions, this indicates the presence of spatial clustering. If no such similarity exists, emissions can be considered spatially uncorrelated.

The idea that nearby locations tend to be more similar than distant ones is a basic principle of spatial analysis and is often summarised by Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" [6]. Spatial autocorrelation refers to the association between values of a variable across adjacent or nearby spatial locations. It shows whether similar or dissimilar values tend to cluster in space [7].

Spatial autocorrelation can be assessed using global summary measures, such as Moran's I. However, these measures aggregate spatial relationships across the entire study area and do not indicate where clustering occurs. In such cases, positive clustering in some areas and opposing spatial patterns in others can cancel out, producing weak or misleading global measures of spatial dependence [8].

For this reason, global spatial autocorrelation measures are best viewed as descriptive diagnostics. To identify where clusters of high and low values occur, local spatial methods are required. These local hot spot methods are discussed in the next subsection.

1.2.2. Local hot spot analysis

To examine local spatial patterns, Getis and Ord [9] introduced a family of local statistics designed to identify clusters of high or low values.

The commonly used statistic from this family is the Getis-Ord G_i^* statistic. Unlike the related G_i statistic, G_i^* includes the value at the focal location itself, in addition to its neighbours [9]. It measures whether high or low values are concentrated at a location and in its surrounding neighbourhood relative to the rest of the study region. Formally, G_i^* is defined as

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j} \tag{1.5}$$

where x_j denotes the value of the variable of interest at location j, and w_{ij} are elements of a spatial weights matrix defining relationships between countries. The statistic is computed separately for each location [9].

High values of G_i^* indicate local clustering of high variable values (hot spots), while low values indicate local clustering of low variable values (cold spots) [9]. For G_i^*, the null hypothesis is that values are randomly distributed across space. This hypothesis is commonly assessed using standardized z scores and corresponding p values derived from a normal distribution. Large positive z scores indicate statistically significant hot spots (for example, $|z| > 1.96$ at the 0.05 significance level), while large negative values indicate cold spots. However, Ord and Getis [7] show that the normal approximation can be unreliable in small samples or when the data are skewed (because the z scores rely on an assumed normal distribution), so later work commonly uses permutation-based methods to assess significance [8].

In this study, because GHG emissions are measured per capita on a production basis, the emission values describe emissions generated within each country rather than atmospheric concentrations. Consequently, identified hot and cold spots reflect similarities between countries in their levels of greenhouse gas production, rather than physical spillover of emissions across national borders.

1.2.3. Spatial weights and neighbourhood definition

Spatial weights matrices are a key element of spatial analysis. They define how spatial units are related to one another and are required to measure spatial autocorrelation and identify spatial clusters using statistics such as Moran's I and the Getis-Ord G_i^* statistic [10, 11]. Because the true structure of spatial dependence is rarely known with certainty (for example, whether spatial dependence operates only through shared borders or extends beyond neighbours through distance), the choice of the spatial weights matrix is considered an important decision in spatial analysis [12].

Several types of spatial weights matrices have been proposed in the literature. Common approaches include contiguity-based weights, distance-based weights, and k-nearest neighbour weights [10]. In contiguity-based weight matrices, neighbouring units share a common border and are assigned a weight of one, while all other pairs receive a weight of zero (binary contiguity). Distance-based weight matrices assign larger weights to nearby units and smaller weights to more distant units. k-nearest neighbour weight matrices connect each unit to a fixed number (k) of closest neighbours. These approaches differ in how neighbourhood relationships are defined, and no single specification is appropriate in all cases.

Some authors note that when strong theoretical guidance is lacking, simpler spatial weights structures are often recommended in exploratory spatial analysis. Griffith [12] recommends low-order spatial models, and Florax and Rey [13] further show that including too many neighbours can reduce the statistical power of spatial tests by introducing irrelevant spatial links.

Spatial weights matrices are often standardized to make results easier to interpret. A common approach is row standardization, where the weights in each row sum to one, so that each spatial unit has the same total influence. In this case, the contribution of neighbouring values represents an average rather than a sum. However, Tiefelsdorf, Griffith, and Boots [14] note that row standardization can place greater weight on neighbours of units with few spatial links, such as units located at the edge of the study region. An alternative is to use non-standardized binary weights, where influence accumulates with the number of neighbours. Cliff and Ord [15] show that the choice of standardization affects how spatial statistics are interpreted rather than whether spatial autocorrelation is detected.

Overall, the specification of the spatial weights matrix should be consistent with the research objective, the meaning of the variable being analysed (for example, whether values represent outcomes measured within spatial units (such as national emissions) rather than flows or spillovers between units), and the size and shape of the spatial units.

1.2.4. Emerging hot spot analysis

Traditional hot and cold spot maps describe spatial clustering at a single point in time. Such information is useful for identifying where high or low values are clustered, but it does not show whether these clusters change over time. As a result, static hot and cold spot analyses are limited

because they cannot distinguish between clusters that are stable and those that are newly emerging, intensifying, or declining [16].

Emerging hot spot analysis (EHSA) was developed to address this limitation by combining information on spatial clustering with temporal trends. In practice, local spatial clustering is evaluated separately for each time step, typically using the Getis-Ord G_i^* statistic. The resulting sequence of local G_i^* z-scores is then examined over time, most commonly using a Mann-Kendall trend test. This allows locations to be classified according to their temporal behaviour into categories such as *new*, *persistent*, *intensifying*, *diminishing*, or *sporadic* hot and cold spots [16].

EHSA was formalized through its implementation in ArcGIS, where it is applied using a space-time cube structure [16]. This implementation made EHSA straightforward to apply and contributed to its widespread use in fields such as environmental research, public health, and urban and regional analysis. Although EHSA is widely associated with ArcGIS and many studies rely on this implementation, it is a general analytical approach that can be implemented in other environments when its core steps and classification rules are clearly specified.

1.2.5. Spatial clustering of emissions in the literature

Many empirical studies use spatial clustering methods to analyse spatial patterns of emissions and spatial dependence between neighbouring units. Most of these studies focus on carbon dioxide (CO_2) rather than total GHG emissions and rely on production-based CO_2 data. At the national scale, spatial clustering is analysed using country-level CO_2 emissions. Rios and Gianmoena [17] analyse per-capita CO_2 emissions for 141 countries using global and local spatial diagnostics. They show that countries with similar CO_2 emission levels are geographically clustered and that convergence occurs within these spatial clusters rather than at the global level. Atalay and Akan [18] also analyse national-level clustering using Moran's I and Local Indicators of Spatial Association (LISA) to identify geographically contiguous clusters of OECD countries with high and low emission levels. Their analysis shows that high-emission countries tend to be located near other high-emission countries, while low-emission countries tend to be located near other low-emission countries. Focusing on Europe, Wang et al. [19] extend descriptive spatial analysis by combining Moran's I with spatial econometric models. They show that CO_2 emissions in one European country are statistically associated with the economic and energy characteristics of neighbouring countries, which they interpret as evidence of spatial dependence between countries.

At sub-national and regional scales, spatial clustering methods are used to analyse the spatial concentration of emissions across regions within countries. Wang et al. [20] apply exploratory spatial data analysis to CO_2 emissions across 30 Chinese provinces and identify high-emission clusters concentrated in eastern coastal provinces using LISA statistics. They show that spatial autocorrelation in emissions across provinces persists over time and associate these spatial patterns with structural characteristics such as industrial composition, energy use, and levels of economic development. In the European context, Vagnini et al. [21] analyse industrial CO_2 emissions across 238 NUTS-2 regions and identify spatial clusters of regions with similar emission trajectories. Some of these clusters extend across national borders and include regions with comparable industrial structures.

Spatial clustering analysis has also been extended to incorporate temporal dynamics. Xia et al. [22] construct a high-resolution (1 km) CO_2 emissions dataset for East Asia and apply a spatiotemporal

emerging hot spot analysis based on the Getis-Ord G_i^* statistic to identify persistent, intensifying, emerging, and diminishing emission clusters over time.

Overall, the literature analyses CO_2 emissions across spatial scales ranging from countries to regions and commonly reports spatial clustering of emissions.

1.3. Panel-data modelling of emissions

This section reviews empirical modelling approaches used in the emissions literature. It covers traditional panel-data models, the use of machine-learning methods in both predictive and causal settings, and introduces GPBoost, a method that explicitly accounts for unit-specific heterogeneity in panel data.

1.3.1. Traditional panel-data approaches in emissions research

Many empirical studies of cross-country emissions use panel data, which consist of repeated observations on the same units over time [39]. To account for unobserved, time-invariant country characteristics that influence emissions, these studies commonly rely on fixed-effects or random-effects estimators, often combined with time effects to capture common shocks.

Early cross-country studies often focused on the relationship between emissions and economic development, most prominently through the Environmental Kuznets Curve (EKC) framework. EKC studies examine whether emissions follow an inverted-U relationship with income by regressing emissions on income and its polynomial terms [40]:

$$GHG_{it} = \beta_1 Y_{it} + \beta_2 Y_{it}^2 + e_{it} \tag{1.6}$$

where GHG_{it} denotes emissions in country i at time t, and Y_{it} is the income per capita. Stern [40] reviewed a large number of EKC studies and showed that EKC findings are not robust, as they depend strongly on model specification choices, and should not be interpreted as evidence that income growth by itself reduces emissions.

Subsequent studies moved beyond single-factor frameworks and adopted multi-factor models that include a broader set of emissions drivers. For example, many such studies use STIRPAT-type regressions estimated with country and time effects [28, 41]. These models focus on identifying associations between emissions and multiple drivers, but they typically assume linear or log-linear relationships between variables.

Some studies also use dynamic panel models to account for persistence in emissions over time. These models typically include lagged emissions as an explanatory variable:

$$GHG_{it} = \rho GHG_{i,t-1} + \beta X_{it} + \alpha_i + e_{it} \tag{1.7}$$

where coefficient ρ measures the dependence of current emissions on past emissions. While dynamic panel models capture adjustment dynamics over time, they usually include a limited number of explanatory variables and do not aim to model complex interactions among drivers [42, 43].

Overall, traditional panel econometric approaches provide interpretable models and explicitly account for unobserved heterogeneity and, in some cases, temporal dependence. However, they

require nonlinearities and interaction effects to be specified explicitly and therefore have limited ability to capture complex nonlinear relationships and high-dimensional interactions.

1.3.2. Machine learning: prediction versus causal inference

Machine Learning (ML) methods are used in empirical research for two related but distinct purposes: causal inference and prediction. In causal applications, the goal is to estimate the effect of an intervention, which requires explicit assumptions about the underlying causal structure and about which variables confound the relationship between the intervention and the outcome [33]. ML methods are then used to adjust for these confounders, but whether an estimated effect can be interpreted as causal depends on the validity of these assumptions rather than on the predictive accuracy of the model itself [34, 33].

Predictive machine learning serves a different purpose. Its aim is to produce accurate predictions for new or unseen data by learning patterns in the observed data. As emphasized by Shmueli [35] and Mullainathan and Spiess [36], predictive models are not designed to estimate causal effects, but to generalise well out of sample. Model inputs are therefore interpreted in terms of their contribution to prediction rather than as estimates of causal effects. This distinction is important for model explanation: feature-attribution methods such as SHAP explain how a fitted predictive model uses its inputs to generate predictions, but they do not estimate causal effects and should not be interpreted as evidence about real-world mechanisms [35, 37].

This thesis focuses on prediction rather than causal inference. Accordingly, SHAP values are used to interpret predictive patterns in the fitted models rather than to draw causal conclusions.

1.3.3. Machine learning panel-data approaches in emissions research

Machine learning (ML) methods have attracted increasing interest in emissions research because they can model complex and nonlinear relationships. This is particularly useful when emissions depend on many interacting variables whose functional form is difficult to specify in advance.

In the emissions literature, ML models are often applied to panel data covering countries, regions, or cities over time. In practice, however, these data are frequently treated in pooled form, with each unit-year observation used as an independent data point. For example, Costantini et al. [44] pool observations for 117 countries over multiple years and train a Random Forest model to predict per-capita CO_2 emissions. Similar pooled applications are common in city and region-level emissions studies [45-47].

Many studies rely on tree-based ensemble models (for example, Random Forest, XGBoost, and Extremely Randomized Trees), which are well suited for handling large numbers of predictors and capturing nonlinear relationships. These models are typically used to predict emissions or emissions-related outcomes and to assess the relative importance of explanatory variables. For instance, Yu et al. [47] apply Extremely Randomized Trees and other tree-based ensemble models to city-level data from China to predict and analyse the determinants of carbon emissions, while Hou and Liu [46] apply a Bayesian-optimised XGBoost ensemble to panel data from China's coastal provinces to predict construction-sector carbon emissions.

Model performance is usually evaluated using standard predictive accuracy measures, including R^2, RMSE, MAE, or MAPE. Evaluation is typically based on random train-test splits or k-fold cross-

validation applied to pooled observations. In many studies, training and testing samples are created using random splits that do not preserve time order or do not hold out entire units.

Several emissions studies also employ post-estimation tools to interpret ML models and summarise variable importance. Common approaches include SHAP values and partial dependence plots. Li et al. [45] use XGBoost combined with SHAP values to rank the contributions of economic, demographic, and infrastructure-related factors to carbon emissions in China's first-tier cities, while Dong et al. [48] apply gradient boosting models together with SHAP and partial dependence plots to analyse the relative importance, nonlinear effects, and interactions of key drivers of carbon emission intensity across Chinese cities.

When ML methods are used for causal inference in emissions research, Double Machine Learning (DML) is commonly used. In these applications, causal machine learning techniques are typically used to evaluate the effects of clearly defined policy interventions, such as environmental regulations or financial incentives [49, 50].

Overall, the literature suggests that ML is primarily used in emissions research as a predictive and exploratory tool. Pooled estimation, standard predictive validation, and post-estimation interpretability tools are common, while panel data features such as unit-specific heterogeneity and temporal dependence are usually not modelled explicitly.

1.3.4. Gaussian Process Boosting (GPBoost)

Gaussian Process Boosting (GPBoost) is a ML framework designed for prediction tasks with dependent observations, such as panel data. It combines gradient-boosted decision trees with Gaussian random effects and, where relevant, Gaussian processes to explicitly model dependence between observations arising from shared unobserved unit-specific characteristics or spatial or temporal correlation [51].

