A Cluster of CO\(_2\) Change Characteristics with GOSAT Observations for Viewing the Spatial Pattern of CO\(_2\) Emission and Absorption

Da Liu \(^1,2\), Liping Lei \(^1,*\), Lijie Guo \(^1,2\) and Zhao-Cheng Zeng\(^1,3\)

\(^1\) Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; E-Mails: liuda@radi.ac.cn (D.L.); leilp@radi.ac.cn (L.L.); guolijie12@gmail.com (L.G.)

\(^2\) University of Chinese Academy of Sciences, Beijing 100049, China; E-Mails: liuda@radi.ac.cn (D.L.); guolijie12@gmail.com (L.G.)

\(^3\) Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin 999077, Hong Kong, China; E-Mail: zzhaoch@gmail.com

* Author to whom correspondence should be addressed; E-Mail: leilp@radi.ac.cn; Tel.: +86-10-8217-8162; Fax: +86-10-8217-8009.

Academic Editor: Robert W. Talbot

Received: 3 September 2015 / Accepted: 22 October 2015 / Published: 5 November 2015

**Abstract:** Satellite observations can be used to detect the changes of CO\(_2\) concentration at global and regional scales. With the column-averaged CO\(_2\) dry-air mole fraction (X\(\text{CO}_\text{2}\)) data derived from satellite observations, the issue is how to extract and assess these changes, which are related to anthropogenic emissions and biosphere absorptions. We propose a k-means cluster analysis to extract the temporally changing features of X\(\text{CO}_\text{2}\) in the Central-Eastern Asia using the data from 2009 to 2013 obtained by Greenhouse Gases Observing Satellite (GOSAT), and assess the effects of anthropogenic emissions and biosphere absorptions on CO\(_2\) changes combining with the data of emission and vegetation net primary production (NPP). As a result, 14 clusters, which are 14 types of X\(\text{CO}_\text{2}\) seasonal changing patterns, are obtained in the study area by using the optimal clustering parameters. These clusters are generally in agreement with the spatial pattern of underlying anthropogenic emissions and vegetation absorptions. According to correlation analysis with emission and NPP, these 14 clusters are divided into three groups: strong emission, strong absorption, and a tendency of balancing between emission and absorption. The proposed clustering approach in this study
provides us with a potential way to better understand how the seasonal changes of CO₂ concentration depend on underlying anthropogenic emissions and vegetation absorptions.

**Keywords:** temporal variation of Xco₂; clustering; emission and absorption; Greenhouse Gases Observing Satellite – Atmospheric CO₂ Observations from Space (GOSAT-ACOS)

1. Introduction

The global carbon cycle has been changed by human activities since the beginning of the industrial era [1]. Anthropogenic emissions of CO₂, especially that from burning fossil fuels, is considered to be a major cause of the continual increase of atmospheric carbon dioxide (CO₂) concentrations [2,3]. This is the leading driving factor of climate change and global warming [4]. It is well known that CO₂ is a long-lived greenhouse gas, and the gradients generated by local fluxes are relatively small compared with background concentrations [5]. To better understand the carbon budget and combat climate change, it is extremely important to know where CO₂ is released into the atmosphere and from where it is removed [6,7]. Therefore, effective approaches for observing atmospheric CO₂ concentrations in high-quality are essential. For a long past time, ground-based observations had been the only reliable way of obtaining stable, highly accurate data of CO₂ concentrations in the atmosphere, which have helped us in understanding the global and latitudinal variations of atmospheric CO₂ concentration [8,9]. However, the sparseness of current ground-based measurement stations [7,9,10] has been limiting our knowledge of the global carbon cycle [11].

With the development of atmospheric remote sensing technology, satellite observation, with high spatial and temporal resolutions has become one of the effective approaches to monitoring the changes of greenhouse gases at regional and global scales [11–14]. Space-based remote sensing observations are expected to complement, rather than replace, ground-based measurements [15] for detecting how CO₂ concentration changes in space and time and where CO₂ is emitted and absorbed [16]. In the past few years, particularly since GOSAT was launched, satellite observations have contributed a large amount of data to help facilitate detection of the changing characteristics of atmospheric CO₂ concentrations at global and regional scales.

