Keywords
Carbon Dioxide Sink, Spatial Autocorrelation, Interpolation, Jackknife, Negev Desert

1. Introduction
Quantitative understanding of the atmospheric carbon budget including biospheric carbon exchanges is crucial for climate mitigation policy in 21st century [1]. Water management planners are facing considerable uncertainties on future demand and availability of water. Climate change and its potential hydrological effects are increasingly contributing to this uncertainty. The Second Assessment of the Intergovernmental Panel on Climate Change [2] states that “increasing concentration of greenhouse gases in the atmosphere is likely to cause an increase in global average temperature of between 1 and 3.5 degrees Celsius over the forthcoming century”. These changes will in turn affect water availability and runoff and thus may affect the discharge regime of rivers [3].

There are certain gases that are very effective at trapping heat in the atmosphere, and warming the Earth’s surface. Carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O) are greenhouse gases that both occur naturally and also are released by human activities [4]. Mitigation measures aiming at short-lived effects can lower the rate of climate change, whereas in order to avoid irreversible climate change, mitigation measures should focus particularly on controlling CO2 level in the atmosphere [5].

Studies on the uptake and release of carbon dioxide gas started around 1990. In these studies, subjects were followed up by two different views. A group studied the trend of gas changes and its related factors. Another group conducted statistical analysis of data on climate maps and estimates the locations where the gas might release or absorb.

Reliable predictions of future levels of atmospheric CO2 require a quantitative understanding of both CO2 emissions and the specific processes and reservoirs responsible for sequestering CO2 [6]. Highly accurate measurements from a surface network are currently monitoring atmospheric CO2 [7].

One of the most important problems in the science of global change is the balancing of the global budget for atmospheric CO2. Roughly half of the CO2 emitted into the atmosphere as a result of burning fossil fuels, remains in the atmosphere and the other half is absorbed into the oceans and the terrestrial biosphere. The partitioning between these two sinks is the subject of considerable debate. Whereas most chemical oceanographers are confident that the oceanic sink is not large enough to account for the entire absorption, many terrestrial ecologists doubt that the land biosphere can be a large carbon sink, particularly given the source to the atmosphere through deforestation, hence, the issue of the “missing” carbon sink [8]. After 30 years of measurements in the atmosphere and the oceans, the global atmospheric CO2 budget is still surprisingly uncertain [9].

[10] says that: “On average, 43% of the total CO2 emissions each year between 1959 and 2008 remained in the atmosphere, but this fraction is subject to very large year-to-year variability (Figure 1(a)). Our estimates of sources and sinks of CO2 were based largely on independent data and methods. Thus, when all the sources and sinks were summed every year they did not necessarily add to zero, because of the errors in the various methods. The sum of all CO2 sources and sinks, which we call the ‘residual’, spanned a range of ±2.1 Pg C yr−1 (Figure 1(b)). This residual was not explained by the atmospheric CO2 growth rate, the CO2 emissions from fossil fuel combustion or the ocean uptake, because the uncertainties in these components were much smaller than the variability of the residual”.

It seems that there are much more source and sink factors which are not well known yet. In this study we focused on spatial characteristics of CO2 measurements. After validating the existence of spatial relationship among the amounts of CO2 gas which have been measured by more than 60 different stations in different parts of the globe, and by implementing spatial analysis based on “Leave One Out” method, we identified spots which had most distinguished absorption effects in their regions, which indicate existence of absorption factor in that area.

2. Material and Methods
The Mauna Loa Observatory (MLO) is an atmospheric baseline station on Mauna Loa volcano, on the big island
of Hawaii. Since 1956 scientists at MLO have been continuously monitoring atmospheric changes and in particular collecting data relating to variation of atmospheric carbon dioxide (CO2). (http://en.wikipedia.org/wiki/Mauna_Loa_Observatory). Since 1956 so many different kinds of atmospheric CO2 measurement stations have been constructed in various locations on the Globe. In March and April of 1983 during an Arctic Gas and Aerosol Sampling Program a set of flask air samples were taken by a number of aircrafts and analyzed for CO2 concentration. Some 18 aircraft samples were collected in electro polished 2.5 stainless steel canisters equipped with Metal Bellows valves [13]. Hall et al. (2007) used a set of CO2 isotope data for CARIBIC aircraft samples and NOAA-Carbon Cycle flask samples. The NOAA-ESRL data provided a representative geographical and seasonal coverage [14]. During summer 2007, two survey cruises covered large sections of both western and eastern Mediterranean basins. Atmospheric CO2 concentrations were measured and 48 discrete air samples were collected in vacuum Pyrex flasks [15]. Nassar et al. (2011) showed the locations of the 59 National Oceanic and Atmospheric Administration-Earth System Research Laboratory Global Monitoring Division (NOAA-ESRLGMD) and Environment Canada (EC) stationary sampling sites as well as NOAA ship-based sample collection locations in the central and western Pacific Ocean, and the Drake Passage. It also showed the sampling locations for aircraft flask CO2 from CONTRAIL and CARIBIC (Civil Aircraft for the Regular Investigation of the atmosphere based on an Instrument Container) [16].