GPBoost is motivated by data structures in which repeated observations of the same units or spatial proximity introduce correlation that can lead to overfitting in standard boosting models. By modelling these dependencies explicitly, GPBoost aims to improve predictive performance and model stability while retaining the flexibility of nonlinear tree-based models [51].

Unlike traditional panel-data econometric models, which are typically specified for parameter estimation and interpretation, GPBoost is designed primarily for prediction. Its focus is on accurately capturing complex relationships in the data while accounting for dependence between observations, rather than on estimating interpretable regression coefficients.

GPBoost has recently been applied to prediction problems with spatial and longitudinal dependence, illustrating how dependence can be modelled explicitly within flexible machine-learning frameworks [54, 55].

1.3.5. SHAP values

ML models such as gradient-boosted decision trees often provide strong predictive performance but are difficult to interpret. To address this issue, post-hoc explanation methods are used to describe how a trained model produces its predictions. One widely used approach is SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee in 2017 [56].

SHAP explains individual predictions by assigning each feature a contribution value that shows how much it increases or decreases the model's prediction relative to a baseline. For tree-based models, such as random forests and gradient-boosted decision trees, Lundberg et al. [53] propose an efficient TreeSHAP method. By aggregating SHAP values across observations, it is possible to summarise global patterns in the fitted model, such as which variables contribute most to predictions and whether relationships appear nonlinear.

While SHAP is useful for explaining how a fitted model makes its predictions, there are several limitations. First, SHAP values describe how the model uses its inputs, not how the outcome is generated in the real world [52, 38]. Second, SHAP explanations depend on the quality of the fitted model. Because SHAP is constructed to be faithful to the model, it will also reflect model errors. For example, if the model overfits the training data, then SHAP will explain overfitting and not patterns that generalize to new data [52, 53]. Third, SHAP can be sensitive to dependence between features, which can complicate interpretation when variables are highly correlated or act as proxies for one another [52].

1.4. Justification of the research topic and tasks of the final project

The literature review shows that emissions are commonly analysed using regression-based panel-data models (such as STIRPAT-type specifications), which are typically estimated with fixed-effects or random-effects to account for unobserved country-specific heterogeneity. While these approaches explicitly control for unobserved, time-invariant country characteristics, they typically impose linear or log-linear functional forms. More flexible machine-learning methods have also been applied in panel-data settings, but they are usually estimated in a pooled manner. In such models, persistent country-specific heterogeneity is not modelled explicitly and can only be absorbed indirectly through observed features. At the same time, the literature on CO_2 emissions consistently shows spatial clustering across countries, indicating that emission levels may not be distributed randomly in geographic space but instead display regional patterns. In addition, the literature identifies a set of core driver groups of emissions, which provide a basis for selecting explanatory variables in empirical analysis.

Based on these findings, the tasks of this project combine exploratory spatial analysis with predictive modelling approaches that differ in how country-specific heterogeneity is handled. Predictive modelling is conducted within a GPBoost framework estimated both with and without explicit country-specific effects, using a country-year panel dataset constructed from variables reflecting the main drivers of emissions identified in the literature. Exploratory spatial analysis is used to describe the geographic distribution of emissions and to assess whether fitted models reduce spatial autocorrelation in the residuals. Model performance, residual behaviour, and SHAP-based feature attribution are used to compare different GPBoost specifications and to evaluate how explicitly accounting for country-specific heterogeneity affects predictions and interpretation.

2. Methodology

2.1. Data

This study uses a panel dataset covering 25 European countries: France, Germany, Italy, Poland, Spain, Sweden, Finland, Belgium, the Netherlands, Austria, the Czech Republic, Greece, Hungary, Portugal, Slovakia, Estonia, Latvia, Lithuania, Slovenia, Denmark, Romania, Bulgaria, Croatia, and Norway. The analysis covers the period 2000-2023. The sample is restricted to countries for which annual data are available for all variables over the study period.

The dependent variable in this study is domestic net GHG emissions per capita, measured in tonnes per capita. Emissions from land use, land-use change, and forestry (LULUCF) are excluded. This variable represents production-based GHG emissions generated within national borders.

The explanatory variables capture key drivers of GHG emissions identified in the literature [1, 23], including energy consumption, economic affluence, climate conditions, energy mix, economic structure, and climate policy. Data for these variables are sourced from Eurostat [57], with the exception of the carbon emissions futures price, which is obtained from Investing.com [58].

Table 2 presents an overview of the variables used in the study, including units of measurement and data sources.

Table 2. Variables used in the analysis

Variable category	**Variable (and its column name in the dataset)**	**Unit**	**Source**
Outcome (dependent)	Domestic net greenhouse gas emissions (*GHG*)	Tonnes per capita	Eurostat (sdg_13_10)
Efficiency	Energy intensity (*Energy_Intensity_CLV*)	Kilograms of oil equivalent (KGOE) per thousand euro (2015 chain-linked volumes)	Eurostat (nrg_ind_ei)
Energy consumption	Final energy consumption per capita (*Energy_Consumption_Per_Capita*)	Terajoules per capita	Eurostat (nrg_bal_s)
Climate conditions	Heating degree days (*Heating_Degree_Days*)	Number of heating degree days	Eurostat (nrg_chdd_a)
Affluence	GDP per capita (PPS-adjusted) (*GDP_Per_Capita_PPP*)	Purchasing power standards (PPS) per capita	Eurostat (prc_ppp_ind)
Policy	Carbon Emissions Futures price (ETS)	EUR per tonne of CO_2-equivalent	Investing.com (CFI2H6)
Energy mix (production)	Renewables and biofuels share in electricity and heat production (*Gross_Renewables_Biofuels_Share*)	Percentage	Eurostat (nrg_bal_peh)
Energy mix (production)	Nuclear share in electricity and heat production (*Gross_Nuclear_Share*)	Percentage	Eurostat (nrg_bal_peh)
Energy mix (production)	Oil and petroleum share in electricity and heat production (*Gross_Liquid_Fuels_Share*)	Percentage	Eurostat (nrg_bal_peh)

Variable category	Variable (and its column name in the dataset)	Unit	Source
Energy mix (production)	Gas share in electricity and heat production (Gross_Gaseous_Fuels_Share)	Percentage	Eurostat (nrg_bal_peh)
Energy mix (production)	Solid fossil fuels share in electricity and heat production (*Gross_Coal_Peat_Shale*)	Percentage	Eurostat (nrg_bal_peh)
Energy mix (consumption)	Renewables and biofuels share in final energy consumption (*Final_Renewables_Biofuels_Share*)	Percentage	Eurostat (nrg_bal_s)
Energy mix (consumption)	Electricity share in final energy consumption (*Final_Electricity_Share*)	Percentage	Eurostat (nrg_bal_s)
Energy mix (consumption)	Oil and petroleum share in final energy consumption (*Final_Liquid_Fuels_Share*)	Percentage	Eurostat (nrg_bal_s)
Energy mix (consumption)	Gas share in final energy consumption (*Final_Gaseous_Fuels_Share*)	Percentage	Eurostat (nrg_bal_s)
Energy mix (consumption)	Solid fossil fuels share in final energy consumption (*Final_Coal_Peat_Shale_Share*)	Percentage	Eurostat (nrg_bal_s)
Economic structure	Construction share of gross value added (*Construction_Value_Added_Gross_CLV_Share*)	Percentage	Eurostat (nama_10_a64)
Economic structure	Agriculture, forestry and fishing share of gross value added (*Agriculture_Value_Added_Gross_CLV_Share*)	Percentage	Eurostat (nama_10_a64)
Economic structure	Mining and quarrying share of gross value added (*Mining_and_Quarrying_Value_Added_Gross_CLV_Share*)	Percentage	Eurostat (nama_10_a64)
Economic structure	Manufacturing share of gross value added (*Manufacturing_Value_Added_Gross_CLV_Share*)	Percentage	Eurostat (nama_10_a64)
Economic structure	Electricity, gas, steam and air conditioning supply share of gross value added (*EGSA_Value_Added_Gross_CLV_Share*)	Percentage	Eurostat (nama_10_a64)

Energy consumption per capita, energy intensity, GDP per capita (PPS), and heating degree days are included to capture differences in energy use, efficiency, economic affluence, and climate-related energy demand across countries. Climate policy is proxied by the EU Emissions Trading System (ETS) carbon price, which varies over time but is common across countries in a given year.

2.1.1. Energy mix variables

Energy mix variables describe how energy is produced and consumed within each country and are expressed as shares to allow comparison across countries of different sizes. Two dimensions are considered: electricity and heat production and final energy consumption.

2.1.1.1. Electricity and heat production

Production-based variables measure the share of electricity and heat production generated from each fuel type. For fuel f, the share is defined as:

$$\text{Production share}_{\text{f}} = \frac{\text{Electricity and heat production from f}}{\text{Total electricity and heat production}} \tag{2.1}$$

These variables describe the composition of fuels used to generate electricity and heat. The analysis includes shares of renewables and biofuels, nuclear energy, oil and petroleum products, gaseous fuels, and solid fossil fuels, which are the main fuels used for electricity and heat production in Europe [59].

2.1.1.2. Final energy consumption

Consumption-based variables measure the share of total final energy consumption by fuel type. For fuel f, the final consumption share is defined as:

$$\text{Final consumption share}_{\text{f}} = \frac{\text{Final energy consumption of f}}{\text{Total final energy consumption}} \tag{2.2}$$

These variables describe the composition of fuels used by end users, including households, transport, and industry. The analysis includes final consumption shares of renewables and biofuels, electricity, oil and petroleum products, gaseous fuels, and solid fossil fuels, which together account for the vast majority of final energy use in Europe [59].

Electricity and heat production and final energy consumption capture different mechanisms that affect emissions. Changes in the electricity and heat production (for example, replacing coal with renewables) reduce emissions even if final energy consumption remains unchanged. In contrast, changes in final energy consumption (for example, electrification of transport) affect emissions even if the electricity and heat production remains unchanged.

2.1.2. Economic structure

Economic structure is captured using sectoral shares of gross value added (GVA) based on NACE Rev. 2, which classifies economic activities into industry sectors. GVA is measured in chain-linked volumes (2015) to remove the effects of price changes over time. For sector s, the GVA share is defined as:

$$\text{GVA share}_{\text{s}} = \frac{\text{Gross value added of sector s}}{\text{Total gross value added}} \tag{2.3}$$

These variables describe the sectoral composition of the economy within each country. The analysis includes shares of manufacturing; mining and quarrying; electricity, gas, and steam supply; construction; and agriculture, as these sectors are energy-intensive and directly relevant for GHG emissions.

2.2. Spatial hot spot and emerging hot spot analysis

To identify spatial clustering and its evolution over time, the study applies Local Getis-Ord hot spot analysis and Emerging Hot Spot Analysis (EHSA). Local Getis-Ord statistics identify countries that are part of statistically significant local clusters of high or low GHG emissions. EHSA summarises how these local clusters evolve over time.

2.2.1. Software and implementation

Spatial hot spot analysis and EHSA are implemented in Python using the esda module from the PySAL library for spatial data analysis. Local Getis-Ord statistics are computed using the G_Local class, and statistical inference is conducted using permutation-based standardized statistics (z_{sim}) and permutation-based p values (p_{sim}) derived from the permutation distributions generated by G_Local. Global Moran's I is computed using the same spatial weights matrix and permutation-based inference and is used to assess global spatial autocorrelation in model residuals.

2.2.2. Spatial weights matrix

Spatial dependence is represented using a binary contiguity-based spatial weights matrix. Neighbours are defined primarily by shared land borders. Two additional non-land neighbour links are included to reflect short sea distances: Estonia-Finland and Sweden-Denmark. All spatial weights are binary and unstandardized.

2.2.3. Local Getis-Ord Statistics

Let y_i denote GHG emissions for country i, and let $N(i)$ denote the set of neighbouring countries defined by the spatial weights matrix. Because the weights are binary, the contribution of neighbouring countries reduces to a simple sum of their emission values.

The local Getis-Ord G_i statistic is computed as

$$G_i = \frac{\sum_{j \in N(i)} y_j}{\sum_{j \neq i} y_j} \tag{2.4}$$

The G_i^* statistic, which includes the focal country, is computed as

$$G_i^* = \frac{y_i + \sum_{j \in N(i)} y_j}{\sum_j y_j} \tag{2.5}$$

2.2.4. Permutation-based inference

Because GHG emissions do not follow a normal distribution, standardization based on normal distribution is not appropriate. Statistical inference is therefore conducted using conditional randomization, in which the observed emission value of each focal country is held fixed while the emission values of all other countries are randomly permuted.

For each country i, the emission value y_i is held fixed, while the remaining $n - 1$ values are randomly permuted across the other countries. For each country and year, $R = 9999$ permutations are

generated. For each permutation r, the corresponding local statistic $G_i^{(r)}$ (or $G_i^{*(r)}$) is computed by using the same set of neighbouring countries but with emissions reassigned at random. This creates a permutation distribution for each country.

A standardized permutation statistic (z value) is then computed as

$$z_{sim,i} = \frac{G_i - \mu_i^{perm}}{\sigma_i^{perm}} \tag{2.6}$$

where G_i denotes the statistic computed using the observed (unpermuted) emission values, and μ_i^{perm} and σ_i^{perm} denote the mean and standard deviation of the permutation distribution.

Permutation-based p values are computed as

$$p_{sim,i} = \frac{1 + min\left(\#\left(G_i^{(r)} \geq G_i\right), \#\left(G_i^{(r)} \leq G_i\right)\right)}{1 + R} \tag{2.7}$$

where $p_{sim,i}$ represents the proportion of permuted statistics $G_i^{(r)}$ that are at least as extreme as the G_i. Smaller values of $p_{sim,i}$ identify locations with unusually high (hot spots) or unusually low (cold spots) local clustering and indicate stronger evidence against the null hypothesis of spatial randomness.

2.2.5. Hot and cold spot identification

Local hot and cold spots are identified by combining the sign of the standardized permutation statistic $z_{sim,i}$ with the permutation-based p value $p_{sim,i}$.

A country i is classified as a hot spot if:

$$z_{sim,i} > 0 \wedge p_{sim,i} < \alpha \tag{2.8}$$

And as a cold spot if:

$$z_{sim,i} < 0 \wedge p_{sim,i} < \alpha \tag{2.9}$$

Where α denotes the chosen significance threshold. In this study, two significance thresholds are used: a strict threshold ($\alpha = 0.05$) and a looser threshold ($\alpha = 0.10$).

