



Article

Discovering Geographical Flock Patterns of CO₂ Emissions in China Using Trajectory Mining Techniques

Pengdong Zhang ^{1,2,*}, Lizhi Miao ^{1,2} , Fei Wang ³ and Xinting Li ¹

¹ School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Wenyuan Road 9, Nanjing 210023, China

² Smart Health Big Data Analysis and Location Services Engineering Research Center of Jiangsu Province, Nanjing University of Posts and Telecommunications, Wenyuan Road 9, Nanjing 210023, China

³ East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China

* Correspondence: pengdong.zhang@njupt.edu.cn

Abstract: Carbon dioxide (CO₂) emissions are considered a significant factor that results in climate change. To better support the formulation of effective policies to reduce CO₂ emissions, specific types of important emission patterns need to be considered. Motivated by the flock pattern that exists in the domain of moving object trajectories, this paper extends this concept to a geographical flock pattern and aims to discover such patterns that might exist in CO₂ emission data. To achieve this, a spatiotemporal graph (STG)-based approach is proposed. Three main parts are involved in the proposed approach: generating attribute trajectories from CO₂ emission data, generating STGs from attribute trajectories, and discovering specific types of geographical flock patterns. Generally, eight different types of geographical flock patterns are derived based on two criteria, i.e., the high–low attribute values criterion and the extreme number–duration values criterion. A case study is conducted based on the CO₂ emission data in China on two levels: the province level and the geographical region level. The results demonstrate the effectiveness of the proposed approach in discovering geographical flock patterns of CO₂ emissions and provide potential suggestions and insights to assist policy making and the coordinated control of carbon emissions.



Citation: Zhang, P.; Miao, L.; Wang, F.; Li, X. Discovering Geographical Flock Patterns of CO₂ Emissions in China Using Trajectory Mining Techniques. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4265. <https://doi.org/10.3390/ijerph20054265>

Academic Editor: Paul B. Tchounwou

Received: 31 December 2022

Revised: 24 February 2023

Accepted: 26 February 2023

Published: 27 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: carbon emission; geographical flock pattern mining; attribute trajectory; bivariate time series; climate change

1. Introduction

Climate change is one of the biggest challenges faced by humankind, and its impact is becoming increasingly significant [1,2]. It can greatly affect the environment, production system, survival and development of humankind [3–5], hence typical extreme events and disasters (e.g., storms, droughts, heatwaves, fires, and floods) have become stronger and more frequent [6]. The 2022 Intergovernmental Panel on Climate Change (IPCC) annual report demonstrates that the world faces unavoidable climate hazards over the next two decades with global warming of 1.5 °C (2.7 °F), and urgent actions are required to deal with increasing risks [7]. One commonly acceptable solution is to make rapid and deep cuts in greenhouse gas emissions, particularly carbon dioxide (CO₂) [8].

The majority of countries in the world have been seeking effective ways to reduce CO₂ emissions. According to previous research, China is currently the largest emitter of CO₂, accounting for 31% of global CO₂ emissions in 2020 due to its rapid economic development and urbanization [9]. To achieve the net-zero or near-zero CO₂ emissions, the Chinese government has launched a long-term mitigation goal, which indicates that China would reach peak emissions by 2030, and achieve carbon neutrality by 2060 [10]. Motivated by this national strategy, emerging research regarding carbon emissions in China has been paid much attention in various domains. Typical research focuses include the modelling, prediction, trading and policy evaluation of carbon emissions [11–16]. The

establishment of effective policies to reduce carbon emissions is critical. Nevertheless, the long-term reduction of carbon emissions remains a key policy challenge for China and the world [17]. To meet the requirements to establish effective policies, specific, precise, and flexible policies must be proposed to different levels of geographical units (e.g., county, city, province, region, and country). Understanding the trends and trajectories of carbon emissions remains challenging in light of uncertainty about world economies and technological breakthroughs [18]. In this regard, the discovery of potentially important emission patterns must be performed from historical carbon emission data. The discovery of emission patterns plays a significant role in guiding the formulation of specific, precise, and flexible policies and the coordinated control of carbon emissions. Therefore, this paper focuses mainly on discovering important patterns from carbon emission data.

Carbon emission data are usually time series data, and time series can be classified into three types, i.e., univariate time series, bivariate time series, and multivariate time series [19,20]. In this paper, we are particularly interested in bivariate time series, because they can provide more abundant information than conventional univariate time series; furthermore, related consumptions (e.g., time) can be saved compared with complex multivariate time series. As a bivariate time series integrates the information of two different attributes into one coordinate system, it can be fully transferred to an attribute trajectory, which was introduced in [21] as a novel kind of trajectory. On this basis, traditional trajectory mining techniques can thus be adopted to discover desired movement related patterns.

Nowadays trajectory data are ubiquitous, having benefited from the proliferation of location aware techniques such as the global navigation satellite system (GNSS), Bluetooth, radio frequency identification (RFID), and Wi-Fi. The discovery of movement patterns has an important place in the domains of trajectory data mining as these patterns can exhibit the rules of individuals' movements and their interactions. Typical movement patterns include the flock pattern [22–26], convoy pattern [27–29], leadership pattern [30–32], moving cluster [33,34], and crew [35]. Among these, the flock pattern has been paid much attention and it commonly plays an important role in many application fields. A flock can be informally depicted as “a group of spatially close objects staying together for a specific time duration”. Correspondingly, assuming geographical units are considered as objects and their attributes staying close for a specific time duration, these geographical units can be regarded as forming a flock. This way, the traditional flock existing between moving objects can be extended to geographical flock existing between geographical units. Given the importance of discovering flock patterns in traditional trajectory data and the shortage of flock pattern discoveries from new attribute trajectory data, this paper aims to discover geographical flock patterns from carbon emission data.

To achieve the discovery of geographical flock patterns, we propose a spatiotemporal graph (STG)-based approach. The proposed approach includes three main parts: first, the carbon emission data are transformed to attribute trajectory data; second, the STGs are generated from attribute trajectory data; third, specific types of geographical flock patterns are discovered from STGs. We adopt two criteria (i.e., the high–low attribute values criterion, and the extreme number–duration values criterion) to derive various types of geographical flock patterns. In short, four corresponding but different types of geographical flock patterns can be derived according to each criterion. A case study is conducted to verify the usefulness and applicability of the proposed approach, in which the original province-level CO₂ emission data in China are employed. The geographical flock patterns of CO₂ emissions are discovered at two levels, i.e., the province level and the geographical region level. The findings of this study may provide us with potential suggestions to assist policy making and the coordinated control of carbon emissions. The framework of this study is shown in Figure 1. According to this framework, important patterns (i.e., the geographical flock patterns in this paper) can be discovered from carbon emission data, and effective policies in reducing CO₂ emissions may thus be formulated with the support of the discovered important patterns.