We found out the coordinates of CO2 station positions from NOAA web site [17], and the required data for CO2 measurements were extracted from ftp web site of NOAA [18]. These measurements were based on monthly intervals. By simple calculations the average annual data were derived from them. The measurements are for years 1980 to 2011. Each site might have values for all or some of the years in that period. A separate study indicated that even with optimal placement, which the present data set does not have, a minimum of about 10 stations per region is needed to obtain estimates with useful accuracy [19]. There were about 100 different sites. By using a world map and indicating the sites on it, the base layers of our working space was prepared as is illustrated in (Figure 2).

Autocorrelation is a very general statistical property of ecological variables observed across geographic space [20]. The first step in understanding ecological processes is to identify patterns. Ecological data are usually characterized by spatial structures due to spatial autocorrelation. Spatial autocorrelation show a pattern in which observations from nearby locations are more likely to have similar magnitudes than by chance alone. Ecologists have shown an increasing interest in geostatistical methods to identify and model spatial patterns. One begins by
estimating parameters that characterize the spatial structure of the data in terms of spatial variance using an experimental variogram, and then use these parameters to interpolate values at ensample locations via kriging method [21]. Our variogram test applied on data from the year 2009 as a sample which shows the existence of exponential pattern type of spatial structure among them (see theory/calculation).

Topographic surfaces are non-stationary, i.e., the roughness of the terrain is not periodic but changes from one land type to another. A regular grid therefore has to be adjusted to the roughest terrain in the model and be highly redundant in smooth terrain (Peucker et al., 1975). “Like several other researchers we realized that a simple means of meeting these specifications is to model the surface as a sheet of triangular facets. We called our implementation of this approach a Triangulated Irregular Network or TIN [22]. In this study, we needed to generate about 700 different TIN models, so it was mandatory to check the validation of applying these models on CO₂ values of measurement stations. This validation has been done for the year 2009 as a sample. For this purpose, each time 15% of data was eliminated randomly and with remaining data a TIN model was generated. Then we calculated the validation via comparing the estimated values of TIN (at those eliminated locations) with their observed values. This operation has been repeated 20 times with different randomly eliminated data. The average error for these tests was about 0.7% which shows a good fitness (see theory/calculation).

The jackknife or “Leave One Out” procedure is a cross-validation technique first developed by Quenouille to estimate the bias of an estimator. John Tukey then expanded the use of the jackknife to include variance estimation and tailored the name of jackknife, like a jackknife this technique can be used as a quick replacement tool for a lot of more sophisticated and specific tools [23]. The jackknife estimation of a parameter is an iterative process. First the parameter is estimated from the whole sample. Then each element is, in turn, dropped from the sample and the parameter of interest is estimated from this smaller sample. This estimation is called a partial estimate (or also a jackknife replication). A pseudo-value is then computed as the difference between the whole sample estimate and the partial estimate [23]. This value may be considered as the real effect of that element in the set, without being hidden below strong neighboring elements (see theory/calculation). In this study we made a main TIN model by entire data of each year. Then for each station in each year, a sub-model TIN was developed. In each sub-model,
one station was eliminated. The difference between sub-models and related main TIN model considered as the “Site activity degree” for those sites.

3. Theory/Calculations

3.1. Spatial Relationship Validation

Since we were supposed to implement spatial analysis to the CO₂ values of measurement sites, it was needed to check whether there exist any spatial relationships among them or not. In order to study the spatial relationships between values of measurement sites we used spatial autocorrelation semivariogram charts. The resulting spatial structure can be studied by examining the patterns of autocorrelation and cross correlations at different spatial distances, which can be described by spatial statistics such as Correlograms or Variogram [24]. Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close coordinates on a two-dimensional surface, introducing a deviation from the independent observational assumptions of classical statistics [25]. To our knowledge, there is no comprehensive overview of the many available spatial statistical methods to take spatial autocorrelation into account in tests of statistical significance [26]. Osborne & Leitaõ (2009) raised an interesting issue, namely that the impact of positional errors on Species distribution models (SDMs) may be understood by examining spatial autocorrelation in predictor variables [27]. One of the first to investigate the concept of spatial autocorrelation was Moran (1947, 1948), and the second one was Geary (1954) who formulated another measure [28]. To formalize the model, let there be \( N \) points for which we have observations \( x_i \) \((i = 1, \ldots, N)\). These points and their relations form a network described by a weight matrix \( W \). The weights \( W_{ij} \) \((i, j = 1, \ldots, N)\) indicate the existence of a relation of point \(i\) with point \(j\), and (optionally) the strength of that relation. The diagonal elements \( W_{ii} \) are zero by definition. The network and the weight matrix \( W \) will be considered as given. Spatial autocorrelation coefficients indicate whether and to what extent the observations \( X_i \) influence each other via the structure of the network. The coefficients Moran’s \( I \) and Geary’s \( C \) are defined as [28]:

