



Carbon dioxide emission typology and policy implications: Evidence from machine learning

Hanjie Wang^a, Xiaohua Yu^{b,c,*}

^a College of Economics and Management, Southwest University, Chongqing City, China

^b Department of Agricultural Economics and Rural Development, University of Goettingen, Platz der Goettinger Sieben 5, 37073 Goettingen, Germany

^c Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY 14853-6201, USA

ARTICLE INFO

Keywords:

CO₂ emission
Emission typology
Machine learning
Carbon peak
Carbon neutrality
Unified National Carbon Market

ABSTRACT

The policy design of carbon dioxide (CO₂) emission mitigation is a hotly debated topic in the context of “Carbon Peak and Carbon Neutrality” in China. This paper contributes to this debate by employing an unsupervised machine learning algorithm to uncover the CO₂ emission typology based on the provincial emission data in China from 2000 to 2018 for a precise design of CO₂ emission mitigation policy for heterogenous regional patterns. The results indicate that we can cluster the provinces into four CO₂ emission patterns: the under-developed pattern, the coal-dominated pattern, the oil-dominated pattern, and the gas-dominated pattern. Notably, both the under-developed pattern and the coal-dominated pattern have a large amount of CO₂ emission from fossil fuels, while the gas-dominated pattern could be regarded as the policy inclination as it relies more on low-carbon fuels. Moreover, we also reveal the transition routes of emission patterns from a dynamic perspective, which could help policymakers better understand the future trend of emission patterns in different regions. On the one hand, the CO₂ emission mitigation policies could have specified priorities in different patterns, ensuring the feasibility during the process of policy implementation. On the other hand, establishing a national unified carbon trade market could facilitate efficient energy transition in China, and prevent carbon leakage cross different regions as well.

1. Introduction

It is a consensus that human activities since the industrial revolution have led to a rapid increase in carbon dioxide (CO₂) emission, aggravating global climate changes (Elheddad, Benjasak, Deljavan, Alharthi, & Almbrok, 2021). China, a developing country with a population of over 1.4 billion, is facing enormous challenges in tackling climate change as an essential part of the global climate campaign. Specifically, with the continuation of urbanization and industrialization, it is undoubted that energy demand, particularly coal-dominated energy, will keep rising over a period of time (Medlock III & Soligo, 2001; Mudakkar, Zaman, Khan, & Ahmad, 2013; Sadorsky, 2014). According to the report issued by Rhodium Group, China emitted 27% of the world’s greenhouse gases in 2019, which surpassed the entire developed world combined, although the per-person emissions in China are still far behind the U.S. In this scenario, the implementation of CO₂ emission mitigation policies has drawn widespread attention (Lin & Tan, 2017).

* Corresponding author at: Agricultural Economics in Developing and Transition Countries, Department of Agricultural Economics and Rural Development, University of Goettingen, Platz der Goettinger Sieben 5, 37073 Goettingen, Germany.

E-mail addresses: whjlee@163.com (H. Wang), xyu@uni-goettingen.de (X. Yu).

<https://doi.org/10.1016/j.chieco.2023.101941>

Received 31 July 2022; Received in revised form 15 February 2023; Accepted 16 February 2023

Available online 20 February 2023

1043-951X/© 2023 Elsevier Inc. All rights reserved.

Against this backdrop, the Chinese government attaches incredible importance to mitigating CO₂ emissions. At the General Debate of the 75th Session of the United Nations General Assembly in 2020, President Xi Jinping claimed that China would scale up the Nationally Determined Contributions (NDCs), aiming to achieve the goals of carbon peak before 2030 and carbon neutrality before 2060. To achieve such goals, the Chinese government has issued a wide range of policy documents, consisting of the “Action Plan for Carbon Dioxide Peaking Before 2030” and “Working Guidance for Carbon Dioxide Peaking and Carbon Neutrality”, as well as “the 14th Five Year Plan” (FYP),¹ which provide a series of specified measures to serve for CO₂ mitigation targets, including low-carbon transition of energy system, carbon emission peaking in industrial and agricultural sectors, promoting green and circular economy, and innovation on low-carbon technology.

However, implementing “Carbon Peak and Carbon Neutrality” policies is undergoing several challenges. The most challenged one is the regional heterogeneity in CO₂ emission patterns (Al Mamun, Sohag, Mia, Uddin, & Ozturk, 2014; Liu et al., 2015), which indicates that a one-size-fits-all policy would not well serve the goal of “Carbon Peak and Carbon Neutrality”, particularly in such a sheer size country. Undoubtedly, different regions in China have various energy sources, consumption structures, and economic development needs, requiring heterogeneous CO₂ mitigation strategies for each area (Chen et al., 2022; Lin & Fei, 2015; Wang & Zhao, 2015) and even for each firm (Xie, 2023). In particular, developed cities such as Beijing, Shanghai, and Tianjin, which have transformed into energy structures with low-carbon dioxide emissions, would be more likely to meet emission mitigation targets. However, underdeveloped provinces such as Guizhou, Shanxi, and Qinghai still heavily rely on traditional fossil fuels and have a high level of CO₂ emission. Implementing the same policies as the developed cities for those areas would be impractical. As such, revealing the CO₂ emission patterns of different regions and then proposing heterogeneous CO₂ mitigation strategies is a crucial issue in the context of “Carbon Peak and Carbon Neutrality”.

Academically, most previous studies mainly focus on the growth of CO₂ emissions and make great efforts to reveal the significant driving factors. On the whole, the growing CO₂ emissions are driven by booming economics, household energy consumption, population growth, capital investment, and growth in exports (Alam, Murad, Noman, & Ozturk, 2016; Guan, Hubacek, Weber, Peters, & Reiner, 2008; O’Neill et al., 2012; Wang, Shi, Fang, & Feng, 2019; Xiao et al., 2019; Zheng et al., 2019). Aside from driving factors of CO₂ emission, recent literature has realized that the recognition of CO₂ emission patterns is a crucial issue in the process of implementation of mitigation policies as a single mitigation strategy would not work well for all regions (Chen et al., 2022; Liu et al., 2015; Shan et al., 2018). For instance, the study of Xiao et al. (2019) discusses the CO₂ emission patterns of shrinking and growing cities and categorizes the cities into six groups with a population index. Also, Gregg, Andres, and Marland (2008) present the seasonal and spatial patterns of CO₂ emissions in China based on fossil fuel combustion and cement manufacturing emissions. Besides, considering the fact that the transport sector contributes most to global CO₂ emission, a number of studies are devoted to analyzing the CO₂ emission patterns in the transport sector (Guo et al., 2014; Solaymani, 2019; Zhang & Nian, 2013). The general conclusions indicate significant regional CO₂ emission differences in transport sectors (Santos, 2017; Timilsina & Shrestha, 2009).