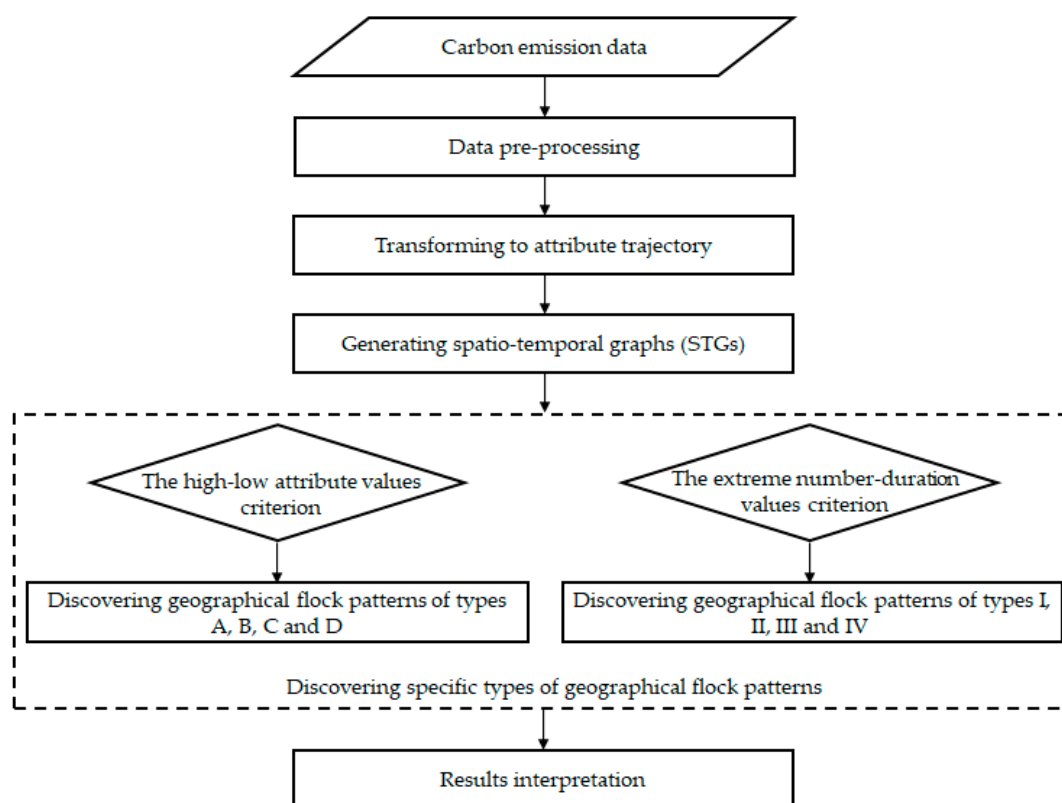


Figure 1. The framework of this study.

2. Materials and Methods

2.1. Basic Concepts

To facilitate understanding, we introduce three basic concepts in the following, namely bivariate time series, attribute trajectory and geographical flock.

2.1.1. Bivariate Time Series

Definition 1 (Time series). Given an object O , assume the value of a specific attribute of O at timestamp t_i is v_i (where $1 \leq i \leq n$), then the continuous segments of $TS = \{(v_1, t_1), (v_2, t_2), \dots, (v_n, t_n)\}$ is considered as a time series.

Definition 2 (Bivariate time series). Given an object O , assume the values of two specific attributes of O at timestamp t_i are u_i and (where $1 \leq i \leq n$), respectively, then the two continuous segments of $TS_1 = \{(u_1, t_1), (u_2, t_2), \dots, (u_n, t_n)\}$ and $TS_2 = \{(v_1, t_1), (v_2, t_2), \dots, (v_n, t_n)\}$ in the same coordinate system is considered as consisting of a bivariate time series.

2.1.2. Attribute Trajectory

Definition 3 (Trajectory). Given a moving object O , assume its spatial location at t_i during the movement path are (x_i, y_i) (where $1 \leq i \leq n$), then $Traj = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)\}$ is considered as a trajectory.

Note that in Definition 3, (x, y) denotes the spatial location (e.g., latitude and longitude) of a moving object. In addition to spatial location, (x, y) can denote other attributes as well. If the 2D spatial locations are replaced by two other attributes in a corresponding attribute space, then an attribute trajectory can be generated. The definition of attribute trajectory is given in Definition 4.

Definition 4 (Attribute trajectory). Given an object O , assume its bivariate time series is $TS_1 = \{(u_1, t_1), (u_2, t_2), \dots, (u_n, t_n)\}$ and $TS_2 = \{(v_1, t_1), (v_2, t_2), \dots, (v_n, t_n)\}$, then $AttrTraj = \{(u_1, v_1, t_1), (u_2, v_2, t_2), \dots, (u_n, v_n, t_n)\}$ is considered as an attribute trajectory.

An illustration of trajectory and attribute trajectory is shown in Figure 2. Note that Figure 2a illustrates a traditional trajectory, and Figure 2b illustrates a novel attribute trajectory. According to Figure 2, we can observe that one of the most significant differences between the two kinds of trajectories is the plane space in the coordinate systems: for the traditional trajectory, its plane space is usually a geographical space, but for the novel attribute trajectory, its coordinate system is generally an attribute space. In the attribute coordinate system, each of the two axes denotes a corresponding attribute. Given the high similarity in the structures of a traditional trajectory and an attribute trajectory, traditional trajectory mining techniques can be adapted to mine information for attribute trajectory.

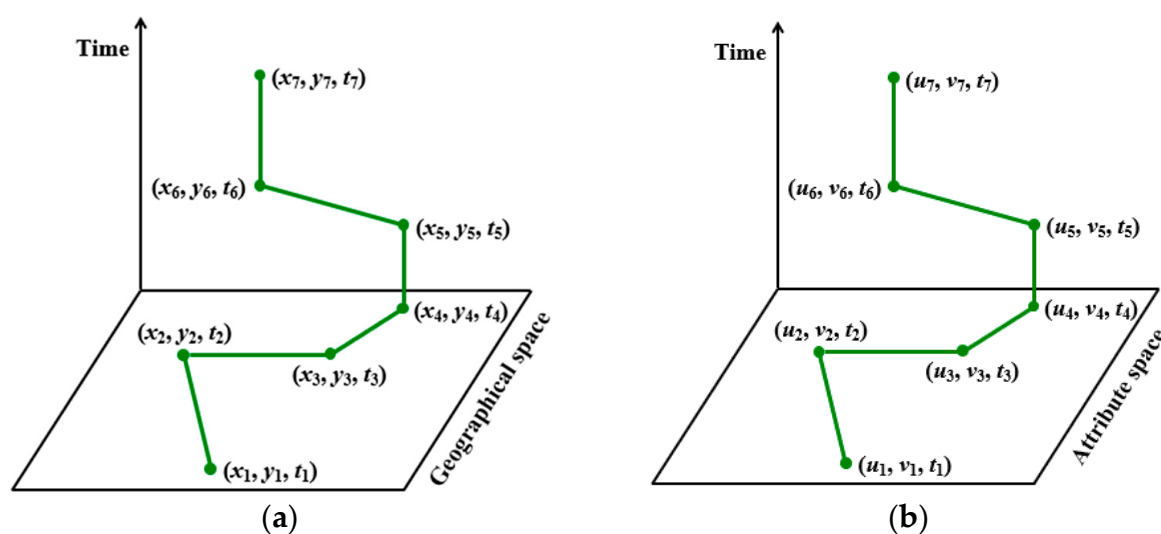


Figure 2. Illustration of traditional trajectory and attribute trajectory: (a) a traditional trajectory, and (b) a novel attribute trajectory.

2.1.3. Geographical Flock

Definition 5 (Flock). Given a set of n trajectories of n moving objects, an (r, m, k) -flock F during a time interval $I = [t_i, t_j]$ (where $j - i + 1 \geq k$) consists of at least m objects such that for each discrete timestamp t_i there exists a disk of radius r containing all m objects.

Definition 6 (Geographical flock). Given a set of n attribute trajectories of n geographical units, an (r, m, k) -geographical flock GF during a time interval $I = [t_i, t_j]$ (where $j - i + 1 \geq k$) consists of at least m geographical units such that for each discrete timestamp t_i there exists a disk of radius r containing all m attribute values of geographical units.