GOSAT was designed to provide views of real spectra from space in the short-wave infrared band (SWIR) and thermal infrared band (TIR), where CO₂ absorption lines are located [17,18]. The CO₂ concentrations, column-averaged volume mixing ratios of CO₂ (Xco₂) [19], are derived from GOSAT observing spectra and auxiliary parameters [11], which are sensitive to the atmospheric boundary layer [10,20]. These parameters have been compared and validated with ground-based measurements [21–23] and model simulations [24,25] in many studies. Satellite observation covering the globe can help us better understand the spatio-temporal changes of atmospheric CO₂ concentrations as to confirm carbon sources and sinks.

Variations of Xco₂ depend on terrestrial biosphere fluxes, anthropogenic emissions, and foreign fluxes transported by atmospheric wind fields [26]. Accordingly, the spatial distributions of anthropogenic emissions and biosphere fluxes can be characterized based on the Xco₂ variations, which can be observed by satellite. At present, there are two main ways, model-driven and data-driven, for detecting sources and sinks. Model-driven way applies the inverse modeling approach, incorporating an atmospheric transport model to deriving surface CO₂ fluxes for sources and sinks from satellite
observations [27–29]. Based on prior fluxes and meteorological data, this method takes advantage of grid cells to quantify CO₂ sources, sinks, and transmission exchanges and help us to know the mechanism of CO₂ changes. However, the uncertainty and low spatial resolutions of initial data used in the model inputs and prior fluxes, and the implicit relationships between emissions/absorptions and concentrations result in our understanding for CO₂ fluxes is not exact in space and time. The data-driven way detects sources and sinks directly by using the multi-source data including satellite observations to detect the changes of CO₂ concentration. This direct analysis of change characteristics of CO₂ concentration using satellite observations based on proper regional divisions has been shown to be an effective way to study CO₂ changes induced by emitting and absorption. For example, Kort et al. [30] explored the enhancement of atmospheric CO₂ concentrations over Los Angeles using GOSAT data. Moreover, Keppel-Aleks et al. [31] used GOSAT data to compare the differences of CO₂ concentrations between emission and upwind regions to analyze fossil fuel emissions. This data-driven way is intuitive and may reveal the real changes since satellite observations provide direct and instantaneous measurements of CO₂ concentrations.

Clustering of data is an effective analysis tool to extract valuable information from data by grouping large datasets according to their similarity [32]. This approach has been widely applied in global climate change analysis [33–35] but is still not applied in studying of CO₂ concentrations. In this paper, we propose a clustering approach of satellite CO₂ observation data based on the temporally changing characteristics of CO₂ concentrations. The objective of this approach is to study the spatial patterns from the clustering that may indicate CO₂ emissions and absorptions. Auxiliary datasets, including bottom-up emission datasets, net primary product (NPP) datasets, and land cover datasets are used to evaluate the performance of the clustering results.

2. Materials and Methods

2.1. Used Data

We chose Central-Eastern Asia (from 18°N to 55°N in latitude, and from 70°E to 140°E in longitude) as the study area. The area covers Eastern China and Northern India, where CO₂ emissions are rapidly increasing [3] on account of high population densities, rapid economic development, and significant energy consumption sharpen [36,37].

(1) Gap-filled Xco₂ Dataset

GOSAT Xco₂ data are irregularly distributed and have many gaps in space and time because of the limitation of GOSAT observation mode, cloudy block and data screening. For viewing the space-time continuous changes of CO₂ concentrations, we applied a spatio-temporal kriging interpolation method [38–40] to fill the gaps and generated a mapping dataset in 1° × 1° grid cells and ten-day interval in time (from 18°N to 55°N in latitude, and from 70°E to 140°E in longitude, from 1 June 2009 to 15 May 2013 in time). We generated two datasets from this gap-filled data. One is the monthly averaged Xco₂ dataset (M-Xco₂) containing 1415 grid cells with 47 month-averaged Xco₂ values for each grid cell. The other is the seasonally averaged Xco₂ dataset (S-Xco₂) containing 1561 grid cells with 15
season-averaged Xco2 values for each grid cell. These two datasets have been filtered based on Xco2 data integrity in time series.