Furthermore, from the perspective of global supply chain networks, Kagawa et al. (2015) distinguish 4756 CO₂ emission clusters and offer some insights on climate policies. According to the existing literature, we can find that the CO₂ emission patterns are often generated by some selected indexes or subjective judgments, which could not well mirror the underlying CO₂ emission patterns. As a result, the conclusions cannot provide a comprehensive understanding of emission typology as the researchers’ bias may exist in the identified typologies. Consequently, the information for supporting the policy design is somewhat limited.

In light of this, the overall aim of this study is to uncover the CO₂ emission patterns based on the similarity of CO₂ emission sources and then propose policy implications concerning regional heterogeneity with the use of unsupervised machine learning techniques, which have been widely used in pattern recognition (Boudet, MacDonald, Robinson, et al., 2020; Graskemper, Yu, & Feil, 2021, 2022; Zeng & Chen, 2022). This study could make two contributions to the existing literature. First of all, although previous studies have substantiated the potential of machine learning techniques in uncovering unlabeled patterns, little attention has been paid to the application of machine learning in CO₂ emission patterns. Correspondingly, the first contribution of this paper is to employ unsupervised machine learning, the K-means clustering algorithm, to analyze the typologies of carbon dioxide emissions. Compared with traditional approaches, unsupervised machine learning could better recognize the underlying typologies as the algorithm could ensure the provinces within the same typologies have high similarity in CO₂ emission patterns and do not rely on subjective judgments. Technically, the K-means algorithm provides an ideal approach to cluster multi-dimensional data (Wang, Feil, & Yu, 2023). With this, we can recognize the CO₂ emission patterns by comprising multi-dimensional metrics rather than a single indicator (Zeng & Chen, 2022). Second, to the best of our knowledge, the transition route of CO₂ emission typologies is so far absent in the existing literature, which limits the understanding of the emission typologies’ transition dynamics and policy designs. To fill this research gap, we investigate the CO₂ emission patterns transition based on the clustering results of machine learning from a dynamic perspective. Our results indicate that there are different routes of transition, which could provide a ground for the designs of future CO₂ mitigation policies.

The rest of the paper is laid out as follows: Section 2 introduces the data and methodology; Section 3 presents the CO₂ emission patterns based on the K-means clustering approach and Section 4 proposes some policy implications given the heterogeneities of CO₂ emission typologies in China; Section 5 is the general conclusion of this paper.

¹ The “14th Five Year Plan” (FYP) sets the goal to limit the increase in coal consumption over the 14th FYP period and phase it down in the 15th FYP period.

2. Data and methodology

2.1. Data

To analyze the CO₂ emission typology, we use the CO₂ emission inventories for 30 provinces from 2000 to 2018, which come from the CEADs (Carbon Emission Accounts & Datasets).² The CEADs is a scientific group gathering experts from the USA, U.K., and China, working on emerging economies' CO₂ emission accounting methods and applications. Given that the Chinese government does not provide official detailed CO₂ emission inventory data, the up-to-date emission data provided by the CEADs is an ideal data source for our research purpose.

There are three prevalent approaches regarding the CO₂ emission accounting method (Usubiaga & Acosta-Fernández, 2015). The first one is territorial emissions, considering the CO₂ emission generated within the national boundary, such as energy consumption and production of goods and services. Compared with the first account method, the second one calculates the indirect emissions generated by electricity heat consumption within the national boundary (Davis & Caldeira, 2010). Apart from these, the third method refers to all other indirect CO₂ emissions of a country during the production of final consumption. Among all the three approaches, the CEADs adopt the first approach, the IPCC's administrative-territorial scope, to calculate the CO₂ emission for 30 provinces in China. This is because the physical CO₂ emission calculated by this approach can provide detailed CO₂ emission information for emission mitigation policies.

Specifically, the CO₂ emission calculated by the CEADs includes two components: fossil fuel combustion and process-related emissions (cement production). For fossil fuel combustion, according to the IPCC guidelines, the CEADs uses the following equation to calculate CO₂ emission:

$$CE_i = AD_i \times NCV_i \times CC_i \times O \quad (1)$$

where CE_i is the CO₂ emission from fossil fuel i , AD_i refers to the combustion volume of fossil fuel i , NCV_i is the heat value per unit from the fossil fuel combustion, CC_i presents the carbon emissions per net calorific value of fossil fuel i , O refers to the oxidation ratio during the process of fossil fuel combustion. Although there are 26 types of fossil fuel in China's Energy Statistical System, the CEADs only uses 17 types as they merge certain types of fossil fuel with small consumption amount and similar quality (see Table 1).

The CEADs only consider cement production in terms of process-related emissions because it accounts for over 70% of process-related emissions in China. Similarly, the IPCC guidelines suggest that the following equation can calculate the process-related emission:

$$CE_t = AD_t \times EF_t \quad (2)$$

where CE_t denotes CO₂ emission from cement production; AD_t refers to the cement production; EF_t is the emission factor, which indicates per tonne of cement production generates 0.2906 t of CO₂. With the combinations of Eq. (1) and Eq. (2), the CEADs provide the CO₂ emission inventories for 30 provinces in China.

As shown in Table 2, the results draw a general picture of CO₂ emissions in China. The average CO₂ emission in 30 provinces from 2000 to 2018 is 242.10 million tonnes. However, we can easily detect the vast difference in CO₂ emission. The minimum emission is only 8.7 million tonnes, while the maximum is up to 912.20 million tonnes. Regarding the CO₂ emission structure, the descriptive statistics paint a more nuanced story. Among all fossil fuels, the CO₂ emissions generated by the combustion of raw coal reach 139.00 million tonnes, accounting for more than half of total CO₂ emissions.

Similarly, the variance also exists in the CO₂ emission of raw coal as the gap between the minimum and the maximum emissions reaches up to 586.40 million tonnes, indicating the different CO₂ emission structures. Likewise, we can draw a similar conclusion for the other types of CO₂ emission sources. In light of this, it is necessary to uncover the underlying CO₂ emission typology to provide policy implications for achieving the "Carbon Peak and Carbon Neutrality" goals.

2.2. K-means clustering

Machine learning techniques have been widely used in energy policy studies, such as energy poverty prediction (Wang, Maruejols, & Yu, 2021) and ethnic differences in energy consumption (Maruejols et al., 2022b). In this study, we use the unsupervised machine learning algorithm to aggregate similar CO₂ emission characteristics and uncover the underlying emission patterns, as the pattern is not labeled ex-ante. As mentioned above, there is a wide range of CO₂ emission resources, which make it difficult to understand the inherent law. The advantage of unsupervised machine learning lies in the fact that this method uses artificial intelligence algorithms to discover data patterns by extracting valuable information or features from input vectors without defined categories or groups (Graskemper et al., 2021, 2022; Maruejols et al., 2022a). In other words, the algorithm allows us just let the data speak without any external guidance during the clustering process. It is purely driven by the data.

Typically, there is a wide range of unsupervised machine learning algorithms, including Partitioning Around Medoids (PAM), K-means clustering, etc. Among all these algorithms, K-means clustering, proposed by Stuart Lloyd of Bell Labs in 1957, is one of the most prevalent and powerful tools in data science. Particularly, K-means clustering works well on clustering continuous quantitative

² <https://www.ceads.net/>

Table 1
Fossil fuels types.