Figure 3 gives an illustration of flock and geographical flock. Figure 3a illustrates a flock, wherein three objects (i.e., O_1 , O_2 and O_3) form a flock during the time period from t_2 to t_4 , and Figure 3b illustrates a geographical flock, which indicates that four geographical units (i.e., G_1 , G_2 , G_3 and G_4) can form a geographical flock (assuming their attribute values at the same timestamp during a time duration are within a given threshold). To assist understanding, a geographical flock can be informally understood as “a group of geographical units with similar attribute values lasting for a specific time duration.”

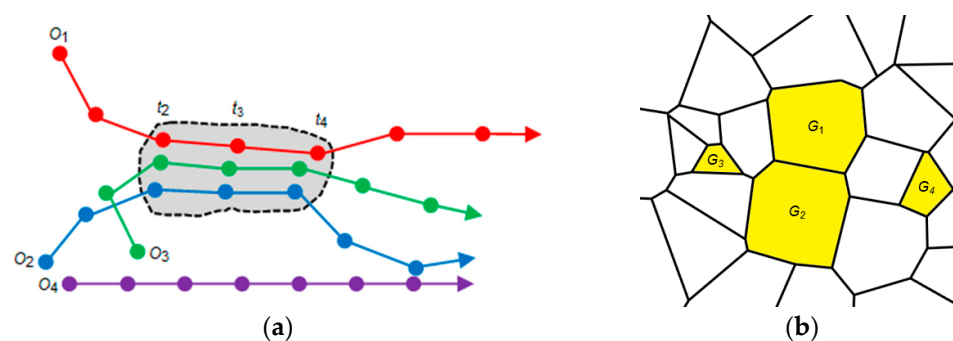


Figure 3. Illustration of flock and geographical flock: (a) a flock, and (b) a geographical flock.

2.2. Study Area and Data

China includes 23 provinces, five autonomous regions, four municipalities, and two special administrative regions. Due to the problem of missing data from the provinces of Tibet, Taiwan, Hong Kong (special administrative region) and Macao (special administrative region), we take only the other remaining provinces/autonomous regions/municipalities as the study area. To facilitate interpretation, we group all of these under the term “provinces”. Therefore, 30 provinces are involved in the study area. As we aim to discover geographical flock patterns to support the coordinated formulation of potential policies in controlling carbon emissions, we also conduct our case study on the level of geographical region, which includes seven altogether. The detailed information of the 30 provinces and the seven geographical regions are listed in Table 1, in which the names and IDs of all the provinces and geographical regions are included. The visualization of the study area is shown in Figure 4, in which Figure 4a denotes the 30 provinces and Figure 4b the seven geographical regions. Note that in Figure 4a, the white color indicates that the corresponding province is not included in the study area.

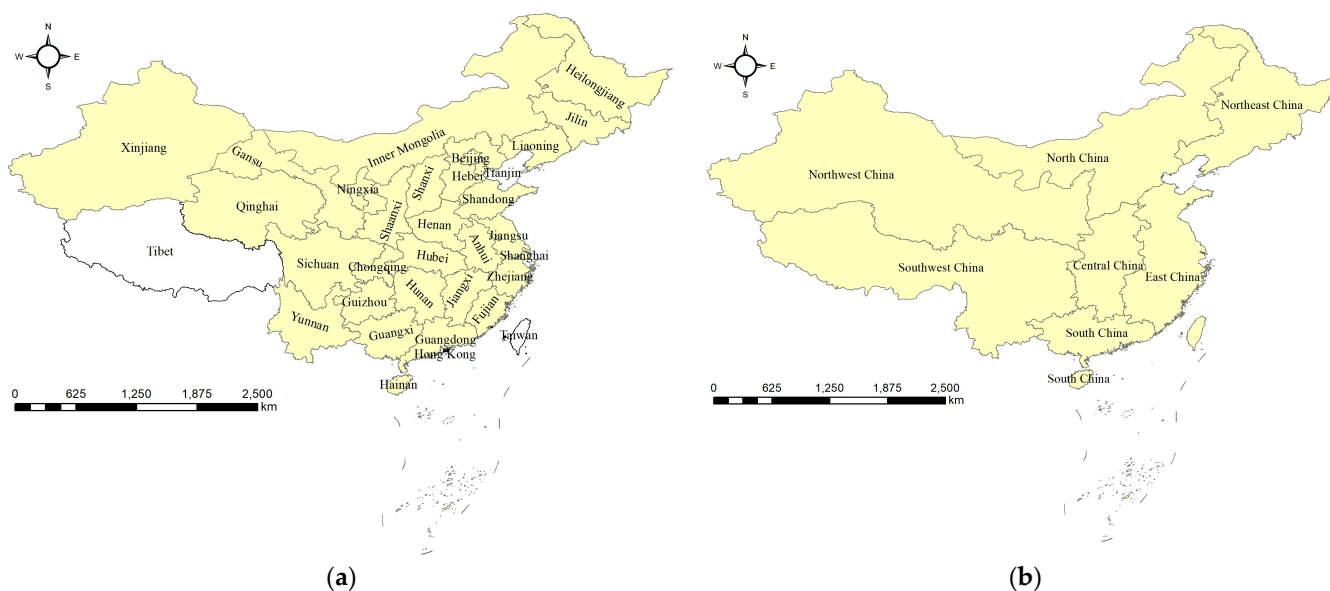


Figure 4. The visualization of the study area: (a) the 30 provinces, and (b) the seven geographical regions.

The data used in this study were acquired from the Carbon Emission Accounts and Datasets for emerging economies (CEADs) [36], which provides datasets related to carbon emissions in China at either province level, prefecture-level city level, or county level with different time span (year as the unit) to the public. The data of province-level CO₂ emissions adopted in this study include the exact information of CO₂ emissions of the 30 provinces. The time span of the data is from 1998 to 2019, and the temporal resolution is one year.

Specifically, the data contain the sectoral CO₂ emissions inventory for all 30 provinces. We use the total consumption of all sectoral CO₂ emissions inventory as the value of the total CO₂ emissions for each province. Note that we adopt the linear interpolation methods to derive the corresponding values for the two missing data. Two important attributes, i.e., the total CO₂ emissions per year and the growth rate of total CO₂ emissions per year, which are studied in a previous work [37], are used to generate the corresponding attribute trajectory of each province and geographical region.

Table 1. The detailed information of the provinces and geographical regions included in the study area.

Province	ID	Geographical Region	ID
Beijing	1	North China	1
Tianjin	2		
Hebei	3		
Shanxi	4		
Inner Mongolia	5		
Liaoning	6	Northeast China	2
Jilin	7		
Heilongjiang	8		
Shanghai	9	East China	3
Jiangsu	10		
Zhejiang	11		
Anhui	12		
Fujian	13		
Jiangxi	14		
Shandong	15		
Henan	16	Central China	4
Hubei	17		
Hunan	18		
Guangdong	19	South China	5
Guangxi	20		
Hainan	21		
Chongqing	22	Southwest China	6
Sichuan	23		
Guizhou	24		
Yunnan	25		
Shaanxi	26	Northwest China	7
Gansu	27		
Qinghai	28		
Ningxia	29		
Xinjiang	30		

2.3. Methodology

We develop an STG-based approach to discover geographical flock patterns from bivariate time series data (e.g., CO₂ emission data). The methodology includes three main parts: first, the CO₂ emission data are transformed to attribute trajectory data; second, the STGs are generated from attribute trajectory data; third, specific types of geographical flock patterns are discovered from the generated STGs. In the following, each part will be described in more detail.

