Proceedings of the 3rd International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2025.GL27830

Structure-Aware Machine Learning for National CO_2 Emission Composition: A Classification Framework

Jiahui Liu

Department of Data Science and Analytics, Georgetown University, Washington, USA
jl3199@georgetown.edu

Abstract. CO_2 structure. The sectoral distribution of national CO_2 emissions is a necessary basis for structural mitigation options and transformation pathways. This paper introduces a structure-based machine learning architecture designed to work with patterns of emission composition rather than the absolute emission trend. We construct normalized sectoral vectors that capture the relative weight of power, industry, transport, residential, and aviation, using data from Carbon Monitor (2019–2025). We perform a PCA to reduce dimensionality and explore the variability of the structure among countries and years. K-Means clustering is subsequently employed to identify archetypal structural types across country-year pairs, and a Support Vector Machine (SVM) classifier is trained to learn the separation boundaries for these types, enabling classification and visual interpretation of the data. We also investigated K-Nearest Neighbors (KNN) for measuring the structural nearest neighbor. Still, KNN was practically useless due to the invariance of the normalized vectors, which often resulted in multiple self-nearest neighbor matches. Yet, the global framework successfully discovers structure-based emission categories and traces structural changes over time. The method can be applied to different scales in space and time and used as a diagnostic tool for monitoring the emission structure, analyzing the type, and designing policy. The combination of unsupervised and supervised learning schemes brings in analytic power as well as the interpretability of a policy.

Keywords: CO_2 structure, Machine learning, PCA, SVM, Classification.

1. Introduction

In the context of global efforts to rein in greenhouse gas emissions, at the national level, mapping carbon dioxide (CO_2) emissions remains one of the essential cornerstones for climate science and policy design. Similar to the traditional approach, which utilizes aggregate indicators such as GDP, population, and energy intensity to identify trends in emissions, this work demonstrates how the interplay between three-part plays at the point is partitioned into productive chain links [1-3]. However, such top-down views do not account for the inner structure of emissions, which varies significantly from country to country and also changes over time. Segmenting emissions in the economy by power generation, industry, transport, residential use, and aviation reveals deeper systemic features, including infrastructure age, the percentage of energy imported, and urbanization levels. These characteristics determine both a country's emissions profile and the type of mitigation

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.GL27830

measures that are feasible and in which direction they can be directed. For example, an economy dominated by industry may require separate policy instruments and transition pathways compared to one where buses and subways account for a higher share of emissions than homes in townships. A few recent studies are beginning to explore the clustering-based identification of energy or emission profiles, particularly within the framework of regional types and temporal stages [4-6]. However, most of these efforts are limited to static unsupervised clustering, and they still lack integration with models under supervision, or do not indicate in terms of image interpretation where the boundary of the emission structure lies. Also, few studies systematically link structure classification with mapping trajectories in time, leaving a gap regarding how national emission patterns evolve. To address these limitations, we propose that by basing the machine learning system on the structure analysis of layers of emission vectors in different parts, it is possible to judge the form classification and boundaries of structural types, as well as transition pathways. We incorporate principal component analysis (PCA) to reduce dimensionality [7], using K-means clustering, an unsupervised method to find patterns in the data [8], and support vector machines (SVM) for classification [9]. We also examine K-nearest neighbors (KNN) for legal neighbor analysis [10]; however, research indicates that KNN performs poorly under conditions of high covariance in data sets. Such a framework highlights analysis results and at the same time serves as an easy-to-use tool for visual interpretation of structures in emissions trends that policy-makers might want to know about.

2. Methodology

To examine the emissions structure of national CO_2 at an interpretable and scalable level, we propose a multi-stage machine learning approach featuring compositional vectorization, dimensionality reduction, unsupervised clustering, and supervised classification. Each stage is tailored to process different perspectives of the emission structure landscape—from pattern recognition to boundary learning—whilst allowing for visual and analytical inspection. The core of our method includes three primary methods (Principal Component Analysis (PCA), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) algorithm) for different kinds of analysis. This is the methodology of this research paper.