The GOSAT Xco2 retrievals used here are Atmospheric CO2 Observations from Space (ACOS) Xco2 retrieval products (v3.3) from June 2009 to May 2013 (http://mirador.gsfc.nasa.gov). The ACOS Xco2 dataset was produced by the Orbiting Carbon Observatory (OCO) team of the US National Aeronautics and Space Administration (NASA) using a full physics algorithm for retrieving data from GOSAT’s onboard Thermal And Near-infrared Sensor for carbon Observation—the Fourier Transform Spectrometer (TANSO-FTS) calibrated spectra measurements (Level 1B). Only data with land high gain were used after screening and systematic bias correction described in the ACOS Level 2 Standard Product Data User’s Guide, v3.3 [41] to ensure high reliability of the data.

(2) Auxiliary Data

In order to evaluate the performance of the clustering results and assess the impact of different underlying surfaces on the variations of CO2 concentrations, we collected bottom-up anthropogenic CO2 emissions, net primary productivity (NPP) data and land cover data.

The 0.1° × 0.1° gridded annual estimates of CO2 emissions from EDGAR 4.2 FT2010 (http://edgar.jrc.ec.europa.eu/) was collected [42]. The EDGAR 4.2 FT2010 database was jointly developed by the Joint Research Center (JRC) and Netherlands Environmental Assessment Agency (PBL). It was generated by applying the emission factors and the calculation method from the 2006 IPCC Guidelines to international statistics on energy production and consumption, industrial manufacturing, agricultural production, waste treatment and disposal, and burning of biomass [43].

We collected the NPP data derived from MODIS/Terra observation (MOD17A3, Gridded 1 Km Annual Net Primary Productivity (https://lpdaac.usgs.gov)). NPP reflects the carbon sequestration capacity of vegetation in terrestrial ecosystems [44,45].

Land cover types in the study area vary. The spatial distributions of land cover in the study area were obtained from MODIS Land Cover Type data (MCD12C1) by the MODIS-derived LAI/FPAR scheme (https://lpdaac.usgs.gov). For statistical analyses purpose, we derived averaged percentages of each land cover type in 1° × 1° grid cells from original data.

2.2. Clustering Approach Based on Multi-Temporal Xco2 Data

To make use of the temporally changing characteristics of Xco2, we used a robust cluster method to classify Xco2 data based on the similarity of CO2 concentrations to temporally changing patterns. K-means is an iterative algorithm used to partition the given dataset into $k$ clusters, where $k$ is a user-specified number [46,47]. The clustering result should meet the conditions that the intra-cluster similarity is high, while the inter-cluster similarity is low. The clustering process was implemented in the following steps:

1. Combine multi-temporal Xco2 into a characteristic vector $x$ for each spatial grid cell. The count of the grid cells is $n$, and “Z-score” measures are used to remove seasonality from the data. “Z-score” is a distance measure of a data point from the mean in terms of the standard deviation, and can be used for standardizing the original data [48];
(2) Randomly select $k$ objects from $n$ grid cells as the initial cluster centroids. The characteristic vector of cluster $C_i$ can be expressed as $m_i$;

(3) Assign or reassign each grid cell to the cluster to which the grid cell is the most similar based on the Euclidean distance criterion (Equation (1)) between the grid cell and cluster;

$$ dis(x, C_i) = \|x - m_i\|_2 $$(1)

(4) Update the cluster centroids by averaging characteristic vector of the grid cells for each cluster;

(5) Repeat Steps (3) and (4) until the sum of square error criterion (Equation (2)) converges when the cluster centroids do not change.

$$ SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (dis(x, C_i))^2 $$ (2)

Because the results of k-means clustering can be influenced by the initial cluster centroids, and the sum of square error (Equation (2)) may fall into the local optimum, the clustering process is repeated for ten times. In addition, the lowest value of SSE was chosen for acquiring stable clustering results.

To test the effects of different time intervals of the mapping Xco$_2$ dataset, we respectively used both the M-Xco$_2$ and S-Xco$_2$ to the clustering process described above. Grid cells with similar Xco$_2$ temporal patterns could be merged to respectively generate the clustering results for different $k$ (from 3 to 40), in which month-interval clustering results are referred to as M-Clusters and season-interval clustering results are referred to as S-Clusters.