2.3.1. Generating Attribute Trajectory Data

As introduced in Definition 2, a bivariate time series has two different attributes. The core of generating a corresponding attribute trajectory from a bivariate time series can be seen in Definition 4. Note that to make effective comparisons between the values of the two attributes (i.e., the total CO₂ emission, and the growth rate of total CO₂ emission), it is

necessary to consider a normalization operation. The adopted normalization method is the Z-Score method, which is represented in Equation (1):

$$v_norm_i = \frac{v_i - E}{\sigma} \quad (1)$$

where v_norm is the normalized value of the original attribute value v , and E and σ are the mean value and the standard deviation of all the original attribute values, respectively. According to this method, the attribute trajectory data of all geographical units (i.e., provinces and geographical regions) can be fully generated.

2.3.2. Generating STGs

The STGs are generated based on regularly sampled attribute trajectory data. Similar to other types of graphs, a STG can also be represented as $G = \{V, E\}$, where $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{v_i v_j\} (1 \leq i \leq n, 1 \leq j \leq n)$ denote the set of vertices and edges, respectively. Two steps are required to construct an STG: the generation of vertices and the construction of edges.

As mentioned in Definition 6, three essential parameters (i.e., r , m and k) are required in a geographical flock. The parameter r is considered when generating vertices. If the “locations” of geographical units in the attribute space at the same timestamp are within a circle of radius r whose value is user-defined, then the corresponding geographical units are considered to be involved in the same vertex. It should be noted that the selection of the base geographical unit (i.e., the geographical unit whose location at a timestamp is regarded as the center of the corresponding circle) is important. In this approach, we propose that for each timestamp, the geographical unit whose distance in the attribute space is closest to the centroid of all geographical units is considered as the base geographical unit. In this way, the vertices at each timestamp can be automatically generated. When constructing the edges, two basic principles are adopted: (1) any two vertices at the same timestamp are not allowed to be connected, and (2) any two vertices at two consecutive timestamps have to be connected if at least one common geographical unit is involved in both vertices. Thus, the edges can be fully constructed.

2.3.3. Discovering Specific Types of Geographical Flock Patterns

Once the STGs of the geographical units are generated, the geographical flock patterns can then be discovered based on the STGs. Before the discovery operation, the STGs whose time durations are less than a user-defined value of k have to be deleted. For the remaining STGs, an iteration method is adopted to find all groups of geographical units, including the IDs of geographical units and the corresponding time interval in each group. Note that for each group, the number of involved geographical units has to be at least m (which is user-defined) and the time duration has to be at least k . Thus, the groups of geographical units which meet the above conditions are considered to be geographical flocks. A geographical flock is represented as $\{IDs\ of\ geographical\ units\} \mid [start\ time, end\ time]$. For example, the geographical flock $\{1, 2, 3\} \mid [0, 5]$ indicates that it includes three geographical units whose IDs are 1, 2 and 3, respectively, and lasts for six continuous timestamps which are 0, 1, 2, 3, 4 and 5, respectively.

The geographical flock patterns can be further classified into different types. Based on more refined information, specific types of geographical flock patterns can be derived, according to which more elaborate insights may be provided in practice. In this paper, we propose to use two specific criteria to derive various types of geographical flock patterns, i.e., the high–low attribute values criterion, and the extreme number–duration values criterion. In detail, the high–low attribute values criterion is adopted to derive types of geographical flocks whose attribute values meet certain conditions (e.g., higher or lower than a pre-defined threshold represented by the corresponding parameter of *high_threshold* and *low_threshold*), and the extreme number–duration values criterion is utilized to derive types of geographical flocks whose number of members (i.e., geographical units) and time

duration have extreme values (e.g., maximum/minimum/longest/shortest). As for the high–low attribute values criterion, since there are two attributes, and the value of each attribute may be higher than or lower than a threshold, four (i.e., 2×2) specific types of geographical flock patterns can be derived. For the extreme number–duration values criterion, the extreme number can be either maximum or minimum, and the extreme duration can be either longest or shortest, thus, four (i.e., 2×2) specific types of geographical flock patterns can be derived in either case. The derived types of geographical flock patterns are listed in detail in Table 2. To avoid confusion, we use two different encoding methods (i.e., A, B, C, D and I, II, III, IV) to distinguish the types of geographical flock patterns, and each type of geographical flock pattern is represented by a corresponding type ID.

Table 2. The specific types of geographical flock patterns.

Criterion	Explanation	Type ID
The high–low attribute values	High X attribute value and high Y attribute value	A
	High X attribute value and low Y attribute value	B
	Low X attribute value and high Y attribute value	C
	Low X attribute value and low Y attribute value	D
The extreme number–duration values	Maximum number and longest duration	I
	Maximum number and shortest duration	II
	Minimum number and longest duration	III
	Minimum number and shortest duration	IV

For types A, B, C and D, a threshold is defined to distinguish the cases of high and low values. We take the percentile method, in which we only give a desired percentile value for distinguishing the high case and the low case, so that the real attribute values for the corresponding high case and low case can be automatically determined. The real attribute values are considered the thresholds. This can effectively avoid the difficulty of giving real attribute values as thresholds in reality. Based on this, the corresponding types of geographical flock patterns meeting the conditions of high and low cases can be extracted. For types I, II, III and IV, we first calculate the number of members and time duration for each discovered geographical flock, we then find the maximum value and minimum value for the number of members and time duration, respectively, and finally extract the geographical flocks meeting the corresponding conditions. Thus, all the specific types of geographical flock patterns can be extracted. Nevertheless, in reality it cannot be guaranteed that all specific types can be discovered simultaneously for each criterion under the same combination of parameter values. One can select the optimal results by adjusting parameter values and considering his/her specific desires.

3. Results and Discussion

The results are presented on two levels: the province level and the geographical region level. As introduced in Definition 6, three essential parameters (i.e., r , m and k) are involved in geographical flock, and different combination of parameter values may lead to different results. Therefore, it is necessary to determine suitable parameter values when discovering geographical flock patterns. An important principle for determining suitable parameter values is that a fit number (i.e., neither too large nor too small) of geographical flocks has to be discovered. This is because, if the number of discovered geographical flocks is too large (or too small), it may provide too much (or insufficient) information, which can lead to corresponding difficulties in interpretation. Based on this, for parameter r , the smaller its value, the better, because a small value indicates that the changes of attributes are in a small fluctuation range; for parameters m and k , larger values are preferred, because large values indicate that the geographical flock patterns that are more meaningful may be discovered. By considering the strategies for selecting potential parameter values mentioned above, we have tested a large number of different combinations of parameter values. Due to

space limitations, only the significant geographical flock patterns from our perspective are presented in detail in the following.

3.1. Geographical Flock Patterns on the Province Level

As mentioned in Section 2.3.3, two criteria are proposed to derive specific types of geographical flock patterns, therefore, the results based on each criterion at the province level will be presented below.

3.1.1. The High–Low Attribute Values

Four representative geographical flocks were discovered based on the criterion of the high–low attribute values under the combination of parameter values for $r = 15$, $m = 3$, $k = 3$, $high_threshold = 70$, $low_threshold = 30$. The detailed information of the four geographical flocks can be seen in Table 3. From Table 3 we can see that, among all the geographical flocks, one belongs to type C and the other three belong to type B, while none were discovered for types A and D. This is reasonable, because it cannot be guaranteed that all types of geographical flock patterns can be discovered simultaneously under the same combination of parameter values.

Table 3. The discovered geographical flock patterns based on the criterion of the high–low attribute values on the province level.