2.1. Sectoral emission structure construction

Each country-year sample is represented as a six-dimensional vector comprising the relative shares of emissions from power, industry, transport, residential use, domestic aviation, and international aviation. These values are normalized to form a unit vector, then standardized (Z-score) to emphasize structural differences. The resulting dataset is a matrix of standardized sector proportions across time and countries.

2.2. Structural mapping and typology formation via PCA and K-Means

Once the standardized emission structure vectors have been constructed for each country-year observation, we use Principal Component Analysis (PCA) to transform the original six-dimensional area into two orthogonal components that retain much of its structural variances. This transformation simplifies the data and allows us to interpret it meaningfully in terms of underlying structure types.

We need to understand that weightings of each sector on principal components reflect trade-offs in national emission profiles, that is, how many emissions come from each sector as opposed to another -- for instance, the first component often represents an opposition between industrial and

DOI: 10.54254/2753-8818/2025.GL27830

residential shares. In contrast, the second captures variations connected to other characteristics like transport or aviation.

To obtain a set of structural patterns representative of the K-means clustered and PCA reduced space, we put into effect the K-Means algorithm. This unsupervised learning method divides observations into K groups while minimizing sample point distances from their assigned cluster center. Since K is a hyperparameter that should be determined experimentally, we make use of an elbow plot to evaluate the trade-off between model simplicity and internal cohesion. The plot, as shown in Figure 1, shows the rate of decrease in the within-cluster sum of squares (inertia) as K grows. An inflection point occurs at K = 3, after which extra clusters only marginally increase the compactness of the data. This means that a three-cluster solution optimally combines differentiation and interpretability.

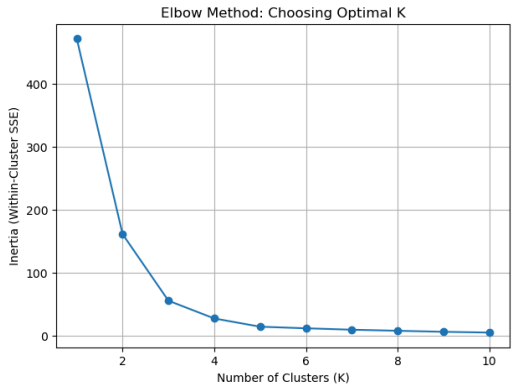


Figure 1. Elbow method plot

Every single cluster that thus emerged now represents a particular model of emission structure. The first group has a high percentage of industrial sector emissions and GNP. This is generally associated with heavy manufacturing economies. Cluster two focuses on increased proportions of emissions coming from travel, indicating the presence of highly mobile systems. Cluster three exhibits more evenly distributed profiles or a residential-weighted structure, perhaps indicative of advanced economies with decarbonized industries but a stable population sector.

These clusters are mapped onto the first two principal components of PCA space, where their spatial distributions show both arthrometric reformate divergence between populations in time-1 of the same country. In particular, countries may move through cluster boundaries over time, dynamically reshaping their emission structure properties around such phases.

Using PCA combined with the K-Means method brings out an understandable, capsule-like, data-driven representation of national emission systems. PCA sharpens both visual and analytical understanding, while K-Means extracts stable groupings embodying fundamental structural commonalities. This typology lays the groundwork for subsequent classification and comparison.

2.3. Structural classification via Support Vector Machine (SVM)

K-means clustering enables country-year observations with similar structures to form aggregated areas. Without a doubt, K-means clustering has limitations. Meanwhile, it can't be explicitly defined where one group's boundaries end and another's begin. This limitation makes it difficult to separate new data points and leaves us without a measure for how well two types are distinct. To address this, we introduce a supervised learning step: namely, Support Vector Machines (SVM). This enables

DOI: 10.54254/2753-8818/2025.GL27830

precise decision boundaries derived from the results of clustering by using a learning algorithm that meets certain conditions.

Support Vector Machines (SVM) is a margin-based classifier. It attempts to find the optimal hyperplane separating categories in a given feature space. It does so by maximizing the distance (also known as 'margin') between itself as a hyperplane and its nearest neighbors from each class of training data points. This approach is particularly suitable for high-dimensional or linearly separable data, as it ensures both classification accuracy and generalization. Furthermore, SVM can be extended by using kernel functions. It is capable of catching non-linear boundaries, and in our case, where PCA-reduced two-dimensional space enables effective linear separation.