No.	Fuels in China's energy statistics	Fuels in CEADs
1	Raw coal	Raw coal
2	Cleaned coal	Cleaned coal
3	Other washed coal	Other washed coal
4	Briquettes	Briquettes
	Gangue	
5	Coke	Coke
6	Coke oven gas	Coke oven gas
7	Blast furnace gas	Other gas
	Converter gas	
	Other gas	
8	Other coking products	Other coking products
9	Crude Oil	Crude Oil
10	Gasoline	Gasoline
11	Kerosene	Kerosene
12	Diesel oil	Diesel oil
13	Fuel oil	Fuel oil
14	Naphtha	Other petroleum products
	Lubricants	
	Paraffin	
	White spirit	
	Bitumen asphalt	
	Petroleum coke	
	Other petroleum products	
15	Liquefied petroleum gas (LPG)	LPG
16	Refinery gas	Refinery gas
17	Natural gas	Natural gas

Source: Shan, Y., Huang, Q., Guan, D., & Hubacek, K. (2020). China CO₂ emission accounts 2016–2017. Scientific data, 7(1), 1–9.

Table 2
Descriptive statistics of variables unit: million tonnes.

CO ₂ Emission	N	Mean	S.D.	Min	Max
Total	570	242.10	182.70	8.70	912.20
Raw Coal	570	139.00	109.30	3.20	589.60
Cleaned Coal	570	2.842	5.129	0	44.70
Other Washed Coal	570	4.277	6.707	0	43
Briquettes	570	1.557	3.615	0	55.90
Coke	570	27.63	37.47	0	268.7
Coke Oven Gas	570	2.722	4.725	0	74.40
Other Gas	570	7.548	14.40	0	200.1
Other Coking Products	570	0.882	1.915	0	19.60
Crude Oil	570	0.887	2.093	0	17
Gasoline	570	8.713	7.604	0	45.48
Kerosene	570	1.857	3.179	0	21.29
Diesel Oil	570	13.85	10.44	0	56.10
Fuel Oil	570	3.481	6.604	0	50.30
LPG	570	2.393	3.483	0	22.57
Refinery Gas	570	1.391	1.626	0	10.22
Other Petroleum Products	570	0.124	0.432	0	6.350
Natural Gas	570	7.193	8.243	0	57.31
Process	570	15.72	13.47	0	56.70

variables that could fit this study's data structure. More importantly, K-means algorithm is a powerful tool for clustering multi-dimensional indicators. In practice, the potential of unsupervised machine learning algorithms in uncovering the typology has been empirically substantiated (Graskemper et al., 2021, 2022).

Technically, the general idea of the K-means clustering algorithm is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. To capture the nearest mean, the K-means clustering algorithm typically uses within-cluster variances, also known as squared Euclidean distances, to ensure the similarity of each cluster.

Mathematically, let a set of observations (x_1, x_2, \dots, x_n) denotes the carbon dioxide emission sources of each province in China. The K-means clustering aims to group the n observations into k sets $S = (S_1, S_2, \dots, S_k)$ by minimizing the within-cluster sum of squares (WSS), that is:

$$\operatorname{argmin} \sum_{i=1}^k \sum_{x \in S_i} (x - u_i)^2 \tag{3}$$

where u_i is the mean of points in the cluster S_i .

Specifically, given an initial set of k means $m_1^{(1)}, m_2^{(1)}, \dots, m_k^{(1)}$, the algorithm proceeds by the following steps:

The first step is the assignment, which assigns each observation to the cluster whose mean has the least squared Euclidean distance.

$$S_i^{(t)} = \left\{ x_p : (x_p - m_i^{(t)})^2 \leq (x_p - m_j^{(t)})^2, \forall j, 1 \leq j \leq k \right\} \tag{4}$$

Based on Eq. (4), we can eventually assign each observation to a specified cluster $S^{(t)}$.

The second step is to update. Given the cluster generated by the first step, we recalculate the centroids for observations and then use the centroids as the new means, that is:

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{5}$$

This algorithm performs an iterative calculation to minimize the within-cluster sum of squares (WSS). Finally, we can obtain the optimal clusters from the dataset.

2.3. Optimal number of clusters

Choosing the optimal number of clusters is a tricky problem when performing clustering algorithms. As the number of clusters is open, selecting an optimal number is vital for the results. To solve this problem, there are many selection methods in the literature (Lesmeister, 2015). Among all these methods, the elbow method is a heuristic approach, which performs very well in determining the optimal cluster numbers in a dataset. Specifically, the elbow method plots the percentage of explained variation as a function of the number of clusters and then chooses the elbow of the curve as the optimal number of clusters. In other words, this method employs mathematical optimization to select an optimal point where diminishing returns are no longer worth the additional cost. In clustering analysis, the optimal number of clusters indicates that adding another cluster cannot obtain much better data modeling. Regarding the measure of variation, we use the Elbow method to judge the optimal number of clusters with Eq. (3), the within-cluster sum of squares.

2.4. Selection of clusters

Using the R program, the results indicate that the optimal cluster number of CO₂ emissions is four. As shown in Fig. 1, the left graph presents that the graph has the steepest slope when the number of clusters reaches four. Additionally, the right chart further confirms the optimal cluster number as the point of cluster four has the highest second-order difference. As a result, the following discussion will

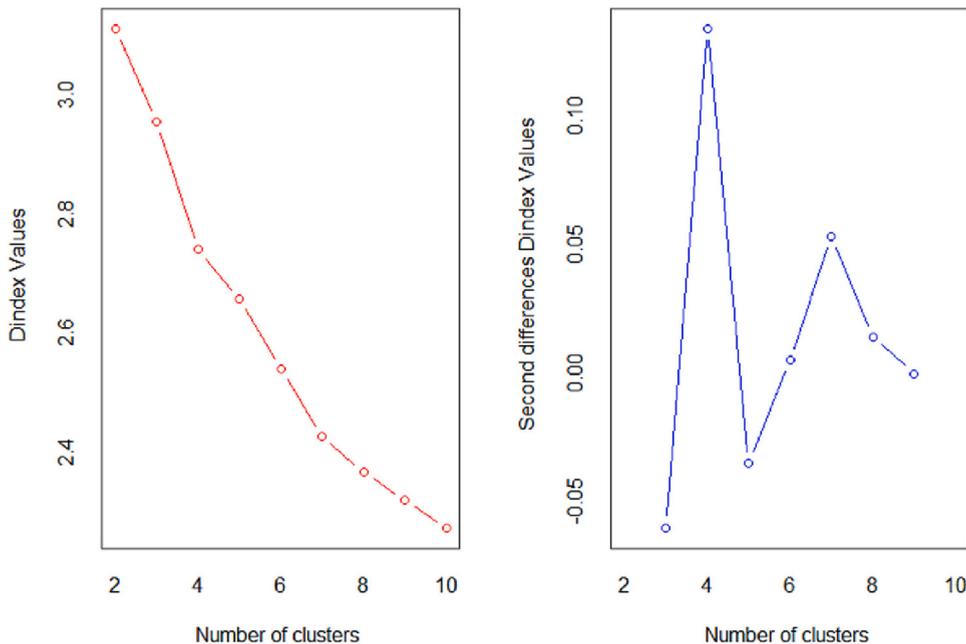


Fig. 1. Optimal cluster number.

be based on the four clusters.