ID	Geographical Flock	Corresponding Provinces	Type ID
1	{11, 18, 19} [2012, 2014]	Zhejiang, Hunan, Guangdong	B
2	{4, 11, 19} [2013, 2015]	Shanxi, Zhejiang, Guangdong	B
3	{6, 11, 17, 23} [2014, 2016]	Liaoning, Zhejiang, Hubei, Sichuan	B
4	{21, 29, 30} [2002, 2004]	Hainan, Ningxia, Xinjiang	C

According to the four geographical flocks, we can see that three groups of provinces had both a high amount and a low growth rate of total CO₂ emission during three continuous years (corresponding to type B). Specifically, the three groups and the corresponding continuous years are the provinces of Zhejiang, Hunan and Guangdong from 2012 to 2014, the provinces of Shanxi, Zhejiang and Guangdong from 2013 to 2015, and the provinces of Liaoning, Zhejiang, Hubei and Sichuan from 2014 to 2016. Generally speaking, the provinces in the same geographical flock performed well in controlling the growth rate of total CO₂ emission in the corresponding years, which indicates that the measures and policies adopted have been effective. However, in the corresponding years, their total amounts of CO₂ emissions were still very high, which demonstrates that more effective measures may be taken to control the total amount of carbon emissions. Secondly, we can see that one group of provinces had both a low amount and a high growth rate of total CO₂ emission (corresponding to type C). The specific provinces and the corresponding years are Hainan, Ningxia and Xinjiang from 2002 to 2004. It can be inferred that the three provinces have performed well in controlling the total amount of CO₂ emission, but that the effects of controlling the growth rate of total CO₂ emissions were not that satisfactory.

To further explore the spatial relations of the provinces in each geographical flock, we visualize each geographical flock on the map by setting a different color. The visualization is shown in Figure 5, in which each figure corresponds to a geographical flock, and the provinces involved in the same geographical flock are denoted the same color. From Figure 5, we can see that there appears to be a stronger spatial relation for the provinces in the geographical flock type B (Figure 5a–c) than type C (Figure 5d), as the provinces involved in the geographical flock type B (Figure 5a–c) are relatively close to each other in space, while the provinces involved in the geographical flock type C (Figure 5d) have a relatively further geographical distance from each other. Additionally, an interesting finding is that the provinces involved in type B (Figure 5a–c) have generally stronger comprehensive strength than those involved in type C (Figure 5d). This indicates that the

related factors (such as economy, population and industry) of the provinces in the same type of geographical flock pattern may be similar, an obvious difference in the related factors may exist in the provinces involved in different types of geographical flock patterns. However, this still needs further exploration. In summary, the geographical flock patterns discovered based on this criterion reveal several interesting findings, which can be fully considered when conducting inter-provincial collaborations and when making coordinated policies to effectively control the amount and growth rate of carbon emissions.

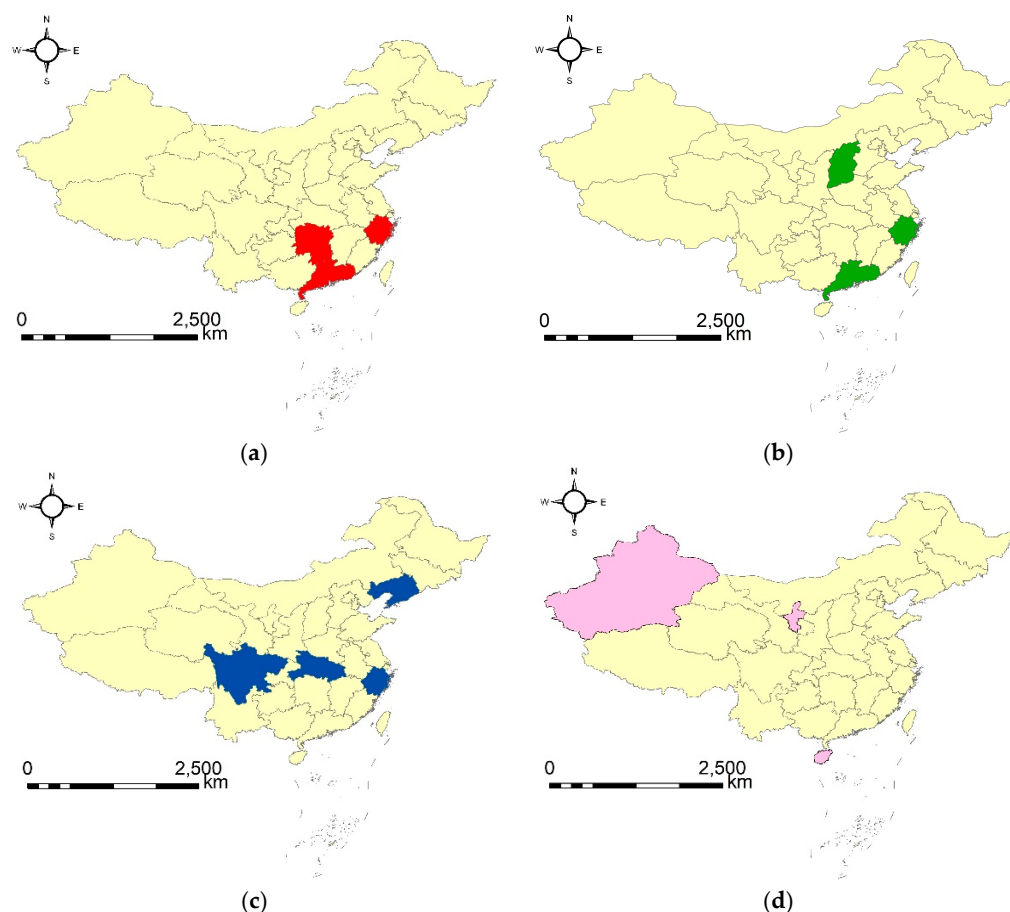


Figure 5. The discovered geographical flock patterns: (a) {11, 18, 19} | [2012, 2014] (type B), (b) {4, 11, 19} | [2013, 2015] (type B), (c) {6, 11, 17, 23} | [2014, 2016] (type B), and (d) {21, 29, 30} | [2002, 2004] (type C).

3.1.2. The Extreme Number–Duration Values

Two significant geographical flocks were discovered based on the criterion of the extreme number–duration values under the combination of parameter values for $r = 10$, $m = 5$, and $k = 3$. The full information of the two geographical flocks is shown in Table 4, from which we can observe that one belongs to type II and the other belongs to type III. As for types I and IV, none has ever been discovered under this specific combination of parameter values.

Table 4. The discovered geographical flock patterns based on the criterion of the extreme number–duration values on the province level.

ID	Geographical Flock	Corresponding Provinces	Type ID
1	{1, 4, 11, 12, 14, 16, 17, 20, 27, 30} [1998, 2000]	Beijing, Shanxi, Zhejiang, Anhui, Jiangxi, Henan, Hubei, Guangxi, Gansu, Xinjiang	II
2	{8, 11, 12, 17, 23} [2014, 2017]	Heilongjiang, Zhejiang, Anhui, Hubei, Sichuan	III

The results show that a maximum number of ten provinces have had similar evolution patterns in both the amount and the growth rate of total CO₂ emission in a shortest duration of three continuous years. The specific provinces and the corresponding years are Beijing, Shanxi, Zhejiang, Anhui, Jiangxi, Henan, Hubei, Guangxi, Gansu and Xinjiang from 1998 to 2000. This demonstrates that the ten provinces formed a maximum group which had similar CO₂ emissions and lasted for three continuous years. This finding would be applicable to meet the needs of detecting the largest number of provinces with similar CO₂ emission so that closer inter-provincial cooperation may be carried out. Secondly, a minimum number of five provinces had similar evolution patterns in both the amount and the growth rate of total CO₂ emission in a maximum duration of four continuous years. The specific provinces and the corresponding years are Heilongjiang, Zhejiang, Anhui, Hubei and Sichuan from 2014 to 2017. This shows that the five provinces have had a similar evolution pattern of CO₂ emission in a maximum duration of four years. Therefore, if one would like to know the provinces which lasted for the longest duration, this finding can provide the ideal answer. In our view, the results can provide useful suggestions to related governmental departments. Furthermore, one can detect the very groups of provinces which have had similar evolution patterns in carbon emissions by adjusting the values of the three parameters to meet his/her specific demands.