In this study, the country-year coordinates in the PCA space are initially used as input features for the SVM classifier, with cluster labels derived from K-means as target classes. The classifier learns to map specific areas of the PCA space onto different structural types. The resultant decision boundaries are all projected onto the same two-dimensional plane and so provide a form of visual proof, dividing emissions structures. Each boundary represents the maximum-margin partition between two different structural regimes, giving geometric and analytical determination of the transition zones.

The addition of SVM provides a critical layer of structure to our typology. While K-means generates cluster membership assignments without defined boundaries, SVM renders these boundaries reproducible from one sample to another and from one time period to the next—in our example, on both country seconds (country-years) as well as third world nations. Thus, it enables a better understanding not only of the spatial location of structural types within our model but also clarifies how to categorize new data points that do not belong to any known category (in terms of their position in PCA space). This also allows us to study cases in which countries make transitions between regions clustering and analyze the dynamics or driving forces behind structural change.

2.4. Supplementary analysis using K-Nearest Neighbors (KNN)

We've distinguished different kinds of structures from one another and know what their boundaries look like. Our next primary step in this study is to determine how closely country-year emission structures align and whether they are mutually comparable. That is turning out to be increasingly tricky for K-Means. PCA distributes these boundary points. Nonetheless, K-Means reveals no information as to whether or not two observations themselves are close structurally. Nor does PCA possess this ability directly, self-contained within its method of creation.

Therefore, to get a closer understanding and quantify similarities between structural types of data from two or more places on Earth, we introduce K-Nearest Neighbors (KNN) as a measure of structural pattern similarity.

KNN is a non-parametric, instance-based algorithm that classifies or associate observations by comparing their similarities with nearby examples in feature space. Unlike global models such as SVM, which learn decision boundaries, KNN works locally: you might say it is hunting within a limited range. Rather than looking at all observations to decide where a target one belongs, KNN identifies the “K” closest points as neighbors—according to a specific distance metric—and decides similarity or class membership based on them. In our case, KNN is an additional method for measuring the proximity of one emission structure profile to another.

We can use KNN to do this by applying the KNN approach to the original sets of PCA-reduced vectors we observed. Instead of Euclidean distance, indices such as cosine similarity are used as distance measures. This measures the angle between two vectors in space. We have chosen cosine similarity because of the way our data is structured. In particular, we observe that the original

DOI: 10.54254/2753-8818/2025.GL27830

vectors represent a sector's fraction of emissions, which is why each vector has different proportions. Constellation meaning becomes more important than its size when considering relative relationships among vectors. Cosine similarity, therefore, is ideally suited to distinguishing similar things, no matter how big or small they are.

KNN adds nuance and depth to the research framework. It also allows us to identify structural analogs in space and time, which could be particularly interesting for policymakers seeking best practices that others have successfully used.