\[
I = \frac{N}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \quad (1)
\]

\[
C = \frac{(N-1)}{2 \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}} \times \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (X_i - X_j)^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \quad (2)
\]

In this study using the Geary coefficient and Equation (3) below a semivariogram chart was produced for CO₂ data of the year 2009 which is shown in (Figure 3).

\[
2\gamma(d) = \frac{1}{n_d} \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} w_{ij} (X_i - X_j)^2 \quad (3)
\]

\( n_d \) is number of samples in predefined distance (lag).

The semivariogram for CO₂ values of the year 2009 (as a sample) shows existence of an exponential spatial structure model among them, which allows implementing spatial analysis over them.

3.2. Kriging Interpolation Validity

Kriging is widely used for obtaining the metamodels. The popularity of kriging is due to the fact that computer models are often deterministic \(i.e.,\) no random error in the output) and thus interpolating metamodels are desirable. Kriging gives an interpolating metamodel and is therefore more suitable than the other common alternatives such as quadratic response surface model [29]. One of the distinctive advantages of Kriging is that it provides not only the prediction of the response at any site, but also the mean square error (or the uncertainty) associated with the prediction [30]. This prediction is defined by:

\[
\hat{Z}(X_p) = \sum_{i=1}^{n} \lambda_i \times Z(X_i) \quad (4)
\]

In which \( \hat{Z}(X_p) \) is the predicted element and \( Z(X_i) \) is the \( i^{th} \) observed element and \( \lambda_i \) is its weight.
The spatial analyses in this research are basically upon Kriging method interpolations. Around 700 different TIN models were generated for consideration. As previously described in “Material and methods” section, we validated the using of this kind of interpolation for CO₂ values by randomly preserving 15% of data and generating a TIN model with remained data, and considering the predicted amounts of those preserved data with their observed values. This operation has been done 20 times with data of the year 2009. The result is shown in Figure 4. As it can be seen, errors are mostly accumulated around zero percent. The average of proportional errors was 0.7%, which shows a good validation. As will be described in “Leave One Out Method” section, from the perspective of this study, those few cases which are quantum leaped from all other ones are most probably the places where the factors of emission or absorption exist.

3.3. Leave One Out Method

The Leave One Out jackknife resampling procedure determines confidence intervals by calculating the variance of the particular statistic over N “Leave One Out” passes through a data set of size N [31]. Let \( y_{all} \) be the statistic (e.g. the variance, the correlation, etc.) calculated over all \( N \) data. Define pseudo values by
\[
Y_j = N Y_{all} - (N - 1)Y_j = 1, 2, \ldots, N
\]
where \( Y_j \) is the statistic calculated using the portion of the data that omits the \( j \)th entry. The estimated mean, \( y^* \), and variance of the mean, \( S^2 \), are given by:
\[
y^* = \frac{\sum_{j=1}^{N} Y_j}{N}
\]
\[
S^2 = \frac{\left( \sum_{j=1}^{N} Y_j^2 - NY_{all}^2 \right)}{(N - 1)}
\]
\[
S^2 = S^2 / N
\]
Given the mean, \( y^* \) and the standard deviation \( s^* \), 95% confidence intervals are calculated from the Student’s parameter, \( t_{95} \): 95% confidence interval on \( y^* \):
\[
[y^* - t_{95} s^*, y^* + t_{95} s^*]
\]
This jackknife method should be used only for statistics, such as the mean or variance that are not narrowly
estimated. It should not be used for order statistics, such as the median, since the \(\gamma\) would then always take one of two or three values. (The jackknife estimates are degraded if the data set has excessively straggling tails, i.e. a few wild outliers) [31].

Figure 5 shows a sample of using this method in our processes. Figure 5(a) shows a TIN model generated upon values of all sites in year 2010. Figure 5(b) is the same model but by eliminating the Negev Desert site. The differences between these two models at the Negev Desert site location may be considered as the amount of gas absorption or emission in that region in the year 2010.