With the increase in cluster number $k$ from 3 to 40, the averages, standard deviations, and non-negative ratios of silhouette values are calculated as measures of how appropriately the data has been clustered. Silhouette values [50,51] are used for validating the clustering results [49,50]. The silhouette is a measure of how similar each grid cell in one cluster is to grid cells in the neighboring clusters. It is given by

$$ s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} $$ (3)

where $a(x)$ is the average distance from $x$ to all other grid cells in the same cluster, and $b(x)$ is the minimum average distance from $x$ to grid cells in any other clusters. $s(x)$ ranges from $-1$ to $+1$. If $s(x)$ is close to $-1$, it indicates that the clustering of $x$ is incorrect and it should be belonged to its neighboring cluster. In contrast, if $s(x)$ is close to 1, the current cluster is suitable for $x$.

2.3. Optimal Number of Clusters

To validate cluster results and choose the best cluster parameters, the numerical evaluation indicators, including average silhouette values, and non-negative ratios of the silhouette values were used. Figure 1 shows the variations of the averages and ratios of $s(x)$, in which the number of clusters $k$ range from 3 to 40. The error bars represent one standard deviation of the silhouette values. The plots show that the incorrect results of the clustering ratios are limited at a low level and that the clustering results are valid for either M-Clusters or S-Clusters. Specifically, the averages remain at approximately 0.4; the standard deviations are approximately the same; and all the ratios are above 95%. Based on Figure 1a,b, we found S-Clusters are slightly better than M-Clusters.
On the other hand, auxiliary data of emissions and absorption were used for proving the effectiveness of the clustering approach. The average and amplitude were considered the two most typical characters of temporal Xco2 variations. Before clustering, we calculated the correlation coefficients: R_{avg,e} between the Xco2 average and emission, R_{avg,n} between the Xco2 average and NPP, and R_{amp,e} between the Xco2 amplitude and emission. In addition, we calculated the correlation coefficient R_{amp,n} between the Xco2 amplitude and NPP based on gridded data, which are introduced in Section 2. R_{avg,e}, R_{avg,n}, R_{amp,e}, and R_{amp,n} results were 0.32, 0.26, 0.13, and 0.52, respectively. These results show that it is difficult to directly obtain and quantify spatial distributions of CO2 emissions using Xco2 data in a grid scale. Moreover, it is obvious that the NPP, which is representative of the absorption capability, showed a relatively strong influence on the Xco2 temporal fluctuation.

After clustering, the new R_{avg,e} and R_{amp,n} in the cluster region scale based on the clustering result were also calculated. They are illustrated in Figure 2. As shown in the figure, these correlation coefficients continuously remain at a relatively high level when k ranges from 3 to 40, although there is a slight decrease associated with the increase of k when a linear or logarithmic model is used. These results respectively prove the strong positive correlations in the cluster scale between Xco2 averages and CO2 emissions, and between Xco2 amplitudes and NPP. Compared to the analysis results before clustering, these results reveal the distribution information on both anthropogenic emissions and NPP.

In general, these results show that the satellite-based observations are correlated with regional CO2 emissions and NPP in a cluster region scale, and that the regional division by temporal clustering plays a role in exploring these relationships. The regional Xco2 averages can be regarded as a potential indicator to quantify the diversity of CO2 emissions intensity in the cluster region scale, while regional Xco2 amplitudes can be regarded as a potential indicator for analyzing NPP.

In addition, p-values were calculated in correlation analyses and applied in significance tests. Here, “p-value” is used for testing the null hypothesis of no correlation against the alternative that there is a nonzero correlation. As shown in Figure 2, where k is greater than 10, the p-values are less than 0.05, and the correlation analysis results can be considered reliable. Based on the above results and analyses, k can be appropriately chosen from the range of 10 to 40, and average silhouette values and non-negative ratios reach their peak when k is equal to 14 in S-Clusters when the minimum standard deviation is
reached. Therefore, $k = 14$ in S-Clusters (average = 0.417, SD = 0.195, and ratio = 98.014%) is regarded as the optimal clustering scheme, and will be described and analyzed in more detail further.

Figure 2. Correlation analysis: (a) $R_{\text{avg,}_c}$ between the Xco$_2$ average and CO$_2$ emissions based on the clustering results; (b) $R_{\text{amp,}_n}$ between the Xco$_2$ amplitude and NPP based on the clustering results.

3. Results and Discussion

Figure 3 presents the results of 14 clusters obtained using the clustering method described in Section 2.2, and the mean Xco$_2$ in each clusters. As shown in the figure, Xco$_2$ data of Clusters 1 and 2 in Central and Eastern China are the maximum, followed by those of Cluster 3 in Xinjiang province in Western China. Xco$_2$ data of Clusters 12, 13, and 14, located around Inner Mongolia, Mongolia, Northeastern Kazakhstan, are the minimum.