3. Result

3.1. Structural clustering in PCA space

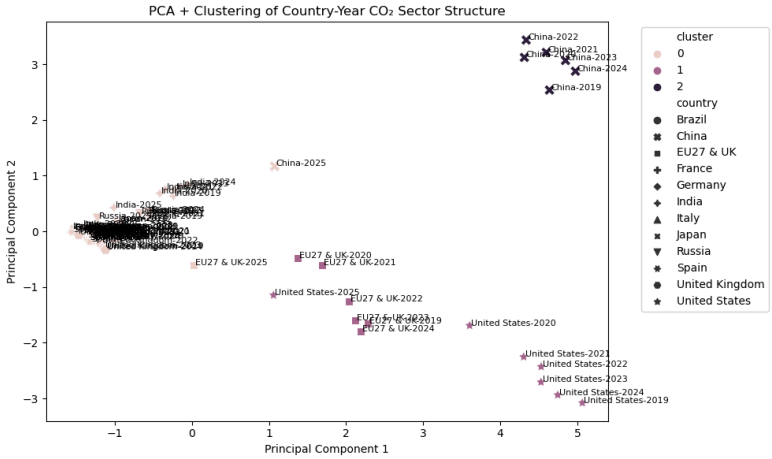


Figure 2. PCA+clustering of country-year CO_2 sector structure

Figure 2 illustrates the two-dimensional projection of all country-year observations using Principal Component Analysis (PCA). This projection consolidates data in a single place, with buildings organized by the standard emission components most commonly found in each location—development vectors span the period 2019-2025. Although each point on the plot represents a unique country-year pair and is thus arranged according to its principal component scores, it does not tell us anything about the set or range of years.

K-Means was employed in the unsupervised clustering of all this PCA-transformed data. A tight cluster appears repeatedly when it is a question of r-squared distribution with ranges large relative to their averages. Cluster 0 consists of observations with relatively high shares of emissions from the residential and transport sectors. Emerging or service-oriented economies, where industrially produced output is not as significant a share of national production as in advanced-industrial nations. More energy is distributed among smaller industries, and energy use is more diffused into other areas. Cluster 1 includes observations with sectoral shares better balanced; at least it suggests transitional or diversified emission structures typical of economies undergoing change or maintaining a balance among major sectors. It is notorious for its substantial, heavy energy-consuming industrial bases and often scores exceptionally high in terms of total emissions. Cluster 2 captures emission structures dominated by industrial and power sector activity. This is seen frequently in countries with a high output level of substantial energy use.

The arrangement of countries in PCA space reveals both cross-sectional heterogeneity and longitudinal trends. Notably, China occupies a consistent position in Cluster 2 from 2019 to 2024,

DOI: 10.54254/2753-8818/2025.GL27830

which means it had a highly industrialized structure of emissions during this period. However, in 2025, China shifts toward Cluster 0, suggesting a significant structural reconfiguration characterized by increasing reliance on residential and transport-related sectors. This movement may be an early sign of the effects energy transition policies are beginning to have, as well as a wider restructuring.

By contrast, the United States follows a stable path, remaining within Cluster 1 throughout the entire 2019–2025 period. This represents a stable internal structure amongst its major sectors, without any changes in composition of the kind visible at point 30 on the y-axis. The EU27 & UK group displays slight shifts across clusters, as it frequently flips back and forth between Clusters 0 and 1. This may in part reflect their dual emphasis on both reducing emissions and spreading responsibility for environmental effects among sectors.

These patterns, in turn, underscore the power of the PCA and K-Means framework to grab both strong systemic distinctions and subtle transitions. By combining both techniques, one can simultaneously read out commonalities in how different economies go about generating emissions, the persistence of particular structures from period to period itself, and trajectories that take distinctive routes through various nations over time.

3.2. Classification boundaries via support vector machine

For a more detailed interpretation of structural types and an evaluation of the type of transition boundaries between them, we used a support vector machine (SVM) classifier that used principal component coordinates and cluster labels obtained from K-Means. The purpose of this step is to produce a reproducible and geometrically reasonable partitioning of the structural space such that, given a sample, we retrieve 'how much' the point belongs to each of the clusters as well as the extent of separation of regions within the structural space.

The decision regions calculated by the SVM model in the PCA-transformed 2-D space are presented in Figure 3. Each colored area cluster represents one of three original emission structure clusters. The implementation of an SVM separates the clusters, i.e., produces a piecewise linear or smooth decision surface that has the maximum possible distance to the clusters. The majority of our country-year samples are deeply seated in their regions, which implies that the grouping produced by the clustering mirrors meaningful and distinct structural patterns in the underlying data.

A significant advantage of this visualization is the possibility to detect so-called borderline and transition states. For example, observations from China between 2022 and 2025 are closer to the boundary between the industry-intensive (Cluster 2) and residential-transport-dominated (Cluster 0). This path is a consistent interpretation of China's gradual structural change from an energy-intensive emissions structure to a diversified or service-oriented emissions structure.