Although the Elbow rule method is rather prevalent in the optimal cluster selection, it is subject to some criticism, such as the ignorance of the inter-cluster similarity. Actually, each method has its advantages and disadvantages. We use the majority rule in this scenario to determine the optimal cluster number. Specifically, apart from the Elbow rule, we employ the “NbClust” package in R software to determine the optimal cluster number. The results show that among 24 indices, 11 of 24 indices propose four as the best number of clusters. According to the majority rule, the best number of clusters is four, which is consistent with our results. The 11 indices are KL, CH, Hartigan, Scott, Marriot, TrCovW, TraceW, Friedman, Rubin, Ratkowsky, and Ball.³

3. Results and discussion

3.1. Clustering of CO₂ emission

The four patterns are identified using the K-means clustering algorithm. We define the four patterns according to the characteristics of the four patterns shown in Table 3 and Fig. 2. Specifically, the first one is identified as the under-developed pattern. The most significant characteristic of this pattern is the low carbon dioxide emission, which indicates the low energy demand. The second one is identified as coal-dominated pattern. It is understandable as the CO₂ emission of this group mainly comes from coal-related energy resources. Similarly, the third one is identified as oil-dominated pattern. This is due to the reason that oil-related energy such as gasoline, diesel oil, fuel oil and so on contribute most to the CO₂ emission of this pattern. Likewise, the fourth one can be identified as the gas-dominated pattern, as CO₂ emission mainly comes from the gas-dominated energy.

1. Pattern I: under-developed pattern. This pattern accounts for 69.47% of the sample. CO₂ emission of this group is relatively low, and almost all provinces belong to this group initially (see Table 4). Although the total emission is relatively low, it mainly comes from raw coal and its related product, such as cleaned coal, other washed coal, etc., which indicate the carbon emissions per net caloric value are relatively high.

2. Pattern II: coal-dominated pattern. This pattern accounts for 21.40% of the sample. CO₂ emissions of this group rank first, suggesting the high energy demand. Regarding the emission structure, the CO₂ emission mainly comes from traditional fossil fuels, including coal-related products and coke-related products, which have a relatively high CO₂ emission.

3. Pattern III: oil-dominated pattern. This pattern accounts for 3.51% of the sample. Although the total CO₂ emission of this pattern is relatively high, different from the coal-dominated pattern, the CO₂ emission mainly comes from oil-related fuels.

4. Pattern IV: gas-dominated pattern. This pattern accounts for 5.61% of the sample. As we can see, the gas-dominated pattern has the lowest CO₂ emissions. Notably, unlike the other three patterns, cleaning energy, natural gas, and refinery gas significantly contribute to the CO₂ emission of this group.

3.2. CO₂ Emission patterns of different clusters

To illustrate the CO₂ emission characteristics of different patterns, we use the one-way ANOVA (Analysis of Variance) to test whether the four patterns' means are equal. As shown in Table 3, the differences in CO₂ emission among the four patterns are statistically significant at 1%, which substantiates the rationality of clustering results. In terms of total CO₂ emission, we find that the coal-dominated pattern has the highest emission level, while the gas-dominated pattern has the lowest.

Crucially, regarding the CO₂ emission sources, the coal-dominated pattern's primary sources come from the high carbon emission fossil fuels such as coal-related and coke-related products. The economic explanation possibly lies in the fact that those high energy-demand provinces would like to choose traditional fossil fuels due to their low price and availability. However, although the CO₂ emission of the oil-dominated pattern is close to the coal-dominated pattern, the CO₂ emission sources are quietly different as it relies on both high-carbon and low-carbon emission fossil fuels.

In addition, it is clear that the under-developed and gas-dominated patterns have similar CO₂ emission levels, but the emission structures are totally different. Specifically, the under-developed pattern has a high CO₂ emission volume from high-carbon emission fossil fuels. In contrast, the CO₂ emission of the gas-dominated pattern mainly comes from low-carbon emission fossil fuels.

We further provide the radar plot to depict the CO₂ emission structures of the four patterns (see Fig. 2). It is apparent that the high-carbon emission fossil fuels dominate the CO₂ emission of the coal-dominated pattern. On the contrary, the oil-dominated pattern relies on both the high-carbon emission fossil fuels and the low-carbon fuels, ensuring the energy demand with a relatively low CO₂ emission. In this scenario, changing the traditional emission patterns into a complex emission pattern could be a CO₂ emission mitigation strategy for the high energy demand provinces. Likewise, emission structures are also different for the low CO₂ emission provinces between the initial and new emission patterns. Explicitly, the under-developed pattern has the lowest CO₂ emission from the low-carbon emission fossil fuels. In contrast, the gas-dominated pattern has the lowest CO₂ emission from the high-carbon emission fossil fuels. In this case, the CO₂ emission structure transformation of the under-developed pattern is necessary for the carbon mitigation policy targets.

³ Charrad, Ghazzali, Boiteau, et al. (2014) introduce the selection indices in detail.

Table 3
CO₂ emission of different patterns unit: million tonnes.

Pattern	I	II	III	IV	p. overall
Total	159.33	508.14	407.12	148.56	< 0.001
Raw Coal	97.28	284.62	208.42	56.93	< 0.001
Cleaned Coal	1.75	6.59	5.18	0.57	< 0.001
Other Washed Coal	3.08	9.92	0.26	0.14	< 0.001
Briquettes	0.82	3.93	3.81	0.26	< 0.001
Coke	15.24	74.25	13.22	12.14	< 0.001
Coke Oven Gas	1.38	7.69	0.77	1.60	< 0.001
Other Gas	3.49	21.32	7.62	5.20	< 0.001
Other Coking Products	0.42	2.54	0.15	0.78	< 0.001
Crude Oil	0.89	1.09	1.01	0.08	0.112
Gasoline	5.20	16.17	26.98	12.31	< 0.001
Kerosene	0.86	1.79	5.94	11.91	< 0.001
Diesel Oil	9.36	24.71	41.24	10.84	< 0.001
Fuel Oil	1.58	4.00	25.55	11.29	< 0.001
LPG	1.35	3.34	17.44	2.34	< 0.001
Refinery Gas	0.91	2.14	2.96	3.54	< 0.001
Other Petroleum Products	0.06	0.10	0.87	0.51	< 0.001
Natural Gas	4.36	13.38	13.77	14.57	< 0.001
Process	11.32	30.54	31.94	3.56	< 0.001