2.2.6. Relationship between G_i and G_i^* when calculating permutation-based z and p values

When permutation-based inference is used, the local Getis-Ord statistics G_i and G_i^* lead to identical standardized permutation statistics (z_{sim}) and identical permutation-based p values (p_{sim}). The equivalence of permutation-based inference for G_i and G_i^* is shown in detail in Appendix 1.

In this study, local hot spot analysis and EHSA use permutation-based standardized statistics computed from G_i^*. However, because permutation-based inference is invariant to the inclusion of the focal country's value, a country with relatively low emissions can still be identified as a hot spot if it is surrounded by high-emission neighbours, and a country with relatively high emissions can be identified as a cold spot if it is surrounded by low-emission neighbours, even though G_i^* is used.

For this reason, countries identified as statistically significant hot or cold spots are further classified as core or non-core. A country is classified as a core hot (cold) spot if its own GHG emission level lies above (below) the median in that year. Countries that are identified as statistically significant hot or cold spots, but whose own emission levels do not exceed (or fall below) the median are classified as non-core. This distinction allows separation between countries that exhibit high or low emission levels themselves and countries that are part of a significant spatial cluster primarily due to neighbouring countries.

2.2.7. Emerging Hot Spot Analysis (EHSA)

To examine the temporal dynamics of spatial clustering, an Emerging Hot Spot Analysis (EHSA) is conducted. EHSA assesses two aspects: temporal trends and temporal patterns.

Temporal trends are assessed using the permutation-based statistic z_{sim} and the nonparametric Mann-Kendall trend test. The test identifies whether z_{sim} values tend to increase, decrease, or show no consistent change over time, and provides a p value for statistical significance.

Temporal patterns are assessed using a rule-based approach. For each country, information on hot and cold spot status is combined with trend information to assign a single EHSA category. Classification into these categories first depends on whether the country is identified as a hot or cold spot in the most recent year. Countries that are not currently hot or cold spots are classified as having a historical pattern if hot or cold spots occurred in the past, or no pattern if no hot or cold spots occurred at any time (they are not shown on the map).

For countries that are currently hot or cold spots, additional rules are applied. A country is classified as a new hot or cold spot if the classification appears for the first time in the most recent year. A consecutive pattern is assigned when the country is classified as a hot or cold spot in several consecutive recent years. A persistent pattern is assigned when the country is classified as a hot or cold spot for large share of the study period (e.g., at least 90% of the years). A sporadic pattern is assigned when hot or cold spots occur in some years but do not meet the criteria for new, consecutive, or persistent patterns.

Finally, if a country is currently a hot or cold spot and has a statistically significant trend in z_{sim}, it is classified as an intensifying or diminishing hot or cold spot depending on the direction of the trend.

2.3. GPBoost

GHG emissions are modelled using contemporaneous variables described in Table 2 in a country-year panel. To capture nonlinear relationships and unobserved country-specific heterogeneity, the study employs GPBoost, which is a hybrid ML framework that integrates gradient boosted decision trees with Gaussian process and random-effects components.

2.3.1. Software and implementation

All models are implemented in Python. GPBoost and linear models are estimated using the gpboost package. Hyperparameter tuning is performed using Optuna.

2.3.2. General model structure

Following Sigrist [51], the GPBoost model can be written as a nonlinear mixed-effects model:

$$y = F(X) + Zb + \epsilon \tag{2.10}$$

with

$$b \sim N(0, \Sigma)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

where:

$y \in \mathbb{R}^n$ is the vector of observed GHG emissions,

$X \in \mathbb{R}^{n \times p}$ is the matrix of observed features,

$F(X)$ is a nonlinear prediction function learned using gradient boosted decision trees,

Zb represents random effects contribution to the prediction,

ϵ is a noise term.

2.3.3. GPBoost specifications

M1: GPBoost with observed features only

In model M1, predictions are generated solely by the nonlinear boosting component:

$$y = F(X) + \epsilon \tag{2.11}$$

M11: GPBoost with observed features and a global time trend

Model M11 extends M1 by including a global linear time trend as an additional observed feature in X:

$$y = F(X, year) + \epsilon \tag{2.12}$$

M2: GPBoost with observed features and country-specific intercepts and time trends

Model M2 augments the nonlinear boosting component with country-specific random effects, implemented as country-specific intercepts and linear time trends:

$$y = F(X) + \alpha_c + \delta_c \cdot year + \epsilon \tag{2.13}$$

where α_c denotes a country-specific intercept and δ_c denotes a country-specific slope on time.

M21: GPBoost with observed features, a global time trend, and country-specific random effects

Model M21 extends M2 by including a global linear time trend as an additional observed feature in X:

$$y = F(X, year) + \alpha_c + \delta_c \cdot year + \epsilon \tag{2.14}$$

2.3.4. Linear benchmark models

L1: Linear pooled regression

The pooled linear regression model is specified as:

$$y = X\beta + \gamma \cdot year + \epsilon \tag{2.15}$$

where all countries share the same coefficients β and a common linear time trend γ.

L2: Linear mixed-effects model

The linear-mixed effects model is specified as:

$$y = X\beta + \gamma \cdot year + \alpha_c + \delta_c \cdot year + \epsilon \tag{2.16}$$

where α_c and δ_c denote country-specific intercepts and linear time trends.

2.3.5. Model training and assessment

2.3.5.1. Cross-validation design

Model training and tuning are conducted using expanding-window cross-validation, which respects the temporal ordering of the data. Three cross-validation folds are defined:

- Fold 1: training: 2000-2009, validation: 2010-2011
- Fold 2: training: 2000-2014, validation: 2015-2016
- Fold 3: training: 2000-2019, validation: 2020-2021

In each fold, the model is trained on all observations up to the end of the training period and validated on subsequent years. Validation periods consist of two consecutive years in order to obtain more stable performance estimates than single-year validation.

The folds are placed in the earlier, middle, and later parts of the sample to assess model performance over time. Validation periods prior to 2010 are not used because they would result in smaller training samples relative to the complexity of the model. Evaluating every possible validation year would be computationally intensive and add little additional information, as adjacent years rely on very similar training samples under an expanding-window design.

2.3.5.2. Hyperparameter tuning

Hyperparameters of the tree-based component are tuned using Optuna, a Bayesian optimization framework. For each candidate hyperparameter configuration, the model is trained and evaluated across all cross-validation folds.

The tuning objective is the mean validation Root Mean Squared Error (RMSE) across these folds.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.17}$$

The search process is guided by the Tree-structured Parzen Estimator (TPE) sampler with a fixed random seed of 42 to ensure reproducibility. The TPE sampler models the relationship between hyperparameters and validation performance and proposes new candidate configurations that are more likely to improve the objective based on past evaluations [60].

For each Optuna trial and each cross-validation fold, models are trained with a maximum of 5000 boosting iterations. Early stopping is applied with a patience of 200 iterations based on validation RMSE.

- The following hyperparameters are tuned:
- Learning rate: learning_rate $\in [0.01, 0.2]$ (logarithmic scale)
- Number of leaves: num_leaves $\in [16, 256]$ (logarithmic scale)
- Maximum tree depth: max_depth $\in [-1, 12]$
- Minimum number of observations per leaf: min_data_in_leaf $\in [10, 100]$ (logarithmic scale)
- Feature fraction: feature_fraction $\in [0.5, 1]$
- L1 regularization: lambda_l1 $\in [10^{-8}, 10]$ (logarithmic scale)
- L2 regularization: lambda_l2 $\in [10^{-8}, 10]$ (logarithmic scale)
- Minimum gain required for a split: min_gain_to_split $\in [0, 1]$
- Bagging fraction: bagging_fraction $\in [0.5, 1]$
- Bagging frequency: bagging_freq $\in [0, 10]$

For model specifications that include random effects component, bagging is disabled by fixing bagging_fraction to 1 and bagging_freq to 0. This ensures that the dependence structure which is explicitly modelled by GPBoost is preserved, as bagging assumes independent observations.

For each model specification, the final hyperparameter configuration that achieves lowest mean validation RMSE across cross-validation folds is selected. The final hyperparameter values selected for each GPBoost model specification are reported in Appendix 2.

2.3.5.3. Final train and test split

After hyperparameter tuning, each model is retrained on the full training sample covering the years from 2000 to 2021. Final model performance is evaluated on a test set covering the years from 2022 to 2023. This test set is not used at any stage of model tuning and provides an unbiased assessment of predictive performance.

2.4. Residual analysis

2.4.1. Residual definition

For all model specifications, residuals are computed on validation samples obtained from the expanding-window cross-validation scheme. Residuals are defined as

$$r_{c,t} = y_{c,t} - \hat{y}_{c,t} \tag{2.18}$$

where $y_{c,t}$ denotes the observed outcome for country c in year t, and $\hat{y}_{c,t}$ is the corresponding model prediction.

2.4.2. Residual global spatial autocorrelation

To assess whether prediction residuals exhibit global spatial dependence, global Moran's I, a measure of global spatial autocorrelation, is computed for residuals in each validation year. For a given year t, Moran's I is calculated using a set of country-level residuals $r_{c,t}$. Statistical inference is conducted using permutation-based testing with $R = 9999$ permutations. The Moran's I statistic and the corresponding permutation-based p value are reported.

2.4.3. Residual local spatial clustering

To assess local spatial patterns in prediction residuals, local Getis-Ord G_i^* statistics are computed for residuals in selected validation years. Residual clusters are displayed using outlines only, without classifying them as hot or cold spots, in order to focus on the spatial location of statistically significant local clustering.

Residual cluster maps are used as a diagnostic tool to examine whether spatial clusters observed in the original GHG emissions persist in prediction residuals and whether their spatial locations remain stable over time. Maps are shown for selected validation years, with one year taken from each cross-validation fold (2011, 2016, and 2021).

2.4.4. Time-specific median residuals

To assess systematic bias over time, the median residual across countries is computed for each validation year. For a given year t, this statistic is defined as

$$\tilde{r}_t = median_c(r_{c,t}) \tag{2.19}$$

Plotting $\tilde{r}_t$ over time allows detection of year-specific overprediction or underprediction.

2.4.5. Cross-country residual dispersion

To assess the dispersion of prediction residuals across countries within a given year, residuals are summarised using percentile bands. For a given year t, the 10th and 90th percentiles of residuals across countries are computed as

$$q_{0.10,t} = Q_{0.10}\left(\{r_{c,t}\}_c\right), \qquad q_{0.90,t} = Q_{0.90}\left(\{r_{c,t}\}_c\right) \tag{2.20}$$

These bounds are plotted together with the median residual to visualise how evenly prediction errors are distributed across countries in each year.

2.4.6. Country-level bias

To assess systematic overprediction or underprediction at the country level, the mean residual for each country is computed over the validation period:

$$\tilde{r}_c = \frac{1}{T}\sum_{t \in T} r_{c,t} \tag{2.21}$$

where T denotes set of validation years. Countries with the most negative and most positive values of $\tilde{r}_c$ are interpreted as being systematically overpredicted and underpredicted.

To summarise country-level bias, the absolute mean residual $|\tilde{r}_c|$ is computed for each country and aggregated across countries using the median.

2.4.7. Within-country residual variance

To assess the temporal stability of prediction residuals within countries, the variance of residuals over time is computed for each country:

$$Var(r_{c,t}) = \frac{1}{T-1}\sum_{t \in T}\left(r_{c,t} - \tilde{r}_c\right)^2 \tag{2.22}$$

This statistic measures the extent to which prediction residuals fluctuate over time within a country. The distribution of within-country residual variances is compared across models using boxplots.

2.5. Variance decomposition of features

To assess whether observed features vary primarily across countries or over time within countries, a variance decomposition is computed for each feature in the dataset. For each country c, the time average $\bar{X}_c$ is calculated. The between-country variance is defined as the sample variance of these country means:

$$Var_{between} = Var_c(\bar{X}_c) \tag{2.23}$$

The within country variance is defined as the average of country-specific sample variances over time:

$$Var_{within} = \frac{1}{C}\sum_{c} Var_t\left(X_{c,t}\right) \tag{2.24}$$

where C is the total number of countries in the sample.

Then, between country variance share is computed as

$$between\ share = \frac{Var_{between}}{Var_{between} + Var_{within}} \tag{2.25}$$

Values closer to 1 indicate that variation is dominated by differences between countries, while lower values indicate greater within-country variation.

3. Results

3.1. GHG emission trends across European countries

Fig. 1 shows changes in GHG emissions over time for all countries included in the analysis.