Figure 3. Clustering result based on S-Xco$_2$ dataset with 14 clusters and the corresponding cluster-averaged Xco$_2$ data from 2010 to 2012.
3.1. Xco₂ and Anthropogenic Emissions

Based on the optimal clustering result shown in Figure 3, an overlay analysis with CO₂ concentrations and emission data, respectively, is implemented. Figure 4a,b show annual Xco₂ and CO₂ emissions in 2010 overlaid with polylines of the 14 clusters. Figure 4b is the average of CO₂ emissions in 1° × 1° grid cells derived from original emission data; the emission units are converted to kg CO₂/m².

Figure 4. (a) Annual Xco₂; (b) Annual CO₂ emissions in 2010 overlapped with the optimal cluster result (black lines).

From a comparison of the spatially distributed borders of clusters shown in Figure 4a with emissions in Figure 4b, it is obvious that the areas of Clusters 1, 2, and 4 with large Xco₂ values coincide with high emissions, whereas Clusters 11, 12, and 13 with small Xco₂ values correspond to the low emission areas in Mongolia, part of Inner Mongolia and Xizang in China. The exception is Cluster 3, located in Xinjiang province in western China, where Xco₂ is very high but emissions are not.

Figure 5. Regression analysis between cluster-averaged CO₂ emissions and Xco₂ in 2010.
In addition, Figure 5 depicts a correlation between Xco2 and CO2 emissions. The results indicate that a positive correlation is shown between emissions and Xco2 at the clustering region scale. Xco2 of Cluster 3, an outlier in Figure 5, is abnormally high but low in emissions.

3.2. Xco2 and NPP

The seasonal changes of CO2 concentrations are mainly caused by vegetation absorptions. The amplitude of seasonal variations in a year depends on vegetation coverage and growth activities in the northern parts of the northern hemisphere [51–53]. Figure 6a presents the seasonal amplitude of Xco2 in 2010, which is calculated by the difference between the maximum and minimum of the monthly averaged Xco2 of each grid cell. Figure 6b presents the annual mean NPP in 2010 of each grid cell, which can indicate the ability of vegetation to absorb CO2. Figure 6b is the average of CO2 fluxes in 1° × 1° grid cells derived from original NPP data.

![Figure 6](image)

**Figure 6.** (a) Seasonal amplitude of Xco2; (b) grid map of annual NPP in 2010 overlapped with the optimal cluster result (black lines).

It is evident that the spatial distribution of Xco2 amplitudes are generally in agreement with the spatial trend of NPP. Clusters 1, 4, 7, and 14 with large amplitudes of Xco2 show significant corresponds to high NPP, while Clusters 3, 5, and 11 with small amplitudes of Xco2 demonstrate low NPP. This indicates that the impact of NPP, a measurement of vegetation activities, on the seasonal changing amplitudes of Xco2.

The correlation coefficients between seasonal amplitudes of Xco2 and NPP for each cluster are shown in Figure 7. It is shown that correlations between the Xco2 amplitude and NPP are mostly positive;
moreover, the larger NPP is, the higher the correlation coefficient is, except for Clusters 1 and 2, which are the maximum emission areas (shown in Figure 4b), and Cluster 9.

Figure 7. Correlation coefficients (R) of all clustering classes (p-value < 0.05) between the Xco2 amplitude and NPP.

Regression analysis between the cluster-averaged NPP and Xco2 amplitude is shown in Figure 8. The coefficient of determination between them is 0.53 for all clusters, which is similar to the coefficient of determination (r^2 = 0.52) between Xco2 and the emissions (shown in Figure 5). The results imply that the variations of Xco2 can be equally explained by the underlying anthropogenic emissions and vegetation absorptions in the whole study area.

Figure 8. Regression analysis between cluster-averaged NPP and Xco2 amplitude in 2010.