The visualization of the two geographical flocks can be seen in Figure 6, which gives an overview of the spatial relations of the provinces involved in the same geographical flock. From Figure 6 we can see that Figure 6a exhibits an overall strong spatial relation for all the involved provinces, and Figure 6b presents a strong spatial relation for most of the involved provinces. Therefore, we can infer that geographical locations may have strong effects on the potential groups of provinces which can form a specific type of geographical flock, but other factors can also have particular effects on the final formulation of potential geographical flocks. The findings indicate that further and finer explorations may be conducted to gain further insight on why the provinces with relatively weak spatial relations can form particular geographical flocks so that more scientific, precise and flexible policies and/or strategies can be made in the future.

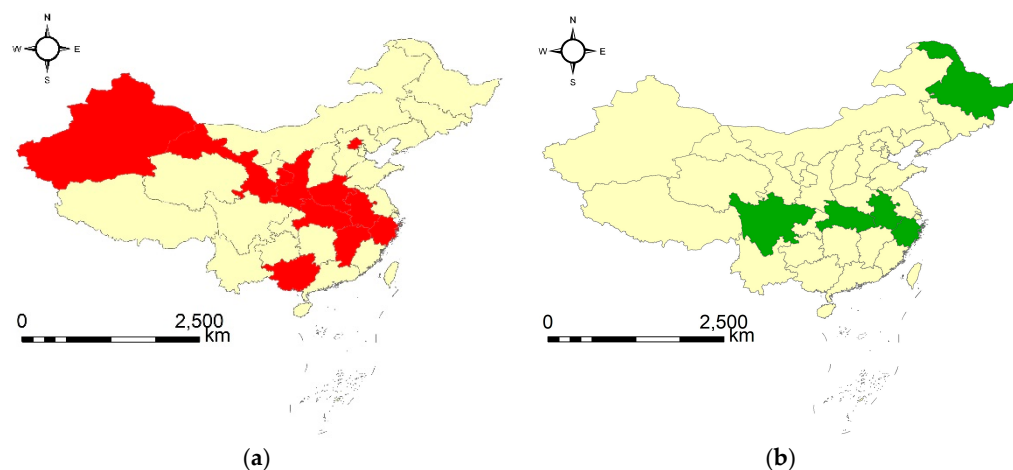


Figure 6. The discovered geographical flock patterns: (a) {1, 4, 11, 12, 14, 16, 17, 20, 27, 30} | [1998, 2000] (type II), and (b) {8, 11, 12, 17, 23} | [2014, 2017] (type III).

3.2. Geographical Flock Patterns on the Geographical Region Level

The discovered geographical flock patterns based on each criterion on the geographical region level will be presented in detail in the following.

3.2.1. The High–Low Attribute Values

Two significant geographical flocks were discovered under the criterion of the high–low attribute values. The exact information of the two geographical flocks is listed in Table 5. From Table 5 we can see that between the two geographical flocks, one belongs to type B and the

other belongs to type C, and none have been discovered for types A and D under this specific combination of parameter values (i.e., $r = 10$, $m = 2$, $k = 3$, $low_threshold = 40$, $high_threshold = 60$).

Table 5. The discovered geographical flock patterns based on the criterion of the high–low attribute values on the geographical region level.

ID	Geographical Flock	Corresponding Provinces	Type ID
1	{2, 4} [2015, 2018]	Northeast China, Central China	B
2	{2, 5} [2005, 2007]	Northeast China, South China	C

The two geographical flocks reveal that one group of geographical regions has had both a high amount and a low growth rate of total CO₂ emission during three continuous years (corresponding to type B). The specific geographical regions and corresponding continuous years are Northeast China and Central China from 2015 to 2018. Secondly, one group of geographical regions has had both a low amount and a high growth rate of total CO₂ emission (corresponding to type C). The corresponding geographical regions and continuous years are Northeast China and South China from 2005 to 2007. Note that a common geographical region involved in both groups is Northeast China. According to the results, it can be inferred that Northeast China may have taken effective measures in controlling the growth rate of CO₂ emissions, as it has been keeping a steady low growth rate in recent years (2015 ~ 2018) while the growth rate was relatively high in years prior to that (2005 ~ 2007). Based on the results, potential suggestions may be provided to related national governmental departments to carry out effective regional cooperation to work out more targeted policies in better controlling carbon emissions.

Figure 7 exhibits the visualization of the two geographical flocks. From Figure 7 we can see that the geographical regions in geographical flock type B (Figure 7a) have relatively stronger spatial relations than those in type C (Figure 7b), which coincides well with the corresponding finding in Section 3.1.1. This may provide us additional useful clues on how to produce more scientific strategies to better control carbon emissions in the future.

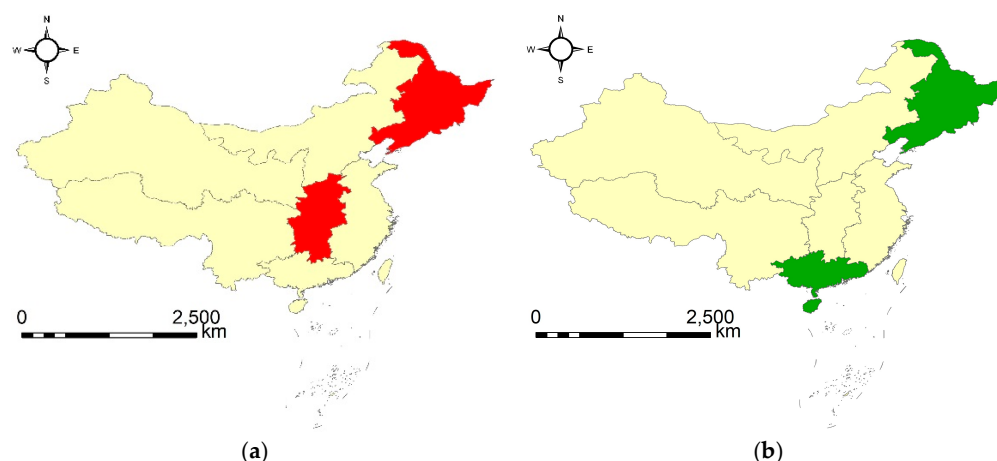


Figure 7. The discovered geographical flock patterns: (a) {2, 4} | [2015, 2018] (type B), and (b) {2, 5} | [2005, 2007] (type C).