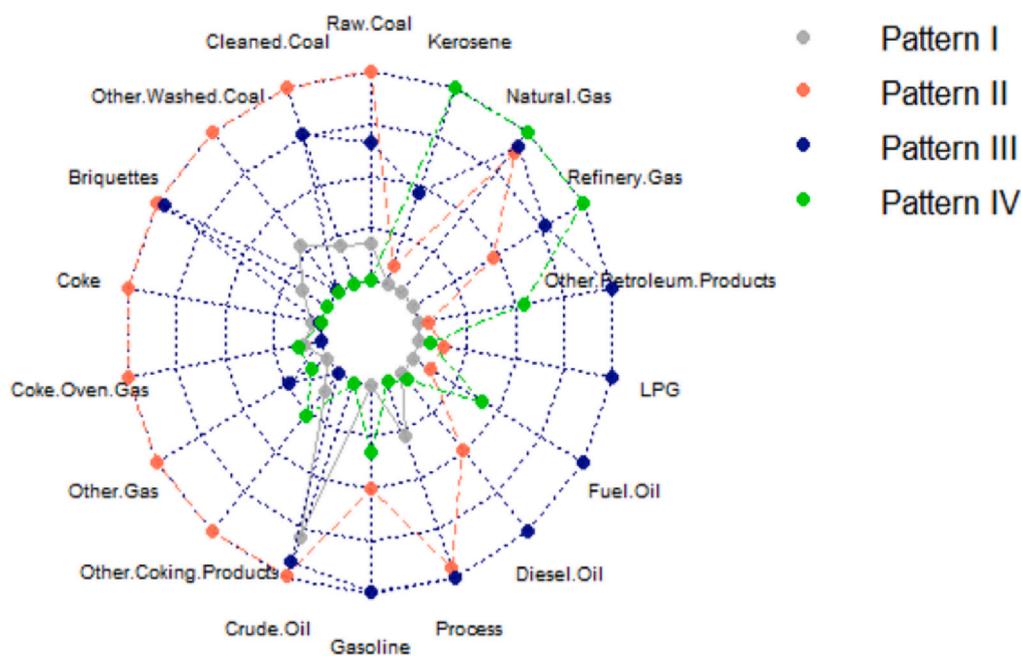


Fig. 2. Radar plot for CO₂ emission patterns.

Table 4
CO₂ emission patterns transition.

Transition Routes	Province
Remain at Pattern I	Jilin, Heilongjiang, Jiangxi, Guangxi, Hainan, Chongqing, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia
Transitions from Pattern I to Pattern IV	Beijing, Tianjin, Shanghai, Fujian
Transitions from Pattern I to Pattern II	Shanxi, Inner Mongolia, Liaoning, Jiangsu, Anhui, Shandong, Henan, Hebei, Hunan, Sichuan, Xinjiang
Transitions from Pattern I to Pattern II, III	Zhejiang
Remain at Pattern III	Guangdong

3.3. Clusters transition

As highlighted in Fig. 2 above, the K-means clustering algorithm detects the four patterns of CO₂ emission. Nevertheless, the question of how the CO₂ emission patterns transition over the period of 2000 to 2018 is still open. For this reason, we then investigate the emission patterns transition routes from a dynamic perspective.

Table 4 reveals four transition routes for 30 provinces in China from 2000 to 2018. As it turns out, almost all provinces transition from the under-developed pattern to other patterns or keep unchanged. It is clear that there are 12 provinces that remain in the under-developed pattern from 2000 to 2018. Notably, most of the 12 provinces locate in western China, which has a comparatively backward economy. As depicted in Fig. 3, there is a positive relationship between energy consumption and economic development. Consequently, those provinces do not have much energy demand leading to the low CO₂ emission level. Moreover, the backward economy also determines the structure of high-carbon emission fossil fuels as such fuels have lower prices.

More importantly, the transition of emission patterns happens in most provinces. Fig. 4 clearly shows the transition routes. We can detect that most provinces transition from under-developed patterns to other patterns. In particular, the four provinces, namely Beijing, Shanghai, Tianjin, and Fujian, transitioned from the under-developed pattern to the gas-dominated pattern. It is well-known that those provinces have an advanced service economy dominated by service industries. Accordingly, those provinces transition to the gas-dominated pattern relying on low-carbon emission fossil fuels. In contrast, there are 11 provinces that transitioned from the under-developed pattern to the coal-dominated pattern. Such shift significantly increases the CO₂ emissions, which are primarily generated by the high-carbon emission fossil fuels such as coal-related and coke-related fuels. This is easy to understand as most of those provinces are heavy industry districts requiring a high energy demand. Another reason for this phenomenon might be due to the reason that some of those provinces are important coal-producing areas in China. For instance, Shanxi and Inner Mongolia account for over half of the coal-producing in 2020.

Furthermore, we also detect that Zhejiang province experiences a two-stage transition route. In the first stage, Zhejiang transitions from the under-developed pattern to the coal-dominated pattern, consistent with the 11 provinces mentioned above. However, after that, Zhejiang continued to transition to the oil-dominated pattern, indicating that some of the high-carbon emission fossil fuels are replaced by low-carbon emission fossil fuels, which could help reduce carbon emissions. Remarkably, among all 30 provinces, only one province, Guangdong, remains in the oil-dominated pattern. As we know, the economies of Guangdong and Zhejiang provinces are highly developed. Understandably, the two provinces need a large number of energy supplies along with relatively high CO₂ emissions. Besides, as a coastal province, Guangdong could easily obtain large amounts of oil from the international oil market by importing. To meet the dual goals of energy demand and carbon emissions mitigation, the CO₂ emission patterns transition to the oil-dominated pattern.

3.4. Robustness check

In this part, we further employ K-median algorithm as a robustness check to ensure the credibility of clustering results. In K-means algorithm, the clustering results are based on the calculation of Euclidean distance between the data points, which indicates that such

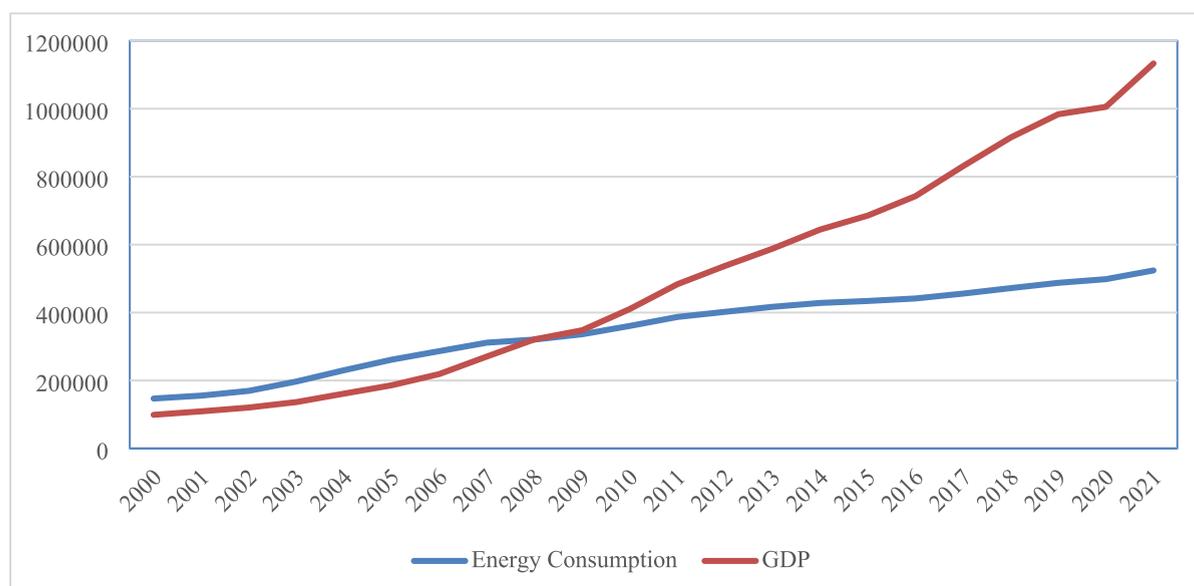


Fig. 3. Energy consumption and GDP in China.