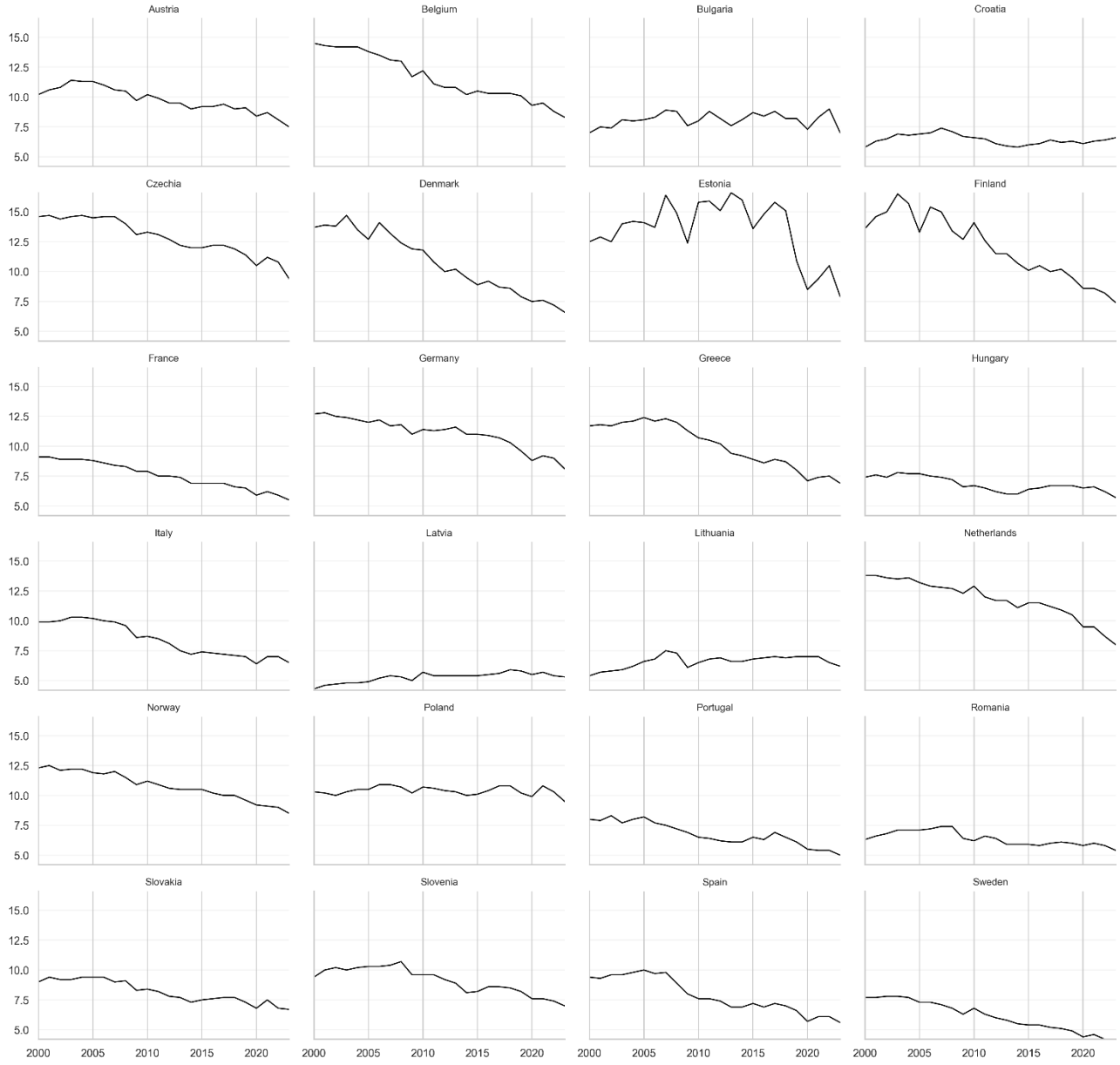


Fig. 1. GHG emissions by country (2000-2023)

Overall, GHG emissions decline in most countries, although the rate of decline differs. In some countries, emissions decrease more gradually, while in others they remain relatively stable over time. Most emission trajectories are smooth, but some countries, such as Estonia, show pronounced fluctuations. To better understand how emissions are spatially distributed across Europe, and whether countries with higher or lower emissions form geographic clusters, the following sections analyse spatial clustering of GHG emissions.

3.2. Local hot spot and emerging hot spot analysis

This section analyses spatial and spatio-temporal clustering of GHG emissions across Europe using local hot spot analysis and Emerging Hot Spot Analysis (EHSA). The analysis focuses on macro-regional clustering across Europe.

3.2.1. Local hot and cold spots of GHG emissions

Local spatial autocorrelation in GHG emissions is examined using the Getis-Ord G_i^* statistic, which identifies statistically significant clusters of high (hot spots) and low (cold spots) values. Hot and cold spots are identified at stricter level ($\alpha = 0.05$, solid outlines) and looser level ($\alpha = 0.10$, dashed outlines) (Fig. 2). Within these clusters, countries are classified as *core* if their own emission lie above (for hot spots) or below (for cold spots) the median in that year, and as *non-core* otherwise. The maps are shown for selected years to illustrate how spatial clustering differs between the beginning, middle, and end of the study period.

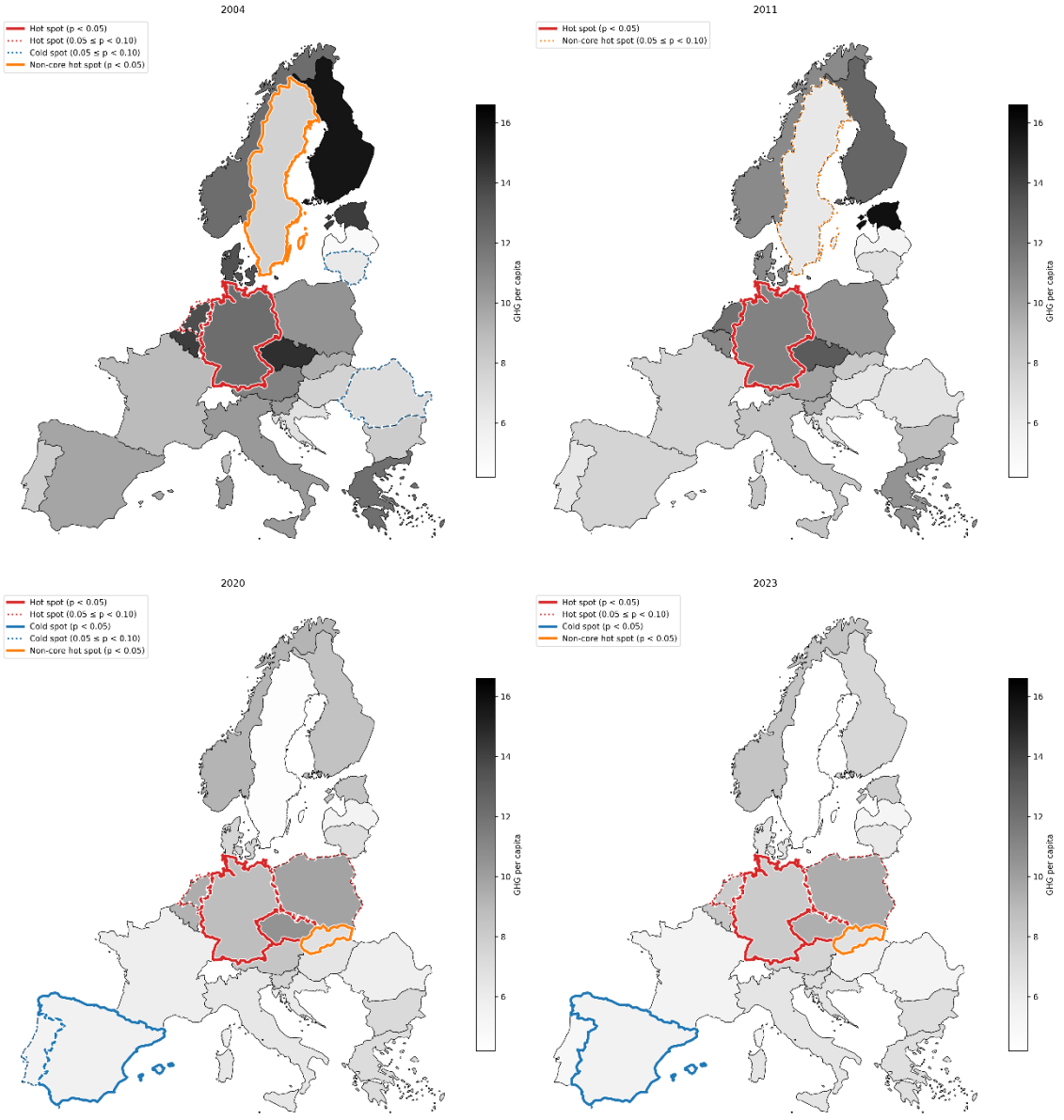


Fig. 2. Local hot and cold spots of GHG emissions in Europe in 2004, 2011, 2020, and 2023 based on the Getis-Ord G_i^* statistic with permutation-based inference. Country shading represents GHG emission levels

(darker shades indicate higher emissions and lighter shades indicate lower emissions). Solid outlines indicate statistically significant hot and cold spots at the $\alpha = 0.05$ level, while dashed outlines indicate significance at the $\alpha = 0.10$ level

In 2004, clusters of high GHG emissions are located in Central and Northern Europe. Germany (Central Europe) and Sweden (Northern Europe) are identified as hot spots at the stricter significance level. Germany is classified as a core hot spot, while Sweden is classified as a non-core hot spot, because Sweden's own emission level is below the median while it is surrounded by high-emitting neighbours. This follows from the fact that permutation-based z and p for the G_i^* statistic are unaffected by the inclusion of the focal country's value, as described in Section 2.2.6. Therefore, identifying Sweden as a non-core hot spot indicates that Sweden lies within a region where emission levels are significantly elevated.

In 2011, Sweden is classified as a non-core hot spot only at the looser significance level, while Germany remains a core hot spot at the stricter level. No statistically significant cold spots are observed.

By 2020 and 2023, overall GHG emission levels have declined, and statistically significant hot spots are concentrated in Central Europe. Germany and Czech Republic are core hot spots, and Slovakia is identified as a non-core hot spot at the stricter significance level. Poland and the Netherlands are identified as core hot spots at the looser significance level. At the same time, Spain is identified as a core cold spot at the stricter significance level, indicating a cluster of relatively low emissions in Southwestern Europe.

Overall, the results show that while GHG emission levels in Europe decline over time, spatial clustering of emissions persists. In the earlier period, statistically significant hot spots are observed in both Northern and Central Europe, while cold spots are weak. In later years, the Northern European cluster weakens, while hot spots remain primarily in Central Europe and a statistically significant cold spot emerges in Southwestern Europe. This change shows uneven emission reductions across regions, with faster declines in Northern Europe and slower declines in parts of Central Europe.

3.2.2. Temporal trends in permutation-based z values

To examine how spatial clustering changes over time, Mann-Kendall trend test is applied to the time series of permutation-based z value for each country. Fig. 3 shows these trends for each country. Countries with a statistically significant increasing trend ($p < 0.05$) are shown in red, countries with a statistically significant decreasing trend are shown in blue, and countries with no statistically significant monotonic trend are shown in grey.

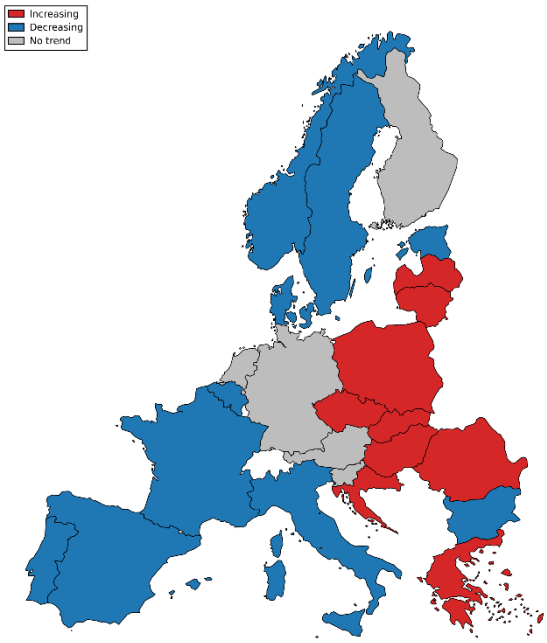


Fig. 3. Statistically significant trends in permutation-based Getis-Ord G_i^* z values across Europe. Red indicates statistically significant increasing trends, blue indicates statistically significant decreasing trends, and grey indicates no statistically significant trend (Mann-Kendall test, $p < 0.05$)

Statistically significant decreasing trends in permutation-based z values are concentrated primarily in Northern and Western Europe, while statistically significant increasing trends are concentrated primarily in Central and Eastern Europe. Permutation-based z values indicate the direction and strength of local spatial clustering, with positive (negative) values corresponding to hot-spot-like (cold-spot-like) patterns, while statistical significance is assessed using the associated permutation-based p values. Decreasing z values in Northern and Western Europe indicate a shift away from hot-spot-like patterns over time, including transitions toward cold spots, strengthening of cold spots, or movement toward the absence of clear spatial clustering. Increasing z values in Central and Eastern Europe indicate a shift toward hot-spot-like patterns, including movement away from cold spots or the strengthening or emergence of hot spots. This shows that, alongside declining GHG emissions across Europe, the spatial structure of emissions evolves differently across regions.

3.2.3. Temporal patterns of GHG emissions hot and cold spots

To summarise the temporal behaviour of local spatial clustering into a single classification for each country, Emerging Hot Spot Analysis (EHSA) is used. Only countries that are classified as hot or cold spots in the most recent year are assigned an EHSA category. The results are shown in Fig. 4.

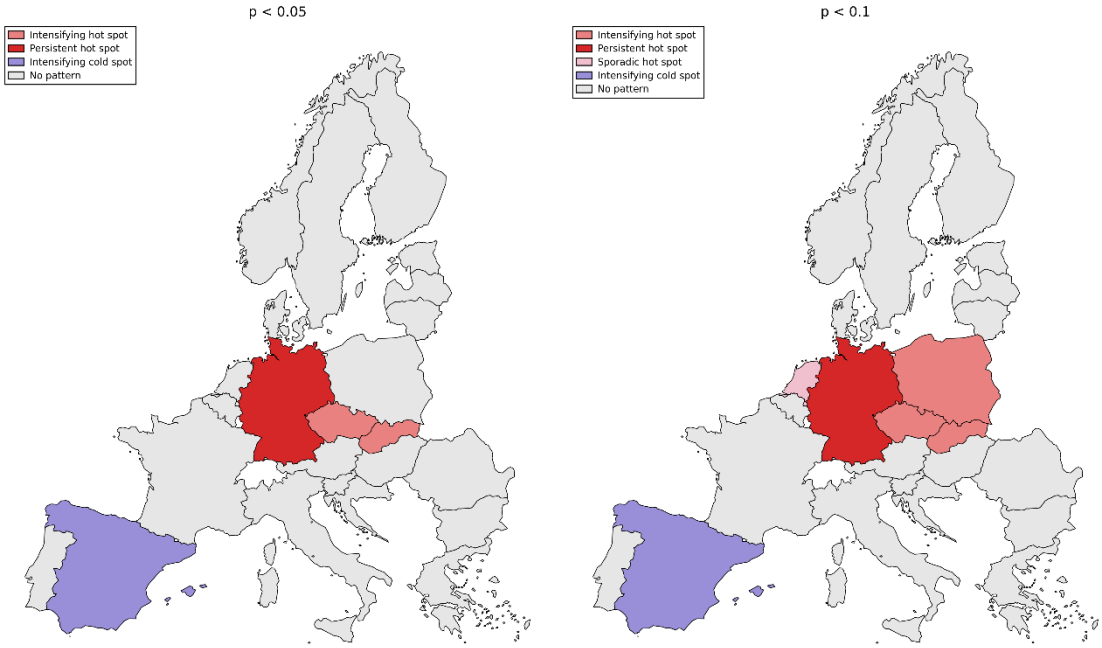


Fig. 4. Emerging Hot Spot Analysis (EHSA) results for GHG emissions across Europe. The map on the left shows EHSA classification based on the stricter significance level ($\alpha = 0.05$), while the map on the right shows classification based on the looser significance level ($\alpha = 0.10$). Countries are coloured according to their EHSA category

At stricter significance level ($\alpha = 0.05$) Germany is classified as a persistent hot spot, meaning that it is identified as a statistically significant hot spot in at least 90% of the study years. The Czech Republic and Slovakia are classified as intensifying hot spots, which means that they are hot spots in most recent year and display statistically significant increasing trends in permutation-based z values over time. Spain is classified as an intensifying cold spot, which means that it is a cold spot in the most recent year and displays a statistically significant decreasing trend in permutation-based z values over time. When a looser significance level is applied ($\alpha = 0.10$), two additional emerging patterns are identified. Poland is classified as an intensifying hot spot, and the Netherlands is classified as a sporadic hot spot, meaning that it is identified as a hot spot in some years but does not meet the criteria for persistent, consecutive, or intensifying patterns.