3.2.2. The Extreme Number–Duration Values

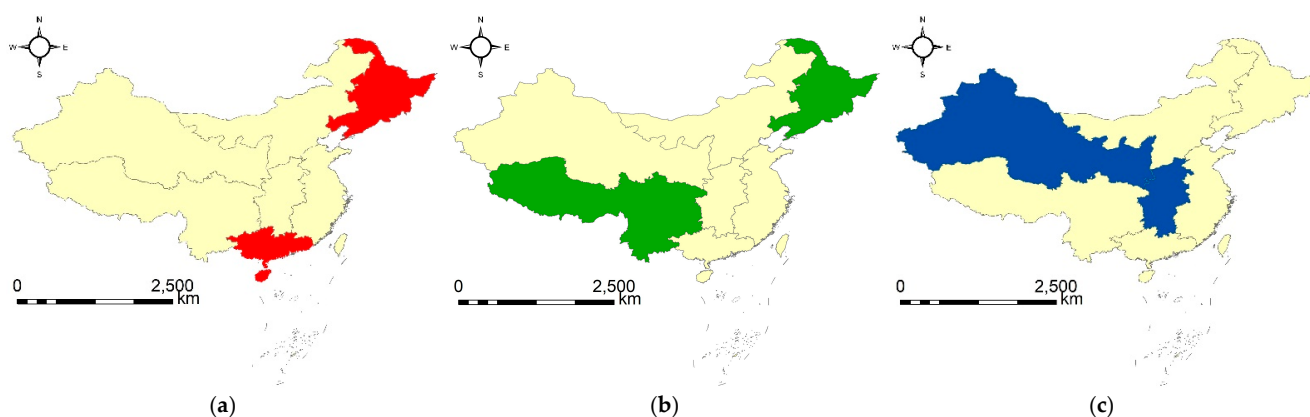
Three typical geographical flocks were discovered under the criterion of the extreme number–duration values. The specific information of all the geographical flocks can be seen in Table 6, from which we can see that, among the three geographical flocks, two belong to both type I and type III, and the other belongs to both type II and type IV. An obvious difference of these results from the previous results is that all types of geographical flock patterns have been discovered under this specific combination of parameter values (i.e., $r = 5$, $m = 2$, $k = 3$).

Table 6. The discovered geographical flock patterns based on the criterion of the extreme number-duration values on the geographical region level.

ID	Geographical Flock	Corresponding Provinces	Type ID
1	{2, 5} [2005, 2008]	Northeast China, South China	I, III
2	{2, 6} [2013, 2016]	Northeast China, Southwest China	I, III
3	{4, 7} [2014, 2016]	Central China, Northwest China	II, IV

The main findings based on the three geographical flocks can be summarized as follows. (1) Two groups of geographical regions have had similar evolution patterns in both the amount and the growth rate of total CO₂ emission in a maximum duration of four continuous years. The number of geographical regions involved in each group is two, which is both the maximum (corresponding to type I) and the minimum (corresponding to type III) among all the discovered geographical flock patterns; the corresponding geographical regions and continuous years are Northeast China and South China from 2005 to 2008, and Northeast China and Southwest China from 2013 to 2016, respectively. (2) One group of geographical regions has had similar evolution patterns in both the amount and the growth rate of total CO₂ emission in a minimum duration of three continuous years. Similar to the other two geographical flocks, the number of geographical regions involved in this geographical flock is both the maximum and the minimum, thus it belongs to both type II and type IV. The detailed information of this geographical flock is the geographical regions of Central China and Northwest China from 2014 to 2016. According to the results, the abovementioned geographical regions involved in the same geographical flock may carry out closer cooperation to determine more scientific measures to better control and release the emissions of carbons. For example, Northeast China may cooperate closely with other regions, such as South China and Southwest China, to explore the specific reasons why similar emission patterns have appeared.

The corresponding visualization of the three geographical flocks is shown in Figure 8, from which the spatial distribution of the geographical regions involved in the same geographical flock can be clearly seen. An interesting finding is that the geographical flock of type II/IV (Figure 8c) exhibits strong spatial relations, while the spatial relations of the geographical flock of type I/III (Figure 8a,b) appear relatively weak. By considering the similar findings in spatial relations in Section 3.1.2, related departments may gain additional insights to propose more elaborate plans so that carbon emissions can be controlled in a more scientific way in the future.

**Figure 8.** The discovered geographical flock patterns: (a) {2, 5} | [2005, 2008] (type I, III), (b) {2, 6} | [2013, 2016] (type I, III), and (c) {4, 7} | [2014, 2016] (type II, IV).

4. Conclusions

Climate change has become one of the greatest global challenges, and one which can greatly affect humankind. A significant solution for mitigating climate change is to reduce

the emissions of greenhouse gas, particularly CO₂. To better support the establishment of effective policies for reducing CO₂, it is crucial to consider specific sorts of important emission patterns that exist between provinces and/or geographical regions. This paper takes geographical flock patterns as the very important kind of emission pattern which deserves further investigation. We propose an STG-based approach to effectively discover geographical flock patterns. The approach mainly includes three steps, i.e., generating attribute trajectories from CO₂ emission data, generating STGs from attribute trajectories, and discovering specific types of geographical flock patterns. In general, eight different types of geographical flock patterns are derived based on two different criteria (i.e., the high–low attribute values criterion and the extreme number–duration values criterion). A case study was conducted on two levels, i.e., the province level and the geographical region level, based on CO₂ emission data in China. The results of the case study demonstrate that the proposed approach is effective in discovering the different types of geographical flock patterns, and potentially useful suggestions and insights can be provided to related departments to assist in policy making and in the coordinated control of carbon emissions in the future.

Given the importance of investigating the evolution patterns of CO₂ emission between different geographical units, we only took the province-level carbon emission data for case studies. Although the results based on the province-level CO₂ emission data appear effective, finer results are still needed. Therefore, fine-granularity CO₂ emission data (e.g., data at the city-level) can be adopted as new datasets for further studies to obtain more precise insight. In addition, other attributes which are meaningful to carbon emission can be used to generate attribute trajectories so that the insights and findings can be further extended. Furthermore, new criteria can be developed and adopted to derive specific types of geographical flock patterns according to one's specific desire, according to which new findings might be acquired.

Author Contributions: Conceptualization, P.Z.; data curation, X.L.; formal analysis, P.Z., L.M. and F.W.; funding acquisition, P.Z. and L.M.; investigation, P.Z.; methodology, P.Z. and L.M.; software, X.L.; validation, P.Z.; visualization, X.L.; writing—original draft, P.Z.; writing—review and editing, L.M. and F.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grants from the Introduction Program of High-Level Innovation and Entrepreneurship Talents in Jiangsu Province (Grant No. CZ0325C20025), the NUPTSF (Grant No. NY219140), the Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province (Grant No. SHEL221004), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19090122), and the Central Public-interest Scientific Institution Basal Research Fund, CAFS (Grant No. 2018HY-ZD0103).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

Conflicts of Interest: The authors declare no competing interest.