In this sample the (Figure 5(b)) TIN model illustrates that if no gas emission or absorption had been occurred in the Negev Desert region, the existence of CO\(_2\) gas in that place might be 401.97 Pg C yr\(^{-1}\). This is while the observed amount for this site in 2010 indicates existence of 391.83 Pg C yr\(^{-1}\) of gas, which is 10.14 units less than expected amount. So it may be true to suppose that Negev Desert region has contained a gas sink factor by the amount of 10.14 units in that period of time. By continuing this process for each of sites in each year and considering the results and trends of them, it will be possible to extract out the most absorption or emission places among them all.

4. Results

4.1. Main TIN models

As described previously, for each of the years between 2000 and 2011, a TIN model has been generated based on the data of all related sites. These models were cited in this study as a basis for observation of CO\(_2\) in each point of the Globe for those years. As it can be seen in Figure 6, the value of the highest amount of gas has been grown from 377.85 Pg C yr\(^{-1}\) in 2000 to 407.44 in 2011, and likewise, the value of the lowest amount of the gas grew from 366.24 Pg C yr\(^{-1}\) in 2000 to 386.73 in 2011, which indicate the growth of Global gas balance by passing time. In a glance it indicates that in all the years, the Black Sea region represents the largest inventory of gas, and consequently there should most probably be a strong and permanent factor of gas emission over there. Another apparent conclusion that can be derived from Figure 6 is that the accumulation of gas in the Southern Hemisphere is less than its Northern counterparts.

To find out more analytical deductions from these elegant data, we applied “Leave One Out” processes over them and generated one special pseudo-TIN model for each one of sites in each year, based on eliminating them, and comparing their subsequent outgrowth with their actual observations. This will be more elaborated in the next section.

4.2. Leave One Out Pseudo-Values

Applying “Leave One Out” method to the explained data, the pseudo-values for each site extracted from them, will represent their actual regional emission or absorption effects. To determine the regional extent of each site, we used “Thiessen polygons” method. In 1911 the climatologist A. H. Thiessen suggested a new method of representing precipitation data from unevenly distributed weather stations. He defined regions based on a set of
Figure 5. Sample of spatial implementation of “Leave One Out” method. (a) Excerpt of the TIN model generated upon the values of all sites in the year 2010; (b) Same as a but eliminating the Negev Desert site.

Figure 6. The main TIN models for the years between 2000 and 2011. In this study each of these was used for estimating the gas amount at any point of the planet within those years.

data points in the plane (weather stations) such that “regions be enclosed by a line midway between the station under consideration and surrounding stations” [32]. Based on this proposal the term Thiessen polygon has since been commonly used in geography to denote polygons defined by proximity criteria with respect to a set of points in the plane [33]. Thiessen diagram is a way of dividing space into a number of regions. A set of points (called seeds, sites, or generators) is specified beforehand and for each seed there will be a corresponding region consisting of all points closer to that seed than to any other. Figure 7 shows our results for most absorption and emission activities during the years 2000 to 2011. As it is observable, in 8 out of 12 years, there exist one most absorption activity for every most emission activity. This confirms that the absorption rate also rises with increased emission rate, but with lower proportion.

By applying the proposed method to data of CO₂ measuring stations during the years 2000 to 2011, the degree of each station activity in each year were calculated and stored. Table 1 shows 10 stations that have the highest degree of positive and negative CO₂ activities. Each year of 2009 and 2011 contain two of most high absorption
Figure 7. The spots of most emission and absorption for the years between 2000 and 2011, based on “Leave One Out” processes. Interesting subjects may be those which two extremes are adjacent to each other.

Table 1. Ten high impacts among all the amount of carbon dioxide between 2000 and 2011.

<table>
<thead>
<tr>
<th>Site</th>
<th>Description</th>
<th>Year</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSC</td>
<td>Black Sea</td>
<td>2010</td>
<td>16.55</td>
</tr>
<tr>
<td>BSC</td>
<td>Black Sea</td>
<td>2011</td>
<td>12.61</td>
</tr>
<tr>
<td>BSC</td>
<td>Black Sea</td>
<td>2009</td>
<td>11.76</td>
</tr>
<tr>
<td>BSC</td>
<td>Black Sea</td>
<td>2005</td>
<td>9.11</td>
</tr>
<tr>
<td>BSC</td>
<td>Black Sea</td>
<td>2006</td>
<td>8.96</td>
</tr>
<tr>
<td>WIS</td>
<td>Negev Desert</td>
<td>2009</td>
<td>-6.59</td>
</tr>
<tr>
<td>LLN</td>
<td>Taiwan</td>
<td>2011</td>
<td>-6.94</td>
</tr>
<tr>
<td>UUM</td>
<td>Mongolia</td>
<td>2009</td>
<td>-7.00</td>
</tr>
<tr>
<td>WIS</td>
<td>Negev Desert</td>
<td>2011</td>
<td>-7.28</td>
</tr>
<tr>
<td>WIS</td>
<td>Negev Desert</td>
<td>2010</td>
<td>-10.14</td>
</tr>
</tbody>
</table>