Overall, the EHSA results reinforce the earlier findings by showing that persistent and intensifying hot spots are concentrated in Central Europe, while an intensifying cold spot emerges in Southwestern Europe.

3.3. GPBoost

This section presents the results of the GPBoost models and compares their predictive performance to linear benchmark specifications. Residual behaviour is analysed across countries and over time, and SHAP values are used to examine feature attributions.

3.3.1. Performance accuracy (RMSE)

To compare different models and assess out-of-sample predictive accuracy, Root Mean Squared Error (RMSE) is used. RMSE is computed on validation samples obtained from an expanding-window

cross-validation scheme and also on a test sample covering the period 2022-2023. Details of model tuning and cross-validation are provided in Section 2.3.5.

The models considered correspond to the specifications defined in Section 2.3.3 and include:

- **M1**: GPBoost with observed features only
- **M11**: GPBoost with observed features and a global linear time trend included in the boosting component
- **M2**: GPBoost with observed features and country-specific intercepts and linear time trends (country-specific effects)
- **M21**: GPBoost with observed features, a global linear time trend, and country-specific effects
- **Linear pooled**: pooled linear regression with observed features and a global time trend
- **Linear mixed-effects**: linear mixed-effects model with country-specific effects

Table 3 reports RMSE values across the three validation folds for each model

Table 3. Cross-validation RMSE across three folds for each model

Model	Fold 1	Fold 2	Fold 3
M1	0.645	0.447	0.482
M2	0.517	0.397	0.514
M11	0.691	0.466	0.495
M21	0.545	0.386	0.474
Linear Mixed	0.463	0.369	0.681
Linear pooled	1.509	1.038	1.243

The pooled linear regression performs worst in all validation folds. This model assumes a single linear relationship between the observed features and emissions that is shared across all countries and over time, which limits its predictive accuracy.

Including country-specific intercepts and linear time trends in the linear mixed-effects model leads to a clear reduction in RMSE compared to the pooled linear specification. Because the linear mixed-effects model is additive, this improvement can be attributed to the inclusion of country-specific effects, which capture systematic differences in emissions and trends across countries.

Among the GPBoost models, model specifications without country-specific effects included (M1 and M11) already achieve relatively low RMSE. This reflects the flexibility of tree-based models, which can represent complex nonlinear relationships and can partially absorb persistent cross-country differences through observed features. Including country-specific intercepts and trends (M2 and M21) leads to slightly lower RMSE on average, although the differences relative to M1 and M11 are small.

Including a global time trend in the boosting component (M11 and M21) leads to RMSE values that are very similar to those of the corresponding specifications without this term (M1 and M2), indicating that overall time dynamics may be captured by other model components. Table 4 reports RMSE calculated on the test sample covering the period 2022-2023.

Table 4. Test sample RMSE for all model specifications

Model	Final 2022	Final 2023
M1	0.442	0.603
M2	0.377	0.603
M11	0.497	0.616
M21	0.343	0.602
Linear Mixed	0.346	0.511
Linear pooled	1.444	1.439

The performance of the models on test sample follows the same pattern as in cross-validation. The pooled linear model again exhibits the largest errors, while both GPBoost models with country-specific effects and the linear mixed-effects model achieve the lowest RMSE. Differences between GPBoost specifications remain small.

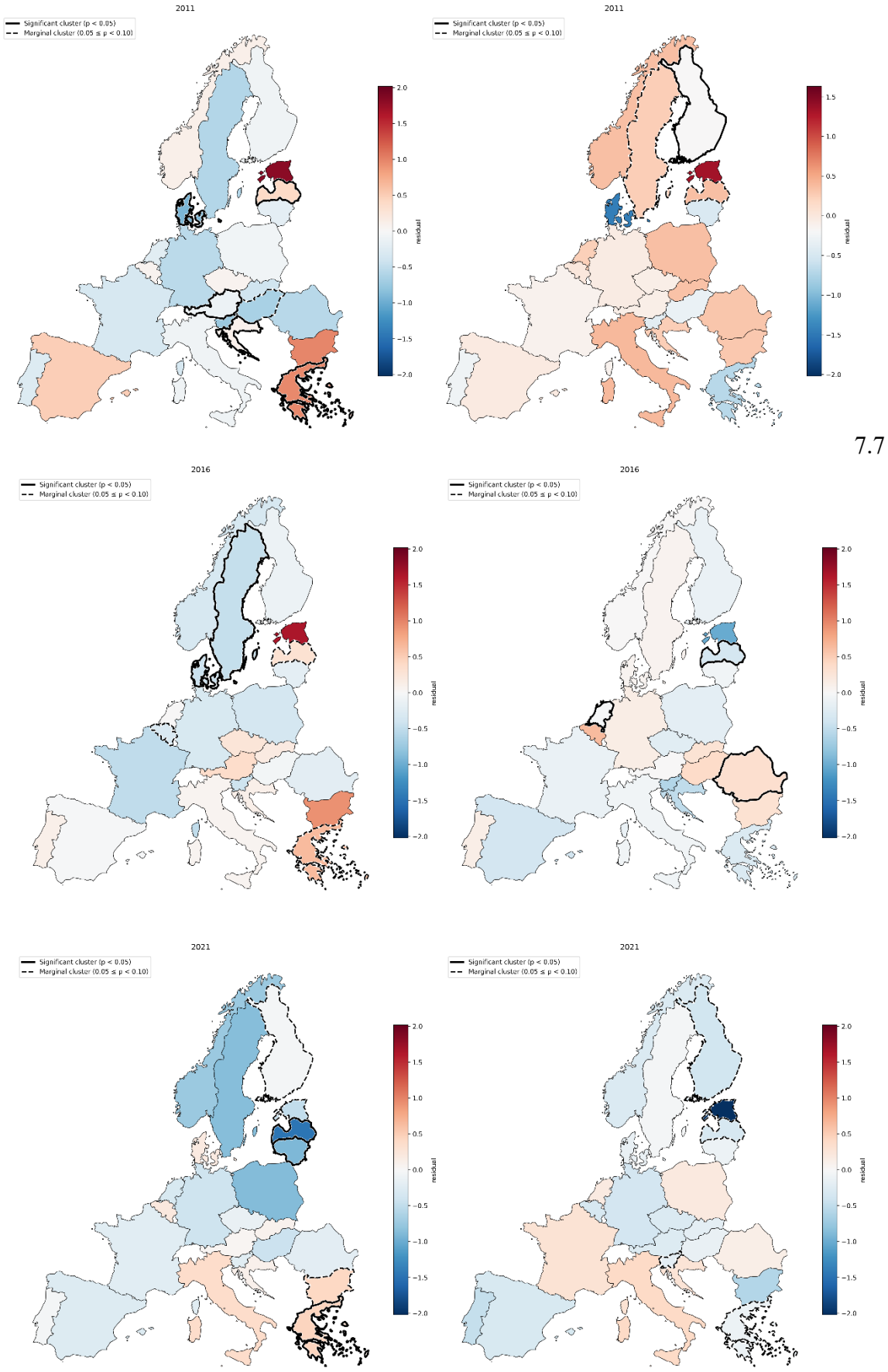
These results show that including country-specific effects strongly improves the predictive performance of the linear model, while in GPBoost the inclusion of such effects does not lead to large improvements in RMSE. However, similar RMSE values across different GPBoost specifications do not imply that country-specific effects are unimportant. Rather, they suggest that the flexible boosting component in GPBoost can capture persistent country-specific heterogeneity through observed features, even when these features were not designed to capture such structure.

3.3.2. Residual analysis

While RMSE provides a summary of average predictive performance, it does not describe how prediction residuals are distributed across countries or how they evolve over time within countries. For example, models with similar RMSE may still differ in how well they capture cross-country heterogeneity and country-specific temporal dynamics. To assess how the inclusion of country-specific effects affects residual behaviour, analyses residual behaviour across four GPBoost model specifications (M1, M11, M2, and M21), with residual spatial autocorrelation analysed for models M1 and M2 only.

3.3.2.1. Residual spatial autocorrelation

To assess whether the models adequately capture the spatial structure of GHG emissions, residual spatial autocorrelation is examined for models M1 and M2. Residuals are analysed for each validation year. For each year, residuals are mapped using a symmetric colour scale centred at zero. Local hot and cold spots identified using the Getis-Ord G_i^* statistic are shown using black outlines only, to simplify the maps and highlight whether clusters persist across years (Fig. 5).



7.7

Fig. 5. Residuals for model M1 (left) and model M2 (right) are shown for the validation years 2011, 2016, and 2021. Black outlines indicate local hot and cold spots based on the Getis-Ord G_i^* statistic ($\alpha = 0.05$ solid, $\alpha = 0.1$ dashed)

In Model M2, local clusters change location across validation years. In a small number of cases (for example Latvia), local clustering is driven by neighbouring countries with extreme residual values, such as Estonia, rather than by stable spatial patterns. Germany, which is identified as a persistent hot spot in the original GHG emission maps, is not identified as a residual hot or cold spot in any validation year.

In model M1, residual clustering is more persistent. Greece is identified as a local residual cluster in all three validation years shown. In addition, some countries that are not classified as local clusters display consistently negative residuals across years (for example Sweden, Germany, and France), indicating systematic overprediction in model M1.

Global Moran's I statistics with permutation-based inference are used to complement the visual inspection (Table 5).

Table 5. Moran's I values and permutation-based p values for residual spatial autocorrelation

	M1		**M2**	
	I	***p***	***I***	***p***
2010	0.258	0.013	0.174	0.074
2011	0.214	0.043	-0.104	0.330
2015	0.064	0.231	0.046	0.255
2016	0.186	0.058	0.126	0.129
2020	0.062	0.239	0.080	0.179
2021	0.204	0.052	0.052	0.215

For model M1, Moran's I is statistically significant or marginally significant in some of the earlier validation years, indicating remaining residual spatial autocorrelation not fully captured by the model. For model M2, Moran's I is not statistically significant in any validation year, and the magnitude of the statistic is consistently lower than for model M1.

Overall, these results indicate that model M2 reduces residual spatial dependence more effectively than model M1. Residual spatial patterns in model M2 are unstable over time and do not show persistent clustering, suggesting that the main spatial structure present in the original emission data is largely captured by model M2.

3.3.2.2. Residual patterns over time across countries

Fig. 6 shows, for each validation year, the median residual across countries. This statistic summarises whether the model tends to overpredict or underpredict emissions in a given year across countries.

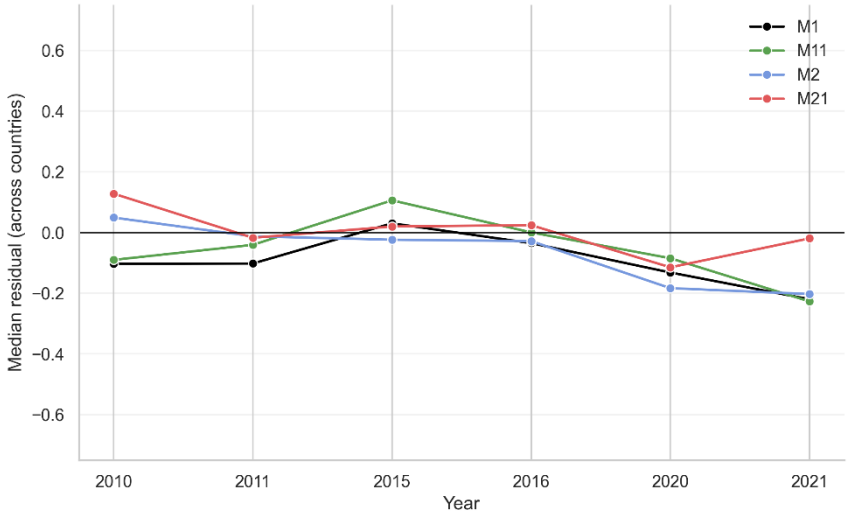


Fig. 6. Median residual across countries for each year in the validation sample, by model specification

In the later validation years, the median residuals across countries become more negative for three out of four models (M1, M11, M2), indicating systematic overprediction of emissions across countries in these years. The similarity of these temporal patterns in residuals across three models suggests that there may be a common temporal component that is not fully captured by the observed features. Model M21, which includes time both as an observed feature in the tree component and as a country-specific effect, shows reduction of this overprediction.

Fig. 7 presents the same median residuals shown in Fig. 6, together with 10^{th}-90^{th} percentile bands that illustrate how dispersed prediction residuals are across countries within each year.