3.3. Attribution of Xco2 Clusters

Figure 9 shows the variation of the monthly mean Xco2 in each cluster. As is shown, each cluster demonstrates a different seasonal changing pattern of Xco2. Clusters 7, 8, 12, and 14 located at the northern part present two high Xco2 peaks, which may be caused by vegetation growth activity and temperature variations. On the other hand, Clusters 3, 5, and 11 show smaller seasonal amplitudes, which can be attributed to sparse vegetation areas in these study regions.
Figure 9. Variation of monthly averaged Xco₂ in each Cluster (number of clusters $k = 14$). Blue, green, and red lines correspond to average, minimum, and maximum Xco₂ values in each cluster, respectively, derived from the gap-filled Xco₂ dataset. Error bars represent one standard deviation of Xco₂, and the grey scatter points correspond to original observations from the ACOS-Xco₂ dataset.
Table 1 demonstrates typical statistics of Xco₂ and auxiliary data in each cluster, including Xco₂ variations, anthropogenic emissions, and NPP. The contrast of Xco₂ in Table 1 is calculated by subtracting the average Xco₂ of the study area (389.10 ppm) from cluster-averaged Xco₂ to indicate the difference from the regional background Xco₂ value in each cluster. Xco₂ amplitude is calculated by the maximum monthly cluster-averaged Xco₂ minus the corresponding minimum in each cluster. The strength of emissions is derived from the emissions in the target cluster divided by the maximum emission value of all clusters (Cluster 1). The strength of NPP is calculated in the same way.

The fraction of land cover is the fraction of low vegetation (L), forest (F), and non-vegetated areas (N), in each cluster derived from land cover data. The correlation coefficients (p-value < 0.05) of the Xco₂ amplitude versus the NPP of all grid cells within a cluster are also listed in this table. The correlation between Xco₂ and emissions of all grid cells within a cluster is very small and not significant, and is therefore not listed in Table 1.

### Table 1. Attribute characteristics of each cluster from the 2010 data.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Contrast of Xco₂ ppm</th>
<th>Xco₂ Amplitude ppm</th>
<th>Xco₂ Amplitude vs. NPP</th>
<th>Strength of Emission</th>
<th>Strength of NPP</th>
<th>Fraction of Land Cover* L:F:N</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>9.3</td>
<td>-</td>
<td>1.00</td>
<td>0.79</td>
<td>70:22:4</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>1.2</td>
<td>7.8</td>
<td>-0.27</td>
<td>0.76</td>
<td>1.00</td>
<td>68:27:2</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>5.4</td>
<td>0.19</td>
<td>0.03</td>
<td>0.15</td>
<td>38:1:60</td>
<td>Δ</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>10.7</td>
<td>0.71</td>
<td>0.45</td>
<td>0.81</td>
<td>58:38:1</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>6.1</td>
<td>0.53</td>
<td>0.05</td>
<td>0.22</td>
<td>38:2:59</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>7.9</td>
<td>-</td>
<td>0.13</td>
<td>0.61</td>
<td>69:10:20</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-0.1</td>
<td>11.3</td>
<td>0.74</td>
<td>0.05</td>
<td>0.78</td>
<td>51:48:0.5</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>-0.2</td>
<td>8.8</td>
<td>0.27</td>
<td>0.03</td>
<td>0.60</td>
<td>44:49:0.5</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-0.4</td>
<td>6.4</td>
<td>-</td>
<td>0.06</td>
<td>0.29</td>
<td>79:0:5:19</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>-0.5</td>
<td>8.7</td>
<td>-</td>
<td>0.09</td>
<td>0.34</td>
<td>89:5:5</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>-0.5</td>
<td>4.6</td>
<td>0.26</td>
<td>0.01</td>
<td>0.08</td>
<td>56:41:2</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>-0.6</td>
<td>7.6</td>
<td>0.61</td>
<td>0.02</td>
<td>0.65</td>
<td>72:26:0.5</td>
<td>A</td>
</tr>
<tr>
<td>13</td>
<td>-0.8</td>
<td>7.6</td>
<td>0.70</td>
<td>0.03</td>
<td>0.33</td>
<td>85:2:12</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>-1.00</td>
<td>8.7</td>
<td>0.56</td>
<td>0.13</td>
<td>0.92</td>
<td>53:45:0.5</td>
<td>A</td>
</tr>
</tbody>
</table>

* The land cover types “grasses/cereal”, “shrubs”, “broadleaf crops” and “savanna” were grouped into type “L”; “evergreen broad-leaf forest”, “deciduous broadleaf forest”, “evergreen needleleaf forest” and “deciduous needle-leaf forest” were grouped into type “F”; “unvegetated” and “urban” areas were grouped into type “N”.