Not so in the United States. All its data points from 2019 to 2025 reside closely in Cluster 1, indicating a similar and continuous emission composition with no noticeable difference during the period. The sparseness of this clustering enhances one’s confidence in the classification and confirms the structural stability of the U.S. emissions network.

By mapping the SVM decision boundary back onto PCA space, we convert the unsupervised clustering output into a predictive classification model. This feature not only enables the classification of new (or future) observations, but may also be used as a diagnostic for the exploration of structural identity and how it changes. Nations closer to the boundaries can be identified as structurally unstable or as being in a state of transition, while those well within a region are robust structural archetypes. Therefore, the SVM model not only brings theoretical soundness but also interpretability to the emission structure categories constructed in this work.

DOI: 10.54254/2753-8818/2025.GL27830

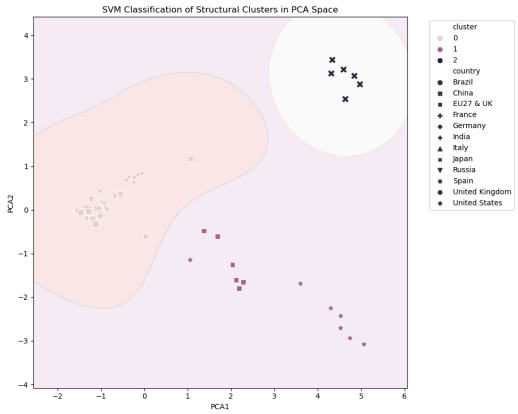


Figure 3. SVM classification of structural clusters in PCA space

3.3. Structural proximity from KNN analysis

To go beyond the static cluster type of this trajectory, we used a K-Nearest Neighbors (KNN) algorithm, in which the original standardized sectoral vectors are applied to the Euclidean distance method. Unlike typological clustering, which divides data among preset categories, KNN offers a continuous measure of similarity that allows for pairwise comparisons of country-year emission structures. This technique is particularly well-suited to studying the extent of intranational consistency and finding potential structural analogues across space and time.

In these studies, we concentrated on a subset of high-emission countries such as China, India, the United States, and the EU27 & UK, which together made up around half of total world CO_2 emissions during 2019–2025. Within each of these four country-years (denoted collectively as 'groups'), we searched for neighbors that were structurally closest to any given observation under vector distance.

The results are consistent across all the sample countries: the closest neighbors invariably come from years immediately preceding or following within individual countries. This consistent pattern is verified in Table 1 for China. For example, China 2021 is most similar to China 2020 (distance = 0.466), and China 2023 corresponds to China 2024 (distance = 0.358). Similar trends are also evident in other countries: the annual structural similarity within each time course of the United States and various years in India means that they fall together naturally. But no major trans-temporal analogs could be found among leading neighbors in any country we checked.

This intense intra-national sequence suggests that change in sectoral composition takes place incrementally, reflecting both inertia in energy systems and the domestic order of a national economy. While the standardization guarantees comparisons are scale-free, it also reveals how significant the difference is between countries structurally, even if emissions are comparable. Checking this for robustness, we also experimented with cosine similarity as an alternative distance metric. Across the board, we found essentially the same pattern: each country's annual emission profile is closest to its own past or future forms. Somewhat surprisingly, there exist some cases where cosine similarity produces an even more linear alignment of intra-country characteristics because it stresses direction rather than magnitude.

These findings suggest that the KNN method, while not ideal for finding external analogues, presents valuable insights into the structural persistence and path-dependent nature of national emission systems. The lack of significant inter-country matches serves once again to underscore

DOI: 10.54254/2753-8818/2025.GL27830

each national history's uniqueness and the difficulty involved in structural convergence in a short time.