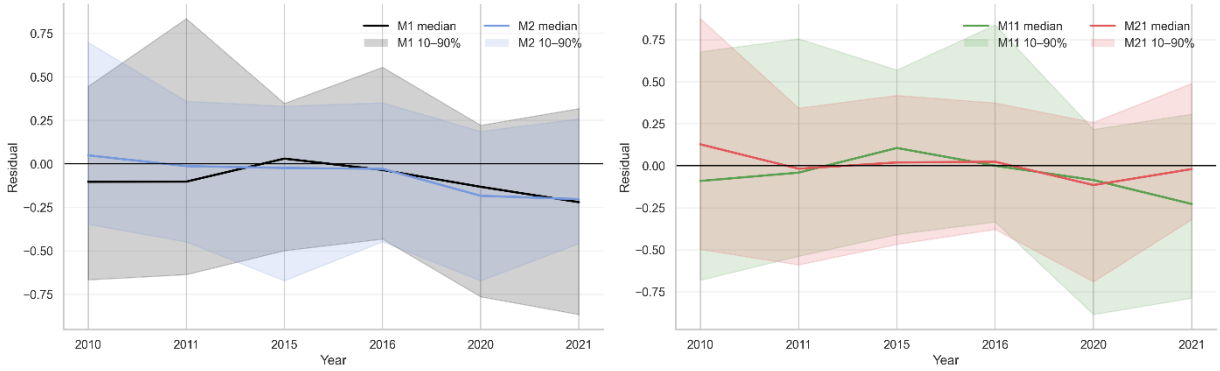


Fig. 7. Dispersion of residuals across countries for each year in the validation sample, by model specification

Models that incorporate country-specific intercepts and time trends (M2 and M21) display narrower dispersion bands than models without country-specific s (M1 and M11). This indicates that, within a given year, prediction errors are more similar across countries when country-specific structure is included.

3.3.2.3. Overpredicted and underpredicted countries

To identify systematic country biases, average residuals are computed for each country over the validation period. Table 6 shows two most overpredicted and two most underpredicted countries for each model specification.

Table 6. Most over-predicted and under-predicted countries based on average validation residuals

Model	Most over-predicted		Most under-predicted	
	Country	$\bar{r}_{c,t}$	**Country**	$\bar{r}_{c,t}$
M1	Sweden	-0.637	Estonia	0.943
	Poland	-0.520	Bulgaria	0.551
M11	Poland	-0.575	Estonia	0.978
	Sweden	-0.559	Bulgaria	0.582
M2	Estonia	-0.600	Italy	0.186
	Greece	-0.447	Romania	0.154
M21	Estonia	-0.448	Netherlands	0.289
	Greece	-0.355	Poland	0.253

The main difference between models without country-specific effects (M1 and M11) and models with country-specific effects (M2 and M21) is the magnitude of underprediction. In models M1 and M11, the most underpredicted country (Estonia) has an average residual close to one unit (0.943 and 0.978, respectively). In models that include country-specific intercepts and time trends, the largest underprediction is much smaller, with the largest value being 0.289 in model M21.

Differences in overprediction are smaller across models. The largest average overpredictions are of similar magnitude in all models (around 0.4-0.6), though they occur in different countries. For example, Estonia shifts from being the most underpredicted country in M1 and M11 to the most overpredicted country in M2 and M21, although the magnitude of Estonia's overprediction is smaller than its earlier underprediction.

To summarise systematic country-level biases, Table 7 shows the median absolute value of country-specific average residuals.

Table 7. Median absolute country-level residuals across the validation period

Model	Median
M1	0.232
M11	0.223
M2	0.126
M21	0.114

Models with country-specific effects (M2 and M21) display lower median values than models without these effects (M1 and M11). Therefore, including country-specific effects reduces systematic country prediction bias.

3.3.2.4. Within-country residual variance

While Fig. 7 shows the spread of residual dispersion across countries at specific points in time, Fig. 8 focuses on within-country residual variance over time. Each observation in the boxplots corresponds to the variance of residuals across years for a single country.

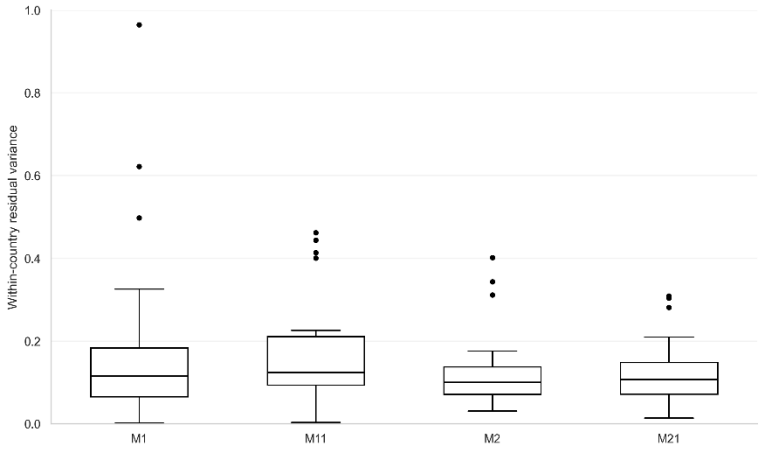


Fig. 8. Distribution of within-country residual variance by model specification

Models that include country-specific effects (M2 and M21) display narrower distributions of within-country residual variance compared to models without these effects (M1 and M11).

Overall, the residual analysis provides diagnostic evidence that including country-specific effects changes the distribution of prediction residuals, even when average predictive performance measured by RMSE is similar across models. The spatial residual analysis shows that model M2 exhibits reduced residual spatial dependence relative to model M1. In addition, models with country-specific intercepts and time trends (M2 and M21) display lower cross-country dispersion of residuals within a given year, smaller country-level underprediction, and lower within-country residual variance over time. At the same time, all model specifications show visible overprediction in later years (except for M21, which includes time both as an observed feature and as a country-specific effect). This suggests that these residuals are driven by common year effects that are not fully captured by the observed features or by country-specific time trends.

3.3.3. Explainability and feature attribution

This subsection analyses feature attributions in GPBoost using SHAP values. The analysis focuses on model M2 rather than model M21. Although model M21 shows slightly more stable residual behaviour in later validation years, it includes time as an explicit feature in the boosting component. Because time is not an observed economic or energy-related characteristic, its inclusion complicates the interpretation of SHAP values when the objective is to analyse associations between emissions and contemporaneous observed features. For this reason, SHAP analysis is conducted using model M2, which excludes time from the tree component.

Feature importance is first compared between models M1 and M2 in order to illustrate how the inclusion of explicit country-specific effects changes SHAP-based feature attribution. SHAP results for model M2 are then analysed in more detail.

3.3.3.1. Feature variation and its implication for SHAP interpretation

Before interpreting feature attributions, it is important to clarify how the observed features vary between countries and within countries over time**.** Table 8 presents between-country variance share for each observed feature.

Table 8. Between-country variance share of observed features

Feature	Between share
Final_Electricity_Share	0.966
Final_Gaseous_Fuels_Share	0.966
Heating_Degree_Days	0.965
Energy_Consumption_Per_Capita	0.958
Mining_and_Quarrying_Value_Added_Gross_CLV_Share	0.957
Gross_Nuclear_Share	0.927
Gross_Coal_Peat_Shale	0.886
Final_Coal_Peat_Shale_Share	0.884
Gross_Gaseous_Fuels_Share	0.872
Manufacturing_Value_Added_Gross_CLV_Share	0.866
Final_Liquid_Fuels_Share	0.865
Agriculture_Value_Added_Gross_CLV_Share	0.842
Final_Renewables_Biofuels_Share	0.820
Energy_Intensity_CLV	0.796
Gross_Renewables_Biofuels_Share	0.785
GDP_Per_Capita_PPP	0.675
EGSA_Value_Added_Gross_CLV_Share	0.663
Construction_Value_Added_Gross_CLV_Share	0.575
Gross_Liquid_Fuels_Share	0.494

For most features, most of the variation is attributable to differences between countries rather than temporal variation within countries. For example, 13 out of 19 features display a between-country variance share exceeding 0.8.

This structure of the input data implies that SHAP values primarily reflect how observed features distinguish higher-emitting countries from lower-emitting countries, rather than how emissions change within a country over time. As a result, SHAP values should be interpreted primarily as capturing cross-country differences in emission levels, rather than within-country changes over time.

3.3.3.2. SHAP values with and without country-specific effects

Models M1 and M2 differ in whether persistent country-specific heterogeneity in emission levels is modelled explicitly. In M1, this heterogeneity can only be reflected indirectly through the observed features, whereas in M2 it is captured by country-specific intercepts and time trends. This difference affects how the tree component allocates importance across features, which is reflected in the corresponding SHAP values.

To illustrate the differences in feature attribution described above, Fig. 9 compares SHAP-based feature importance between models M1 and M2.

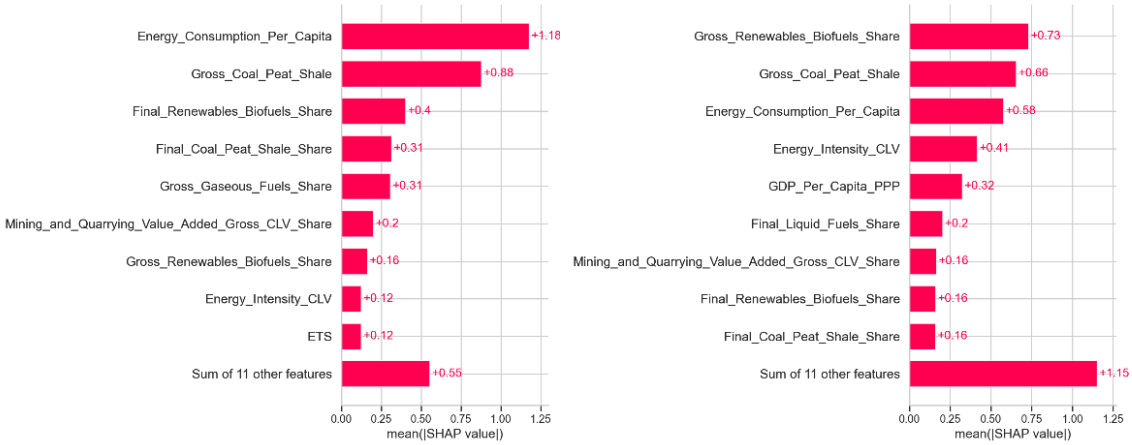


Fig. 9. SHAP-based feature importance for GPBoost models M1 (left) and M2 (right)

In model M1, feature importance is concentrated in a small number of variables, such as energy consumption per capita and gross coal, peat, and shale share. Together, these two features account for approximately 50% of total SHAP importance.

In model M2, energy consumption per capita and gross coal, peat, and shale share remain among the three most important predictors, but their SHAP contributions are smaller. At the same time, the gross renewables and biofuels share becomes the most important feature, whereas in model M1 it ranked seventh. Overall, feature importance in model M2 is more evenly distributed across variables: the two most important features are less dominant relative to the rest and they account for approximately 30% of total SHAP importance, compared to about 50% in M1. This pattern can also be seen in the combined contribution of less important features (those that are outside the top ten): in model M1, the sum of SHAP contributions for features outside the top ten is 0.55, compared to 1.15 in model M2.

3.3.3.3. Local SHAP

While the previous section compares feature attribution between M1 and M2 using aggregated SHAP values, this subsection shows how the same differences appear in individual country predictions. Fig. 10 shows local SHAP explanations for Germany (a persistent hot spot) and Spain (an intensifying cold spot) for the year 2022.

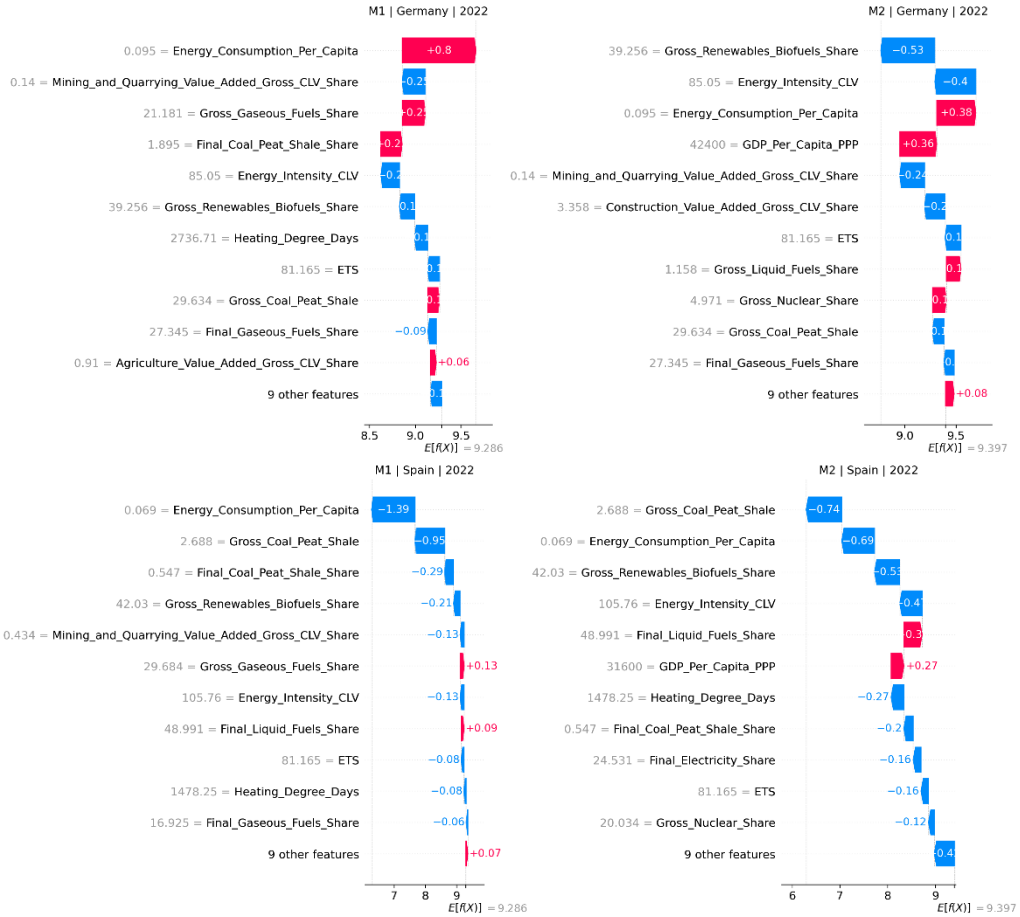


Fig. 10. Local SHAP explanations for Germany (top) and Spain (bottom) in 2022, comparing model M1 (left) and model M2 (right)

In model M1, predictions for both countries are dominated by energy consumption per capita, with at most one additional feature making a noticeable contribution, while remaining features play a minor role. In model M2, contributions are more evenly distributed across several features, rather than being dominated by a single variable. This mirrors the differences observed in aggregate SHAP results.

Overall, the comparison between M1 and M2 shows that SHAP-based feature importance depends on whether country-specific heterogeneity is modelled explicitly. By absorbing persistent country-level differences outside the tree component, model M2 reduces the extent to which observed variables act as proxies for country identity, leading to more balanced and interpretable SHAP explanations.