References

1. Zhang, X.; Zhao, X.; Jiang, Z.; Shao, S. How to achieve the 2030 CO₂ emission-reduction targets for China's industrial sector: Retrospective decomposition and prospective trajectories. *Glob. Environ. Chang.* **2017**, *44*, 83–97. [[CrossRef](#)]
2. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A. Tackling climate change with machine learning. *ACM Comput. Surv.* **2022**, *55*, 1–96. [[CrossRef](#)]
3. Abbas, A.; Zhao, C.; Waseem, M.; Ahmad, R. Analysis of energy input–output of farms and assessment of greenhouse gas emissions: A case study of cotton growers. *Front. Environ. Sci.* **2022**, *725*, 826838. [[CrossRef](#)]
4. Elahi, E.; Khalid, Z.; Tauni, M.; Zhang, H.; Lirong, X. Extreme weather events risk to crop-production and the adaptation of innovative management strategies to mitigate the risk: A retrospective survey of rural Punjab, Pakistan. *Technovation* **2022**, *117*, 102255. [[CrossRef](#)]

5. Elahi, E.; Khalid, Z. Estimating smart energy inputs packages using hybrid optimisation technique to mitigate environmental emissions of commercial fish farms. *Appl. Energy* **2022**, *326*, 119602. [\[CrossRef\]](#)
6. Lahsen, M.; Ribot, J. Politics of attributing extreme events and disasters to climate change. *Wiley Interdiscip. Rev. Clim. Chang.* **2022**, *13*, e750. [\[CrossRef\]](#)
7. IPCC. Intergovernmental Panel on Climate Change. 2022. Available online: <https://www.ipcc.ch/sr15/> (accessed on 13 December 2022).
8. Ren, L.; Zhou, S.; Peng, T.; Ou, X. A review of CO₂ emissions reduction technologies and low-carbon development in the iron and steel industry focusing on China. *Renew. Sustain. Energy Rev.* **2021**, *143*, 110846. [\[CrossRef\]](#)
9. Friedlingstein, P.; Jones, M.W.; O'Sullivan, M.; Andrew, R.M.; Bakker, D.C.; Hauck, J.; Le Quéré, C.; Peters, G.P.; Peters, W.; Pongratz, J. Global carbon budget 2021. *Earth Syst. Sci. Data* **2022**, *14*, 1917–2005. [\[CrossRef\]](#)
10. Li, J.; Ho, M.S.; Xie, C.; Stern, N. China's flexibility challenge in achieving carbon neutrality by 2060. *Renew. Sustain. Energy Rev.* **2022**, *158*, 112112. [\[CrossRef\]](#)
11. Huang, Z.; Du, X. Toward green development? Impact of the carbon emissions trading system on local governments' land supply in energy-intensive industries in China. *Sci. Total Environ.* **2020**, *738*, 139769. [\[CrossRef\]](#)
12. Wen, L.; Yuan, X. Forecasting CO₂ emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Sci. Total Environ.* **2020**, *718*, 137194. [\[CrossRef\]](#)
13. Xuan, D.; Ma, X.; Shang, Y. Can China's policy of carbon emission trading promote carbon emission reduction? *J. Clean. Prod.* **2020**, *270*, 122383. [\[CrossRef\]](#)
14. Chen, X.; Lin, B. Towards carbon neutrality by implementing carbon emissions trading scheme: Policy evaluation in China. *Energy Policy* **2021**, *157*, 112510. [\[CrossRef\]](#)
15. Ye, L.; Yang, D.; Dang, Y.; Wang, J. An enhanced multivariable dynamic time-delay discrete grey forecasting model for predicting China's carbon emissions. *Energy* **2022**, *249*, 123681. [\[CrossRef\]](#)
16. Cai, M.; Ren, C.; Shi, Y.; Chen, G.; Xie, J.; Ng, E. Modeling spatiotemporal carbon emissions for two mega-urban regions in China using urban form and panel data analysis. *Sci. Total Environ.* **2023**, *857*, 159612. [\[CrossRef\]](#)
17. Liu, Z.; Deng, Z.; He, G.; Wang, H.; Zhang, X.; Lin, J.; Qi, Y.; Liang, X. Challenges and opportunities for carbon neutrality in China. *Nat. Rev. Earth Environ.* **2022**, *3*, 141–155. [\[CrossRef\]](#)
18. Heil, M.T.; Selden, T.M. Carbon emissions and economic development: Future trajectories based on historical experience. *Environ. Dev. Econ.* **2001**, *6*, 63–83. [\[CrossRef\]](#)
19. Silvestrini, A.; Veredas, D. Temporal aggregation of univariate and multivariate time series models: A survey. *J. Econ. Surv.* **2008**, *22*, 458–497. [\[CrossRef\]](#)
20. Edinburg, T.; Eglen, S.J.; Ercole, A. Causality indices for bivariate time series data: A comparative review of performance. *Chaos* **2021**, *31*, 083111. [\[CrossRef\]](#)
21. Zhang, L.; Weghe, N.V.D. Attribute trajectory analysis: A framework to analyse attribute changes using trajectory analysis techniques. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1043–1059. [\[CrossRef\]](#)
22. Benkert, M.; Gudmundsson, J.; Hübner, F.; Wolle, T. Reporting flock patterns. *Comput. Geom.* **2008**, *41*, 111–125. [\[CrossRef\]](#)
23. Fort, M.; Sellares, J.A.; Valladares, N. A parallel GPU-based approach for reporting flock patterns. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1877–1903. [\[CrossRef\]](#)
24. Turdukulov, U.; Romero, A.; Huisman, O.; Retsios, V. Visual mining of moving flock patterns in large spatio-temporal data sets using a frequent pattern approach. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2013–2029. [\[CrossRef\]](#)
25. Sanches, D.E.; Alvares, L.O.; Bogorny, V.; Vieira, M.R.; Kaster, D.S. A top-down algorithm with free distance parameter for mining top-k flock patterns. In Proceedings of the 21st AGILE Conference on Geographic Information Science, Lund, Sweden, 12–15 June 2018; pp. 233–249.
26. Zhang, T.; He, W.; Huang, J.; He, Z.; Li, J. Interactive visual analytics of moving passenger flocks using massive smart card data. *Cart. Geogr. Inf. Sci.* **2022**, *49*, 354–369. [\[CrossRef\]](#)
27. Jeung, H.; Yiu, M.L.; Zhou, X.; Jensen, C.S.; Shen, H.T. Discovery of Convoys in Trajectory Databases. In Proceedings of the VLDB Endowment, Auckland, New Zealand, 23–28 August 2008; pp. 1068–1080.
28. Yeoman, J.; Duckham, M. Decentralized detection and monitoring of convoy patterns. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 993–1011. [\[CrossRef\]](#)
29. Liu, Y.; Dai, H.; Li, J.; Chen, Y.; Yang, G.; Wang, J. BP-Model-based convoy mining algorithms for moving objects. *Expert Syst. Appl.* **2023**, *213*, 118860. [\[CrossRef\]](#)
30. Gudmundsson, J.; Kreveld, M.V.; Speckmann, B. Efficient Detection of Patterns in 2D Trajectories of Moving Points. *Geoinformatica* **2007**, *11*, 195–215. [\[CrossRef\]](#)
31. Solera, F.; Calderara, S.; Cucchiara, R. Learning to identify leaders in crowd. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 43–48.
32. Amornbunchornvej, C.; Berger-Wolf, T.Y. Mining and modeling complex leadership–followership dynamics of movement data. *Soc. Netw. Anal. Min.* **2019**, *9*, 58. [\[CrossRef\]](#)
33. Kalnis, P.; Mamoulis, N.; Bakiras, S. On discovering moving clusters in spatio-temporal data. In Proceedings of the Advances in Spatial and Temporal Databases, Angra dos Reis, Brazil, 22–24 August 2005; pp. 364–381.

34. Patel, D. On discovery of spatiotemporal influence-based moving clusters. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 1–23. [[CrossRef](#)]
35. Loglisci, C. Using interactions and dynamics for mining groups of moving objects from trajectory data. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1436–1468. [[CrossRef](#)]
36. Carbon Emission Accounts and Datasets for Emerging Economies. Available online: <https://ceads.net> (accessed on 1 September 2022).
37. Zhang, P.; Li, H. Visually Exploring the Carbon Emissions in China from a Multi-Temporal Scale Perspective: A Case Study on the Province-Level CO₂ Emissions. In Proceedings of the 29th International Conference on Geoinformatics, Beijing, China, 15–18 August 2022; pp. 1–6.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.