It is obvious that the station “Black Sea” is a more prominent and dominant emission place than all other ones. This station was in all first places of five categories accounted for the highest activities and can be used as an option for the location of the CO₂ emission source. Former researches have confirmed huge amount of gas emission in BSC site location [34]. In the Black Sea, measurements were performed on R/V Professor Vodyanitsky during the two CRIMEA cruises, May-June 2003 and 2004. In the north western Black Sea, hundreds of active gas seeps were detected along the shelf and slope of the Crimea Peninsula at water depths between 35 and 800 m. Active gas seeps down to 2100 m water column were also detected. These transport mechanisms allow us to estimate the conditions and means by which gas released from seeps reaches the surface. In dispersed seeps, which typically occur at the shallow sites, gas is released intermittently as bubbles. Vertical transport is therefore by bubble dissolution in the water column, or gaseous release to the atmosphere. The model, which includes argon, nitrogen, oxygen, methane and CO₂, was verified using air bubbles in shallow conditions.

On the opposite side according to table1 amounts, the station (WIS) at Negev Desert seems to be as a place with high potential of absorbing gas. By observing the physical locations of both the BSC and WIS stations, it can be seen that the vicinity of poor gas adsorption (WIS) with a strong regional emissions (BSC), caused it to be misdiagnosis. In all the years that have been mentioned in table1, the WIS station values were among the 15
most high amounts, which shows large inventory of gas. However, Leave One Out method indicated that there existed about 391.83 Pg C∙yr⁻¹ of gas in the year 2010 in Negev Desert. If no absorption had been taken place over there in that period of time, there might have been 401.97 Pg C∙yr⁻¹ gas, due to existence of 407.55 Pg C∙yr⁻¹ CO₂ over its neighbor at Black Sea site. Researches show that the plants in desert arid lands which have high density of CO₂ gas in atmosphere may act as high sequester of atmosphere CO₂ [35]. Possible explanations for the high sequestration performance involve the ability of the trees to absorb carbon dioxide without opening pores excessively, which would compromise the tree’s water balance, through evaporation. Presumably, the increased CO₂ in the atmosphere makes it easier for trees to efficiently absorb CO₂. In addition, the conifers planted are specially selected for their ability to thrive under drought and relatively saline conditions. Yatir forest planted at arid lands of the Negev Desert 40 years ago is expanding at unexpectedly high rates. From this study’s point of view the high amount of CO₂ gas emitted from Black sea causes the Yatir forest to grow in an arid land while absorbing higher amount of CO₂.

5. Conclusions

The objective of this work has been to show that by using appropriate spatial analysis (in this case “Leave One Out” method), some CO₂ sink spots may be extracted out from the amounts of gas inventory in different parts of the Globe. Identifying the places of gas absorption and their ecological characteristics, may lead us to find new factors of gas adsorption. By implementing these absorption factors besides those emission sources which mostly for economic reasons cannot be stopped, we can reduce their environmental impacts. Result of the study has identified that Negev Desert region consists of a strong factor of carbon dioxide absorption. The main ecological characteristic of that location is the growth of Yatir forest in such arid land. Studies have conceived researchers that the trees which are planted in dry and arid deserts may begin to grow and develop if the concentration of CO₂ in their atmosphere is high enough.

According to this study, the required carbon dioxide for planted Yatir forest in arid Negev Desert has been provided through its neighborhood by the high quantity of CO₂ gas released from the Black Sea region. Therefore it could be possible to plant a band of suitable trees around the center of a CO₂ diffuser with no need of irrigation and at the same time reduce the excessive carbon dioxide in the atmosphere.

Acknowledgements

We wish to thank Dr. Babak Naimi for his informative classes which provided us the validation techniques which were used in this study.

References


[17] NOAA (2011) CCGG Measurement Sites. http://www.esrl.noaa.gov/gmd/ccgg/tables/?columns=code%7Cname%7Ccountry%7Clat%7Clon%7Calt%7Ccoop_na me&table_type=active&project=all&param=CO&pos_label=decimal


Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.