3.3.3.4. SHAP global importance

The remainder of this section focuses on interpreting the SHAP results for model M2. Fig. 11 shows the SHAP bee swarm summary plot for model M2. The plot shows the distribution of SHAP values for each feature across the test observations. The features are ordered by their average absolute SHAP value. The horizontal position indicates whether a feature contributes positively or negatively to the predicted emissions, while the colour reflects the feature value.

Fig. 11. SHAP bee swarm summary plot for GPBoost model M2

In model M2, the most important features fall into the main categories of emission drivers discussed in the literature. These include energy use (energy consumption per capita), energy efficiency (energy intensity), carbon intensity (gross renewables and biofuels share; gross coal, peat, and shale share), and economic affluence (GDP per capita). These categories also correspond to the components of the Kaya identity.

The beeswarm plot also shows that higher values of coal, peat and shale share, energy consumption per capita, and energy intensity are generally associated with higher predicted emissions, while higher values of gross renewables and biofuels share are associated with lower predicted emissions. GDP per capita does not display a clear directional association with predicted emissions once differences in energy use, efficiency, and carbon intensity are accounted for.

3.3.3.5. SHAP feature-level behaviour

Fig. 12 and Fig. 13 show SHAP dependence plots for the most important features identified in the global importance analysis. In each dependence plot, SHAP values are shown against feature values, with additional colouring by GDP per capita. The colouring is used to visualize how relationships between feature and prediction differ across countries with different income levels.

Fig. 12 shows SHAP dependence plots for energy consumption per capita and energy intensity.

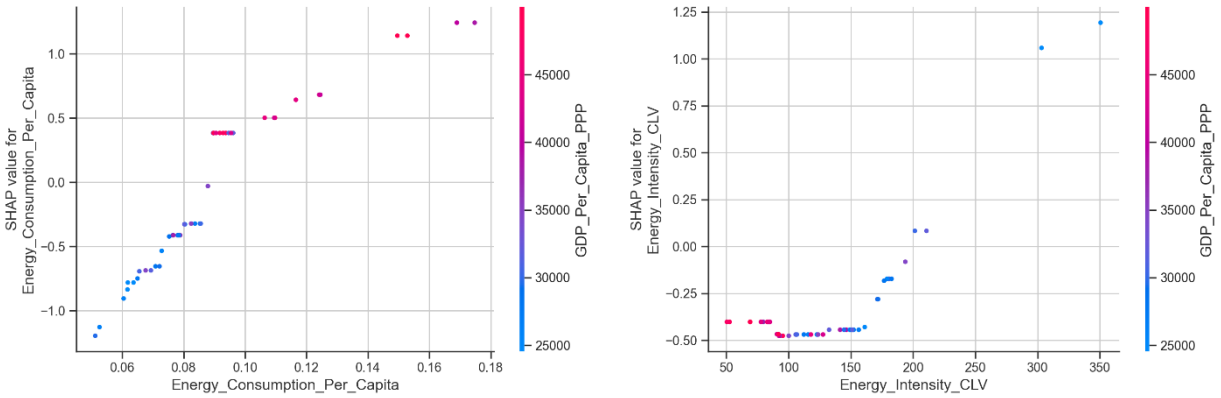


Fig. 12. SHAP dependence plots for energy consumption per capita (left) and energy intensity (right) in GPBoost model M2, with observations coloured by GDP per capita

These plots show visible differences in GDP per capita across the feature ranges. Higher values of energy consumption per capita are more frequently observed among countries with higher GDP per capita, while higher values of energy intensity are more frequently observed among countries with lower GDP per capita.

Fig. 13 shows SHAP dependence plots for energy source variables, including, gross renewables and biofuels share, gross coal, peat and shale share, and final liquid fuel share.

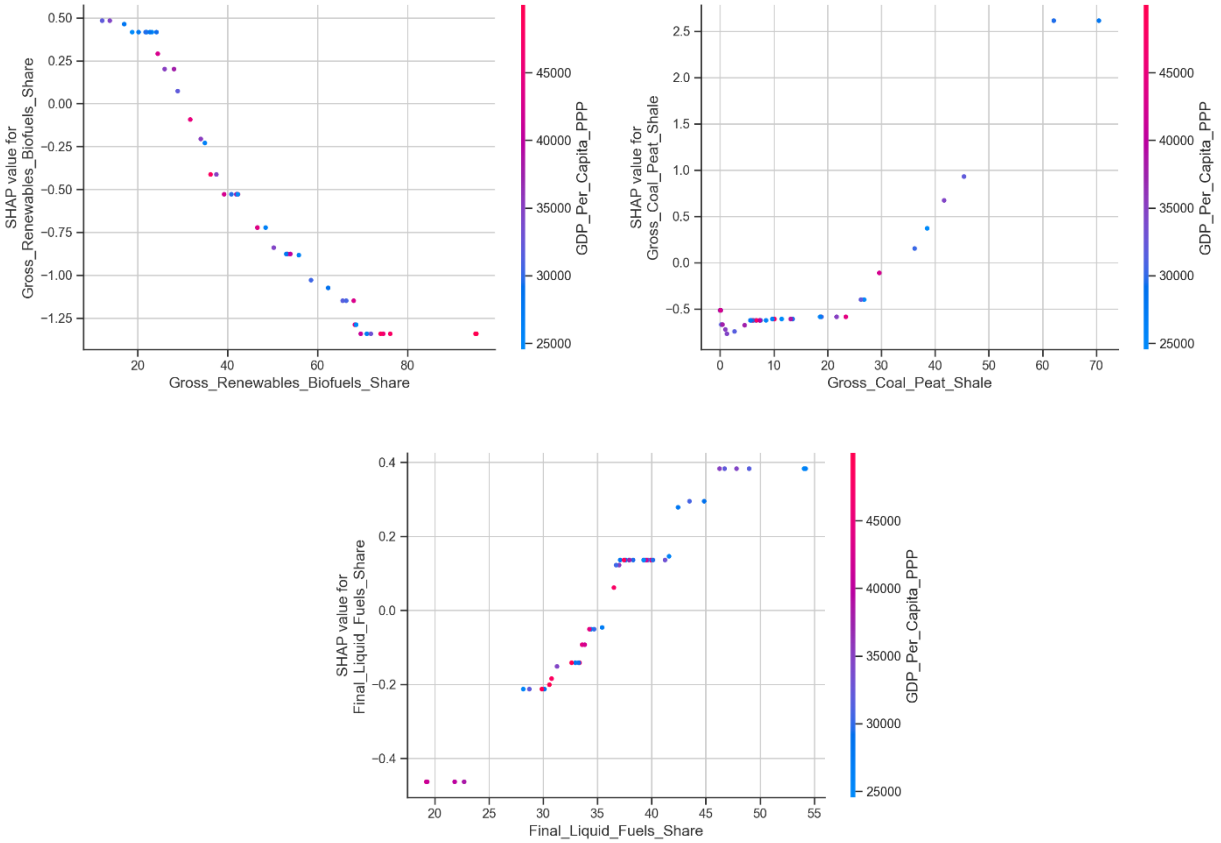


Fig. 13. SHAP dependence plots for gross renewables and biofuels share (top left), gross coal, peat and shale share (top right), and final liquid fuel share (bottom) in GPBoost model M2, with observations coloured by GDP per capita

For renewables and biofuels share, the GDP colouring is mixed across the feature range, with no clear pattern. For coal, peat, and shale use, higher SHAP values tend to occur more often for countries with lower GDP per capita. A similar but less pronounced pattern is observed for final liquid fuel share, where higher SHAP values are also more frequently associated with lower GDP per capita.

In summary, the SHAP analysis shows that the GPBoost model assigns the largest importance on energy consumption per capita, energy intensity, and energy production fuel mix variables, particularly coal and renewables shares. GDP per capita has a smaller individual contribution once these energy-related features are included.

Conclusions

1. Exploratory spatial and spatiotemporal analysis reveals that GHG emissions decline across most European countries, but spatial clustering persists, even though the location of clusters may change over time. In the early years, statistically significant ($p < 0.05$) hot spots are observed in both Northern and Central Europe, while in later years they are concentrated mainly in Central Europe. At the same time, a cold-spot cluster emerges in Southwestern Europe. These results indicate that relative differences in emission levels across European regions persist and that spatial clustering in emissions remains despite overall declines in GHG emissions.
2. Predictive accuracy depends on how cross-country heterogeneity is treated, particularly in linear models. The pooled linear model, which does not include country-specific effects and therefore does not account for persistent country-level heterogeneity, exhibits large prediction errors (final test RMSE is 1.44), while linear mixed-effects model with country specific intercepts and time trends performs clearly better (final test RMSE is 0.43). In contrast, GPBoost models achieve low RMSE both with and without country-specific structure (final test RMSE is 0.49 and 0.52, respectively). This indicates that flexible machine learning models can achieve good predictive performance even without explicitly modelling country-level heterogeneity.
3. Residual diagnostics reveal differences between GPBoost specifications that are not captured by RMSE alone. While average predictive accuracy is similar across GPBoost models, GPBoost specifications with country-specific intercepts and time trends show reduced residual spatial autocorrelation, lower cross-country dispersion of residuals within a given year, decreased systematic country-level prediction bias, and reduced within-country residual variance over time. These results indicate that GPBoost models with explicit country-specific effects more effectively account for persistent cross-country heterogeneity.
4. SHAP-based feature attribution in GPBoost models is also sensitive to whether country-level heterogeneity is modelled explicitly. In GPBoost specifications without country-specific effects, SHAP importance is concentrated in a small number of variables (with the top two accounting for roughly half of total mean absolute SHAP), whereas including country-specific intercepts and time trends reduces this concentration (with the top two accounting for around one third) and leads to more evenly distributed feature importance.

Limitations

1. Most explanatory variables vary much more between countries than within countries over time. For 13 out of 19 features, more than 80% of the total variation is attributable to between-country differences. As a result, both the GPBoost models and the SHAP analysis primarily capture cross-country differences in emissions levels rather than within-country dynamics. This limits the ability of the analysis to explain how changes in energy mix, efficiency or economic structure within individual countries affect emissions over time.
2. The analysis is conducted at the country level, which requires aggregating emissions and explanatory variables to national averages. This masks within-country variation in emissions, energy use, and economic structure. As a result, the identified spatial patterns describe similarities and differences between countries, rather than finer-scale differences within individual countries.
3. The models describe statistical associations between emissions and observed features but do not identify causal effects of, for example, policy interventions or technological change. SHAP values explain how features contribute to model predictions within a given model specification and therefore do not provide estimates of the causal impact of changing individual variables.

List of references

1. IPCC. Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 2022. Available from: https://doi.org/10.1017/9781009157926.001.
2. YANG, Bingduo; Wei LONG; Zongwu CAI. Machine Learning Based Panel Data Models. Preprint. University of Kansas, Department of Economics, 2024. Available from: https://ideas.repec.org/p/kan/wpaper/202402.html. [viewed 2025-12-02].
3. FILONCHYK, Mikalai; Michael P. PETERSON; Lifeng ZHANG; Volha HURYNOVICH. *Greenhouse gases emissions and global climate change: Examining the influence of CO_2, CH_4, and N_2O.* Science of the Total Environment, 2024, 935, 173359. Available from: https://doi.org/10.1016/j.scitotenv.2024.173359.
4. IPCC. Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 2021. Available from: https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/.
5. EUROSTAT. *Domestic net greenhouse gas emissions. Eurostat Data Browser*. Web site. Available from: https://ec.europa.eu/eurostat/databrowser/view/sdg_13_10__custom_19529893/default/table. [viewed 2025-10-18].
6. TOBLER, Waldo R. *A computer movie simulating urban growth in the Detroit region*. Economic Geography, 1970, 46, 234-240..
7. ORD, J. Keith and GETIS, Arthur. *Local spatial autocorrelation statistics: Distributional issues and an application*. Geographical Analysis, 1995, 27(4), 286-306. Available from: https://doi.org/10.1111/j.1538-4632.1995.tb00912.x.
8. ANSELIN, Luc. *Local indicators of spatial association-LISA. Geographical Analysis*, vol. 27 (1995), no. 2, pp. 93-115. Available from: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
9. GETIS, Arthur and ORD, J. Keith. *The analysis of spatial association by use of distance statistics*. Geographical Analysis, 1992, 24(3), 189-206. Available from: https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.
10. GETIS, Arthur and ALDSTADT, Jared. *Constructing the spatial weights matrix using a local statistic*. Geographical Analysis, 2004, 36(2), 90-104. Available from: https://doi.org/10.1111/j.1538-4632.2004.tb01127.x.
11. GETIS, Arthur. *Spatial weights matrices*. Geographical Analysis, 2009, 41(4), 404-410. Available from: https://doi.org/10.1111/j.1538-4632.2009.00768.x.
12. GRIFFITH, Daniel A. *Spatial autocorrelation and eigenfunctions of the geographic weights matrix*. The Canadian Geographer / Le Géographe canadien, 1996, 40(4), 351-367. Available from: https://doi.org/10.1111/j.1541-0064.1996.tb00462.x.
13. FLORAX, Raymond J. G. M. and REY, Serge. *The impacts of misspecified spatial interaction in linear regression models*. In: ANSELIN, Luc and FLORAX, Raymond J. G. M. (eds.). New Directions in Spatial Econometrics. Berlin: Springer, 1995, pp. 111-135.
14. TIEFELSDORF, Michael; GRIFFITH, Daniel A. and BOOTS, Barry. *A variance-stabilizing coding scheme for spatial link matrices*. Environment and Planning A, 1998, 30(1), 165-180.
15. CLIFF, Andrew D. and ORD, J. Keith. *Spatial autocorrelation*. London: Pion, 1973.