Table 1. SVM classification of structural clusters in PCA space

source_country	source_year	neighbor_country	neighbor_year	distance
China	2019	China	2020	0.8709962863823050
China	2020	China	2021	0.4658128206487130
China	2021	China	2020	0.4658128206487130
China	2022	China	2021	0.6402798068954180
China	2023	China	2024	0.3580461135210040
China	2024	China	2023	0.3580461135210040
China	2025	India	2024	1.4188739457609100
United States	2019	United States	2024	0.5151093983100120
United States	2020	United States	2021	1.231593669262730
United States	2021	United States	2022	0.6742901981209510
United States	2022	United States	2023	0.5384498915170740
United States	2023	United States	2024	0.4108939158580670
United States	2024	United States	2023	0.4108939158580670
United States	2025	EU27 & UK	2020	2.11893403257128
India	2019	India	2022	0.1920204768595410
India	2020	India	2021	0.1721354503103570
India	2021	India	2022	0.1485029278554570
India	2022	India	2021	0.1485029278554570
India	2023	India	2024	0.1187201583763180
India	2024	India	2023	0.1187201583763180
India	2025	Japan	2020	0.3396964685954170
EU27 & UK	2019	EU27 & UK	2024	0.301719900208061
EU27 & UK	2020	EU27 & UK	2021	0.4257682758818530
EU27 & UK	2021	EU27 & UK	2020	0.4257682758818530
EU27 & UK	2022	EU27 & UK	2023	0.5512980914826740
EU27 & UK	2023	EU27 & UK	2024	0.3236105200431700
EU27 & UK	2024	EU27 & UK	2019	0.301719900208061
EU27 & UK	2025	Germany	2019	1.4247573595744100

4. Discussion

4.1. Interpretation of structural patterns

This research employs a mechanism learning architectural style to examine the underlying composition of CO_2 emissions in a selected set of countries from 2019 to 2025. Because the framework combines unsupervised learning techniques—namely PCA and K-Means clustering—

DOI: 10.54254/2753-8818/2025.GL27830

with supervised classification via SVM, what emerges is a structural typology that's both understandable and temporally responsive.

The clustering that ensues from this analysis shows that emission structures are not randomly dispersed. Instead, it shares a close correspondence with the industrial orientation and policy environment of a given country, as well as its developmental stage.

For example, China's structural transitions can be seen. From 2019 to 2022, it lay firmly in a cluster of energy-intensive production-oriented nations. However, starting in 2023, China turns slowly but steadily towards a more residential and transport-weighted pattern. This passageway is part of the government's strategy to diversify its energy mix, reduce coal dependence, and promote the electrification of residential and transport systems. In contrast, throughout the observed period, America remains structurally stable. Classed consistently in one cluster with balanced shares among all sectors of industry, this persistent classification indicates a high degree of both institutional and infrastructural inertia.

SVM analysis adds a layer of interpretability to the mix. By drawing clear boundaries between structural types in PCA space, this model pinpoints the crucial tipping points at which behavioral transitions may occur. Cases lying close to these thresholds—like China from 2023 to 2025—indicate a period in which structural readjustment may take place. This might be brought about by macroeconomic shifts or by particular measures introduced under policy. What's more, the SVM classifier lays out an empirically reproducible rule that can be used in real-time applications to segregate new country-year combinations or predict structural alignments around a range of hypothetical scenarios. Taking the view offered by boundaries into account indicates how far along a country is from shifting its emissions structure to a new state.

In sum, these findings reveal the value of combining structural typology with classification to investigate that carbon emissions are not just scalar quantities but configurations spanning various industries, indicating deeper development logics as well.

4.2. Methodological constraints

Despite the advantages of the proposed framework, we also want to acknowledge some of its methodological limitations.

The first three editions of this book covered a period that started at the turn of the 21st century and ended ten years later, making the series a practical yardstick for tracking the capital markets in different historical periods. But according to both editors and authors, those time frames were not arbitrary. There was actual economic data to support them: such factors as market capitalization and credit growth could be compared with tangible changes in society, such as people's incomes and the number of households.

Second, the analytical focus on normalized shares introduces trade-offs. While standardization is essential for cross-country comparability and pattern recognition, it also eliminates absolute scale differences, which are often critical to evaluating policies. For example, suppose two countries have the same normalized structural shares, but their total emissions are dramatically different. In that case, it may be difficult for funders who call themselves realists not to consider this abstraction. This problem particularly affects sectors with high emissions, such as industry and power generation. Here, one industrial structure may be functionally equivalent to another structurally for emissions purposes, but not reciprocally.