Data source: China Statistical Yearbook (2001–2022).

Notes: the unit of energy consumption is ten thousand tons of coal equivalent; the unit of GDP is 100 million yuan.

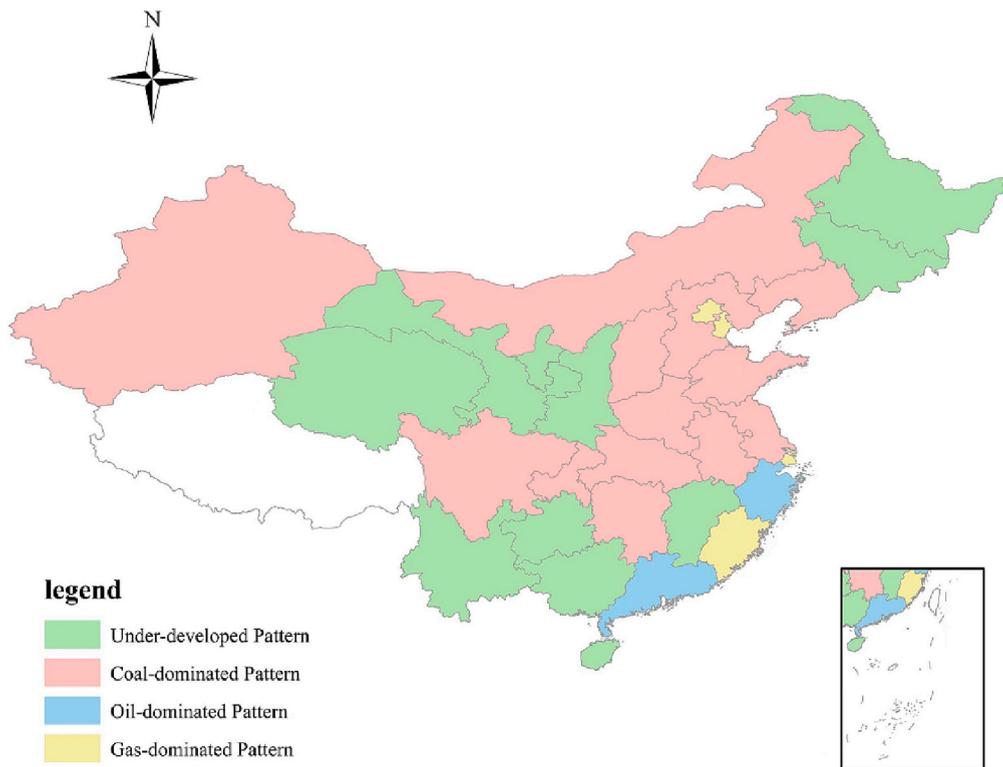


Fig. 4. CO₂ Emission patterns in 2018.

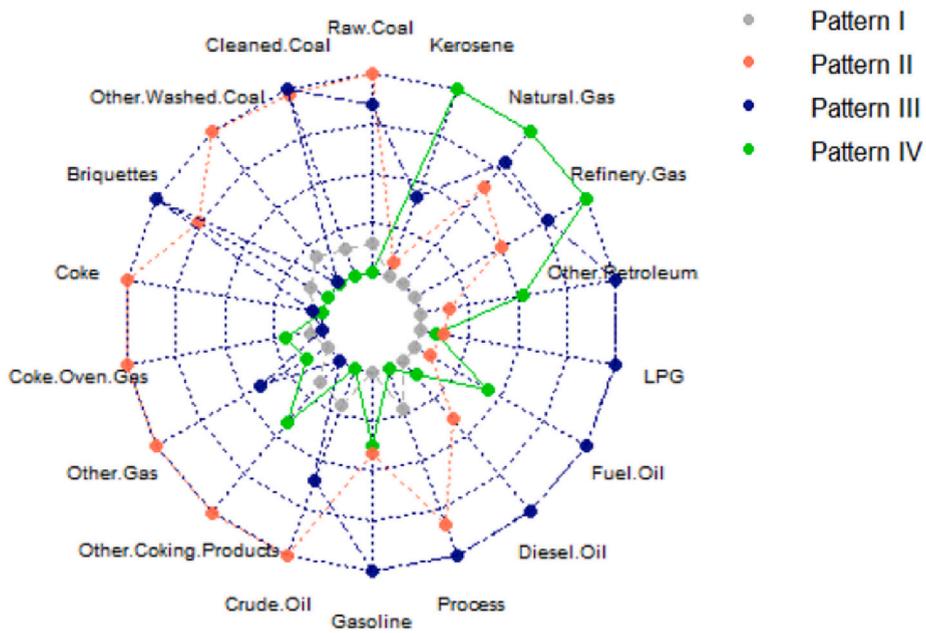


Fig. 5. Robustness check.

an approach might magnify the influence of outliers. On the contrary, K-median algorithm uses the Manhattan distance, which could significantly reduce the influence of outliers (Bradley, Mangasarian, & Street, 1996). As such, we use K-median algorithm to reveal the CO₂ emission patterns. Fig. 5 clearly shows that the four patterns we recognize by K-median is consistent with K-means, substantiating our results' robustness.

4. Policy implications

This section aims to provide some policy implications for CO₂ emission mitigation targets and discuss the inequality problem during this process. The four patterns identified by the K-means clustering algorithm substantiate the heterogeneity of CO₂ emissions across 30 provinces in China. The different characteristics of patterns mirror the differences in CO₂ emission structure, which require the specified policy intervention. Hence, the “one-size-fits-all” approach may not achieve policy targets. The policy design should consider all four CO₂ emission patterns to reach the “Carbon Peak and Carbon Neutrality” aims. For this reason, the following discussion will base on the four CO₂ emission patterns.