16. ESRI. *Emerging Hot Spot Analysis (Space Time Pattern Mining)*. Web site. Available from: https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/emerginghotspots.htm. [viewed 2025-12-11].
17. RIOS, Vicente and Lisa GIANMOENA. *Convergence in CO2 emissions: A spatial economic analysis with cross-country interactions*. Energy Economics, 2018, 75, 222-238. Available from: https://doi.org/10.1016/j.eneco.2018.08.009.
18. ATALAY, Ayse C. and Yusuf AKAN. *The spatial analysis of green economy indicators of OECD countries*. Frontiers in Environmental Science, 2023, 11, 1-19. Available from: https://doi.org/10.3389/fenvs.2023.1243278.
19. WANG, Jing-he; Jamal MAMKHEZRI; Mohsen KHEZRI; Mohammad Sharif KARIMI; Yousaf Ali KHAN. Insights from European nations on the spatial impacts of renewable energy sources on CO_2 emissions. Energy Reports, 2022, 8, 5620-5630. Available from: https://doi.org/10.1016/j.egyr.2022.04.005.
20. WANG, Bin; Zheng QIUXIA; Ao SUN; Jie BAO; Dianting WU. *Spatio-temporal patterns of CO_2 emissions and influencing factors in China using ESDA and PLS-SEM*. Mathematics, 2021, 9(21), 2711. Available from: https://doi.org/10.3390/math9212711
21. VAGNINI, Chiara; Leticia Canal VIEIRA; Mariolina LONGO, Matteo MURA. *Chasing net zero: An exploratory space-time analysis of European regions' industrial carbon emissions*. Journal of Environmental Management, 2025, 391, 126466. Available from: https://doi.org/10.1016/j.jenvman.2025.126466
22. XIA, Lei; Rui LIU; Wenxuan FAN; Changxu REN. *Emerging carbon dioxide hotspots in East Asia identified by a top-down inventory*. Communications Earth and Environment, 2025, 6, 10. Available from: https://doi.org/10.1038/s43247-024-01991-7
23. ROSA, Eugene A. and DIETZ, Thomas. *Human drivers of national greenhouse-gas emissions*. Nature Climate Change, 2012, 2, 581-586. Available from: https://doi.org/10.1038/nclimate1506.
24. KAYA, Yoichi. Impact of carbon dioxide emission control on GNP growth: interpretation of proposed scenarios. *IPCC Energy and Industry Subgroup, Response Strategies Working Group*. Paris, 1990.
25. HWANG, Youngseok; Jung-Sup UM; Junhwa HWANG; Stephan SCHLÜTER. *Evaluating the causal relations between the Kaya identity index and ODIAC-based fossil fuel CO_2 flux*. Energies, 2020, 13(22), 6009. Available from: https://doi.org/10.3390/en13226009.
26. HOLDREN, John P. *A brief history of "IPAT"*. Journal of Population and Sustainability, 2016, 2(2), 66–74. Available from: https://doi.org/10.3197/jps.2018.2.2.66.
27. EHRLICH, Paul R. and John P. HOLDREN. *Critique*. Bulletin of the Atomic Scientists, 1972, 28(5), 16–27. Available from: https://doi.org/10.1080/00963402.1972.11457930.
28. YORK, Richard; Eugene A. ROSA; Thomas DIETZ. *STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts*. Ecological Economics, 2003, 46(3), 351–365. Available from: https://doi.org/10.1016/S0921-8009(03)00188-5.
29. DIETZ, Thomas and Eugene A. ROSA. *Rethinking the environmental impacts of population, affluence and technology*. Human Ecology Review, 1994, 1(2), 277–300.
30. ANG, B.W. and F.Q. ZHANG. *A survey of index decomposition analysis in energy and environmental studies*. Energy, 2000, 25(12), 1149-1176. Available from: https://doi.org/10.1016/S0360-5442(00)00039-6.

31. ANG, B.W. *LMDI decomposition approach: A guide for implementation*. Energy Policy, 2015, 86, 233-238. Available from: https://doi.org/10.1016/j.enpol.2015.07.007.
32. ANG, B.W. *Decomposition methodology in industrial energy demand analysis: The energy intensity approach*. Energy economics, 1994, 16(3), 163-174. Available from: https://doi.org/10.1016/0140-9883(94)90030-2.
33. ATHEY, Susan and Guido W. IMBENS. *Machine learning methods that economists should know about*. Annual Review of Economics, 2019, 11, 685-725. Available from: https://doi.org/10.1146/annurev-economics-080217-053433.
34. BLAKELY, Tony; John LYNCH; Koen SIMONS; Rebecca BENTLEY and Sherri ROSE. *Reflection on modern methods: when worlds collide-prediction, machine learning and causal inference*. International Journal of Epidemiology, 2019, 49(6), 2058-2064. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8453374/. [viewed 2025-12-14].
35. SHMUELI, Galit. *To explain or to predict?.* Statistical Science, 2010, 25(3), 289-310. Available from: https://doi.org/10.1214/10-STS330.
36. MULLAINATHAN, Sendhil and Jann SPIESS. *Machine learning: an applied econometric approach*. Journal of Economic Perspectives, 2017, 31(2), 87-106. Available from: https://doi.org/10.1257/jep.31.2.87.
37. BECKERS, Sander. *Causal explanations and XAI*. arXiv preprint, 2022. Available from: https://doi.org/10.48550/arXiv.2201.13169
38. COUPLAND, Helen et al. *Exploring the potential and limitations of deep learning and explainable AI for longitudinal life course analysis*. BMC Public Health, 2025, 25, 1520. Available from: https://link.springer.com/article/10.1186/s12889-025-22705-4. [viewed 2025-12-14].
39. HSIAO, Cheng. *Analysis of Panel Data*. 3rd ed. Cambridge: Cambridge University Press, 2014.
40. STERN, David I. *The rise and fall of the Environmental Kuznets Curve*. World Development, 2004, 32(8), 1419-1439. Available from: https://steadystate.org/wp-content/uploads/Stern_KuznetsCurve.pdf. [viewed 2025-12-08].
41. POUMANYVONG, Phetkeo and Shinji KANEKO. *Does urbanization lead to less energy use and lower CO_2 emissions? A cross-country analysis*. Ecological Economics, 2010, 70(2), 434-444. Available from: https://doi.org/10.1016/j.ecolecon.2010.09.029.
42. SHARMA, S. *Determinants of carbon dioxide emissions: empirical evidence from 69 countries*. Applied Energy, 2011, 88(1), 376-382. Available from: https://doi.org/10.1016/j.apenergy.2010.07.022.
43. KAIS, Saidi and Sami HAMMAMI. *An econometric study of the impact of economic growth and energy use on carbon emissions: Panel data evidence from fifty-eight countries*. Renewable and Sustainable Energy Reviews, 2016, 59, 1101-1110. Available from: https://doi.org/10.1016/j.rser.2016.01.054.
44. COSTANTINI, Lorenzo; Francesco LAIO; Manuel S. MARIANI; Luca RIDOLFI; Carla SCIARRA. *Forecasting national CO2 emissions worldwide*. Scientific Reports, 2024, 14, 22438. Available from: https://doi.org/10.1038/s41598-024-73060-0.
45. LI, Yi; Yujuan HE; Fei YANG; Hongyu AN; Jiayu LI; Yizhang XIE. *An XGBoost-SHAP analysis of the driving factors of carbon emissions in China's first-tier cities*. Scientific Reports, 2025. Available from: https://doi.org/10.1038/s41598-025-31260-2.

46. HOU, Yunfei and Shouwei LIU. Predictive Modelling and Validation of Carbon Emissions from China's Coastal Construction Industry: A BO-XGBoost Ensemble Approach. Sustainability, 2024, 16, 4215. Available from: https://doi.org/10.3390/su16104215.
47. YU, Wenmei; Lina XIA; Qiang CAO. *A machine learning algorithm to explore the drivers of carbon emissions in Chinese cities*. Scientific Reports, 2024, 14, 23609. Available from: https://doi.org/10.1038/s41598-024-75753-y.
48. DONG, Zhaoyingzi; Jiayan SHI; Sheng PAN. The interactions of carbon emission driving forces: Analysis based on interpretable machine learning. Urban climate, 2025, 59, 102323. Available from: https://doi.org/10.1016/j.uclim.2025.102323.
49. DONG, Yuling and Lili GU. *Can policy-based agricultural insurance promote agricultural carbon emission reduction? Causal inference based on double machine learning*. Sustainability, 2025, 17(9), 4086. Available from: https://doi.org/10.3390/su17094086.
50. TANG, Weixin and Qihao ZHOU. *Can green financial policy drive urban carbon unlocking efficiency? A causal inference approach based on double machine learning*. Frontiers in Environmental Science, 2025, Article 1608475. Available from: https://doi.org/10.3389/fenvs.2025.1608475.
51. SIGRIST, Fabio. *Gaussian process boosting. Journal of Machine Learning Research*, 2022, 23, 1-46. Available from: https://www.jmlr.org/papers/v23/20-322.html [viewed: 2025-11-25].
52. MOLNAR, Christoph. Interpretable Machine Learning. 2nd ed. Online book, 2022. Available from: https://christophm.github.io/interpretable-ml-book.
53. LUNDBERG, Scott M.; Gabriel, ERION; Hugh CHEN; Alex DEGRAVE; Jordan M. PRUTKIN. *From local explanations to global understanding with explainable AI for trees*. Nature Machine Intelligence, 2020, 2, 56-67.
54. WU, J.; P. FENG; T. LI; H. KUANG. *Spatiotemporal analysis of built environment restrained traffic carbon emissions and policy implications*. Transportation Research Part D: Transport and Environment. 2023, 121, 103839. Available from: https://doi.org/10.1016/j.trd.2023.103839.
55. ARUNKUMAR, K. E.; Nathan E. BLAKE; Matthew WALKER; Tylor J. Yost; Domingo Mata PADRINO et al. *Predicting dry matter intake in cattle at scale using gradient boosting regression techniques and Gaussian process boosting regression with Shapley additive explanation explainable artificial intelligence, MLflow, and its containerization*. Journal of Animal Science. 2025, 103, skaf041. Available from: https://doi.org/10.1093/jas/skaf041.
56. LUNDBERG, Scott M. and Su-In LEE. *A unified approach to interpreting model predictions*. 2017. In: I. GUYON et al. (eds.). Advances in Neural Information Processing Systems 30. p. 4766-477. Red Hook, NY: Curran Associates, 2017.
57. EUROSTAT. *Statistical database. European Commission*. Available from: https://ec.europa.eu/eurostat. [viewed 2025-11-12].
58. INVESTING.COM. Carbon Emissions Futures (EU ETS) historical data. Web site. Available from: https://www.investing.com/commodities/carbon-emissions-historical-data. [viewed 2025-11-14].
59. EUROSTAT. *Energy in Europe – 2025 edition*. Web site. Available from: https://ec.europa.eu/eurostat/web/interactive-publications/energy-2025. [viewed 2025-11-13].
60. OPTUNA. *optuna.samplers.TPESampler*. Web site. Available from: https://optuna.readthedocs.io/en/stable/reference/samplers/generated/optuna.samplers.TPESampler.html. [viewed 2025-12-12]

Appendices

Appendix 1. Equivalence of permutation-based inference for G_i and G_i^*

This appendix shows that, when permutation-based inference with conditional randomization is used (where the observed value at location i is held fixed), the local Getis-Ord statistics G_i and G_i^* yield identical standardized permutation statistics and permutation-based p values for a given location. That is, the inclusion of the focal country's emission value in G_i^* does not affect statistical inference.

Let y_i denote the observed GHG emissions for a country i, let $N(i)$ denote the set of its neighbouring countries, and let

$$S_i = \sum_{j \in N(i)} y_j$$

denote the sum of emissions in the neighbourhood of country i. Let

$$T = \sum_{j=1}^{n} y_j$$

denote total emissions across all n countries.

The local Getis-Ord statistics are then defined as

$$G_i = \frac{S_i}{T - y_i}, \qquad G_i^* = \frac{S_i + y_i}{T}$$

Using these definitions, G_i^* can be rewritten as

$$G_i^* = \frac{S_i + y_i}{T} = \frac{(T - y_i)G_i + y_i}{T} = \frac{T - y_i}{T} G_i + \frac{y_i}{T} = a_i G_i + b_i$$

where

$$a_i = \frac{T - y_i}{T} > 0, \qquad b_i = \frac{y_i}{T} > 0$$

When permutations are computed for location i, the observed value y_i for that location and total sum T remain fixed. The only quantity that varies across permutations is the neighbour sum

$$S_i^{(r)} = \sum_{j \in N(i)} y_j^{(r)}$$

Therefore, for every permutation r,

$$G_i^{*(r)} = a_i G_i^{(r)} + b_i$$

Let μ_i^{perm} and σ_i^{perm} denote the mean and standard deviation of the permutation distribution of G_i. The corresponding mean and standard deviation for G_i^* then satisfy

$$\mu_i^{*,perm} = a_i \mu_i^{perm} + b_i, \qquad \sigma_i^{*,perm} = a_i \sigma_i^{perm}$$

Substituting these expressions into the standardized permutation statistic for G_i^* gives

$$z_{sim}(G_i^*) = \frac{G_i^* - \mu_i^{*\,perm}}{\sigma_i^{*\,perm}} = \frac{a_i G_i + b_i - \left(a_i \mu_i^{perm} + b_i\right)}{a_i \sigma_i^{perm}} = \frac{G_i - \mu_i^{perm}}{\sigma_i^{perm}} = z_{sim}(G_i)$$

Because converting G_i into G_i^* rescales the statistic and shifts it by a constant, this transformation preserves the comparison between each permuted statistic and the observed statistic:

$$G_i^{(r)} \geq G_i \iff G_i^{*(r)} \geq G_i^*$$

Therefore, when the same permutations are used for both G_i and G_i^*, the number of permuted values that are at least as extreme as the observed value are identical. As a result, the permutation-based p values are equal:

$$p_{sim}(G_i^*) = p_{sim}(G_i)$$

Appendix 2. Tuned GPBoost hyperparameters (Optuna best trial)

Table 9. Optimal hyperparameter values selected by Optuna for each GPBoost model specification

Hyperparameter	**M1**	**M11**	**M2**	**M21**
learning_rate	0.015	0.043	0.011	0.026
num_leaves	185	69	68	60
max_depth	2	2	1	1
min_data_in_leaf	13	17	25	33
feature_fraction	0.868	0.879	0.520	0.883
lambda_l1	4.18×10^{-4}	3.52×10^{-7}	1.68×10^{-6}	1.32×10^{-3}
lambda_l2	7.02×10^{-5}	8.97×10^{-6}	6.43×10^{-2}	3.86×10^{-6}
min_gain_to_split	0.049	0.053	0.047	0.787
bagging_fraction	0.611	0.830	1	1
bagging_freq	4	1	0	0