In this section, the anthropogenic emissions and vegetation absorptions described in Sections 3.1 and 3.2, and the seasonal changing patterns of Xco₂ of each cluster shown in Figure 9, will be combined to analyze the attribute characteristics of the 14 clusters listed in Table 1. As a result, the 14 clusters can be divided into three flux types, as outlined below.

(1) Strong Emission Type: Clusters 1 and 2

Clusters 1 and 2 as a group, strongly tends to present sources regions as they show the highest Xco₂ (390.6 ppm and 390.3 ppm in average, as shown in Table 1), and the largest positive contrast (1.5 ppm and 1.2 ppm). It can be known from bottom-up investigation of carbon emissions (EDGAR) that in these
two cluster regions there are very intensive big power fuel plants and dense populations which corresponds to large emissions (Table 1). The strong anthropogenic emissions can disturb the temporal variations of $X_{CO_2}$. Consequently, the correlations between the $X_{CO_2}$ amplitude and NPP do not show a significant correlation although they show a large $X_{CO_2}$ amplitude (9.3 ppm and 7.8 ppm), which is likewise due to vegetation coverage of more than 90% in these regions.

These results indicate that both Clusters 1 and 2 can be attributed to a source with strong emissions. These clusters show different temporal variations of $X_{CO_2}$ (Figure 9), even if both of them belong to the strong emission type. Cluster 1 is located in the major grain producing areas mixed with many residential areas in China. Intensive farming and crop vegetation growth have resulted in a significant seasonal amplitude of $X_{CO_2}$ (9.3 ppm in Table 1) due to crop CO$_2$ absorption. Temporal variations of $X_{CO_2}$ show that $X_{CO_2}$ sharply increased from August to October over this cluster compared with Cluster 2, which is likely due to the crops being harvested and large amounts of straw being burned during this period. Cluster 2, located at the paddy fields mixed with the fragmented forests, residential areas, shows that a 7.8 ppm of $X_{CO_2}$ amplitude, which is slightly less than that in Cluster 1. Their different temporally changing patterns of $X_{CO_2}$ indicate the effects of different anthropogenic emitting actions on variations of $X_{CO_2}$.

(2) Strong Absorption Type: Clusters 7, 12, 13, and 14

Clusters 7, 12, 13, and 14, as a group, tend to be strong sinks, according to the negative contrasts and strong correlations between $X_{CO_2}$ and NPP, high fractions of vegetation, and less anthropogenic emissions shown in Table 1. $X_{CO_2}$ values over these cluster regions are lower than the overall average; moreover, the contrasts are $-0.1$ ppm, $-0.6$ ppm, $-0.8$ ppm, and $-1.0$ ppm, respectively (Table 1). These clusters are located at north of latitude 45°N, and are covered with dense vegetation and forest. Accordingly, the correlations between $X_{CO_2}$ amplitudes and NPP are significantly larger than 0.55 over these clusters. This demonstrates the effects of vegetation absorption on $X_{CO_2}$ variations. With strong vegetation absorptions in addition to low anthropogenic emissions over these clusters (Figure 4), they tend to be sinks of atmospheric CO$_2$. These clusters show different temporally changing patterns of $X_{CO_2}$ (Figure 9), which demonstrate the effects of vegetation absorption. Among the clusters, Clusters 7 and 14 show the highest amplitude of $X_{CO_2}$ seasonal variation (11.3 ppm and 8.7 ppm, respectively, shown in Table 1), which is likely induced by the highest coverage of mixed broad-leaf with needle-leaf forest. Clusters 12 and 13 show a strong correlation between $X_{CO_2}$ amplitude and NPP, which is 0.61 and 0.70, respectively, although they have less forest coverage and a high fraction of grasslands. The values of $X_{CO_2}$ over these clusters as shown in Figure 9, moreover, are almost unchanged from October to January, or they present a low peak around December, which is likely due to the low temperature in the high latitude region.