First of all, for the under-developed pattern, although those provinces do not have high CO₂ emissions, it mainly comes from high-carbon emission fossil fuels. To reduce the CO₂ emission of this pattern, the strategies could be two folds. On the one hand, given the reality that the under-developed pattern and the gas-dominated pattern have similar CO₂ emissions, it is feasible for the under-developed pattern to transition to the gas-dominated pattern, which means that those provinces could reshape the energy structure changing from high-carbon emission fossil fuels to low-carbon fuels such as natural gas. On the other hand, as mentioned above, most of those provinces are located in western China, which owns affluent renewable resources based on natural conditions. Accordingly, they have a promising prospect for renewable energy development.

Additionally, for the coal-dominated pattern, in response to the CO₂ emissions from heavy industries, those provinces should consider the strategies concerning the transition of the industrial structure. Notably, the critical point of this strategy refers to the new environment-friendly technology adoption and equipment replacement but not simply closing the heavy industries. Meanwhile, it is vital to transition the energy structure from high-carbon fossil fuels to low-carbon ones, promoting energy efficiency.

Besides, although the oil-dominated pattern has a lower CO₂ emission energy structure than the coal-dominated pattern, it still has a great potential to reduce CO₂ emission further. As shown in Table 4, the oil-dominated pattern only refers to two developed provinces, Zhejiang and Guangdong, both coastal areas. Due to the geographical advantage, the two provinces could take full advantage of wind and ocean energy, converting fossil fuels into renewable ones.

Regarding the CO₂ emission routes, in the context of “Carbon Peak and Carbon Neutrality”, the policy priority should pay more attention to those provinces that transition to the coal-dominated pattern. On the one hand, the developed province like Jiangsu could gradually transfer the energy structure from coal-dominated energy to environment-friendly energy. On the other hand, some provinces, such as Xinjiang and Inner Mongolia, could take full advantage of geographical location and utilize natural energy such as wind and solar. The detection of four different regional patterns also provides a benchmark for carbon trade across different provinces under different carbon endowments. Establishing a national unified carbon trade market could facilitate energy transition in China and prevent carbon leakage across different regions.

Notably, another concern is the social impact of CO₂ emission mitigation policies. Theoretically, reducing CO₂ emissions could limit the negative impact of climate change on social inequality. However, the mitigation policies design should also consider the consequences of inequalities. The existing literature has substantiated that climate change mitigation policies might worsen social inequality (Sanna & Annela, 2019). For instance, transitioning from the high carbon dioxide emission industries to low carbon could lead to job losses, particularly for the low-income population. Moreover, the infrastructure and technology investments are expensive, which would be unaffordable for economically backward provinces, expanding regional economic inequality. Therefore, the mitigation policies should serve the broader objectives including poverty and inequality reduction, but not only the purpose of improvement of environment. Given the greater historical contribution, the rich provinces should take more responsibilities. More specifically, in the process of CO₂ emission mitigation policies implication, policymakers should offer more support for the low-income population and less developed areas by transfer payment and preferential policies to offset the negative impacts on the poor.

5. Conclusions

CO₂ emission mitigation policy design is a much-debated topic in the context of “Carbon Peak and Carbon Neutrality”. Due to the regional differences with regard to economic development and industrial structure, the CO₂ emission patterns vary across provinces. In this context, the “one-size-fits-all” policy design may be unable to meet the needs of policy objectives. Although some literature has been devoted to revealing the different CO₂ emission patterns, most of the classifications only depend on common knowledge but not the similarity of CO₂ emission structure, which cannot scientifically reveal the carbon dioxide patterns of the real world. In light of this, this study aims to employ an unsupervised machine learning approach, the K-means clustering algorithm, to analyze the CO₂ emission patterns of 30 provinces in China and their transition ways.

The first contribution is that we use the unsupervised machine learning technique, the K-means clustering algorithm, to identify the CO₂ emission patterns. The clustering results show that the CO₂ emission patterns of 30 provinces can be clustered into four patterns according to their emission characteristics: the under-developed pattern, the coal-dominated pattern, the gas-dominated pattern, and the oil-dominated pattern. In particular, both the under-developed pattern and the coal-dominated pattern have a large amount of CO₂ emission from high-carbon emission fossil fuels such as coal-related fuels and coke-related fuels, although the coal-dominated pattern has a higher total CO₂ emission level than the under-developed pattern. In contrast, the gas-dominated pattern could be regarded as the

policy inclination because it relies more on low-carbon fuels, including natural gas and refinery gas, than other patterns. Besides, the CO₂ emission of the oil-dominated pattern comes from both the high-carbon and low-carbon emission fuels. This could be a promising pattern in policy design for the high energy demand area as it meets the dual goals of energy demand and carbon emissions mitigation.

Another contribution of this study is that we investigate the CO₂ emission patterns transition. The results reveal that there are four routes of transition. The areas with a backward economy remain in the under-developed pattern, which means that although those areas do not have too much energy demand, they rely on high-carbon emission fuels. The areas with heavy industries or coal-producing areas transition from the under-developed pattern to the coal-dominated pattern, indicating the high energy demand of those areas is accompanied by increased CO₂ emissions. The areas with advanced service economies transition from the under-developed pattern to the gas-dominated pattern, which means that low-carbon fuels replace high-carbon emission fuels. The areas with highly developed economies either remain at or transition to the oil-dominated emission pattern. Doing so could meet the dual goals of energy demand and carbon emissions mitigation.

On the whole, this study provides a new perspective for CO₂ emission mitigation policy designs. The policy implication of this study is that the CO₂ emission patterns should be taken into consideration under China's "Carbon Peak and Carbon Neutrality" targets. For different patterns, the CO₂ emission mitigation policies could have specified priorities, ensuring feasibility during the policy implementation process.

Finally, this paper also provides evidence to support carbon trade across different provinces and patterns under different carbon endowments. Establishing a national unified carbon trade market could effectively facilitate energy transition while preventing carbon leakage across different regions with heterogenous energy patterns and policy designs.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the National Social Science Foundation of China (20CSH048, 22CJY037, 20AZD080, 21ZDA062), the Humanities and Social Sciences Project Funded by the Chinese Ministry of Education (21YJC790110), the Social Science Foundation of Chongqing (2022YC004, 2021NDQN34), and the Innovation Research 2035 Pilot Plan of Southwest University (SWUPilotPlan026).