Third, the framework uses a simplified aggregation of six sectors (power, industry, transport, residential, domestic aviation, and international aviation). While this simplification enhances interpretability, it may also gloss over substantial variation within more general categories of

DOI: 10.54254/2753-8818/2025.GL27830

emitters. For example, within the transport sector, shipment of road freight or passenger car driving and shipping are all emissions classes with quite different origins as well as policy implications. While it is similarly misleading to pool "industry" activities without regard to the marked differences between heavy manufacturing, for instance, construction, and extractive sectors. A finer classification would yield greater structural detail and may indeed detect further patterns.

Fourthly, although these data cover 13 major source countries of CO_2 emissions, whose high output and availability make them a necessary inclusion for policy analysis, they also have an implication on space that is not entirely insignificant. Destitute regions are bypassed, as well as countries in transition to the modern era and high-emitting states. Future editions of this model should involve broader regional coverage and perhaps be designed in panel format with lower-tier or regional units built in.

4.3. Future directions

To build upon the current approach, future research could integrate causal inference techniques or econometric frameworks to explain the drivers of structural change. Expanding the dataset to include subnational units or extending the timeline could further enhance the analytical power. This structure-based lens has potential for supporting climate governance tools that aim to identify transition pathways, evaluate peer group alignment, or benchmark decarbonization strategies.

5. Conclusion

This paper presents a comprehensive framework for analyzing CO_2 emissions structures at the country level, encompassing dimensionality reduction, clustering, classification, and proximity analysis. The approach converts sectoral emissions into standardized vectors, facilitating the identification of types of latent structure and their changes across countries and over time.

Three main types of structures (industry-intensive, balanced, and transport-residential) representing different economic orientations were identified. Some countries, including China, undergo dynamic changes in their structure, while others, such as the United States, exhibit persistence in their structure. The SVM model enables the interpretation of these boundaries, and the KNN analysis confirms the consistency of national emission patterns.

Limitations include the short duration of data coverage, proportional data abstractions, and a limited spatial scope. Despite this, the approach is sound and can be used elsewhere. With richer data, this approach can enable the construction of structure-aware diagnostic instruments.

Crucially, this method provides a framework for comparing the structure of emissions that transcends basic national totals. Instead of sorting countries by emissions, it allows for sorting by structure, grouping them according to shared developmental patterns or similar transitions. This structural view of compositional development complements traditional intensity- or scale-oriented measures, providing a focus on internal composition and transformational potential.

From a more applied perspective, such information could help guide the design of more targeted decarbonization policies. Policymakers can then find their peers, not just according to economic status, but in terms of structural similarity, which enables cross-national learning, the transfer of technology, or the emulation of policy. Given the growing prominence that differentiated responsibility and just transition planning have in global carbon governance, a structure-based empiricism of this kind might provide new tools for guiding cooperation and resource allocation.

Ultimately, the integration of machine learning and structural emissions analysis offers a scalable and interpretable method that holds great promise for providing empirical support for policies. Its

DOI: 10.54254/2753-8818/2025.GL27830

potential applications include multiple-scale climate planning, international coordination, and structural analysis within global emission frameworks.

References

[1] Raupach, M. R., et al. (2007). Sharing CO_2 emission reductions among one billion high emitters. PNAS, 104(24), 10288–10293.
[2] Peters, G. P., et al. (2011). Growth in emission transfers via international trade from 1990 to 2008. PNAS, 108(21), 8903–8908.
[3] Le Quéré, C., et al. (2018). Global Carbon Budget 2018. Earth System Science Data, 10(4), 2141–2194.
[4] Zhang, Y., et al. (2021). Cluster analysis of energy consumption patterns in OECD countries. Energy Reports, 7, 101–113.
[5] Xie, R. & Wang, S. (2022). Clustering CO_2 emission pathways in emerging economies. Journal of Cleaner Production, 344, 131056.
[6] Li, M., et al. (2020). Exploring emission profiles via multidimensional indicators: A PCA-based clustering approach. Environmental Research Letters, 15(8), 085005.
[7] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
[8] Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137.
[9] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297.
[10] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.