(3) Tending to Balance Type: Clusters 4, 5, 6, 8, 9, 10 and 11

Clusters 4, 5, 6, 8, 9, 10 and 11 are defined as being a “tending to balance” type. All of their mean $X_{CO_2}$ values are close to the average level of the whole study area. In addition, Clusters 5, 6, 8, 9, 10 and 11 do not show significant correlations between $X_{CO_2}$ amplitude and NPP, and there are the small anthropogenic emissions over these regions. Cluster 4 shows slightly higher $X_{CO_2}$ than the average level,
with a 0.4 ppm contrast, mainly due to significant large anthropogenic emission. However, the correlation between the \( \text{Xco}_2 \) amplitude and NPP is 0.71, which implies that the CO\(_2\) enhancement from the anthropogenic emission may be in equilibrium with the vegetation absorption. Cluster 11, located on the Qinghai-Tibet Plateau, shows lower \( \text{Xco}_2 \) amplitude than the other regions, which are probably owing to its low NPP and high altitude.

Cluster 3 is indistinctly grouped as its abnormally high \( \text{Xco}_2 \) while there is less anthropogenic emissions (Figure 4). Cluster 3, located at a desert area in China, shows high \( \text{Xco}_2 \) values, which are likely owing to uncertainties of \( \text{Xco}_2 \) retrievals from satellite over these high lightness desert [9], and further verification is needed in this region.

4. Conclusions

In this paper, a k-means cluster analysis method based on the temporally changing features of \( \text{Xco}_2 \) was proposed for application to the gap-filled ACOS \( \text{Xco}_2 \) dataset to view spatial pattern of CO\(_2\) emissions and absorption in Central-Eastern Asia. 14 clusters were obtained by optimizing the clustering results and evaluated using the characteristics of \( \text{Xco}_2 \) variations combined with emissions data, NPP data, and land cover data. The clustering result demonstrates that each cluster can be related to the typical features of anthropogenic emissions and vegetation absorptions.

The relationships between seasonal variations of \( \text{Xco}_2 \), and underlying anthropogenic emissions and vegetation absorption were analyzed respectively. Cluster-averaged \( \text{Xco}_2 \) tend to correlate with regional emissions, while seasonal amplitude of \( \text{Xco}_2 \) is highly related to vegetation NPP. Strong anthropogenic emissions may disturb the relationship between the seasonal amplitude of \( \text{Xco}_2 \) and NPP data. Consequently, the 14 clusters can be divided into three types: strong emission, strong absorption, and a “tending to balance” type. Different clusters, corresponding to different temporally changing patterns of \( \text{Xco}_2 \), indicate that the \( \text{Xco}_2 \) values increase with anthropogenic activities and that \( \text{Xco}_2 \) reduction is caused by vegetation absorption on a regional scale.

This study shows that the developed cluster-analysis approach based on temporal variation of \( \text{Xco}_2 \) can effectively provide a way to reveal the spatial patterns of underlying anthropogenic emissions and vegetation absorptions, and therefore enable to us to better understand how the seasonally changing pattern of CO\(_2\) concentrations is affected by anthropogenic emissions and vegetation absorptions. The result of clustering can provide the significant monitoring targets of anthropogenic emissions and vegetation absorption to support the implement of regional emissions reduction of carbon.

Acknowledgments

This research was supported by the “Strategic Priority Research Program—Climate Change: Carbon Budget and Relevant Issues” of the Chinese Academy of Sciences, Grant No. XDA05040401, and the National High Technology Research and Development Program of China (2012AA12A301). The authors are immensely grateful to the anonymous reviewers and editor for their excellent, valuable comments. The ACOS-GOSAT v3.3 data were produced by the ACOS/OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and were obtained from the ACOS/OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. We would also like to acknowledge the GOSAT Project for acquiring the spectra, and the EDGAR data
(http://edgar.jrc.ec.europa.eu/) provided by the joint project of the European Commission Joint Research Center (JRC) and the Netherlands Environmental Assessment Agency (PBL). We also acknowledge that the Terra MODIS GPP/NPP Product (MOD17A3) and Land Cover Type Climate Modeling Grid (CMG) product (MCD12C1) were retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool.

**Author Contributions**

Da Liu and Liping Lei conceived and designed the experiments; Da Liu performed the experiments; Da Liu and Liping Lei analyzed the data; Lijie Guo contributed materials; and Da Liu, Liping Lei, and Zhaocheng Zeng wrote the paper.

**Conflicts of Interest**

The authors declare no conflicts of interest.

**References**


© 2015 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 license (http://creativecommons.org/licenses/by/4.0/).