References

- Al Mamun, M., Sohag, K., Mia, M. A. H., Uddin, G. S., & Ozturk, I. (2014). Regional differences in the dynamic linkage between CO₂ emissions, sectoral output and economic growth. *Renewable and Sustainable Energy Reviews*, 38, 1–11.
- Alam, M. M., Murad, M. W., Noman, A. H. M., & Ozturk, I. (2016). Relationships among carbon emissions, economic growth, energy consumption and population growth: Testing environmental Kuznets Curve hypothesis for Brazil, China, India and Indonesia. *Ecological Indicators*, 70, 466–479.
- Boudet, F., MacDonald, G. K., Robinson, B. E., et al. (2020). Rural-urban connectivity and agricultural land management across the Global South[J]. *Global Environmental Change*, 60, Article 101982.
- Bradley, P., Mangasarian, O., & Street, W. (1996). Clustering via concave minimization. *Advances in Neural Information Processing Systems*, 9.
- Charrad, M., Ghazzali, N., Boiteau, V., et al. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61, 1–36.
- Chen, J., et al. (2022). Carbon sequestration capacity of terrestrial vegetation in China based on satellite data. *Journal of Chinese Economics and Business Studies*, 20(1), 109–124.
- Davis, S. J., & Caldeira, K. (2010). Consumption-based accounting of CO₂ emissions. *Proceedings of the National Academy of Sciences*, 107(12), 5687–5692.
- Elheddad, M., Benjasak, C., Deljavan, R., Alharthi, M., & Almbrok, J. M. (2021). The effect of the Fourth Industrial Revolution on the environment: The relationship between electronic finance and pollution in OECD countries. *Technological Forecasting and Social Change*, 163, Article 120485.
- Graskemper, V., Yu, X., & Feil, J. H. (2021). Farmer typology and implications for policy design—an unsupervised machine learning approach. *Land Use Policy*, 103, Article 105328.
- Graskemper, V., Yu, X., & Feil, J. H. (2022). Values of farmers—evidence from Germany. *Journal of Rural Studies*, 89, 13–24.
- Gregg, J. S., Andres, R. J., & Marland, G. (2008). China: Emissions pattern of the world leader in CO₂ emissions from fossil fuel consumption and cement production. *Geophysical Research Letters*, 35(8).
- Guan, D., Hubacek, K., Weber, C. L., Peters, G. P., & Reiner, D. M. (2008). The drivers of Chinese CO₂ emissions from 1980 to 2030. *Global Environmental Change*, 18(4), 626–634.
- Guo, B., Geng, Y., Franke, B., Hao, H., Liu, Y., & Chiu, A. (2014). Uncovering China's transport CO₂ emission patterns at the regional level. *Energy Policy*, 74, 134–146.
- Kagawa, S., Suh, S., Hubacek, K., Wiedmann, T., Nansai, K., & Minx, J. (2015). CO₂ emission clusters within global supply chain networks: Implications for climate change mitigation. *Global Environmental Change*, 35, 486–496.
- Lesmeister, C. (2015). *Mastering machine learning with R*. Packt Publishing Ltd.
- Lin, B., & Fei, R. (2015). Regional differences of CO₂ emissions performance in China's agricultural sector: A Malmquist index approach. *European Journal of Agronomy*, 70, 33–40.
- Lin, B., & Tan, R. (2017). Sustainable development of China's energy intensive industries: From the aspect of carbon dioxide emissions reduction. *Renewable and Sustainable Energy Reviews*, 77, 386–394.
- Liu, Z., Guan, D., Wei, W., Davis, S. J., Ciais, P., Bai, J., ... He, K. (2015). Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature*, 524(7565), 335–338.
- Maruejols, L., et al. (2022a). Comparison of machine learning predictions of subjective poverty in rural China. *Forthcoming in China Agricultural Economic Review*. <https://doi.org/10.1108/CAER-03-2022-0051>
- Maruejols, L., et al. (2022b). Vietnam between Economic Growth and Ethnic Divergence: A LASSO examination of income-mediated energy consumption. *Forthcoming. Energy Economics*. <https://doi.org/10.1016/j.eneco.2022.106222>
- Medlock, K. B., III, & Soligo, R. (2001). Economic development and end-use energy demand. *The Energy Journal*, 22(2).
- Mudakkar, S. R., Zaman, K., Khan, M. M., & Ahmad, M. (2013). Energy for economic growth, industrialization, environment and natural resources: Living with just enough. *Renewable and Sustainable Energy Reviews*, 25, 580–595.

- O'Neill, B. C., Liddle, B., Jiang, L., Smith, K. R., Pachauri, S., Dalton, M., & Fuchs, R. (2012). Demographic change and carbon dioxide emissions. *The Lancet*, 380 (9837), 157–164.
- Sadorsky, P. (2014). The effect of urbanization and industrialization on energy use in emerging economies: Implications for sustainable development. *American Journal of Economics and Sociology*, 73(2), 392–409.
- Sanna, M., & Annala, A. (2019). Social impacts of climate change mitigation policies and their implications for inequality. *Climate Policy*, 19(7), 827–844.
- Santos, G. (2017). Road transport and CO₂ emissions: What are the challenges? *Transport Policy*, 59, 71–74.
- Shan, Y., Guan, D., Hubacek, K., Zheng, B., Davis, S. J., Jia, L., ... Schellnhuber, H. J. (2018). City-level climate change mitigation in China. *Science Advances*, 4(6), eaaq0390.
- Solaymani, S. (2019). CO₂ emissions patterns in 7 top carbon emitter economies: The case of transport sector. *Energy*, 168, 989–1001.
- Timilsina, G. R., & Shrestha, A. (2009). Transport sector CO₂ emissions growth in Asia: Underlying factors and policy options. *Energy Policy*, 37(11), 4523–4539.
- Usubiaga, A., & Acosta-Fernández, J. (2015). Carbon emission accounting in MRIO models: The territory vs. the residence principle. *Economic Systems Research*, 27(4), 458–477.
- Wang, H., Feil, J. H., & Yu, X. (2023). Let the data speak about the cut-off values for multidimensional index: Classification of human development index with machine learning. *Socio-Economic Planning Sciences*, 101523.
- Wang, H., Maruejols, L., & Yu, X. (2021). Predicting energy poverty with combinations of remote-sensing and socioeconomic survey data in India: Evidence from machine learning. *Energy Economics*, 102, Article 105510.
- Wang, S., Shi, C., Fang, C., & Feng, K. (2019). Examining the spatial variations of determinants of energy-related CO₂ emissions in China at the city level using Geographically Weighted Regression Model. *Applied Energy*, 235, 95–105.
- Wang, Y., & Zhao, T. (2015). Impacts of energy-related CO₂ emissions: Evidence from under developed, developing and highly developed regions in China. *Ecological Indicators*, 50, 186–195.
- Xiao, H., Duan, Z., Zhou, Y., Zhang, N., Shan, Y., Lin, X., & Liu, G. (2019). CO₂ emission patterns in shrinking and growing cities: A case study of Northeast China and the Yangtze River Delta. *Applied Energy*, 251, Article 113384.
- Xie, X. (2023). Does carbon neutrality commitment enhance firm value? *Forthcoming in Journal of Chinese Economics and Business Studies*. <https://doi.org/10.1080/14765284.2022.2161171>
- Zeng, Q., & Chen, X. (2022). Identification of urban-rural integration types in China-an unsupervised machine learning approach. *China Agricultural Economic Review*. <https://doi.org/10.1108/CAER-03-2022-0045>
- Zhang, C., & Nian, J. (2013). Panel estimation for transport sector CO₂ emissions and its affecting factors: A regional analysis in China. *Energy Policy*, 63, 918–926.
- Zheng, J., Mi, Z., Coffman, D. M., Milcheva, S., Shan, Y., Guan, D., & Wang, S. (2019). Regional development and carbon emissions in China. *Energy Economics*, 81, 25–36.